COPULA-BASED STATISTICAL MODELLING OF SYNOPTIC-SCALE CLIMATE INDICES FOR QUANTIFYING AND MANAGING AGRICULTURAL RISKS IN AUSTRALIA

A thesis submitted by

Thong Huy Nguyen

BEng (Mapping & Surveying)

MSc (Spatial Science Technology)

For the award of

Doctor of Philosophy

2018

Abstract

Australia is an agricultural nation characterised by one of the most naturally diverse climates in the world, which translates into significant sources of risk for agricultural production and subsequent farm revenues. Extreme climatic events have been significantly affecting large parts of Australia in recent decades, contributing to an increase in the vulnerability of crops, and leading to subsequent higher risk to a large number of agricultural producers. However, attempts at better managing climate-related risks in the agricultural sector have confronted many challenges.

First, crop insurance products, including classical claim-based and index-based insurance, are among the financial implements that allow exposed individuals to pool resources to spread their risk. The classical claim-based insurance indemnifies according to a claim of crop loss from the insured customer, and so can easily manage idiosyncratic risk, which is the case where the loss occurs independently. Nevertheless, the existence of systemic weather risk (covariate risk), which is the spread of extreme events over locations and times (*e.g.*, droughts and floods), has been identified as the main reason for the failure of private insurance markets, such as the classical multi-peril crop insurance, for agricultural crops. The index-based insurance is appropriate to handle systemic but not idiosyncratic risk. The indemnity payments of the index-based insurance are triggered by a predefined threshold of an index (*e.g.*, rainfall), which is related to such losses. Since the covariate nature of a climatic event, it sanctions the insurers to predict losses and ascertain indemnifications for a huge number of insured customers across a wide geographical area. However, basis risk, which is related to the strength of the relationship between the predefined indices used to estimate the average loss by the insured community and the actual loss of insured assets by an individual, is a major barrier that hinders uptake of the index-based insurance. Clearly, the high basis risk, which is a weak relationship between the index and loss, destroys the willingness of potential customers to purchase this insurance product.

Second, the impact of multiple synoptic-scale climate mode indices (*e.g.*, Southern Oscillation Index (SOI) and Indian Ocean Index (IOD)) on precipitation and crop yield is not identical in different spatial locations and at different times or seasons across the Australian continent since the influence of large-scale climate

heterogeneous over the different regions. The occurrence, role, and amplitude of synoptic-scale climate modes contributing to the variability of seasonal crop production have shifted in recent decades. These variables generally complicate the climate and crop yield relationship that cannot be captured by traditional modelling and analysis approaches commonly found in published agronomic literature such as linear regression. In addition, the traditional linear analysis is not able to model the nonlinear and asymmetric interdependence between extreme insurance losses, which may occur in the case of systemic risk. Relying on the linear method may lead to the problem that different behaviour may be observed from joint distributions, particularly in the upper and lower regions, with the same correlation coefficient. As a result, the likelihood of extreme insurance losses can be underestimated or overestimated that lead to inaccuracies in the pricing of insurance policies. Another alternative is the use of the multivariate normal distribution, where the joint distribution is uniquely defined using the marginal distributions of variables and their correlation matrix. However, phenomena are not always normally distributed in practice.

It is therefore important to develop new, scientifically verified, strategic measures to solve the challenges as mentioned above in order to support mitigating the influences of the climate-related risk in the agricultural sector. Copulas provide an advanced statistical approach to model the joint distribution of multivariate random variables. This technique allows estimating the marginal distributions of individual variables independently with their dependence structures. It is clear that the copula method is superior to the conventional linear regression since it does not require variables have to be normally distributed and their correlation can be either linear or non-linear.

This doctoral thesis therefore adopts the advanced copula technique within a statistical modelling framework that aims to model: (1) The compound influence of synoptic-scale climate indices (*i.e.*, SOI and IOD) and climate variables (*i.e.*, precipitation) to develop a probabilistic precipitation forecasting system where the integrated role of different factors that govern precipitation dynamics are considered; (2) The compound influence of synoptic-scale climate indices on wheat yield; (3) The scholastic interdependencies of systemic weather risks where potential adaptation strategies are evaluated accordingly; and (4) The risk-reduction efficiencies of geographical diversifications in wheat farming portfolio optimisation. The study areas

are Australia's agro-ecological (*i.e.*, wheat belt) zones where major seasonal wheat and other cereal crops are grown. The results from the first and second objectives can be used for not only forecasting purposes but also understanding the basis risk in the case of pricing climate index-based insurance products. The third and fourth objectives assess the interactions of drought events across different locations and in different seasons and feasible adaptation tools. The findings of these studies can provide useful information for decision-makers in the agricultural sector.

The first study found the significant relationship between SOI, IOD, and precipitation. The results suggest that spring precipitation in Australia, except for the western part, can be probabilistically forecasted three months ahead. It is more interesting that the combination of SOI and IOD as the predictors will improve the performance of the forecast model. Similarly, the second study indicated that the large-scale climate indices could provide knowledge of wheat crops up to six months in advance. However, it is noted that the influence of different climate indices varies over locations and times. Furthermore, the findings derived from the third study demonstrated the spatio-temporally stochastic dependence of the drought events. The results also prove that time diversification is potentially more effective in reducing the systemic weather risk compared to spatially diversifying strategy. Finally, the fourth objective revealed that wheat-farming portfolio could be effectively optimised through the geographical diversification.

The outcomes of this study will lead to the new application of advanced statistical tools that provide a better understanding of the compound influence of synoptic-scale climatic conditions on seasonal precipitation, and therefore on wheat crops in key regions over the Australian continent. Furthermore, a comprehensive analysis of systemic weather risks performed through advanced copula-statistical models can help improve and develop novel agricultural adaptation strategies in not only the selected study region but also globally, where climate extreme events pose a serious threat to the sustainability and survival of the agricultural industry. Finally, the evaluation of the effectiveness of diversification strategies implemented in this study reveals new evidence on whether the risk pooling methods could potentially mitigate climate risks for the agricultural sector and subsequently, help farmers in prior preparation for uncertain climatic events.

Certification of Thesis

This thesis is the work of ***Thong Huy Nguyen*** except where otherwise acknowledged, with the majority of the authorship of the papers presented as a Thesis by Publication undertaken by the Student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Thong Huy Nguyen	*20 July 2018*
PhD Candidate	Date

Endorsement

Dr Ravinesh C. Deo	*20 July 2018*
Principal Supervisor	Date

Assoc Prof Shahbaz Mushtaq	*20 July 2018*
Associate Supervisor	Date

Prof Shahjahan Khan	*20 July 2018*
Associate Supervisor	Date

Student and supervisors' signatures of endorsement are held at USQ.

Statements of contributions

The articles produced from this study were a joint contribution of the student and supported by the supervisors. The details of the scientific contribution of each author in journal publications and a book chapter are provided as follows.

- **Article I:** ***Thong Nguyen-Huy***, Ravinesh C. Deo, Duc-Anh An-Vo, Shahbaz Mushtaq, and Shahjahan Khan. "Copula-statistical precipitation forecasting model in Australia's agro-ecological zones". *Agricultural Water Management*, 191 (2017): 153-172. [***Impact Factor: 2.848, SNIP: 1.814, Scopus Ranked Q1, 93rd percentile in Water Sc & Technology***].

  DOI: https://doi.org/10.1016/j.agwat.2017.06.010

  The overall contributions in this paper are: Thong Nguyen-Huy 70%, Ravinesh C. Deo 15%, Duc-Anh An-Vo 5%, Shahbaz Mushtaq 5%, and Shahjahan Khan 5%.

Author	**Tasks Performed**
Thong Nguyen-Huy (PhD Candidate)	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo (Principle Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Duc-Anh An-Vo	Comments on the final manuscript.
Shahbaz Mushtaq (Associated Supervisor)	Supervised and provided comments on the manuscript.
Shahjahan Khan (Associated Supervisor)	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.

- **Article II:** ***Thong Nguyen-Huy***, Ravinesh C. Deo, Shahbaz Mushtaq, Duc-Anh An-Vo, and Shahjahan Khan. "Modelling the joint influence of multiple

climate mode indices on Australian wheat yield using a vine copula-based approach." *European Journal of Agronomy*, 98 (2018): 65-81. [***Impact Factor: 3.757, SNIP: 1.828, Scopus Rated Q1, 94th percentile in Agronomy & Crop Sc***.].

DOI: https://doi.org/10.1016/j.eja.2018.05.006

The overall contributions in this paper are: Thong Nguyen-Huy 70%, Ravinesh C. Deo 15%, Shahbaz Mushtaq 5%, Duc-Anh An-Vo 5% and Shahjahan Khan 5%.

Author	**Tasks Performed**
Thong Nguyen-Huy (PhD Candidate)	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo (Principle Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Shahbaz Mushtaq (Associated Supervisor)	Supervised and provided comments on the manuscript.
Duc-Anh An-Vo	Comments on the final manuscript.
Shahjahan Khan (Associated Supervisor)	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.

- **Article III:** ***Thong Nguyen-Huy***, Ravinesh C. Deo, Shahbaz Mushtaq, Jarrod Kath and Shahjahan Khan. "Copula statistical models for analyzing stochastic dependencies of systemic drought risk". *Stochastic Environmental Research and Risk Assessment* (in review). [***Impact Factor = 2.629, Scopus Rated Q1, 85th percentile in Safety, Risk, Reliability & Quality***].

  Ref. No.: SERR-D-18-00328

The overall contributions in this paper are: Thong Nguyen-Huy 70%, Ravinesh C. Deo 15%, Shahbaz Mushtaq 5%, Jarrod Kath 5%, and Shahjahan Khan 5%.

Author	**Tasks Performed**
Thong Nguyen-Huy (PhD Candidate)	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo (Principle Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Shahbaz Mushtaq (Associated Supervisor)	Supervised and provided comments on the manuscript.
Jarrod Kath	Discussion and comments on the manuscript.
Shahjahan Khan (Associated Supervisor)	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.

- **Article IV:** ***Thong Nguyen-Huy***, Ravinesh C. Deo, Shahbaz Mushtaq, Jarrod Kath and Shahjahan Khan. "Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management". *Weather and Climate Extremes*. [***Impact Factor = 4.21, SNIP: 2.428, Scopus Rated Q1, 98th percentile in Geography, Planning & Development***].

  DOI: https://doi.org/10.1016/j.wace.2018.07.002

  The overall contributions in this paper are: Thong Nguyen-Huy 70%, Ravinesh C. Deo 15%, Shahbaz Mushtaq 5%, Jarrod Kath 5%, and Shahjahan Khan 5%.

Author	**Tasks Performed**
Thong Nguyen-Huy (PhD Candidate)	Method development, programming, data analysis, preparation of tables and figures,

	compilation, writing, and revision of the manuscript.
Ravinesh C. Deo (Principle Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Shahbaz Mushtaq (Associated Supervisor)	Supervised and provided comments on the manuscript.
Jarrod Kath	Discussion and comments on the manuscript.
Shahjahan Khan (Associated Supervisor)	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.

- **Book chapter:** ***Thong Nguyen-Huy***, Ravinesh C. Deo, Shahbaz Mushtaq and Shahjahan Khan. "Probabilistic seasonal rainfall forecasts using semi-parametric D-vine copula-based quantile regression". In: Samui P, Bui DT, Chakraborty S, Deo RC. (Eds.) (2018-in progress). *Handbook of Probabilistic Models*. New York, NY: Elsevier. Publication Date January 2019.

  The overall contributions in this paper are: Thong Nguyen-Huy 75%, Ravinesh C. Deo 15%, Shahbaz Mushtaq 5%, and Shahjahan Khan 5%.

Author	**Tasks Performed**
Thong Nguyen-Huy (PhD Candidate)	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo (Principle Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Shahbaz Mushtaq (Associated Supervisor)	Supervised and provided comments on the manuscript.

Shahjahan Khan (Associated Supervisor)	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.

Acknowledgements

Please accept my great appreciation to these individuals and organisations for their whole-hearted assistance, scientific guidance, valuable and useful advice, and for generously sharing their expertise. It is undoubted that without their support, motivation and encouragement it would not have been possible for me to complete the present study. In particular, I would like to express my gratitude and sincere appreciation to:

- *Dr* ***Ravinesh C. Deo***, my principal supervisor, for his motivation and guidance, and for patiently guiding me while I learn the ropes of advanced statistical modelling and publishing in high-quality journals. His quick response and easy approach helped me to carry out my study within schedule. Obviously, several sentences are not enough to acknowledge his contribution to my achievement.
- *Associate Professor* ***Shahbaz Mushtaq***, my associate supervisor, to whom I am grateful for sharing his valuable suggestions, plentiful comments and criticisms during different stages of my research. His critical analysis helped me to figure out practical applications and direction of future research. He played a vital role in building up my research foundation and contributing additional research funding.
- *Professor* ***Shahjahan Khan***, my associate supervisor, for his scientific advice and expertise on statistical analysis, interpretation of results and useful comments in the journal articles as well as in the thesis.
- *Dr* ***Jarrod Kath*** for his contribution and comments in the research and journal articles.
- *Dr* ***Barbara Harmes*****;** the key representative of the *English Angels Program* within Open Access College for her special support in proofreading and improving my writing skills.
- *The* ***University of Southern Queensland*** through the USQ Postgraduate Research Scholarships for providing me with a full scholarship to pursue my PhD degree in Australia.
- *Professor* ***Geoff Cockfield***: who approved USQ's Strategic Research Funds (SRF) for 2015-2016 to support living costs.

- *Advanced Data Analytics:* Environmental Modelling and Simulation Group, led by the Principal Supervisor for contributions and my colleagues in sharing ideas.
- ***School of Agricultural, Computational and Environmental Sciences*** and ***Drought and Climate Adaptation*** (DCAP) Projects (Producing Enhanced Crop Insurance Systems and Associated Financial Decision Support Tools) for 2017-2018 to support living costs.
- *The* ***International Centre for Applied Climate Sciences*** for providing my work station and space, and for contributing additional research funds.
- *My wife* ***Duy Ngoc*** and *daughter* ***Anh Bao***, my biggest motivations. In particular, I thank my wife ***Duy*** for her support and the sacrifices she has made so that I could spend many hours a day over many years to pursue this research.
- *My* ***parents*** and ***friends***, who have always supported and encouraged me in difficult times and hard situations.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Changing climate, as an uncertain variable, is one of the greatest sources of risk to the farming community, significantly influencing inter-annual agricultural production and consequently received revenues. In particular, recent extreme climatic conditions have been associated with enormous losses in agricultural production, in both developed and developing countries (Barriopedro et al. 2011; Coumou and Rahmstorf 2012; Herold et al. 2018). According to the FAO (2015), extreme climate-related events have been identified as the reason for approximately one-quarter of the reduced agricultural production in developing nations. Furthermore, Lesk et al. (2016) state that extreme drought and heat events have caused a significant decline in national cereal production of about 9 – 10% worldwide during 1964 – 2007. However, these authors also emphasise that developed agricultural nations have 8 – 11% more losses in comparison with developing countries. It is clear that extreme weather disasters may affect the agricultural sector even more severely in countries that have high crop yields and advanced agri-technologies. Subsequently, profitability of growing crops received by farmers can vary substantially across farming zones and over years, postulating a financial protection against weather variations (Odening and Shen 2014).

Notwithstanding a matter of urgency to mitigate climate-related risks, a number of challenges have trodden in the development of efficient adaptation instruments used to transfer climate-related risks in the agricultural sector. Crop insurance, for example, is one of the financial instruments, which allows exposed individuals to pool resources to spread their risk. Nevertheless, the existence of systemic weather risk (covariate risk), which is the spread of extreme events over locations and times (*e.g.*, drought and floods), has been identified as the main reason for the failure of private insurance markets, such as the classical multi-peril crop insurance, for agricultural crops (Duncan and Myers 2000; Miranda and Glauber 1997; Xu et al. 2010).

Classical claim-based insurance indemnifies according to a claim of crop loss from the insured customer, and so can easily manage idiosyncratic risk, which is the case where the loss occurs independently. However, in the situation of systemic risk, insurance companies often encounter with a large number of simultaneous claims, and thus pressing their ability to solve all claims properly. In that case, it becomes a challenge to authenticate all claims in time before verifiable evidence of damage and loss has vanished. Further, insurers can face ruin when confronted with massive indemnity payments exceeding their solvability (Okhrin et al. 2013).

Index-based insurance is appropriate to handle systemic but not idiosyncratic risk. The indemnity payments of the index-based insurance are triggered by a predefined threshold of an index (*e.g.*, rainfall), which is related to such losses. Since the covariate nature of a climatic event, it sanctions the insurers to predict losses and ascertain indemnifications for a huge number of insured customers across a wide geographical area. However, basis risk is a major barrier that hinders uptake of the index-based insurance (Barnett 2004). It is defined as the situation that an insured customer sustains no loss but receiving indemnification or vice versa. Basis risk is related to the strength of the relationship between the predefined index used to estimate the average loss by the insured community and the actual loss of insured assets by an individual. Clearly, the high basis risk, which is a weak relationship between the index and loss, destroys the willingness of potential customers to purchase this insurance product.

Climate variables such as precipitation and temperature have been employed to analyse crop yield variations and forecast crop yield worldwide, often at the shire level (Asseng et al. 2011; Bannayan et al. ; Palosuo et al. 2011). Averaged or gridded data derived from a number of meteorological stations may be used to examine the climate-yield relationship at broader scale levels (Lobell and Burke 2010; Lobell et al. 2007; Revadekar and Preethi 2012). However, such interpolated data relies on the density of the meteorological station network which may have questions related to the data quality, the continuous collection of weather data and the convenience issues related to regular monitoring of data acquisition systems (Harris et al. 2014; Schepen et al. 2012). For these reasons, it is essential to understand the requirements of the use of larger scale variables (*e.g.*, climate modes) to forecast crop yield large-scale

regions. This can be performed through a teleconnection between synoptic-scale climate indices, their variables such as precipitation, and crop yield.

Previous published studies have identified a complicated nonlinear relationship between climate variables and crop yield (Schlenker and Roberts 2006; Schlenker and Roberts 2009). However, conventional statistical methods such as linear regression analysis and correlations have failed to obtain a complete understanding of the complex impact of climate variables on crop yields (Bokusheva 2011). This is attributable to the spread and the low coefficient of determination between seasonal data (Biscoe and Gallagher 1977; French and Schultz 1984). In addition, typical models used classical families of bivariate distribution (*e.g.*, normal, lognormal, gamma, and extreme value) to describe the pairwise dependency between two variables based on the assumption that they belonged to the same parametric family of univariate distributions (Genest and Favre 2007). As a result, the likelihood of extreme insurance losses can be underestimated or overestimated, which lead to inaccuracies in the pricing of insurance policies.

Multivariate copula functions, introduced by Sklar (1959), have recently become powerful tools for understanding relationships among random variables (Frees and Valdez 1998). By analysing the marginal distribution of each variable and extracting its parameters, copula functions can sanction one to link the univariate marginal (of single variables) to their full multivariate (*e.g.*, bi or trivariate) distributions. Therefore, the copula approach, compared to the conventional linear correlation-based model, is a more robust methodological framework for modelling the dependence structure, in particular the tail dependencies, of a set of multivariate distributions. These are important for quantifying risks due to extreme climatic events. Because of their usefulness in modelling joint behaviours of two or more inter-related variables, in recent decades, copula functions have been extensively applied in many fields involving hydrology (Genest et al. 2007; Salvadori and De Michele 2004), drought studies (Kao and Govindaraju 2010; Shiau and Modarres 2009; Shiau 2006) and finance and risk insurance (da Costa Dias 2004; Frees and Valdez 2008; Zhu et al. 2008). However, applications of the copula theory in the agricultural sector are very limited in the published literature. In particular, the copula-based model of crop productivity using synoptic-scale climate indices has not yet been performed in Australia's wheat growing regions.

A better understanding of the compound impacts of synoptic-scale climate indices, in particular the co-occurrence of extreme events, on precipitation and wheat crops, can provide an accurate forecast of both precipitation and wheat yield. In particular, using a novel statistical copula-based model to capture fully the dependence structure between variables, we may be able to implement probabilistic forecasts of wheat yields at different quantile levels with sufficient time ahead. It can therefore provide useful information to agricultural managers in developing potential adaptation strategies that can be implemented to reduce climate-related risks. Broadly speaking, the developed copula model can provide better support for food security and climate index-based insurance while it is possible to extend this support to other fields.

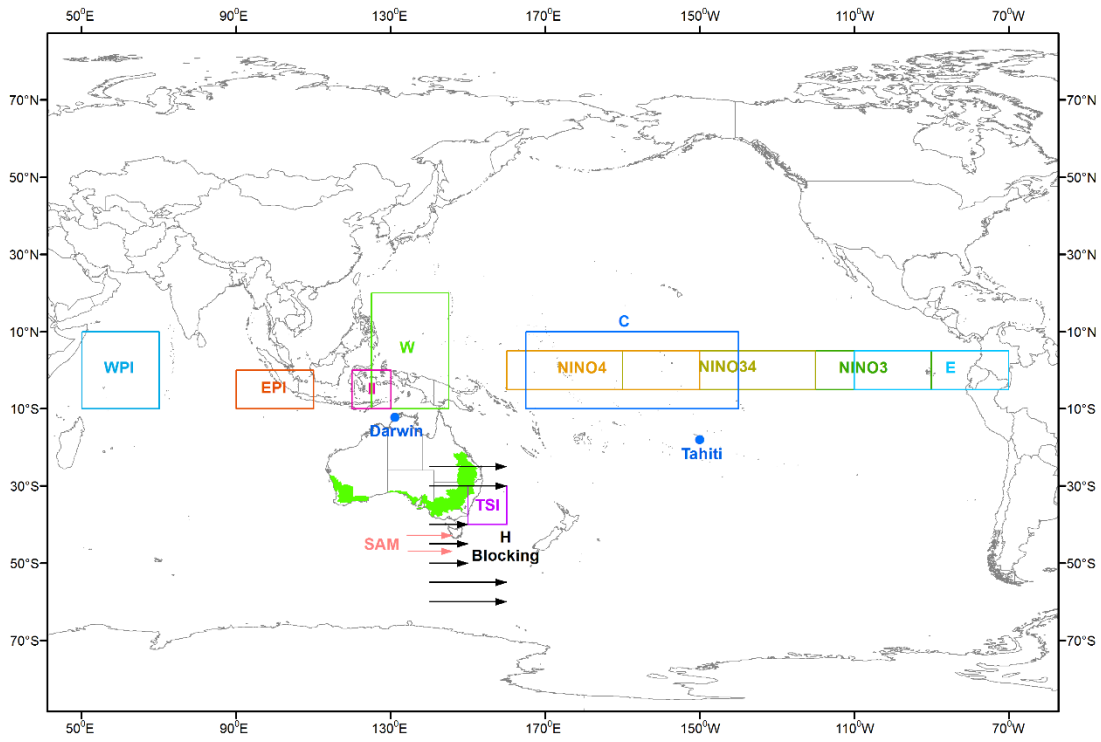
1.2 Statement of the Problem

Motived by the reasons mentioned above, this doctoral thesis focuses on analysing the interdependencies between synoptic-scale climate indices, precipitations and yield of wheat crops in Australia using the copula approach. The thesis then assessed the spread of climate events over space and time (seasons) and evaluated the effectiveness of potential adaptation instruments. However, it was essential to first identify a number of relevant research problems. While these established questions are extensively represented in the following sections, the research gaps have been also identified at the same time based on published literature.

1.2.1 The importance of understanding the influence of synoptic-scale climate indices on precipitation

As mentioned above, wheat crop in Australia is mostly cultivated in drylands, so precipitation plays a vital role in the year-to-year success of crops. This is evidenced by the fact that the association between precipitation and wheat yield has been extensively investigated and published in literature (Hochman et al. 2009; Nicholls 1997; Sadras et al. 2002; Stephens and Lyons 1998). Simultaneously, a number of different large-scale climate indices have been recognised as the main responses to the precipitation pattern in Australia, depending on the regions and the seasons (Ashok et al. 2003; Min et al. 2013; Schepen et al. 2012; Taschetto and England 2009). It is

therefore not surprising that better understanding of how synoptic-scale climate indices modulate precipitation variability is important in the agricultural sector.



Source: Nguyen-Huy et al. (2018)

Figure 1. Main climate drivers of precipitation and agricultural yield variability in Australia.

Anomalously periodic fluctuations in winds and sea surface temperatures over the tropical Pacific region, popularly known as El Niño Southern Oscillation (ENSO) phenomenon have been identified as the main factors which affect the Australian precipitation pattern (Nicholls et al. 1996; Suppiah 2004). In particular, the associations between opposite phases of ENSO and precipitation variability have been investigated since the early ‘80s (McBride and Nicholls 1983; Nicholls 1983; Pittock 1975). In general, La Niña phases may bring more precipitation often in eastern Australia meanwhile El Niño phases are often associated with broad-scale drought events (Yuan and Yamagata 2015). It is clear that the ENSO plays an important role in the precipitation variability over much of Australia; however, its impact differs due to the difference in times and locations. Risbey et al. (2009) studied the simultaneous correlation between ENSO and precipitation patterns for the four standard seasons in Australia. The findings pointed out that while the highest influence of ENSO is on the

eastern and north-eastern areas, particularly in winter and spring, precipitation in most regions of Australia have a significant correlation with ENSO in at least one season.

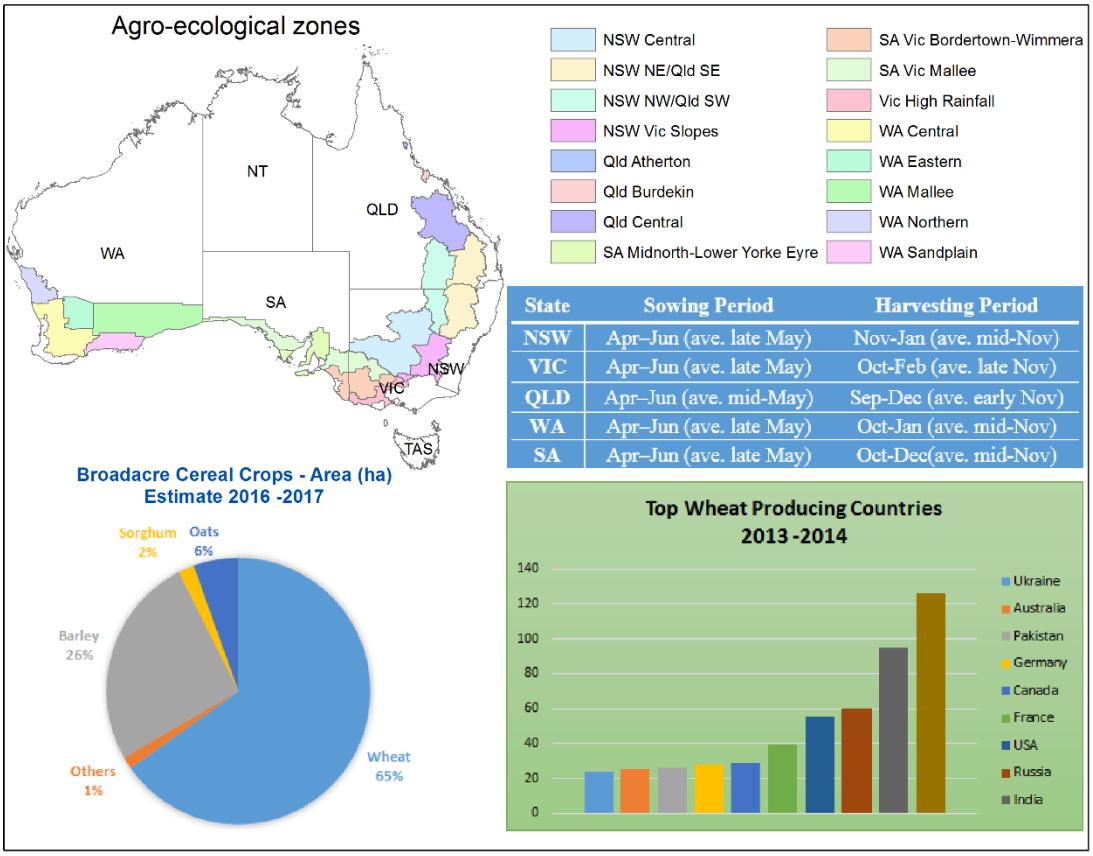
Recent studies found in literature have also paid much attention on the role of the Indian Ocean Dipole (IOD), known as Indian Niño, in the variability of precipitation worldwide (Chan et al. 2008; Zubair et al. 2003). Following these analyses, a number of studies have investigated the influences of IOD on precipitation in Australia (England et al. 2006; Ummenhofer et al. 2009). Like ENSO, the influence of IOD on precipitation also varies according to different phases, locations and times. For example, negative IOD phases are associated with above average precipitation over southern Australia (Pook et al. 2006). According to Ashok et al. (2003), IOD has significant negative partial correlations with precipitation over the western and southern regions of Australia. Furthermore, Cai et al. (2009) also found that the changes of IOD potentially account for much of the observed austral winter and spring precipitation reduction since 1950.

Other synoptic-scale climate indices have been demonstrated to have an impact on the variability of precipitation in Australia. These are Southern Annular Mode (SAM) (Hendon et al. 2007; Meneghini et al. 2007), Madden–Julian oscillation (MJO) (Barlow et al. 2005; Donald et al. 2006; Pohl et al. 2007; Wheeler et al. 2009) and Quasi-biennial Oscillation (QBO) (Phelps and Wang 2014; Seo et al. 2013; Williams and Stone 2009). For example, the greatest precipitation impact of the MJO occurs in northern Australia in summer, although in every season precipitation impacts of various magnitude were found in most locations, associated with corresponding circulation anomalies (Wheeler et al. 2009).

From published literature, there has been a lack of research effort on the joint influences of multiple climate drivers on precipitation variability in Australia. Traditional analyses of precipitation variability related to large-scale climate mode indices as mentioned above tend to focus on single variables such as ENSO or IOD. However, in many practical situations, methods based on univariate extremes are insufficient to identify and detect all interactions that lead to a significant influence (Fischer and Knutti 2013; Leonard et al. 2014). Therefore, any management of drought-related risk on wheat production requires knowledge of compound impact triggered by the joint occurrence of extreme climatic events.

1.2.2 Why study models for wheat crop?

Wheat is the main cereal crop in Australia, accounting for 56% of Australian grain production and 10 – 15 % of global wheat exports (Yuan and Yamagata 2015). The majority of wheat crops are grown in rain-fed conditions distributed across the Agro-ecological zones (also known as the Australia's wheat belt) (Fig. 1). This means that climate and weather variables such as precipitation and temperature are the key factors directly affecting crop production. In fact, the timing of wheat planting, fertiliser applications and harvesting dates vary from region to region and year to year, depending on the precipitation pattern, and daytime temperatures (Fig. 1).



Source: Author. Data collected from Australian Bureau of Statistics

Figure 2. Information on wheat crop farming in Australia.

However, as an agricultural nation, Australia suffers one of the world's most variable climate conditions (Portmann et al. 2010; Turner 2004). According to Best et al. (2007), Australia has a large mean inter-annual variability in climatic conditions by about 15-18% compared to other major agricultural countries. Furthermore,

extreme weather events such as droughts, floods, hails, and extreme temperatures have often caused significant impacts on the wheat crop across the whole continent. For example, very dry conditions that occurred in the Western Australian wheat belt in 2010 resulted in a 43% reduction of wheat and other winter crop production compared to the previous season (ABARES 2011b). A projected decline in precipitation and higher temperatures by 2050 could lead to about 30% losses in grain production including wheat (Gool 2009). Therefore, there a need of improvements in the understanding of climate-yield relationship in order to minimise the climate-related impact and support agricultural managers in strategy development and decision making.

1.2.3 How does a variable and changing climate affect wheat yield?

It is evident that large-scale climate mode indices have a strong impact on weather variables such as precipitation and temperature, so they may indirectly influence the crop yield variability. Globally, a number of studies have investigated the associations between large-scale climate drivers and variability of crop production at regional or national levels (Royce et al. 2011; Shuai et al. 2013; y Garcia et al. 2010), and at continental or global scales (Anderson et al. 2017; Ceglar et al. 2017; Gutierrez 2017; Iizumi et al. 2014). In Australia, ENSO has been commonly employed to analyse the inter-annual variability of wheat yield (Potgieter et al. 2002; Potgieter et al. 2005; Yuan and Yamagata 2015). In general, these results show that La Niña (El Niño) events are associated with the increase (decrease) of wheat yield. This pattern is the same with the climate-precipitation relationship.

In many cases, the use of large-scale climate indices in crop yield forecasts may be more favourable than using weather variables as predictors. Weather variables such as precipitation and temperature are commonly employed to predict crop production worldwide, often at the shire scale (Asseng et al. 2011; Palosuo et al. 2011). Gridded data interpolated from a number of meteorological stations may be used in broader levels (Lobell et al., 2007; Revadekar and Preethi, 2012). However, relying on data derived from weather station networks may lead to issues related to data quality, continuous collection, and regular monitoring of data acquisition systems (Harris et al. 2014; Schepen et al. 2012). By contrast, large-scale climate information is often measured at a small number of stations or by satellites that may provide

reliable data and reduce data uncertainties. For example, Southern Oscillation Index (SOI), an indicator of ENSO phenomenon, is acquired based on the difference of monthly sea-level pressures between Tahiti and Darwin, Australia (Fig. 2) (McBride and Nicholls 1983; Stone et al. 1996). In addition, the current development of climate models allows achieving advanced knowledge of large-scale climate mode indices such as ENSO from six months to one year ahead (Jin et al. 2008; Ludescher et al. 2013). Therefore, one may use ENSO to forecast crop production with sufficient time in advance. For example, information of the April–May ENSO can be potentially employed as an early forecasting tool for several seasonal crops, including wheat, in the following growing season (Potgieter et al. 2002). These facts suggest that large-scale climate modes are useful for studying the associations with crop yields in terms of large-scale study areas, long time lag, and reliable and available datasets.

However, in conjunction with the same issues mentioned above for precipitation forecast, the compound influences on crop yield when extreme events co-occur have been rarely explored in published literature. In fact, interactions between different synoptic-scale climate indices during the co-occurrence of extreme climate events may modulate the individual influence on weather variables (Li et al. 2016; Lim et al. 2016; Weller and Cai 2013). Min et al. (2013) reported that there were anomalously drier and hotter conditions occurring across north-eastern and southern coastal Australia during the co-occurrence of El Niño and positive IOD phases in the cold seasons, whereas wetter and cooler conditions appeared during the presence of both La Niña and negative IOD phases. However, minimal research effort has attempted to justify the concurrent impacts of multiple synoptic-scale climate indices on Australian crop yield (Jarvis et al. 2018; Yuan and Yamagata 2015), regardless of the co-occurrence of extreme climate events that may subsequently affect crop yield. For these reasons, it is emphasised that there is a need for comprehensive research on the impact of compound influences of different synoptic-scale climate indices on crop yield.

1.2.4 Systemic weather and climate risks and potential agricultural adaptation strategies

Climate variability can be seen as one of the greatest sources of risk affecting agricultural producers and their revenues even in countries with high crop yield and

advanced agri-technologies (Barriopedro et al. 2011; Coumou and Rahmstorf 2012). Lesk et al. (2016) has explained that drought and heat extreme events considerably decreased national cereal production by 9 – 10% worldwide during 1964 – 2007. These authors also found that developed countries have 8 – 11% more damage compared to that in developing countries. Clearly, the agricultural sector requires significant financial protection against climate variability (Odening and Shen 2014).

This problem can be solved through an efficient and affordable instrument, such as weather index-based insurance, for transferring systemic weather risks. However, the existence of systemic weather risk has been determined as the leading reason for the failure of private insurance markets for agricultural crops (Duncan and Myers 2000; Miranda and Glauber 1997). Systemic weather risk is defined as when a weather event such as drought occurs over a considerable area and affect a large number of farmers (Odening and Shen 2014; Xu et al. 2010). This means that many farmers are affected at the same time, resulting in a huge number of simultaneous insurance claims and subsequently leading to the bankruptcy problems for insurance companies.

High systemic weather risk, as mentioned above, is a major hindrance to feasible crop insurance and leads to the failure of an unsubsidised private insurance market (Skees and Barnett 1999; Vedenov and Barnett 2004). However, there are several possible implements allowing managing systemic risks such as reinsurance and weather derivatives (Musshoff et al. 2011; Skees et al. 2007). Alternatively, insurers may spatially diversify the systemic weather risk by extending its trading area (Okhrin et al. 2013; Xu et al. 2010). According to Odening and Shen (2014), the level of covariate risk depends on the size of the risk pool and thus it seems natural to reduce the systemic weather risk by increasing the regional dissemination of insurance products. For example, a drought event may be highly correlated within a small region but possibly independent at a broader scale. Furthermore, the impact of geographical basis risk on the hedging effectiveness of weather derivatives has been regularly highlighted in the published literature (Ritter et al. 2014; Woodard and Garcia 2008). Clearly, quantifying the dependence between weather events occurring concurrently across different locations is essential to the measurement of joint weather-related losses and the hedging effectiveness of weather derivatives that insurance companies may wish to sell to farmers.

1.2.5 Limitations of common statistical multivariate models

The use of a simple univariate approach can lead to severe underestimation of potential risks associated with the interdependencies between events (Favre et al. 2004). Therefore, better capturing the joint distribution between random variables (here, the compound impact of multiple synoptic-scale climate indices on precipitation or wheat yield) is a fundamental and prime part in any modelling study.

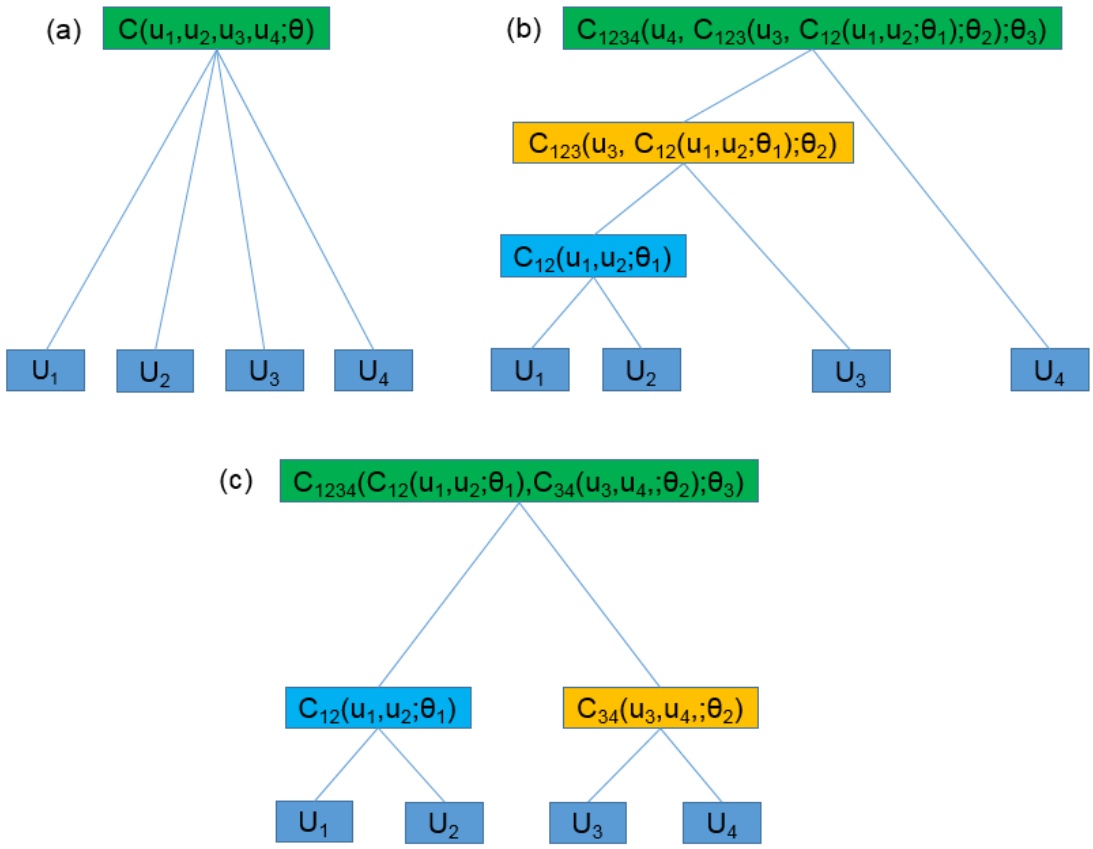
In practice, several types of models are available, for example to forecast wheat yield using synoptic-scale climate indices, including empirical models and eco-bio-physiological simulation (or process-based) models. The empirical models that do not rely on eco-bio-physiological equations may use statistical (Jarvis et al. 2018; Yuan and Yamagata 2015) or machine learning methods (Deo et al. 2017; Deo and Şahin 2016). In particular, statistical models deliver yield forecasts using the historical relationship between synoptic-scale climate indices and crop yields. Clearly, the main advantage of statistical models is that they do not require many crop parameters. Furthermore, they are generally easier to develop and more suitable for forecasting crop yield over a large scale (Matsumura et al. 2015). Finally, statistical models may be used to estimate the uncertainty in the simulation process which is often difficult to achieve in process-based models (Lobell et al. 2006).

However, it is worth noting that the statistical multivariate models applied in previous published studies (Jarvis et al. 2018; Yuan and Yamagata 2015) often assumed a linear relationship between synoptic-scale climate indices and crop yield, implying the joint distribution among these variables to be a normal distribution. It is undoubted that linear regression models are simple to build and can provide a quick conclusion on general trends (*i.e.*, the fitted straight line) between response variable given the values of the explanatory variables. However, the model specification may be strongly affected by outliers (*e.g.*, the occurrence of extreme events) resulting in a poor measure of actual dependencies between variables (Ghosh et al. 2011; Hassani 2016; Tang and Valdez 2009). Therefore, the assumption that the joint distribution between variables is normally distributed may not always hold in many practical cases and lead to an incorrect interpretation of findings.

1.2.6 Why should we use copulas in statistical models?

Copulas (Sklar 1959) deliver a better approach to handling dependence structures between random variables. In the copula technique, joint distributions of variables are modelled independently to the choice of the marginal distributions. Clearly, copula-based models can overcome the above shortcomings of traditional regression models. For example, variables in the model can be fitted parametrically using any hypothetical distribution (*e.g.*, Gamma or Weibull) or non-parametrically using a kernel density function. Another advantage of copula-based models relative to other statistical methods is that the data series can be of different lengths, which is useful to apply for practical data (Patton 2001; Zhang and Singh 2014).

In general, copulas can be categorised into families including, but not limited to, empirical, Archimedean, extreme value, elliptical, vine, and entropy copulas based on their construction. Here, we briefly describe the drawbacks of the elliptical and Archimedean copulas and the advantages of vine copulas due to their wide applications (Abdul Rauf and Zeephongsekul 2013; Brechmann et al. 2013; Fang and Madsen 2013; Grimaldi and Serinaldi 2006; Pham et al. 2016). Within elliptical families, the Gaussian copula is not able to model the tail dependencies while the Student's t copula can capture symmetric dependence in the upper and lower tails. Archimedean families can measure both symmetric and asymmetric dependence structures between variables but have some limitations. The symmetric Archimedean copulas (*e.g.*, Clayton, Gumbel) use a single parameter to model dependence structures between variables, which may not be appropriate (Fig. 3a). The asymmetric Archimedean copulas constructed by hierarchical structures assume the same copula functions for each level in construction where the parameters estimated for higher levels must be smaller than those for lower levels (Fig. 3b-c). Again, this assumption may not be reasonable in practice (Hao and Singh 2016; Zhang and Singh 2014).

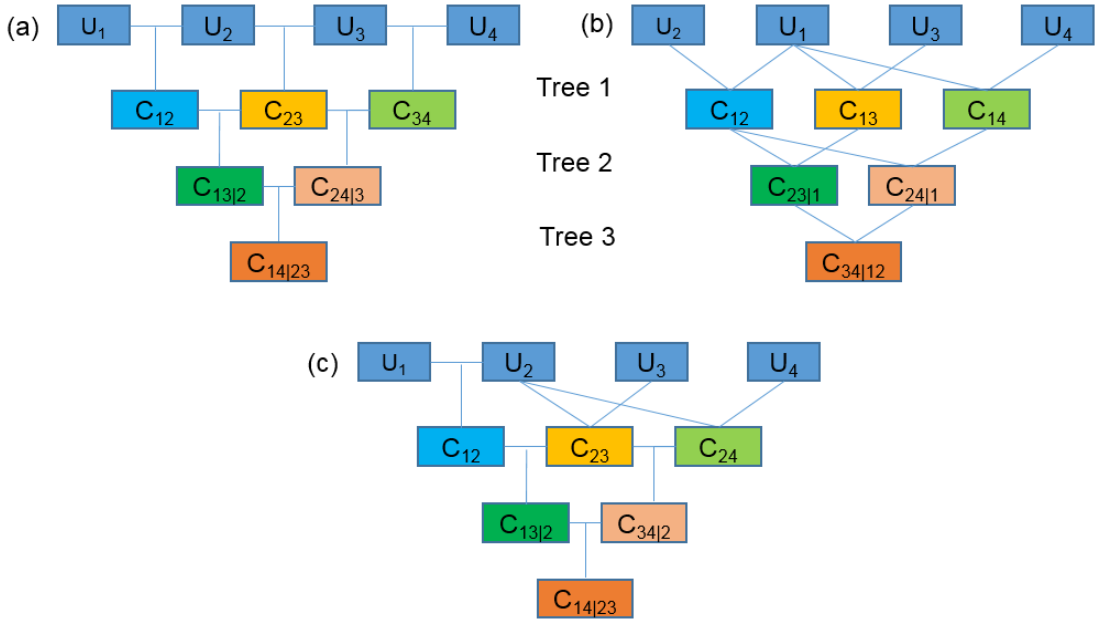


(Source: Author)

Figure 3. Four-dimensional Archimedean copulas constructed by symmetric (a), fully nested (b) and partial nested (c) structures.

The vine copulas are graphical dependency models decomposing joint distributions of multivariate variables into a cascade of bivariate copulas, also known as pair-copulas (Bedford and Cooke 2001; Bedford and Cooke 2002; Brechmann and Schepsmeier 2013; Joe 1996). It is worth noting that every pair-copula can be modelled flexibly using any bivariate copula function. Clearly, a variety of dependence structures including asymmetries and tail dependencies can be taken into consideration through vine copula-based models (AghaKouchak et al. 2010). For example, the vine copula can comprise the pair-copulas of Clayton and Gumbel, which exhibit strong left- and right-tail dependence between variables, respectively (Bokusheva 2011), into a joint multivariate model. This construction is extremely useful for capturing fully the compound influence of extreme events on precipitation and wheat yield which is potentially modelled in an inappropriate manner. The regular vine (R-vine) is a general form of vine copulas (Kurowicka and Cooke 2006) (Fig. 5c)

while the statistical inference techniques are divided into special classes of drawable vine (D-vine) (Fig. 4a) and canonical vine (C-vine) (Fig. 4b) copulas (Aas et al. 2009; Brechmann 2010) which are mainly used in this study.



Source: Author

Figure 4. Four-dimensional drawable vine (D-vine) (a), canonical vine (C-vine) (b), and regular vine (R-vine) copulas.

Since the copula technique is advanced, which can model complex dependence structures between variables; recent years have enabled an extensive application of copula-based models in a variety of research fields. Many authors have applied copula methods for joint modelling of the hydrological properties (*e.g.*, peak streamflow, streamflow volumes and drought durations, severity, and intensity) (Grimaldi et al. 2016; Grimaldi and Serinaldi 2006), drought monitoring (Wong et al. 2010; Wong et al. 2013; Yang 2010). In finance and insurance, the copula method can be seen as a standard and popular instrument for multi asset pricing (Tankov 2011; Van Den Goorbergh et al. 2005), credit portfolio modelling (Frey and McNeil 2003; Frey et al. 2001), and risk management (Embrechts et al. 2002; Fang and Madsen 2013; Jaworski et al. 2013; Ouyang et al. 2009).

However, applications of the copula approach in agricultural economics, and particularly in risk managements are limited. Several studies have been implemented worldwide (Goodwin and Hungerford 2014; Larsen et al. 2015; Okhrin et al. 2013;

Xu et al. 2010) but not in Australia. Furthermore, these studies use multivariate elliptical or Archimedean copulas, which have some restrictions, as mentioned above. To the best of our knowledge, no previous studies have applied the vine copula technique to model the compound influence of climate mode indices on precipitation and wheat yield as well as to analyse weather systemic risk in the wheat belt region in Australia.

1.3 Thesis Research Aims and Objectives

With the research gaps already identified, the primary purpose of this doctoral thesis is to develop vine copula-based models for analysing the inter-association between synoptic-scale climate indices, precipitation and wheat yield in Australia's wheat belt. Subsequently, the copula approach is applied to analyse the stochastic dependence of weather systemic risk and to evaluate potential adaptation strategies. In particular, this thesis will adopt the copula technique within a statistical modelling framework to achieve the following specific objectives:

i. Modelling the compound influences of multiple synoptic-scale climate indices (*e.g.*, SOI and IOD) on weather variables (*e.g.*, precipitation). The developed copula model is able to provide probabilistic 'forecasting' information of seasonal precipitation.

ii. Modelling the joint influences of multiple synoptic-scale climate indices on the wheat yield at different times and locations. Subsequently, the wheat yield can be forecasted at different quantile levels using the most powerful copula-based model.

iii. Analysing the stochastic dependence of weather systemic risk in different locations and crop growing seasons and how it affects a hypothetical weather index-based insurance. Based on this analysis, the efficiency of different potential diversification strategies is evaluated and commented on, using regional, national and temporal scales.

iv. Modelling the joint return of wheat crops growing at different locations to evaluate the efficiency of the geographical diversification strategy. This objective attempts to optimise the expected return, given different levels of

risk, using the information randomly simulated from the developed copula-based model.

1.4 Scope and Limitations

To achieve these goals, the workflow has been partitioned into a number of separate tasks and the scope of the study are summarised as follows:

i. The potential compound impact due to co-occurrence of extreme climatic events on precipitation and wheat yield focusing on the Australian's wheat belt region are extensively reviewed. The stochastic interdependencies of weather events leading to systemic risks in agriculture and economics and potential adaptation strategies are also reviewed.

ii. The limitations of traditional statistical models such as linear regression in modelling multivariate distributions are identified leading to an understanding of the copula technique, particularly vine copulas and their advantages, and how it can overcome the challenges experienced in the conventional statistical methods.

iii. The study on the influence of climatic conditions is implemented using major synoptic-scale climate indices have been examined to affect precipitation and crop yield in Australia. The implementation is also conducted in different locations and at different times to analyse the spatio-temporal characteristics of the climate-related impact.

iv. While vine copulas are the core approach, the study also applies the conventional multivariate normal methods and other copula families for a comparison of modelling capacity. The marginal distributions can be fitted parametrically or non-parametrically. Copulas are estimated using a number of parametric copula functions and implemented in **R** software (Team 2013).

All the codes serving this study are implemented in **R** software, an open source and a free environment for statistical computing and graphics. Furthermore, the following factors may play an essential role in studies of climatic impact on agricultural sector; however, they are beyond the scope of this study due to the time limitations including:

- Dynamic impact and mechanism under the interaction of climate mode indices;

- Other copula methods such as empirical and entropy copulas;
- Other costs may occurred in implementing diversification strategies;
- Other potential adaptation strategies.

1.5 Organisation of the Thesis

This thesis comprises four major studies, presented as a PhD by publications, which cover the four objectives and a book chapter, and a conclusion that summarises the challenges, findings, significance and scientific contributions of this study, and recommendations for future works.

Four high quality journal articles produced from this study are represented below:

- **Article from Objective I:** Thong Nguyen-Huy, Ravinesh C. Deo, Duc-Anh An-Vo, Shahbaz Mushtaq, and Shahjahan Khan. "Copula-statistical precipitation forecasting model in Australia's agro-ecological zones". ***Agricultural Water Management***, 191 (2017): 153-172. [Impact Factor: 2.848, SNIP: 1.814, Scopus Rated Q1, 93rd percentile in Water Sc & Technology].

  DOI: https://doi.org/10.1016/j.agwat.2017.06.010

- **Article from Objective II:** Thong Nguyen-Huy, Ravinesh C. Deo, Duc-Anh An-Vo, Shahbaz Mushtaq, and Shahjahan Khan. "Modelling the joint influence of multiple climate mode indices on Australian wheat yield using a vine copula-based approach." ***European Journal of Agronomy***, 98 (2018): 65-81. [Impact Factor: 3.757, SNIP: 1.828, Scopus Rated Q1, 94th percentile in Agronomy & Crop Sc.].

  DOI: https://doi.org/10.1016/j.eja.2018.05.006

- **Article from Objective III:** Thong Nguyen-Huy, Ravinesh C. Deo, Shahbaz Mushtaq, Jarrod Kath and Shahjahan Khan. "Copula statistical models for analyzing stochastic dependencies of systemic drought risk". ***Stochastic Environmental Research and Risk Assessment*** (in review). [Impact Factor =

2.629, Scopus Rated Q1, 85th percentile in Safety, Risk, Reliability & Quality].

Ref. No.: SERR-D-18-00328

- **Article from Objective IV:** Thong Nguyen-Huy, Ravinesh C. Deo, Shahbaz Mushtaq, Jarrod Kath and Shahjahan Khan. "Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management". ***Weather and Climate Extremes***. [Impact Factor = 4.21, SNIP: 2.428, Scopus Rated Q1, 98th percentile in Geography, Planning & Development].

  DOI: https://doi.org/10.1016/j.wace.2018.07.002

Furthermore, the **book chapter** and the copyright information of the published articles are given in ***Appendix***.

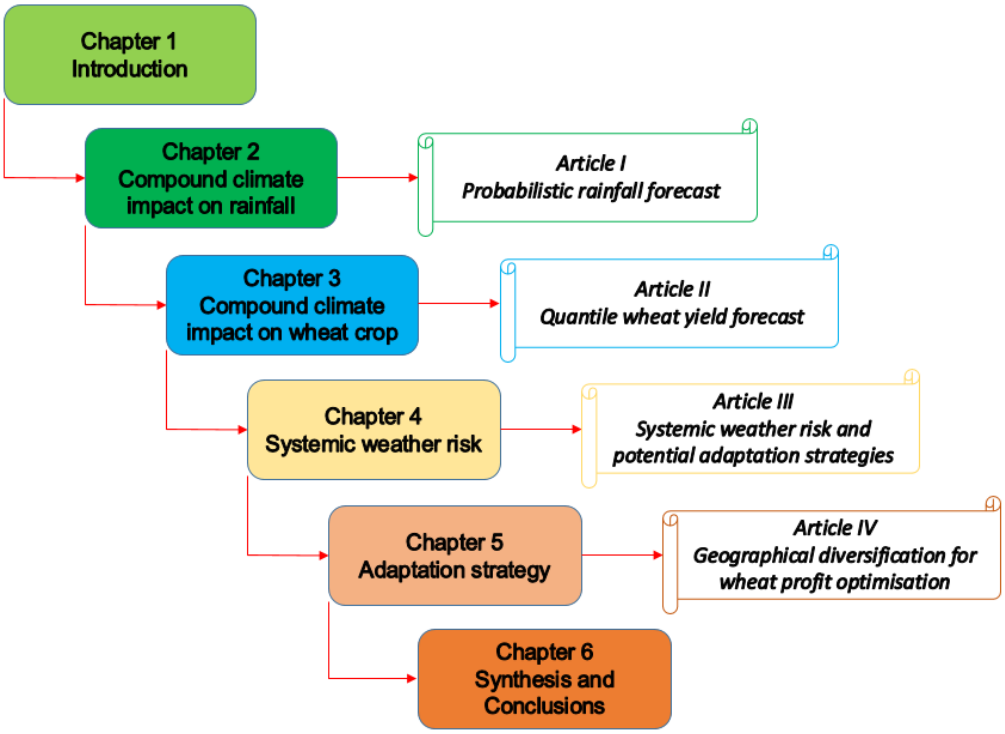
The ***first objective*** is to understand the compound influences of different phases between the inter-annual synoptic-scale climate driver ENSO and Inter-decadal Pacific Oscillation (IPO) Tripole Index (TPI) on spring precipitation forecast in Australia's wheat belt. The findings and analyses are represented in ***Article I***. The results show that using the vine copula technique, the trivariate models can provide a better accuracy of precipitation forecast than the bivariate models in the east and southeast wheat belt region. The trivariate forecasting models are also found to improve the forecast during the La Niña and negative TPI phases.

The ***second objective*** is to investigate the spatio-temporal influence of multiple large-scale climate drivers on the variability of wheat yield that are addressed in ***Article II***. Twelve large-scale climate indices, which have been examined as major drivers of Australian precipitation variability, are employed to investigate their joint influence on the seasonal wheat yield across five major wheat-producing states. The results indicate that wheat yield can be skilfully forecast 3–6 months ahead, supporting early decision-making in regard to precision agriculture. Generally, the developed D-vine quantile regression model provide greater accuracy for the wheat yield forecast given different quantile levels compared to the traditional linear quantile regression (LQR) method.

The ***third objective*** uses C-vine copula-based models to investigate the stochastic dependencies of systemic weather risk and evaluate the effectiveness of spatial and temporal diversification strategies, which can be acquired from ***Article III***. This study calculates the buffer fund, which is used as a reserve to handle payouts and avoid bankruptcy during widespread systemic losses, at regional, national and temporal levels. The results indicate that diversification strategies are feasible to mitigate systemic weather risks in Australia. Furthermore, diversifying risk over time potentially achieves more effectiveness than over space.

The ***fourth objective*** is to assess the performance of a geographical diversification wheat farming portfolio in Australia, which is described in ***Article IV***. Conditional Value-at-Risk (CVaR) and the joint copula model are employed to optimise the effectiveness of geographical diversification. The study indicates that the copula-based mean-CVaR model is seen to better simulate extreme losses compared to the conventional multivariate-normal models, which underestimates the minimum risk levels at a given target of expected return.

For better understanding the connection among the studies and articles, the flow story of the thesis is graphically represented in Fig. 5.



(Source: Author)

Figure 5. Flow diagram of the thesis.

1.6 Summary

Extreme climate events and their co-occurrence have a significant impact on weather variables and crop productions. Agricultural producers and financial managers are looking for a powerful tool to understand and quantify climate-related risks. Although a number of statistical models such as multivariate linear regression have been developed worldwide, they often fail to measure a complex dependence structure. The copula technique provides a better way to describe joint behaviour of compound events. The main objective of this research is to promote the valuable application of vine copulas in managing climate risks in the agricultural sector.

Chapter 2

Probabilistic precipitation forecasting using synoptic-scale climate indices

---

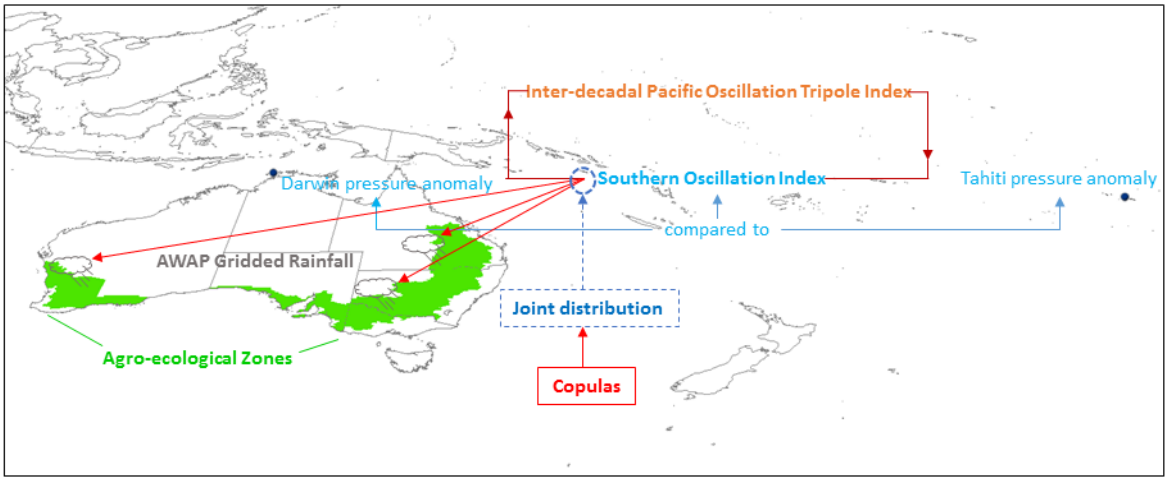
Article I: Copula-statistical precipitation forecasting model in Australia's agro-ecological zones

Summary:

This study adopts vine copulas to investigate the association of three-month (June – August) average values of ENSO and TPI on spring precipitation (September – November) forecast across the Agro-ecological Zones (AEZs) (Australia's wheat belt). Gridded monthly precipitation data (0.05^0 x 0.05^0) are collected from Australian Water Availability Project (1900–2013). Furthermore, bivariate and trivariate copula models are developed to measure the influences of single (ENSO) and dual predictors (ENSO & TPI) on seasonal precipitation forecasts. A set of hypothetical parametric distribution functions is used for fitting the marginal process. A total of ten one- and two-parameter bivariate copulas and their rotated versions ranging from elliptical to Archimedean functions are employed to fit bivariate models between pairwise variables. Both fitting procedures are enriched with graphical and statistical goodness-of-fit tests. Fig. 6 displays the graphical abstract of this study on compound impact of synoptic-scale climate indices on precipitation in Australia.

The analysis indicates that most of the AEZs exhibit statistically significant dependence between spring precipitation and synoptic-scale climate indices, except for the western AEZs. Stronger dependence in the upper tail observed in bivariate models implies that the influence of ENSO on precipitation forecast during La Niña phases is more evident than during El Niño phases. In regard to trivariate models, while the inclusion of TPI into bivariate models generally results in a notable reduction in the mean values of simulated precipitation, it depicts a general improvement in the median values. In particular, the Spearman correlation coefficients between observed and forecasted anomalies in bivariate cases are approximately 0.59, 0.26, 0.48 and 0.49 in Zones 1, 7, 8, and 12, respectively. However, if TPI is included into bivariate models these correlation coefficients drop to 0.50 in Zone 1 but increase to 0.52, 0.49

and 0.52, respectively, in Zones 7, 8 and 12. Furthermore, the correlation coefficients in the upper right and lower left quadrants (when both SOI and TPI are in the same phases) are significantly improved in trivariate models. The correlation coefficients in the upper right (lower left) quadrant increase from 0.72 (0.24), 0.27 (0.37), 0.27 and 0.10 (bivariate models) to 0.74 (0.52), 0.29 (0.54), 0.32 and 0.64 for Zones 1, 7, 8, and 12, respectively.



Source: Author

Figure 6. Graphical display of the study on compound impact of synoptic-scale climate indices on precipitation in Australia.

The results demonstrate that trivariate models can provide a better accuracy of precipitation forecast than bivariate models in the eastern and south-eastern AEZs. In addition, the trivariate forecast models are found to improve the precipitation forecast during the La Niña and negative TPI phases. This study determines the success of copula-based models for investigating the joint behaviour between multiple synoptic-scale climate indices and seasonal precipitation. The forecast information and respective models can provide useful tools for water resources and crop health management including better ways to adapt and implement viable agricultural solutions in the face of climatic challenges in major agricultural hubs, such as Australia's wheat belt.

Agricultural Water Management 191 (2017) 153–172

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Copula-statistical precipitation forecasting model in Australia's agro-ecological zones

Thong Nguyen-Huy[a,b,d], Ravinesh C. Deo[a,b,d,*], Duc-Anh An-Vo[b,c,d], Shahbaz Mushtaq[b,d], Shahjahan Khan[a,b,d]

[a] School of Agricultural, Computational and Environmental Sciences, QLD 4300, Australia
[b] International Centre for Applied Climate Sciences (ICACS), QLD 4300, Australia
[c] Computational Engineering and Science Research Centre (CESRC), QLD 4300, Australia
[d] University of Southern Queensland, Institute of Agriculture and Environment (IAg&E), QLD 4300, Australia

ARTICLE INFO

Article history:
Received 2 January 2017
Received in revised form 24 May 2017
Accepted 18 June 2017

Keywords:
Copula-statistical models
Seasonal precipitation forecasting
Vine copulas
Joint distribution
Goodness of fit
Climate indices

ABSTRACT

Vine copulas are employed to explore the influence of multi-synoptic-scale climate drivers – El Niño Southern Oscillation (ENSO) and Inter-decadal Pacific Oscillation (IPO) Tripole Index (TPI) – on spring precipitation forecasting at Agro-ecological Zones (AEZs) of the Australia's wheat belt. To forecast spring precipitation, significant seasonal lagged correlation of ENSO and TPI with precipitation anomalies in AEZs using data from Australian Water Availability Project (1900–2013) was established. Most of the AEZs exhibit statistically significant dependence of precipitation and climate indices, except for the western AEZs. Bivariate and trivariate copula models were applied to capture single (ENSO) and dual predictor (ENSO & TPI) influence, respectively, on seasonal forecasting. To perform a comprehensive evaluation of the developed copula-statistical models, a total of ten *one*- and *two*-parameter bivariate copulas ranging from elliptical to Archimedean families were examined. Stronger upper tail dependence is visible in the bivariate model, suggesting that the influence of ENSO on precipitation forecasting during a La Niña event is more evident than during an El Niño event. In general, while the inclusion of TPI as a synoptic-scale driver into the models leads to a notable reduction in the mean simulated precipitation, it depicts a general improvement in the median values. The forecasting results showed that the trivariate forecasting model can yield a better accuracy than the bivariate model for the east and southeast AEZs. The trivariate forecasting model was found to improve the forecasting during the La Niña and negative TPI. This study ascertains the success of copula-statistical models for investigating the joint behaviour of seasonal precipitation modelled with multiple climate indices. The forecasting information and respective models have significant implications for water resources and crop health management including better ways to adapt and implement viable agricultural solutions in the face of climatic challenges in major agricultural hubs, such as Australia's wheat belt.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Australia, an agricultural nation, has a relatively high inter-annual variability in climatic properties (including annual and seasonal precipitation); which is about 15–18% higher than any other major agricultural nation (Best et al., 2007; Cleugh et al., 2011; Mekanik and Imteaz, 2013; Walker and Mason, 2015). The Australian precipitation is influenced by the synoptic-scale processes mainly of tropical oceanic and atmospheric origin, with the primary attention largely paid to the different phase of the El Nino Southern Oscillation (ENSO) (Risbey et al., 2009; Schepen et al., 2012). An association between ENSO and the precipitation have been investigated since the early '80s (McBride and Nicholls, 1983), which ENSO phenomenon producing a strong impact on seasonal precipitation from July–March. The magnitude and timing of the effect varies considerably with the sites and coincides with major cropping periods (McBride and Nicholls, 1983; Stone et al., 1996). Interdecadal Pacific Oscillation (IPO) is also associated with climate variations across the Pacific, albeit on decadal timescales, acting to modulate the interannual variation of ENSO-related effects (Salinger et al., 2001), and is directly attributable to the shift in the intertropical convergence zone (Folland et al., 2002). However,

* Corresponding author at: School of Agricultural, Computational and Environmental Sciences, QLD 4300, Australia.
E-mail addresses: ThongHuy.Nguyen@usq.edu.au (T. Nguyen-Huy), ravinesh.deo@usq.edu.au, physrcd@yahoo.com (R.C. Deo).

http://dx.doi.org/10.1016/j.agwat.2017.06.010
0378-3774/© 2017 Elsevier B.V. All rights reserved.

Southern Oscillation Index (SOI) features in Australia are reinforced in negative IPO phases in some areas, with lesser impacts in others (Chiew and Leahy, 2003). Also, individual ENSO events have stronger and more predictable impacts across Australia during the negative (cool) IPO phases (Kirono et al., 2010). Power et al. (1999) found no significant relationship between interannual climate variability and ENSO in the positive IPO phases, although a relationship between IPO and precipitation, river flow and wheat yield, were identified in negative phases. While a better understanding of ENSO's impact in different IPO phases is creating an opportunity for improving the performance of forecasting model, the mechanism of interdecadal modulation of IPO on ENSO-related climatology, including precipitation, is complicated. Precipitation forecasting, however, is an important task for agricultural water management and agricultural economics (An-Vo et al., 2015; Deo et al., 2017), affecting subsistence and commercial aspects of Australia's agricultural industry (Anwar et al., 2007; Best et al., 2007; Chiew et al., 2003; Montazerolghaem et al., 2016).

In published literature, Risbey et al. (2009) has identified the key rivers of Australian precipitation variability based on concurrent relationships between synoptic-scale climate indices and precipitation. However, to forecast future precipitation, lagged relationships between climate indices in the current and the (next) seasonal precipitation are important. Further, a strong concurrent relationship does not always lead to a strong lagged relationship (Schepen et al., 2012). Although such lagged relationships have been widely applied to forecast the precipitation in Australia, the authors utilised either a single climate index (Chiew et al., 1998; Stone et al., 1996; Taschetto and England, 2009) or assessed the impact of each climate index separately (Hasan and Dunn, 2012; Kirono et al., 2010; Schepen et al., 2012). In general, the influence of different climate indices varies across seasons and regions. For example, in north-eastern Australia, spring precipitation was found to exhibit the highest correlation with Niño-4.0 and thermocline properties, while optimal predictors for summer precipitation were Niño-4.0 and Dipole Mode Index (Kirono et al., 2010). Such temporal and spatial variations of the impact of different climatic indices have been also confirmed in an extensive research by Schepen et al. (2012). However, in spite of an acceptable level of model performance for seasonal forecasting, utilisation of a single predictor can hinder forecasting ability since relationships between climate indices and precipitation can be relatively complex (Rasel et al., 2016; Wang and Hendon, 2007).

In Australia, the impacts of different climate indices on precipitation amounts vary according to the continent's geographic diversity (Chowdhury and Beecham, 2010; Deo et al., 2017). In spite of this, only a handful of research has considered the joint influence of multiple indices on seasonal precipitation forecasting. Mekanik and Imteaz (2013) combined ENSO and Indian Ocean Dipole to develop a model for spring precipitation in south-eastern Australia. Rasel et al. (2016) incorporated SOI and Southern Annular Mode to demonstrate 63% better prediction accuracy of spring precipitation in South Australia compared to a single index. However, these studies employed a regression model where the dependence structure between an index and precipitation was measured by Pearson's correlation, whilst assuming linearity and normal distributions. Precipitation data, however, exhibits a skewed distribution and its relationship with climate indices is nonlinear (Schepen et al., 2012) which invalidates the use of Pearson's correlation and normal assumption. As the dependence structure between the predictors of precipitation and the predictand (i.e., precipitation) is governed by the marginal distributions of these variables that can help decision-makers to capture the 'cause and effect' relationships,, a robust forecasting model must allow the establishment of the linear or nonlinear dependence between the predictors (*e.g.*, climate indices) and predictand (*e.g.*, precipitation) with the marginal distributions being derived from diverse distribution families. Here, we aim to achieve such a novel forecasting method by employing copula-statistical models, which are yet to be applied for precipitation forecasting research in the present study region.

Copula-statistical models (Sklar, 1996; Sklar, 1959) that utilise ranked Spearman or Kendall tau coefficients provide viable alternatives for modelling non-linear dependences, and have attracted much attention in bivariate and trivariate based modelling (De Michele and Salvadori, 2003; Evin and Favre, 2008; Hao and Singh, 2016; Rauf and Zeephongsekul, 2014; Zhang and Singh, 2007). Copula functions allow to model the dependence structure independently from the marginal selection. Further, they overcome issues associated with joint dependences between rare events (*e.g.*, precipitation extremes) by considering tail dependences (*i.e.*, asymmetric dependence structure) which is impossible with simplistic statistical models. Since their advent, recent years have witnessed an extensive application of copulas for: insurance and financial risk (Fang and Madsen, 2013; Jaworski et al., 2013; Trede and Savu, 2013), hydrology and water resources (Favre et al., 2004; Hao and Singh, 2013; Wong, 2013), drought (Wong et al., 2010; Wong et al., 2013; Yang, 2010), flood (Chowdhary et al., 2011; Favre et al., 2004) and streamflow (Hao and Singh, 2012; Lee and Salas, 2011). Khedun et al. (2014) applied multivariate Gaussian and Archimedean copulas for modelling the effect of ENSO and PDO on precipitation anomalies in Texas. However, multivariate Gaussian copulas, as applied in that study, can have a drawback by restricting the symmetric dependence associated with elliptical copulas. Also, multivariate Archimedean copulas employ a single parameter on pairs of variables which assumes the same dependence structure for variable pairs. Such an assumption might be unrealistic (Hao and Singh, 2016). In addition, to the best of our knowledge, copula-statistical models have not been explored for probabilistic forecasting of seasonal precipitation in Australia, despite that the challenges and a need for a robust model for spring precipitation forecasting at Agro-ecological Zones (AEZs) of the Australia's wheat belt.

In this paper, we model the joint influence of ENSO (through the phases of the Southern Oscillation Index, SOI) and IPO Tripole Index (TPI) on seasonal precipitation at the AEZs using vine copulas. The TPI (which is a robust and stable representation of the IPO with less variance in the decadal than the shorter timescales compared to Niño 3.4 due to an inclusion of off-equatorial sea surface temperature (SST)) (Henley et al., 2015) exhibits similar characteristics to the Pacific Decadal Oscillation (PDO) in terms of SST but theirs influences are spatially disparate. TPI utilises SSTs in the South of 20^0N, associated with a 'tripole' pattern and three centres of action and variations, stipulated in the second principal component of a low-pass filtered global SST. Further, vine copulas allows to unite different bivariate copulas for modelling the flexible dependence among pairwise variables independently with the marginal selection (Bedford and Cooke, 2001, 2002; Kurowicka and Cooke, 2006). Vine copulas have not been tested for precipitation forecasting although they were verified for precipitation refinement studies (Liu et al., 2015), stochastic modelling (Verhoest et al., 2015) and daily precipitation disaggregation (Gyasi-Agyei, 2011).

Considering the need for precipitation forecast model to be developed at Australia's Agro-ecological Zones (AEZs), the aims of this study are threefold: (1) to develop a copula-based bivariate and trivariate models describing the joint impact of SOI and TPI on spring seasonal precipitation variability; (2) to evaluate statistically the influence mechanism of the considered climate indices on precipitation through a comparison of the prescribed bivariate and trivariate models; and (3) to evaluate the utility of copula-based conditional model for forecasting seasonal precipitation. The paper is structured as follows. Section 2 covers the theory of copulas; Section 3 presents data, methodology and model development; Section

4 outlines the results and general discussion. Further discussions are presented in Section 5, and conclusions are presented in Section 6.

2. Theory of copula-statistical model

Sklar (1959) introduced the copulas, as a multivariate distribution with all univariate marginals being standard uniform $U[0, 1]$ (Nelsen, 2006; Sklar, 1959). Let F be the n-dimensional joint cumulative distribution function (CDF) of an n-dimensional random vector $X = [x_1, ..., x_n]^T$ with marginal CDF F_1,..., F_n. There exists a copula $C : [0, 1]^n \rightarrow [0, 1]$ to satisfy Eq. (1) (Sklar, 1996; Sklar, 1959):

$$\begin{aligned} &F(x_1, ..., x_n) = P(X_1 \le x_1, ..., X_n \le x_n) \\ &= C[F_1(x_1), ..., F_n(x_n)] = C(u_1, ..., u_n), \end{aligned} \tag{1}$$

where $P(X_i \le x_i) = F_i(x_i) = u_i$ for $i = 1, ..., n$ with $U_i \sim U[0, 1]$. If F_1, F_2, ..., F_n are continuous distributions, C is unique; otherwise, C is uniquely determined on $RanF_1 \times RanF_2 \times ... \times RanF_n$. Conversely, if C is an n-copula and $F_1, F_2, ..., F_n$ are distribution functions, then the function F is an n-dimensional distribution function with marginal distributions $F_1, F_2, ..., F_n$ (Sklar, 1996).

The multivariate density function, $f(x_1, x_2, ..., x_n)$ can be expressed as (Sklar, 1996):

$$f(x_1, ..., x_n) = \left[\prod_{i=1}^{n} f_i(x_i)\right] c(u_1, ..., u_n), \tag{2}$$

where $c = \frac{\partial^n C}{\partial F_1, ..., \partial F_n}$ is the copula density and $f_i(x_i)$ is the marginal density.

Copulas are generally classified into families including, but not limited to, empirical, Archimedean, extreme value, elliptical, vine, and entropy copulas. More details of elliptical and Archimedean copulas are found in Sections A.1 and A.2 of the Supplementary material.

2.1. Vine copulas

Vine copulas, which were introduced by Joe (1996) and applied more comprehensively by Bedford and Cooke (2001), as graphical dependency models for describing multivariate variables using Markov trees and construction of a cascade of bivariate copulas; they are the so-called pair-copulas (Brechmann and Schepsmeier, 2013). A multivariate probability density is decomposed into bivariate cases where a selection of each pair-copulas is independent. The vine copulas perform a variety of modelling where asymmetries and tail dependence, as evident in precipitation forecasting problems (Aghakouchak et al., 2010), are taken into account. The statistical inference techniques are divided into special classes of Canonical- (C-) and D- vines (Aas et al., 2009). In a C-vine, pairs and associated pair-copula for n variables can be constructed as (Bedford and Cooke, 2001):

$$\begin{aligned} &(1, 2), (1, 3), (1, 4), ..., (1, n) \qquad (Tree1) \\ &(2, 3|1), (2, 4|1), ..., (2, n|1) \qquad (Tree2) \\ &..., \\ &(n-1, n|1, ..., n-2) \qquad (Treen-2). \end{aligned}$$

The pair-copula construction requires a computation of marginal conditional distribution functions, $F(x|v)$ for an m-dimensional vector $v = (v_1, ..., v_m)$. Aas et al. (2009) showed that, for every$j$, the conditional distribution function is:

$$h(x|v) = F(x|v) = \frac{\partial C_{x, v_j|v_{-j}}\left[F\left(x|v_{-j}\right), F\left(v_j|v_{-j}\right)\right]}{\partial F\left(v_j|v_{-j}\right)}, \tag{3}$$

Note that v_j denotes an arbitrary component of v, and v_{-j} denotes the vector v excluding element v_j, and $C_{x, v_j|v_{-j}}$ is a bivariate copula distribution function.

If v is univariate, Eq. (3) can be written as

$$h(x|v) = F(x|v) = P(X \le x| = v) = \frac{\partial C_{xv}[F(x), F(v)]}{\partial F(v)}. \tag{4}$$

The hierarchical construction of the conditioning sequence for the C-vine is that the variable 1 is conditioned on first, then variable 2, and so on (Brechmann and Schepsmeier, 2013; Liu et al., 2015). According to Aas et al. (2009), the n-dimensional density corresponding to C-vine is written as:

$$\begin{aligned} f(x_1, ..., x_n) = &\prod_{k=1}^{n} f_k(x_k) \times \prod_{i=1}^{n-1}\prod_{j=1}^{n-i} c_{i,i+j|1:(i-1)} \\ &\left[F(x_i|x_1, ..., x_{i-1}), F\left(x_{i+j}|x_1, ..., x_{i-1}\right)\right] \end{aligned} \tag{5}$$

For the three random variables X_1–X_3, the construction of C-vine is shown in Fig. 2.1 of Supplementary material. The conditional density derived from Eq. (5) is

$$\begin{aligned} &f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ &\cdot c_{12}[F_1(x_1), F_2(x_2)] \cdot c_{13}[F_1(x_1), F_3(x_3)] \\ &\cdot c_{23|1}[F(x_2|x_1), F(x_3|x_1)], \end{aligned} \tag{6}$$

where $F(x_2|x_1) = h(u_2|u_1)$ and $F(x_3|x_1) = h(u_3|u_1)$ can be found from Eq. (4).

It is also noted that for the three-dimensional case, the C-vine copula is the D-vine copula where the center variable is identified (Zhang and Singh, 2014). The next Section presents the method that applies the conditional distribution function to develop the forecasting model.

2.2. Copula-based conditional forecasting model

The inverse form of the conditional distribution function was used to construct a forecasting model (Chen et al., 2009; Liu et al., 2015). In bivariate case of two random variables (x_1, x_2), with a given conditional distribution function $h(u_1|u_2)$, the modeller's aim is to obtain u_2 based on the information of u_1. For known probabilities $\mathcal{P} \in (0, 1)$, u_2 can be derived by solving $u_2 = C_{2|1}^{-1}(\mathcal{P}, u_1) = h^{-1}(\mathcal{P}|u_1)$, where $C_{2|1}^{-1}$ is the inverse of the copula $C_{2|1}$. Variable x_2 can be obtained by solving the $\mathcal{P}^{th}$ conditional quantile function as:

$$Q_{x_2}(\mathcal{P}|x_1) = F^{-1}\left[C_{2|1}^{-1}(\mathcal{P}, x_1)\right] = F^{-1}\left[h^{-1}(\mathcal{P}|x_1)\right], \tag{7}$$

where F^{-1} is the quantile function of u_2. For the three variables case, the $\mathcal{P}^{th}$ conational quantile function $Q_{x_3}(\mathcal{P}|x_1, x_2)$ of x_3 can be obtained by recursive computations following Eq. (7)

$$Q_{x_3}(\mathcal{P}|x_1, x_2) = F^{-1}(u_3) = F^{-1}\left\{h^{-1}\left[h^{-1}(\mathcal{P}|h(u_2, u_1))\right]|u_1\right\}. \tag{8}$$

Hence, variable x_3 is forecasted based on the given information of x_1, x_2 (*i.e.*, bivariate data).

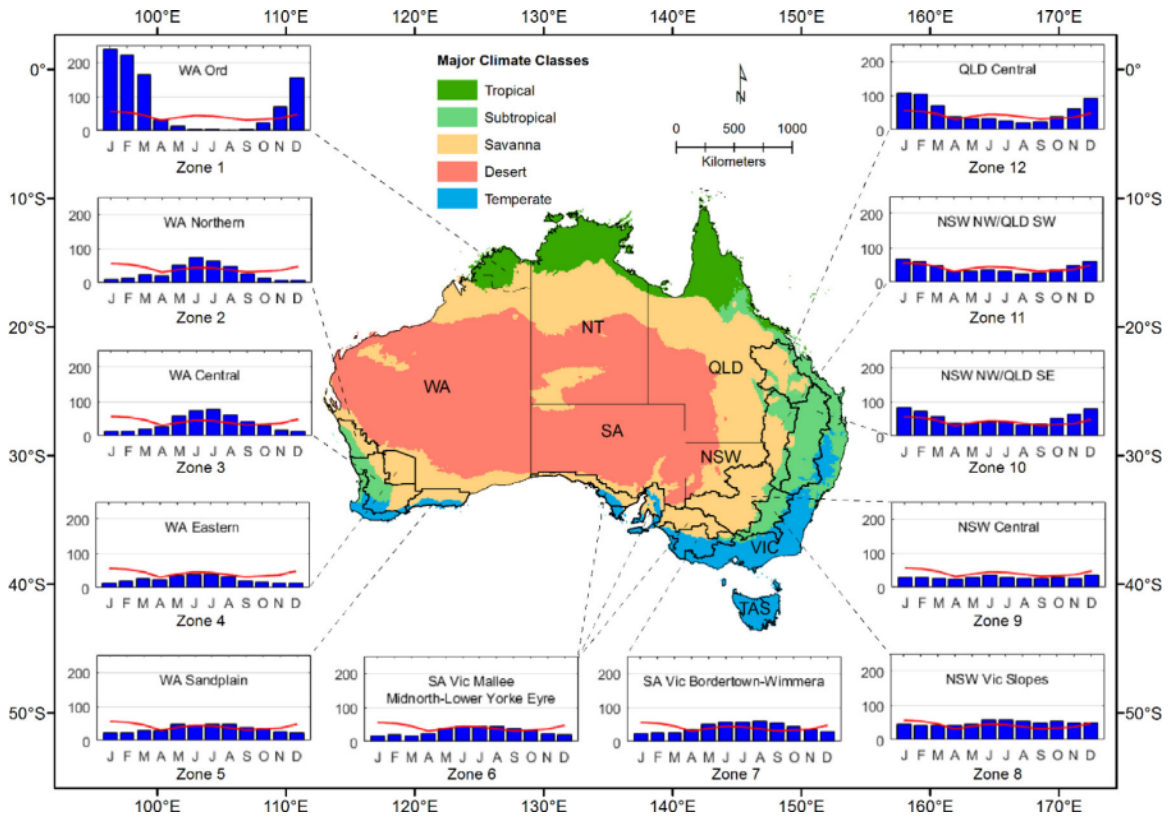


Fig. 1. Map of Agro-ecological Zones (AEZs). Bar chart shows monthly average precipitation of each AEZ from 1900 to 2013 plotted with the monthly average precipitation of all Zones (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Materials and method

3.1. Study area

In this paper, we validate the utility of copula-statistical models for forecasting seasonal precipitation in AEZs located on coastal edge of the Australian continent from the western to the south/south-eastern end, and to the eastern Australia. Grains Research and Development Corporation (GRDC) classified cropping zones in wheat growing (Agro-ecological) sites based on similar climatic conditions (GRDC, 2012; Murray and Brennan, 2009).

