FORECASTING SEASONAL RAINFALL WITH COPULA MODELLING APPROACH FOR AGRICULTURAL STATIONS IN PAPUA NEW GUINEA

A Thesis Submitted by

Kingsten Okka

BSc

For the award of

Master of Science Research **(**MSCR**)**

2019

ABSTRACT

Developing innovative forecasting tools is important to address issues related to climate change, agriculture, and economy of small Pacific Island nations. Papua New Guinea, PNG is a developing nation that is vulnerable to the imminent threats of climate change and influences agricultural sector that supports a majority of its citizens. Accurate modeling and forecasting methods for both monthly and seasonal rainfall (that influences agricultural and other human activities) by employing large-scale climate mode indices (linked to rainfall events) are significant predictive tools for developing climate resilience and productivity in agricultural activities.

Copula statistical models, developed in this Master's study, are considered as viable alternative tools to fulfill this objective. This Masters by Research Thesis utilizes the D-vine copula-based quantile regression methods that are developed to create a model between statistically significant lagged relationships and joint influences of large-scale climate mode indices such as the El-Niño Southern Oscillation (ENSO) and Indian Ocean Dipole- on seasonal rainfall data across four major agricultural-based weather stations. Copula techniques allow the respective model to fully capture the dependence structure between input(s) and the target variable regardless of the marginal distribution of each variable. The D-vine copula-based quantile approach, used in this study, through Akaike information criterion (AIC)-corrected conditional log-likelihood (cll^{AIC}) can also enable researchers to identify the most influent predictor variables for seasonal rainfall forecasting.

To forecast the monthly and the respective seasonal rainfall for PNG, an agricultural-reliant nation, the statistically significant lagged correlations between ENSO indicators (e.g., SOI, Nino3.0, etc.) and the IOD indicator (i.e., DMI) with a three-monthly total rainfall were established for up to 7 months ahead time. For example, in a 'lead-0' timescale case study for seasonal rainfall forecasting, this study has utilized the January to March average SOI (as a model input) relative to the April to June total rainfall (as the target variable) deduced by the Kendall rank correlation coefficients established between the input and the target variable.

In terms of the results of this study, a correlation analysis performed between the most optimal lead times considering climate mode indices and the three-monthly total rainfall were found to be consistent with the most influent predictor variables identified from the D-vine copula-based quantile model (as a basis to generate bivariate models that captured ENSO impacts on rainfall). To further explore any improvements in rainfall forecast model accuracy, particularly,

the extreme rainfall events, the study has also considered the impact of Indian Ocean Dipole (IOD) index by embedding the DMI into the bivariate model to finally construct a trivariate forecast models that accounts for compound effects of ENSO and IOD on extreme rainfall events.

To ascertain the versatility of the proposed copula-based forecast models as a major contribution of this study, a number of statistical score metrics based on the Willmott's Index (*d*), Nash–Sutcliffe Efficiency (E_{NS}), Legates-McCabe's Index (*L*), root-mean-square-error (*RMSE*), and mean absolute error (*MAE*), including the Relative Root Mean Square Error (*RRMSE*) and Mean Absolute Percentage Error (*MAPE*) are computed from forecasted and observed rainfall data in the testing phase. It was evident that the station Aiyura attained the best result for both the bivariate and the trivariate model, exhibiting $r = 0.63$, $RMSE = 105.99$, $MAE = 89.75$, $E_{NS} = 0.63$, $d = 0.38$, $L=0.20$ with, the $RRMSE = 15.39\%$ for the bivariate study, whereas the trivariate model evaluations generated a score metric of 0.68, 0.42, 0.28 and 14.84%, respectively.

In summary, the copula statistical modelling approaches contributed by this study, can be enabling mechanisms for climate change resilience, measuring and implementing risk management strategies. These predictive tools can have significant implications for applications in many socioeconomic sectors such as water resources management, better farming practices for crop health, and other agricultural management not only in the present study region but also in the other agricultural-reliant nations where rainfall prediction is often challenging task.

CERTIFICATION OF THESIS

This thesis is entirely the work of KINGSTEN OKKA except where it is otherwise acknowledged, for any research results presented is undertaken by the Student. This work is original and has not previously been submitted for any other award, except where acknowledged.

Endorsements

Dr. Ravinesh C. Deo	*20 September 2019*
Principal Supervisor	Date
Dr. Thong Nguyen-Huy	*20 September 2019*
Associate Supervisor	Date
Professor Armando Apan	*20 September 2019*
Associate Supervisor	Date

Student and supervisors' signatures of endorsement are held at the University.

ACKNOWLEDGEMENT

Foremost, I would like to express my deepest gratitude to my Principal Supervisor, Dr. Ravinesh C Deo for his endless commitments to directing the research and invaluable guidance and support throughout my research. His continuous support and guidance throughout the scholarship application phase (Australia Awards Scholarship) to the completion and submission of this thesis. His own zeal for perfection, courage, and conviction has always inspired me to do more.

I am also very thankful to my Associate Supervisors, Dr. Thong Nguyen-Huy for his undivided attentions in providing me technical support with the copula model development ideas including his technical expertise during this research. I am also very appreciative and would like to thank Professor Armando Apan for significant help in my thesis editing and providing valuable feedback.

I fully acknowledge and am grateful to the Australia Awards PNG scheme for the scholarship to pursue this research study. Also, I would like to thank the Coffee Industry Corporation Ltd for releasing me from work to undertake this study while providing part funding.

I also would like to take this time to acknowledge the University of Southern Queensland for research support and the travel grant to attend a conference in Fiji. The grant enabled me to present my thesis results at the *"International Symposium on Renewable Energy, Environment and Climate Change"* as well as provided me confidence to represent at public forums and disseminate research outcomes.

I am also very grateful to PNG National Weather Service special mention to Kasis Inape and New Britain Palm Oil Ltd (Dami & Ramu) for the rainfall data for this research.

Furthermore, I also acknowledge the *Advanced Data Analytics: Environmental Modelling and Simulation Research* group for sharing research idea in advanced data analysis technique through seminars, and discussions.

Finally, my heartfelt gratitude is bestowed upon my life partner, Rachael, for her courage and strength for keeping the house in order while I was in a distant land, away from my beloved children: Dereck, Soge Hannah, Kaelah Kename, Petra and Aleko. Importantly I dedicate this thesis to my late father, Peter Okka and my precious mother, none of this research work would be possible without their help and support.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AIC	Akaike's information criterion
AEZ	agro-ecological zones
BIC	Bayesian information criterion
CIC	Coffee Industry Corporation
CDF	Cumulative Distribution Function
d	Willmott's Index of Agreement
DMI	Dipole Mode Index
EHP	Eastern Highlands Province
EMI	El Niño Modoki Index
E_{NS}	Nash-Sutcliffe Coefficient
ENSO	El Niño–Southern Oscillation
ESRL	Earth Systems Research Laboratory
GOF	Goodness of Fit
ITCZ	Inter-tropical convergence zone
IOD	Indian Ocean Dipole
λ	Lambda
L	Legates and McCabes index
m asl	meters above sea level
MJO	Madden Julian Oscillation
NOAA	National Oceanic and Atmospheric Administration

NBPOL	New Britain Palm Oil Ltd.
PNG	Papua New Guinea
PNG NWS	Papua New Guinea National Weather Services
POM	Port Moresby
PDF	Probability Distribution Function
PDO	Pacific Decadal Oscillation
r	Correlation coefficient
RMAE	Relative Mean Absolute Error
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
SST	Sea Surface Temperatures
SOI	Southern Oscillation Index

CHAPTER 1: INTRODUCTION

1.1 Background

The synoptic (large-scale) climate drivers, such as El Niño–Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) events, are important phenomena that have a profound impact on precipitation around the world (McBride & Nicholls 1983; Walsh et al. 2001; Smith, I. et al. 2013; Khedun et al. 2014; Ubilava & Abdolrahimi 2019) as well as other documented natural and social disasters (Schmidt et al. 2001). Papua New Guinea, the present study region, is no exception.

It is well known that ENSO, as defined by Dogar et al. (2019), is an inter-annual climate pattern linked with the ocean-atmospheric system with a random, periodic fluctuation in the wind, pressure, and sea surface temperatures (SST) over the tropical Pacific Ocean. ENSO is linked with the warmer than usual (termed El Niño) or, the colder than usual (termed La Niña) water in the central and eastern Pacific Ocean and is often associated with the lower (higher) than normal rainfalls respectively (Cobon et al. 2016). The two most extreme El Nino events recorded in recent times is the 1997-98 and 2015-16 event (Paek et al. 2017).

The Southern Oscillation Index (SOI) is the most commonly used indicator of the ENSO event (Ropelewski & Jones 1987), which gives an indication of the intensity El (La) events in the Pacific Ocean and, is calculated using the atmospheric pressure difference between Darwin , Australia (12.4ºS, 130.9ºE) and Tahiti (17.5ºS, 149.6ºW) (Wolter & Timlin 1998; Khedun *et al.* 2014). The opportunity to generate rainfall predictions based on the behavior of ENSO events is an exhilarating problem of investigation.

In addition to the SOI, four other naturally occurring large-scale phenomena including the sea surface temperatures (SST) fluctuations, has been recognized as an essential component of the ENSO behavior, and as such, SSTs have been used in the tropical Pacific to study rainfall (Rasmusson & Carpenter 1982; Nalley et al. 2019). The Niño1+2 covers the South American coastal SST while Niño3 and Niño4 are located in the eastern and central equatorial Pacific and covers a broader region with, Niño3.4 (most recent) overlapping and located between Niño3 and Niño4 (Barros et al. 1997) have been linked to seasonal rainfall anomalies in different regions globally (Drosdowsky & Chambers 2001; Walsh *et al.* 2001; Smith, I. *et al.* 2013; Zhao et al. 2019).

The occurrence of different types ENSO teleconnections has led to a large number of studies separating these into two types known as canonical ENSO and ENSO Modoki (Ashok et al.

2007; Yuan & Yang 2012; Dogar *et al.* 2019). These ENSO Modoki events have had a profound influence on the temperature and precipitation globally as well (Ashok *et al.* 2007; Feng et al. 2011; Wang & Wang 2013). The large-scale SST which have above-normal SST characterizes the El Niño Modoki in the central equatorial Pacific and below normal SST in both east and west and is quantified by an El Niño Modoki Index (EMI) (Ashok *et al.* 2007) which capture the zonal SST gradients in the Pacific (eastern and western). These indicate that a model to forecast rainfall must include the behavior of SST including the ENSO Modoki index, together with SOI discussed earlier.

The long-lived El Niño like pattern in the Pacific is described as the Pacific Decadal Oscillation and is one of the dominant modes of climate variability in the North Pacific Ocean (Zhang et al. 1997; Vishnu et al. 2018). The positive (negative) phases during the El Niño (La Niño) phases of PDO modulates climate unpredictability globally and is linked to precipitation patterns especially when both are in the same phase (Khedun *et al.* 2014). The warm and cold phases of PDO can last up to 30 years and has an influence on precipitation like ENSO (Goodrich 2007) which have been studied in the USA (Goodrich 2007; Khedun *et al.* 2014) and China (Chan & Zhou 2005; Lyu et al. 2019) to name a few.

IOD also influences rainfall and affects climates of countries like Australia, Indonesia, and other countries which surround the Indian Ocean (Ashok, Karumuri et al. 2003; Nur'utami & Hidayat 2016). IOD is defined as the difference in low sea surface temperatures of Sumatra and high sea surface temperature in the western Indian Ocean. IOD accompanying wind and precipitation anomalies play a vital role in Australia's climate, mainly its agriculture as it happens together with the winter crop-growing season (Ashok, Karumuri *et al.* 2003; Yuan, Chaoxia & Yamagata, Toshio 2015).

This Master of Science Research thesis is focussed on monthly rainfall forecasting, one of the most important climatic factors that influence the success (or failures) of the agricultural activities of any nation.

Rainfall is very vital in certain critical stages of agricultural crop productions, and a shortfall of rain can have a profound impact on crop production (Lobell & Burke 2008). It is also a vital source of drinking water for countries who have annual rainfalls of less than 500 mm (Howard & Bartram 2010; Danladi et al. 2018). Extreme climate events such as the El Niño Southern Oscillation (ENSO) are projected to cause widespread drought and extensive flooding in different areas (Wara et al. 2005; Marengo 2015) hindering accessibility to cleaner water (Connor 2015) as well as causing havoc to infrastructures (Ball et al. 2016) in developing

countries. Rainfall forecasting is very useful for agricultural reliant countries as it will significantly assist in making informed and strategic planning of resources (Htike & Khalifa 2010) especially in the context of climate change (Black & Thompson 1978; Drake 1994; Rosenzweig et al. 2001).

The most stable and dominant climate driver that influences the rainfall of PNG is the El Niño Southern Oscillation (ENSO) with the Inter-Tropical Convergence Zone (ITCZ) and Madden Julian Oscillation (MJO)(Madden & Julian 1971) also exerting some influence in PNG climate (Smith, I. *et al.* 2013). Studies conducted on the effects of ENSO driven El Niño drought in 1997 have had a profound impact on agriculture production as well as natural resource management in many parts of PNG (Allen & Bourke 2001; Hombunaka & von Enden 2001; Cobon *et al.* 2016).

Advancing new tools for the forecasting of rainfall using large-scale climate indices, which have recently been seen to be linked to rainfall events (Khedun *et al.* 2014; Nguyen-Huy et al. 2017), is very vital to PNG's agriculture and other sectors as it would help facilitate decisions and policy-making in the sector. Investigations done by Smith, I. *et al.* (2013) on ENSO related rainfall changes in New Guinea, the Southern Oscillation Index (SOI) and Niño 3.4 Sea Surface Temperature (SST) had influence over rainfall with Niño 3.4 SST having a better influence statistically in terms of spatial influence as well as temporal influence.

However, studies on rainfall forecasting in PNG remain rather scarce, although the work of Smith et al., (2013) is one of the few recent studies demonstrating the need for a rainfall forecasting model incorporating large-scale climate indices. The current forecasting technique used by the PNGNWS to do rainfall forecasting is the statistical software called SCOPIC (Seasonal Climate Outlook for the Pacific Island Countries). This statistical model (SCOPIC) uses discriminant analysis (multiple linear regression) and the relationships of Southern Oscillation Index (predictors) or the sea surface temperatures (SST) and monthly rainfall (predictands) to predict rainfall at various lead times, however one of SCOPIC limitations to using only one predictor (SOI, SST) (Cottrill et al. 2013). Further, the linear discriminant analysis assumes that the data follows a normal distribution with each attribute has the same variance, i.e. values of each variable vary around the mean by the same amount on average. However, it is noted that the relationship between ENSO and rainfall has been indicated to be nonlinear in New Guinea regions, as shown in Smith, Ian et al. (2013).

One prominent approach for capturing the nonlinear relationship between rainfall and climate indices without any assumption on the distribution of variables is through a copula-statistical

model. Copula models are essential for this area since they can utilize several multivariate datasets to develop probabilistic predictions, which can increase the confidence in rainfall forecasting (Khedun *et al.* 2014; Nguyen-Huy *et al.* 2017). Recently, these studies successfully modeled seasonal rainfall in Australia and Texas (USA) and provided a new approach to jointly forecast rainfall utilizing large-scale climate mode indices. However, the application of copula models for forecasting rainfall in PNG has not been explored.

This study aims to develop new copula-statistical models to forecast seasonal and annual rainfall in PNG given that no such predictive models have so far been developed for this important region where rainfall patterns can significantly affect important agricultural crops such as coffee, coffee, oil, cocoa, copra, tea, rubber, and sugar and many other important commodities.

1.2 Research Problem

The significance of statistical forecasting models, such as copula-based approaches, in sustainable agricultural practices is evolving very rapidly in the 21st century.

Developing highly innovative agricultural and rainfall forecasting tools is very important to address the issue of climate threats to agriculture and the economy of PNG. Forecasting rainfall is very vital for policymakers, governments, resource managers, and farmers in decision making. A review of previous studies showed that copulas could be used in precision agriculture, and countries like Australia (Nguyen-Huy *et al.* 2017; Nguyen-Huy et al. 2018) have used this method for rainfall forecasting. Decision-makers can also adopt a copula-statistical model for evaluation of uncertainties in water resources and other sectors, and also to develop cross-cutting research initiatives in PNG's coffee and other industries. However, forecast models using copulas have never been explored in PNG.

Since the behavior and rainfall variability is mostly dependent on multiple large-scale climate indicators (e.g., SOI, Niño3.4 SST, Niño3.0 SST, Niño4.0 SST, etc.), which can have a joint moderating effect on rainfall received, the use of several input variables to model rainfall remains an open problem of interest. Hence, multivariate input-based modeling of rainfall considering the contributions from each driver is an essential problem for decision-makers.

Forecast of climate variations with a seasonal lead time is possible as the atmosphere responds to the more slowly varying ocean and land surfaces (Nobre et al. 2019).

This Master of Science Research thesis aims to seek answers to the key research questions, which are fundamentally associated with each objective of this study, as follows:

1) Can we develop the most appropriate probabilistic bivariate copula-statistical model to forecast seasonal rainfall using synoptic-scale climate mode indices relevant to PNG?

2) Can we improve the bivariate models (e.g., rainfall vs. SOI) by the inclusion of additional variable – IOD?

3)What is the performance of bivariate and multivariate models over key regions in PNG's agricultural research stations (Aiyura, Ramu, and Dami) including Port Moresby?

1.3 Research Aim and Research Question

The overall research question associated with this Masters's thesis is connected with the following research questions:

Rainfall is very significant in PNG's Rain-fed agriculture, and accurate forecasting of rainfall is an integral part of PNG's food security. An important question associated with this study is, what is the best copula-based bivariate and trivariate model for forecasting seasonal rainfall in PNG's agricultural stations, including Port Moresby?

1.4 Research Objectives

In this thesis, the influence of ENSO and IOD on rainfall in four rainfall stations in PNG (three are agricultural stations) are modeled using multivariate vine copulas functions. The recently discovered IOD is a physical entity, and many IOD events are not related to ENSO (Ashok, K. et al. 2003) is another important manifestation of the tropical air-sea interaction, which covers the sea surface temperature between the tropical western Indian Ocean (50ºE-70ºE, 10ºS-10ºN) and the tropical southeastern Indian Ocean (90ºE-110ºE, 10ºS).

Copulas Sklar (1959) provides a feasible substitute for modeling non-linear dependence through the utilization of ranked Spearman or Kendall tau coefficients. Rainfall forecasting is essential in addressing the issue of climate threats to agriculture and the economy of PNG.

The study objectives of this Masters thesis are as follows:

1) To develop a bivariate probabilistic copula statistical model to forecast rainfall using synoptic-scale climate mode indices.
2) To develop a multivariate copula statistical model to forecast rainfall using synoptic-scale climate mode indices

To fulfill these objectives, PNG's Rain-fed agricultural sites are considered as case study sites.

1.5 Organisation of Thesis

This Master of Science Research thesis is organized into the following chapters:

Chapter 1 presents the introduction with the background and scientific research gaps, including research questions.

Chapter 2 is the overview of the literature concerning the application of the copula statistical forecasting models and their uses in areas such as hydrology and water resources, agriculture, and meteorology.

Chapter 3 describes the data sources and the methodology of the models. It also presents the model development, including equations of the models, which have been used in this research.

Chapter 4 presents the research results for the bivariate models for forecasting rainfall covers the first objective of this research.

Chapter 5 present the research results for the trivariate models and covers the second objective of this study.

Chapter 6 covers the conclusions and also discusses the limitations and the recommendations for future research in the field of rainfall forecasting.

CHAPTER 2 LITERATURE REVIEW

This chapter focuses on the literature review of large-scale climate mode indices (used as input variables in this) and their influences on rainfall, including a discussion on the conventional statistical-based rainfall forecasting models and their limitations, copula methods and discusses the advantages of using bivariate and trivariate copula statistical methods compared to other rainfall forecasting approaches.

A comprehensive literature review on the application of copula statistical methods is also covered in this chapter.

2.1 Large-scale climate mode indices and its influence on rainfall

The influence of large-scale climate mode ENSO is one of the most critical climatic process affecting rainfall variability in the South Pacific (Kumar et al. 2014). According to Smith, I. *et al.* (2013), ENSO through the phase of SOI and Niño3.4 SST both indices have shown influence the rainfall and the climate of PNG, with Niño3.4 SST having a better influence statistically in terms of spatial influence as well as temporal influence including a more stable influence than SOI. Lagged and extended lead time relationships between climate indices (current) and the next seasonal rainfall are essential to forecast future rainfall, and several studies have done so (Stone et al. 1996). With the current advancement of climatic models using large-scale climate modes indices such as the ENSO, climate information can be forecasted accurately at least six months to one year ahead (Jin et al. 2008; Ludescher et al. 2013). Fiji, for instance, is taking advantage of the strong relationship between SOI and its rainfall and have developed seasonal rainfall forecasting scheme using, the three-month mean of the SOI to forecast the following three-month total rainfall (Walsh *et al.* 2001).

Wide range of large-scale climate mode index has also been documented and identified to have an influence on the climate variability in Australia across different regions as well as the seasons (Ashok, K. *et al.* 2003; Nguyen-Huy *et al.* 2017). The well documented ENSO phenomenon has a significant influence on precipitation over the north and east end of the continent with El Niño associated with drought conditions, whereas substantial precipitation is linked to La Niña events (Yuan, C. & Yamagata, T. 2015). IOD also has a profound influence on Australia's climate variability, and in recent decades it has been identified as the main driver behind the major drought which occurred in the southeast of the continent and not ENSO as

presumed (Cai et al. 2012). There are three phases in IOD (positive, neutral and negative) with the positive phase this results in drier and hotter conditions stirring across southern an northeastern Australia during EL Nino while the negative phase of IOD brings in above-average rainfall including colder conditions during La Niña (Min et al. 2013; Nguyen-Huy *et al.* 2018). Furthermore, as indicated by Ashok *et al.* (2007), the ENSO Modoki phenomenon, have displayed different teleconnection patterns to Australia's climate compared to ENSO. Hence, the finding reinforces the impact of large-scale climate mode indices on the weather.

2.2 Common statistical-based rainfall forecasting models and their limitation

Long-range rainfall forecasting is extremely valuable and can be used as a planning tool to mitigate the harmful effects of drought and water resource management. Conventional statistical-based rainfall forecast models, including the regression forecast models, have been applied in rainfall forecasting as shown by Kim and Kim (2010), who developed an empirical statistical model using climate indices (ENSO) as predictors to forecast monthly precipitation in China with a two to twelve month lead time, Korecha and Barnston (2007) applied regression forecast models as well to forecast Ethiopia's June-September rainfall using ENSO. Furthermore, in South Australia Rasel et al. (2016) applied multiple regression models to investigate the relationship between legged climate predictors ENSO (SOI) and Southern Annular Mode with spring precipitation and demonstrated 63% prediction accuracy using combined climate predictors.

However, with regression models as used in these studies, the dependence structure was measured using Pearson's correlation between the rainfall (precipitation) and the predictor climate index while assuming Gaussian distribution and linearity of precipitation data, models assume the random variable follows the normal distribution which may not always hold in practice. Pearson's correlation assesses how well a linear model fits the data. However, rainfall data are generally skewed in distribution, and this invalidates the use of Pearson's correlation while (Schepen et al. 2012). To overcome these, copula statistical models (Sklar 1996) provide a practical alternative for modelling non-linear dependence by utilizing the Spearman or Kendall coefficient and have been applied by forecasting method by employing copula statistical models, which are yet to be applied in rainfall forecasting (Khedun *et al.* 2014; Nguyen-Huy *et al.* 2017).

2.4 What are copulas?

The origin of the word **copula** is a Latin word **copulare**, which means 'to join, link, tie together.

Copula was first introduced as Sklar's theorem (Sklar. 1959) and was investigated further by statisticians like Genest and Mackay (1986), Genest and Rivest (1993), (2003), Christian and Anne-Catherine (2007), Czado (2010) and Schepsmeier (2010). Copulas were revitalized in the 1980s and have since attracted significant attention in the fields of agricultural science (Ribeiro et al. 2019), finance (Nagler et al. 2019), hydrology (Grimaldi et al. 2016), drought (Vazifehkhah et al. 2019), flood (Chen & Guo 2019b), and streamflow (Liu et al. 2015).

Copula statistical models provide a viable alternative for modeling non-linear dependences and are based on rank dependence (i.e., Kendall's tau correlations). Kendall's measures dependence, based on the ranks of the data and provides, arguably, the best alternatives to the linear correlation coefficient as a measure of dependence for non-elliptical distributions, for which the linear correlation coefficient is unsuitable and often confusing (Manner et al. 2019). Copula enables modeling of multivariate data through advanced techniques, which are handy in studying scale-free measures of dependence, including constructing families of bivariate/multivariate distributions, which have non-linear dependence. Further, the marginal distributions (i.e., the distributions of each variable) do not have to follow a normal distribution.

2.5 Copula Theory

This section outlines necessary steps in defining copulas; additional information can be found in the works of Joe (1997) and Nelsen (2007) including Kraus and Czado (2017) and Nguyen-Huy *et al.* (2018).

A copula is a multivariate distribution whose one-dimensional margins, each of which is marginally uniformly distributed over (0,1). Copula models capture the dependence structure between random variables irrespective of their marginal distributions. Sklar (1959) through Sklar's Theorem revealed that suppose each continuous variable X_i (i.e., the climate mode index and the rainfall used in this study) has its own marginal probability density function (PDF), and marginal cumulative distribution function (CDF) denoted as $f_k(x_k)$ and $F_k(x_k)$ respectively. The CDF of a *d*-dimensional variable $(X_1,...,X_d)$ can be written as :

$$F(x_1,x_2,...,x_d) = C\left[F_1(x_1),F_2(x_2),...,F_d(x_d)\right], \quad (1)$$

with the corresponding joint PDF (Sklar 1996):

$$f(x_1,...,x_d)=\left[\prod_{i=1}^{d} f_i(x_i)\right]c\left[F_1(x_1),...,F_d(x_d)\right]. \quad (2)$$

The *d*-variate distribution function associated with the unit hypercube $C:[0,1]^d \rightarrow [0,1]$ with uniform marginal distribution function is unique, called *copula*, and the corresponding copula density $c=\frac{\partial^d}{\partial_1...\partial_d}C(u_1,...,u_d)$ were $u_i=F_i(x_i)$ with $u_i \in (0,1)$ is known as a *probability integral transform* (PIT), (Kraus & Czado 2017). Any copula density c can be decomposed into a product of $d(d-1)/2$ bivariate (conditional) copula densities. It is evident from equation (1) that the marginal distribution can be modeled separately from the dependence structure, which is an advantage of the copula approach.

2.6 Types of copula functions

There are many families of copulas, including elliptical, Archimedean, vine, extreme value, and other families (Yan et al. 2010). Archimedean copulas have either symmetric or asymmetric forms. While the Archimedean (symmetric or asymmetric) and multivariate-elliptical copulas are two most widely used copula families, however, they do have some challenges and constraints. For instance, the elliptical copulas tend to exhibit symmetric dependence in their tails (Frahm et al. 2003) while Archimedean copulas have one or two dependence parameters (i.e., $d \geq 3$) which can limit their flexibility in higher dimensions. Fortunately, these limitations can be overcome through the use of vine copulas by decomposing the multivariate copulas and constructing several bivariate copulas (Aas et al. 2009; Fischer et al. 2017).

2.6.1 Elliptical copulas

Copulas that corresponds to the elliptical distribution are known as the elliptical copulas, and one of their advantage over the Archimedean family is, they can specify the correlation between each pair of marginals. The Gaussian copula and t-copula are the most commonly used elliptical copulas. The Gaussian copula is often used for modeling in finance (Chakrabarti & Sen 2019), t-copulas have recently been used for modeling hydrological extremes, like flood and drought events (Dodangeh et al. 2019; Vazifehkhah *et al.* 2019), owing to their ability to characterize tails of a distribution. The Gaussian copula, resulting from a multivariate Gaussian distribution, is perhaps the most popular. It is expressed as:

$$C_{\Sigma}\left(u_1,...,u_n\right)=\Phi_{\Sigma}\left[\Phi^{-1}\left(u_1\right),...,\Phi^{-1}\left(u_n\right)\right] \quad (3)$$

where Φ is the distribution function of a standard normal variable $N(0,1)$ and Σ is the correlation matrix with $n(n-1)/2$ parameters satisfying the positive semi definiteness constraint. Φ_{Σ} is the *n*-variate standard normal distribution with mean 0 and covariance matrix Σ. i.e., $\Phi_{\Sigma} \sim N_n(0, \Sigma)$.

The Student's *t* copula can be written as (Khedun *et al.* 2014)

$$C_{\Theta}\left(u_1,...,u_n\right)=t_{v,\Sigma}\left[t_v^{-1}\left(u_1\right),...,t^{-1}\left(u_n\right)\right] \quad (4)$$

Where $\Theta=\left\{(v,\Sigma): v\in(1,\infty),\Sigma\in\Re^{n\times n}\right\}$ and t_v represents one variate t distribution v degrees of freedom. The multivariate Student's t distribution is $t_{v\Sigma}$ with correlation matrix Σ and v degrees of freedom.

2.6.2 Symmetric Archimedean copulas

The Archimedean family is the most common copula family employed in hydrological analyses because of its ease of construction and a wide range of choices for the strength of dependence. Symmetric (or exchangeable) Archimedean copulas refers to one-parameter copulas, and the general form of the Archimedean family is expressed as;

$$C(u)=C\left(u_1,...,u_d;\psi\right)=\psi\left(\psi^{-1}\left(u_d\right)\right), u\in[0,1]^d \ , \quad (5)$$

where $\psi:[0,1]\rightarrow[0,\infty]$ is a strict Archimedean copula generator function and its inverse ψ^{-1} is completely monotonic on $[0,\infty)$. The generator is a decreasing function and is termed strict, and the resulting copula a strict copula, when $\psi(0)=\infty$ and $\psi(1)=0$. The dependence parameter ψ is embedded in the generating function ψ (Hofert 2008). Frank, Clayton, and Gumbel copulas are the most commonly used one-parameter copulas, where their generator and the inverse of generator functions are described in (Yan 2007; Hofert 2008) and shown in **Table 1**

.

Table 1: Constructors of Clayton, Frank, and Gumbel copulas three one-parameter (symmetric) Archimedean copulas

Family	**Generator** $\psi(t)$	**Generator Inverse**
Clayton	$t^{-\vartheta}-1$	$(1+t)^{-1/\vartheta}$
Frank	$-\ln\frac{e^{-\psi t}-1}{e^{-\vartheta}-1}$	$-\vartheta^{-1}\ln\left(1+e^{-t}\left(e^{-\vartheta}-1\right)\right)$
Gumbel	$(-\ln t)^{\vartheta}$	$\exp\left(-s^{\frac{1}{\vartheta}}\right)$

2.6.3 Asymmetric Archimedean copulas

One of the unique characteristics in exchangeable Archimedean copulas is the strong restriction, especially with the dependence of larger dimensions. This is because the dependence among all components is identical or in other words; all standard marginal distribution is of equal dimensions. To allow for asymmetry in exchangeable Archimedean copulas, an important property called associativity is utilized (Hofert 2008).

The asymmetric Archimedean copulas can be recursively structured, called nested (or hierarchical) Archimedean copulas, in two common forms including fully nested Archimedean copula and partially nested Archimedean copulas.

Extending the equation introduced in (5) can be written as;

$$C(u_1,...,u_d;\psi_0,...,\psi_{d-2})=\psi_0\left(\psi_0^{-1}(u_1)+\psi_0^{-1}\left(C(u_2,...,u_d;\psi_1,...,\psi_{d-2})\right)\right) \quad (6)$$

Where $d\geq 3$, the joint CDF in equation (16) in three-dimension variables constructed by the asymmetric Archimedean, can be is expressed as (Hofert 2008):

$$C(u)=C_2\left\{C_1\left[C_1(u_1),C_2(u_2);\psi_1\right],C_3(u_3);\psi_2\right\}, \quad (7)$$

Which is also known as the hierarchical Archimedean copula (HAC), were C_1 and C_2 denotes the same Archimedean copula family with the corresponding parameters $\psi_1\geq\psi_2$. The combination of equation (5) and (6) results in the partially nested Archimedean copula and is expressed as (Hofert 2008);

$$C(u)=C\left(C\left(u_{11},...,u_{1d_1};\psi_1\right),...,C\left(u_{s1},...,sd_s;\psi_s\right);\psi_0\right) \quad (8)$$

$$= \psi_0\left(\psi_0^{-1}\left(\psi_1\left(\psi_1^{-1}\left(u_{11}\right)+\ldots+\psi_1^{-1}\left(u_{1d_1}\right)\right)\right)+\ldots+\psi_0^{-1}\left(\psi_s\left(\psi_s^{-1}\left(u_{s1}\right)+\ldots+\psi_s^{-1}\left(u_{sd_s}\right)\right)\right)\right)$$

$$= \psi_0\left(\sum_{i=1}^{s}\psi_0^{-1}\left(\psi_i\left(\sum_{j=1}^{d_i}\psi_i^{-1}\left(u_{ij}\right)\right)\right)\right)$$

Based on the structures, fully and partially nested Archimedean copulas are referred to as nested (or hierarchical) Archimedean copulas.

2.6.4 Vine copulas

There are three primary forms of vine copulas: (Drawable) D-vine, (Regular) R-vine, and (Canonical) C-vine. The vine copulas were initiated by Joe (1997) which was further evaluated and applied more comprehensively by Bedford and Cooke (2001) as graphical reliance models using Markov trees for describing multivariate variables and the construction of bivariate copulas. A variety of modeling is performed using the vine copulas were asymmetries, and tail dependence poses problems when taken into account, as evident in precipitation forecasting (AghaKouchak et al. 2010; Nguyen-Huy *et al.* 2017).

The application of the vine copulas decomposes the joint PDF in equation (2) into the copula densities (bivariate conditional) with its marginal densities expressed as (Fischer *et al.* 2017; Kraus & Czado 2017);

$$f\left(x_1,\ldots,x_d\right) = \prod_{k=1}^{d} f_k\left(x_k\right)\prod_{i=1}^{d-1}\prod_{j=i+1}^{d} c_{ij;i+1,\ldots,j-1}\left[F_{i|i+1,\ldots,j-1}\left(x_i \mid x_{i+1,\ldots,}x_{j-1}\right),\right.$$
$$\left. F_{j|i+1,\ldots,j-1}\left(x_j \mid x_{i+1,\ldots,}x_{j-1}\right)\right] \quad (9)$$

The bivariate copula densities in this regard, the pair copula is independent of each other with the expression called D-vine copula if all marginal distributions are uniform (Nguyen-Huy *et al.* 2018). The various bivariate copula families, which are essential building blocks of the models, are shown in **Table 2** (Wang et al. 2018).

