

Contents lists available at ScienceDirect

# Agricultural and Forest Meteorology

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

# Quantifying the impacts of pre-occurred ENSO signals on wheat yield variation using machine learning in Australia



Bin Wang<sup>a,b,#</sup>, Puyu Feng<sup>a,b,#</sup>, Cathy Waters<sup>c</sup>, James Cleverly<sup>d</sup>, De Li Liu<sup>b,e</sup>, Qiang Yu<sup>a,d,f,\*</sup>

<sup>a</sup> State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest A&F University, Yangling 712100, Shaanxi, China

<sup>b</sup> NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Pine Gully Road Wagga Wagga, NSW 2650, Australia

<sup>c</sup> NSW Department of Primary Industries, Dubbo NSW 2830, Australia

<sup>d</sup> School of Life Sciences, Faculty of Science, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

<sup>e</sup> Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia

<sup>f</sup> College of Resources and Environment, University of Chinese Academy of Science, Beijing 100049, China

# ARTICLE INFO

Keywords: spring rainfall wheat yield variation ENSO climate indices random forest

# ABSTRACT

Australia is one of the top wheat exporting countries in the world and the reliable prediction of wheat production plays a key role in ensuring regional and global food security. However, wheat yield in Australia is highly exposed to the impacts of climate variability, especially seasonal rainfall, as wheat is mostly grown in the drylands. Previous studies showed that El Niño Southern Oscillation (ENSO) has a strong influence on Australia's climate and found the ENSO-related phenomena have prognostic features for future climatic conditions. Therefore, we examined the predictability of state-scale variation in Australian wheat yields based on ENSOrelated large-scale climate precursors using machine learning techniques. Here, we firstly established a set of random forest (RF, a machine learning method) models based on pre-occurred climate indices to forecast spring rainfall for the four major wheat producing states of Australia, the forecasted rainfall was then combined with selected precedent climate drivers to predict yield variations using another set of RF models for each state. We explored the most influential variables in determining spring rainfall and yield variation. We found that the first set of RF models accounted for 43-59% of the change in spring rainfall across the four states. By incorporating forecasted spring rainfall with selected ENSO climate indices, the RF model accounted for 33-66% of the variation in yield which was greater than the 22-50% of yield variations explained by ENSO-related indices alone. The results suggest that wheat yield variation at a state level could be reliably forecasted at lead-times of three months prior to the commencement of harvest. We also found that forecasted spring rainfall and precedent Southern Oscillation Index (SOI) in July were the most important factors in estimation of crop yield in the winter dominant rainfall states. ENSO climate indices are easy to obtain and can be rapidly used to drive the forecasting model. Therefore, we believe the proposed models for predicting wheat yield variations at three-month lead time would be helpful for state governments and policy makers to develop effective planning to reduce monetary loss and ensure food security.

## 1. Introduction

The El Niño-Southern Oscillation (ENSO) is a quasi-periodic interannual variation in atmospheric and oceanic circulation patterns occurring across the equatorial Pacific Ocean. In general, ENSO phenomena can be categorized into three phases, El Niño (the warm phase), La Niña (the cool phase), and Neutral. El Niño and La Niña are opposite phases that swing back and forth every 3-7 years on average (https://www.climate.gov/enso), leading to substantially different climate conditions in affected zones. Rain-fed crop production in affected zones is vulnerable to the change in ENSO phases as this alters the magnitude of rainfall. Teleconnections between ENSO states and crop production have been reported worldwide in the late 1980s and 1990s, e.g., Argentina (Podestá et al., 1999), Australia (Nicholls, 1986), Mexico (Adams et al., 2003), North America (Phillips et al., 1999) and Zimbabwe (Cane et al., 1994; Phillips et al., 1998). There have been

E-mail address: yuq@nwafu.edu.cn (Q. Yu).

https://doi.org/10.1016/j.agrformet.2020.108043

<sup>\*</sup> Corresponding author: State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>&</sup>lt;sup>#</sup> Bin Wang and Puyu Feng contributed equally to this work.

Received 23 November 2019; Received in revised form 28 April 2020; Accepted 20 May 2020 0168-1923/ © 2020 Elsevier B.V. All rights reserved.

significant progress towards a more robust understanding of the predictive ability of ENSO teleconnections, and the effects of ENSO-induced climate variability on crop yield have been further examined more recently in America (Anderson et al., 2017; Yu et al., 2018), Australia (Nguyen-Huy et al., 2018; Yuan and Yamagata, 2015) and China (Liu et al., 2014; Shuai et al., 2016).

30-50% of wheat yield variation in Australia can be attributed to climate variability (Wang et al., 2015) and yield is particularly susceptible to adverse climate conditions. Seasonal rainfall variation and water shortage, in particular during spring are the major causes of the variation in Australian wheat yield (Anwar et al., 2008; Yu et al., 2014). Therefore, early detection of spring rainfall variation can assist in improving yield by targeted management to minimize the potential losses due to unfavourable seasonal conditions. There are currently two government-based seasonal rainfall forecasting programs widely used in Australia (Abbot and Marohasy, 2014). The first one is based on the strong relationship between ENSO phenomena and seasonal rainfall variability. Here, the ENSO cycle is divided into 5 phases (rapid rising, rapid falling, consistently negative, consistently positive and near zero) according to the change of Southern Oscillation Index (SOI) over two consecutive months (Stone et al., 1996). The possibility of exceeding the median (50th percentile) of rainfall distribution for next three months during a certain phase can be calculated based on historical observations of the same phase (Long Paddock, 2019). This SOI phase forecasting program has already been used widely by the Queensland State government to assist in making better tactical management decisions, resulting in some economic benefits to Queensland farmers (Hammer et al., 1996). A second seasonal forecasting approach has been developed by the Australian Bureau of Meteorology (BOM) using various rainfall predicting systems, e.g. the Predictive Ocean Atmosphere Model for Australia (POAMA) in 2013 and the Australian Community Climate Earth-System Simulator-Seasonal (ACCESS-S) in 2018, which both provide the probability of above median rainfall for future months (Hudson et al., 2017). However, both models are physicsbased and require a large amount of climatic and geographical input information to develop rainfall outlooks.

Large-scale climatic phenomena will influence future monthly rainfall as it has been demonstrated by numerous researchers (Ashok et al., 2003; Cai et al., 2011; Kirono et al., 2010; Risbey et al., 2009). Recently, machine learning techniques have proven to be powerful tools for forecasting seasonal rainfall using large-scale climate precursors, e.g. SOI, Multivariate ENSO Index (MEI), sea surface temperature based index of ENSO (NINO3.4) and Indian Ocean Dipole (IOD) (Abbot and Marohasy, 2014; Hossain et al., 2019; Mekanik et al., 2016). For example, Abbot and Marohasy (2014) used Artificial Neural Networks (ANNs) with different climate indices to forecast monthly rainfall in three geographically distinct regions in Queensland and found ANN produced more practical and reliable rainfall forecasts than the POAMA model used by BOM. Mekanik et al. (2016) developed adaptive network-based fuzzy inference systems (ANFIS) models in south-east Australia to forecast spring rainfall with different climate signals and found that ANFIS performed better compared to conventional ANN. Therefore, machine learning-based seasonal rainfall forecasting system is a promising tool as this approach is able to achieve high accuracy using minimal input variables with less developing time (Mekanik et al., 2016).

