Modelling impacts of climate and weather extremes on wheat cropping systems across New South Wales

Puyu Feng

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Life Sciences, Faculty of Science

University of Technology Sydney

Australia

May 2020

Certificate of Original Authorship

I, Puyu Feng declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Life Sciences/Faculty of Sciences at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Signature of student: Production Note: Signature removed prior to publication.

Date: 21 May 2020

Acknowledgement

Undertaking this PhD is a challenging but life-changing experience for me. It would not have been possible to reach the ending point without the support and guidance that I received from many people.

First and foremost, I would like to express my sincere gratitude to my principal supervisor Professor Qiang Yu. It was him who encouraged me to pursue PhD at University of Technology Sydney and helped me apply for scholarships. It was also him who suggested me to conduct cooperative research at Wagga Wagga agricultural institute, New South Wales Department of Primary Industries (NSW DPI). His professional guidance as well as general support and encouragement helped me in all the time of research.

I would like to express my special appreciation and thank to my co-supervisor, principal research scientist Dr. De Li Liu, who provided me the chance to study at NSW DPI. During the four years in NSW DPI, Dr. De Li Liu has always been holding me to a high research standard by providing insightful comments on my work and constructive criticisms at different stages of my PhD study. I could not have imagined having a better advisor and mentor for my PhD study. His enthusiasm and rigorous manner for scientific research affects deeply on me and will have profound impact on my future career.

I am deeply grateful to Dr. Bin Wang at NSW DPI. He has been acting as a trustworthy brother, a responsible colleague, and an awe-inspiring teacher during my 4-year PhD study. Whenever I meet with difficulties, no matter from work or life, he is always there, ready to help me. He leads me into an amazing world of science and without his step-by-step guidance, insightful criticisms, constructive suggestions as well as constant encouragements, I could not make it to this point. His guidance plays a key role in the completion of this thesis.

I thank my fellows from UTS, Lijie Shi, Mingxi Zhang, Hong Zhang, Siyi Li, Dr. Jie He, Dr. Rong Gan, Dr. Qinggaozi Zhu, and Dr. Jianxiu Shen and fellows or visiting scholars from NSW DPI, Dr. Hongtao Xing, Dr. Xiaojun Shen, Dr. Hongyan Ruan, Dr. Yue Chen, Dr. Weiwei Xiao, Dr. Dengpan Xiao, for the kindness in your hearts, for the encouragement and support, , and for all the memorable happy and difficult time we share. I am thankful to my co-authors, Dr. Cathy Waters, Dr. Fei Ji, Dr. Ian Macadam, for their tremendous efforts in improving the early manuscripts that have been published. I am also grateful to staff in UTS and NSW DPI for their administrative assistance in various ways, which ensured a pleasant environment and convenient workflow during my PhD study.

I would especially like to thank my parents, my father Guangzhi Feng and my mother Guiling Wang, who supported my PhD study and always encourage and help me out of difficulties. I am

the only child to my parents. It has been hard to them since I decided to study overseas four years ago, but they still supported me without a word of complaints.

Most importantly, I would like to thank my beloved partner Ting Li. I feel so sorry that I left China without a word four years ago and keep you waiting alone. If it weren't for me, you would have a man to accompany you and live a happier life. You chose to keep our long-distance relationship with a virtual boyfriend by video call. You have always been standing by me through all my travails, giving me unconditional support, sharing my smiles and tears during this journey. Now, we make it and we survive the long-distance. I owe you everything and will keep you in my love throughout my rest life.

I wish to gratefully acknowledge that the Chinese Scholarship Council provided the scholarship and the NSW DPI provided facilities for conducting this work.

Publications arising from this thesis

Journal papers directly included in this thesis

Feng, P., Wang, B., Liu, D.L., Xing, H., Ji, F., Macadam, I., Ruan, H. and Yu, Q., 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic change, 147(3-4), pp.555-569. (Chapter 2)

Feng, P., Wang, B., Liu, D.L. and Yu, Q., 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, pp.303-316. (Chapter 3)

Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., and Yu, Q., 2019. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agricultural and Forest Meteorology, 285: 107922. (Chapter 4)

Feng, P., Liu, D.L., Wang, B., Waters, C., Zhang, M. and Yu, Q., 2019. Projected changes in drought across the wheat belt of southeastern Australia using a downscaled climate ensemble. International Journal of Climatology, 39(2), pp.1041-1053. (Chapter 5)

Feng, P., Wang, B., Liu, D.L., Waters, C. and Yu, Q., 2019. 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, pp.100-113. (Chapter 6)

Peer-reviewed International Conference proceedings

Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., and Yu, Q., 2019. Developing a machine learning-based model to forecast growing season rainfall for Australian wheat belt. The 23rd International Congress on Modelling and Simulation (MODSIM), Canberra, Australia, 3-8 December 2019. (Extended Abstract, Accepted).

Contents

List of Figures

Chapter 4

List of Tables

Glossary

APSIM	*Agricultural Production System sIMulator*
ARID	*Agricultural Reference Index for Drought*
BOM	*Bureau of Meteorology*
BRF	*bias-corrected random forest*
CDD	*consecutive dry days*
CDF	*cumulative distribution function*
CMIP5	*Coupled Model Intercomparison Project phase 5*
CV	*coefficient of variation*
ECEs	*extreme climate events*
ENSO	*El Niño Southern Oscillation*
GCM	*global climate models*
GEE	*Google Earth Engine*
GHG	*greenhouse gas*
IDW	*Inverse Distance Weighted*
IPCC	*Intergovernmental Panel on Climate Change*
LCCC	*Lin's concordance correlation coefficient*
MAE	*mean absolute prediction error*
MAPE	*Mean Absolute Percentage Error*
MLP	*multi-layer perceptron neural network*
MLR	*multiple linear regression*
MODIS	*Moderate Resolution Imaging Spectroradiometer*
mtry	*the number of randomly selected predictor variables at each node*
NDVI	*Normalized Difference Drought Index*
NSW	*New South Wales*
ntree	*the number of trees to grow in the forest*
PET	*potential evapotranspiration*
R^2	*coefficient of determination*
RCP	*Representative Concentration Pathway*
RF	*random forest*
RMSE	*root mean square error*
ROC	*receiver operating characteristic*
SA2	*Statistical Areas Level 2*
SILO	*Scientific Information for Land Owners*
SPEI	*Standardized Precipitation Evapotranspiration Index*
SPI	*Standardized precipitation index*
SVM	*support vector machine*
Tmax	*maximum land surface temperature*
Tmin	*minimum land surface temperature*
TRMM	*Tropical Rainfall Measuring Mission*
VIF	*variance inflation factor*

Abstract

Australian wheat production is crucial to global food security, as Australia is one of the world's major grain exporters. The NSW wheat belt is a main wheat production area in south-eastern Australia. Interannual wheat yields in the NSW wheat belt are highly variable, as the rainfed wheat cropping systems are significantly affected by recurrent climate and weather extremes. Ongoing climate change is projected to induce more extremes events, thereby leading to more unfavourable climate conditions for wheat production.

This thesis aims to quantify the impacts of various climate and weather extremes on wheat yield in the present and explore their potential impacts in the future, thereby enhancing the capability of stakeholders to reduce yield losses. Five inter-related studies based on statistical regression-based models, process-based crop models, or the integration of both models were conducted in the NSW wheatbelt. Consistent findings demonstrate that: (1) Inter-annual variability of rainfall in winter and spring was largely responsible for wheat yield variation. (2) Seasonal agricultural drought conditions could be well monitored for the wheat belt using remote sensing information and machine learning-based statistical models. (3) APSIM simulated biomass, multiple climate extremes indices, NDVI, and SPEI were incorporated into the RF model to develop a hybrid model for improved modelling of impacts of climate extremes. Drought events throughout the growing season were identified as the main factor causing yield losses. (4) The wheat belt was expected to experience drier conditions in spring and winter but had little change in summer and autumn. By the end of the 21st century, over half of the wheat belt was at a high risk of experiencing spring and winter drought. (5) The hybrid model was used to assess the impacts of future climate and weather extremes on wheat yield. Increasing drought and heat events around reproductive stages were identified to be major threats causing yield losses in the future.

This project enhanced systematic understanding of impacts of present and future climate and weather extremes on wheat yield and their likely changes in the future. However, certain aspects such as new crop cultivars, efficient management practices, pests and weed, were not explicitly considered in the modelling methods. Therefore, these findings should be further reconfirmed by models involving more influential information to guide agricultural production.

Key words: climate and weather extremes; climate change; wheat yield; machine learning; process-based crop models; Australia

Chapter 1. Introduction

1.1 Research background

Climate and weather extremes (e.g. heat and drought) are drawing attention among researchers and decision makers because of their adverse impacts on human activities, especially crop production. Crop growth and development can be severely affected and result in great yield losses. Moreover, under the background of global warming, climate and weather extremes are projected to increase around the world and lead to more unfavourable conditions for crop production. Therefore, assessing the impacts of various climate and weather extremes on crop yield and identifying their likely change and potential impacts in the future are important to develop effective agronomic practices to reduce yield losses. This section will provide a comprehensive review of literatures on climate change and variability, climate and weather extremes, impacts of climate extremes on crops, and methods to explore climate-yield relationships. It is intended to provide useful background information for this Ph.D. research.

1.1.1 Climate change

It is certain that climate is changing all over the world according to large-scale and long-term observations. The following are explicit conclusions from the IPCC (Intergovernmental Panel on Climate Change) report (IPCC, 2014).

• Global mean surface temperature has been surely increasing during last two centuries, with increasing rate being faster and faster.

• Precipitation change is inconsistent around the world. In mid-latitude regions, averaged precipitation has been increasing in recent decades.

• All factors of the climate system will continue changing if emissions of greenhouse gases are not well controlled.

For Australia, according to "State of the Climate 2016" (BOM and CSIRO, 2016), the climate has warmed in mean surface air temperature by around 1 °C since 1910, especially since 1950. Daily maximum and minimum temperature have increased by 0.8 °C and 1.1 °C, respectively. Annual total rainfall has slightly increased across Australia from 1900. However, seasonal distribution of rainfall is increasingly uneven. In the south of Australia, autumn and winter rainfall has reduced more than 10 percent since 1970. These trends will keep up and even aggravate in the future based on current research results.

1.1.2 Climate variability

Climate factors, such as precipitation and temperature, are fluctuant around mean values in long-term observations (Folland et al., 2002). This phenomenon is defined as climate variability. For example, consecutive winters, surely contain some wetter and some drier climate conditions than the long-term mean in a given region. This variability is mainly the result of natural and large-scale features of the climate (Latif and Barnett, 1994).

Australia has a seriously variable climate. A main cause of inter-annual variability in Australia is said to be the El Niño Southern Oscillation (ENSO) (Murphy and Timbal, 2008). ENSO is an irregularly periodic variation in winds and sea surface temperatures over the tropical eastern Pacific Ocean, affecting much of the tropics and subtropics (L'Heureux, 2014). ENSO affects intensities and positions of main climate features, resulting in variability in precipitation, temperature, winds, cyclone activity, ocean currents and sea level. In addition, ENSO is characterised with alternations of El Niño and La Niña patterns. Two patterns alternate over periods of three to eight years. El Niño and La Niña can lead to drier and wetter conditions, respectively, in most of Australia (Watterson, 2009).

Climate variability is a common phenomenon in all climate systems in the world. For a given climate, if long-term means of climate variables do not alter significantly, the climate is considered to be stable. However, climate change, which is increasing temperature and altering rainfall amount, is breaking this stable status. Statistically-significant trends of climate variables means are superimposed on natural climate variability, thereby enhancing natural variability and leading to more violent fluctuations (Easterling et al., 2000).

1.1.3 Extreme climate and weather events

Climate and weather extremes, such as heat and drought (Dai, 2013; Teixeira et al., 2013), have always been hotspots among researchers and decision makers, because of their adverse impacts on agriculture and economics. According to the IPCC definition (IPCC, 2012), extreme climate and weather events include heatwave (Perkins et al., 2015; Seneviratne et al., 2006), cold extreme (Peterson et al., 2013), drought and flood (Whetton et al., 1993) and so on.

In recent years, economic losses caused by climate extremes are increasing (Coumou and Rahmstorf, 2012). As some extreme events, such as heatwave and drought, are closely related to temperature, a suspicion has proposed on whether the increased losses are caused by climate change. Planton et al. (2008) emphasized that climate change will surely result in increases in climate extremes because the shift of the distribution of a climatic factor towards one side would indeed enhance occurrences of extremes on the same side. Taking temperature as an example, Figure 1-1 shows three patterns of changes (IPCC, 2012). Each change pattern will no doubt cause

an increase of extreme hot or chilly days. Decreases or increases in mean precipitation should have same impacts on flood or drought episodes.

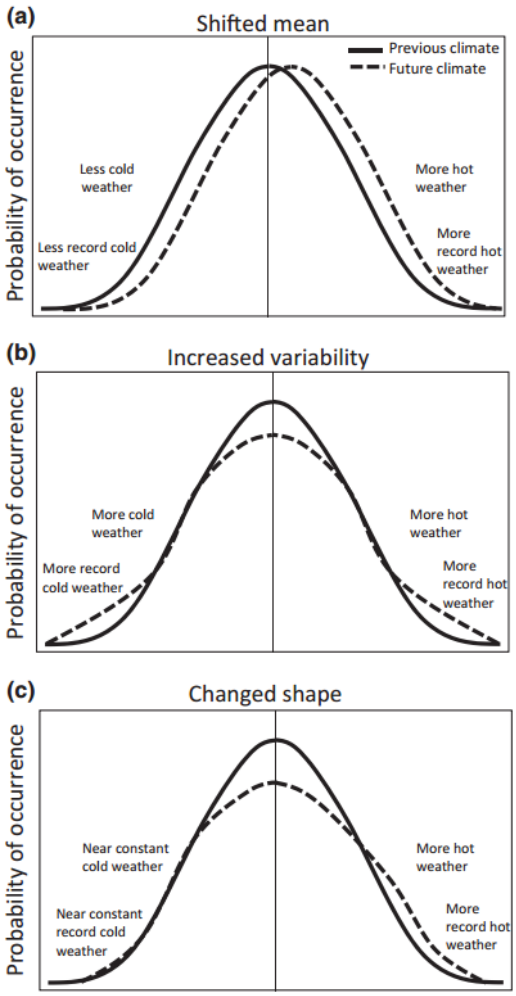


Figure 1-1 Impacts of shifts in temperature distribution on temperature extremes (IPCC, 2012). Three possible change patterns are plotted.

Heat events are likely to increase in frequency and intensity due to the increase in mean temperature (IPCC, 2012) or temperature variability (Schär et al., 2004). On the other hand, cold extremes are correspondingly decreasing. The Coupled Model Intercomparison Project Phase 5 (CMIP5) climate models ensemble indicates that in all regions of the world, there are likely to be more heatwaves and less cold extremes (Kharin et al., 2013). By the end of 21st century, increases in daily maximum land surface temperature (T_{max}) are 1.2 °C, 2.4 °C and 5.4 °C, respectively under three RCPs (Representative Concentration Pathways 2.6, 4.5 and 8.5), compared to a baseline of 1986-2005. Corresponding increases in daily minimum land surface temperature (T_{min}) are 1.7 °C, 3.2 °C, and 6.2 °C, which are all higher than T_{max} changes. Heatwaves are also associated with soil-moisture-temperature and soil-moisture-precipitation feedbacks (Perkins et al., 2015). Zhao et al. (2015) considered 3 heatwave indices (based on both temperature and

moisture) in CMIP5 global climate model ensemble. Their results showed that temperature-humidity related heatwave indices detect a greater increase over the world than that expected due to warming alone. In general, the intensity of heatwave is projected to reach unprecedented levels under RCP8.5 scenario by the end of 21st century, especially over tropical and subtropical areas.

Precipitation increase is more common around the world for CMIP5 model ensemble projections. Wang and Chen (2014) developed a high-resolution multi-model ensemble projection based on 35 GCMs under RCP4.5 and RCP8.5. Results showed that across most areas of China, a significant increase (2%-20%) in rainfall is projected by the end of 21st century compared to a baseline of 1961-1990. Similar results are also discovered in North American (Maloney et al., 2014). It should be noted that increases in precipitation means usually accompany more uneven seasonal distribution. An increasing trend of year-to-year variability of seasonal mean rainfall has been found in most of the world using the UK Hadley Centre for Climate Prediction and Research (HadCM2) simulations (Giorgi and Francisco, 2000). In another study using the Canadian Centre for Climate Modelling and Analysis (CGCM1) simulations, mean rainfall and inter-annual variability also have positive correlations (Boer et al., 2000).

It is inevitable that intensity and frequency of extremely heavy rainfall events will increase on the condition of increasing precipitation means in the future. Region-scale studies have demonstrated this result in Europe, North America, East Asia, the Sahel, southern Africa, Australia and the South Pacific (Beniston et al., 2007; Goswami et al., 2006; Panthou et al., 2014; Ummenhofer et al., 2015; Villarini et al., 2013) as well as in global studies by Kharin and Zwiers (2000) and Zwiers and Kharin (1998). For example, Hennessy et al. (1997) presented that heavy rainfall events of one-year return period in Europe, Australia, India and the USA are likely to increase in intensity by 10 to 25% according to the UKHI and CSIRO9 models.

Despite projections of precipitation increases, occurrences of dry spells or droughts also shows an increasing tendency under enhanced greenhouse gas (GHG) simulations. This is mainly due to two reasons: uneven distribution of precipitation and increased temperature. Rainfall events are likely to decrease, but heavy rainfall events are projected to increase, thus consecutive dry days will increase (Li et al., 2012). Moreover, heavy rainfall amount often surpasses bearing capacities of soils and excessive water forms runoff which cannot be used by rainfall areas (Yasufuku et al., 2015). On the other hand, increased temperature will inevitably lead to increased evaporation, which can aggravate drought occurrence (Venkataraman et al., 2016). Ahmadalipour et al. (2017) assessed impacts of climate change on drought features across the United States using 21 downscaled GCMs. They revealed a significant increase in intensity and frequency of future summer drought in most of the United States. This increase can be largely attributed to increased potential evapotranspiration caused by increased temperature. Significant increases of seasonal drought are also discovered by Dai (2013) across Australia. Increased temperature and a reduction

in number of precipitation events rather than a reduction in average precipitation are also main reasons.

Australia is a continent which is frequently subjected to various climate and weather extremes. For example, according to records, severe droughts occur on an average of every 18 years (Anderson, 2014). Climate can be viewed to be steady on condition that long-term means does not alter apparently along with time variation. Nevertheless, changing climate can enhance natural variabilities leading to more violent fluctuations (Easterling et al., 2000). According to BOM and CSIRO (2016), Australia climate is going to be warmer and warmer. Heatwaves have become more frequent, hotter, and occur earlier (Perkins-Kirkpatrick and Gibson, 2017; Perkins et al., 2012a). Since 1950, annual number of hot days (>35°C) has doubled across large parts of Australia. While for droughts, they are projected to be more frequent and severe in southeast Australia through this century, consistent with expected declines in winter and spring rainfall (BOM and CSIRO, 2016).

1.1.4 Impacts of climate and weather extremes on crops

Previous studies have paid much attention on impacts of mean conditions of climate factors on crops (IPCC, 2012). However, crop growth receives more adverse impacts from climate and weather extremes. Climate and weather extremes often spread quickly, cause direct damage to crop organs, and cannot be accurately predicted in advance. Understanding impacts of various climate extremes on crop growth is necessary for long-term climate-yield relationships studies.

1.1.4.1 Heat

Heat stress has long been a big challenge to agricultural producers. In recent years, heat stress has caused many significant yield reductions. For example, a record-breaking heatwave affected the European continent in summer 2003. With mean summer temperatures exceeding the 1961-1990 mean by about 3 °C over large areas and by over 5°C regionally, it was very likely the hottest European summer over the past 500 years (Fischer et al., 2007). These summer heat events in Europe caused severe damage to crop and pasture yields and cut production by a third in those affected regions. Direct economic losses for agriculture sector were estimated at 35 billion euros (IPCC, 2012). Similar events also happened in the United States Severe yield reductions of soybean and corn were reported which were also induced by heatwaves (Schlenker and Roberts, 2009).

Generally, heat stress events can be divided into two categories: (1) above long-term mean temperatures for a long time (10-60 days), and (2) short-term (1-3 days) heat waves with extremely high temperatures. Both two events can seriously affect physiological processes of crop growth. Short term heat stress can adversely reduce enzymatic activity and genetic expression, while long-term heat events can cut down carbon sequestration and thus reduce growth velocity.

Owing to these reasons, heat waves tend to accelerate maturation process. Lobell et al. (2012) used 9 years of satellite measurements of wheat growth in northern India to monitor rates of wheat senescence following exposure to temperatures greater than 34 °C. Results show a statistically significant acceleration (up to 7 days shortening of whole wheat growth periods) of senescence from extreme heat, above and beyond the effects of increased average temperatures. Shortenings of growth periods directly lead to reductions of carbon assimilation, which will lead to insufficient supply of photo-assimilates for grains (Stone et al., 1994). In addition, heat stress can result in infertile grains and reduce grain numbers, thereby reducing yield (Liu et al., 2016).

Actual effects of heat waves on crops are also dependent on occurrence time because endurance abilities of crops against high temperatures are different at different growth stages. For example, wheat plants are most vulnerable to heat stress at reproductive stages, especially from anthesis to grain filling. Winter wheat is generally flowering at late spring and heat stress is most common during this period (Yang et al., 2017). Wollenweber et al. (2003) reported that wheat yield would decrease seriously if wheat plants experience a short-term heat stress (35 °C) during anthesis. According to 27 years of experimental data, Blumenthal et al. (1991) reported that heat waves at grain filling period could seriously reduce wheat yield.

1.1.4.2 Drought

Drought is currently one of main constraints to crop production in rainfed systems throughout the world. Generally, more than 50% of the area under wheat cultivation is affected by periodic drought (Rajaram, 2001). In major wheat growing areas of the world, particularly with a Mediterranean climate, mean pan evaporation often surpasses average precipitation especially during grain filling stage, leading to drought during reproductive and grain-filling phases (Reynolds et al., 2015).

Drought adversely affects crop growth mainly through water deficit. Crop growth can be restrained in all growing periods, but extent of damage caused by drought mainly depends on its occurrence time. For wheat, terminal droughts during reproductive stages are most harmful (Pradhan et al., 2012). Reproductive stages of wheat are key stages that determine eventual yields. Wheat plants during this stage require large amounts of water for carbon sequestration and assimilate translocation (Ji et al., 2012). Water deficit during these stages can not only reduce production of photosynthate through accelerating processes of leaf senescence and lessening photosynthesis (Chaves et al., 2002), but also decrease the distribution of assimilates to grains through reducing the rate of translocating and sink capacity (Liang et al., 2001).

1.1.4.3 Frost

As with drought and heat, frost can also restrain crop growth at all growing periods. For wheat, frost is more critical during post-head-emergence stage. Post-head-emergence frosts occur mainly

in mid-latitude regions, especially in Mediterranean climate areas, including Australia (Crimp et al., 2017), Central Europe (Zhao et al., 2013). Post-head-emergence frost is a big threat to wheat growth. Wheat stems and even whole spike can be completely destroyed during a single frost event (Frederiks et al., 2012). Generally, frost-induced wheat yield reduction usually reaches 10 percent, whereas this number can increase to 80% in frost-vulnerable areas, such as in Australia (Boer et al., 1993).