Fig. 1 plots the study area. AEZs exhibits a wide range of climatic conditions with four out of the five major climatic classes typically prevalent in Australia, including the tropical (Zone 1), subtropical (Zones 2, 3, 8 and 10), savanna (Zones 4, 5, 9, 11 and 12) and temperate (Zones 6 and 7) region (excluding some small Zones). Climate classifications in this paper are based on the Köppen-Geiger proposition, following Kriticos et al. (2012) applied to the 5′ resolution WorldClim global climatology (www.worldclim.org; Version 1.4, release 3; (Hijmans et al., 2005). The system is downloadable from CliMond climate data products (www.climond.org; Kriticos et al. (2012)). Average annual precipitation varies from 300 to 1000 mm where the eastern and southern ends receive the highest and the lowest total precipitation, respectively.

Fig. A.1 (Supplementary material) displays a violin plot of monthly precipitation; as a combination of boxplot with a kernel density plot, rotated and added on each side to show the data distribution for the climatological period (1900–2013). It is noticeable that the probability of the months with zero precipitation is highest for Zones 2, 4 and 8 and the lowest for Zones 7, 9, and 11. Also, Zones located in the north-western and eastern ends are more skewed and kurtosis for precipitation distribution compared to those in the south. Average precipitation (Fig. 1) shows a unimodal pattern in western and southern Zones with more precipitation in winter (June–August) compared to the others. Zones 10–12 (eastern end) have a bimodal pattern, while Zones 8 and 9 receive almost identical precipitation. Given the wide climatic conditions, evaluation of a probabilistic model is important in this economical region (Australia's wheat belt).

3.2. Data

3.2.1. Precipitation

To fit the most accurate copula-statistical models, gridded monthly precipitation data from January 1900 to December 2013 with a spatial resolution of $0.05^0 \times 0.05^0$ were acquired from Australian Water Availability Project (AWAP) (Raupach et al., 2008; Raupach et al., 2009). The AWAP dataset was obtained by interpolating daily precipitation measurements from over 7000 (the number reported in the early 1970s) stations across Australia. Three dimensional smoothing splines were employed to create gridded climatological averages. Analyses of the daily and monthly anomalies were performed using the Barnes successive correction method (Jones et al., 2009).

In terms of the data quality, the spatial accuracy of the fields is low in central-western Australia where meteorological stations are sparse (Schepen et al., 2012). As the horizontal resolution of the present data are relatively good (5 km), monthly average precipitation for each Zone was computed by averaging grid cells within a zonal boundary. Long-term spring (September–November (SON)) precipitation was subtracted by SON precipitation (i.e., removed the seasonality) of each year to determine SON anomalies for the AEZs. Fig. A.2 (Supplementary Material) plots time series of monthly precipitation for each Zone, smoothed with a 13 month centred moving average window and the associated mean values.

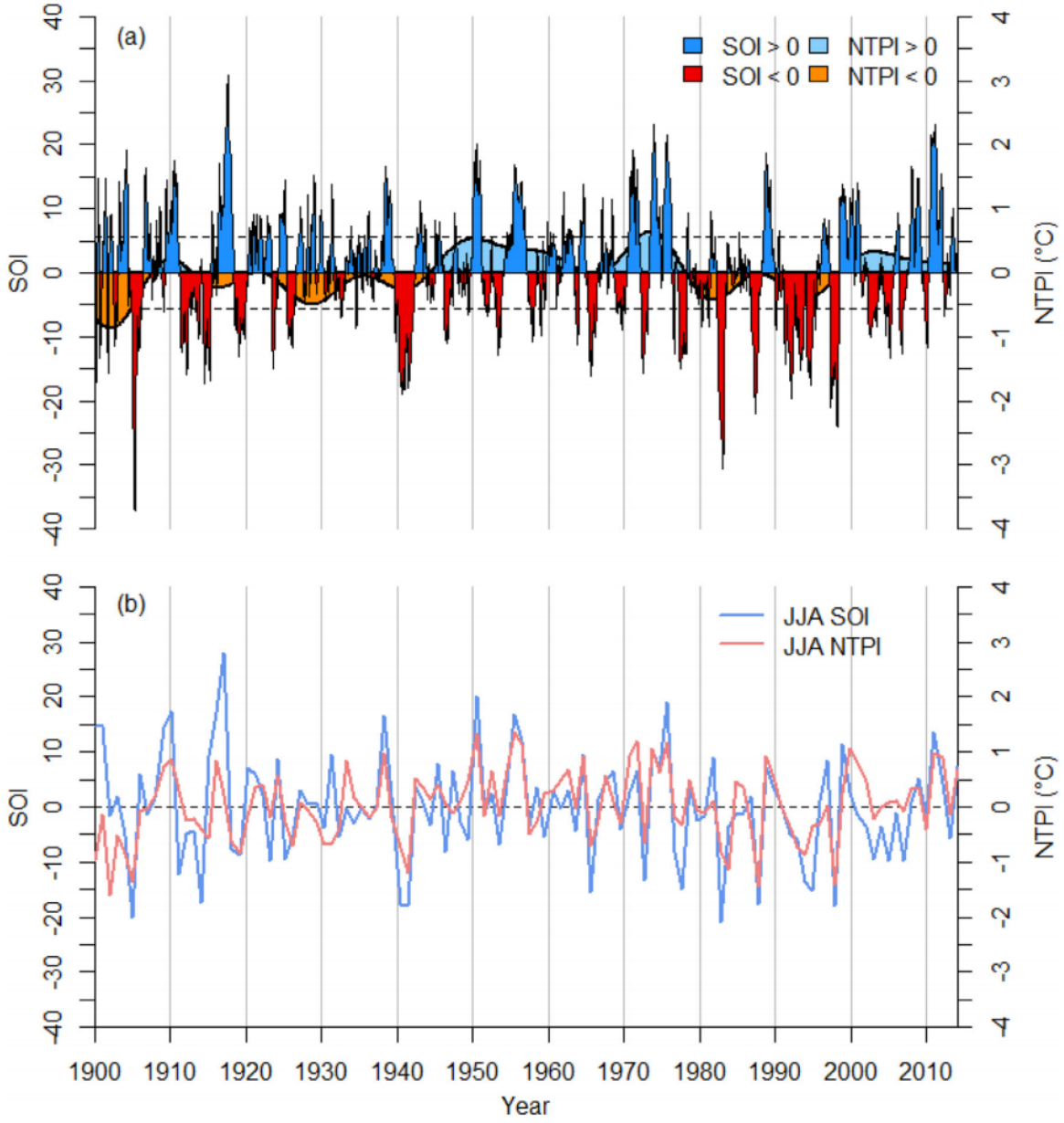


Fig. 2. (a) Time series of monthly SOI and NTPI. The SOI series has been overlaid with a 3-month centred moving average filter and ± 5.5 threshold. The NTPI series has been smoothed with 13-year Chebyshev low-pass filter. (b) JJA SOI versus JJA NTP with an overall correlation coefficient of 0.64.

3.2.2. Climate indices

In the present work, precipitation forecasting was based on the notion that the principal synoptic-scale circulation drivers of inter-annual precipitation variability in Australia was the ENSO phenomenon (Chowdhury and Beecham, 2010; Mekanik and Imteaz, 2013; Schepen et al., 2012). There are a number of indices used to characterise ENSO such as SOI and Niño-3 index. This paper employs SOI to depict the correlation of ENSO with precipitation variability, since it is a broadly available and simple metric, being closely associated with the process of precipitation occurrences through its relationship with synoptic-scale surface pressures (Risbey et al., 2009). SOI data accorded to the Troup (1965) method with standardised anomaly of sea level pressure difference between Tahiti (17.5^0S–149.6^0W) and Darwin (12.4^0S–130.9^0E) multiplied by a factor of 10, were acquired from the Bureau of Meteorology. It is noteworthy that the impact of SOI on Australian precipitation was mainly due to its oscillating phases, where in general, the positive SOI (>5.5; La Niña phase) brings more precipitation, while the negative SOI (<−5.5, El Niño phase) is associated with a drought probability in major parts of Australia (McKeon et al., 2004). The lagged SOI-precipitation correlation, is reasonable for seasonal precipitation forecasting as stipulated in other studies (Kirono et al., 2010; Stone et al., 1996). In particular, such lagged relationship has been found useful for peak predictability of precipitation during August–November period in the eastern and northern Australia (Chiew et al., 1998; McBride and Nicholls, 1983; Schepen et al., 2012). Hence, this study utilises June-July-August averaged SOI (JJA SOI) to forecast precipitation in the spring (SON) season. JJA SOI adopted here is an attempt to incorporate the effects of different phases (*i.e.*, La Niña and El Niño events) on precipitation into copula-statistical forecasting models.

As a supplementary metric, we explored a relatively new climate index: Interdecadal Pacific Oscillation (IPO)-based Tripole Index (TPI) introduced by Henley et al. (2015), acquired from Earth

Table 1
Rank-based correlation coefficients and *p*-value of Kendall's *tau* and Spearman's *rho* between average JJA SOI, NTPI and average SON precipitation anomaly for Australia's Agro-ecological Zones (AEZs).

Agro-ecological Zone #	SOI				NTPI			
	Kendall	*p*-value	Spearman	*p*-value	Kendall	*p*-value	Spearman	*p*-value
1	0.34	0.57e-7	0.51	0.74e-8	0.23	0.27e-3	0.35	0.15e-3
2	0.02	0.75	0.03	0.75	−0.07	0.28	−0.10	0.31
3	0.03	0.67	0.04	0.69	−0.07	0.29	−0.09	0.32
4	0.02	0.80	0.03	0.72	−0.04	0.51	−0.07	0.45
5	0.04	0.57	0.06	0.53	−0.03	0.67	−0.03	0.73
6	0.17	0.01	0.25	0.01	0.09	0.17	0.14	0.15
7	0.23	0.36e − 3	0.32	0.56e − 3	0.21	0.12e − 2	0.29	0.17e − 2
8	0.26	0.31e − 4	0.37	0.42e − 4	0.25	0.85e − 4	0.37	0.61e − 4
9	0.27	0.16e − 4	0.39	0.19e − 4	0.27	0.19e − 4	0.40	0.11e − 4
10	0.21	0.11e − 2	0.30	0.11e − 2	0.25	0.92e − 4	0.36	0.71e − 4
11	0.22	0.70e − 3	0.31	0.65e − 3	0.26	0.38e − 4	0.39	0.19e − 4
12	0.27	0.18e − 4	0.40	0.87e − 5	0.27	0.17e − 4	0.40	0.10e − 4

System Research Laboratory, Physical Science Division. TPI is an SST-based index for IPO according to anomalies in three geographic regions of the Pacific Ocean considering SST anomalies (SSTA) averaged over the central equatorial Pacific and the SSTA in Northwest and Southwest Pacific. TPI is reported as a native temperature anomaly providing a simple metric for IPO evolution and is consistent with indices used to track ENSO, *e.g.*, Niño-3.4 index (Henley et al., 2015). Verdon et al. (2004) found that IPO had a strong impact on precipitation not only in New South Wales and Queensland but also further to the south such as Victoria. It should be noted that, while SOI has a positive relationship with precipitation, IPO exhibits an opposite trend where positive IPO is associated with lower precipitation, and negative IPO is associated with higher precipitation (Power et al., 1999). Previous work also indicated that IPO modulates the magnitude as well as the occurrence frequencies of ENSO events (Kiem and Franks, 2004; Power et al., 1999). In fact, the ENSO-hydroclimate relationship is relatively stronger for a negative IPO phase compared to when it is positive (*e.g.*, Chiew and Leahy, 2003). They all suggest that IPO (TPI in this study), in the combination of ENSO, should be considered in developing forecasting models, particularly for long lead-times.

Fig. 2a–b plots SOI and TPI, showing an anti-phasic pattern (top panel) in the amplitude of the index where a negative SOI is mirrored by a positive TPI at the same temporal scale, and vice versa. Correlations between JJA TPI and SON precipitation (not shown) were negative. It is noteworthy, however, that negative correlation can have an impact on the range of copulas evaluated in the study. In fact, we can use the rotated Archimedean copulas to model the negative dependences; however, this would double the number of copulas to be analysed. An offset was thus applied where the TPI data were multiplied by − 1 to generate negative TPI (denoted 'NTPI' hereafter) to show a positive correlation with precipitation, to accord to the correlation sign of the SOI data on precipitation.

Fig. 2b shows the fluctuations of NTPI time series, where a significant correlation between JJA SOI and JJA NTPI exists. SOI time series are overlaid with a seasonal (3-month) running mean and a ± 5.5 threshold to identify extreme SOI values associated with El Niño and La Niña events as a result of the synoptic-scale changes. In general, La Niña (El Niño) years occur concurrently with the positive (negative) phase of the NTPI (Fig. 2a).

3.3. Forecasting model development

As a prior step for constructing a model, the influence of climate variability on spring precipitation was examined using two common rank correlation coefficients, Kendall's *tau* and Spearman's *rho* (Pham et al., 2016). Both measure the monotonicity relationship among variables and are bounded by [−1,1] demonstrating perfect negative and perfect positive associations, respectively. The significance of the acquired values for both measures of dependence were invested by Genest and Favre (2007). Next, the SOI, NTPI, and precipitation anomalies were fitted to their probability distribution functions based on graphical analysis and statistical goodness-of-fit (GOF) test. Such combinatorial approach was also used to select the best fitting copula model, independently to the marginal selection, for the dataset. Finally, selected copula model was employed to generate the forecasted precipitation anomalies given the SOI and NTPI data.

3.3.1. Marginal distribution selection

We applied a combination of statistical GOF test and graphical analysis to a set of theoretical probability distributions to select the best fitting margins for the predictors (SOI, NTPI) and predictand (precipitation). The graphical assessment includes the density, CDF, quantile–quantile (Q-Q), and probability–probability (P-P) plots. While the density and CDF plots are the basic classical GOF plots, the Q–Q and P–P plots emphasise the lack-of-fit for the distribution tails and centres, respectively. The chi-squared statistic was employed comparing the GOF of assumed distributions to the theoretical distributions (Khedun et al., 2014; Liu et al., 2015) with a null hypothesis that the observed data emerged from a specified distribution. A larger *p*-values supported the tenability of the null hypothesis, viz the test statistic (Khedun et al., 2014):

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \tag{9}$$

where O_i and $E_i$are the observed and expected frequency, respectively, for the bin i, and $k$is the number of bins based on Sturges' formula ($k = \log_2 N + 1$). $E_i = N(F(Y_u) - F(Y_l))$ where $F$is the CDF for the distribution, Y_u and Y_i are the upper and lower limits for class i, respectively, and$N$is the sample size. The test statistic is distributed as a χ^2 random variable with $k - p - 1^\circ$ of freedom, p being the number of estimated parameters (Fischer et al., 2009).

3.3.2. Copula selection

We applied a similar procedure in Section 3.3.1 to select the best fitting copula models for the paired variables (e.g., SOI-precipitation), *i.e.*, using both graphical tools and formal statistical GOF tests. Firstly, dependence between predictors and predictand was analysed using rank-based, Kendall (K) and chi plots. Such plots were explained in detailed by Fisher and Switzer (2001) and Genest and Favre (2007). The interpretation of K plot is similar to that of the Q–Q plot, where the pair variables are said to be independent if their points fall on the diagonal line ($y = x$) and points above (below) the diagonal line indicate positive (negative) dependence. Chi plots are constructed based on the control charts and chi-square statistic

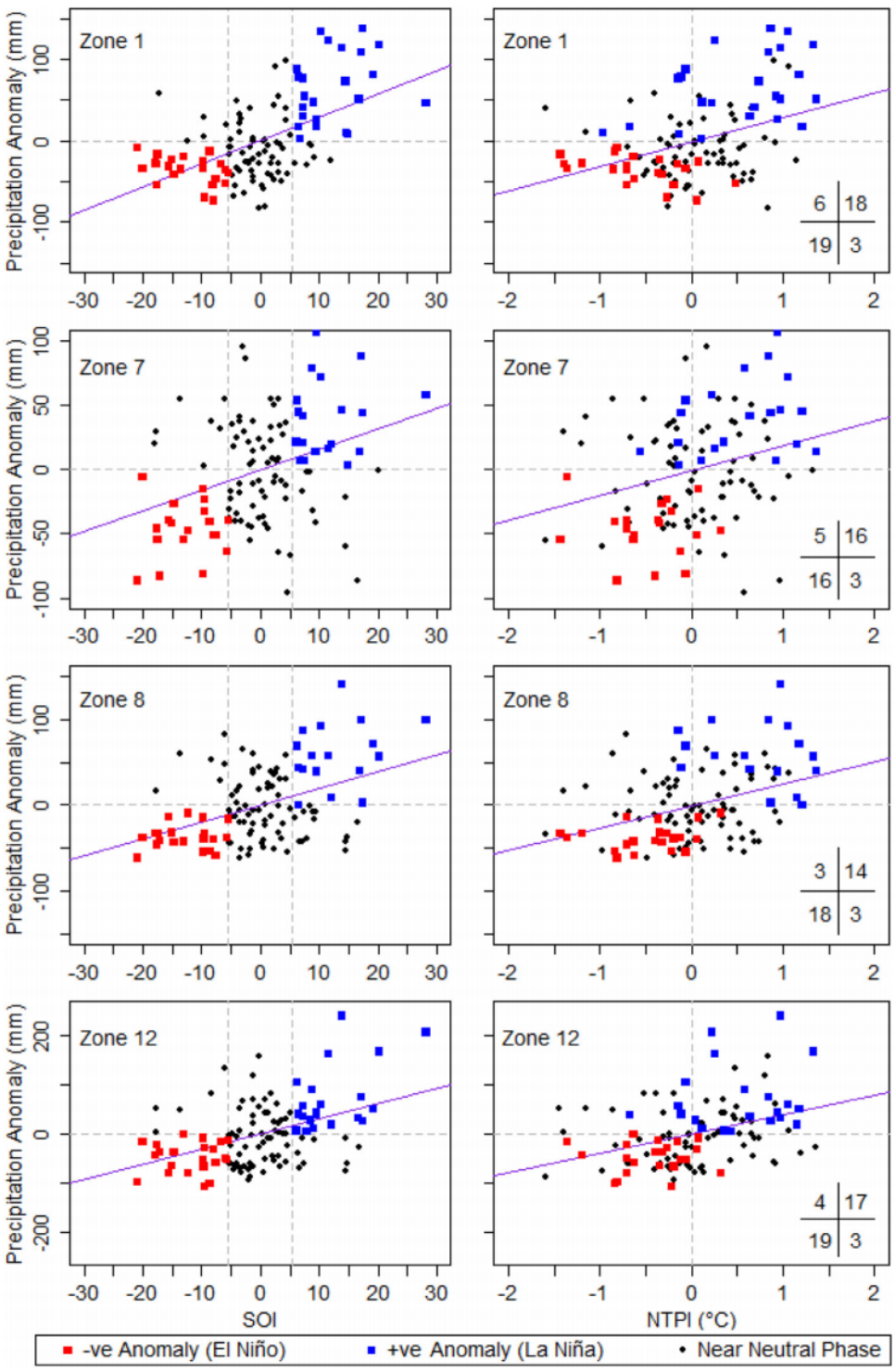


Fig. 3. Scatter plot of monthly average SON precipitation versus average JJA SOI (left column) and JJA NTPI (right column) indices for Zones 1, 7, 8, and 12. Positive (negative) anomalies during La Niña (El Niño) phases of preceding events are shown in blue (red). The numbers of blue and red events in the four quadrants are given in the lower corner of the right column. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for independence in a two-way table. It is a scatter graph of the pairs (λ_i, χ_i) with λ_i being the distance measured from the data point to the center of the dataset. If there is no relationship between the pair variables, approximately 95% of the data points should plot within the two horizontal control lines. Further, chi plots for upper and lower quadrants also allow to explore the tail dependence.

We also employed the scatterplot of observed data (from empirical copula C_n) overlapped on the randomly simulated data from copula C_{θ_n} to assess the preciseness of the fitted copulas. In accordance with Genest and Favre (2007), a large sample simulated from C_{θ_n} was generated to avoid any arbitrariness due to sampling variability in displaying the range of the distributions (Chowdhary et al., 2011). Further, lambda (λ) plot was applied to compare the empirical and theoretical λ-function of given bivariate copula data where λ-function was computed for each bivariate copula defined by Kendall's distribution (Schepsmeier, 2010). Next, the Cramér-

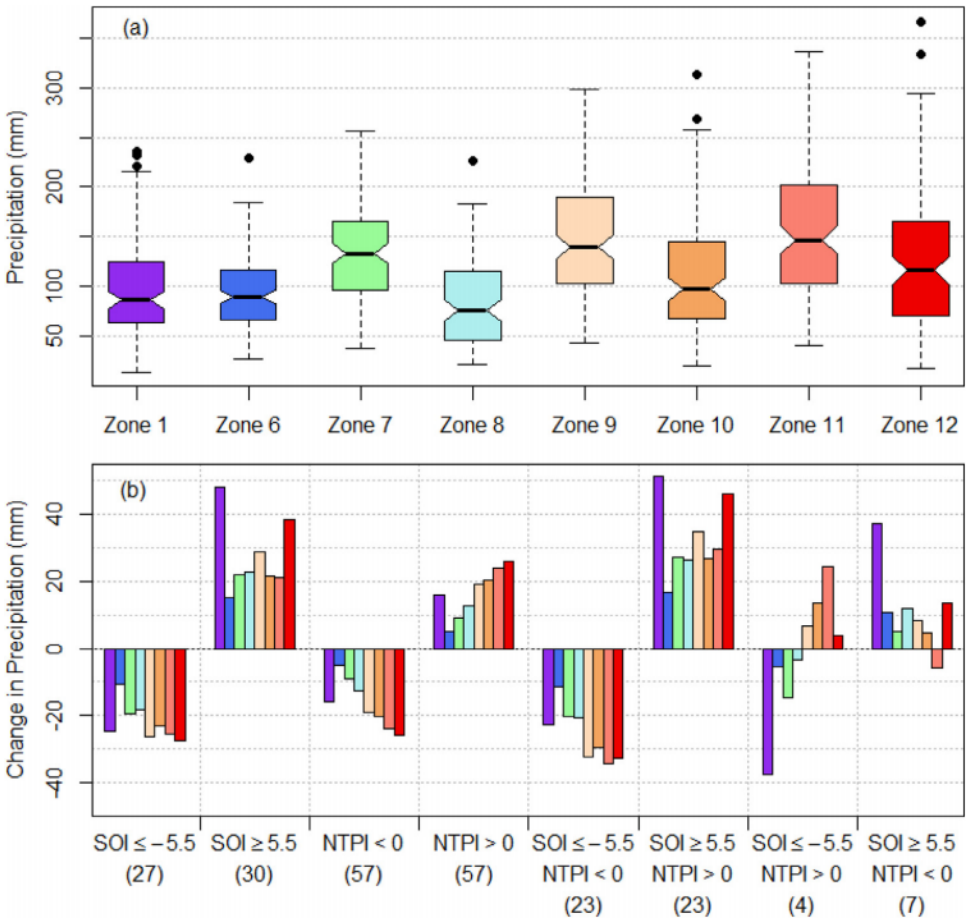


Fig. 4. (a) Boxplot of average SON precipitation at Australian Agro-ecological Zones. (b) Change in average SON precipitation anomalies for different phases of JJA SOI and JJA NTPI. Colours in Fig. 4b match boxplot colours in Fig. 4a. Numbers in parentheses show such events in the dataset.

von Mises (S_n) and Kolmogorov-Smirnov (T_n) statistics, proposed by (Genest et al., 2006; Genest and Rivest, 1993), were applied as the statistical GOF tests. Also, the Kendall's *tau*, maximized log-likelihood (ll_{max}) and Akaike and Bayesian Information Criteria (AIC and BIC, respectively) were determined to choose optimal copulas (Brechmann, 2010).

Copula parameters were estimated using maximum pseudo-likelihood method (Chowdhary et al., 2011), requiring observation pairs, $x_i = (x_{i1}, \ldots, x_{in})^T$, $i = 1,\ldots, n$, transformed to pseudo-observations, $\hat{u}_i = \left(\hat{u}_{i1}, \ldots, \hat{u}_{in}\right)$, $i = 1, \ldots, n$, *i.e.*, in the unit hyper-cube. Note that $\hat{u}_{i,j} = \left(1/(n+1)\right) rank\left(x_{i,j}\right)$, where $rank\left(x_{i,j}\right)$ is the rank, in ascending order between $i = 1, \ldots, n$ (Genest and Favre, 2007; Khedun et al., 2014). It is important to note that this ranking ensured that the dependence structure between pairwise data was independent of the marginal distribution (Genest and Favre, 2007). For a copula $C(u_1, \ldots, u_n; \theta)$, with density $c(u_1, \ldots, u_n; \theta)$, the parameter $\theta$was:

$$\hat{\theta} = \underset{\theta \in \Theta}{arg\max} \sum_{i=1}^{n} \log c\left(\hat{u}_{i1}, \ldots, \hat{u}_{in}; \theta\right). \tag{10}$$

In this study, elliptical (*i.e.*, Gaussian and *t*-copula with different degrees of freedom) and Archimedean (*i.e.*, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, and BB8) copulas were considered. Mixed two-parameter copula BB1, BB6, BB7, and BB8 were Clayton-Gumbel, Joe-Gumbel, Joe-Clayton, and Joe-Frank, respectively. We selected two-parameter copulas as they captured more than one type of dependence, *e.g.*, one parameter for the upper tail and lower tail dependence each, or one parameter for concordance while the other captures the lower tail dependence (Joe, 1996). Subsequently, these were used to investigate trivariate cases (SOI ∪ TPI *vs.* SON precipitation) based C-vine copulas. The more details of copula families, parameters, nonparametric dependence measures, and lower and upper tail dependence were fully described in the study of Brechmann and Schepsmeier (2013).

3.3.3. *Forecasting* via *optimal copulas*

114-years of data were partitioned in two sets; *viz* constructing and validation of the model. The constructing set contained 70% data where the average JJA values of SOI and NTPI, and the average SON precipitation anomaly matrices were stratified based on SOI phases followed by the NTPI phases, and representative samples were randomly selected from each stratum. This procedure, followed Khedun et al. (2014), avoiding a bias in the model, ensuring that the dependence structure was not impaired while enough datum points were obtained. Extraction of validation data once, and not repeating the statistical procedure separately for each climate division, allowed a comparison of the model performance across different climatic conditions. Copulas were fitted with the same procedure applied to the same set of copulas. The maximum pseudo-likelihood method was applied to ensure the estimated parameters reflected the multivariate dependence structure and was statistically independent of the selected marginal distribution. After selection of copulas, the response variable, conditioned upon the explanatory variable(s), was obtained for bivariate and trivariate model viz Eqs. (7) and (8), respectively.

Table 2
Selected marginal distributions with parameters, chi-square statistics, and *p*-values for average JJA SOI, NTPI and average SON precipitation anomaly.
2. Mathematical equations are:

Logistic: $f(x) = \frac{e^{\left(\frac{x-\mu}{\sigma}\right)}}{\sigma\left(1+e^{\left(\frac{x-\mu}{\sigma}\right)}\right)^2}$ Normal: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)}{2\sigma^2}}$

GEV: $f(x) = \frac{1}{\sigma} e^{-\left(1+k\frac{x-\mu}{\sigma}\right)^{\frac{-1}{k}}} \left(1+k\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{k}}$ Gamma: $f(x) = \frac{(x-\mu)^{k-1}}{\sigma^k \Gamma(k)} e^{-\frac{x-\mu}{\sigma}}$

Weibull: $f(x) = \frac{k}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{k-1} e^{-\left(\frac{x-\mu}{\sigma}\right)^k}$.

Climate Index	Distribution	Parameters	Chi-square	*p*-value
SOI	Logistic	σ = 5.24 μ = 0.04	4.16	0.84
NTPI	Normal	σ = 0.63 μ = 0.04	3.48	0.90

SON rainfall at Agro-ecological Zone #	Distribution	Parameters	Chi-square	*p*-value
1	GEV	k = 0.04 σ = 37.03 μ = −22.86	9.14	0.24
6	Gamma	k = 12.71 σ = 7.91 μ = −100.55	6.83	0.45
7	Weibull	k = 2.28 σ = 108.64 μ = −97.36	9.70	0.21
8	Gamma	k = 1.88 σ = 34.29 μ = −64.49	12.29	0.09
9	GEV	k = −0.13 σ = 52.83 μ = −25.08	12.62	0.08
10	Gamma	k = 2.46 σ = 39.23 μ = −96.50	3.97	0.78
11	Weibull	k = 1.94 σ = 135.30 μ = −119.96	4.10	0.77

4. Results and general discussion

The results are presented in two phases. We firstly present influences of synoptic-scale climate indices that moderate spring seasonal precipitation in the AEZs, followed by precipitation simulation using bivariate (pairwise SOI or TPI *vs.* precipitation) and trivariate (SOI & TPI *vs.* precipitation) models. Although published literature (*e.g.*, (McKeon et al., 2004)) showed that standard La Niña and El Niño phases are classified based on average SOI from June to November, we utilised average SOI from June to August since our aim was to forecast spring precipitation coinciding with the wheat growing season in the AEZs. The utilisation of three-month average SOI is an attempt to account for the influence of ENSO phases, as much as possible, and reduce the noise of SOI data instead of using the one-month lagged SOI that commonly has the highest correlation with precipitation (Schepen et al., 2012). Based on this, for El Niño events, there were three non-El Niño years in standard classification (average SOI based on June–November period) that were classified as El Niño years in our classification (average SOI based on June-August period), and seven El Niño years that became non-El Niño years. In a similar manner, such numbers of years for La Niña events were nine and five, respectively.

4.1. Effect of climate indices on seasonal precipitation

Table 1 shows Kendall and Spearman rank correlation coefficients for average JJA SOI and NTPI *vs.* average SON precipitation anomalies. Reflecting the spatial variabilities of climatic conditions and heterogeneous influences of synoptic-scale drivers, the correlation, bounded by [0.2, 0.5], was distinguishable across the AEZs. The highest correlation of SOI-precipitation anomalies appeared at Zone 1 in the northwest, while Zone 6 in the southwest showed the lowest level of such correlation and no statistically significant correlation with NTPI. It was noticed also that there were not any significant correlations between both indices and precipitation anomalies in the western Zones (Zones 2–5), to concur with previous observations (Kirono et al., 2010; Schepen et al., 2012). Interestingly, the correlation between NTPI and precipitation anomalies was lower than that in the case of SOI but more consistent with a correlation coefficient of about 0.4 across most of the wheat belt (excepting Zones 2–5, 7). This is expected since the variation of NTPI was on decadal time scales, and thus it is not sensitive to the seasonal changes of precipitation like SOI.

To assess precipitation anomalies in different phases of both indices, the average SON precipitation anomalies were plotted against the average JJA SOI and NTPI (Fig. 3). Zones 1, 7, 8, and 12 were selected for illustration since they were located in different regions from the north, east to the south representing four distinct climates. Fitted regressions (violet line) show the trend among the presented data. Visually, a significant correlation between precipitation anomalies and SOI in quadrant I (upper right quadrant) for all considered Zones was evident, which meant that if the preceding SOI was positive, there was a more likely chance of above normal precipitation. Further, only Zones 7 and 8 showed an association among the data in quadrant III (lower left quadrant), *i.e.*, if the preceding SOI occupied a negative value, the precipitation was below the climatological average. Further, the correlation pattern was dispersive in the neutral phase of SOI.

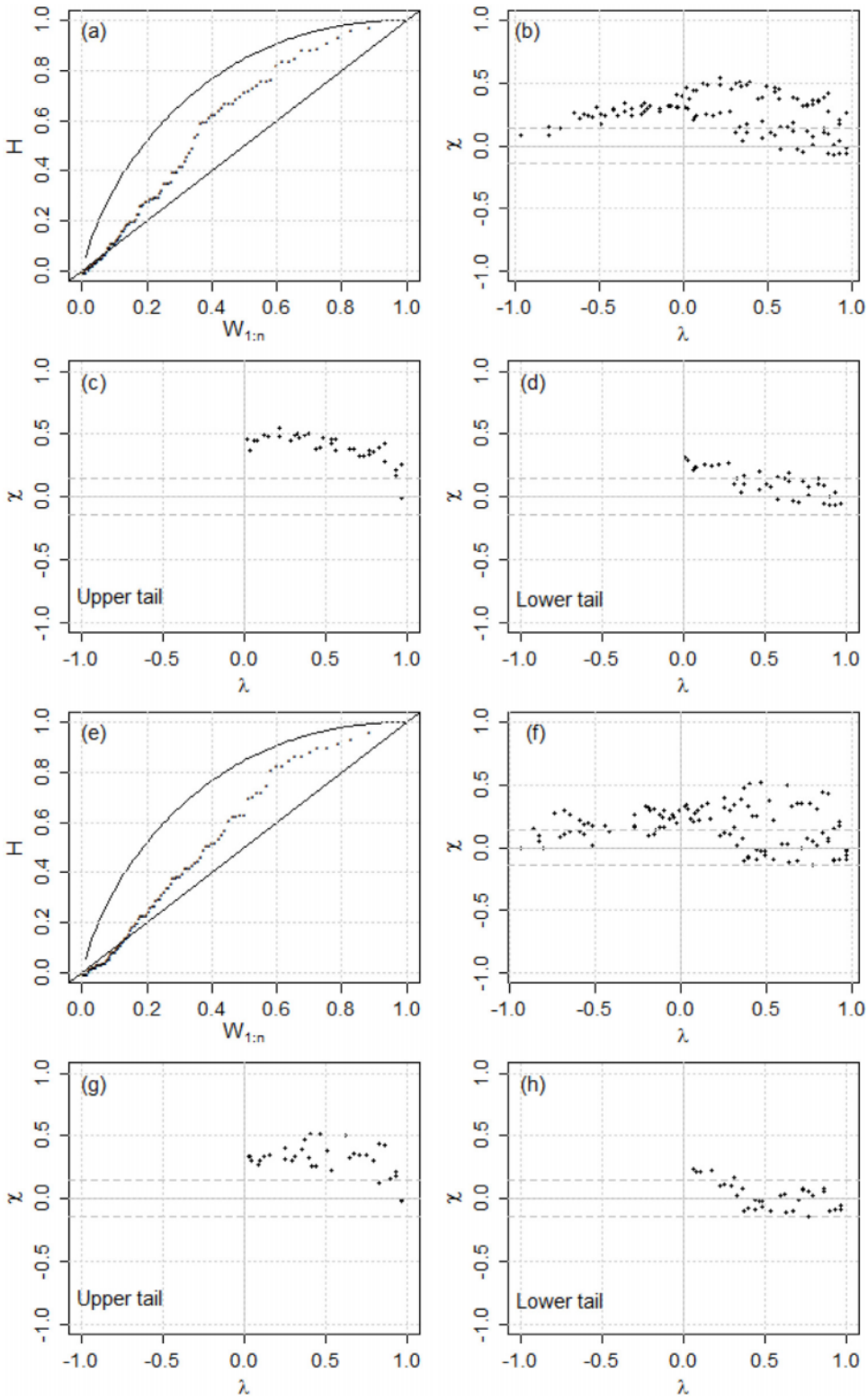


Fig. 5. Dependence between average JJA SOI, JJA NTPI and average SON precipitation anomaly for Zone 1 illustrated through (a) and (e) Kendall's plot, and (b–d) and (f–h) chi plots, respectively.

NTPI exhibited a lower correlation, indicated by large scatter (Fig. 3, right column). More interesting was the fact that there was a higher frequency of above or below average precipitation when both indices were in the same, compared to in opposite the phases (see the hit scores for each event in the four quadrants plotted in the lower right corners of the left column). An ostensible deduction, requiring a verification based on joint analysis of indices and precipitation thus stands; that the influence of ENSO and NTPI on precipitation patterns were not independent, in concurrence with McKeon et al. (2004) that synoptic-scale variation in precipitation across major parts in Australia's rangelands are related to both the ENSO and the IPO (*i.e.*, NTPI represented in our work). There-

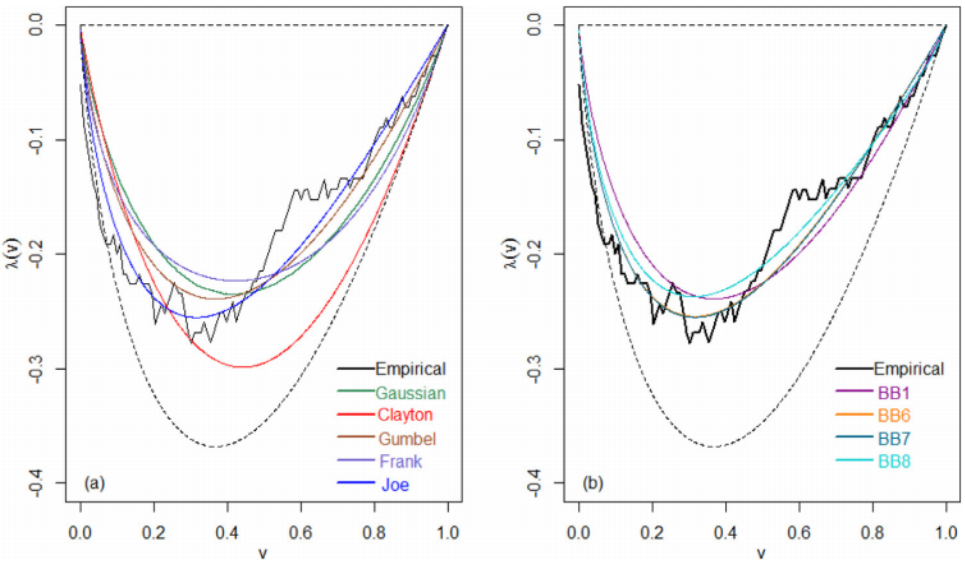


Fig. 6. Copula dependence between average JJA SOI and average SON precipitation anomaly for Zone 1 indicated through a comparison of empirical and theoretical λ-function of *one*- (a) and *two*-parameter (b) copula families. Dashed lines are limits corresponding to independence (Kendall's $\tau = 0$) and comonotonicity (Kendall's $\tau = 1$ ($\lambda = 0$)).

fore, copula-statistical models that jointly simulate precipitation data can be helpful in decision-making in regard to identifying the pivotal role of climate drivers on the hydrology of the region.

Fig. 4a shows a boxplot of average SON precipitation anomalies for all phases of JJA SOI and NTPI at each considered Zone. The number of times SOI was in El Niño and La Niña phases was different (Fig. 4b). When the state of SOI and NTPI were considered jointly, an equal number of events that both indices fell in the same phase was noted. The frequency of indices in the same phase, for example, SOI $\geq$ 5.5 and NTPI > 0, was threefold higher than their occurrence in different phases (*i.e.*, SOI $\geq$ 5.5; NTPI < 0) (Fig. 4b).

To better understand how different SOI and NTPI phases affect precipitation, we observe the bar plots showing precipitation anomalies for the respective individual or joint influences of climate indices (Fig. 4b). Considering Zone 12, for example; when a La Niña event was considered separately (ignoring NTPI), a positive anomaly (*i.e.*, surplus of precipitation) was indicated with a precipitation higher than average by 38.4 mm. Similarly, when an El Niño event alone was considered, an average deficit of 27.3 mm was noted. However, when the SOI and NTPI were positive, a surplus in average precipitation of about 46.0 mm was evident and when both were negative, an average deficit of 32.7 mm was evident. Other Zones showed a similar pattern except for Zone 1. When both climate indices were in negative phase, there was a reduction in the precipitation deficit by a small margin in Zone 1. Interestingly, Zone 6 exhibited no change in precipitation anomaly when NTPI was included, as the NTPI-precipitation correlation was not statistically significant (Table 1). When both climate indices occurred simultaneously, the change in precipitation was noticeably greater, emphasising the need for a joint probabilistic forecasting model (Section 4.2).

When both indices were in an opposite phase, the anomaly in precipitation exhibited a different pattern. In case of El Niño years and a positive value of the NTPI; the precipitation data showed a deficit for Zones 1, 6, 7 and 8, although an opposite result was attained for Zones 9, 10, 11 and 12. However, it is imperative to note that the La Niña years along with a negative phase of the NTPI facilitated more precipitation to all Zones, exccept for Zone 11. Importantly, Fig. 4b shows that when NTPI was negative (positive), this led to a negative (positive) precipitation anomaly but with negative NTPI and positive SOI, the precipitation anomaly was positive. This implies that the SOI (during a La Niña event) had a stronger effect on precipitation anomaly than the NTPI.

4.2. Copula-statistical bivariate and trivariate model

4.2.1. Marginal selection

In Table 2, we show the fitted marginal distributions with copula parameters, chi-square statistics and *p*-values for JJA SOI, NTPI and SON precipitation anomalies. Based on the chi-square statistic, JJA SOI and NTPI data can be appropriately modelled by Logistic and Normal distributions, respectively, confirmed by GOF plots (Fig. A.3 of Supplementary material).

Fig. A4 (Supplementary material) shows a density plot of the selected margins for SON precipitation anomaly at each Zone. It was evident that the marginal distributions varied with climatic conditions dominating each Zone. It was also interesting that all selected marginal distributions for SON precipitation anomalies were right-skewed with lower medians than the mean, indicating that precipitation in the spring season was generally below the climatological value. It was also evident that the Zones in the eastern region had a smaller peak, *i.e.*, lower kurtosis, implying more variability in the spring seasonal precipitation.

4.2.2. Bivariate models

We used K (a & e) and chi (b–d & f–h) plots (Fig. 5) to validate bivariate copulas based on average JJA SOI or NTPI *vs.* average SON precipitation anomalies. Zone 1 was selected for illustration since its SON precipitation anomalies had the highest correlation with JJA SOI (Table 1). Evidently, datum points located above the diagonal exhibit a significant positive dependence between pairwise average JJA SOI- and NTPI-average SON precipitation anomalies. This is confirmed in chi plots where the datum points in both cases were above the control boundary around zero (Fig. 5b & f). However, SOI-precipitation points had a larger deviation than NTPI-precipitation anomalies data, indicating that the degree of dependency of the former pairwise data was stronger. Some data points also fell into the two control bounds, and around zero, meaning that there was an insignificant dependence between pairwise variables since the values of χ_i were close to zero corresponding to $F\left(u_{i,1}, u_{i,2}\right) = F\left(u_{i,1}\right) F\left(u_{i,2}\right)$ (Eq. (11)). A stronger upper tail dependence was visible, pointing out that the influence of ENSO in a La Niña event was more evident than during an El Niño event. A

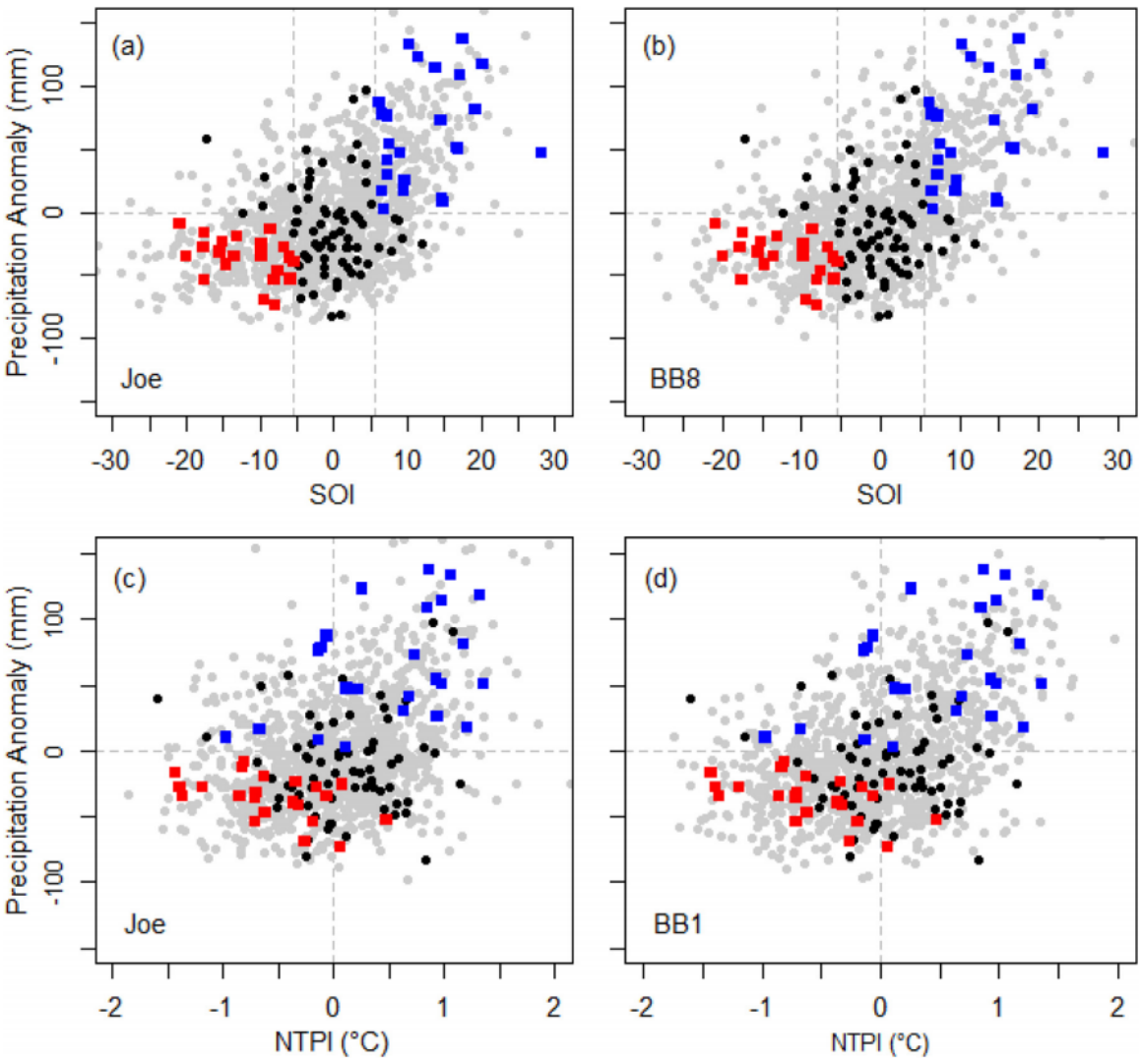


Fig. 7. Comparison of observed data with 1000 random samples generated from the Joe (a, c), BB8 (b), and BB1 (d) copulas (solid light grey dots) for Zone 1. Observed positive (negative) anomalies during La Niña (El Niño) events are shown in blue (red) and other events are shown in solid black dots (similar to Fig. 3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Copula parameters, Kendall's tau (τ), maximum log-likelihood (ll_{max}), AIC, BIC and Cramer-von Mises (S_n) and Kolmogorov-Smirnov (T_n) goodness-of-fit statistics along with their respective p-values, for each copula between average JJA SOI and average SON precipitation anomaly in Zone 1.

Copula Type	$\hat{\rho}$ or $\hat{\theta}$	$\hat{\delta}$	τ	ll_{max}	AIC	BIC	S_n	$p(S_n)$	T_n	$p(T_n)$
Elliptical Copula										
Gaussian	0.52		0.35	16.62	−31.24	−28.51	0.16	0.08	0.83	0.21
t	0.52		0.35	16.36	−28.70	−23.23	0.17	0.05	0.82	0.22
Archimedean Copula										
Clayton	0.48		0.19	5.33	−8.66	−5.93	0.39	0.00	1.51	0.00
Gumbel	1.54		0.35	19.63	−37.26	−34.53	0.08	0.32	0.73	0.32
Frank	3.43		0.34	16.11	−30.21	−27.48	0.15	0.04	0.81	0.16
Joe	1.88		0.33	21.26	−40.53	−37.79	2.09	0.52	2.33	0.57
BB 1	0.00	1.54	0.35	19.62	−35.24	−29.77	0.08	0.23	0.73	0.24
BB 6	1.89	1.00	0.33	21.26	−38.52	−33.05	0.04	0.84	0.66	0.41
BB 7	1.88	0.00	0.33	21.26	−38.51	−33.04	0.04	0.83	0.66	0.41
BB 8	2.25	0.96	0.36	22.31	−40.61	−35.14	0.03	0.91	0.55	0.79

similar pattern was attained for NTPI-precipitation cases. Hence, it is construed that the copulas (*i.e.*, Gumbel, Joe and mixed copulas) emphasising the upper tail dependence might be more appropriate for modelling the joint dependences.

The λ-function as characteristic of the copula family (Section 3.3.2) was adopted as the graphical tool to select bivariate models (Fig. 6). Zone 1 was chosen again for this illustration. Accordingly, Joe and BB8 (a mixed copula of Joe and Frank family) copulas seem to be appropriate for modelling the dependence between SOI and precipitation since their theoretical λ-functions are the closest to the empirical curve.

Table 4
Parameter ($\hat{\rho}, \hat{\theta}, \boldsymbol{\delta}$), Kendall's tau ($\boldsymbol{\tau}$), and p-values, for copula selected for modelling the dependence between average JJA SOI and average SON precipitation anomaly for selected Agro-ecological Zones.

Agro-ecological Zone #	Copula	$\hat{\rho}$ or $\hat{\theta}$	$\hat{\delta}$	τ	$p(S_n)$	$p(T_n)$
6	Gaussian	0.28		0.18	0.99	0.97
7	*t*-copula	0.38		0.25	0.96	0.81
8	Gumbel	1.38		0.28	0.63	0.35
9	BB7	1.37	0.43	0.29	0.93	0.73
10	Gumbel	1.25		0.20	0.72	0.64
11	Gumbel	1.28		0.22	0.29	0.36
12	Gumbel	1.38		0.27	0.34	0.37

The choice of copulas was cross-examined by the formal GOF tests (Section 3.3.2). Table 3 shows the statistic and p-values, computed using bootstrapping procedure based on random samples ($n = 1000$). It was clear that the Clayton and Frank copulas should be rejected at $p = 0.05$, which also concurs with maximum log-likelihood, AIC and BIC. As confirmed by Fig. 6, Joe and BB8 copulas are suitable for modelling the dependence structure between average JJA SOI and average SON precipitation anomaly in Zone 1. It is also noted that, according to AIC and BIC (Brechmann, 2010), the penalty for *two*-parameter families was stronger than *one*-parameter families using BIC compared to the AIC.

Fig. 7 illustrates the scatter plots of observed data with 1000 random samples generated via Joe, BB1, and BB8 copulas for Zone 1. Positive (negative) anomalies, similar to Fig. 3, during El Niño (La Niña) were identified. The spread of different copulas was different although they all enveloped the observations but exhibited different tail behaviours. Joe copula was somewhat restricted in its range of correlations and upper tail dependence, while BB8 copula was close to the Joe copula but had a larger spread in the upper tail. Fig. 7c-d compares Joe and BB1 copulas for JJA NTPI – SON precipitation anomaly pair. By the combination of graphical tool and statistical GOF tests, the BB8 and Joe copulas were selected to model the dependence structure of the pairwise JJA SOI – SON precipitation anomaly and JJA NTPI – SON precipitation anomaly, respectively, for Zone 1.

Table 4 shows optimal copulas and associated GOF statistics for the rest of the AEZs. While most copula families had similar forms in their central part, they exhibited significant differences in the tails (or extreme ends). As depicted in Fig. 1, each AEZ has diverse climatic conditions, so the degree of dependence between SON precipitation anomalies and JJA SOI is different. It leads to the difference of copula families selected to model such a dependence relationship at each Zone.

Gaussian and *t*-copula were suitable for modelling the dependence between JJA SOI and SON precipitation anomalies for Zones 6 and 7, respectively. The results implies that the impact of extreme ENSO events on the spring precipitation anomalies in these two AEZs is weak, so no tail dependence exists. It is noted that the difference between Gaussian and *t*-copula was in the tail areas, wherein a *t*-copula exhibited symmetric tail dependence but a Gaussian copula did not. However, if the degree of freedom was large, a *t*-copula is mirrored like a Gaussian copula (Khedun et al., 2014). In terms of the results, the Gumbel copula was most suitable for Zones 8, 10, 11, and 12, located in the east and southeast of the wheat belt, exhibiting a strong upper tail dependence, presumably linked to the dominant influence of La Niña events on SON precipitation anomalies. For Zone 9, BB7 (*i.e.*, mixed Joe-Clayton) was suitable, emphasising the upper as well as the lower dependence.

The same set of copulas considered above were applied for modelling the dependence structure between average JJA NTPI and average SON precipitation anomalies. Consequently, Frank and Gaussian copulas, exhibiting symmetric dependences, were suitable for modelling such a dependence across all Zones, accept for Zones 1 and 10 which indicated Joe and Gumbel, respectively, as the optimal copulas. Differences between Gaussian and Frank copulas are mainly in the central part, where the latter has a stronger dependence. Gaussian copula, however, is stronger at the tail.

Fig. 8 shows boxplots of observed and simulated spring precipitation anomaly for different JJA SOI ranges for Zones 1, 7, 8, and 12. Notched boxplots were selected as a graphical tool to illustrate the significance between observed and simulated data. Considering the entire range ($-30 < \text{SOI} < 30$), the interquartile and whisker ranges and medians of the observed and simulated data were approximate at the same level, except for the upper quartile of Zone 8. It is noted that the medians of precipitation anomalies in all considered Zones were lower than the long-term means (*i.e.*, below zero value), showing a higher probability of precipitation deficit. In general, there was a sound agreement of the distributions between the simulated values and observations. Further, if a specific range of JJA SOI was considered, *e.g.*, $\text{SOI} \leq -5.5$, there was a better agreement of values between simulation and observations than in case of $\text{SOI} \geq 5.5$ for Zone 7, 8 and 12. Zone 1 had a better result in case of $\text{SOI} \geq 5.5$ compared to that in case of $\text{SOI} \leq -5.5$, which concurs with a stronger dependence for the upper tail depicted in Fig. 5b.

4.2.3. Trivariate models

Seven AEZs with significant correlations in terms of both indices were selected for building the trivariate models. NTPI was selected as the first root node, followed by SOI and precipitation anomaly to

Table 5
Parameters ($\hat{\rho}, \hat{\theta}, \hat{\delta}$), Kendall's tau, maximum log-likelihood (ll_{max}), AIC, BIC, and independence test with p-values, for C-vine copula selected for modelling the dependence between average JJA SOI, JJA NTPI and average SON precipitation anomaly for selected Agro-ecological Zones.

	Zone 1		**Zone 7**	**Zone 8**	**Zone 9**	**Zone 10**	**Zone 11**	**Zone 12**
Tree 1	**Joe**	***t*** [a]	**Frank**	**Gaussian**	**Frank**	**Gumbel**	**Frank**	**Frank**
$\hat{\rho}, \hat{\theta}, \hat{\delta}$	1.60	0.68	1.93	0.39	2.69	1.31	2.52	2.68
Kendall's tau	0.25	0.48	0.21	0.26	0.48	0.24	0.26	0.28
ll_{max}	12.61	35.97	5.21	8.51	9.97	8.51	9.03	9.85
AIC	−23.21	−67.94	−8.42	−15.02	−17.94	−15.03	−16.05	−17.71
BIC	−20.48	−62.47	−5.68	−12.29	−15.20	−12.29	−13.31	−14.97
p-value	2.66e − 4	6.53e − 14	1.16e − 3	8.29e − 5	1.87e − 5	8.99e − 5	3.67e − 5	1.64e − 5
Tree 2	**BB8**		**Clayton**	**Gumbel**	**Gaussian**	N/A	N/A	**Gumbel**
$\hat{\rho}$ or $\hat{\theta}$ ($\hat{\delta}$)	2.11 (0.89)		3.62	1.20	0.28	N/A	N/A	1.20
Kendall's tau	0.48		0.15	0.17	0.18	N/A	N/A	0.16
ll_{max}	12.70		3.22	5.01	4.41	N/A	N/A	4.98
AIC	−21.40		−4.45	−8.02	−6.83	N/A	N/A	−7.97
BIC	−15.93		−1.71	−5.28	−4.09	N/A	N/A	−5.23
p value	3.45e − 5		0.01	9.78e − 3	1.72e − 2	0.24	0.20	1.67e − 2

[a] Student's *t* copula selected for modelling dependence between average JJA SOI and average JJA NTPI in Tree 1 is the same for every considered Zone.

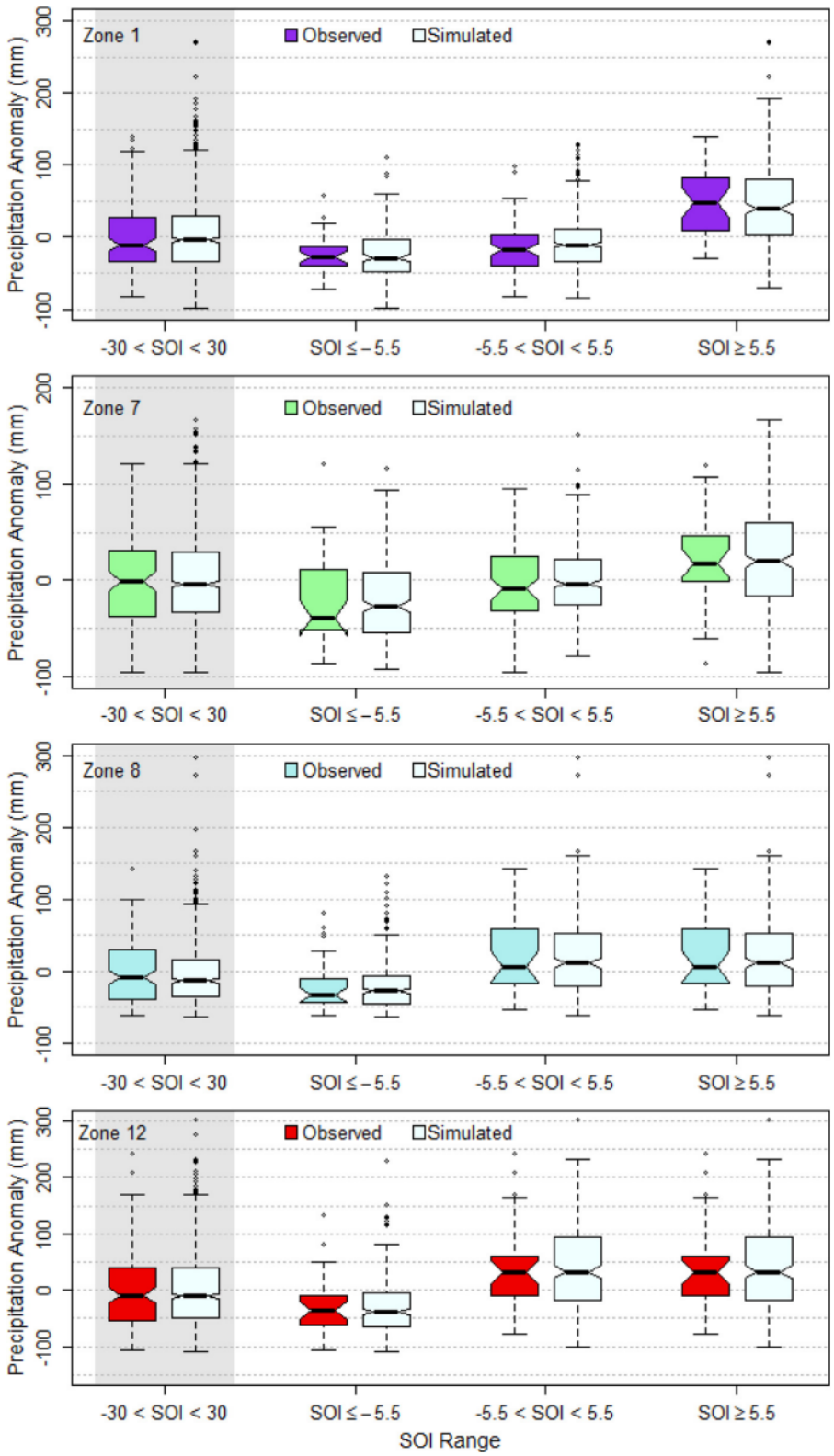


Fig. 8. Boxplots of observed and simulated spring precipitation anomalies for different SOI ranges for Agro-ecological Zones 1, 7, 8, and 12. Boxplots with grey background are for the entire SOI range.

structure the C-vine trees. Importantly, C-vine struture was based on the stacking of ordered conditional bivariate copulas (Aas et al., 2009), so this procedure allowed the conditional copula evaluation at SOI for given NTPIs to account for the inter-decadal modulations induced by the NTPI phases on the SOI effects on precipitation anomalies.

Table 5 shows the parameter estimates, log-likelihood, AIC and BIC, and *p*-value of the independent test. Here, the independence test was a simple bivariate tool based on Kendall's τ (Genest and

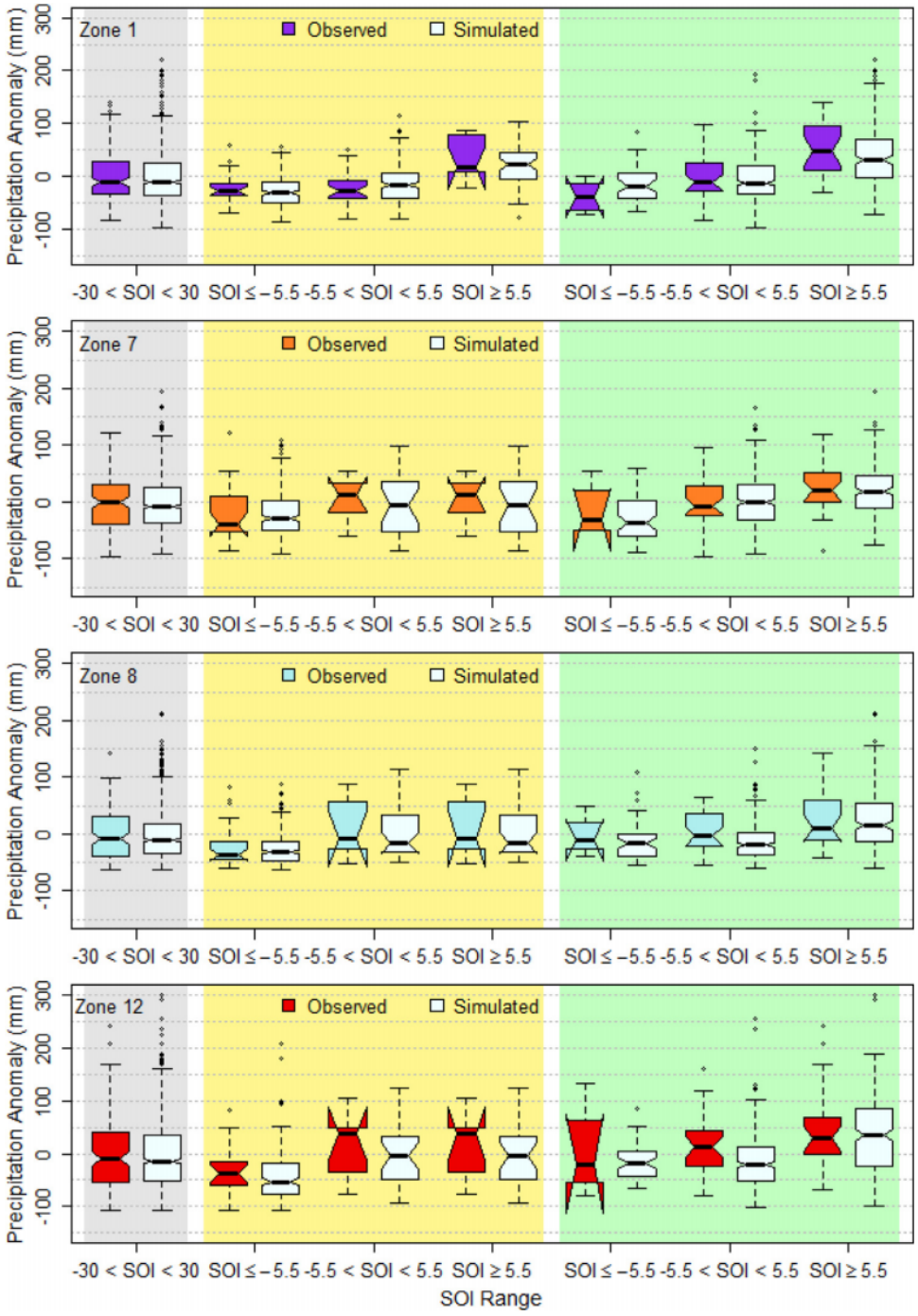


Fig. 9. Boxplots of observed and simulated spring precipitation anomalies for different JJA SOI and JJA NTPI ranges for Agro-ecological Zones 1, 7, 8 and 12. Boxplots with the light grey back ground are for entire JJA SOI and JJA NTPI ranges and light yellow (green) is for different SOI ranges and negative (positive) NTPI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Favre, 2007), taking the advantage of the asymptoticity of the test statistic $T = \sqrt{[9n(n-1)]/[2(2n+5)]}|\hat{\tau}|$, where n = number of observations, $\hat{\tau}$ = empirical Kendall's τ and p-value of null hypothesis of bivariate independence was $p = 2 \times [1 - \Phi(T)]$. Evidently, Zones 10 and 11 were not considered for trivariate modelling based on the independence test, at a confidence level of 95%.

Fig. 9 is similar to Fig. 8; however, including the impact of NTPI. Again, there is no significant difference for the distributions if considering the entire range of SOI independently with NTPI. When data were stratified into different cases for negative and positive NTPI values (represented in light yellow and light green, respectively), the influences of NTPI in model performance varied for both NTPI phases. The absolute difference between mean and median values of observed and simulated SON precipitation anomalies in both models are summarised (Table 6). In general, a trivariate model did not appear to improve the mean simulated value, except for Zones 8 and 12 when a La Niña event occurred concurrently with the positive phases of NTPI. However, median values simulated from the trivariate model were lower than those from the bivariate model for all Zones except for Zone 12 (in all cases) and Zone 1 (in case of negative phases of both climate indices).

Table 6
Comparison of means and medians for observed and simulated SON precipitation anomaly (mm) from the bivariate and trivariate models for the selected Agro-ecological Zones.

SOI and NTPI range	Bivariate Model						Trivariate Model					
	SOI ≤ −5.5			SOI ≥ 5.5			SOI ≤ −5.5 and NTPI < 0			SOI ≥ 5.5 and NTPI > 0		
	Ob.	Sim.	AD	Ob.	Sim.	AD	Ob.	Sim.	AD	Ob.	Sim.	AD
Zone 1												
Mean	−24.74	−25.11	0.37	47.92	42.44	5.48	−22.53	−28.03	5.50	51.18	43.73	7.45
Median	−27.44	−28.91	1.47	47.77	37.05	10.72	−27.44	−31.33	3.89	48.10	37.90	10.20
Zone 7												
Mean	−19.41	−19.64	0.23	21.96	20.09	1.87	−20.27	−23.41	3.14	27.09	21.26	5.83
Median	−38.97	−27.58	11.39	17.85	20.89	3.04	−38.97	−28.86	10.11	19.51	19.05	0.46
Zone 8												
Mean	−18.24	−19.70	1.46	22.85	29.34	6.49	−20.84	−25.05	4.21	26.13	30.47	4.34
Median	−32.71	−27.55	5.16	6.74	18.72	11.98	−37.03	−31.98	5.05	9.97	20.91	10.94
Zone 12												
Mean	−27.29	−29.04	1.75	38.39	41.70	3.31	−32.70	−39.07	6.37	45.98	43.02	2.96
Median	−36.82	−38.22	1.40	31.35	37.12	5.77	−37.35	−48.24	10.89	28.86	38.39	9.53

Ob.: Observation; Sim.: Simulation; AD: Absolute Difference.

4.2.4. Spring precipitation forecast

In Figs. 10 and 11, we show a joint probabilistic forecasting model, expressed via conditional copulas for bivariate and trivariate cases, respectively, at Zone 1. An interpretation of this is relatively straightforward. Considering the bivariate case, for example; given the SOI = −10 (*i.e.*, extreme El Niño event) and SOI = 5 (neutral), the probability for precipitation anomaly of 25 mm below the average could be as much as 0.57 and 0.21, respectively. This indicates that there was a greater chance for a marked deficit in precipitation in the El Niño year. Similarly, the probability of the precipitation anomaly lower 50 mm can be only 0.33 given an SOI value of 20 (*i.e.*, La Niña event) while it could be as much as 0.81 given an SOI value of zero (neural phase).

Using bivariate and trivariate copulas, forecasted seasonal precipitation anomalies conditional upon different ranges of NTPI and SOI data were acquired using Eq. (8). The forecasted data against the observed data for the bivariate (*i.e.*, SOI *vs.* precipitation anomaly) and the trivariate (*i.e.*, SOI, NTPI *vs* precipitation anomaly) cases, in Zones 1, 7, 8 and 12 are shown as Fig. 12a–b. For bivariate cases, the Spearman correlation coefficients between observed and forecasted anomaly were approximately 0.59, 0.26, 0.48 and 0.49 in Zones 1, 7, 8, and 12, respectively. When NTPI was included, however, such correlation coefficients dropped to 0.50 in Zone 1 but increased to 0.52, 0.49 and 0.52, respectively, in Zones 7, 8 and 12. Further, it is more interesting that such correlation in the upper right and lower left quadrants (when both SOI and NTPI are in the same phases) was significantly improved in the trivariate model. The correlation coefficients in the upper right (lower left) quadrant increased from 0.72 (0.24), 0.27 (0.37), 0.27 and 0.10 (bivariate model) to 0.74 (0.52), 0.29 (0.54), 0.32 and 0.64 for Zones 1, 7, 8, and 12, respectively. This indicated that trivariate model was effectively better for forecasting the spring precipitation anomalies based on the stages of average JJA SOI and TPI. Also, it is clear that the influence of ENSO on these Zones remained more prevalent than the IPO (TPI) in AEZs considered in this paper. Thus, this verifies the primary role of ENSO to be used in seasonal precipitation forecasting (*e.g.*, (Meinke and Hochman, 2000)), although the inclusion of other synoptic-scale indices could further enhance the accuracy.

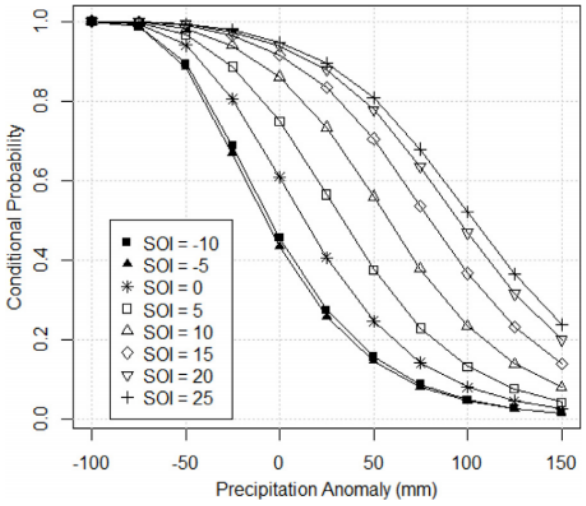


Fig. 10. Conditional distributions of precipitation anomalies of Zone 1 given different SOI values.

5. Further discussion

Probabilistic-based precipitation forecast models such as that adopted in this study are considered as essential tenets for end-users like agriculturalists, irrigators, resource managers and drought planners to develop management strategies for informed decision-making (Goddard et al., 2001). In this paper, the utility of vine copulas (Brechmann and Schepsmeier, 2013), that so far remained unexplored for precipitation forecasting in Australia's wheat belt Agro-ecological Zones (AEZs), has been demonstrated. Following evidence of a lagged relationship between climate indices and precipitation in Schepen et al. (2012), we have demonstrated the capability of vine copulas to jointly model the relationship between two dependent major synoptic-scale, lagged climate drivers (SOI and TPI) on spring precipitation anomaly. Lagged SOI has been found to be useful for forecasting Australian seasonal precipitation in some regions, and some seasons, if not all (Chiew et al., 2003; McBride and Nicholls, 1983; Stone et al., 1996). A strong evidence of forecasting northern and eastern Australian precipitation from August–October to November–January and a weak evidence from March–May to May–July with lagged climate indices were noted (Schepen et al., 2012). To attain comprehensively evaluated results, both the bivariate (SOI *vs.* precipitation; TPI *vs.* precipitation) and trivariate (conjoint SOI and TPI *vs.* precipitation) models were formulated. In accordance with evaluation metrics, the results captured emphatically the joint dependence structure between predictors and predictand, highlighting the importance of jointly forecasting precipitation where correlation

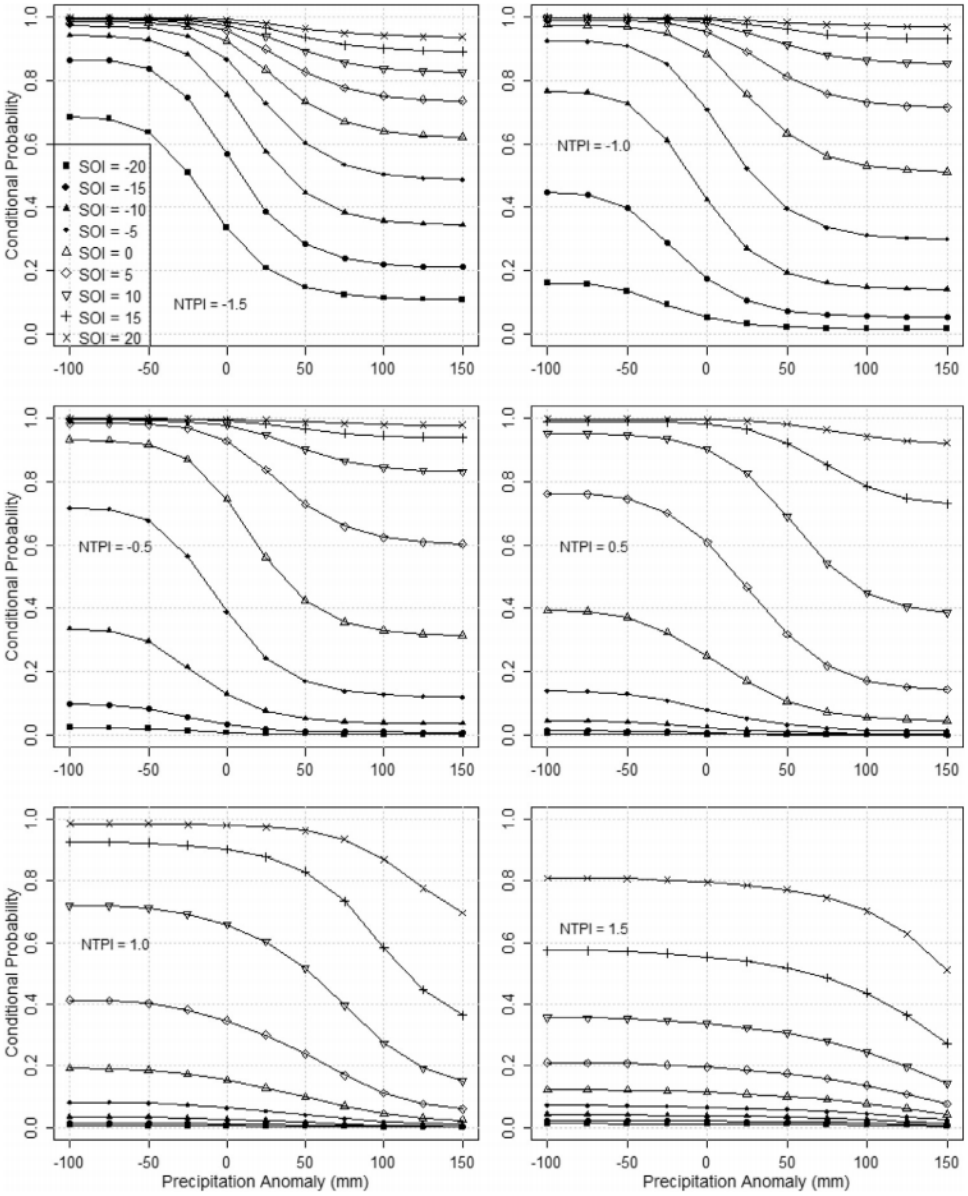


Fig. 11. Conditional distributions of SON anomalies of Zone 1 given different SOI and NTPI values.

between a set of climate indices and the precipitation data can be established.

We emphasise that this paper was a first research to adopt vine copulas to forecast precipitation anomalies in Australia's Agro-ecological Zones; and in fact, yielded very satisfactory simulations by utilising the most relevant, yet a limited set of indices (*i.e.*, only SOI and TPI). It is possible, however, that the other climate indices not investigated in this paper could be tested as alternative variables for seasonal forecasting; providing a better response of the models. This study, which applied vine copulas no doubt established their prime importance to create a pathway for a follow-up work with other synoptic-scale indices (*e.g.*, Madden-Julian Oscillation Index, Quasi-biennial Oscillation Index; ENSO Modoki Index, etc.) (Hudson et al., 2011; Lau and Sheu, 1988; Marshall et al., 2016). Follow-up studies could apply the developed models to data from other seasons (*e.g.*, December–February; March–May and June–August that exhibit different climatic conditions).

The study has focused on Australia's wheat belt where seasonal forecasting is considered a very important aspect of sustainable agricultural management (Moeller et al., 2008; Stephens et al., 1994). In order to achieve this, La Niña and El Niño events were classified based on a three monthly (June–August) average. It should be noted that a formal classification of the ESNO events in Australia follows the six monthly (June–November) SOI average. Future work could apply the averaged six-month SOI and extract precipitation data associated with each ENSO event based on the available classification scheme by the Australian Bureau of Meteorology, and thus develop a wide range of dependence-based models for other periods that were not covered in this study.

The results are highly significant as the adopted vine copulas offer benefits by modelling high-dimensional dependency of bivariate or trivariate predictor-predictand data compared to 'standard' copulas. Vine copulas have better computational tractability (Bedford and Cooke, 2002; Joe, 1996), are flexible in deducing bivariate family copulas for each edge (Gyasi-Agyei, 2011) and can identify the dependence structure for simulating tail dependence (Joe et al., 2010) which must be considered in precipitation *vs.* climate index models. By decomposing high-dimensional multivariate density into bivariate copulas, the association between, say X_1–X_3, can be statistically modelled by the bivariate copula

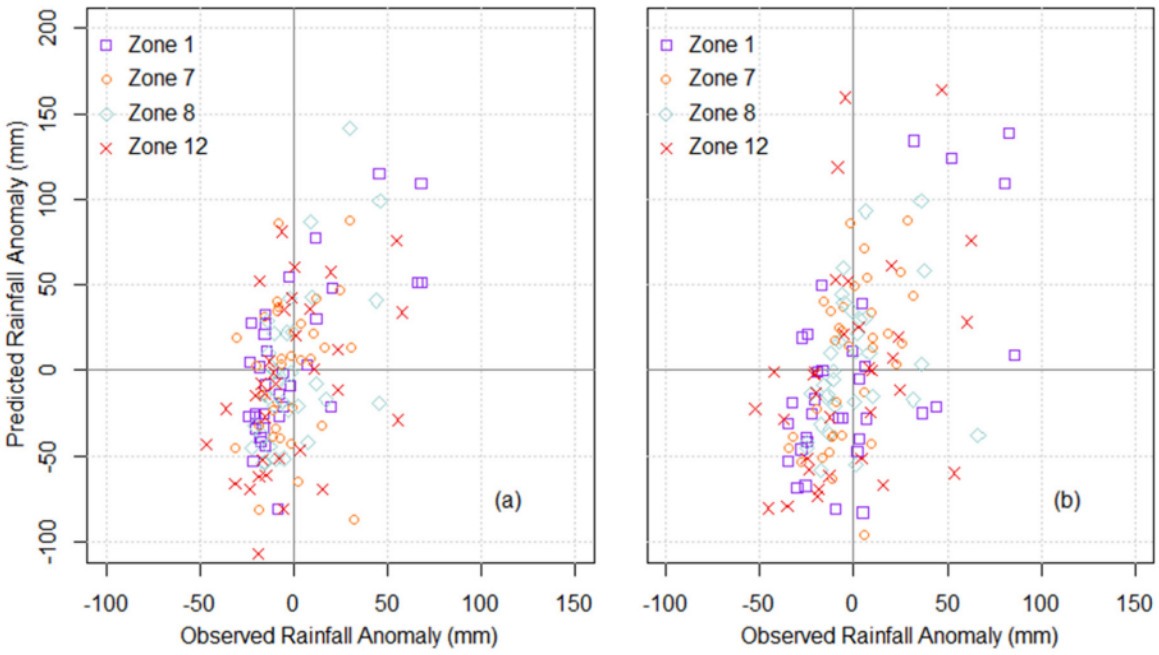


Fig. 12. Scatter plot of predicted versus observed average spring precipitation anomaly for (a) SOI and precipitation anomaly models and (b) SOI, NTPI, and precipitation anomaly models.

density $c_{12}[F(x_1), F(x_2)]$ of X_1 and X_2, and another copula density $c_{23}[F(x_2), F(x_3)]$ of X_2 and X_3. Opportunities exist in applying vine bivariate copulas with tree structures, allowing modellers to examine in a flexible manner the tail dependence in precipitation forecasting not only in Australia but elsewhere. It is construed that an implementation of vine copulas offers a potential for dependence-based modelling in other regions, including an exploring of interactions between predictors and predictand from a joint influence perspective. Our study has set a baseline for incorporating multiple climate indices to model extreme events (*e.g.*, drought or floods) (Shafaei et al., 2016) and dependence structure between temperature variables (*e.g.*, heatwaves) (Erhardt, 2013) where tail dependence is a usual phenomenon.

In summary, decomposition of vine copula was based on ordered bivariate conditional copulas, in which the roots of each tree need to be identified. According to Aas et al. (2009), there exist, for example, $d!/2$ different C-vines on d nodes, indicating that it is important to select an appropriate structure for dependence-based modelling when vine copula functions are adopted. Furthermore, a computation of joint CDF of a vine copula is definitely a challenging task, except for some special cases where all paired copulas are nominally Gaussian.

6. Conclusions

Seasonal precipitation in Agro-ecological Zones (AEZs) is influenced by synoptic-scale oceanic atmospheric-oceanic circulation patterns, where a dominant role of the Southern Oscillation Index is well-established (Chowdhury and Beecham, 2010; Mekanik and Imteaz, 2013; Schepen et al., 2012). In this paper we employed SOI, coupled with IPO Tripole Index (TPI) which is a temperature anomaly identifying for IPO evolution used to track ENSO, *e.g.*, Niño-3.4 index (Henley et al., 2015), as predictors to jointly model spring seasonal precipitation in AEZs. To illustrate the usefulness of indices for conditional-based precipitation forecasting, vine copula models were evaluated for all AEZs, except those in the western area, where statistically significant lagged correlations existed between June-August SOI and NTPI, and spring (SON) precipitation. Spatial correlation of climate indices across different AEZs was not identical, where a higher correlation with SOI was recorded compared to a lower, negative correlation with TPI. However, results showed that the impact of NTPI can be relatively significant for precipitation forecasting when a dependence structure between precipitation and two synoptic-scale drivers in a trivariate copula model is established.

In general, NTPI is known to modulate the strength of ENSO, and thus, is expected to lead to a stronger response when the two synoptic-scale events are coincident, particularly when they are both positive, rather than being out of phase. In this study, a copula-based model was assessed in terms of how well ENSO conditions by themselves could forecast seasonal precipitation and if an inclusion of the phasal state of NTPI did improve the results. Copulas were applied to model bivariate dependence between SOI and precipitation anomaly and trivariate dependence on precipitation anomaly and joint effects of SOI and NTPI. Stronger upper tail dependence was clearly visible, suggesting that the influence of ENSO during a La Niña was more evident on precipitation modulation than during an El Niño event.

To comprehensively evaluate the models, ten copulas ranging from elliptical to Archimedean were explored via goodness-of-fit and visual measures. Four of the copulas were *two*-parameter copulas, as they captured more than one type of dependence structure in predictor-predictand data. Copulas were fitted using the maximum pseudo-likelihood method, ensuring an independence of the choice of the marginal distribution of the optimal copulas. Not surprisingly, different copulas were found to be suitable for different regions within the AEZs, and the choice of the forecasting model was found to be strongly driven by the upper tail dependence associated with the La Niña conditions and positive precipitation anomalies.

Bivariate and trivariate models showed that when the entire SOI and NTPI data were considered, both models generated interquartile ranges and statistics similar to observations. However, when data were stratified in different SOI cases, in the bivariate model; a better agreement in the negative range was met compared to the positive range. It was evident that inclusion of NTPI in the trivariate model improved the results in terms of Spearman's rank coefficient

for three out of four AEZs (*i.e.*, Zones 7, 8 & 12) whereas the performance was worse for Zone 1 implying that NTPI is not suitable for precipitation forecasting in this Zone (*i.e.*, northwest Australia). However, simulations with a bivariate model (employing SOI *vs.* precipitation) attained good performance in Zone 1. Further, the trivariate model has been found to be effective for forecasting the spring precipitation anomalies associated with the stages of average JJA SOI and JJA TPI. It is imperative to note that the TPI representing long-term climate variability, shifts its phases on at least an inter-decadal time scale, usually 20–30 years. We ascertain that JJA SOI can be used in conjunction with TPI as a forewarning index prior to the start of the spring season, for many of the AEZs, if not all.