The El Niño state corresponds to weaker trade winds and low rainfall in Australia. A seesaw shift in rainfall may be found in the west Pacific. Therefore, spatial variation of ENSO signals plays an important role in determining agricultural production patterns in Australia. Knowledge of the relationship between large-scale climate signals and crop yields is essential to provide useful information to the agriculture sector as this will allow growers to cope with potential negative impacts of climate variability earlier (Nguyen-Huy et al., 2018; Podestá et al., 2002; Zhang et al., 2008). Numerous methods including simulation modelling analysis, linear regression analysis or probability/cluster analysis have been applied to investigate the impacts of ENSO events (Meinke et al., 1996; Nguyen-Huy et al., 2018; Potgieter et al., 2002; Potgieter et al., 2005; Zhang et al., 2008). For example, using agroclimatic model simulated wheat yield, Potgieter et al. (2002, 2005) proposed that the strength of El Niño years could be categorized into three patterns in the Australian wheat belt. The relationship between long-term detrended yield and SOI may provide direct clues of ENSO impacts. Recently, Nguyen-Huy et al. (2018) developed a statistical regression model to assess the effects of multiple large-scale climate indices on variation in Australian wheat yields between 1983-2013. They found fluctuations in the Indian Ocean had major effects on wheat vield in all states except Western Australia, and the impacts from oceanic conditions in the Pacific were much stronger in Oueensland. A similar conclusion was supported by Yuan and Yamagata (2015). This is because ENSO effects in Australia are due to direct baroclinic effects in Queensland, but indirectly influence the moisture through Rossby wave trains elsewhere (Cleverly et al., 2016). These studies provided the basis for quantifying the impacts of different large-scale climate indices on variation in Australian wheat yields. However, unlike ENSO-related seasonal rainfall forecasts, wheat yield forecasts at a long lead time (3 months) relying on advanced machine-learning techniques and multiple ENSO climate indices, like NINO3.4, MEI, SOI and SOI phase, have not been fully investigated.

The main objectives of our study were to (1) characterize the relationship between ENSO-related large scale climate precursors and spring rainfall as well as wheat yield variation at a state level (2) identify the relative contribution of antecedent ENSO-related climate signals on spring rainfall and yield variability and (3) quantify the dependence of the most influential variables in determining rainfall and yield variability. The major contribution of this research was to establish and validate the suitability of a machine learning approach for the forecasting of wheat yield based on large-scale ENSO climate signals. We expect improving understanding in the relationship of yield and climate drives to be able to assist state government and policymakers in making decisions. This paper is organized in five sections. The introduction is followed by methodology in section 2, which describes the data, study area and includes a brief introduction of the random forest model as well as the forecast verification metrics that are employed to evaluate the generated predictions. Results and discussion are shown in section 3 and 4 to present our model performance and make comparisons with previous studies. Finally, the major conclusions are provided in section 5.

#### 2. Materials and methods

## 2.1. Historical state yield and rainfall data

Our study was conducted across four main wheat production states, Queensland (QLD), New South Wales (NSW), Victoria (VIC) and South Australia (SA), in south and east Australia (Fig. 1). We did not include Western Australia because previous study has shown that the impacts of ENSO are weak (Rimmington and Nicholls, 1993) with ENSO having only a direct effect on Australian rainfall within 25 degrees of the equator. This means ENSO's baroclinic (direct) effects are confined to near-tropical eastern Australia (Cai et al., 2011). Nationally, the Australian cropping areas or croplands within the four states are confined to a relatively narrow band along the south and east coast (129.0°-152.5°E and 21.0°-38.5°S) with a Mediterranean or temperate climate. Annual wheat production across the four states contributes more than 50% of Australia's total wheat production (ABS, 2019). Wheat production occurs under rain-fed conditions with sowing in April to June and harvesting from November to December. Historical long-term annual wheat yields from 1891 to 2016 across the four states were derived from Australia Bureau of Statistics (ABS, 2019). State wheat yield has significantly increased with time (Fig. 1) due to technological innovation in cultivar improvement, fallowing, rotation with lucerne, and



**Fig. 1.** The spatial extent of the south-eastern Australian wheat belt and long-term trends in annual wheat yield (kg ha<sup>-1</sup>) of four major wheat-producing states: Queensland (QLD), New South Wales (NSW), Victoria (VIC) and South Australia (SA) in 1891-2016. The black line in each chart denotes linear regression line of yield series. Regression equation and R<sup>2</sup> are given inside each chart. The red line represents nine-year moving average yield series in each chart.

fertiliser application, as well as use of machinery and technology advances over the past hundred years. Generally, across Australia there has been a great increasing trend in wheat yields in all states except Queensland (Fig. 1).

To separately characterize the relationship between climatic factors and wheat yield, yield trend caused by factors other than climate need to be excluded. Numerous methods have been widely used to remove yield trend caused by non-climatic factors (Lu et al., 2017; Wang et al., 2015; Yu et al., 2001). In our study, wheat yields were detrended by the moving average method (Reilly et al., 2003). This method is data selfadaptive and can identify the overall trend and pattern in a time series by smoothing high-frequency variation and irregular roughness. We adopted this method by calculating the residual yields from observations and a nine-year backward moving average, which meant all values from the previous nine years were averaged. The largest variation of detrended yield deviations was found in NSW and VIC, following by QLD and SA (Fig. A1b). Detrended yields were highly correlated with state pairs except VIC/SA versus QLD (Fig. A2). It was worth mentioning that yield anomalies were largely caused by drought events influenced by strong El Niño events. The years with yield reduction matched well with drought occurrence (e.g. the years of 1915, 1945, 1968, 2003, 2007 and 2008). In addition, La Niña years with severe storm and flood events have adverse influence on yield as well.