Table 2 : Detail inference of bivariate copula families

Copula	$C(u,u^*)$	**Generator** $\varphi(t)$	**Tail Dep. (lower, Upper)**	**Parameter Range**
Gaussian			0	$\theta > (-1,1)$
Student-t			$2t_{v+1}\left(-\sqrt{v+1}\sqrt{\frac{1-\theta}{1+\theta}}\right)$	$\theta > (-1,1), v > 2$
Clayton Copula	$\max\left[\left(u^{-\theta}+u^{*-\theta}-1\right)^{-\frac{1}{\theta}}, 0\right]$	$\frac{1}{\theta}\left(t^{-\theta}-1\right)$	$\left(2^{-\frac{1}{\theta}}, 0\right)$	$\theta > 0$
Gumbel Copula	$\exp\left\{-\left[\left(-\ln u\right)^{\theta}+\left(-\ln u^*\right)^{\theta}\right]^{\frac{1}{\theta}}\right\}$	$(-\ln t)^{\theta}$	$\left(0, 2-2^{\frac{1}{0}}\right)$	$\theta \geq 1$
Frank Copula	$-\frac{1}{\theta}\ln\left[1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta u^*}-1\right)}{e^{-\theta}-1}\right]$	$\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(0,0)$	$\theta \in R\{0\}$
Joe		$-\ln\left[1-(1-t)^{\theta}\right]$	$\left(0, 2-2^{\frac{1}{\theta}}\right)$	$0>1$
Clayton-Gumbel		$\left(t^{-\theta}-1\right)^{\delta}$	$\left(2^{-\frac{1}{\theta\delta}}, 2-2^{\frac{1}{\theta}}\right)$	$\theta > 0, \delta \geq 1$
Joe-Gumbel		$\left(-\ln\left[1-(1-t)^{\theta}\right]\right)^{\delta}$	$\left(0, 2-2^{\frac{1}{\theta\delta}}\right)$	$\theta \geq 1, \delta \geq 1$
Joe-Clayton		$\left(1-(1-t)^{\theta}\right)^{-\delta}-1$	$\left(2^{\frac{1}{\theta}}, 2-2^{\frac{1}{\theta}}\right)$	$\theta \geq 1, \delta > 0$
Joe-Frank		$-\ln\left[\frac{1-(\delta t)^{\theta}}{1-(\delta)^{\theta}}\right]$	$(0,0)$	$\theta \geq 1, \delta \in (0,1)$

2.6.5 Marginal Distribution Models

Marginal distributions can be modeled using parametric and non-parametric approaches. The parametric approaches make assumptions about the parameters of the data to hypothesis distributions. Before using the multivariate data for statistical copula modeling, the data is fitted individually to their respective marginal distributions to obtain the marginal parameters. The parametric method fits the data to a given set of theoretical probability distributions and selects the most appropriate distribution function for each variable by applying statistical goodness-of-fit tests (GOF) and graphical analysis. There are three famous GOF tests that can be applied to either discrete or continuous distribution. The Kolmogorov-Smirnov and

Anderson-Darling tests are applied to continuous distributions while Chi-Square goodness of fit test can be applied either to discrete or continuous distribution (Ricci 2005).

The Chi-Square goodness-of-fit test (χ^2) is used in order to discriminate between theoretical distributions applied to binned data (i.e., data put into classes). The (χ^2) test validates the null hypothesis by explaining the direction of data distribution. With higher *p*-values supporting the null hypothesis, in other words, the test statistics (Khedun *et al.* 2014; Nguyen-Huy *et al.* 2017). The chi-squared test is given by:

$$x^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{10}$$

In Equation (9), O_i is observed frequency and E_i is expected frequency, was i and k is the number of bins according to Sturges formula $k = \log_2 N + 1.$ $E_i = N(D(Y_a) - D(Y_b))$ where D denotes CDF for the distribution. The upper and lower limits are denoted by Y_a and Y_b class, i respectively, with N being the sample size. The test statistics are distributed as X^2 a random variable with $k - d - 1$ degrees of freedom, d being the number of estimated parameters.

The Kolmogorov-Smirnov GOF (D) test is used to compare the empirical cumulative distribution function (ECDF) with the theoretical CDF from the stated distribution and is based on the most significant vertical difference between the two functions and is defined as:

$$D = \frac{\max}{1 \le i \le n} \left(F(\chi_i) - \frac{i-1}{n}, \frac{i}{n} - F(\chi_i) \right) \tag{11}$$

Where the empirical CDF denoted as $F_n(x) = \frac{1}{n}.[Obs \le x]$ with Obs = Number of observations.

The null hypothesis is rejected (in corresponds to a given significance level (α)) if the absolute (maximum) value difference between the theoretical CDF and ECDF is more significant than the given significance level. D test performs better than the chi-square test provided the sample size is not too large (Ricci 2005).

Finally, the Anderson-Darling GOF test (A^2) is a general test to examine whether the data is from a given probability distribution and is defined as:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1).\left[\ln F(X_i) + \ln\left(1 - F(X_{n-i+1})\right) \right] \tag{12}$$

The A^2 gives more weight to tails compared to the D test. The distributions hypothesis is rejected from a given significance level (α) if A^2 is higher than the set critical level. The hypothesis regarding the distributional form is rejected at the chosen significance level () if the test statistic, A2, is higher than the critical value obtained

In addition to the GOF test described above, graphical assessment is also employed to select the best fitting distributions, including the PDF, CDF, quantile-quantile (Q-Q), probability-probability (P-P). Basic classical GOF plots include the density and Cumulative Distribution Function (CDF) while the Q-Q and P-P plot reveal how well a specific distribution fits the observed data, as shown in **Figure 1**.

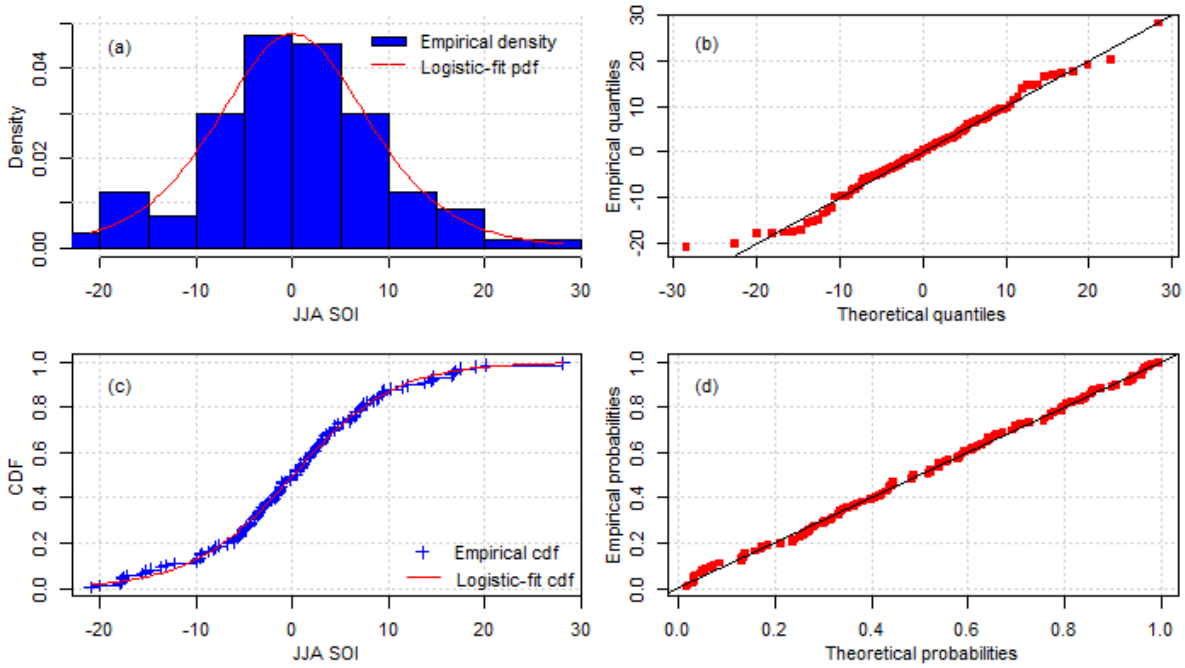


Figure 1: Four Goodness-of-fit plots. (a) shows the density plots with the histogram of the data, (b) Q-Q plot, (c) CDF plot and (d) P-P plot (source: (Nguyen-Huy *et al.* 2017))

The non-parametric approach estimates the probability density function using a weighting function (a kernel). Kernel estimators have engrained tools for nonparametric density estimations and are widely used for data smoothing techniques. Noh et al. (2013) stated that modeling the marginals, including the copula parametrically can cause the resulting parametric estimator to be inconsistent and biased if one parametric model is misspecified. The kernel density estimation is defined by Kraus and Czado (2017);

$$F(x) = n^{-1}\sum_{i=1}^{n} K\left(\frac{x - x^{(i)}}{h}\right),\ x \in \Re \tag{13}$$

The given sample is $\left(x^{(i)}\right)_{i=1,\ldots,n}$, were $K(x) := \int_{-\infty}^{x} k(t)\,dt$ being the symmetric PDF with $k(.)$ and $h > 0$ the bandwidth parameter (which can be implemented with package `ks` (Duong 2007)). This transforms the observed data to pseudo copula data. This non-parametric approach is used in this study.

2.6.6 Copula Models

There are several methods which can be applied to get the estimates of copula parameters: the exact maximum likelihood method, the moment-like method (MOM), which is based on the inversion of the nonparametric dependence measure (e.g., Kendall's tau), and the canonical or maximum pseudo-likelihood (MPL) method and exact maximum likelihood (EML) method.
The MOM approach is based on the reason that the bivariate dependence structure is fully defined by the relative ranks of the integral variables. Based on Kendall's tau τ and Spearman's rho ρ_s obtained from Chowdhary et al. (2011) the relationship is given as

$$\tau = 4\int_{[0,1]^2} C(u,v)c_\theta(u,v)du\ dv - 1 \tag{14}$$

$$\rho_s = 12\int_{[0,1]^2} C(u,v)du\ dv - 3 \tag{15}$$

This relationship, including dependence parameters among τ and ρ_s is specified as $\tau = \frac{2\theta}{9}$ and $\rho_s = \frac{\theta}{3}$ where $-1 \le \theta \le 1$ which has a restricted dependence for Kendall's at $-0.22 \le \tau \le 0.22$ and Spearmen $-0.33 \le \rho_s \le 0.33$. Finally, the moment-based estimate is obtained as $\hat{\theta} = \frac{9\hat{\tau}}{2}$ and $\hat{\theta} = 3\hat{\rho}$.

The other method used to estimate the dependence structure of copulas is the MPL method. With this method, the dependence structure is entirely independent of the represented margins by scaled ranks non-parametrically. The dependence parameter is obtained by maximizing the likelihood function. The real observation data pairs, $x_k = (x_{k1},\ldots,x_{kn})^T, k = 1,\ldots,n$, will be transformed into pseudo-observations, $\hat{u}_k = (\hat{u}_{k1} =,\ldots..,\hat{u}_{kn})$, $k = 1,\ldots,n$, i.e., in the unit hypercube (Nguyen-Huy *et al.* 2017). In this conversion, $\hat{u}_{kj} = (1/(n+1))\ rank\ (x_{k,j})$

Where: $rank(x_{k,j})$ is the rank in ascending order between 1 to n (Khedun *et al.* 2014).

For a copula $C(u_1,...,u_n;\theta)$, with density $c(u_1,...,u_n;\theta)$, the parameter θ is deduced to be:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{k=1}^{n} \log c(\hat{u}_{k1},...,\hat{u}_{kn};\theta) \tag{16}$$

Finally, for the EML method embraces the log-likelihood function in all parameters are simultaneously estimated as reported by Chowdhary *et al.* (2011)

$$l(\theta,\beta,\omega) = \sum_{i=l}^{n} \log\left\{C_\theta\left[F(x;\beta),G(y;\omega)\right]\right\} \tag{17}$$

Where, $\theta$is the association parameter while β and ω denotes the vector parameter of the marginals $F(x;\beta)$ and $G(y;\omega)$.

2.6.7 Goodness-of-Fit Test

To select the best copula pair, GOF tests, criterion and graphical assessment can be used in to determine the best fitting copula models for variables. The standard statistical GOF tests for copula model selections include White and Kendall method.

White method

The White information matrix equality (White 1982) method was introduced through a study by Huang and Prokhorov (2014). The contribution of the study under correct copula specification is that the Fisher information can be computed as minus the expected Hessian or outer product matrix of the score function $-H_0 = C(\theta)$. The null hypothesis is expressed as

$$H_0 : H(\theta) + C(\theta) = 0$$

and the alternative hypothesis as

$$H_0 : H(\theta) + C(\theta) \neq 0,$$

where $C(\theta)$is the expected outer product and $H(\theta)$ the expected Hessian matrix of the score function. This correction in regard to the covariance matrix enables the uncertainty in margins of the test statistics for two-dimensional integrals to be skipped (Huang & Prokhorov 2014).

Kendall method

The Kendall's method based on Kendall's transform process where the p-values associated with the test statistics, are computed through bootstrapping using the Cramer-von Mises and

Kolmogorov-Smirnov test statistic (Wang & Wells 2000; Genest et al. 2006) and are expressed respectively as shown in Genest *et al.* (2006)

$$S_v = \frac{v}{3} + v\sum_{j=1}^{v-1} K_v^2\left(\frac{j}{v}\right)\left\{K\left(\theta_v, \frac{j+1}{v}\right) - K\left(\theta_n, \frac{j}{v}\right)\right\} - v\sum_{j=1}^{v-1} K_v\left(\frac{j}{v}\right)\left\{K^2\left(\theta_v, \frac{j+1}{v}\right) - K^2\left(\theta_n, \frac{j}{v}\right)\right\} \quad (18)$$

and

$$T_v = \frac{\sup}{0 \le w \le 1} |\kappa_v(w) = \sqrt{v}\frac{\max}{i = 0,1; 0 \le j \le n-1}\left\{|K_v\left(\frac{j}{v}\right) - K\left(\theta_v, \frac{j+i}{v}\right)|\right\}. \quad (19)$$

Where $K_v(w) = \sqrt{v}\left\{\mathrm{K}_v(w) - \mathrm{K}_{\theta v}(w)\right\}$, $\mathrm{K}_{\theta v}(w) = P\left\{C_{\theta v}(U,V) \le w\right\}$, and $k_{\theta v}(w) = d\mathrm{K}_{\theta v}(w)/dw$.

Criterion

A quantitative valuation of the performance of various copula families can be made by comparing maximized loglikelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. The two popular GOF indices are AIC and BIC. With *L* representing the log-likelihood, $AIC = -2L + 2d$ and $BIC = -2L + d\ln(N)$. The AIC and BIC are used for model selection, i.e., the model with the lowest index is selected (Maydeu-Olivares & Garcia-Forero 2010). If the interpretation for log-likelihood is higher, the better were as the AIC and BIC should be small on an excellent copula.

Graphical tools

Lambda plot can be used to support selecting the best bivariate copula model by comparing the empirical and theoretical lambda-function of different copula families (Nguyen-Huy *et al.* 2017). Further, the contour plot allows to visualize copula data and compare with overlays of density contours or simulated data from different copula families with fitted parameters (Schepsmeier et al. 2018).

2.7 Modelling Rainfall using Copulas-Statistical Models

Forecasting rainfall with sufficient lead-time with more exceptional skill of reasonable accuracy given the large-scale circulation patterns is extremely valuable for modeling of whether distributions will allow farmers in PNG to make better decisions to reduce their exposures to weather risks or take advantage of the favorable climatic relationship.

Copulas have the advantage of modelling independently the dependence structure of the marginal distributions of variables compared to linear models correlation-based models and are fast becoming a standard tool for multivariate analysis of rainfall and Standard Precipitation Index (SPI) (Rauf & Zeephongsekul 2014b, 2014a) including hydro-logical change and human activities (Li et al. 2019).

In the published literature, Nguyen-Huy *et al.* (2017) investigated the influence of ENSO and Inter-decadal Pacific Oscillation (IPO) Tripole Index (TPI) ─ on spring precipitation forecasting at agro-ecological zones (AEZs) in Australia's wheat belt. The study forecasted spring precipitation, with a significant seasonal lagged correlation of ENSO and TPI and precipitation anomalies in AEZs. Most of the AEZs exhibited statistically significant dependence of precipitation and climate indices, hence bivariate, and trivariate copula models were applied to capture single (ENSO) and dual predictor (ENSO & TPI) influence, respectively, on seasonal forecasting.

Khedun *et al.* (2014) explored the influence of two large-scale circulation patterns ENSO and the PDO, and the effect of ENSO on precipitation in the state of Texas by applying copula statistical models to inspect the dependence structure through bivariate models for ENSO and precipitation, and trivariate models for ENSO, PDO, and precipitation. The study, in general, discovered the inclusion of PDO was found to improve simulation results.

Rainfall forecasting is an essential task for agricultural management of water resources (Deo, Kisi, et al. 2017), and most of PNG's agriculture and food production systems are adapted to cope with much soil-water, should the soil dry up, crops will fail.

2.8 Applications of Copulas-Statistical Model in Other Areas

Since the advent of copula statistical models many years ago (Sklar 1959), its application in the number of fields in recent years has increased. Studies have been conducted in the field of hydrology and water resources (Chen & Guo 2019c; Valle & Kaplan 2019), drought and crop insurance (Wong et al. 2009; Nguyen-Huy et al. 2019), health (Klein et al. 2018), streamflow (Kong et al. 2018) and flood (Chen & Guo 2019a; Moftakhari et al. 2019). Copulas have also been applied to economics, agriculture, and climate forecasts (AghaKouchak 2014; Fousekis & Grigoriadis 2017; Zhang et al. 2018).

In agriculture economics, Nguyen-Huy *et al.* (2018), applied twelve large-scale climate mode indices to model their influence on seasonal wheat crop yield in Australia using the vine-based copula approach. The study revealed that wheat yield could skillfully forecast 3-6 months lead using the IOD and are a fundamental strategy for planning purposes that potentially benefits

farmers by informing risk-management processes. In revenue insurance in agriculture, Okhrin et al. (2013) modeled the systemic weather risk of agricultural production regions in China by using a larger time-series length of data derived from daily weather models.

2.9 Statement of Gaps in Existing Literature

A review of the literature shows that the copula statistical models used in this thesis have been used in countries like Australia (Nguyen-Huy *et al.* 2017) and the USA (Khedun *et al.* 2014) to forecast precipitation. Khedun *et al.* (2014) used Gaussian and Archimedean copulas to model the effect of ENSO and PDO on precipitation anomalies in Texas while Nguyen-Huy *et al.* (2017) applied vine copulas to model the joint influence ENSO (SOI) and IPO Tripole Index (TPI) to forecast seasonal precipitation in Australia. However, copula-statistical models have not been used for probabilistic of forecasting of rainfall in PNG. Therefore, this research study primarily focuses on addressing this gap.

CHAPTER 3 MATERIALS AND METHOD

3.1 Study Area

The study area is in four different regions of PNG; Ramu- Momase, Aiyura- Highlands, Port Moresby -Southern and Dami - Islands. The country experiences relatively high annual rainfall of around 2500-3500 mm although a few lowland areas such as Port Moresby do receive less than 1000 mm and seasonality can be observed in this particular region (McAlpine 1983).

PNG has two main seasons wet and dry seasons with the wet season, encompassing the period generally from November to April (**Figure 4**), and the dry season from May to October. Rainfall varies seasonally in most areas, but the degree of seasonality is not high.

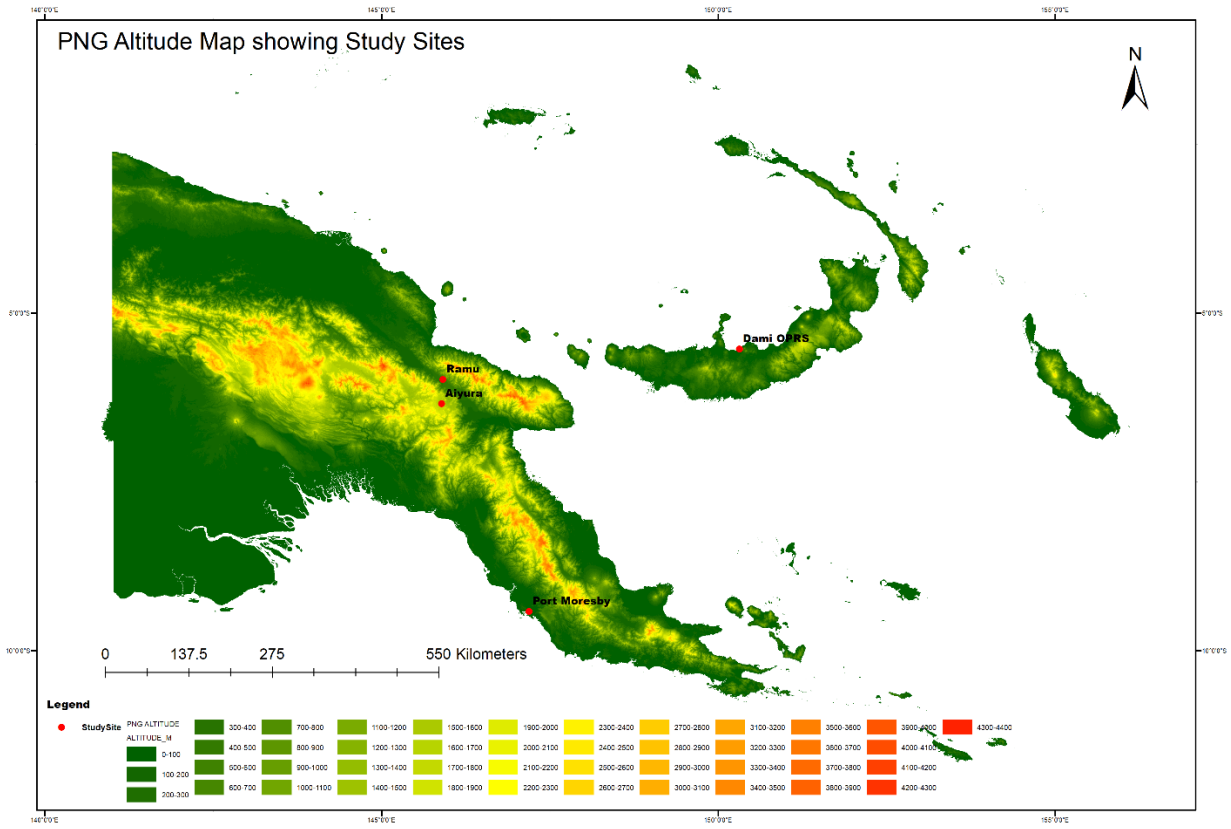


Figure 2: Map of PNG showing the locations of the four rainfall stations used in this study (Source: James Vuvu).

A brief description of these study area are as follows;

1) Port Moresby - Weather Office (9.4º S, 147.2 º E); This meteorological station is the primary climate station, is located at Jacksons Airport near Port Moresby in PNG's capital city and managed by the Papua New Guinea National Weather Services

(PNGNWS), a government institution which provides meteorological and climate information the country.

2) Aiyura (6.3º S, 145.9 º E); This weather station is located in the highlands region of PNG and is managed by Coffee Industry Corporation Ltd.'s (CIC) – Research and Growers Services Division (R&GSD) – which mainly conducts research and provides extension services on coffee in PNG.
3) Ramu (5.9º S, 145.8 º E); This weather station is in the Ramu Valley sits astride the Morobe – Madang provincial boundary and is managed by Ramu Agri Industries (New Britain Palm Oil Ltd) and primarily conducts research on Sugarcane including oil palm (*Elaeis guineensis*)
4) Dami (5.5º S, 150.3 º E); this weather station managed by New Britain Palm Oil Ltd (NBPOL) is mainly responsible for oil palm research and provides plantation managers with the necessary technical information needed for sustainable management and production of oil palm.

3.2 Data

3.2.1 Rainfall Pattern and data

PNGNWS has around 18 weather stations which carry out several multiple observations within 24 hours: two in the Highlands, four synoptic stations in Momase, six in the New Guinea Islands and six in the Southern region. In addition to the multiple observation stations, there are18 single observation rainfall stations and three single observation climate stations. The main issues with these rainfall stations are the missing data and are mainly attributed to funding to replace aging equipment or lack of staffing.

The only reliable primary climate station was Port Moresby Weather Office, located at Jacksons International Airport. For the other three agricultural weather stations (AIY, DMI, and RMU) used in this study, are from vital agricultural institutions in PNG As shown in **Figure 2**, the study area is situated at different elevations throughout the rugged terrains of PNG, with Port Moresby, Aiyura, Ramu, and Dami stations located at an elevation of 50, 1658, 450 and 30 m above sea level respectively.

The average annual rainfall varies noticeably, as shown in **Table 3** is around 98.25 -305.97 mm, with the median between 62-305.97 mm and a standard deviation between 93.45-221.10 mm. From the four rainfall stations, two have a positive kurtosis value (POM and DMI) that

indicates the distribution of data have heavy tails, including outliers. All the four stations have positive skewness which indicates the data distribution is right-skewed with the skewness of the distribution curve having tails at a value above the median and mean is on the right of the peak value.

Table 3: Station name and statistics for monthly rainfall station in PNG from 1990-2017

Station Name	Mean(mm)	Median(mm)	Standard Deviation(mm)	Skewness	Kurtosis
POM	98.25	62.00	101.90	1.18	0.70
AIY	172.09	159.00	93.45	0.48	-0.46
RMU	169.20	159.35	114.15	0.47	-0.60
DMI	305.97	232.40	221.10	1.37	1.72

Seasonality is most evident in POM where it is quite dry, but there is no period nil monthly rainfall, which is found to be true in 'monsoon' climates (McAlpine 1983). This is quite noticeable in **Figure 3** showing the widest sections of the violin plot representing POM having the highest probability of having rainfall between the 60 mm region and shows the median monthly rainfall is also lower, this is not the case for the stations. A violin plot is a kernel density plot combined with boxplots to show the probability distribution of the data set.

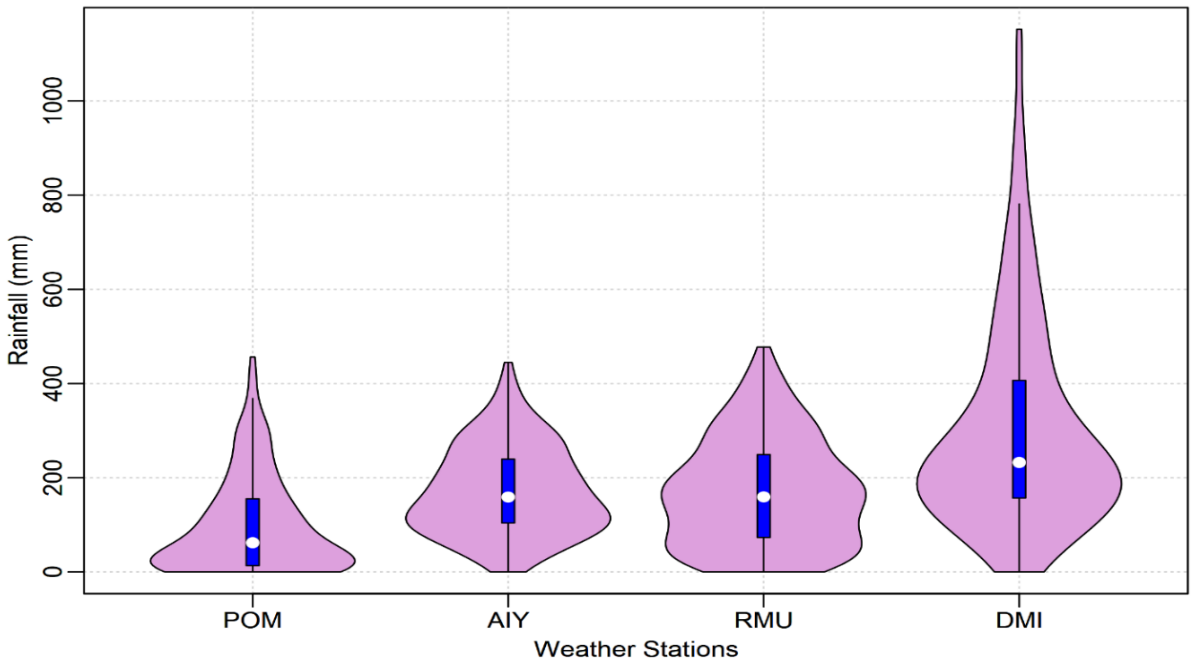


Figure 3: Violin plots of monthly rainfall in each rainfall station used in this study; POM (Port Moresby), AIY (Aiyura), RMU (Ramu) and DMI (Dami). The thick blue

line and white dot indicate the 25th and 75th percentile range and median, respectively, and the thin black line shows the 5th and 95th percentile range.

The mean monthly rainfall in each station, along with that of the whole study site, is given in **Figure 4**. The rainfall pattern is unimodal with a very distinct unimodal rainfall pattern observed in Port Moresby. The average monthly rainfall varies from the lowest of 22 mm to the 580 mm of rainfall across the four stations with Dami and Port Moresby receiving the highest and lowest, respectively with the wettest months from January to March (although it starts around October).

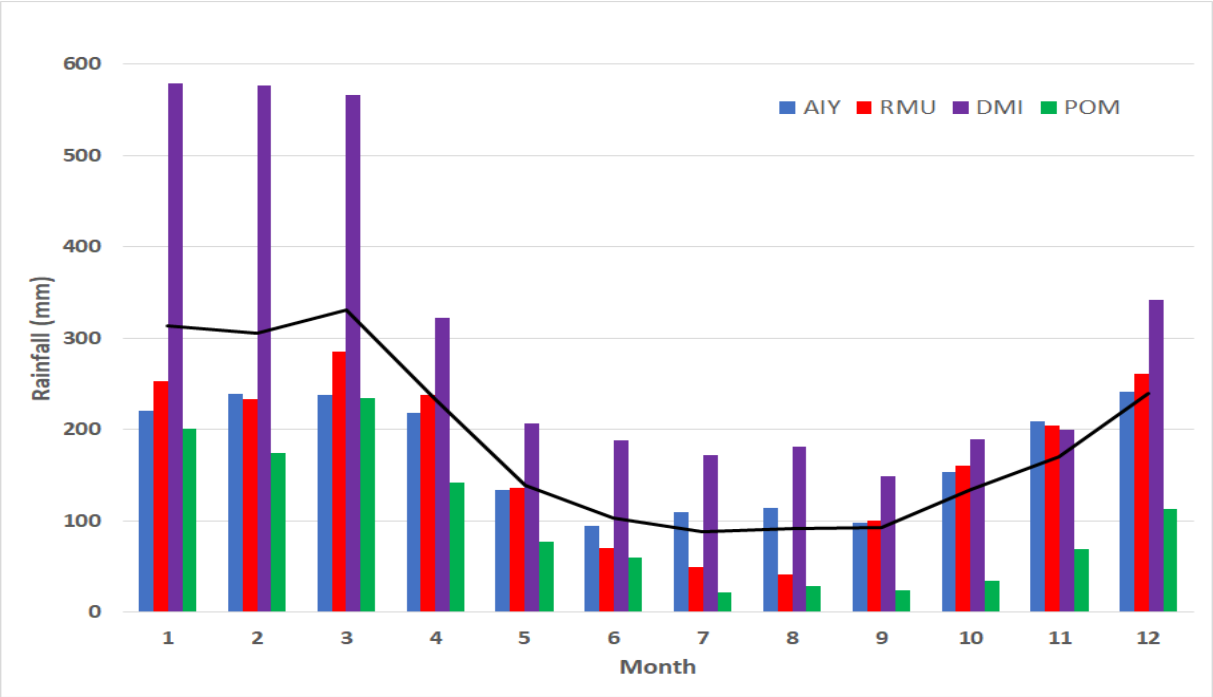


Figure 4*:* Mean monthly rainfall (mm) for the study sites. The solid black line represents the monthly mean rainfall for all the stations based on data for 1990–2017. POM (Port Moresby), AIY (Aiyura), RMU (Ramu) and DMI (Dami)

With regards to the quality of the data, these weather stations have engaged whether attendants who attend to the daily recordings fulltime, hence as shown in **Table 4** the missing day's data percentage is minimal, although Aiyura has 2.39%, it is still less than 10% (Simolo et al. 2010).

Table 4*:* Meteorological stations with record period, latitude, longitude, altitude and missing data details

Station	Record Period	Latitude S	Longitude W	Days Missing	Missing days (%)
Ramu		5.9	145.8	4	0.04
Dami		5.5	150.3	9	0.1
Aiyura	1990-2017	6.3	145.9	244	2.39
Port Moresby		9.4	147.2	10	0.1

3.2.2 Climate Mode Indices

The present study will consider previous works of Nguyen-Huy *et al.* (2017), and Khedun *et al.* (2014) applied for Australia and USA, respectively, who have developed copula models to forecast seasonal rainfall. Six climate indices are used to investigate their impacts on rainfall (**Table 5**) in PNG. These climate indices have been identified to have a relationship with monthly and seasonal rainfall patterns in PNG and the Pacific (Smith, I. *et al.* 2013). Independent correlation performed for various climate-mode indices were also tested but performed poorly.

Several different climate indices are used to measure and monitor the ENSO phenomena such as the EMI (ocean-atmosphere index), SOI (air pressure index) and the sea surface temperature indices (Niño-3.0, Niño-3.4, and Niño-4.0). Also included in this study is the IOD indices (Dipole Mode Index (DMI)) is used to consider its overall synoptic influence, on rainfall in PNG.

Table 5. The six climate mode indices investigated as inputs (predictors) for the four rainfall stations in PNG. Data Sources: Monthly sea surface temperature (SST) downloaded from the Physical Science Division (PSD, NOAA). SOI data was downloaded from the Bureau of Meteorology, Australia (BOM).

Climate Mode Index (Predictor)	Description	Region
Niño3.0	Average SSTA over 150°–90 °W and 5 °N–5 °S	Pacific
Niño3.4	Average SSTA over 170 °E–120 °W and 5 °N–5 °S	Pacific
Niño4	Average SSTA over 160 °E–150 °W and 5 °N–5 °S	Pacific
SOI	The difference in pressure between Tahiti and Darwin (Troup 1965).	Pacific
EMI	C – 0.5 x (E+W) Where the components are average SSTA over C: 165 °E–140 °W and 10 °N–10 °S E: 110°–70°W and 5 °N–15°S W: 125°–145 °E and 20 °N–10 °S	Pacific
DMI	WPI – EPI	Indian

The source of the monthly SST for the period of January 1989 –December 2017 derived from Physical Science Division (PSD, NOAA) using the Hadley Centre sea ice and sea surface temperature (SST) data set (HadISST1) version 1 (Rayner et al. 2003). The SOI data was obtained from the Australian Bureau of Meteorology (BOM) and the method used in this study is the Troup SOI (Troup 1965) with the standardized anomaly of the monthly mean sea-level pressure difference between Tahiti (175ºS, 149.6 ºW) and Darwin, Australia (12.4ºS, 130.9ºE). The DMI also obtained from BOM were calculated by subtracting the West Pole Index (WPI) from the East Pole Index (EPI) to determine the overall polarization effect. The method used by the Australian BOM. The EMI, which was defined by Ashok *et al.* (2007) to monitor El Niño Modoki activity (summarised in

Table 5), the data was from the Japan Agency for Marine Earth-Science and Technology (JAMSTEC). Three-month averages of these climate indices mentioned were used in this study; January to March (JFM), April to June (AMJ), July to September (JAS) and October to December (OND).

3.3 Forecasting Model Development

The focus of this research study is to investigate the lagged relationships between rainfall and large-scale climate mode indices.

Throughout this study, a 'lead 0, 1, 2...' terminology is applied. For instance, a 'lead-0' is defined as the average large-scale climate mode index (SOI, SST, EMI or DMI) for each three-month block compared to the average rainfall for the following three-month block.

> *For example, in a 'lead-0' study January to March SOI is compared with April to June rainfall. For 'lead-1', July to September SOI would be compared with October to December rainfall, a 'lead-2' November to January rainfall and so on.*

For this study, comparisons are made for forecasts up to 'lead-6'.