Inclusion of TPI (as a recent index for IPO) for the first time in a precipitation forecasting problem was successful in attaining more accurate forecasts for three out of the four Zones. Verdon et al. (2004) found that the negative phase of IPO enhances the impact of La Niña on precipitation and streamflow in eastern New South Wales and Victoria, and in particular, the impact of ENSO alone is weak in Victoria but it is significantly enhanced in the negative phase. Other work (*e.g.*, Kiem and Franks (2004)) found that IPO modulate the frequency of ENSO (in particular the occurrence of La Niña), noting the fact that the negative IPO phase increases the magnitude and variability of precipitation and runoff compared to a positive phase. Consideration of TPI as an evolution of IPO can, therefore, be included in precipitation forecasting problems in agricultural regions where this index can act to modulate the ENSO phenomenon (*e.g.*, Australia's wheat belt).

Acknowledgments

The project was financed by University of Southern Queensland Postgraduate Research Scholarship (USQPRS 2015–2017), School of Agricultural, Computational and Environmental Sciences and Strategic Research Funding (SRF) Projects (Resilient Landscapes SRF and Computational Models SRF). The authors would like to acknowledge constructive comments from the reviewers and the helpful proofreading work from Dr Barbara Harmes.

The data were acquired as follows: Australian Bureau of Meteorology (SOI), Commonwealth Scientific and Industrial Research Organisation (AWAP) and Earth System Research Laboratory, Physical Science Division (TPI), which are greatly acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agwat.2017.06.010.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Math. Econ. 44, 182–198.
Aghakouchak, A., Ciach, G., Habib, E., 2010. Estimation of tail dependence coefficient in rainfall accumulation fields. Adv. Water Resour. 33, 1142–1149.
An-Vo, D.-A., Mushtaq, S., Nguyen-Ky, T., Bundschuh, J., Tran-Cong, T., Maraseni, T., Reardon-Smith, K., 2015. Nonlinear optimisation using production functions to estimate economic benefit of conjunctive water use for multicrop production. Water Resour. Manage. 29, 2153–2170.
Anwar, M.R., O'leary, G., McNeil, D., Hossain, H., Nelson, R., 2007. Climate change impact on rainfed wheat in south-eastern Australia. Field Crops Res. 104, 139–147.
Bedford, T., Cooke, R.M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. intell. 32, 245–268.
Bedford, T., Cooke, R.M., 2002. Vines: a new graphical model for dependent random variables. Ann. Stat., 1031–1068.
Best, P., Stone, R., Sosenko, O., 2007. Climate risk management based on climate modes and indices-the potential in Australian agribusinesses. 101st Seminar July 5–6 2007 Berlin Germany. European Association of Agricultural Economists.
Brechmann, E.C., Schepsmeier, U., 2013. Modeling dependence with C-and D-vine copulas: the R-package CDVine. J. Stat. Softw. 52, 1–27.
Brechmann, E.C., 2010. Truncated and Simplified Regular Vines and Their Applications. Diploma Thesis. University of Technology, Munich, Germany (URL http://mediatum.ub.tum.de/doc/1079285/1079285.pdf).
Chen, X., Koenker, R., Xiao, Z., 2009. Copula-based nonlinear quantile autoregression. Econ. J. 12, S50–S67.
Chiew, F., Leahy, M., 2003. Inter-decadal Pacific Oscillation modulation of the impact of El Nino/Southern Oscillation on Australian rainfall and streamflow. Proceedings of the International Congress on Modelling and Simulation (MODSIM 2003), 100–105.
Chiew, F.H., Piechota, T.C., Dracup, J.A., McMahon, T.A., 1998. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. J. Hydrol. 204, 138–149.
Chiew, F., Zhou, S., McMahon, T., 2003. Use of seasonal streamflow forecasts in water resources management. J. Hydrol. 270, 135–144.
Chowdhary, H., Escobar, L.A., Singh, V.P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. Hydrol. Res. 42, 193–216.
Chowdhury, R., Beecham, S., 2010. Australian rainfall trends and their relation to the southern oscillation index. Hydrol. Processes 24, 504–514.
Cleugh, H., Cleugh, H., Smith, M.S., Battaglia, M., Graham, P., 2011. Climate Change: Science and Solutions for Australia. CSIRO.
Deo, R.C., Kisi, O., Singh, V.P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175.
De Michele, C., Salvadori, G., 2003. A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. J. Geophys. Res.: Atmospheres, 108.
Erhardt, T.M., 2013. Predicting Temperature Time Series Using Spatial Vine Copulae. Master's Thesis. Technische Universität, München http://mediatum.ub.tum.de/node.
Evin, G., Favre, A.C., 2008. A new rainfall model based on the Neyman-Scott process using cubic copulas. Water Resour. Res., 44.
Fang, Y., Madsen, L., 2013. Modified Gaussian pseudo-copula: applications in insurance and finance. Insurance: Math. Econ. 53, 292–301.
Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resour. Res. 40.
Fischer, M., Köck, C., Schlüter, S., Weigert, F., 2009. An empirical analysis of multivariate copula models. Quant. Finance 9, 839–854.
Fisher, N., Switzer, P., 2001. Graphical assessment of dependence: is a picture worth 100 tests? Am. Stat. 55, 233–239.
Folland, C., Renwick, J., Salinger, M., Mullan, A., 2002. Relative influences of the interdecadal Pacific oscillation and ENSO on the South Pacific convergence zone. Geophys. Res. Lett. 29.
GRDC, 2012. GRDC Agroecological Zones Grains Research and Development Corporation.
Genest, C., Favre, A.-C., 2007. Everything you always wanted to know about Copula modeling but were afraid to ask. J. Hydrol. Eng. 12, 347–368.
Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. J. Am. Stat. Assoc. 88, 1034–1043.
Genest, C., Quessy, J.F., Rémillard, B., 2006. Goodness-of-fit procedures for copula models based on the probability integral transformation. Scand. J. Stat. 33, 337–366.
Goddard, L., Mason, S.J., Zebiak, S.E., Ropelewski, C.F., Basher, R., Cane, M.A., 2001. Current approaches to seasonal to interannual climate predictions. Int. J. Climatol. 21, 1111–1152.
Gyasi-Agyei, Y., 2011. Copula-based daily rainfall disaggregation model. Water Resour. Res., 47.
Hao, Z., Singh, V., 2012. Entropy-copula method for single-site monthly streamflow simulation. Water Resour. Res., 48.
Hao, Z., Singh, V., 2013. Modeling multisite streamflow dependence with maximum entropy copula. Water Resour. Res. 49, 7139–7143.
Hao, Z., Singh, V.P., 2016. Review of dependence modeling in hydrology and water resources. Prog. Phys. Geogr. 40 (4) (0309133316632460).
Hasan, M., Dunn, P.K., 2012. Understanding the effect of climatology on monthly rainfall amounts in Australia using Tweedie GLMs. Int. J. Climatol. 32, 1006–1017.
Henley, B.J., Gergis, J., Karoly, D.J., Power, S., Kennedy, J., Folland, C.K., 2015. A tripole index for the interdecadal Pacific oscillation. Clim. Dyn. 45, 3077–3090.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.
Hudson, D., Alves, O., Hendon, H.H., Marshall, A.G., 2011. Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. Q. J. R. Meteorolog. Soc. 137, 673–689.
Jaworski, P., Durante, F., Härdle, W.K., 2013. Copulae in mathematical and quantitative finance. In: Lecture Notes in Statistics—Proceedings. Springer, Berlin.
Joe, H., Li, H., Nikoloulopoulos, A.K., 2010. Tail dependence functions and vine copulas. J. Multivar. Anal. 101, 252–270.
Joe, H., 1996. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. Lecture Notes-Monogr. Ser., 120–141.
Jones, D.A., Wang, W., Fawcett, R., 2009. High-quality spatial climate data-sets for Australia. Aust. Meteorol. Oceanogr. J. 58, 233.

Khedun, C.P., Mishra, A.K., Singh, V.P., Giardino, J.R., 2014. A copula-based precipitation forecasting model: investigating the interdecadal modulation of ENSO's impacts on monthly precipitation. Water Resour. Res. 50, 580–600.
Kiem, A.S., Franks, S.W., 2004. Multi-decadal variability of drought risk, eastern Australia. Hydrol. Processes 18, 2039–2050.
Kirono, D.G., Chiew, F.H., Kent, D.M., 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. Hydrol. Processes 24, 1237–1247.
Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J., Scott, J.K., 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. Methods Ecol. Evol. 3, 53–64.
Kurowicka, D., Cooke, R.M., 2006. Uncertainty Analysis with High Dimensional Dependence Modelling. John Wiley & Sons.
Lau, K.M., Sheu, P., 1988. Annual cycle, quasi-biennial oscillation, and southern oscillation in global precipitation. J. Geophys. Res. Atmospheres 93, 10975–10988.
Lee, T., Salas, J.D., 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. Hydrol. Res. 42, 318–330.
Liu, Z., Zhou, P., Chen, X., Guan, Y., 2015. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. J. Geophys. Res.: Atmospheres 120.
Marshall, A.G., Hendon, H.H., Son, S.-W., Lim, Y., 2016. Impact of the quasi-biennial oscillation on predictability of the Madden–Julian oscillation. Clim. Dyn., 1–13.
McBride, J.L., Nicholls, N., 1983. Seasonal relationships between Australian rainfall and the Southern Oscillation. Mon. Weather Rev. 111, 1998–2004.
McKeon, G., Hall, W., Henry, B., Stone, G., Watson, I., 2004. Pasture degradation and recovery in Australia's rangelands: learning from history.
Meinke, H., Hochman, Z., 2000. Using Seasonal Climate Forecasts to Manage Dryland Crops in Northern Australia—experiences from the 1997/98 Seasons Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems. Springer, pp. 149–165.
Mekanik, F., Imteaz, M., 2013. Analysing lagged ENSO and IOD as potential predictors for long-term rainfall forecasting using multiple regression modelling. In: 20th International Congress on Modelling and Simulation, Adelaide, Australia.
Moeller, C., Smith, I., Asseng, S., Ludwig, F., Telcik, N., 2008. The potential value of seasonal forecasts of rainfall categories—case studies from the wheatbelt in Western Australia's Mediterranean region. Agric. Forest Meteorol. 148, 606–618.
Montazerolghaem, M., Vervoort, W., Minasny, B., McBratney, A., 2016. Spatiotemporal monthly rainfall forecasts for south-eastern and eastern Australia using climatic indices. Theor. Appl. Climatol. 124, 1045–1063.
Murray, G.M., Brennan, J.P., 2009. Estimating disease losses to the Australian wheat industry. Austral. Plant Pathol. 38, 558–570.
Nelsen, R.B., 2006. An Introduction to Copulas, 2nd ed. Springer.
Nicholls, N., 1983. Predicting Indian monsoon rainfall from sea-surface temperature in the Indonesia–north Australia area. Nature 306, 576–577.
Pham, M.T., Vernieuwe, H., De Baets, B., Willems, P., Verhoest, N., 2016. Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas. Stochastic Environ. Res. Risk Assessment 30, 2197–2214.
Power, S., Casey, T., Folland, C., Colman, A., Mehta, V., 1999. Inter-decadal modulation of the impact of ENSO on Australia. Clim. Dyn. 15, 319–324.
Rasel, H., Imteaz, M., Mekanik, F., 2016. Investigating the influence of remote climate drivers as the predictors in forecasting south australian spring rainfall. Int. J. Environ. Res. 10, 1–12.
Rauf, U.F.A., Zeephongsekul, P., 2014. Analysis of rainfall severity and duration in victoria: Australia using non-parametric copulas and marginal distributions. Water. Resour. Manag. 28, 4835–4856.
Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., Trudinger, C., 2008. Australian water availability project (AWAP). In: CSIRO Marine and Atmospheric Research Component:Final Report for Phase 2., pp. 38.
Raupach, M.R., Briggs, P., Haverd, V., King, E., Paget, M., Trudinger, C., 2009. Australian water availability project (AWAP). In: CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. Bureau of Meteorology and CSIRO.
Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C., Hendon, H.H., 2009. On the remote drivers of rainfall variability in Australia. Mon. Weather Rev. 137, 3233–3253.
Salinger, M., Renwick, J., Mullan, A., 2001. Interdecadal Pacific oscillation and south Pacific climate. Int. J. Climatol. 21, 1705–1721.
Schepen, A., Wang, Q., Robertson, D., 2012. Evidence for using lagged climate indices to forecast Australian seasonal rainfall. J. Clim. 25, 1230–1246.
Schepsmeier, U., 2010. Maximum Likelihood Estimation of C-vine Pair-copula Constructions Based on Bivariate Copulas from Different Families. Master's Thesis. Technische Universität München.
Shafaei, M., Fakheri-Fard, A., Dinpashoh, Y., Mirabbasi, R., De Michele, C., 2016. Modeling flood event characteristics using D-vine structures. Theor. Appl. Climatol., 1–12.
Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris, pp. 8.
Sklar, A., 1996. Random variables, distribution functions, and copulas: a personal look backward and forward. Lecture Notes-Monogr. Ser., 1–14.
Stephens, D., Walker, G., Lyons, T., 1994. Forecasting Australian wheat yields with a weighted rainfall index. Agric. Forest Meteorol. 71, 247–263.
Stone, R.C., Hammer, G.L., Marcussen, T., 1996. Prediction of Global Rainfall Probabilities Using Phases of the Southern Oscillation Index.
Taschetto, A.S., England, M.H., 2009. El Niño modoki impacts on australian rainfall. J. Clim. 22, 3167–3174.
Trede, M., Savu, C., 2013. Do stock returns have an Archimedean copula? J. Appl. Statist. 40, 1764–1778.
Verdon, D.C., Wyatt, A.M., Kiem, A.S., Franks, S.W., 2004. Multidecadal variability of rainfall and streamflow: eastern Australia. Water Resour. Res. 40.
Verhoest, N., Vernieuwe, H., Pham, M.T., Willems, P., De Baets, B., 2015. A copula-based stochastic generator for coupled precipitation and evaporation time series. EGU General Assembly Conference Abstracts, p. 3690.
Walker, R., Mason, W., 2015. Climate Change Adaptation for Health and Social Services. Csiro Publishing.
Wang, G., Hendon, H.H., 2007. Sensitivity of Australian rainfall to inter-El Nino variations. J. Climate 20, 4211–4226.
Wong, G., Lambert, F., Leonard, M., Metcalfe, A.V., 2010. Drought analysis using trivariate copulas conditional on climatic states. J. Hydrol. Eng. 15, 129–141.
Wong, G., van Lanen, H.A.J., Torfs, P.J.J.F., 2013. Probabilistic analysis of hydrological drought characteristics using meteorological drought. Hydrol. Sci. J. 58, 253–270.
Wong, G., 2013. A comparison between the Gumbel-Hougaard and distorted Frank copulas for drought frequency analysis. Int. J. Hydrol. Sci. Technol. 3, 77–91.
Yang, W., 2010. Drought Analysis Under Climate Change by Application of Drought Indices and Copulas, Civil and Environmental Engineering. Portland State University.
Zhang, L., Singh, V.P., 2007. Bivariate rainfall frequency distributions using Archimedean copulas. J. Hydrol. 332, 93–109.
Zhang, L., Singh, V.P., 2014. Trivariate flood frequency analysis using discharge time series with possible different lengths: cuyahoga river case study. J. Hydrol. Eng. 19, 05014012.

Supplementary Material

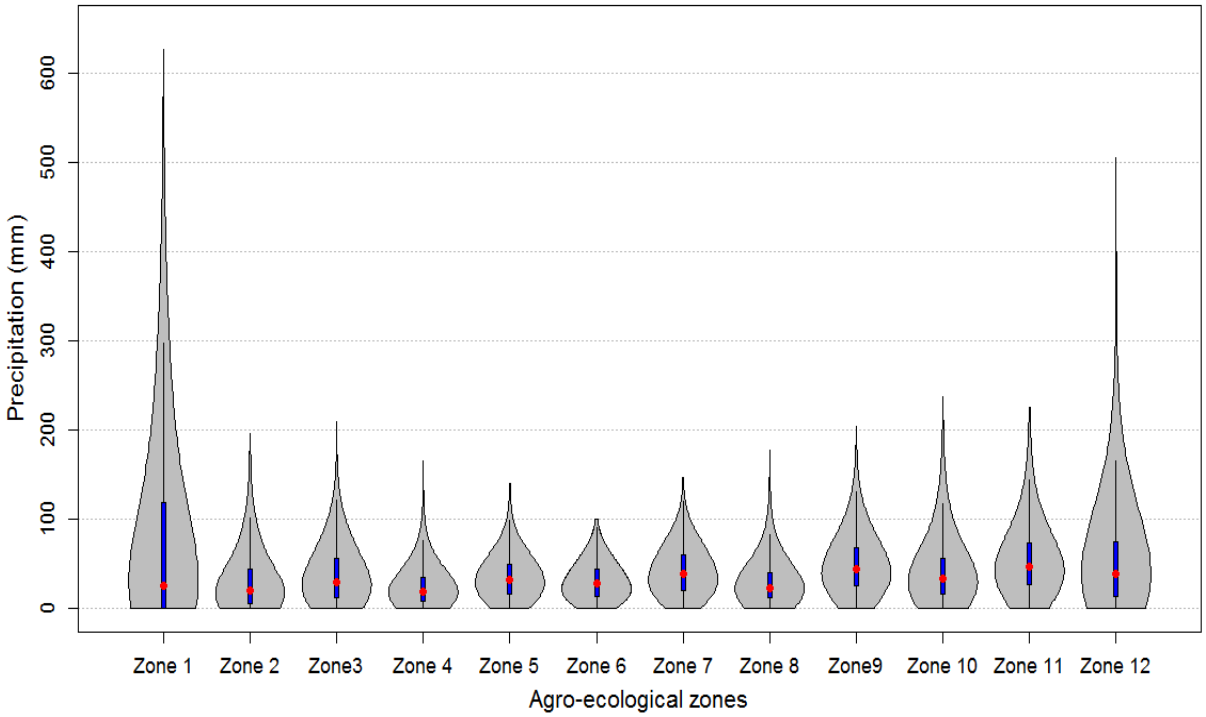


Fig. A.1. Violin plot of monthly precipitation (1900-2013) at each AEZ. Thick blue line and red dot show the 25th and 75th percentile ranges and median, respectively, and the thin black line shows the 5th and 95th percentile ranges.

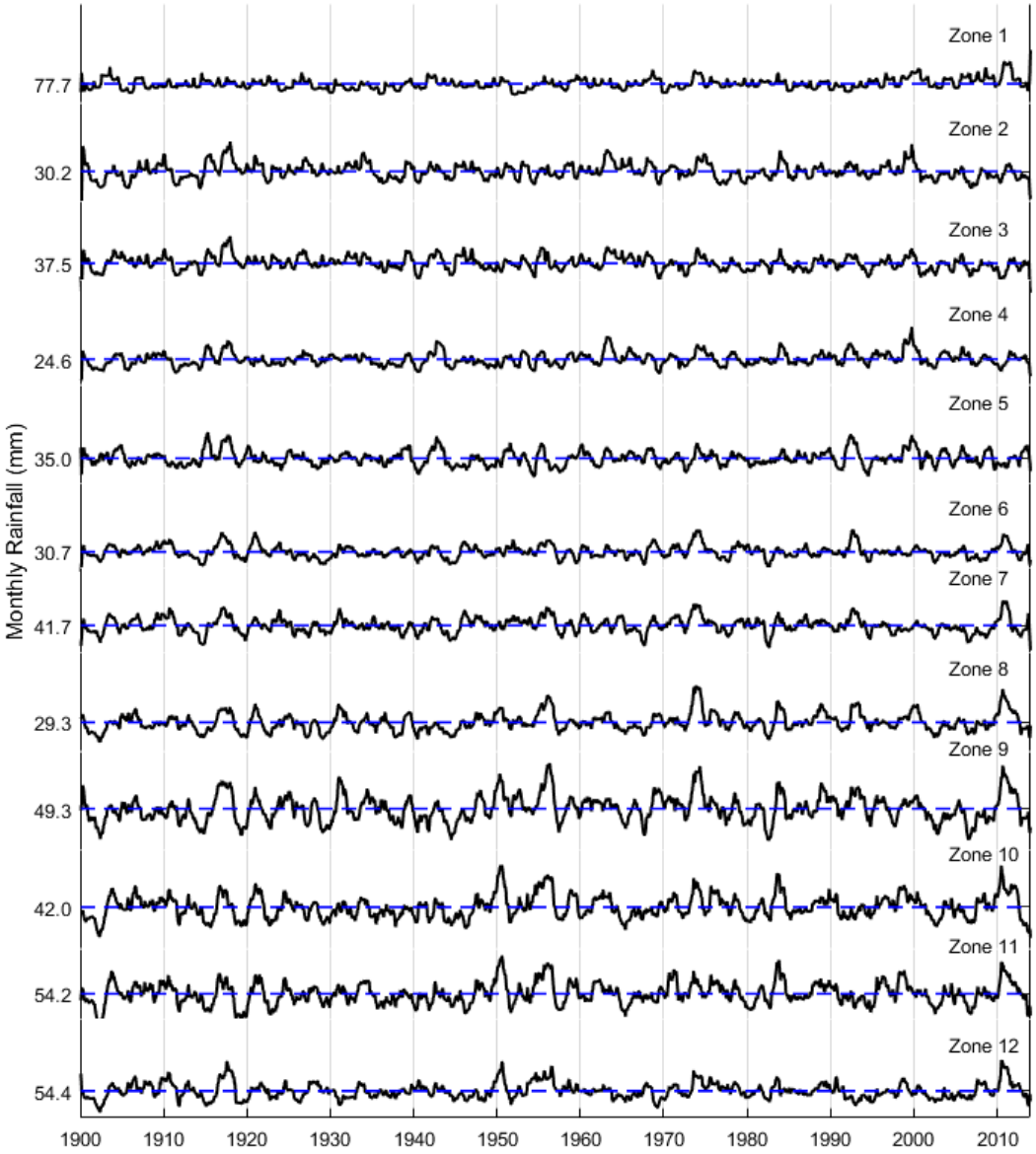


Fig. A.2. Time series of monthly precipitation for each Zone. The time series have been smoothed with a 13 month centered moving average filter. The blue dashed line represents the means for the dataset (mm).

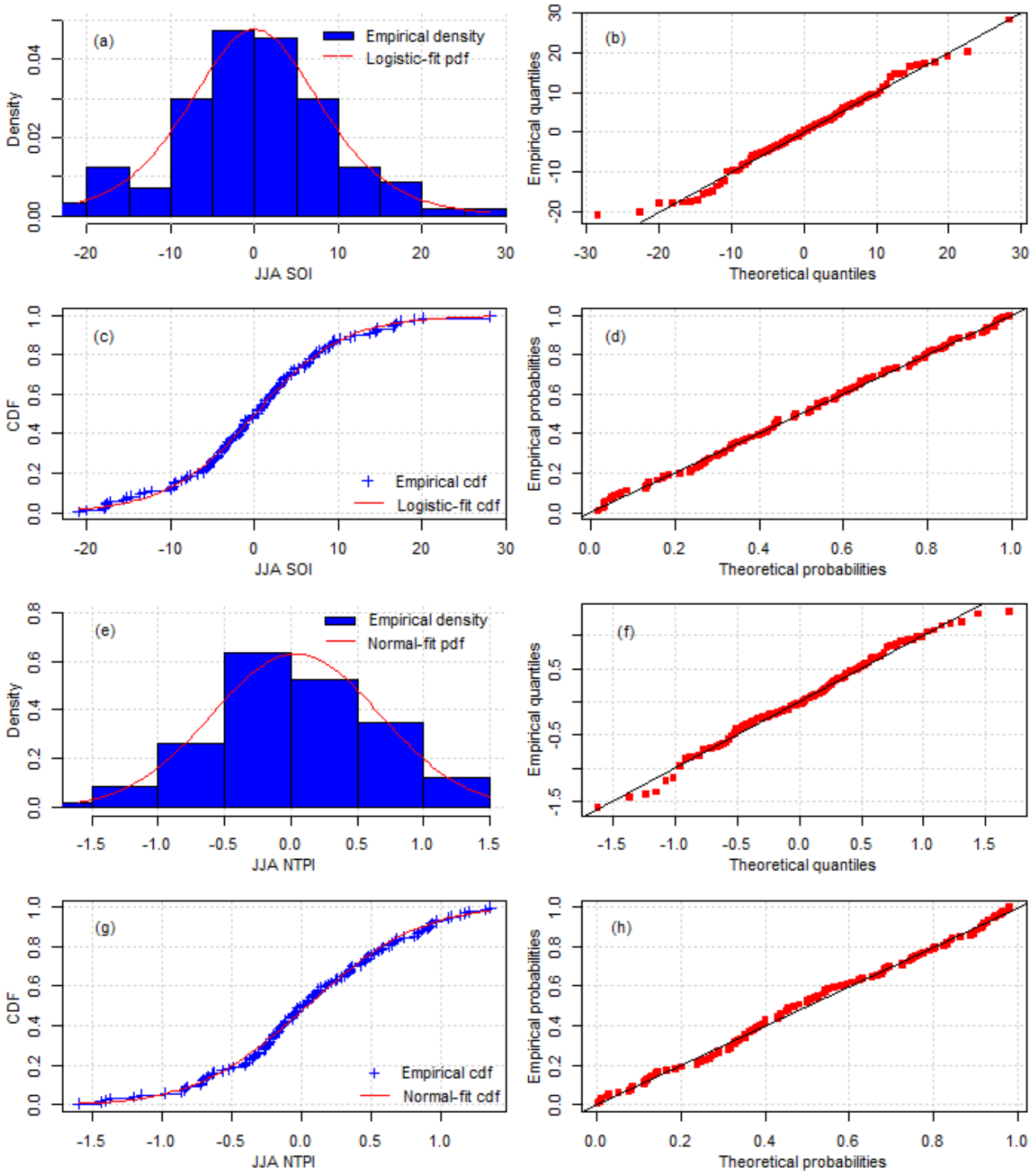


Fig. A.3. Four Good-of-Fit plots of JJA SOI (a-d) and JJA NTPI (e-h). SOI and NTPI follow logistic and normal distribution, respectively. (a) and (e) revealed the density patterns of fitted distributions with the histogram of the data. (c) and (g) compare the CDF between the empirical distribution and the fitted distribution using the Hazen's rule for the empirical distribution with the probability of n points defined as $(1:n-0.5)/n$. (b) and (f) illustrate the quantiles of the

theoretical distribution (x-axis) versus the empirical quantiles (y-axis) defined as the same to the CDF plot. (d) and (h) compare the probability of the fitted distribution (x-axis) against the empirical probability (y-axis) with the same rule (Delignette-Muller and Dutang, 2015).

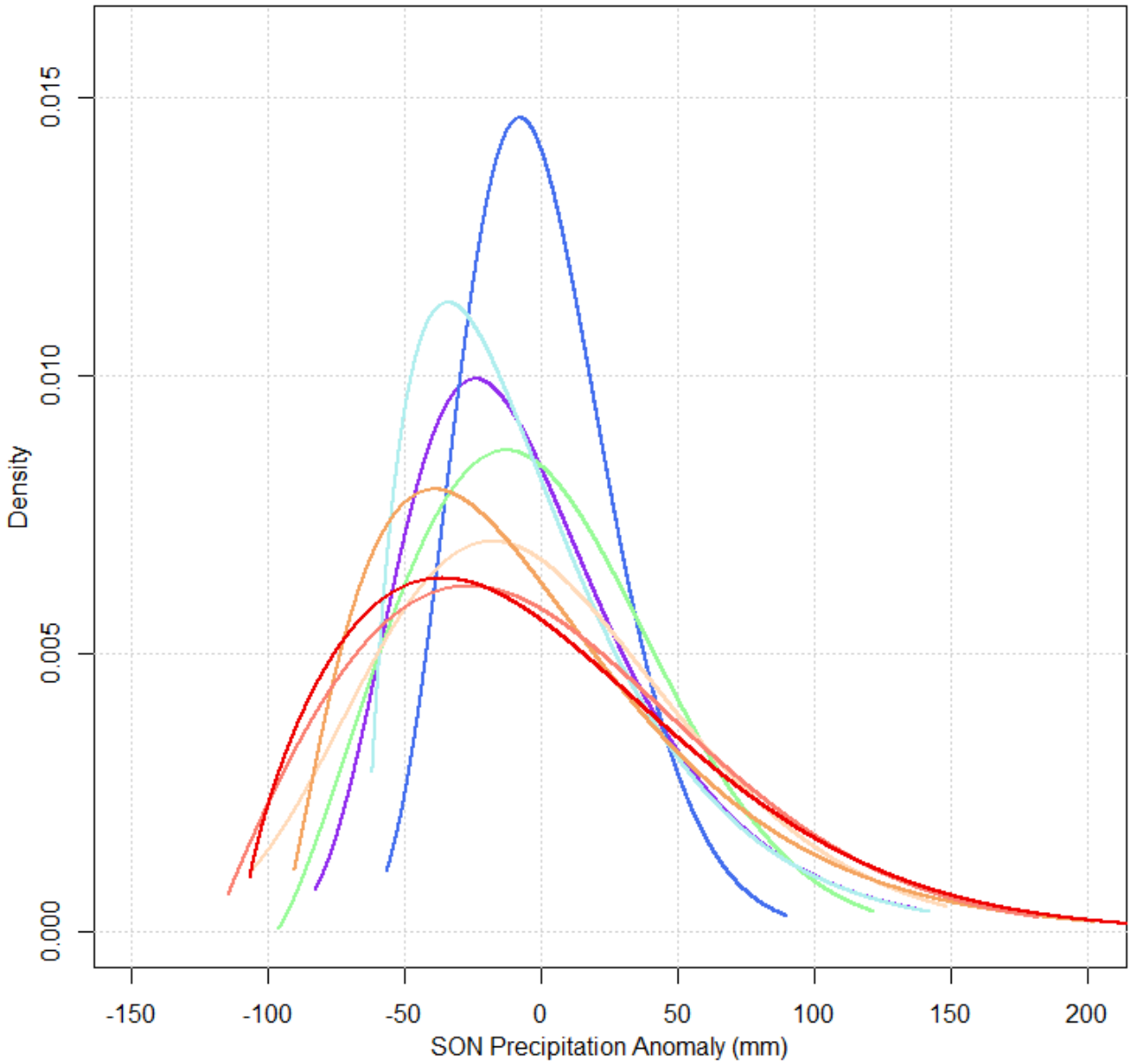


Fig. A.4. Selected marginal distributions of SON precipitation anomalies of each Agro-ecological Zone. Colours in Fig. A.4 match colours in Fig. 4.

Chapter 3

Quantile wheat yield forecasting using synoptic-scale climate indices

---

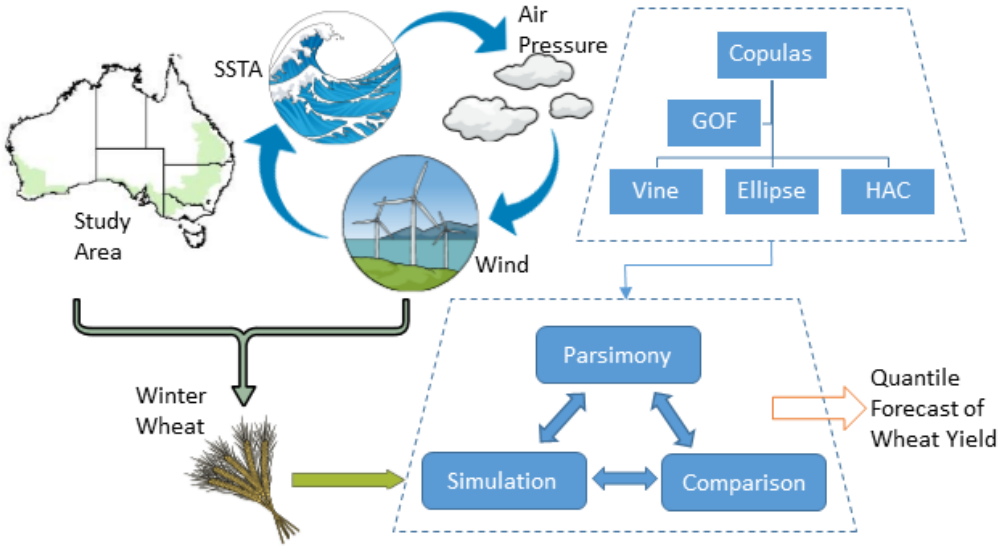
Article II: Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach

Summary:

This study uses twelve large-scale climate drivers to investigate their spatio-temporal influence on wheat yield, highlighting the variability in five major wheat-producing states across Australia using data for the period 1983–2013. These synoptic-scale climate indices are mainly derived from the data on relevant variables in the Pacific and Indian Ocean region. D-vine quantile regression models are developed to forecast wheat yield at preselected quantile levels. Models based on traditional linear quantile regression (LQR) are also established at the same time for the purpose of comparison. The out-of-sample accuracy of both statistical forecast models is estimated through the five-fold cross-validation approach. Fig. 7 illustrates the graphical abstract of this study on compound impact of synoptic-scale climate indices on wheat crop in Australia.

The influences of the Indian Ocean on wheat crops are generally dominant in all states except Western Australia, while the Pacific region has much stronger impact in Queensland. In particular, the results suggest that the wheat yield can be skilfully forecasted 3 – 6 months ahead using large-scale climate information, supporting early decision-making. The co-occurrence of extreme events is likely to enhance the impacts of climate mode and this can be quantified probabilistically through conditional copula-based models. For example, given the highly positive anomaly of DMI = 0.8 (*i.e.*, extremely positive phase), the probability of wheat yield anomaly in Queensland being 40% lower than average is about 62% in the bivariate model. Given the same DMI condition, the co-occurrence of extremely positive SOI = −20 increases that probability to 68%. Furthermore, the developed D-vine quantile regression models generally provide greater accuracy of wheat yield forecast at pre-selected quantile

levels compared to the LQR method. This study improves the quantification of the influences of large-scale climate drivers on the wheat yield that can allow a development of suitable planning processes and crop production strategies designed to optimise the yield and agricultural profit.



Source: Nguyen-Huy et al. (2018)

Figure 7. Graphical display of the study on compound impact of synoptic-scale climate indices on wheat crop in Australia.

European Journal of Agronomy 98 (2018) 65–81

Contents lists available at ScienceDirect

European Journal of Agronomy

journal homepage: www.elsevier.com/locate/eja

Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach

Thong Nguyen-Huy[a,b,d,*], Ravinesh C Deo[a,b,d,*], Shahbaz Mushtaq[b,d], Duc-Anh An-Vo[b,c], Shahjahan Khan[a,b,d]

[a] School of Agricultural, Computational and Environmental Sciences, Australia
[b] International Centre for Applied Climate Sciences (ICACS), Australia
[c] Computational Engineering and Science Research Centre (CESRC), Australia
[d] University of Southern Queensland, Institute of Agriculture and Environment (IAg&E), QLD, 4350, Australia

ARTICLE INFO

Keywords:
Crop modeling
Australian wheat
Crop yield forecasting
Multiple climate indices
Copula models
D-vine copulas
Quantile regression
Joint distribution
Conditional probability
Food security

ABSTRACT

Twelve large-scale climate drivers are employed to investigate their spatio-temporal influence on the variability of seasonal wheat yield in five major wheat-producing states across Australia using data for the period 1983–2013. Generally, the fluctuations in the Indian Ocean appear to have a dominant effect on the Australian wheat crop in all states except Western Australia, while the impact of oceanic conditions in the Pacific region is much stronger in Queensland. The results show a statistically significant negative correlation between the Indian Ocean Dipole (IOD) and the anomalous wheat yield in the early growing stage of the crop in the eastern and southeastern wheat belt regions. This correlation suggests that the wheat yield can be skillfully forecast 3–6 months ahead, supporting early decision-making in regard to precision agriculture. In this study, we use vine copula models to capture climate-yield dependence structures, including the occurrence of extreme events (i.e., the tail dependences). The co-occurrence of extreme events is likely to enhance the impacts of climate mode and this can be quantified probabilistically through conditional copula-based models. Generally, the developed D-vine quantile regression model provide greater accuracy for the forecasting of wheat yield at given different confidence levels compared to the traditional linear quantile regression (LQR) method. A five-fold cross-validation approach is also used to estimate the out-of-sample accuracy of both copula-statistical forecasting models. These findings provide a comprehensive analysis of the spatio-temporal impacts of different climate mode indices on Australian wheat crops. Improved quantification of the impacts of large-scale climate drivers on the wheat yield can allow a development of suitable planning processes and crop production strategies designed to optimize the yield and agricultural profit.

1. Introduction

Wheat is a major cereal crop in Australia, accounting for more than half of the approximately 23 M ha of Australian grain crops annually (Potgieter et al., 2013). Wheat is grown mostly in the drylands (i.e., rainfed crops); in Australia's wheat belt region, which experiences one of the world's most variable climate conditions (Portmann et al., 2010; Rimmington and Nicholls, 1993; Turner, 2004). According to Matsumura et al. (2015), although soil type and fertilizer improvements are important influencing factors, agricultural production is strongly influenced by climate conditions, especially precipitation and temperature, even in highly developed countries such as Australia. Therefore, improvements in the understanding and estimation of spatio-temporal climate-related variabilities that drive wheat yield are extremely important for agricultural management and food security.

Climate variables (e.g., precipitation and temperature) are commonly used to forecast crop yields worldwide, often at the site level (Asseng et al., 2011; Bannayan et al., 2003; Palosuo et al., 2011). Averaged or gridded data derived from a number of weather stations may be applied to the broader scale levels (Lobell et al., 2007; Revadekar and Preethi, 2012). However, relying on data from the weather station networks may create issues in relation to data quality, continuous collection of such data and the convenience issues related to regular monitoring of data acquisition systems (Harris et al., 2014; Schepen et al., 2012). Furthermore, with the current development and maturation of climate models, large-scale climate modes such as the El Nino-Southern Oscillation (ENSO) can be used to accurately forecast climate information at least six months to one year in advance (Cane,

* Corresponding author at: University of Southern Queensland, Institute of Agriculture and Environment (IAg&E), QLD, 4350, Australia.
E-mail addresses: thonghuy.nguyen@usq.edu.au (T. Nguyen-Huy), ravinesh.deo@usq.edu.au (R.C. Deo).

https://doi.org/10.1016/j.eja.2018.05.006
Received 18 November 2017; Received in revised form 15 May 2018; Accepted 20 May 2018
1161-0301/ © 2018 Elsevier B.V. All rights reserved.

2005; Jin et al., 2008; Ludescher et al., 2013). In addition, many studies have examined the relationship between large-scale climate variability, rainfall and agricultural production at regional/national (Garcia y Garcia et al., 2010; Royce et al., 2011; Shuai et al., 2013; Nguyen-Huy et al., 2017), and at continental/global (Anderson et al., 2017; Ceglar et al., 2017; Gutierrez, 2017; Iizumi et al., 2014) scales. All these facts suggest that large-scale climate modes may be more appropriate for analyzing the association with crop yields over large scale study areas with appropriately identified time lag, and reliable and available datasets.

In Australia, a range of different synoptic-scale climate indices have been identified as suitable responses to climate variability, depending on the regions and the seasons (Ashok et al., 2003; Min et al., 2013; Schepen et al., 2012; Taschetto and England, 2009; Nguyen-Huy et al., 2017). For example, the ENSO phenomenon continues to have a significant impact on precipitation over much of the Australia continent, especially in the north and east of the continent, with regional differences in different seasons (Risbey et al., 2009). In particular, La Niña events may bring substantial precipitation (often in eastern Australia), while El Niño events are often associated with broad-scale drought conditions (Yuan and Yamagata, 2015). Therefore, ENSO indices have been widely applied to explain the interannual variabilities in Australian wheat yields (Potgieter et al., 2005, 2002; Yuan and Yamagata, 2015). In general, these results show that La Niña (El Niño) events are related to increased (decreased) wheat yield. In addition, information about April–May ENSO indices can potentially act as an early forecasting tool for several seasonal crops, including wheat, in the following growing season (Potgieter et al., 2002).

In conjunction with the above issues, teleconnections or interactions among different climate modes are likely to modify the impact of individual drivers on climate conditions, in particular, during extreme climate events (Li et al., 2016; Lim et al., 2016; Nguyen-Huy et al., 2017; Weller and Cai, 2013). For example, Min et al. (2013) observed that there were anomalously drier and hotter conditions occurring across north-eastern and southern coastal Australia during El Niño and positive phases of IOD in the cold seasons, whereas wetter and cooler conditions appeared during La Niña and negative phases of IOD. Furthermore, in recent decades, the occurrence, role, and amplitude of the impact of climate modes have been found to have shifted. For example, increased occurrences of positive IOD events have been identified as the main driver of major 20th century droughts in southeast Australia, not ENSO conditions as commonly assumed (Cai et al., 2012; Ummenhofer et al., 2009). Furthermore, the ENSO Modoki phenomenon, which is a coupled ocean-atmosphere mode of variability in the tropical Pacific, appears to exhibit different teleconnection patterns to Australia's climate compared to the canonical (traditional) ENSO (Ashok et al., 2007; Ashok and Yamagata, 2009). Hence, it is clear that these factors and their changes over the growing season could modulate the relationship between large-scale climate modes, rainfall (Nguyen-Huy et al., 2017) and therefore, the crop yield. However, only a few studies have investigated the simultaneous impacts of multiple climate drivers on Australian crop yield (Jarvis et al., 2018; Yuan and Yamagata, 2015), although a previous study has develop probabilistic models for rainfall forecasting in Australia's agro-ecological zones (Nguyen-Huy et al., 2017). Considering the paucity of this essential information required for agricultural management, there is a need for a comprehensive study on the association between the joint influences of major climate indices and crop yield.

Several types of models are used to forecast wheat yield, including empirical models and biophysical simulation models. The empirical models, that do not utilize physical equations, can be classified into statistically-based or machine learning methods developed using artificial intelligence tools (Deo et al., 2017; Deo and Şahin, 2016). In contrast to biophysical models (Hansen et al., 2004; Mushtaq et al., 2017), statistical models are able to use historical relationships between climate indices and crop yield in order to forecast future crop yield. Since statistical models rely on historical data, they are not able to be used to simulate scenarios that have not previously occurred and so cannot be easily adjusted to accommodate changes in climate, crop genetics or cultivation practices. However, the main advantage of statistical models are that they do not consider the underlying eco-biophysiological processes, and hence do not require the significant crop parameterization used in biophysical models. In addition, they are generally easier to construct and more suited for forecasting crop yield, especially over regional scales where a relationship between yield and climate modes can be identified (Matsumura et al., 2015). Last but not least, the copula-statistical models are able to estimate uncertainties which are often difficult to acquire by process-based models (Lobell et al., 2006).

Statistical models applied in previous works (i.e., Jarvis et al., 2018; Yuan and Yamagata, 2015; Nguyen-Huy et al., 2017) employed linear relationships between a small number of climate indices and the crop yield, implying the assumption of normal joint distribution among these variables. While a linear regression model is more simple and provides a quick overview of the general trends (i.e., the fitted straight line) of the fitted response variable given the values of the explanatory variables, this model might be strongly influenced by outliers (e.g., extreme events) resulting in a spurious correlation between the considered variables (Hassani, 2016). Hence, assumption of normal distribution of the response variable might not always be realistic in practice.

To overcome these limitations, this study aims to employ multivariate copula functions (Sklar, 1959) to model the joint influence of climate modes on Australian (winter) wheat yield. This study advanced our early research performed where copula-statistical models were developed for rainfall forecasting in Australia's agro-ecological zones (Nguyen-Huy et al., 2017). The premise of the copula model, as also stipulated in earlier study (Nguyen-Huy et al., 2017), is that it has the ability to analyze the correlation structures between predictor-target variables, and provides a powerful and flexible tool to model the dependence structures between such complex and jointly correlated variables (Schepsmeier, 2015). Therefore, copulas have been broadly applied for statistical modeling and forecasting in several fields such as energy (Bessa et al., 2012; Zhang et al., 2014), financial risks (Huang et al., 2009; Lu et al., 2014), rainfall and climate predictions (Nguyen-Huy et al., 2017) and hydrology (Kao and Govindaraju, 2010; Liu et al., 2015). However, to the best of the authors' knowledge, the copula method has not yet been employed for analyzing relationships between multiple large-scale climate modes and wheat yield.

The aims of this study are, therefore: (1) to explore with the advanced statistical methodologies, for the first time, the spatio-temporal influence of well-known large-scale climate indices on seasonal wheat yield forecasting in different Australian states; (2) to compare the ability of the vine copula, which is a specific class of conventional copulas, with the other copula families for seasonal wheat yield forecasting; (3) to probabilistically quantify the variation in wheat yield conditions on climate modes; and (iv) to evaluate the forecasting skill of copula-based model against the conventional LQR method. The primary contribution of this research work is to establish and validate the suitability of a copula-statistical methodology for the forecasting of wheat yield based on large-scale climate mode influences and the implications in agronomic decision-making.

2. Materials and methods

2.1. Winter wheat yield

In Australia, the planting and the harvesting seasons of winter wheat crops vary from April to June and from October to January, respectively, and are mainly dependent on the winter-dominant rainfall patterns in each agricultural region. The mean wheat yield data of five major wheat producing states including Queensland (QLD), New South Wales (NSW), Victoria (VIC), South Australia (SA) and Western

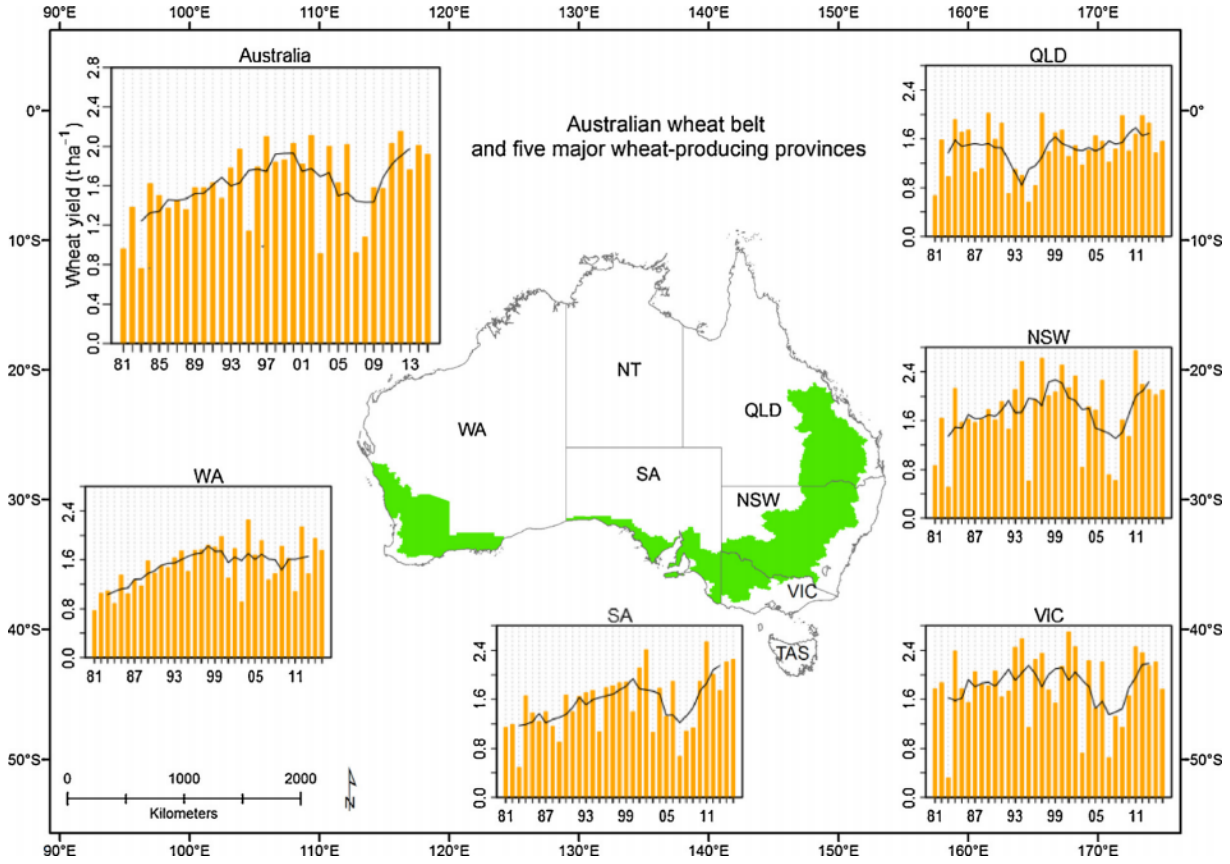


Fig. 1. Map of the Australian wheat belt region (green shaded area). Charts show the annual wheat yield (t ha^{-1}) of five major wheat-producing provinces: Queensland (QLD), New South Wales (NSW), Victoria (VIC), South Australia (SA) and Western Australia (WA) compared to those of the national level in 1981–2015 period overlaid by the 5-year running average (black lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Australia (WA) from 1981/1982 to 2014/2015 were downloaded from the Australian Bureau of Statistics. Fig. 1 plots the map of the Australian wheat belt with included charts showing the annual wheat yield for each state.

Several methods can be used to estimate the yield trend such as the single spectrum analysis or the local polynomial regression (Zhang et al., 2014). However, the trend fitting technique is not the focus of this study, and thus, for simplicity, the five-year running mean method is used to remove the trends (black lines shown in Fig. 1) in the data as in Yuan and Yamagata (2015):

$$Y = \frac{Y' - \overline{Y}}{\overline{Y}} \times 100, \quad (1)$$

where Y is the year-to-year variation of the annual wheat yield, Y′ the annual wheat yield, and $\overline{Y}$ the five-year running average. We used the ratio (i.e., division) instead of only the subtraction of the overall trend since we were interested in the anomalous percentage of the wheat yield (Fig. B2 in Supplementary Materials). By doing so, the changes in wheat yield caused by the short-term factors primarily, the climatic conditions, can be highlighted, although the others factors such as the demand, prices, and technology may also have a certain influence (Iizumi et al., 2014).

2.2. Climate indices

Twelve climate indices are used to comprehensively investigate their impacts on annual wheat yield data (Table 1). These climate indices have been identified as having a concurrent or a lagged relationship with monthly and seasonal rainfall patterns in Australia (Risbey et al., 2009; Schepen et al., 2012; Nguyen-Huy et al., 2017). Some of them, such as ENSO and IOD, have been used to forecast wheat yield (Power et al., 1999; Shuai et al., 2013; Yuan and Yamagata, 2015). This published evidence has reinforced the idea of utilizing all such climate mode indices to investigate their spatio-temporal impact on wheat yield at different times and locations.

ENSO phenomena are measured and monitored by several different indicators consisting of the Niño-3.0, Niño-3.4 and Niño-4.0 (i.e., sea surface temperature indices), Southern Oscillation Index (SOI) (air pressure index), and the EMI (coupled ocean-atmosphere index). The Dipole Mode Index (DMI) represents the intensity of IOD across the tropical Indian Ocean. Furthermore, the Indonesian Index (II), the Blocking Index (B140), Southern Annular Mode (SAM) and the Tasman Sea Index are also included in this study to consider their overall synoptic influence, particularly of extratropical and tropical origins.

Detailed information about these indices is provided in Table 1.

In terms of the origin of these data, the monthly sea surface temperature (SST) for the period of January 1983–December 2013 derived from Optimum Interpolation SST, version 2 (OISST v2) were downloaded from the Climate Prediction Center (CPC, NOAA). These post-satellite era data were based on the incorporation of satellite data, in situ observations and proxy SSTs derived from sea ice concentrations (Reynolds et al., 2007; Worley et al., 2005). We have used this data source to provide an analysis consistent with the work of Yuan and Yamagata (2015). Furthermore, these data have been widely applied in studying the interaction between climate mode indices (Izumo et al., 2010; Luo et al., 2010) or investigating climate-precipitation relationships (Kumar et al., 2013; Omondi et al., 2012).

The SST anomalies (SSTA) in this study were defined by subtracting each three-month average index from their corresponding 31-year average to indicate the inter-annual variations whereas the EMI was

Table 1
The twelve climate indices used as predictors for wheat yield (Y). Data Sources: Monthly sea surface temperature (SST) in different oceanic regions derived from Optimum Interpolation SST, version 2 (OISST v2) were downloaded from Climate Prediction Center (CPC, NOAA); SST anomalies (SSTA) are defined by subtracting each three-month-mean index from their corresponding 31-year mean indicating the inter-annual variations. SOI data were downloaded from Bureau of Meteorology, Australia (BOM) and the objective variable, Y data were obtained from Australian Bureau of Statistics.

Predictor Variables	Notation	Description	Region	Sources
Nino3.0	N3	Average SSTA over 150°–90 °W and 5 °N–5 °S	Pacific	OISST v2, NOAA
Nino3.4	N34	Average SSTA over 170 °E–120 °W and 5 °N–5 °S	Pacific	OISST v2, NOAA
Nino4.0	N4	Average SSTA over 160 °E–150 °W and 5 °N–5 °S	Pacific	OISST v2, NOAA
EMI	E	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	OISST v2, NOAA
SOI	S	Pressure difference between Tahiti and Darwin as defined by Troup (1965)	Pacific	BOM
WPI	WP	Average SSTA over 50°–70 °E and 10 °N–10 °S	Indian	OISST v2, NOAA
EPI	EP	Average SSTA over 90°–110 °E and 0 °N–10 °S	Indian	OISST v2, NOAA
DMI	D	WPI – EPI	Indian	OISST v2, NOAA
II	II	Average SSTA over 120°–130 °E and 0 °N–10 °S	Indian	OISST v2, NOAA
B140	B	0.5 × [U(25) + U(30) − U(40) − 2 × U(45) − U(50) + U (55) + U(60)]	Extratropical	NCEP
AAO	A	Antarctic Oscillation (Southern Annular Mode (SAM)) anomaly	Extratropical	NCEP
TSI	T	Average SSTA over 150°–160 °E and 30 °S–40 °S	Tropical	OISST v2, NOAA
Objective Variable	Y	Total yield (t ha^{-1})	Five agronomic States & Australian total yield value	Australian Bureau of Statistics

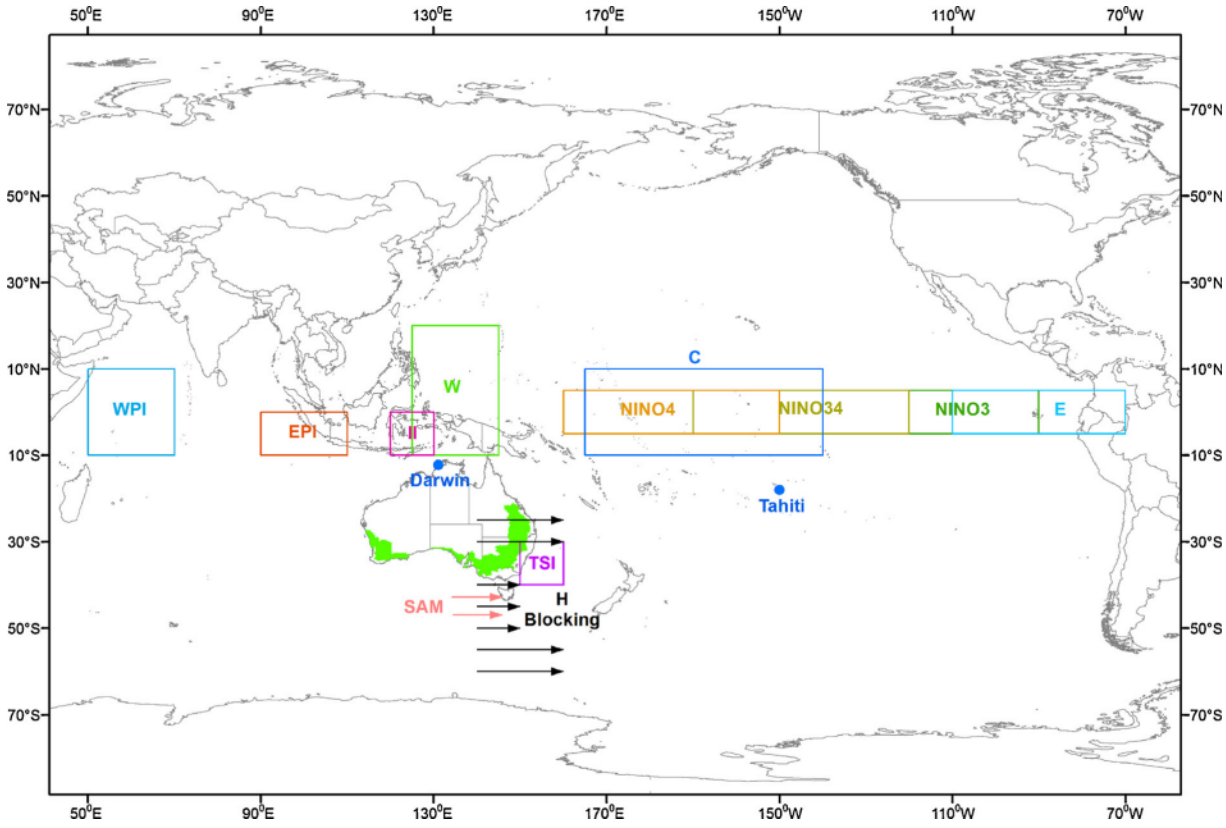


Fig. 2. Map of the study region and the oceanic representation used to calculate the twelve climate mode indices. Notations and equations for climate indices can be found in Table 1. (For a color version of this figure, the reader is referred to the web version of this article).

calculated following the equations in Table 1. The SOI data were downloaded from the Australian Bureau of Meteorology (BOM) and the DMI were calculated by subtracting the West Pole Index (WPI) from the East Pole Index (EPI) to determine the overall polarization effect. The blocking index data centered at 140 °E (B140) were developed by the BOM whereas the monthly total 500 hPa zonal wind data at different latitude X°S (U_X) were downloaded from the National Center for Environmental Prediction (NCEP) in order to calculate the B140 index, following the equations in Table 1. Fig. 2 shows the locations of the oceanic regions that were used to calculate the different climate mode indices in the present study. Hereafter, we use climate index or climate mode to refer to both the climate index anomalies and other climate indices (such as SOI, B140).

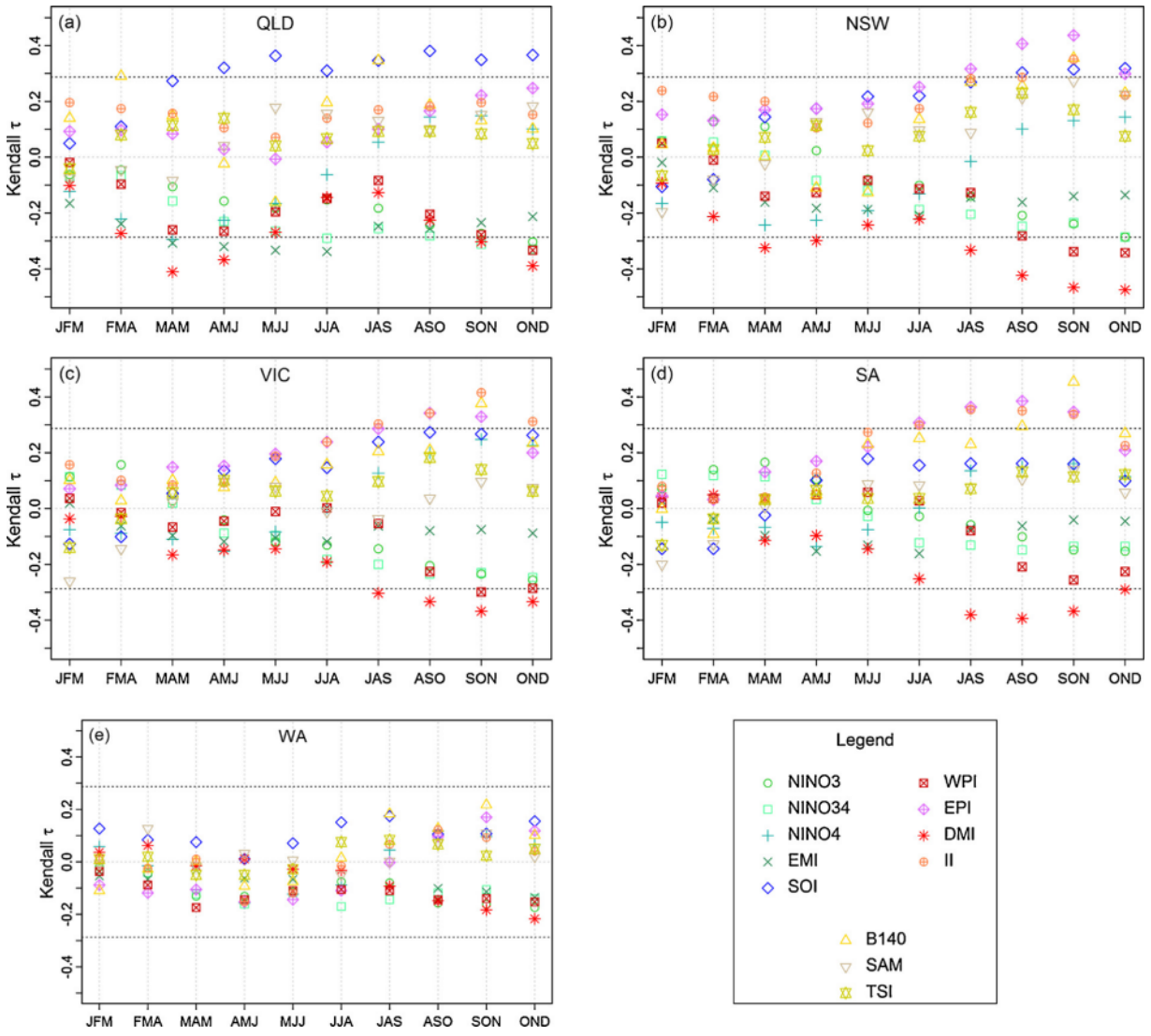


Fig. 3. Kendall's Tau-A correlation between the climate mode indices and the percentage wheat yield anomaly of the five major wheat-producing states: QLD, NSW, VIC, SA, and WA. (Horizontal dotted lines indicate significant at 5%). (For a color version of this figure, the reader is referred to the web version of this article).

2.3. Forecasting wheat yield with D-vine quantile regression model

This study aims to forecast the conditional quantile q_α of a response variable, Y (wheat yield) at some arbitrary quantile level $\alpha \in (0,1)$ for given covariates $(X_1,\ldots,X_d)$ (i.e., climate mode indices). To achieve this, following Kraus and Czado (2017) the inverse of the conditional distribution can be expressed as:

$$q_\alpha(x_1,\ldots,x_d) = F^{-1}_{Y|X_1,\ldots,X_d}(\alpha\,|x_1,\ldots,x_d). \quad (2)$$

Applying the Sklar's (1959) theorem with the probability integral transform, $V = F_Y(Y)$ and $U_j = F_j(X_j)$, the right-hand side of the Eq. (2) can be rewritten as:

$$F^{-1}_{Y|X_1,\ldots,X_d}(\alpha\,|x_1,\ldots,x_d) = F^{-1}_Y[C^{-1}_{V|U_1,\ldots,U_d}(\alpha\,|u_1,\ldots,u_d)]. \quad (3)$$

To preserve the desired advantages of the vine copula model and to make the estimation of the conditional copula quantile function $C^{-1}_{V|U_1,\ldots,U_d}(\alpha\,|u_1,\ldots,u_d)$ calculable and easy to implement in a real forecasting problem, a drawable vine (D-vine) copula is then fitted to the predictor-target data, in order to construct the D-vine quantile regression model automatically as proposed by Kraus and Czado (2017). In Section A (Supplementary Materials), we briefly describe the vine theory and approach and then illustrate it with an example for constructing a three-dimensional D-vine model. Readers can refer to the work of Kraus and Czado (2017) for more details. The advantage of this approach is mainly to be able to choose the most influential covariates that are likely to permit each covariate to be added into the model. This implementation ultimately improves the explanatory power of the resulting model. This algorithm is also able to solve the common issue in terms of regression, including collinearity, transformation and the inclusion and exclusion of covariates.

To model the influence of multiple climate mode indices on Australian wheat yield data using a D-vine copula-based approach, this study has employed a total of ten one and two-parameter copula families. This includes their rotated versions to fully assess the various forms of joint dependence structures, resulting in a total of thirty-two copula functions used to fit the bivariate models (Brechmann, 2010). Except for the forecasting section, the entire 31-year data set was used for the calculation of other remaining sections. All statistical computations were performed using the **VineCopula** package (Schepsmeier et al., 2017) available in **R** software (R Development Core Team, 2013) and **ArcGIS** was used to create the map (ESRI, 2016).

Table 2
Construction of the joint D-vine copula models with the corrected Akaike Information Criterion (AIC) conditional log-likelihood (cll_{AIC}) for QLD, NSW, VIC and SA. Notations for climate mode indices are stated in Table 1 using entire data from 1983 to 2013 (e.g. Y – B – E represents D-vine copula with B140 and EMI used to jointly model the yield, Y). The excluded periods (shown in green) did not have a statistically significant relationship with the wheat yield data (For interpretation of the references to colour in this table, the reader is referred to the web version of this article).

	JFM	*FMA*	*MAM*	*AMJ*	*MJJ*	*JJA*	*JAS*	*ASO*	*SON*	*OND*
QLD		Y – B – E	Y – D – S	Y – D	Y – S – N4	Y – E	Y – S	Y – S	Y – S – T	Y – S
cll_{AIC}		-8.47	-23.67	-14.27	-10.74	-7.70	-8.30	-9.14	-10.40	-9.97
NSW			Y – D	Y – D		Y – EI	Y – B – EI	Y – D	Y – D	Y – D
cll_{AIC}			-9.81	-10.99		-5.70	-20.28	-13.48	-15.08	-14.25
VIC	Y – A						Y – II	Y – II	Y – II	Y – S
cll_{AIC}	-4.87						-7.04	-10.17	-12.46	-6.94
SA					Y – II	Y – II	Y – II	Y – II – B	Y – B – II – N3	Y – D
cll_{AIC}					-3.87	-6.34	-9.25	-12.44	-22.94	-5.32

3. Results

3.1. Spatio-temporal impacts of climate modes on Australian wheat yield

The wheat yield variation for Australia and five major states is depicted in Fig. 1. In general, there is a similar pattern of tendency between the nation and states except for Queensland. However, the individual significant reductions of wheat yield for each state are different in terms of time and amplitude. These findings are in agreement with those from Yuan and Yamagata (2015) (see also in Fig. B2 in Supplementary Materials). As mentioned in Section 1, these reductions may be associated with extreme climate conditions and these differences may reflect the spatio-temporal impacts of climate modes on wheat crops. For example, according to the Australian Bureau of Meteorology (BOM), the 1982/1983 drought was possibly the worst climate event in Australia during the 20th century. Our results reveal that, during this period, except for WA, wheat yield was reduced by up to 80% in VIC, followed by 60% in NSW and SA; however, in QLD the yield fell by only 28%. However, a subsequent severe drought in the 1991/1995 period in QLD resulted in a wheat yield reduction of about 45% (1991/1992) and 50% (1994/1995). These extreme dry years also caused a 69% decrease in 1994/1995 wheat yield in NSW (even worse than that in 1981/1982). Yield reductions of about 47% and 34% occurred in VIC and SA, respectively, in the same years. Similar drought impacts on wheat yield in five Australian states have been found from two other dry years occurring in 2002/2003 and 2006/2007. It is worth mentioning that these drought events are influenced by strong El Niño events and are associated with higher than normal temperatures (Chiew et al., 1998; Deo et al., 2009). In particular, the dry-year period of 1991/1995 was consistent with the four consecutive El Niño years classified by the Science Division, Department of Science, Information Technology and Innovation (DSITI) (www.LongPaddock.qld.gov.au).

The spatio-temporal impacts of climate indices on percentage wheat yield anomalies are explored further through the popular statistical tools involving Pearson and rank-based Kendall's correlation coefficients. The former approach aims to measure the strength of a linear association between any two quantitative variables while latter can provide the association between the two sets of ranks given to the data from the same set of variables (Abdi, 2007). Since the dependence measures rely on the rank only, the rank-based approach has the advantage of capturing the outlying observations compared to the linear approach. The results shown in Fig. 3 (for rank-based method) and B3 in Supplementary Materials (for linear method) indicate that the impacts of climate modes on wheat yield have different spatio-temporal signatures. It is found that none of the climate indices has a statistically significant correlation to WA's wheat yield from both methods. In general, the results derived from the linear measure for DMI, Niño-3, and EMI are again in agreement with the study of Yuan and Yamagata (2015) although their study period was 1982/2010, and ours was 1983/2013. In the following sections, we describe the climate-yield relationship in term of non-linear dependence derived from the rank-based method since it will be used in the copula model development later.

3.1.1. Impacts from the Indian Ocean

Among the different climate indices in the Indian Ocean region, DMI has a dominant impact on Australian wheat productivity, as indicated by the significant negative correlation with the percentage wheat yield anomalies in all agronomical states. It is worth noting that the likely longer lead time of the significant correlations are clearly observable in QLD (Fig. 3a) and NSW (Fig. 3b). The effect of DMI on QLD wheat yield is found to be the strongest during the March–May (MAM; hereafter we use this notation style to refer a consecutive three months) period, i.e., before and during the sowing season. However, the influence of DMI is found to be strongest one month (JAS period) prior and during the harvesting for other agronomical states. None of the other climate indices derived from the Indian Ocean (i.e., EPI, WPI, and II) has a significant lagged relationship with QLD's wheat yield. However, such indices reveal a stronger influence in regions extending from southeast to south Australia. In particular, EPI and II are found to have a significant positive relationship with SA's wheat yield in JJA meaning two months before the harvesting season in October. The variabilities of WPI mode can be used to explain the wheat yield at least one month ahead of harvesting across NSW, VIC, and SA. Therefore, using the information derived from the Indian Ocean could provide useful forecasting proficiency early at the growing stage or before the harvesting season.

3.1.2. Impacts from the Pacific Ocean

Based on the present analysis, the results emphasize the important role of the ENSO on the wheat yield in the east (QLD). Niño-4.0 and EMI figures can be used to forecast wheat crops in QLD from MAM

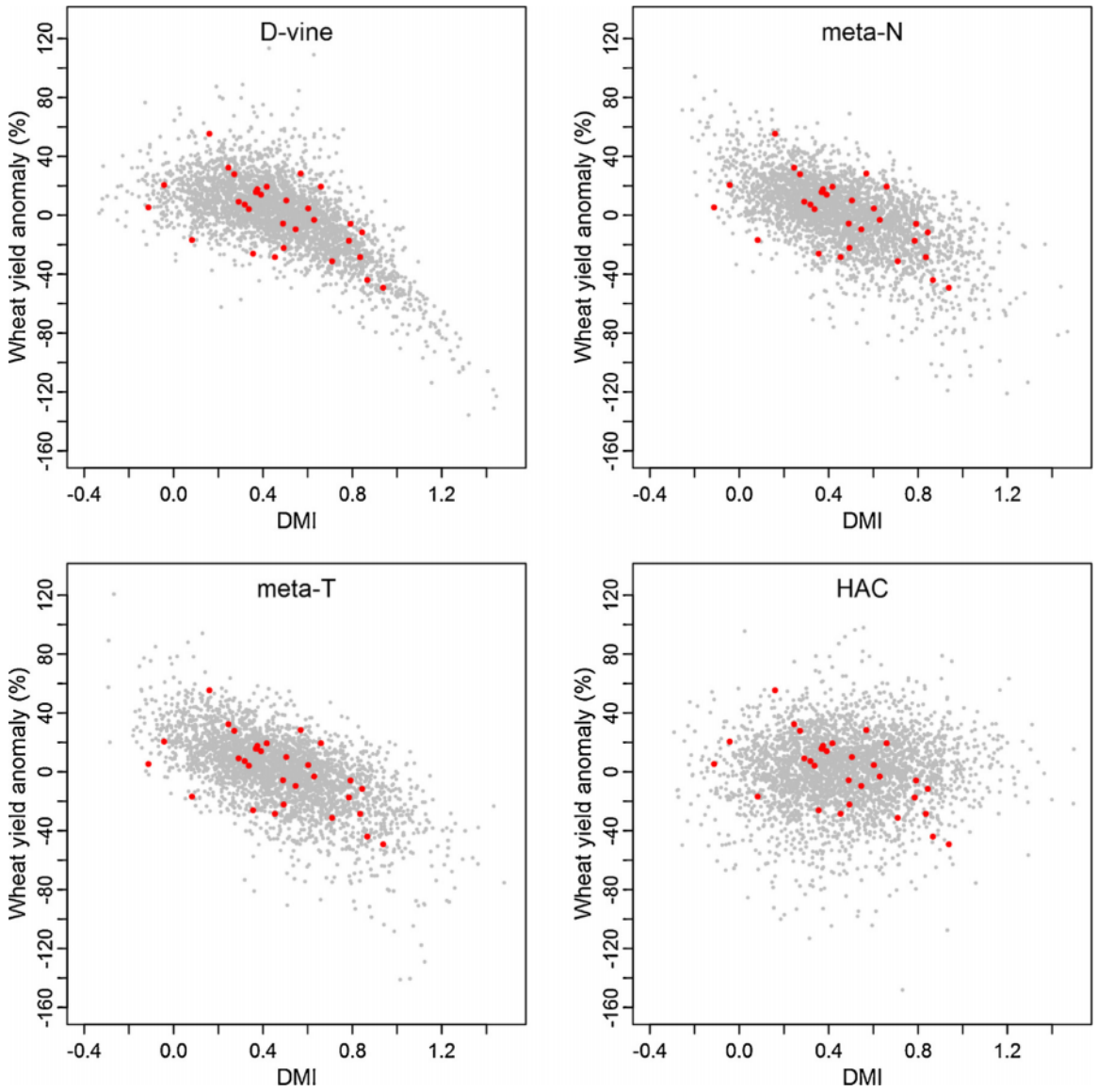


Fig. 4. Comparisons between the observed and simulated averaged March–May DMI and QLD percentage wheat yield anomalies for different three-dimensional copulas: vine (D-vine), meta-Gaussian (meta-N), meta-Student T (meta-T), and hierarchical Archimedean copula (HAC). (For a color version of this figure, the reader is referred to the web version of this article).

illustrated by significant negative correlation coefficients with wheat yield. This finding is similar to that of DMI, however, the impact of the Indian Ocean region in terms of amplitude is considerably higher than that of the Pacific region. SOI values show a stable strong impact on QLD's wheat yield from growing to harvesting stage and on NSW's wheat yield during the harvesting stage. Therefore, SOI and EMI can be utilized to forecast the wheat yield in QLD as the alternative predictors of DMI for the period May–October. Furthermore, it is found that none of ENSO indicators affect the wheat yield in the remaining states. These findings are found to be in agreement with those reported in a number of previous works (Hansen et al., 2004; Potgieter et al., 2002; Rimmington and Nicholls, 1993). Thus, it is expected that there is little change in the influence of ENSO (in particular the popular indicator, SOI) on the wheat crop grown in QLD in recent decades.

3.1.3. Impacts from tropical and extra-tropical regions

Our analysis reveals that the influence of tropical and extratropical conditions on the wheat yield variability appear to be indistinct for the different agronomic states. Closely observing these influences, the index B140 has the strongest positive correlation with wheat yield before the growing (FMA) and harvesting season (JAS) in QLD, and during the early months of harvesting (SON) in NSW, VIC, and SA. The correlation coefficients between the TSI and SAM data and the wheat yield are low across the different agronomical states. It is important to note that TSI and SAM were shown to have a significant impact on the VIC and SA rainfall patterns (Schepen et al., 2012), and our results also found that there is a significant climate-yield relationship using a linear approach (see Fig. B3 in Supplementary Materials). However, both TSI and SAM do not exhibit a notable impact on the wheat yield using the non-linear method, as indicated by insignificant rank-based correlation coefficients. This discrepancy could be the result of the differences in climate data versions, the period considered, and the methods (i.e., linear versus rank-based correlations) used to measure the dependence structure. However, this difference could also indicate that the influence of large-scale climate drivers on Australian wheat yield may vary from time to time and from region to region.

3.1.4. Impacts from co-occurrences of extreme events

Since the relationship between climate modes and wheat yield is evident from the analysis presented so far, it is also important to model

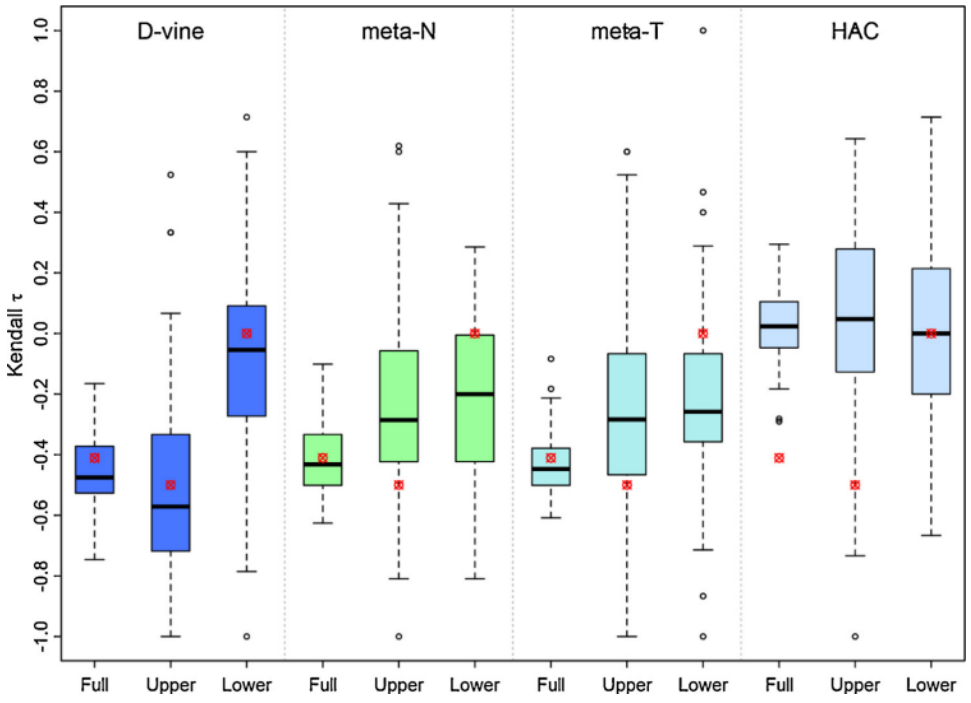


Fig. 5. Comparisons of the observed and simulated Kendall's tau between averaged March-April-May DMI and QLD percentage wheat yield anomalies from different three-dimensional copulas as in Fig. 4. Cases are considered for the whole data (Full), the upper tail (Upper) and the lower tail (Lower). (For a color version of this figure, the reader is referred to the web version of this article).

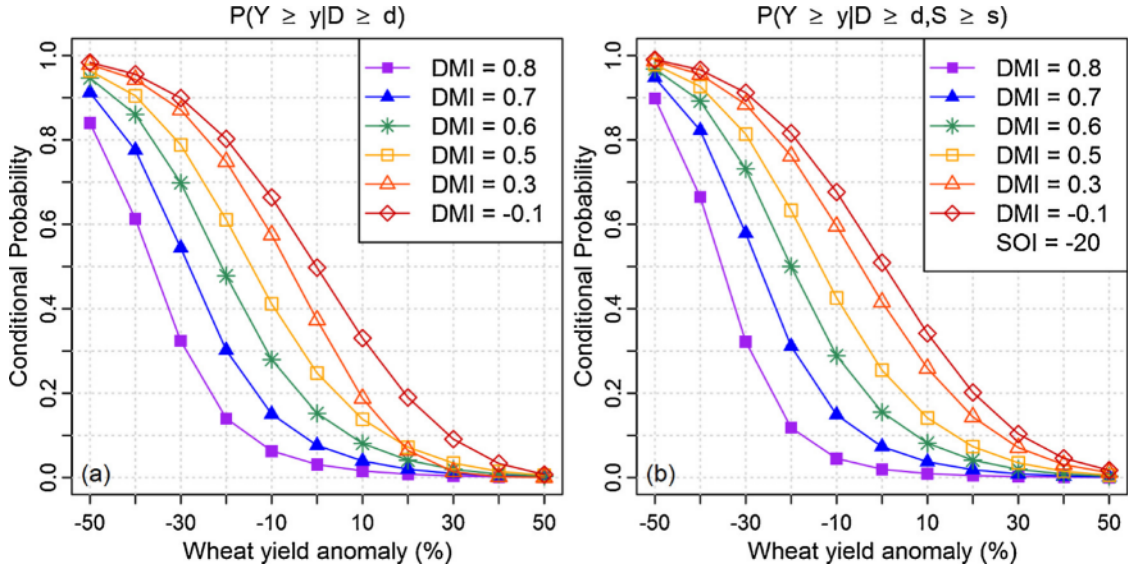


Fig. 6. Conditional probabilities of the percentage wheat yield anomaly (Y) of QLD given the different average of DMI (D) values (a) and co-occurrence of such values of DMI and extreme negative SOI (b) for the March-April-May period. (For a color version of this figure, the reader is referred to the web version of this article).

the extent to which the co-occurrence of extreme climatic events are likely to influence wheat yield. This analysis is an important aspect of agricultural risk and farming decisions since climate events have been demonstrated to bring more or less rainfall in major agricultural regions (Mekanik and Imteaz, 2014; Nguyen-Huy et al., 2017), and thus, might exhibit similar impacts on wheat yield. In a former work, Yuan and Yamagata (2015) identified the phases of a climate index, querying whether the monthly values of that index cross the 0.7 standard deviation of its time series. Here, the three-phase classification system was identified by comparing the three-monthly average values of a climate index to the quartile values computed from the 31-year time series of that index data. A climate index was then assigned to the positive and negative phase if its three-monthly average value fell into the highest 25% and the lowest 25% of that climate time series, respectively. Hence, the neutral phase is identified by the remaining range value of the climate index. Note that this scheme is different from the other systems, such as the ENSO events being classified based on the six-monthly mean of the SOI values (e.g., the classification from DSITI mentioned in Section 1). Hence, one phase of a climate index should be understood as the phase of those three-month average values, not the phase of that year.

Fig. B4 (in the Supplementary Materials) shows a plot of the QLD percentage wheat yield anomalies in relation to three phases of climate indices. Here, QLD has been selected for illustration purposes, since the wheat yield in this region showed the significant relationship with many climate indices with respect to the long lagged time period. One of the interesting results which can be integrated from this figure is the changes in the patterns of climate indices in a year. For example, the TSI data show a consistent positive phase during the January–March period, a neutral phase during the April–July period and a negative

phase during the August–October period. WPI reaches the peak during February-June before turning into the negative phase in June–September and becoming neutral after that. EPI shows a similar pattern to WPI, but acquires the positive and negative phase later and is in negative phase longer.