Frost can harm wheat growth and development in all growing periods from seedling to harvest. During early stages, some types of wheat seedlings are vulnerable to very cold conditions. Leaves might be withered and seedlings might even die (Fuller et al., 2007). This will eventually lead to yield reduction. The magnitude of yield reduction caused by frost from pre-flowering to ripening is much larger than at any other growing periods. During different periods of reproductive stage, frost can damage various parts of plant, which all can result in serious effects on crop growth. Around booting, stem is the most vulnerable to frost harm (Crimp et al., 2017). The most common injure to stem occurs below head which might lead to loss of head. Stem injury is irreclaimable and translocation of photoassimilates in later period will be affected greatly. When spike has formed but before flowering, a single frost event can lead to injury to spike by sterile flowers, thereby decreasing grain numbers (Barlow et al., 2015). During heading period, frost can also lead to sterile flowers through damaging anthers and embryos (Al-Issawi et al., 2013; Marcellos and Single, 1984). At grain filling stage, frost can lead to shriveled and shrunken grains through killing partially filled grains (Perry et al., 2017). It should be noted that although climate warming has significantly reduced occurrences of cold extremes, wheat is still highly vulnerable to frost. This is mainly because higher mean temperature can speed up plants' growth rate and shorten whole growth duration, resulting in the shift of reproductive stages to early dates (He et al., 2015). Thus, frost-induced yield reduction should be focused highly now as well as in the future.

1.1.4.4 Heavy rainfall and flooding

Excessive water is also a big threat to crop yield. Impacts of heavy rainfall can be divided into two categories: environmental impacts and operational impacts. Environmental impacts include physical damage and diseases. Rains with high intensity for short times will cause direct harm to crop canopies, especially during seedling or flowering stage. In particular, floods can even destroy whole plants across large areas. A short-term of soil waterlogging caused by excessive water can result in root damage. On the other hand, during grain filling stage long-term humid weather is likely to result in fungal disease. Operational influence is mainly delayed field management due to inability to operate machinery (Falloon et al., 2010). Muddy ground may break down farming machines. Based on above impacts, heavy rainfall and flooding can seriously reduce crop yield. Yield losses caused by extremely wetness have been reported across the world (Van der Velde et al., 2012).

1.1.5 Global climate models

To study the likely change of climate and weather extremes and assess their impacts on crops, long term climate data are needed to carry out this task. Historical climate data can be obtained from climate observation stations, while future climate projections are usually acquired from atmosphere-ocean global climate models (GCM). Nearly 60 GCMs have been created by various scientific research institutions around the world. These GCMs are based on internal mechanisms of various climatic phenomenon and are capable of offering a comprehensive assessment of climate dynamics. GCMs are able to reproduce historical climate and project future climates under different GHG emission scenarios (IPCC, 2013). Thus, GCMs offer great convenience for exploring climate change and its impacts (Perkins et al., 2007). In recent years, numerous studies have been done to project future climates and assess climate change impacts (Hewitt et al., 2001; Kloster et al., 2010; Prudhomme and Davies, 2009) using GCMs.

There are significant disagreements of climate projections among those GCMs. For instance, projections of regional precipitation change often vary greatly among those GCMs, because different GCMs adopt different internal climatological mechanisms to reproduce multi-scale climatic dynamics (Perkins and Pitman, 2009; Woldemeskel et al., 2014). This phenomenon is called climate projection uncertainty. Climate change impact studies should take this uncertainty into consideration. If using a single model, significant bias of impact assessment will generate due to uncertainties (IPCC, 2013). Nowadays, studies have shown that projections based on multiple models are better than any single model and can overcome uncertainties to some degree (Perkins et al., 2012b). Thus, multi-model ensembles projections of climate factors have been used for impact assessment on future crop growth (Özdoğan, 2011).

Raw GCMs often have a spatial resolution of 2° and a temporal resolution of 1-month. This limits their application in region-scales and in combination with other tools, such as crop models which need daily climate variables. In last few decades, researchers have developed several downscaling techniques to overcome limitations (Benestad, 2010; Charles et al., 2010; Mearns et al., 2003). Most commonly used techniques are change factors, dynamical downscaling and statistical downscaling. Statistical downscaling contains three sections: regression analysis, weather classification methods and stochastic weather generators. Regression analysis capture relationships between local meteorological factors and large-scale atmospheric factors. Weather classification methods or weather typing schemes gather days that are of similar climatic conditions into a limited number of discrete weather patterns. Weather generators are statistical methods which provide sequences of climatic variables through simulating key properties of observed meteorological records (i.e. daily means, frequencies, extremes, variance, and covariance) (Kilsby et al., 2007). This technique is based on 2 key assumptions that (1) predictor

variables should be characterized well by GCMs; (2) these observed, an empirical relationship is assumed to be stationary under future climate conditions. Once the 2 assumptions hold, GCM projections can be used to generate smaller-scale climate projections which are more credible and of higher resolution. In comparison with dynamical downscaling, main merits of statistical downscaling are: (1) it is cheap to compute; (2) it can generate site-scale meteorological factors; (3) it is based on recognized computational processes and statistical relationships can be explained rather easily; and (4) it can be conveniently used in other areas and provide point-scale projections (Fowler et al., 2007). Despite above advantages, some disadvantages should be taken into consideration in practice. Firstly, it usually needs long-term and reliable observed historical climate variables series for calibrating and validating statistical relationships. Then, it depends upon choices of predictors and GCM boundary forcings. Thirdly, it supposes that statistical relationships will remain constant in the future.

1.1.6 Modelling methods

Crop production is affected by multiple factors, such as climate, soil, and farming management. To explore relationships between climate and crop yield, complex impacts of climate and non-climate factors on crop growth should be systematically analysed. Process-based crop models and statistical models are usually two principal methods to carry out this task.

1.1.6.1 Process-based crop models

Long-term climate impacts on crops are of so much spatio-temporal complexity that field experiments are often lack of abilities to analyse these impacts. Crop models have abilities to diagnose crop growth, evaluate environmental impacts and predict crop yield at multiple spatial-temporal scales (Rosenzweig et al., 2014; Rosenzweig et al., 2013). In recent decades, many crop models have been developed for research and production management. For instance, APSIM (Agricultural Production Systems sIMulator) from Australia, CERES (Crop Environment Resource Synthesis) from the US, WOFOST (WOrld FOod STudies) from Europe have been widely used across the world. At present, crop model is the most powerful method to systematically assess complicated effects of many variables on crops.

Crop growth is affected by both climate and non-climate factors. Non-climate factors mainly include soil characteristics and field management practices. It is difficult to quantify effects of future climate, because of complicated effects originating from non-climate factors. Crop models show enormous potential in solving this problem. Many studies have demonstrated that crop models can not only assess individual impacts of warmer temperature, changed precipitation, and elevated CO_2 but also explore interactive relationships among various climate variables (Ludwig and Asseng, 2006). For example, Asseng et al. (2011) used APSIM to quantify how inter-annual temperature variability affected wheat yields in Australia. The result presented that observed

variability in mean growing-season temperature of ±2 °C can lead to declines of wheat yields by half when other climate factors remained unchanged during study years. Potgieter et al. (2013) assessed interactive effects of increased temperature and changed precipitation under two CO_2 emission scenarios projected for the periods 2020 and 2050 in Australia. They found that considering CO_2 fertilization effects attenuated adverse effect caused by warmer temperatures and decreased precipitation by 2050. Although crop models cannot perfectly reproduce real crop growth and development, they can generate hypotheses for testing and synthesis our existing knowledge and extend it beyond experimental sites and testing years so that general trends and relationships can be obtained. This is very important when determining optimal response strategies for the future across many sites.

Most crop models perform poorly in handling impacts of climate and weather extremes on plant development and growth (Palosuo et al., 2011). For example, most crop models rarely consider impacts of short-term heat stress on crop growth after post-anthesis (Wheeler et al., 2000). There is large uncertainty existing in climate impacts analysis using crop simulation models, especially the uncertainty related to model structure. Palosuo et al. (2011) compared 8 popular crop models on their abilities to predict wheat yields under various climate conditions in Europe. They concluded that simulations from these crop models were largely different, even though each model had been strictly calibrated. In addition, crop models when used in a new location often require long-term observed data to calibrate (Wallach, 2013), which is difficult to achieve in many regions.

1.1.6.2 Statistical models

Statistical models are also useful tools to explore climate-yield relationships (Tebaldi and Lobell, 2008). Statistical models are models that use observations of weather and crop yields to develop linear or non-linear relationships that functionally relate the former to the latter (Lobell and Asseng, 2017). Statistical models estimate functional relationships between historical observations of weather and yields, whereas crop models attempt to represent key dynamic processes affecting crop yields. Statistical models are easy to handle and cheap to compute. Statistical models usually have excellent performances, when observed data are sufficient and of high quality. Statistical models sometimes even performed better on yield predictions compared to crop models (Lobell, 2010). Many efforts have been made to empirically evaluate climatic effects on crop yield by combining climate observations with crop yields from long-term site-specific field observations to regional, national and global scales (Hansen and Indeje, 2004; Kucharik and Serbin, 2008). Thus, statistical models are powerful alternative tools to explore climate-crop relationships.

Linear statistical models are limited in detecting nonlinear relationships or identifying factors with multicollinearity (Sheehy et al., 2006). Separating contributions of solar radiation, rainfall and temperature to yield cannot be achieved using simple correlation because of the co-linearity problem between predictor variables (e.g. temperature and precipitation, temperature and radiation). Furthermore, some statistical models do not explicitly take into account extreme climate events, which makes them obtain overestimated simulation results during years with climate and weather extremes (Ceglar et al., 2016).

Machine learning methods which originate from artificial intelligence (Mitchell et al., 2003), are a family of statistical techniques. Machine learning theory is mainly to design and analyse algorithms that computer can automatically "learn". These algorithms can make automatic analysis based on known data and obtain the law, and then use the law to make predictions. These new emerged algorithms include artificial neural networks (ANN) (Schalkoff, 1997), genetic algorithms (GA) (Goldberg, 2006), decision tree (DT) (Quinlan, 1986), random forest (RF) (Breiman, 2001), support vector machines (SVM) (Chang and Lin, 2011).

The growing use of machine methods in recent years is the direct result of their ability to model complex, nonlinear relationships in input data without having to satisfy the restrictive assumptions required by conventional, parametric approaches (e.g., generalized linear models) (Olden et al., 2008). As a result, machine learning methods usually have higher accuracy in prediction compared to traditional statistical models. Those algorithms have gradually attracted wide attention and have been applied in many fields. Were et al. (2015) evaluated performances of SVM, ANN, and RF models in predicting and mapping soil organic carbon the Eastern Mau Forest Reserve, Kenya. All models obtained >0.5 coefficient of determination. Everingham et al. (2015) reported benefits that a machine learning method offers over contemporary, time-honored methods. They used the random forest modeling method to investigate how climate attributes relate to sugarcane productivity in Australia. Results showed that the random forest method was robust and gave more accurate predictions than traditional methods. Jeong et al. (2016) applied the random forest and multiple linear regressions to predict crop yields through blending of multiple biophysical variables. The random forest was highly capable of predicting crop yields and outperformed multiple linear regressions in all performance statistics that were compared. More studies using machine learning methods to explore inner relationships can be expected in the future.

1.2 Research questions and objectives

Agriculture is the leading industry in Australia. Australia's agricultural output as a proportion of the economy remains among the highest in Organisation for Economic Cooperation and Development (ABS, 2012). In grain production, wheat is the most vital cereal for planting area

and contribution value to the economy. Australia produces just three percent of world's wheat (about 25 million tonnes per annum) but accounts for more than 10 percent of world's annual wheat trade (AEGIC, 2016). Thus, wheat production in Australia is not only crucial to nation's economy but also to global food security. The NSW wheat belt is a major production area accounting for nearly 30% of national wheat area. Wheat production in this wheat belt is crucial to nation's food supply and even global food security as Australia is a major food exporter globally (AEGIC, 2016). However, wheat yield shows significant year-to-year fluctuation in the NSW wheat belt. According to ABS (2013), during 1991-2011, annual wheat production in this area varied from 875 kt to 10 500 kt and cultivated area ranged from 3000 to 4300 kha, resulting in yields produced per hectare varied greatly from 0.61 to 2.75 t/ha. These variations in wheat yield can be largely attributed to frequent climate and weather extremes, such as drought, flood, heat, and frost caused by climate variability (BOM and CSIRO, 2016). However, impacts of diverse types of climate extremes on wheat yields are complex and still poorly understood. Further, Australia is regarded as one of most sensitive areas worldwide that suffer significant impacts from climate change. Occurrences of climate and weather extremes is most likely to increase due to climate change (BOM and CSIRO, 2016; IPCC, 2012). Thus, wheat growth will suffer more harm from climate extremes in the future, but it is still unclear on expected effects in the whole wheat belt.

This study will make a systematic exploration of spatio-temporal patterns of climate and weather extremes and their effects on wheat productivity over the NSW wheat belt from last century to the future (~2100). Therefore, this study will provide answers to following important questions:

- How did various climate and weather extremes affect wheat yield in last few decades?
- How will climate and weather extremes (e.g. drought and heat) evolve in the future?
- How will wheat yield respond to more frequent climate and weather extremes in the future?

Specific aims of this study are to:

- Quantify the impacts of various climate and weather extremes on wheat yield at shire scales over past decades using machine learning technique.
- Monitor extreme climate and weather events by integrating various remotely-sensed drought factors using machine learning technique.
- Develop a farm-level wheat yield forecasting system by coupling APSIM crop model with machine learning technique.
- Project changes in spatio-temporal characteristics of extreme climate and weather events through statistical downscaled GCM projections.

• Evaluate the impacts of future climate and weather extremes on wheat yield using APSIM crop model and machine learning technique driven by statistical downscaled GCM projections.

1.3 Statement of significance

As the global population and living standards increase, demand for stable foods such as wheat is expected to increase by 60% towards the middle of 21st century. Sustainably improving crop production is urgently needed to meet this demand. However, the impacts of ongoing climate change will increase the risk of meeting this demand for crop production. In particular, climate change induced increases of extreme climate and weather events are considered as the major threat to crop production.

Accurately estimating historical and future yield losses caused by extreme climate and weather events is of great importance to maintain the sustainability of our agricultural production. Briefly, investigating the relationships between historical climate and weather extremes and crop yields can help us understand how unfavourable climatic conditions affect crop production. Coupled with timely and reliable monitoring of current climate and weather extremes, effective agronomic practices can then be developed to reduce yield losses. On the other hand, long-term strategies can also be established based on future climate projections to deal with increasing climate and weather extremes.

Modelling methods are commonly used tools for the estimation of climate impacts on crop yield at regional and multi-year scales. Two distinct modelling methods, i.e. process-based crop models and statistical models, have been widely used to examine climate-yield relationships. Process-based crop models are capable of describing key physical and physiological processes by capturing the effects of the complex interactions between crop, soil, climate, and management practices. However, crop models are usually limited in modelling the effects of climate and weather extremes due to oversimplification or vague description of certain process and uncertainties in parameterization.

Statistical models are simple and easy to understand and need fewer parameter settings, thereby making them widely-used around the world. As the observed data are increasing in both quantity and quality in recent years, statistical models usually present satisfactory performance. Moreover, newly emerging machine learning algorithms can improve the ability of statistical models to explore climate-yield relationships, as they are capable of disentangling the effects of co-linear climate variables and analysing hierarchical and nonlinear relationships between the predictors and the response variable. However, a major limitation of statistical models is that they usually only provide a simple evaluation of impacts, rather than provide a deeper understanding of physiological constraints required to inform adaptation strategies.

The value of combining both process-based crop models and statistical models is gaining recognition. Integrating the two kinds of models is expected to achieve complementary advantages and offer new insights of the effects of current and future climate and weather extremes on crop yield. Outcomes of this study will enhance the capabilities of farmers and policymakers to alleviate adverse effects of climate and weather extremes on crops under a changing climate.

1.4 Reference

ABS, 2012. Ausrtalian farming and farmers, viewed 9 April 2017 < http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Main+Features10Dec+2012 >.

ABS, 2013. Historical Selected Agriculture Commodities, by State (1861 to Present), 2010-11, viewed 9 April 2017, <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/7124.0Chapter72010-11>.

AEGIC, 2016. Australian grain production – a snapshot, viewed 9 April 2017, ≤http://aegic.org.au/australian-grain-production-a-snapshot/>.

Ahmadalipour, A., Moradkhani, H. and Svoboda, M., 2017. Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble. International Journal of Climatology, 37(5): 2477-2491.

Al-Issawi, M., Rihan, H.Z., El-Sarkassy, N. and Fuller, M.P., 2013. Frost Hardiness Expression and Characterisation in Wheat at Ear Emergence. Journal of Agronomy & Crop Science, 199(1): 66–74.

Anderson, D., 2014. Endurance: Australian Stories of Drought.

Asseng, S., Foster, I. and Turner, N.C., 2011. The impact of temperature variability on wheat yields. Global Change Biology, 17(2): 997–1012.

Barlow, K.M., Christy, B.P., O'Leary, G.J., Riffkin, P.A. and Nuttall, J.G., 2015. Simulating the impact of extreme heat and frost events on wheat crop production: A review. Field Crops Research, 171: 109-119.

Benestad, R.E., 2010. Downscaling precipitation extremes. Theoretical & Applied Climatology, 100(1-2): 1-21.

Beniston, M. et al., 2007. Future extreme events in European climate: an exploration of regional climate model projections. Climatic Change, 81: 71-95.

Blumenthal, C.S., Batey, I.L., Bekes, F., Wrigley, C.W. and Barlow, E.W.R., 1991. Seasonal-Changes in Wheat-Grain Quality Associated with High-Temperatures during Grain Filling. Australian Journal of Agricultural Research, 42(1): 21-30.

Boer, G.J., Flato, G., Reader, M.C. and Ramsden, D., 2000. A transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the twentieth century. Clim Dynam, 16(6): 405-425.

Boer, R., Campbell, L.C. and Fletcher, D.J., 1993. Characteristics of Frost in a Major Wheat-Growing Region of Australia. Australian Journal of Agricultural Research, 44(8): 1731-1743.

BOM and CSIRO, 2016. State of the Climate.

Breiman, L., 2001. Random Forest. Machine Learning, 45: 5-32.

Ceglar, A., Toreti, A., Lecerf, R., Van der Velde, M. and Dentener, F., 2016. Impact of meteorological drivers on regional inter-annual crop yield variability in France. Agr Forest Meteorol, 216: 58-67.

Chang, C.C. and Lin, C.J., 2011. LIBSVM: A library for support vector machines. Acm Transactions on Intelligent Systems & Technology, 2(3): 27.

Charles, S.P., Bari, M.A., Kitsios, A. and Bates, B.C., 2010. Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia. International Journal of Climatology, 27(12): 1673-1690.

Chaves, M.M. et al., 2002. How Plants Cope with Water Stress in the Field? Photosynthesis and Growth. Ann Bot-London, 89(7): 907-916.

Coumou, D. and Rahmstorf, S., 2012. A decade of weather extremes. Nature Climate Change, 2(7): 491-496.

Crimp, S.J. et al., 2017. Recent changes in southern Australian frost occurrence: implications for wheat production risk. Crop & Pasture Science, 67(8): 801-811.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nature Climate Change, 3(1): 52-58.

Easterling, D.R. et al., 2000. Observed Variability and Trends in Extreme Climate Events: A Brief Review *. Bulletin of the American Meteorological Society, 81(3): 417-426.

Everingham, Y., Sexton, J. and Robson, A., 2015. A statistical approach for identifying important climatic influences on sugarcane yields. In: Proceedings of the 37th Annual Conference of the Australian Society of Sugar Cane Technologists, 37(37): 8-15.

Falloon, P., Betts, R., Macleod, C.J.A. and Haygarth, P.M., 2010. Climate impacts on European agriculture and water management in the context of adaptation and mitigation - the importance of an integrated approach. Sci Total Environ, 408(23): 5667-87.

Fischer, E.M., Seneviratne, S.I., Vidale, P.L., Lüthi, D. and Schär, C., 2007. Soil moisture–atmosphere interactions during the 2003 European summer heat wave. Journal of Climate, 20(20): 5081-5099.

Folland, C.K., Karl, T.R. and Salinger, M.J., 2002. Observed climate variability and change. Weather, 57(8): 269–278.

Fowler, H.J., Blenkinsop, S. and Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27(12): 1547–1578.

Frederiks, T.M., Christopher, J.T., Harvey, G.L., Sutherland, M.W. and Borrell, A.K., 2012. Current and emerging screening methods to identify post-head-emergence frost adaptation in wheat and barley. Journal of Experimental Botany, 63(15): 5405-5416.

Fuller, M.P., Fuller, A.M., Kaniouras, S., Christophers, J. and Fredericks, T., 2007. The freezing characteristics of wheat at ear emergence. European Journal of Agronomy, 26(4): 435-441.

Giorgi, F. and Francisco, R., 2000. Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. Clim Dynam, 16(2-3): 169-182.

Goldberg, D.E., 2006. Genetic algorithms. Pearson Education India.

Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S. and Xavier, P.K., 2006. Increasing trend of extreme rain events over India in a warming environment. Science, 314(5804): 1442-1445.

Hansen, J.W. and Indeje, M., 2004. Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agr Forest Meteorol, 125(1-2): 143-157.

He, L. et al., 2015. Impacts of recent climate warming, cultivar changes, and crop management on winter wheat phenology across the Loess Plateau of China. Agricultural & Forest Meteorology, 200(4): 135-143.

Hennessy, K.J., Gregory, J.M. and Mitchell, J.F.B., 1997. Changes in daily precipitation under enhanced greenhouse conditions. Clim Dynam, 13(9): 667-680.

Hewitt, C.D., Senior, C.A. and Mitchell, J.F.B., 2001. The impact of dynamic sea-ice on the climatology and climate sensitivity of a GCM: a study of past, present, and future climates. Clim Dynam, 17(9): 655-668.

IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental Panel on Climate Change. Journal of Clinical Endocrinology & Metabolism, 18(6): 586-599.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia,V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Jeong, J.H. et al., 2016. Random Forests for Global and Regional Crop Yield Predictions. Plos One, 11(6): e0156571.

Ji, K. et al., 2012. Drought-responsive mechanisms in rice genotypes with contrasting drought tolerance during reproductive stage. Journal of Plant Physiology, 169(4): 336-344.

Kharin, V.V. and Zwiers, F.W., 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. Journal of Climate, 13(21): 3760-3788.

Kharin, V.V., Zwiers, F.W., Zhang, X. and Hegerl, G.C., 2013. Changes in Temperature and Precipitation Extremes in the IPCC Ensemble of Global Coupled Model Simulations. Climatic Change, 119(2): 345-357.

Kilsby, C.G. et al., 2007. A daily weather generator for use in climate change studies. Environ Modell Softw, 22(12): 1705-1719.

Kloster, S. et al., 2010. A GCM study of future climate response to aerosol pollution reductions. Clim Dynam, 34(7-8): 1177-1194.

Kucharik, C.J. and Serbin, S.P., 2008. Impacts of recent climate change on Wisconsin corn and soybean yield trends. Environmental Research Letters, 3(3).

L'Heureux, M., 2014. What is the El Niño–Southern Oscillation (ENSO) in a nutshell. https://www.climate.gov/news-features/blogs/enso/what-el-ni%C3%B1o%E2%80%93southern-oscillation-enso-nutshell. assessed at 13 Jan 2018.

Latif, M. and Barnett, T.P., 1994. Causes of Decadal Climate Variability over the North Pacific and North-America. Science, 266(5185): 634-637.

Li, Z., Zheng, F.L., Liu, W.Z. and Jiang, D.J., 2012. Spatially downscaling GCMs outputs to project changes in extreme precipitation and temperature events on the Loess Plateau of China during the 21st Century. Global Planet Change, 82-83: 65-73.

Liang, J., Zhang, J. and Cao, X., 2001. Grain sink strength may be related to the poor grain filling of indica-japonica rice (Oryza sativa) hybrids. Physiologia Plantarum, 112(4): 470–477.

Liu, B. et al., 2016. Testing the responses of four wheat crop models to heat stress at anthesis and grain filling. Global Change Biology, 22(5): 1890.

Lobell, D., 2010. Crop Responses to Climate: Time-Series Models. Springer Netherlands, 85-98 pp.

Lobell, D.B. and Asseng, S., 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environmental Research Letters, 12(1): 015001.

Lobell, D.B., Sibley, A. and Ortiz-Monasterio, J.I., 2012. Extreme heat effects on wheat senescence in India. Nature Climate Change, 2(3): 186-189.