3.3.1 Correlations between Predictor Variables and the Objective Variable (Rainfall)

Correlations coefficients were calculated to assess the statistical significance of the strength of the relationship between two variables. Despite the extensive use of the Pearson correlation coefficient, several undesirable characteristics associated with its use, due to which we select Kendall's tau as the alternative for measuring dependence.

Kendall's tau is a well-accepted measure of dependence, which assesses statistical associations based on the ranks of the data from the same set of variables (Abdi 2007). It offers the best substitutes to the linear correlation coefficient as a measure of dependence for non-elliptical distributions, for which the linear correlation coefficient is inappropriate and often misleading.

In addition, since the dependence measures depend on the rank only, the approach (rank-based) has the benefit of capturing the outlying observations compared to the linear approach. Further, Kendall's tau plays an essential role in the parametrization of copula functions (Manner *et al.* 2019). In this study, we define 'strong' correlations as those over 0.28.

3.3.2 Forecasting rainfall with *D-vine* based quantile regression model

Suppose a data set (Y, X_1, X_2) which is three-dimensional where Y is the response variable (i.e. rainfall) in this study and (X_1,X_2) are the two covariates (e.g., climate mode indices), then all the three variables using their empirical distribution functions (or probability integral transform (PIT) as described earlier) are transferred to the pseudo-copula data (V,U_1,U_2).

For the purpose of this study, we are interested in forecasting the conditional quantiles q_α of a response variable Y at a random quantile level $\alpha \in (0,1)$ for given covariates $(X_1, X_2, ..., X_d)$. In order to achieve this, the inverse of the conditional distribution can be expressed following Kraus and Czado (2017);

$$q_\alpha(x_1,...,x_d) := F^{-1}_{Y|X_{1,...,X_d}}(\alpha \mid x_1,...,x_d) \tag{20}$$

Applying the PIT with the Sklar (1959) theorem, the response variable $V = F_Y(Y)$ conditioned on predictor variables $U_j = F_j(X_j), ..., U_d = F_d(X_d)$ the equation (3.1) on the right with the inversion can be written as;

$$F^{-1}_{Y|X_{1,...,X_d}}(\alpha \mid x_1,...,x_d) = F_Y^{-1}\left(C^{-1}_{V|U_1,...,U_d(\alpha|u_1,...,u_d)}\right) \tag{21}$$

Hence, given a D-vine copula with response V as the first variable, the conditional copula quantile function $C^{-1}_{V|U_1,...,U_d(\alpha|u_1,...,u_d)}$ can only be expressed analytically with the pair-copulas of D-vine which is not possible for arbitrary regular vines, therefore warranting the use of D-vine copula (see Kraus and Czado (2017) for more details). **Figure 5** shows the three-dimensional D-vine model.

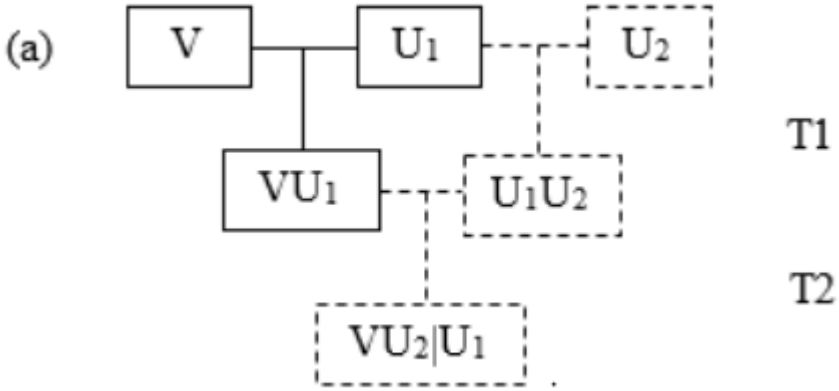


Figure 5: Figure (a) is the model construction framework of a three-dimensional D-vine model (source (Nguyen-Huy *et al.* 2018))

Figure 5, shows an acyclic system in which the variables (V, U1, U2) are called the "nodes" of the D-vine copula model, and two notes are connected together by a unique path. These paths are named the "edges" defined by a corresponding bivariate copula function of those two variables (two nodes). The dashed path implies that whether the inclusion of the variable U2 into the bivariate model of V-U1 increases the cll^{AIC} or not. If the inclusion of U2 improves the cll^{AIC} of the bivariate model, then a trivariate copula model is formed, i.e. V─U1─U2. The algorithm for sampling values (v, u_1, u_2) from a three-dimensional D-vine copula is straightforward. First, sample (w_1, w_2, w_3) are independently drawn from a uniform distribution on $[0,1]$. Then, set:

$$u_1 = w_1,$$

$$u_2 = F_{2|1}^{-1}(w_2|u_1),$$

$$v = F_{3|1}^{-1}\left[F_{3|12}^{-1}(w_3|u_1, u_2)\right],$$

where the conditional cumulative distribution functions $F_{2|1}, F_{3|1}$ and $F_{3|12}$ are defined:

$$F_{2|1}(x_2|x_1) = \frac{\partial}{\partial u_1} C_{12}(u_1, u_2),$$

$$F_{3|1}(x_3|x_1) = \frac{\partial}{\partial u_1} C_{13}(u_1, u_3),$$

$$F_{3|12}(x_3|x_1, x_2) = \frac{\partial C_{23|1}\left[F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right]}{\partial F_{2|1}(x_2|x_1)}.$$

All statistical computations were performed using the VineCopula package (Schepsmeier et al. 2017) available through the R software (Team 2018), Matlab and Q-GIS 3.4 was used to create the map.

3.4 Forecast Model Performance Evaluation

The model performance for rainfall forecasting for the rainfall stations in PNG was ascertained by using the model performance evaluation process.

A robust model assessment requires evaluations from a wide range of statistical matrices as no single statistical measure is final (Willmott 1982; Deo et al. 2018). Pearson's correlation coefficient (*r*), has been identified as an unsuitable measure in hydrologic model evaluation in (Legates & McCabe 1999). A good alternative to the Pearson's correlation is the Nash-Sutcliffe Efficiency E_{NS} (Nash & Sutcliffe 1970) as a "goodness-of-fit" or relative error measure since it is sensitive to differences in the observed and forecasted means and variances (Sumi et al. 2012). A wide range of statistical metrics is used to assess the performance of the developed models used in this study.

The following statistical parameters and definition of the mathematical formulas were considered for the purpose;

I. Correlation coefficient (r);

$$R^2 = \left[\frac{\sum_{i=1}^{n} \left(O_i - \bar{O} \right) \left(S_i - \bar{S} \right)}{\sqrt{\sum_{i=1}^{n} \left(O_i - \bar{O} \right)^2} \sqrt{\sum_{i=1}^{n} \left(S_i - \bar{S} \right)^2}} \right]^2 ; \left(-1 \le r \le 1 \right) \tag{22}$$

II. Root mean square error (*RMSE*) (Chai & Draxler 2014);

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (S_i - O_i)^2} ; \left(0 \le RMSE \le 1 \right) \tag{23}$$

III. Mean absolute error (*MAE*) (Chai & Draxler 2014);

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | \left(S_i - O_i \right) | ; \left(0 \le MAE \le 1 \right) \tag{24}$$

IV. Nash-Sutcliffe Efficiency E_{NS} (Nash & Sutcliffe 1970);

$$E_{NS} = 1 - \left(\frac{\sum_{i=1}^{n} \left(O_i - S_i \right)^2}{\sum_{i=1}^{n} \left(O_i - \bar{O} \right)^2} \right) ; \left(-\infty \le E_{NS} \le 1 \right) \tag{25}$$

V. Willmott's Index of agreement $\left(d \right)$ (Willmott 1982);

$$d = 1 - \left[\frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} \left(|S_i - \bar{S}| + |O_i - \bar{O}| \right)^2} \right]; 0 \le d \le 1 \quad (26)$$

VI. Legates-McCabe's Index (L) (Legates & McCabe 1999);

$$L = 1 - \left[\frac{\sum_{i=1}^{n} |(S_i - O_i)|}{\sum_{i=1}^{n} |O_i - \bar{S}|} \right]; (-\infty \le L \le 1) \quad (27)$$

VII. Relative Root Mean Square Error $(RRMSE;\%)$

$$RRMSE = \frac{\sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} (S_i - O_i)^2}}{\frac{1}{n} \sum_{i=1}^{n} O_i} \text{ x } 100 \quad (28)$$

VIII. Mean Absolute Percentage Error $(MAPE;\%)$

$$RMAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(S_i - O_i)}{O_i} \right| \text{ x } 100 \quad (29)$$

Where O_i = Observed rainfall, $\bar{O}$ =mean of Observed rainfall, S_i = Simulated rainfall and $\bar{S}$ = mean of Simulated rainfall. The first evaluation metric used to assess the models is the Pearson's correlation coefficient (*r*), the *r* is useful in determining if there is a relationship between two different variables, i.e., the Observed O_i and Simulated (S_i) rainfall values. The value of r lies between -1 to 1 with a value of $r = 0$ indicating that the variables have no relation and if *r* = 1 then they have perfect positive relation, i.e., the change in one variable indicates the change in the other variable in the same direction with the same ratio. However, the equation is based on a linear relationship, and it is inappropriate to state that such a measure is statistically significant (Willmott 1982).

The *RMSE* and *MAE* are absolute error measures that express the overall error in the rainfall forecast model and are both used in evaluating the model. The *MAE*, by definition, is never larger the *RMSE* when both matrices are calculated (Chai & Draxler 2014). The *RMSE* represents the sample standard deviation of the difference between predicted values and observed rainfall values. It follows an assumption that errors follow a Gaussian distribution and can evaluate the model with a higher level of skill compared to the correlation coefficient. For a perfect model, *RMSE* should be equivalent to zero; a small *RMSE* value indicates a more efficient model.

The *MAE* calculates the errors as an average of absolute differences between the observed values (rainfall), and the simulated were all individual differences have equal weights (Legates & McCabe 1999). It is used for non-Gaussian cases since it has a linear scoring rule that describes only the average errors, overlooking their direction of variation from the measured values. Thus, *MAE* always is ≥ 0, where *MAE* = zero signify a perfect fit while it ∞ represents the worst fit model.

The Nash–Sutcliffe Efficiency (E_{NS}) is widely used and is the most reliable statistic for assessing the goodness-of-fit of hydrologic models; it determines the relative magnitude of the remaining variance of simulated data in comparison to the measured variance, which in this case is the mean of the observed rainfall values (Prasad et al. 2017). The E_{NS} indicates how the plots of observed versus simulated data fits the 1:1 line, 1 = it perfectly matches modeled simulation to the observed data; 0 = no predictive advantage as it indicates that the model forecasts are as accurate as the mean of the observed data; and negative values when forecasted values deviate (Legates & McCabe 1999). However, the E_{NS}, according to Legates and McCabe (1999) overestimates the higher values, and the lower values are ignored. In order to overcome the disadvantage of the insensitivity issues associated with E_{NS}, Willmott's Index (d) of the agreement was introduced.

The d is a standardized measure and is used to calculate the degree of the prediction error of the model with regards to the observed values of the data by considering the ratio of mean squared error. d = 1 signifies a perfect agreement between the simulated rainfall data and the observed rainfall data with a decrease from d 1 to 0, reduces their agreement as well (Legates & McCabe 1999).

To further compliment the model's evaluation skills, and overcome the over sensitiveness to extreme values the Legates-McCabe's Index (L) (Legates & McCabe 1999) is used as well to evaluate the model's performance since both d and E_{NS} cannot handle. The L value lies between $(-\infty,1)$ and has an absolute value at 1.

Since the four rainfall stations are from four geographically diverse sites, which shows in the different amount of rainfall received, the relative errors; *RRMSE* and *MAPE* were included to evaluate the model performance (Prasad *et al.* 2017). The *RRMSE* is computed by dividing the *RMSE* with the mean of the observed rainfall data, while the *MAPE* calculates the absolute value of average between simulated and observed rainfall data and is measured in percentage error, e.g.; if $RRMSE < 10\%$ is excellent, $10\% - 20\%$ reflects good, $20\% - 30\%$ fair and more than 30% poor (Mohammadi, K. et al. 2015; Despotovic et al. 2016). It is important to note

that the absolute error measures (*RMSE* and *MAE*) have a dissimilarity with the computation of the residual in *RMSE* squared, while in *MAE*, it is not. Also, since the absolute errors are in real units, this limits their ability to assess the model performance across the different sites. Therefore, to get a better understanding of the individual model performance assessments, various diagnostic plots, e.g., forecasting error histogram, boxplots and scatter plot, are also created.

CHAPTER 4 BIVARIATE FORECASTING MODEL

This chapter focuses on developing and evaluating the bivariate copula forecasting techniques with the two different types of El Niño/Southern Oscillation (ENSO): canonical ENSO and ENSO Modoki, using bivariate copula models.

The relationship between different three-month-average ENSO indicators and three-month-total rainfall was first analyzed using rank-based Kendall's correlation coefficients, and then the ENSO indicator with the highest correlation with rainfall was selected to develop bivariate copula models further. This chapter also discusses the results of the models performance evaluated via statistical means such as Willmott's Index (*d*), Nash–Sutcliffe Efficiency (E_{NS}), Legates-McCabe's Index (*L*), root-mean-square-error (*RMSE*), and mean absolute error (*MAE*), plus the mean absolute percentage error (*MAPE*) and the relative root mean square error (*RRMSE*) which concludes this chapter.

This chapter is arranged as 4.1 Introductory note, 4.2 Methods, 4.3 Results and Discussion, 4.4 Comparison of all sites, and 4.5 Concluding remarks.

4.1 Introductory Note

A continuous shortage of rainfall can have a devastating bearing on the environment, society, and economy, particularly in agricultural regions such as PNG. Accurate modeling of rainfall distribution enables farmers to make informed decisions to reduce their exposure to weather risks as well as taking advantage of favorable climatic relationships. Large-scale climate mode indices have been are known to have a noticeable influence on rainfall and have been used in recent years to forecast rainfall using copula statistical approaches (Khedun *et al.* 2014; Nguyen-Huy *et al.* 2017).

ENSO has a profound influence on rainfall in PNG (Smith, I. *et al.* 2013) and will be used to develop the bivariate models to forecast rainfall in this chapter. At a larger scale, the impact of ENSO on rainfall patterns is generally well known, but at the regional scale, it is less known, which is the case over much of the New Guinea region. From the four weather stations used in this study, three stations are from research stations which research coffee (Aiyura), Sugar Cane (Ramu) and Oil Palm with the final central climate station located in Port Moresby.

4.2 Methods

4.2.1 Data and Explanation of Modelling Time Steps

In this study, the three-month mean values of the ENSO indicators (SOI, Niño 3.4 SST, Niño 3.0 SST, Niño 4.0 SST, and EMI) are used to forecast the following three-month block total rainfall (e.g., March-May mean SOI is predicting June-August total rainfall).

Here we define the different lead times as lead-0, lead-1, lead-2 to lead-6. For example, in a 'lead-0', the March-May (MAM; hereafter we use this notation style to refer a consecutive three months) SOI is used to compare the total June-August (JJA) rainfall in each rainfall stations.

For a 'lead-1', February-April (FMA) SOI is compared with JJA rainfall, 'lead-2', January-March (JFM) SOI is compared to JJA rainfall. Comparisons are made up to "lead 6'.

The same for other ENSO indicators, e.g., to forecast FMA rainfall with Niño 4.0: Lead-0 = JFM, Lead-1=DJF, Lead-2=NDJ, Lead-3 = OND, and Lead-4 = SON.

4.2.2 Correlations

The importance of copula models is that it can select the best correlation structure between the predictor-target variables and offers a more robust tool to model dependence structures with jointly correlated variables (Schepsmeier 2015). The rank-based Kendall's correlation coefficient was used to examine the relationship between the two different types of El Niño/Southern Oscillation (ENSO): canonical ENSO and ENSO Modoki on the rainfall across the four study sites in PNG. The climate-rainfall correlation magnitude required for statistical significance is ± 0.28 (at a significance level of 0.05). **Table 6** provides a summary for the three-monthly rainfall forecasting period and the different ENSO lead times based on the correlations. From the table models M3, M4, M10, and M11 for Aiyura, M4, and M10 for Ramu and M4 for Dami displayed weaker correlations and were not further used in the bivariate copula model development.

Table 6: Selections based on the correlations for each of the study sites. The mean three-month climate indices with the lead times were selected based on the highest correlations against the three-monthly total rainfall to use in the construction of the bivariate copula model development. The periods (shown in green) did not have a statistically significant relationship (at a significance level of 0.05). *RFP is the rainfall forecasting period, and CI denotes ENSO Climate Index.

		Port Moresby			Aiyura			Ramu			Dami		
Model Name	**RFP**	**CI**	**Lead (0-6)**	**Kendall**	**CI**	**Lead (0-6)**	**Kendall**	**CI**	**Lead (0-6)**	**Kendall**	**CI**	**Lead (0-6)**	**Kendall**
M1	JFM	Niño 3.4	6	-0.52	EMI	1	0.49	Niño 4.0	0	0.38	Niño 3.0	3	0.45
M2	FMA	Niño 4.0	6	-0.34	Niño 4.0	4	0.38	EMI	2	0.39	Niño 3.0	6	0.37
M3	MAM	Niño 3.4	0	-0.32	SOI	5	-0.19	EMI	3	0.29	Niño 4.0	0	0.33
M4	AMJ	Niño 3.4	0	-0.40	Niño 3.4	0	-0.19	Niño 3.0	4	-0.14	EMI	2	-0.26
M5	MJJ	Niño 3.0	0	-0.28	Niño 4.0	0	-0.39	EMI	0	-0.50	EMI	3	-0.40
M6	JJA	EMI	6	-0.38	Niño 4.0	0	-0.41	Niño 4.0	0	-0.47	EMI	3	-0.32
M7	JAS	SOI	0	0.32	SOI	0	0.43	Niño 3.0	1	-0.51	Niño 4.0	0	-0.40
M8	ASO	SOI	0	0.39	SOI	0	0.36	Niño 3.4	0	-0.45	Niño 3.4	0	-0.49
M9	SON	SOI	1	0.46	SOI	0	0.38	SOI	0	0.45	SOI	1	0.56
M10	OND	SOI	0	0.62	Niño 4.0	3	-0.14	Niño 4.0	5	0.20	SOI	2	0.30
M11	NDJ	SOI	1	0.63	Niño 3.0	0	0.17	Niño 3.0	2	0.29	Niño 4.0	0	0.30
M12	DJF	SOI	1	0.50	SOI	0	-0.37	Niño 4.0	3	0.32	Niño 3.0	4	0.39

4.2.3 Model Development

To fully assess the various forms of joint dependence structures, the rotated bivariate copula versions were also used to fit the bivariate models (Brechmann 2010). The bivariate copula models were constructed using the **vinereg** package, which performs the sequential estimation of a regression D-vine for quantile prediction as described in Kraus and Czado (2017). A total

of ten one and two-parameter copula families were applied to develop the bivariate copula model to assess the impact of multiple climate mode indices on rainfall.

Following this, a one-fold cross-validation method was used to assess the out-of-sample performance of forecasting models at given quantile levels by applying the cross-validation (Nguyen-Huy *et al.* 2018) to the observed and forecasted values using R (Team 2018).

The observed and forecasted values were then assessed using the performance matrices *d*, E_{NS}, *L*, *RMSE*, and *MAE, RRMSE*, and *MAPE* to decide on the best model.

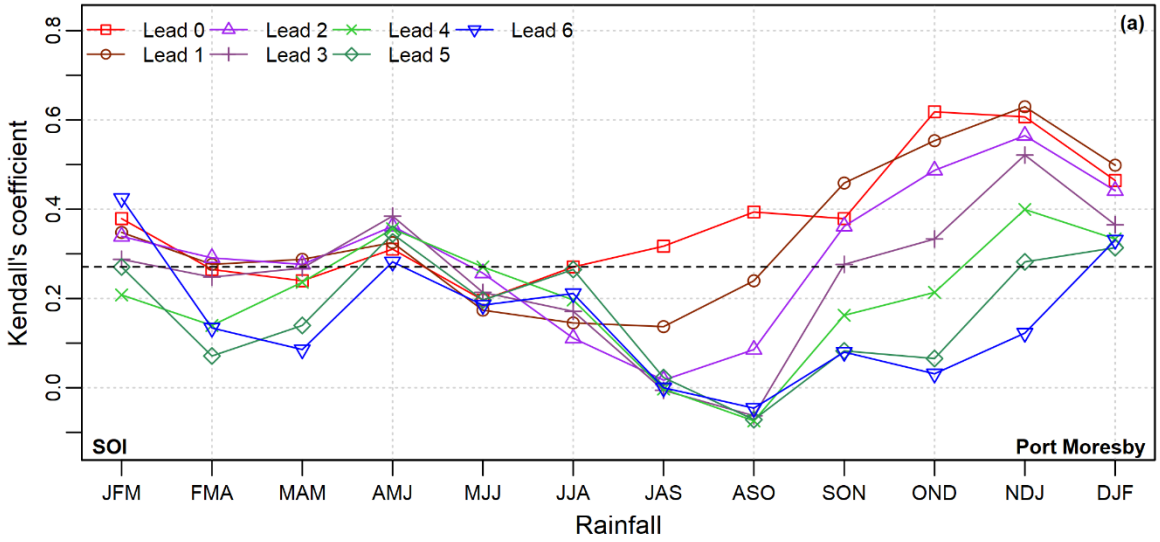
4.3 Results and Discussion

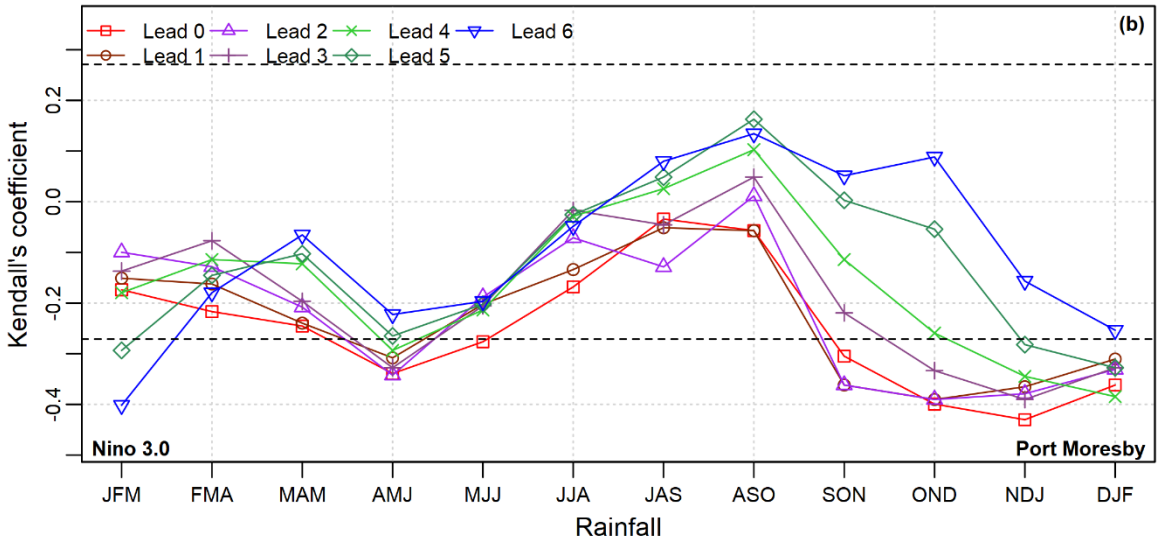
4.3.1 Station 1 – Port Moresby (POM)

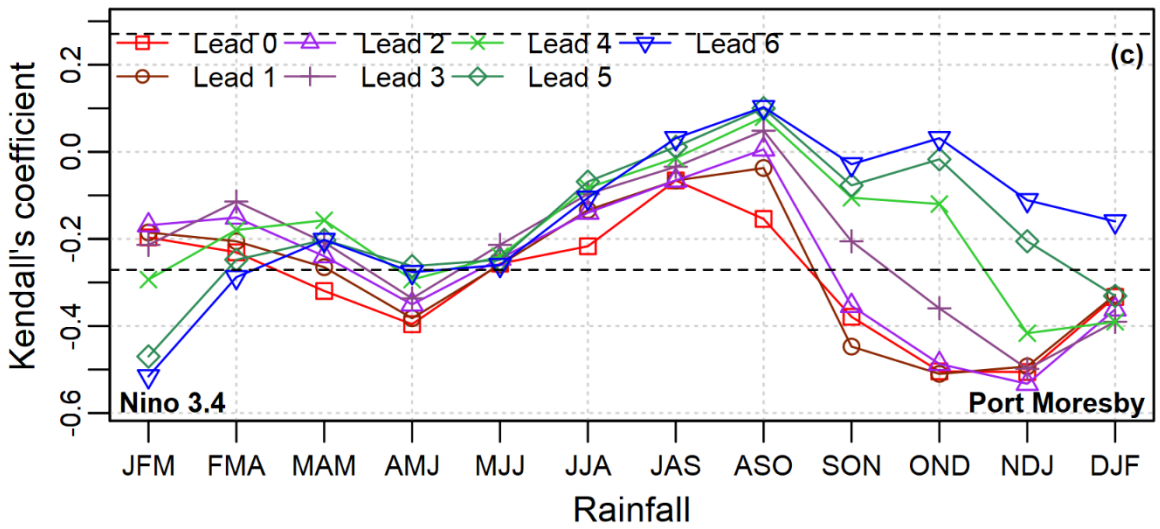
Port Moresby has a clear wet/dry season (**Figure 4**) with the dry season occurring around May to October and the wet season from November to April. **Figure 6** shows Kendall's correlations between the two different types of ENSO and rainfall in Port Moresby.

It is interesting to note that SOI has a strong relationship with rainfall in Port Moresby between the three-monthly periods of JAS to DJF (M7-M12) (Figure 5-(a)), while the SST's exerted a significant relationship between SON-JFM (M1-M5) (**Figure 6**-(b-d)) and EMI on the JJA (M6), OND and NDJ rainfall (**Figure 6**-(e)) (see also **Table 6**). The correlation between the NDJ rainfall and SOI lead-1 had the highest with 0.63, followed by correlation the between OND rainfall and SOI lead-0 (0.62). Niño 3.4 lead-6 had the highest negative correlations (-0.52) with JFM rainfall **Figure 6**(c); these correlations are backed up by Smith, I. *et al.* (2013) while investigating correlations between seasonal average Niño 3.4 and seasonal rainfall over the New Guinea region.

The findings are expected since both the El Niño, and La Niña develops around March to May, which explains why the SST's have good correlations during that period (JFM-MJJ). The El Niño and La Niña gradually strengthen around the November-February period, which explains the significant SOI-rainfall correlations for M10 and M11.


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(a)
Kendall's coefficient
0.8
0.6
0.4
0.2
0.0
SOI
Port Moresby
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(b)
Kendall's coefficient
0.2
0.0
-0.2
-0.4
Nino 3.0
Port Moresby
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(c)
Kendall's coefficient
0.2
0.0
-0.2
-0.4
-0.6
Nino 3.4
Port Moresby
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall

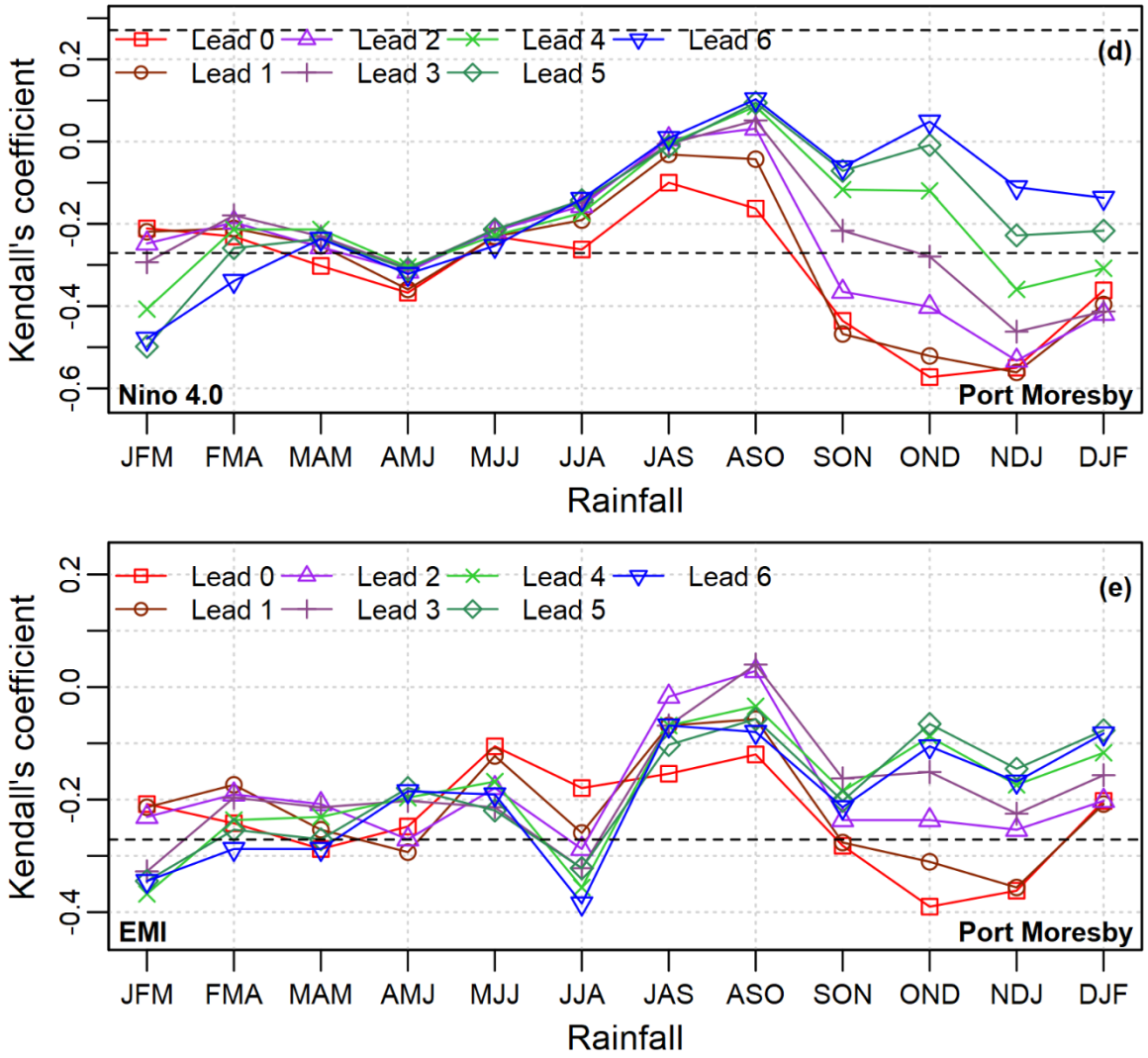


Figure 6: Kendall's Tau correlation between the climate mode indices and the rainfall for Port Moresby, each plot represents different indices, (a) SOI, (b) Niño 3.0, (c) Niño 3.4, (d) Niño 4.0 and (e) EMI (Horizontal dotted lines indicate significant at 5%).

4.3.1.1 Model evaluation

From the selections in **Table 6**, each models performance were evaluated through different visual evaluation such scatter-plots and the coefficient of determination R^2, Boxplots (Observed and Simulated), Frequency plot of absolute forecasting error |FE|, as well through the performance metrics equations [Eq: (22) to (29)] were used to evaluate the ability of ENSO climate index to forecast rainfall. The models M8 and M9 (highlighted in **Table 7**) forecasted negative values in the year 1997, which is an extremely dry year indicating by very low rainfall of 7.8 mm observed for both (ASO and SON and the corresponding rainfall values for the forecasted was -17.94 and -16.85 respectively. The negative forecasted values may lie in the transform step where the quantile function was applied to convert the simulated copula data in [0,1] back to the observational scale. However, this also implies that the copula model has well captured the low part in the dependence structure between the two variables resulting in a correspondingly low value of realization.

From the performance evaluation metrics presented in **Table 7**, the bivariate rainfall forecasting model with SOI lead-1 for NDJ rainfall (M11; hereafter we use this notation style to refer to the model name as provided in **Table 6**) rainfall returned the highest d for all the three months rainfall forecasting periods with 0.85 followed by the M10 (0.68), M1 (0.53), and M9 0.47 as the top-ranked and is further supported with the L = 0.48, 0.31, 0.28 and 0.20 respectively. Overall the M11 model with indicator metrics r = 0.82, $RMSE$ = 0.35,MAE = 64.09, E_{NS} = 0.67, d = 0.85 and L = 0.48 returned better results and is consistent with Kendall's correlation discussed earlier of which M11 had the highest. The next best models include M10, M12 and M1.

Table 7: The performance of the observed and simulated rainfall for Port Moresby, based on Pearson's correlation coefficient (r), *RMSE* (root mean square error), *MAE* (mean absolute error), E_{NS} (Nash–Sutcliffe Efficiency), *d* (Wilmott Index) and *L* (Legates-McCabe's index). The highlighted periods (green) is M8 and M9 model which had forecasted negative values *RFP is the rainfall forecasting period

Model Name	RFP	CI	Lead (0-6)	Lead Months	r	*RMSE* (mm)	*MAE* (mm)	E_{NS} [-∞, 1]	*d* [0, 1]	*L* [-∞, 1]
M1	**JFM**	Niño 3.4	6	AMJ	0.63	130.72	99.79	0.38	0.53	0.28
M2	**FMA**	Niño 4.0	6	MJJ	0.49	145.66	118.57	0.22	0.44	0.09
M3	**MAM**	Niño 3.4	0	DJF	0.20	177.80	142.24	-0.04	0.24	-0.05
M4	**AMJ**	Niño 3.4	0	JFM	0.28	136.45	96.17	-0.01	0.33	0.08
M5	**MJJ**	Niño 3.0	0	FMA	0.14	120.92	93.72	-0.06	0.14	-0.02
M6	**JJA**	EMI	6	SON	0.23	97.83	64.04	0.01	0.31	0.11
M7	**JAS**	SOI	0	AMJ	0.30	46.68	35.83	0.05	0.12	0.11
M8	**ASO**	SOI	0	MJJ	0.45	51.05	39.12	0.18	0.45	0.15
M9	**SON**	SOI	1	MJJ	0.46	94.18	59.34	0.19	0.47	0.20
M10	**OND**	SOI	0	JAS	0.69	89.02	69.53	0.47	0.68	0.31
M11	**NDJ**	SOI	**1**	JAS	0.82	83.35	64.09	0.67	0.85	0.48
M12	**DJF**	SOI	**1**	ASO	0.52	118.59	87.15	0.26	0.43	0.19

The *RMSE* and *MAE* are crucial evaluations for model performance- lower the number, better the model**,** this was observed M7 model for Port Moresby which had the lowest value (*RMSE* = 46.65 mm and *MAE* = 35.83 mm), however, the models performed poorly with the E_{NS} = 0.05, *d* = 0.12 and *L* = 0.11.