Furthermore, it is clear that the co-occurrence of the extreme events is likely to enhance the profit or the loss in terms of the wheat yield anomaly; however, this also depends on the time lag and the climate modes that are being considered, for example, the case of very early forecasting in AMJ, where DMI and SOI show a significant correlation with the wheat yield. The extreme losses of wheat yield in 1992 (45%) and 1995 (50%) were associated with the co-occurrence of the positive DMI phase and the negative SOI phase. However, when the extreme events occur alone, the impacts on wheat yield can be diminished. For example, the positive DMI occurred with neutral SOI phases reducing 29% (1983) and 20% (2010) of wheat yield; meanwhile, the negative SOI with neutral DMI phases caused a reduction of wheat yield by 10% (1998) and 17% (2003). These results also indicate the dominant role of the Indian Ocean in relation to the wheat yield as mentioned above. However, the impact from the co-occurrence of extreme events may be different to other considered times and regions due to the spatio-temporal characteristic of the climate impact. These facts suggest that the inter-modulation of yield data in respect to the co-occurrence of different climate indices should be considered further. This requires a vigorous model to be able to capture the joint impacts of climate drivers on Australian wheat yield. Therefore, the next section will introduce a

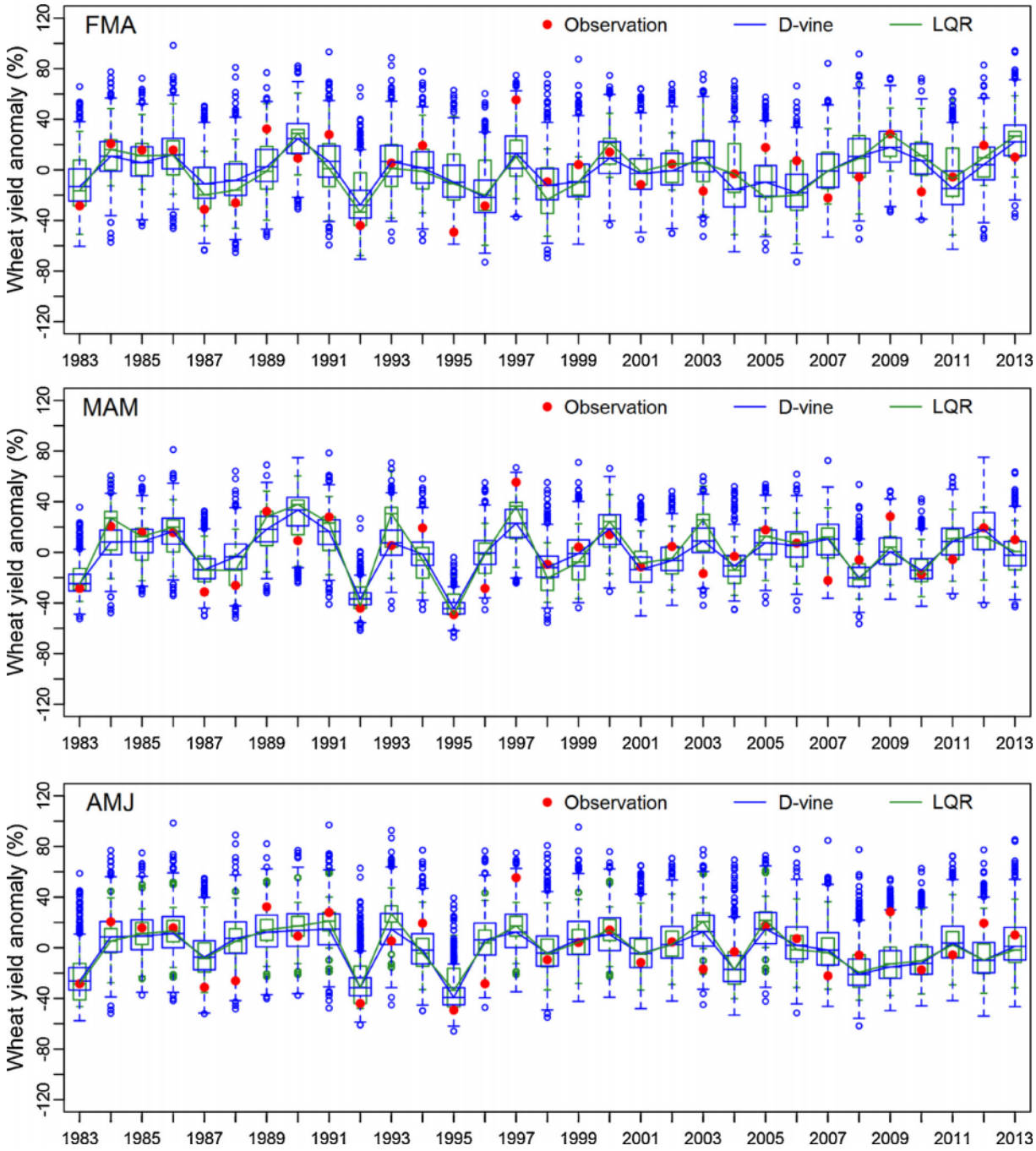


Fig. 7. Comparison of the median value of QLD observed and the 1-month to 8-month lead percentage wheat yield anomaly simulated from the D-vine copula (blue line) and linear quantile regression (LQR) (green line). Blue and green box plots show the uncertainties corresponding to both models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

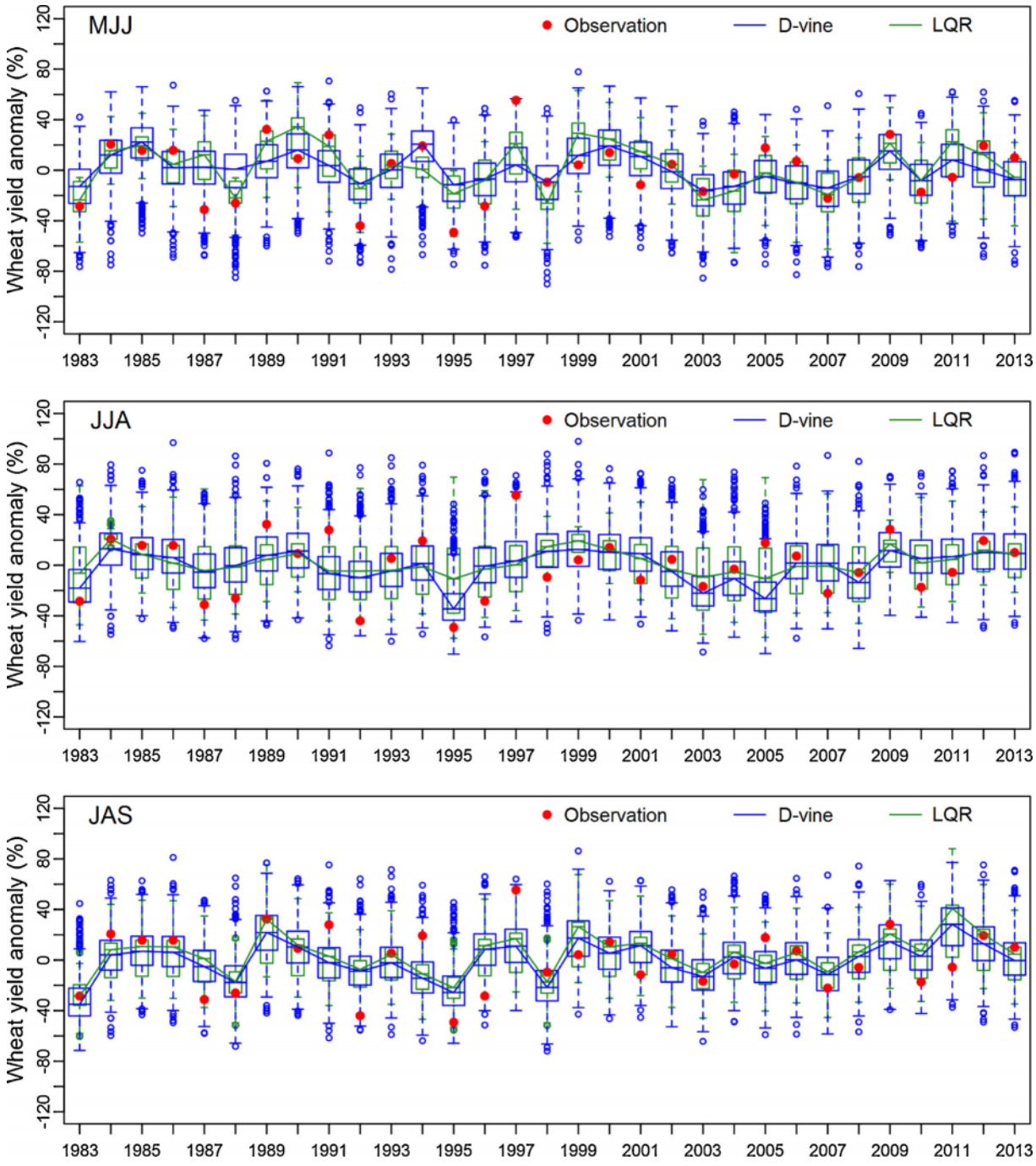


Fig. 7. (continued)

robust copula modeling framework for wheat yield forecasting.

3.2. Modelling dependence structures of climate and yield

3.2.1. D-vine copula models

The results of the fitted D-vine copula model with the corrected Akaike Information Criterion (AIC) conditional log-likelihood cll_{AIC} (Nguyen-Huy et al., 2017) for QLD, NSW, VIC, and SA are illustrated in Table 2. The model construction follows the steps described in Supplementary Material A. It can be seen that the structures and covariates included in each copula model have a significant difference for the different periods and regions of study. These differences are expected to be induced by the fluctuations and influences from different climate mode indices to the wheat yield data, and the interaction among the different indices in various agronomic zones. The first order, i.e., the covariate added to the model after the wheat yield Y, of the D-vine model, indicates the key driver with the strongest correlation to the wheat yield (i.e., this pair results in the minimum cll_{AIC}). These results reflect the relationship between climate mode indices and wheat yield investigated in Section 3.1.

The rest of the orders of the D-vine model is identified based on whether the addition of those climate indices improves the cll_{AIC} or not, as described in Supplementary Material A. Table C2 (in Supplementary Materials) gives an example of the construction of a four-dimensional D-vine model (Y – B – II – N3) in each tree for South Australia (SA) during SON (see Table 2). The bivariate model of percentage wheat yield anomaly-B140 (Y – B) resulted in the lowest cll_{AIC} (-15.46) identified in the first edge of D-vine. The additions of II and Niño-3

Table 3
Comparison of the average value of the absolute mean and median differences between the simulated and observed wheat yield (Y, t ha^{-1}) anomalies with an ensemble of 1000 different scenarios from the two models: vine copula (D-vine) and linear quantile regression (LQR) using entire data from 1983 to 2013 (For interpretation of the references to colour in this table, the reader is referred to the web version of this article).

	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND
	Mean / *Median*									
QLD										
D-vine		16.13	12.29	14.69	15.48	16.63	16.33	16.11	13.81	15.95
		16.23	*12.47*	*14.68*	*15.62*	*16.87*	*16.51*	*16.22*	*13.92*	*16.05*
LQR		15.88	13.22	15.05	14.71	17.59	16.21	16.00	14.92	16.71
		15.50	*12.68*	*14.82*	*14.72*	*16.99*	*15.34*	*15.27*	*14.15*	*16.60*
NSW										
D-vine			20.95	20.48		21.65	16.56	20.09	19.39	19.35
			20.90	*20.42*		*21.93*	*16.44*	*20.20*	*19.49*	*19.31*
LQR			20.44	20.41		22.42	17.76	19.70	18.48	19.12
			20.41	*20.32*		*21.63*	*17.39*	*19.57*	*18.44*	*18.75*
VIC										
D-vine	20.29						19.66	18.36	17.67	20.60
	20.49						*19.34*	*18.06*	*17.53*	*20.61*
LQR	20.78						20.86	19.72	18.13	22.40
	20.49						*20.16*	*18.68*	*17.92*	*21.90*
SA										
D-vine					18.38	17.79	16.66	15.02	12.11	17.88
					18.27	*17.70*	*16.42*	*14.99*	*11.92*	*17.69*
LQR					18.54	17.80	17.07	15.88	14.80	18.81
					18.31	*17.52*	*16.65*	*15.49*	*13.34*	*18.23*

indices improved the total cll_{AIC} (-22.94) by the new pair-copulas in the edge (1,3;2) and (1,4;2,3), respectively. After that, the addition of the remaining climate indices into the model does not improve the total cll_{AIC}, so the algorithm is stopped and the D-vine model is established.

3.2.2. D-vines better model tail dependence structures

It is worth noting that the tail dependences should be modeled in an efficient way in order to have an accurate assessment of the impact of extreme conditions on wheat crops. Thus, a comparison between the fitted D-vine copula models with the meta-Gaussian (meta-N), meta-Student T (meta-T), and hierarchical Archimedean copula (HAC) has been made for the simulation skill. Fig. 4 shows the simulation results from 1000 scenarios of DMI and wheat yield from four copula models, overlaid with the observed QLD wheat yield (Y) and average DMI (D) during MAM. We apply the simulation approach for the D-vine as shown in (Zhang and Singh, 2014). It is visualized that the dependences structure of Y – D in the upper tail (i.e., the losses of wheat yield corresponding to the positive DMI phase) are modeled better in the D-vine model compared to others.

Another analysis has been performed by comparing the dependence structure of Y – D indicated by Kendall's tau correlation measured for the whole time series, and for positive (upper) and negative (lower) phase as shown in Fig. 5. Although the overall dependence (full data range) is clearly obtained in three trivariate copula models, excepting HAC model, the dependence in the upper/lower (DMI values greater than its Q3 and smaller than its Q1, respectively) tail is better captured by D-vine model. This result is in agreement with the study by Zhang and Singh (2014) who performed the same comparison with the maximum daily discharge. Therefore, only D-vine copula is employed for further analysis in the following sections.

3.2.3. Conditional probability of wheat yield given values of climate indices

Since the joint CDF between variables is modeled through the copula function C, the conditional distributions of one variable given other variables can be explored. From the selected D-vine models, this study also quantifies the probabilities of wheat yield (Y) conditioned on different values (phases) of climate indices such as DMI (D) and SOI (S). The results from MAM D-vine model (Y – D – S), as in the previous example for QLD, are represented by bivariate and trivariate models. For example, one might be interested in evaluating the exceedance probabilities of Y given that D peak exceeds certain thresholds d as:

$$P(Y \geq y \mid D \geq d) = \frac{P(Y \geq y, D \geq d)}{P(D \geq d)} = \frac{1 - F_D(d) - F_Y(y) + F_{YD}(y, d)}{1 - F_D(d)}. \quad (4)$$

In the trivariate case, the conditional probabilities of Y given D and S are written as:

$$P(Y \geq y \mid D \geq d, S \geq s) = \frac{P(Y \geq y, D \geq d, S \geq s)}{P(D \geq d, S \geq s)} = \frac{1 - F_D(d) - F_S(s) - F_Y(y) + F_{DS}(d,s) + F_{YD}(y,d) + F_{YS}(y,s) - F_{YDS}(y,d,s)}{1 - F_D(d) - F_S(s) + F_{DS}(d,s)}, \quad (5)$$

where the joint CDF can be expressed in term of D-vine copula (Zhang and Singh, 2014):

$$F_{YD}(y, d) = C[F_Y(y), F_D(d)], \quad (6)$$

and:

$$F_{YDS}(y, d, s) = F_D(d) F_{YS|D}(y, s \mid d) = F_D(d) C_{YS|D}[F_{Y|D}(y \mid d), F_{S|D}(s \mid d)]. \quad (7)$$

In Fig. 6, we show the conditional probability of wheat yield given different phases of DMI and co-occurrence of both DMI and SOI, respectively. The interpretation of these is relatively straightforward.

Table 4
D-vine copula models constructed with 5-fold cross-validation. To overcome data length limitations, models designated as K1 to K5 represent the different periods for training and validation (i.e., Fig. B5 in Supplementary Materials). The approach evaluates out-of-sample performance by splitting data for different periods into a training set (25 years) fitting the quantile regression model (Eq. (2)) and a testing set (6 years). Notations for climate mode indices are as per Table 1. The excluded periods (shown in green) did not have a statistically significant relationship with the wheat yield data (For interpretation of the references to colour in this table, the reader is referred to the web version of this article).

Region & Cross-Validation Model	*Training Data*	*Testing Data*	*JFM*	*FMA*	*MAM*	*AMJ*	*MJJ*	*JJA*	*JAS*
QLD									
K1	1983 –2007	2008 – 2013		Y – B	Y – D – S	Y – D	Y – D	Y – E	Y – S – A
K2	1983 – 2000 & 2008 – 2013	2001 – 2007		Y – B – D	Y – D – S	Y – D	Y – E – B – S	Y – E	Y – S
K3	1983 – 1994 & 2001 – 2013	1995 - 2000		Y – N4	Y – D – E	Y – D	Y – E – N3 – B	Y – N34	Y – S – A
K4	1983 – 1988 & 1995 – 2013	1989 – 1994		Y – E – B – EP	Y – D – N34	Y – D	Y – E	Y – E	Y – S
K5	1989 – 2013	1983 – 1988			Y – D – N34	Y – D	Y – S	Y – S	Y – B
NSW									
K1	1983 –2007	2008 – 2013	Y – II – A		Y – D		Y – EP – N4	Y – EP	Y – EP
K2	1983 – 2000 & 2008 – 2013	2001 – 2007	Y – N4	Y – D	Y – D	Y – D	Y – D		Y – S – N4
K3	1983 – 1994 & 2001 – 2013	1995 - 2000	Y – II	Y – II – N4	Y – D	Y – EP – N4			Y – D
K4	1983 – 1988 & 1995 – 2013	1989 – 1994			Y – D – S	Y – D	Y – S	Y – S	Y – S
K5	1989 – 2013	1983 – 1988			Y – D	Y – D		Y – D	Y – B – EP
VIC									
K1	1983 –2007	2008 – 2013							Y – S
K2	1983 – 2000 & 2008 – 2013	2001 – 2007	Y – A	Y – A	Y – D	Y – D	Y – D	Y – EP	Y – D
K3	1983 – 1994 & 2001 – 2013	1995 - 2000	Y – A	Y – S					Y – II
K4	1983 – 1988 & 1995 – 2013	1989 – 1994					Y – S	Y – II	Y – II
K5	1989 – 2013	1983 – 1988							Y – D
SA									
K1	1983 –2007	2008 – 2013							Y – II
K2	1983 – 2000 & 2008 – 2013	2001 – 2007	Y – A	Y – A			Y – B	Y – EP	Y – D
K3	1983 – 1994 & 2001 – 2013	1995 - 2000					Y – II	Y – II – N4	Y – II – D
K4	1983 – 1988 & 1995 – 2013	1989 – 1994					Y – II	Y – II	Y – II
K5	1989 – 2013	1983 – 1988					Y – B	Y – EP	Y – D

Considering the bivariate case, for example, given the highly positive anomaly of DMI = 0.8 (i.e., extremely positive phase), the probability of QLD wheat yield exceeding the losses of 40% is about 62% (Fig. 6(a)). Given such conditions of DMI anomaly, the co-occurrence of extremely positive SOI increases that probability to 68% (Fig. 6(b)). However, the inclusion of SOI into the model does not give many changes. These results again indicate the dominant role of the Indian Ocean region in the QLD wheat yield.

3.3. Comparison of forecasting wheat yield to linear quantile regression

In this section, the quantile regression models are employed to forecast the quantile of the wheat yield variations Y_α given its historical values, and the historical and current values of associated climate indices. Because of the limitation of data length (31 years), we tested the forecasting model in two strategies. Firstly, we used the entire data to compute the absolute mean and median differences between the observed and simulated wheat yield anomalies from 1000 scenarios. Then, the five-fold cross-validation approach was used to evaluate the out-of-sample performance of forecasting models at given quantile levels (Refaeilzadeh et al., 2009). In the out-of-sample test, we split the dataset at a certain time point into a training data set (25 years), based on which we fitted the quantile regression model using the Eq. (2), and a testing data set (6 years) was used to assess the performance of the method. Therefore, the forecasting was repeated five times with different sets of training and testing data (see Table 4 and Fig. B5 in Supplementary Materials).

We also compared the D-vine model to the linear quantile regression (LQR) model where the predicted conditional quantiles are assumed to be linearly represented by predictors, i.e.,:

$$q_\alpha(x_1^{(i)},\ldots,x_d^{(i)}) = \widehat{\beta}_0 + \sum_{j=1}^{d} \widehat{\beta}_j x_j^{(i)}. \tag{8}$$

Performances of the copula-based quantile regression model in forecasting wheat yield anomalies, Y_α, is statistically evaluated using three prediction score metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Willmott's Index of Agreement (d):

$$MAE = \frac{1}{K}\left[\frac{1}{N}\sum_{i=1}^{N} |(Y_{\alpha,i} - Y_{o,i})|\right], \tag{9}$$

$$RMSE = \frac{1}{K}\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N} (Y_{\alpha,i} - Y_{o,i})^2}\right] \tag{10}$$

$$d = \frac{1}{K}\left\{1 - \left[\frac{\sum_{i=1}^{N} (Y_{o,i} - Y_{\alpha,i})^2}{\sum_{i=1}^{N} (|Y_{\alpha,i} - \overline{Y_o}| + |Y_{o,i} - \overline{Y_o}|)^2}\right]\right\}, \tag{11}$$

where K denotes the number of forecasting periods repeated (K = 5 in this study), $Y_{\alpha,i}$ and $Y_{o,i}$ are the forecast yields at the α quantile level and observed wheat yield anomaly respectively, and N = 6 is the length of the test set.

For the first strategy, the fitted D-vine and LQR models were applied for the quantile levels $\alpha \in (0, 1)$ illustrated for QLD as an example. The

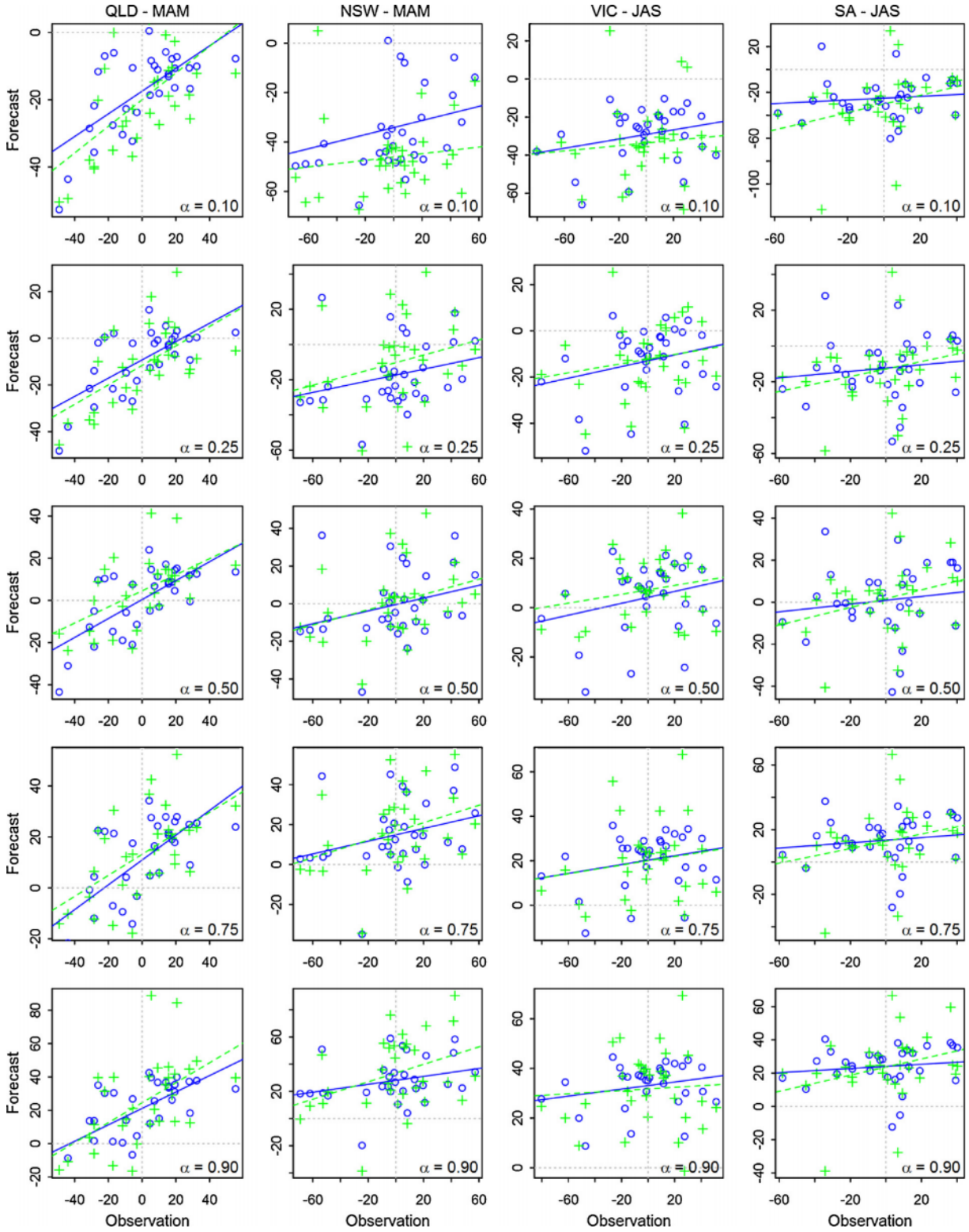


Fig. 8. Scatter plot of forecast versus observed wheat yield anomaly (%) from D-vine (blue circle) and LQR (green plus) and their corresponding fitted line at different quantile levels (α = 0.10, 0.25, 0.50, 0.75, 0.90) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

time considered is from January to September, just before the harvesting season. Given the historical observations, the simulation of wheat yield anomalies is repeated 1000 times for each model to account for the stochastic effects. The average of mean and median for the 1000 scenarios is then compared to those from observations to achieve the absolute difference. Visual inspection shows that D-vine models perform better than LQR in capturing the observed extreme values (Fig. 7). It is also noted that the simulations consist of only a single realization and the length of data set is limited (31 years), so several simulations will show the variability. The computable results summarized in Table 3 show that the forecasting skill of mean and median values varies from 12% to 22%. The D-vine copula models, in general,

Table 5
Evaluation of the Willmott's Index of Agreement (d), Relative Mean Absolute Error (RMAE) and the Relative Root-Mean-Square Error (RRMSE) between D-vine copula and linear quantile regression model at different quantile for the case of simulated wheat yields with an average of the 5-fold averaged data. The excluded periods (shown in green) did not have a statistically significant relationship with the wheat yield data (For interpretation of the references to colour in this table, the reader is referred to the web version of this article).

	JFM			FMA			MAM			AMJ			MJJ			JJA			JAS		
	d	RMAE	RRMSE	d	RMAE	RRMSE	d	RMAE	RRMSE	d	RMAE	RRMSE	d	RMAE	RRMSE	d	RMAE	RRMSE	d	RMAE	RRMSE
QLD																					
0.10				0.44	1.00	1.00	0.59	1.11	1.10	0.54	1.03	1.04	0.47	1.06	1.07	0.48	1.05	1.04	0.51	1.18	1.09
0.25				0.42	1.03	1.05	0.65	1.08	1.10	0.61	0.93	0.96	0.48	1.01	1.01	0.52	1.16	1.14	0.53	1.06	1.06
0.50				0.36	1.00	1.01	0.74	1.18	1.16	0.62	0.99	1.00	0.48	1.11	1.10	0.52	1.05	1.05	0.50	0.96	1.02
0.75				0.42	1.02	1.05	0.73	1.21	1.16	0.54	1.02	0.98	0.37	1.10	1.11	0.49	1.04	1.05	0.49	1.08	1.07
0.90				0.44	0.98	1.01	0.65	1.17	1.18	0.48	1.26	1.21	0.43	1.03	1.08	0.48	1.11	1.11	0.48	1.18	1.17
NSW																					
0.10	0.40	1.32	1.25	0.33	1.08	1.11	0.46	1.26	1.23	0.36	0.92	0.91	0.38	1.07	1.04	0.37	1.19	1.19	0.48	1.25	1.26
0.25	0.40	0.92	0.94	0.29	0.87	0.90	0.54	0.92	0.94	0.37	0.87	0.89	0.36	1.10	1.10	0.34	1.02	1.10	0.50	1.18	1.15
0.50	0.36	0.90	0.91	0.31	0.83	0.88	0.54	1.04	0.99	0.37	0.83	0.84	0.30	0.93	0.93	0.27	1.06	1.01	0.52	0.99	1.00
0.75	0.33	0.97	1.00	0.33	0.92	0.90	0.51	1.08	1.01	0.37	0.93	0.95	0.28	0.91	0.95	0.23	0.98	0.98	0.47	0.91	0.89
0.90	0.41	1.06	1.07	0.38	1.17	1.11	0.49	1.22	1.11	0.39	1.17	1.14	0.34	1.17	1.09	0.32	1.07	1.04	0.42	0.87	0.91
VIC																					
0.10	0.47	1.46	1.37	0.31	1.09	1.05	0.23	1.13	1.15	0.34	1.03	1.01	0.33	0.94	1.00	0.39	1.32	1.24	0.48	1.22	1.14
0.25	0.45	1.00	0.98	0.16	0.99	0.98	0.17	0.99	0.98	0.22	0.91	0.92	0.28	0.98	0.98	0.30	1.01	1.02	0.46	1.06	1.05
0.50	0.44	0.98	0.97	0.16	1.02	1.04	0.13	0.94	0.93	0.07	0.90	0.92	0.20	0.92	0.91	0.30	0.93	0.89	0.45	1.05	1.03
0.75	0.45	0.90	0.89	0.29	0.97	0.96	0.22	1.01	0.98	0.22	1.01	0.99	0.22	0.95	0.94	0.35	0.92	0.90	0.44	1.07	1.05
0.90	0.44	1.11	1.08	0.37	1.09	1.05	0.32	1.01	0.99	0.35	1.09	1.03	0.36	1.08	1.09	0.41	1.04	0.96	0.43	1.08	1.03
SA																					
0.10	0.42	0.95	0.98	0.42	1.00	0.98							0.42	1.13	1.14	0.43	1.11	1.10	0.42	1.12	1.14
0.25	0.32	0.99	0.99	0.35	1.03	1.00							0.38	1.02	1.01	0.39	0.96	0.98	0.44	1.01	0.98
0.50	0.13	0.87	0.91	0.12	0.94	0.96							0.33	0.96	0.97	0.39	1.02	1.02	0.46	0.93	0.93
0.75	0.34	0.85	0.86	0.20	0.86	0.90							0.39	0.92	0.92	0.46	1.02	1.02	0.46	1.02	0.98
0.90	0.42	0.95	0.89	0.36	1.11	1.14							0.43	0.93	0.93	0.45	0.96	0.98	0.44	1.04	1.00

estimate a lower difference of absolute mean and median whereas the LQR method performs better in the case of NSW. The models with two or three predictors yield a greater accuracy than those with only one predictor, as expected.

In the second strategy, we investigate the relationship between wheat yield anomalies and climate indices at different levels of quantiles, i.e., at α = (0.10, 0.25, 0.50, 0.75, 0.90) for all states except WA. The median forecast (i.e., α = 0.50) provides a general idea about the wheat yield while the extreme values in the lower tail ($\alpha \leq 0.25$) or upper tail ($\alpha \geq 0.75$) represent the worst-case scenarios. Table 4 illustrates the constructions of D-vine models fitted to a training set at five-fold (K = 5) for four states. It is important to note that there is a variation in the dimension and predictors in the constructed models for different training sets. However, it is clear that the Indian Ocean (DMI, EP, and II) shows the significant effect on the wheat yield in all considered states. ENSO has more impact on the QLD wheat yield. Also, SAM has some associations with the wheat yield in VIC and SA early in the year. These results are expected by the spatio-temporal impact of the climate indices on the wheat yields.

Fig. 8 (in Supplementary Materials) represents a comparison of the fitted lines at different quantiles and five-fold K for both quantile regression models during MAM for QLD. It is clear that the D-vine model provides a better fit compared to LQR in most cases. It is confirmed by the results as shown in Table 5 and Fig. 8. The comparisons are evaluated using the Relative Mean Absolute Error (RMAE) and Relative Root Mean Square Error (RRMSE) where RMAE = MAE^{LQR}/MAE^{D-vine} and RRMSE = $RMSE^{LQR}/RMSE^{D-vine}$ in which MAE and RMSE acquired from the Eqs. (10) and (11) corresponding to each model. The value number is greater than one, indicating that the performance of the D-vine model is better.

The forecasting skill of the D-vine copula model is also evaluated by applying Willmott's Index d. The results show a moderate agreement between simulated and observed wheat yield anomaly. However, the results are expected given the limitation of data length. In general, D-vine models show a greater accuracy at all considered quantile levels compared to LQR method. A closer observation also reveals that the D-vine approach performs better at extreme events (i.e., quantile levels of 0.10 and 0.90) while the traditional LQR provides higher accuracy in some cases of the neutral phase (i.e., quantile levels in between 0.25 and 0.75).

4. Discussions

This study has indicated comprehensively the dominant role of the Indian Ocean compared to the Pacific and tropical regions in the interannual variation of the Australian wheat yield. The conclusions attained in this study in respect to the role of synoptic-scale climate mode indices on wheat yield is consistent with our earier study (Nguyen-Huy et al., 2017) where a similar causal role was identified on probabalistic predictions of rainfall in Australian agro-ecological zone. Firstly, in this study, we have explored the spatio-temporal influence of individual climate drivers as well as their co-occurrences with respect to wheat yield in the five major wheat producing states in Australia. Then, we have probabilistically quantified the changes in wheat yield

conditioned on respective climate modes. Next, we have presented the results for the construction of the D-vine copula models, revealing that the differences in the model structure appear to be consistent with the spatio-temporal influences of climate mode indices on the wheat yield. We have also indicated the advantage of vine copulas in modeling the tail dependences compared to other copula methods. Finally, we have demonstrated a potential skill in forecasting the wheat yield variabilities using the vine copula model based on observed climate modes. The vine copula-based quantile regression model has been indicated to be better than the conventional LQR.

The results in regard to the impacts from the co-occurrence of extreme events in this study should be interpreted carefully. It is because the phases of a climate index are classified based on the three-month average anomalies/values compared to the average values of 31-year data of that climate index. Therefore, a climate index can have three different phases in a year. Further, the impacts of climate modes are analyzed based on the rank-based correlation measure which possibly yields a different result compared to the linear method used in other studies. For example, while Niño-3 and Niño-3.4 may have a significant linear relationship with wheat yield in QLD, NSW, and VIC (Fig. B3 in Supplementary Materials), those do not hold in the case of a rank-based approach.

We also highlight that the common linear (quantile) regression approach still play an important role in either agriculture or other sectors, although the vine copula approach generally shows a greater accuracy in this study. It can be observed from the published literature that the linear regression model has been widely applied in many studies to investigate the effect of climate conditions on agricultural crops (Barnwal and Kotani, 2013; Lobell et al., 2006; Lobell and Burke, 2010) or analyze yield gap (Laborte et al., 2012). In addition, the LQR model displays a higher accuracy in forecasting, for example mean and median of wheat yields in NSW (see Table 3), compared to vine copula method. Therefore, researcher should not eliminate the use of the linear (quantile) regression without any reasonable evidence.

It is also noted that the D-vine model in the present study is not extended to more than four dimensions. It is expected since we take the independence copula into consideration with a confidence level of 95%. Furthermore, the correlation coefficients between new pairs of variable reduce in each tree by the conditional rule whereas the correlation coefficients among variables are moderate (< 0.5). Also, the length of time series is limited and that may affect the accuracy of the parameter estimation. Moreover, the structure of D-vine models varies depending on the correlation among variables reflecting the spatio-temporal impacts of climate modes on Australian wheat crops.

The joint influences of multiple climate modes on wheat yield can be observed in this study, providing a mechanism to predict rainfall (Nguyen-Huy et al., 2017) and Australian wheat yield (as indicated in the present study. However, the mechanism of the interaction among climate mode indices and yield cannot be explained by the present statistical method. Furthermore, the wheat yield forecasts is found to be skillfully made 3–6 months ahead based on indices derived Pacific and Indian Ocean. However, recent studies have indicated that the variation in the tropical Atlantic region could trigger the occurrence of ENSO (Ham et al., 2013; Polo et al., 2014). Hence, it may be suggested that the SST anomalies in the Atlantic region could also have indirect impacts on the Australian wheat yield via Pacific region (Yuan and Yamagata, 2015). Moreover, while this study utilized the concurrent relationship between climate indices, the lagged relationship of the variation between the Indian and Pacific regions has been identified recently (Izumo et al., 2010). This reveals that the wheat yield can be potentially forecast at the longer lagged time by an efficient forecasting of ENSO phases 14 months ahead using the conditions in the Indian Ocean.

5. Conclusion

A total of twelve common climate drivers of Australian rainfall and agricultural production have been employed to investigate their relationship with the winter wheat yield. The influences of considered climate indices on the wheat yield anomaly exhibit a spatio-temporal change. The Indian Ocean region plays a key role in all the wheat belt regions, except for the Western part. The Pacific indices have more impact on the wheat growing regions in Queensland while the extra-tropical and tropical climate patterns affect mostly southeast and southern Australia. The co-occurrence of extreme events might enhance the fluctuations of the wheat yield. The forecasting information and respective models have significant applications for enhancing food security by enabling an earlier planning and design of agricultural strategies and policies to optimize the wheat yield.

The D-vine copula model is able to capture fully the joint dependence between different climate indices and wheat yield. Compared to asymmetric Archimedean and meta-elliptical copulas, D-vine performs better in extreme events (tail dependence) because it is allowed to consist of flexibly bivariate copula families. Further, the construction of D-vine based on the proposed approach helps to automatically choose the influential predictors and ignore the superfluous variables. It results in the improvement of the strength of forecasting the response. Compared to the common regression methods such as linear quantile regression, the D-vine model shows more accuracy of forecasting at different levels of quantiles.

Although this framework is performed for the wheat crop at the state scale, it can be potentially applied to other rainfed crops at a smaller scale. Regarding the site level, research should be undertaken into consideration of the teleconnection between large-scale climate modes and regional synoptic patterns (such as cut-off lows and easterly dips) that actually deliver the precipitation (Verdon-Kidd and Kiem, 2009). Future research might develop the nonparametric copula function for establishing a global forecasting model of Australian wheat yield using multiple climate drivers. Other potential research projects might consider the influences of new climate indices (e.g., derived from the tropical Atlantic) and the interactions between such new indices and well-known indices (e.g., SOI) on crop yields. Finally, researchers may explore further whether the lagged relationship between climate indices can be used to improve the crop yield forecasts in terms of the time lag and accuracy.

Acknowledgments

The project was financed by a University of Southern Queensland Post Graduate Research Scholarship (USQPRS 2015-2017), School of Agricultural, Computational and Environmental Sciences, Strategic Research Funding (SRF) Projects (Resilient Landscapes SRF and Computational Models SRF) and Climate Adaptation [DCAP] Projects (Producing Enhanced Crop Insurance Systems and Associated Financial Decision Support Tools). The authors would like to acknowledge constructive comments from the reviewers and the helpful proofreading work from Dr Barbara Harmes and Dr Kathryn Reardon-Smith. We are grateful to both reviewers and the Editor-in-Chief for their critical comments based on which we have reached a clear and concise high-quality version of the paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.eja.2018.05.006.

References

Abdi, H., 2007. The Kendall rank correlation coefficient. Encycl. Meas. Stat. Sage, Thousand Oaks, CA, pp. 508–510.

Anderson, W., Seager, R., Baethgen, W., Cane, M., 2017. Crop production variability in North and South America forced by life-cycles of the El Niño Southern Oscillation. Agric. For. Meteorol. 239, 151–165. http://dx.doi.org/10.1016/j.agrformet.2017.03.008.

Ashok, K., Behera, S.K., Rao, S.A., Weng, H., Yamagata, T., 2007. El Niño Modoki and its possible teleconnection. J. Geophys. Res. Ocean 112. http://dx.doi.org/10.1029/2006JC003798.

Ashok, K., Guan, Z., Yamagata, T., 2003. Influence of the Indian Ocean Dipole on the Australian winter rainfall. Geophys. Res. Lett. 30. http://dx.doi.org/10.1029/2003GL017926.

Ashok, K., Yamagata, T., 2009. The El Nino with a difference. Nature 461, 481–483.

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Change Biol. 17, 997–1012. http://dx.doi.org/10.1111/j.1365-2486.2010.02262.x.

Bannayan, M., Crout, N.M.J., Hoogenboom, G., 2003. Application of the CERES-wheat model for within-season prediction of winter wheat yield in the United Kingdom. Agron. J. 114–125. http://dx.doi.org/10.2134/agronj2003.0114.

Barnwal, P., Kotani, K., 2013. Climatic impacts across agricultural crop yield distributions: an application of quantile regression on rice crops in Andhra Pradesh. India Ecol. Econ. 87, 95–109. http://dx.doi.org/10.1016/j.ecolecon.2012.11.024.

Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J., 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renew. Energy 40, 29–39. http://dx.doi.org/10.1016/j.renene.2011.08.015.

Brechmann, E.C., 2010. Truncated and Simplified Regular Vines and Their Applications. Diploma.

Cai, W., Van Rensch, P., Cowan, T., Hendon, H.H., 2012. An asymmetry in the IOD and ENSO teleconnection pathway and its impact on Australian climate. J. Clim. 25, 6318–6329. http://dx.doi.org/10.1175/JCLI-D-11-00501.1.

Cane, M.A., 2005. The evolution of El Niño, past and future. Earth Planet Sci. Lett. 230 (3-4), 227–240. http://dx.doi.org/10.1016/j.epsl.2004.12.003.

Ceglar, A., Turco, M., Toreti, A., Doblas-Reyes, F.J., 2017. Linking crop yield anomalies to large-scale atmospheric circulation in Europe. Agric. For. Meteorol. 240–241, 35–45. http://dx.doi.org/10.1016/j.agrformet.2017.03.019.

Chiew, F.H.S., Piechota, T.C., Dracup, J.A., McMahon, T.A., 1998. El Nino/Southern oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. J. Hydrol. 204, 138–149. http://dx.doi.org/10.1016/S0022-1694(97)00121-2.

Deo, R.C., Byun, H.-R., Adamowski, J.F., Begum, K., 2017. Application of effective drought index for quantification of meteorological drought events: a case study in Australia. Theor. Appl. Climatol. 128, 359–379. http://dx.doi.org/10.1007/s00704-015-1706-5.

Deo, R.C., Şahin, M., 2016. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ. Monit. Assess. 188, 90. http://dx.doi.org/10.1007/s10661-016-5094-9.

Deo, R.C., Syktus, J.I., McAlpine, C.A., Lawrence, P.J., McGowan, H.A., Phinn, S.R., 2009. Impact of historical land cover change on daily indices of climate extremes including droughts in eastern Australia. Geophys. Res. Lett. 36. http://dx.doi.org/10.1029/2009GL037666.

ESRI, 2016. ArcGIS Desktop: Release 10.5. Redlands CA. .

Garcia y Garcia, A., Persson, T., Paz, J.O., Fraisse, C., Hoogenboom, G., 2010. ENSO-based climate variability affects water use efficiency of rainfed cotton grown in the southeastern USA. Agric. Ecosyst. Environ. 139, 629–635. http://dx.doi.org/10.1016/j.agee.2010.10.009.

Gutierrez, L., 2017. Impacts of El Niño-Southern oscillation on the wheat market: a global dynamic analysis. PLoS One 12. http://dx.doi.org/10.1371/journal.pone.0179086.

Ham, Y.G., Kug, J.S., Park, J.Y., Jin, F.F., 2013. Sea surface temperature in the north tropical Atlantic as a trigger for El Niño/Southern oscillation events. Nat. Genet. 6, 112–116. http://dx.doi.org/10.1038/ngeo1686.

Hansen, J.W., Potgieter, A., Tippett, M.K., 2004. Using a general circulation model to forecast regional wheat yields in northeast Australia. Agric. For. Meteorol. 127, 77–92. http://dx.doi.org/10.1016/j.agrformet.2004.07.005.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int. J. Climatol. 34, 623–642. http://dx.doi.org/10.1002/joc.3711.

Hassani, B.K., 2016. Dependencies and relationships between variables. Scenario Analysis in Risk Management. Springer, pp. 141–158. http://dx.doi.org/10.1007/978-3-319-25056-4_11.

Huang, J.-J., Lee, K.-J., Liang, H., Lin, W.-F., 2009. Estimating value at risk of portfolio by conditional copula-GARCH method. Insur. Math. Econ. 45, 315–324. http://dx.doi.org/10.1016/j.insmatheco.2009.09.009.

Iizumi, T., Luo, J.-J., Challinor, A.J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M.E., Yamagata, T., 2014. Impacts of El Niño Southern oscillation on the global yields of major crops. Nat. Commun. 5, 3712. http://dx.doi.org/10.1038/ncomms4712.

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S.K., Luo, J.-J., Cravatte, S., Masson, S., Yamagata, T., 2010. Influence of the state of the Indian Ocean Dipole on the following year's El Niño. Nat. Genet. 3, 168–172. http://dx.doi.org/10.1038/ngeo760.

Jarvis, C., Darbyshire, R., Eckard, R., Goodwin, I., Barlow, E., 2018. Influence of El Niño-Southern oscillation and the Indian Ocean Dipole on winegrape maturity in Australia. Agric. For. Meteorol. 248, 502–510. http://dx.doi.org/10.1016/j.agrformet.2017.10.021.

Jin, E.K., Kinter, J.L., Wang, B., Park, C.K., Kang, I.S., Kirtman, B.P., Kug, J.S., Kumar, A., Luo, J.J., Schemm, J., Shukla, J., Yamagata, T., 2008. Current status of ENSO prediction skill in coupled ocean-atmosphere models. Clim. Dyn. 31, 647–664. http://dx.doi.org/10.1007/s00382-008-0397-3.

Kao, S.-C., Govindaraju, R.S., 2010. A copula-based joint deficit index for droughts. J. Hydrol. 380, 121–134. http://dx.doi.org/10.1016/j.jhydrol.2009.10.029.

Kraus, D., Czado, C., 2017. D-vine copula based quantile regression. Comput. Stat. Data Anal. 110, 1–18. http://dx.doi.org/10.1016/j.csda.2016.12.009.

Kumar, A., Zhang, L., Wang, W., 2013. Sea surface temperature–precipitation relationship in different reanalyses. Mon. Weather Rev. 141, 1118–1123. http://dx.doi.org/10.1175/MWR-D-12-00214.1.

Laborte, A.G., de Bie, K.C.A.J.M., Smaling, E.M.A., Moya, P.F., Boling, A.A., Van Ittersum, M.K., 2012. Rice yields and yield gaps in Southeast Asia: past trends and future outlook. Eur. J. Agron. 36, 9–20. http://dx.doi.org/10.1016/j.eja.2011.08.005.

Li, Z., Cai, W., Lin, X., 2016. Dynamics of changing impacts of tropical Indo-Pacific variability on Indian and Australian rainfall. Sci. Rep. 6. http://dx.doi.org/10.1038/srep31767.

Lim, E.-P., Hendon, H.H., Zhao, M., Yin, Y., 2016. Inter-decadal variations in the linkages between ENSO, the IOD and south-eastern Australian springtime rainfall in the past 30 years. Clim. Dyn. 1–16.

Liu, Z., Zhou, P., Chen, X., Guan, Y., 2015. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. J. Geophys. Res. Atmos. 120. http://dx.doi.org/10.1002/2015JD023787.

Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 150, 1443–1452.

Lobell, D.B., Cahill, K.N., Field, C.B., 2007. Historical effects of temperature and precipitation on California crop yields. Clim. Change 81, 187–203. http://dx.doi.org/10.1007/s10584-006-9141-3.

Lobell, D.B., Field, C.B., Cahill, K.N., Bonfils, C., 2006. Impacts of future climate change on California perennial crop yields: model projections with climate and crop uncertainties. Agric. For. Meteorol. 141, 208–218. http://dx.doi.org/10.1016/j.agrformet.2006.10.006.

Lu, X.F., Lai, K.K., Liang, L., 2014. Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model. Ann. Oper. Res. 219, 333–357. http://dx.doi.org/10.1007/s10479-011-0900-9.

Ludescher, J., Gozolchiani, A., Bogachev, M.I., Bunde, A., Havlin, S., Schellnhuber, H.J., Ludescher®, J., Gozolchianib, A., Bogacheva, M.I., Bunde, A., Havlinb, S., Schellnhuberd, H.J., 2013. Improved El Niño forecasting by cooperativity detection source. Proc. Natl. Acad. Sci. U. S. A. 110, 11742–11745. http://dx.doi.org/10.1073/pnas.1309353110.

Luo, J.-J., Zhang, R., Behera, S.K., Masumoto, Y., Jin, F.-F., Lukas, R., Yamagata, T., 2010. Interaction between El Nino and extreme Indian ocean dipole. J. Clim. 23, 726–742. http://dx.doi.org/10.1175/2009JCLI3104.1.

Matsumura, K., Gaitan, C.F., Sugimoto, K., Cannon, A.J., Hsieh, W.W., 2015. Maize yield forecasting by linear regression and artificial neural networks in Jilin. China J. Agric. Sci. 153, 399–410. http://dx.doi.org/10.1017/S0021859614000392.

Min, S.K., Cai, W., Whetton, P., 2013. Influence of climate variability on seasonal extremes over Australia. J. Geophys. Res. Atmos. 118, 643–654. http://dx.doi.org/10.1002/jgrd.50164.

Mushtaq, S., An-Vo, D.-A., Christopher, M., Zheng, B., Chenu, K., Chapman, S.C., Christopher, J.T., Stone, R.C., Frederiks, T.M., Alam, G.M.M., 2017. Economic assessment of wheat breeding options for potential improved levels of post head-emergence frost tolerance. F. Crop. Res. 213, 75–88. http://dx.doi.org/10.1016/j.fcr.2017.07.021.

Nguyen-Huy, T., Deo, R.C., An-Vo, D.-A., Mushtaq, S., Khan, S., 2017. Copula-statistical precipitation forecasting model in Australia's agro-ecological zones. Agric. Water Manag. 191. http://dx.doi.org/10.1016/j.agwat.2017.06.010.

Omondi, P., Awange, J.L., Ogallo, L.A., Okoola, R.A., Forootan, E., 2012. Decadal rainfall variability modes in observed rainfall records over East Africa and their relations to historical sea surface temperature changes. J. Hydrol. 464, 140–156. http://dx.doi.org/10.1016/j.jhydrol.2012.07.003.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Şaylan, L., Šiška, B., Rötter, R., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35. http://dx.doi.org/10.1016/j.eja.2011.05.001.

Polo, I., Martin-Rey, M., Rodriguez-Fonseca, B., Kucharski, F., Mechoso, C.R., 2014. Processes in the Pacific La Niña onset triggered by the Atlantic Niño. Clim. Dyn. 44, 115–131. http://dx.doi.org/10.1007/s00382-014-2354-7.

Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000-global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Glob. Biogeochem. Cycles 24http://dx.doi.org/10.1029/2008GB003435. n/a-n/a.

Potgieter, A., Meinke, H., Doherty, A., Sadras, V.O., Hammer, G., Crimp, S., Rodriguez, D., 2013. Spatial impact of projected changes in rainfall and temperature on wheat yields in Australia. Clim. Change 117, 163–179. http://dx.doi.org/10.1007/s10584-012-0543-0.

Potgieter, A.B., Hammer, G.L., Butler, D., 2002. Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. Crop Pasture Sci. 53, 77–89. http://dx.doi.org/10.1071/AR01002.

Potgieter, A.B., Hammer, G.L., Meinke, H., Stone, R.C., Goddard, L., 2005. Three putative types of El Niño revealed by spatial variability in impact on Australian wheat yield. J. Clim. 18, 1566–1574. http://dx.doi.org/10.1175/JCLI3349.1.

Power, S., Casey, T., Folland, C., Colman, A., Mehta, V., 1999. Inter-decadal modulation of the impact of ENSO on Australia. Clim. Dyn. 15, 319–324.

R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL. http://www.R-project.org/.

Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-validation. Encyclopedia Database Syst. 532–538. http://dx.doi.org/10.1007/978-0-387-39940-9_565.

Revadekar, J.V., Preethi, B., 2012. Statistical analysis of the relationship between summer

monsoon precipitation extremes and foodgrain yield over India. Int. J. Climatol. 32, 419–429. http://dx.doi.org/10.1002/joc.2282.
Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. J. Clim. 20, 5473–5496. http://dx.doi.org/10.1175/2007JCLI1824.1.
Rimmington, G.M., Nicholls, N., 1993. Forecasting wheat yields in Australia with the Southern oscillation index. Crop Pasture Sci. 44, 625–632. http://dx.doi.org/10.1071/AR9930625.
Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C., Hendon, H.H., 2009. On the remote drivers of rainfall variability in Australia. Mon. Weather Rev. 137, 3233–3253. http://dx.doi.org/10.1175/2009MWR2861.1.
Royce, F.S., Fraisse, C.W., Baigorria, G.A., 2011. ENSO classification indices and summer crop yields in the Southeastern USA. Agric. For. Meteorol. 151, 817–826. http://dx.doi.org/10.1016/j.agrformet.2011.01.017.
Schepen, A., Wang, Q.J., Robertson, D., 2012. Evidence for using lagged climate indices to forecast Australian seasonal rainfall. J. Clim. 25, 1230–1246. http://dx.doi.org/10.1175/JCLI-D-11-00156.1.
Schepsmeier, U., 2015. Efficient information based goodness-of-fit tests for vine copula models with fixed margins: a comprehensive review. J. Multivar. Anal. 138, 34–52. http://dx.doi.org/10.1016/j.jmva.2015.01.001.
Schepsmeier, U., Stoeber, J., Brechmann, E.C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., 2017. Package "VineCopula.".
Shuai, J., Zhang, Z., Sun, D.Z., Tao, F., Shi, P., 2013. ENSO, climate variability and crop yields in China. Clim. Res. 58, 133–148. http://dx.doi.org/10.3354/cr01194.
Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges 8 Université Paris.
Taschetto, A.S., England, M.H., 2009. El Niño Modoki impacts on Australian rainfall. J. Clim. 22, 3167–3174. http://dx.doi.org/10.1175/2008JCLI2589.1.
Turner, N.C., 2004. Agronomic options for improving rainfall-use efficiency of crops in dryland farming systems. J. Exp. Bot. 55, 2413–2425. http://dx.doi.org/10.1093/jxb/erh154.
Ummenhofer, C.C., England, M.H., McIntosh, P.C., Meyers, G.A., Pook, M.J., Risbey, J.S., Sen Gupta, A., Taschetto, A.S., 2009. What causes southeast Australia's worst droughts? Geophys. Res. Lett. 36. http://dx.doi.org/10.1029/2008GL036801.
Verdon-Kidd, D.C., Kiem, A.S., 2009. On the relationship between large-scale climate modes and regional synoptic patterns that drive Victorian rainfall. Hydrol. Earth Syst. Sci. 13, 467–479. http://dx.doi.org/10.5194/hess-13-467-2009.
Weller, E., Cai, W., 2013. Asymmetry in the IOD and ENSO teleconnection in a CMIP5 model ensemble and its relevance to regional rainfall. J. Clim. 26, 5139–5149. http://dx.doi.org/10.1175/JCLI-D-12-00789.1.
Worley, S.J., Woodruff, S.D., Reynolds, R.W., Lubker, S.J., Lott, N., 2005. ICOADS release 2.1 data and products. Int. J. Climatol. 25, 823–842. http://dx.doi.org/10.1002/joc.1166.
Yuan, C., Yamagata, T., 2015. Impacts of IOD, ENSO and ENSO Modoki on the Australian winter wheat yields in recent decades. Sci. Rep. 5. http://dx.doi.org/10.1038/srep17252.
Zhang, L., Singh, V.P., 2014. Trivariate flood frequency analysis using discharge time series with possible different lengths: Cuyahoga river case study. J. Hydrol. Eng. 19, 5014012. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001003.
Zhang, N., Kang, C., Xia, Q., Liang, J., 2014. Modeling conditional forecast error for wind power in generation scheduling. IEEE Trans. Power Syst. 29, 1316–1324. http://dx.doi.org/10.1109/TPWRS.2013.2287766.

Supplementary Materials

A. Copula models

A.1 Vine copulas

A copula is a multivariate function that captures the dependence structure between random variables irrespective of their marginal distributions. Furthermore, the copula models are based on rank dependence (i.e. Kendall's tau correlations), so they do not suffer from the distributional assumptions made in linear models and as such, they are able to model the non-linear relationships between different variables. In this section, the copula theory is introduced briefly but readers can refer to the works of Joe (1997) and Nelsen (2006) for more theoretical details.

Suppose each continuous variable X_i (i.e., the climate mode index and the wheat yield used in this study) has its own cumulative distribution function (CDF) and the probability density function (PDF) (marginal CDF and PDF) denoted as $F_i(x_i)$ and $f_i(x_i)$, respectively. The joint CDF of a *d*-dimensional random variable $(X_1,...,X_d)$ can be written as (cf. Sklar (1959)):

$$F(x_1,...,x_d) = C\left[F_1(x_1),...,F_d(x_d)\right], \tag{A.1}$$

and the corresponding joint PDF (cf. Sklar (1996)):

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

The associated function $C:[0,1]^d \to [0,1]$ used to joint *d* marginal distributions is unique, so-called *copula*, and $c = \dfrac{\partial^d C(u_1,...,u_d)}{\partial u_1,...,\partial u_d}$ is the corresponding *copula density* where $u_i = F_i(x_i)$ with $u_i \in [0,1]$ is a probability integral transform (PIT). It is clear that, from the equation (A.1), the dependence structure can be modelled separately with the marginal distributions indicating the advantage of copula approach compared to other methods mentioned above.

While there are several common copula-based models denoted as symmetric, asymmetric Archimedean and multivariate elliptical copulas, they all present some challenges and constraints. Firstly, in regards to the symmetric model, only a few of the bivariate Archimedean copulas can be extended to the higher dimensions (i.e., $d \geq 3$) and this can limit their practical functionalities. Furthermore, for the asymmetric methods, the copula functions in each level must belong to the same family where the estimated parameters in higher levels are smaller than those in the lower levels (Zhang and Singh, 2014). For example, the joint CDF in equation (A.1) of three variables constructed by the asymmetric Archimedean, known as the hierarchical Archimedean copula (HAC), is expressed as:

$$F\left(x_1, x_2, x_3\right) = C_2\left\{C_1\left[F_1\left(x_1\right), F_2\left(x_2\right); \theta_1\right], F_3\left(x_3\right); \theta_2\right\}, \tag{A.3}$$

where C_1 and C_2 are the same Archimedean copula family with the corresponding parameters $\theta_1 \geq \theta_2$ (Later, we show in Figure 1(b) to illustrate the construction of the HAC model). However, this assumption, which is data-specific, may not be reasonable in reality (Zhang and Singh, 2014). Finally, the tail dependences present in some data may not be captured well in the case of multivariate elliptical copulas (Genest and Favre, 2007; Zhang and Singh, 2014). For example, the meta-elliptical copulas including the meta-Gaussian (meta-N) and the meta-Student t (meta-T) (Fang et al., 2002; Song and Singh, 2010) could encounter this issue where the former copula function has no tail dependence, while the latter copula function is likely to have the symmetric tail dependence.

Fortunately, these limitations, however, can be overcome by the use of vine copulas that were first introduced by Joe (1996) and further redeveloped extensively by Bedford and Cooke (2002, 2001). The vine copula can be expressed in three basic forms: regular (R)-vine, canonical (C)-vine and D-vine copulas. In short, the vine approach decomposes the joint PDF represented in the equation (A.1) into the (conditional) bivariate copula densities (i.e., the so-called pair-copulas) and its marginal densities are then expressed as (Kraus and Czado, 2017):

$$f\left(x_1, ..., x_d\right) = \prod_{k=1}^{d} f_k\left(x_k\right) \prod_{i=1}^{d-1} \prod_{j=i+1}^{d} c_{ij;i+1,...,j-1}\left[F_{i|i+1,...,j-1}\left(x_i \middle| x_{i+1}, ..., x_{j-1}\right), F_{j|i+1,...,j-1}\left(x_j \middle| x_{i+1}, ..., x_{j-1}\right)\right]. \tag{A.4}$$

In this regards, the selection of each pair-copula is then independent of each other. This expression is called the D-vine copula if all marginal distributions are uniform. By such a construction of the joint distribution function, the vine copulas are quite flexible in modelling the asymmetric and tail dependences of high-dimensional random variables, thus addressing the limitations of conventional copula functions (Aas et al., 2009; Haff et al., 2010; Stoeber et al., 2013).

A.2 Model construction

Suppose we have three-dimensional data set $\left(Y, X_1, X_2\right)$ where Y denotes the response (*i.e.*, wheat yield in this study) and $\left(X_1, X_2\right)$ are the two covariates (*e.g.*, climate mode indices). All three variables are transferred to the pseudo-copula data $\left(V, U_1, U_2\right)$ using their empirical distribution functions (i.e., 'PIT' as mentioned before). It is important to mention that this nonparametric transformation allows the modeler to estimate the yet to be jointly distributed copula parameters irrespectively of the marginal distributions using the maximum pseudo-likelihood (Chowdhary et al., 2011; Genest and Favre, 2007). Since we later need the inverse of the estimated marginal distributions for the quantile regression model (as shown in the

equation (2) and (3)), we also fit the yield data to the parametric distribution. Thus, we apply both graphical tools and statistical test (i.e., the Kolmogorov-Smirnov statistic (KS)) to a set of theoretical distribution functions to select the parsimonious approximate fitting of the data. If the p-value of the KS test is greater than 0.05, we cannot reject the null hypothesis that the data follow that specific distribution. Then, the distribution with a lower Akaike Information Criterion (AIC) is selected for that data. The selected distributions with the estimated parameters, AIC, and p-value for the wheat yields data of five states are represented in the table C1.

To illustrate the process, Figure B1 (a) represents the three-dimensional D-vine copula model where the dashed box implies whether or not the covariate U_2 should be included into the copula model. As a benchmark measure to the three-dimensional D-vine, the present study also employed the asymmetric and meta-elliptical copulas in order to compare the different types of copulas used for the simulation of wheat yield.

The asymmetric Archimedean, referred hereafter the hierarchical Archimedean copula (HAC), is constructed in accordance with Figure B1 (b) where the dashed box corresponds to the addition of the covariates for the D-vine model in Figure B1 (a).

The primary steps employed to select the most parsimonious D-vine copula model is stated as follows:

1. Thirty-two bivariate copula functions (further explained later) are first fitted to the two pairs (V,U_1) and (V,U_2). The most appropriate bivariate copula model selected for each pair is based on the AIC.
2. To select most influential covariate, i.e. the first covariate after V in the D-vine model, we compute the corrected AIC conditional log-likelihood (cll_{AIC}) for the two pair-copulas (V,U_1) and (V,U_2) in accordance with the proposed technique of (Kraus and Czado, 2017):

$$cll_{AIC}\left(l,F,\theta,V,U\right)=-2cll\left(l,F,\theta,V,U\right)+2\left|\theta\right| \tag{A.5}$$

in which

$$cll\left(l,F,\theta,V,U\right)=\sum_{i=1}^{n}\ln c_{V|U}\left(\hat{v}^{(i)}\middle|\mathrm{u}^{(i)};l,F,\theta\right). \tag{A.6}$$

It is imperative to mention that the term $l=(\ell_1,...,\ell_d)$ is the ordering of the *d*-dimensional D-vine copula, in this case, F being the estimated parametric pair-copula families with the corresponding copula parameters θ given the pseudo-copula data (V,U). The conditional copula density $c_{V|U}$ is expressed in term of the product over all pair-copulas of the D-vine model that include V (Kraus and Czado, 2017):

$$c_{V|\mathrm{U}}\left(\hat{v}^{(i)}\middle|\mathrm{u}^{(i)};l;F;\theta\right)=c_{VU_{\ell_1}}\left(\hat{v}^{(i)},u_{\ell_1}^{(i)};F_{VU_{\ell_1}},\theta_{VU_{\ell_1}}\right)$$
$$\times\prod_{j=2}^{d}c_{VU_{\ell_j};U_{\ell_1},\ldots,U_{\ell_{j-1}}}\left[C_{V|U_{\ell_1},\ldots,U_{\ell_{j-1}}}\left(\hat{v}^{(i)}\middle|u_{\ell_1}^{(i)},\ldots,u_{\ell_{j-1}}^{(i)}\right),\right. \tag{A.7}$$
$$\left.C_{U_{\ell j}|U_{\ell_1},\ldots,U_{\ell_{j-1}}}\left(u_{\ell_j}^{(i)}\middle|u_{\ell_1}^{(i)},\ldots,u_{\ell_{j-1}}^{(i)}\right);F_{VU_{\ell j};U_{\ell_1},\ldots,U_{\ell_{j-1}}},\theta_{VU_{\ell j};U_{\ell_1},\ldots,U_{\ell_{j-1}}}\right].$$

Thus, in this example, the conditional copula density $c_{V|U}$ can be expressed for the bivariate case as

$$c_{V|\mathrm{U}}\left(\hat{v}^{(i)}\middle|\mathrm{u}^{(i)};l;F;\theta\right)=c_{VU_{\ell_1}}\left(\hat{v}^{(i)},u_{\ell_1}^{(i)};F_{VU_{\ell_1}},\theta_{VU_{\ell_1}}\right). \tag{A.8}$$

If the absolute value of cll_{AIC} of the pair (V,U_1) is larger than that of (V,U_2), then the first order of the tree 1 (T1) is denoted as U_1.

3. Next, we add the remaining variable U_2 to the current D-vine model, i.e., $V-U_1$, and then check whether this addition leads to an improvement of the cll_{AIC} value of the resulting model. The conditional copula density $c_{V|U}$ is written as:

$$c_{V|\mathrm{U}}\left(\hat{v}^{(i)}\middle|\mathrm{u}^{(i)};l;F;\theta\right)=c_{VU_{\ell_1}}\left(\hat{v}^{(i)},u_{\ell_1}^{(i)};F_{VU_{\ell_1}},\theta_{VU_{\ell_1}}\right)$$
$$\times c_{VU_{\ell_2};U_{\ell_1}}\left[C_{V|U_{\ell_1}}\left(\hat{v}^{(i)}\middle|u_{\ell_1}^{(i)}\right),C_{U_{\ell_2}|U_{\ell_1}}\left(u_{\ell_2}^{(i)}\middle|u_{\ell_1}^{(i)}\right);F_{VU_{\ell_2};U_{\ell_1}},\theta_{VU_{\ell_2};U_{\ell_1}}\right]. \tag{A.9}$$

It is visible that equation (A.9) contains only the two pair-copulas $C_{VU_{\ell_1}}$ and $C_{U_{\ell_2}U_{\ell_1}}$. Then, following (Aas et al., 2009), the conditional distribution function can be applied as:

$$C_{V|U}=h_{V|U}=\frac{\partial C_{VU}(V,U)}{\partial U} \tag{A.10}$$

4. Finally, if the cll_{AIC} of the current model is improved, then the D-vine copula becomes the three-dimensional copula model with the structure of $V-U_1-U_2$. On the other hand, if the cll_{AIC} of the current model is not improved, the algorithm execution is stopped and the $V-U_1$ model is finally selected.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insur. Math. Econ. 44, 182–198. https://doi.org/https://doi.org/10.1016/j.insmatheco.2007.02.001

Bedford, T., Cooke, R.M., 2002. Vines: A new graphical model for dependent random variables. Ann. Stat. 1031–1068. https://doi.org/https://doi.org/10.1214/aos/1031689016

Bedford, T., Cooke, R.M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. 32, 245–268. https://doi.org/https://doi.org/10.1023/A:1016725902970

Fang, H.-B., Fang, K.-T., Kotz, S., 2002. The meta-elliptical distributions with given marginals. J. Multivar. Anal. 82, 1–16.

Genest, C., Favre, A.-C., 2007. Everything you always wanted to know about Copula modeling but were afraid to ask. J. Hydrol. Eng. 12, 347–368. https://doi.org/https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Haff, I.H., Aas, K., Frigessi, A., 2010. On the simplified pair-copula construction—Simply useful or too simplistic? J. Multivar. Anal. 101, 1296–1310. https://doi.org/https://doi.org/10.1016/j.jmva.2009.12.001

Joe, H., 1997. Multivariate models and multivariate dependence concepts. CRC Press. https://doi.org/https://doi.org/10.1201/b13150

Kraus, D., Czado, C., 2017. D-vine copula based quantile regression. Comput. Stat. Data Anal. 110, 1–18. https://doi.org/https://doi.org/10.1016/j.csda.2016.12.009

Nelsen, R.B., 2006. An introduction to copulas, 2nd ed. Springer.

Sklar, A., 1996. Random variables, distribution functions, and copulas: a personal look backward and forward. Lect. notes-monograph Ser. 1–14. https://doi.org/https://doi.org/10.1214/lnms/1215452606

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8.

Song, S., Singh, V.P., 2010. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stoch. Environ. Res. Risk Assess. 24, 425–444.

Stoeber, J., Joe, H., Czado, C., 2013. Simplified pair copula constructions—limitations and extensions. J. Multivar. Anal. 119, 101–118. https://doi.org/https://doi.org/10.1016/j.jmva.2013.04.014

Zhang, L., Singh, V.P., 2014. Trivariate flood frequency analysis using discharge time series with possible different lengths: Cuyahoga river case study. J. Hydrol. Eng. 19, 5014012. https://doi.org/https://doi.org/10.1061/(ASCE)HE.1943-5584.0001003

B. List of Figures

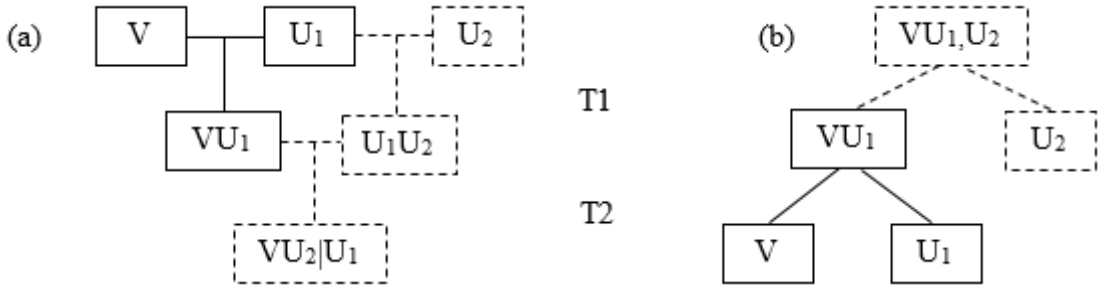


Figure B1. Model construction framework of three-dimensional *D*-vine (a) and asymmetric copulas (b) by adding the successively U_2 into the final model. The new pair-copulas indicated in the dashed box have to be estimated corresponding to the addition of the U_2.

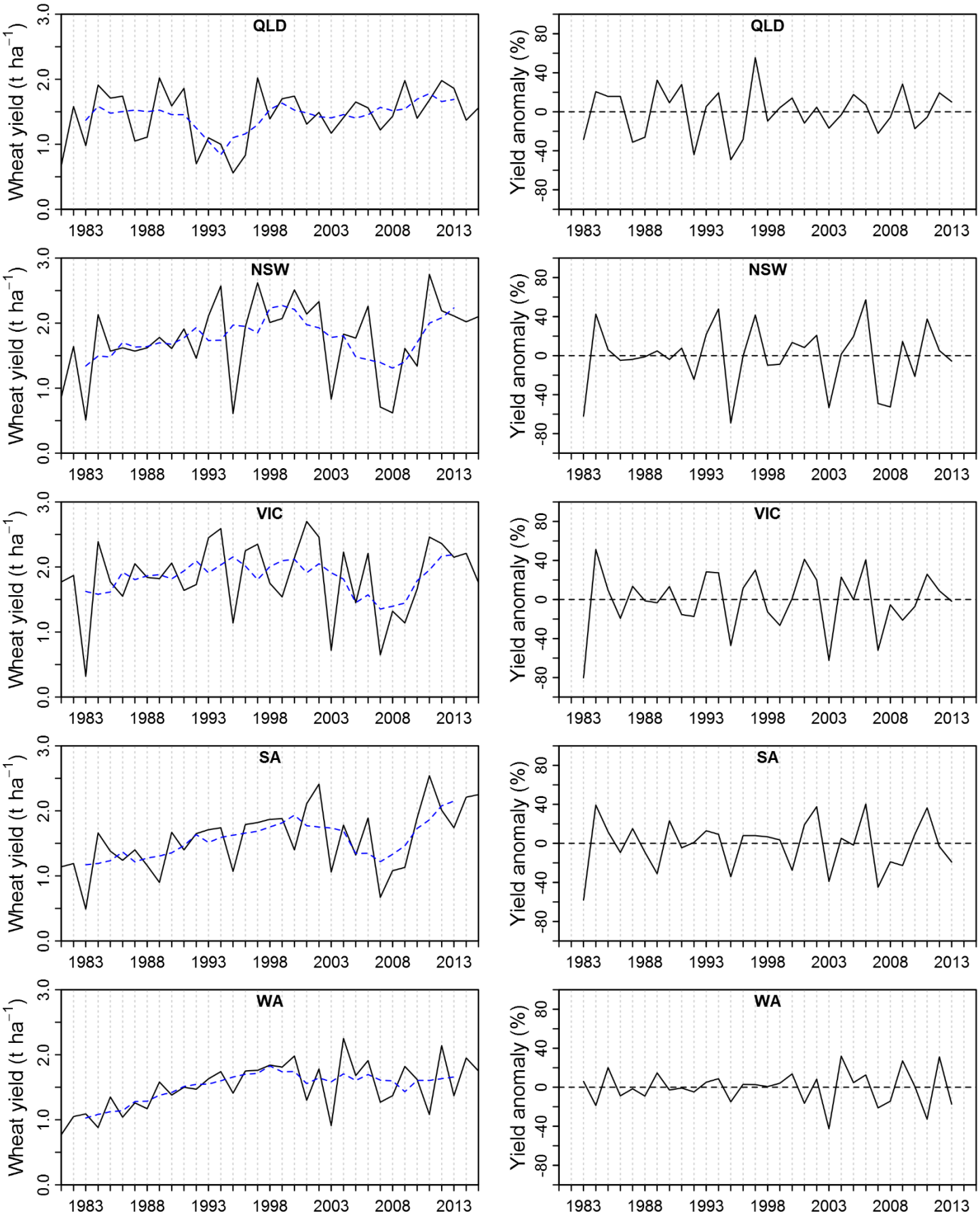
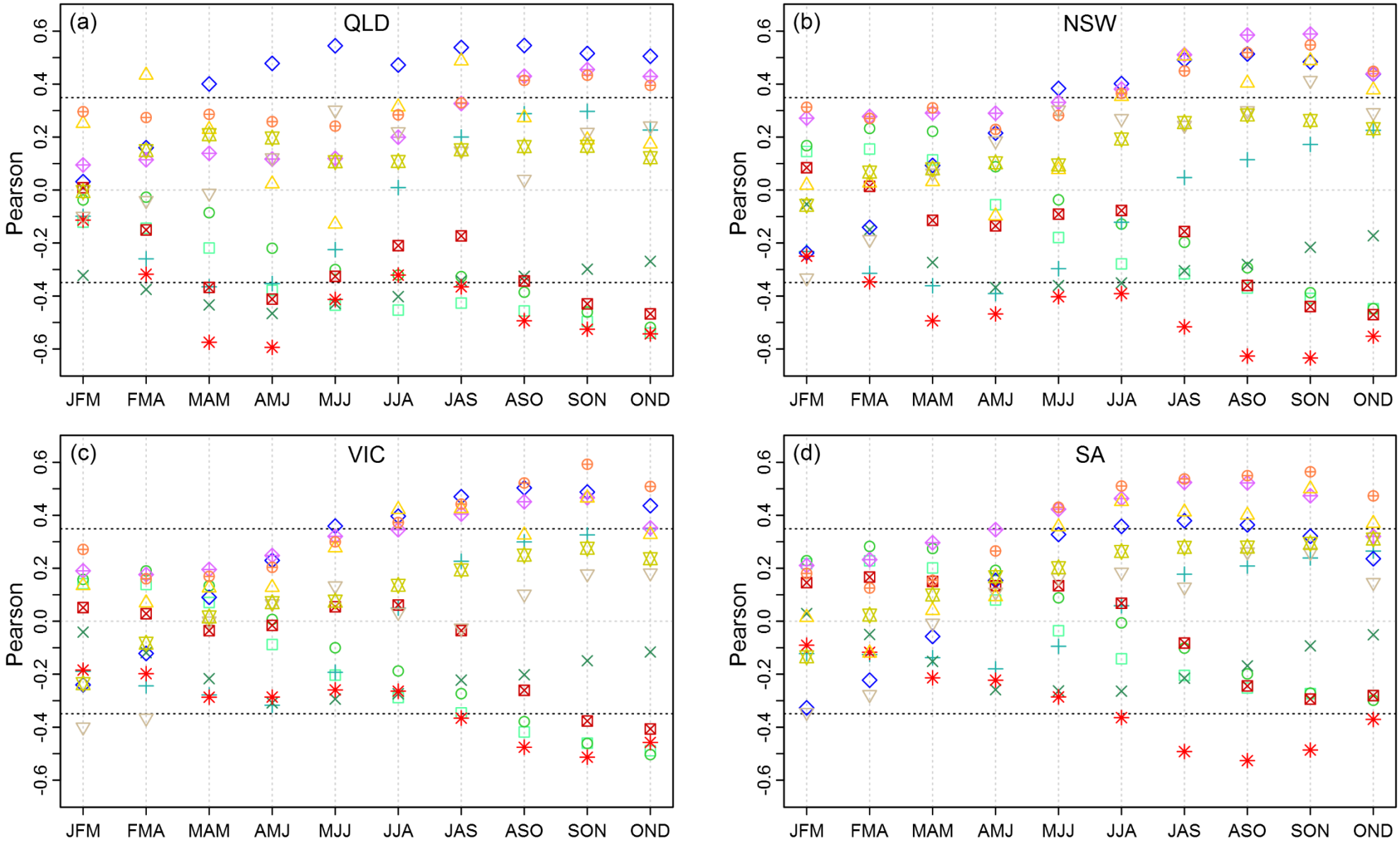


Figure B2. Time series of winter wheat yields with their trend (blue dashed lines) (left panel) and the corresponding anomalous percentages (detrended data) (right panel) in five major wheat-producing states, Australia.


(a)
QLD
(b)
NSW
(c)
VIC
(d)
SA
Pearson
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
JFM
FMA
MAM
AMJ
MJJ
JJA
JAS
ASO
SON
OND

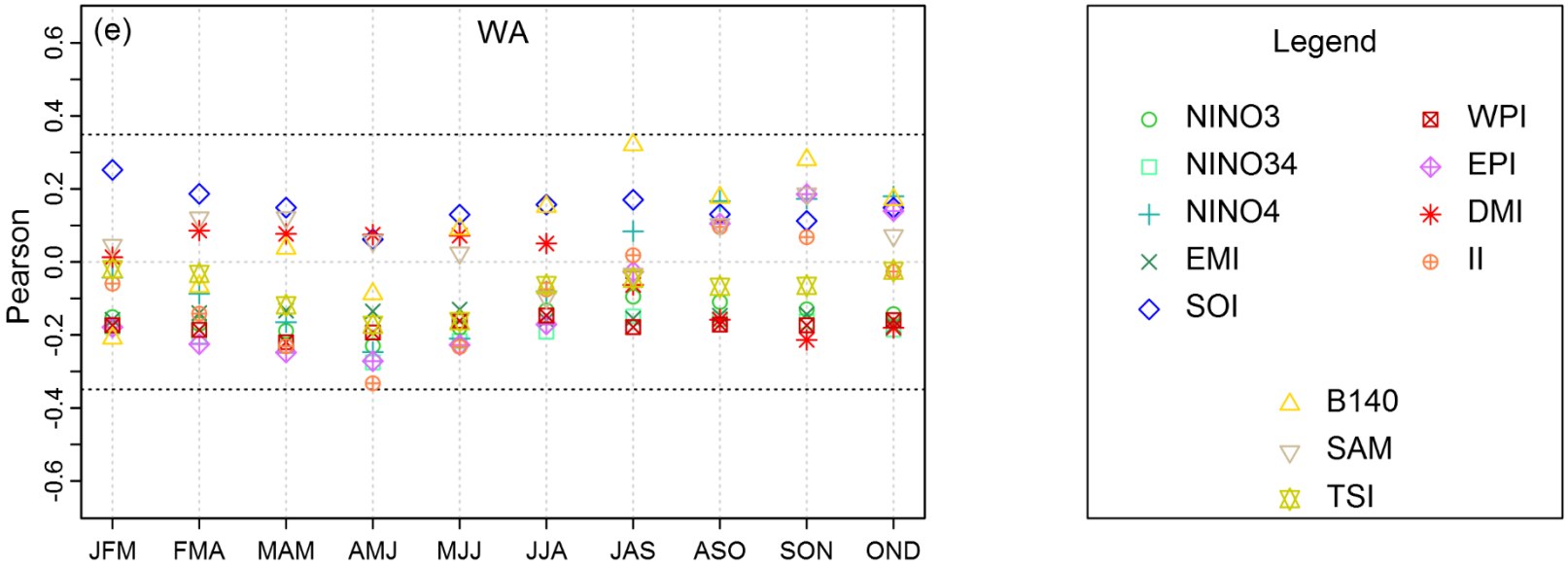
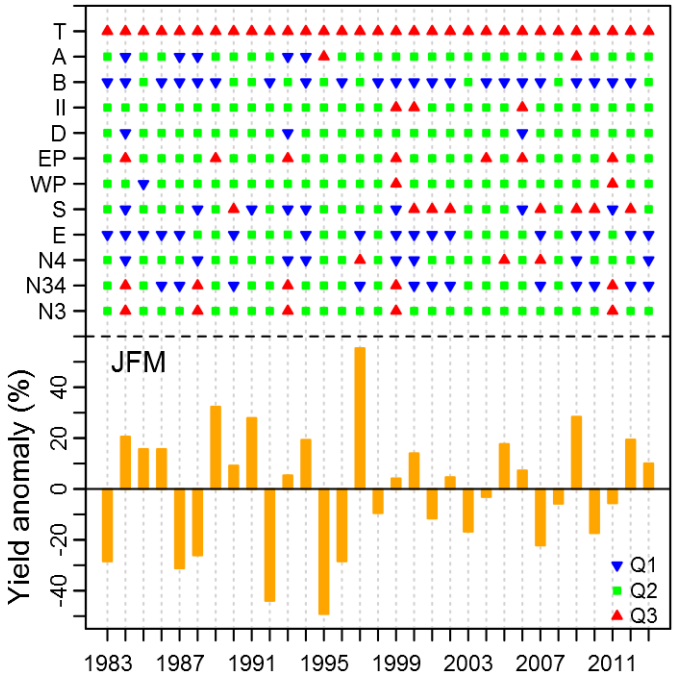
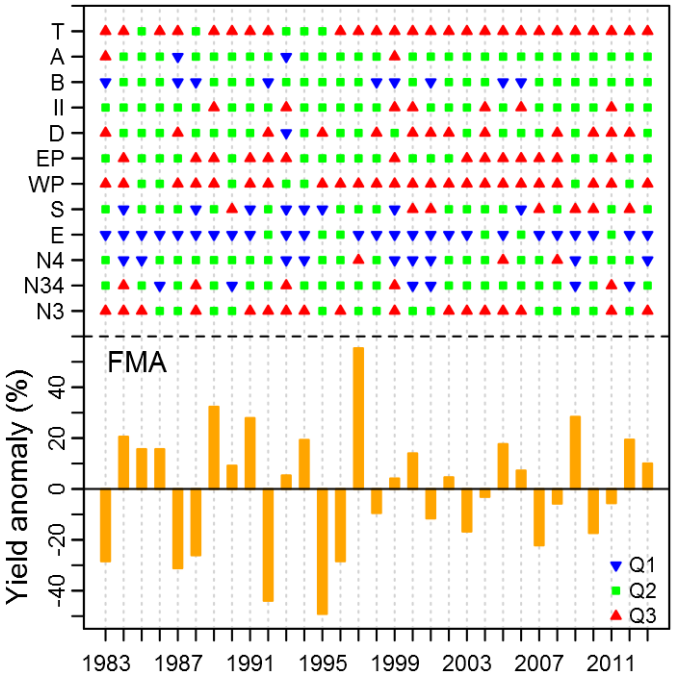
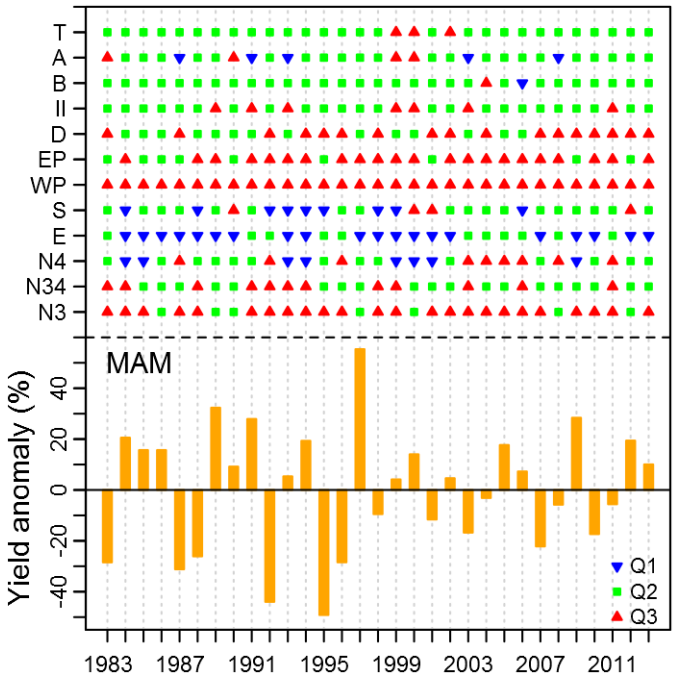
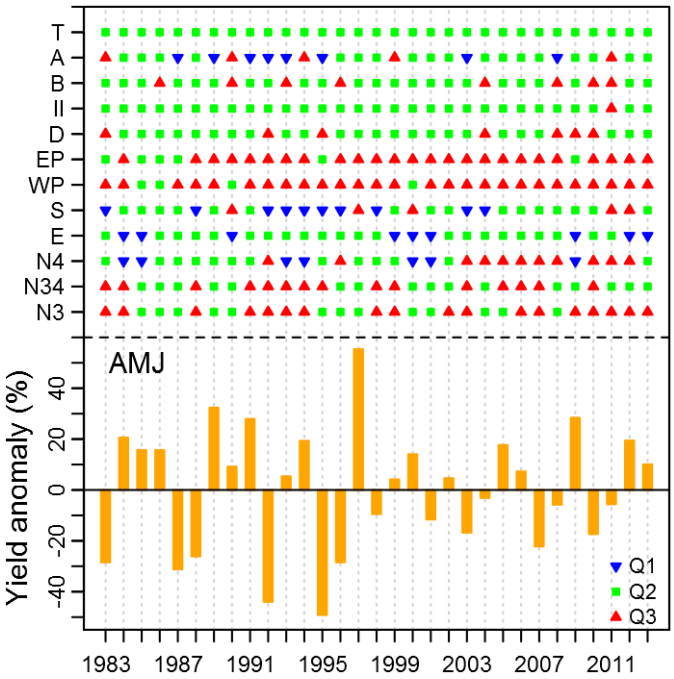


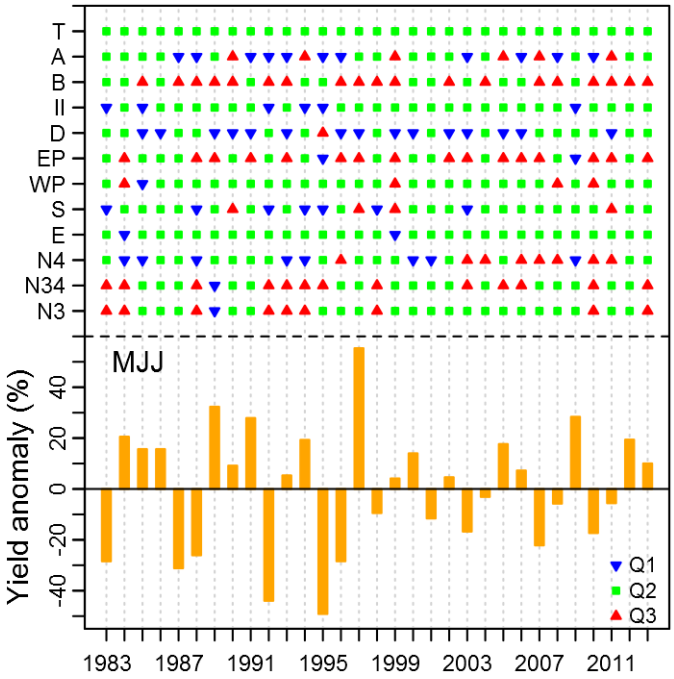
Figure B3. Pearson correlation between twelve climate indices and percentage wheat yield anomaly of five major wheat-producing states: QLD, NSW, VIC, SA, and WA. Horizontal dotted lines indicate significant at 95%.