Monthly rainfall anomaly data during the 1900-2018 in four states were obtained from Australian Bureau of Meteorology (BOM, 2019). ENSO variability affects crop production mainly through its impacts on climatic variability (Lee et al., 2018). In the Australian wheat belt, austral spring (September to November) rainfall variation has been shown to be one of the most important variables in determining yield, especially in winter dominant rainfall regions, e.g. NSW, VIC and SA (Feng et al., 2018; Wang et al., 2015; Yu et al., 2014). Thus, we adopted austral spring rainfall anomalies (1900-2018) for the four states (Fig. A1) and used them in subsequent analysis. Rainfall anomalies were calculated based on a 30-year average climatology (1961-1990) (http://www.bom.gov.au/climate/change/index.shtml#tabs = Tracker). Long-term spring rainfall anomalies were highly correlated between each two states with SA having the lowest deviation of rainfall

# 2.2. Historical large-scale climate signals

departure (Figs. A1a and A2).

Previous studies have found ENSO-related large-scale climate indices in Pacific regions had strong effects on seasonal rainfall and yield variability in south and east Australia (Anwar et al., 2008; Chiew et al., 1998; Kirono et al., 2010). Thus, three different types of indicators (SOI, NINO3.4 and MEI) representing the ENSO phenomena, were selected to investigate their effects in this study. ENSO's features are usually indicated by two kinds of fluctuations originating from the Pacific, i.e. sea level pressure fluctuations and sea surface temperature fluctuations. SOI is a measurement of the anomalies of sea level pressure from Darwin and Tahiti, whereas NINO3.4 indicates sea surface temperature anomalies measured from the equatorial Pacific Ocean (5°N-5°S, 170°W-120°W). MEI is a multivariate index that combines both sea level pressure and sea surface temperature as well as atmospheric anomalies to capture a more holistic representation of ENSO events (Wolter and Timlin, 1998). Additionally, we also used SOI phase (Stone et al., 1996), a derivative of the SOI index, which was derived from the values of SOI during two consecutive months. Table 1 shows the detailed description on the selected indices.

Monthly SOI data sources included the values and phases of SOI from 1900 to 2018. They were downloaded from the Queensland government Long Paddock website (Long Paddock, 2019). Monthly NINO3.4 and MEI were extracted from the Climate Data Guide (CDS, 2019).

Table 1

ENSO-related large-scale climate indices used in our study.

Indices	Definition	Data source
SOI SOI phase	Anomalies of the seas level pressures from Darwin to Tahiti The phases of the SOI were defined by Dr Roger Stone, who used cluster analysis to group all sequential two-month pairs of the SOI into five clusters (consistently negative, consistently positive, falling, rising, consistently near zero) (Stone et al., 1996).	https://www.longpaddock.qld.gov.au/soi/soi-data-files/ https://www.longpaddock.qld.gov.au/soi/soi-data-files/
NINO3.4	Sea surface temperature anomalies over 5°N-5°S and 170°W-120°W	https://climatedataguide.ucar.edu/climate-data/nino-sst- indices-nino-12-3-34-4-oni-and-tni
MEI	Using 6 variables as proxies for ENSO relevant atmosphere and ocean conditions (Wolter and Timlin, 1998).	https://climatedataguide.ucar.edu/climate-data/ multivariate-enso-index

## 2.3. Random forest model

Random forest (RF) was used as the regression method to quantify the relationship between ENSO-related large-scale climate indices and austral spring rainfall anomaly as well as wheat yield for each of the four states in south and east Australia. RF is a tree-based ensemble machine learning approach which can be used to build predictive models for both classification and regression purposes Breiman, 2001). An ensemble approach is an algorithm that obtains averaged results from multiple learning models. RF first builds a forest of decision trees within the training procedure. Each tree is independently created based on randomized subsets of the predictors generated from a bootstrap aggregating method (Heung et al., 2014). This method enables the RF to avoid overfitting in comparison with decision trees. All trees in the forest grow to maximum size without pruning and the average of the outputs from all trees is regarded as the final outcome (Cutler et al., 2007). Thus, the RF is capable of effectively reducing the variance as it is a majority-votes model. RF has been used in various agrometeorological studies and shown better performance in disentangling the complex relationships between crop yield and climate factors in comparison to conventional linear models (Feng et al., 2018; Feng et al., 2019a). Additionally, RF has only two parameters, ((1) m<sub>try</sub> (the number of randomly selected predictor variables at each node) and (2)  $n_{\mathrm{tree}}$  (the number of trees to grow in the forest). To fit a RF regression model, default values of  $\ensuremath{m_{try}}$  (one third of the total number of the predictor variables) was commonly used (Ließ et al., 2016; Were et al., 2015; Yang et al., 2016). The default of n<sub>tree</sub> was 500, which has been proven to be insufficient to yield stable results for estimating variable importance (Grimm et al., 2008). Here we used  $n_{tree}$  with 1000 (Guo et al., 2015; Ließ et al., 2016). There were two reasons that we did not tune the parameters of  $m_{try}$  and  $n_{tree}\xspace$  Firstly, we have conducted preliminary analyses by adopting a trial and error method to determine the value of m<sub>trv</sub> as our dataset were not large. We found that the model performance did not increase very much compared to that with default m<sub>trv</sub>. This was consistent with previous studies showing the parameters of RF are not overly sensitive to the particular values they take (Ahmad et al., 2018; Immitzer et al., 2012). Secondly, we developed two kinds of RF models for yield prediction at each state, i.e. with and without forecasted spring rainfall anomaly. To make the models comparable, we used same values of mtry and ntree for each model to avoid the difference of model performance caused by RF parameters.

A useful characteristic of the RF is the ability to evaluate the relative importance of each predictor variable in the model. We adopted the accuracy-based importance metric in the RF model. During model construction, each tree had its own out-of-bag sample of data, which was left out of the bootstrap samples (around 33% of the total training data) (Wang et al., 2018a). The mean decrease in prediction accuracy when the values of a variable in the OOB sample were randomly shuffled, was defined as the importance value of the variable (Heung et al., 2014). It was expressed as the mean square error (MSE<sub>OOB</sub>) as follows

$$MSE_{OOB} = \frac{1}{n} \sum_{k=1}^{n} (O_i - \bar{P}_{kOOB})^2$$
(1)

where n denotes the number of observations,  $O_i$  indicates observed value, and  $\bar{P}_{kOOB}$  represents the average of all OOB predictions across all trees.

# 2.4. Model performance evaluation

The detailed forecasting scheme is shown in the workflow chart (Fig. 2). Given yield is sensitive to spring rainfall, we firstly used the RF to build spring rainfall forecast model with different pre-occurred climate indicators (SOI, SOI phase, NINO3.4 and MEI) that measured ENSO phenomena. Then we used forecasted spring rainfall (FSR) as an additional predictor to integrate with significantly yield-correlated ENSO indicators to drive RF to forecast yield. For model calibration and validation, 80% of each dataset was randomly selected for model training and the rest (20%) of the dataset was used for model performance evaluation. This procedure was implemented for 100 times to evaluate the stability of each model. We used two performance measurements, coefficients of determination ( $R^2$ ) and root mean squared error (RMSE), which are defined as follows:

$$R^{2} = \left(\frac{\sum_{i=1}^{n} (O_{i} - \bar{O})(P_{i} - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_{i} - \bar{O})^{2}} \sqrt{\sum_{i=1}^{n} (P_{i} - \bar{P})^{2}}}\right)^{2}$$
(2)

RMSE = 
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2}$$
 (3)

where  $P_i$  and  $O_i$  denote the predicted and observed spring rainfall or detrended yield;  $\bar{P}$  and  $\bar{O}$  represent the means for the predicted and observed spring rainfall or detrended yield; n is the number of samples. In general, the model with higher  $R^2$  and lower RMSE is identified to be the more accurate model.