Ludwig, F. and Asseng, S., 2006. Climate change impacts on wheat production in a Mediterranean environment in Western Australia. Agricultural Systems, 90(1-3): 159-179.

Maloney, E.D. et al., 2014. North American Climate in CMIP5 Experiments: Part III: Assessment of Twenty-First-Century Projections*. Journal of Climate, 27(6): 2230-2270.

Marcellos, H. and Single, W.V., 1984. Frost Injury in Wheat Ears after Ear Emergence. Aust J Plant Physiol, 11(1-2): 7-15.

Mearns, L.O. et al., 2003. Guidelines for Use of Climate Scenarios Developed from Regional Climate Model Experiments.

Mitchell, T.M., Carbonell, J.G. and Michalski, R.S., 2003. Machine Learning. China Machine Press ;McGraw-Hill Education (Asia), 417-433 pp.

Murphy, B.F. and Timbal, B., 2008. A review of recent climate variability and climate change in southeastern Australia. International Journal of Climatology, 28(7): 859-879.

Olden, J.D., Lawler, J.J. and Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly review of biology, 83(2): 171-193.

Özdoğan, 2011. Modeling the impacts of climate change on wheat yields in Northwestern Turkey. Agriculture Ecosystems & Environment, 141(1–2): 1-12.

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(3): 103–114.

Panthou, G., Vischel, T. and Lebel, T., 2014. Recent trends in the regime of extreme rainfall in the Central Sahel. International Journal of Climatology, 34(15): 3998-4006.

Perkins-Kirkpatrick, S. and Gibson, P., 2017. Changes in regional heatwave characteristics as a function of increasing global temperature. Scientific Reports, 7(1): 12256.

Perkins, S., Alexander, L. and Nairn, J., 2012a. Increasing frequency, intensity and duration of observed global heatwaves and warm spells. Geophysical Research Letters, 39(20).

Perkins, S. and Pitman, A., 2009. Do weak AR4 models bias projections of future climate changes over Australia? Climatic change, 93(3-4): 527-558.

Perkins, S.E., Argüeso, D. and White, C.J., 2015. Relationships between climate variability, soil moisture, and Australian heatwaves. Journal of Geophysical Research: Atmospheres, 120(16): 8144-8164.

Perkins, S.E. et al., 2012b. CMIP3 ensemble climate projections over the western tropical Pacific based on model skill. Climate Research, 51(1): 35-58.

Perry, E.M., Nuttall, J.G., Wallace, A.J. and Fitzgerald, G.J., 2017. In-field methods for rapid detection of frost damage in Australian dryland wheat during the reproductive and grain-filling phase. Crop & Pasture Science, 68.

Peterson, T.C. et al., 2013. Monitoring and Understanding Changes in Heat Waves, Cold Waves, Floods, and Droughts in the United States: State of Knowledge. Bulletin of the American Meteorological Society, 94(6): 821-834.

Planton, S., Déqué, M., Chauvin, F. and Terray, L., 2008. Expected impacts of climate change on extreme climate events. Comptes Rendus Geosciences, 340(9): 564-574.

Potgieter, A. et al., 2013. Spatial impact of projected changes in rainfall and temperature on wheat yields in Australia. Climatic Change, 117(1-2): 163-179.

Pradhan, G.P., Prasad, P.V.V., Fritz, A.K., Kirkham, M.B. and Gill, B.S., 2012. Effects of drought and high temperature stress on synthetic hexaploid wheat. Functional Plant Biology, 39(3): 190-198.

Prudhomme, C. and Davies, H., 2009. Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: Baseline climate. Climatic Change, 93(1-2): 177-195.

Quinlan, J.R., 1986. Induction on decision tree. Machine Learning, 1(1): 81-106.

Rajaram, S., 2001. Prospects and promise of wheat breeding in the 21st century. Euphytica, 119(1-2): 3-15.

Reynolds, M.P., Mujeebkazi, A. and Sawkins, M., 2015. Prospects for utilising plant-adaptive mechanisms to improve wheat and other crops in drought- and salinity-prone environments. Annals of Applied Biology, 146(2): 239-259.

Rosenzweig, C. et al., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proceedings of the National Academy of Sciences of the United States of America, 111(9): 3268-73.

Rosenzweig, C. et al., 2013. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. Agr Forest Meteorol, 170: 166-182.

Schalkoff, R.J., 1997. Artificial neural networks, 1. McGraw-Hill New York.

Schär, C. et al., 2004. The role of increasing temperature variability in European summer heatwaves. Nature, 427(6972): 332-336.

Schlenker, W. and Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences of the United States of America, 106(37): 15594-15598.

Seneviratne, S.I., Lüthi, D., Litschi, M. and Schär, C., 2006. Land–atmosphere coupling and climate change in Europe. Nature, 443(7108): 205-209.

Sheehy, J.E., Mitchell, P.L. and Ferrer, A.B., 2006. Decline in rice grain yields with temperature: Models and correlations can give different estimates. Field Crops Research, 98(2): 151-156.

Stone, P.J., Nicolas, M.E., Stone, P.J. and Nicolas, M.E., 1994. Wheat Cultivars Vary Widely in Their Responses of Grain Yield and Quality to Short Periods of Post-Anthesis Heat Stress. Functional Plant Biology, 21(6): 887-900.

Tebaldi, C. and Lobell, D.B., 2008. Towards probabilistic projections of climate change impacts on global crop yields. Geophysical Research Letters, 35(8).

Teixeira, E.I., Fischer, G., Velthuizen, H.V., Walter, C. and Ewert, F., 2013. Global hot-spots of heat stress on agricultural crops due to climate change. Agricultural & Forest Meteorology, 170(2): 206-215.

Ummenhofer, C.C. et al., 2015. How did ocean warming affect Australian rainfall extremes during the 2010/2011 La Nina event? Geophysical Research Letters, 42(22): 9942-9951.

Van der Velde, M., Tubiello, F.N., Vrieling, A. and Bouraoui, F., 2012. Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. Climatic Change, 113(3-4): 751-765.

Venkataraman, K., Tummuri, S., Medina, A. and Perry, J., 2016. 21st century drought outlook for major climate divisions of Texas based on CMIP5 multimodel ensemble: Implications for water resource management. Journal of Hydrology, 534: 300-316.

Villarini, G., Smith, J.A. and Vecchi, G.A., 2013. Changing Frequency of Heavy Rainfall over the Central United States. Journal of Climate, 26(1): 351-357.

Wallach, D., 2013. Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications - Daniel Wallach.

Wang, L. and Chen, W., 2014. A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. International Journal of Climatology, 34(6): 2059-2078.

Watterson, I., G, 2009. Components of precipitation and temperature anomalies and change associated with modes of the Southern Hemisphere. International Journal of Climatology, 29(6): 809-826.

Were, K., Bui, D.T., Dick, O.B. and Singh, B.R., 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. Ecol Indic, 52: 394-403.

Wheeler, T.R., Craufurd, P.Q., Ellis, R.H., Porter, J.R. and Prasad, P.V.V., 2000. Temperature variability and the yield of annual crops. Agriculture Ecosystems & Environment, 82(1-3): 159-167.

Whetton, P., Fowler, A., Haylock, M. and Pittock, A., 1993. Implications of climate change due to the enhanced greenhouse effect on floods and droughts in Australia. Climatic Change, 25(3): 289-317.

Woldemeskel, F.M., Sharma, A., Sivakumar, B. and Mehrotra, R., 2014. A framework to quantify GCM uncertainties for use in impact assessment studies. Journal of Hydrology, 519: 1453-1465.

Wollenweber, B., Porter, J.R. and Schellberg, J., 2003. Lack of Interaction between Extreme High-Temperature Events at Vegetative and Reproductive Growth Stages in Wheat. Journal of Agronomy and Crop Science, 189(3): 142–150.

Yang, X. et al., 2017. The impacts of increased heat stress events on wheat yield under climate change in China. Climatic Change, 140(3).

Yasufuku, N., Araki, K. and Omine, K., 2015. Evaluation of Inhibitory Effect by Adaptation Measures for Red Soil Runoff from Farmland due to Heavy Rainfall (Special Issue on Adaptation Measures for Disasters due to Climate Change). Journal of Disaster Research, 10: 457-466.

Zhao, Y., Ducharne, A., Sultan, B., Braconnot, P. and Vautard, R., 2015. Estimating heat stress from climate-based indicators: present-day biases and future spreads in the CMIP5 global climate model ensemble. Environmental Research Letters, 10(8).

Zhao, Y. et al., 2013. Dissecting the genetic architecture of frost tolerance in Central European winter wheat. Journal of Experimental Botany, 64(14): 4453.

Zwiers, F.W. and Kharin, V.V., 1998. Changes in the Extremes of the Climate Simulated by CCC GCM2 under CO2 Doubling. Journal of Climate, 11(9): 2200-2222.

Chapter 2. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia

This chapter is based on the following publication (journal paper):

Feng, P., Wang, B., Liu, D.L., Xing, H., Ji, F., Macadam, I., Ruan, H. and Yu, Q., 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic change, 147(3-4), pp.555-569.

Abstract

Investigating the relationships between climate extremes and crop yield can help us understand how unfavourable climatic conditions affect crop production. In this study, two statistical models, multiple linear regression and random forest, were used to identify rainfall extremes indices affecting wheat yield in 3 different regions of the New South Wales wheat belt. The results show that the random forest model explained 41%-67% of the year-to-year yield variation, whereas the multiple linear regression model explained 34%-58%. In the two models, 3-month timescale Standardized precipitation index of Jun.-Aug. (SPI_{JJA}), Sep.-Nov. (SPI_{SON}), and consecutive dry days (CDD) were identified as the three most important indices which can explain yield variability for most of the wheat belt. Our results indicated that the inter-annual variability of rainfall in winter and spring was largely responsible for wheat yield variation, and pre-growing season rainfall played a secondary role. Frequent shortages of rainfall posed a greater threat to crop growth than excessive rainfall in eastern Australia. We concluded that the comparison between multiple linear regression and machine learning algorithm proposed in the present study would be useful to provide robust prediction of yields and new insights of the effects of various rainfall extremes, when suitable climate and yield datasets are available.

Key words: Rainfall variation; Wheat yield variation; NSW wheat belt; Multiple linear regression; Random forest

2.1 Introduction

The special report of the Intergovernmental Panel on Climate Change has emphasized the adverse effect of extreme climate events on crop yield (IPCC, 2012). In particular, rainfall variability is widely considered to be the direct cause of crop yield fluctuation in semi-arid environments across the world (Gichangi et al., 2015). Large variability in the intensity, frequency, and timing of annual or seasonal rainfall creates significant challenges for farmers. Understanding the relationship between rainfall extremes and historical crop yields is vital for assessing the sustainability of our agricultural production.

The New South Wales (NSW) wheat belt in eastern Australia is a major wheat production area and accounts for nearly 30% of the national wheat area. Wheat production in NSW is crucial to the nation's food supply and contributes to global food security because Australia is one of the world's major food exporters (http://aegic.org.au/australian-grain-production-a-snapshot/). However, wheat yields are highly variable in the NSW wheat belt. According to the Australia Bureau of Statistics (http://www.abs.gov.au/Agriculture), during 1991-2011, annual wheat production in this area varied from 875 kt to 10 488 kt, with wheat yields ranging between 0.61 to 2.75 $t \cdot ha^{-1}$. To explore the reason for yield variability, Ray et al. (2015) found that in the NSW wheat belt, inter-annual rainfall variability could explain ~40% of the total yield variations. Potgieter et al. (2016) also blamed wheat yield fluctuations during 1975-2010 on frequent rainfall shortage in this area. However, the lack of consideration on the contribution of rainfall extremes limits further understanding of how wheat yields respond to rainfall variability.

The principle methods to explore climate-yield relationships are crop modelling and statistical analysis. Crop models that account for multiple climatic factors, in addition to crop, soil, and management parameters, can promote a better understanding of crop response to climate (Rosenzweig et al., 2014). The main advantage of using a crop model is that it creates a comprehensive characterization of the cropping system. If crop models are accurately calibrated with observed data, they can be applied to simulate possible management interactions to better cope with anticipated changes in climate (Liu et al., 2010). However, most crop models perform poorly in handling the effects of extreme climate events on crop growth and development (Moriondo et al., 2011). Some of this poor performance relates to the simple description of certain processes, which can lead to inaccurate results. Also, crop models require several years of experimental data to train and calibrate in the local environment (Chen et al., 2010) and a recalibration needs to be conducted when they are used in other regions.

Because of these limitations in crop models, some linear statistical models, such as multiple linear regression, have been widely used in characterizing the relationship between yields and climate variables (Tebaldi and Lobell, 2008). Linear models are easy to handle and cheap to compute. With the increasing availability and improved quality of observed data, linear models usually perform well. Innes et al. (2015) suggested a superior performance of linear models in comparison with crop models to identify climate-yield relationships. However, linear models are incapable of detecting nonlinear relationships or identifying factors with multicollinearity. Multicollinearity arises when two or more explanatory variables in a multiple regression model are highly linearly correlated, which will result in incorrect estimates of coefficient in the multiple regression (Farrar and Glauber, 1967).

In recent decades, machine learning algorithms have gradually attracted wide attention and have been applied in many fields. Machine learning methods can investigate nonlinear and hierarchical

relationships between the predictors and the response using an ensemble learning approach (Shalev-Shwartz and Ben-David, 2014). They usually perform well in prediction compared to the traditional linear regression model. Everingham et al. (2015) reported that a machine learning method is superior to contemporary, time-honoured methods. They used a random forest (RF) modelling method (Breiman, 2001) to investigate how climate attributes relate to sugarcane productivity in Australia. The results showed that the RF method was robust and provided accurate predictions of sugarcane yield. Jeong et al. (2016) applied the RF and multiple linear regressions (MLR) to predict crop yields through the blending of multiple biophysical variables. The RF was highly capable of predicting crop yields and outperformed the MLR in all performance statistics that were compared. Despite this, few studies have used machine learning methods to identify the impacts of rainfall extremes on crop yield.

In this study, we used rainfall extremes indices to investigate the relationship between extreme rainfall events and shire-level wheat yields in 3 geographical-climate sub-regions across the NSW wheat belt. The RF and MLR models were used to evaluate the effects of extreme rainfall events on wheat yields. The main objectives of this study were to (1) quantify the relationship between extreme rainfall and observed wheat grain yields for the period 1922-2000, (2) identify the contribution of each extreme rainfall index to wheat yield and (3) analyze the temporal variation of the key indices.

2.2 Materials and methods

2.2.1 Study area

The NSW wheat belt covers 66 shires, most of which are dominated by a Mediterranean climate with large rainfall variation. Based on geography and topography, we divided the wheat belt into 3 sub-regions: (I) eastern slopes, (II) northern plains and (III) southern plains (Figure 2-1). Region I is the eastern part of the wheat belt and mainly occupied by mountains, with an elevation of more than 500 m. Region I has more than 400 mm growing season rainfall and less than 12 °C growing season temperature, making it the wettest region in the wheat belt. Region II and III are the northwestern and southwestern part, respectively, and mainly plains. Region II has less than 250 mm growing season rainfall and more than 14 °C growing season temperature, compared to 300 mm and 12.5 °C in region III. In addition, the 3 sub-regions consist of 20, 12, and 34 shires, respectively. According to OEH (2017), the dominant soil types in the 3 sub-regions are sandy clay, clay and sandy loam, respectively. The clay content of 0-60 cm soil in region II is more than 45%, compared to 20% and 30% in region I and region III.

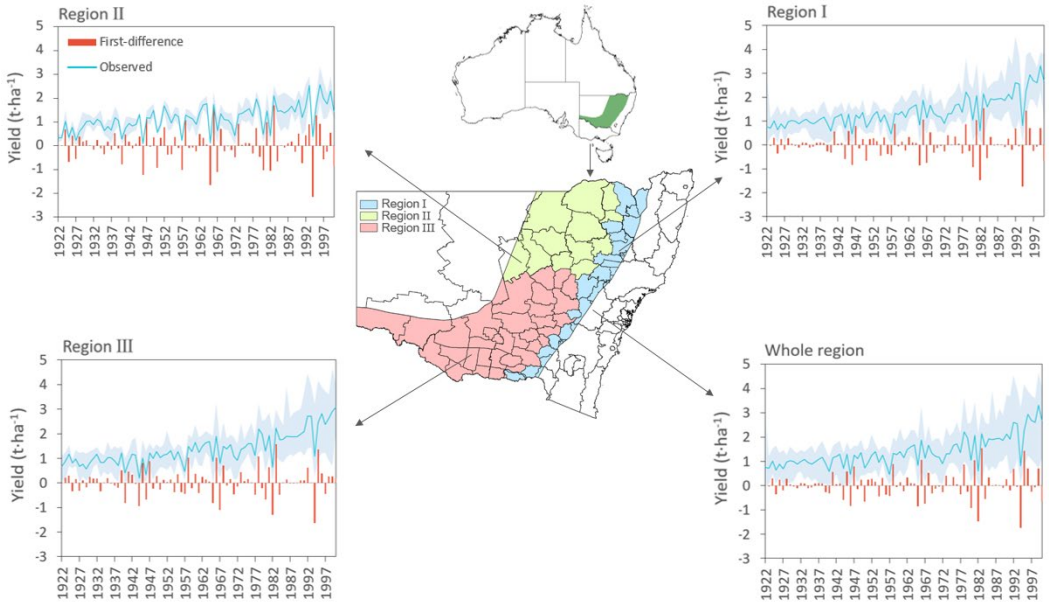


Figure 2-1 Averaged observed (blue, 1922-2000) and first-difference (red, 1923-2000) wheat yields in three sub-regions and the whole region of the New South Wales wheat belt. The top and bottom bounds of the shaded area are the maximum and minimum yields in each region.

2.2.2 Climate and yield data

In this study, historical daily climate data (1922-2000) were downloaded from SILO (Scientific Information for Land Owners) patched point dataset (http://www.longpaddock.qld.gov.au/silo/ppd/index.php) (Jeffrey et al., 2001). Daily rainfall data at 932 weather stations were used for the analysis. These stations are quite evenly distributed across the NSW wheat belt. Wheat yields for each shire during 1922-2000 across the NSW wheat belt were obtained from Fitzsimmons (2001).

2.2.3 Rainfall extreme indices

Seven rainfall extreme indices were selected for the study (Table 2-1) from those listed by the World Meteorological Organization Expert Team on Sector-specific Climate Indices (WMO ET-SCI, http://www.wmo.int/pages/prog/wcp/ccl/opace/opace4/ET-SCI-4-1.php). These indices were calculated using daily rainfall data. Some are threshold based indices, including CDD, CWD, Rx1day, Rx5day, SDII, and R10mm. For these indices, thresholds are the same for all stations. While for R99pTOT, the threshold varies from location to location and is typically defined as a percentile of the data for specified recent base period. In this study, the whole study period (1922-2000) was set as the base period. All of these indices were calculated only for growing season (Apr.-Nov.).

Table 2-1 List of 11 rainfall extremes indices.

ID	Indicator name	Definition	Unit
CDD	consecutive dry days	maximum number of consecutive days with RR* < 1 mm	days
CWD	consecutive wet days	maximum number of consecutive days with RR ⩾ 1mm	days
Rx1day	max 1-day precipitation total	maximum 1-day precipitation	mm
Rx5day	max 5-days precipitation total	maximum 5-day precipitation	mm
SDII	simple daily intensity index	total precipitation divided by the number of wet days (when total precipitation ⩾ 1.0 mm) during the growing season	mm
R10mm	number of heavy precipitation days	count of days when precipitation ⩾ 10 mm during the growing season	days
R99pTOT	extremely wet days	total precipitation when RR > 99th percentile during the growing season	mm
SPI_{DJF}	3-month time-scale standardized precipitation index of Dec.-Feb.	rainfall probability distribution that reflects dry/wet conditions for Dec-Feb	
SPI_{MAM}	3-month time-scale standardized precipitation index of Mar.-May	rainfall probability distribution that reflects dry/wet conditions for Mar-May	
SPI_{JJA}	3-month time-scale standardized precipitation index of Jun.-Aug.	rainfall probability distribution that reflects dry/wet conditions for Jun-Aug	
SPI_{SON}	3-month time-scale standardized precipitation index of Sep.-Nov.	rainfall probability distribution that reflects dry/wet conditions for Sep-Nov	

*Note: *RR* is the daily rainfall amount. Growing season is from the beginning of April to the end of November.

Standardized precipitation index (SPI) which is also listed by ET-SCI, is a probability-based indicator that depicts the degree to which accumulative precipitation for a specific time period departs from the average state (Mckee et al., 1993). Since the SPI is standardized, an index of 0 indicates the median precipitation amount (i.e., normal conditions), while dry conditions are indicated by negative values (i.e., -2 for extremely dry) and wet conditions are indicated by positive values (i.e., 2 for extremely wet). It should be noted that dryness and wetness are relative

to the historical average rather than the absolute total of rainfall at a certain station. For example, a given amount of rainfall at a wet station that produces negative SPI may produce positive SPI values at a dry station.

An important aspect of SPI is its ability to track dry and wet events on various timescales from 1 to 48 months. It is widely accepted that SPI at 1 to 6-month scale can be applied to agricultural interests, whereas SPIs at 6 up to 24-month scale are more suitable for hydrological drought analyses and applications (Raziei et al., 2009). In this study, four 3-month scale SPIs which represent rainfall conditions for four seasons were used (Table 2-1). For example, SPI_{DJF} is the SPI for Dec.-Feb., reflecting dry/wet conditions for austral summer. In addition, the baseline period was set as 1922-2000 when computing SPIs.

All the 11 indices for each weather station were calculated using the R software (R Core Team, 2016). Then, the indices at the weather stations within each shire were averaged to produce shire-average indices.

2.2.4 De-trending method

Crop yield is affected by both climatic and non-climatic factors. Non-climatic factors, including breeding, fertilizer, pesticide application etc., are main drivers of yield increase. To separately assess the effect of climate on yield variation, the yield increase by factors other than climate should be excluded. In this study, a first-difference method (Lobell and Asner, 2003) was used. This method is easy to implement and can minimize the influence of non-climatic factors, enabling the explanation of climate-crop yield relationships. All time series of yields and rainfall extreme indices are calculated using first differences approach through the following equation:

$$\Delta X_{(t)} = X_{(t)} - X_{(t-1)}, \; t = 1923, 1924, \ldots, 2000 \qquad (2\text{-}1)$$

where $\Delta X_{(t)}$ denotes the first difference of X at year t, $\Delta X_{(t)}$ denotes the value of time series X at year t and $X_{(t\text{-}1)}$ is the value for the (t-1)th year.

2.2.5 Models

Multiple linear regression (MLR) is used to explain the relationship between one continuous dependent variable and two or more independent variables. The MLR was performed using the R package 'Rattle' (Williams, 2011). The contribution of each independent variable was assessed through relative importance measures calculated with the R package "relaimpo" (Gromping, 2006). The metric "lmg" was used. It represents the R^2 (coefficient of determination) contribution of each variable.

Random forest (RF) is an ensemble learning algorithm developed by Breiman (2001). It is a nonparametric technique based on classification and regression trees. The RF consists of

numerous independent trees, where each tree is generated by bootstrap samples, leaving about a third of the overall sample for validation. Each split of the tree is determined using a randomized subset of the predictors at each node. The final outcome is the average of the results of all the trees. The RF can explore nonlinear and hierarchical relationships between the predictors and the response. It has been applied in agricultural studies (Jeong et al., 2016), showing high accuracy and ability to model complex interactions between variables. However, this method behaves as a "black box" since the individual trees cannot be examined separately and it does not calculate regression coefficients nor confidence intervals (Cutler et al., 2007). Nevertheless, it produces a variable importance list that can be compared to other regression models. The RF was also implemented through the 'Rattle' package. We applied the RF model with default settings, ntree (the number of trees to grow in the forest) = 500 and mtry (the number of randomly selected predictor variables at each node) = 3. The relative importance of variables was estimated using the "%IncMSE" metric. The %IncMSE indicates the mean increase of mean square error in nodes that use a variable in the RF model, when values of the variable are randomly permuted.