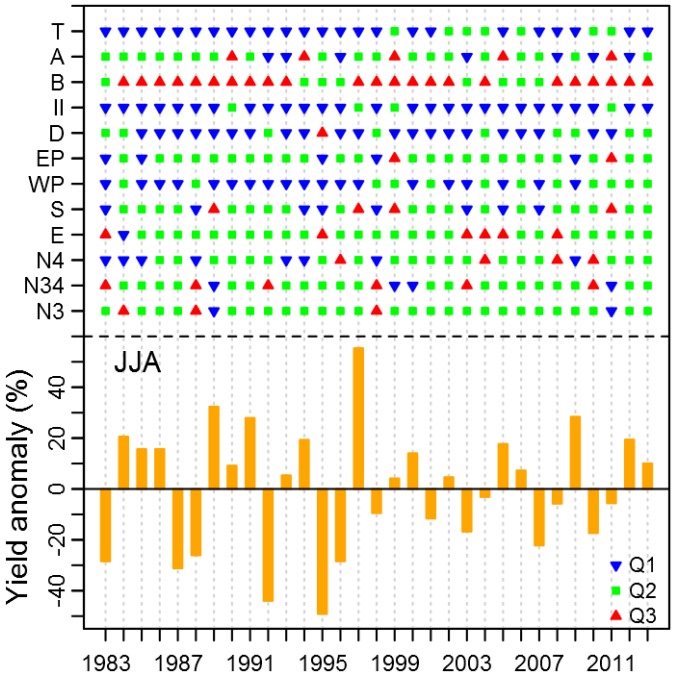

T
A
B
II
D
EP
WP
S
E
N4
N34
N3
JFM
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011


T
A
B
II
D
EP
WP
S
E
N4
N34
N3
FMA
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011


T
A
B
II
D
EP
WP
S
E
N4
N34
N3
MAM
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011


T
A
B
II
D
EP
WP
S
E
N4
N34
N3
AMJ
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011


T
A
B
II
D
EP
WP
S
E
N4
N34
N3
MJJ
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011


T
A
B
II
D
EP
WP
S
E
N4
N34
N3
JJA
Yield anomaly (%)
40
20
0
-20
-40
Q1
Q2
Q3
1983
1987
1991
1995
1999
2003
2007
2011

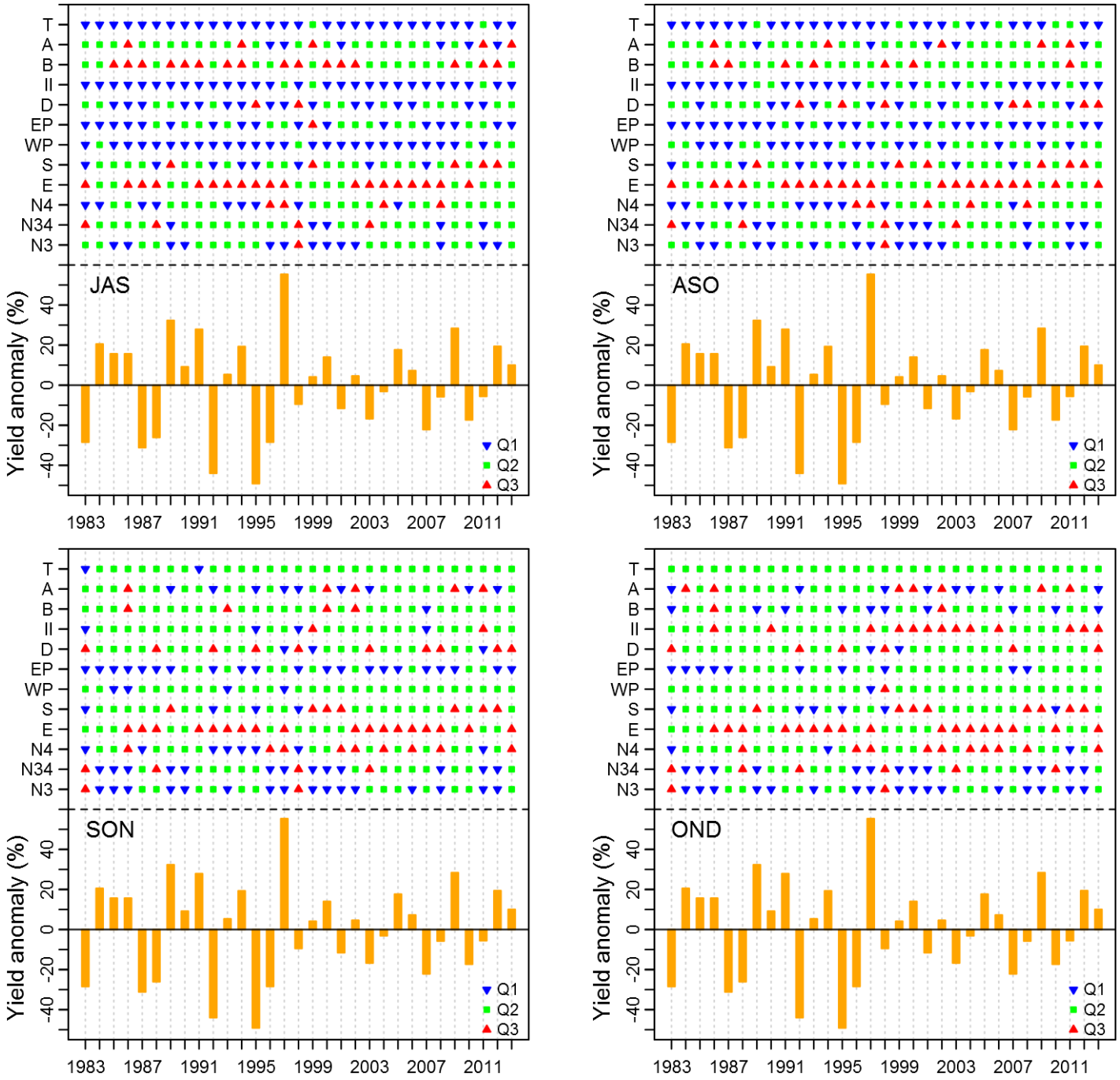
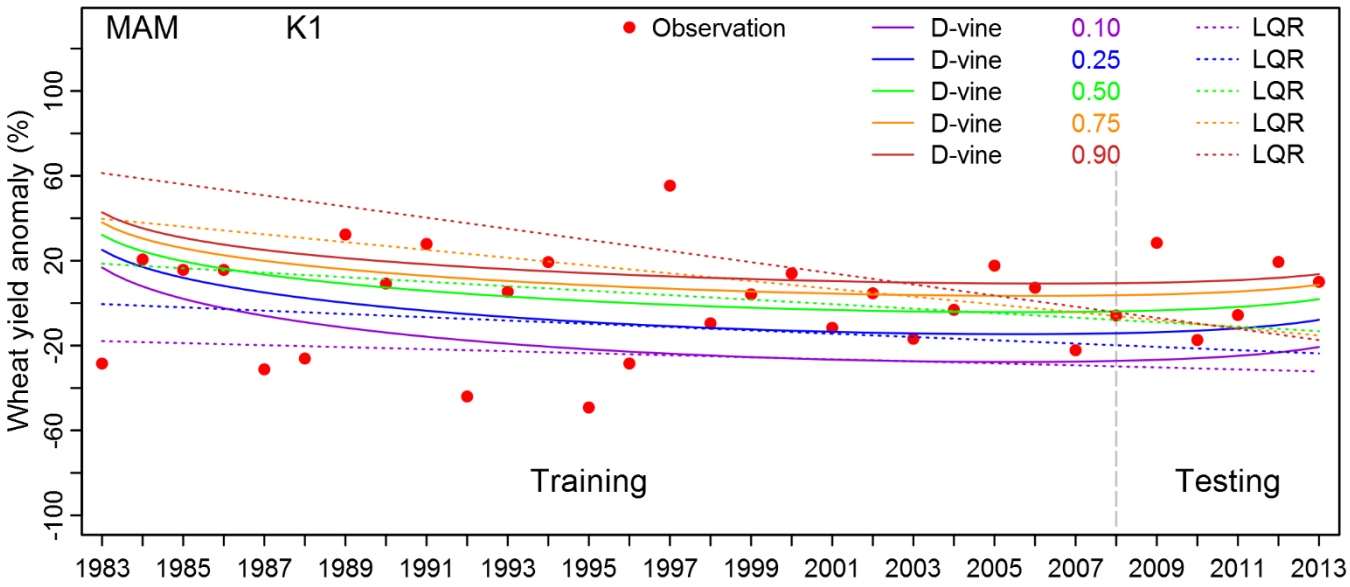
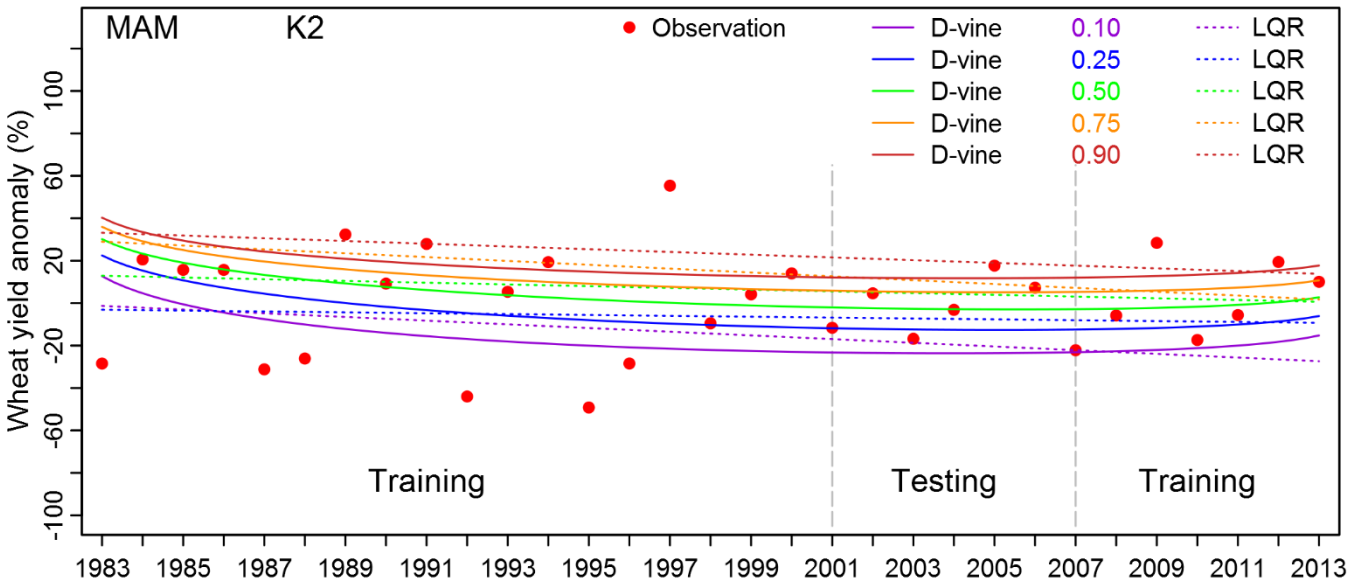
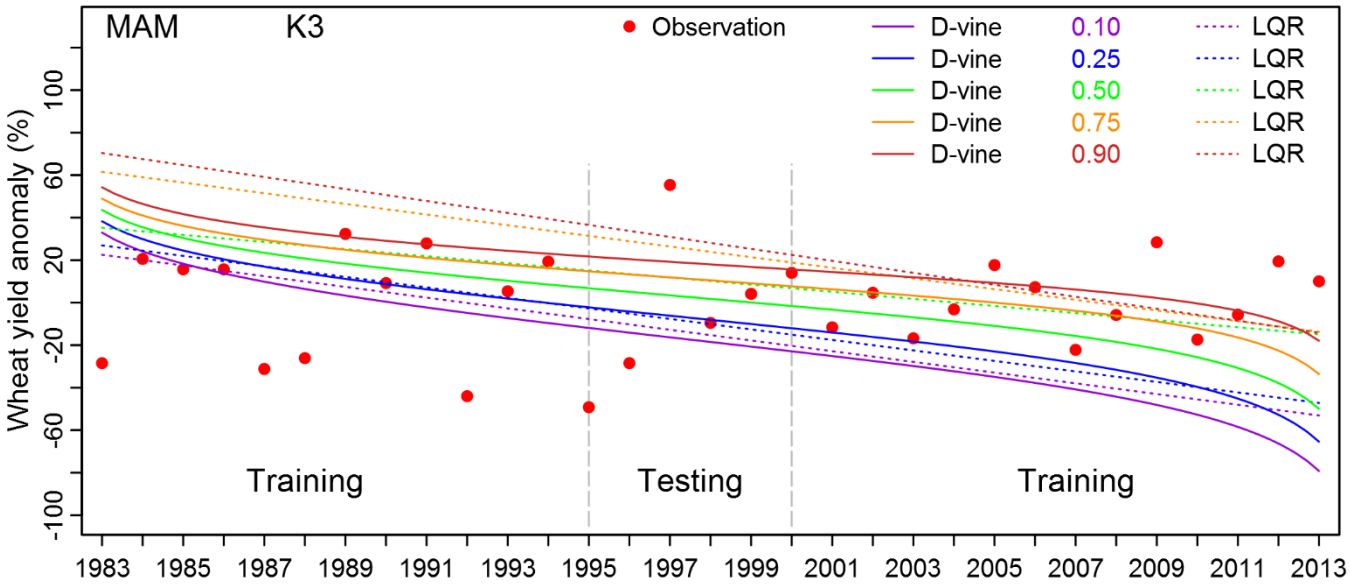


Figure B4. QLD percentage wheat yield anomalies in relation to three phases of climate indices classified as: negative phase (Q1), neutral phase (Q2), and positive phase (Q3). The positive and negative phase are assigned if the three-monthly average value of a climate index fell into the highest 25% and the lowest 25% of that climate time series, respectively. The neutral phase is identified by the remaining range value of the climate index. Notations for climate indices can be found in Table 1.


MAM
K1
Observation
D-vine 0.10
D-vine 0.25
D-vine 0.50
D-vine 0.75
D-vine 0.90
LQR
Wheat yield anomaly (%)
Training
Testing


MAM
K2
Observation
D-vine 0.10
D-vine 0.25
D-vine 0.50
D-vine 0.75
D-vine 0.90
LQR
Wheat yield anomaly (%)
Training
Testing
Training


MAM
K3
Observation
D-vine 0.10
D-vine 0.25
D-vine 0.50
D-vine 0.75
D-vine 0.90
LQR
Wheat yield anomaly (%)
Training
Testing
Training

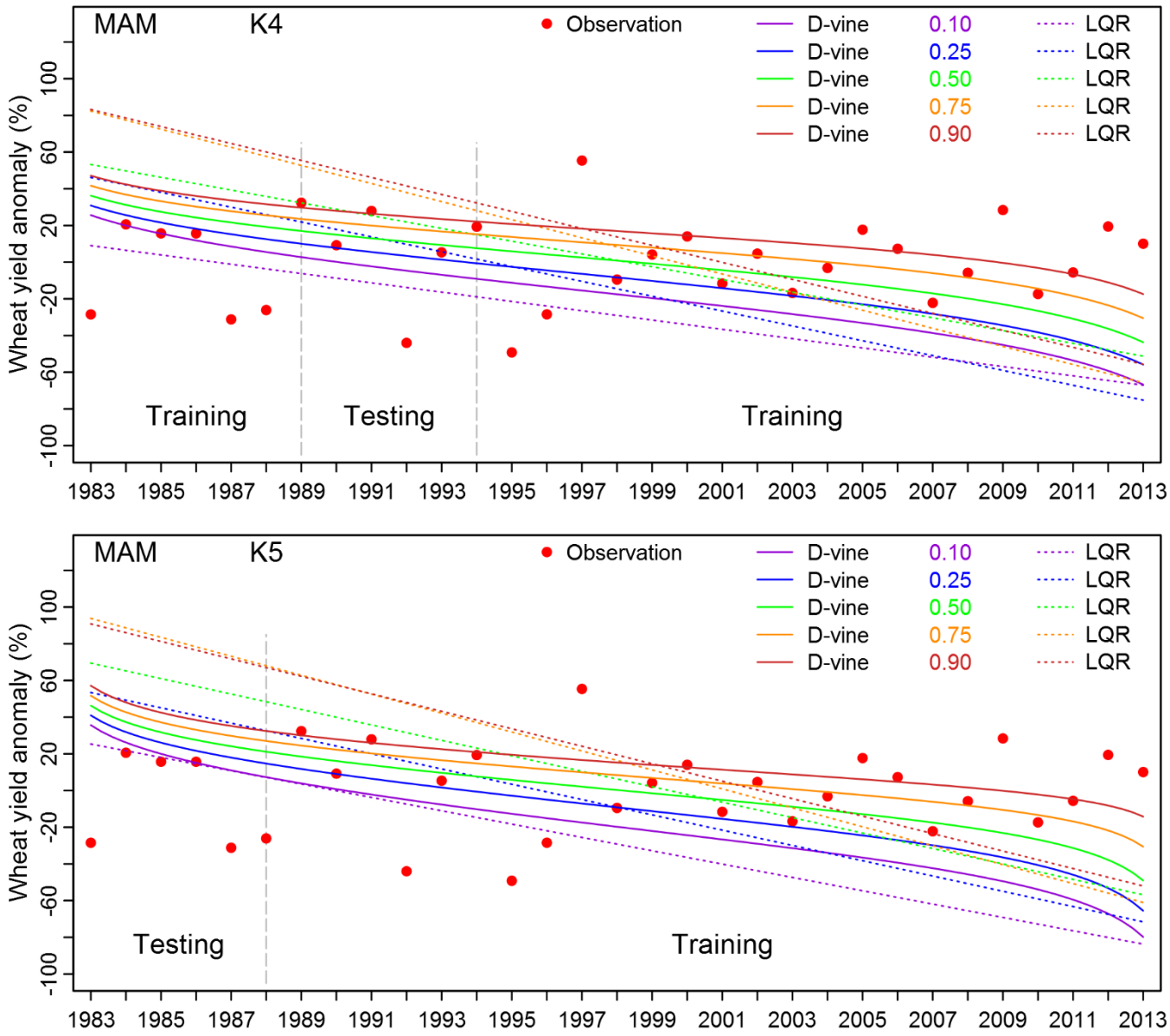


Figure B5. Comparison of fitted models between the D-vine copula (solid line) and linear quantile regression (LQR) (dotted line) at different quantiles. Grey dashed lines show division of training and testing dataset in 5-fold cross-validation. K1 to K5 represent the number of forecasting repeated.

C. List of Tables

Table C1. Selected marginal distributions of percentage wheat yield anomaly with parameters, Akaike's Information Criterion (AIC), and p-values of Kolmogorov-Smirnov test (p-KS) in five states.

States	*Distribution*	*Parameters*			*AIC*	*p-KS*
QLD	Normal	mean = 0.25	sd = 23.55		287.83	0.77
NSW	Normal	mean = -0.64	sd = 31.27		306.67	0.38
VIC	Generalised extreme value	shape = -0.51	scale = 32.73	location = -8.19	301.97	1.00
SA	Pert	min = -85.76	mode = 7.96	max = 48.57	289.40	0.89
WA	Normal	mean = -0.34	sd = 16.87		267.13	0.96

Table C2. Summary of the construction of D-vine copula model for Southern Australia (SA) during September-October-November with the name of copula family, estimated parameters ($\hat{\theta}1$, $\hat{\theta}2$), Kendall's tau (τ), upper and lower tail dependence coefficient (Utd and Ltd), conditional log-likelihood (*cll*) and AIC-corrected (cll_{AIC}) for each tree.

Tree	*Edge*	*Copula*	$\hat{\theta}1$	$\hat{\theta}2$	τ	*Utd*	*Ltd*	*cll*	cll_{AIC}
1	1,2	J_{180}	2.44	0.00	0.44	n\a	0.67	8.73	-15.46
	2,3	I	0.00	0.00	0.00	n\a	n\a	0.00	0.00
	3,4	C_{270}	-2.01	0.00	-0.50	0.00	0.00	0.00	0.00
2	1,3;2	J	1.71	0.00	0.28	0.50	n\a	3.55	-5.10
	2,4;3	I	0.00	0.00	0.00	n\a	n\a	0.00	0.00
3	1,4;2,3	C	0.53	0.00	0.21	n\a	0.27	2.19	-2.38
							Total	14.47	-22.94

Z_X : rotated Z copula X degrees

J: Joe

C: Clayton

I: Independence

Chapter 4

Systemic weather risk prediction and potential adaptation strategies

---

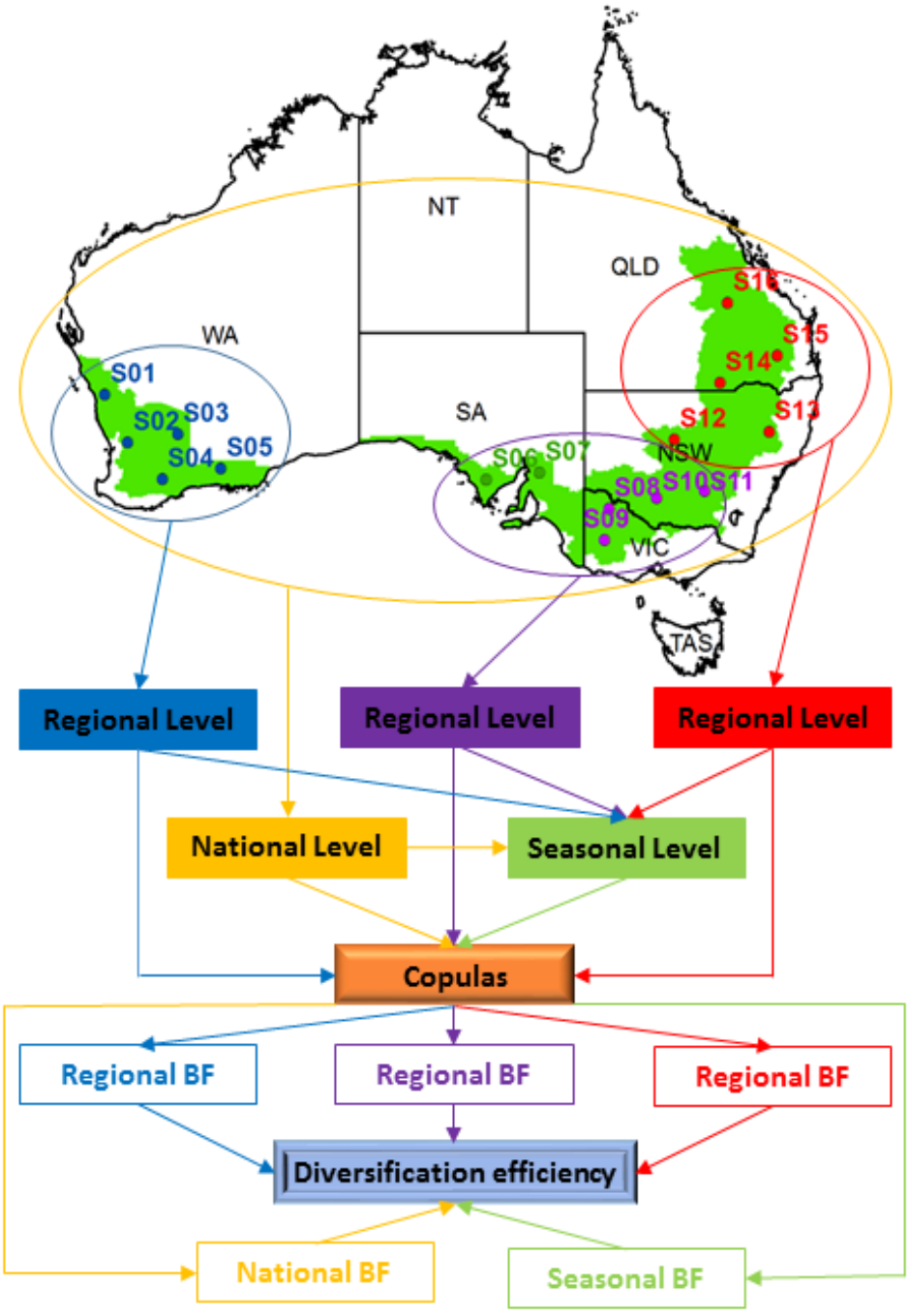
Article III: Copula statistical models for analysing stochastic dependencies of systemic drought risks and potential adaptation strategies

Summary:

In many cases, extreme weather events, such as droughts and floods, tend to cover a large area, affecting a large number of farmers. This phenomenon is therefore known as a systemic risk. This study develops C-vine copula-based models to measure the joint weather-related losses in insurance caused by insufficient precipitation occurring concurrently in different locations, or consecutively in different growing stages, in Australia. This modelling approach is enriched by a clustering analysis method known as the multidimensional Kruskal-Shephard scaling. The technique utilises the dissimilarity measure calculated by the empirical pairwise Kendall's τ of cumulative rainfall index (CRI). The daily precipitation data (1889 – 2012) are recorded in sixteen meteorological stations across the Australia's wheat belt spanning over different climatic conditions. Fig. 8 is the graphical display of this study on systemic weather risk and potential adaptation strategies for wheat crops in Australia.

On a regional scale, the study finds that drought events occurring in the west are more scattered during the October – December period, and for April – June and October – December in the eastern, south-eastern and southern regions. On a national scale, drought events in the east may spread out to the south-east and south but not to the west. The results also reveal that drought events in different seasons may not be perfectly correlated. Therefore, we conclude that applying spatial and temporal diversification strategies can feasibly reduce the systemic weather risks in Australia. In particular, the average risk-reducing effectiveness of the entire insurance area in regional, national and temporal scales ranges between 0.62 – 0.94, 0.48 – 0.76, and 0.25 – 0.33 corresponding to 5% and 25% strike levels. The findings indicate that diversifying risk over time potentially achieves a greater effectiveness than over space.

The findings of this study may act as an efficient tool for risk reduction mainly for farmers, but at the same time, it may also generate useful information for the pricing of weather index-based insurance products.



Source: Author

Figure 8. Graphical display of the study on systemic weather risk and potential adaptation strategies for wheat crops in Australia.

Manuscript
Click here to access/download;Manuscript;Nguyen-Huy et al 2018 Systemic Risk Text_revised.docx

Copula statistical models for analyzing stochastic dependencies of systemic drought risk and potential adaptation strategies

Thong Nguyen-Huy[1, 2, 3, 4], Ravinesh C Deo[1, 2, 3, *], Shahbaz Mushtaq[2, 3], Jarrod Kath [2, 3], Shahjahan Khan[1, 2, 3]

[1] School of Agricultural, Computational and Environmental Sciences

[2] Centre for Applied Climate Sciences (CACS)

[3] University of Southern Queensland, Institute of Agriculture and Environment (IAg&E)

QLD 4350, AUSTRALIA

[4]Vietnam National Space Center (VNSC), Vietnam Academy of Science and Technology (VAST)

Email: thonghuy.nguyen@usq.edu.au

* Corresponding Author (Dr R C Deo): ravinesh.deo@usq.edu.au

Abstract

Development of advanced model approaches for analyzing stochastic dependencies of systemic weather risk can help farmers, agricultural policy-makers and financial agents to address potential risk adaptation strategies and mitigation of threats to the agricultural industry. This study develops copula-based statistical models to provide a better understanding of systemic weather risks with agricultural and weather event data from Australia. In particular, we adopt a C-vine approach to model the joint insurance losses caused by drought events occurring simultaneously across the different spatial locations, and consecutively in different growing stages. This modelling approach is enriched by a clustering analysis process through the multidimensional Kruskal-Shephard scaling method. Daily rainfall data (1889–2012) recorded in sixteen meteorological stations across Australia's wheat belt spanning different climatic

conditions are employed. On a regional scale, droughts occurring in the west are more scattered during the October – December period and for April – June and October – December in the eastern, south-eastern and southern regions. On a national scale, drought events in the east are likely to spread out to the south-east and the south but not to the west. The results also reveal that the drought events in different seasons may not be perfectly correlated. Therefore, we conclude that the spatial and temporal diversification strategies are likely to feasibly reduce the systemic weather risk in Australia. In particular, the average risk-reducing effect of the entire insurance area in regional, national and temporal scales range in between 0.62–0.94, 0.48–0.76, and 0.25–0.33, corresponding to 5% and 25% strike levels. The findings asertain that diversifying the risk over time can potentially achieve a greater effect than over space. The outcomes may also act as an efficient tool for agricultural risk reduction, but simultaneously, it may also generate immensely useful information for suitable pricing of weather index-based insurance products.

Keywords: joint insurance losses; clustering; C-vine copulas; index-based insurance; weather systemic risk; diversification.

1. Introduction

Climate variability is a key risk affecting agricultural producers globally and, as a consequence, it can also affect their net revenues. Extreme weather conditions can lead to a partial or a complete loss of crops (Barriopedro et al. 2011; Coumou and Rahmstorf 2012). For example, climate-related disasters such as droughts, floods and tropical storms account for approximately 25% of all damage and losses in the agricultural sector in developing countries (FAO 2015). Furthermore, Lesk et al. (2016) have explained that a combination of drought and extreme heat events decreased national cereal production by 9 – 10% worldwide. In addition, this study also emphasizes that recent droughts cause about 7% greater production damage and

developed countries suffer 8 – 11% more damage than in developing countries. The research indicates that extreme weather disasters can be seen as sources of risk in the agricultural sectors across the globe even in developed countries with advanced technology and high agricultural yield.

It is clear that agriculture is one of the most weather sensitive sectors, requiring a financial protection in case of weather variabilities (Odening and Shen 2014). However, the existence of systemic weather risk has been determined as the leading reason for the failure of private insurance markets for agricultural crops (Duncan and Myers 2000; Miranda and Glauber 1997). Thus, financial-related problems cannot be easily solved unless efficient and affordable instruments for transferring systemic weather risks are available.

Published literature has noted that traditional multi-peril crop insurance has failed to provide affordable and comprehensive crop insurance in private insurance markets (Glauber et al. 2002; Goodwin 2001; Vedenov and Barnett 2004). The main cause of this failure is that weather risks often violate classical requirements for insurability, namely individual risks are independent or covariance risk is small (Okhrin et al., 2013). However, although this independence assumption may hold for some types of weather perils such as hail damage, it does not hold for other types (at least for the regional level). For example, a widespread drought is a slowly developing weather peril that is spatially correlated and causes systemic risks (Odening and Shen 2014; Xu et al. 2010). It means that many farmers may be affected at the same time, leading to a large number of concurrent insurance claims. For an insurance company, it is crucial to allocate separately sufficient capital (buffer fund) and thus add to the cost of insurance (premium loading) to avoid bankruptcy during high widespread systemic losses. Therefore, a question arises of how to model the joint insurance losses occur during a widespread weather event.

As mentioned above, high systemic weather risk and intermediation problems are often major impediments to viable crop insurance and may lead to a breakdown of an unsubsidized private

insurance market (Skees and Barnett 1999; Vedenov and Barnett 2004). However, there are several possible tools that allow for the administration of systemic risks such as reinsurance and weather derivatives (Musshoff et al. 2011; Skees et al. 2007). Insurers may spatially diversify the systemic weather risk through increasing its trading area (Xu et al. 2010). According to Odening and Shen (2014), the level of covariate risk depends on the size of the risk pool and thus it seems natural to reduce the systemic weather risk by increasing the regional dissemination of insurance products. For example, drought occurring within a small region may be highly correlated while these dependencies possibly disappear at a broader scale. Therefore, the question arises of how to quantify the spatio-temporal changes of the dependence structure among weather events.

The probabilistic quantification of large payouts due to the joint occurrence of unfavorable weather events at different locations is a particularly interesting topic in the context of weather insurance (Xu et al. 2010). The common approaches are based on simple correlation coefficients between weather indices measured at different weather stations. Subsequently, the correlation of weather variables can be easily described as a function of the distance among weather locations using a decorrelation process. This technique has been applied to measure the spatial dependence of weather events and evaluate the efficiency of risk pooling for cropping systems in the USA (Goodwin 2001; Holly Wang and Zhang 2003; Woodard and Garcia 2008).

While the use of linear correlation in risk context is computationally appealing, there are significant drawbacks (McNeil et al. 2005). Linear approaches are not able to capture nonlinear dependency between weather variables, and therefore, often do not contain information on the dependence structure of weather risks. As a result, different behavior can be observed from joint distributions, particularly in the upper and lower regions, with the same correlation coefficient. Therefore, the likelihood of extreme insurance losses can be either underestimated

or overestimated leading to inaccuracies in the pricing of insurance policies. Another alternative is the use of the multivariate normal distribution, where the joint distribution is uniquely defined using the marginal distributions of variables and their correlation matrix. However, weather indices and/or crop yield data are not always normally distributed (Nguyen-Huy et al. 2017; Nguyen-Huy et al. 2018a; Odening et al. 2007). A model that is able to handle the pitfalls of linear correlation and the estimation of multivariate distributions is required for better quantification of the relationship between weather events at different locations.

Multivariate copula-based models have recent become powerful instruments for analyzing the dependences structure between random variables. A copula is a function linking all individual univariate marginal distributions of variables into a full multivariate distribution (Sklar 1959). In this approach, the marginal distributions of the risk variables are determined independently with the copula estimation. Clearly, the copula approach allows for greater flexibility in modeling the dependence structures of weather risks compared to the simple linear correlation-based model (Nguyen-Huy et al. 2017; Nguyen-Huy et al. 2018a; Serinaldi 2009). Therefore, it is a more robust methodological framework for modeling spatial relationships of weather events, particularly the tail dependencies, that are important for quantifying risks during natural disasters (Embrechts et al. 2002; McNeil et al. 2005; Xu et al. 2017).

Although the copula theorem was introduced many years ago (Sklar 1959), it has been rediscovered recently with extensive applications in a number of fields. In finance and insurance, it is a standard and popular tool for multi asset pricing (Tankov 2011; Van Den Goorbergh et al. 2005), credit portfolio modeling (Frey and McNeil 2003; Frey et al. 2001) and risk management (Carreau and Bouvier 2016; Sak et al. 2017; Zhang et al. 2013). Copula-based models have also been successfully applied in the area of agriculture, climate prediction and other types of meteorological research (AghaKouchak et al. 2014; Reddy and Singh 2014; Schoelzel and Friederichs 2008; Song and Singh 2010; Zhang et al.).

However, applications of the copula technique in agricultural economics are very limited in the published literature. Nguyen-Huy et al. (2018a) investigated the dependence of multiple climate mode indices and Australian wheat yield at different quantile levels that potentially benefits to farmers and risk-management processes. In regards to revenue insurance, Zhu et al. (2008) inspected the dependence structure of prices and yields whereas Vedenov (2008) studied the association between individual farm yields and area yields. In the context of systemic weather risk, Xu et al. (2010) first applied the knowledge of copulas to estimate spatial dependence between weather events in Germany. In that study, only thirty-four observations were available for the estimation of a four-dimensional copula-based model that may result in poor statistical reliability.

Okhrin et al. (2013) extended that study by using a greater length of time series data derived from daily weather models to model systemic weather risk in China. These works, however, use either multivariate symmetric (Xu et al. 2010) or asymmetric Archimedean (Okhrin et al. 2013) copulas that have some restrictions in practice. The multivariate symmetric Archimedean copulas use a single parameter for all variable pairs, implying that they have the same dependence structure. The asymmetric Archimedean copulas are able to overcome the restriction of symmetric methods allowing modeling the tail dependencies. However, while asymmetric Archimedean copulas assume that all variable pairs can be modeled by the same copula function, these assumptions may be unrealistic in practice (Musafer and Thompson 2017; Nguyen-Huy et al. 2017; Nguyen-Huy et al. 2018a).

Motivated by the reasons mentioned above, this study utilizes vine copulas to analyze the spatial and temporal dependence structures of weather events across different zones in Australia and the associated joint losses of a hypothetical crop insurance written on these weather perils. Subsequently, the risk-reducing efficiency of spatial and temporal diversification strategies is assessed by comparing the buffer load (BL) of a joint insurance of

n stations and the average BL of all those single stations. The diversification strategy and its effect are implemented and evaluated at three levels: regional, national and temporal. In particular, for the regional scale, the trading area is extended by adding each weather station within that region to each corresponding aggregation level. For the national scale, one weather station is randomly selected from each region to join in the risk pooling strategy. Finally, temporal diversification is performed in each station where weather events occurring at different growing stages are aggregated together.

The importance and contribution of the present study are emphasized in the following ways. Firstly, Australia is an agricultural nation suffering one of the world's most variable climate conditions (Nguyen-Huy et al. 2018a). However, to the best of our knowledge, the statistical copula-based models have not been applied for the quantification of weather systemic risk in the agricultural sector in Australia. Secondly, the vine copulas overpower other copula methods by allowing the modeling dependence structure of variable pairs with different copula functions and taking asymmetric tail dependencies into account. Thirdly, the stochastic dependence of weather events is analyzed either spatially or temporally. Furthermore, the present study uses historical rainfall data recorded at sixteen weather stations covering the period 1889 – 2012. These weather stations span different climatic conditions across Australia's wheat belt. These long time series data are selected in order to improve the estimation of copula models.. Therefore, this comprehensive study on the stochastic dependence of systemic weather risk will provide useful information and a robust tool to better appreciate and quantify climatic risks and joint losses in the agricultural sector in Australia.

Based on the primary aims and objectives, the remaining parts of the present paper are structured as follows. Section 2 concisely presents some basic properties of the copula theorem and its special form, the vine copulas. Section 3 describes materials and the application of copulas for simulating the weather-related insurance losses. The results and discussion are

presented in sections 4 and 5, respectively. The study ends with some conclusions on the effectiveness of the copula-based measurement of stochastically dependent risks and risk-reducing strategies.

2. Copula-based statistical model

2.1. Copula functions

Suppose an *n*-dimensional random vector $X=\left(X_1,...,X_n\right)$ with a joint density function $f\left(x_1,...,x_n\right)$. Based on Rosenblatt's transform (Rosenblatt 1952), the density function can be factorized as (Aas et al. 2009):

$$\begin{aligned} f\left(x_1,...,x_n\right) = f_n\left(x_n\right).f\left(x_{n-1}\middle|x_n\right) \\ .f\left(x_{n-2}\middle|x_{n-1},x_n\right)...f\left(x_1\middle|x_2,...,x_n\right). \end{aligned} \tag{1}$$

Clearly, every joint distribution function comprises a description of the individual margin of each variable and a description of their dependency structure. Copulas provide an efficient way to separate the description of their dependency structure.

Sklar's theorem (Sklar 1959) states that if F is a multivariate distribution function with margins $F_1\left(x_1\right),...,F_n\left(x_n\right)$, then there exists a copula C such that:

$$F\left(x_1,...,x_n\right)=C\left[F_1\left(x_1\right),...,F_n\left(x_n\right)\right]=C\left(u_1,...,u_n\right), \tag{2}$$

where $F_i\left(x_i\right)=u_i$ for $i=1,...,n$ with $U_i\sim U\left[0,1\right]$ is the univariate probability integral transformation (PIT). If all margins are continuous, then C is uniquely defined on $RanF_1\times...\times RanF_n$. Conversely, the copula from Eq. (1) has the expression:

$$C\left(u_1,...,u_n\right)=F\left[F_1^{-1}\left(u_1\right),...,F_n^{-1}\left(x_n\right)\right], \tag{3}$$

where $F_i^{-1}(u_i)$ denotes the inverse distribution function of the margin. Clearly, the copula $C(u_1,...,u_n)$ is a multivariate distribution function with all margins being uniformly distributed on unit square. Also, if the corresponding copula density is defined as:

$$c(u_1,...,u_n) = \frac{\partial^n C(u_1,...,u_n)}{\partial u_1 ... \partial u_n}, \tag{4}$$

and F strictly increases, then the multivariate density function is expressed by:

$$\begin{aligned} f(x_1,...,x_n) &= c_{1...n}\left[F_1(x_1),...,F_n(x_n)\right].f_1(x_1)...f_n(x_n) \\ &= \left[\prod_{i=1}^{n} f_i(x_i)\right] c(u_1,...,u_n), \end{aligned} \tag{5}$$

where $f_i(x_i)$ is the marginal density. Based on the properties and construction methods, copula functions can be generally categorized into different groups such as symmetric elliptical, symmetric and asymmetric Archimedean, vine, empirical and entropy copulas. Taking the advantages of vine copulas mentioned in section 1 into consideration, and its recent applications in rainfall and agricultural production forecasting (Nguyen-Huy et al. 2017; Nguyen-Huy et al. 2018a; Pham et al. 2016), this study makes an application of the canonical vine (*i.e.*, C-vine) approach (Aas et al. 2009; Kurowicka 2005), a special class of vine copulas (Bedford and Cooke 2001; Bedford and Cooke 2002), to model the stochastically spatial dependence of weather events for the specific case of Australia.

2.2. C-vine copulas

The C-vine is a graphical model that decomposes the copula density in Eq. (5) in a specific nested set of trees and nodes. Each tree T_j, $j = 1,...,n$ has $n+1-j$ nodes linked together by $n-j$ edges where each edge is associated with a pair-copula density. The *n*-dimensional copula density expressed in a C-vine form may be generally written as:

$$c\left(u_1,...,u_n\right)=\prod_{j=1}^{n-1}\prod_{i=1}^{n-j}c_{j,j+i|1,...,j-1}\left[C_{j|1,...,j-1}\left(u_j\left|u_1,...,u_{j-1}\right.\right),C_{j+i|1,...,j-1}\left(u_{j+i}\left|u_1,...,u_{j-1}\right.\right)\right]. \quad (6)$$

Here, we illustrate the C-vine model for five random variables as an example (Fig. 1). As seen in Fig. 1, the pairwise dependences between the five variables U_1, U_2, U_3, U_4 and U_5 are captured by the four bivariate copulas (pair-copulas) C_{12}, C_{13}, C_{14} and C_{15}. These pair-copulas can be conditioned on the root variable U_1 using partial differentiation to obtain the corresponding conditional copula functions $C_{2|1}, C_{3|1}, C_{4|1}$ and $C_{5|1}$. Then, three new bivariate copulas are fitted to these four conditioned pseudo data resulting in the three pair-copulas $C_{23|1}, C_{24|1}$ and $C_{25|1}$. This process is implemented sequentially to obtain $C_{45|123}$.

<Figure 1>

The general expression of the five-dimensional density distribution in the C-vine structure is:

$$\begin{aligned}f\left(x_1,x_2,x_3,x_4,x_5\right)= & f_1\left(x_1\right).f_2\left(x_2\right).f_3\left(x_3\right).f_4\left(x_4\right).f_5\left(x_5\right)\\ & .c_{12}\left[C_{12}\left(u_1,u_2\right)\right].c_{13}\left[C_{13}\left(u_1,u_3\right)\right].c_{14}\left[C_{14}\left(u_1,u_4\right)\right].c_{15}\left[C_{15}\left(u_1,u_5\right)\right]\\ & .c_{23|1}\left[C_{2|1}\left(u_2\left|u_1\right.\right),C_{3|1}\left(u_3\left|u_1\right.\right)\right].c_{24|1}\left[C_{2|1}\left(u_2\left|u_1\right.\right),C_{4|1}\left(u_4\left|u_1\right.\right)\right]\\ & .c_{25|1}\left[C_{2|1}\left(u_2\left|u_1\right.\right),C_{5|1}\left(u_5\left|u_1\right.\right)\right].c_{34|12}\left[C_{3|12}\left(u_3\left|u_1,u_2\right.\right),C_{4|12}\left(u_4\left|u_1,u_2\right.\right)\right]\\ & .c_{35|12}\left[C_{3|12}\left(u_3\left|u_1,u_2\right.\right),C_{5|12}\left(u_5\left|u_1,u_2\right.\right)\right]\\ & .c_{45|123}\left[C_{4|123}\left(u_4\left|u_1,u_2,u_3\right.\right),C_{5|123}\left(u_5\left|u_1,u_2,u_3\right.\right)\right],\end{aligned} \quad (7)$$

where the conditional copula function can be derived sequentially such as:

$$C_{2|1}\left(u_2\left|u_1\right.\right)=\frac{\partial C_{12}\left(u_1,u_2\right)}{\partial u_1}, \quad (8)$$

$$C_{3|12}\left(u_3\left|u_1,u_2\right.\right)=\frac{\partial C_{23|1}\left[C_{2|1}\left(u_2\left|u_1\right.\right),C_{3|1}\left(u_3\left|u_1\right.\right)\right]}{\partial C_{2|1}\left(u_2\left|u_1\right.\right)}, \quad (9)$$

$$C_{4|123}\left(u_4|u_1,u_2,u_3\right)=\frac{\partial C_{34|12}\left[C_{3|12}\left(u_3|u_1,u_2\right),C_{4|12}\left(u_4|u_1,u_2\right)\right]}{\partial C_{3|12}\left(u_3|u_1,u_2\right)}. \quad (10)$$

The construction of the C-vine model is beneficial to the purpose of the present study where an arbitrary weather station will be selected for investigating the spread of extreme events. In particular, the selected station is located at the root of the C-vine model similar to variable one in Fig. 1. Other weather stations are then conditioned on the variable 1 given some specific quantiles.

3. Materials and Method

This section represent how the data are collected, computed indices and used in corresponding steps of the tests and copula-based approach in different locations and times. A summary description of the main steps is illustrated in Fig. 2.

<Figure 2>

3.1. Insurance losses

In the context of weather-related insurance, particular attention is paid to quantifying the probability of significant large claims for indemnities made when unfavorable weather events occur at the same time at different places. Therefore, for an insurance company, it is important to estimate sufficient the buffer fund (BF) as reserve to handle indemnity payments and avoid bankruptcy during widespread systemic losses. BF is defined as the Value-at-Risk (VaR) of the net losses of the insurer, i.e., the total indemnity payments minus the insurance premium (Okhrin et al. 2013; Xu et al. 2010):

$$BF=P\left[\sum_{i=1}^{n}w_i\left\{L\left(X_i\right)-\pi_i\right\}\geq BF\right]=1-\gamma, \quad (11)$$

where $L(X_i)$ is the weather-related indemnity payment for trading area or stage i. Here, π_i is the corresponding (fair) insurance premium defined as $E[L(X_i)]$. w_i is the weight of the ith insurance contract and $1-\gamma$ is the ruin probability. Dividing the BF by the number of contracts m yields the buffer load $BL = BF / m$.

The risk-reducing efficiency achieved by spatial or temporal diversification is assessed by dividing the BL for a joint insurance of *n* regions or stages to the average BL for a single region or stage (Okhrin et al. 2013):

$$DE = \frac{BL_n^*}{\left(\sum_{i=1}^{n} BL_i\right) n^{-1}}, \tag{12}$$

where *DE* and BL_n^* denote the diversification efficiency over the whole aggregation scenario, respectively, and BL_i is the buffer load of location or stage i.

3.2. Weather indices

In this study, the cumulative rainfall index (CRI), which represents drought risk (Martin et al. 2001; Xu et al. 2010), is employed as a hypothetical index-based insurance policy in Australia. Sixteen weather stations across the Australia's grain belt are selected for the analysis of dependencies of weather events (Fig. 3).

<Figure 3>

Cumulative rainfall index measures the rainfall total in a particular period as (Xu et al. 2010):

$$CRI_{i,t} = \sum_{j=T_A}^{T_J} P_{j,t,i}\,, \tag{13}$$

where $P_{j,t,i}$ is the daily precipitation in millimeter (mm) at day j in year t and station i, $i = 1,...,n$. T_A and T_J denote the beginning and the end of the three main stages of the wheat

cropping season (the main cereal in Australia), respectively, which are April 1 – June 30 (sowing stage), July 1 – September 30 (vegetation stage), and October 1 – 31 December (harvest stage). The mean and standard deviation of the CRI at sixteen weather stations in the three cropping stages are represented in Table 1. For better analysis of spatial and temporal dependencies between weather events, we also plot the median and quantiles of CRI in these stations as in Fig. 4

<Table 1>

<Figure 4>

Indemnities are paid if CRI falls below a predefined trigger level K_i^{CRI}:

$$L_{CRI_{i,t}} = \max\left\{0, K_i^{CRI} - CRI_{i,t}\right\} V, \tag{14}$$

where V is the tick size converting physical units into monetary terms. As the present study does not attempt an optimal contract design in terms of hedging efficiency, we simply set $V = 1$ AUD. Another assumption is that no policy limits apply (Xu et al. 2010). Since we want to investigate the buffer load in different drought conditions, the 5%- (extreme drought) and the 25%-quantiles (moderate drought) of the CRI distribution are assumed for the two strike levels (the strike level is the point at which a payout is triggered for an index insurance contract).

3.3. Measuring spatio-temporal weather dependencies using C-vine copulas

The calculation of the BF in Eq. (11) requires knowledge of the joint distribution of losses for all stations or stages considered in the insurance contracts. With the copula tool at hand, the joint distribution can thus be estimated in a straightforward way. As mentioned in section 2, this implementation includes two separate performances: first, the estimate of the marginal loss distributions for the station i and second, the estimate of the copulas covering dependency structure between losses in different stations. The developed C-vine model is then used to

simulate the joint losses for different scenarios of spatial and temporal diversification. We briefly describe these processes in the following sections.

3.3.1. Marginal distribution estimation

The marginal distributions can be estimated using either a parametric or non-parametric method. However, since the copulas will be estimated parametrically in the later step, this study prefers the non-parametric method for modeling the margins. Therefore, this finally results in a semi-parametric C-vine model that can minimize the bias and inconsistency problems often faced in a fully parametric model if the estimation in one of the parametric processes is mis-specified (Noh et al. 2013).

This study applies the continuous kernel smoothing estimator introduced by Parzen (1962) to model the marginal distributions. The estimated marginal distribution is given by (Duong 2016b):

$$F\left(x\right)=\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right), \tag{15}$$

where h denotes a bandwidth parameter and $K\left(x\right)=\int_{-\infty}^{x}k\left(t\right)dt$ with $k\left(.\right)$ is a symmetric probability density function. The estimation is implemented using the function **kcde** available in the R-package **ks** (Duong 2016a). In this setting, a Gaussian kernel is selected and the plug-in bandwidth developed in Duong (2016b). These estimated marginal distributions are then used to derive pseudo copula data from observed data. These pseudo copula data are an approximately i.i.d sample through the PIT process as mentioned in section 2.1 (Kraus and Czado 2017) and will be employed to estimate the C-vine in the next step.

3.3.2. Copula parameter estimation

With the pseudo copula data obtained in the earlier step, copula parameters are estimated using the maximum-pseudo likelihood method, which is (Kim et al. 2007):

$$\theta = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log\left[c\left(u_{i1},...,u_{in};\theta\right)\right]. \quad (16)$$

With a number of advantages in construction, the C-vine approach does not restrict copula functions for each pair-copula. Therefore, a total of forty copula functions are used for bivariate copula selection (Schepsmeier et al. 2018). The bivariate copula function that yields the lowest Akaike Information Criterion (AIC) is selected for the corresponding pair-copula. The selection is verified with graphical tools such as quantile-quantile and contour plots (not shown here), which has been performed in our previous studies (Nguyen-Huy et al. 2017; Nguyen-Huy et al. 2018b). There are other ways to select the bivariate copulas, which may lead to different results, for example, based on their theoretical properties for better modelling the tail dependence. However, the selection needs to be double-checked with other goodness-of-fit tests to ensure the data are approximately modelled and the model correctly reflects the characteristics of the data. The assumption of using Clayton copula, for example, to focus on the lower tail dependence without verification may result in the misspecification issue since the data used may have a weak dependence in the lower part. Furthermore, one may select the bivariate copulas according to the Bayesian Information Criteria (BIC), however, it should be noted that the penalty for two-parameter copulas (e.g., Student's t, BB1, BB7, etc.) is greater than that based on the AIC value (Schepsmeier et al. 2017).

It is also noted that for an *n*-dimensional multivariate random variables, it is possible to construct *n(n-1)/2* unique C-vine copulas. Therefore, the tree structure is selected using the maximum spanning tree method where the absolute values of the empirical pairwise Kendall's tau as weights (Dissmann et al. 2013). Furthermore, a hypothesis test for the independence of the two variables is performed before bivariate copula selection. The independence copula is selected if the null hypothesis of independence cannot be rejected at significance level of 5%.

3.3.3. C-vine copula-based simulation of insurance losses

This study follows the simulation algorithm described in the studies of Kurowicka and Cooke (2007) and Aas et al. (2009), to draw random samples from a C-vine copula-based model. For the example illustrated in section 2.2, the implement of sampling values $\left(u_1, u_2, u_3, u_4, u_5\right)$ out of a five-dimensional C-vine copula-based model is straightforward and easily calculable. First, five random values $\left(w_1, w_2, w_3, w_4, w_5\right)$ are independently generated such that they are uniformly distributed on [0,1]. These values are subsequently used as probability levels of the conditional functions in sequentially recursive processes. In particular, sampling values are obtained by set:

$$u_1 = w_1, \tag{17}$$

$$u_2 = C_{2|1}^{-1}\left(w_2 \middle| u_1\right), \tag{18}$$

$$u_3 = C_{3|1}^{-1}\left[C_{3|12}^{-1}\left(w_3 \middle| u_1, u_2\right)\right], \tag{19}$$

$$u_4 = C_{4|1}^{-1}\left\{C_{4|12}^{-1}\left[C_{4|123}^{-1}\left(w_4 \middle| u_1, u_2, u_3\right)\right]\right\}, \tag{20}$$

$$u_5 = C_{5|1}^{-1}\left\{C_{5|12}^{-1}\left\{C_{5|123}^{-1}\left[C_{5|1234}^{-1}\left(w_5 \middle| u_1, u_2, u_3, u_4\right)\right]\right\}\right\}. \tag{21}$$

These sampling values are then transformed back to the real scale using the inverse distribution function estimated in section 3.3.1. Finally, the buffer fund and risk-reducing efficiency are acquired following section 3.1.

The BL is derived from the BF that is defined as the 99% quantile of loss distribution that the insurer subtracts to the fair premium. The loss distribution is the result of 10,000 random simulations generated from the estimated margins and the estimated C-vine copula models. The BL is represented for the increasing size of the trading area. In particular, one station or

stage is added sequentially corresponding to each aggregation level. For example, the regional aggregation trading areas include S01, S01 + S02,…, S01 + …+ S05 corresponding to the aggregation levels L1, L2, …, L5 or the seasonal aggregation rainfall in AMJ, AMJ + JAS, AMJ + JAS + OND respective to L1, L2, and L3. Two different trigger levels 5% and 25% quantile of index distribution are used for insurance contracts.

Furthermore, the weights of the insurance contracts are ideally selected based on the cultivated agricultural area in the region and should reflect the potential demand of producers for index-based insurance products (Okhrin et al. 2013). However, the weights w_i are set to one in this study for simplicity. In temporal analysis, the weights w_i are random draws uniformly distributed on (0,1) such that $w_1 + w_2 + w_3 = 1$. A total of 1,000 scenarios of weights is generated randomly.

4. Results and discussion

In this section, we show the spatial and temporal dependencies of weather events and the efficiency of the corresponding risk diversification strategies at regional, national and temporal levels in Australia. The analysis is supported by inspecting the differences in probabilistic properties of rainfall index in different locations (*i.e.*, weather stations) and times (*i.e.*, growing states). The results of the BL and risk-reducing efficiency are reported subsequently.

4.1. Regional study

4.1.1. Spatial interdependencies of weather events

Table 1 and Fig. 4 provide the probabilistic information of the cumulative rainfalls recorded at sixteen weather stations across the Australia's grain belt. The average cumulative rainfalls at all weather stations are approximated together during AMJ period except S01 and S02 but the stations in the west have the highest variability in general. The average cumulative rainfalls during JAS significantly vary between weather stations. In the OND period, weather stations

in the east receive the highest rainfall with a considerable difference between locations. The rainfall and variability in the southern and south-eastern regions during this time is lower compared to the west. Evidently, the western region has the driest conditions with the lowest rainfall variability. Furthermore, the cumulative rainfalls across stations at different times generally have right skewed distributions reflecting the asymmetry of the index.

The spatial interdependencies among of the cumulative rainfalls at different weather stations are explored further through a multidimensional Kruskal-Shephard scaling plot (Fig. 5). This method embeds the CRIs at different stations in the plane using the dissimilarity measured by $\left(1-\hat{\tau}\right)$ where $\hat{\tau}$ is the empirical pairwise Kendall's τ values for all pairs (Brechmann et al. 2013). This means that the closer two stations are to each other, the stronger is the dependence on their rainfalls. The multidimensional scaling reveals that there is significant geographical clustering existence of rainfalls between stations. The stations in the west (S01 – 05) can be found on the left of Fig. 5, stations in the south and south-east (S06 – 11) in the upper right corner, and stations in the east (S12 – 16) in the lower right corner. It is also worth pointing out that for the regional scale, the cumulative rainfalls in the west are more divergent in AMJ and OND periods than in JAS. The cumulative rainfalls in the south and south-east have a dispersive pattern in OND while the cumulative rainfalls in the east scatter in all three different times. These initial analyses may reveal the first impression that, for example, geographical aggregation of weather risks are more effective for weather stations in the eastern region.

<Figure 5>

Finally, we inspect the interdependencies in the lower tails of weather stations in each region. In order to do that, the joint distributions of all stations in each region are modeled through C-vine copulas. Then different climatic events are assumed to occur at one station to assess how these extreme events correlate with other stations in that region. This means three quantile

levels 25%, 15% and 5% of CRI distribution are set to the selected station implying the moderate, severe, and extreme drought conditions, respectively. The drought spreads are assessed through the predicted median values of other stations conditioned on the selected station in the joint distributions modeled by C-vine copulas. It can be seen clearly in Fig. 6 that if drought events occur at S05 in AMJ or at S01 in JAS, they will have a systemic effect on other stations in the west or south and south-east regions, respectively. By contrast, the drought events occurring at a station (e.g., S05 in the west or S16 in the east) in OND are more local and do not spread out to other stations. This evidence agrees with the analyses above that provide an initial assessment of geographical diversifying the weather systemic risk.

<Figure 6>

4.1.2. Geographical diversification effect

In the spatial aggregation strategy for each region, each weather station is selected respectively for L0 (starting point) and the remainders are randomly added to each aggregation level. This is implemented by creating a matrix of $(n-1)!$ scenarios where n is the number of stations. For the case of five stations in the west as an example, S01 is selected for the BL calculation of a single insurance contract. S02 – 05 are aggregated arbitrarily into L2 – 5 through 24 scenarios.

The main results of regional aggregation in the west, south and south-east, and east of Australia are represented in Figs. 7, 8, and 9, respectively. These figures display the average BL and risk-reducing effect in different trigger and aggregation levels for regions under consideration. The stochastic fluctuations are indicated by the shading area with the maximum and minimum bounds. Clearly, significant differences can be observed between locations. According to Xu et al. (2010), the findings are expected whenever analyzing the influences of aggregating trading areas for weather-related insurance.

<Figure 7>

<Figure 8>

<Figure 9>

Generally, the results indicate that the BL may decline or increase depending on the weather station selected for the initial single insurance contract (i.e., L0 of aggregation level). However, it is noted that the slope of the decrease and increase is rather small if the trading area becomes larger (i.e., the number of stations increases) or the weather events become more extreme (i.e., at 5% quantile level). These findings contrast to the initial expectation of diversification efficiency but agree with the results of the study by Xu et al. (2010). This phenomena can be explained by the fact that weather indices pooled in the trading areas are heterogeneous, which is, the index distributions may be dependent or not identically distributed. In addition, it is clear that the BL greatly relies on the trigger level. The BL values are considerably smaller for an insurance contract with a strike level of 5% compared to one with a strike level of 25%.

To achieve a more accurate assessment of the diversification efficiency, the BL of the joint insurance of all stations is compared to the average BL associated with separate insurance for each station. In general, the findings (see Figs. 6, 7 and 8) indicate that weather-related risks can be reduced through increasing the size of the trading area. Furthermore, the risk pooling strategy is more effective in the case of extreme weather events (*i.e.*, 5% quantile level) compared to moderate weather events (*i.e.*, 25% quantile level). The average diversification effects for entire insurance area in three aggregation levels: regional, national and temporal are summaried in Table 2.

<Table 2>

According to Xu et al. (2010), it is expected that the BL in the case of normal i.i.d losses drops with $1/\sqrt{n}$, which means that a reduction of the BL would be approximately 45% if n = 5.

However, the reduction is smaller in the present analysis of insurance contract indicating the existence of systemic weather risk, except for temporal diversification (Table 2).

For the western region, while the risk-reducing effects through the extent of aggregation levels are significant for the period OND, a contrast pattern can be observed during AMJ and JAS (Fig. 7). However, the diversification strategy is still potentially beneficial for reducing the impact of systemic weather risk since the DE (diversification efficiency) values are below one for the AMJ and JAS cases up to L4 (Table 2). For the southern and south-eastern areas, there is evidence supporting the usefulness of risk diversification during AMJ and OND but growing the trading area does not increase the risk-reducing effect in the period JAS for both strike levels (Fig. 8). It is interesting that the efficiency of the risk pooling strategy applied in the eastern region is directly proportional to the increase of weather station numbers in all of periods considered. It is important to highlight that these results are consistent with the initial analyses of spatial interdependencies between weather stations mentioned above.

We also represent the C-vine models developed for three different cases, as in Table 3. The first model is the case when the risk-reducing effect is indirectly proportional to the increasing of aggregation levels (see Fig. 7, AMJ, and L0 = S02). The second is when there is not much change in the risk-reducing effect associated with the increase of aggregation levels (see Fig. 8, JAS, L0 = S08). The final model is the case where the risk-reducing effect is directly proportional to the extent of trading area. It can be clearly seen that there more dependencies in the lower tails between pair weather stations indicated by lower coefficients derived from the survival Gumbel copula for the first and second cases than the third one. This may imply that the risk pooling is more effective for the cases having less interdependencies in the tails.

<Table 3>

4.2. National study

4.2.1. Spatial interdependencies of weather events

The same procedure mentioned in the regional study is applied to analyze and evaluate the spatial interdependencies and the risk-reducing efficiency of the risk pooling strategy on the national scale. This is implemented by modeling the joint distribution of cumulative rainfalls in all sixteen stations through the C-vine copula method. Following the same procedure before, the three drought levels are arbitrarily set to S16 in the eastern region and the results for different times are displayed in Fig. 10. Generally speaking, if drought events occur in the eastern region, it is likely they will affect much of the eastern, south-eastern and southern regions. However, these drought events may have less impact on the western region implying that the systemic risk may be reduced efficiently by pooling with stations in the west. This finding is consistent with the analysis of multidimensional Kruskal-Shephard scaling shown in Fig. 5 where stations in the west are clustered separately to others in the east and south and south-east.

<Figure 10>

4.2.2. Geographical diversification effect

For the illustrative purpose of risk aggregation at national scale, we randomly select one station in each region resulting in a total of four stations. In general, the BL declines when all stations are taken into account except for the case when S01 is selected for L0 in OND period. The considerable low values observed in the bar charts of risk-reducing effect in AMJ and OND may be explained by the analyses of drought spread characteristic mentioned above. This means that the risk pooling strategy is more efficient when initial stations include the stations (e.g., in the west) that has a high independent relationship with them.

<Figure 11>

4.3. Temporal study

4.3.1. Temporal interdependencies of weather events

An alternative strategy that may reduce loadings of weather systemic risk is a time diversification approach. The foundation is similar to spatial diversification, but risks are pooled over time instead of across locations (Odening and Shen 2014). In the context of insurance, these products are provided as multi-year or long-term contracts. They are based on the fact that if the losses in different years occur independently, it has the potential to diminish the volatility of average yearly insurance indemnities. The multi-year insurance contract has been implemented in the studies undertaken by Kleindorfer et al. (2012), Botzen et al. (2013) or Osipenko et al. (2015).

This study proposes to diversify risks over different periods corresponding to the crop growing season. It is motivated by the fact that we recognized that the cumulative rainfalls in different seasons are not perfectly correlated when analyzing the spatial interdependencies between stations. For example, the probabilistic information derived from Table 1 and Fig. 4 indicates that there much differences in the rainfall pattern between AMJ, JAS and OND periods. For example, the western Australia receives more rainfall during sowing (AMJ) and vegetation (JAS) stages and extreme low during harvest months (OND) in the end of year. A similar rainfall pattern can be observed in the weather stations in the south where rainfall peaks in JAS before decreasing in OND. By contrast, the eastern part has the highest rainfall during the three-month period at the end of the year and the lowest rainfall during JAS. These evidences mean that it is possible to reduce widespread systemic losses by a multi-stage insurance contract at each station.

We therefore use the same procedure used for spatial diversification to analyze and evaluate the effect of risk reduction through the multi-stage risk pooling strategy. The examination of seasonal systemic losses are shown in Fig. 12 where a time period (i.e., AMJ or OND) is set

with different climatic events and others are conditional on the selected time. In general, the insufficient rainfall during the sowing stage (i.e., AMJ) does not mean that drought conditions will occur subsequently in other stages.

<Figure 12>

4.3.2. Time diversification effect

The findings of risk reduction for the insurance contracts based on seasonal aggregation are somewhat different to those with a spatial context (Fig. 13). For all stations, the BL is significantly reduced for the aggregation level one (L1) while it does not cause much change in the L2. It is interesting that the average value of risk-reducing effect ranges from 25% to 33% if the insurance contracts are designed within the 5% and 25% quantile of the index distributions, respectively. The results agree with the initial analyses of temporal interdependencies between the cropping stages above. In addition, the BL and risk-reducing effect are found to be greater compared to the spatial risk pooling strategy.

<Figure 13>

5. Further discussion

It is not surprising that there are different results on the effect of index-based insurance products and diversification strategies (Okhrin et al. 2013; Xu et al. 2010). At the worst scenario, if the weather events are perfectly correlated, there is no diversification benefit. Mahul (1999) indicated that index-based insurance products cannot solve the problem of risk pooling and therefore cannot diversify the systemic risk which occurs when a natural risk strikes simultaneously among a large number of farmers. Xu et al. (2010) showed that, as one example, systemic weather risk cannot be regionally diversified for Germany. By contrast, Okhrin et al. (2013) found that the spatial diversification effect is significant in China where BLs drop by more than 50% if insurance losses are aggregated over several provinces. However, the spatial

manners of systemic weather risk vary worldwide because of differences of geographical topography and climatic conditions (Odening & Shen 2014). The results found in this study demonstrate that both spatial and temporal aggregation strategies potentially reduce the weather systemic risk in Australia where temporal diversification effects range from 25% to 30% depending on the strike level.

However, the present study demonstrates that some challenges still remain that require the reading of the results with care. First, the study used the historical rainfall data observed in meteorological stations; the results thus reflect the dependency structure of weather events that occurred in the past. Second, while the arbitrary choice of 99% quantile of loss distribution is quite common for the calculation of VaR in the financial field, this selection greatly influences the levels of the BF in this application (Xu et al. 2010). Third, the study relies on several assumptions that are similar to the study of Okhrin et al. (2013) and Xu et al. (2010). That means only a single period was considered while the product diversification of insurer and equity reverses cumulated annually with premium surpluses were not taken into account. Finally, even if weather conditions are the same, regional differences in soil quality or cultivated technology, for instance, may result in dissimilarities in crop yield.

Furthermore, risk diversification by pooling individual risks may not be effective for skewed distributions as in the case of normal distributions (Wang 2000). Because VaR is a quantile-based measure, it is subadditive for elliptical distributions. In the context of finance, subadditivity is one of the four axioms characterizing coherent measures of risk including the exhibition of sub-additivity (or diversification), translation invariance, positive homogeneity, and monotonicity. VaR satisfies subadditive conditions if the risk of holding two assets x_1 and x_2 simultaneously is less than or equal to the sum of their single risks given as $\rho(x_1 + x_2) \leq \rho(x_1) + \rho(x_2)$ with $\rho(x)$ denoting some risk measure. However, VaR lacks

subadditivity in cases of non-elliptical distributions. In addition, VaR is also non-subadditive if the loss distributions of assets are smooth and symmetric, but their dependency structure is asymmetric, and when underlying risk factors are independent but heavy-tailed.

Although the systemic weather risk can be mitigated by regional diversification, it is nevertheless still high. We highlight that much more research is needed to evaluate the promise of weather index insurance. For example, the use of supplementary tools for reducing systemic weather risk, such as securitization, are recommended by Odening and Shen (2014). A comparison of the efficiency of regional diversification and securitization via weather bonds can be found in Shen and Odening (2013). Furthermore, future research may use a coherent risk-measure method, which is Conditional Value-at-Risk (CVaR), as an alternative instrument for calculating the losses (Rockafellar and Uryasev 2002). Finally, the approach that combines a vegetation index derived from satellite–based data with an in situ assessment of crop damages may be a subject for potential ongoing research.

6. Conclusion

The study assesses the joint losses of a hypothetical index-based insurance at different aggregation scenarios using the C-vine copula-based model. Extensive analyses of spatial and temporal interdependencies between weather events at different locations and times have been made. The C-vine is able to capture different dependency structures between weather events where its construction is particularly useful for testing widespread systemic losses. The results reveal that either spatial or temporal diversification strategies may potentially reduce the systemic weather risk in Australia. It can be seen that the dependencies between insurance indemnities at different locations and times becomes smaller when reducing the strike level. Furthermore, the cumulative rainfalls in different seasons are found to be not perfectly

correlated, therefore the time diversification is more efficient for systemic risks compared to the case of spatial pooling.

Insurance products have been crucial in the agricultural sector as one of the formal risk-mitigation strategies. In particular, weather index-based insurance has been considered a valuable alternative to traditional crop insurance. While the index-based insurance product may not be a direct insurance for farmers because of its high basis risk, the analysis of weather risks is an interesting topic to insurers for several reasons. First, covariate yield risk is mainly driven by weather conditions. Second, weather derivatives could be useful for transferring systemic risk from insurers to reinsurers or to the capital market (Xu et al. 2010). This application helps solve the questions to what extent weather-related risk exposure at different locations and times can be spatially or temporally diversified by increasing the trading area or season of the contracts. The study therefore provides an efficient tool for risk management and supports pricing of weather index-based insurance products.

Acknowledgement

The project was financed by the University of Southern Queensland Post Graduate Research Scholarship (USQPRS 2015-2018); School of Agricultural, Computational and Environmental Sciences and the Drought and Climate Adaptation (DCAP) Project (Producing Enhanced Crop Insurance Systems and Associated Financial Decision Support Tools). The authors would like to acknowledge constructive comments from the reviewers.

References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence Insurance: Mathematics and economics 44:182-198 doi:https://doi.org/10.1016/j.insmatheco.2007.02.001

AghaKouchak A, Cheng L, Mazdiyasni O, Farahmand A (2014) Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought Geophysical Research Letters 41:8847-8852 doi:https://doi.org/10.1002/2014GL062308

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the temperature record map of Europe Science 332:220-224

Bedford T, Cooke RM (2001) Probability density decomposition for conditionally dependent random variables modeled by vines Annals of Mathematics and Artificial intelligence 32:245-268 doi:https://doi.org/10.1023/A:1016725902970

Bedford T, Cooke RM (2002) Vines: A new graphical model for dependent random variables Annals of Statistics:1031-1068 doi:https://doi.org/10.1214/aos/1031689016

Botzen WW, de Boer J, Terpstra T (2013) Framing of risk and preferences for annual and multi-year flood insurance Journal of economic psychology 39:357-375 doi:https://doi.org/10.1016/j.joep.2013.05.007

Brechmann EC, Hendrich K, Czado C (2013) Conditional copula simulation for systemic risk stress testing Insurance: Mathematics and Economics 53:722-732 doi:https://doi.org/10.1016/j.insmatheco.2013.09.009

Carreau J, Bouvier C (2016) Multivariate density model comparison for multi-site flood-risk rainfall in the French Mediterranean area Stochastic Environmental Research and Risk Assessment 30:1591-1612

Coumou D, Rahmstorf S (2012) A decade of weather extremes Nature climate change 2:491 doi:https://doi.org/10.1038/nclimate1452

Dissmann J, Brechmann EC, Czado C, Kurowicka D (2013) Selecting and estimating regular vine copulae and application to financial returns Computational Statistics & Data Analysis 59:52-69 doi:https://doi.org/10.1016/j.csda.2012.08.010

Duncan J, Myers RJ (2000) Crop insurance under catastrophic risk American Journal of Agricultural Economics 82:842-855 doi:https://doi.org/10.1111/0002-9092.00085

Duong T (2016a) ks: Kernel Smoothing, r package version 1.10. 4.

Duong T (2016b) Non-parametric smoothed estimation of multivariate cumulative distribution and survival functions, and receiver operating characteristic curves Journal of the Korean Statistical Society 45:33-50 doi:https://doi.org/10.1016/j.jkss.2015.06.002

Embrechts P, McNeil A, Straumann D (2002) Correlation and dependence in risk management: properties and pitfalls Risk management: value at risk and beyond 176223

Frey R, McNeil AJ (2003) Dependent defaults in models of portfolio credit risk Journal of Risk 6:59-92 doi:https://doi.org/10.21314/JOR.2003.089

Frey R, McNeil AJ, Nyfeler M (2001) Copulas and credit models Risk 10

Glauber JW, Collins KJ, Barry PJ (2002) Crop insurance, disaster assistance, and the role of the federal government in providing catastrophic risk protection Agricultural Finance Review 62:81-101 doi:https://doi.org/10.1108/00214900280001131

Goodwin BK (2001) Problems with market insurance in agriculture American Journal of Agricultural Economics 83:643-649 doi:https://doi.org/10.1111/0002-9092.00184

Holly Wang H, Zhang H (2003) On the possibility of a private crop insurance market: A spatial statistics approach Journal of Risk and Insurance 70:111-124 doi:https://doi.org/10.1111/1539-6975.00051

Kim G, Silvapulle MJ, Silvapulle P (2007) Comparison of semiparametric and parametric methods for estimating copulas Computational Statistics & Data Analysis 51:2836-2850 doi:https://doi.org/10.1016/j.csda.2006.10.009

Kleindorfer PR, Kunreuther H, Ou-Yang C (2012) Single-year and multi-year insurance policies in a competitive market Journal of Risk and Uncertainty 45:51-78 doi:https://doi.org/10.1007/s11166-012-9148-2

Kraus D, Czado C (2017) D-vine copula based quantile regression Computational Statistics & Data Analysis 110:1-18 doi:https://doi.org/10.1016/j.csda.2016.12.009

Kurowicka D (2005) Distribution-free continuous bayesian belief Modern statistical and mathematical methods in reliability 10:309

Kurowicka D, Cooke RM (2007) Sampling algorithms for generating joint uniform distributions using the vine-copula method Computational statistics & data analysis 51:2889-2906 doi:https://doi.org/10.1016/j.csda.2006.11.043

Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production Nature 529:84 doi:https://doi.org/10.1038/nature16467

Mahul O (1999) Optimum area yield crop insurance American Journal of Agricultural Economics 81:75-82 doi:https://doi.org/10.2307/1244451

Martin SW, Barnett BJ, Coble KH (2001) Developing and pricing precipitation insurance Journal of Agricultural and Resource Economics:261-274

McNeil A, Frey R, Paul E (2005) Quantitative risk management: Concepts, techniques and tools. Princeton University Press, Princeton, New Jersey

Miranda MJ, Glauber JW (1997) Systemic risk, reinsurance, and the failure of crop insurance markets American Journal of Agricultural Economics 79:206-215 doi:https://doi.org/10.2307/1243954

Musafer GN, Thompson MH (2017) Non-linear optimal multivariate spatial design using spatial vine copulas Stochastic environmental research and risk assessment 31:551-570

Musshoff O, Odening M, Xu W (2011) Management of climate risks in agriculture–will weather derivatives permeate? Applied economics 43:1067-1077 doi:https://doi.org/10.1080/00036840802600210

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 191:153-172 doi:https://doi.org/10.1016/j.agwat.2017.06.010

Nguyen-Huy T, Deo RC, Mushtaq S, An-Vo D-A, Khan S (2018a) 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 98:65-81 doi:https://doi.org/10.1016/j.eja.2018.05.006

Nguyen-Huy T, Deo RC, Mushtaq S, Kath J, Khan S (2018b) Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management Weather and climate extremes 21:76-89

Noh H, Ghouch AE, Bouezmarni T (2013) Copula-based regression estimation and inference Journal of the American Statistical Association 108:676-688 doi:https://doi.org/10.1080/01621459.2013.783842

Odening M, Mußhoff O, Xu W (2007) Analysis of rainfall derivatives using daily precipitation models: Opportunities and pitfalls Agricultural Finance Review 67:135-156 doi:https://doi.org/10.1108/00214660780001202

Odening M, Shen Z (2014) Challenges of insuring weather risk in agriculture Agricultural Finance Review 74:188-199 doi:https://doi.org/10.1108/AFR-11-2013-0039

Okhrin O, Odening M, Xu W (2013) Systemic weather risk and crop insurance: the case of China Journal of Risk and Insurance 80:351-372 doi:https://doi.org/10.1111/j.1539-6975.2012.01476.x

Osipenko M, Shen Z, Odening M (2015) Is there a demand for multi-year crop insurance? Agricultural Finance Review 75:92-102 doi:https://doi.org/10.1108/AFR-12-2014-0043

Parzen E (1962) On estimation of a probability density function and mode The annals of mathematical statistics 33:1065-1076 doi:https://doi.org/10.1214/aoms/1177704472

Pham MT, Vernieuwe H, De Baets B, Willems P, Verhoest N (2016) Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas Stochastic Environmental Research and Risk Assessment 30:2197-2214

Reddy MJ, Singh VP (2014) Multivariate modeling of droughts using copulas and meta-heuristic methods Stochastic environmental research and risk assessment 28:475-489

Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions Journal of banking & finance 26:1443-1471 doi:https://doi.org/10.1016/S0378-4266(02)00271-6

Rosenblatt M (1952) Remarks on a multivariate transformation The annals of mathematical statistics 23:470-472 doi:https://doi.org/10.1214/aoms/1177729394

Sak H, Yang G, Li B, Li W (2017) A copula-based model for air pollution portfolio risk and its efficient simulation Stochastic Environmental Research and Risk Assessment 31:2607-2616

Schepsmeier U et al. (2017) Package 'VineCopula'

Schepsmeier U et al. (2018) Package 'VineCopula'

Schoelzel C, Friederichs P (2008) Multivariate non-normally distributed random variables in climate research–introduction to the copula approach Nonlinear Processes in Geophysics 15:761-772 doi:https://doi.org/10.5194/npg-15-761-2008

Serinaldi F (2009) Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective Stochastic environmental research and risk assessment 23:677-693

Skees JR, Barnett BJ (1999) Conceptual and practical considerations for sharing catastrophic/systemic risks Review of Agricultural Economics 21:424-441 doi:https://doi.org/10.2307/1349889

Skees JR, Hartell J, Murphy AG (2007) Using index-based risk transfer products to facilitate micro lending in Peru and Vietnam American Journal of Agricultural Economics 89:1255-1261

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. Université Paris 8,

Song S, Singh VP (2010) Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data Stochastic Environmental Research and Risk Assessment 24:425-444

Tankov P (2011) Improved Fréchet bounds and model-free pricing of multi-asset options Journal of Applied Probability 48:389-403

Van Den Goorbergh RW, Genest C, Werker BJ (2005) Bivariate option pricing using dynamic copula models Insurance: Mathematics and Economics 37:101-114 doi:https://doi.org/10.1016/j.insmatheco.2005.01.008

Vedenov D Application of copulas to estimation of joint crop yield distributions. In: American Agricultural Economics Association Annual Meeting, Orlando, FL, 2008. pp 27-29

Vedenov DV, Barnett BJ (2004) Efficiency of weather derivatives as primary crop insurance instruments Journal of Agricultural and Resource Economics:387-403

Wang SS (2000) A class of distortion operators for pricing financial and insurance risks Journal of risk and insurance:15-36 doi:https://doi.org/10.2307/253675

Woodard JD, Garcia P (2008) Basis risk and weather hedging effectiveness Agricultural Finance Review 68:99-117 doi:https://doi.org/10.1108/00214660880001221

Xu W, Filler G, Odening M, Okhrin O (2010) On the systemic nature of weather risk Agricultural Finance Review 70:267-284 doi:https://doi.org/10.1108/00021461011065283

Xu Y, Huang G, Fan Y (2017) Multivariate flood risk analysis for Wei River Stochastic environmental research and risk assessment 31:225-242

Zhang L, Yang B, Guo A, Huang D, Huo Z Multivariate probabilistic estimates of heat stress for rice across China Stochastic Environmental Research and Risk Assessment:1-14

Zhang Q, Xiao M, Singh VP, Chen X (2013) Copula-based risk evaluation of hydrological droughts in the East River basin, China Stochastic environmental research and risk assessment 27:1397-1406

Zhu Y, Ghosh SK, Goodwin BK Modeling Dependence in the Design of Whole Farm Insurance Contract,| A Copula-Based Model Approach. In: Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, July, 2008. pp 27-29

Click here to access/download;colour figure;Nguyen-Huy et al 2018 Systemic Risk Figures_revised.docx

List of Figures

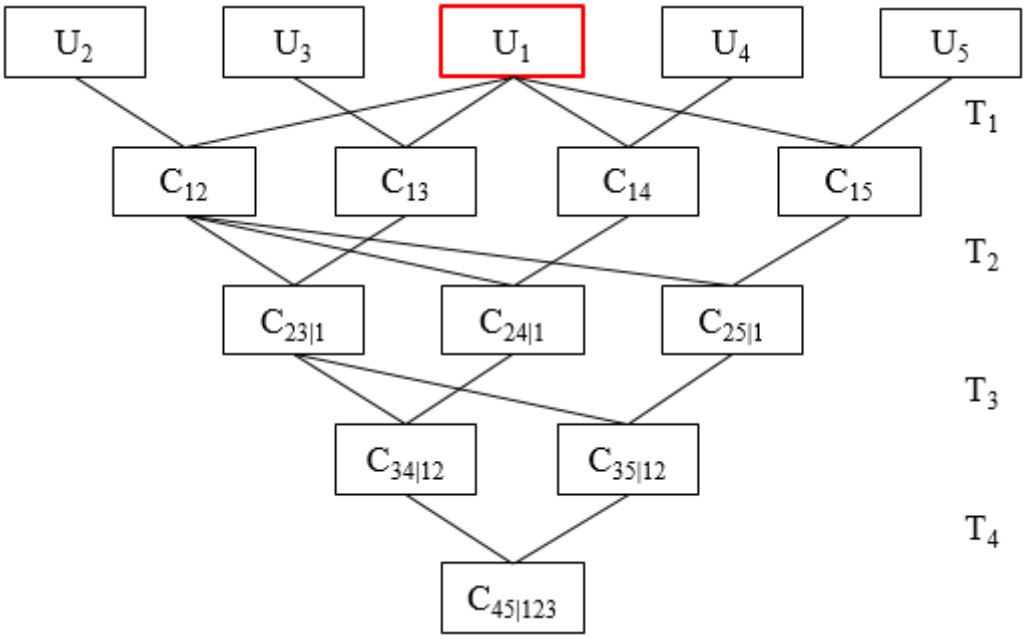


Figure 1. An example of the five-dimensional canonical (C)-vine copula with variable one (U_1, *e.g.*, cumulative rainfall index at station 1) as the root node (indicated in red color) and the pairwise copulas (C) in each tree (T). The numbers, for example, 23|1 denote variable U_2 and U_3 conditioned U_1 modeled by $C_{23|1}$.