The statistical metrics construe the excellent forecasting capabilities of M11 rainfall model compared to the other rainfall models, and this is further supported using additional diagnostic plots in **Figure 7**. The coefficient of determination (R^2) and the linear fit equation also provided. Whereby the R^2 values close to 1 indicate better fit (Hora & Campos 2015). The scatter plot clearly shows the NDJ-M11 bivariate models outperforms the other three models

(OND-M10, JFM-M1, and SON-M9) as their scatter points are diverted from the $y = x$ linear form. The scatterplot of the site Port Moresby is in congruence with the outcomes of Willmott's Index (d), Nash–Sutcliffe Efficiency (E_{NS}) and Legates-McCabe's index (L) which confirms the superior performance of M11 bivariate rainfall model. The R^2 of NDJ model has the highest R^2 = 0.67, and it can be assuredly established that an overall 67% of the observed values could be forecasted for M11 rainfall. Again, the second-best model was M10 with a value of R^2 = 0.47 followed by M1 (R^2 = 0.39) and M12 (R^2 = 0.27). On the other hand, the gradient (m) which is an alternative model performance metric of the linear fit, was observed in M11 to be very closer to unity (i.e., 0.69) which reaffirms the model's superior performance.

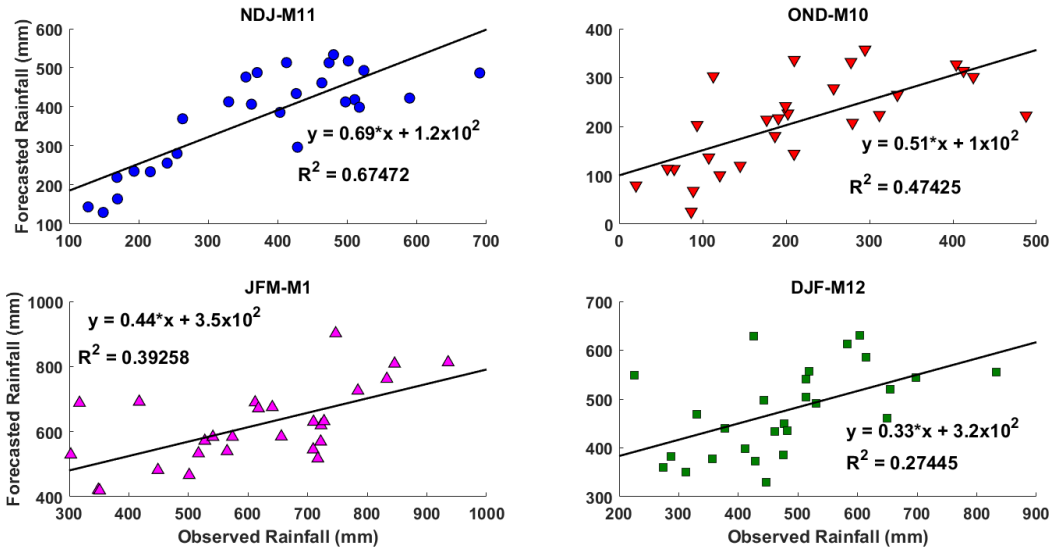


Figure 7: Bivariate copula models developed using lead-1 SOI, lead-0 SOI, and lead6 Niño3.4 as the predictor variable respectively for three monthly cumulative rainfall for NDJ, OND, JFM, and DJF for Port Moresby (POM) rainfall station (PNG). In each panel, a least-squares regression line of the form $R_{Pred} = m\ R_{Obs} + C$ with a coefficient of determination R^2 included, to assess the level of agreement between the forecasted and observed values is included.

To further evaluate the bivariate rainfall forecast models, analysis of the forecasting errors (FE) was applied to assess the competence of the best models for Port Moresby for forecasting the rainfall. The difference between the forecasted and observed rainfall is the FE. **Figure 8** is a histogram plot showing the percentage frequency of errors of various error brackets of ± 100 step sizes, was used to give a better understanding of the model's accuracy and also for practical applications (Deo et al. 2016). A histogram of the frequency error for the four best performing models for Port Moresby is prepared (**Figure 8**). Consistent with the results provided earlier, the most suitable forecast model is M1l, with 89 % of the forecasting errors observed within

the lowest error margin 0 < error ≤ 100. The next best models are the M10 with 81% in the same error bracket ± (0 < error ≤ 100). However, when the next error bracket (100 ≤ error ≤ 200) was assessed M10 shows around 15% (i.e., 96% summed with the first bracket), whereas M10 model 11% (i.e., 100% summed with the first bracket). This demonstrates that the M11 model is better than M10. The next best model (after M11 and M10) is M12 which displayed the highest percentage of errors in the smallest error bracket 70% to be precise compared to M1 with 67%.

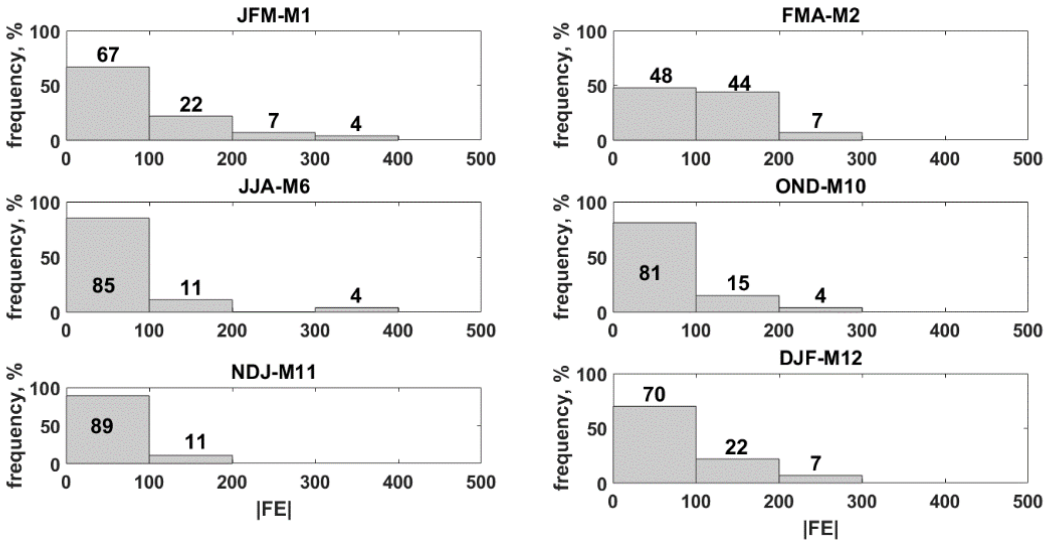


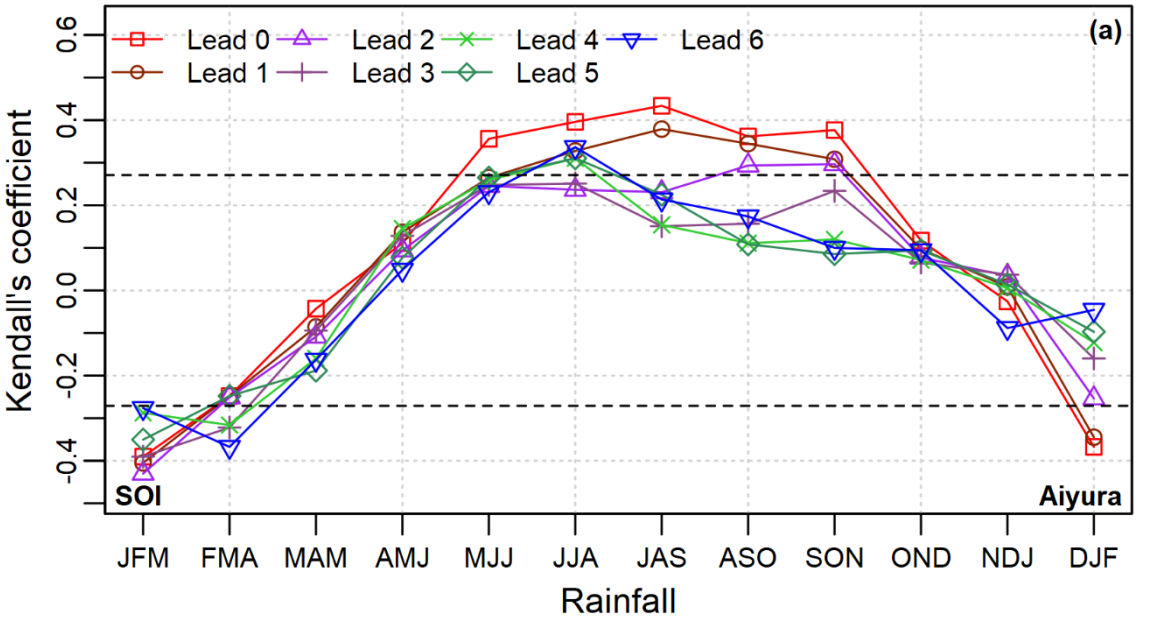
Figure 8: Histogram of the frequency distribution (Freq %) of forecasted errors (FE) generated by the bivariate copula statistical model for rainfall in Port Moresby. The number of datum points in each error bin is shown.

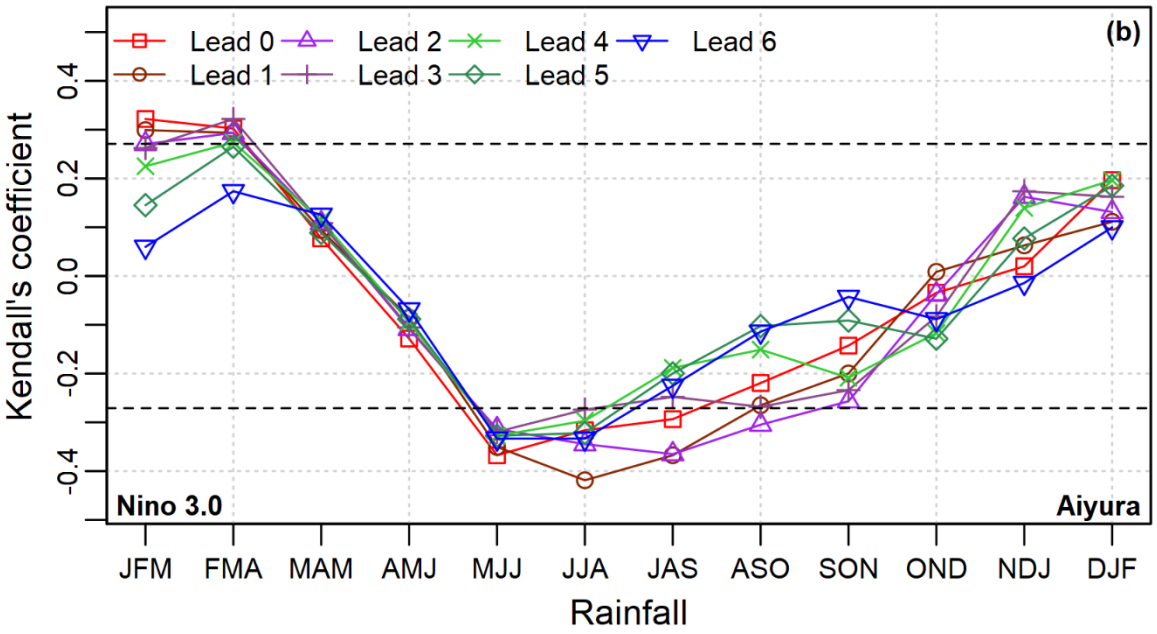
4.3.2 Station 2 – Aiyura

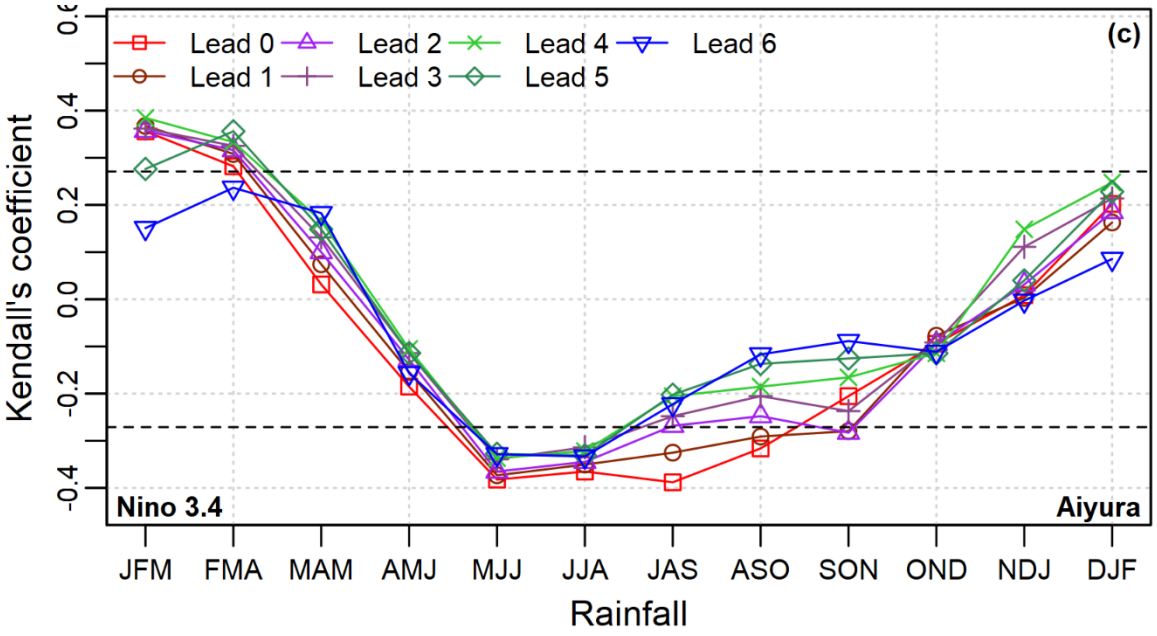
Rainfall data from Aiyura (AIY) - first highlands rainfall stations in PNG established in 1937 - was also used to develop bivariate models. As we have seen in the previous section, the ranked based correlations (Kendall tau) were performed between the three months cumulative total and three-month average climate indices with different lead times to select the best relationship as summarized in **Table 6**.

Before constructing the copula-based bivariate model, the rank-based correlations were performed to identify statistically significant lagged correlations, based on which climate index that had the highest correlation with the 3-month total rainfall. The SOI lead-0 had strong positive correlations with the three months rainfall periods of JAS, ASO, OND, as shown in **Figure 9**(a). The JFM rainfall correlations with EMI leads 1-6 all displayed strong correlations ranging between 0.38-0.49 with lead-1 having the highest (**Figure 9** (e)). It was also found that

none of the climate indices has a statistically significant correlation to the MAM, AMJ, OND, and NDJ rainfall with the correlation values falling between the horizontal dotted lines. The negative correlations displayed by ENSO occurs around the dry season for Aiyura, which is around May to September (less rainfall **Figure 4**).


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(a)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
SOI
Aiyura
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(b)
Kendall's coefficient
0.4
0.2
0.0
-0.2
-0.4
Nino 3.0
Aiyura
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(c)
Kendall's coefficient
0.4
0.2
0.0
-0.2
-0.4
Nino 3.4
Aiyura
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall

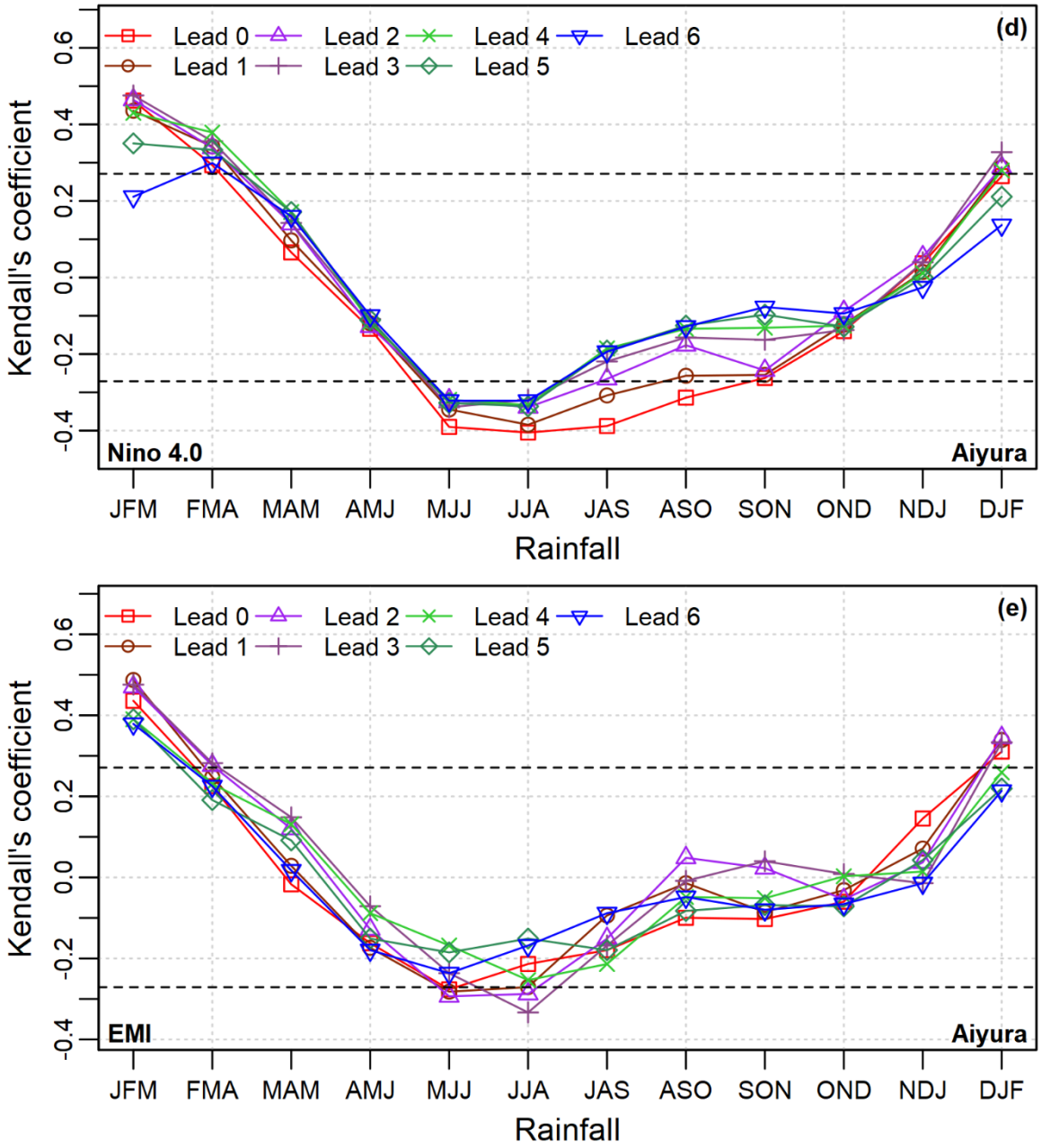


Figure 9: Ranked-based Kendall's Tau correlation between the climate mode indices and the rainfall for Aiyura, each plot represents different indices, (a) SOI, (b) Niño 3.0, (c) Niño 3.4, (d) Niño 4.0 and (e) EMI (Horizontal dotted lines indicate significant at 5%).

4.3.2.1 Model evaluation

Evaluation of Aiyura rainfall models was carried out after generating the forecasted rainfall values. **Table 8** provides the evaluation matrices that were employed to evaluate each model's performance through the conventional matrices *r, RMSE,* and *MAE*, and the goodness of fit matrices E_{NS}, *d* and *L*. The M1 copula-based bivariate model which was developed using JFM total rainfall (predictand) and EMI lead-1 (predictor) registered the highest *r*-value (0.63) and is trailed by M12 (0.51), M6 (0.48), M5 (0.47), M9 and M7 (0.41 apiece), M2 (0.38) and M8 (0.27) in descending order. Further assessment using the absolute error measures, however, indicated that M5 attained the lowest possible *RMSE* and *MSE* values. Although M1 had the next best *RMSE* (105.99 mm), the MAE (89.75 mm) was outdone by M6 (*MAE* = 86.41 mm). Similarly, M5 and M6 outclassed M12 although the model had the second-best r value when compared with the absolute error measures. The results discussed so far are obscured implying an additional combination of metrics to further asses model performance.

Statistical evaluation of model performances using Nash–Sutcliffe Efficiency (E_{NS}), Willmott's Index (*d*), and Legates-McCabe's Index (L), which generally require unity for perfect models, again M1 performed better yielding E_{NS} = 0.63, d = 0.38 and L = 0.20 and reaffirming its superior performance which is consistent with the *r*-value. M12 again displayed the second-best results through the goodness of fit assessment. The other two models, M6 and M5 displayed exceptional results with M6 outclassing M5 with a noticeable difference between the *r,* E_{NS}, and *d* while M6 performed better with the *L* index as shown in **Table 8**. In addition, although M2 had a slightly lower correlation compared to the models discussed so far although it performed better with the E_{NS} index value of 0.37 which is also the same value with M7 and M9.

Table 8: The performance metrics of the observed and simulated rainfall for Aiyura, based on Pearson's correlation coefficient (*r*), *RMSE* (root mean square error), *MAE* (mean absolute error), Nash–Sutcliffe Efficiency (E_{NS}), Willmott's Index (d) and Legates-McCabe's index (*L*). *RFP is the rainfall forecasting period, LTM is the lead time in months.

Model Name	RFP	CI	Lead (0-6)	LTM	*r*	*RMSE* (mm)	*MAE* (mm)	E_{NS} [-∞, 1]	*d* [0, 1]	*L* [-∞, 1]
M1	JFM	EMI	1	SON	0.63	105.99	89.75	0.63	0.38	0.20
M2	FMA	Niño 4.0	4	JAS	0.38	157.14	126.45	0.37	0.10	0.09
M3	MAM	SOI	5	JAS						
M4	AMJ	Niño 3.4	0	JFM						
M5	MJJ	Niño 4.0	0	FMA	0.47	104.40	81.19	0.26	0.20	0.13
M6	JJA	Niño 4.0	0	MAM	0.48	111.26	86.41	0.30	0.21	0.10
M7	JAS	SOI	0	AMJ	0.41	136.11	107.88	0.37	0.14	0.07
M8	ASO	SOI	0	MJJ	0.27	176.82	130.72	0.24	-0.08	-0.01
M9	SON	SOI	0	JJA	0.41	151.12	116.02	0.37	0.16	0.09
M10	OND	Niño 4.0	3	AMJ						
M11	NDJ	Niño 3.0	**3**	MJJ						
M12	DJF	SOI	0	SON	0.51	117.29	97.22	0.47	0.26	0.10

Moreover, to further verify suitable rainfall models, the scatter plots of observed and predicted rainfall was employed. Consistent with the results discussed so far, the two best performing rainfall models was reaffirmed to be M1 and M12 (**Figure 10**). The plot clearly shows that the M1 rainfall model performed better with R^2 (coefficient of determination) of 0.39 which is then trailed by M12 (R^2 = 0.26), M6 (R^2 = 0.23), M5 (R^2 = 0.22), M9 (R^2 = 0.16) and M2 (R^2 = 0.14). The scatterplot of the Aiyura is in agreement with the outcomes of d , E_{NS} and *L*, which still confirms the superior performance of M1. An alternative model performance metric is the gradient (m) of the linear fit, and for a good model the gradient should be close to unity and based on **Figure 10** it was reaffirmed again that M1 had the uppermost value of 0.47. However,

it is interesting to note that the gradient for M5 and M6 (0.29 apiece) performed better compared to M12 (0.27).

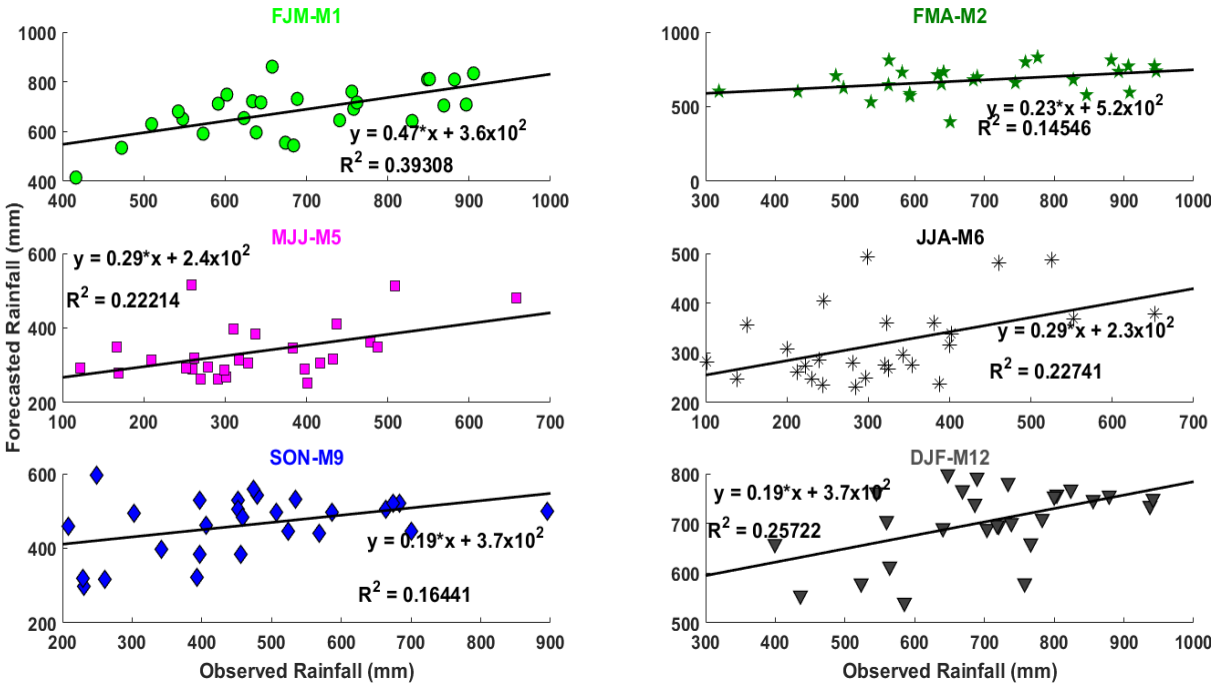


Figure 10: Scatter plot of the six-best performing bivariate copula rainfall models for Aiyura rainfall station (PNG). M1 and M2 (top panel L-R), M5 and M6 (middle panel L-R) and M9 and M1 (bottom panel L-R). In each panel, a least-squares regression line (black) of the form $R_{Pred} = m\ R_{Obs} + C$ with a coefficient of determination (R^2) included, to assess the level of agreement between the forecasted and observed value.

In addition to the scatter plots, the model's preciseness was also evaluated using histograms of absolute forecasting errors plots which has a much more in-depth analysis of actual errors (Deo *et al.* 2016). **Figure 11** shows the histogram plot with numbers displayed on each bar representing the actual percent-age values for the error brackets of 100 step-sizes. The M1 model again maintained its dominance over all the other models registering the uppermost percentage of forecasting errors (74%) in the first bin ($0 < \text{error} \leq 100$) followed by M6 (67%), M9 (63%), M5 and M12 (56% apiece) and finally M2 (48%). Although there is a tie between M5, and M12, the accumulation of percentages indicates that the total forecasting error of M5 when compared to the second error bin $100 \leq \text{error} \leq 200$, was calculated to be 96% while M12 scored a total of 89% of errors which indicates M5 outperforming M12.

There is no doubt M1 performs better, having accumulated all percentage errors within the first and second error bins. Although M12 performed better (second best) in *r* metrics including the

d, E_{NS} and L, its performance with the frequency error plots contradicts the results. One obvious limitation of the *r* metric is it only evaluates the general patterns between the observed and forecasted (modelled) values and not the absolute differences (Legates & Davis 1997). In addition, *r* ranges from 1.0 (perfect model) to 0.0 (poor model) and records the level of agreement between observed and forecasted data. However, it is limited as it based on linear relationships since it standardizes the variance of the observed and forecasted data (DeoTiwari, et al. 2017). On the other hand, the E_{NS}, *d*, and *L* are based on 'average' performance were all forecasted, and observed data is used, and individual error assessment is not made (Willmott 1982; Legates & McCabe 1999). Frequency plots are more useful as it again provides a more in-depth analysis of actual errors of individual model performance.

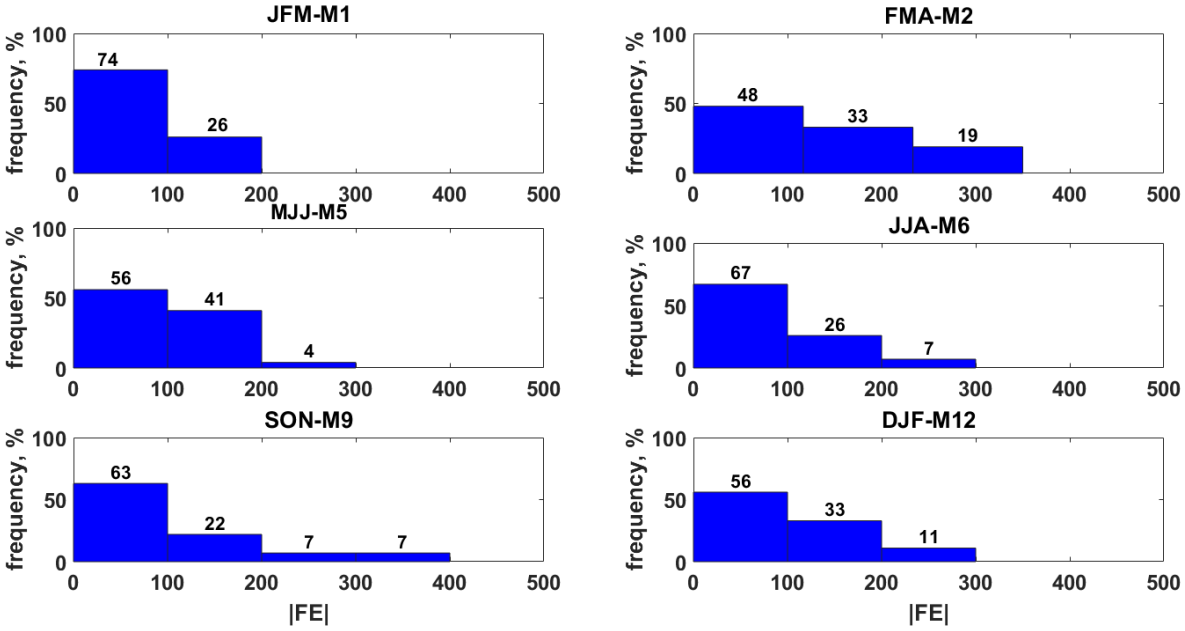
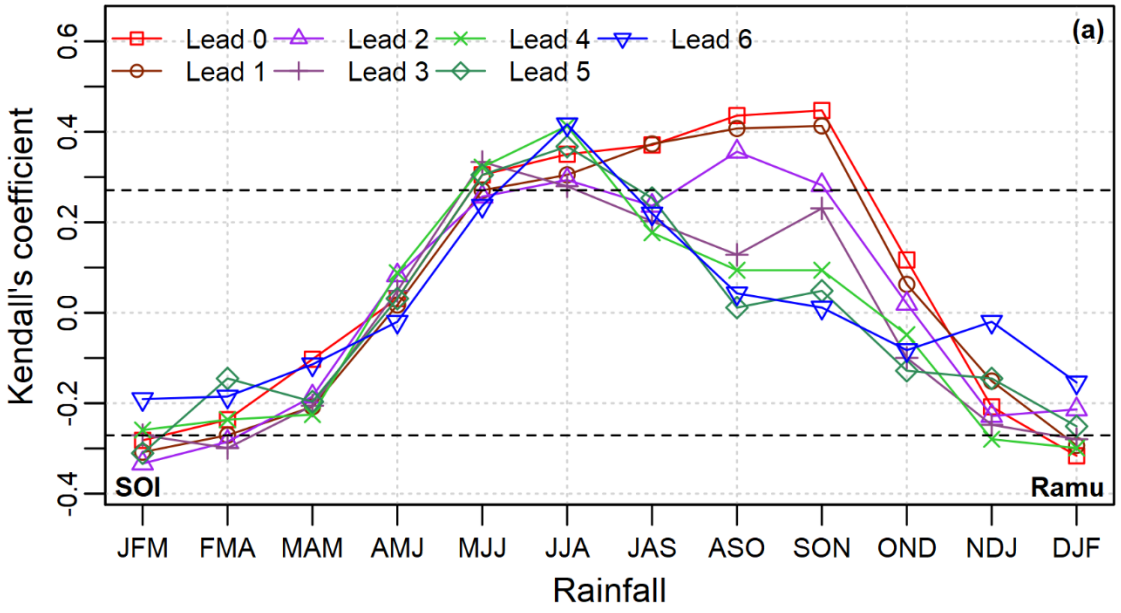


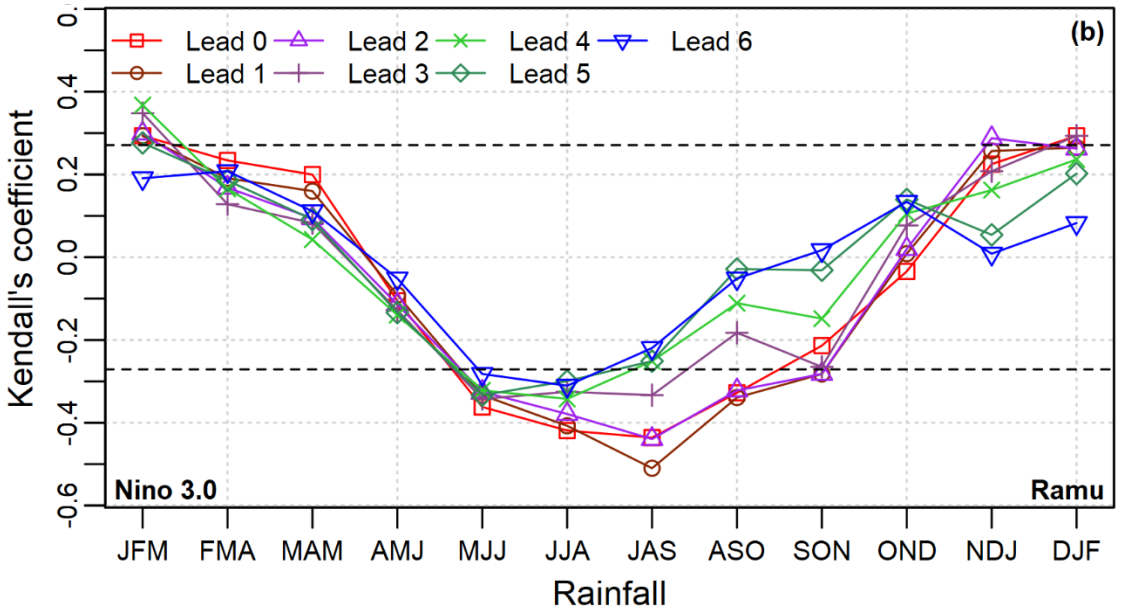
Figure 11: Histogram illustrating the percentage frequency of absolute forecasting errors (|FE|) of the six rainfall forecasting models generated through the copula-based approach (bivariate) for Aiyura study site. Models M1 and M2 (top pane L-R), M5 and M6 (middle pane L-R) and M9 and M12 (bottom pane L-R). Each bar shows the percentage counts.