The performance of the MLR and RF models was evaluated using the same procedure. 80% of the dataset was randomly selected to calibrate a model, and the remaining data were used to validate the model. This procedure was executed 100 times with different randomly selected calibration and validation datasets to evaluate model stability. Four validation measurements were calculated: mean absolute prediction error (MAE), root mean square error (RMSE), coefficient of determination (R^2) and Lin's concordance correlation coefficient (LCCC) (Lin, 1989). These indices were calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right| \tag{2-2}$$

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

$$R^2 = \left(\frac{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2}\sqrt{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2}}\right)^2 \tag{2-4}$$

$$LCCC = \frac{2r\sigma_O\sigma_P}{\sigma_O^2 + \sigma_P^2 + \left(\overline{O} - \overline{P}\right)^2} \tag{2-5}$$

where P_i and O_i are the predicted and observed values, respectively; $\overline{O}$ and $\overline{P}$ express the mean of observed and predicted values, respectively; n is the number of samples; σ_P and σ_O are the variances of predicted and observed values; and r is the Pearson correlation coefficient between

the predicted and observed values. MAE measures the average prediction bias, and RMSE represents the sample standard deviation of the differences between predicted and observed values. LCCC represents the degree to which the predicted and observed values follow the 45° line through the origin. Predictions become increasingly accurate as MAE and RMSE approach 0 and R^2 and LCCC approach 1.

2.3 Results

2.3.1 Descriptive statistics of wheat yield

The continual innovation of agricultural technology and introduction of new cultivars combined with dynamical changes in fertilization normally result in a long-term increasing trend across all regions (Figure 2-1). However, large inter-annual variations of yield were observed in each region. The coefficient of variation (CV) for observed yield series for region I, II, III, and the whole region were 44.6%, 47.8%, 46.7% and 44.0%, indicating high yield variations. Furthermore, the first-difference yield (Δyield) also clearly showed the temporal yield variation. Frequent and severe ups and downs of Δyield existed in all regions, with region III had the largest fluctuation range, from -2.14 to 1.70 $t \cdot ha^{-1}$. The Δyield excludes the effects of non-climatic factors, so its variability is mostly related to climate variability. In addition, the fluctuation range of Δyield amplified in the late 20^{th} century (1980-2000), indicating an increasingly adverse effect on wheat yield caused by climate variability.

2.3.2 Model performance

Before model fitting, to minimize the influence of multicollinearity amongst the independent variables, we calculated the variance inflation factor (VIF) and iteratively removed variables with >10 VIF values, because a VIF value of <10 is acceptable threshold in climate related studies (Doetterl et al., 2015). In this process, we discarded the variable R10mm and kept other 10 variables with <10 VIF values (Table 2-S1). Figure 2-2 shows the predictive performance evaluation of the RF and MLR models based on the validation procedure. The results suggest that the RF performed better than the MLR in all 3 sub-regions and the whole NSW wheat area based on the 4 validation measurements. R^2 measures the covariance of the observed and predicted values. For region I, the RF model can explain more than 40% of the yield variation, compared to approximately 35% for the MLR model. The RF model can explain 60%-70% of the yield variation, versus 40%-60% for the MLR model, for region II, III, and the whole region. Both models performed poorly in region I because this region had relatively high rainfall during main growing seasons (winter and spring, Figure 2-S1) and rarely suffered from water shortage. Rainfall was not a limiting factor in region I and other climate factors, such as temperature or radiation, might contribute greatly to yield fluctuation. Modelling uncertainty can be evaluated

by the range between the 10th percentile and 90th percentile of each set of 100 validation results. The RF and MLR models produced similar 10th-90th intervals of MAE, RMSE, R^2, and LCCC, indicating that RF and MLR models have similar stability in their predictive ability.

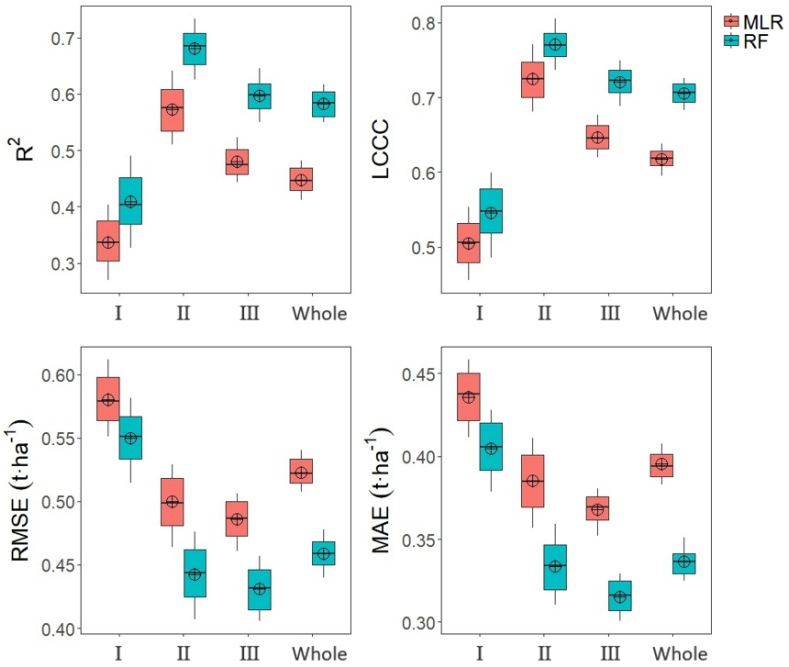


Figure 2-2 Summary statistics of the predictive quality of random forest (RF) and multiple linear regression (MLR) models for the yield variation with 100 runs in 3 sub-regions and the whole New South Wales wheat belt. Mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and Lin's concordance correlation coefficient (LCCC) 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.

2.3.3 Relative importance of rainfall extreme indices

The relative importance of each index generated by the MLR and RF models is shown in Figure 2-3. We normalized the %IncMSE of variables in the RF model to sum to 100% to provide a simple basis for comparison with the MLR model. The relative importance of indices varied between the two models in each region. However, both models indicated that ΔSPI_{JJA}, ΔSPI_{SON} and CDD can be regarded as the most important indices for any of three regions. Both models also simultaneously attached lower importance values to several indices, such as ΔRx5day and ΔCWD. Other indices were not consistent for the two models. For example, in each region, ΔSPI_{DJF} and ΔSPI_{MAM} were identified as having high importance in the RF model, but low in the MLR model. The opposite was true for ΔSDII and ΔRx1day. In addition, indices also ranked differently in different regions. ΔSPI_{DJF} had high importance in region II but lower in region I. ΔSPI_{SON} ranked relatively low in region II, compared to its superior performance in other regions.

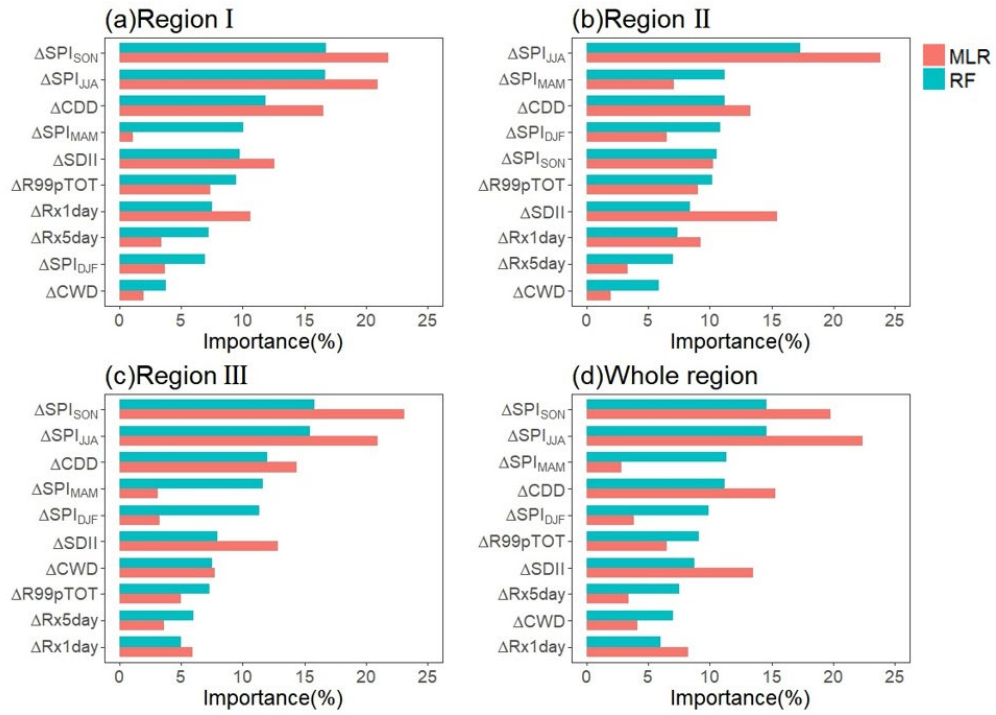


Figure 2-3 Relative importance of rainfall extreme indices as determined from the random forest (RF) and multiple linear regression (MLR) models for each region. The results are normalized to sum to 100% and shown in decreasing order of RF importance.

For the MLR model, the relationships between first-difference wheat yield and rainfall extreme indices are summarized in Table 2-2. The four SPI indices all had positive effects on wheat yield, and ΔSPI_{JJA} and ΔSPI_{SON} were more important than the other two. However, SPI_{DJF} and SPI_{MAM} had greater effects in region II when compared with other regions, indicating the great importance of pre-growing season rainfall in region II. This is mainly because region II has a summer-dominant rainfall pattern (Figure 2-S1). Summer rainfall can contribute to stored soil water, which can be used for water uptake of wheat during growing season. ΔCDD denotes the maximum duration of a rain-free period and it had a negative effect on wheat yield. ΔR99pTOT and ΔRx5day, which reflect heavy rainfall events, also had negative effects on wheat yield. However, ΔRx1day and ΔSDII had positive effects. Extreme heavy rainfall usually adversely affects crop growth, but the effect of each heavy rainfall event depends on its intensity and timing.

Table 2-2 Regression coefficients of first-difference wheat yield (ΔYield, $kg \cdot ha^{-1}$) with the 11 rainfall extreme indices in the multiple linear regression (MLR) model ($\Delta Yield = a\Delta SPI_{MAM} + b\Delta SPI_{JJA} + c\Delta SPI_{SON} + d\Delta SPI_{DJF} + e\Delta CDD + f\Delta CWD + g\Delta R99pTOT + h\Delta Rx1day + i\Delta Rx5day + j\Delta SDII$). Only significant ($P<0.05$) variables are shown. Values shown in brackets are standard error.

Regression coefficient	Region I	Region II	Region III	Whole region
a ($kg \cdot ha^{-1}$)	-	149.4(18.1)	25.1(9.6)	41.6(7.1)

b (kg·ha^{-1})	123.2(13.3)	171.9(14.6)	123.6(9.3)	134.2(6.7)
c (kg·ha^{-1})	133.0(15.3)	112.7(19.1)	147.0(10.9)	133.4(8.0)
d (kg·ha^{-1})	39.4(11.5)	87.8(15.1)	41.7(6.7)	47.2(5.5)
e (kg·ha^{-1}·d^{-1})	-8.8(1.5)	-7.2(1.4)	-8.4(1.1)	-8.5(0.8)
f (kg·ha^{-1}·d^{-1})	-	-	33.9(8.8)	-
g (kg·ha^{-1}·mm^{-1})	-5.6(0.6)	-9.1(0.7)	-5.6(0.5)	-6.3(0.3)
h (kg·ha^{-1}·mm^{-1})	17.6(1.8)	16.8(2.4)	14.5(1.5)	16.7(1.0)
i (kg·ha^{-1}·mm^{-1})	-4.0(0.9)	-2.7(1.2)	-4.4(0.7)	-4.2(0.5)
j (kg·ha^{-1}·mm^{-1})	58.7(12.9)	92.8(13.8)	86.7(10.3)	78.2(6.8)

2.3.4 Periodic variation of SPI indices

According to the above results, the wheat yield variability in the NSW wheat belt could be largely associated with rainfall extremes. Rainfall extreme indices had different impacts on wheat yield. As the yield variability was mainly caused by climate variability, we could infer that severe inter-annual fluctuation existed in these indices, especially in high ranking indices. SPI_{JJA} can be viewed as the most important index evaluated by the two models. Figure 2-4 shows its temporal patterns across the NSW wheat belt. During 1922-2000, there were obvious variations between multi-annual wet and dry periods. These were coherent across the three sub-regions studied. Based on the Jun.-Aug. rainfall condition of the whole wheat belt, we delineated our study period into dry, wet and normal periods. As shown in Figure 2-4, the three kinds of periods appeared alternately with an interval of ~10 years. Temporal patterns of the other three SPI indices are shown in Figure 2-S2.

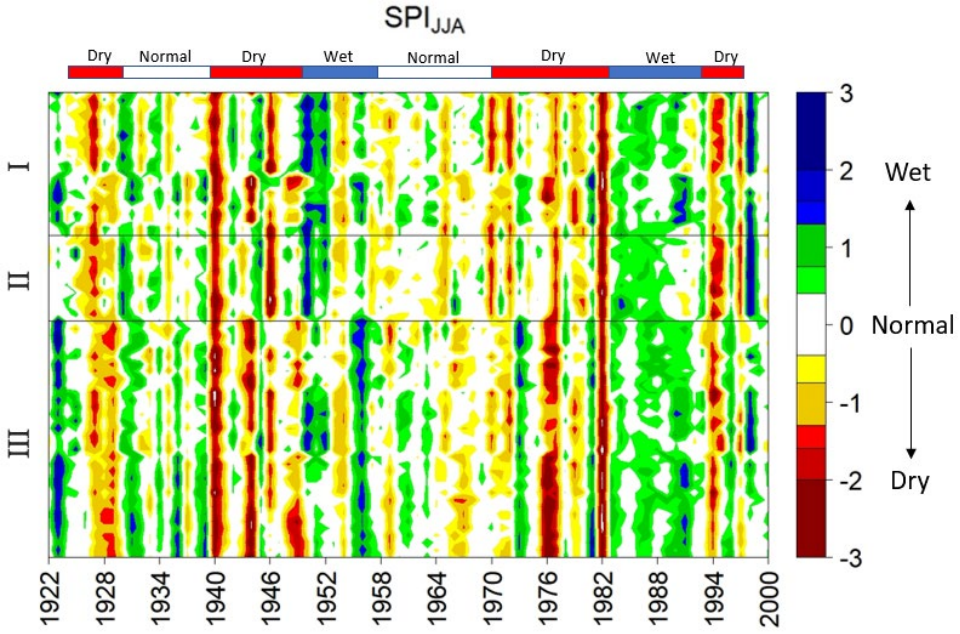


Figure 2-4 Temporal patterns of the standardized precipitation index (SPI) of Jun.-Aug. in three sub-regions of the New South Wales wheat belt. The three sub-regions consist of 20, 12, and 34 shires respectively, so they occupy different widths in the plot.

Figure 2-4 and Figure 2-1 show that there was an obvious and direct correspondence between SPI_{JJA} and wheat yields. Wheat yields decreased seriously around extreme dry years, such as 1940-1946 and 1975-1985. During consecutive years of stable rainfall conditions, yield increased gradually, such as 1985-1990. This again demonstrated the importance of SPI_{JJA} in reflecting wheat yield variability.

2.4 Discussion

Inter-annual variability of crop yields is well known to depend on climate factors (Chen et al., 2004). Rainfall, temperature and solar radiation all can lead to fluctuating crop yields. In the NSW wheat belt, which is a typical semi-arid area (Nicholls, 2004), inter-annual dry-wet alternate conditions can be viewed as the major cause of crop yield variability (Potgieter et al., 2013). In our study, the SPI turned out to be a powerful index in tracking wetness/dryness, considering its high performance in the two models (Figure 2-2 and 2-3). We obtained higher fitness (R^2) between observed and predicted yield when considering the SPI as a covariate in the model compared to a previous study (Wang et al., 2015) conducted in the same area. That study used absolute rainfall amount rather than the SPI. Murphy et al. (2001) also obtained a higher significant coefficient of correlation ($R^2 = 0.78$) by correlating the SPI values with the productivity of a corn crop in Argentina. Overall, it is clear that the SPI is superior in its ability to quantify rainfall conditions and reflect rainfall impacts on crops, compared to the absolute amount of rainfall.

Previous studies have emphasized the importance of seasonal rainfall variability on affecting crop growth (Meinke and Stone, 2005), because crops can be subjected to different degrees of damage from extremes of rainfall during different growth stages. Our study has shown that seasonal SPIs can represent the influence of inter-annual variability in rainfall extremes on the NSW wheat yields. Our results showed that four 3-month timescale SPIs, which quantify rainfall conditions in four seasons, had high importance rankings in the different regions of the NSW wheat belt (Figure 2-3). For the whole NSW wheat belt, SPI_{JJA} and SPI_{SON} were recognized as the two most important indices by the RF and MLR models. SPI_{JJA} can reflect the rainfall condition during the wheat over-wintering stage. Specifically, winter rainfall plays a key role in boosting wheat growth for the rain-fed environment across the NSW wheat belt. Cornish (1950) pointed out that, in Australia, rainfall in the winter months is more effective in producing high wheat yields than rainfall at other times of the year. Stephens and Lyons (1998) also highlighted the increasingly positive effect of winter rainfall on wheat yield for drier regions in Australia. The SPI_{JJA} which had high rankings in our study again proved the importance of winter rainfall on wheat yield in

Australia. SPI_{SON} reflects rainfall in September-November when booting and heading-flowering stages occur. Rainfall shortage during this period can enhance leaf senescence and reduce photosynthesis, thus reducing final grain yield (Yang et al., 2001). However, in region II, SPI_{SON} ranked relatively low, whereas SPI_{DJF} had a higher ranking. This is mainly because region II had a summer-dominant rainfall pattern (Figure 2-S1). Summer rainfall can contribute to stored soil water that can be used for water uptake of wheat during the growing season (Angus et al., 1980). In addition, soils in region II have high clay content, which means high water holding capacity (Reynolds et al., 2000).

The SPI is a widely-used drought index (Ahmadalipour et al., 2017). It is well known that the El Niño Southern Oscillation (ENSO) has the strongest influence on the occurrence of drought in Australia (Watterson, 2009). According to historical records, significant El Niño events happen roughly every 10 years, such as 1963-64, 1972-73, 1982-83 and 1994-95 (Yu and Zou, 2013), which were also years of substantial drought in Australia (Nicholls, 2004). The SPI_{JJA} (Figure 2-4) corresponded well to historical drought records, meaning that the index reflected the ENSO-induced rainfall variability in Australia. Previous studies have used ENSO to predict long-term wheat yields (Rimmington and Nicholls, 1993). Here, we suggest the 3-month timescale SPI for its short-term prediction ability. Based on its value before harvesting, the wheat yield can be estimated. The SPI_{JJA} has a direct correspondence with the wheat yield in any given year. In addition, adaptation measures can be worked out in advance. For example, in region II, if the SPI_{DJF} reaches a low level, a drought-resistant wheat cultivar, or even fallow, could be adopted.

While for other rainfall-related extreme indices, their effects on yield variations should also be highly-regarded. Those indices can be divided into two categories, i.e. extreme dryness and extreme wetness indices. Both extreme dryness and extreme wetness can affect crop growth via negative effects on plant physiological processes and direct physical damage, as well as by affecting the timing and conditions of field operations (Van der Velde et al., 2012). Wijeratne (1996) found that tea yield is sensitive to water shortage and heavy rainfall. An increase in the frequency of extreme rainfall events could result in a decrease in tea yield. Van der Velde et al. (2012) reported low wheat yields in France in 2003 and 2007 caused by extremely dry and extremely wet conditions, respectively. In our results, rainfall-related extreme indices ranked differently and some of them, such as CDD and SDII, had high rankings (Figure 2-3). Nevertheless, most heavy rainfall indices had low importance rankings. Adverse effects of heavy rainfall are often related to waterlogging, which could severely limit crop growth. However, heavy rainfall events are rare in a dry climate of the study area, especially during wheat growing seasons (Gallant et al., 2007). This explains the reason for the effects of heavy rainfall on yield variability in our analysis being relatively low. Therefore, in general, wheat cropping in NSW suffers more

harm from dry than from wet conditions. Adaptive measures should focus on drought resistance, such as drought-tolerant varieties and improvement of soil water retention capacity.

The comparison between the results obtained from the MLR and RF models showed that the RF model is better at predicting wheat yield variation based on rainfall extremes. The RF model outperformed the MLR model in all 3 regions based on the 4 validation measurements used. The importance rankings of the different indices differed to some extent between the two models (Figure 2-3). In the RF model, there was just a small difference (around 8%) between the highest and the lowest importance values of the different indices. However, differences are much larger for the MLR model (almost 20%). The MLR model tended to distribute higher importance values (~15%) to 2-3 indices and largely disregarded the remaining indices. This could be because multicollinearity might still exist, and one index might mask the contribution from another (Chatterjee and Hadi, 2012). Since the RF is a nonlinear algorithm, its result is not affected by multicollinearity. So, the MLR results should be interpreted with great caution. However, this does not affect the capacity of the MLR model to obtain a good fit with regression, or the quality of predictions from the regression (Dielman, 2005). Furthermore, compared to the RF, the MLR model can also quantitatively estimate variable contribution (Table 2-2). Consequently, both the two methods have their own strengths in exploring climate-yield relationships.

Our study obtained greater accuracy in analyzing wheat yield variation using rainfall extreme indices through the RF model than through MLR. However, it should be noted that, compared to MLR, the RF model generates only the importance order of variables, but cannot be used for quantifying factor contributions, which limits its use in further analysis. In addition, this study only focused on rainfall extremes. Other climate extremes (Trnka et al., 2014), such as frost (Frederiks et al., 2015; Fuller et al., 2007) and heat stress (Asseng et al., 2015), can also contribute to variability in crop yields. Studies taking temperature and radiation into consideration will be done in the future.

2.5 Conclusion

This study provided the first look at the impacts of various facets of rainfall extremes on wheat yields in eastern Australia using both machine learning and multiple linear regression methods. The introduction of machine learning significantly improved the precision of climate-yield relationship diagnostics, overcame the shortcomings of linear model in dealing with correlated predictors, revealed new insights of different effects of similar climate factors on crop yields. In addition, the comparison between both machine learning and linear model ensured the robustness of our results.

Our results showed that rainfall extremes were dominant factors affecting wheat yield variation and could explain more than half yield variability in eastern Australia. In the eastern slopes and southern plains, growing season rainfall and consecutive dry days were major factors causing yield variation. By contrast, in the northern plains, pre-growing season rainfall was included as one of the most important factors. Overall, wheat yield variability in the study area was mainly caused by frequent water shortage, while extreme wetness within growing season had a small effect as it occurred less frequently. This is a relevant result because Eastern Australia is an El Niño sensitive region and each El Niño induced water shortage will adversely affect wheat yield.

2.6 Supporting information

Table 2-S1 Multicollinearity diagnostics of input variables.