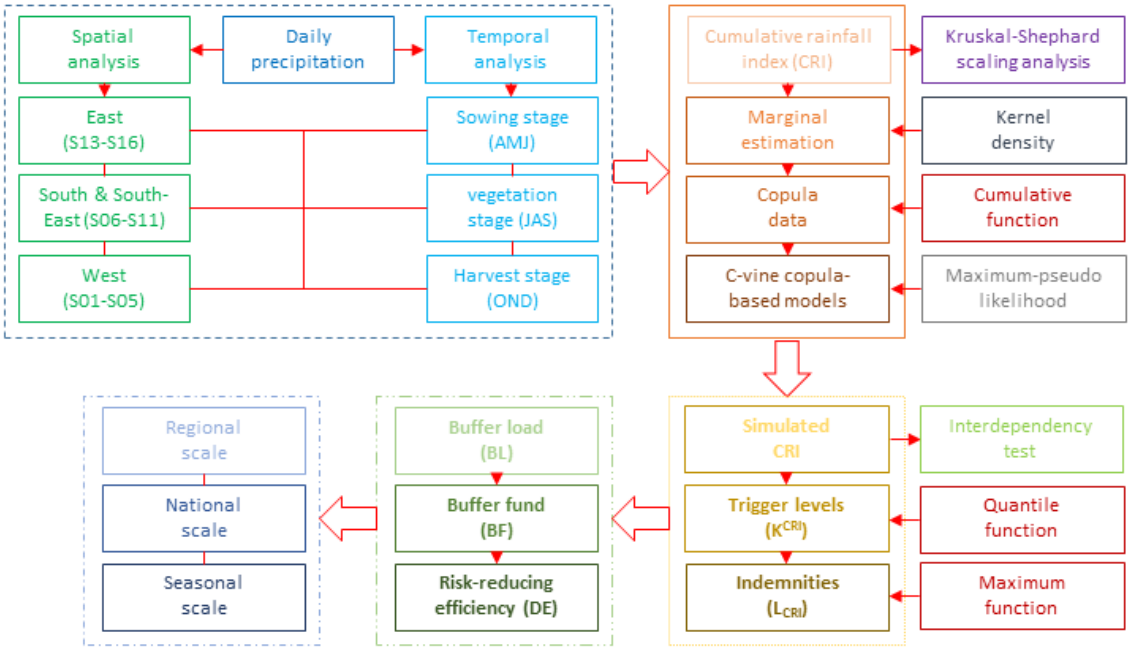


Figure 2. Study flowchart describes about data used and main steps of tests and copula-based method.

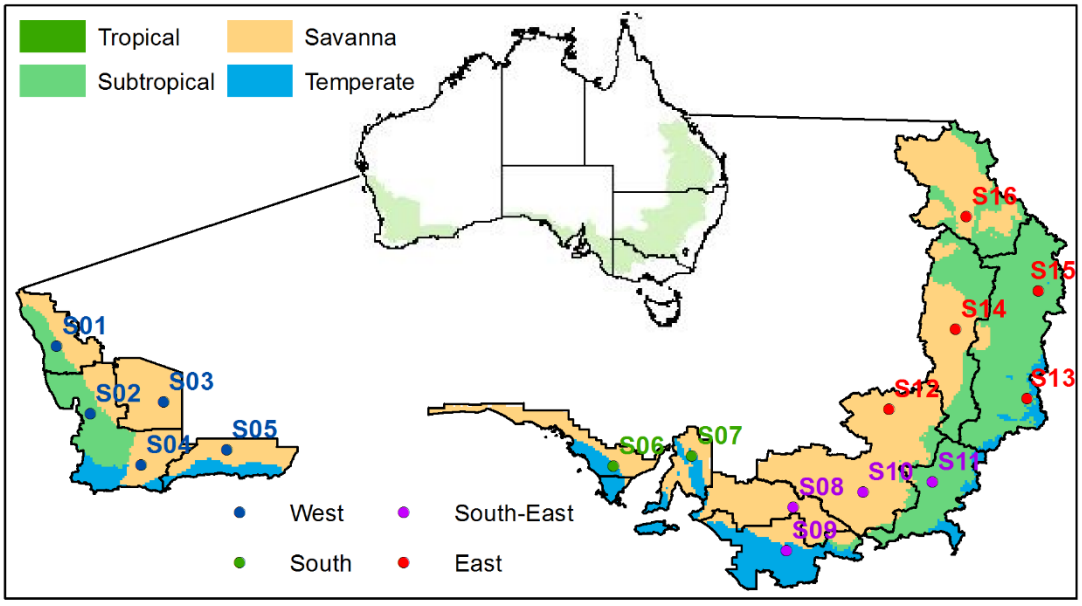


Figure 3. Map of selected weather stations across Australia's wheat belt spanning different climatic conditions.

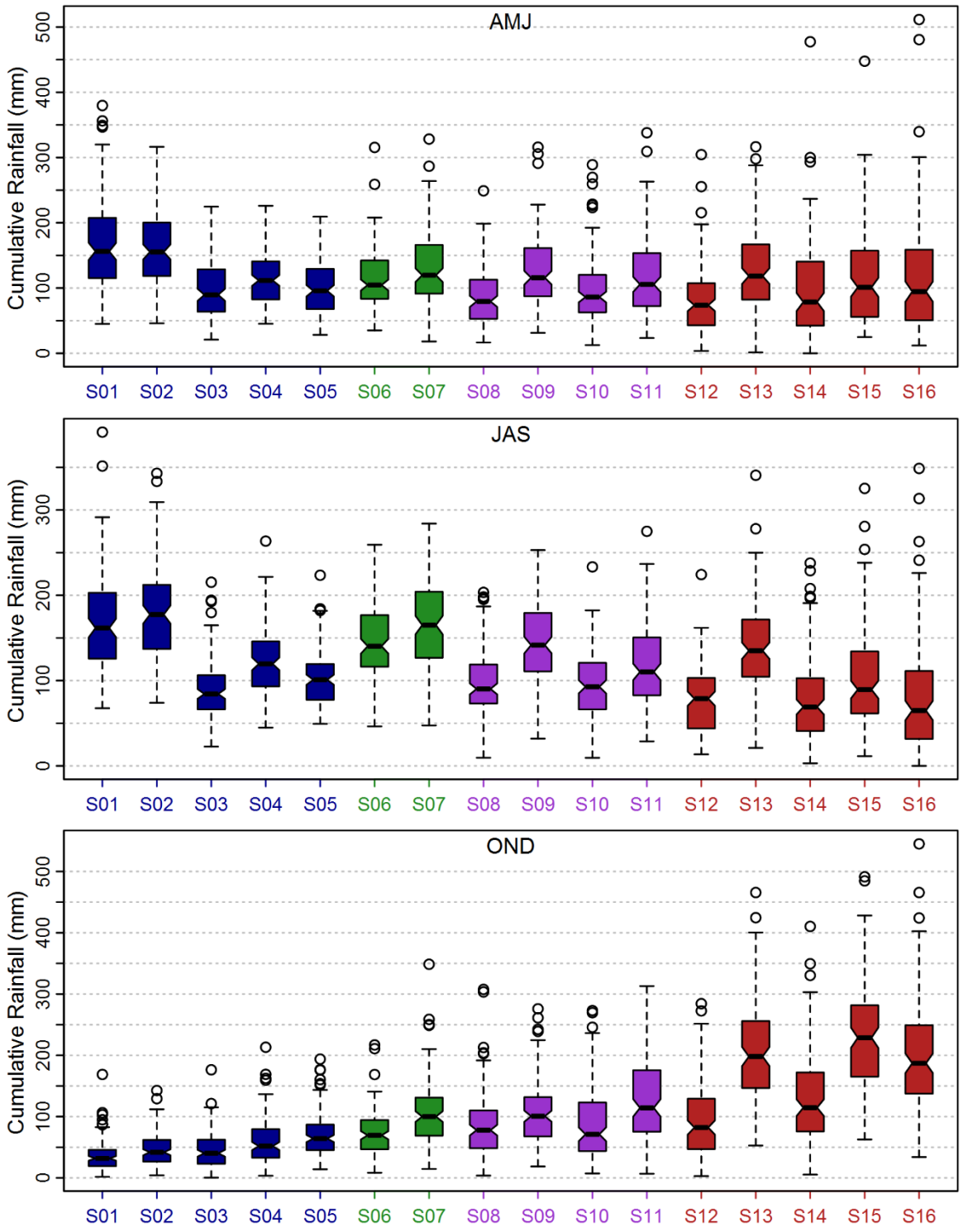


Figure 4. Boxplot of the cumulative rainfall index (CRI) in different weather stations and separately for three different seasons. Colors matched to Fig. 2.

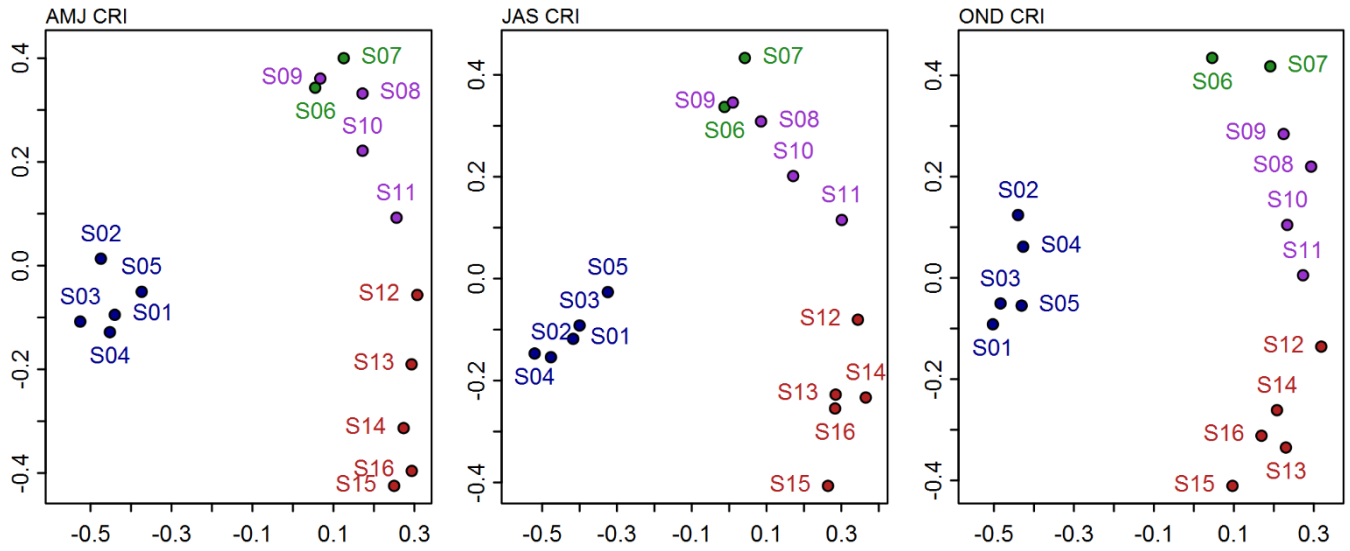


Figure 5. Multidimensional Kruskal-Shephard scaling (non-dimensional unit) (Brechmann et al., 2013) of the weather stations using dissimilarity measure calculated by empirical pairwise Kendall's τ of cumulative rainfall index (CRI). Colors matched to Fig. 2.

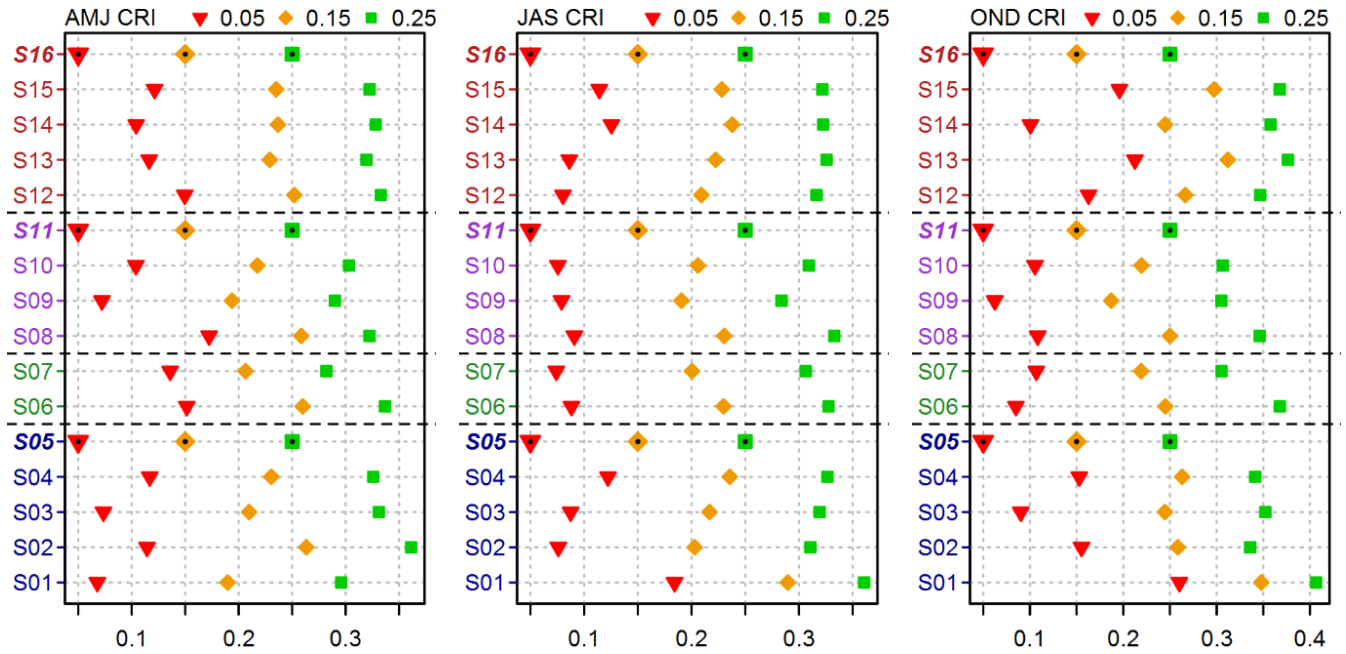


Figure 6. Tests of interdependencies between weather stations using cumulative rainfall index (CRI) in different growing stages of wheat crop. The predicted median values (50%-quantile) for the three extreme events of each station conditioned on S05, S11 and S16 given extreme (5%-quantile), severe (15%-quantile), and moderate (25%-quantile) drought. The conditioning stations marked with bigger symbol overlaid by small points.

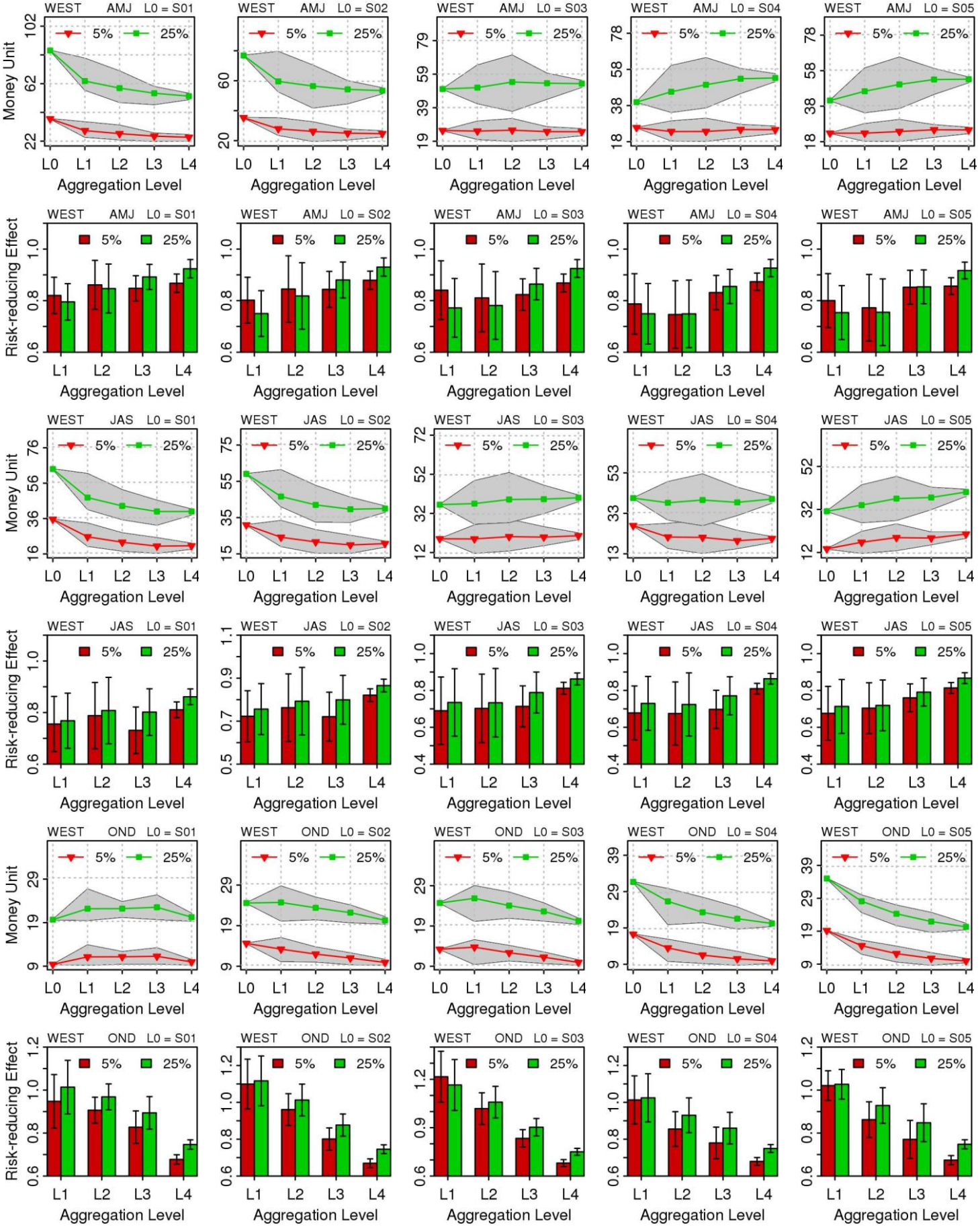
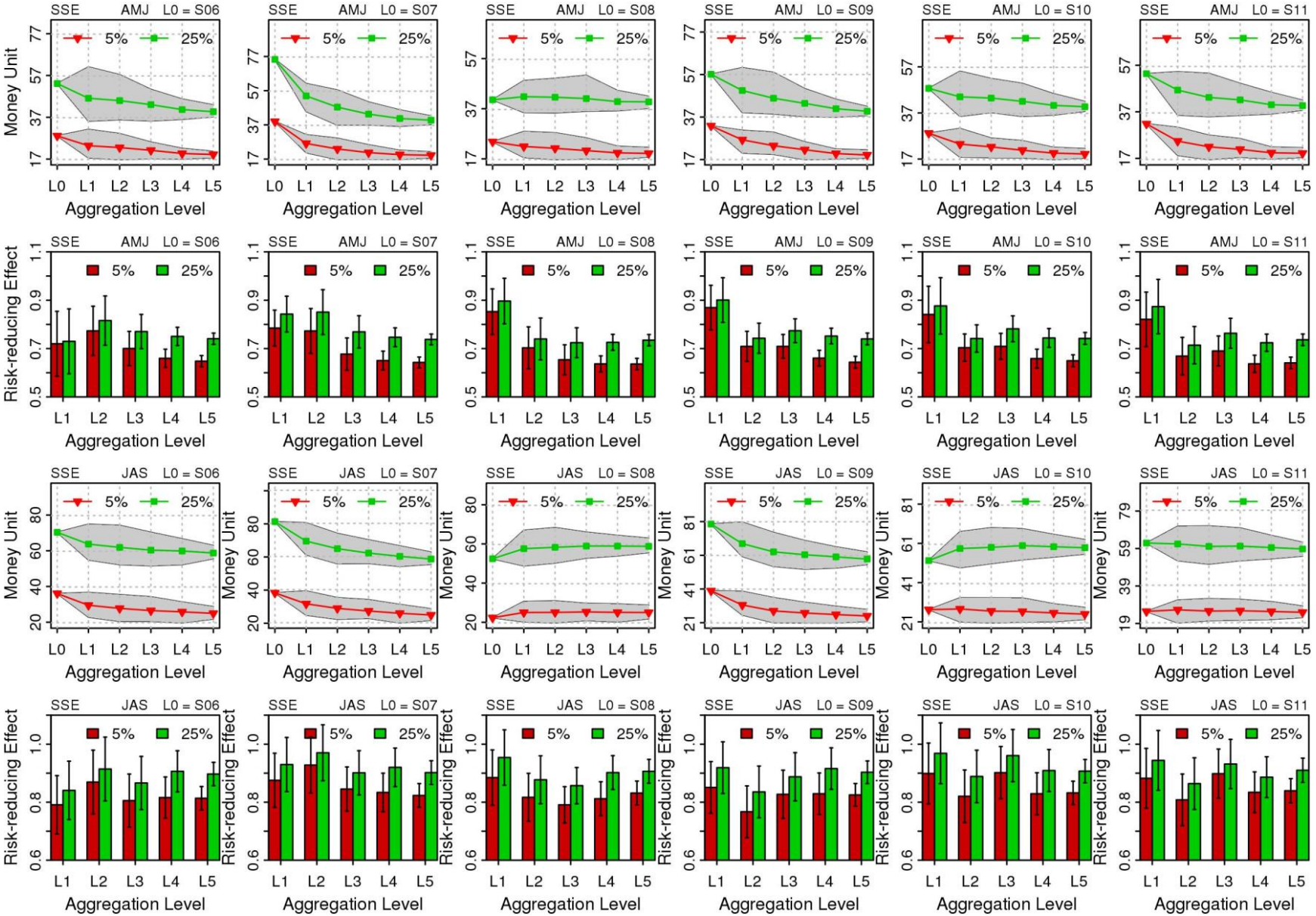


Figure 7. Buffer loads at 99% quantile of loss distributions with different strike (5% and 25%) and aggregation levels (L1 – 4) using the cumulative rainfall index (CRI) and the corresponding risk-reducing effects for five weather stations in western Australia.


SSE AMJ L0 = S06
SSE AMJ L0 = S07
SSE AMJ L0 = S08
SSE AMJ L0 = S09
SSE AMJ L0 = S10
SSE AMJ L0 = S11
SSE JAS L0 = S06
SSE JAS L0 = S07
SSE JAS L0 = S08
SSE JAS L0 = S09
SSE JAS L0 = S10
SSE JAS L0 = S11
5%
25%
Money Unit
Risk-reducing Effect
Aggregation Level
L0 L1 L2 L3 L4 L5
L1 L2 L3 L4 L5

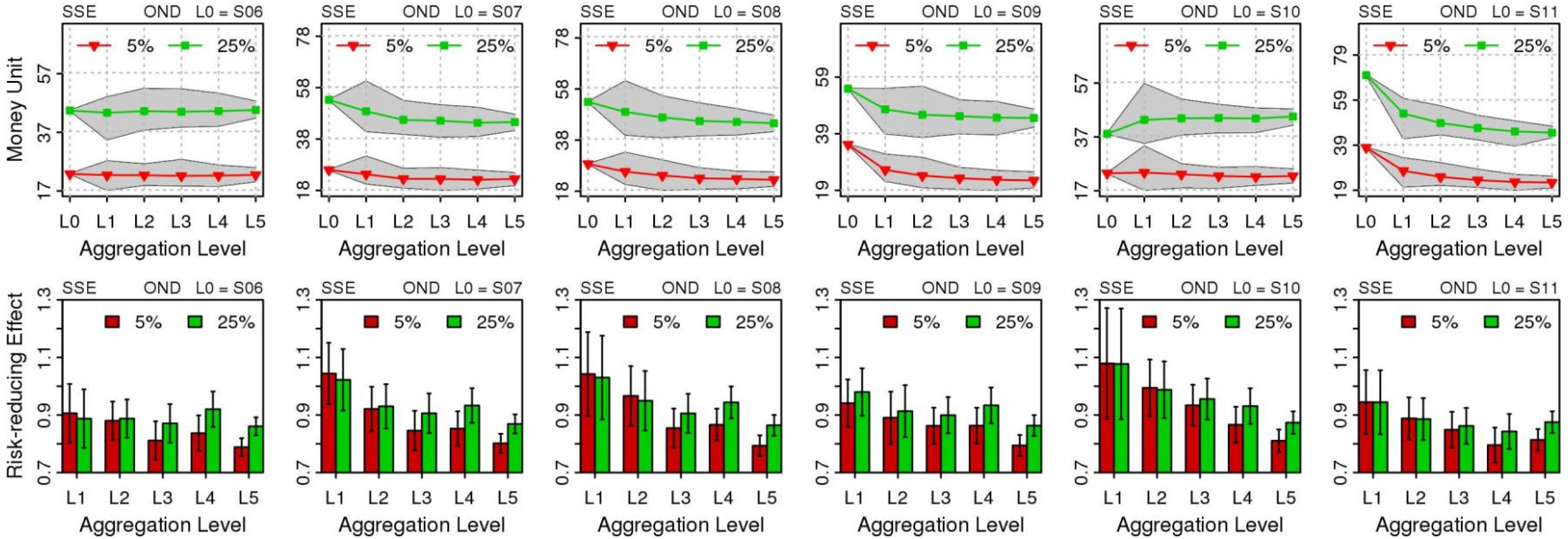


Figure 8. Similar to Fig. 7 but for the south and south-eastern Australia.

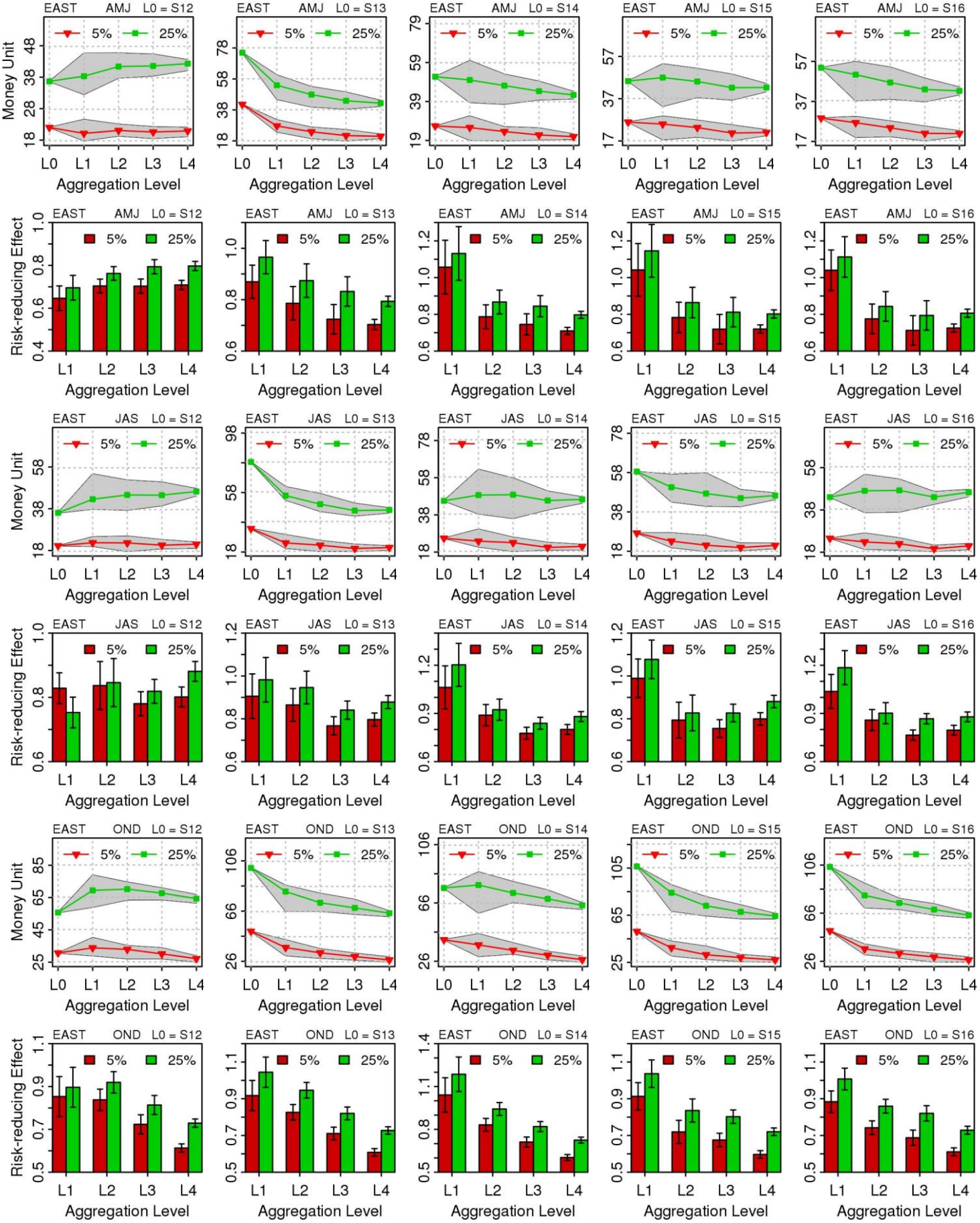


Figure 9. Similar to Fig. 7 but for the eastern Australia.

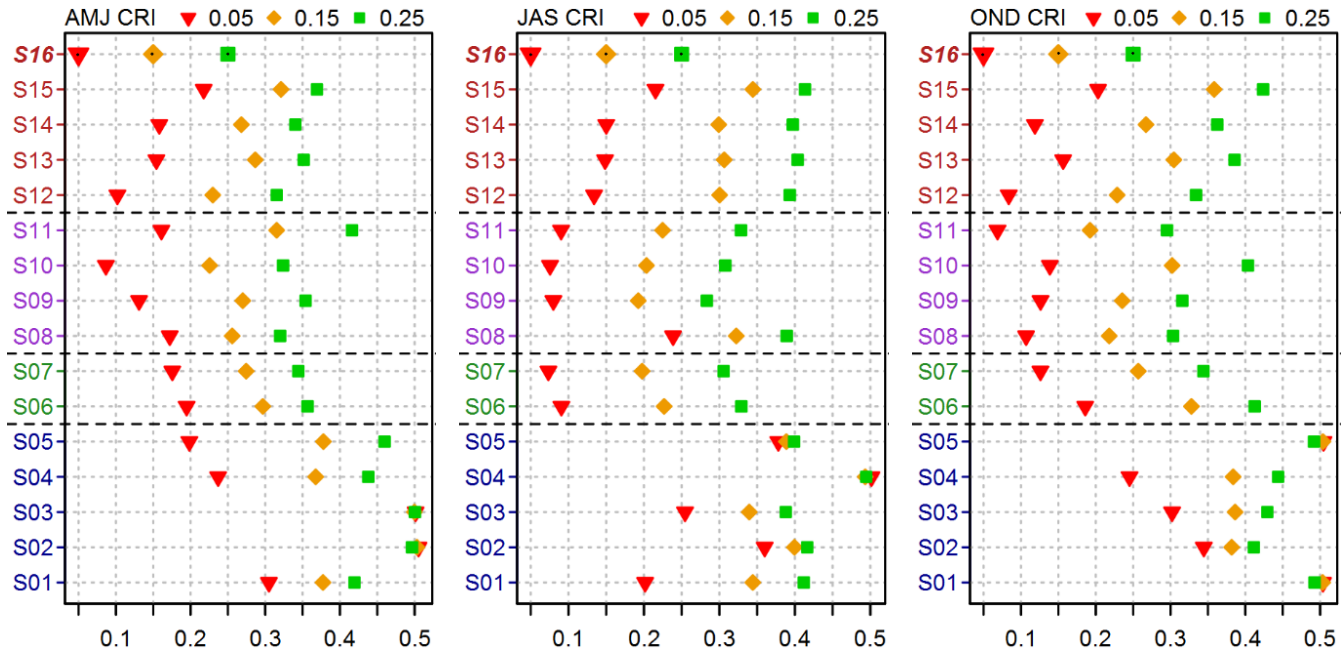


Figure 10. Similar to Fig. 6 but for the entire Australia's grain belt.

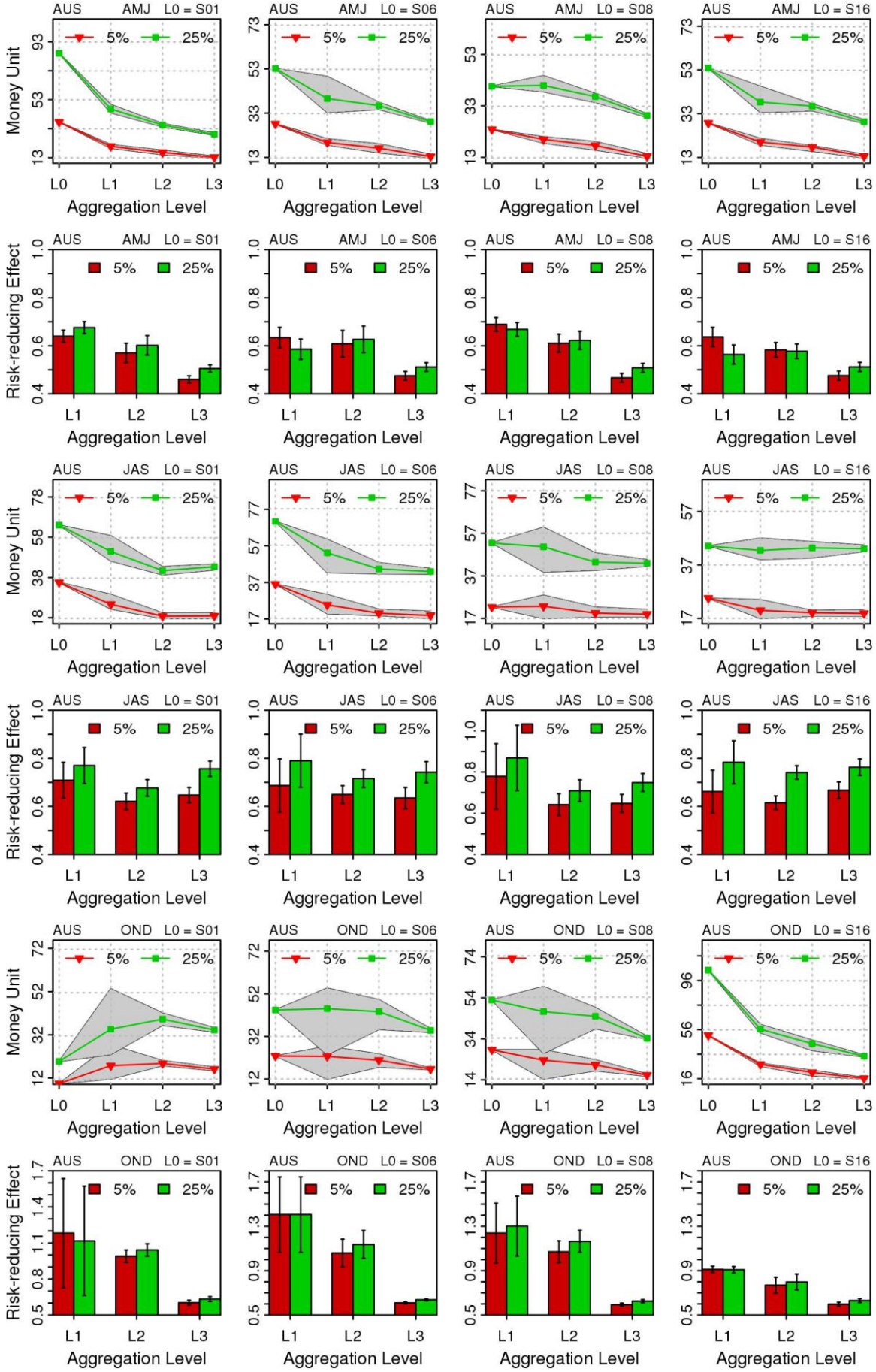


Figure 2. Similar to Fig. 7 but for the entire Australia's grain belt.

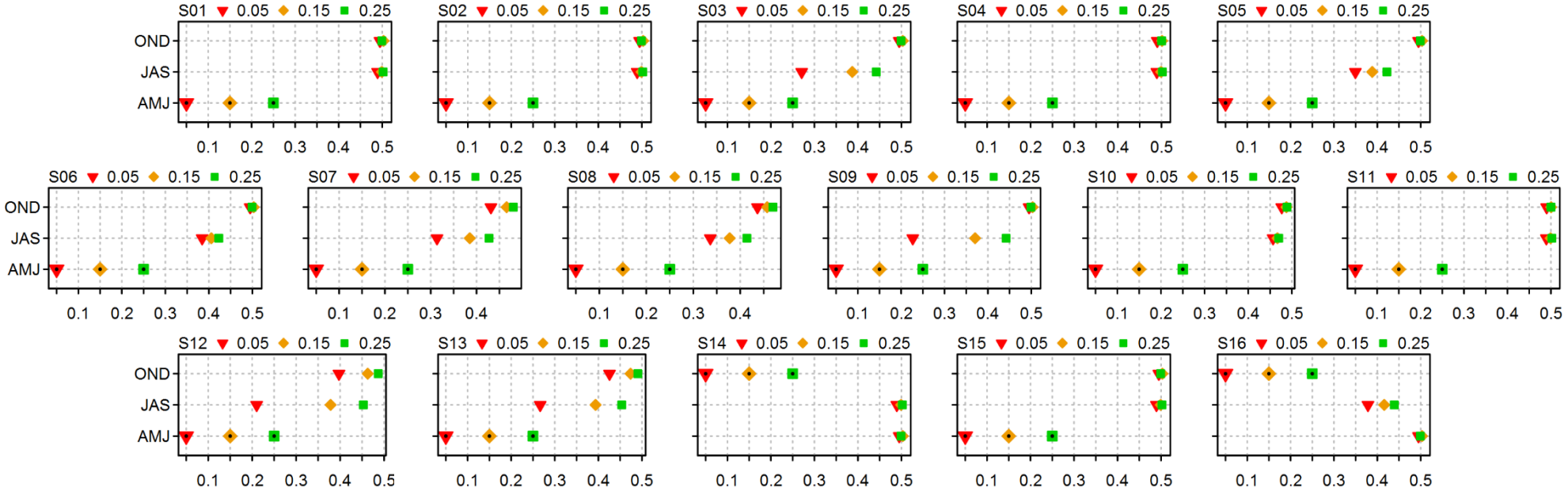
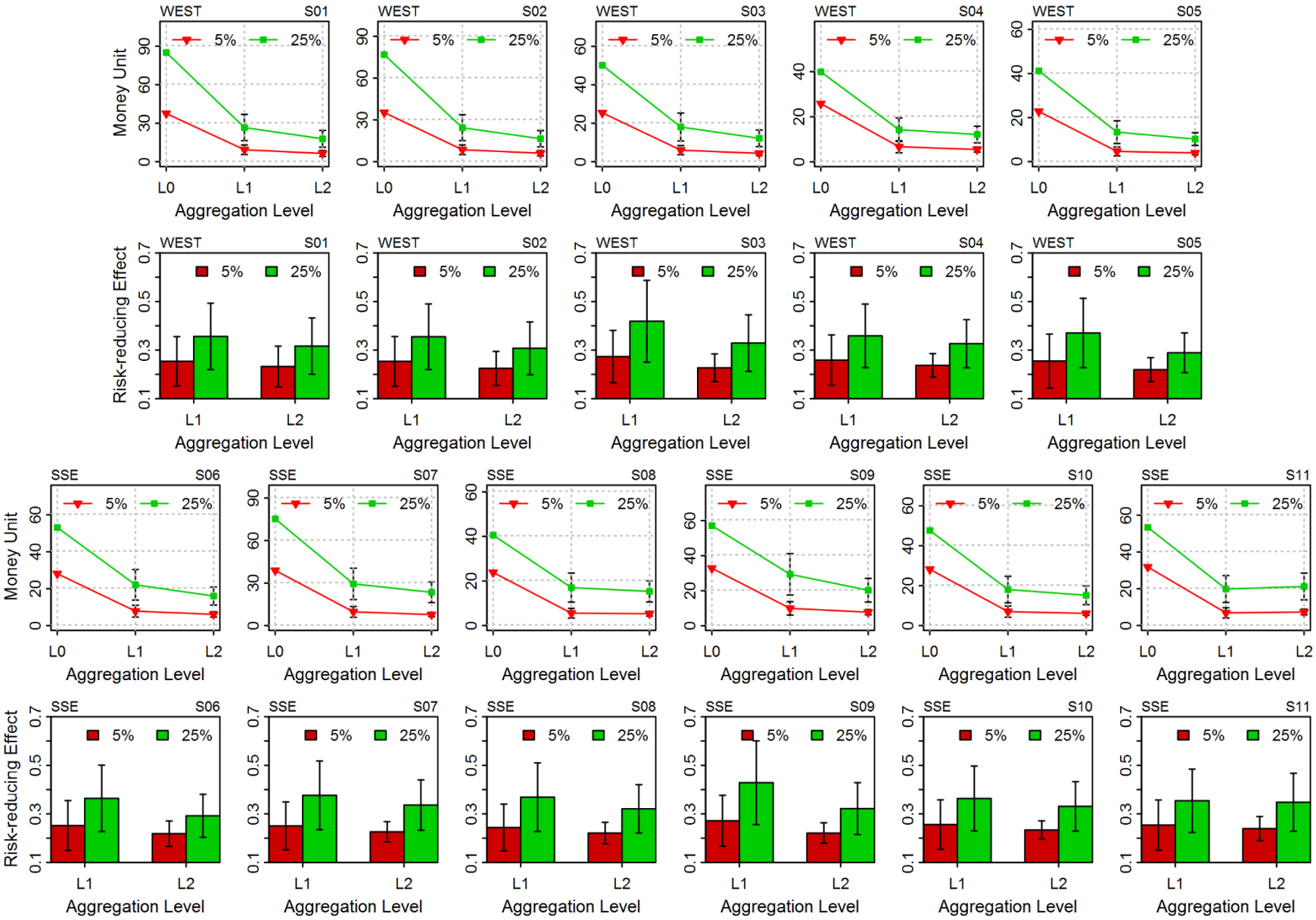


Figure 3. Similar to Fig. 6 but for different seasons.


WEST
S01
S02
S03
S04
S05
SSE
S06
S07
S08
S09
S10
S11
5%
25%
Money Unit
Risk-reducing Effect
Aggregation Level
L0
L1
L2

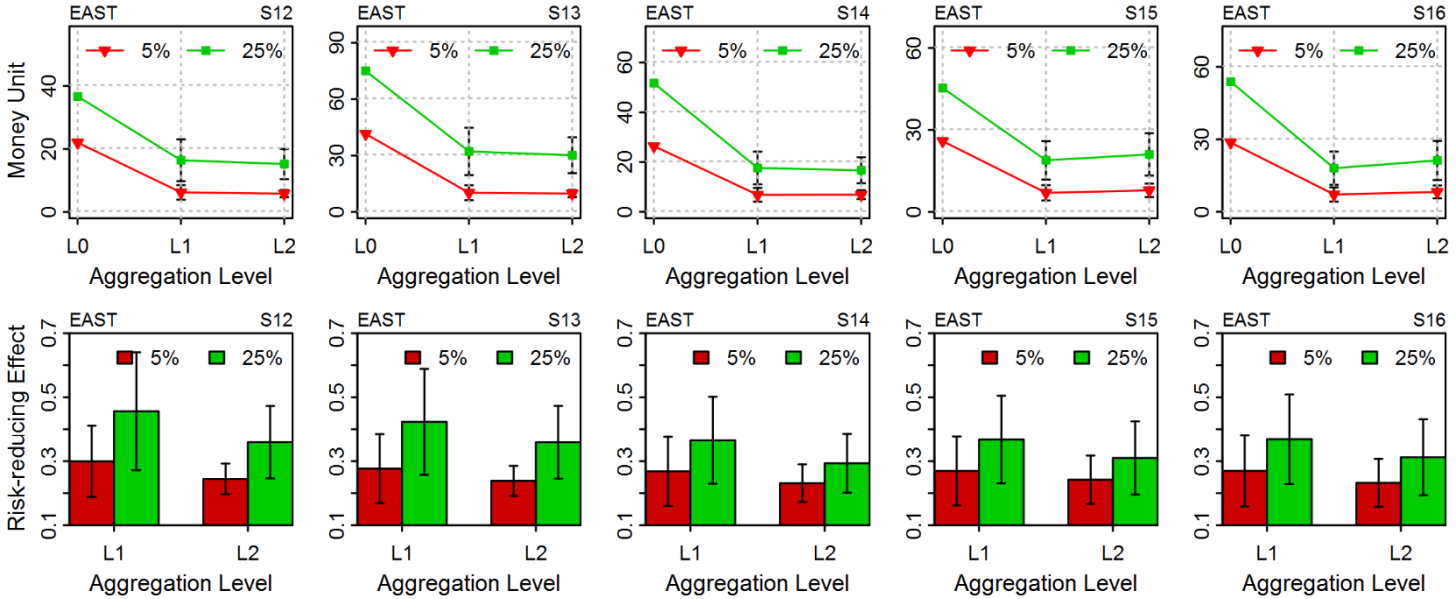


Figure 13. Similar to Fig. 7 but for different seasons.

Click here to access/download;table;Nguyen-Huy et al 2018 Systemic Risk Tables.docx

List of Tables

Table 1. Mean and standard deviation (SD) of cumulative rainfall index (mm) for the period of April – June (AMJ), July – September (JAS), and October – December (OND) for sixteen weather stations considered in this study.

Code for Australian region	Station ID	Name	Long – Lat	AMJ		JAS		OND	
				Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
West									
S01	08088	Mingenew	115.44^{0}E – 29.19^{0}S	163.92	70.36	167.02	54.92	36.04	24.78
S02	10111	Northam	116.66^{0}E – 31.65^{0}S	159.67	58.50	180.27	55.29	47.71	27.94
S03	12074	Southern Cross	119.33^{0}E – 31.23^{0}S	98.80	45.55	89.27	34.16	44.79	29.24
S04	10627	Pingrup	118.51^{0}E – 33.53^{0}S	117.55	43.06	123.36	37.28	60.74	37.69
S05	12070	Salmon Gums	121.64^{0}E – 32.98^{0}S	100.51	39.38	103.17	33.43	70.01	35.40
South									
S06	18064	Lock	135.76^{0}E – 33.57^{0}S	116.14	47.56	145.62	45.21	73.32	37.68
S07	21027	Jamestown	138.61^{0}E – 33.20^{0}S	129.48	56.72	166.44	52.74	106.97	54.12
South-East									
S08	76047	Ouyen	142.32^{0}E – 35.07^{0}S	86.55	43.78	95.82	38.65	87.20	53.40
S09	79023	Horsham Polkemmet	142.07^{0}E – 36.66^{0}S	127.76	53.57	144.59	47.87	107.74	51.73
S10	75031	Hay	144.85^{0}E – 34.52^{0}S	98.56	51.55	95.12	41.67	89.21	57.06
S11	73000	Barmedman	147.39^{0}E – 34.14^{0}S	115.49	59.15	116.59	48.07	124.46	62.56
East									
S12	48030	Cobar	145.80^{0}E – 31.50^{0}S	84.72	53.99	77.00	37.74	97.19	63.40
S13	55054	Tamworth	150.85^{0}E – 31.09^{0}S	130.14	67.54	137.86	55.56	204.95	78.99
S14	44030	Dirranbandi	148.23^{0}E – 28.58^{0}S	96.63	71.25	77.36	50.83	132.90	79.16
S15	41023	Dalby	151.26^{0}E – 27.18^{0}S	114.25	71.08	101.43	60.02	231.55	88.05
S16	35059	Rolleston	148.63^{0}E – 24.46^{0}S	113.33	86.80	80.29	65.71	200.03	89.22

Table 2. Average diversification effects for the entire insurance area in three different aggregation levels: regional (*i.e.*, west, south and south-east, and east), national (entire wheat belt region) and temporal (April – June, July – September, and October – December), with two strike levels: 5% and 25%.

Period	**Regional**			**National**
AMJ	*WEST*	*SSE*	*EAST*	
5%	0.88	0.67	0.72	0.48
25%	0.94	0.78	0.8	0.52
JAS				
5%	0.83	0.85	0.81	0.68
25%	0.89	0.93	0.88	0.76
OND				
5%	0.69	0.81	0.62	0.61
25%	0.78	0.88	0.73	0.65
	Temporal			
5%	0.25	0.25	0.26	
25%	0.32	0.33	0.33	

Table 3. Example of selected C-vine copulas for the west region in April – June (AMJ), south and southeast (SSE) region in July – September (JAS), and east region in October – December (OND) with the corresponding root stations (e.g., L0 = S02), parameters (θ_1 and θ_2), Kendall's tau (τ), upper and lower tail coefficients (λ_U and λ_L), maximum log-likelihood (ll_{max}), and Akaike Information Criterion (AIC). Copulas have no upper and/or lower tail denoted as na.

Tree	Edge	Copula	θ_1	θ_2	τ	λ_U	λ_L
WEST AMJ: L0 = S02; ll_{max} = 175.43; AIC = -336.85							
1	2,3	Clayton	1.58	0.00	0.44	na	0.64
	2,1	S. Gumbel	2.38	0.00	0.58	na	0.66
	2,4	Gaussian	0.71	0.00	0.51	na	na
	5,2	S. Gumbel	1.58	0.00	0.37	na	0.45
2	5,3;2	Gaussian	0.51	0.00	0.34	na	na
	5,1;2	S. Gumbel	1.26	0.00	0.20	na	0.26
	5,4;2	Frank	2.46	0.00	0.26	na	na
3	1,3;5,2	I	na	na	0.00	na	na
	4,1;5,2	I	na	na	0.00	na	na
4	4,3;1,5,2	I	na	na	0.00	na	na
SSE JAS: L0 = S08; ll_{max} = 280.01; AIC = -544.02							
1	3,4	Gaussian	0.85	0.00	0.65	na	na
	3,1	S. Gumbel	1.82	0.00	0.45	na	0.54
	3,5	S. Gumbel	2.1	0.00	0.52	na	0.61
	3,2	S. Gumbel	2.16	0.00	0.54	na	0.62
	6,3	S. Gumbel	1.82	0.00	0.45	na	0.54
2	2,4;3	Frank	2.2	0.00	0.23	na	na
	2,1;3	S. Gumbel	1.6	0.00	0.37	na	0.46
	2,5;3	I	na	na	0.00	na	na
	6,2;3	I	na	na	0.00	na	na
3	5,4;2,3	I	na	na	0.00	na	na
	5,1;2,3	I	na	na	0.00	na	na
	6,5;2,3	Gumbel	1.59	0.00	0.37	0.45	na
4	1,4;5,2,3	I	na	na	0.00	na	na
	6,1;5,2,3	I	na	na	0.00	na	na
5	6,4;1,5,2,3	I	na	na	0.00	na	na
EAST OND: L0 = S13; ll_{max} = 102.92; AIC = -183.83							
1	2,1	Gaussian	0.59	0.00	0.4	na	na
	2,4	Gaussian	0.52	0.00	0.35	na	na
	2,3	S. Gumbel	1.68	0.00	0.41	na	0.49
	5,2	Gaussian	0.47	0.00	0.31	na	na
2	3,1;2	Joe-Clayton	1.17	0.57	0.28	0.19	0.30
	3,4;2	Frank	1.5	0.00	0.16	na	na
	5,3;2	Joe-Frank	1.93	0.78	0.18	na	na
3	5,1;3,2	Gaussian	0.25	0.00	0.16	na	na
	5,4;3,2	Gaussian	0.36	0.00	0.24	na	na
4	4,1;5,3,2	I	na	na	0.00	na	na

Chapter 5

Diversification for wheat farming portfolio optimisation

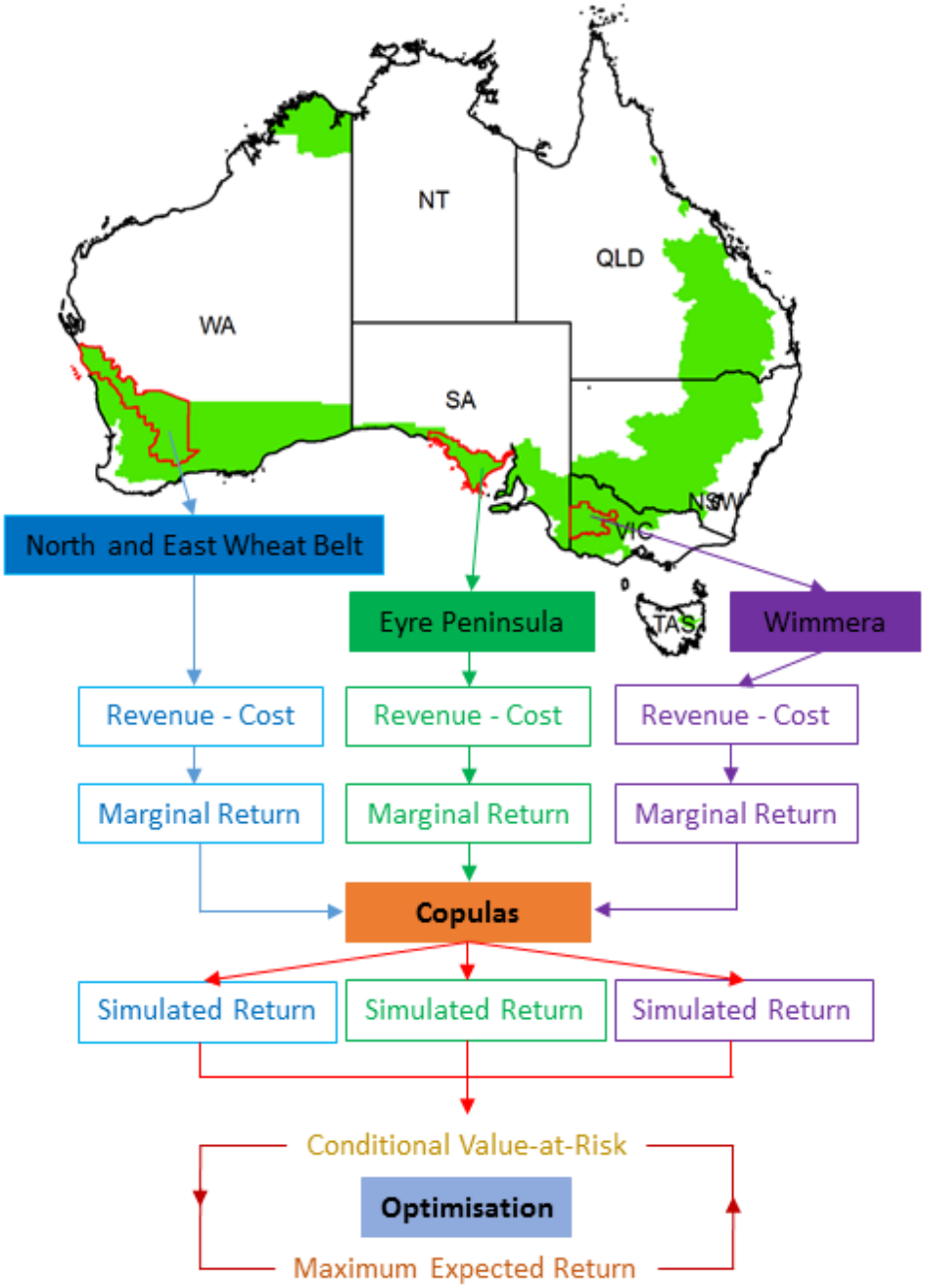
Article IV: Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management

Summary:

Geographical diversification has been identified as a potential farmer adaptation and decision support tool that could assist producers to reduce unfavourable financial impacts due to variability in crop price and yield, associated with climate variations. This study investigates whether a wheat farm portfolio that is geographically diversified over three climate rain-fed locations could potentially reduce financial risks related to climate variability for producers in Australia's wheat belt. We propose a new statistical approach: a set of popular and statistically robust tools commonly applied in finance and statistical theories including the Conditional Value-at-Risk (CVaR) and copula models are combined to evaluate the effectiveness of geographical diversification. CVaR is utilised to benchmark the loss (*i.e.*, downside risk), while the copula function is employed to model joint distribution among marginal returns (*i.e.*, profit in each zone). Fig. 9 reveal the graphical abstract of this study on geographical diversification for wheat portfolio optimisation in Australia.

The results of mean-CVaR optimisations indicate that geographical diversification is a feasible agricultural risk instrument for wheat-farming portfolio managers in achieving their optimised expected returns while controlling the risks (*i.e.*, targeting levels of risk). For example, by allocating about 10% of their production area to VIC, wheat producers in the SA region can adjust their expected profitability in the worst 5% of the cases from approximately 33.98% to 33.69% (*i.e.*, a reduction of 0.29%), which in turn can reduce the downside risk from approximately 14.70% to 11.51% (*i.e.*, a risk reduction of 3.19%). This is because the average marginal return (and the standard deviation) in SA region is just 3.67% higher (and 1.09% lower) than that in VIC region. The kurtosis (and skewness) in the SA region is also 22.59% higher (and 27.06% lower) than that in the VIC region. Furthermore, the copula-based mean-

CVaR model is found to better simulate extreme losses compared to the traditional multivariate-normal models, which underestimate the minimum risk levels at a given target of expected return. Among the suite of tested copula-based models, the vine copula is found to be superior in capturing the tail dependencies compared to the other multivariate copula models investigated.



Source: Author

Figure 9. Graphical display of the study on geographical diversification for wheat portfolio optimisation in Australia.

Weather and Climate Extremes 21 (2018) 76–89

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management

Thong Nguyen-Huy[a,b,c,d,**], Ravinesh C. Deo[a,b,c,*], Shahbaz Mushtaq[b,c], Jarrod Kath[b,c], Shahjahan Khan[a,b,c]

[a] School of Agricultural, Computational and Environmental Sciences, QLD 4350, Australia
[b] Centre for Applied Climate Sciences, QLD 4350, Australia
[c] University of Southern Queensland, Institute of Agriculture and Environment, QLD 4350, Australia
[d] Vietnam National Space Center (VNSC), Vietnam Academy of Science and Technology (VAST), Australia

ARTICLE INFO

Keywords:
Copula models
Portfolio optimisation
Conditional value-at-risk
Agriculture management
Crop decision
Geographical diversification

ABSTRACT

An agricultural producer's crop yield and the subsequent farming revenues are affected by many complex factors, including price fluctuations, government policy and climate (e.g., rainfall and temperature) extremes. Geographical diversification is identified as a potential farmer adaptation and decision support tool that could assist producers to reduce unfavourable financial impacts due to the variabilities in crop price and yield, associated with climate variations. There has been limited research performed on the effectiveness of this strategy. This paper proposes a new statistical approach to investigate whether the geographical spread of wheat farm portfolios across three climate broad-acre (i.e., rain-fed) zones could potentially reduce financial risks for producers in the Australian agro-ecological zones. A suite of popular and statistically robust tools applied in the financial sector based on the well-established statistical theories, comprised of the Conditional Value-at-Risk (CVaR) and the joint copula models were employed to evaluate the effectiveness geographical diversification. CVaR is utilised to benchmark the losses (i.e., the downside risk), while the copula function is employed to model the joint distribution among marginal returns (i.e., profit in each zone). The mean-CVaR optimisations indicate that geographical diversification could be a feasible agricultural risk management approach for wheat farm portfolio managers in achieving their optimised expected returns while controlling the risks (i.e., target levels of risk). Further, in this study, the copula-based mean-CVaR model is seen to better simulate extreme losses compared to the conventional multivariate-normal models, which underestimate the minimum risk levels at a given target of expected return. Among the suite of tested copula-based models, the vine copula in this study is found to be a superior in capturing the tail dependencies compared to the other multivariate copula models investigated. The present study provides innovative solutions to agricultural risk management with advanced statistical models using Australia as a case study region, also with broader implications to other regions where farming revenues may be optimized through copula-statistical models.

1. Introduction

Climate variability significantly influences agricultural production and the subsequent revenues received from the sale of various crops. However, recent extreme climatic events have been linked to large losses in agricultural production, in both developing and developed nations (Barriopedro et al., 2011; Coumou and Rahmstorf, 2012; Herold et al., 2018). For instance, about one-quarter of the damaged agricultural production in developing nations has been associated with extreme climate-related disasters (FAO, 2015). In addition, the study of Lesk et al. (2016) reported that extreme drought and heat events have also caused a significant decline in cereal production ranging from 9 to 21% in both developed and developing nations. To mitigate and possibly, to reduce agricultural yields and the associated financial losses that could be triggered by extreme climate events, agricultural adaptation strategies are required.

Portfolio theory suggests that the geographical diversification strategy could assist farmers in reducing the impacts of the variabilities faced in respect to the crop yield and prices associated with climate variabilities and the changes in other types of factors (Bradshaw et al.,

* Corresponding author. School of Agricultural, Computational and Environmental Sciences, QLD 4350, Australia.
** Corresponding author. School of Agricultural, Computational and Environmental Sciences, QLD 4350, Australia.
E-mail addresses: thonghuy.nguyen@usq.edu.au (T. Nguyen-Huy), ravinesh.deo@usq.edu.au (R.C. Deo).

https://doi.org/10.1016/j.wace.2018.07.002
Received 22 March 2018; Received in revised form 21 June 2018; Accepted 20 July 2018
Available online 20 July 2018
2212-0947/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

2004; Mishra et al., 2004). This means that farming systems are diversified over space to reduce the impact of systemic risks. However, the effectiveness of geographical diversification strategies in agriculture is to date poorly studied (see Larsen et al., 2015). To address this need, especially for agricultural reliant nations (e.g., Australia), this study aims to investigate the utility of geographical diversification in portfolio management of wheat farming, an important grain crop for Australia's agricultural sector (Murray and Brennan, 2009).

In classical Markowitz mean-variance (MV) portfolio optimisation, efficient portfolios are optimised to minimise their variances and to reduce overall financial risk (Markowitz, 1952). Hence, each portfolio along the efficient frontier must have a minimum variance for that level of return. However, despite its popularity, the MV method has limitations. For example, the variance metric is a symmetrical measure that does not take into consideration the direction of the co-movement. Minimising the variance penalises the downside risk in a manner appearing the same as the upside risk of the portfolio return distribution. This is an issue since an asset that experiences better than the expected return is deemed to be a risky scenario relative to an asset that is suffering from a lower than expected return. To address this issue, alternative risk-based measures such as the Value-at-Risk (VaR) and the Conditional Value-at-Risk (CVaR) have been introduced to replace the MV method.

Rockafellar and Uryasev (2000) have recommended CVaR as a measure of alternative risk that is preferred to the common VaR concept. A CVaR-based optimised portfolio only penalises for the loss (i.e., the downside risk), and not the gain (i.e., upside risk) in the portfolio return distribution. It is related to but is superior to the VaR for optimisation applications for several reasons. Firstly, the CVaR tends to satisfy the four properties of a coherent risk measure; translation invariance, monotonicity, subadditivity and positive homogeneity (Pflug, 2000). Secondly, the VaR is able to describe a loss of X or greater than this, and thus, this last clause tends to be omitted in most cases when people quote the VaR. CVaR, on the contrary, is an estimate of the size of the tail loss, which gives a more accurate estimate of the associated risk.

In the existing literature, common methods of calculating the CVaR normally consists of the variance-covariance, historical and the Monte Carlo simulation (Chernozhukov and Umantsev, 2001; Zhu and Fukushima, 2009). Calculating CVaR also involves an estimation of the tails of the joint distribution among the marginal returns (i.e., the profit of each farm that is considered in the problem). However, the variance-covariance and historical simulation method have some degree of restrictions, which might not be always reasonable, and necessarily true in practice. For example, the variance-covariance method assumes the returns to be normally distributed, which can be problematic from a practical point of view. This is because many financial returns have elongated and broadened tails in the dataset so a normal distribution assumption can seriously underestimate the size (and the pivotal role) of the tail end of the data (Ang and Chen, 2002; Embrechts et al., 2001; Longin and Solnik, 2001). Simulations based on historical data also assumes that the distributions of the returns in the future are similar to those in the past. Furthermore, in most cases, there are relatively few data points that are present in, for example, the 0–5th percentile or extreme tail of the distribution. The Monte Carlo method is therefore preferred in such circumstances since it is able to calculate the CVaR in a similar fashion to historical simulation, while also being based on the randomly generating scenarios from a model whose parameters are acquired from the historical data.

As mentioned above, the non-linear interdependence at the tails between the marginal returns need to be captured more effectively relative to conventional approaches in order to obtain accurate estimations of CVaR. This requires a robust multivariate prediction model that is capable of fully capturing the joint dependence structure among the related variables. A conventional approach commonly relies on the utilization of a multivariate-normal distribution that assumes a normality of the considered variable(s). However, there is no doubt that the agricultural prices and crop yields have been shown to be non-normally distributed (e.g., Goodwin and Ker, 2002), and therefore, any approach that does not consider this important data limitation aspect can lead to erroneous conclusions. Fortunately, copula functions (that can analyse non-linearity in multivariate data) is able to provide an alternative statistical approach to modelling the joint distribution of multivariate datasets, allowing one to specify the marginal distribution among the tested variable and their dependence structures independently. Due to their distinct merits in modelling multivariate joint distributions, copula-based models have been applied extensively in many fields such as insurance and financial risk modelling (Hu, 2006; Kole et al., 2007), hydrology and water resources (Chowdhary et al., 2011; Favre et al., 2004), drought studies, agricultural and precipitation forecasting (Bessa et al., 2012; Ganguli and Reddy, 2012; Janga Reddy and Ganguli, 2012; Nguyen-Huy et al., 2017; Vergni et al., 2015; Nguyen-Huy et al., 2018).

Although copula method is a popular tool in financial risk literature in general and also in portfolio analysis (Boubaker and Sghaier, 2013; Huang et al., 2009; Kresta and Tichý, 2012), its application in agricultural risk management and crop insurance aspects are relatively recent (Bokusheva, 2014; Goodwin and Hungerford, 2014; Nguyen-Huy et al., 2018; Okhrin et al., 2013; Vedenov, 2008). Furthermore, the published literature in this area shows limited research has been undertaken regarding the application of copulas in geographically diversifying risks in agriculture. In spite of this, some studies are particularly notable, for example, Larsen et al. (2015) proposed a copula-based mean-CVaR model to inspect the potential benefits of risk reduction using a geographical diversification strategy for the case of a US wheat farming scenario. The authors applied multivariate Archimedean copula model and compared it with a traditional multivariate-normal model as a benchmark tool. The mean-CVaR optimisation results indicated the effectiveness of geographical diversification in risk management strategy from a farm's marginal return viewpoint. It was not surprising to note that the multivariate-normal model led to an underestimation of the minimum level of associated risk faced by the wheat farmer at a given level of agricultural profitability. Importantly, the study concluded the copula-based model performed more appropriately for extreme losses of the farm profitability. However, the multivariate Archimedean copulas assume the same dependence parameter among the pair of variables. This sort of assumption is unrealistic in practical scenario (Hao and Singh, 2016; Zhang and Singh, 2014; Nguyen-Huy et al., 2018).

In this paper, we focus on wheat, a primary cereal crop in Australia. However, wheat is mostly grown in drylands in Australia (i.e., as a rainfed crop) that exhibits one of the world's most extreme variable climate conditions (Portmann et al., 2010; Turner, 2004). However, to the best of the authors' knowledge, the effectiveness of geographical diversification including the mean-CVaR optimisation in risk management strategy has not been examined in Australian farming contexts. The present study, therefore, utilises the contemporary vine copula method in Monte Carlo simulation approach for calculating the corresponding value of CVaR. This approach allows to randomly generate the scenarios of the marginal returns of wheat farming based on their joint distribution. The primary merit of vine copula model (Nguyen-Huy et al., 2017, 2018) (in comparison to the other types of multivariate copulas) is that it allows the integration of different bivariate copulas for the modelling of the flexible dependence among the pairwise variable disregarding the marginal selections differences (Bedford and Cooke, 2002).

By extending previous studies in the context of agricultural yield modelling and seasonal precipitation forecasting studies in Australia (Nguyen-Huy et al., 2017, 2018), the aims of the present study are as follows. (1) To investigate the effectiveness of the geographical diversification strategy in reducing risks in agricultural operations. (2) To demonstrate a robust statistical method, the vine copula-based mean-

CVaR model, for quantifying optimum amount of diversification needed for given level of risk. (3) To compare the traditional multivariate-normal, multivariate Archimedean and vine copula model in simulating the extreme losses. The vine copula-based mean-CVaR approach is expected to perform better and provide further insights into improving conventional multivariate-normal models that underestimate the minimum risk levels at a given target of profitability.

2. Materials and method

2.1. Data

In this study, we used aggregated yield and financial data from three of Australia's key wheat producing zones collected from the Department of Agriculture and Water Resources, Australian Government (AgSurf) (http://apps.daff.gov.au/agsurf/) for the period 1990–2016. The three broad-acre wheat zones include Wimmera (Victoria), Eyre Peninsula (South Australia), and the North and East Wheat Belt (Western Australia) where the "()" show the respective wheat growing States. For conciseness and consistency, the study site names henceforth are based on Australian States (i.e., VIC, SA & WA, respectively) for each of the wheat growing zone. These zones, reported in previous studies (Nguyen-Huy et al., 2017, 2018) have been selected as they are geographically distinct spanning across a wide range of climatic and wheat growing conditions and so are expected to expose to different risks at different times (Fig. 1).

The data are as per farm averages, including the wheat receipts ($), the total area sown (ha) and the costs per hectare ($/ha). The total cost consists of the contracts, chemicals, electricity, fertiliser, fuel, interest paid, water charges, repairs, seed, insurance, labour and some of the other related expenses. Marginal returns measured at the farm profitability are expressed as the percentage of the gross revenue exceeds the total cost. The marginal return of the ith zone r_i, (i = 1,2,3) is calculated as follows (Larsen et al., 2015):

$$gross\ \ revenue = wheat\ \ receipts/toal\ \ area\ \ shown \tag{1}$$

$$r_i = \frac{gross\ \ revenue - total\ \ \text{cost}}{gross\ \ revenue} \tag{2}$$

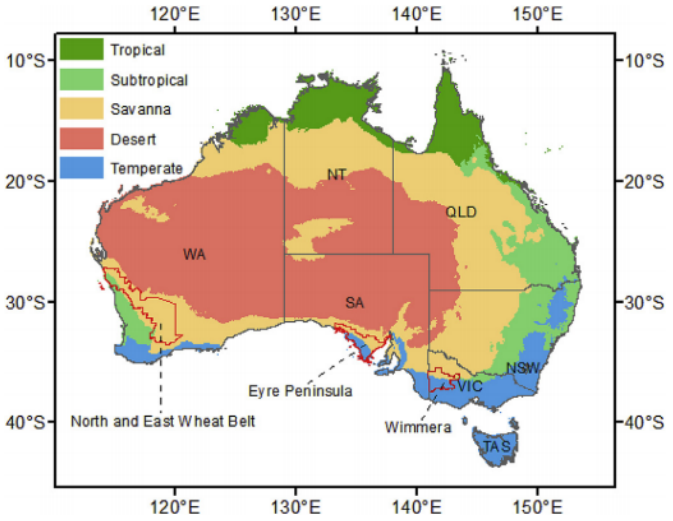


Fig. 1. Location of the broad-acre wheat zones in Australia that spans across different growing conditions. Wheat is grown mostly in temperate climate condition in Wimmera (Victoria, VIC). Eyre Peninsula (South Australia, SA) exhibits a mixture of the temperate and savanna while the entire North and East Wheat Belt (Western Australia, WA) is dominated by savanna.

3. Method

3.1. Conditional Value-at-risk

Suppose that $f(x, y)$ denotes a loss function depending upon the decision x, to be chosen from a feasible set of a realistic portfolio X, and a random vector y. Let $\Psi(x, \alpha)$ be the probability that the loss $f(x, y)$ does not exceed some threshold value α. The VaR function $\alpha_\beta(x)$, which is the percentile of the loss distribution at the confidence level β, is formally defined as (Rockafellar and Uryasev, 2000):

$$\alpha_\beta(x) = \min\{\alpha \in \mathbb{R} | \Psi(x, \alpha) \geq \beta\}. \tag{3}$$

By this definition, *CVaR* is able to measure the conditional expectation of the losses greater than that amount α. Therefore, the CVaR function $\phi_\beta(x)$ is defined mathematically as follows (Rockafellar and Uryasev, 2000):

$$\phi_\beta(x) = (1-\beta)^{-1} \int_{f(x,y) > \alpha_\beta(x)} f(x, y) p(y) dy, \tag{4}$$

Where $p(y)$ is the probability density function of the random vector y. It is clear that the CVaR is a greater bound for the VaR at the same confidence level. Also, with many advantages stated in the previous section, CVaR offers a more consistent risk measure than VaR and generally results more efficient in the context of portfolio optimisation (Mulvey and Erkan, 2006). In addition, CVaR can be expressed as a convex function allowing the construction of the portfolio optimisation problem which can be efficiently solved by linear programming techniques as shown in (Rockafellar and Uryasev, 2000) and will be described in the forthcoming method section. Although VaR plays a role in the optimal portfolio approach, it exposes some inherent restrictions as mentioned above. Therefore, the risk of high losses could be reduced through minimising CVaR rather than minimising VaR since a portfolio with low CVaR will necessary have low VaR as well (Rockafellar and Uryasev, 2000).

3.2. Portfolio optimisation problem

Suppose a portfolio consists of n production zones with a random percentage of the marginal returns $r_1, ..., r_n$, the marginal expected return $E[r_i]$ and w_i is a share of the total hectares allocated to the production zone (i.e., the decision vector or weight). The farmer's portfolio optimisation problem, in the context of the agricultural sector, is to maximise the expected returns (sum of all marginal expected returns multiply with the corresponding weights) of the portfolio given a specified risk level β. The portfolio optimisation problem can then be formulated as (Larsen et al., 2015):

$$maximise - - \sum_{i=1}^{n} w_i E[r_i], \tag{5}$$

$$subject,\ to - - \begin{cases} \phi_\beta(w_i) \leq \phi \\ \sum_{i=1}^{n} w_i = 1 \end{cases}, \tag{6}$$

where ϕ is defined as the target risk (CVaR) levels.

3.3. Calculating Conditional Value-at-risk

To solve the subject function in Eq. (6), the CVaR function in Eq. (4) can be expressed as (Rockafellar and Uryasev, 2000):

$$F_\beta(w, \alpha) = \alpha + (1-\beta)^{-1} \int [f(w, r) - \alpha]^+ p(r) dr, \tag{7}$$

where the indicator function:

$$[I]^+ = \begin{cases} I & when \quad I > 0 \\ 0 & when \quad I \leq 0 \end{cases}. \tag{8}$$

The integral in Eq (7) can be approximated further by sampling the

probability distribution of r based on its density $p(r)$ as (Rockafellar and Uryasev, 2000):

$$\widetilde{F}_{\beta}(w, \alpha) = \alpha + \frac{1}{m(1-\beta)} \sum_{j=1}^{m} [f(w, r_j) - \alpha]^{+}. \tag{9}$$

Therefore, the portfolio optimisation problem, as shown in Section 2.2.2, can be alternately formulated as the following linear programming problem:

$$minimise - -\alpha + \frac{1}{m(1-\beta)} \sum_{j=1}^{m} u_k, \tag{10}$$

$$subject, to - -\begin{cases} \sum_{k=1}^{n} w_i E[r_i] \geq R \\ f(w, r_j) - \alpha \leq u_k \\ 0 \leq u_k \\ \sum_{i=1}^{n} w_i = 1 \end{cases}, \tag{11}$$

where $[f(w, r_j) - \alpha]^{+} = u_k$ and R denotes the target expected returns. The sampling of vector r based on the copula methods is introduced in the next section. The linear optimisation problem was solved using the R-package **fPortfolio** (Würtz et al., 2009).

3.4. Copulas

As stipulated above, the calculation of the CVaR using the Monte Carlo simulation method requires a knowledge of the joint distribution of all marginal returns involved in the portfolio. To fulfill this requirement, Sklar (1959) theorem suggests that the joint distribution $F(x_1,...,x_n)$ can be expressed as:

$$F(x_1,...,x_n) = C[F_1(x_1),...,F_n(x_n)], \tag{12}$$

where C: $[0,1]^n \rightarrow [0,1]$ is a unique copula function and $F_i(x_i)$ are marginal distributions (margins) of variables of interest. Note that Eq. (12) implies that the unknown joint distribution can be constructed by two separate parts, including the copula function and the marginal distributions of the historical marginal returns.

Regarding the most suitable copula function, in this study, we have considered several copulas that are commonly classified into different families based on their construction methods, comprising, but not limited to, the elliptical, Archimedean, vine, empirical, extreme value, and the entropy copulas. For more details on the full suite of copula functions, readers are referred to the studies of Joe (1996), Nelsen (2006), and Bedford and Cooke (2002). In this paper, the first three families including the elliptical, Archimedean, and vine copula are tested and compared. The estimation and usage of these functions are described in the next section.

3.5. Construction of the copula-based model

We employ the vine copula approach that was previously utilised in our earlier published work (Nguyen-Huy et al., 2017, 2018) to develop vine copula-based models for this study. Here, we briefly describe the main steps of the vine copula model construction procedure. The first step in constructing the copula model is to select the theoretical distribution functions that are able to approximately describe the historical marginal returns. This study adopts the parametric approach to fit the historical marginal returns since later in the simulating process, the reverse distribution function needs to be used to transform the copula-modelled data back to the real scale values.

A set of twenty-five theoretical probability distributions are fitted to the marginal return data, which follows earlier studies (Nguyen-Huy et al., 2017, 2018). The candidate distribution is selected based on a statistical assessment of the goodness-of-fit test, i.e., the Kolmogorov-Smirnov statistic (KS). If the p-value of the KS test is greater than 0.05, we cannot reject the null hypothesis that the observed data follow that specific distribution. Then, the distribution with a lower Akaike Information Criterion (AIC) is selected for that data. Further, the graphical analysis is also performed to support selecting the most appropriate distribution function as in our previous works (Nguyen-Huy et al., 2017, 2018).

In the second step, the copula parameters are estimated using the maximum pseudo-likelihood method (Chowdhary et al., 2011), requiring the marginal return data to be transformed in the unit hypercube. In general, this transformation can be performed by applying either the fitted distribution (selected in the first step) or the empirical distribution. Here we utilise the empirical method (Genest and Favre, 2007) to ensure that the dependence structure between the pairwise data is independent of the marginal distributions. Thus, the marginal returns are transformed into the pseudo-data using the corresponding empirical distribution function $F(.)$ as $u_i = F(r_i)$. Henceforth, the copula parameters θ are estimated through the maximum pseudo-likelihood estimation method (Chowdhary et al., 2011):

$$\widehat{\theta} = \underset{\theta \in \Theta}{\arg\max} \sum_{t=1}^{T} \ln c(u_{1t},...,u_{nt}; \theta), \tag{13}$$

where $c(.)$ is the copula density. The most accurately fitted copula model is selected based on the Akaike Information Criterion $AIC = -2\ln(ll_{\max}) + 2k$ as the function of the maximum log-likelihood value ($ll_{\max}$) and the number of estimated parameters k.

Subsequently, a random vector $(u_1,...,u_n)$ whose marginal distributions follow a uniform distribution is generated using the selected copulas. The steps in randomly generating the data samples from the fitted copulas are in accordance with the study of Brechmann (2010). Finally, the simulated realizations of the marginal return for each zone are obtained by inverse transformation following $(r_1,...,r_n) = [F_1^{-1}(u_1),...,F_n^{-1}(u_n)]$. The six popular copula functions and their rotated functions were employed in this analysis including Gaussian, Student's t (symmetric but heavier tails), Clayton, Gumbel, Frank, and Joe. These copula functions are employed in the construction of both multivariate Archimedian and vine copula models. Readers may refer to the previously published study of Zhang and Singh (2014) for more details on the multivariate elliptical and Archimedean copulas, including vine copulas. The computations are performed using several packages, including: **copula** (Yan, 2007) and **VineCopula** (Schepsmeier et al., 2017) available in **R** software (R Core Team, 2016).

Further applications of a vine copula model in climate extreme event prediction and agricultural yield simulation can be found in our earlier studies (e.g. (Nguyen-Huy et al., 2017; Nguyen-Huy et al., 2018)).

4. Results and discussion

In this section, the modelled results generated to solve the problem of farming portfolio-optimisation based on optimal copula-statistical model are provided with a physical interpretation in context of the applied models and the problem of interest. Fluctuations in marginal returns potentially associated with extreme climate conditions are firstly represented. Henceforth, the results of the copula model selection are described using multivariate copulas and vine copula functions. The conventional multivariate-normal model is also developed, for a comparison of the results with multivariate copulas and vine copula models. Finally, we discuss the mean-CVaR optimisations and optimal portfolio allocation results derived from models at different confidence levels.

4.1. Variations in the marginal return

Fig. 2 illustrates the historical marginal returns of each wheat growing zones in Australia. The pattern of marginal return at SA appears to be most stable, except for 2007–8. The extreme losses occurring in all zones for the period 2006–7 may be associated with one of

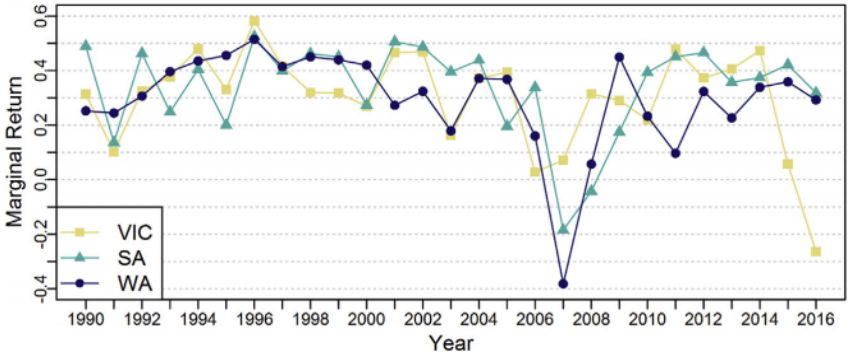


Fig. 2. Historical marginal returns over the period study 1990–2016 at the three wheat production zones in Australia: VIC, SA and WA.

Table 1
The degree of dependence of the farm-level return margins across the different wheat growing study sites across Victoria (VIC), South Australia (SA) and Western Australia (WA) measured by the Pearson's correlation coefficient, Spearman's rho, and the Kendall's τ parameters.

Tested Study Site	VIC	SA	WA
Pearson's correlation coefficient			
VIC	1.0000	0.3643	0.3585
SA		1.0000	0.5770
WA			1.0000
Spearman's rho			
VIC	1.0000	0.4438	0.3358
SA		1.0000	0.1978
WA			1.0000
Kendall's τ			
VIC	1.0000	0.3105	0.2422
SA		1.0000	0.1339
WA			1.0000

Table 2
Summary statistics for the return margins at the three wheat zones: VIC, SA and WA.

Statistical Property	VIC	SA	WA
Mean	0.3018	0.3385	0.2962
Maximum	0.5825	0.5251	0.5150
Minimum	−0.2630	−0.1849	−0.3816
SD	0.1808	0.1699	0.1776
Skewness	−1.2370	−1.5076	−2.1309
Kurtosis	1.7342	1.9601	5.9013
Shapiro-Wilk test	0.8991	0.8436	0.8071
p-value	0.0128	0.0009	0.0002

Table 3
Selected marginal distributions with their parameters, Akaike Information Criterion (AIC), and the p-value of the Kolmogorov-Smirnov statistic for marginal returns.

Zones	Distribution	Parameters	AIC	p-value
VIC	Generalised Logistic	location = 0.4533 scale = 0.0421 shape = 0.2529	−17.9163	0.9498
SA	Generalised Logistic	location = 0.5047 scale = 0.0130 shape = 0.0775	−30.1221	0.9975
WA	Generalised Logistic	location = 0.4586 scale = 0.0217 shape = 0.1302	−26.0397	0.9942

most severe drought conditions on record, caused by the El Niño event across most of Australia (Bureau of Meteorology). However, it is noted that the marginal return at each zone generally moves in an opposite direction to that in other zones. It can be observed clearly during the El Niño year of 2006–07, while the marginal returns at the SA and the WA farms dropped severely, that at the VIC farm increased considerably. Moreover, the marginal returns at VIC and SA farms are seen to fluctuate during the five consecutive El Niño years of 1991–5 (https://www.longpaddock.qld.gov.au/), however, that of the WA farm during the same period either remained stable or increased. If we study the data further, the opposite co-movement of the marginal returns is also indicated by the generally low correlation coefficients and the different degrees of dependence between the marginal returns of each study zone pair (see Table 1). The stochastic nature of the marginal returns at these study zones clearly suggests that the geographical diversification can be considered as a feasible risk management strategy to possibly assist the wheat farmers in reducing their losses.

In Table 2, we summarize the basic statistics of the historical marginal returns. The difference between the highest (i.e., SA) and the lowest (i.e., WA) average marginal return is found to be approximately 14%. Notably, VIC is seen to have the widest range of marginal return that varies from −26% (loss) to 58% (gain), while the marginal return at SA is seen to be the smallest, ranging from −19% to 53%. The highest marginal return at WA is approximately 52%, whereas the lowest is approximate −38%. It is worth pointing out that the maximum and the minimum values of the marginal returns at the VIC and WA study sites suggest that these farming zones might potentially yield a high profitability but they may also potentially have an extremely low profitability. Both of these zones have the highest standard deviation, as expected. Conversely, the SA farming region does not exhibit extreme values of marginal return accompanying the lowest standard deviation. Therefore, a visual conclusion derived from the analysis of the summary statistics is that the growing of wheat in SA is likely to gain a more stable benefit and a reduction in some risks. However, the skewness and kurtosis also expose VIC has the lowest outliers in the lower tail (extreme losses).

