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/ 

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