# 3. Results

# 3.1. Correlation of precedent ENSO climate indices with spring rainfall and wheat yield variation

Linear correlations between austral spring rainfall anomaly and quantitative climatic predictors were derived and are shown in Table 2. All three precedent ENSO indices in June, July and August were highly (p < 0.01) correlated with rainfall anomaly across each of the four states except SOI of June in SA. Overall, spring rainfall was positively correlated with June, July and August SOI, but negatively correlated with NINO3.4 and MEI in June-August. Previous studies identified that spring rainfall had significant correlations with pre-occurred ENSO climate signals at a regional and site level (Hossain et al., 2018; Mekanik et al., 2016; Nguyen-Huy et al., 2017). This was further explored in our correlation analysis at a state level.

We also investigated the relationship between yield variation and precedent SOI, MEI and NINO3.4 at each state (Table 3). A Pearson

# Spring rainfall forecasting model



Fig. 2. An overview of the workflow to develop a methodological approach to forecast austral spring rainfall and wheat yield variation based on random forest model with large-scale ENSO climate precursors.

#### Table 2

Relationships between ENSO climate indices in austral winter (June-August) and spring (September-November) rainfall anomaly across four Australian states. Values are Pearson correlation coefficients of for each ENSO climate index.

Precedent climate indices	States				
	QLD	NSW	VIC	SA	
SOI <sub>Jun</sub>	0.33**	0.25**	0.23*	0.17	
SOI <sub>Jul</sub>	0.51**	0.43**	0.42**	0.41**	
SOI <sub>Aug</sub>	0.53**	0.44**	0.39**	0.34**	
NINO3.4 <sub>Jun</sub>	-0.44**	-0.38**	-0.31**	-0.26**	
NINO3.4 <sub>Jul</sub>	-0.50**	-0.45**	-0.42**	-0.32**	
NINO3.4 <sub>Aug</sub>	-0.49**	-0.43**	-0.44**	-0.36**	
MEI <sub>Jun</sub>	-0.35**	-0.30**	-0.26**	-0.18*	
MEI <sub>Jul</sub>	-0.47**	-0.41**	-0.37**	-0.30**	
MEI <sub>Aug</sub>	-0.53**	-0.47**	-0.44**	-0.35**	

\* p<0.05 (significant), \*\* p<0.01 (highly significant)

correlation analysis was conducted between yield variation and monthly values of these climate signals from January to August (prior to grain forming stage). Yield variation had significantly (p<0.05) positive relationships with SOI in the period from May to August across four states. We found significantly (p<0.05) negative correlations between yield and MEI and NINO3.4 of July and August in QLD, NSW and VIC. In contrast, only August MEI and NINO3.4 were significantly correlated with yield variation in SA. We also found the weakest correlation was around March and April in all cases. The ENSO climate indices that had statistically significant correlations with yield were further used as input variables in the subsequent random forest yield forecasting models. We also included monthly SOI phases corresponding with SOI in the same month. The dimension and predictors used in model construction varied over states.

# 3.2. Spring rainfall anomaly and detrended yield forecasting

To assess the uncertainty of RF model based on different training data, boxplots were employed to visualise the distribution of  $R^2$  and RMSE based on the 100 runs for spring rainfall anomaly and detrended yield over four states, respectively. The model uncertainty could be due to the variation in observed values for target variables (Wang et al., 2018b). The performance of RF for forecasting spring rainfall is shown in Fig. 3. The austral winter (Jun-Aug) ENSO indices had acceptable

#### Table 3

Relationships between yield variation and ENSO climate indices three months prior to harvest (November) over four states. Values are Pearson correlation coefficients for each of ENSO climate index.

Precedent climate indices	States			
	QLD	NSW	VIC	SA
SOI <sub>Jan</sub>	-0.006	-0.194*	-0.177	-0.134
SOI <sub>Feb</sub>	-0.011	-0.256**	-0.222*	-0.196*
SOI <sub>Mar</sub>	0.096	-0.074	-0.037	-0.161
SOI <sub>Apr</sub>	0.120	0.036	0.053	0.011
SOI <sub>May</sub>	0.293**	0.276**	0.248**	0.191*
SOI <sub>Jun</sub>	0.206*	0.209*	0.208*	0.237*
SOI <sub>Jul</sub>	0.325**	0.383**	0.345**	0.362**
SOI <sub>Aug</sub>	0.264**	0.304**	0.282**	0.252**
NINO3.4 <sub>Jan</sub>	0.025	0.179	0.090	0.096
NINO3.4 <sub>Feb</sub>	0.004	0.165	0.074	0.074
NINO3.4 <sub>Mar</sub>	-0.033	0.178	0.083	0.097
NINO3.4 <sub>Apr</sub>	-0.143	0.103	0.023	0.077
NINO3.4 <sub>May</sub>	-0.285**	-0.055	-0.107	0.019
NINO3.4 <sub>Jun</sub>	-0.334**	-0.207*	-0.208*	-0.105
NINO3.4 <sub>Jul</sub>	-0.268**	-0.237*	-0.259**	-0.157
NINO3.4 <sub>Aug</sub>	-0.257**	-0.295**	-0.320**	-0.222*
MEI <sub>Jan</sub>	0.029	0.164	0.083	0.090
MEI <sub>Feb</sub>	0.026	0.185*	0.096	0.095
MEI <sub>Mar</sub>	0.001	0.176	0.084	0.103
MEI <sub>Apr</sub>	-0.082	0.121	0.055	0.104
MEI <sub>May</sub>	-0.156	0.034	-0.019	0.044
MEI <sub>Jun</sub>	-0.213*	-0.074	-0.139	-0.053
MEI <sub>Jul</sub>	-0.271**	-0.208*	-0.228*	-0.145
MEI <sub>Aug</sub>	-0.283**	-0.267**	-0.281**	-0.197*

\* p<0.05 (significant), \*\* p<0.01 (highly significant).

forecasting capability of spring rainfall with mean  $R^2$  values of 0.59, 0.46, 0.44 and 0.43 for QLD, NSW, VIC and SA, respectively (Fig. 3a). The  $R^2$  was highest in QLD and lowest in SA, which suggests the greatest impact of Pacific Ocean conditions in QLD and weakest influence in SA. The values of RMSE were also calculated for each state for further assessment of model accuracy. The models with higher  $R^2$  normally had the lower RMSE. This was the case in QLD, NSW and VIC. For example, the mean values of RMSE from the 100 runs in QLD was 37 mm, followed by 40 mm and 44 mm in NSW and VIC respectively (Fig. 3b). By contrast, SA had the lowest RMSE of 22 mm mainly due to its smaller rainfall variation in spring (Fig. A1a).