VIF	Region I	Region II	Region III	Whole region
SPI_{MAM}	1.34	1.34	1.68	1.43
SPI_{JJA}	1.67	1.68	1.96	1.75
SPI_{SON}	1.97	2.08	2.09	1.98
SPI_{DJF}	1.08	1.09	1.10	1.08
CDD	1.47	1.62	1.89	1.64
CWD	1.57	1.94	1.89	1.72
R99pTOT	4.92	5.13	4.80	4.72
Rx1day	5.17	7.33	5.67	5.55
Rx5day	3.90	5.76	3.90	4.10
SDII	3.20	3.91	3.34	3.24

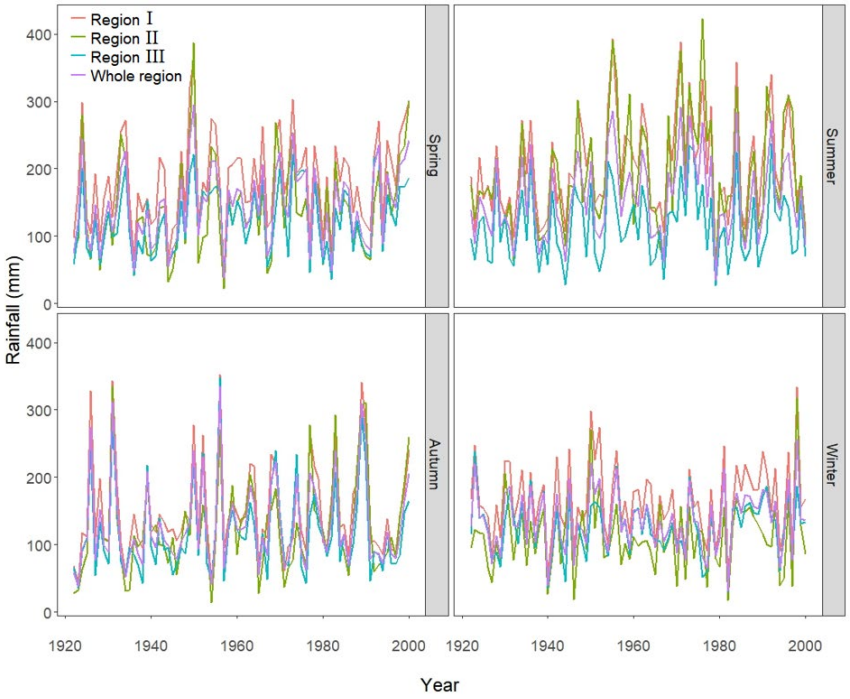
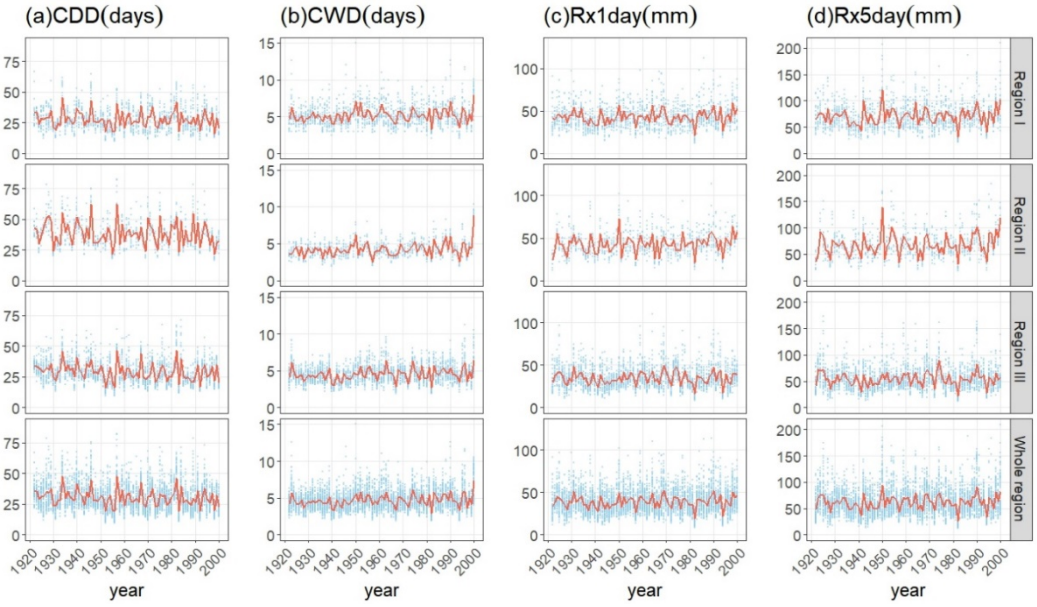


Figure 2-S1 Seasonal rainfall in 3 sub-regions and the whole New South Wales wheat belt during 1922-2000.



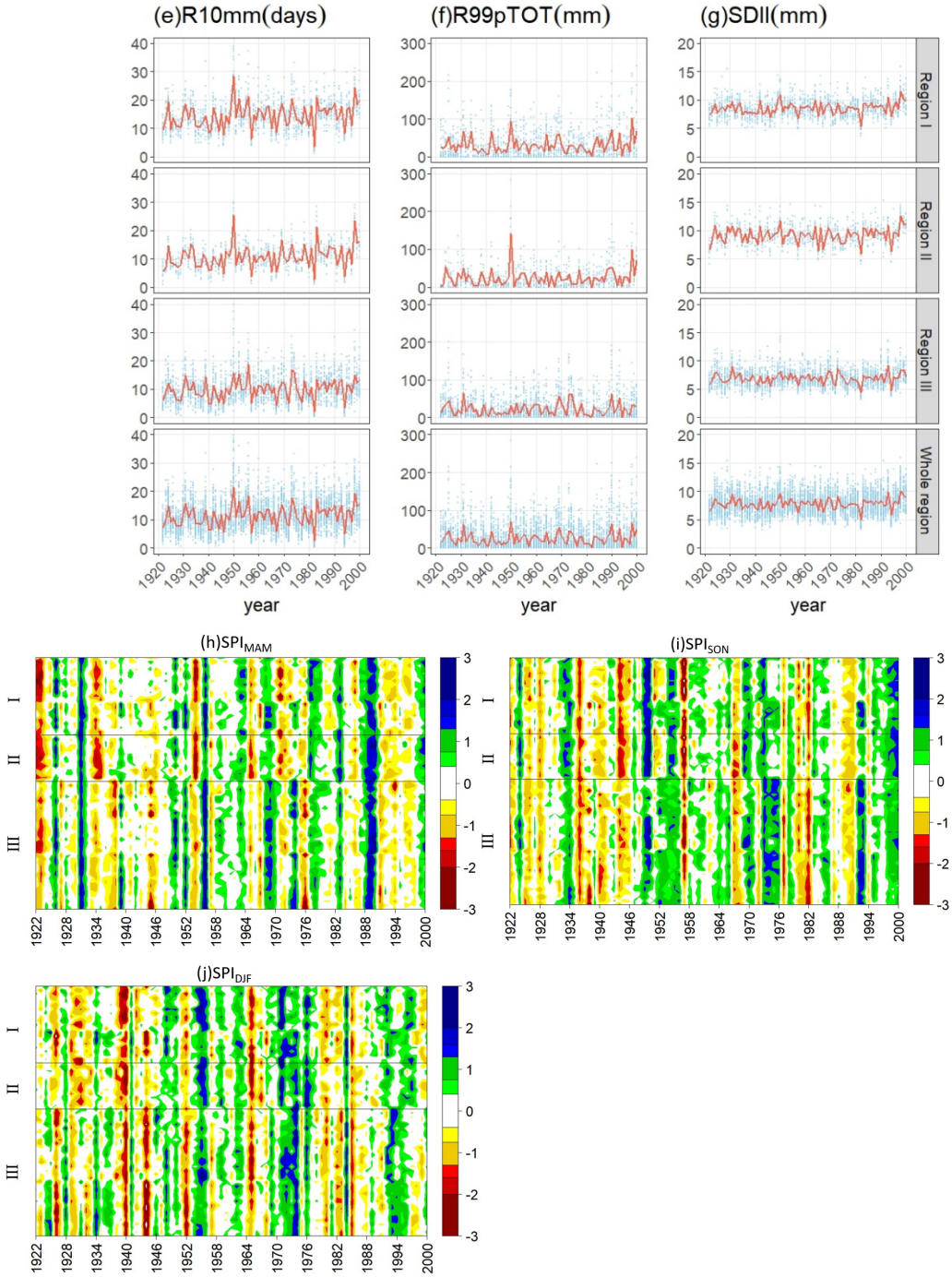


Figure 2-S2 Temporal trends of rainfall extreme indices in three sub-regions and the whole New South Wales wheat belt during 1922-2000. The three sub-regions consist of 20, 12, and 34 shires respectively, so they occupy different widths in (h-j). In (a-g), blue points represent values for each shire and red lines are shire-averaged values for each region.

2.7 References

Ahmadalipour, A., Moradkhani, H. and Svoboda, M., 2017. Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble. International Journal of Climatology, 37(5): 2477-2491.

Alexander, L. and Herold, N., 2015. climpact2: indices and software. Technical report of the World Meteorological Organisation Commission for climatology expert team on sector-specific climate indices. World Meteorological Organisation. https://github.com/ARCCSS-extremes/climpact2.

Angus, J.F., Nix, H.A., Russell, J.S. and Kruizinga, J.E., 1980. Water use, growth and yield of wheat in a subtropical environment. Australian Journal of Agricultural Research, 31(5): 873-886.

Asseng, S. and Pannell, D.J., 2013. Adapting dryland agriculture to climate change: Farming implications and research and development needs in Western Australia. Climatic Change, 118(2): 167-181.

Breiman, L., 2001. Random Forest. Machine Learning, 45: 5-32.

Chatterjee, S. and Hadi, A.S., 2012. Regression Analysis by Example. 5th Edition. John Wiley & Sons Inc.

Chen, C. et al., 2010. Modelling the effects of climate variability and water management on crop water productivity and water balance in the North China Plain. Agr Water Manage, 97(8): 1175-1184.

Chen, C.C., Mccarl, B.A. and Schimmelpfennig, D.E., 2004. Yield Variability as Influenced by Climate: A Statistical Investigation. Climatic Change, 66(1): 239-261.

Cornish, E.A., 1950. The Influence of Rainfall on the Yield of Wheat in South Australia. Australian Journal of Biological Sciences, 3(2): 178-218.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A. and Hess, K.T., 2007. Random forests for classification in ecology. Ecology, 88(11): 2783-2792.

Dielman, T.E., 2005. Applied regression analysis : a second course in business and economic statistics. Brooks/Cole Thomson Learning, 367 pp.

Everingham, Y., Sexton, J. and Robson, A., 2015. A statistical approach for identifying important climatic influences on sugarcane yields. In: Proceedings of the 37th Annual Conference of the Australian Society of Sugar Cane Technologists, 37(37): 8-15.

Farrar, D.E. and Glauber, R.R., 1967. Multicollinearity in Regression Analysis: The Problem Revisited. Review of Economics & Statistics, 49(1): 92-107.

Fitzsimmons, R., 2001. Winter cereal production statistics, NSW 1922−1999: wheat, oats, barley: area production and yield: NSW by local government areas, individual years plus 5 and 10 year averages, 6th edn. Australian Institute of Agricultural Science and Technology,Wahroonga.

Gichangi, E.M. et al., 2015. Assessment of climate variability and change in semi-arid eastern Kenya. Climatic Change, 130(2): 287-297.

Gromping, U., 2006. Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software, 17(1): 925-933.

Gromping, U., 2009. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. Am Stat, 63(4): 308-319.

Guttman, N.B., 1999. Accepting the standardized precipitation index: A calculation algorithm. J Am Water Resour As, 35(2): 311-322.

Huang, J., Zhang, F., Xue, Y. and Li, Q., 2016. Recent changes of extreme dryness/wetness pattern and its possible impact on rice productivity in Jiangsu Province, southeast China. Natural Hazards Journal of the International Society for the Prevention & Mitigation of Natural Hazards, 84(3): 1-13.

Ingram, J., 2011. A food systems approach to researching food security and its interactions with global environmental change. Food Secur, 3(4): 417-431.

IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental Panel on Climate Change. Journal of Clinical Endocrinology & Metabolism, 18(6): 586-599.

Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ Modell Softw, 16(4): 309-330.

Jeong, J.H. et al., 2016. Random Forests for Global and Regional Crop Yield Predictions. Plos One, 11(6): e0156571.

King, A.D. et al., 2014. Extreme Rainfall Variability in Australia: Patterns, Drivers, and Predictability. Journal of Climate, 27(15): 6035-6050.

Lin, L.I., 1989. A Concordance Correlation-Coefficient to Evaluate Reproducibility. Biometrics, 45(1): 255-268.

Liu, Y.A., Wang, E.L., Yang, X.G. and Wang, J., 2010. Contributions of climatic and crop varietal changes to crop production in the North China Plain, since 1980s. Global Change Biology, 16(8): 2287-2299.

Lobell, D., 2010. Crop Responses to Climate: Time-Series Models. Springer Netherlands, 85-98 pp.

Lobell, D.B. and Asner, G.P., 2003. Climate and management contributions to recent trends in U.S. agricultural yields. Science, 299(5609): 1032.

Lobell, D.B. and Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agricultural & Forest Meteorology, 150(11): 1443-1452.

Lobell, D.B. and Field, C.B., 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environmental Research Letters, 2(1): 625-630.

Mckee, T.B., Doesken, N.J. and Kleist, J., 1993. The Relationship of Drought Frequency and Duration to Time Scales.

Meinke, H. and Stone, R., 2005. Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Climatic Change, 70(1-2): 221-253.

Michel, L. and Makowski, D., 2013. Comparison of Statistical Models for Analyzing Wheat Yield Time Series. Plos One, 8(10).

Mohammad, B., Sarah, S., Amin, A., Ssadeghi, L. and Azadeh, M., 2010. Association between climate indices, aridity index, and rainfed crop yield in northeast of Iran. Field Crops Research, 118(2): 105-114.

Murphy, G., Hurtado, R., Fernandez Long, M. and Spescha, L., 2001. The standard precipitation index for the yield of corn in the junin region (Provincia de Buenos Aires), Anais XII Congresso Brasileiro de Agrometeorologia e III Reunião Latino Americana de Agrometeorologia, pp. 15-17.

Nicholls, N., 2004. The changing nature of Australian droughts. Climatic Change, 63(3): 323-336.

O'Meagher, B., du Pisani, L.G. and White, D.H., 1998. Evolution of drought policy and related science in Australia and South Africa. Agricultural Systems, 57(3): 231-258.

Osborne, T., Rose, G. and Wheeler, T., 2013. Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. Agricultural & Forest Meteorology, 170(3): 183-194.

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(3): 103–114.

Potgieter, A. et al., 2013. Spatial impact of projected changes in rainfall and temperature on wheat yields in Australia. Climatic Change, 117(1-2): 163-179.

Potgieter, A.B., Hammer, G.L. and Butler, D., 2002. Spatial and temporal patterns in Australian wheat yield and their relationship with ENSO. Australian Journal of Agricultural Research, 53(1): 77-89.

Potgieter, A.B. et al., 2016. Yield trends under varying environmental conditions for sorghum and wheat across Australia. Agr Forest Meteorol, 228: 276-285.

Prasad, A.M., Iverson, L.R. and Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems, 9(2): 181-199.

Ray, D.K., Gerber, J.S., Macdonald, G.K. and West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nature Communications, 6(5989): 5989-5989.

Raziei, T., Saghafian, B., Paulo, A.A., Pereira, L.S. and Bordi, I., 2009. Spatial Patterns and Temporal Variability of Drought in Western Iran. Water Resources Management, 23(3): 439-455.

Richards, R.A., Hunt, J.R., Kirkegaard, J.A. and Passioura, J.B., 2014. Yield improvement and adaptation of wheat to water-limited environments in Australia-a case study. Crop & Pasture Science, 65(7): 676-689.

Rimmington, G.M. and Nicholls, N., 1993. Forecasting Wheat Yields in Australia with the Southern Oscillation Index. Australian Journal of Agricultural Research, 44(4): 625-632.

Rosenzweig, C. et al., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proceedings of the National Academy of Sciences of the United States of America, 111(9): 3268-73.

Rotter, R.P., Carter, T.R., Olesen, J.E. and Porter, J.R., 2011. Crop-climate models need an overhaul. Nature Climate Change, 1(4): 175-177.

Rowhani, P., Lobell, D.B., Linderman, M. and Ramankutty, N., 2011. Climate variability and crop production in Tanzania. Agricultural & Forest Meteorology, 151(4): 449-460.

Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.

Sheehy, J.E., Mitchell, P.L. and Ferrer, A.B., 2006. Decline in rice grain yields with temperature: Models and correlations can give different estimates. Field Crops Research, 98(2): 151-156.

Stagge, J.H., Tallaksen, L.M., Gudmundsson, L., Loon, A.F.V. and Stahl, K., 2015. Candidate Distributions for Climatological Drought Indices (SPI and SPEI). International Journal of Climatology, 35(13): 4027–4040.

Stephens, D.J. and Lyons, T.J., 1998. Rainfall-yield relationships across the Australian wheatbelt. Australian Journal of Agricultural Research, 49(2): 211-223.

Takahashi, K., Montecinos, A., Goubanova, K. and Dewitte, B., 2011. ENSO regimes: Reinterpreting the canonical and Modoki El Nino. Geophysical Research Letters, 38.

Tebaldi, C. and Lobell, D.B., 2008. Towards probabilistic projections of climate change impacts on global crop yields. Geophysical Research Letters, 35(8).

Traore, B., Corbeels, M., Wijk, M.T.V., Rufino, M.C. and Giller, K.E., 2013. Effects of climate variability and climate change on crop production in southern Mali. European Journal of Agronomy, 49(49): 115-125.

Van der Velde, M., Tubiello, F.N., Vrieling, A. and Bouraoui, F., 2012. Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. Climatic Change, 113(3-4): 751-765.

Wang, B. et al., 2015. Effects of climate trends and variability on wheat yields in eastern Australia. Climate Research, 64(2).

Wang, H. et al., 2016. Monitoring winter wheat drought threat in Northern China using multiple climate-based drought indices and soil moisture during 2000–2013. Agricultural & Forest Meteorology, s 228–229: 1-12.

Watterson, I., G, 2009. Components of precipitation and temperature anomalies and change associated with modes of the Southern Hemisphere. International Journal of Climatology, 29(6): 809-826.

Were, K., Bui, D.T., Dick, O.B. and Singh, B.R., 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. Ecol Indic, 52: 394-403.

Wijeratne, M.A., 1996. Vulnerability of Sri Lanka tea production to global climate change. Water Air Soil Poll, 92(1-2): 87-94.

Williams, G.J., 2011. Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery.

Yang, J.C., Zhang, J.H., Wang, Z.Q., Zhu, Q.S. and Wang, W., 2001. Remobilization of carbon reserves in response to water deficit during grain filling of rice. Field Crops Research, 71(1): 47-55.

Yu, J.Y. and Zou, Y.H., 2013. The enhanced drying effect of Central-Pacific El Nino on US winter. Environmental Research Letters, 8(1).

Yuan, L., Wang, E., Yang, X. and Jing, W., 2010. Contributions of climatic and crop varietal changes to crop production in the North China Plain, since 1980s. Global Change Biology, 16(8): 2287-2299.

Chapter 3. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia

This chapter is based on the following manuscript:

Feng, P., Wang, B., Liu, D.L. and Yu, Q., 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, pp.303-316.

Highlights

• Remotely-sensed drought factors were used to estimate agricultural drought through three machine learning models.

• The bias-correct random forest model outperformed the other two models in monitoring agricultural drought.

• The importance of various remotely-sensed factors was examined for different climate regions.

• Model-predicted drought distribution maps showed strong visual and statistical agreement with station-based drought maps.

Abstract

Agricultural drought is a natural hazard arising from insufficient crop water supply. Many drought indices have been developed to characterize agricultural drought, relying on either ground-based climate data or various remotely-sensed drought proxies. Ground-based drought indices are more accurate but limited in coverage, while remote sensing drought indices cover large areas but have poor precision. Application of advanced data fusion approaches based on remotely-sensed data to estimate ground-based drought indices may help fill this gap. The overall objective of this study was to determine whether various remotely-sensed drought factors could be effectively used for monitoring agricultural drought in south-eastern Australia. In this study, thirty remotely-sensed drought factors from the Tropical Rainfall Measuring Mission (TRMM) and the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensors were used to reproduce a ground-based drought index, SPEI (Standardized Precipitation Evapotranspiration Index) during 2001-2017 for the New South Wales wheat belt in south-eastern Australia. Three advanced machine learning methods, i.e. bias-corrected random forest, support vector machine, and multi-layer perceptron neural network, were adopted as the regression models in this procedure. A station-based historical climate dataset and observed wheat yields were used as reference data to

evaluate the performance of the model-predicted SPEI in reflecting agricultural drought. Results show that the bias-corrected random forest model outperformed the other two models for SPEI prediction, as quantified by the lowest root mean square error (RMSE) and the highest R^2 values (<0.28 and ~0.9, respectively). Drought distribution maps produced by the bias-corrected random forest model were then compared with the station-based drought maps, showing strong visual and statistical agreement. Furthermore, the model-predicted SPEI values were more highly correlated with observed wheat yields than the station-based SPEI. The method used in this study is effective and fast, and based on data that are readily available. It can be easily extended to other cropping areas to produce a rapid overview of drought conditions and to enhance the present capabilities of real-time drought monitoring.

Keywords: agricultural drought; Standardized Precipitation Evapotranspiration Index; remote sensing; machine learning

3.1 Introduction

Drought, one of the most complex and devasting natural disasters, is a recurring feature of nearly every climatic zone (Zarch et al., 2015). It can be broadly classified into four common types, i.e. meteorological drought, agricultural drought, hydrological drought, and socio-economic drought (Wilhite, 2005). These types of drought are interrelated. Assessing agricultural drought is of great importance as it is viewed as the most serious problem in most countries in terms of food security, economy, and social stability (Hazaymeh and Hassan, 2016; He et al., 2013; Mottaleb et al., 2015).

Agricultural drought is a period of soil water deficit that results from below normal precipitation, and/or above-average evaporation and transpiration (Dai, 2011; Quiring and Papakryiakou, 2003). It is a region- and period-specific disaster (Mishra and Singh, 2010). Therefore, identifying the onset and termination of an agricultural drought event is normally difficult in a given region. Moreover, the severity and likely temporal and spatial variations can vary significantly by periods and regions (Touma et al., 2015). Agricultural drought, therefore, is a major threat to crop production, causing considerable agricultural losses around the world (Piao et al., 2010; Wang et al., 2018b). For example, during the 2006 drought in Australia, winter cereal crop production decreased by 36%, causing about AUD$3.5 billion economic loss (Wong et al., 2009). Therefore, accurate identification of agricultural drought, especially real-time or near real-time drought monitoring, are urgently needed to provide essential information for decision makers to institute timely response actions to reduce the negative impacts of drought.

The most important method of monitoring and analysing agricultural drought is applying appropriate drought indices. Generally, a drought index is a prime variable for determining the duration and intensity of drought and for assessing the impacts of drought. However, due to the

complexity of drought definitions, it is difficult to evaluate drought characteristics comprehensively and systematically in a given region through only a single index. Thus, over 160 drought indices have been developed by researchers (Niemeyer, 2008). These indices can be classified into two categories: ground-based and remotely-sensed indices. The ground-based drought indices are usually derived from ground measurements of meteorological variables such as precipitation and temperature. They are able to accurately and effectively monitor drought around climate stations (Rhee et al., 2010). Some of the examples include the Palmer Drought Severity Index (Palmer, 1965), Crop Drought Identification Index (Wu et al., 2018), the standardized precipitation index (SPI) (Mckee et al., 1993), and the SPEI (Vicente-Serrano et al., 2010). Among them, the SPEI, which accounts for both precipitation and temperature, is capable of monitoring different types of drought in various regions. It has been widely accepted among researchers (Gao et al., 2017; Wang et al., 2015). However, these ground-based indices are station-based and limited in characterizing detailed spatial distributions of drought at regional scales (Park et al., 2016). Although the advanced geographic information system (GIS)-based spatial interpolation techniques (e.g., inverse distance and kriging) may help estimate the drought conditions at unmeasured locations, uncertainties may exist in interpolated areas due to the interpolation algorithms used and complex topographic conditions (Swain et al., 2011).

In order to address the spatial extent of ground-based drought indices, remotely-sensed indices have been developed based on satellite data. Satellite remote sensing data include precipitation, temperature, evapotranspiration, and vegetation information which are all of continuous coverage and real-time. Remotely-sensed drought indices are therefore able to capture detailed spatial characteristics of drought. In recent decades, many remotely-sensed drought indices have been developed, such as the Normalized Difference Vegetation Index (NDVI) (Rouse Jr. et al., 1974), the Normalized Difference Drought Index (NDDI) (Gu et al., 2007), and the Normalized Multiband Drought Index (NMDI) (Wang and Qu, 2007). Although these indices overcome the deficiencies of conventional ground-based drought indices in spatial coverage, they cannot completely replace ground-based observations in drought monitoring because of their relatively short data records (Lai et al., 2019). Furthermore, the quality of remotely-sensed drought indices can be limited by atmospheric conditions (e.g., clouds) and retrieval algorithms (Zhang and Jia, 2013). Therefore, the accuracy and reliability of remotely-sensed drought indices are still in question (Alizadeh and Nikoo, 2018).