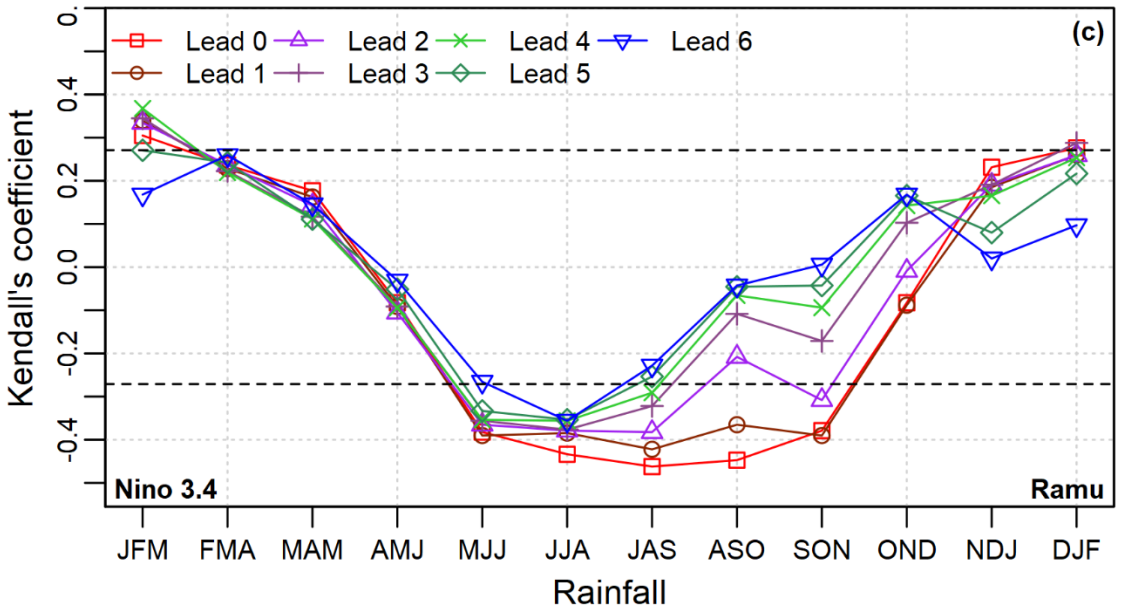
4.3.3 Station 3 - Ramu

Ramu is an important agricultural station and located in the Ramu valley in PNG and is involved in the production of sugar, beef and recently oil palm. Correlations were performed to assess the relationship between the climate index (lead 0-6) and rainfall. Selections of the predictors (climate index) and predictand (rainfall) to develop the copula-based bivariate models were based on the best (highest) correlation, as shown in **Table 6**.

The influence of the SST's over rainfall in Ramu is obvious; Niño 3.0 - JAS and NDJ, Niño 3.4 - ASO, and Niño 4.0 - JFM, JJA, and DJF. The SOI lead-0 exerted its influence on the SON rainfall period with a correlation value of 0.45 as shown in **Figure 12 (a)**. The SST based mode indices displayed a strong negative correlation between JJA – ASO rainfall periods with the highest recorded Niño 3.0 lead-0 (-0.51) for the JAS rainfall while the EMI had a significant association with rainfall between FMA, MAM, and MJJ with the latter having the strongest correlation (-0.51) with lead-0 (**Figure 12**(e)). The EMI correlation with Ramu rainfall is noticeable around the March to May period, and this is consistent with studies conducted in Australia which shows the north-western Australian autumn rainfall is correlated with El Nino Modoki (EMI) (Taschetto & England 2009). On the other hand, the positive correlation recorded between the SON – JFM rainfall on ENSO is consistent with the period in which ENSO's peak intensity (December-April) and weakens during May-July which justifies the negative correlation values (**Table 6**).


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(a)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
SOI
Ramu
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(b)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
Nino 3.0
Ramu
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(c)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
Nino 3.4
Ramu
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall

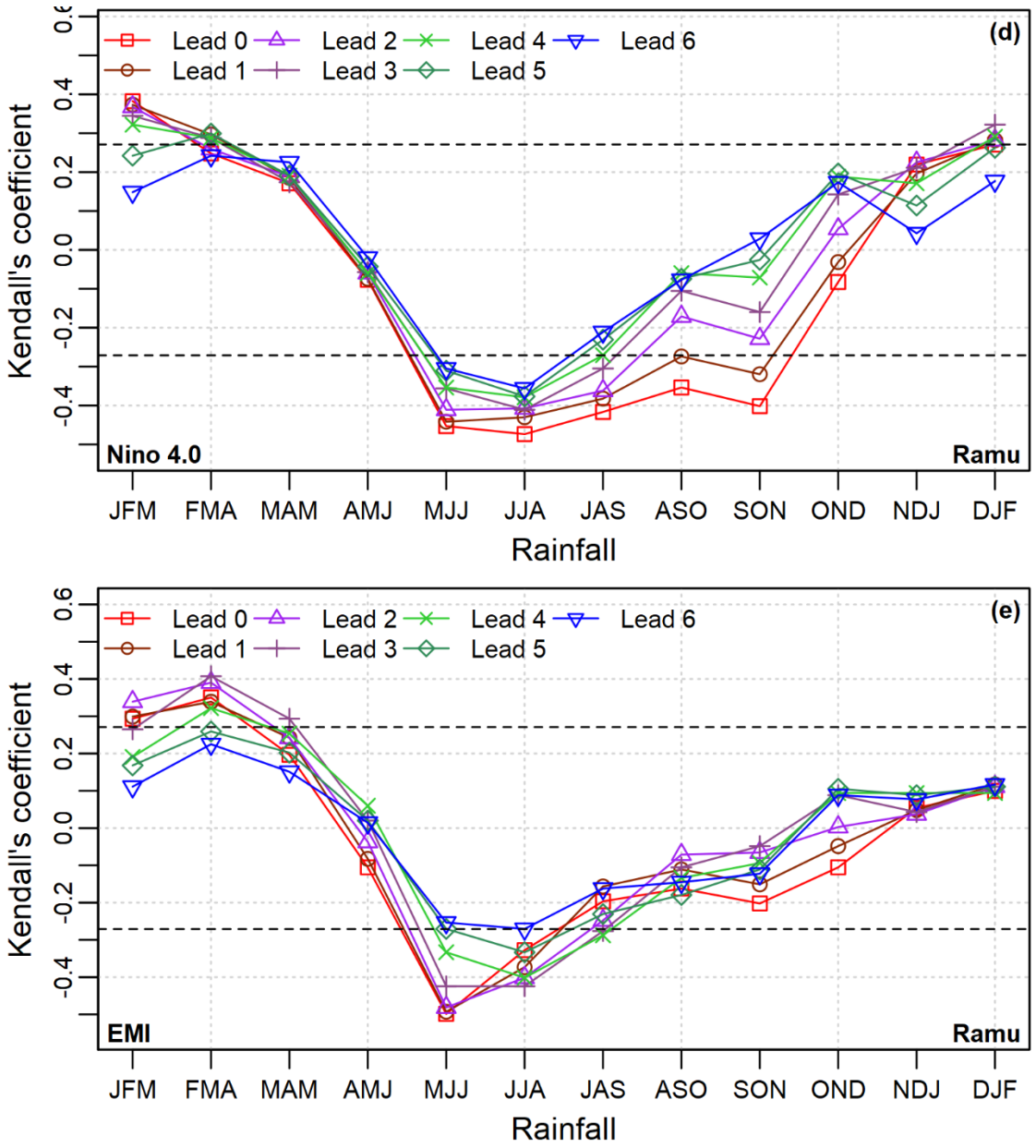


Figure 12: Kendall's Tau correlation between the climate mode indices and the rainfall for Ramu. Plot (a) represent SOI, Niño 3.0 (b) Niño 3.4 (c), Niño 4.0 (d) and EMI (e). (The horizontal dotted line represents significance at 5%).

4.3.3.1 Model evaluation

From the ten different bivariate copula-based rainfall models developed, it is evident that M9 (SON) had a higher significant r value followed by M6, M8, M5 and M7 with respective values of 0.59, 0.57, 0.56 and 0.54 as shown in **Table 9.**

The next evaluation metrics considered was the *RMSE* and the *MAE*. Although M9 had a higher r value its *RMSE* = 107.33 was not as good compared to M5, M6, M7 and M8 of which had a value of *RMSE* = 90.25, 74.39, 89.84 and 104.14 respectively as shown in **Figure 13**. The order changed when comparing metrics MAE, M5 performed better with 60.71, followed by M6 (61.75), M7 (73.69), M8 (84.22) and M9 (85.80). However, we cannot decide the best models based on *RMSE* and *MAE* evaluations. Instead, a combination of other metrics, including *RMSEs* and *MAEs*, are often essential to assess model performance.

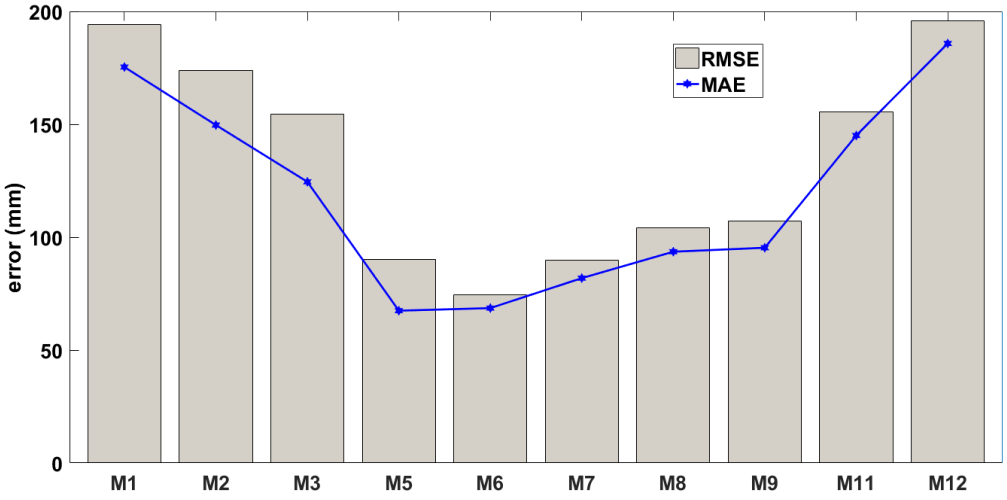


Figure 13: The root means square error *RMSE* (mm) and the mean absolute error *MAE* (mm) for the ten-rainfall forecasting model for Ramu station.

Additional assessment of model performance using the Nash-Sutcliffe Efficiency (E_{NS}), Willmott's Index (d), and the Legate-McCabe's Index (L), which ideally for perfect models requires unity, are performed to evaluate each of the ten copula-based bivariate rainfall models (**Table 9**). Based on the performance metrics mentioned, M9 model demonstrated a substantial improvement in the performance displaying the highest value of E_{NS} and d compared to all the other models for Ramu, with a value of 0.70 and 0.39 respectively. M9 however, had the second-best performance with L = 0.22 compared to M5 (L = 0.33). The next best performing model based on d is M6 (0.34) followed by M8 (0.32) with the lowest-performing model M3 (0.21). Other models which performed poorly based on *WI*, L and E_{NS} included model M1, M2,

M3, and M12 all have negative values indicating the forecasted values diverging, whereas values closer to 0 have no predictive advantage (Legates & McCabe 1999).

Table 9: The performance evaluation metrics for the ten different bivariate copula rainfall models for Ramu, based on Pearson's correlation coefficient (r), *RMSE* (root mean square error), *MAE* (mean absolute error), Nash–Sutcliffe Efficiency (E_{NS}) and Legates-McCabe's index (L). *RFP – The three-monthly total rainfall forecasting period, CI – three moths average Climate Index (Predictor), LTM – Lead time months (Predictor)*

Model Name	**RFP**	**CI**	**Lead (0-6)**	**LTM**	**r**	***RMSE* (mm)**	***MAE* (mm)**	E_{NS} **[-∞ 1]**	d **[0, 1]**	L **[-∞ 1]**
M1	JFM	Niño 4.0	0	OND	0.29	194.22	157.81	0.23	-0.08	-0.04
M2	FMA	EMI	2	SON	0.33	173.99	134.70	0.24	-0.10	0.03
M3	MAM	EMI	3	SON	0.21	154.37	112.08	0.16	0.01	0.01
M4	AMJ	Niño 3.0	4	SON						
M5	MJJ	EMI	0	FMA	0.56	90.25	60.71	0.53	0.30	0.33
M6	JJA	Niño 4.0	0	MAM	0.59	74.39	61.75	0.50	0.34	0.18
M7	JAS	Niño 3.0	1	MAM	0.54	89.84	73.69	0.54	0.26	0.17
M8	ASO	Niño 3.4	0	MJJ	0.57	104.14	84.22	0.59	0.32	0.18
M9	SON	SOI	0	JJA	0.64	107.33	85.80	0.70	0.39	0.22
M10	OND	Niño 4.0	4	MAM						
M11	NDJ	Niño 3.0	2	JJA	0.36	155.63	130.55	0.39	0.11	0.03
M12	DJF	SOI	0	SON	0.36	195.69	167.21	0.45	0.03	-0.03

Moreover, comparisons were made to evaluate the best rainfall models for Ramu using scatterplots of the observed and simulated values, as shown in **Figure 14.** From the four best models, the plots reaffirm M9 superior performance as the scatter points are clustered close to the y = x linear form with the maximum R^2 = 0.41 indicating an overall 41% of the observed rainfall could be forecasted using that bivariate model. The next best model was M6 (R^2 = 0.34) followed by M8 (R^2 = 0.33) with M5 as the least performing having R^2 = 0.31. An alternative model performance metric is the gradient (m) of the linear fit. For a suitable model the gradient should be close to unity, based on the models evaluated M9 was close to idyllic magnitudes further backing up the outcomes of the predictor metrics with a m = 0.41, which was again trailed M6.

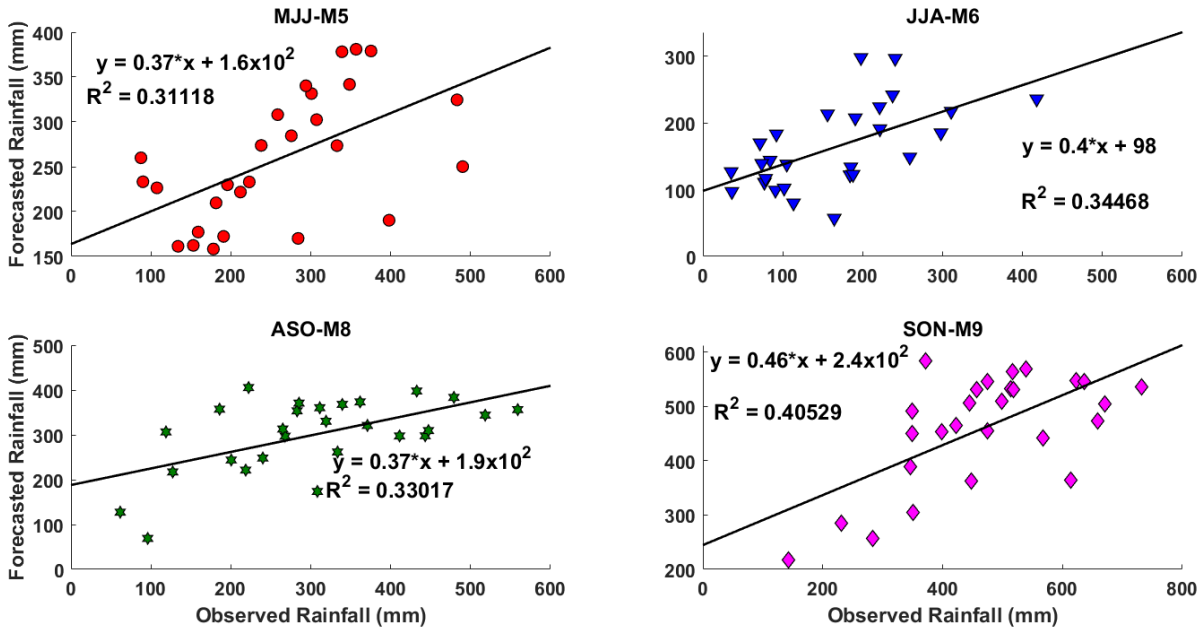


Figure 14: Comparisons of the four best performing copula-based bivariate rainfall models for Ramu rainfall station (PNG). In each panel, a least-squares regression line of the form $R_{Pred} = m\ R_{Obs} + C$ with a coefficient of determination R^2 included, to assess the level of agreement between the forecasted and observed values are included.

Furthermore, the bivariate copula-based rainfall forecast models were further evaluated using the frequency error plots as shown in **Figure 15** where the absolute value of error accumulated in separate error bins of ±100 of various error brackets was used to give a better understanding of the model accuracy for practical applications. For each error bin, the percentage count is shown at the top of each error bar. It is interesting to note that the rainfall predictions using M6 yielded the highest frequency of errors in the smallest error bracket ($0 < \text{error} \leq 100$) with 81% of all data points in this smallest error bracket, followed by M5 (74%) M5 (70%) and M8 (67%). Evidently, **Figure 15** indicates that M6 is the best model since a more significant portion of errors lies within smallest error bracket ($0 < \text{error} \leq 100$) with a smaller portion of the datum points redistributed in larger error bracket (precisely 19%). This signifies that M6 tends to generate more accurate predictions.

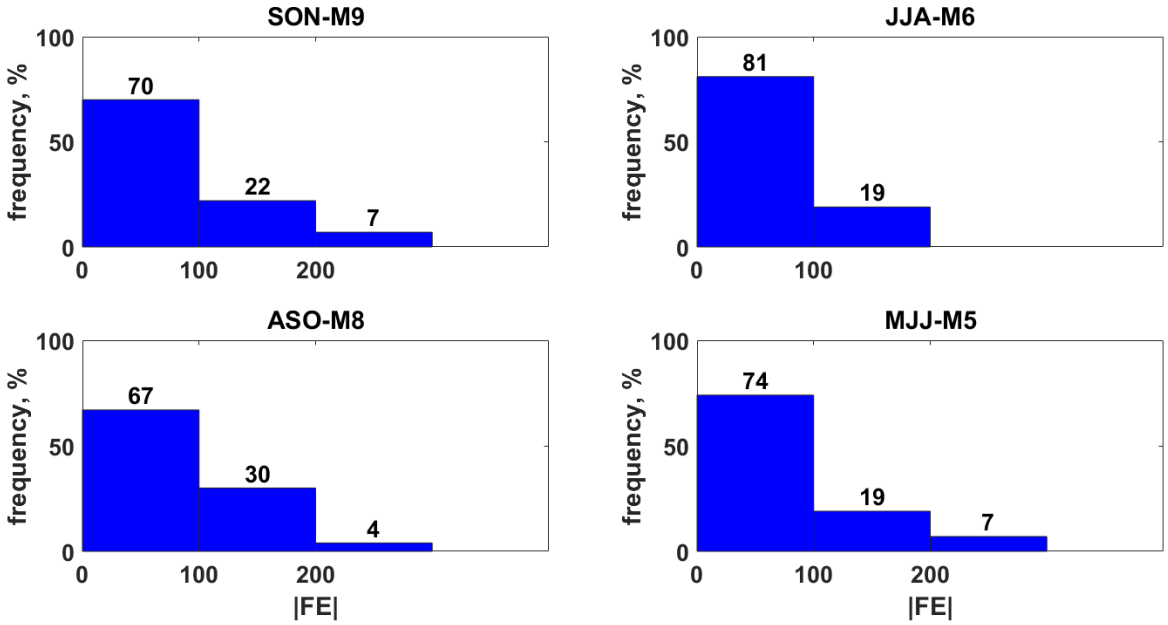


Figure 15: Frequency plot of absolute forecast errors (Freq %) generated by the bivariate copula statistical model for the best performing rainfall models for Ramu. The number of datum points in each error bin is shown

However, when compared with the scatterplot (**Figure 14**), the overall covariance between forecasted and observed data indicates that M9 has better performance, and this conflicts the results. However the major limitation of *r* is it standardizes means and variance of the observed and forecasted datum which was bounded by [−1, 1] and only measures the overall correspondence (Deo, Kisi, *et al.* 2017; DeoTiwari, *et al.* 2017).

Also, based on the E_{NS}, *d*, and *L* indicated the better model as M9 as discussed; however they are based on 'average' performance where all forecasted and observed data is used and individual error assessment are not made. E_{NS} assesses the differences between observed and forecasted data and calculates as average square values, and as a consequence, larger values in the data can be over-estimated as well as smaller values neglected (Legates & McCabe 1999). The Willmott index (*d*) (Willmott 1982) overcomes the insensitivity created by E_{NS} since it evaluates the precision of the forecasted values concerning observed values of the data by taking into account the ratio of mean squared error instead of the square of the error differences. Whereas L does not overestimate since the absolute values are considered and give appropriate weights to the errors and differences (Legates & McCabe 1999),

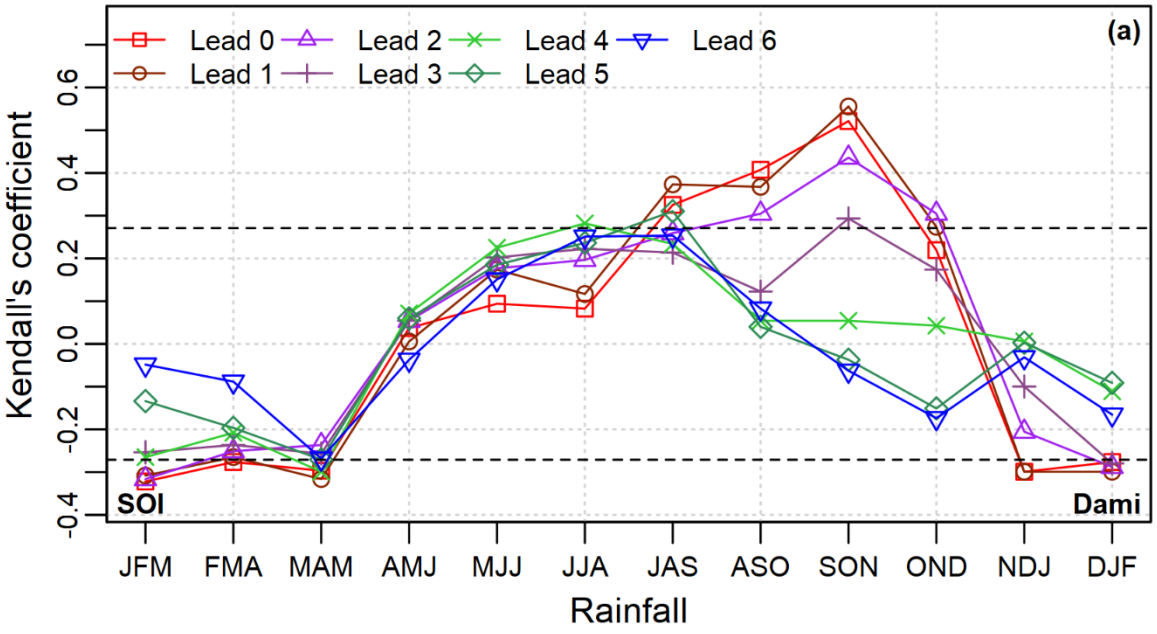
The frequency plots are more meaningful as it evaluates the individual error assessment of the models and therefore M6 is the best models since 81% of errors lies within smallest error

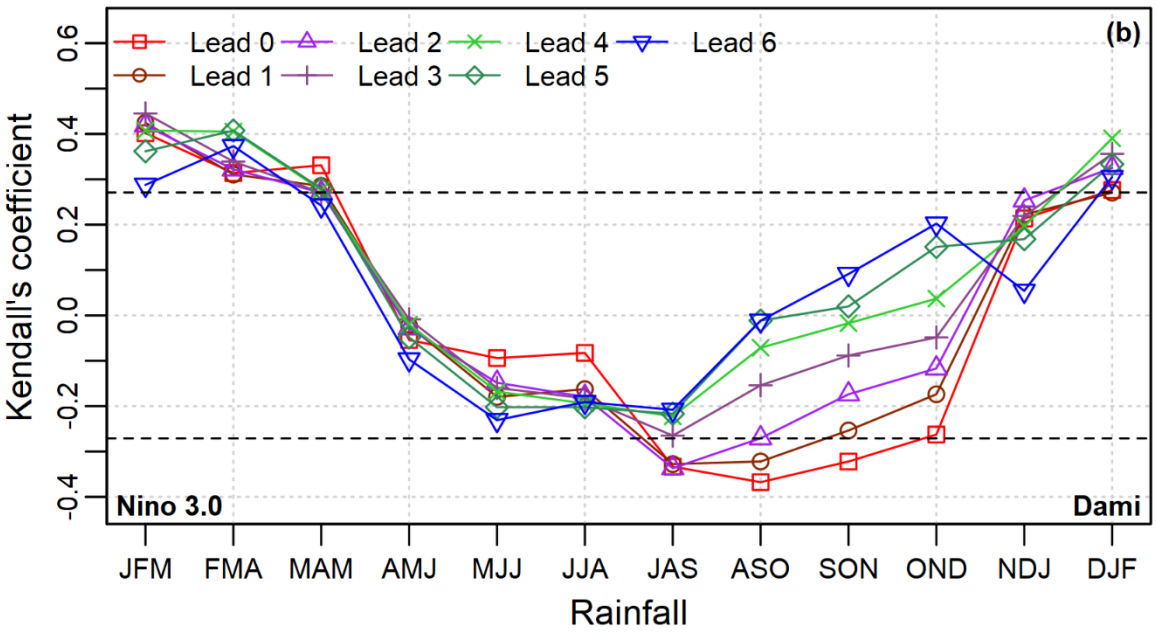
bracket ($0 < \text{error} \leq 100$) with the remaining 19% of the datum points redistributed in larger error bracket.

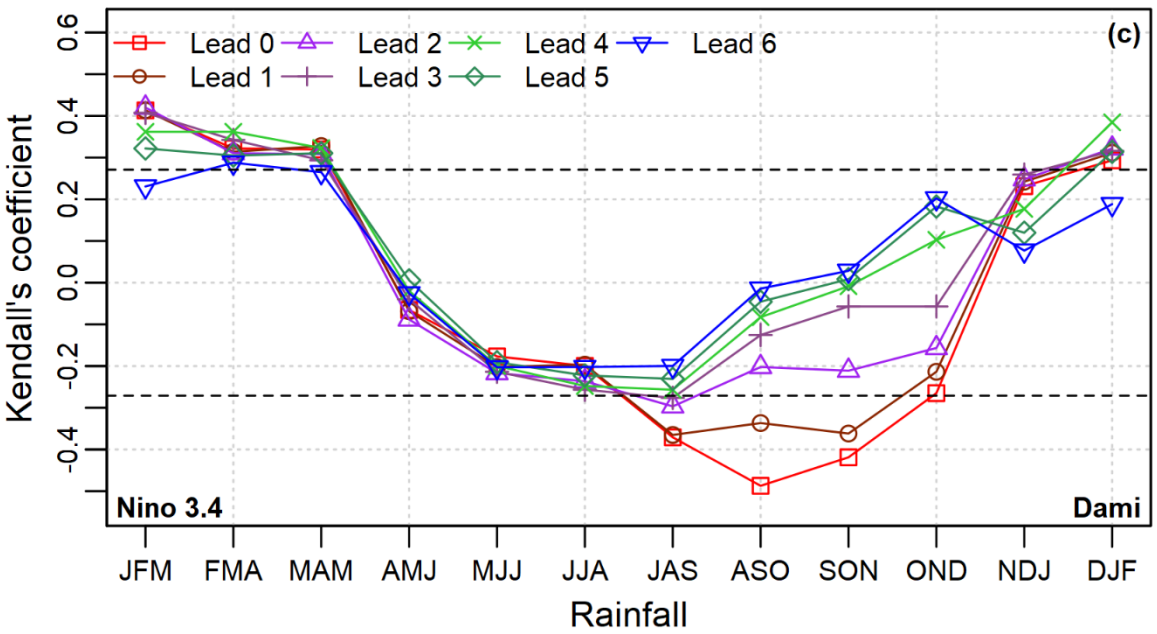
4.3.4 Station 4 – Dami

The fourth station used in this study is Dami research station, which is in West New Britain Province, PNG with an estimated terrain elevation above sea level of 30 m. As a prior step for constructing a model, correlations were done using Kendall tau correlations to assess how strongly the relationship of the three-monthly average climate index (predictor) and three-monthly total rainfall. These were done using the rank-based method since this approach has the advantage of capturing far-off observations hence in deciding the best associations with the predictor and the predictand (i.e., rainfall) for the model development using copula (Nguyen-Huy *et al.* 2018).

The correlations plots in **Figure 16** showed low correlations with the ENSO indices during the period AMJ with all values falling between the dotted line. The EMI had a better correlation with MJJ and JJA rainfall. The three months mean SOI lead-1 correlation with the SON total rainfall displayed the highest positive significant correlation (0.56) compared to all the other models. The SOI also displayed good correlations (0.30) with the OND rainfall while the remaining three-monthly rainfall periods correlated well with the SST's as shown in **Table 6** and **Figure 16**. The EMI negative correlation values associated with Dami's rainfall occurs around June – August which is similar to negative EMI-rainfall correlations observed in Australia (Taschetto & England 2009; Yuan, C. & Yamagata, T. 2015). On the other hand, again and as similar to Ramu station the positive correlation recorded between the SON – JFM rainfall on ENSO is consistent with the period in which ENSO's is at peak intensity (December-April) and weakens during around May-August which supports the negative correlation values.


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(a)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
SOI
Dami
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(b)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
Nino 3.0
Dami
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
(c)
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
Nino 3.4
Dami
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall

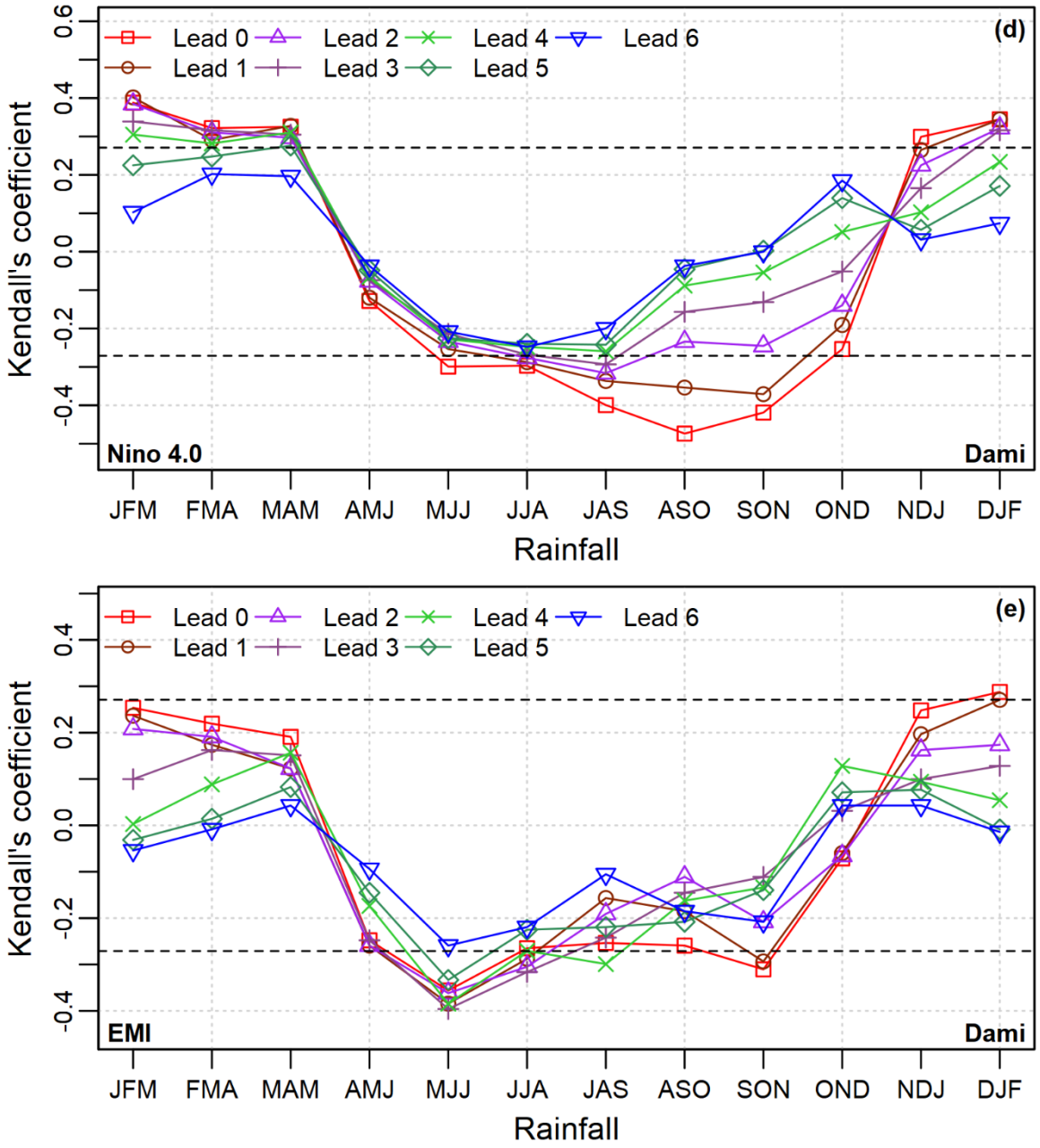


Figure 16: Kendall's Tau correlation between the climate mode indices and the rainfall for Dami station. Plot (a) represent SOI, Niño 3.0 (b) Niño 3.4 (c), Niño 4.0 (d) and EMI (e). (The horizontal dotted line represents significance at 5%).

4.3.4.1 Model evaluation

The evaluation of each model's performance was carried out after forecasting the rainfall values, which was generated through the copula-based method. Each model performance was assessed to validate their performance using the evaluation metrics, as shown in **Table 10**.

Model M4 (MJJ) as discussed in the previous section, was excluded since it did not have a statistically significant relationship with rainfall and the climate mode index. Based on the Pearson's correlation coefficient (**Table 10**), M8 had the maximum r-value of 0.72 and is followed by M9 (0.66), M1 (0.48), M5 (0.40), M2 (0.39), M12 (0.31), M6 (0.30), M3 (0.26), M7 (0.24), M11 (0.12) and finally M10 (0.05). The *r* metric quantifies the strength of linear associations between the observed and forecasted rainfall values, and as observed from the previous evaluation (Ramu and Port Moresby), although models with mediocre or poor models some could achieve higher correlations.

The next evaluation metrics considered was the absolute error measures *RMSE* and the *MAE*. The lower the absolute error measure, the better the model (ideal values = 0). The smallest *RMSE* and *MAE* values were again recorded by M8, and the next best was M9 which so far based on the conventional metrics (*r*, *RMSE*, and *MAE*) both consistently displayed superior performance. Intriguingly, M1 and M2 performed poorly compared to M10 although both had higher *r* values.

In addition to the conventional metrics (*r*, *RMSE*, and *MAE*) additional evaluation of model performance using the goodness-of-fit metrics E_{NS}, *d*, and *L*, which ideally for perfect models requires unity. The value of *d* was also high again high for M8 with 0.77 followed by M9 (0.68), M1 (0.46) and M12 (0.31), this order of the top-performing model is also observed with the E_{NS} and *L*. Furthermore, models which performed poorly based on *d*, *L* and E_{NS} included model M7 and M11, have negative values indicating the forecasted values diverging, whereas values closer to 0 (M6, M10, and M3) have no predictive advantage (Legates & McCabe 1999). Based on the evaluations discussed, the results from **Table 10**, reaffirms M8 provided the best results for rainfall forecasting for the ASO rainfall scheme followed by the M9 (SON).