The performance of RF model for yield prediction varied over states



**Fig. 3.** Evaluation of random forest (RF) model to forecast austral spring rainfall anomaly (mm) across four states. The coefficient of determination ( $R^2$ ) and root mean squared error (RMSE, mm) are used to evaluate model performance. The black lines within the box indicate the medians with 100 runs while crosshairs indicate means. Box boundaries indicate the 25th and 75th percentiles, whiskers below and above the box indicate the 10th and 90th percentiles.

(Fig. 4), which may be attributed to heterogeneous contribution from ENSO climate drivers in determining yield. Our results showed that the RF model in predicting yield variations with pre-occurred climate drivers captured 22-50% of the variation in yield and resulted in a RMSE

of 347-435 kg ha<sup>-1</sup> across the four states. The largest  $R^2$  was found in NSW, followed by QLD, VIC and SA. Integrating FSR and ENSO climate indices showed a great potential to improve the RF model performance. The results showed that the model achieved the highest estimation



**Fig. 4.** Evaluation of random forest (RF) model to forecast wheat yield variation (kg ha<sup>-1</sup>) across four states for the different input predictor variables (no\_FSR: only selected ENSO-related climate indices without forecasted spring rainfall (FSR); with\_FSR: ENSO-related climate indices including FSR). The coefficient of determination (R<sup>2</sup>) and root mean squared error (RMSE) are used to evaluate model performance. The black lines within the box indicate the medians with 100 runs while cross-hairs indicate means. Box boundaries indicate the 25th and 75th percentiles, whiskers below and above the box indicate the 10th and 90th percentiles.



Fig. 5. The relative importance of the top eight important predictor variables used in random forest (RF) model to predict austral spring rainfall anomaly for four states of Australia.

accuracy ( $R^2 = 0.66$  and RMSE = 349 kg ha<sup>-1</sup>) in VIC (Fig. 4). The value of  $R^2$  increased by 34% and RMSE reduced by 86 kg ha<sup>-1</sup> compared to using climate indices alone. By contrast, the RF model with FSR input accounted for 59% of the variation in yield with  $R^2$  increasing by 9% and RMSE reducing by 24 kg ha<sup>-1</sup> in NSW. The  $R^2$  increased to 33% when the FSR was incorporated in SA, though the impacts of ENSO phenomena on yield anomalies in SA were weak. Incorporation of FSR into the RF model moderately improved  $R^2$  in QLD, which indicated austral spring rainfall was not important in determining yield in northeast Australia. This was not surprising because QLD has a summer dominant rainfall pattern which contrasts to other states with a winter dominant rainfall.

# 3.3. Variable importance in forecasting spring rainfall and yield variation

We explored the relative importance of each predictor variable in determining austral spring rainfall (Fig. 5) and yield (Fig. 6) from the RF model. The most important variables influencing spring rainfall were August MEI and July NINO3.4 in QLD and NSW (Figs. 5a and b), August NINO3.4 and August MEI in VIC and SA (Figs. 5c and d). SOI

phase had marginal effects on rainfall anomaly, which were not included in the Figure 5. We also examined the partial dependence plots (Fig. 7) to understand the response of rainfall anomaly to the two dominant climate variables at four states. These figures help to explain how important variables interactively affect the target variable. RF model produced similar patterns of predictor interaction on austral spring rainfall. For example, spring rainfall in QLD (Fig. 7a) tended to increase by more than 20 mm as the value of August MEI approached zero and decreased more, and July NINO3.4 declined around 26.8°C. Similar patterns could be found in NSW, VIC and SA, though the range of rainfall anomaly varied over states. The values of MEI to the right of zero, and values of NINO3.4 in July/August above around 27.5°C clearly corresponded to dry conditions due to warm events.

Figure 6 shows the eight most important variables from our RF yield forecasting model with FSR integrated. The relative importance of the variables differed between states, though FSR had the highest relative importance in regions extending from southeast (NSW and VIC) to South Australia (SA). By contrast, ENSO had stronger impacts in QLD. We could also see that many of the most important variables were relevant to SOI in SA, despite this not being the case in the other three



Fig. 6. The relative importance of the top eight important predictor variables used in random forest (RF) model to predict the yield variation for four states of Australia.



Fig. 7. Partial dependence plot for the influence of two most important predictors on austral spring rainfall anomaly for four states of Australia.

states. Comparing Figure 6a and 6b, July SOI and SOI phase, FSR and NINO3.4 in August were the four most important variables within the top five important predictors in QLD and NSW.

We then used the two dominant variables to explore their interactive effects on detrended yield (Fig. 8). MEI in August and SOI in July for QLD, and FSR and July SOI for NSW and SA were used to show the partial dependence plots. July MEI (Fig. 6c) was selected as it has comparable importance with SOI phase in July for VIC to make consistent heatmaps. In QLD, yield would increase most when August MEI approached zero and July SOI was positive (Fig. 8a). The values of yield increase became larger as FSR values approached the maximum and SOI in July was above 5 in NSW and SA (Fig. 8b and d). In VIC, it was apparent that yield variation was sensitive to FSR and irrelevant to the distribution of July MEI (Fig. 8c).

# 4. Discussion

Agricultural industry in south-eastern and southern Australia relies on austral spring rainfall. Our spring rainfall forecast based on a random forest model with three types of large-scale ENSO climate precursors accounted for 43-59% of spring rainfall variation. This is a comparable performance to those studies using similar climate indices. For example, Mekanik et al. (2016) developed ANFIS model with different inputs of climate signals to forecast spring rainfall at nine sites in VIC and their best model yielded R<sup>2</sup> values of 8-44%. These authors also found that machine learning based rainfall forecasting model was more promising as the approach could produce comparable accuracy using fewer input variables and less computation time than dynamic climatology forecasts (Abbot and Marohasy, 2014).