Table 2 also provides information regarding the higher moments of the marginal return data indicating the unreality of the normal assumption of marginal returns. It can be seen that WA study site has the highest absolute values of the skewness (2.13) and the kurtosis (5.90) factors, following by SA (1.5 and 1.9, respectively), meanwhile those are the lowest at VIC (1.24 and 1.73, respectively). According to Curran et al. (1996), a normal distribution has the skewness equal to 0 and the kurtosis equal to 3. It is clear that the skewness and kurtosis of all the three zones are significantly different to those of normal distribution. Therefore, it is suggested that the distributions of the marginal returns at three zones are non-normal and asymmetric. The results from the Shapiro-Wilk normality test also reject the hypothesis that the marginal return data are normally distributed with p-values less than 0.1. These results, therefore, question the practice of the linear correlation analysis and normal assumptions in previous studies, to justify the use of the non-linear copula approach that is pursued in this study.

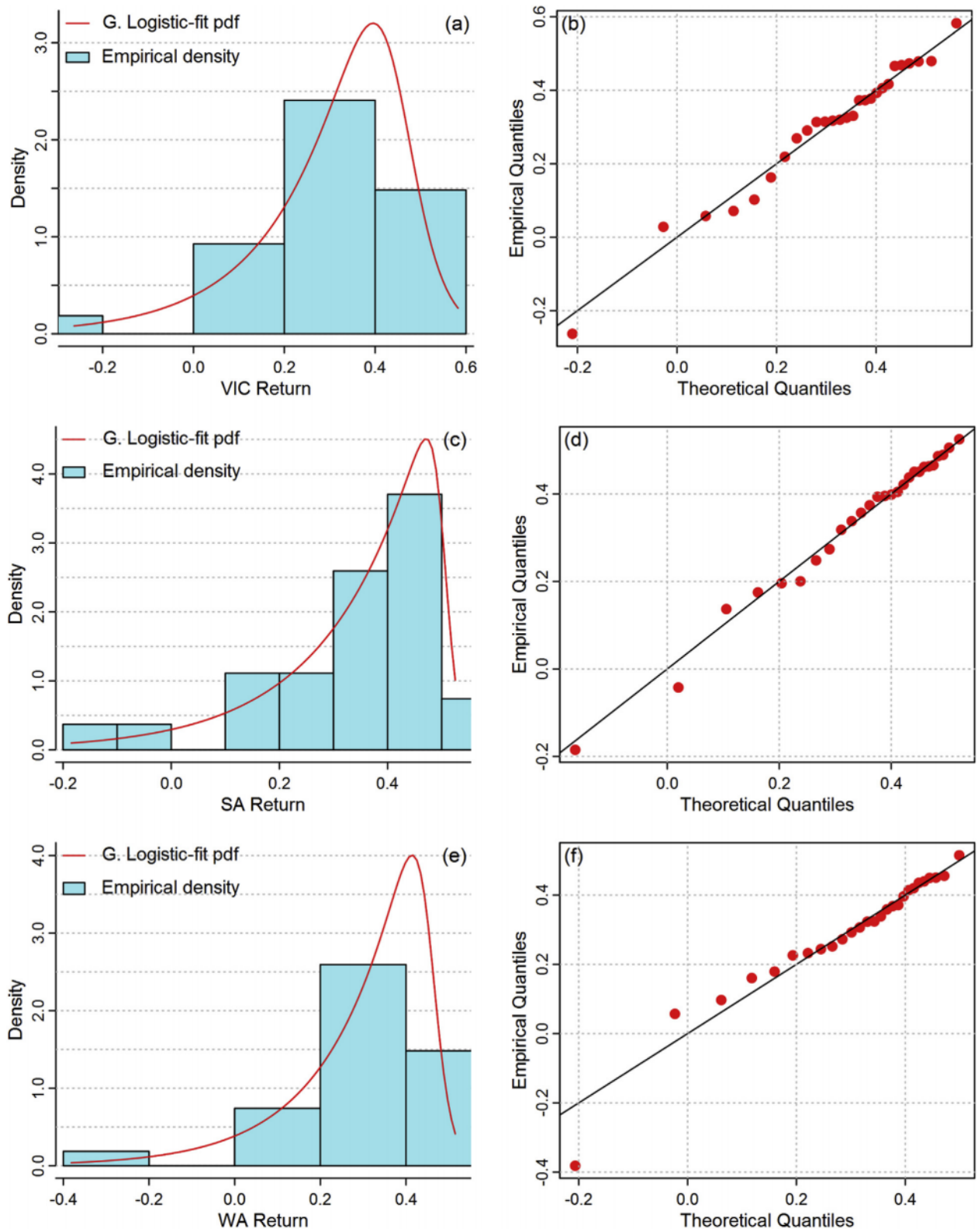


Fig. 3. Graphical analysis of goodness-of-fit for selecting marginal distributions approximate to VIC (a–b), SA (c–d), and WA (e–f) returns with density and quantile-quantile plots.

4.2. Copula model

As the first step of the model construction, the historical marginal returns are fitted to the theoretical distribution curves (Nguyen-Huy et al., 2017, 2018). All of the three historical marginal return data can be appropriately described by the generalised logistic distribution with the estimated parameters shown in Table 3. The graphical assessment involves the density, cumulative distribution function, quantile-quantile, and probability-probability plots, which are analysed to confirm the marginal distribution results. Fig. 3 displays the density and

Table 4
Copula parameters, maximum log-likelihood (ll_{max}), and the Akaike Information Criterion (AIC).

Copula function	Parameters	ll_{max}	AIC
Gaussian	ρ_1 = 5526, ρ_2 = 0.453, ρ_3 = 0.395	6.060	−6.121
Student's t	ρ_1 = 0.417, ρ_2 = 0.371, ρ_3 = 0.282, ν = 4.000	5.526	−5.052
Clayton	θ = 0.655	4.980	−7.959
Gumbel	θ = 1.365	5.688	−9.376
Frank	θ = 2.189	3.679	−5.358
Joe	θ = 1.512	4.827	−7.653
Survival Clayton	θ = 0.654	4.722	−7.444
Survival Gumbel	θ = 1.342	4.980	−7.960
Survival Joe	θ = 1.482	4.379	−6.759

Table 5
Structure of vine copula model with parameters, maximum log-likelihood (ll_{max}), and Akaike Information Criterion (AIC).

Tree level	Edge	Copula function	Parameter	ll_{max}	AIC
SA as center: VIC – SA – WA					
T_1	VIC, SA	Survival Clayton	θ = 0.949	8.389	−8.779
	WA, SA	Student's t	ρ = 1.604, ν = 2.000		
T_2	VIC|SA, WA|SA	Survival Joe	θ = 1.560		
VIC as center: SA – VIC – WA					
T_1	SA, VIC	Survival Clayton	θ = 0.949	6.040	−6.079
	WA, VIC	Gumbel	θ = 1.370		
T_2	SA|VIC, WA|VIC	Gumbel	θ = 1.082		
WA as center: SA – WA – VIC					
T_1	SA, WA	Student's t	ρ = 1.604, ν = 2.000	7.330	−6.660
	VIC, WA	Gumbel	θ = 1.370		
T_2	SA|WA, VIC|WA	Survival Clayton	θ = 0.547		

quantile-quantile plots (as for example), while graphical analysis of goodness-of-fit in conjunction with statistical test in Table 3 support the selection of the generalised logistic distribution for fitting the returns in VIC, SA, and WA.

Table 4 represents the summary results of the multivariate copula functions with the corresponding parameters, maximum log-likelihood (ll_{max}), and AIC. Based on the AIC, the results show that the Gumbel copula is the most appropriate copula model regarding the case of multivariate copulas. The same set of copula functions are employed for the vine copula development and the selected vine copula model is illustrated in Table 5. Similar to the procedure adopted for fitting the marginal distributions, in this study we also applied graphical tools to support the selection of the most suitable copula. Fig. 4 plots the contours of the selected bivariate copulas for each pairs of the returns, superimposed on empirical observations and simulated data derived from the corresponding copulas.

Since there are three zones in this study it is pertinent to construct three unique drawable D- and canonical C-vine copulas (Aas et al., 2009). The vine copula with the construction data of VIC – SA – WA combinations is selected among the three cases since this construction yields the lowest AIC. It is noteworthy that the zone names imply the nodes of the copula model with the corresponding and the respective order whereas the dashed symbols denote the edges of the first tree of the vine copula model construction.

Following the construction of optimal copula-statistical models, we apply the copula-based Monte Carlo simulation and obtain 2700 simulations (i.e., simulation is repeated in 100 times for the sample size of 27 points) of the marginal returns for each zone from the chosen Gumbel and the vine copula models (Nguyen-Huy et al., 2017, 2018). For the purpose of comparison, the traditional multivariate-normal distribution is also used in this study to generate another a set of simulated data using the Monte Carlo simulation technique. In this case, the marginal returns are assumed to follow a multivariate-normal distribution (i.e., the individual marginal return distributions and their dependences are assumed to be normal). These three sets of randomly simulated data (have been transformed back to the real values) are finally employed in the following geographical diversification analysis and interpretation.

4.3. Mean-CVaR efficient frontiers

This section describes the mean-CVaR optimisations where the expected return of wheat farmer's portfolio are maximised subject to the target risk (CVaR) constraint. Table 6 displays the examples of optimal portfolios at three common confidence levels (i.e., 90%, 95%, and 99%) from copula-based and conventional multivariate normal models. It is noticed that, by definition, the CVaR risk measure evaluates the outcomes versus the zero and, consequently, it is likely to have positive and negative values. The reported values of the positive or greater than zero CVaR (similar to the positive VaR) refers to the certain negative outcomes (i.e., losses), and the negative CVaR correspond to certain positive outcomes (i.e., the gains or the returns). For example, a value of 95% CVaR of 0.10 (a positive value) refers to the scenario that the expected return of the 135 worst scenarios (i.e., 5%*2700) is equal to −10%, and conversely, a value of 95% CVaR of −0.10 (a negative value) refers to the scenarios that the expected return of the 135 worst

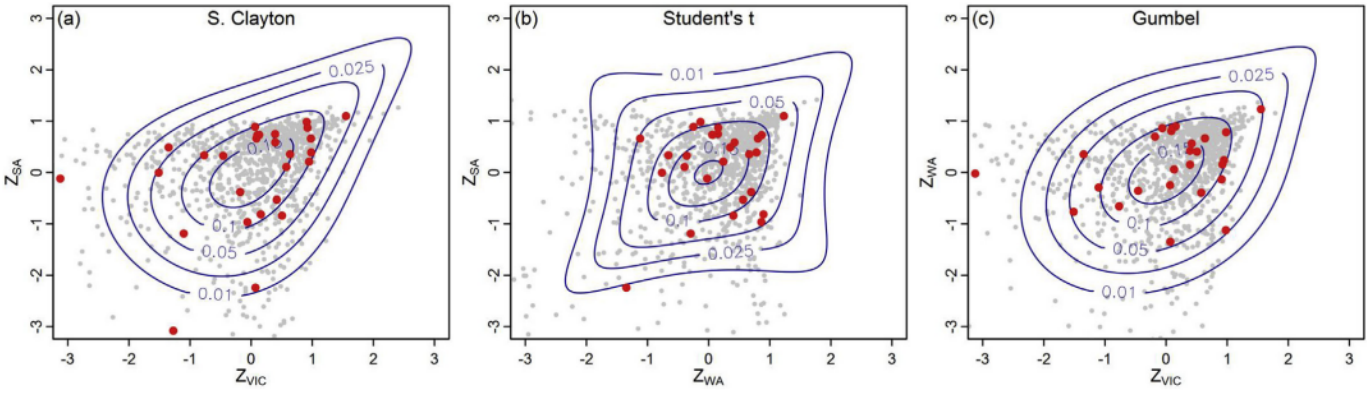


Fig. 4. Contour plots of selected bivariate copulas for each pairs of returns superimposed with standardized empirical observations (red points) and 1000 simulated data (smaller grey points) derived from the corresponding survival Clayton, Student's t, and Gumbel copulas. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 6
Three examples of the optimal portfolios with the conditional value-at-risk (CVaR) and the target returns at 90%, 95%, and 99% confidence levels for the case of the vine, Gumbel, and multivariate-normal (M-Normal) portfolios.

Copula Type	90%		95%		99%	
	Target Return	Mean-CVaR	Target Return	Mean-CVaR	Target Return	Mean-CVaR
C-Vine	0.332	−0.0177	0.332	0.0680	0.332	0.2628
Gumbel	0.332	−0.0029	0.332	−0.0090	0.332	0.2611
M-Normal	0.332	−0.0651	0.332	−0.0229	0.332	0.0520

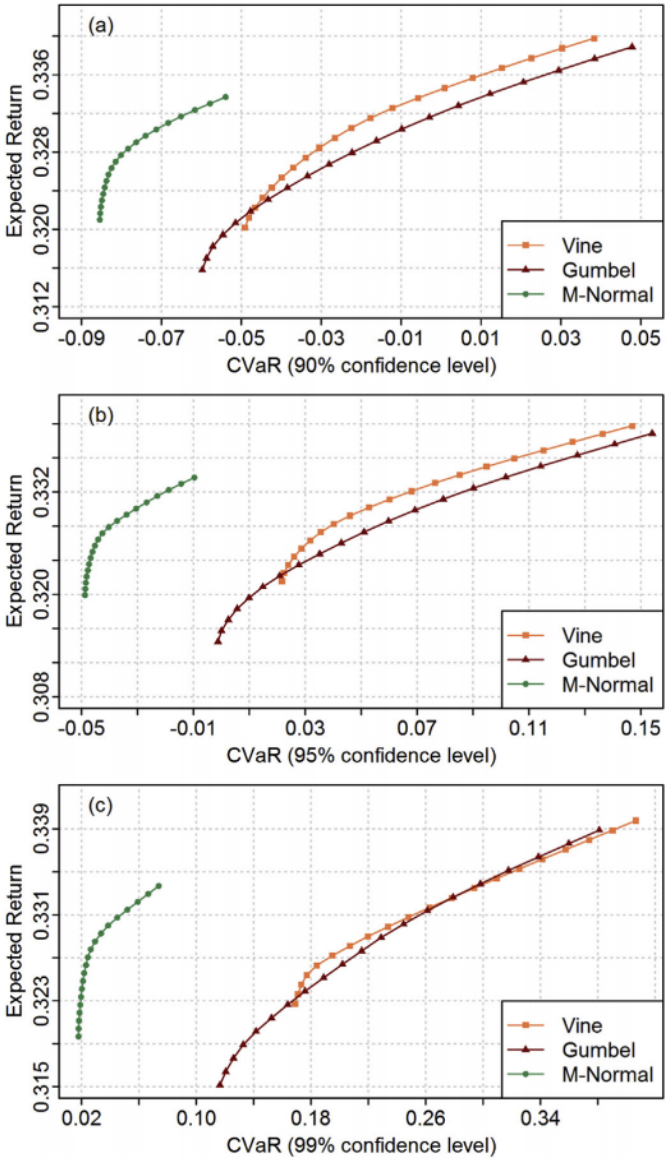


Fig. 5. Mean-CVaR efficient frontiers from the vine, Gumbel, and multivariate-normal (i.e., M-Normal) model at confidence levels of 90%, 95% and 99%.

scenarios is equal to 10%.

In order to compare the optimised values of mean-CVaR under different distribution assumptions, the same targets of the expected returns are selected for each confidence level. The two copula-based portfolios produce a higher mean-CVaR value than the conventional multivariate-normal portfolio. Thus, the results in Table 6 indicate that if the joint distribution of the marginal returns is followed properly by a non-normal distribution modelled as by the copulas, the wheat farmers are likely to underestimate the minimum level of the risk measured by mean-CVaR for a given expected return using the multivariate-normal method. Since the marginal returns clearly do not follow the normal distribution as shown in section 3.1 and Table 3, the risk level should be measured based on the copula model.

The underestimation of risk under the assumption of a multivariate-normal distribution is displayed clearly in Fig. 5. The mean-CVaR efficient frontier acquired from the traditional multivariate-normal portfolio is plotted against those from the copula-based portfolios for different confidence levels. As it can be seen from Fig. 5, the significantly higher values of the frontiers can be observed from the copula-based models compared to the multivariate-normal model. This is because the copula-based models are able to account for the tails dependences whereas the multivariate-normal distribution assumes the coefficient of the tail dependence is zero, and therefore, it ignores the co-movement in the tail of the joint distributions. As such, the portfolio optimisation method relying on the conventional multivariate-normal assumption might be less protective, whereas copula-based models are more appropriate for farmers who are concerned with the extreme losses of their farm profitability.

Regarding the copula-based portfolios, we can infer that the vine copula is able to measure the risk much better than the Gumbel copula for all considered confidence levels. It is because, by the construction method, the vine copula models the dependences of each variable pairs more flexible than the multivariate Archimedean copula (Bedford and Cooke, 2002; Zhang and Singh, 2014). To examine this, we also inspect the preservative capacity of the three model for modelling the dependences among variable pairs. Fig. 6 displays a comparison of simulated and observed rank-based correlation coefficients (Kendall's τ) for the three models. It is clear that the vine approach is able to reserve the dependences of all variable pairs compared to the multivariate Gumbel and multivariate-normal model. Therefore, the Gumbel model may overestimate the risks given the same target expected returns in comparison to the vine model.

The single portfolios of each zone relative to the vine copula-based frontiers are shown in Fig. 7. This figure reveals how risk reduction can be achieved by a geographical diversification strategy. It can be seen that the farmer's profitability currently growing wheat at VIC and WA zones is below the efficient frontiers level whereas those for SA are on the frontier curve. Geographical diversification is likely to improve the profitability in both the VIC and the WA zones, but not in the SA farming area for a given level of downside risk. Growing wheat in SA could, therefore, face the maximum risk since it is located at the highest point of the frontier curve, however, it has the possibility of reaching the highest profitability as well. In addition, in the circumstances, the producers could decide to be slightly less profitability by geographically diversifying in order to reduce a relatively large downside risk. For example, by allocating about 10% of their production area to VIC, wheat producers in the SA region can adjust their expected profitability in the worst 5% of the cases from approximately 33.98%–33.69% (i.e., a reduction of 0.29%), which in turn can reduce the downside risk from approximately 14.70%–11.51% (i.e., a risk reduction of 3.19%). This is because the average marginal return (and the standard deviation) in SA is just 3.67% higher (and 1.09% lower) than in VIC. The kurtosis (and skewness) in the SA region is also 22.59% higher (and 27.06% lower) than that in the VIC region (Table 3). By definition, the kurtosis factor is able to measure whether the data are heavy-tailed or light-tailed

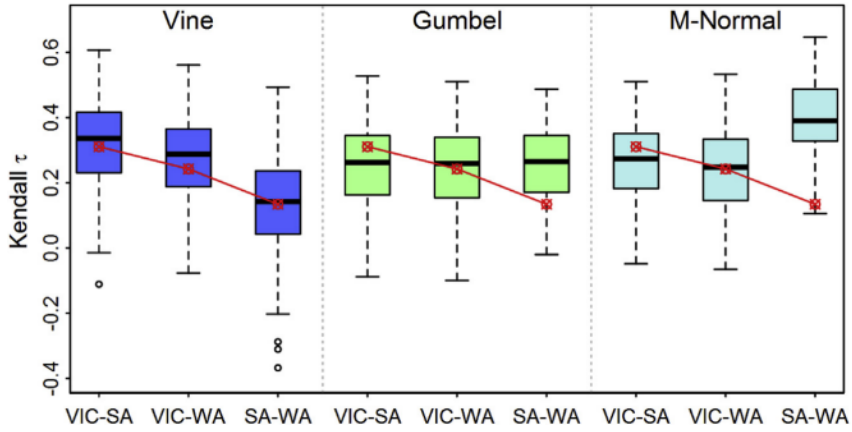


Fig. 6. Comparison of the simulated (in box plots) and the observed (as red points) values of the Kendall's τ for the vine, Gumbel, and multivariate-normal (M-Normal) model. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

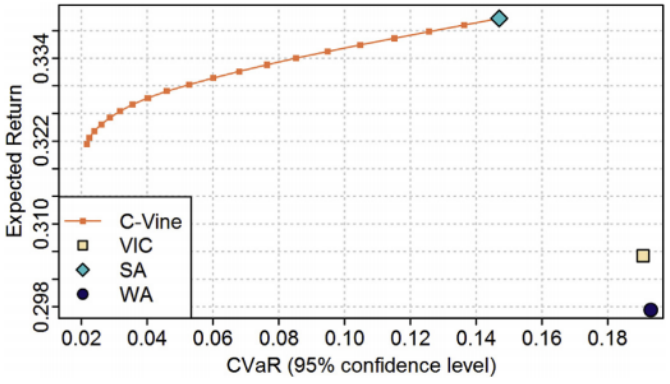


Fig. 7. Mean-CVaR efficient frontiers at the 95% confidence level for the vine copula model and single portfolios.

Table 7
Comparison of equal weight feasible and efficient CVaR portfolios (at 95% confidence level).

Allocation and Risk Level	VIC	SA	WA	Expected return	CVaR
Equal weight feasible portfolio					
Hectare allocation	0.3333	0.3333	0.3333	0.3142	0.0345
Covariance risk budget	0.3702	0.2962	0.3336		
Efficient CVaR portfolio					
Hectare allocation	0.3978	0.3214	0.2808	0.3142	0.0337
Covariance risk budget	0.4603	0.2768	0.2630		

relative to a normal distribution. Thus, we deduce that the SA region is likely to have higher heavy tails or outliers in the lower tail (extreme losses) since the high negative skewness implies the asymmetry to the left of its marginal return distribution.

In accordance with the results, the ratios of the trade-off between target risks and expected returns changes along the efficient frontiers. In contrast to the high targets of the expected returns, the wheat producers can increase their expected returns without exposing themselves to higher risk through the geographical spread of wheat farms at the lower levels of expected profitability. This is possible by balancing the hectares allocated to the SA and VIC regions, and allocating a small part to the WA region. This result is expected in terms of the reasons mentioned above between the SA and VIC zones. Importantly, WA has the lowest average marginal return and the highest kurtosis and (absolute) skewness. Therefore, the major benefits from growing in WA are derived mostly from the low relationship (or opposite co-movement) of the marginal returns with VIC and SA (see Table 1 and Fig. 2).

4.4. Optimal portfolio allocation

In this section, we analyse the optimal percentage allocation among three growing zones. Firstly, we investigate the differences between a feasible portfolio with equal weight (i.e., the total hectare is divided equally into three zones) and an efficient CVaR portfolio. This comparison is performed by specifying the target expected return and then optimising the portfolio which has the lowest risk for both cases. The results illustrated in Table 7 indicate that the risk of the optimised efficient CVaR portfolio has been lowered from 3.45% to 3.37% for the same target return.

We further explore on the optimal hectare allocation with the mean-CVaR efficient frontiers. Fig. 8 represents the efficient allocation (i.e., optimal weight) (a), weighted returns (b), and the covariance risk budgets (c) corresponding to different targets of the mean-CVaR efficient frontiers (for 95% confidence level) for the vine copula-based portfolios. Since the weighted return is the product of the optimal weight (i.e., the hectare allocated) and corresponding marginal return, its value illustrates the proportion of each zone contributing to the expected marginal return. Thus, these figures appear to show a similar pattern to figure (a).

It is clear that the optimal share allocated to each growing zone varies depending on the different expected marginal returns and risk levels. As expected, the optimal decision is to allocate all production to the zone with the highest expected marginal return, i.e., SA in this case, resulting in the maximum risk level. The optimal choice for the minimum CVaR portfolio is to operate in all the three zones with the highest proportion of growing land allocated to SA (50%), followed by VIC (40%) and WA (10%). In order to achieve a medium to high level of expected profitability, wheat should be grown mostly in SA and not at all in WA. It is also optimal to allocate the majority of the land to SA and VIC, and less than 10% to WA when targeting low to medium levels of profitability and risk.

Figs. 9 and 10 are similar to Fig. 8, however, for the confidence levels of 90% and 99%, respectively. It can be seen clearly that the patterns of hectare allocation are different corresponding to the interested confidence levels. For the very worst cases (i.e., at the confidence level of 99%), to optimize the minimum risk, the total hectare should be allocated more in SA (55%) and lesser in WA (5%) since SA has the lowest standard deviation.

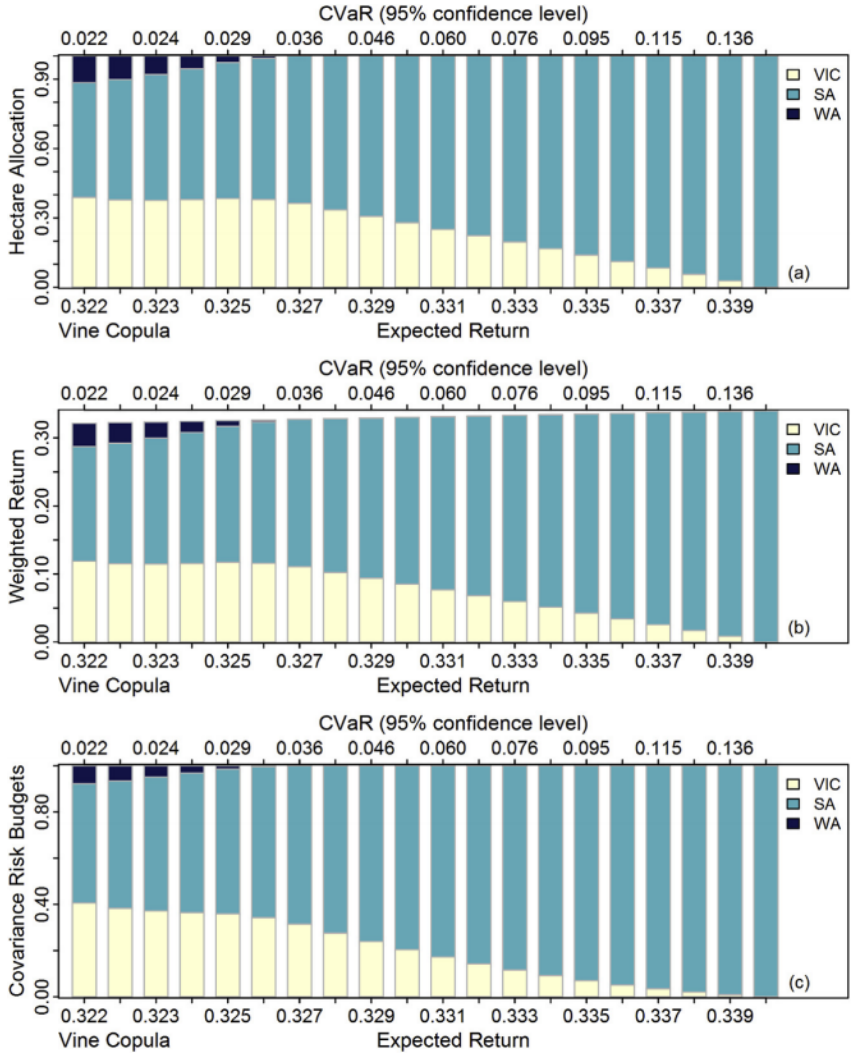


Fig. 8. The percentage of hectare allocation among the three wheat zones at the 95% confidence level for the vine copula-based portfolios.

5. Discussion

It is not surprising that there is an argument on improving the efficiency of diversification strategies in agriculture. In the worst case when a series of weather events are highly correlated, it is obvious there may be no benefit of diversification. According to Mahul (1999), we cannot diversify systemic risk if natural disasters occur concurrently among a large number of farming systems. Some relevance may be drawn from the study of Xu et al. (2010) in Germany, stating that systemic weather risks are not possible to be diversified a regional scale. Based on a study in the United States, Holly Wang and Zhang (2003) stated that a wheat-cropping system can be geographically diversified at the county level. Accordingly, the behavior of systemic weather risks may be different over a global scale because of the differences in geographical topography and climatic conditions (Odening and Shen, 2014). In this study, geographical diversification has been examined as a potentially effective strategy for risk reduction in an Australian farming system. This study is important since portfolio managers can achieve an optimal portfolio with specifically required target risks and expected returns through the proposed copula-based mean-CVaR approach. This can be performed by adjusting the proportion of the total growing hectare to acquire an optimal return-risk trade-off.

In regards to the methodology, the copula-based model is found to be superior to the conventional multivariate-normal approach. It is expected since the distribution of the marginal returns is not normal and our results are in agreement with the study of Larsen et al. (2015). However, while that author applied only the multivariate copulas with lower tail, our study is employed copula functions that have either lower tail or upper tail for more flexible and appropriate description of data dependences. Furthermore, the vine copula is found to be better than the multivariate copula (as used in the study of Larsen et al., 2013) in modelling the dependence structures of the joint distribution by reserving the dependences among variable pairs. This finding reconfirms the advantages of the vine copulas stated in Brechmann (2010) and found by Zhang and Singh (2014).

This study points out several challenges in copula model development that could form the subject of further investigation to address these limitations. One such challenge is that underlying uncertainties in the model that could influence of result when estimating the copula parameters, including the potential sources of error that are derived from data management and model structures generated by a purely statistical approach. This could lead to major issues, where some of the copula parameters may equally fit the statistical goodness-of-fit test (Sadegh et al., 2018; Vrugt et al., 2003) but may in fact carry errors within them to confound the overall accuracy of the simulated data. This problem could also affect the process of finding a unique combination of copula parameters that are considerably superior to the others. Furthermore, one combination of copula parameters may be either be better than the others based on the goodness-of-fit measure or it may be worse in respect to another parameter. For example, if a

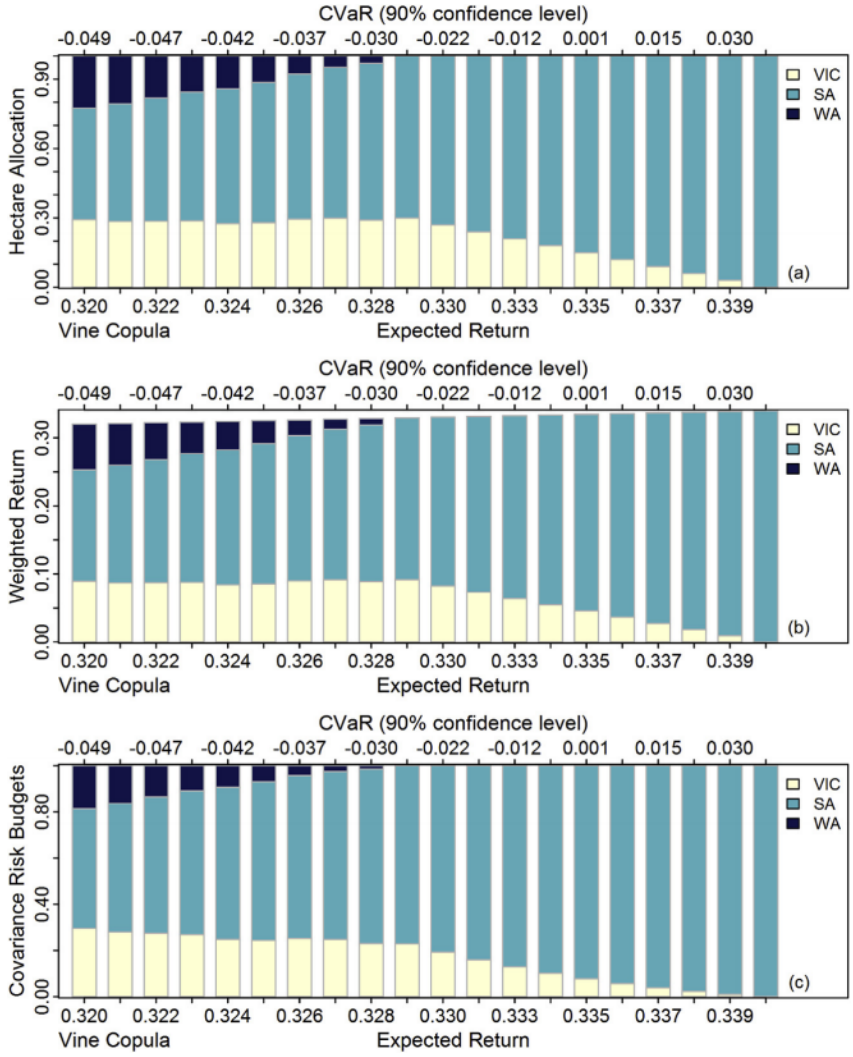


Fig. 9. The percentage of hectare allocation among the three wheat zones at the 90% confidence level for the vine copula-based portfolios.

copula family is selected according to the Bayesian Information Criteria (BIC), the penalty for a two-parameter copula (e.g., Student's t, BB1, BB6, etc.) could be greater than that based on the AIC value (Schepsmeier et al., 2017).

It is also worth noting that the estimation of copula parameters relies on the period of observed data (Nguyen-Huy et al., 2018; Sadegh et al., 2018). This means that the dependence structure between any observations could vary with the time factor, resulting in different selection of copulas for modelling the relationship between the same objects. For example, in our previous study (Nguyen-Huy et al., 2018), the copula combination was different in each k-fold cross-validation process where the dataset was split into different training and testing subsamples. Therefore, the use of an acceptable group of samples to reflect more information about the system behavior is encouraged rather than finding the best parameter combination which is implied as the true representative of the system (Sadegh et al., 2018). In addition, according to Sadegh and Vrugt (2014), choosing the best copula parameter combination may lead to an underestimation of the uncertainties of the entire system. Finally, the limited length of the data can plausibly affect the accuracy of the parameter estimation by increasing the uncertainties (Bevacqua et al., 2017). All these reasons, and others, warrant a further investigation to mitigate the complications in selecting the best copula model as well as the best parameters of the optimal copula function.

The present study also comes with common assumptions that have been reported in published literature. First, this study does not account for the cost of growing crops in different zones (Larsen et al., 2015). Second, it is assumed that the marginal distribution does not change over the passage of time (Sadegh et al., 2018; Sadegh and Vrugt, 2014). Finally, since the statistical model was developed using historical data, this data is not able to account for the scenarios which have not been occurred before. This means the model cannot be easily adjusted to accommodate for the changes in factors such as climate, technology, and cultivation practices. Therefore, in order to achieve more robustness diversification benefits, it is important to incorporate the impacts of all the costs that may occur in geographical distributing the farm system as well as performing the model with under many projected scenarios.

6. Conclusion

In this study, we have demonstrated the effectiveness of applying a geographical diversification strategy to agricultural risk management. The mean-CVaR, the most popular and appropriate measure of downside risk, was calculated using the copula-based approach. Compared to the traditional multivariate-normal model, the copula-statistical approach was able to flexibly model the joint distribution of different types of marginal datasets including those of the non-normal distributions. Furthermore, the study revealed that the vine copula-statistical models were able to capture the full range of different dependence

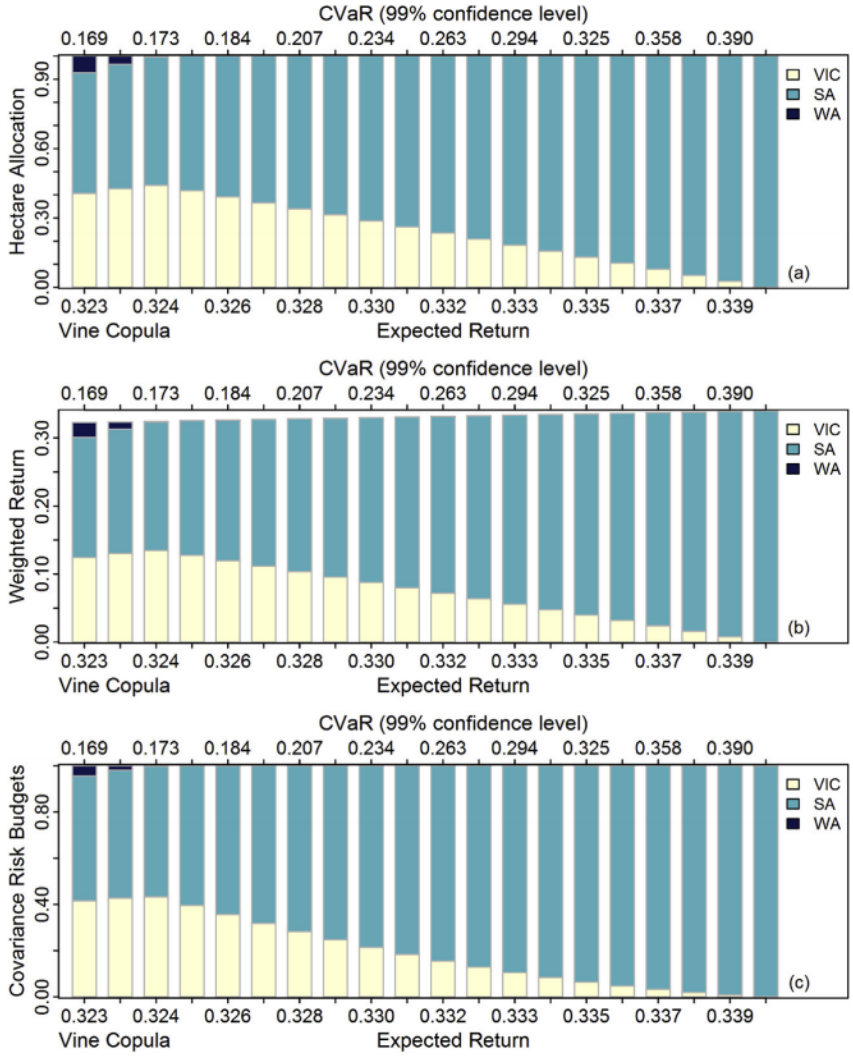


Fig. 10. The percentage of hectare allocation among the three wheat zones at the 99% confidence level for the vine copula-based portfolios.

structures and in particular the case where the joint distribution of marginal returns exhibits the tail dependence, as also revealed in earlier studies on precipitation and wheat yield forecasting (Nguyen-Huy et al., 2017, 2018).

Although the results have useful implications for three major wheat growing zones in Australia including VIC Mallee, SA Eyre Peninsula, and WA North and East Wheat Belt, the approach is applicable to other agricultural regions and crops outside of Australia. This is because the models have a good ability to analyse joint dependences, and able to examine the potential assistance that can be offered to the farmers as part of the optimised geographical strategy in agricultural risk reduction. The approach is fairly justified to be used as a broad method for modelling such problems since the multivariate joint distribution of the marginal returns was constructed by the copula function and then evaluated against the multivariate-normal approach for comparison purposes. To optimize the method, the CVaR criteria were calculated using scenarios from Monte Carlo simulation methods and the portfolio optimisation was attained by maximising the expected marginal return for given target levels of CVaR.

The optimised mean-CVaR results, as described by the corresponding efficient frontier and optimal hectare allocation, indicated that using geographical diversification to downside risk is viable. To be more specific, the risk can be reduced for wheat producers in VIC and WA region since both regions are located below the efficient frontiers. To explain this, we consider SA, which was located on the frontier curve, and therefore meant that zone was able to obtain the least benefit from geographical diversification. Nevertheless, it was also evident that SA was able to gain a relatively large risk reduction by reducing the marginal return in a subtle way from the geographical diversification since it was located at the riskiest point of the frontier curve. In general, three optimal portfolio models in this study showed that the geographical diversification strategy was an achievable tool for agricultural risk modelling and management. However, the optimal share of the hectares allocated to each zone varied depending on the target risk and the profitability that the wheat producers expect.

The results in this paper also indicate the advantages of the copula method in addressing the lower tail dependence of the joint return distribution. That is, if the marginal returns are not normally distributed (as it is the case in this study), the multivariate-normal model is likely to underestimate the minimum level of the downside risk at a given target of expected marginal return by discounting the existence of the lower tail dependence in the model. In this case, the copula approach developed in this paper is more appropriate and can be used to analyse the benefits of the geographical diversification strategy. It was evident that the vine copula performed better than the Gumbel copula since it allowed each variable pairs to be modelled by different copula functions.

Considering the results and their interpretation it is concluded that wheat producers could possibly achieve a higher expected return given the same level of downside risk by dividing the crops among the three

zones. While the results are at a regional scale, the method can be extended to a farm level as well as to the other crops. This study, however, was unable to account for the costs that could possibly occur when growing in different places, a dataset that could add value to the modelling strategy followed in this paper. Thus, a follow-up study could take into account the cost-related components in the performance of geographical diversification strategy. Finally, a potential avenue of future research could also be to consider the spatio-temporal impact of climate conditions on the marginal returns across the different zones.

Acknowledgment

The project was financed by University of Southern Queensland Post Graduate Research Scholarship [USQPRS 2015-2018]; School of Agricultural, Computational and Environmental Sciences and Drought and Climate Adaptation [DCAP] Projects (Producing Enhanced Crop Insurance Systems and Associated Financial Decision Support Tools). The authors would like to acknowledge all constructive comments from the reviewers and the journal Editor-in-Chief.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.wace.2018.07.002.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insur. Math. Econ. 44, 182–198. https://doi.org/https://doi.org/10.1016/j.insmatheco.2007.02.001.

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. J. Financ. Econ. 63, 443–494. https://doi.org/10.1016/S0304-405X(02)00068-5.

Barriopedro, D., Fischer, E.M., Luterbacher, J., Trigo, R.M., García-Herrera, R., 2011. The hot summer of 2010: redrawing the temperature record map of Europe. Science. https://doi.org/10.1126/science.1201224.

Bedford, T., Cooke, R.M., 2002. Vines: a new graphical model for dependent random variables. Ann. Stat. 1031–1068. https://doi.org/https://doi.org/10.1214/aos/1031689016.

Bessa, R.J., Miranda, V., Botterud, A., Zhou, Z., Wang, J., 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renew. Energy 40, 29–39. https://doi.org/https://doi.org/10.1016/j.renene.2011.08.015.

Bevacqua, E., Maraun, D., Haff, I.H., Widmann, M., Vrac, M., 2017. Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy). Hydrol. Earth Syst. Sci. 21, 2701.

Bokusheva, R., 2014. Improving the Effectiveness of Weather-based Insurance: an Application of Copula Approach. University Library of Munich, Germany.

Boubaker, H., Sghaier, N., 2013. Portfolio optimization in the presence of dependent financial returns with long memory: a copula based approach. J. Bank. Finance 37, 361–377.

Bradshaw, B., Dolan, H., Smit, B., 2004. Farm-level adaptation to climatic variability and change: crop diversification in the Canadian prairies. Climatic Change 67, 119–141. https://doi.org/10.1007/s10584-004-0710-z.

Brechmann, E.C., 2010. Truncated and Simplified Regular Vines and Their Applications. Diploma.

Chernozhukov, V., Umantsev, L., 2001. Conditional value-at-risk: aspects of modeling and estimation. Empir. Econ. 26, 271–292. https://doi.org/10.1007/s001810000062.

Chowdhary, H., Escobar, L.A., Singh, V.P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. Nord. Hydrol 42, 193–216. https://doi.org/https://doi.org/10.2166/nh.2011.065.

Coumou, D., Rahmstorf, S., 2012. A decade of weather extremes. Nat. Clim. Chang. https://doi.org/10.1038/nclimate1452.

Curran, P.J., West, S.G., Finch, J.F., 1996. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychol. Methods 1, 16.

Embrechts, P., Lindskog, F., McNeil, A., 2001. Modelling Dependence with Copulas. Rapp. Tech. Département mathématiques, Inst. Fédéral Technol, Zurich, Zurich.

FAO, 2015. The Impact of Natural Hazards and Disasters on Agriculture and Food and Nutrition Security – a Call for Action to Build Resilient Livelihoods.

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002456.

Ganguli, P., Reddy, M.J., 2012. Risk assessment of droughts in Gujarat using bivariate copulas. Water Resour. Manag. https://doi.org/10.1007/s11269-012-0073-6.

Genest, C., Favre, A.-C., 2007. Everything you always wanted to know about Copula modeling but were afraid to ask. J. Hydrol. Eng. 12, 347–368. https://doi.org/https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347).

Goodwin, B.K., Hungerford, A., 2014. Copula-based models of systemic risk in US agriculture: implications for crop insurance and reinsurance contracts. Am. J. Agric. Econ. 97, 879–896.

Goodwin, B.K., Ker, A.P., 2002. Modeling price and yield risk. A Comprehensive Assessment of the Role of Risk in US Agriculture. Springer, pp. 289–323.

Hao, Z., Singh, V.P., 2016. Review of dependence modeling in hydrology and water resources. Prog. Phys. Geogr 0309133316632460.

Herold, N., Ekström, M., Kala, J., Goldie, J., Evans, J.P., 2018. Australian climate extremes in the 21st century according to a regional climate model ensemble: implications for health and agriculture. Weather Clim. Extrem 20, 54–68.

Holly Wang, H., Zhang, H., 2003. On the possibility of a private crop insurance market: a spatial statistics approach. J. Risk Insur. https://doi.org/10.1111/1539-6975.00051.

Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. Appl. Financ. Econ. 16, 717–729.

Huang, J.-J., Lee, K.-J., Liang, H., Lin, W.-F., 2009. Estimating value at risk of portfolio by conditional copula-GARCH method. Insur. Math. Econ. 45, 315–324. https://doi.org/https://doi.org/10.1016/j.insmatheco.2009.09.009.

Janga Reddy, M., Ganguli, P., 2012. Application of copulas for derivation of drought severity-duration-frequency curves. Hydrol. Process. https://doi.org/10.1002/hyp.8287.

Joe, H., 1996. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. Lect. notes-monograph Ser 120–141. https://doi.org/https://doi.org/10.1214/lnms/1215452614.

Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. J. Bank. Finance 31, 2405–2423.

Kresta, A., Tichý, T., 2012. International Equity Portfolio Risk Modeling: the Case of the NIG Model and Ordinary Copula Functions.

Larsen, R., Leatham, D., Sukcharoen, K., 2015. Geographical diversification in wheat farming: a copula-based CVaR framework. Agric. Finance Rev. 75, 368–384.

Larsen, R., Mjelde, W., Klinefelter, D., Wolfley, J., 2013. The use of copulas in explaining crop yield dependence structures for use in geographic diversification. Agric. Finance Rev. 73, 469–492.

Lesk, C., Rowhani, P., Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production. Nature. https://doi.org/10.1038/nature16467.

Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–676. https://doi.org/10.1111/0022-1082.00340.

Mahul, O., 1999. Optimum area yield crop insurance. Am. J. Agric. Econ. https://doi.org/10.1111/j.l467-8276.2009.01368.x.

Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91.

Mishra, A.K., El-Osta, H.S., Sandretto, C.L., 2004. Factors affecting farm enterprise diversification. Agric. Finance Rev. 64, 151–166. https://doi.org/10.1108/00214660480001160.

Mulvey, J.M., Erkan, H.G., 2006. Applying CVaR for decentralized risk management of financial companies. J. Bank. Finance 30, 627–644.

Murray, G.M., Brennan, J.P., 2009. Estimating disease losses to the Australian wheat industry. Australas. Plant Pathol. 38, 558–570.

Nelsen, R.B., 2006. An Introduction to Copulas, second ed. Springer.

Nguyen-Huy, T., Deo, R.C., An-Vo, D.-A., Mushtaq, S., Khan, S., 2017. Copula-statistical precipitation forecasting model in Australia's agro-ecological zones. Agric. Water Manag. 191. https://doi.org/10.1016/j.agwat.2017.06.010.

Nguyen-Huy, T., Deo, R.C., Mushtaq, S., An-Vo, D.-A., Khan, S., 2018. Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach. Eur. J. Agron. 98. https://doi.org/10.1016/j.eja.2018.05.006.

Odening, M., Shen, Z., 2014. Challenges of insuring weather risk in agriculture. Agric. Finance Rev. 74, 188–199.

Okhrin, O., Odening, M., Xu, W., 2013. Systemic weather risk and crop insurance: the case of China. J. Risk Insur. 80, 351–372.

Pflug, G.C., 2000. Some remarks on the value-at-risk and the conditional value-at-risk. Probabilistic constrained Optim. Methodol. Appl 272–281. https://doi.org/10.1007/978-1-4757-3150-7_15.

Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24 n/a-n/a. https://doi.org/10.1029/2008GB003435.

R Core Team, 2016. R development Core Team. R A Lang. Environ. Stat. Comput. https://doi.org/http://www.R-project.org.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. J. Risk 2, 21–42.

Sadegh, M., Moftakhari, H., Gupta, H.V., Ragno, E., Mazdiyasni, O., Sanders, B., Matthew, R., AghaKouchak, A., 2018. Multi-hazard scenarios for analysis of compound extreme events. Geophys. Res. Lett.

Sadegh, M., Vrugt, J.A., 2014. Approximate bayesian computation using Markov chain Monte Carlo simulation. Water Resour. Res. https://doi.org/10.1002/2014WR015386.Received.

Schepsmeier, U., Stoeber, J., Brechmann, E.C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., 2017. Package 'VineCopula.'.

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Université Paris, pp. 8.

Turner, N.C., 2004. Agronomic options for improving rainfall-use efficiency of crops in dryland farming systems. J. Exp. Bot. 55, 2413–2425. https://doi.org/https://doi.org/10.1093/jxb/erh154.

Vedenov, D., 2008. Application of copulas to estimation of joint crop yield distributions. In: American Agricultural Economics Association Annual Meeting, pp. 27–29 Orlando, FL.

Vergni, L., Todisco, F., Mannocchi, F., 2015. Analysis of agricultural drought characteristics through a two-dimensional copula. Water Resour. Manag. https://doi.org/10.1007/s11269-015-0972-4.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. https://doi.org/10.1029/2002WR001642.
Würtz, D., Chalabi, Y., Chen, W., Ellis, A., 2009. Portfolio Optimization with R/Rmetrics. Rmetrics.
Xu, W., Filler, G., Odening, M., Okhrin, O., 2010. On the systemic nature of weather risk. Agric. Finance Rev. 70, 267–284.
Yan, J., 2007. Enjoy the joy of copulas: with a package copula. J. Stat. Software 21, 1–21.
Zhang, L., Singh, V.P., 2014. Trivariate flood frequency analysis using discharge time series with possible different lengths: cuyahoga river case study. J. Hydrol. Eng. 19, 5014012. https://doi.org/https://doi.org/10.1061/(ASCE)HE.1943-5584.0001003.
Zhu, S., Fukushima, M., 2009. Worst-case conditional value-at-risk with application to robust portfolio management. Oper. Res. 57, 1155–1168. https://doi.org/10.1287/opre.1080.0684.

Chapter 6

Synthesis and Conclusions

The present study has conducted a consecutive analysis on the compound influences of synoptic-scale atmospheric-oceanic modes on weather variables and agricultural production in Australia. In particular, the climatic impact on the precipitation variability is investigated first followed by a research of such influence on wheat yield. These studies have developed probabilistic forecast models, which can provide advanced information of precipitation and wheat yield with sufficient time ahead through observing climate conditions. The problems of weather systemic risk associated with widespread extreme events and potential adaptation strategies are implemented in subsequent studies. The findings of this study can therefore provide useful information and powerful tools to producers and risk managers. However, the study, with limited time, retains several challenges that need to be investigated further. This conclusion will briefly represent the difficulties faced in developing copula-based models. The important results and significant contributions of this study are summarised in the next subsections. Finally, potentially interesting paths for future research are suggested at the end of this chapter.

6.1 Challenges in the development of copula-based models

Copula technique has been proved as a useful tool for studying the joint distribution among random variables. However, our published studies always contain a discussion section where the difficulties and limitations experienced during the research implementation are discussed in more details. This section summarises several challenges in both developments and applications of copula models that could potentially form interesting subjects for further investigation to address these limitations in the future.

In general, the development of copula-based models include two separate parts: fitting marginal distributions; and fitting copula functions. Both procedures can be processed through parametrical or non-parametrical methods (Kraus and Czado 2017). Therefore, one such challenge is that underlying uncertainties in the model

when estimating the marginal distributions and copula parameters that could influence the result. These uncertainties include the potential sources of error that are derived from data management, estimating methods and model structures generated by a purely statistical approach (*i.e.*, goodness-of-fit tests).

In regard to the error related to the data management, there is the fact that the estimation of marginal distributions or copula parameters relies on the period of observed data (Nguyen-Huy et al. 2018; Sadegh et al. 2018). This means that the dependence structure within univariate or between variables could vary with the change of time factor, resulting in different selection of marginal distributions and copulas for modelling the relationship between the same objects. For example, in our previous study (Nguyen-Huy et al. 2018) where the dataset was split into different training and testing subsamples, the selected copula combination was different in each k-fold cross-validation process. Therefore, the use of an acceptable group of samples to reflect more information about the system behaviour is encouraged rather than finding the best parameter combination which is implied as the true representative of the system (Sadegh et al. 2018). In addition, according to Sadegh and Vrugt (2014), choosing the best copula parameter combination may lead to an underestimation of the uncertainties of the entire system. Finally, the limited length of the data can plausibly affect the accuracy of the parameter estimation by increasing the uncertainties (Bevacqua et al. 2017).

Observations can be parametrically fitted to univariate distributions (margins) using maximum likelihood, moment matching, quantile matching or maximizing goodness-of-fit estimation (or minimizing distance estimation) methods (Cullen et al. 1999; Delignette-Muller and Dutang 2015; Venables and Ripley 2013; Vose 2008). Clearly, the best fitting distribution selected for a variable may be different depending on the method used. Furthermore, marginal distributions can be fitted non-parametrically using the kernel density estimator (Duong 2016; Geenens 2014; Geenens and Wang 2018; Sheather and Jones 1991). However, this method also relies on selection of the density function (*e.g.*, Gaussian kernel), plug-in bandwidth parameter, lower and upper bound, and degree of the polynomial (*e.g.*, log-constant, log-linear or log-quadratic fitting) (Nagler 2017; Nagler 2018). These selections may lead to different results of the marginal fitting process, and thus contribute to the uncertainty of final models.

Copula parameter can be estimated using different approaches ranging from fully parametric, semi-parametric to non-parametric methods, which potentially generate similar problems, mentioned above for fitting marginal distributions. The first method estimates the copula parameter jointly with the parameters of margins, known as exact maximum likelihood (ML) (or one-stage parametric maximum likelihood) (Cherubini et al. 2004). This method assumes that models are correctly specified and thus the log-likelihood function is correct. By contrast, if models are misspecified leading to the incorrect log-likelihood, then the maximizer is not the maximum likelihood and thus it may lose its desirable status (Kim et al. 2007). Alternatively, copula parameters can be estimated in a two-stage (or multi-stage) procedure called the inference function for margins (IFM) method (Joe 1997; Joe 2005). In the IFM method, margins are first identified separately and then copula parameters are estimated by treating the given parameters of the margins in the log-likelihood and afterwards maximizing the resulting function. This IFM technique is also a fully parametric method, incorrect estimate of margins may thus has an influence on the performance of the estimator (Kim et al. 2007). Furthermore, copula parameters can be estimated semi-parametrically using the pseudo maximum likelihood (PML) method as proposed by Genest et al. (1995), also known as canonical maximum likelihood (Cherubini et al. 2004; McNeil et al. 2005). The implementation of the PML method is similar to the IFM but the difference lies in the fact that the margins are estimated non-parametrically through empirical distribution functions of samples. While this method is not restricted by specific forms of parametric margins, it is generally not as efficient as the ML technique (Genest et al. 1995). Also, the PLM method is asymptotically efficient under a specific condition proposed by Genest and Werker (2002) but not to all common parametric copula models. Finally, copulas can be estimated non-parametrically based on kernel methods (Fermanian 2005; Gijbels and Mielniczuk 1990). However, the non-parametric estimation of copulas faces some challenges such as the boundary bias related to kernel curve estimation, and smooth factors and bandwidth selection (Chen and Huang 2007). While the comparison between these methods is not a part of this discussion, it is clear that different estimators may result in different copula selections, which is likely to be another source of error to the final model.

A purely statistical approach can lead to major issues, where some of the copula parameters may equally fit the goodness-of-fit test (Sadegh et al. 2018; Vrugt et al. 2003) but may in fact carry errors within them due to the estimator, and thus the overall accuracy of the simulated data is confounded. This problem can also have impacts on the process of finding a unique combination of copula parameters, which are noticeably better to the others. Additionally, one combination of copula parameters may either be superior to others based on a statistical goodness-of-fit test or inferior in regard to another measure. For example, when a copula family is chosen using the Bayesian Information Criteria (BIC), the penalty for two-parameter copulas (*e.g.*, Student's t, BB1, BB6, etc.) may be larger than that based on the AIC value (Schepsmeier et al. 2018). All these reasons require a further examination to diminish the complications in the selection of the best copula model along with the best set of parameters of the optimal copula function.

The construction of vine copulas, which is based on ordered bivariate conditional copulas, can also contribute to the model uncertainty. For example, an *n*-dimensional multivariate random vector can be constructed *n(n-1)/2* unique C- and D-vine copulas (Aas et al. 2009), it is required to optimise at least the same number of parameters where good starting values are important for the optimisation process. Aas et al. (2009) suggest a sequential estimation algorithm for the pair-copula constructions which is further investigated in Min and Czado (2010) and Smith et al. (2010). However, it is possible to construct a significantly large number $\binom{n}{2}\times(n-2)!\times 2^{\binom{n-2}{2}}$ of R-vine copulas (Morales Napoles et al. 2010). Dissmann et al. (2013) develops an automated algorithm to select the best R-vine structure using a maximum spanning tree method based on edge weights such as the absolute value of empirical Kendall's tau or Spearman's rho, AIC or corrected AIC, or BIC. Clearly, the selection of vine structures varies depending on the edge weight used.

In short, a number of error sources may contribute to the uncertainty of statistical copula-based models. While the importance of good copula choice has been noted in the studies undertaken by (Garcia and Tsafack 2011; Woodard et al. 2011), research on selection and comparisons and of the best copula models may be complicated in relation to the reasons and uncertainties which have been discussed above. Developing such algorithms will be a promising direction for future research.

6.2 Summary of Important Findings

This study found that the synoptic-scale atmospheric-oceanic circulation patterns have an important role on the variability of precipitation in Australia. In particular, the results indicate that there is a significant lagged correlation between average values of June – August Southern Oscillation Index (SOI) and Inter-decadal Pacific Oscillation (IPO) Tripole Index (TPI) with spring seasonal precipitation (September – November) across wheat belt zones, except for western regions. In general, the SOI-precipitation relationship is stronger in the upper tail implying that the influence of SOI during a La Niña event is dominant to that during an El Niño event. Furthermore, the associations between the two drivers and precipitation are opposite meaning that SOI has a positive relationship and TPI a negative relationship with spring precipitation in most of the Agro-ecological zones. However, IPO is known to modulate the frequency and impact of ENSO on precipitation variability in Australia; in particular there is a co-occurrence of a negative IPO phase and a La Niña event (Kiem and Franks 2004; Power et al. 1999). This study also found that the inclusion of TPI into the bivariate forecast models (*i.e.*, models use SOI to forecast precipitation) may have an impact on the present models but this modulation is not the same for different locations and phases. For example, when TPI is added into the forecast models, the Spearman's rank coefficients between simulated and observed samples are improved for the zones in the east and south-east but reduced in the north-west. The results also found that the correlation coefficients in the upper right quadrant (*i.e.*, impact of negative IPO phases and La Niña events) are greatly improved compared to others. The findings of this study are in agreement with previous studies published in the literature (Kiem and Franks 2004; Power et al. 1999; Verdon et al. 2004).

This study has explored the individual and compound influences of large-scale climate drivers on the Australian wheat yield over different locations and times. The dominant role of the Indian Ocean in recent years in the inter-annual variability of the wheat yield compared to the Pacific and tropical regions has been confirmed in regard to previous work (Yuan and Yamagata 2015). The statistically significant correlation coefficients between wheat yield anomalies and averaged values over March – May of Dipole Mode Index (DMI) obtained from Indian region reveal that wheat yield can be forecasted at a very early stage (*i.e.*, planting stage) in Queensland and New South

Wales. The climate drivers derived from the Pacific Ocean show more influences on wheat crops grown in the east (*i.e.*, Queensland) while the tropical and extra-tropical synoptic-scale climate indices have an impact mostly on the south and south-east. The results also confirm that the co-occurrence of extreme climatic events may increase the variability of the wheat yield. Vine technique, as expected, is found to be superior to other copula approaches in modelling the tail dependencies. Furthermore, the D-vine copula-based quantile regression model provides better accuracy of wheat yield forecast compared to the traditional linear quantile regression (LQR) method. The forecast information and developed models indicate noteworthy applications for improving food security by supporting an earlier planning of agricultural strategies and policies to optimise the profits of wheat crops.

The study has evaluated the joint losses of a hypothetical drought index insurance caused by systemic weather risks in different aggregation scales through the C-vine copula-based model. The spatial and temporal interdependencies between weather events are investigated in different locations and seasons. The construction of C-vine copulas is found to be particularly useful for modelling different dependency structures between weather events and analysing systemic insurance losses. The findings reveal that both spatial and temporal diversification strategies possibly diminish the systemic weather risk in Australia. The dependencies between insurance payouts in different locations and seasons becomes smaller when decreasing the strike level. Furthermore, the precipitation deficiency in different seasons is found to be not perfectly correlated, therefore the systemic weather risk diversified over seasons may be more effective than over space. While insurance has been important in the agricultural sector as one of the formal risk-mitigation instruments, weather index-based products have been considered a feasible alternative to overcome restrictions of the traditional crop insurance. Although the results of this study may not be a direct insurance product for farmers due to the problems related to the high basis risk, the assessment of systemic weather risks is an attractive topic to insurers for several reasons. First, covariate yield risk is largely driven by weather conditions. Second, weather derivatives could be beneficial for transferring systemic risk from insurers to reinsurers or to the capital market (Xu et al. 2010). This application answers the question of to what extent systemic drought risk exposure in different locations and seasons can be diversified spatially or temporally by increasing the aggregation level

of the insurance contracts. The results of this study therefore provide an efficient instrument for climate-related risk management and support pricing of weather index-based insurance products.

This study has demonstrated the effectiveness of geographical diversification strategy in mitigating weather-related risks in agriculture through a combination of the two powerful techniques in the fields of finance and statistics, which are Conditional Value-at-Risk (CVaR), and copulas. The optimised mean-CVaR results derived from copula models confirm that using geographical diversification to downside risk is feasible as indicated by the corresponding efficient frontier and optimal hectare allocation. In particular, the risk can be mitigated for wheat producers in Victoria (VIC) and Western Australia (WA) since both areas are located below the efficient frontiers. By contrast, the farms in South Australia (SA) are located on the frontier curve, indicating that they may achieve the least benefit from geographical diversification. However, it is found that SA may obtain a relatively large risk reduction by decreasing the marginal return in a subtle way from the geographical diversification because it is located at the riskiest point of the frontier curve. Three optimal portfolio models in this study generally reveal that the geographical diversification strategy is a feasible instrument for agricultural risk management. However, the optimal share of the total hectares allocated to each location may vary depending on the profitability and target risk that wheat producers expect. The findings of this paper also indicate the benefits of the vine copula approach in modelling the lower tail dependence of the joint return distribution. That means when the marginal returns are not normally distributed (as it is the case in this study), the traditional multivariate-normal model may underestimate the minimum level of the downside risk at a given target of expected marginal return by omitting the existence of the lower tail dependence in the model. We conclude that wheat producers can possibly reach a higher expected return given the same level of downside risk by distributing the crops across the three location as in this study. Although the results are found on a regional scale, the method can potentially be extended to a shire level, and to other crops and locations.

In regard to the methodology, copula-based models, compared to other approaches, exhibit many advantages in modelling joint distributions of high-dimensional multivariate time series. Copula technique allows modelling joint

distributions among multiple variables where each marginal separately with their dependence structure. In addition, copula parameters can be estimated using standard maximum likelihood procedures. In short, copula-based models provide a unit-free and non-parametric tool for measuring dependence and are free of influences of the marginal distributions and linear correlation. Copula models can provide further advantages by enabling the estimate of the multivariate joint probability showing different strengths across the joint distribution and conditional probability. Among copula models, the construction of high-dimensional copulas using the vine technique shows prevails over others by decomposing joint distributions of multivariate variables into a conditional consequence of bivariate copulas, also known as pair-copulas. It is noticed that all pair-copulas can be any bivariate copula family, and therefore the full dependence structure comprising of asymmetries and tail dependencies can be taken into consideration via vine models. For example, a vine copula-based model can involve the pair of Clayton and Gumbel copulas accounting for strong left- and right-tail dependence among variables, respectively, into the joint multivariate distribution. Therefore, vine copulas provides much flexibility in modelling practical and high-dimensional data encountered in many fields.

6.3 Significance and Scientific Contribution of the Study

This research has firstly apply the copula theorem and the respective models for investigating the compound impact of climate mode indices on precipitation and wheat crop in Australia. The results can be used to verify critical assumptions of the linear and other forms of dependencies such as the symmetric relationship between synoptic-scale climate indices and precipitation or what yield. The study also provides value-based information for agricultural risk reduction and insurance contracts by applying better tools to measure the dependence between multiple random variables. The study also represents a more adequate methodological framework for modelling tail dependencies of multivariate distributions since such studies have not been performed in the present study region. The comprehensive analyses of systemic weather risk and potential adaptation strategies are expected to support the development of climate index-based insurance products. Therefore, the knowledge achieved from this study will be important for better understanding the influences of the climate mode indices on precipitation and wheat yield in Australia. The method evolved from this study will

be applicable to either agricultural planning or insurance product designing and may be extended to other regions or crops.

6.4 Recommendations for Future Works

The statistical copula-based framework has been successfully developed in this study for investigating the influence of climate conditions on precipitation and wheat crop in Australia. However, the results need to be interpreted with careful attention since the applications share common assumptions with the published literature. Furthermore, the present study faced several challenges on the uncertainty in developing copula models mentioned above that have not been solved completely due to limited time. Therefore, the following are proposed, which could be promising directions, for future research to enhance the understanding of climate impact and extent the use of copula methods:

- *Interactions between large-scale and local climatic drivers*: the influence of synoptic-scale drivers on weather variables such as precipitation has been demonstrated in literature and in this study. However, regional synoptic patterns such as cut-offlows and easterly dips may have a strong influence on the precipitation in small areas (Verdon-Kidd and Kiem 2009). Future research may apply the copula technique to model the teleconnection between large-scale and local climatic factors to downscale the precipitation forecast model.
- *Development of hybrid models*: the common limitation of statistical models is that the simulations acquired from developed models are based on the joint behaviour between variables recorded in the history. This implies that these relationships are stable in future but this assumption may not be always hold in practice. Furthermore, statistical models are not able to explain the underlying mechanism of the interaction between variables, and therefore they are not easy to justify in the future. Clearly, it is desirable to develop a model that can integrate statistical techniques and eco-bio-physical equation in a hybrid models.
- *Reduction of uncertainties in the development of copula-based models*: The construction of a copula-based model includes two separate parts, namely marginal distributions and copula functions. As mentioned above, both

procedures may have errors in their estimating processes by either parametric or non-parametric method. Therefore, it is essential to have more attempts in the future to mitigate these uncertainties in copula models, for example, comparison between estimation methods or optimisation of final models using the most robust statistical test.

List of References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence Insurance: Mathematics and economics 44:182-198 doi:https://doi.org/10.1016/j.insmatheco.2007.02.001

ABARES (2011b) Australian crop report. ABARES, Canberra

Abdul Rauf UF, Zeephongsekul P (2013) Copula based analysis of rainfall severity and duration: a case study Theoretical and Applied Climatology 115:153-166 doi:10.1007/s00704-013-0877-1

AghaKouchak A, Bárdossy A, Habib E (2010) Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula Advances in Water resources 33:624-634

Anderson W, Seager R, Baethgen W, Cane M (2017) Crop production variability in North and South America forced by life-cycles of the El Niño Southern Oscillation Agricultural and Forest Meteorology 239:151-165 doi:10.1016/j.agrformet.2017.03.008

Ashok K, Guan Z, Yamagata T (2003) Influence of the Indian Ocean Dipole on the Australian winter rainfall Geophysical Research Letters 30 doi:https://doi.org/10.1029/2003GL017926

Asseng S, Foster I, Turner NC (2011) The impact of temperature variability on wheat yields Global Change Biology 17:997-1012 doi:10.1111/j.1365-2486.2010.02262.x

Bannayan M, Crout NMJ, Hoogenboom G Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom. In, 2003. pp 114-125. doi:10.2134/agronj2003.0114

Barlow M, Wheeler M, Lyon B, Cullen H (2005) Modulation of daily precipitation over southwest Asia by the Madden-Julian oscillation Monthly weather review 133:3579-3594

Barnett BJ Agricultural index insurance products: strengths and limitations. In: Agricultural Outlook Forum, 2004. Citeseer,

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the temperature record map of Europe Science 332:220-224

Bedford T, Cooke RM (2001) Probability density decomposition for conditionally dependent random variables modeled by vines Annals of Mathematics and Artificial intelligence 32:245-268 doi:https://doi.org/10.1023/A:1016725902970

Bedford T, Cooke RM (2002) Vines: A new graphical model for dependent random variables Annals of Statistics:1031-1068 doi:https://doi.org/10.1214/aos/1031689016

Best P, Stone R, Sosenko O Climate risk management based on climate modes and indices-the potential in Australian agribusinesses. In: 101st Seminar, July 5-6,

2007, Berlin Germany, 2007. vol 9257. European Association of Agricultural Economists,

Bevacqua E, Maraun D, Haff IH, Widmann M, Vrac M (2017) Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy) Hydrology and Earth System Sciences 21:2701

Biscoe PV, Gallagher JN (1977) Weather, dry matter production and yield Environmental Effects on Crop Physiology:75-100

Bokusheva R (2011) Measuring dependence in joint distributions of yield and weather variables Agricultural Finance Review 71:120-141 doi:https://doi.org/10.1108/00021461111128192

Brechmann EC (2010) Truncated and simplified regular vines and their applications. Diploma thesis, University of Technology, Munich, Germany. URL http://mediatum. ub. tum. de/doc/1079285/1079285. pdf

Brechmann EC, Hendrich K, Czado C (2013) Conditional copula simulation for systemic risk stress testing Insurance: Mathematics and Economics 53:722-732 doi:https://doi.org/10.1016/j.insmatheco.2013.09.009

Brechmann EC, Schepsmeier U (2013) Modeling dependence with C-and D-vine copulas: The R-package CDVine Journal of Statistical Software 52:1-27

Cai W, Cowan T, Sullivan A (2009) Recent unprecedented skewness towards positive Indian Ocean Dipole occurrences and its impact on Australian rainfall Geophysical Research Letters 36

Ceglar A, Turco M, Toreti A, Doblas-Reyes FJ (2017) Linking crop yield anomalies to large-scale atmospheric circulation in Europe Agricultural and Forest Meteorology 240-241:35-45 doi:10.1016/j.agrformet.2017.03.019

Chan SC, Behera SK, Yamagata T (2008) Indian Ocean dipole influence on South American rainfall Geophysical Research Letters 35

Chen SX, Huang TM (2007) Nonparametric estimation of copula functions for dependence modelling Canadian Journal of Statistics 35:265-282

Cherubini U, Luciano E, Vecchiato W (2004) Copula methods in finance. John Wiley & Sons,

Coumou D, Rahmstorf S (2012) A decade of weather extremes Nature climate change 2:491 doi:https://doi.org/10.1038/nclimate1452

Cullen AC, Frey HC, Frey CH (1999) Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. Springer Science & Business Media,

da Costa Dias A (2004) Copula inference for finance and insurance. Diss., Mathematische Wissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 15283, 2004

Delignette-Muller ML, Dutang C (2015) fitdistrplus: An R package for fitting distributions Journal of Statistical Software 64:1-34

Deo RC, Byun H-R, Adamowski JF, Begum K (2017) Application of effective drought index for quantification of meteorological drought events: a case study in

Australia Theoretical and Applied Climatology 128:359-379 doi:https://doi.org/10.1007/s00704-015-1706-5

Deo RC, Şahin M (2016) An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland Environmental monitoring and assessment 188:90 doi:https://doi.org/10.1007/s10661-016-5094-9

Dissmann J, Brechmann EC, Czado C, Kurowicka D (2013) Selecting and estimating regular vine copulae and application to financial returns Computational Statistics & Data Analysis 59:52-69 doi:https://doi.org/10.1016/j.csda.2012.08.010

Donald A et al. (2006) Near-global impact of the Madden-Julian Oscillation on rainfall Geophysical Research Letters 33

Duncan J, Myers RJ (2000) Crop insurance under catastrophic risk American Journal of Agricultural Economics 82:842-855 doi:https://doi.org/10.1111/0002-9092.00085

Duong T (2016) Non-parametric smoothed estimation of multivariate cumulative distribution and survival functions, and receiver operating characteristic curves Journal of the Korean Statistical Society 45:33-50 doi:https://doi.org/10.1016/j.jkss.2015.06.002

Embrechts P, McNeil A, Straumann D (2002) Correlation and dependence in risk management: properties and pitfalls Risk management: value at risk and beyond 176223

England MH, Ummenhofer CC, Santoso A (2006) Interannual rainfall extremes over southwest Western Australia linked to Indian Ocean climate variability Journal of Climate 19:1948-1969 doi:10.1175/JCLI3700.1

Fang Y, Madsen L (2013) Modified Gaussian pseudo-copula: Applications in insurance and finance Insurance: Mathematics and Economics 53:292-301

FAO (2015) The Impact of Natural Hazards and Disasters on Agriculture and Food and Nutrition Security – A Call for Action to Build Resilient Livelihoods Food and Agriculture Organization of the United Nations

Favre A-C, El Adlouni S, Perreault L, Thiémonge N, Bobée B (2004) Multivariate hydrological frequency analysis using copulas Water Resources Research 40 doi:10.1029/2003WR002456

Fermanian J-D (2005) Goodness-of-fit tests for copulas Journal of multivariate analysis 95:119-152

Fischer EM, Knutti R (2013) Robust projections of combined humidity and temperature extremes Nature Climate Change 3:126

Frees EW, Valdez EA (1998) Understanding relationships using copulas North American actuarial journal 2:1-25

Frees EW, Valdez EA (2008) Hierarchical insurance claims modeling Journal of the American Statistical Association 103:1457-1469

French R, Schultz J (1984) Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate Crop and Pasture Science 35:743-764

Frey R, McNeil AJ (2003) Dependent defaults in models of portfolio credit risk Journal of Risk 6:59-92 doi:https://doi.org/10.21314/JOR.2003.089

Frey R, McNeil AJ, Nyfeler M (2001) Copulas and credit models Risk 10

Garcia R, Tsafack G (2011) Dependence structure and extreme comovements in international equity and bond markets Journal of Banking & Finance 35:1954-1970

Geenens G (2014) Probit transformation for kernel density estimation on the unit interval Journal of the American Statistical Association 109:346-358

Geenens G, Wang C (2018) Local-likelihood transformation kernel density estimation for positive random variables Journal of Computational and Graphical Statistics doi:https://doi.org/10.1080/10618600.2018.1424636

Genest C, Favre A-C (2007) Everything you always wanted to know about Copula modeling but were afraid to ask Journal of Hydrologic Engineering 12:347-368 doi:https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest C, Favre AC, Béliveau J, Jacques C (2007) Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data Water Resources Research 43 doi:10.1029/2006WR005275

Genest C, Ghoudi K, Rivest L-P (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions Biometrika 82:543-552

Genest C, Werker BJ (2002) Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In: Distributions with given marginals and statistical modelling. Springer, pp 103-112

Ghosh S, Woodard JD, Vedenov DV Efficient estimation of copula mixture model: an application to the rating of crop revenue insurance. In: 2011 Annual Meeting, July 24-26, 2011, Pittsburgh, Pennsylvania, 2011. vol 103738. Agricultural and Applied Economics Association,

Gijbels I, Mielniczuk J (1990) Estimating the density of a copula function Communications in Statistics-Theory and Methods 19:445-464

Goodwin BK, Hungerford A (2014) Copula-based models of systemic risk in US agriculture: implications for crop insurance and reinsurance contracts American Journal of Agricultural Economics 97:879-896

Gool Dv (2009) Climate Change Effects on WA Grain Production. Western Australia, Department of Agriculture and Food,

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 90:116-133 doi:https://doi.org/10.1016/j.advwatres.2016.02.003

Grimaldi S, Serinaldi F (2006) Asymmetric copula in multivariate flood frequency analysis Advances in Water Resources 29:1155-1167

Gutierrez L (2017) Impacts of El Niño-Southern Oscillation on the wheat market: A global dynamic analysis PLoS ONE 12 doi:10.1371/journal.pone.0179086

Hao Z, Singh VP (2016) Review of dependence modeling in hydrology and water resources Progress in Physical Geography:0309133316632460

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset International Journal of Climatology 34:623-642 doi:10.1002/joc.3711

Hassani BK (2016) Dependencies and Relationships Between Variables. In: Scenario Analysis in Risk Management. Springer, pp 141-158. doi:https://doi.org/10.1007/978-3-319-25056-4_11

Hendon HH, Thompson DW, Wheeler MC (2007) Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode Journal of Climate 20:2452-2467

Herold N, Ekström M, Kala J, Goldie J, Evans J (2018) Australian climate extremes in the 21st century according to a regional climate model ensemble: Implications for health and agriculture Weather and Climate Extremes 20:54-68

Hochman Z, Holzworth D, Hunt J (2009) Potential to improve on-farm wheat yield and WUE in Australia Crop and Pasture Science 60:708-716

Iizumi T et al. (2014) Impacts of El Niño Southern Oscillation on the global yields of major crops Nature communications 5:3712 doi:https://doi.org/10.1038/ncomms4712

Jarvis C, Darbyshire R, Eckard R, Goodwin I, Barlow E (2018) Influence of El Niño-Southern Oscillation and the Indian Ocean Dipole on winegrape maturity in Australia Agricultural and Forest Meteorology 248:502-510 doi:10.1016/j.agrformet.2017.10.021

Jaworski P, Durante F, Härdle WK (2013) Copulae in mathematical and quantitative finance Lecture Notes in Statistics—Proceedings Springer, Berlin

Jin EK et al. (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models Climate Dynamics 31:647-664 doi:10.1007/s00382-008-0397-3

Joe H (1996) Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters Lecture Notes-Monograph Series:120-141 doi:https://doi.org/10.1214/lnms/1215452614

Joe H (1997) Multivariate models and multivariate dependence concepts. CRC Press. doi:https://doi.org/10.1201/b13150

Joe H (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models Journal of Multivariate Analysis 94:401-419

Kao S-C, Govindaraju RS (2010) A copula-based joint deficit index for droughts Journal of Hydrology 380:121-134 doi:10.1016/j.jhydrol.2009.10.029

Kiem AS, Franks SW (2004) Multi-decadal variability of drought risk, eastern Australia Hydrological Processes 18:2039-2050

Kim G, Silvapulle MJ, Silvapulle P (2007) Comparison of semiparametric and parametric methods for estimating copulas Computational Statistics & Data Analysis 51:2836-2850 doi:https://doi.org/10.1016/j.csda.2006.10.009

Kraus D, Czado C (2017) D-vine copula based quantile regression Computational Statistics & Data Analysis 110:1-18 doi:https://doi.org/10.1016/j.csda.2016.12.009

Kurowicka D, Cooke RM (2006) Uncertainty analysis with high dimensional dependence modelling. John Wiley & Sons,

Larsen R, Leatham D, Sukcharoen K (2015) Geographical diversification in wheat farming: a copula-based CVaR framework Agricultural Finance Review 75:368-384 doi:https://doi.org/10.1108/AFR-07-2014-0020

Leonard M et al. (2014) A compound event framework for understanding extreme impacts Wiley Interdisciplinary Reviews: Climate Change 5:113-128

Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production Nature 529:84 doi:https://doi.org/10.1038/nature16467

Li Z, Cai W, Lin X (2016) Dynamics of changing impacts of tropical Indo-Pacific variability on Indian and Australian rainfall Scientific reports 6 doi:https://doi.org/10.1038/srep31767

Lim E-P, Hendon HH, Zhao M, Yin Y (2016) Inter-decadal variations in the linkages between ENSO, the IOD and south-eastern Australian springtime rainfall in the past 30 years Climate Dynamics:1-16

Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change Agricultural and Forest Meteorology 150:1443-1452

Lobell DB, Cahill KN, Field CB (2007) Historical effects of temperature and precipitation on California crop yields Climatic Change 81:187-203 doi:10.1007/s10584-006-9141-3

Lobell DB, Field CB, Cahill KN, Bonfils C (2006) Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties Agricultural and Forest Meteorology 141:208-218 doi:10.1016/j.agrformet.2006.10.006

Ludescher J et al. (2013) Improved El Niño forecasting by cooperativity detection Source: Proceedings of the National Academy of Sciences of the United States of America 110:11742-11745 doi:10.1073/pnas.1309353110

Matsumura K, Gaitan C, Sugimoto K, Cannon A, Hsieh W (2015) Maize yield forecasting by linear regression and artificial neural networks in Jilin, China The Journal of Agricultural Science 153:399-410 doi:https://doi.org/10.1017/S0021859614000392

McBride JL, Nicholls N (1983) Seasonal relationships between Australian rainfall and the Southern Oscillation Monthly Weather Review 111:1998-2004 doi:https://doi.org/10.1175/1520-0493(1983)111<1998:SRBARA>2.0.CO;2

McNeil A, Frey R, Paul E (2005) Quantitative risk management: Concepts, techniques and tools. Princeton University Press, Princeton, New Jersey

Meneghini B, Simmonds I, Smith IN (2007) Association between Australian rainfall and the southern annular mode International Journal of Climatology 27:109-121 doi:https://doi.org/10.1002/joc.1370

Min A, Czado C (2010) Bayesian inference for multivariate copulas using pair-copula constructions Journal of Financial Econometrics 8:511-546

Min SK, Cai W, Whetton P (2013) Influence of climate variability on seasonal extremes over Australia Journal of Geophysical Research Atmospheres 118:643-654 doi:https://doi.org/10.1002/jgrd.50164

Miranda MJ, Glauber JW (1997) Systemic risk, reinsurance, and the failure of crop insurance markets American Journal of Agricultural Economics 79:206-215 doi:https://doi.org/10.2307/1243954

Morales Napoles O, Cooke RM, Kurowicka D (2010) About the number of vines and regular vines on n nodes

Musshoff O, Odening M, Xu W (2011) Management of climate risks in agriculture–will weather derivatives permeate? Applied economics 43:1067-1077 doi:https://doi.org/10.1080/00036840802600210

Nagler T (2017) A generic approach to nonparametric function estimation with mixed data arXiv preprint arXiv:170407457

Nagler T (2018) Asymptotic analysis of the jittering kernel density estimator Mathematical Methods of Statistics 27:32-46

Nguyen-Huy T, Deo RC, Mushtaq S, An-Vo D-A, 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 98:65-81 doi:https://doi.org/10.1016/j.eja.2018.05.006

Nicholls N (1983) Predicting Indian monsoon rainfall from sea-surface temperature in the Indonesia–north Australia area Nature 306:576-577

Nicholls N (1997) Increased Australian wheat yield due to recent climate trends Nature 387:484-485

Nicholls N, Lavery B, Frederiksen C, Drosdowsky W, Torok S (1996) Recent apparent changes in relationships between the El Niño-Southern Oscillation and Australian rainfall and temperature Geophysical Research Letters 23:3357-3360 doi:https://doi.org/10.1029/96GL03166

Odening M, Shen Z (2014) Challenges of insuring weather risk in agriculture Agricultural Finance Review 74:188-199 doi:https://doi.org/10.1108/AFR-11-2013-0039

Okhrin O, Odening M, Xu W (2013) Systemic weather risk and crop insurance: the case of China Journal of Risk and Insurance 80:351-372 doi:https://doi.org/10.1111/j.1539-6975.2012.01476.x

Ouyang Z-s, Liao H, Yang X-q (2009) Modeling dependence based on mixture copulas and its application in risk management Applied Mathematics-A Journal of Chinese Universities 24:393-401 doi:https://doi.org/10.1007/s11766-009-2249-2

Palosuo T et al. (2011) Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models European Journal of Agronomy 35 doi:10.1016/j.eja.2011.05.001

Patton AJ (2001) Estimation of copula models for time series of possibly different lengths U of California, Econ Disc Paper

Pham MT, Vernieuwe H, De Baets B, Willems P, Verhoest N (2016) Stochastic simulation of precipitation-consistent daily reference evapotranspiration using vine copulas Stochastic Environmental Research and Risk Assessment 30:2197-2214

Phelps J, Wang S (2014) Extreme spring precipitation events in Intermountain West influenced by quasi-biennial oscillation

Pittock A (1975) Climatic change and the patterns of variation in Australian rainfall

Pohl B, Richard Y, Fauchereau N (2007) Influence of the Madden-Julian Oscillation on southern African summer rainfall Journal of Climate 20:4227-4242

Pook MJ, McIntosh PC, Meyers GA (2006) The synoptic decomposition of cool-season rainfall in the southeastern Australian cropping region Journal of Applied Meteorology and Climatology 45:1156-1170

Portmann FT, Siebert S, Döll P (2010) MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling Global Biogeochemical Cycles 24:n/a-n/a doi:10.1029/2008GB003435

Potgieter A, Hammer G, Butler D (2002) Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO Crop and Pasture Science 53:77-89 doi:https://doi.org/10.1071/AR01002

Potgieter A, Hammer G, Doherty A, De Voil P (2005) A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia Agricultural and Forest Meteorology 132:143-153

Power S, Casey T, Folland C, Colman A, Mehta V (1999) Inter-decadal modulation of the impact of ENSO on Australia Climate Dynamics 15:319-324

Revadekar JV, Preethi B (2012) Statistical analysis of the relationship between summer monsoon precipitation extremes and foodgrain yield over India International Journal of Climatology 32:419-429 doi:10.1002/joc.2282

Risbey JS, Pook MJ, McIntosh PC, Wheeler MC, Hendon HH (2009) On the remote drivers of rainfall variability in Australia Monthly Weather Review 137:3233-3253 doi:https://doi.org/10.1175/2009MWR2861.1

Ritter M, Musshoff O, Odening M (2014) Minimizing geographical basis risk of weather derivatives using a multi-site rainfall model Computational Economics 44:67-86

Royce FS, Fraisse CW, Baigorria GA (2011) ENSO classification indices and summer crop yields in the Southeastern USA Agricultural and Forest Meteorology 151:817-826 doi:10.1016/j.agrformet.2011.01.017

Sadegh M et al. (2018) Multi-hazard scenarios for analysis of compound extreme events Geophysical Research Letters

Sadegh M, Vrugt JA (2014) Approximate bayesian computation using markov chain monte carlo simulation: Dream (abc) Water Resources Research 50:6767-6787

Sadras V, Roget D, O'Leary G (2002) On-farm assessment of environmental and management constraints to wheat yield and efficiency in the use of rainfall in the Mallee Australian Journal of Agricultural Research 53:587-598

Salvadori G, De Michele C (2004) Frequency analysis via copulas: Theoretical aspects and applications to hydrological events Water Resources Research 40 doi:10.1029/2004WR003133

Schepen A, Wang Q, Robertson D (2012) Evidence for using lagged climate indices to forecast Australian seasonal rainfall Journal of Climate 25:1230-1246 doi:https://doi.org/10.1175/JCLI-D-11-00156.1

Schepsmeier U et al. (2018) Package 'VineCopula'

Schlenker W, Roberts MJ (2006) Nonlinear effects of weather on corn yields Applied Economic Perspectives and Policy 28:391-398

Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to US crop yields under climate change Proceedings of the National Academy of sciences 106:15594-15598

Seo J, Choi W, Youn D, Park DSR, Kim JY (2013) Relationship between the stratospheric quasi-biennial oscillation and the spring rainfall in the western North Pacific Geophysical Research Letters 40:5949-5953

Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation Journal of the Royal Statistical Society Series B (Methodological):683-690

Shiau J-T, Modarres R (2009) Copula-based drought severity-duration-frequency analysis in Iran Meteorological Applications 16:481-489

Shiau J (2006) Fitting drought duration and severity with two-dimensional copulas Water resources management 20:795-815

Shuai J, Zhang Z, Sun DZ, Tao F, Shi P (2013) ENSO, climate variability and crop yields in China Climate Research 58:133-148 doi:10.3354/cr01194

Skees JR, Barnett BJ (1999) Conceptual and practical considerations for sharing catastrophic/systemic risks Review of Agricultural Economics 21:424-441 doi:https://doi.org/10.2307/1349889

Skees JR, Hartell J, Murphy AG (2007) Using index-based risk transfer products to facilitate micro lending in Peru and Vietnam American Journal of Agricultural Economics 89:1255-1261

Sklar M (1959) Fonctions de répartition à n dimensions et leurs marges. Université Paris 8,

Smith M, Min A, Almeida C, Czado C (2010) Modeling longitudinal data using a pair-copula decomposition of serial dependence Journal of the American Statistical Association 105:1467-1479

Stephens D, Lyons T (1998) Rainfall-yield relationships across the Australian wheatbelt Australian Journal of Agricultural Research 49:211-224

Stone RC, Hammer GL, Marcussen T (1996) Prediction of global rainfall probabilities using phases of the Southern Oscillation Index

Suppiah R (2004) Trends in the southern oscillation phenomenon and Australian rainfall and changes in their relationship International Journal of Climatology 24:269-290

Tang A, Valdez EA (2009) Economic capital and the aggregation of risks using copulas Available at SSRN 1347675

Tankov P (2011) Improved Fréchet bounds and model-free pricing of multi-asset options Journal of Applied Probability 48:389-403

Taschetto AS, England MH (2009) El Niño Modoki impacts on Australian rainfall Journal of Climate 22:3167-3174

Team RC (2013) R: A language and environment for statistical computing

Turner NC (2004) Agronomic options for improving rainfall-use efficiency of crops in dryland farming systems Journal of Experimental Botany 55:2413-2425 doi:https://doi.org/10.1093/jxb/erh154

Ummenhofer CC et al. (2009) What causes southeast Australia's worst droughts? Geophysical Research Letters 36 doi:https://doi.org/10.1029/2008GL036801

Van Den Goorbergh RW, Genest C, Werker BJ (2005) Bivariate option pricing using dynamic copula models Insurance: Mathematics and Economics 37:101-114 doi:https://doi.org/10.1016/j.insmatheco.2005.01.008

Vedenov DV, Barnett BJ (2004) Efficiency of weather derivatives as primary crop insurance instruments Journal of Agricultural and Resource Economics:387-403

Venables WN, Ripley BD (2013) Modern applied statistics with S-PLUS. Springer Science & Business Media,

Verdon-Kidd DC, Kiem AS (2009) On the relationship between large-scale climate modes and regional synoptic patterns that drive Victorian rainfall Hydrology and Earth System Sciences 13:467-479 doi:10.5194/hess-13-467-2009

Verdon DC, Wyatt AM, Kiem AS, Franks SW (2004) Multidecadal variability of rainfall and streamflow: Eastern Australia Water Resources Research 40

Vose D (2008) Risk analysis: a quantitative guide. John Wiley & Sons,

Vrugt JA, Gupta HV, Bouten W, Sorooshian S (2003) A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters Water resources research 39

Weller E, Cai W (2013) Asymmetry in the IOD and ENSO teleconnection in a CMIP5 model ensemble and its relevance to regional rainfall Journal of Climate 26:5139-5149 doi:https://doi.org/10.1175/JCLI-D-12-00789.1

Wheeler MC, Hendon HH, Cleland S, Meinke H, Donald A (2009) Impacts of the Madden-Julian oscillation on Australian rainfall and circulation Journal of Climate 22:1482-1498

Williams AA, Stone RC (2009) An assessment of relationships between the Australian subtropical ridge, rainfall variability, and high-latitude circulation patterns International Journal of Climatology 29:691-709

Wong G, Lambert F, Leonard M, Metcalfe AV (2010) Drought Analysis Using Trivariate Copulas Conditional on Climatic States Journal of Hydrology Engineering 15:129-141 doi:10.1061//asce/he.1943-5584.0000169

Wong G, van Lanen HAJ, Torfs PJJF (2013) Probabilistic analysis of hydrological drought characteristics using meteorological drought Hydrological Sciences Journal 58:253-270 doi:10.1080/02626667.2012.753147

Woodard JD, Garcia P (2008) Basis risk and weather hedging effectiveness Agricultural Finance Review 68:99-117 doi:https://doi.org/10.1108/00214660880001221

Woodard JD, Paulson ND, Vedenov D, Power GJ (2011) Impact of copula choice on the modeling of crop yield basis risk Agricultural Economics 42:101-112

Xu W, Filler G, Odening M, Okhrin O (2010) On the systemic nature of weather risk Agricultural Finance Review 70:267-284 doi:https://doi.org/10.1108/00021461011065283

y Garcia AG, Persson T, Paz JO, Fraisse C, Hoogenboom G (2010) ENSO-based climate variability affects water use efficiency of rainfed cotton grown in the southeastern USA Agriculture, ecosystems & environment 139:629-635

Yang W (2010) Drought Analysis under Climate Change by Application of Drought Indices and Copulas. Portland State University

Yuan C, Yamagata T (2015) Impacts of IOD, ENSO and ENSO Modoki on the Australian winter wheat yields in recent decades Scientific reports 5 doi:https://doi.org/10.1038/srep17252

Zhang L, Singh VP (2014) Trivariate flood frequency analysis using discharge time series with possible different lengths: Cuyahoga river case study Journal of Hydrologic Engineering 19:05014012 doi:https://doi.org/10.1061/(ASCE)HE.1943-5584.0001003

Zhu Y, Ghosh SK, Goodwin BK Modeling Dependence in the Design of Whole Farm Insurance Contract---A Copula-Based Model Approach. In: annual meetings of the American Agricultural Economics Association, Orlando, FL, 2008. Citeseer, pp 27-29

Zubair L, Rao SA, Yamagata T (2003) Modulation of Sri Lankan Maha rainfall by the Indian Ocean dipole Geophysical Research Letters 30

Appendix A

Seasonal rainfall forecasts using D-vine copula-based quantile regression

---

Book chapter: Probabilistic seasonal rainfall forecasts using semi-parametric D-vine copula-based quantile regression

Summary:

This chapter applies an advanced statistical D-vine copula to forecast the seasonal cumulative rainfall in sixteen weather stations distributed over the Australia's wheat belt. These stations span different climate conditions recording historical data for the period 1889 – 2012. The seasonal rainfalls are forecast in different quantile levels using different climate predictor datasets derived from eight synoptic-scale climate indices. The five-fold cross-validation was employed to evaluate the out-of-sample performance of both models. The corrected Akaike Information Criterion (AIC)-conditional log-likelihood is employed to identify the most influential covariates to be additively incorporated into the multivariate probabilistic forecast model, resulting in a parsimonious predictive model.