Table 10: The performance evaluation metrics for the bivariate copula rainfall models for Dami, based *r*, *RMSE*, *MAE*, E_{NS}, *d*, and *L*. *RFP – The three-monthly total rainfall forecasting period, CI – three moths average Climate Index (Predictor), LTM – Lead time months (Predictor)*

Model Name	RFP	CI	Lead (0-6)	LTM	r	*RMSE* (mm)	*MAE* (mm)	E_{NS} [-∞, 1]	*d* [0, 1]	*L* [-∞, 1]
M1	JFM	Niño 3.0	3	JAS	0.48	319.76	263.64	0.21	0.46	0.07
M2	FMA	Niño 3.0	6	MJJ	0.39	320.27	278.79	0.14	0.34	0.06
M3	MAM	Niño 4.0	0	DJF	0.26	273.00	216.86	0.03	0.19	0.10
M4	AMJ	EMI	2	NDJ						
M5	MJJ	EMI	3	NDJ	0.40	149.89	116.08	0.15	0.36	0.12
M6	JJA	Niño 4.0	0	MAM	0.30	150.40	119.83	0.05	0.26	0.01
M7	JAS	Niño 4.0	0	AMJ	0.24	161.49	111.12	-0.01	0.28	0.03
M8	ASO	Niño 3.4	0	MJJ	0.72	132.01	98.67	0.51	0.77	0.34
M9	SON	SOI	1	MJJ	0.66	136.91	104.42	0.42	0.68	0.27
M10	OND	SOI	2	MJJ	0.05	289.55	217.17	-0.45	-0.16	-0.22
M11	NDJ	Niño4.0	0	ASO	0.12	337.48	248.11	-0.05	0.21	0.05
M12	DJF	Niño3.0	4	MJJ	0.31	313.80	238.42	0.05	0.31	0.11

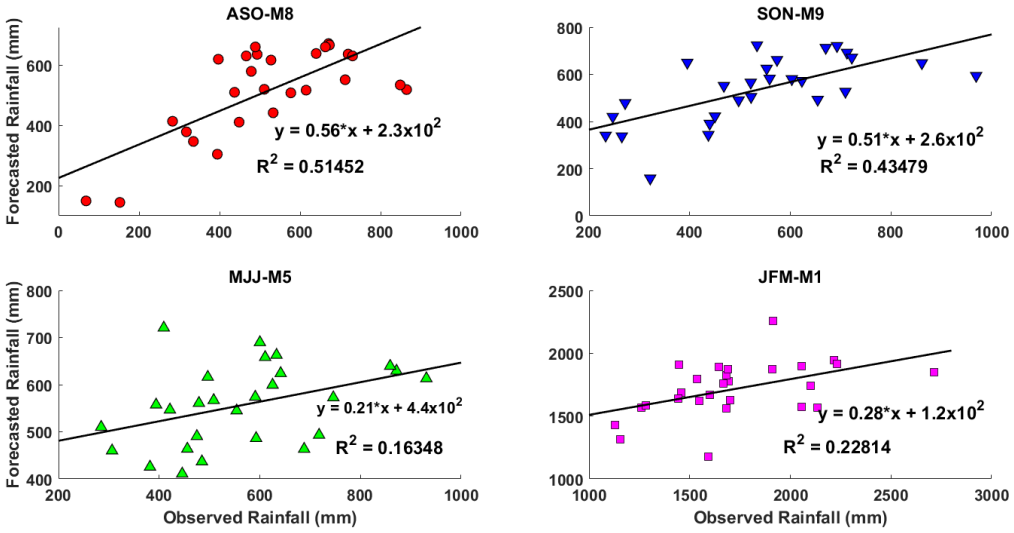


Figure 17: Scatter plots of the four best bivariate copula rainfall models for Dami rainfall station (PNG). ASO-M8 (top-left), SON-M9 (top-right), MJJ-M5 (bottom-left) and JFM-M1 (bottom-right). In each panel, a least-squares regression line of the form $R_{Pred} = m\ R_{Obs} + C$ with a coefficient of determination (R^2) included, to assess the level of agreement between the forecasted and observed values are included.

Moreover, comparisons were further made to evaluate the best rainfall models for Dami using scatterplots of the observed and forecasted values, as shown in **Figure 17.** From the four best models selected based on model performance, the plots reaffirm M8 superior performance, with the maximum R^2 = 0.51 indicating an overall 51% of the observed rainfall could be forecasted using that bivariate copula-based model. The next best model was M9 (R^2 = 0.43) followed by M1 (R^2 = 0.22) with a lower performance displayed by M5 having R^2 = 0.16. An alternative model performance metric is the gradient (m) of the linear fit and based on the models evaluated M9 was close to idyllic magnitudes (unity) further backing up the outcomes of the predictor metrics with an m = 0.56, with, both showing consistent results across all performance evaluations so far.

Furthermore, model accuracy was evaluated using the frequency percentage error plot by examining the magnitude of errors for each model with different error ranges of ± 100. The results in **Figure 18** reaffirmed again M8 superior performance since 70% of errors lies within the smallest error bracket (0 < error ≤ 100). The next best model was again M9 with 63% of errors in the smallest magnitude (0 < error ≤ 100) error bracket. This signifies that M8 tends to generate more accurate predictions since a more in-depth analysis of actual errors is explained by frequency plot and is consistent with the results discussed earlier.

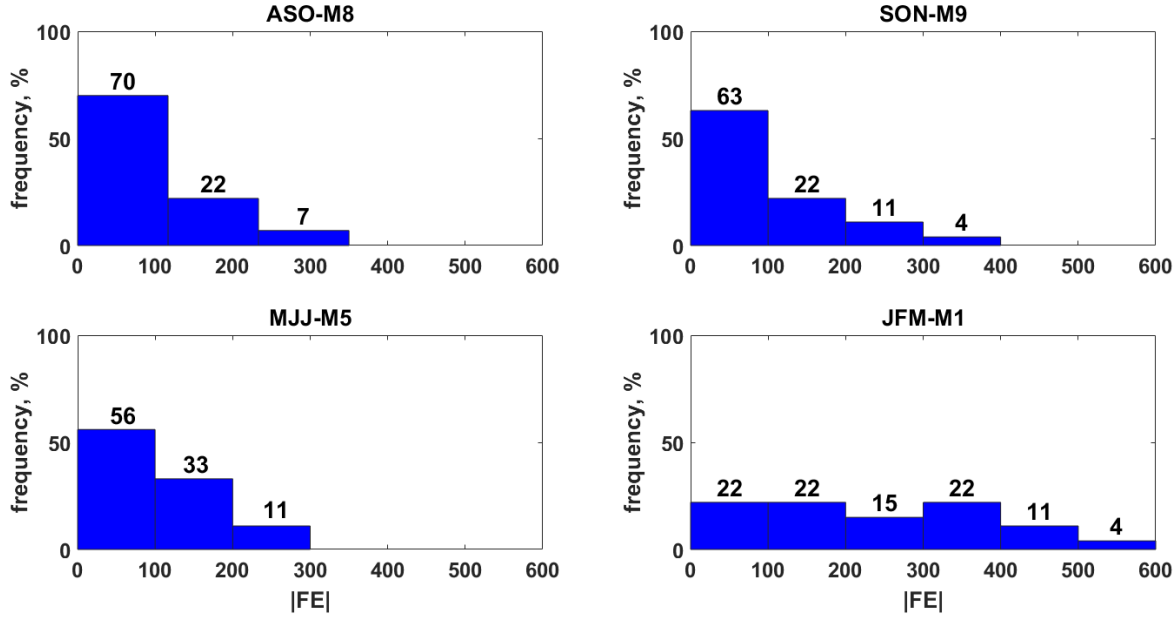


Figure 18: Frequency plot of absolute forecasting error |FE| created by the bivariate copula models M8, M9, M5, and M12 for Dami rainfall. The number of datum points in each error bin is shown on top of each histogram.

4.4 Comparisons of All sites

Finally, a geographical comparison of the best performing model for each rainfall sites was compared using relative root mean square error (*RRMSE*) and the mean absolute percentage error (*MAPE*). According to (Mohammadi, K. *et al.* 2015) the different ranges of *RRMSE* is defined as follows; excellent if % *RRMSE* < 10%; good if 10% < % *RRMSE* < 20%; fair if 20% < % *RRMSE* < 30%; and poor if % *RRMSE* ≥ 30%.

Corresponding to the results presented in **Table 11** and **Figure 19**, the *RRMSE* and *MAPE* values ranged between 21.54 - 87.48% and 16.88 – 141.42% respectively. Based on the *RRMSE* ranges, M1, M2, M11, and M12 models were found between the 20% < *RRMSE* < 30% range indicating the fair model precision while the remaining displayed poor performance. The performance of M10, although performed well with the performance evaluation metrics (*r*, *RMSE*, *MAE*, E_{NS}, *d*, and *L*) performed poorly based on relative errors. The M11, M12, and M1, however, displayed consistent results for Port Moresby and are preferred over the other models.

For Aiyura, the lowest relative percentage errors were generated by M1 (*RRMSE* = 15.39% and *MAPE* = 13.23%) with the lowest followed by M12 (*RRMSE* = 22.89% and *MAPE* = 13%) as seen in **Figure 19** and both models displayed good (10% < *RRMSE* < 20%) model precision. The next category (20% < *RRMSE* < 30%), only M2 fitted that category (RRMSE = 22.89% and *MAPE* = 20.27%) and can be described as a fair model. The next station Ramu, the relative percentage errors for M1, M2, M3, M9, M11, and M12 can all be categorized as fair models (20% < *RRMSE* < 30%) (Mohammadi, K. *et al.* 2015) with the M9 again displaying consistent results throughout the model evaluation and is preferred for Ramu. While for Dami, the lowest percentage of errors was observed in M1. However, the forecasted errors generated in **Figure 18** for the smallest magnitude (0 < error ≤ 100) error bracket is 22% which is poor model performance. M8, M9, and M5 as discussed in the previous section had 70%, 63%, and 56% (respectively) of their errors in the smallest magnitude (0 < error ≤ 100) error bracket and based on the *RRMSE* and *MAPE*, they registered their relative error between the (20% < *RRMSE* < 30%) and are preferred models for Ramu with high esteem in M8 model.

Table 11: Performance evaluation of the study sites' using the relative measures, *RRMSE*, and *MAPE* for each site. The excluded periods (shown in green) did not have a statistically significant relationship with the wheat yield, whereas M8 and M9 for Port Moresby generated errors (negative forecasted rainfall values) during model development.

	Port Moresby		Aiyura		Ramu		Dami	
Model Name	**RRMSE, %**	**RMAE, %**	**RRMSE, %**	**RMAE, %**	**RRMSE, %**	**RMAE, %**	**RRMSE, %**	**RMAE, %**
M1	21.54	20.00	15.39	13.23	25.42	21.82	18.54	15.22
M2	26.40	25.55	22.89	20.27	23.08	19.86	21.80	20.64
M3	39.04	34.68			23.37	17.92	25.04	22.55
M4	47.82	37.51						
M5	73.69	94.37	31.19	29.95	34.84	31.33	26.62	22.03
M6	87.48	129.01	35.05	34.66	45.33	53.30	27.70	27.87
M7	61.82	141.42	42.32	41.72	45.98	67.21	32.15	35.32
M8	61.01	102.61	47.83	50.70	34.26	35.09	25.37	22.44
M9	74.06	94.77	32.58	29.02	22.94	19.28	25.48	22.90
M10	41.75	48.65					40.48	33.79
M11	22.25	16.88			21.91	19.25	30.67	22.57
M12	24.77	21.03	16.75	15.10	26.33	25.29	21.20	17.03

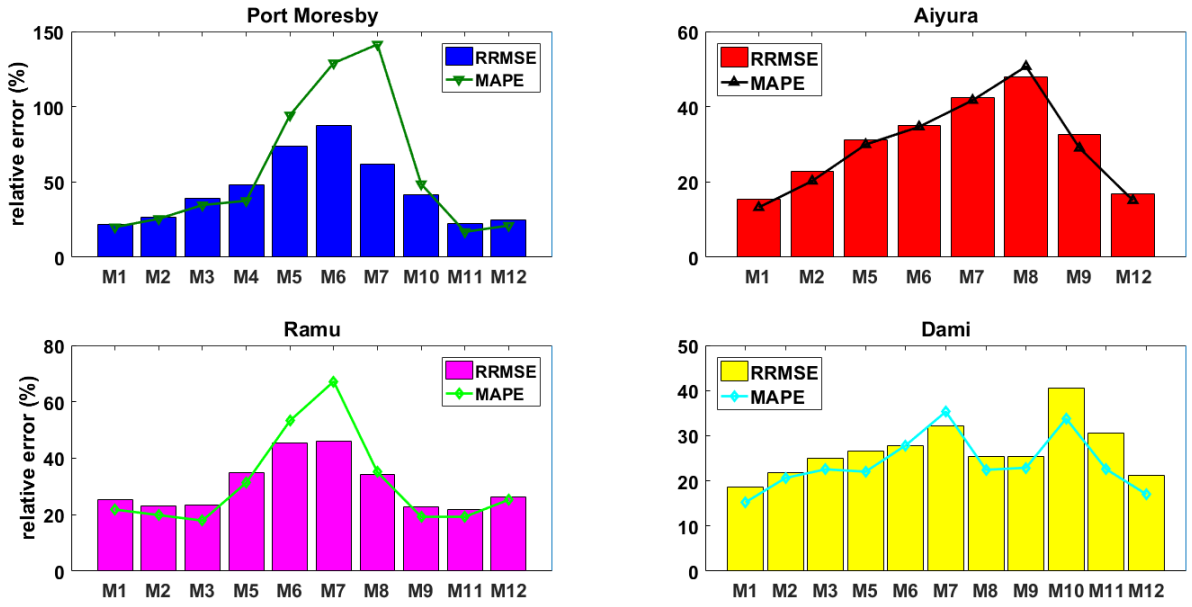


Figure 19: The mean absolute percentage error (*MAPE*) and relative root mean square error (*RMSE*) for the different bivariate copula-based rainfall models for the four study sites.

4.5 Concluding Remarks

Correlations performed between the best lead times for the climate mode indices and the three-month total rainfall, which forms a basis for bivariate models applied to capture ENSO impact are summarised in **Table 6**. Based on the copula-based bivariate rainfall models developed and discussed so far in this chapter, several models for each station can be used to forecast rainfall for different three-monthly schemes. For Aiyura, M1 and M12 can be used to forecast the total rainfall for the JFM and DJF. For Ramu, M9 which displayed consistent results throughout the model evaluation and is preferred to forecast rainfall for June-July-August, while Dami station, M8, and M9 are equally exceptional models that can be used to forecast the total ASO and SON rainfall. The final rainfall site Port Moresby, M11 and M1 can be used to forecast the cumulative rainfall for NDJ and JFM.

CHAPTER 5 TRIVARIATE MODELS

5.1 Introduction

This chapter focuses on developing and evaluating the trivariate copula forecasting techniques using rainfall as the response variable and the two covariates canonical (ENSO)/ENSO Modoki and the IOD (Dipole Mode Index (DMI)).

Based on the selections done with the bivariate models, DMI was added to the copula models to evaluate whether the performance of the model is improved or not.

This chapter also discusses the results of the model performance evaluated via statistical means using the Pearson's correlation coefficient *r*, Willmott's Index (*d*), Nash–Sutcliffe Efficiency (E_{NS}), Legates-McCabe's Index (*L*), root-mean-square-error (*RMSE*), and mean absolute error (*MAE*), plus the relative root mean square error (*RRMSE*) and mean absolute percentage error (*MAPE*).

5.2 Methods

5.2.1 Data

The rainfall data (1990-2017) was arranged in three-monthly totals, JFM, FMA, MAM, AMJ, MJJ, JJA, JAS, ASO, OND, NDJ and DJF for the four study sites. The Dipole Mode Index (DMI) which represent the Indian Ocean Dipole (IOD) (details in **Table 5**), was also arranged in three monthly averages. Here we define the different lead times as lead-0, lead-1, lead-2 to lead-6. For example, in a 'lead-0', the February-April (FMA) DMI is used to forecast the May-June (MJJ) rainfall. For a 'lead-1', February-April (FMA) DMI is compared with JJA rainfall, 'lead-2', January-March (JFM) DMI is compared to JJA rainfall. Comparisons are made up to "lead 6'.

5.2.2 Model Development

The D-vine regression models were constructed using the **vinereg** package, which performs the sequential estimation of a regression D-vine for quantile prediction as described in Kraus and Czado (2017). The common form of the trivariate D-vine regression models is $Y{-}X_1{-}X_2$, where Y denotes the three-month total rainfall, X_1 is one of the ENSO indicators used to forecast Y and defined in Chapter 4 (**Figure 5**), and X_2 refers to the corresponding IOD indicator (i.e. DMI). By the definition of vine copulas, these variables are called the nodes, and any two nodes are connected by a unique path "─", which is associated with a bivariate pair-

copula. Then, the one-fold cross-validation method was used to assess the out-of-sample performance of forecasting models at given quantile levels by applying the cross-validation (Nguyen-Huy *et al.* 2018) to the observed and forecasted values.

The observed and forecasted values were then assessed using the performance matrices *d*, E_{NS}, *L, RMSE, MAE, RRMSE,* and *MAPE* to decide on the best model.

5.3 Results and Discussion

5.3.1 Station 1- Port Moresby (POM)

The inclusion of IOD (DMI) index to the bivariate models discussed in the preceding chapter for Port Moresby rainfall is evaluated using the evaluation metrics, as shown in **Table 12**.

Table 12 shows M11 had the highest correlation between the observed and forecasted rainfall values with $r = 0.76$ followed by M10 $r = 0.64$, while M3 had the lowest value $r = 0.11$. When compared with the bivariate result in **Table 7** the trivariate copula-based model improved the *r* performance of M4, M5, and M6. Further evaluations assessment using the absolute error measures *RMSE* (mm) and *MAE* (mm) which are wieldy used in evaluating forecasting model accuracy were also used to evaluate the trivariate copula models as depicted in **Figure 20**.

The range of the *RMSE* was in the array of 48.14 to 186.33, of which M7 had the lowest value and M3 the highest. Similarly, the *MAE* was lowest for M7 followed by M6 and M10, while the highest was M12. Based on the *RMSE* and *MAE* the values must be as smaller to indicate small deviations of the forecasted values from the observations, and this was displayed by M7 which performed better followed by M6, M10, and M11 in ascending order. When compared to the bivariate results (**Table 7**) for *RMSE* and *MAE* the trivariate results performed lesser.

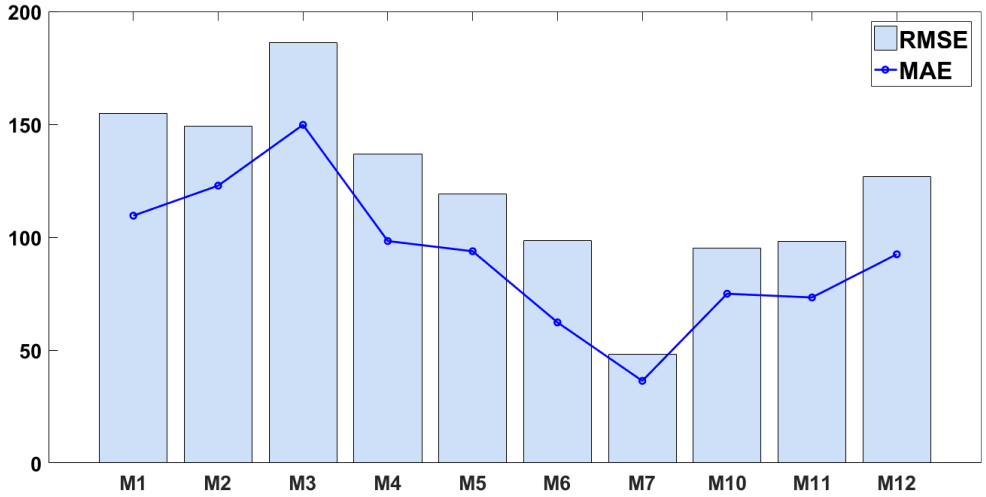


Figure 20: The *RMSE* (mm) and *MAE* (mm) for the copula based trivariate rainfall models for Port Moresby rainfall station.

Table 12: The performance evaluation metrics for the trivariate copula rainfall models for Port Moresby (POM), based on correlation coefficient (*r*), *RMSE, MAE, E_{NS}, d* and *L*. The models M8 and M9 forecasted negative values in the year 1997 (drought), which as expected the ASO and SON observed rainfall was 7.8 mm for both (SON and ASO) and the corresponding rainfall values for the forecasted was -17.94 and -16.85 respectively. *RFP – The three-monthly total rainfall forecasting period, CI –Climate Index (Predictor), LTM – Lead time months (ENSO and IOD).

Model Name	**RFP**	**Lead (0-6)**	**LTM**	***r***	***RMSE* (mm)**	***MAE* (mm)**	***E_{NS}* [-∞, 1]**	***d* [0, 1]**	***L* [-∞, 1]**
M1	JFM	6	AMJ	0.50	154.80	109.54	0.13	0.47	0.21
M2	FMA	6	MJJ	0.46	149.14	122.84	0.18	0.38	0.06
M3	MAM	0	DJF	0.11	186.33	149.76	-0.15	0.15	-0.10
M4	AMJ	0	JFM	0.29	136.71	98.30	-0.02	0.32	0.05
M5	MJJ	0	FMA	0.18	119.07	93.78	-0.03	0.17	-0.02
M6	JJA	6	SON	0.28	98.40	62.29	0.00	0.39	0.13
M7	JAS	0	AMJ	0.26	48.14	36.38	-0.01	0.15	0.10
M8	ASO	0	MJJ	0.40	53.03	40.68	0.12	0.43	0.11
M9	SON	1	MJJ	0.49	93.00	60.04	0.21	0.51	0.19
M10	OND	0	JAS	0.64	95.15	74.96	0.39	0.65	0.25
M11	NDJ	1	JAS	0.76	98.03	73.32	0.55	0.80	0.41
M12	DJF	1	ASO	0.44	126.75	92.45	0.16	0.38	0.14

Additional assessment of the trivariate copula rainfall models was also done using the goodness-of-fit assessment E_{NS}, *d* and *L* index to further evaluate the models' performance. In order to attain the best model, the values of E_{NS}, and *d* should be close to unity for the best fit (DeoTiwari, *et al.* 2017). Based on the results depicted in **Table 12**, M11 had the uppermost E_{NS} value with 0.55 followed by M10 with 0.39; this was consistent with the d wherein M11 had the highest value of 0.80 followed by M10 (0.65). Further assessment using *L* was evaluated and again, M11 had the uppermost value of 0.41 followed by M10. Given the models' performance evaluation it is clear that M11 and M10 performed exceptionally well compared with the other models outperforming them in r, E_{NS}, *d* and *L*. The evaluation metric evaluated

so far is consistent with the bivariate results indicating M11 and M10 as the two best models for Port Moresby.

Furthermore, the trivariate rainfall models were further appraised using the relative frequency of absolute forecast error to determine the best models as shown in **Figure 11** were each error bins ± 100, the percentage count is displayed at the top of each bar. It is interesting to note that M10 yielded the third highest frequency of errors in the smallest error bracket ($0 < \text{error} \leq 100$) recording 81% (**Figure 8**) in the bivariate model, however with the trivariate model it recorded 78%, which was the highest % error score. Also, M11 models had the highest with 89% of errors in the smallest error bracket ($0 < \text{error} \leq 100$) in the bivariate model, a decrease of 22% (i.e., from 67%) was recorded in the trivariate. The M12 model displayed no changes to its score for both models with 70% in the $0 < \text{error} \leq 100$), 22% in the $100 \leq \text{error} \leq 200$ and 7% in the $200 \leq \text{error} \leq 300$ error brackets. From the results discussed so far (**Figure 8** and **Figure 11**), the inclusion of DMI clearly improved the forecasting error of M1, M6, M11 and M10 in the first error bracket ($0 < \text{error} \leq 100$) $100 \leq \text{error} \leq 200$ brackets.

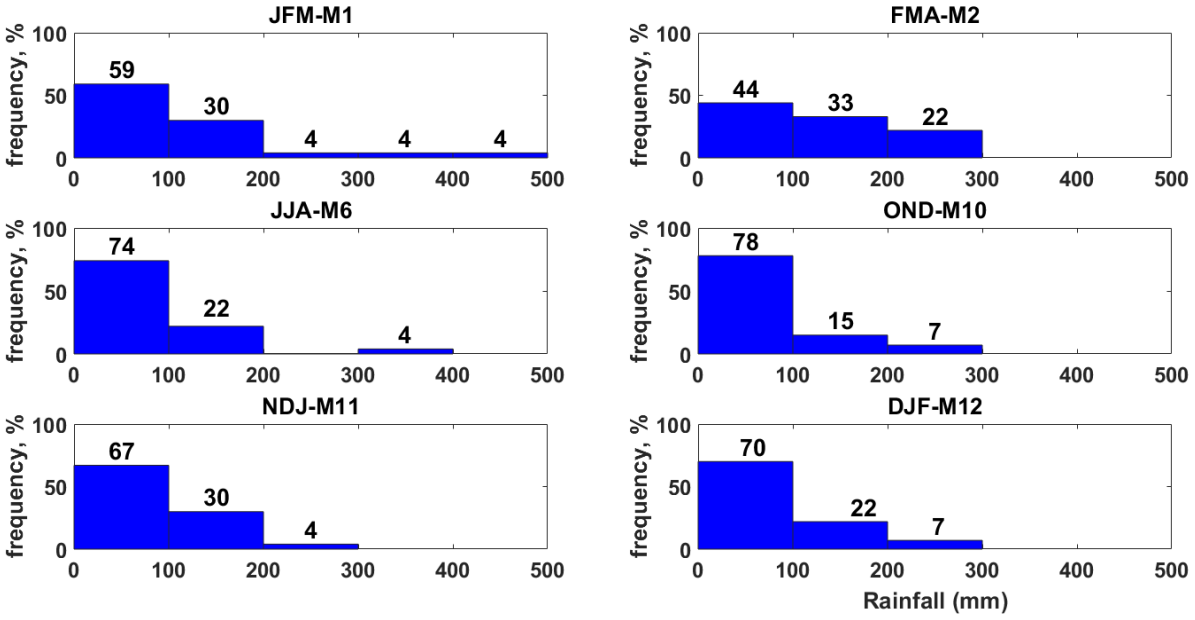


Figure 21: A histogram of the relative frequency of absolute forecasting error |FE| (mm) for copula-based trivariate rainfall models for Port Moresby. The number of datum points in each error bin is shown on top of each histogram

5.3.2 Station 2- Aiyura

The performance evaluation of the copula-based quantile regression models in forecasting rainfall for Aiyura was statistically evaluated using the *r*, *RMSE*, *MAE*, E_{NS}, *d* and *L*. In terms

of the conventional metrics *r*, *RMSE* and *MAE* as shown in **Table 13**, the absolute error measures *RMSE* (mm) and *MAE* (mm) for M1 was the lowest with 99.64 mm and 79.33 mm respectively, the second-best result was displayed by M5 followed by M2, while M9 performed poorly. Further evaluations using the correlation measures to determine the strength and direction between the observed and forecasted values of each copula-based quantile regression models. M1 again displayed the highest linear relationship with $r = 0.68$ followed by M2 ($r = 0.62$) and M5 ($r = 0.45$).

Furthermore, models performance using the goodness-of-fit assessment E_{NS}, d and L index further supported the performance of model M1 with the highest E_{NS}, d, and L with 0.45, 0.69 and 0.30 scores while the next best model was M2 with 0.37, 0.64, and 0.20 respectively. This performance confirmed that M1 and M2 provide an ideal combination of evaluations with the lowest errors (*MAE/RMSE*) including high performance of *d* and *L* which reaffirms both models performance. The next best results after M1 and M2 was displayed by M5. According to Legates and McCabe (1999) to achieve a perfect model E_{NS} should be equal to 1 and, for this study M1 displayed the highest value (0.45). However, with the d of Agreement as explained by Willmott et al. (2012), values approaching +1 represents a better model performance, the negative values displayed by M6, M8, M9, and M12 only indicates that the model is worse in performance (Willmott *et al.* 2012).

Moreover, comparisons were further made to evaluate the best rainfall models for Aiyura using scatterplots of the observed and forecasted values, as shown in **Figure 22.** From the three models discussed so far, the plots reaffirm M1 superior performance, with the maximum correlation coefficient value of $R^2 = 0.46$ indicating an overall 46% of the observed rainfall could be forecasted using that trivariate copula-based model, an increase of 7% when compared to the bivariate model (**Figure 10**). The next best model M2 recorded $R^2 = 0.39$ followed by M5 ($R^2 = 0.20$) with a lower performance displayed by M9. An alternative model performance metric is the gradient (m) of the linear fit and again M1 was close to idyllic magnitudes (unity) further backing up the outcomes of the predictor metrics with an m = 0.46, showing consistent results across all performance evaluations so far.

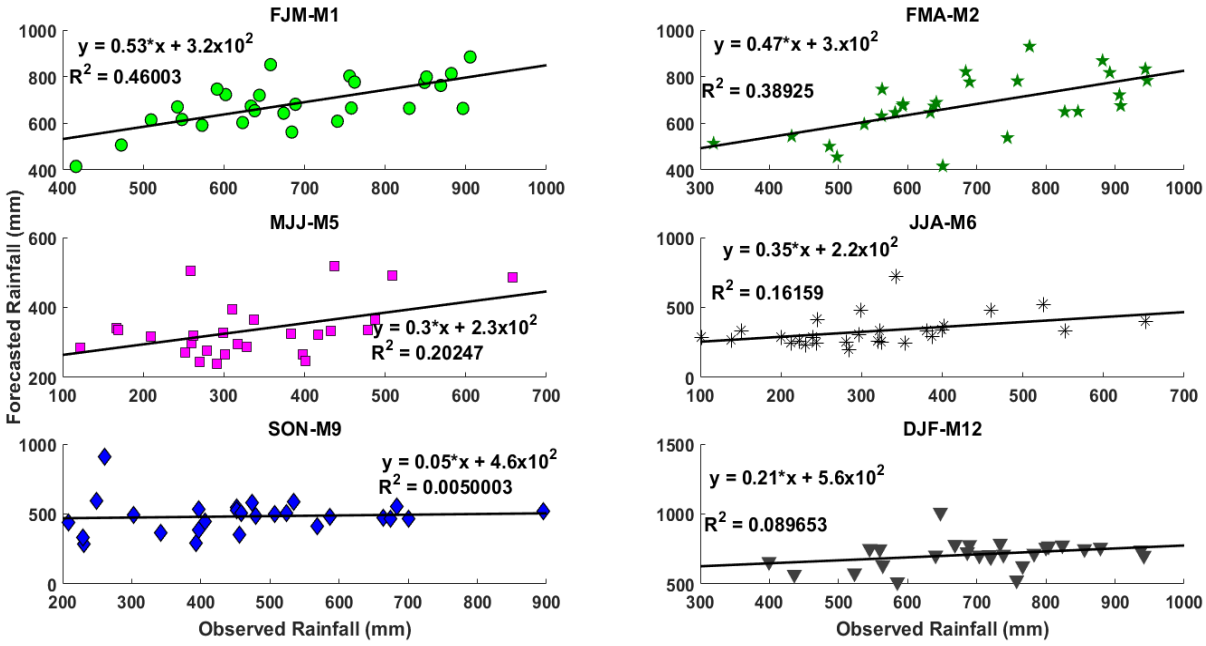


Figure 22: Scatter plots of the trivariate copula-based quantile regression models for the observed and forecasted rainfall for Aiyura station (PNG). (Note: The straight blue line is the least square fit line for each scatter plots of the form $R_{Pred} = m\ R_{Obs} + C$).

Table 13: The performance evaluation metrics for the three different D-vine copula-based quantile regression rainfall models for Aiyura, based on *r*, *RMSE*, *MAE*, E_{NS}, *d* and *L*. *RFP – The three-monthly total rainfall forecasting period, CI – three moths average Climate Index (Predictor), LTM – Lead time months (ENSO and IOD) (Predictor)*

Model Name	RFP	Lead (0-6)	LTM	r	RMSE (mm)	MAE (mm)	E_{NS} [-∞, 1]	*d* [0, 1]	*L* [-∞, 1]
M1	JFM	1	SON	0.68	99.64	79.33	0.45	0.69	0.30
M2	FMA	4	JAS	0.62	131.63	111.18	0.37	0.64	0.20
M5	MJJ	0	FMA	0.45	107.56	87.60	0.15	0.23	0.06
M6	JJA	0	MAM	0.40	130.01	93.92	-0.08	0.27	0.02
M7	JAS	0	AMJ	0.43	135.87	107.05	0.14	0.40	0.08
M8	ASO	0	MJJ	0.30	194.16	132.52	-0.30	0.17	-0.02
M9	SON	0	JJA	0.07	196.51	140.37	-0.42	-0.08	-0.10
M12	DJF	0	SON	0.30	141.66	110.77	-0.08	0.32	-0.03

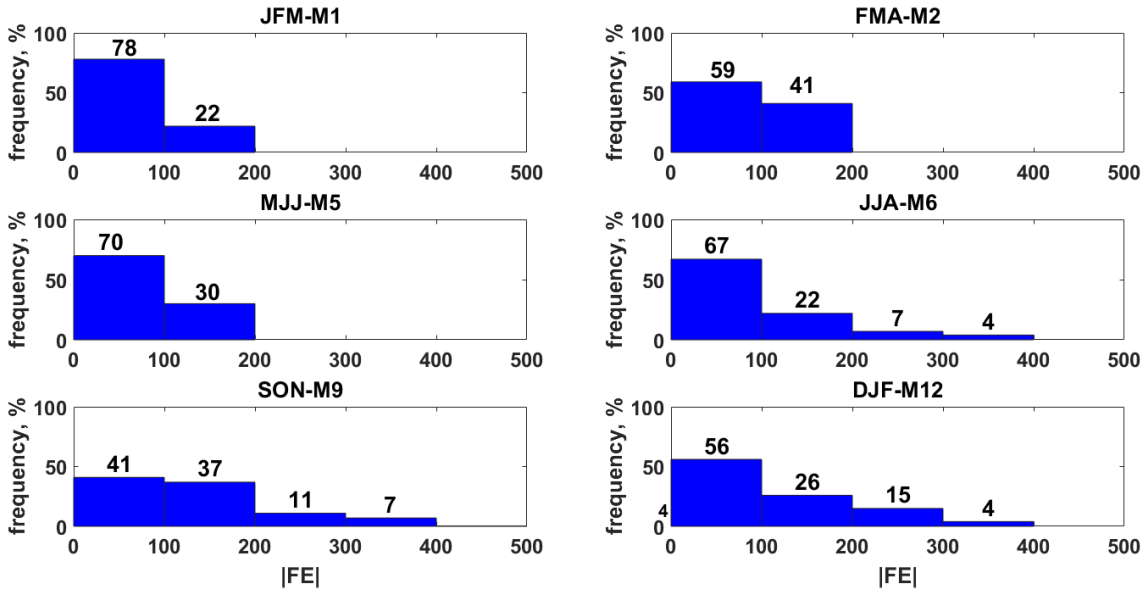


Figure 23: A histogram of the relative frequency of absolute forecasting error |FE| (mm) for the three copula-based trivariate rainfall models for Aiyura. The number of datum points in each error bin is shown on top of each histogram.