The RF model accounted for 33-66% of the variation in yield for testing data across four states by integrating forecasted spring rainfall and ENSO climate indices early at the growing stage. This was an



Fig. 8. Partial dependence plot for the influence of two most important predictors on wheat yield variation for four states of Australia.

improvement from 22-50% of yield variation explained by ENSO-based indices alone. This is expected as wheat grain yield is typically sensitive to rainfall variability in austral spring, especially for winter dominant rainfall states (NSW, VIC and SA). Therefore, providing better spring rainfall forecast appears to be most important step in our workflow. However, we only considered ENSO-related climate indices and neglected complex relationships between other climate drivers and Australian seasonal rainfall (Cai et al., 2011). Thus, additional largescale drivers from other climatic modes, such as the Southern Annular Mode and IOD, may be helpful in improving rainfall predictability in SA and VIC (Nguyen-Huy et al., 2017; Risbey et al., 2009).

It should be noted that large-scale climate index is only one of many variables affecting crop yield variability. Wheat yield is also affected by other factors such as pest, disease, soil and management decisions and their impacts may have a greater impact than climate variability in some cases (Wang et al., 2019). Overall, the performance of our yield-forecasting model relying solely on large-scale climate precursors in long-term historical observations is comparable to those simulation

studies using process-based biophysical models and statistical models forced with growing season climate variables at different spatial levels (Cai et al., 2019; Wang et al., 2015; Zhang et al., 2008). For example, Yuan and Yamagata (2015) developed a multiple linear regression model and found that ENSO and IOD climate modes could explain 44% inter-annual national yield variation of Australia between 1965-1995. We believe that our predictions of yield variation could be further improved with the use of other climate drivers including IOD, Southern Annular Mode (SAM), thermocline, and Madden-Julian oscillation (MJO), to forecast wheat growing season (April-November) rainfall (Kirono et al., 2010). Studies have found that the key driver of major droughts over past few decades (1895-1902, 1937-1945 and 1995-2008) in southeast Australia was IOD variability (Ummenhofer et al., 2009) and they also reported that IOD plays a dominant role in recent yield fluctuations in Australia (Nguyen-Huy et al., 2018; Yuan and Yamagata, 2015). Therefore, different climate modes may have dominant effects at a regional level in a given time series due to complex atmospheric circulation patterns (Anderson et al., 2019; Cane et al.,



(a) Spring rainfall



Fig. A1. Spring rainfall anomaly (a) in 1900-2018 and detrended yield (b) in 1900-2016 for four states in southern and eastern Australia.

1994; Risbey et al., 2009). It is also important to note that observed historical yields used for evaluation are also inherently uncertain. The yield statistics cannot be used directly as non-climate factors have a large contribution to long-term trends. Year-to-year yield variation that was detrended by different methods may result in variance of the correlations between yield and climate drivers. Heino et al. (2018) developed a framework to quantify the relationships between large-scale climate signals and simulated crop productivity at the sub-country scale. They believe a simulation model is able to isolate the impacts of climate variability on yield without detrending observed data.

We found the dominant climate signal in determining spring rainfall is August MEI and July NINO3.4 in QLD and NSW, August MEI and NINO3.4 in VIC and SA (Fig. 5). Our study focused on state-level spring rainfall anomalies, and the most important predictors may be different

from other studies derived at a site level. For example, Hossain et al. (2019) and Mekanik et al. (2016) reported that IOD and NINO3.4 signals are important in determine spring rainfall in selected sites in VIC and Western Australia. Our study firstly used the dependence of spring rainfall on two important ENSO indicators to present how they interact with rainfall. The suitable range of each variable leading to increased rainfall can be gleaned from the partial dependence plot (Fig. 7), which provides useful information for rainfall prediction at a state scale, based on precedent climate indices. Additionally, assessing the importance of predictor variables was able to provide a reference value to reduce the impacts of climate variability on yield loss. The RF model with FSR showed forecasted spring rainfall was the most influential predictor in determining yield variation of NSW, VIC and SA as FSR did reflect some information of winter ENSO



**Fig. A2.** Pearson correlation coefficients for the relations between each pair of state spring rainfall and detrended yield. The lower triangle is for spring rainfall and upper triangle for detrended yield. \*p < 0.01, \*\*p < 0.001.

predictors that were not included in the RF model without FSR. Spring rainfall was not an important variable in influencing wheat yield in QLD where there was a summer dominant rainfall pattern, although including FSR did improve the predictive ability of yield change estimation in QLD. In southern Queensland, wheat is predominantly grown on moisture stored in the soil profile from the preceding summer in fine-textured soils with high water-holding capacity. July SOI was another important predictor which influenced yield in QLD, NSW and SA. Positive SOI values in July (La Niña events) were more favourable for wheat growth. This corresponded well with the findings of Rimmington and Nicholls (1993) using the data of 1948-1988.

Expanding these analyses to farm-scale may provide reliable seasonal climate forecast. Advances in seasonal climate forecasts will offer considerable opportunities to reduce climate risk (Tao et al., 2004). At a more fundamental level, the reliable seasonal rainfall forecasting model would allow the growers to identify appropriate adaptive management under changing weather patterns e.g. making tactical fertiliser applications (top-dressing) during early August (Anwar et al., 2008; Hammer et al., 1996; Mann et al., 2019). However, yield forecasting at a farm level is more complex as it involves climatic, edaphic, and biological processes, as well as complexities of local management (Anwar et al., 2008). Crop simulation models are able to account for biophysical relationship between crop, environment, and some management options. Thus, Feng et al. (2020) developed a hybrid model by integrating crop phenology, biomass, meteorology, and remote sensing data to estimate site-level wheat yields and found satisfactory forecasts occurred up to two months prior to harvest. Additional investigation is required as to whether including large-scale climate signals in this hybrid model could improve model accuracy.

Based on R<sup>2</sup> and RMSE, our model showed comparable results compared to historical observations and is a valuable contribution to yield forecasting. Despite its value, we acknowledge that the statistical approaches such as the RF model presented here have some limitations in capturing extreme values (Feng et al., 2019b; Wang et al., 2018c). Exploring additional advanced machine learning methods (e.g. deep learning) and adopting other climate indices such as developed MEI, the bi-monthly Multivariate El Niño/Southern Oscillation (ENSO) index MEI.v2 to capture ENSO's seasonality may reduce effects of higher frequency intra-seasonal variability and lead to more robust model performance.

# 5. Conclusions

Our study developed machine learning techniques to predict spring rainfall anomaly and wheat yield variations at a state level in Australia by using precedent ENSO climate drivers. The overall performance of our model is acceptable based on two statistical indices used. By using RF model, we also quantified the interactive effects of two dominant predictor variables on yield variation, which provided valuable reference information on yield forecasting. Our model showed ENSO phenomena had stronger impacts on austral spring rainfall in QLD and NSW, but weak impacts on VIC and SA, which is supported by previous studies. We also used the RF model to predict detrended yield with ENSO climate indices at up to three months prior to harvest. Our RF model showed that the contribution of ENSO climate indices to yield variation diminished from northeast (QLD) toward southeast (NSW and VIC) and south (SA) Australia. By integrating FSR in RF model we significantly improved model performance in predicting yield, especially in VIC and NSW. In contrast, FSR had moderate impacts on yield prediction in QLD due to its characteristics of summer dominant rainfall.