The result found a statistically significant correlation between El Niño Southern Oscillation (ENDO) and Indian Oscillation Dipole (IOD) indices and seasonal rainfall. In general, the concurrent relationships between climate mode indices and rainfall are strongest. With the development of climate forecast systems, climate information like prominent El Niño events could be successfully predicted up to two years ahead. Therefore, it is possible to achieve seasonal rainfall forecasts with high accuracy and longer lagged time using statistical models and climate drivers. In addition, lagged climate indices also expose significant evidence to support the seasonal rainfall forecast. Furthermore, the D-vine copula model is superior to the traditional quantile regression methods in forecasting rainfall in the median and the upper quantile levels.

Climate indices derived from oceanic and atmospheric variability in the Pacific region, such as SOI, exhibit strong evidence for forecasting seasonal rainfall over much of the Australian wheat belt. The impact of the Indian Ocean on the seasonal

rainfall represents similar evidence in supporting the forecast of all weather stations. The extratropical region also shows a significant relationship with rainfall in some regions and during some seasons. The strongest and most spatially widespread evidence supporting JAS rainfall forecast comes from the JAS climate indices. Since the spread of the rainfall extremes is modelled through a joint distribution using a robust vine copula approach, these results will potentially improve risk-management and crop insurance.

Probabilistic seasonal rainfall forecasts using semi-parametric D-vine copula-based quantile regression

[Thong Nguyen-Huy, Ravinesh C Deo, Shahbaz Mushtaq, Shahjahan Khan]

[NON PRINT ITEMS]

Abstract:

Skillful probabilistic seasonal rainfall forecasts play a vital role in supporting water resource users, developing agricultural risk-management plans and improving decision-making processes. This chapter applies a novel statistical copula-based approach to develop a probabilistic seasonal rainfall forecast model using multiple large-scale oceanic and atmospheric climate indices. Here, a D-vine copula is employed to forecast the seasonal cumulative rainfall in sixteen weather stations across the Australia's wheat belt. These stations span different climate conditions recording historical data for the period 1889 – 2012. The seasonal rainfalls are forecast in different quantile levels using different climate predictor datasets. The corrected Akaike Information Criterion (AIC)-conditional log-likelihood is then used to screen the most influential covariates to be additively incorporated into the multivariate probabilistic forecast model, resulting in a parsimonious predictive model. The mutually inclusive correlations between El Niño Southern Oscillation (ENDO) and Indian Oscillation Dipole (IOD) indices and seasonal rainfall are found to be statistically significant. Therefore, using the climate information, skillful rainfall forecasts can be made three to six months ahead. The D-vine copula model is found to outperform the traditional quantile regression methods in forecasting rainfall in the median and the upper levels. The information from lagged,

concurrent and combined climate indices is therefore demonstrated to be a potentially useful predictor for forecasting seasonal rainfall in Australia's wheat belt region.

Key words: vine copulas; quantile regression; climate indices; rainfall prediction; conditional forecast.

[Chapter Starts Here]

1. Introduction

Skillful probabilistic seasonal rainfall forecasting plays an important role in supporting water resource users, developing agricultural risk-management plans and improving decision-making processes. The use of climate information in explaining rainfall variability, and its application to managing climate risks, has been well documented globally (Nicholson & Kim 1997; Corte-Real et al. 1998; Enfield et al. 2001). However, traditional probabilistic forecasting approaches, focusing on the mean values, are unable to quantify the tail dependence when extreme events occur. In this context, the quantile regression method, as proposed by Koenker and Bassett (1978), is an essential tool for capturing the full dependency structure between the climate indices and seasonal rainfall. This approach measures the association of the predictor variables with a conditional quantile of a dependent variable without any specific assumption on the conditional distributions. Therefore, quantile regression models are useful for quantifying the dependencies between variables in the outer regions of the conditional distribution.

This chapter develops a novel copula-based quantile regression method for investigating the impacts of various climate indices on rainfall variability, particularly when extreme events occur. In order to provide a clear focus, the Australian wheat belt will be used

as a case study. Australia is an agricultural nation with climate variability that is more spatially and temporally diverse than any other country (Nicholls et al. 1997; Best et al. 2007). The remote, synoptic-scale drivers, including El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) modes, are the principal factors influencing the inter-annual and inter-seasonal rainfall variabilities. ENSO and IOD are representative of the synoptic-scale processes of the air-sea interaction over the tropical Pacific and the Indian Ocean regions, respectively. Although many local factors such as atmospheric blocking and the subtropical ridge also influence the variability of Australian rainfall (King et al. 2014), the relationship between the remote drivers and Australian rainfall variation is the primary subject of discussion in this chapter.

The effects of ENSO on Australian rainfall fluctuation have been extensively investigated since the early 1980s (McBride & Nicholls 1983; Nicholls et al. 1996). It is well known that the oscillating phases of ENSO are the main factors explaining Australia's rainfall variability, in particular during the period July – March. Risbey et al. (2009) reported that ENSO has the strongest relationship to rainfall in the east of Australia, where generally La Niña phases bring more rainfall and El Niño phases are linked to decreased rainfall. McKeon et al. (2004) found that the El Niño phases were associated with drought events over much of the Australia continent. Furthermore, ENSO also has a significant influence on the rainfall patterns in north and northeast Australia (Holland 1986; Brown et al. 2011). The ENSO-rainfall relationship varies across the Australian continent (Power et al. 2006; Nguyen-Huy et al. 2017), even within particular regions such as southeast Queensland (Cai et al. 2010) and southeast Australia (King et al. 2013). In general, the influence of ENSO on rainfall during the La Niña phase is stronger than during the El Niño phase.

The Indian Ocean dipole, similar to ENSO, is associated with the variability of Australian rainfall depending on seasons and times. IOD mainly modulates inter-annual rainfall

in western and southern Australia during the winter and spring seasons (Risbey et al. 2009). The influence of IOD on the climatic conditions of southeast Australia has also been observed in some studies (Meyers et al. 2007) where it has been associated with drought events in this region (Ummenhofer et al. 2009; Ummenhofer et al. 2011). Moreover, it was also observed throughout the twentieth century that the increased occurrences of the positive IOD phases were key drivers of major drought events in south-east Australia, where ENSO conditions are not usually assumed (Cai & Rensch 2012).

Several studies have identified the relative roles of climate mode indices on Australian rainfall variability within either particular regions (Gallant et al. 2007; Klingaman et al. 2013) or a whole country (Risbey et al. 2009; Schepen et al. 2012; Min et al. 2013). However, there has been less attention paid to the relationship between joint climate drivers and extreme rainfall. In one of the first studies, Nguyen-Huy et al. (2017) found that the rainfall forecast was significantly improved in the upper and lower tails using the combination of ENSO and Inter-decadal Pacific Oscillation (IPO) Tripole Index (TPI). Therefore, further studies of the association between rainfall and multiple climate modes are required to better understand how these remote drivers are able to modulate extreme rainfall over a monthly and seasonal timescale.

In this chapter, we adopt the copula theorem (Sklar 1959) as a way to provide a powerful approach for modeling the non-linear dependencies among bivariate, trivariate and multivariate random variables. In a copula-based joint distribution, the associations between the relevant variables are modeled independently with the individual marginal distribution of each variable. As a result, statistical copula-based models can overcome the issues of normal and symmetric assumptions in traditional forecast methods. Therefore, recent years have witnessed extensive applications of copula-based modeling in a wide range of fields such as economics and finance (de Melo Mendes et al. 2010; Nguyen & Bhatti 2012), water resources

and hydrology (Hao & Singh 2012; Grimaldi et al. 2016), agriculture (Bokusheva 2011; Nguyen-Huy et al. 2018), and environment (Kao & Govindaraju 2010; Sraj et al. 2015).

This chapter aims to develop new understandings and applications of copula models by investigating the teleconnections related to climate variability between the different remote synoptic-scale climate drivers and extreme seasonal rainfall across the Australian wheat belt. Comparisons are made between the novel D-vine quantile regression and traditional quantile regression. Five-fold cross-validation is also applied to evaluate their out-of-sample performance and observe the sensitivity of the predictor set. The primary contribution of this chapter is to develop and validate the suitability of a copula-statistical methodology for the quantile-based forecasting of rainfall using large-scale climate mode influences and the implications of the model in agricultural risk-management and decision-making.

A brief description of the data used and methodologies applied is presented in Section 2. Results and analysis of climate-rainfall relationship and model performance are described in Section 3. Discussion of the results and future works and the conclusions are given in Sections 4 and 5, respectively.

2. Data and Methodology

2.1 Cumulative rainfall index

The monthly and seasonal total precipitation data employed in this chapter were obtained from the daily rainfall data covering the period from January 1, 1889 until December 31, 2012. These datasets are available from Scientific Information for Land Owners (SILO) and can be downloaded via the website of The Long Paddock, Queensland Government (https://legacy.longpaddock.qld.gov.au/silo/). These SILO databases are constructed from historical observational climate records provided by the Bureau of Meteorology (BOM). These

time series are acquired for sixteen weather stations that are spread over the Australian wheat belt and span different climate regimes (Fig. 1).

[Insert Figure 1 here]

The cumulative rainfall index (CRI) is derived from measurements based on these daily observations. It is commonly used as the hypothetical underlying for agricultural weather insurance. In this study, CRI measures the rainfall within the two main vegetation periods of wheat crops that last from April 1 until June 30 and from July 1 until September 30. The index is calculated as follows (Xu et al. 2010):

$$CRI_{i,t} = \sum_{j=T_B}^{j=T_E} P_{j,t,i}, \quad (1)$$

where $P_{j,t,i}$ is the daily precipitation (mm) observed at day j in year $t = 1,...,124$ and station $i = 1,...,16$. T_B and T_E denote the beginning and the end of the considered period, respectively. In Australia, wheat is the main grain crop grown in the wheat belt where the sowing season is commonly from the mid-April to June, depending the rainfall pattern. The harvest season is normally from mid-October until January of the next year. Therefore, two cumulative rainfall indices are derived for two periods, namely April – June (AMJ CRI) (sowing stage) and July – September (JAS CRI) (before harvesting). In the context of weather index-based insurance, this index addresses drought risk (Martin et al. 2001; Xu et al. 2010). The information and statistics of these selected weather stations are described in Table 1.

[Insert Table 1 here]

It is clear that weather and climate regimes vary over the Australian wheat belt. In particular, the eastern part of the wheat belt has a subtropical and savanna climate with the average rainfall during AMJ higher than during JAS months. Furthermore, the rainfall variability in this region is generally higher than other remaining sites in both summer and

winter. The south, southwest and west have a subtropical, savanna and temperate climate, experiencing more rainfall in the winter season. In addition, the western part of the wheat belt receives the highest rainfall on average. Remote drivers influence these weather systems mentioned previously in a complicated way, resulting in rainfall variability in Australia.

2.2 Climate indices

Eight synoptic-scale climate indices are used for a comprehensive analysis of their influence on seasonal rainfall variability. These climate mode indices have been well documented in many studies investigating the climate-rainfall relationship in Australia (Risbey et al. 2009; Kirono et al. 2010; Schepen et al. 2012). ENSO is represented by several different indicators including Niño3.0, Niño3.4, and Niño4.0 (*i.e.*, sea surface temperature (SST) representative), Southern Oscillation Index (SOI) (air pressure) and El Niño Modoki (EMI) (coupled ocean-atmospheric). The Dipole Mode Index (DMI) characterizes the intensity and Indonesian Index (II) the individual pole of IOD over the tropical Indian Ocean. The Tasman Sea Index (TSI) is included in this research to consider a potential link between extratropical SST and rainfall variability.

In terms of the origin of these data, the monthly SST anomalies for the period from January 1, 1889 to December 31, 2012 are derived from NOAA Extended Reconstructed Sea Surface Temperature Anomalies (SSTA) data, version 4, downloaded from the Asia-Pacific Data Research Center (APDRC). Monthly SOI data were acquired from the Bureau of Meteorology, Australia (BOM). The seasonal climate indices are calculated as the average of three-month values. Table 2 summarizes the description of these climate indices including the formula of EMI and DMI calculations and their components.

2.3 Methodology

2.3.1 Copula theorem

A copula, as explained by Sklar (1959), is a function used to link multiple univariate marginal distributions of random variables into a multivariate distribution. In brief, suppose a *d*-dimensional random vector $X = (x_1, ..., x_d)^T$ has its marginal cumulative distribution functions (CDFs) $F_1(x_1), ..., F_d(x_d)$ and probability density functions (PDFs) $f_1(x_1), ..., f_d(x_d)$. Their joint CDF is expressed as (Sklar 1959):

$$F(x_1, ..., x_d) = C\left[F_1(x_1), ..., F_d(x_d)\right] = C(u_1, ..., u_d) \tag{2}$$

and the corresponding joint PDF:

$$f(x_1, ..., x_d) = \left[\prod_{i=1}^{d} f_i(x_i)\right] c(u_1, ..., u_d), \tag{3}$$

where C denotes the copula function and $c = \frac{\partial^d C}{\partial F_1, ..., \partial F_d}$ the corresponding copula density. If marginal distributions are continuous, then C is unique and $u_i = F_i(x_i), i = 1, ..., d$ being the univariate probability integral transformation (PIT). Copula families are generally distinguished as empirical, elliptical, Archimedean, extreme value, vine, and entropy copulas. This study focuses on the use of the D-vine copula approach, described below, which serves a quantile forecast purpose.

2.3.2 D-vine copulas

D-vine copula, a special form of vine family, was first proposed by Joe (1997) and further developed by Bedford and Cooke (2001, 2002). In short, the copula density in Eq. (3) is decomposed into the conditional and unconditional bivariate densities, so-called bivariate pair-copulas, as follows (Czado 2010):

$$c(u_1,...,u_d)=\prod_{i=1}^{d-1}\prod_{j=i+1}^{d}c_{ij;i+1,...,j-1}\left[C_{i|i+1,...,j-1}\left(u_i\left|u_{i+1},...,u_{j-1}\right.\right),C_{j|i+1,...,j-1}\left(u_j\left|u_{i+1},...,u_{j-1}\right.\right)\right] \quad (4)$$

In this construction, each pair-copula is selected independently from the others, allowing the flexible model of full dependence structure of high-dimensional random variables existing as the characteristics of asymmetric and tail dependences. Therefore, the D-vine approach can address the limitations of other copula families such as meta-elliptical or symmetric Archimedean copulas.

2.3.3 Semi-parametric D-vine quantile regression

Equation (2) reveals that the construction of copula-based models commonly includes fitting the marginal distributions and fitting the copulas. In general, these both procedures can be fitted either parametrically or non-parametrically. In this study, the non-parametric approach is employed to fit marginal distributions and the copulas are fitted parametrically, resulting in a semi-parametric quantile regression model. Constructing the model in this way can minimize the bias and inconsistency issues often faced by the fully parametric model when one of the parametric components is misspecified (Noh et al. 2013).

As the first step, marginal distributions are fitted non-parametrically using the univariate local-polynomial likelihood density estimation method (Nagler 2017). Given a sample $x^i, i=1,...,n$ with unknown PDF, the estimated kernel density is defined as (Geenens & Wang 2018):

$$f(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x^i}{h}\right), \quad (5)$$

where h is the bandwidth parameter and $K(x)$ the smoothing kernel function. In this study, the Gaussian kernel and the plug-in bandwidth are used as in the methodology developed by

Sheather and Jones (1991). The degree of the polynomial is selected as the log-quadratic fitting (Nagler 2017).

The estimated marginal distribution functions F_Y and F_j can be then obtained for the response variable Y and predictor variables $X_1,...,X_d$, $j=1,...,d$, respectively. These functions are used to convert the observed data to pseudo-copula data, which are, $\hat{v}^i = F_Y\left(y^i\right)$ and $u_j^i = F_j\left(x_j^i\right)$. These pseudo-copula data $\mathrm{v}=\left(\hat{v}^i\right)$ and $\upsilon=\left(u_j^i\right)$ approximate to an i.i.d sample from the PIT vector $\left(V,U_1,...,U_d\right)^T$ and are therefore able to be used for the D-vine copula estimate in the next step (Kraus & Czado 2017).

These pseudo-copula data are fitted to a D-vine model with an order $V-U_{l_1}-...-U_{l_d}$, where $L=\left(l_1,...,l_d\right)^T$ is the arbitrary ordering resulting in *d*! possible models. Therefore, this study applies the new algorithm proposed by Kraus and Czado (2017) to automatically select the parsimonious D-vine model. In short, only the most influential predictors are added into the model in an order that minimizes the Akaike Information Criterion (AIC)-corrected conditional log-likelihood cll^{AIC}. As a result, the conditional quantile prediction model has the highest explanation for the response variable. Furthermore, this algorithm overcomes the common issue in terms of conventional regression, involving collinearity, transformation and the inclusion and exclusion of covariates.

Finally, for quantile levels $\alpha\in(0,1)$, the quantile q_α of a response variable Y given predictor variables $X_1,...,X_d$ can be obtained using the inverse forms of the marginal distribution function F_Y^{-1} and the conditional copula function $C^{-1}_{V|U_1,...,U_d}$ conditional on $u_1,...,u_d$ which is defined as (Kraus & Czado 2017):

$$q_\alpha(x_1,...,x_d) = F_Y^{-1}\left[C^{-1}_{V|U_1,...,U_d}\left(\alpha \middle| u_1,...,u_d\right)\right]. \quad (6)$$

More details of this approach can be found in Kraus and Czado (2017) and Schallhorn et al. (2017).

2.3.4 Linear quantile regression

For the purpose of comparison, this study also utilizes the traditional linear quantile regression (LQR) model to predict rainfall with the same predictor sets for the D-vine copula model. The LQR approach, first introduced by Koenker and Bassett (1978), assumed the conditional quantile of the predicted variables to be linear in the predictors. This assumption can be expressed as (Schallhorn et al. 2017):

$$q_\alpha(x_1,...,x_d) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j, \quad (7)$$

where the estimates of the regression coefficients β_j are acquired by solving the minimization problem:

$$\min_{\beta \in \mathrm{R}^{d+1}} \left[\alpha \sum_{i=1}^{n} \left(y^i - \beta_0 - \sum_{j=1}^{d} \beta_j x_j^i \right)^+ + (1-\alpha) \sum_{i=1}^{n} \left(\beta_0 + \sum_{j=1}^{d} \beta_j x_j^i - y^i \right)^+ \right]. \quad (8)$$

The LQR method has a number of limitations such as a very restrictive assumption of normal margins and the changeable slopes at different quantile levels (Bernard & Czado 2015; Kraus & Czado 2017; Schallhorn et al. 2017).

2.3.5 Evaluation of model performance

In order to assess the forecast performance of semi-parametric D-vine and linear quantile regression, this study applies a five-fold cross-validation test to evaluate the out-of-sample performance. Therefore, the total 124 data points (1889 – 2012) are split into five folds where each fold will become an evaluation data set respectively. The remaining data corresponding

to each fold are used as training datasets. As a result, all data points are joined in training and testing processes.

Since in this out-of-sample test the true regression is unknown (Kraus & Czado 2017), only a realization for each seasonal rainfall can be obtained, and an averaged cross-validated tick-losses function $L_{\alpha,m}^{j}$ is employed to evaluate the forecasted α-quantiles for $\alpha \in (0,1)$. The expression of this computation is expressed as follows (Komunjer 2013; Kraus & Czado 2017; Schallhorn et al. 2017):

$$L_{\alpha,m}^{j} = \frac{1}{n_{eval}} \sum_{i=1}^{n_{eval}} \rho_{\alpha} \left(y^{i} - \overset{i}{q}_{\alpha,m} \right), \tag{9}$$

where y^{i} and $\overset{i}{q}_{\alpha,m}$ denote the observation and forecast quantile using method m at the station j and a point i in a sample size n_{eval}. The function $\rho_{\alpha} = y\left[\alpha - \mathrm{I}(y<0)\right]$ is the check or tick function. The lower values of the averaged cross-validated tick-losses function imply better performance of the forecast model.

3. Results

3.1 Climate-rainfall relationships

The influence of climate indices on rainfall variability is inspected as an initial analysis prior to the development of the probabilistic models. The Kendall statistic has been used to estimate the rank-based measure of association between climate indices and CRI at different lag times in the entire dataset for different weather stations. The concurrent relationship of climate indices and CRI has been explored as well. Such simultaneous relationships may benefit the seasonal rainfall forecast models since the recent maturity of climate forecasting systems allows information of climate mode indices to be forecast with sufficient lead time and

accuracy (Chen et al. 2004). The correlation coefficients between AMJ CRI and climate conditions during the period JFM and AMJ are illustrated in Table 3. All this climate information is used together with JAS climate indices to analyze JAS CRI, and the results are represented in Table 4.

It can be seen from Table 3 that all JFM climate indices provide very limited information for forecasting of AMJ CRI. This is to be expected since the impact of remote drivers on rainfall variability, as mentioned before, is generally strong from July of the year being considered to March of the next year. The results agree with findings from a study undertaken by Schepen et al. (2012). The concurrent relationship between a range of climate indices and rainfall is stronger in AMJ. It is worth pointing out that the ENSO plays an important role in the variations of AMJ rainfall over the Australian wheat belt, as indicated by the significant coefficients between SOI and rainfall across most of the weather stations. According to Lo et al. (2007), SST-based indices are more useful as the predictors or drivers of Australian rainfall at longer times ahead compared to SOI. However, our observations show that the SOI can be potentially used to forecast rainfall at the same lead timescale with SST-based indices depending on locations and seasons (see Tables 3 and 4). The mechanism for this relationship may be explained by the fact that the SOI is related to the large-scale surface pressure, and therefore its variability is more closely associated with the rainfall process. Furthermore, the collection of SOI data is based on consistent pressure values observed from two stations providing more confidence in the early record than SST-based indices that are interpolated from the observations of sparse stations (Risbey et al. 2009).

The Indian Ocean dipole, similar to ENSO, is an index representing a coupled interaction of ocean–atmosphere phenomena in the equatorial Indian Ocean (Saji et al. 1999). In this study, we found that the DMI and II anomalies derived from the Indian Ocean have a similar impact on Australian rainfall compared to that of the Pacific region. There is minimum

evidence encouraging the use of any lagged climate drivers (i.e., JFM DMI and JFM II) to forecast AMJ CRI. However, variations of AMJ DMI are related to AMJ CRI along the east to the south-east and south Australia, but excluding the western region. The impact of II and extratropical index TSI can be useful as a predictor to forecast AMJ rainfall in some stations, but are very limited in general using either lagged or simultaneous information.

[Insert Table 3 here]

Table 4 represents the usefulness of lag ENSO information in forecasting seasonal rainfall, in particular in the east and south-east regions, during the period of July – August. However, it is interesting that the influence of different ENSO indicators on JAS rainfall varies at different locations. Niño4.0 affects the rainfall in the east region only while there is no significant correlation between EMI and rainfall at all weather stations. On the other hand, Niño3.4, SOI, and Niño3.0 extend their effects on JAS CRI to the south-east, south and west regions, respectively. The influence from the Indian Ocean anomalies on rainfall patterns is similar to ENSO, where the lag information of AMJ DMI and AMJ II is potentially useful for forecasting JAS rainfall from the east to the south and south-west regions. The influence from the extratropical region on rainfall can be observed in the south-east, south-west and western Australia where AMJ TSI has a significant correlation with S10-11, S04 and S01, respectively.

[Insert Table 4 here]

As expected, simultaneous correlation coefficients between climate drivers and rainfall of the JAS period are stronger than the lag coefficients. In regard to ENSO phenomena, while EMI influences rainfall in the eastern region only (i.e., S11-16) and the impact from all Niño indices does not cover some regions in the south and west-east (i.e., S03 and S05-06), the information of JAS SOI can be used to skillfully forecast JAS rainfall over much of the Australian wheat belt. Taking into consideration the impact from the Indian Ocean, DMI affects

most of the weather stations except S03 and S05. In addition, II can be used to compensate for the lack of forecasting information in these weather stations. These results agree with a study undertaken by Risbey et al. (2009) where IOD generally peaks in spring (September – October), but can be observed from May to November. Furthermore, there is no evidence supporting the use of TSI as a predictor for rainfall forecast over the wheat belt regions except for the western region where it is useful for explaining JAS rainfall in three out of five stations.

The results indicate that probabilistic seasonal rainfall forecasts can be performed efficiently in all regions with sufficient lead time using multiple climate drivers. For these reasons, it is obvious that a robust rainfall forecasting model should take multiple climate drivers and lag information into account to achieve better performance, at least in terms of time sufficiency and spatial coverage. In addition, if the IOD and ENSO events occur together they can reinforce each other (Kirono et al. 2010), although this need not necessarily happen (Meyers et al. 2007; Risbey et al. 2009). The question of whether these combinations can improve the accuracy of rainfall forecasts will be addressed in the following section.

3.2 Rainfall quantile forecast

We now present the forecast of seasonal rainfall and the evaluation of D-vine, and benchmark these results with an LQR model performance. The rainfall forecast is made at three quantile levels: 0.05 (lower tail), 0.50 (median) and 0.95 (upper tail) for two periods of AMJ and JAS using various combinations of climate drivers as predictor variables. In particular, three predictor sets including climate indices observed in JFM (i.e., eight predictors), AMJ (i.e., eight predictors) and JFM+AMJ (i.e., sixteen predictors) are used to forecast AMJ CRI. Similarly, JAS CRI at sixteen weather stations is forecast using six predictor sets consisting of the former sets and three predictor sets of JAS, AMJ+JAS, and JFM+AMJ+JAS (i.e., twenty-four predictors).

Figure 2 displays the results of the AMJ CRI forecast at the three alpha levels derived from all predictor sets and both models. The first visual inspection indicates that the results of rainfall forecast and model performance vary across the study regions depending on alpha levels, predictor sets and models used. In general, the D-vine copula model provides better accuracy than LQR for the west-west region (S01-02) at the lower tail and for the south (S06-07) and south-east (S09-11) regions at the median and upper tail for all predictor sets. These findings imply that the impact of climate indices on these stations are more scattered and non-linear at the median and upper extreme events which cannot be captured by the traditional LQR method. Furthermore, both models reveal that the use of simultaneous information or its combination with lag information does not always improve the forecasting performance. These outcomes reflect the spatio-temporal characteristic of influences of climate indices on Australian rainfall.

[Insert Figure 2 here]

This spatio-temporal variability affecting rainfall of climate drivers is especially emphasized in Fig. 3 where the D-vine copula model outperforms the LQR approach in most cases at the median and upper extreme levels. This highlights the usefulness of the copula-based model in forecasting JAS rainfall above the median level. Furthermore, it is clear that the impact of climate on Australian seasonal rainfall is asymmetric where the upper tail is more scattered and non-linear. In regard to the use of different predictor sets, the results again show an inconsistent pattern of using information from lag, concurrent and their combination. For example, JAS climate information yields the best performance of JAS CRI forecast in the S16 below the median levels. However, to forecast JAS CRI in the upper quantiles, the climate information of lagged and concurrent times and their combination are seen to provide almost the same results.

[Insert Figure 3 here]

The differences in the forecasting performance of D-vine and LQR are illustrated as an example in Fig. 4 for S16. In this figure, the observations of AMJ SOI and JAS CRI are represented by dotted points overlaid with locally weighted regression lines of forecasted JAS CRI from five-fold cross-validation in three quantile levels (0.05, 0.50, and 0.95). It is clear that the relationship in the upper tail between observed AMJ SOI and JAS CRI is scattered and non-linear. This empirical pattern of dependence may explain the reason for the outperformance of the D-vine copula-based model in the quantile level of 0.95 where the LQR method is inadequate. Both models yield a small difference in the forecast lines of the lower and median quantile levels. However, there is large divergence between forecast lines in the upper quantile which is in agreement with the results derived from the average cross-validated tick-losses function (see Fig. 3).

[Insert Figure 4 here]

Finally, the results (not shown here) also spell out the fact that the influence of climate indices on seasonal rainfall can vary over a decadal timescale in the present study region. This was indicated by the change of predictor sets selected for the training model in each fold of the cross-validation process. Furthermore, the maximum number of selected predictors was four in all considered cases. These findings question whether it is possible to build up a certain predictor set of climate indices for rainfall forecast in each region. However, the answer to this problem, although will it bring interesting insights, is out of the scope of this chapter.

4. Discussion

This chapter has explored the association between a number of climate mode indices and rainfall observed in many weather stations across Australia's wheat belt regions. The stations selected were those that geographically distribute over the Australian wheat belt and experience different climate conditions. The lagged and concurrent information derived from the Pacific

and Indian Oceans were found to be useful for rainfall forecast systems. Based on this analysis, the chapter also developed a quantile rainfall forecast model using the vine copula approach. The semi-parametric D-vine copula-based model used in this chapter showed better performance for rainfall forecasting in the median and upper levels. To minimize the model misspecification further (Noh et al. 2013), future research may apply the fully non-parametric copula-based approach (Schallhorn et al. 2017) meaning that both estimates of marginal distributions and copulas are non-parametric. Furthermore, the results from this chapter also showed that the LQR provides better agreement of rainfall forecast in the lower tail. Therefore, other several quantile regression models such as boosting additive (Koenker 2011) or non-parametric quantile regression (Li et al. 2013) may be used in future work for the purpose of comparison.

It is worth noting that the predictor datasets chosen for the training models in each fold of the cross-validation method and each location differed from each other. These changes reflected the spatio-temporal characteristics of the impact of various climate mode indices on Australian rainfall. Therefore, further research could assist in building up a certain predictor dataset of climate indices for rainfall forecast corresponding to each study site. This work can be done by using a comparison of model performance between a fixed predictor set and exchangeable predictor sets using a k-fold cross-validation approach. In order to provide a more comprehensive analysis, the rainfall forecast may be conducted at more time scale points.

The Australian climate is also affected by many local factors such as atmospheric blocking and the subtropical ridge, which were not considered in this study. Atmospheric blocking to the southeast of Australia has been examined as the driver of rainfall increase across a large part of Australia while the position and intensity of the subtropical ridge affect rainfall in the east of Australia (Risbey et al. 2009; Cai et al. 2011; Schepen et al. 2012). According to Zscheischler et al. (2018), extreme events are often the result of the processes that many drivers

interact together and have spatio-temporal dependencies. As a result, the risk assessment is potentially underestimated. Therefore, investigating the joint influence of large-scale and local drivers to improve risk management is an important consideration for future research.

In practice, probabilistic seasonal rainfall forecasts can be derived from both empirical and dynamic climate forecasting models up to a year ahead (Goddard et al. 2001; Schepen et al. 2012). However, they all have their own advantages and limitations. The empirical models might be categorized into statistically-based (Rajeevan et al. 2007; Nguyen-Huy et al. 2017) or machine learning methods (Ramirez et al. 2005). Empirical models use the empirical relationships between historically observed variables and therefore depend on the availability of recorded data length and assume stationary relationships between variables. On the other hand, dynamic forecasting models (Druce 2001; Vieux et al. 2004) rely on numerical simulations directly modeling physical processes; however, they often cost more than statistical models in terms of implementation and operation. Therefore, a hybrid integrated forecasting system is preferred to provide greater accuracy and precision of rainfall forecasts, while being more economically viable.

5. Conclusions

This chapter has demonstrated that the information derived from large-scale oceanic-atmospheric processes is potentially useful for seasonal rainfall forecasting in Australia. In general, the simultaneous relationships between climate mode indices and rainfall are strongest. this finding agrees with results from studies undertaken by Risbey et al. (2009). With the development of climate forecast systems, climate information like prominent El Niño events could be successfully predicted up to two years ahead (Chen et al. 2004). Therefore, it is possible to achieve seasonal rainfall forecasts with high accuracy and longer lagged time using

statistical models and climate drivers. In addition, lagged climate indices also expose significant evidence to support the seasonal rainfall forecast.

Climate drivers have an asymmetric influence on seasonal rainfall in Australia. Climate indices derived from oceanic and atmospheric variability in the Pacific region such as SOI exhibit strong evidence for forecasting seasonal rainfall over much of the Australian wheat belt. The impact of the Indian Ocean on the seasonal rainfall represents similar evidence in supporting the forecast of all weather stations. The extratropical region also shows a significant relationship with rainfall in some regions and during some seasons. The strongest and most spatially widespread evidence supporting JAS rainfall forecast comes from the JAS climate indices. Furthermore, the joint occurrence of extreme climate events may reinforce rainfall fluctuation (Nguyen-Huy et al. 2017) and may subsequently affect crop yield (Nguyen-Huy et al. 2018). Therefore, using a copula-based model with multiple climate indices as predictors could improve the forecast of seasonal rainfall ahead.

The copula-based joint probability modeling method was applied to forecast the seasonal cumulative rainfall across Australian wheat belt using different predictor sets of lagged and concurrent climate indices. In addition, the traditional linear quantile regression is simultaneously implemented for a comparison. The five-fold cross-validation was employed to evaluate the out-of-sample performance of both models. Furthermore, the most influential predictors were selected based on the AIC-corrected conditional log-likelihood to form the parsimonious model. In general, the D-vine copula-based model shows greater potential for forecasting rainfall above the median level. The results imply that the impact of climate indices on rainfall is non-linear in the upper quantiles where they may be unable to be measured by the traditional LQR.

The usefulness of lagged, concurrent or combined climate information for seasonal rainfall forecast varies with locations and times. The performance of seasonal rainfall quantile

forecasts using the information of lagged climate indices may be higher than that of using simultaneous predictor sets. Furthermore, the selected parsimonious predictor sets are different from each other for each training model in each fold of the cross-validation method. Therefore, a potential study is to test the performance of seasonal rainfall using certain predictor sets for each location. In addition, future research may be conducted with more climate indices at more time scales using fully non-parametric models.

Probabilistic seasonal rainfall forecasts derived from statistical models can provide important information to a variety of users related to water resource in regard to planning and decision-making processes. For example, seasonal rainfall forecasts may assist water managers to make operational decisions on water allocation for rival users (Kirono et al. 2010). In addition, seasonal rainfall forecasting is one of the most effective means to adapt and diminish the vagaries of adverse weather and support the development of risk-management strategies. For example, skillful quantification of seasonal rainfall in extreme cases with a sufficient time lag can support agricultural producers geographically diversifying farming systems to minimize climate risk and optimize profitability (Larsen et al. 2015). We are currently studying the use of a copula-based approach for evaluating the weather (rainfall) systemic risk (Xu et al. 2010; Okhrin et al. 2013) in Australia. Since the spread of the rainfall extremes is modeled through a joint distribution using a robust vine copula approach, these results will potentially improve risk-management and crop insurance.

Acknowledgments

The project was financed by a University of Southern Queensland Post Graduate Research Scholarship (USQPRS 2015-2018), School of Agricultural, Computational and Environmental Sciences and Drought and Climate Adaptation Program (DCAP) (Producing Enhanced Crop

Insurance Systems and Associated Financial Decision Support Tools). The authors would like to acknowledge Dr. Louis Kouadio for his support in collecting rainfall data.

References

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.
Bedford, T & Cooke, RM 2002, 'Vines: A new graphical model for dependent random variables', *Annals of Statistics*, pp. 1031-68.
Bernard, C & Czado, C 2015, 'Conditional quantiles and tail dependence', *Journal of multivariate analysis*, vol. 138, pp. 104-26.
Best, P, Stone, R & Sosenko, O 2007, 'Climate risk management based on climate modes and indices-the potential in Australian agribusinesses', in *101st Seminar, July 5-6, 2007, Berlin Germany*, European Association of Agricultural Economists.
Bokusheva, R 2011, 'Measuring dependence in joint distributions of yield and weather variables', *Agricultural Finance Review*, vol. 71, no. 1, pp. 120-41.
Brown, JR, Power, SB, Delage, FP, Colman, RA, Moise, AF & Murphy, BF 2011, 'Evaluation of the South Pacific Convergence Zone in IPCC AR4 climate model simulations of the twentieth century', *Journal of Climate*, vol. 24, no. 6, pp. 1565-82.
Cai, W & Rensch, P 2012, 'The 2011 southeast Queensland extreme summer rainfall: a confirmation of a negative Pacific Decadal Oscillation phase?', *Geophysical Research Letters*, vol. 39, no. 8.
Cai, W, Van Rensch, P & Cowan, T 2011, 'Influence of global-scale variability on the subtropical ridge over southeast Australia', *Journal of Climate*, vol. 24, no. 23, pp. 6035-53.
Cai, W, van Rensch, P, Cowan, T & Sullivan, A 2010, 'Asymmetry in ENSO teleconnection with regional rainfall, its multidecadal variability, and impact', *Journal of Climate*, vol. 23, no. 18, pp. 4944-55.
Chen, D, Cane, MA, Kaplan, A, Zebiak, SE & Huang, D 2004, 'Predictability of El Niño over the past 148 years', *Nature*, vol. 428, no. 6984, p. 733.
Corte-Real, J, Qian, B & Xu, H 1998, 'Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation', *International Journal of Climatology*, vol. 18, no. 6, pp. 619-35.
Czado, C 2010, 'Pair-copula constructions of multivariate copulas', in *Copula theory and its applications*, Springer, pp. 93-109.
de Melo Mendes, BV, Semeraro, MM & Leal, RPC 2010, 'Pair-copulas modeling in finance', *Financial Markets and Portfolio Management*, vol. 24, no. 2, pp. 193-213.
Druce, DJ 2001, 'Insights from a history of seasonal inflow forecasting with a conceptual hydrologic model', *Journal of Hydrology*, vol. 249, no. 1-4, pp. 102-12.
Enfield, DB, Mestas-Nuñez, AM & Trimble, PJ 2001, 'The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US', *Geophysical Research Letters*, vol. 28, no. 10, pp. 2077-80.
Gallant, AJ, Hennessy, KJ & Risbey, J 2007, 'Trends in rainfall indices for six Australian regions: 1910-2005', *Australian Meteorological Magazine*, vol. 56, no. 4, pp. 223-41.
Geenens, G & Wang, C 2018, 'Local-likelihood transformation kernel density estimation for positive random variables', *Journal of Computational and Graphical Statistics*, no. just-accepted.

Goddard, L, Mason, SJ, Zebiak, SE, Ropelewski, CF, Basher, R & Cane, MA 2001, 'Current approaches to seasonal to interannual climate predictions', *International Journal of Climatology*, vol. 21, no. 9, pp. 1111-52.
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.
Hao, Z & Singh, VP 2012, 'Entropy-copula method for single-site monthly streamflow simulation', *Water Resources Research*, vol. 48, no. 6.
Holland, GJ 1986, 'Interannual variability of the Australian summer monsoon at Darwin: 1952–82', *Monthly Weather Review*, vol. 114, no. 3, pp. 594-604.
Joe, H 1997, *Multivariate models and multivariate dependence concepts*, CRC Press.
Kao, S-C & Govindaraju, RS 2010, 'A copula-based joint deficit index for droughts', *Journal of Hydrology*, vol. 380, no. 1, pp. 121-34.
King, AD, Alexander, LV & Donat, MG 2013, 'Asymmetry in the response of eastern Australia extreme rainfall to low-frequency Pacific variability', *Geophysical Research Letters*, vol. 40, no. 10, pp. 2271-7.
King, AD, Klingaman, NP, Alexander, LV, Donat, MG, Jourdain, NC & Maher, P 2014, 'Extreme rainfall variability in Australia: Patterns, drivers, and predictability', *Journal of Climate*, vol. 27, no. 15, pp. 6035-50.
Kirono, DG, Chiew, FH & Kent, DM 2010, 'Identification of best predictors for forecasting seasonal rainfall and runoff in Australia', *Hydrological Processes*, vol. 24, no. 10, pp. 1237-47.
Klingaman, NP, Woolnough, S & Syktus, J 2013, 'On the drivers of inter-annual and decadal rainfall variability in Queensland, Australia', *International Journal of Climatology*, vol. 33, no. 10, pp. 2413-30.
Koenker, R 2011, 'Additive models for quantile regression: Model selection and confidence bandaids', *Brazilian Journal of Probability and Statistics*, vol. 25, no. 3, pp. 239-62.
Koenker, R & Bassett, G 1978, 'Regression Quantiles', *Econometrica*, vol. 46, no. 1, pp. 33-50.
Komunjer, I 2013, 'Quantile prediction', in *Handbook of Economic Forecasting*, Elsevier, vol. 2, pp. 961-94.
Kraus, D & Czado, C 2017, 'D-vine copula based quantile regression', *Computational Statistics & Data Analysis*, vol. 110, pp. 1-18.
Larsen, R, Leatham, D & Sukcharoen, K 2015, 'Geographical diversification in wheat farming: a copula-based CVaR framework', *Agricultural Finance Review*, vol. 75, no. 3, pp. 368-84.
Li, Q, Lin, J & Racine, JS 2013, 'Optimal bandwidth selection for nonparametric conditional distribution and quantile functions', *Journal of Business & Economic Statistics*, vol. 31, no. 1, pp. 57-65.
Lo, F, Wheeler, MC, Meinke, H & Donald, A 2007, 'Probabilistic forecasts of the onset of the north Australian wet season', *Monthly Weather Review*, vol. 135, no. 10, pp. 3506-20.
Martin, SW, Barnett, BJ & Coble, KH 2001, 'Developing and pricing precipitation insurance', *Journal of Agricultural and Resource Economics*, pp. 261-74.
McBride, JL & Nicholls, N 1983, 'Seasonal relationships between Australian rainfall and the Southern Oscillation', *Monthly Weather Review*, vol. 111, no. 10, pp. 1998-2004.
McKeon, G, Hall, W, Henry, B, Stone, G & Watson, I 2004, 'Pasture degradation and recovery in Australia's rangelands: learning from history'.
Meyers, G, McIntosh, P, Pigot, L & Pook, M 2007, 'The years of El Niño, La Niña, and interactions with the tropical Indian Ocean', *Journal of Climate*, vol. 20, no. 13, pp. 2872-80.
Min, SK, Cai, W & Whetton, P 2013, 'Influence of climate variability on seasonal extremes over Australia', *Journal of Geophysical Research Atmospheres*, vol. 118, no. 2, pp. 643-54.

Nagler, T 2017, 'A generic approach to nonparametric function estimation with mixed data', *arXiv preprint arXiv:1704.07457*.
Nguyen-Huy, T, Deo, RC, An-Vo, DA, Mushtaq, S & Khan, S 2017, 'Copula-statistical precipitation forecasting model in Australia's agro-ecological zones', *Agricultural Water Management*, vol. 191.
Nguyen-Huy, T, Deo, RC, Mushtaq, S, An-Vo, D-A & 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, CC & Bhatti, MI 2012, 'Copula model dependency between oil prices and stock markets: Evidence from China and Vietnam', *Journal of International Financial Markets, Institutions and Money*, vol. 22, no. 4, pp. 758-73.
Nicholls, N, Drosdowsky, W & Lavery, B 1997, 'Australian rainfall variability and change', *Weather*, vol. 52, no. 3, pp. 66-72.
Nicholls, N, Lavery, B, Frederiksen, C, Drosdowsky, W & Torok, S 1996, 'Recent apparent changes in relationships between the El Niño-Southern Oscillation and Australian rainfall and temperature', *Geophysical Research Letters*, vol. 23, no. 23, pp. 3357-60.
Nicholson, SE & Kim, J 1997, 'The relationship of the El Niño–southern oscillation to African rainfall', *International Journal of Climatology*, vol. 17, no. 2, pp. 117-35.
Noh, H, Ghouch, AE & Bouezmarni, T 2013, 'Copula-based regression estimation and inference', *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 676-88.
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.
Power, S, Haylock, M, Colman, R & Wang, X 2006, 'The predictability of interdecadal changes in ENSO activity and ENSO teleconnections', *Journal of Climate*, vol. 19, no. 19, pp. 4755-71.
Rajeevan, M, Pai, D, Kumar, RA & Lal, B 2007, 'New statistical models for long-range forecasting of southwest monsoon rainfall over India', *Climate dynamics*, vol. 28, no. 7-8, pp. 813-28.
Ramirez, MCV, de Campos Velho, HF & Ferreira, NJ 2005, 'Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region', *Journal of Hydrology*, vol. 301, no. 1-4, pp. 146-62.
Risbey, JS, Pook, MJ, McIntosh, PC, Wheeler, MC & Hendon, HH 2009, 'On the remote drivers of rainfall variability in Australia', *Monthly Weather Review*, vol. 137, no. 10, pp. 3233-53.
Saji, N, Goswami, B, Vinayachandran, P & Yamagata, T 1999, 'A dipole mode in the tropical Indian Ocean', *Nature*, vol. 401, no. 6751, pp. 360-3.
Schallhorn, N, Kraus, D, Nagler, T & Czado, C 2017, 'D-vine quantile regression with discrete variables', *arXiv preprint arXiv:1705.08310*.
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.
Sheather, SJ & Jones, MC 1991, 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 683-90.
Sklar, M 1959, *Fonctions de répartition à n dimensions et leurs marges*, Université Paris 8.
Sraj, M, Bezak, N & Brilly, M 2015, 'Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River', *Hydrological Processes*, vol. 29, no. 2, pp. 225-38.
Ummenhofer, CC, England, MH, McIntosh, PC, Meyers, GA, Pook, MJ, Risbey, JS, Gupta, AS & Taschetto, AS 2009, 'What causes southeast Australia's worst droughts?', *Geophysical Research Letters*, vol. 36, no. 4.

Ummenhofer, CC, Sen Gupta, A, Briggs, PR, England, MH, McIntosh, PC, Meyers, GA, Pook, MJ, Raupach, MR & Risbey, JS 2011, 'Indian and Pacific Ocean influences on southeast Australian drought and soil moisture', *Journal of Climate*, vol. 24, no. 5, pp. 1313-36.
Vieux, BE, Cui, Z & Gaur, A 2004, 'Evaluation of a physics-based distributed hydrologic model for flood forecasting', *Journal of Hydrology*, vol. 298, no. 1-4, pp. 155-77.
Xu, W, Filler, G, Odening, M & Okhrin, O 2010, 'On the systemic nature of weather risk', *Agricultural Finance Review*, vol. 70, no. 2, pp. 267-84.
Zscheischler, J, Westra, S, Hurk, BJ, Seneviratne, SI, Ward, PJ, Pitman, A, AghaKouchak, A, Bresch, DN, Leonard, M & Wahl, T 2018, 'Future climate risk from compound events', *Nature Climate Change*, p. 1.

List of Figures

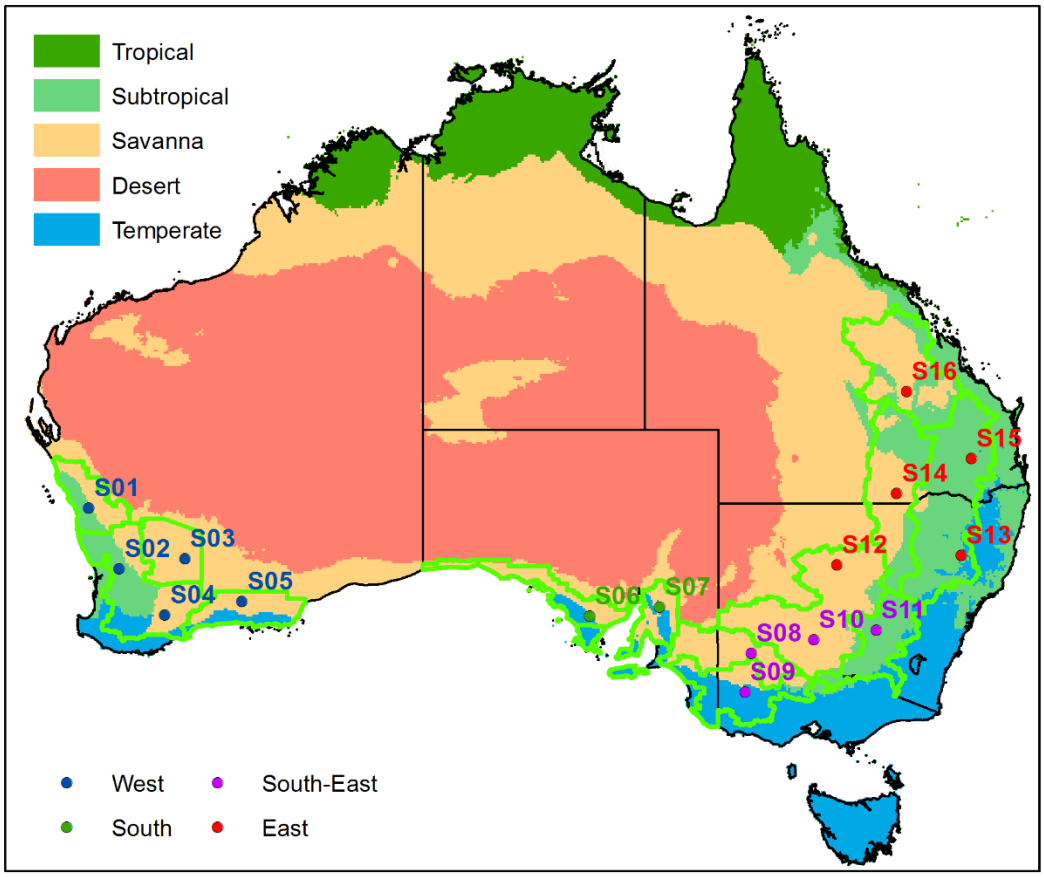


Figure 1. Selected weather stations across the Australian wheat belt (green border) spanning different climate conditions.

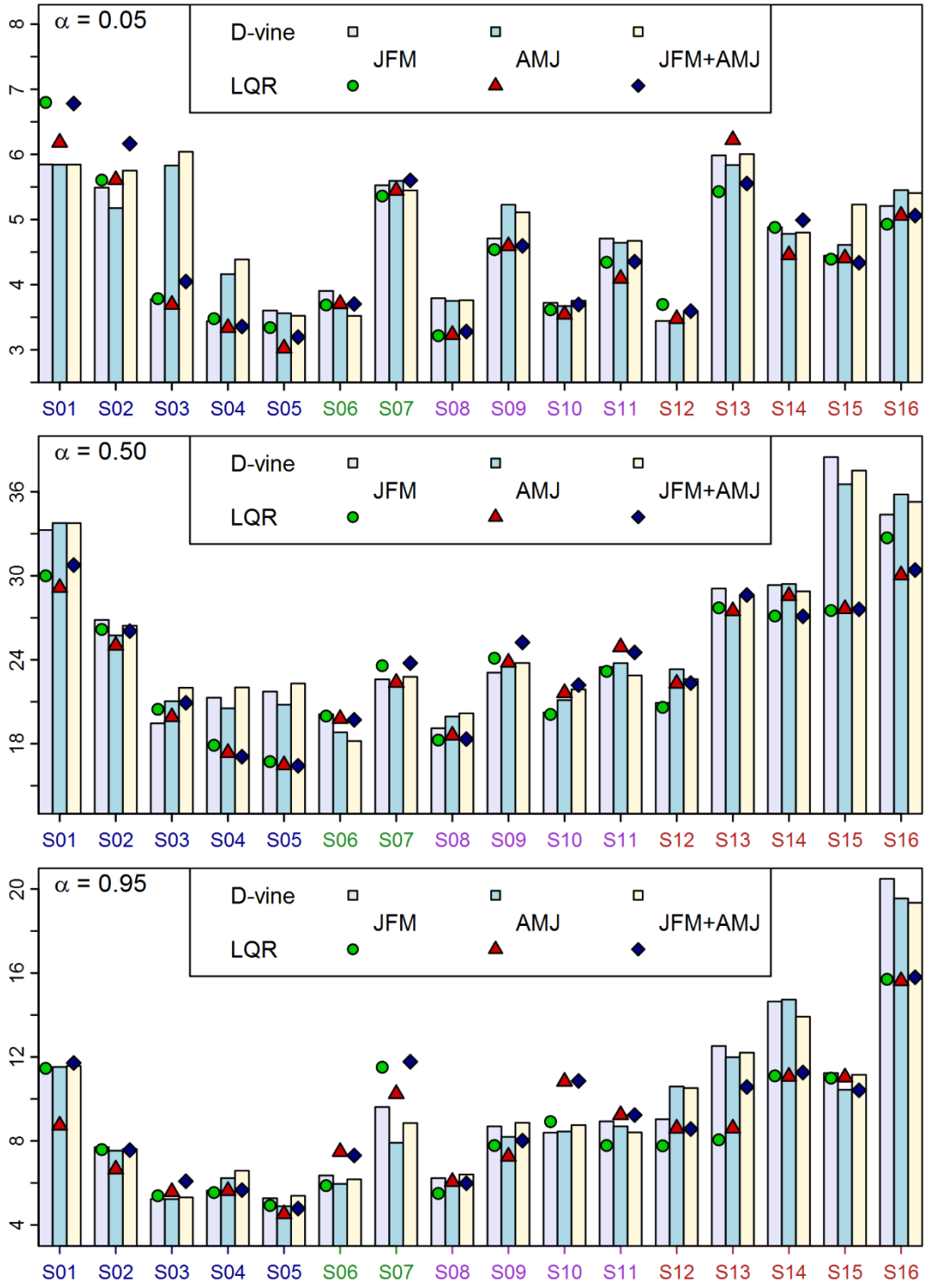


Figure 2. Averaged five-fold cross-validated tick-losses of April – June cumulative rainfall index (AMJ CRI) forecast at sixteen stations using different sets of predictors January – March (JFM), AMJ, and JFM+AMJ and using D-vine (bar charts) and linear (symbols) quantile regression (LQR) for different quantile levels.

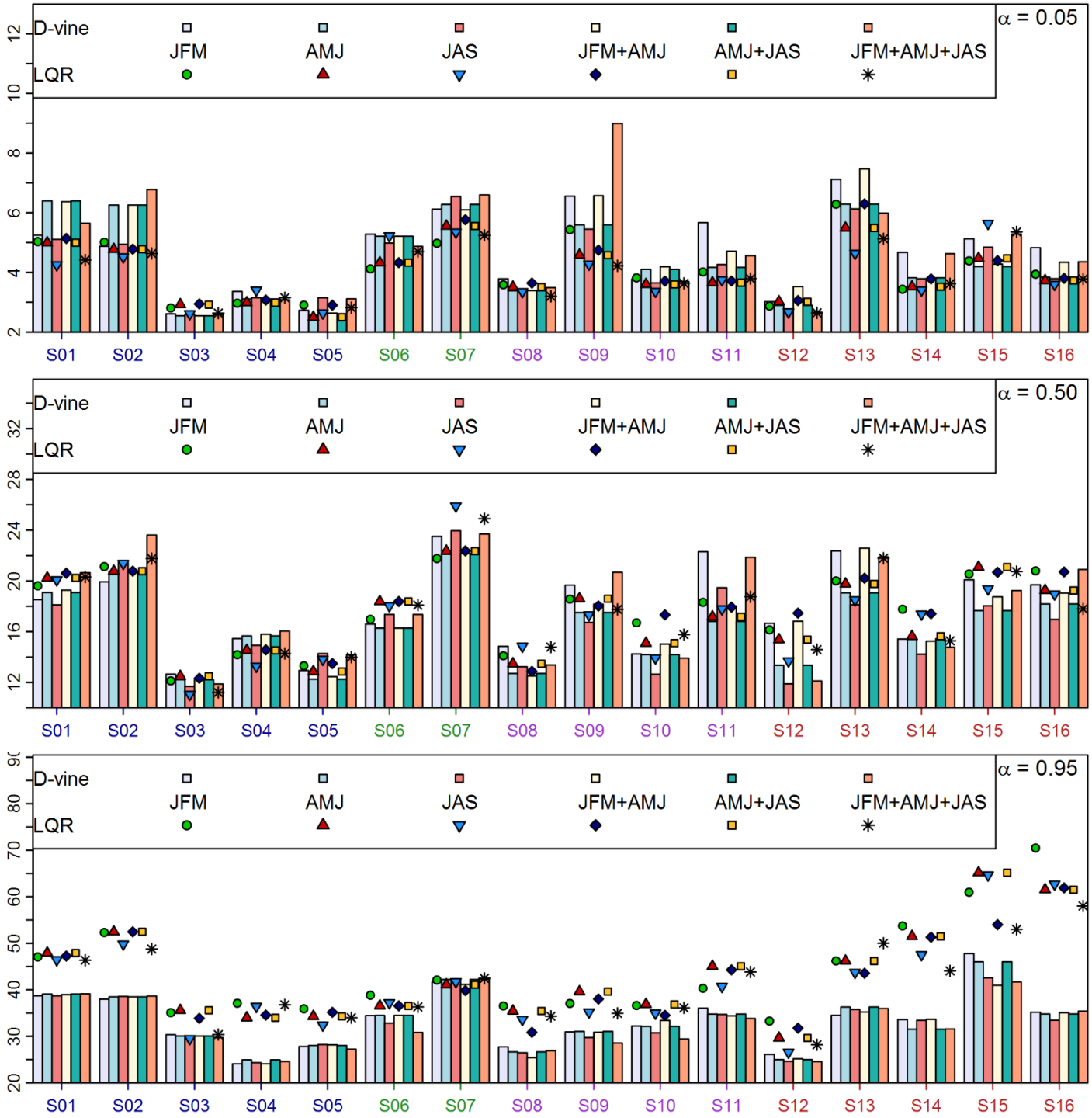


Figure 3. Averaged five-fold cross-validated tick-losses of July – September cumulative rainfall index (JAS CRI) forecast at sixteen stations using different sets of predictors January – March (JFM), April – June (AMJ), JAS, JFM+AMJ, AMJ+JAS, and JFM+AMJ+JAS and using D-vine (bar charts) and linear (symbols) quantile regression (LQR) for different quantile levels.

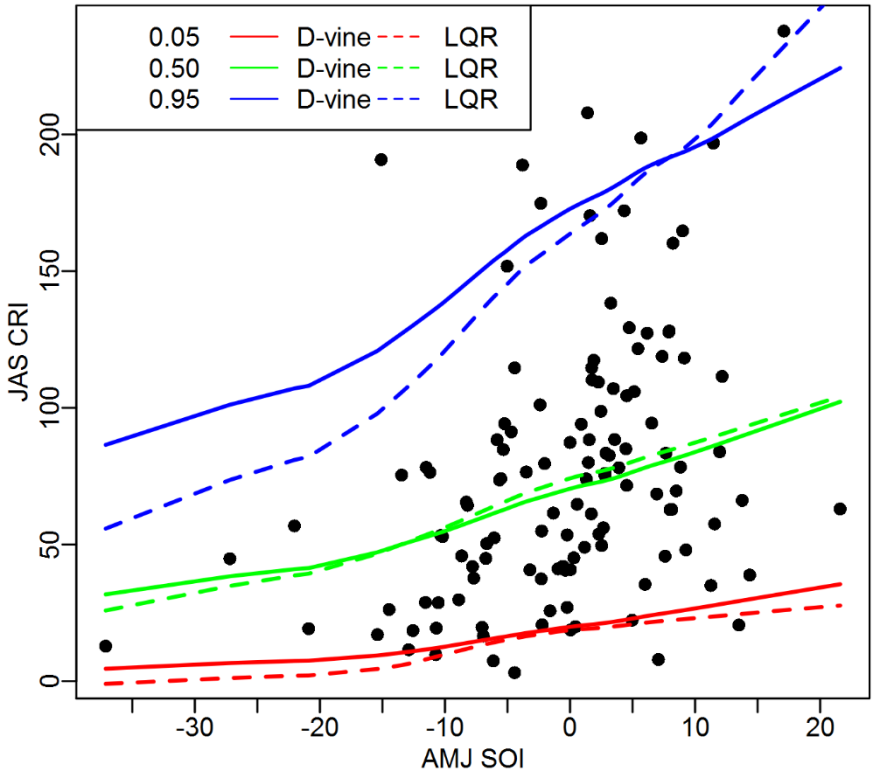


Figure 4. Exemplified scatterplot between observed July – September cumulative rainfall index (JAS CRI) and April – June SOI (AMJ SOI) (dotted points) overlaid with locally weighted regression lines of forecasted JAS CRI from five-fold cross-validation in three quantile levels (0.05, 0.50, and 0.95).

List of Tables

Table 1. Mean and standard deviation of cumulative rainfall index (CRI) in the period of April – June (AMJ) and July – September (JAS) for sixteen weather stations.

Code	*ID*	*Name*	*Coordinates*	*AMJ CRI*		*JAS CRI*	
				Mean	*Standard Deviation*	*Mean*	*Standard Deviation*
West							
S01	08088	Mingenew	115.44^0E – 29.19^0S	163.921	70.357	167.024	54.916
S02	10111	Northam	116.66^0E – 31.65^0S	159.671	58.497	180.270	55.290
S03	12074	Southern Cross	119.33^0E – 31.23^0S	98.798	45.547	89.271	34.163
S04	10627	Pingrup	118.51^0E – 33.53^0S	117.550	43.055	123.364	37.276
S05	12070	Salmon Gums	121.64^0E – 32.98^0S	100.510	39.377	103.171	33.431
South							
S06	18064	Lock	135.76^0E – 33.57^0S	116.136	47.562	145.624	45.207
S07	21027	Jamestown	138.61^0E – 33.20^0S	129.480	56.715	166.443	52.743
South-East							
S08	76047	Ouyen	142.32^0E – 35.07^0S	86.548	43.783	95.815	38.653
S09	79023	Horsham Polkemmet	142.07^0E – 36.66^0S	127.763	53.566	144.585	47.872
S10	75031	Hay	144.85^0E – 34.52^0S	98.559	51.546	95.123	41.671
S11	73000	Barmedman	147.39^0E – 34.14^0S	115.492	59.153	116.589	48.070
East							
S12	48030	Cobar	145.80^0E – 31.50^0S	84.715	53.988	76.996	37.743
S13	55054	Tamworth	150.85^0E – 31.09^0S	130.139	67.544	137.860	55.555
S14	44030	Dirranbandi	148.23^0E – 28.58^0S	96.633	71.250	77.356	50.829
S15	41023	Dalby	151.26^0E – 27.18^0S	114.245	71.076	101.434	60.020
S16	35059	Rolleston	148.63^0E – 24.46^0S	113.331	86.799	80.285	65.712

Table 2. Climate mode indices derived from NOAA Extended Reconstructed Sea Surface Temperature Anomalies (SSTA) data, version 4, and downloaded from Asia-Pacific Data Research Center (APDRC). SOI data acquired from Bureau of Meteorology, Australia (BOM).

Predictor Variables	*Description*	*Region*
Niño3.0	Average SSTA over 150^0–90^0W and 5^0N–5^0S	Pacific
Niño3.4	Average SSTA over 170^0E–120^0W and 5^0N–5^0S	Pacific
Niño4.0	Average SSTA over 160^0E–150^0W and 5^0N–5^0S	Pacific
EMI	C – 0.5 x (E + W) Where the components are average SSTA over C: 165^0E–140^0W and 10^0N–10^0S E: 110^0–70^0W and 5^0N–15^0S W: 125^0–145^0E and 20^0N–10^0S	Pacific
SOI	Pressure difference between Tahiti and Darwin as defined by Troup (1965)	Pacific
DMI	WPI – EPI Where the components are average SSTA over WPI: 50^0–70^0E and 10^0N–10^0S EPI: 90^0–110^0E and 0^0N–10^0S	Indian
II	Average SSTA over 120^0–130^0E and 0^0N–10^0S	Indian
TSI	Average SSTA over 150^0–160^0E and 30^0S–40^0S	Extratropical

Table 3. Kendall-tau correlation coefficients with significant p-values at 10% (bold) and 5% (underlined bold) significance levels between January – March (JFM) and April – June (AMJ) climate indices and AMJ cumulative rainfall index (CRI) in sixteen weather stations.

	S01	*S02*	*S03*	*S04*	*S05*	*S06*	*S07*	*S08*	*S09*	*S10*	*S11*	*S12*	*S13*	*S14*	*S15*	*S16*
JFM Niño3.0	**-0.105**	**-0.132**	0.009	-0.084	-0.069	-0.075	-0.059	-0.025	-0.036	-0.028	-0.022	0.041	0.055	-0.007	-0.019	-0.055
JFM Niño3.4	-0.091	-0.099	0.030	-0.044	-0.017	-0.075	-0.026	-0.002	-0.020	-0.003	-0.007	0.021	0.035	-0.014	-0.037	-0.053
JFM Niño4.0	-0.090	-0.099	0.047	-0.039	0.002	-0.070	-0.025	-0.001	-0.016	0.007	0.001	0.006	0.023	-0.021	-0.049	-0.059
JFM SOI	0.042	0.056	-0.030	0.005	-0.020	0.013	-0.034	-0.061	-0.059	-0.090	-0.086	**-0.103**	-0.096	-0.035	-0.013	-0.006
JFM EMI	-0.047	-0.026	0.050	0.030	0.049	-0.028	0.020	0.007	-0.025	-0.017	-0.051	-0.085	-0.016	-0.072	**-0.100**	**-0.104**
JFM DMI	0.027	0.006	0.046	0.023	-0.047	-0.005	-0.012	0.023	-0.011	0.007	0.012	0.001	-0.097	0.000	-0.052	-0.094
JFM II	-0.092	**-0.136**	0.029	**-0.128**	-0.031	0.006	0.000	0.003	-0.006	0.010	-0.002	0.081	**0.118**	0.038	0.016	0.076
JFM TSI	**-0.137**	-0.090	0.026	**-0.136**	-0.047	0.075	0.078	0.007	-0.008	0.023	-0.037	0.024	0.000	-0.042	**-0.123**	0.024
AMJ Niño3.0	-0.038	-0.071	-0.007	-0.093	-0.099	**-0.107**	**-0.103**	-0.079	**-0.139**	-0.066	-0.033	-0.012	-0.044	-0.043	-0.067	-0.057
AMJ Niño3.4	-0.067	-0.089	0.002	**-0.122**	-0.075	-0.087	-0.082	-0.032	-0.064	-0.065	-0.018	-0.018	-0.033	-0.038	-0.062	-0.092
AMJ Niño4.0	-0.086	-0.096	0.012	-0.114	-0.077	-0.077	-0.049	-0.026	-0.054	-0.086	-0.018	-0.041	-0.033	-0.047	-0.072	-0.098
AMJ SOI	**0.164**	**0.127**	0.037	**0.135**	**0.135**	**0.201**	**0.163**	**0.141**	**0.175**	**0.101**	**0.131**	0.071	**0.132**	**0.118**	**0.129**	**0.149**
AMJ EMI	-0.076	-0.054	0.003	-0.099	-0.026	0.027	0.041	0.060	0.037	-0.062	-0.029	-0.069	-0.007	-0.062	-0.055	-0.095
AMJ DMI	-0.098	-0.009	-0.014	-0.039	-0.072	**-0.109**	**-0.151**	-0.078	**-0.114**	**-0.102**	**-0.127**	-0.096	**-0.157**	**-0.123**	**-0.126**	**-0.196**
AMJ II	-0.035	-0.065	0.053	**-0.126**	0.046	0.070	0.094	0.061	0.045	0.084	0.057	**0.111**	**0.143**	0.081	-0.006	0.092
AMJ TSI	-0.062	-0.078	0.068	**-0.109**	0.060	0.068	0.083	0.069	0.039	0.081	0.068	**0.107**	0.085	**0.121**	0.034	**0.126**

Table 4. Kendall-tau correlation coefficients with significant p-values at 10% (bold) and 5% (underlined bold) significance levels between January – March (JFM), April – June (AMJ), and July – September (JAS) climate indices and JAS cumulative rainfall index (CRI) in sixteen weather stations.

	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16
JFM Niño3.0	-0.013	-0.071	-0.001	0.015	-0.035	-0.056	-0.072	-0.010	-0.046	0.036	-0.016	-0.078	-0.065	-0.065	-0.068	**-0.100**
JFM Niño3.4	-0.001	-0.057	0.007	0.000	-0.034	-0.042	-0.054	0.023	-0.016	0.046	0.019	-0.036	-0.027	-0.027	-0.034	-0.080
JFM Niño4.0	-0.004	-0.051	0.013	-0.017	-0.036	-0.036	-0.055	0.020	-0.017	0.019	0.009	-0.035	-0.026	-0.016	-0.052	-0.087
JFM SOI	-0.041	0.024	-0.051	-0.039	0.018	-0.030	-0.072	**-0.107**	-0.060	**-0.123**	-0.091	-0.027	-0.093	-0.035	-0.049	-0.022
JFM EMI	-0.023	-0.068	-0.016	-0.045	-0.035	-0.029	-0.065	0.054	0.022	-0.004	0.017	-0.013	-0.018	-0.006	-0.012	-0.082
JFM DMI	**-0.113**	-0.051	-0.019	-0.085	-0.081	-0.031	-0.008	-0.019	0.016	-0.015	-0.049	0.012	-0.026	0.004	-0.068	-0.034
JFM II	0.002	-0.001	0.067	-0.039	0.063	-0.028	0.029	0.088	0.027	**0.108**	0.069	-0.010	0.047	0.071	0.007	0.027
JFM TSI	-0.048	0.018	0.065	-0.082	0.000	-0.028	-0.003	0.039	0.034	0.077	0.058	-0.002	0.005	0.035	-0.066	-0.031
AMJ Niño3.0	**-0.106**	**-0.119**	-0.078	-0.091	**-0.128**	-0.046	-0.084	**-0.163**	**-0.145**	-0.063	**-0.137**	**-0.172**	**-0.186**	**-0.240**	**-0.242**	**-0.274**
AMJ Niño3.4	-0.080	-0.090	-0.041	-0.055	-0.080	-0.029	-0.079	-0.100	**-0.111**	-0.060	**-0.109**	**-0.138**	**-0.154**	**-0.188**	**-0.210**	**-0.240**
AMJ Niño4.0	-0.085	-0.094	-0.032	-0.070	-0.051	-0.026	-0.051	-0.065	-0.080	-0.059	-0.077	**-0.115**	**-0.119**	**-0.138**	**-0.192**	**-0.214**
AMJ SOI	0.082	**0.106**	0.071	0.044	0.048	**0.163**	**0.127**	**0.198**	**0.152**	**0.172**	**0.199**	**0.193**	**0.222**	**0.275**	**0.250**	**0.223**
AMJ EMI	-0.024	-0.008	0.030	0.007	0.010	0.003	-0.034	0.066	0.023	-0.055	-0.026	-0.042	-0.030	-0.022	-0.057	-0.068
AMJ DMI	-0.091	-0.071	-0.036	**-0.136**	-0.090	-0.084	**-0.141**	**-0.132**	**-0.117**	**-0.177**	**-0.118**	**-0.113**	**-0.127**	**-0.100**	**-0.101**	-0.072
AMJ II	-0.054	-0.060	0.028	**-0.148**	0.033	-0.013	0.067	**0.127**	0.067	**0.155**	**0.107**	0.018	0.054	0.046	-0.024	0.039
AMJ TSI	**-0.130**	-0.071	-0.018	**-0.169**	0.004	-0.004	0.023	0.059	0.033	**0.133**	**0.108**	0.024	0.035	0.053	-0.052	-0.012
JAS Niño3.0	**-0.139**	**-0.139**	-0.080	**-0.126**	-0.075	-0.081	**-0.132**	**-0.212**	**-0.149**	**-0.117**	**-0.185**	**-0.156**	**-0.231**	**-0.222**	**-0.283**	**-0.289**
JAS Niño3.4	**-0.127**	**-0.141**	-0.060	-0.100	-0.074	-0.084	**-0.112**	**-0.204**	**-0.144**	**-0.131**	**-0.206**	**-0.191**	**-0.279**	**-0.262**	**-0.342**	**-0.320**
JAS Niño4.0	**-0.102**	**-0.140**	-0.018	-0.083	-0.052	**-0.108**	-0.097	**-0.188**	**-0.151**	**-0.131**	**-0.186**	**-0.209**	**-0.274**	**-0.259**	**-0.352**	**-0.335**
JAS SOI	**0.192**	**0.195**	**0.154**	**0.118**	**0.124**	**0.225**	**0.208**	**0.299**	**0.271**	**0.249**	**0.285**	**0.290**	**0.317**	**0.310**	**0.318**	**0.351**
JAS EMI	-0.081	-0.081	0.019	-0.030	-0.024	-0.077	-0.034	-0.009	-0.037	-0.089	**-0.106**	**-0.144**	**-0.143**	**-0.124**	**-0.225**	**-0.187**
JAS DMI	**-0.158**	**-0.140**	-0.093	**-0.132**	-0.083	**-0.228**	**-0.257**	**-0.276**	**-0.249**	**-0.215**	**-0.173**	**-0.162**	**-0.196**	**-0.201**	**-0.191**	**-0.221**
JAS II	0.088	0.053	**0.159**	-0.010	**0.109**	**0.219**	**0.246**	**0.298**	**0.265**	**0.321**	**0.293**	**0.198**	**0.254**	**0.221**	**0.154**	**0.237**
JAS TSI	**-0.111**	**-0.100**	0.065	**-0.148**	0.002	-0.067	0.023	-0.011	0.003	0.093	0.081	-0.029	-0.007	-0.002	-0.075	-0.088

Appendix B

Copyright information

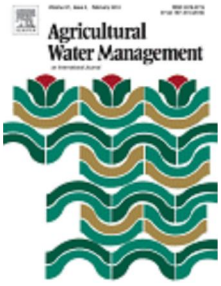
Article I: Copula-statistical precipitation forecasting model in Australia's agro-ecological zones

Thong Nguyen-Huy, Ravinesh C. Deo, Duc-Anh An-Vo, Shahbaz Mushtaq, and Shahjahan Khan. "Copula-statistical precipitation forecasting model in Australia's agro-ecological zones". *Agricultural Water Management*, 191 (2017): 153-172. [Impact Factor: 2.848, SNIP: 1.814 , Scopus Rated Q1, 93rd percentile in Water Sc & Technology].

DOI: https://doi.org/10.1016/j.agwat.2017.06.010

Copyright Clearance Center RightsLink®

Home Account Info Help



Title: Copula-statistical precipitation forecasting model in Australia's agro-ecological zones

Author: Thong Nguyen-Huy,Ravinesh C. Deo,Duc-Anh An-Vo,Shahbaz Mushtaq,Shahjahan Khan

Publication: Agricultural Water Management

Publisher: Elsevier

Date: September 2017

© 2017 Elsevier B.V. All rights reserved.

Logged in as:
Thong Nguyen
Account #:
3001304107

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Article II: Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach

Article II: ***Thong Nguyen-Huy***, Ravinesh C. Deo, Duc-Anh An-Vo, Shahbaz Mushtaq, and Shahjahan Khan. "Modelling the joint influence of multiple climate mode indices on Australian wheat yield using a vine copula-based approach." *European Journal of Agronomy*, 98 (2018): 65-81. [Impact Factor: 3.757, SNIP: 1.828, Scopus Rated Q1, 94th percentile in Agronomy & Crop Sc.].

DOI: https://doi.org/10.1016/j.eja.2018.05.006

Copyright Clearance Center RightsLink®

Home Account Info Help

Title: Modeling the joint influence of multiple synoptic-scale, climate mode indices on Australian wheat yield using a vine copula-based approach

Author: Thong Nguyen-Huy,Ravinesh C Deo,Shahbaz Mushtaq,Duc-Anh An-Vo,Shahjahan Khan

Publication: European Journal of Agronomy

Publisher: Elsevier

Date: August 2018

© 2018 Elsevier B.V. All rights reserved.

Logged in as:
Thong Nguyen
Account #:
3001304107

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Article IV: Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management

Thong Nguyen-Huy, Ravinesh C. Deo, Shahbaz Mushtaq, Jarrod Kath and Shahjahan Khan. "Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management". *Weather and Climate Extremes*. [Impact Factor = 4.21, SNIP: 2.428, Scopus Rated Q1, 98th percentile in Geography, Planning & Development].

DOI: https://doi.org/10.1016/j.wace.2018.07.002

Copyright Clearance Center RightsLink®

Home Account Info Help

ELSEVIER

Title: Copula-based agricultural conditional value-at-risk modelling for geographical diversifications in wheat farming portfolio management

Author: Thong Nguyen-Huy,Ravinesh C. Deo,Shahbaz Mushtaq,Jarrod Kath,Shahjahan Khan

Publication: Weather and Climate Extremes

Publisher: Elsevier

Date: Available online 20 July 2018

© 2018 Published by Elsevier B.V.

Logged in as:
Thong Nguyen
Account #:
3001304107

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com