In agricultural drought monitoring, ground-based drought indices are usually considered to be more accurate than remotely-sensed drought indices. This is mainly because climate variables derived from ground measurements are normally more accurate. Furthermore, most ground-based drought indices directly monitor the soil water balance rather than indirect vegetative information. Many newly developed remotely-sensed drought indices usually need to be validated using

ground-based indices to prove their reliability (Rhee et al., 2010). However, little agreement is typically found when common remotely-sensed drought indices are compared to ground-based drought indices (Bayarjargal et al., 2006). For example, NDVI-derived VCI (Vegetation Condition Index, Kogan (1995)) and ground-based drought indices were poorly correlated for counties in east Texas, USA, with coefficients of determination (R^2) around 0.1 (Quiring and Ganesh, 2010).

To better monitor drought, researchers have attempted to fuse data from various sources to reproduce ground-based drought indices using data-driven models, such as artificial neural networks (Morid et al., 2007) and autoregressive integrated moving average models (Belayneh et al., 2014). However, these models are usually limited in dealing with nonlinearities or non-stationarities in drought estimations. With the development of artificial intelligence, more advanced and adaptive machine learning methods are gaining recognition. Machine learning methods are capable of disentangling the effects of co-linear variables and analysing hierarchical and nonlinear relationships between the independent variables and the dependent variable, and usually result in better performance compared with conventional linear regression models (Belayneh et al., 2014; Guzmán et al., 2017). Park et al. (2016) applied three machine learning approaches to fuse sixteen remotely-sensed drought factors targeting a meteorological drought index (SPI). Results showed that machine learning approaches were able to capture more than 90% of the SPI variation in the southern USA. Alizadeh and Nikoo (2018) obtained similar results in Iran using five individual machine learning methods. Fusing remote sensing data to reproduce ground-based drought indices can extend point-based indices to an entire region, promoting a deeper understanding of the spatial characteristics of a drought event. However, to our knowledge, few studies have used similar methods to monitor drought in the drought-vulnerable continent of Australia.

Drought is a recurring event in Australia, with the current 2018 drought now having lasted for months. This severe drought has affected most of south-eastern Australia, which may reduce winter wheat production by almost 45% (https://www.dpi.nsw.gov.au/about-us/publications/pdi/2018/wheat). Regional and timely drought monitoring is therefore urgently needed to develop appropriate strategies to minimise damage. Therefore, the overall objective of this study was to determine whether various remotely-sensed drought factors could be effectively used to characterize agricultural drought in south-eastern Australia. The specific objectives of this study were to 1) compare the performance of three machine learning methods, i.e., bias-corrected random forest (BRF), multi-layer perceptron neural network (MLP), and support vector machine (SVM), in reproducing ground-based SPEI based on various remotely-sensed drought factors in dry and wet areas; 2) identify the relative importance of remotely-sensed drought factors in

determining SPEI; 3) compare model-predicted drought maps with corresponding ground-based drought maps.

3.2 Materials and methods

3.2.1 Study area

The study area was the New South Wales (NSW) wheat belt (Figure 3-1) in south-eastern Australia, with its western border bounded by the arid interior of Australia. It covers an area of ~360 000 km^2, accounting for about 28% of the wheat-planted area in Australia (www.abares.gov.au, 2013-14). Topography and climatic characteristics vary widely among different parts of the wheat belt. Topographically, the western part of the wheat belt is mostly plains with elevation lower than 300 m, while the eastern part is occupied by mountains of up to 1100 m. Climatically, the wheat belt is characterized by warm and dry conditions in the western part and cold and wet conditions in the eastern part (Feng et al., 2018). Average annual precipitation ranges from 200 mm in the southwest to 1000 mm in the southeast, while average temperature ranges from 11 °C in the southeast to 20 °C in the northwest.

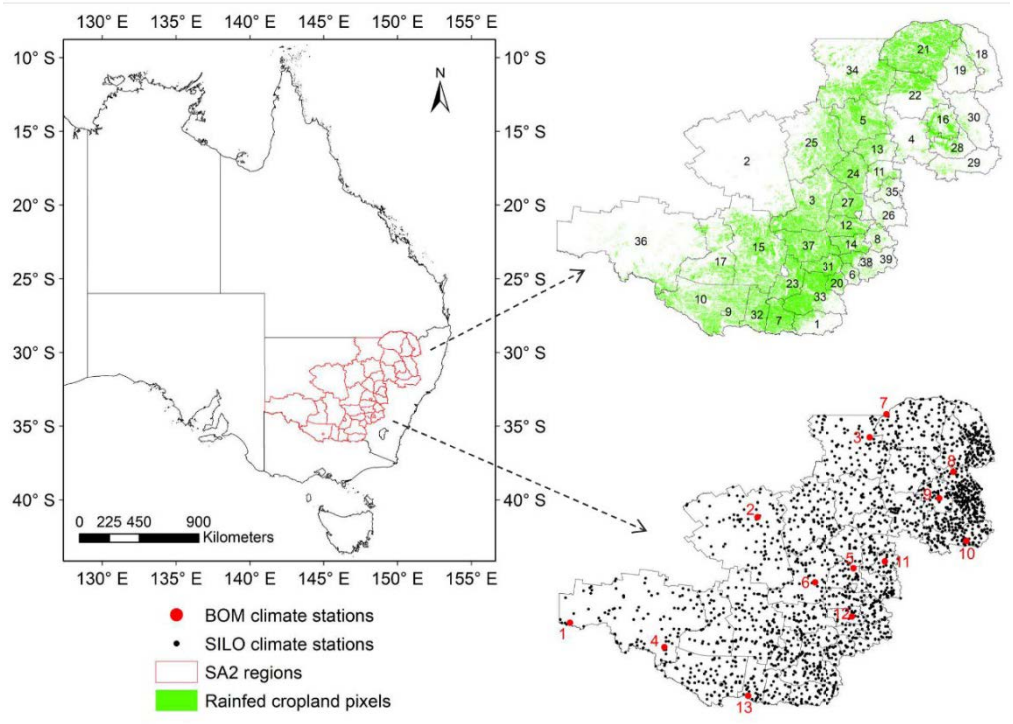


Figure 3-1 The study area consisting of 39 Statistical Areas Level 2 (SA2) regions in New South Wales in south-eastern Australia, which are clipped from a yield gap map (http://yieldgapaustralia.com.au/). The green areas are the gridded rainfed cropland derived from DLCD (Dynamic Land Cover Dataset, http://www.ga.gov.au/) land cover map. Black points and

red points indicate the locations of 2242 SILO weather stations and 13 BOM weather stations, respectively.

In order to make our models more precise regarding agricultural drought, we separated out the rainfed cropland from other land types. The DLCD (Dynamic Land Cover Dataset, http://www.ga.gov.au/) land cover map was used to distinguish the rainfed cropland. This map was developed by Geoscience Australia according to an analysis of the enhanced vegetation index (EVI) from MODIS products during 2000-2008 (Lymburner et al., 2010). It is highly consistent (93%) with independent field-based observations. This map was originally at 250 m resolution, but we upscaled the resolution to 500 m in order to reduce computational costs. The green area in Figure 3-1 illustrates the rainfed cropland over the NSW wheat belt. Our subsequent calculations and results visualizations were both performed on these rainfed cropland pixels.

3.2.2 Data

3.2.2.1 Remote sensing data

Surface reflectance, evapotranspiration (ET), potential evapotranspiration (PET), and land surface temperature (LST) data (2001-2017, actually to February 2018 as the first two months of 2018 still belonged to the summer of 2017 in Australia) were acquired from the MODIS satellite sensor. The surface reflectance data (bands 1-7) were extracted from MODIS MOD09GA.006, with 500 m spatial resolution and 1-day temporal resolution. Afterwards, NDVI, EVI, NMDI, NDWI, and NDDI were calculated using the reflectance data (Table 3-1). In particular, the NDWI was calculated using three shortwave infrared bands 5, 6, and 7 separately, resulting in three different NDWIs, i.e. NDWI5, NDWI6, and NDWI7. ET and PET data were extracted from MODIS MOD16A2.006 at 500 m spatial resolution and 8-day temporal resolution. LST data were extracted from MODIS MOD11A1.006 at 1 km spatial resolution and 1-day temporal resolution.

Table 3-1 Remote sensing indices used in the prediction of agricultural drought in the New South Wales wheat belt, Australia.

Indices	Definition	Formula	Resolution	Source
Pre	precipitation	-	0.25°	TRMM/3B43
PET	potential evapotranspiration	-	500 m	MODIS
ET	evapotranspiration	-	500 m	MODIS
LSTd	Land surface temperature-day	-	1 km	MODIS
LSTn	Land surface temperature-night	-	1 km	MODIS
LSTm	Land surface temperature-mean	(LSTd+LSTn)/2	1 km	MODIS
NDVI	Normalized difference vegetation index	(b2–b1)/(b2+b1)	500 m	Rouse Jr et al. (1974)
EVI	Enhanced vegetation index	2.5*((b2-b1)/(b2+6*b1-7.5*b3+1))	500 m	Huete et al. (2002)
NMDI	Normalized multi-band drought index	(b2-(b6-b7))/(b2+(b6-b7))	500 m	Wang and Qu (2007)
NDWI5	Normalized difference water index	(b2-b5)/(b2+b5)	500 m	Gao (1996)
NDWI6	Normalized difference water index	(b2-b6)/(b2+b6)	500 m	Gao (1996)
NDWI7	Normalized difference water index	(b2-b7)/(b2+b7)	500 m	Gao (1996)
NDDI5	Normalized difference drought index	(NDVI-NDWI5)/(NDVI+NDWI5)	500 m	Gu et al. (2007)
NDDI6	Normalized difference drought index	(NDVI-NDWI6)/(NDVI+NDWI6)	500 m	Gu et al. (2007)
NDDI7	Normalized difference drought index	(NDVI-NDWI7)/(NDVI+NDWI7)	500 m	Gu et al. (2007)

Monthly precipitation data from 2001 to 2017 at 0.25-degree resolution were acquired from the 3B43 product of the TRMM satellite sensor. The data unit of measure was mm hr^{-1} but was subsequently converted to mm $month^{-1}$. All of the remote sensing data used in this study were acquired from the Google Earth Engine (GEE, https://earthengine.google.com) platform.

For all of these indices, 1- and 3-month time-scale means were then calculated to evaluate the lagged responses. Since data from different locations were pooled together and used for the same regression function for each region or for each climate division, all remote sensing variable values were scaled using equation (1) for each pixel during 2001-2017 to distinguish weather components from ecosystem components (Kogan, 1995; Rhee et al., 2010). Thus, a total of 30 scaled indices were used as input variables.

$$x_s = \frac{x - \overline{x}}{\sigma} \tag{3-1}$$

where x_s represents the scaled indices, x denotes the actual values of the indices, $\overline{x}$ and σ indicate the mean and the variance of the indices, respectively.

3.2.2.2 Ground-based climate data

The station-based observational climate data for the study area were obtained from the Bureau of Meteorology (BOM, http://www.bom.gov.au). We selected 13 weather stations with full records of daily rainfall and temperature for the past 30 years (1988-2017). These stations are scattered throughout most of the wheat belt (Figure 3-1). The data from these stations were used for model calibration and validation in our study. A brief description of the 13 BOM stations, including location, elevation, annual mean temperature, and annual precipitation, is shown in Table 3-2.

Table 3-2 A brief description of 13 climate observational stations from the Bureau of Meteorology in New South Wales wheat belt, Australia, including location, elevation, annual mean precipitation (AP), and annual mean temperature (AT).

ID	Latitude	Longitude	Elevation (m)	AP (mm)	AT (°C)
1	-34.0	141.3	26	258	17.7
2	-31.5	145.8	260	389	19.4
3	-29.5	148.6	145	532	20.3
4	-34.6	143.6	61	336	17.3
5	-32.7	148.2	285	589	18.4
6	-33.1	147.2	195	460	17.7
7	-29.0	149.0	160	502	20.8
8	-30.4	150.6	500	699	16.6
9	-31.0	150.3	307	640	18.8
10	-32.1	150.9	216	623	17.7
11	-32.6	149.0	305	644	17.1
12	-33.9	148.2	390	630	16.3
13	-35.8	145.6	114	459	16.2

Another Australian station-based climate dataset is available from SILO (Scientific Information for Land Owners, https://silo.longpaddock.qld.gov.au). It contains daily climate data from 1889 to the current time for a variety of climate variables including rainfall, maximum and minimum temperature. This dataset is constructed from the observational data obtained from BOM. As the BOM dataset contains a large number of missing values, SILO estimates these missing values through interpolation methods (Jeffrey et al., 2001). Thus, the SILO dataset can be viewed as a product of estimated climate data for thousands of climate stations across Australia. We eventually selected 2242 SILO climate stations that were identified as being located in the NSW wheat belt (Figure 3-1). These data were used as historical data to compare with our model predictions.

3.2.2.3 Wheat yield data

Crop yield is often used as a proxy variable reflecting agricultural drought conditions (Yagci et al., 2015). As our study area was rainfed cropland, it is likely that crop yield is highly correlated with agricultural drought conditions. Thus, we used region-level wheat yield records to evaluate the model-predicted drought and to explore its potential applications. These yield records were obtained from the Australian yield gap map (http://yieldgapaustralia.com.au/). This map divides the NSW wheat belt into 39 Statistical Areas Level 2 (SA2) regions for yield statistics. Wheat yield records from 2001 to 2014 were available for each of the 39 regions and were used for subsequent data analysis. A brief description of the 39 SA2 regions, including location, annual mean temperature, annual precipitation, and annual mean wheat yield, is shown in Table 3-3.

Table 3-3 A brief description of the 39 Statistical Areas Level 2 (SA2) regions in New South Wales wheat belt, Australia, including location, annual mean precipitation (AP), annual mean temperature (AT), and annual mean wheat yield during 2001-2014.

ID	Region	Latitude	Longitude	AP (mm)	AT (°C)	Yield ($t\ ha^{-1}$)
1	Albury Region	-36.0	146.9	697	15.2	2.31
2	Cobar	-31.5	145.8	394	18.8	1.15
3	Condobolin	-33.1	147.1	441	17.5	1.33
4	Coonabarabran	-31.3	149.3	783	15.8	1.86
5	Coonamble	-31.0	148.6	543	18.8	1.41
6	Cootamundra	-34.6	148.0	642	15.3	2.67
7	Corowa Region	-36.0	146.4	558	15.1	2.09
8	Cowra Region	-33.8	148.7	618	15.9	2.17
9	Deniliquin	-35.5	145.0	398	16.3	1.78
10	Deniliquin Region	-35.3	144.3	369	16.5	1.77
11	Dubbo Region	-32.3	148.6	609	17.3	1.80
12	Forbes	-33.7	148.0	535	16.7	1.63
13	Gilgandra	-31.7	148.7	580	19.1	1.51
14	Grenfell	-33.9	148.1	616	16.2	1.89
15	Griffith Region	-34.3	146.0	427	17.0	1.87
16	Gunnedah Region	-31.0	150.3	645	18.5	2.45

17	Hay	-34.7	144.8	355	17.0	1.88
18	Inverell Region-East	-29.6	151.2	776	16.0	1.94
19	Inverell Region-West	-29.9	150.6	746	18.0	2.25
20	Junee	-34.9	147.6	544	15.6	2.47
21	Moree Region	-29.5	149.8	600	19.3	1.98
22	Narrabri Region	-30.3	149.8	637	18.7	2.14
23	Narrandera	-34.7	146.6	431	16.5	1.91
24	Narromine	-32.3	148.2	609	17.3	1.73
25	Nyngan-Warren	-31.6	147.5	475	18.8	1.37
26	Orange Region	-33.3	149.1	649	16.5	2.01
27	Parkes Region	-33.1	148.2	576	17.2	1.56
28	Quirindi	-31.5	150.7	662	17.3	2.88
29	Scone Region	-32.1	150.9	662	17.3	2.33
30	Tamworth Region	-31.1	150.9	681	17.2	2.17
31	Temora	-34.5	147.5	514	15.5	2.11
32	Tocumwal-Finley-Jerilderie	-35.6	145.6	430	16.2	2.27
33	Wagga Wagga Region	-35.1	147.4	575	15.7	2.15
34	Walgett-Lightning Ridge	-29.7	148.0	468	20.2	1.19
35	Wellington	-32.6	148.9	629	16.5	1.97
36	Wentworth-Balranald Region	-34.1	142.6	287	17.3	1.36
37	West Wyalong	-33.9	147.2	483	16.9	1.66
38	Young	-34.4	148.3	696	14.4	2.40
39	Young Region	-34.4	148.7	715	13.7	2.38

Crop growth is influenced by both climatic and non-climatic factors. As our study was mainly focused on agricultural drought, yield variations caused by non-climatic factors, e.g., breeding and farming practice developments, should be excluded. A first-difference approach introduced by Nicholls (1997) was used in this present study. This approach is able to effectively minimize the impacts of non-climatic factors, enabling the evaluation of climate impacts separately. The approach was implemented according to the following equation:

$$\Delta X_t = X_t - X_{t-1},\ t = 2001, 2002, \ldots, 2014 \quad (3\text{-}2)$$

where ΔX_t represents the first difference of X at year t, X_t and $X_{t\text{-}1}$ represent the values of X at year t and year t-1, respectively.

3.2.3 Modelling methodology

The framework presented in Figure 3-2 illustrates all of the procedures used in this study for agricultural drought estimation based on remote sensing data. Three different advanced machine learning methods were adopted to analyze agricultural drought using remotely-sensed drought factors to estimate the SPEI index. The NSW wheat belt is a large area influenced by different climate patterns. Different remote sensing predictors may have different contributions to the drought index in dry or wet areas. Therefore, at the beginning, the K-means clustering algorithm (Hartigan and Wong, 1979) was utilized to classify BOM climate stations into two coherent clusters (Table 3-4) according to multi-year precipitation conditions. Cluster 1 had average annual rainfall of 419 mm and mean temperature of 18.5 °C, while the respective values for cluster 2

were 638 mm and 17.5 °C. Thus, stations in cluster 2 were relatively more humid than stations in cluster 1. Then for each cluster, three different machine learning models, namely BRF, SVM, and MLP, were developed for agricultural drought estimation using remotely-sensed drought factors and the best model was determined. In the second phase, the NSW cropland grids were divided into two clusters based on remotely-sensed precipitation data and the K-means algorithm developed in the first phase. Then the NSW cropland drought maps were created using the best model for each cluster. The performance of the model-predicted drought maps was evaluated through the agreement with the SILO SPEI drought maps and observed wheat yields. The major parts of the proposed procedures are given in detail below.

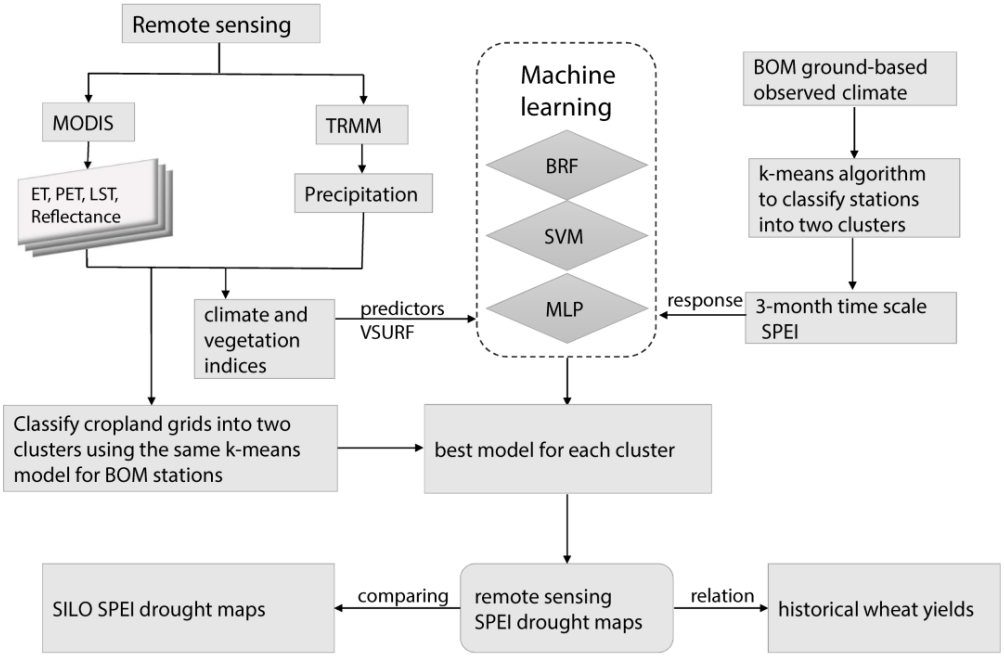


Figure 3-2 Framework of the procedures used in this study, where MODIS: Moderate Resolution Imaging Spectroradiometer, TRMM: Tropical Rainfall Measuring Mission, ET: evapotranspiration, PET: potential evapotranspiration, LST: land surface temperature, BOM: Bureau of Meteorology, SILO: Scientific Information for Land Owners, SPEI: Standardized Precipitation Evapotranspiration Index, BRF: bias-corrected random forest, SVM: support vector machine, MLP: multi-layer perceptron neural network, VSURF: An R package for variable selection.

Table 3-4 Results of K-means clustering algorithm for 13 climate stations from the Bureau of Meteorology (BOM). The station IDs can be found in Table 3-2. Annual mean precipitation (AP) and annual mean temperature (AT) for each cluster are given.

Clusters based on K-means algorithm	BOM station's ID	AP (mm)	AT (°C)
Cluster 1	1, 2, 3, 4, 6, 7, 13	419	18.5
Cluster 2	5, 8, 9, 10, 11, 12	638	17.5

3.2.3.1 Standardized Precipitation Evapotranspiration Index

All of the climate data from BOM and SILO sites were used to calculate the ground-based SPEI. SPEI was developed by Vicente-Serrano et al. (2010) and has attracted widely attention in drought analysis. Briefly, SPEI characterizes drought through standardizing the difference between precipitation and potential evapotranspiration (P-PET). A value of 0 represents normal conditions, while negative and positive values denote dry and wet conditions, respectively. In general, values <-1 are considered to be drought conditions. In addition, according to specific aims, the SPEI can be calculated on different timescales. 1- to 6-month time scale SPEIs are suitable for meteorological and agricultural drought, while longer time scale SPEIs are suitable for hydrological drought. In the present study, the 3-month time scale SPEI was used in order to analyze agricultural and seasonal drought characteristics. Seasonal drought monitoring is important for guiding agricultural production in the study area. The SPEI was calculated using the "SPEI" package coded in R software and the "Thornthwaite" equation (Thornthwaite, 1948) was adopted when calculating PET.

3.2.3.2 Feature selection

We initially built 30 predictor variables for the drought monitoring models. However, 30 variables produced a large dataset that may be susceptible to increased computation costs and over-fitting problems caused by correlated variables and the "curse of dimensionality" (Wang et al., 2018). The correlation analysis showed that significant relationships existed among the candidate predictors, especially in cluster 2 (Figure 3-3). In this study, we applied the VSURF method (Genuer et al., 2015) to exclude redundant predictors. This method is widely applicable and can efficiently select optimal variables from multiple dimensions of data for regression purposes. This variable selection procedure is based on random forest and includes three steps. In the first step, irrelevant variables are eliminated from the dataset. In the second step, all variables related to the response are selected for interpretation purpose. In the third step, redundant variables in the set of variables selected in the second step are eliminated for prediction purposes. Detailed descriptions of the VSURF method can be found in Genuer et al. (2015). These steps were executed prior to developing drought monitoring models through the "VSURF" package in R software.