As Australia accounts for a large portion of global wheat trade, more accurate forecasting of Australian wheat yields will be crucial not only for farmers but also export market security. We believe this forecasting information will have significant implications for domestic and global food security by adopting strategic agronomic management decisions early to optimize the wheat yield at a national level in response to interdecadal climate variability.

Although this modelling approach is undertaken at a state level, it has potential to be applied at a regional scale. Incorporating large-scale climate drivers with more localised environmental predictors (e.g. soil moisture, vegetation reflection index and extreme climate events) using machine learning techniques may allow farm-scale decision support to optimize the yield and maximize farm profit. Additionally, the proposed method may be implemented in other similar rain-fed regions because the input data are easily accessible. Therefore, this modelling approach appears to have broad application in understanding the contribution of climate oscillation on crop yield globally.

# Acknowledgement

Puyu Feng acknowledges the China Scholarship Council (CSC) for

the financial support for his PhD study. Bernie Dominiak reviewed and improved an early version of the manuscript. The authors also thank the

# Appendix

# Fig. A1, Fig.A2

#### References

- Abbot, J, Marohasy, J, 2014. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmospheric Research 138, 166–178.
- ABS (2019). Australian Bureau of Statistics, Australian Government. Available at: https:// www.abs.gov.au. Accessed 6 Jun. 2019.
- Adams, RM, Houston, LL, McCarl, BA, Tiscareño L, M, Matus G, J, Weiher, RF, 2003. The benefits to Mexican agriculture of an El Niño-southern oscillation (ENSO) early warning system. Agricultural and Forest Meteorology 115, 183–194.
- Ahmad, MW, Reynolds, J, Rezgui, Y, 2018. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production 203, 810–821.
- Anderson, W, Seager, R, Baethgen, W, Cane, M, 2017. Life cycles of agriculturally relevant ENSO teleconnections in North and South America. International Journal of Climatology 37, 3297–3318.
- Anderson WB, Seager R, Baethgen W, Cane M, You L (2019) Synchronous crop failures and climate-forced production variability. Science Advances 5:eaaw1976.
- Anwar, MR, Rodriguez, D, Liu, DL, Power, S, O'Leary, GJ, 2008. Quality and potential utility of ENSO-based forecasts of spring rainfall and wheat yield in south-eastern Australia. Australian Journal of Agricultural Research 59, 112–126.
- Ashok, K, Guan, Z, Yamagata, T, 2003. Influence of the Indian Ocean Dipole on the Australian winter rainfall. Geophysical Research Letters 30.
- BOM (2019). Bureau of Meteorology, Australian Government. Available at: http://www. bom.gov.au/climate. Accessed 15 Jun. 2019.
- Breiman, L, 2001. Random Forest. Machine Learning 45, 5–32.
- Cai, W, van Rensch, P, Cowan, T, Hendon, HH, 2011. Teleconnection Pathways of ENSO and the IOD and the Mechanisms for Impacts on Australian Rainfall. Journal of Climate 24, 3910–3923.
- Cai, Y, Guan, K, Lobell, D, Potgieter, AB, Wang, S, Peng, J, Xu, T, Asseng, S, Zhang, Y, You, L, Peng, B, 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology 274, 144–159.
- Cane, MA, Eshel, G, Buckland, RW, 1994. Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature. Nature 370, 204–205.
- CDS (2019). Climate Data Sets, NCAR-Climate Data Guide. Available at: https:// climatedataguide.ucar.edu/climate-data. Accessed 6 Jun. 2019.
- Chiew, FHS, Piechota, TC, Dracup, JA, McMahon, TA, 1998. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. Journal of Hydrology 204, 138–149.
- Cleverly, J, Eamus, D, Luo, Q, Restrepo Coupe, N, Kljun, N, Ma, X, Ewenz, C, Li, L, Yu, Q, Huete, A, 2016. The importance of interacting climate modes on Australia's contribution to global carbon cycle extremes. Scientific Reports 6, 23113.
- Cutler, DR, Edwards, TC, Beard, KH, Cutler, A, Hess, KT, 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
- Feng, P, Wang, B, Li Liu, D, Xing, H, Ji, F, Macadam, I, Ruan, H, Yu, Q, 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic Change 147, 555–569.
- Feng, P, Wang, B, Liu, DL, Waters, C, Xiao, D, Shi, L, Yu, Q, 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agricultural and Forest Meteorology 285-286, 107922.
- Feng, P, Wang, B, Liu, DL, Waters, C, Yu, Q, 2019a. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agricultural and Forest Meteorology 275, 100–113.
- Feng, P, Wang, B, Liu, DL, Yu, Q, 2019b. Machine learning-based integration of remotelysensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems 173, 303–316.
- Grimm, R, Behrens, T, Märker, M, Elsenbeer, H, 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102–113.
- Guo, P-T, Li, M-F, Luo, W, Tang, Q-F, Liu, Z-W, Lin, Z-M, 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. Geoderma 49–59 237–238.
- Hammer, GL, Holzworth, DP, Stone, R, 1996. The value of skill in seasonal climate forecasting to wheat crop management in a region with high climatic variability. Australian Journal of Agricultural Research 47, 717–737.
  Heino, M, Puma, MJ, Ward, PJ, Gerten, D, Heck, V, Siebert, S, Kummu, M, 2018. Two-
- Heino, M, Puma, MJ, Ward, PJ, Gerten, D, Heck, V, Siebert, S, Kummu, M, 2018. Twothirds of global cropland area impacted by climate oscillations. Nature Communications 9, 1257.
- Heung, B, Bulmer, CE, Schmidt, MG, 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. Geoderma 214, 141–154.
- Hossain, I, Rasel, HM, Imteaz, MA, Mekanik, F, 2018. Long-term seasonal rainfall forecasting: efficiency of linear modelling technique. Environmental Earth Sciences 77, 280.

two anonymous reviewers for their constructive comments to improve the earlier version of this manuscript.