Model accuracy was further evaluated through examining the frequency of absolute value of the forecasting errors within the different ranges of errors of up to ± 100, as shown in the figure above (**Figure 23**). The M1 model again performed exceptionally well compared to the other two models for Aiyura, since it had 78% of all the model's errors falling in the smallest magnitude ($0 < \text{error} \leq 100$) error bracket, whereas for models M2 and M5 in this category had 59% and 70% respectively. Furthermore, the M1 model when compared to the bivariate results in **Figure 11** (M1), the trivariate model performed better in the smallest error bracket ($0 < \text{error} \leq 100$) with 78% whereas the bivariate had 74% in the same range while 22% and 26 % of errors in the next error bracket ($100 \leq \text{error} \leq 200$). The inclusion of DMI in the trivariate models greatly improved the forecasting error for M9 models performance from 41% (bivariate) to 63% of errors in the first error bracket compared indicating ENSO and IOD enhances the model's performance.

5.3.3 Station 3- Ramu

Correlations performed between the best lead times for the climate mode indices and the three-month total rainfall, which forms a basis for bivariate models applied to capture ENSO impact. IOD indices (DMI) embedded into bivariate models (i.e. the trivariate models) to account for the compound effect of ENSO and IOD on the extreme events. Each model developed were evaluated using the *r*, *RMSE*, *MAE*, E_{NS}, *d*, and *L* to identify the best trivariate copula-based

regression models for rainfall forecasting in Ramu. The correlation, as shown in **Table 14** varied between $r = 0.20$ to $r = 0.57$ in correspondence with the linear agreement between the observed and forecasted rainfall datum. Model M8 had the strongest linear agreement, and the weakest was displayed by M1 while the next best two models M5 and M6 scored r = 0.54 apiece.

The models forecasting capabilities were further evaluated through the *RMSE* and *MAE*. When evaluating the model performance via *RMSE* and *MAE,* the smaller the magnitude of metrics, the better the models to reflect the deviation from forecasted data and the observed data. Therefore, based on **Table 14** the highest values for both the *RMSE* and *MAE* values were displayed by M1, M2 and M12 indicating poor performance whist M5 and M6 had the lowest which indicates both as best model based on *RMSE* and *MAE* metric.

Further evaluation using the E_{NS}, d and L index was also used to determine the performance of the best trivariate copula-based quantile regression models, and the results are shown in **Table 14**. The E_{NS} index logged in the highest value through M8 E_{NS} = 0.31, followed by M5 and M6 both with E_{NS} = 0.0.27. However with the d index as explained by Willmott et al. (2012), values approaching +1 represents a better model performance, although the results varied from 0.62 to 0.04, M8 had the highest index score of d = 16 on the other hand, M2 had $d = .04$ which indicates that the M2 performed poorly (Willmott *et al.* 2012). Furthermore, the results based on Legates-McCabe's index L the highest value was M5 and M6 had the best results at L = 0.32, 0.15 respectively while M1 and M2 performed poorly with negative values. From the model evaluations so far (r, *RMSE, MAE*, $E_{NS,}$ d and L), models M5, M6, M7 and M8 displayed exceptional results.

Table 14: Table showing performance evaluation metrics for the D-vine copula-based quantile regression rainfall models for Ramu, based on correlation coefficient (r), RMSE, MAE, E_{NS}, d and L. *RFP – The three-monthly total rainfall forecasting period, CI – three moths average Climate Index (Predictor), LTM – Lead time months (Predictor)*

Model Name	RFP	DMI Lead (0-6)	DMI LTM	r	RMSE (mm)	MAE (mm)	E_{NS} [-∞, 1]	d [0, 1]	L [-∞, 1]
M1	JFM	0	OND	0.20	222.06	170.24	-0.42	0.19	-0.12
M2	FMA	2	SON	0.24	199.68	142.67	-0.46	0.04	-0.02
M3	MAM	3	SON	0.11	160.75	117.13	-0.07	0.05	-0.03
M5	MJJ	0	FMA	0.54	91.97	61.52	0.27	0.52	0.32
M6	JJA	0	MAM	0.54	77.94	64.45	0.27	0.48	0.15
M7	JAS	1	MAM	0.51	93.52	76.45	0.20	0.53	0.14
M8	ASO	0	MJJ	0.57	105.28	89.06	0.31	0.62	0.13
M9	SON	0	JJA	0.45	127.08	101.15	0.15	0.52	0.08
M11	NDJ	2	JJA	0.30	161.55	133.82	0.04	0.37	0.01
M12	DJF	0	SON	0.42	198.18	160.94	0.01	0.48	0.00

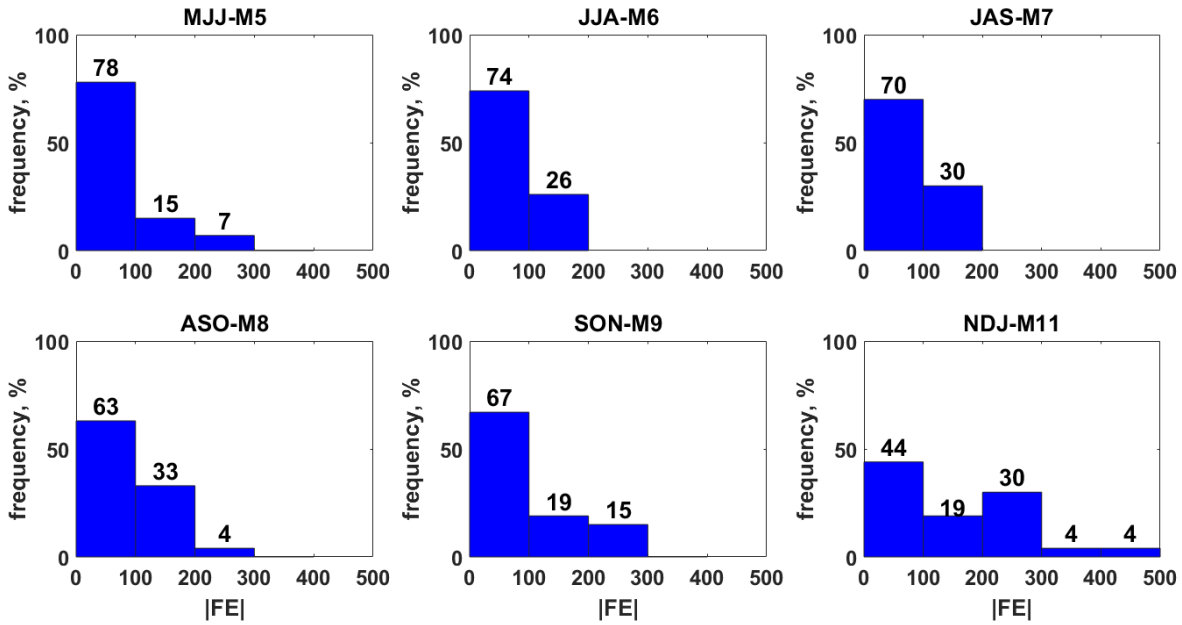


Figure 24: A histogram of the relative frequency of absolute forecasting error |FE| (mm) for trivariate rainfall models for Ramu. The number of datum points in each error bin is shown on top of each histogram.

The forecasting capabilities of the best four models as indicated through the evaluation process, were further evaluated using the frequency of absolute error plots through examining the frequency of the forecasting errors within the different ranges up to ± 100 as shown in **Figure 24**. As seen in the above plot there is no doubt M5 is the top model since a larger portion of errors (≈ 78%) lies within the smallest error bracket ($0 < \text{error} \leq 100$) while M6 had the second-highest errors (≈ 74%). The inclusion of DMI improved the rainfall forecast errors in the smallest error bracket for M5-JJA from 74% in the bivariate model to 78% in the current model.

5.3.4 Station 4- Dami

The D-vine copula-based quantile approach was also applied to develop the copula-based trivariate models for rainfall at Dami station. IOD indices (DMI) was inserted into bivariate models (i.e., the trivariate models). For the eleven different trivariate models, a robust model assessment require a combination of evaluations as no single statistical measure is conclusive, hence a wide range of statistical metrics are used to evaluate the models and includes the Pearson correlation coefficient (*r*), *RMSE*, *MAE*, E_{NS}, *d* and *L* including diagnostic plots to identify the best trivariate copula-based regression models for rainfall forecasting in Dami.

Table 15: Table showing performance evaluation metrics D-vine copula-based quantile regression rainfall models for Dami, based on correlation coefficient (r), *RMSE*, *MAE*, E_{NS}, *d* and *L*. *RFP – The three-monthly total rainfall forecasting period, CI – three moths average Climate Index (Predictor), LTM – Lead time months (ENSO and IOD) (Predictor)*

Model Name	RFP	Lead (0-6)	LTM	r	RMSE (mm)	MAE (mm)	E_{NS} [-∞, 1]	*d* [0, 1]	*L* [-∞, 1]
M1	JFM	3	JAS	0.48	321.46	267.35	0.21	0.48	0.05
M2	FMA	6	MJJ	0.25	357.12	308.99	-0.07	0.26	-0.04
M3	MAM	0	DJF	0.22	279.58	218.84	-0.01	0.21	0.09
M5	MJJ	3	NDJ	0.34	155.99	120.84	0.08	0.32	0.09
M6	JJA	3	DJF	0.25	156.19	125.52	-0.02	0.26	-0.03
M7	JAS	0	AMJ	0.00	199.08	129.89	-0.54	-0.03	-0.14
M8	ASO	0	MJJ	0.54	172.95	122.25	0.16	0.62	0.18
M9	SON	1	MJJ	0.61	145.16	106.90	0.35	0.64	0.25
M10	OND	2	MJJ	-0.23	299.36	223.13	-0.55	-0.63	-0.25
M11	NDJ	0	ASO	0.07	350.22	254.79	-0.13	0.17	0.03

M12	DJF	4	MJJ	0.32	331.47	258.06	-0.06	0.38	0.04

The performance of error indicators *RMSE* and *MAE* as shown in **Table 15**, ranged from 114.16 mm to 357.12 mm and 145.16 mm to 308.99 mm, respectively. Based on the error indicators the best model is M9 since it has the lowest values of *RMSE* (145.16 mm) and *MAE* (106.90 mm), while the inferior model having the highest error indicators is M2 (*RMSE* = 357.12 mm, *MAE* = 308.99 mm). It is worth mentioning that the smaller the values of *RMSE* and *MAE* it represents further preciseness of the rainfall estimation and in an ideal case for a perfect model value should be zero. The M9 model enhanced performance was further supported through the measure of linear relationship *r* reaffirming the model's superiority over the other models with the highest value of 0.61 and is followed by M8 with *r* = 054. So far based on the conventional metrics, *r*, *RMSE,* and *MAE* the model 9 performed better.

Furthermore, the copula-based trivariate models' performance was further evaluated using the E_{NS}, *d,* and *L* index. The M9 again performed remarkably well having the highest scores E_{NS} = 0.35, *d* = 0.64 and *L* = 0.25 including all performance evaluation measures evaluated so far. The second-best model clearly is M8 with E_{NS} = 0.16, *d* = 0.62 and *L* = 0.18 index values. M1 and M5 make up the best four models based on the evaluation metrics in **Table 15**.

Additionally, analogies were further made to assess the top rainfall models for Dami using scatterplots of the observed and forecasted values, as shown in **Figure 25.** From the four models, the plots reaffirm M9 superior performance, with the highest correlation coefficient value of $R^2 = 0.37$ demonstrating an overall 37% of the observed rainfall could be forecasted using that trivariate copula-based model. The subsequent model M8 recorded $R^2 = 0.29$ followed by M1 ($R^2 = 0.23$) with a lesser performance displayed by M5 having $R^2 = 0.11$. In addition, an alternative model performance metric is the gradient (m) of the linear fit and as seen from the models evaluated M8 was close to idyllic magnitudes (unity) further reaffirming the outcomes of the metrics with a value of m = 0.37, showing consistent results across all performance evaluations so far.

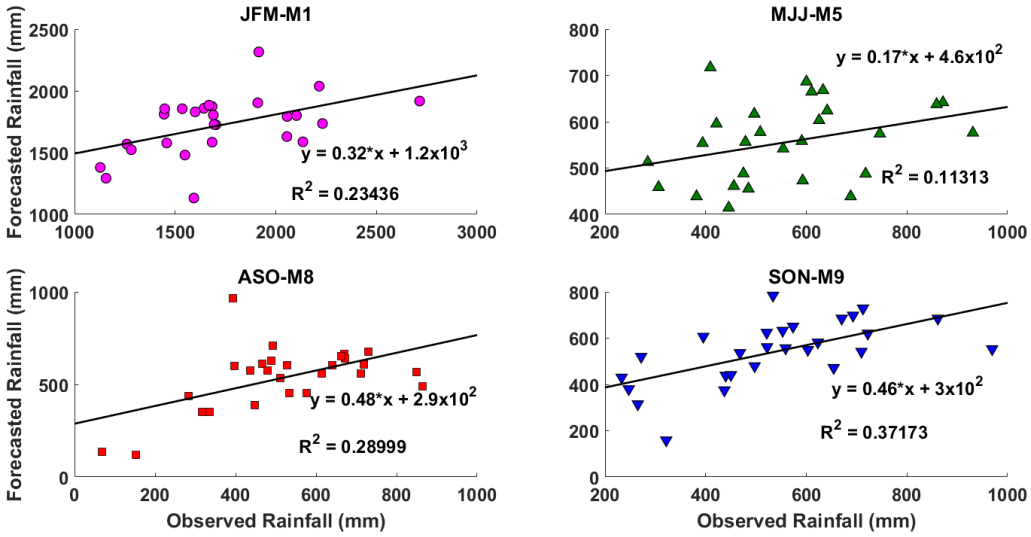


Figure 25: Scatter plots of the trivariate copula-based quantile regression models for the observed and forecasted rainfall for Dami station (PNG). (Note: The straight blue line is the least square fit line for each scatter plots of the form $R_{Pred} = m\ R_{Obs} + C$).

Finally, **Figure 26** shows a histogram plot displaying the percentage frequency distribution of forecasting error calculated in error brackets of ± 100 step-sizes which has a deeper analysis of actual errors and could aid in the interpretation of model accuracy for practical applications (Deo *et al.* 2016). No doubt the best performing model is M9 since a larger share of errors (≈ 63%) lies within the smallest error bracket (± 100) followed by M8 with 56%. This indicated the superiority of M9 over M8, M5 (52%) and M1 (19%) and models. The results are consistent with the earlier results in **Table 15**.

Furthermore, when comparing the histograms of forecasting errors in both bivariate (**Figure 18**) and trivariate models, the bivariate model for M9 displayed consistent results of 63% of forecasting error in the smallest error bin ($0 < \text{error} \leq 100$) in both the bi/trivariate models, however there was a noticeable increase in the $100 \leq \text{error} \leq 200$ error bracket observed in the trivariate model (26%) compared to 22% in the bivariate model indicating that DMI, when included, increased the M9 performance. M8 although performed well in the bivariate model with 70% in the smallest error bracket, with the trivariate model the performance dropped to 56% indicating no improvement in the trivariate model with the inclusion of IOD, however when compared to the next error bracket ($100 \leq \text{error} \leq 200$) the trivariate model improved the forecasting error from 22% (bivariate) to 26%. The bivariate model also performed better with the M9 model with 63% of errors in the smallest error bin while the trivariate model had 52%.

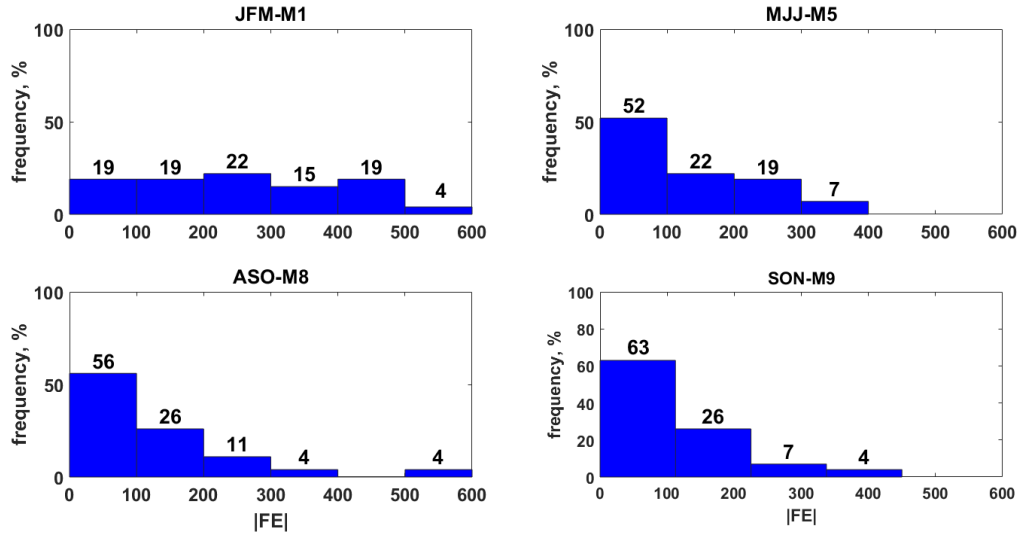


Figure 26: A histogram of the relative frequency distribution of absolute forecasting error |FE| (mm) for copula-based trivariate rainfall models for Dami station. The number of datum points in each error bin is shown on top of each histogram

5.4 Comparison of all sites

The models were finally evaluated using geographical comparison for the models using relative root mean squared error (*RRMSE*) and mean absolute percentage error (*MAPE*) for the study sites (Port Moresby, Aiyura, Ramu and Dami) was made. According to Mohammadi, K. *et al.* (2015) the different ranges of *RRMSE* is defined as follows; excellent if %RMSE < 10%; good if 10% < %RMSE < 20%; fair if 20% < % *RMSE* < 30%; and poor if % *RMSE* ≥ 30%.

Based on **Table 16** and **Figure 27**, the model with the lowest relative percentage errors (*RRMSE* and *MAPE*) across all the other trivariate rainfall models was M11 (*RRMSE* = 24.49% and *MAPE* = 19.69%) and M1 (*RRMSE* = 25.50% and *MAPE* = 20.99%) for Port Moresby followed by M12 (*RRMSE* = 30.33% and *MAPE* = 25.63%). In addition, M2 and M12 also performed well having both *RRMSE* and *MAPE* % between the 20% < *RRMSE* < 30% category. In contrast, models M6 and M7 generated the highest *RRMSE* and *MAPE* across all stations, with the highest *MAPE* observed M7 (135.46%). This was because the *MAPE* (the most common measure used to forecast error) works best if there are no extremes to the data and this scenario is expected since Port Moresby experiences very little rainfall around the July to September period. The accuracy of models M11, M1, M2, and M12 are considered as fair models since 20% < *RRMSE* < 30% whereas the other performed poorly since *RRMSE* > 30% (Despotovic *et al.* 2016)

For Aiyura, the lowest relative percentage errors were generated by M1 (*RRMSE* = 14.46% and *MAPE* = 11.54%) and M2 (*RRMSE* = 19.18% and *MAPE* = 16.97%). Overall M1 again displayed its lead over the other models followed by M2 and both can be classified as good models since both laid between the 10% < *RRMSE* < 20% range (Despotovic *et al.* 2016). The results for M1 is consistent with the results for the bivariate model (M1) of which this model performed slightly better in all evaluation metrics indicating the index DMI improves forecasting during the JFM period.

For the third rainfall station Ramu, M11 displayed the lowest *RRMSE* (22.74%), and *MAPE* (19.61%) including M3 with 24.34% and 19.01% respectively. M12 and M1 also displayed exceptional *RRMSE* and *MAPE* results of 26.67% and 22.24% and 29.06% and 23.27% respectively. Overall the four trivariate copula model performed within the ranges of a fair model since all the *RRMSE* was within the 20% < % *RRMSE* < 30% (Mohammadi, Kasra et al. 2015). Contrastingly, the models M5, M6, M7, and M8 which performed better with the evaluation metrics (*r*, *RMSE*, *MAE*, $E_{NS,}$ and *d*) did poorly with the *RRMSE* and *MAPE* all having scored ≥ 30%. Although M11, M12, M1, and M3 scored within the 'fair' model category (20% < % *RRMSE* < 30%) there performance in the evaluation metrics was not as good compared to M5-M8. However, M9 displayed consistent results throughout the evaluation, and based on the RRMSE and MAPE, it is a 'fair' model and is preferred for Ramu.

Finally, the *RRMSE* and *MAPE* metrics were applied to evaluate the trivariate copula models for Dami station. Based on **Figure 27**, the M1 is a good model since it had the lowest *RRMSE* and *MAPE* values of 18.64% and 15.39% respectively and is within the range 10% < %RMSE < 20%. The other model which scored between the 20% < % *RRMSE* < 30% (fair) included M2, M3, M5, M6, M9, and M12.

Table 16: A comparison of the copula-based trivariate models' performance at different study sites' using the relative error measures, the relative root mean square error (*RRMSE*) and mean absolute percentage error (*MAPE*).

	Port Moresby		**Aiyura**		**Ramu**		**Dami**	
Model Name	RRMSE, %	MAPE, %	RRMSE, %	MAPE, %	RRMSE, %	MAPE, %	RRMSE, %	MAPE, %
M1	25.50	20.99	14.46	11.54	29.06	23.27	18.64	15.39
M2	27.03	26.89	19.18	16.97	26.48	21.50	24.31	22.72
M3	40.91	36.37	26.71	24.53	24.34	19.01	25.65	22.50
M4	47.91	38.48	29.97	30.69	24.64	19.85	30.21	24.65
M5	72.56	87.35	32.13	32.07	35.50	32.04	27.70	22.87
M6	87.99	83.71	40.95	36.73	47.49	55.13	28.77	28.86
M7	63.76	135.46	42.24	40.40	47.87	70.33	39.63	49.78
M8	63.37	106.31	52.52	51.68	34.64	36.55	33.24	28.02
M9	73.14	114.05	42.37	37.68	27.16	24.55	27.01	23.67
M10	44.63	51.42	29.43	22.32	25.49	23.14	41.85	36.37
M11	26.17	19.02	19.04	15.60	22.74	19.61	31.83	23.45
M12	26.48	22.12	20.23	17.21	26.67	22.24	22.39	18.54

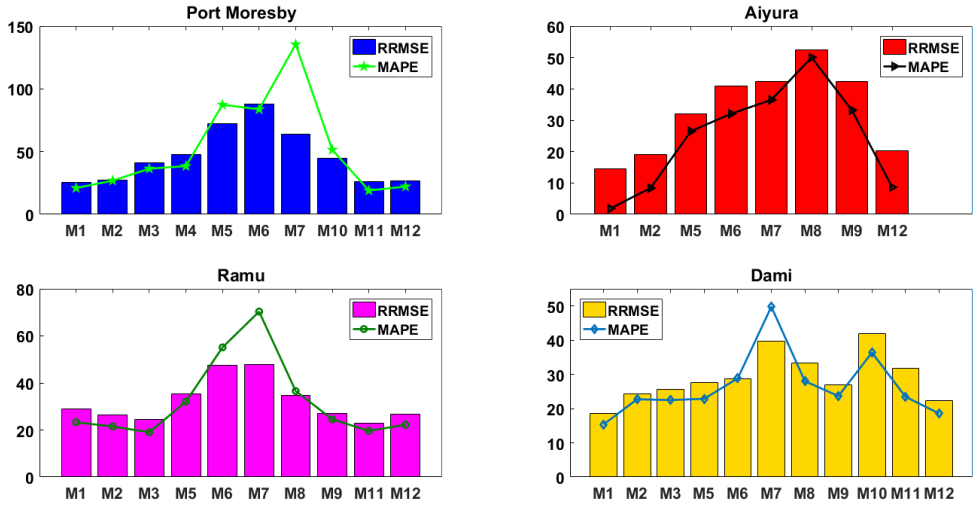


Figure 27: The mean absolute percentage error (*MAPE*) and relative root mean square error (*RMSE*) for the different trivariate copula-based rainfall models for the four different study sites.

5.5 Concluding Remarks

This chapter explored the spatio-temporal influence of the climate driver IOD (DMI) influence with respect to rainfall in four locations in PNG. Based on the bivariate copula model defined in Chapter 4, the trivariate copula models were developed through the D-vine copula-based quantile approach to account for the joint influence of ENSO and IOD on rainfall. From the results discussed in this chapter, for Port Moresby the M11 as it displayed consistent results. For Aiyura, M1 and M2 displayed consistent and exceptional results and can be used to forecast the total rainfall for JFM and FMA period. Furthermore, the best performing model for Ramu, M9 which displayed consistent results throughout the model evaluation is preferred, while Dami station, M1, and M9 are equally exceptional models that can be used to forecast the total ASO and SON rainfall.

From the results discussed so far (**Figure 8** and **Figure 11**) for Port Moresby, the inclusion of DMI clearly improved the forecasting error of M1, M6, M11 and M10 in the first error bracket ($0 < \text{error} \leq 100$) and the second error bracket ($100 \leq \text{error} \leq 200$) brackets when compared. The performance of the trivariate models when compared to the bivariate models, there was not much improvement in model performance. Improvements were observed in the Aiyura with the trivariate models with noticeable increase in M1, M2 outperforming the results displayed by the bivariate models. Whereas for Ramu, the inclusion of DMI improved the rainfall forecast errors in the smallest error bracket for M5-JJA from 74% in the bivariate model to 78% in the current model, however there was not much improvement other model performance when DMI was added to the trivariate model. Finally, for Dami the best models M8 and M9 when compared to the bivariate models, the performance of both models in the trivariate models was noticeably lower indicating no improvement in model performance with the inclusion of DMI in the trivariate models.

CHAPTER 6 CONCLUSION

6.1 Summary of the Findings

Rainfall forecasting is an essential task in precision agriculture as subsistence agricultural activities are associated with the number of water resources available for better-sustained crop health. The paramountcy statistical models and artificial intelligence-based big data analytic techniques in sustainable agricultural practices are evolving very rapidly in the 21st century. This study was primarily focused on applying copula-based statistical approaches to forecast rainfall in four different rainfall stations (Port Moresby, Aiyura, Ramu and Dami) in PNG.

This study has employed the two different types of El Niño/Southern Oscillation (ENSO): canonical ENSO and ENSO Modoki to model rainfall (bivariate) as well as the former coupled with IOD (DMI) to jointly model rainfall (trivariate). The rainfall data (monthly) used in this study was from the years 1990-2017 whist the climate modes index data from 1989 to 2017. Also, the impact of climate mode index on rainfall was investigated using the rank-based correlation measure.

To illustrate the usefulness of climate mode indices for conditional-based rainfall forecasting through the D-vine copula approach, statistically significant lagged correlations were identified based on which climate index that had the highest correlation with the 3-month total rainfall. Spatial correlation of the climate indices across different study sites was not identical across all four stations. In order to completely understand the nature of the rainfall-ENSO relations, it is important to appreciate that rainfall at any site is reliant on several factors which are not examined in detail in this thesis, but most likely reflect the interaction such as topography and the Western Pacific monsoon.

The D-vine copula-based quantile approach allows identifying the most influent predictors for rainfall forecast through the AIC-corrected conditional log-likelihood (cll^{AIC}). To forecast rainfall, the lagged correlations of IOD indicator (DMI) with three-month total rainfall were established up to 7 months ahead (e.g., in a 'lead-0' study January to March average SOI is compared with April to June total rainfall) using the Kendall rank correlation coefficients. Correlations performed between the best lead times for the climate mode indices and the three-month total rainfall forms the basis for bivariate models applied to capture ENSO impact. IOD indices (DMI) are then embedded into the bivariate models (i.e., the trivariate models) to account for the compound effect of ENSO and IOD on the extreme events.

6.1.1 Bivariate Models

i. Port Moresby: Evaluations based on the predictor metrics also indicated M11 to be the best model, displaying the best results compared to all model with r = 0.82, *RMSE* = 91.74, *MAE* =73.15, E_{NS} =0.61, WI = 081, L =0.41. RRMSE results for M11 laid between 20% < % *RRMSE* < 30% indicating a fair rainfall forecasting model. Additionally, the M11 which represents the NDJ-SOI lead-1 also displayed the highest correlation (0.63) for Port Moresby, and the performance evaluations reaffirmed the model's performance.

ii. Aiyura: The forecasting skill displayed by JFM and FMA rainfall period with EMI and Niño 4.0 respectively showed consistent results throughout the results. Both had statistically significant Kendall's correlation of 0.49 and 0.38 respectively, and both displayed consistent superior performance during the course of the model's evaluations. The best performing model was no doubt M1, displaying with r = 0.63, *RMSE* = 105.99, *MAE* =89.75, E_{NS} =0.63, WI = 0.38, L =0.20. Additionally, the RRMSE indicate that M1 is a good model with 15.39% (good if; 10% < % *RRMSE* < 20%). Additional evaluations using the plots further reaffirms M1 superior performance.

iii. Ramu: The EMI correlation with Ramu rainfall is noticeable around the March to May period, and this is consistent with studies conducted in Australia which shows north-western Australian autumn rainfall is correlated with El Nino Modoki (EMI). On the other hand, the positive correlation recorded between the SON – JFM rainfall on ENSO is consistent with the period in which ENSO's peak intensity (December-April) and weakens during May-July which justifies the negative correlation values. The best rainfall model after evaluating the performance metrics was M6 and was further supported by the frequency plots, 81% of errors lies within smallest error bracket (0 < error ≤ 100), however with the *RRMSE* and *MAPE* it performed poorly (45.33% and 53.30 respectively). M9 model is preferred.

iv. Dami: The EMI negative correlation values associated with Dami's rainfall occurs around June – August which is similar to negative EMI-rainfall correlations observed in Australia. On the other hand, again and as similar to Ramu station the positive correlation recorded between the SON – JFM rainfall on ENSO is consistent with the period in which ENSO's is at peak intensity (December-April) and weakens during around May-August which supports the negative correlation values

6.1.1 Trivariate Models

The trivariate copula-based models constructed through the D-vine copula-based quantile regression approach consists of the IOD indices (DMI) embedded into the bivariate models to account for the compound effect of ENSO and IOD on the rainfall. The findings are as follows;

i. Port Moresby: Model M11 and M1 are 'fair' models as indicated through the percentage errors with M11 (*RRMSE* = 24.49% and *MAPE* = 19.69%) and M1 (*RRMSE* = 25.50% and *MAPE* = 20.99%) but based on the performance evaluation the M11 is preferred.
ii. Aiyura: Based the performance evaluation metrics, M1 model had the highest E_{NS}, *WI*, and *L* with 0.45, 0.69 and 0.30 scores while the next best model M2 with 0.37, 0.64 and 0.20 respectively. In addition, both models displayed good results with the relative percentage errors, M1 (*RRMSE* = 14.46% and *MAPE* = 11.54%) and M2 (*RRMSE* = 19.18% and *MAPE* = 16.97%). Both can be classified as good models since both scores are between the 10% < *RRMSE* < 20% range
iii. Ramu: From the evaluations made to assess the model's performance, M9 displayed consistent results throughout the model evaluation is preferred
iv. Dami: Based on the *RRMSE* and *MAE*, M1 is a good model since the models *RRMSE* and value was within the range 10% < % *RRMSE* < 20%. The other model which scored between the 20% < % *RRMSE* < 30% (fair) included M2, M3, M5, M6, M9 and M12, however model M9 is preferred as it displayed consistent results throughout the model evaluation.

6.2 Synthesis

Based on the results presented in this study, for the bivariate and trivariate copula-based rainfall models, bivariate models for Aiyura displayed consistent results with models M1 and M2, the inclusion of DMI in the trivariate model further helped improved the overall performance for both models. For Port Moresby, the inclusion of DMI in the trivariate models when compared to the bivariate models, there was not much improvement in model performance evaluations, however the forecasting error for M1, M6, M11 and M10 was improved. The best rainfall model identified for Port Moresby is the bivariate M11 model. For Ramu, the M9 model displayed consistent results throughout the model evaluation and is preferred to forecast rainfall for June-July-August, while Dami station, the bivariate model M8, and M9 are equally exceptional models that can be used to forecast the total ASO and SON rainfall.

6.2 Recommendations for Future Work

Since this study was limited in terms of its scope as a Master's project, and the objectives that could be fulfilled in this period, the following recommendations have been made for further research opportunities:

- **i.** Port Moresby rainfall experiences very-low rainfalls, negative values in predicted rainfall in M8-ASO and M9-SON in the years where rainfall amount is low, and these negative values contributed to the reduction of model performance. Future study should explore further this issue related to a marginal fitting method to improve the forecast skill using copula-based models.
- **ii.** The present study considers the same lagged time of predictors (e.g., JJA-SOI and JJA-DMI) to forecast the seasonal rainfall. Future work may combine different lagged time of predictors (e.g. JJA-SOI and ASO DMI). In addition, inclusions of other large-scale climate mode indices are expected to improve the performance of the forecast model.
- **iii.** The methods of fitting marginal distributions need to be investigated further to avoid the occurrence of the negative forecasted values and hence to improve the performance of the forecast model.
- **iv.** Statistical models, in general, rely on the relationship between predictand and predictors. The limitation of data length used in this thesis may affect the performance of the forecast models since it does not fully reflect the characteristic in the relationship between large-scale climate mode indices and rainfall. Further, due to the short record, the co-occurrences between different phases of ENSO and IOD are rare, which may affect the analysis of the compound influences of such two indices on extreme rainfall events.

One of the limitations of this Master's study (performed over 1.5 years) was that it was restricted to developing only a copula-based approach for rainfall forecasting. However, it is recommended that a further study be performed that compares copula models with other kinds of models, such as machine learning, time-series models or other statistical models used in rainfall forecasting areas. This was beyond the scope of the present study but can form useful subject for further investigation.

REFERENCE

Aas, K, Czado, C, Frigessi, A & Bakken, H 2009, 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 182-98.

Abdi, H 2007, 'The Kendall rank correlation coefficient', *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pp. 508-10.

AghaKouchak, A 2014, 'Entropy–Copula in Hydrology and Climatology', *Journal of Hydrometeorology*, vol. 15, no. 6, pp. 2176-89.

AghaKouchak, A, Bárdossy, A & Habib, E 2010, 'Copula-based uncertainty modelling: application to multisensor precipitation estimates', *Hydrological Processes*, pp. n/a-n/a.

Allen, BJ & Bourke, RM 2001, 'The 1997 drought and frost in PNG: overview and policy implications', *Food Security for Papua New Guinea*, vol. 11, p. 155.

Ashok, K, Guan, ZY & Yamagata, T 2003, 'A look at the relationship between the ENSO and the Indian Ocean Dipole', *Journal of the Meteorological Society of Japan*, vol. 81, no. 1, pp. 41-56.

Ashok, K, Guan, Z & Yamagata, T 2003, 'Influence of the Indian Ocean Dipole on the Australian winter rainfall', *Geophysical Research Letters*, vol. 30, no. 15.

Ashok, K, Behera, SK, Rao, SA, Weng, H & Yamagata, T 2007, 'El Niño Modoki and its possible teleconnection', *Journal of Geophysical Research: Oceans*, vol. 112, no. C11.

Ball, J, Babister, M, Nathan, R, Weinmann, P, Weeks, W, Retallick, M & Testoni, I 2016, *Australian Rainfall and Runoff-A guide to flood estimation*, Commonwealth of Australia, 1925297071, <http://hdl.handle.net/10453/85297>.

Barros, RS, da Se Mota, JW, Da Matta, FM & Maestri, M 1997, 'Decline of vegetative growth in Coffea arabica L. in relation to leaf temperature, water potential and stomatal conductance', *Field Crops Research*, vol. 54, no. 1, pp. 65-72.