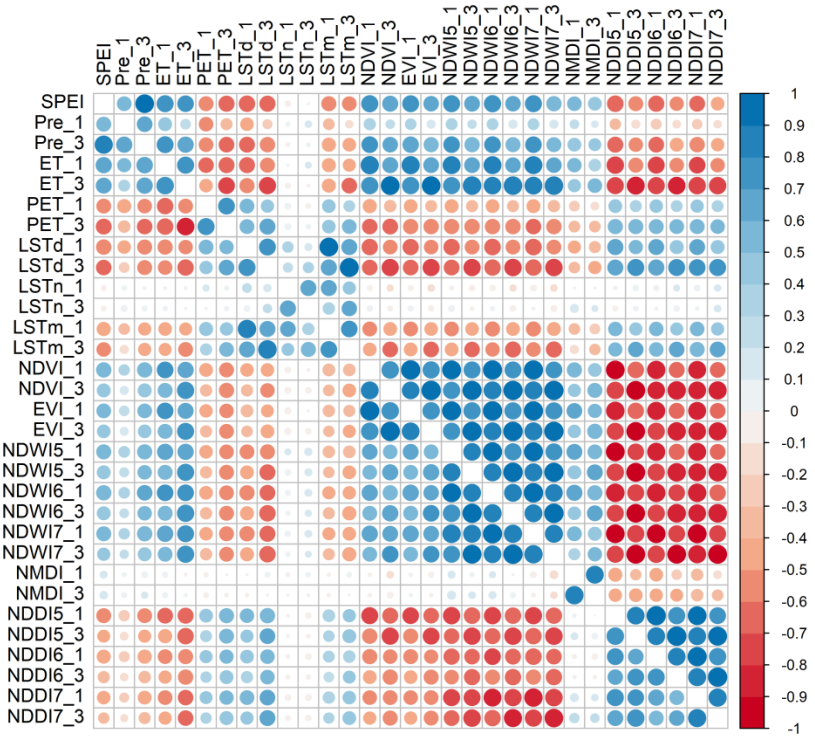


Figure 3-3 Pearson correlation coefficients for the relation between the 3-month time scale SPEI and 30 predictor variables used in this study. The "1" and "3" designations appended to the end of the variable names designate 1-month and 3-month time scale means. The lower and upper triangular parts of the matrix indicate the correlation results for cluster 1 and cluster 2, respectively. Positive correlations are displayed in blue and negative correlations in red colour. Colour intensity and the size of the circle are proportional to the correlation coefficients.

3.2.3.3 Bias-corrected random forest

Random forest (RF) is a popular algorithm for regression and classification purpose (Breiman, 2001). It is a tree-based algorithm and uses de-correlated trees to effectively reduce the variance of predictor variables. The RF algorithm first generates numerous independent trees using bootstrap samples. In the meantime, about 33% of the entire sample is kept out for subsequent out-of-bag validations. For each tree, a randomized subset of the predictor variables is used to determine its results. The average of the results from all of the trees is determined to be the final outcome (Cutler et al., 2007). By doing this, RF can reduce variance, resulting in more accurate predictions in comparison with ordinary tree-based algorithms (Hastie et al., 2009).

The prediction of the RF is basically the average of all trees' results, which is a good feature to avoid producing unreasonable predicted values. However, it may result in bias in dealing with very large or very small observations (Song, 2015). Predictions tend to be overestimated when observations are small, while predictions tend to be underestimated when observations are large. Consequently, bias is not negligible and bias correction is necessary. In this study, we applied a

method developed by Zhang and Lu (2012) to estimate and correct bias in RF for regression. This is a simple and efficient method and the performance with real data is good (Zhang and Lu, 2012). The detailed procedure of this bias-correction method is summarized below.

(1) Fit an RF model with the training dataset, $Y_{train} = \mathrm{RF}(X_{train})$, where X_{train} and Y_{train} represent the independent and dependent variables from the training dataset.

(2) Calculate the predicted values and residuals, $r_{train} = Y_{train} - \hat{Y}_{train}$, where r_{train} represents the residuals and $\hat{Y}_{train}$ represent predicted values.

(3) Fit an RF model with the residuals as the dependent variable and the training dataset as independent variables, $r_{train} = \mathrm{RF}_{res}(X_{train}, Y_{train})$.

(4) Calculate the predicted values (Y_{test}) with a test dataset using the RF model from step 1.

(5) Calculate the estimated residuals using the RF_{res} model with predicted values from step 4 and independent variables in the test dataset, $r_{test} = \mathrm{RF}_{res}(X_{test}, Y_{test})$.

(6) Add the estimated residuals to the predicted values to correct bias, $Y_{bias-correction} = Y_{test} + r_{test}$.

3.2.3.4 Support vector machine

Support vector machine (SVM) is another popular machine learning model with associated learning algorithms for classification and regression purposes (Cortes and Vapnik, 1995). The principle of SVM is to find a hyperplane or multiple hyperplanes to divide a high-dimensional space into multiple different classes. New examples then can be placed into the same space and assigned to belong to one class or another according to which side of the gap they fall. SVM has a better learning capability and usually requires less grid-searching to get a reasonably accurate model. Moreover, SVMs tend to be resistant to over-fitting due to the use of regularization, which aims to choose a low-error but simpler fitting model. Given its powerful classification and regression capacity, SVM has been widely used in image classification and handwriting recognition as well as in the remote sensing field (Melgani and Bruzzone, 2004; Mountrakis et al., 2011). In this study, we implemented the SVM model via the "e1071" package coded in R software. There are four common options of kernel functions in SVM. We selected the Radial Basis Function kernel as it is designed to deal with non-linear problems.

3.2.3.5 Multi-layer perceptron neutral network

A multilayer perceptron (MLP) is a type of feedforward artificial neural network wherein connections between the nodes do not form a cycle (Zell et al., 1994). Generally, there are three layers of nodes in an MLP, namely input layer, hidden layer, and output layer. Both of the last

two nodes are neurons that use a nonlinear activation function. Meanwhile, a supervised learning method, namely backpropagation, is used for training in MLP. These characteristics distinguish MLP from a linear perceptron and enable it to differentiate data that is not linearly separable (Cybenko, 1989). Alizadeh and Nikoo (2018) used five individual artificial intelligence models to estimate a ground-based drought index (SPI) using remote sensing factors and found that MLP had the best performance. Therefore, the MLP was also used in our study to test its performance in Australia. We implemented the MLP model via the "monmlp" package coded in R software.

3.2.3.6 Model performance assessment

In order to improve the performance of the proposed machine learning models, a trial and error analysis was adopted to determine the values of model's influential parameters. Then, the three machine learning models were calibrated and validated using 80% and 20% of the dataset, respectively. The dataset was randomly sampled and split it into calibration and validation datasets. This procedure was implemented 100 times to evaluate the stability of each model. The coefficient of determination (R^2) and root mean square error (RMSE) were used for the evaluation of model performance:

$$R^2 = \left(\frac{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \sqrt{\sum_{i=1}^{n} \left(P_i - \overline{P}\right)^2}} \right)^2 \tag{3-3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(O_i - P_i\right)^2}{n}} \tag{3-4}$$

where n is the number of samples, O_i and P_i denote observed and predicted values, $\overline{O}$ and $\overline{P}$ represent the mean of observed and predicted values. Generally, the model with higher R^2 and lower RMSE is considered to be the more accurate model. In addition, we also conducted a "leave-one-station-out" cross validation for each cluster of BOM stations to determine the performance of each model in continuous time series estimation of drought conditions.

3.3 Results

3.3.1 Model performance assessment

The VSURF feature selection method was used prior to developing drought prediction models. This method used ground-based SPEI values as the target variable and all ancillary variables as inputs, and resulted in 13 drought predictor variables for cluster 1 and 9 predictors for cluster 2 (Table 3-5, the "1" and "3" designations appended to the end of the variable names designate 1-month and 3-month time scale means).

Table 3-5 Drought predictor variables selected by VSURF for two data clusters in New South Wales, Australia. Predictor variable definitions are given in Table1. The "1" and "3" designations appended to the end of the variable names designate 1-month and 3-month time scale means.

Clusters	Selected variables
Cluster 1	Pre_3, PET_1, PET_3, LSTd_1, LSTd_3, LSTm_3, NDVI_3, NMDI_3, NDWI6_1, NDWI6_3, NDWI7_1, NDWI7_3, NDDI7_3
Cluster 2	Pre_3, PET_3, ET_1, LSTd_3, NMDI_1, NDWI6_1, NDDI5_1, NDDI7_1, NDDI7_3,

We used the selected predictors for each cluster to train the BRF, SVM, and MLP models for each cluster separately. The performance evaluation criteria (R^2 and RMSE) of the three models for the 100 runs are shown in Figure 3-4. Overall, the BRF model performed better than the other two models based on the two validation measurements, regardless of clusters. The BRF model explained about 90% of the SPEI variation with smaller prediction errors (RMSE<0.28) for the two clusters. The SVM and MLP models had similar performance, but the SVM model was slightly better. In addition, all three models performed better in cluster 2 (the wetter and colder stations), compared with the drier and warmer cluster 1 stations.

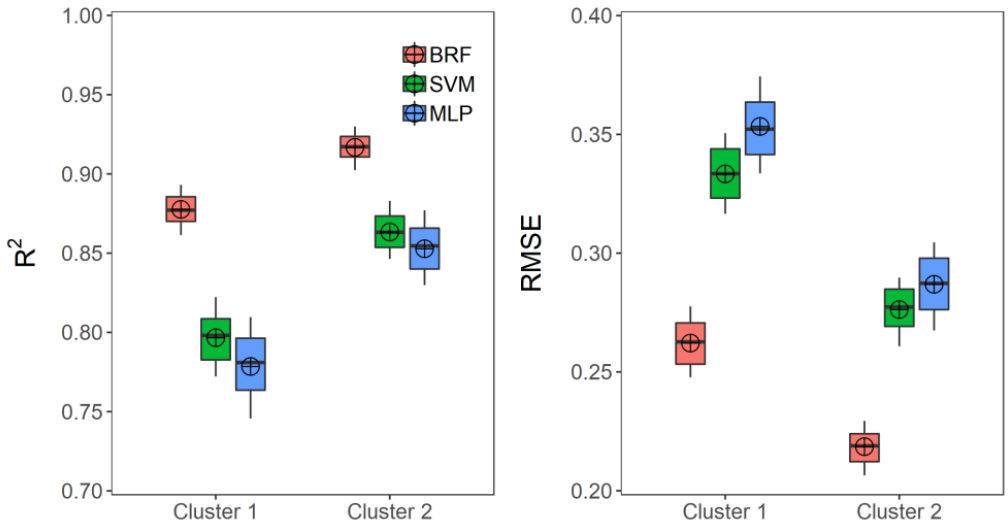


Figure 3-4 Boxplots of model performance measurements (coefficient of determination (R^2) and root mean squared error (RMSE)) for prediction of SPEI using bias-corrected random forest (BRF), support vector machine (SVM), and multi-layer perceptron neutral network (MLP) models for 100 model runs. Results were obtained from independent validation datasets which were randomly selected from the entire dataset. The black lines within each box indicate the

medians for 100 runs and crosshairs indicate means. Box boundaries indicate the 25th and 75th percentiles, whiskers below and above the box indicate the 10th and 90th percentiles.

We then conducted a "leave-one-station-out" cross validation for each cluster to determine the performance of each model in continuous time series estimation of SPEI. We selected results of two weather stations to show in the main text (Figure 3-5) and results for other stations can be found in the supplementary material (Figure 3-S1). Figure 3-5 shows the time series of observed and predicted 3-month time scale SPEI from 2001-2017 for station 4 (cluster 1) and station 5 (cluster 2). Similar to the above result, all three models had better performance at station 5 in terms of both pattern and intensity of drought conditions. However, all three models had relatively poorer performance at station 4 (the relatively warmer and drier station). Nevertheless, the BRF model still performed better than the other two models (i.e., better agreement between estimations and observations).

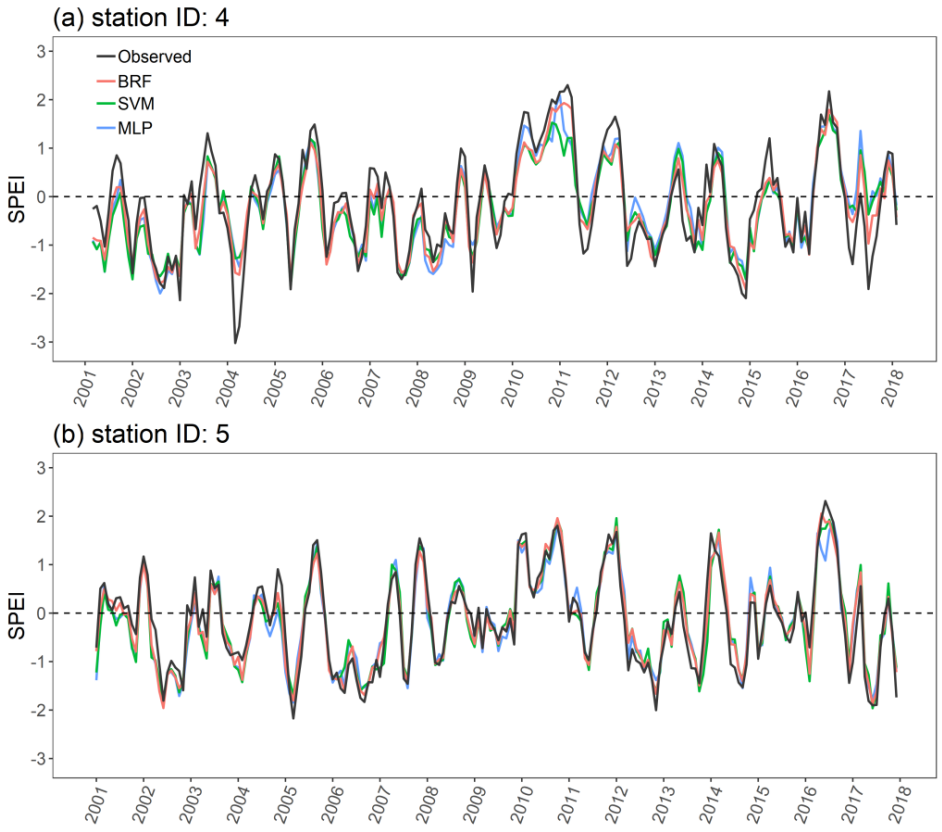


Figure 3-5 Predicted 3-month time scale SPEI (2001-2017) calculated from ground-based observations and from bias-corrected random forest (BRF), support vector machine (SVM), and multi-layer perceptron neutral network (MLP) models at two stations in New South Wales, Australia.

3.3.2 Relative importance of selected remotely-sensed drought factors

The BRF model can generate a measure of variable importance based on each predictor's relative influence (Were et al., 2015). We obtained the importance ranking of selected remotely-sensed drought factors based on a run of the BRF model using all of the data for each cluster (Figure 3-6). Generally, Pre_3 was ranked the most important in both clusters, accounting for 56% and 60% of the overall relative importance for cluster 1 and cluster 2, respectively. The importance magnitudes of other predictors were all near or lower than 10%. In cluster 1, the top four predictors were all meteorological factors, while vegetation factors all ranked relatively low. However, in cluster 2, vegetation indices had more impact on model accuracy, with NDWI6_1 and NDDI5_1 ranking 2nd and 3rd. Given the fact that cluster 2 stations are wetter and colder than cluster 1 stations, vegetation indices tended to be relatively more important for drought monitoring in humid environments.

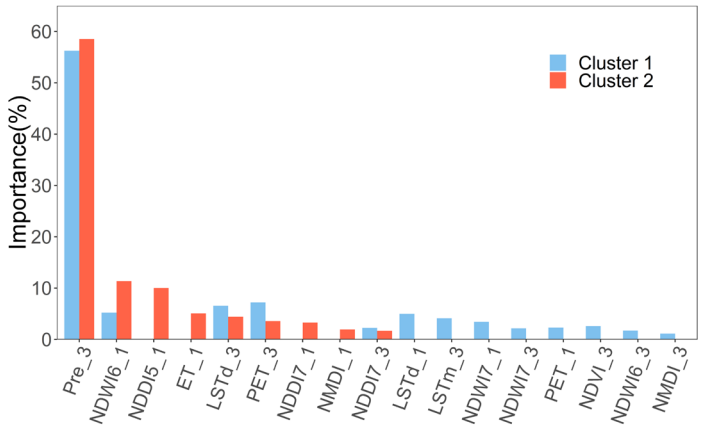


Figure 3-6 Relative importance of remotely-sensed drought factors as determined from the bias-corrected random forest (BRF) model for each cluster. There are 13 and 9 drought predictor variables in cluster 1 and cluster 2, respectively. The results are normalized to sum to 100% for each cluster.

3.3.3 Comparison with SILO drought maps

According to the above results, the BRF model was selected as the optimal model for drought monitoring in the study area. Thus, we mapped the spatial distribution of seasonal (spring: Sep.-Nov., summer: Dec.-Feb., autumn: Mar.-May, winter: Jun.-Aug.) drought conditions across the NSW wheat belt from 2001 to 2017. These maps were of predicted 3-month time scale SPEI values and could be directly used to evaluate agricultural drought conditions. The SILO dataset is a widely used station-based climate dataset to evaluate drought conditions in Australia. We were interested to see whether the SPEI drought maps derived from our BRF model were consistent with the SILO-drought maps. Thus, we also used the SILO dataset to calculate SPEI and interpolated the values calculated from 2242 stations to the NSW wheat belt using the inverse

distance weighting method. Then the SILO SPEI map was clipped by the boundary of cropland grids to match with BRF predicted drought maps. We selected a normal year (2001), a dry year (2002), and a wet year (2010) to show typical results in the main text (Figure 3-7, 3-8 and 3-9).

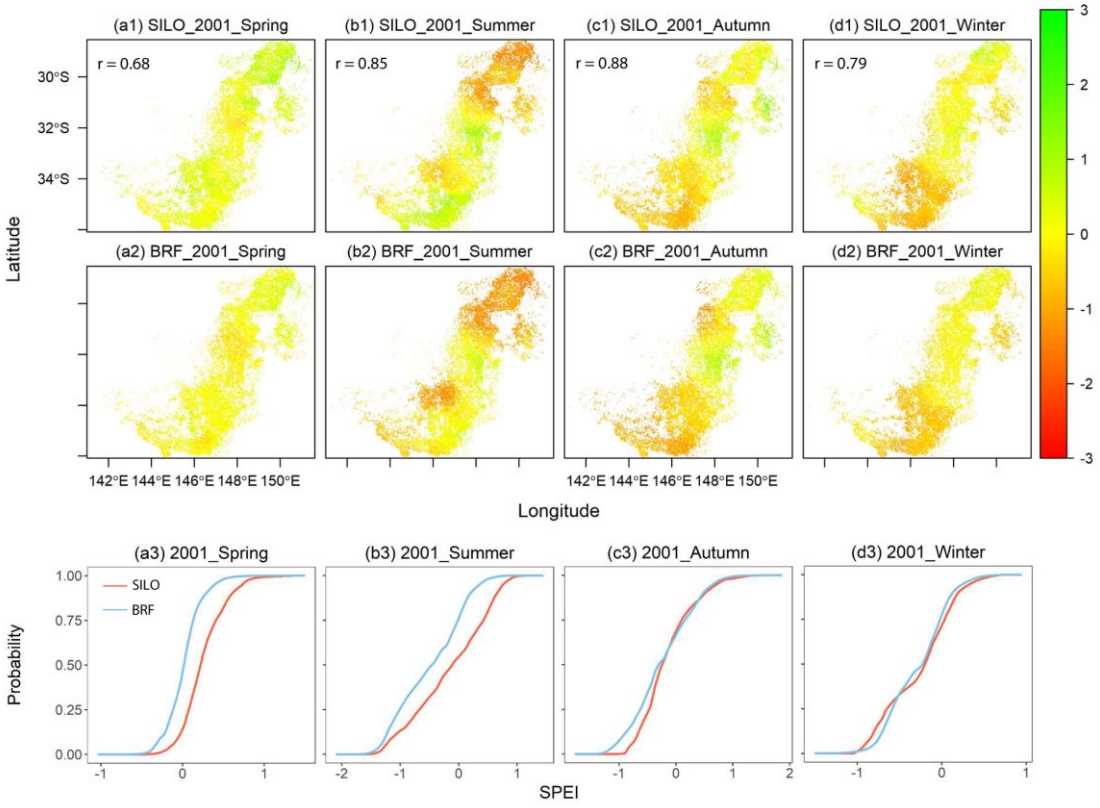


Figure 3-7 Comparison of SILO-observed and BRF-predicted SPEI drought maps for the four seasons of 2001. Spatial correlation values (Pearson correlation) between predictions from the two methods are annotated in the top four plots. The bottom four plots show the probability (cumulative distribution function) of SILO-observed and BRF-predicted SPEI values across the New South Wales wheat belt.

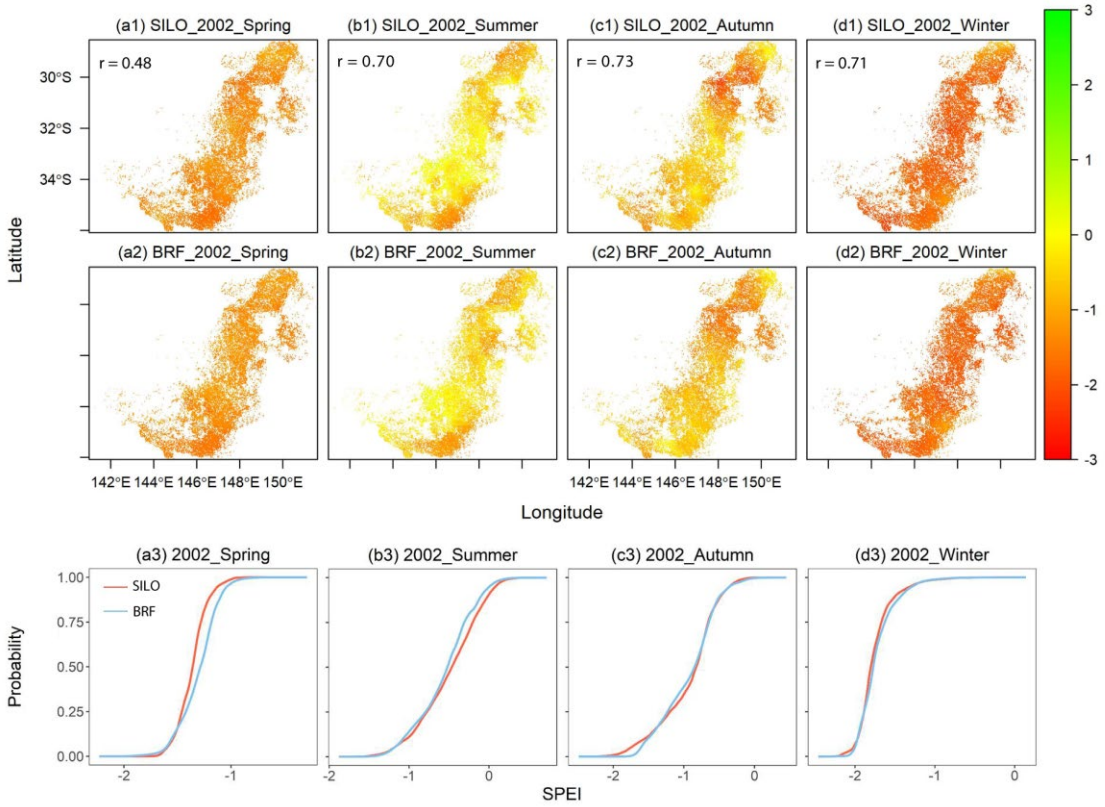


Figure 3-8 Comparison of SILO-observed and BRF-predicted SPEI drought maps for the four seasons of 2002. Spatial correlation values (Pearson correlation) between predictions from the two methods are annotated in the top four plots. The bottom four plots show the probability (cumulative distribution function) of SILO-observed and BRF-predicted SPEI values across the New South Wales wheat belt.