- Hossain, I, Rasel, HM, Imteaz, MA, Mekanik, F, 2019. Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics.
- Hudson, D, Alves, O, Hendon, HH, Lim, EP, Liu, GQ, Luo, JJ, MacLachlan, C, Marshall, AG, Shi, L, Wang, GM, Wedd, R, Young, G, Zhao, M, Zhou, XB, 2017. ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. J. South Hemisph. Earth Syst. Sci. 67, 132–159.
- Immitzer, M, Atzberger, C, Koukal, T, 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. Remote Sensing 4, 2661–2693.
- Kirono, DGC, Chiew, FHS, Kent, DM, 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. Hydrological Processes 24, 1237–1247.
- Lee, E-J, Kim, M, Kim, K-H, 2018. Connecting ENSO-related climatic variations with a long-term crop supply data to enhance agro-meteorological capability of Tongan stakeholders. International Journal of Climatology 38, e18–e27.
- Ließ, M, Schmidt, J, Glaser, B, 2016. Improving the Spatial Prediction of Soil Organic Carbon Stocks in a Complex Tropical Mountain Landscape by Methodological Specifications in Machine Learning Approaches. PLOS ONE 11, e0153673.
- Liu, Y, Yang, X, Wang, E, Xue, C, 2014. Climate and crop yields impacted by ENSO episodes on the North China Plain: 1956–2006. Regional Environmental Change 14, 49–59.
- Long Paddock (2019). The Long Paddock, Queensland Government. Available at: https:// www.longpaddock.qld.gov.au. Accessed 17 Jun. 2019.
- Lu, J, Carbone, GJ, Gao, P, 2017. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. Agricultural and Forest Meteorology 237–238, 196–208.
- Mann, ML, Warner, JM, Malik, AS, 2019. Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia. Climatic Change 154, 211–227.
- Meinke, H, Stone, RC, Hammer, GL, 1996. SOI phases and climatic risk to peanut production: a case study for northern Australia. International Journal of Climatology 16, 783–789.
- Mekanik, F, Imteaz, MA, Talei, A, 2016. Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals. Climate Dynamics 46, 3097–3111.
- 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.
- 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.
- Nicholls, N, 1986. Use of the Southern Oscillation to predict Australian sorghum yield. Agricultural and Forest Meteorology 38, 9–15.
- Phillips, J, Rajagopalan, B, Cane, M, Rosenzweig, C, 1999. The role of ENSO in determining climate and maize yield variability in the U.S. combelt. International Journal of Climatology 19, 877–888.
- Phillips, JG, Cane, MA, Rosenzweig, C, 1998. ENSO, seasonal rainfall patterns and simulated maize yield variability in Zimbabwe. Agricultural and Forest Meteorology 90, 39–50.
- Podestá, G, Letson, D, Messina, C, Royce, F, Ferreyra, RA, Jones, J, Hansen, J, Llovet, I, Mn, Grondona, O'Brien, JJ, 2002. Use of ENSO-related climate information in agricultural decision making in Argentina: a pilot experience. Agricultural Systems 74, 371–392.
- Podestá, GP, Messina, CD, Grondona, MO, Magrin, GO, 1999. Associations between Grain Crop Yields in Central-Eastern Argentina and El Niño–Southern Oscillation. Journal of Applied Meteorology 38, 1488–1498.
- Potgieter, AB, Hammer, GL, Butler, D, 2002. Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. Australian Journal of Agricultural Research 53, 77–89.
- Potgieter, AB, Hammer, GL, Meinke, H, Stone, RC, Goddard, L, 2005. Three Putative Types of El Niño Revealed by Spatial Variability in Impact on Australian Wheat Yield. Journal of Climate 18, 1566–1574.
- Reilly, J, Tubiello, F, McCarl, B, Abler, D, Darwin, R, Fuglie, K, Hollinger, S, Izaurralde, C, Jagtap, S, Jones, J, Mearns, L, Ojima, D, Paul, E, Paustian, K, Riha, S, Rosenberg, N, Rosenzweig, C, 2003. U.S. Agriculture and Climate Change: New Results. Climatic Change 57, 43–67.
- Rimmington, GM, Nicholls, N, 1993. Forecasting wheat yields in Australia with the Southern Oscillation Index. Australian Journal of Agricultural Research 44, 625–632.
- 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.
- Shuai, J, Zhang, Z, Tao, F, Shi, P, 2016. How ENSO affects maize yields in China: understanding the impact mechanisms using a process-based crop model. International Journal of Climatology 36, 424–438.

Stone, RC, Hammer, GL, Marcussen, T, 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation Index. Nature 384, 252–255.

- Tao, F, Masayuki, Y, Zhao, Z, Yousay, H, Hartmut, G, Congbin, F, 2004. Variability in climatology and agricultural production in China in association with the East Asian summer monsoon and Niño southern oscillation. Climate Research 28, 23–30.
- 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 36.
- Wang, B, Chen, C, Liu, DL, Asseng, S, Yu, Q, Yang, X, 2015. Effects of climate trends and variability on wheat yield variability in eastern Australia. Climate Research 64, 173–186.
- Wang, B, Deveson, ED, Waters, C, Spessa, A, Lawton, D, Feng, P, Liu, DL, 2019. Future climate change likely to reduce the Australian plague locust (Chortoicetes terminifera) seasonal outbreaks. Science of The Total Environment 668, 947–957.
- Wang, B, Waters, C, Orgill, S, Cowie, A, Clark, A, Li Liu, D, Simpson, M, McGowen, I, Sides, T, 2018a. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecological Indicators 88, 425–438.
- Wang, B, Waters, C, Orgill, S, Gray, J, Cowie, A, Clark, A, Liu, DL, 2018b. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Science of The Total Environment 630, 367–378.
- Wang, B, Zheng, L, Liu, DL, Ji, F, Clark, A, Yu, Q, 2018c. Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. International Journal of Climatology 38, 4891–4902.

- Were, K, Bui, DT, ØB, Dick, Singh, BR, 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators 52, 394–403.
- Wolter, K, Timlin, MS, 1998. Measuring the strength of ENSO events: How does 1997/98 rank? Weather 53, 315–324.
- Yang, R-M, Zhang, G-L, Liu, F, Lu, Y-Y, Yang, F, Yang, F, Yang, M, Zhao, Y-G, Li, D-C, 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecological Indicators 60, 870–878.
- Yu, B, Qian, B, Zhang, Y, Anderson, K, 2018. ENSO and Sea Surface Temperature Anomalies in Association with Canadian Wheat Yield Variability. Atmosphere-Ocean 56, 28–39.
- Yu, Q, Hengsdijk, H, Liu, JD, 2001. Application of a progressive-difference method to identify climatic factors causing variation in the rice yield in the Yangtze Delta. China. International Journal of Biometeorology 45, 53–58.
- Yu, Q, Li, L, Luo, Q, Eamus, D, Xu, S, Chen, C, Wang, E, Liu, J, Nielsen, DC, 2014. Year patterns of climate impact on wheat yields. International Journal of Climatology 34, 518–528.
- Yuan, C, Yamagata, T, 2015. Impacts of IOD, ENSO and ENSO Modoki on the Australian Winter Wheat Yields in Recent Decades. Scientific Reports 5, 17252.
- Zhang, T, Zhu, J, Yang, X, Zhang, X, 2008. Correlation changes between rice yields in North and Northwest China and ENSO from 1960 to 2004. Agricultural and Forest Meteorology 148, 1021–1033.