Bedford, T & Cooke, RM 2001, 'Probability density decomposition for conditionally dependent random variables modeled by vines', *Annals of Mathematics and Artificial intelligence*, vol. 32, no. 1-4, pp. 245-68.

Black, JR & Thompson, SR 1978, 'Some evidence on weather-crop-yield interaction', *American Journal of Agricultural Economics*, vol. 60, no. 3, pp. 540-3.

Brechmann, E 2010, 'Truncated and simplified regular vines and their applications'.

Cai, WJ, van Rensch, P, Cowan, T & Hendon, HH 2012, 'An Asymmetry in the IOD and ENSO Teleconnection Pathway and Its Impact on Australian Climate', *Journal of Climate*, vol. 25, no. 18, pp. 6318-29.

Chai, T & Draxler, RR 2014, 'Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature', *Geoscientific model development*, vol. 7, no. 3, pp. 1247-50.

Chakrabarti, A & Sen, R 2019, 'Copula estimation for nonsynchronous financial data', *arXiv preprint arXiv:1904.10182*.

Chan, JCL & Zhou, W 2005, 'PDO, ENSO and the early summer monsoon rainfall over south China', *Geophysical Research Letters*, vol. 32, no. 8.

Chen, L & Guo, S 2019a, 'Copula-Based Seasonal Design Flood Estimation', in *Copulas and Its Application in Hydrology and Water Resources*, Springer, pp. 73-96.

Chen, L & Guo, S 2019b, 'Copula-Based Uncertainty Evolution Model for Flood Forecasting', in *Copulas and Its Application in Hydrology and Water Resources*, Springer, pp. 211-35.

Chen, L & Guo, S 2019c, *Copulas and Its Application in Hydrology and Water Resources*, Springer.

Chowdhary, H, Escobar, LA & Singh, VP 2011, 'Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data', *Hydrology Research*, vol. 42, no. 2-3, pp. 193-216.

Christian, G & Anne-Catherine, F 2007, 'Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask'.

Cobon, DH, Ewai, M, Inape, K & Bourke, RM 2016, 'Food shortages are associated with droughts, floods, frosts and ENSO in Papua New Guinea', *Agricultural Systems*, vol. 145, pp. 150-64.

Connor, R 2015, *The United Nations world water development report 2015: water for a sustainable world*, vol. 1, UNESCO Publishing.

Cottrill, A, Charles, A & Kuleshov, Y 2013, 'An analysis of seasonal forecasts from POAMA and SCOPIC in the Pacific region', in *EGU General Assembly Conference Abstracts*.

Czado, C 2010, 'Pair-copula constructions of multivariate copulas', in *Copula theory and its applications*, Springer, pp. 93-109.

Danladi, A, Stephen, M, Aliyu, BM, Gaya, GK, Silikwa, NW & Machael, Y 2018, 'Assessing the influence of weather parameters on rainfall to forecast river discharge based on short-term', *Alexandria Engineering Journal*, vol. 57, no. 2, pp. 1157-62.

Deo, RC, Wen, X & Qi, F 2016, 'A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset', *Applied Energy*, vol. 168, pp. 568-93.

Deo, RC, Kisi, O & Singh, VP 2017, 'Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model', *Atmospheric Research*, vol. 184, pp. 149-75.

Deo, RC, Tiwari, MK, Adamowski, JF & Quilty, JM 2017, 'Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model', *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 5, pp. 1211-40.

Deo, RC, Ghorbani, MA, Samadianfard, S, Maraseni, T, Bilgili, M & Biazar, M 2018, 'Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data', *Renewable energy*, vol. 116, pp. 309-23.

Despotovic, M, Nedic, V, Despotovic, D & Cvetanovic, S 2016, 'Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation', *Renewable & Sustainable Energy Reviews*, vol. 56, pp. 246-60.

Dodangeh, E, Shahedi, K, Solaimani, K, Shiau, JT & Abraham, J 2019, 'Data-based bivariate uncertainty assessment of extreme rainfall-runoff using copulas: comparison between annual maximum series (AMS) and peaks over threshold (POT)', *Environmental monitoring and assessment*, vol. 191, no. 2, p. 67.

Dogar, MM, Kucharski, F, Sato, T, Mehmood, S, Ali, S, Gong, Z, Das, D & Arraut, J 2019, 'Towards understanding the global and regional climatic impacts of Modoki magnitude', *Global and planetary change*.

Drake, V 1994, 'The influence of weather and climate on agriculturally important insects: an Australian perspective', *Australian Journal of Agricultural Research*, vol. 45, no. 3, pp. 487-509.

Drosdowsky, W & Chambers, LE 2001, 'Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall', *Journal of Climate*, vol. 14, no. 7, p. 1677.

Duong, T 2007, 'ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R', *Journal of Statistical Software*, vol. 21, no. 7, pp. 1-16.

Feng, J, Chen, W, Tam, CY & Zhou, W 2011, 'Different impacts of El Niño and El Niño Modoki on China rainfall in the decaying phases', *International Journal of Climatology*, vol. 31, no. 14, pp. 2091-101.

Fischer, M, Kraus, D, Pfeuffer, M & Czado, C 2017, 'Stress Testing German Industry Sectors: Results from a Vine Copula Based Quantile Regression', *Risks*, vol. 5, no. 3, p. 38.

Fousekis, P & Grigoriadis, V 2017, 'Joint price dynamics of quality differentiated commodities: copula evidence from coffee varieties', *European Review of Agricultural Economics*, vol. 44, no. 2, pp. 337-57.

Frahm, G, Junker, M & Szimayer, A 2003, 'Elliptical copulas: applicability and limitations', *Statistics & Probability Letters*, vol. 63, no. 3, pp. 275-86.

Genest, C & Mackay, J 1986, 'The Joy of Copulas - Bivariate Distributions with Uniform Marginals', *American Statistician*, vol. 40, no. 4, pp. 280-3.

Genest, C & Rivest, LP 1993, 'Statistical-Inference Procedures for Bivariate Archimedean Copulas', *Journal of the American statistical Association*, vol. 88, no. 423, pp. 1034-43.

Genest, C, Quessy, J-F & Remillard, B 2006, 'Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation', *Scandinavian Journal of Statistics*, vol. 33, no. 2, pp. 337-66.

Goodrich, GB 2007, 'Influence of the Pacific decadal oscillation on winter precipitation and drought during years of neutral ENSO in the western United States', *Weather and forecasting*, vol. 22, no. 1, pp. 116-24.

Grimaldi, S, Petroselli, A, Salvadori, G & De Michele, C 2016, 'Catchment compatibility via copulas: A non-parametric study of the dependence structures of hydrological responses', *Advances in water Resources*, vol. 90, pp. 116-33.

Hofert, M 2008, 'Sampling Archimedean copulas', *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5163-74.

Hombunaka, P & von Enden, J 2001, 'The Influence of Available Water in 1997 on Yield of Arabica Coffee in 1998 at Aiyura, Eastern Highlands Province1', *Food Security for Papua New Guinea*, vol. 11, p. 242.

Hora, J & Campos, P 2015, 'A review of performance criteria to validate simulation models', *Expert Systems*, vol. 32, no. 5, pp. 578-95.

Howard, G & Bartram, J 2010, *The resilience of water supply and sanitation in the face of climate change*, Vision 2030, Technical report, World Health Organization, Geneva, Switzerland.

Htike, KK & Khalifa, OO 2010, 'Rainfall Forecasting Models Using Focused Time-Delay Neural Networks'.

Huang, W & Prokhorov, A 2014, 'A goodness-of-fit test for copulas', *Econometric Reviews*, vol. 33, no. 7, pp. 751-71.

Jin, EK, Kinter, JL, Wang, B, Park, CK, Kang, IS, Kirtman, BP, Kug, JS, Kumar, A, Luo, JJ, Schemm, J, Shukla, J & Yamagata, T 2008, 'Current status of ENSO prediction skill in coupled ocean–atmosphere models', *Climate Dynamics*, vol. 31, no. 6, pp. 647-64.

Joe, H 1997, *Multivariate models and multivariate dependence concepts*, Chapman and Hall/CRC.

Khedun, CP, Mishra, AK, Singh, VP & Giardino, JR 2014, 'A copula-based precipitation forecasting model: Investigating the interdecadal modulation of ENSO's impacts on monthly precipitation', *Water Resources Research*, vol. 50, no. 1, pp. 580-600.

Kim, M-K & Kim, Y-H 2010, 'Seasonal prediction of monthly precipitation in China using large-scale climate indices', *Advances in Atmospheric Sciences*, vol. 27, no. 1, p. 47.

Klein, N, Kneib, T, Marra, G, Radice, R, Rokicki, S & McGovern, M 2018, 'Mixed binary-continuous copula regression models with application to adverse birth outcomes', *Statistics in Medicine*, vol. 38, no. 3, pp. 413-36.

Kong, X, Zeng, X, Chen, C, Fan, Y, Huang, G, Li, Y & Wang, C 2018, 'Development of a Maximum Entropy-Archimedean Copula-Based Bayesian Network Method for Streamflow Frequency Analysis—A Case Study of the Kaidu River Basin, China', *Water*, vol. 11, no. 1, p. 42.

Korecha, D & Barnston, AG 2007, 'Predictability of june–september rainfall in Ethiopia', *Monthly Weather Review*, vol. 135, no. 2, pp. 628-50.

Kraus, D & Czado, C 2017, 'D-vine copula based quantile regression', *Computational Statistics & Data Analysis*, vol. 110, pp. 1-18.

Kumar, R, Stephens, M & Weir, T 2014, 'Rainfall trends in Fiji', *International Journal of Climatology*, vol. 34, no. 5, pp. 1501-10.

Legates, DR & Davis, RE 1997, 'The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches', *Geophysical Research Letters*, vol. 24, no. 18, pp. 2319-22.

Legates, DR & McCabe, GJ 1999, 'Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation', *Water Resources Research*, vol. 35, no. 1, pp. 233-41.

Li, H, Wang, D, Singh, VP, Wang, Y, Wu, J, Wu, J, Liu, J, Zou, Y, He, R & Zhang, J 2019, 'Non-stationary frequency analysis of annual extreme rainfall volume and intensity using Archimedean copulas: A case study in eastern China', *Journal of Hydrology*, vol. 571, pp. 114-31.

Liu, Z, Zhou, P, Chen, X & Guan, Y 2015, 'A multivariate conditional model for streamflow prediction and spatial precipitation refinement', *Journal of Geophysical Research: Atmospheres*, vol. 120, no. 19.

Lobell, DB & Burke, MB 2008, 'Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation', *Environmental Research Letters*, vol. 3, no. 3.

Ludescher, J, Gozolchiani, A, Bogachev, MI, Bunde, A, Havlin, S & Schellnhuber, HJ 2013, 'Improved El Niño forecasting by cooperativity detection', *Proceedings of the National Academy of Sciences*, vol. 110, no. 29, pp. 11742-5.

Lyu, L, Büntgen, U, Treydte, K, Yu, K, Liang, H, Reinig, F, Nievergelt, D, Li, M-H & Cherubini, P 2019, 'Tree rings reveal hydroclimatic fingerprints of the Pacific Decadal Oscillation on the Tibetan Plateau', *Climate Dynamics*, pp. 1-15.

Madden, RA & Julian, PR 1971, 'Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific', *Journal of the atmospheric sciences*, vol. 28, no. 5, pp. 702-8.

Manner, H, Fard, FA, Pourkhanali, A & Tafakori, L 2019, 'Forecasting the joint distribution of Australian electricity prices using dynamic vine copulae', *Energy Economics*, vol. 78, pp. 143-64.

Marengo, J 2015, 'Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts', in *AGU Fall Meeting Abstracts*.

Maydeu-Olivares, A & Garcia-Forero, C 2010, 'Goodness-of-fit testing', *International encyclopedia of education*, vol. 7, no. 1, pp. 190-6.

McAlpine, JR 1983, *Climate of Papua New Guinea*, Canberra, ACT: Commonwealth Scientific and Industrial Research Organization ….

McBride, JL & Nicholls, N 1983, 'Seasonal relationships between Australian rainfall and the Southern Oscillation', *Monthly Weather Review*, vol. 111, no. 10, pp. 1998-2004.

Min, SK, Cai, WJ & Whetton, P 2013, 'Influence of climate variability on seasonal extremes over Australia', *Journal of Geophysical Research-Atmospheres*, vol. 118, no. 2, pp. 643-54.

Moftakhari, H, Schubert, JE, AghaKouchak, A, Matthew, RA & Sanders, BF 2019, 'Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries', *Advances in water Resources*, vol. 128, pp. 28-38.

Mohammadi, K, Shamshirband, S, Anisi, MH, Alam, KA & Petkovic, D 2015, 'Support vector regression based prediction of global solar radiation on a horizontal surface', *Energy Conversion and Management*, vol. 91, pp. 433-41.

Mohammadi, K, Shamshirband, S, Tong, CW, Arif, M, Petković, D & Ch, S 2015, 'A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation', *Energy Conversion and Management*, vol. 92, pp. 162-71.

Nagler, T, Bumann, C & Czado, C 2019, 'Model selection in sparse high-dimensional vine copula models with an application to portfolio risk', *Journal of Multivariate Analysis*, vol. 172, pp. 180-92.

Nalley, D, Adamowski, J, Biswas, A, Gharabaghi, B & Hu, W 2019, 'A multiscale and multivariate analysis of precipitation and streamflow variability in relation to ENSO, NAO and PDO', *Journal of Hydrology*.

Nash, JE & Sutcliffe, JV 1970, 'River flow forecasting through conceptual models part I—A discussion of principles', *Journal of Hydrology*, vol. 10, no. 3, pp. 282-90.

Nelsen, RB 2003, 'Properties and applications of copulas: A brief survey', in *Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance (J. Dhaene, N. Kolev, and P. Morettin, eds.), Institute of Mathematics and Statistics, University of Sao Paulo*, Citeseer, p. 10.

Nelsen, RB 2007, *An introduction to copulas*, Springer Science & Business Media.

Nguyen-Huy, T, Deo, RC, An-Vo, D-A, Mushtaq, S & Khan, S 2017, 'Copula-statistical precipitation forecasting model in Australia's agro-ecological zones', *Agricultural Water Management*, vol. 191, pp. 153-72.

Nguyen-Huy, T, Deo, RC, Mushtaq, S, An-Vo, DA & Khan, S 2018, 'Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach', *European Journal of Agronomy*, vol. 98, pp. 65-81.

Nguyen-Huy, T, Deo, RC, Mushtaq, S, Kath, J & Khan, S 2019, 'Copula statistical models for analyzing stochastic dependencies of systemic drought risk and potential adaptation strategies', *Stochastic Environmental Research and Risk Assessment*, vol. 33, no. 3, pp. 779-99.

Nobre, GG, Hunink, JE, Baruth, B, Aerts, JC & Ward, PJ 2019, 'Translating large-scale climate variability into crop production forecast in Europe', *Scientific reports*, vol. 9, no. 1, p. 1277.

Noh, H, El Ghouch, A & Bouezmarni, T 2013, 'Copula-Based Regression Estimation and Inference', *Journal of the American statistical Association*, vol. 108, no. 502, pp. 676-88.

Nur'utami, MN & Hidayat, R 2016, 'Influences of IOD and ENSO to Indonesian rainfall variability: role of atmosphere-ocean interaction in the Indo-Pacific sector', *Procedia Environmental Sciences*, vol. 33, pp. 196-203.

Okhrin, O, Odening, M & Xu, W 2013, 'Systemic Weather Risk and Crop Insurance: The Case of China', *Journal of Risk and Insurance*, vol. 80, no. 2, pp. 351-72.

Paek, H, Yu, J-Y & Qian, C 2017, 'Why were the 2015/2016 and 1997/1998 extreme El Niños different', *Geophys. Res. Lett*, vol. 44, p. 2.

Prasad, R, Deo, RC, Li, Y & Maraseni, T 2017, 'Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm', *Atmospheric Research*, vol. 197, pp. 42-63.

Rasel, HM, Imteaz, MA & Mekanik, F 2016, 'Investigating the Influence of Remote Climate Drivers as the Predictors in Forecasting South Australian Spring Rainfall', *International Journal of Environmental Research*, vol. 10, no. 1, pp. 1-12.

Rasmusson, EM & Carpenter, TH 1982, 'Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño'.

Rauf, UFA & Zeephongsekul, P 2014a, 'Copula based analysis of rainfall severity and duration: a case study', *Theoretical and Applied Climatology*, vol. 115, no. 1-2, pp. 153-66.

Rauf, UFA & Zeephongsekul, P 2014b, 'Analysis of rainfall severity and duration in Victoria, Australia using non-parametric copulas and marginal distributions', *Water resources management*, vol. 28, no. 13, pp. 4835-56.

Rayner, NA, Parker, DE, Horton, EB, Folland, CK, Alexander, LV, Rowell, DP, Kent, EC & Kaplan, A 2003, 'Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century', *Journal of Geophysical Research-Atmospheres*, vol. 108, no. D14.

Ribeiro, AF, Russo, A, Gouveia, CM & Páscoa, P 2019, 'Copula-based agricultural drought risk of rainfed cropping systems', *Agricultural Water Management*, vol. 223, p. 105689.

Ricci, V 2005, 'Fitting distributions with R', *Contributed Documentation available on CRAN*, vol. 96.

Ropelewski, CF & Jones, PD 1987, 'An extension of the Tahiti–Darwin southern oscillation index', *Monthly Weather Review*, vol. 115, no. 9, pp. 2161-5.

Rosenzweig, C, Iglesias, A, Yang, X, Epstein, PR & Chivian, E 2001, 'Climate change and extreme weather events; implications for food production, plant diseases, and pests', *Global change & human health*, vol. 2, no. 2, pp. 90-104.

Schepen, A, Wang, QJ & Robertson, D 2012, 'Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall', *Journal of Climate*, vol. 25, no. 4, pp. 1230-46.

Schepsmeier, U 2010, 'Maximum likelihood estimation of C-vine pair-copula constructions based on bivariate copulas from different families'.

Schepsmeier, U 2015, 'Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review', *Journal of Multivariate Analysis*, vol. 138, pp. 34-52.

Schepsmeier, U, Stoeber, J, Brechmann, E & Graeler, B 2017, 'R package 'VineCopula': Statistical inference of vine copulas', *Available at: cran. r-project. org: https://cran. r-project. org/web/packages/VineCopula/VineCopula. pdf*.

Schepsmeier, U, Stoeber, J, Brechmann, EC, Graeler, B, Nagler, T, Erhardt, T, Almeida, C, Min, A, Czado, C & Hofmann, M 2018, 'Package 'VineCopula''.

Schmidt, N, Lipp, E, Rose, J & Luther, M 2001, 'ENSO influences on seasonal rainfall and river discharge in Florida', *Journal of Climate*, vol. 14, no. 4, p. 615.

Simolo, C, Brunetti, M, Maugeri, M & Nanni, T 2010, 'Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach', *International Journal of Climatology*, vol. 30, no. 10, pp. 1564-76.

Sklar, A 1996, 'Random variables, distribution functions, and copulas: a personal look backward and forward', *Lecture Notes-Monograph Series*, pp. 1-14.

Sklar, M 1959, 'Fonctions de repartition an dimensions et leurs marges', *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229-31.

Smith, I, Moise, A, Inape, K, Murphy, B, Colman, R, Power, S & Chung, C 2013, 'ENSO-related rainfall changes over the New Guinea region', *Journal of Geophysical Research-Atmospheres*, vol. 118, no. 19, pp. 10665-75.

Smith, I, Moise, A, Inape, K, Murphy, B, Colman, R, Power, S & Chung, C 2013, 'ENSO-related rainfall changes over the New Guinea region', *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 19, pp. 10,665-10,75.

Stone, RC, Hammer, GL & Marcussen, T 1996, 'Prediction of global rainfall probabilities using phases of the southern oscillation index', *Nature*, vol. 384, no. 6606, pp. 252-5.

Sumi, SM, Zaman, MF & Hirose, H 2012, 'A Rainfall Forecasting Method Using Machine Learning Models and Its Application to the Fukuoka City Case', *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 4, pp. 841-54.

Taschetto, AS & England, MH 2009, 'El Niño Modoki Impacts on Australian Rainfall', *Journal of Climate*, vol. 22, no. 11, pp. 3167-74.

Team, RC 2018, 'R Foundation for Statistical Computing; Vienna, Austria: 2014', *R: A language and environment for statistical computing*, p. 2013.

Troup, AJ 1965, 'The 'southern oscillation'', *Quarterly Journal of the Royal Meteorological Society*, vol. 91, no. 390, pp. 490-506.

Ubilava, D & Abdolrahimi, M 2019, 'The El Nino impact on maize yields is amplified in lower income teleconnected countries', *Environmental Research Letters*.

Valle, D & Kaplan, D 2019, 'Quantifying the impacts of dams on riverine hydrology under non-stationary conditions using incomplete data and Gaussian copula models', *Sci Total Environ*, vol. 677, pp. 599-611.

Vazifehkhah, S, Tosunoglu, F & Kahya, E 2019, 'Bivariate risk analysis of the droughts using a nonparametric multivariate standardized drought index and copulas', *J. Hydrol. Eng*, vol. 24, no. 5, p. 05019006.

Vishnu, S, Francis, PA, Shenoi, SC & Ramakrishna, SSVS 2018, 'On the relationship between the Pacific Decadal Oscillation and monsoon depressions over the Bay of Bengal', *Atmospheric Science Letters*, vol. 19, no. 7, p. e825.

Walsh, KJE, Bettio, L, Power, S, Fawcett, R & Pahalad, J 2001, 'Extended seasonal prediction of precipitation in Fiji', *Australian Meteorological Magazine*, vol. 50, no. 3, pp. 195-203.

Wang, C & Wang, X 2013, 'Classifying El Niño Modoki I and II by different impacts on rainfall in southern China and typhoon tracks', *Journal of Climate*, vol. 26, no. 4, pp. 1322-38.

Wang, W & Wells, MT 2000, 'Model Selection and Semiparametric Inference for Bivariate Failure-Time Data', *Journal of the American statistical Association*, vol. 95, no. 449, pp. 62-72.

Wang, W, Dong, Z, Zhu, F, Cao, Q, Chen, J & Yu, X 2018, 'A Stochastic Simulation Model for Monthly River Flow in Dry Season', *Water*, vol. 10, no. 11, p. 1654.

Wara, MW, Ravelo, AC & Delaney, ML 2005, 'Permanent El Niño-like conditions during the Pliocene warm period', *Science*, vol. 309, no. 5735, pp. 758-61.

White, H 1982, 'Maximum likelihood estimation of misspecified models', *Econometrica: Journal of the Econometric Society*, pp. 1-25.

Willmott, CJ 1982, 'Some comments on the evaluation of model performance', *Bulletin of the American Meteorological Society*, vol. 63, no. 11, pp. 1309-13.

Willmott, CJ, Robeson, SM & Matsuura, K 2012, 'A refined index of model performance', *International Journal of Climatology*, vol. 32, no. 13, pp. 2088-94.

Wolter, K & Timlin, MS 1998, 'Measuring the strength of ENSO events: How does 1997/98 rank?', *Weather*, vol. 53, no. 9, pp. 315-24.

Wong, G, Lambert, M, Leonard, M & Metcalfe, A 2009, 'Drought analysis using trivariate copulas conditional on climatic states', *Journal of Hydrologic Engineering*, vol. 15, no. 2, pp. 129-41.

Yan, J 2007, 'Enjoy the joy of copulas: With a package copula', *Journal of Statistical Software*, vol. 21, no. 4, pp. 1-21.

Yan, J, Kojadinovic, I & Yan, MJ 2010, 'Package ‘copula’', *Version 0.9-7, May*, vol. 28, p. 2010.

Yuan, C & Yamagata, T 2015, 'Impacts of IOD, ENSO and ENSO Modoki on the Australian Winter Wheat Yields in Recent Decades', *Sci Rep*, vol. 5, p. 17252.

Yuan, C & Yamagata, T 2015, 'Impacts of IOD, ENSO and ENSO Modoki on the Australian winter wheat yields in recent decades', *Scientific reports*, vol. 5, p. 17252.

Yuan, Y & Yang, S 2012, 'Impacts of different types of El Niño on the East Asian climate: Focus on ENSO cycles', *Journal of Climate*, vol. 25, no. 21, pp. 7702-22.

Zhang, L, Yang, BY, Guo, AH, Huang, DP & Huo, ZG 2018, 'Multivariate probabilistic estimates of heat stress for rice across China', *Stochastic Environmental Research and Risk Assessment*, vol. 32, no. 11, pp. 3137-50.

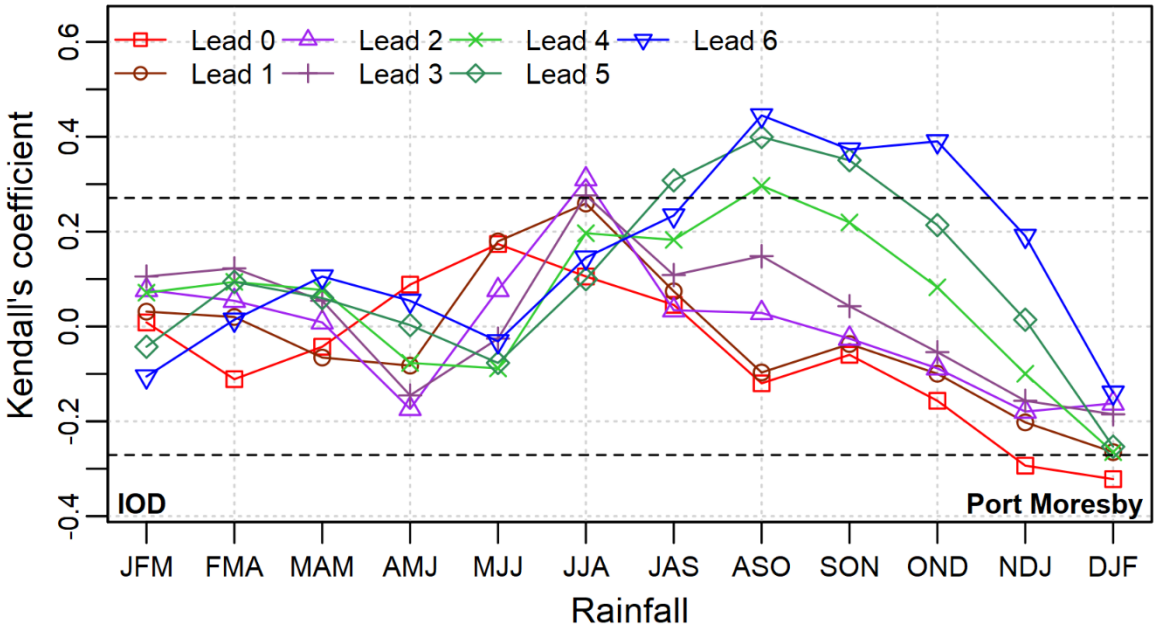
Zhang, Y, Wallace, JM & Battisti, DS 1997, 'ENSO-like Interdecadal Variability: 1900-93', *Journal of Climate*, vol. 10, pp. 1004-20.

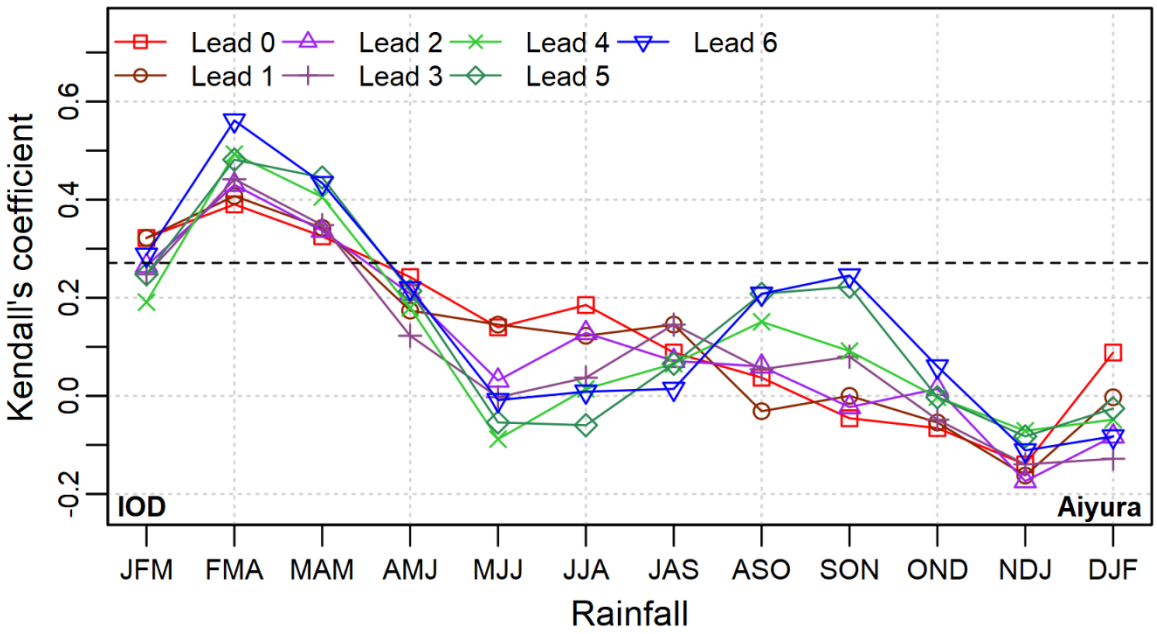
Zhao, M, Zhang, H & Dharssi, I 2019, 'On the soil moisture memory and influence on coupled seasonal forecasts over Australia', *Climate Dynamics*, pp. 1-25.

APPENDICES

The highest positive correlation for Aiyura was between FMA rainfall and IOD lead-6 (May-June-July) with 0.56, followed by MAM rainfall and IOD lead-5 (JAS) trailed by JFM and IOD lead-0 (OND) on 0.45 and 0.32 respectively. Overall the IOD had a significant correlation with Aiyura rainfall between the three-rainfall period mentioned, and the correlations were below the signature line between AMJ through to DJF. Whereas in Port Moresby the rainfall periods JJA, JAS, ASO, SON, and OND have a significant positive correlation with IOD (lead-2, 5, 6, 6, 6) with ASO rainfall showing a reliable correlation between IOD lead-6 (NDJ) with 0.45. Furthermore, the strong correlation between IOD and Port Moresby rainfall occurs between the August to October period. These months are generally low rainfall months, and also the IOD usually intensifies around August to October and decay rapidly when the monsoon arrives which justifies the correlation (Yuan, C. & Yamagata, T. 2015).

The total JFM, FMA and NDJ rainfall for Ramu displayed strong correlation of 0.43, 0.36 and 0.30 with DMI lead-0 including DMI lead-6 with the SON rainfall with a correlation of 0.28 while all the other remaining three months rainfall period had none statistical significance with the climate mode index IOD (Dipole Mode Index (DMI)). Highest correlation for Dami is observed between the MAM rainfall and IOD lead-2 (OND) with 0.43, followed by the correlation between IOD lead-6 and lead-5 for AMJ and ASO rainfall with 0.35 and 0.30 respectively, whilst the weak correlations OND-DJF period is expected since IOD influence in the southern hemisphere decreases around end of spring when the monsoon arrives.


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
-0.4
IOD
Port Moresby
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall


Lead 0
Lead 1
Lead 2
Lead 3
Lead 4
Lead 5
Lead 6
Kendall's coefficient
0.6
0.4
0.2
0.0
-0.2
IOD
Aiyura
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND
NDJ
DJF
Rainfall

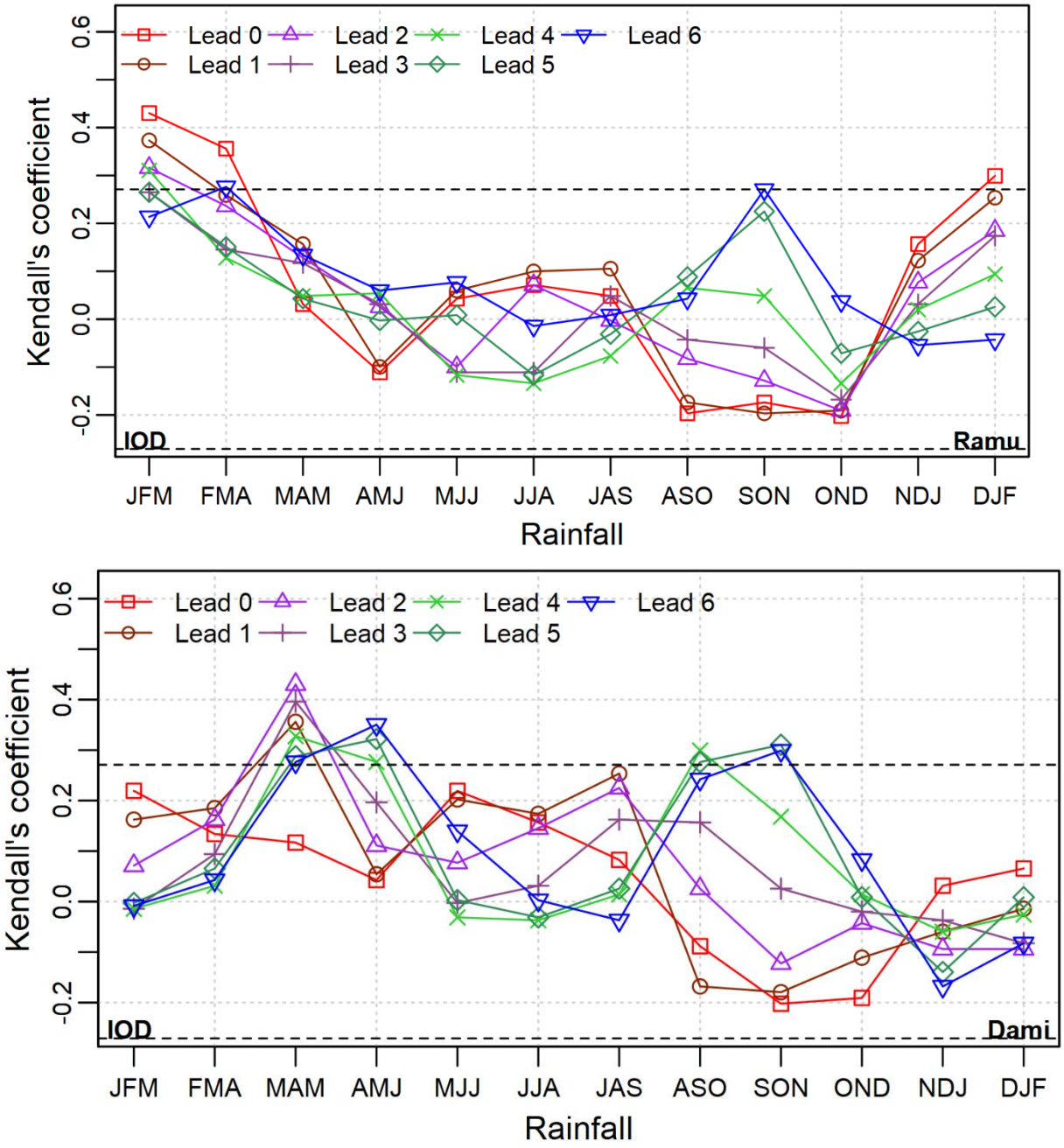


Figure 28: Ranked based Kendall's Tau correlation between the DMI and the rainfall for Port Moresby, Aiyura, Ramu and Dami. (Horizontal dotted lines indicate significant at 5%).