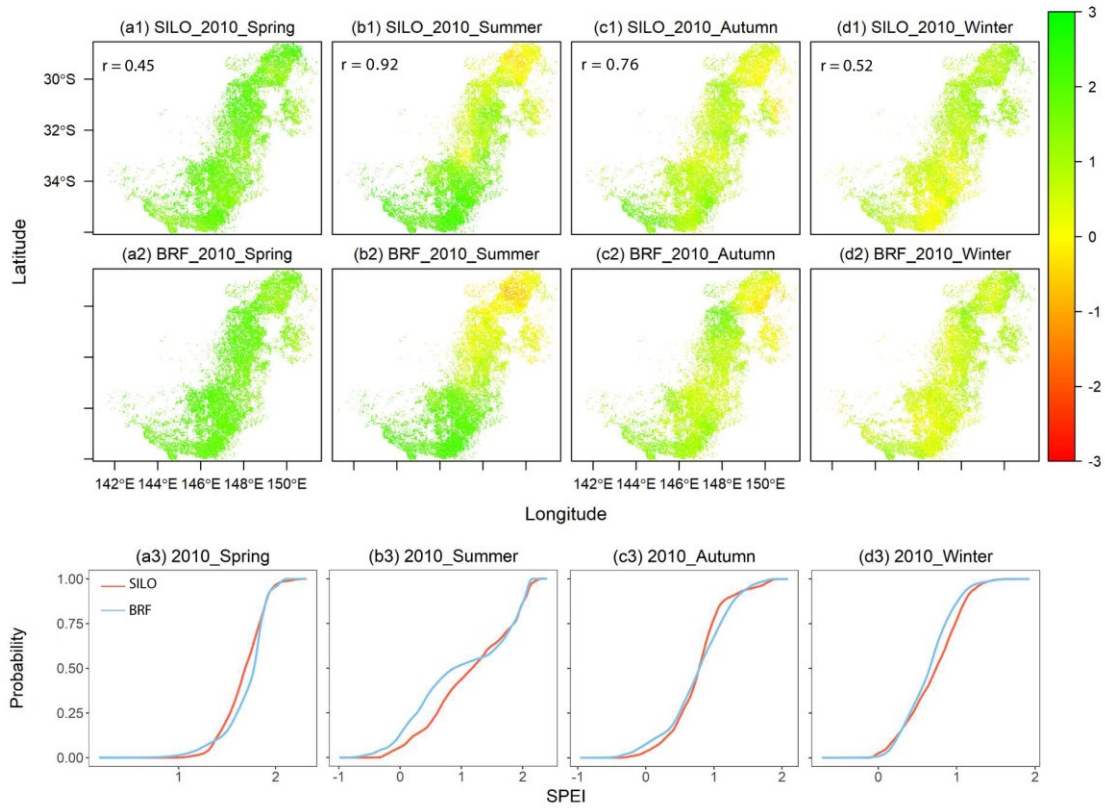


Figure 3-9 Comparison of SILO-observed and BRF-predicted SPEI drought maps for the four seasons of 2010. Spatial correlation values (Pearson correlation) between predictions from the two methods are annotated in the top four plots. The bottom four plots show the probability (cumulative distribution function) of SILO-observed and BRF-predicted SPEI values across the New South Wales wheat belt.

Generally, the SPEI drought maps derived from our BRF model were consistent with SILO-observed drought maps during normal, dry, and wet years. The BRF model successfully identified drought seasons and drought affected zones. We then used two measurements, i.e. spatial correlation and cumulative distribution function (CDF), to compare the two types of drought maps. Spatial correlation results ranged from 0.45 to 0.92 with a mean level of 0.71, meaning that the 500 m-resolution drought estimations from the two types of drought maps were highly correlated. Meanwhile, the CDF results also suggested a good match between the two drought maps in each season of 2001, 2002, and 2010. For example, in the dry year (2002, Figure 3-8), the CDF lines from the two maps nearly overlapped each other in each of the four seasons.

3.3.4 Monitoring agricultural drought

Crop yield is an important measured quantity reflecting agricultural drought. Yields of wheat grown under rainfed conditions in the NSW wheat belt are likely to be highly correlated with drought conditions. We compared seasonal SPEI predictions from the two types of drought

estimations (BRF and SILO) to measured wheat yields in 39 SA2 regions from 2001 to 2014 (Figure 3-10). Generally, drought estimations from the SILO dataset and the BRF model were similar for each year and each season. Drought conditions in the wheat belt had high inter-annual variability with obvious variations between wet and dry years. Wheat yields also varied greatly from year to year. As winter and spring are the main winter wheat growing seasons, there was an obvious and direct correspondence between wheat yields and drought conditions in these two seasons. For example, wheat yields were low in 2002 when spring and winter droughts occurred but were high in 2005 when spring and winter were wet. The last three years, 2012, 2013, and 2014, experienced spring and winter droughts, but still achieved medium to high yields, which may be due to farming practice improvement.

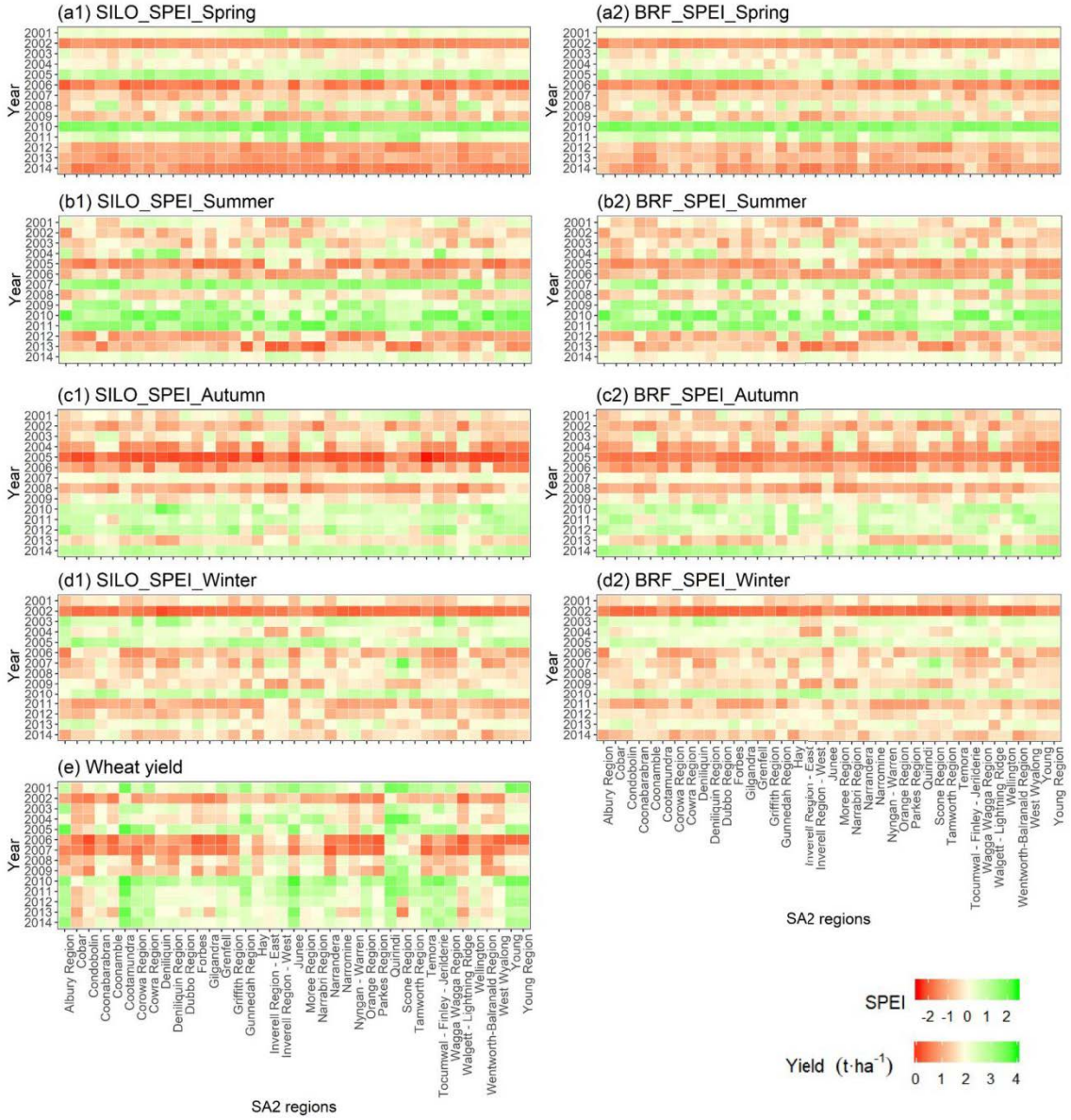


Figure 3-10 Seasonal SPEI values and wheat yields for 39 SA2 regions from 2001 to 2014. Drought estimations from both the SILO dataset and the BRF model are given in the figure. Region-level SPEI values are averaged values of cropland grids located in each region.

We calculated Pearson correlation between wheat yields and SPEI values from the two types of seasonal drought estimations to compare their performance in reflecting wheat yield (Figure 3-11). A first-difference method was applied to remove the possible confounding effects of non-climatic factors prior to calculation. Results showed that wheat yields had the highest correlation with spring drought conditions (r = 0.81). We also observed that the BRF-predicted SPEI values were slightly better correlated with wheat yields than were the SPEI values generated from the SILO dataset in all four seasons.

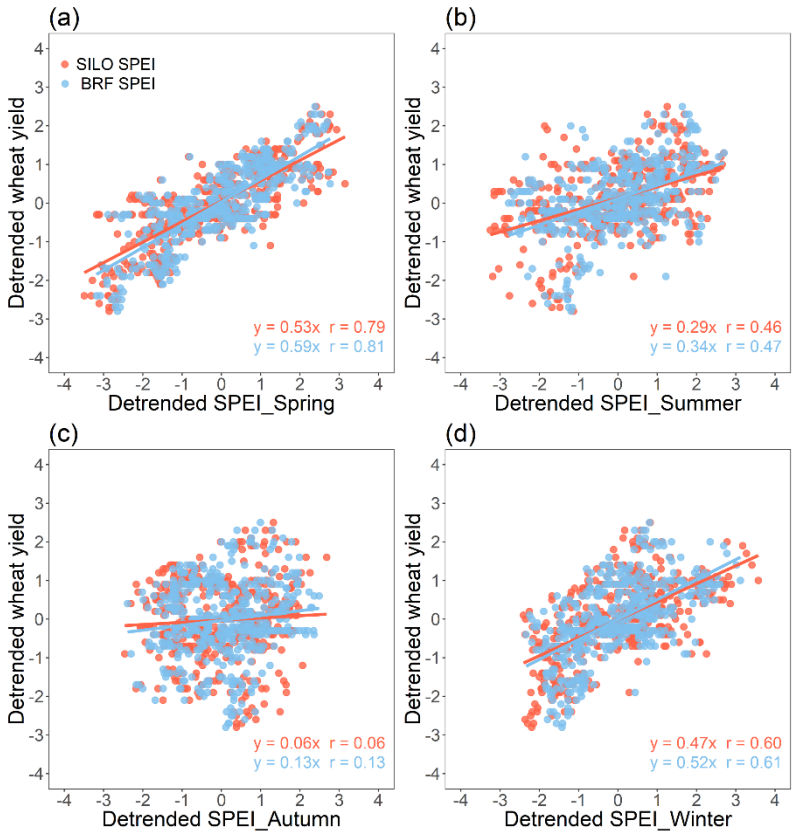


Figure 3-11 Comparisons of detrended wheat yields and detrended seasonal SPEI values (from both BRF and SILO) for the 39 SA2 regions from 2001-2014. The linear regression slopes and Pearson's correlation coefficients are given in the figure.

3.4 Discussion

A number of previous studies suggested that data-driven models can be very effective in drought monitoring (Alizadeh and Nikoo, 2018; Park et al., 2016; Park et al., 2017). Our study also produced satisfactory drought estimations using machine learning models based on remotely-sensed drought factors. In particular, the BRF model outperformed the SVM and MLP models in

reproducing the ground-based SPEI drought index, as quantified by larger R^2 and smaller prediction errors. This result is inconsistent with that reported by Alizadeh and Nikoo (2018) who demonstrated that MLP improved SPI estimations in Iran compared with four other machine learning models. This may be due to the differences in study areas, data sources, and the setup of model input and output. The superior estimation results derived from the BRF model in our study are likely attributable to reduced susceptibility to over-fitting and superior ability in dealing with the hierarchical and non-linear relationships that may exist between the ground-based drought index and remotely-sensed drought factors. Furthermore, the bias-corrected method (Zhang and Lu, 2012) used in the present study also helped improve the performance of the original RF model. The original RF model usually performs poorly in reproducing extreme observations, with large values underestimated and small values overestimated (Breiman, 2001). This may result in bias in drought monitoring as drought is usually defined as abnormally dry conditions compared with normal conditions. The bias-correction method used in our study significantly reduced the prediction bias with smaller RMSE compared with the results of Park et al. (2016). Thus, the combination of the RF and bias-correction methods outlined in our study is a promising approach for drought monitoring and can be extended to other cropping regions to obtain new insights to guide agricultural practices.

The validation results (Figure 3-4) showed that all three machine learning methods performed better for cluster 2 than cluster 1 in assessing drought conditions (i.e. SPEI). Generally, the cluster 1 stations represent a relatively arid environment, while the cluster 2 stations represent a more humid environment. Thus, the machine learning models tended to have better performance in the humid environment in our study. However, Park et al. (2016) obtained an opposite result suggesting that the machine learning models performed better for the arid region than the humid region in the USA. In their study, the average annual precipitation of the arid region was 323 mm, while the annual precipitation of the humid region was 1105 mm. In our study, these two values were 419 mm and 638 mm (Table 3-4), respectively. However, the model performance is not simply based on precipitation conditions. In our study, we also found that the SPEI values had higher correlation coefficients with vegetation indices in cluster 2 than cluster 1 (Figure 3-3). Moreover, vegetation indices had higher importance rankings in cluster 2 than cluster 1 (Figure 3-6). In other words, vegetation conditions in cluster 2 were more sensitive to 3-month time scale drought. The results reported by Park et al. (2016) are comparable to ours and they demonstrated that vegetation indices had higher importance rankings in the region where the machine learning models performed better. Generally, there is a time-lag relationship between drought and vegetation responses (Gessner et al., 2013; Piao et al., 2003). Lagged responses of vegetation to drought mainly depend on the characteristics of the region (Gessner et al., 2013). It is highly likely that vegetation in cluster 2 and in the arid study region of Park et al. (2016) reacted to water

shortage and adapted to limited water availability more rapidly during short-term droughts (e.g., 3 months). Therefore, the machine learning-based remotely-sensed drought monitoring is more suitable for semi-arid and vegetation-sensitive environments.

Since the main purpose of this study was to estimate 3-month scale SPEI using machine learning models and remotely-sensed data, a comparison was made between ground-based SPEI and estimations derived from machine learning models. There was good agreement between both method's SPEI estimations in terms of both pattern and intensity (Figure 3-7, 3-8 and 3-9). Thus, the machine learning-based fusion models did improve the accuracy of remotely-sensed drought estimations. Zhang et al. (2017) compared three different ground-based drought indices and six different remote sensing drought factors to demonstrate that remotely-sensed drought factors and ground-based drought indices usually spatially disagree, but even a simple fusion method can improve the correlation with ground-based indices. Thus, fusing various remotely-sensed drought factors is necessary in monitoring drought to fully explain the complexity and diversity of drought events (Mizzell, 2008; Wardlow et al., 2012). In this study, we accurately predicted the ground-based drought index, SPEI, for unmeasured areas based on remotely-sensed drought factors and the BRF model. This method can be seen as a drought monitoring product that can be extended to many other areas. This product has several advantages. First, the drought factors used in the present study are based on global remote sensing data (e.g., TRMM and MODIS) which covers from 50°S to 50°N, so the proposed approach can be applied to any vegetated region in the world. Second, this product can generate results comparable to the ground-based database. Third, this product is cost-free as the remotely-sensed data are easily accessible over the internet.

The proposed drought monitoring in the present study is also very useful for guiding agricultural practices. In rainfed croplands across the world, drought is widely viewed as the major threat causing yield losses. Timely drought warning systems can help farmers develop optimal strategies to ultimately reduce drought damage. In our study, high correlation coefficients were found between observed wheat yields and model-predicted seasonal SPEIs (spring, 0.81; winter, 0.61; and summer, 0.47). Thus, the remotely-sensed, fused SPEI could be a good tool to predict wheat yield based on estimated SPEI, ultimately providing invaluable information for grain price estimations. In addition, appropriate drought-response measures could be worked out in advance. For example, if SPEI_Summer reaches a low level, a drought-resistant wheat cultivar could be adopted in the NSW wheat belt to mitigate yield losses, or a fallow season could be recommended to limit production costs in a year in which very low yields were anticipated due to severe drought.

3.5 Conclusions

This study examined three advanced machine learning methods (namely BRF, SVM, and MLP) for estimating an agricultural drought index (3-month time scale SPEI) using various remotely-

sensed drought factors from MODIS and TRMM satellite sensors for the NSW wheat belt in south-eastern Australia. A station-based climate dataset (SILO) and crop yield data were used as reference data to evaluate the performance of the model-predicted SPEI in reflecting agricultural drought. The results indicated that the bias-corrected RF model is the most promising tool for monitoring agricultural drought. The bias-corrected RF model successfully produced agricultural maps of SPEI which were consistent with drought maps of SPEI derived from the station-based dataset. The results suggested that machine learning-based remotely-sensed drought monitoring is more suitable for semi-arid and vegetation-sensitive environments. Nevertheless, the BRF model can still provide satisfactory results in other types of environments. As such, the approach proposed in this study can be extended to any vegetated region where remotely-sensed data are available, even in areas with limited in situ data availability, to provide detailed spatial information regarding drought extent and severity.

We acknowledge that the spatial resolution of 500 m used in the present study is somewhat coarse. Drought estimations could be improved with use of remotely-sensed data of finer spatial resolution (e.g., 30 m). Furthermore, the BRF model still had some limitations in predicting extreme conditions of drought. Future research should be conducted using more advanced fusion models and more drought-related factors from more detailed data sources to achieve improved performance in drought estimation.

3.6 Supporting information

We conducted a "leave-one-station-out" cross validation for each cluster to determine the performance of each model in continuous time series estimation of SPEI. Figure 3-S1 below shows the results of predicted 3-month time scale SPEI (2001-2017) calculated from ground-based observations and from bias-corrected random forest (BRF), support vector machine (SVM), and multi-layer perceptron neutral network (MLP) models at BOM stations in New South Wales, Australia.

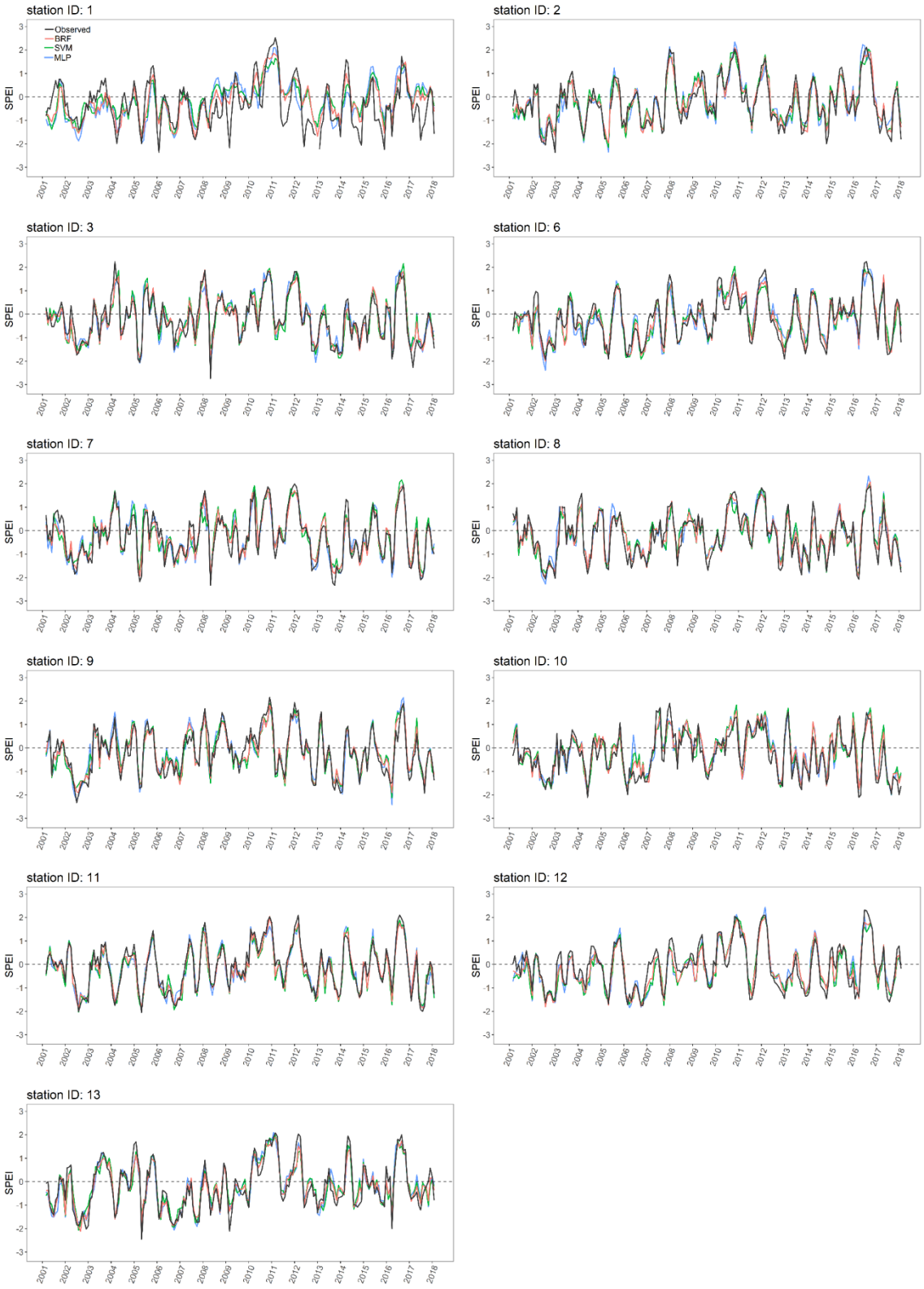


Figure 3-S1. Predicted 3-month time scale SPEI (2001-2017) calculated from ground-based observations and from bias-corrected random forest (BRF), support vector machine (SVM), and multi-layer perceptron neutral network (MLP) models at BOM stations in New South Wales, Australia.

3.7 Reference

Alizadeh, M.R., Nikoo, M.R., 2018. A fusion-based methodology for meteorological drought estimation using remote sensing data. Remote Sens Environ, 211: 229-247.

Bayarjargal, Y., Karnieli, A., Bayasgalan, M., Khudulmur, S., Gandush, C., Tucker, C. J., 2006. A comparative study of NOAA–AVHRR derived drought indices using change vector analysis. Remote Sens Environ, 105(1): 9-22.

Belayneh, A., Adamowski, J., Khalil, B., Ozga-Zielinski, B., 2014. Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology, 508: 418-429.

Breiman, L., 2001. Random Forest. Machine Learning, 45: 5-32.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3): 273-297.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., 2007. Random forests for classification in ecology. Ecology, 88(11): 2783-2792.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4): 303-314.

Dai, A., 2011. Erratum: Drought under global warming: a review. Wiley Interdisciplinary Reviews Climate Change, 2(1): 45-65.

Feng, P., Liu, D.L., Wang, B., Waters, C., Zhang, M., Yu, Q., 2019. Projected changes in drought across the wheat belt of southeastern Australia using a downscaled climate ensemble. International Journal of Climatology, 39(2), pp.1041-1053.

Gao, B.C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ, 58(3): 257-266.

Gao, X., Zhao, Q., Zhao, X., Wu, P., Pan, W., Gao, X., Sun, M., 2017. Temporal and spatial evolution of the standardized precipitation evapotranspiration index (SPEI) in the Loess Plateau under climate change from 2001 to 2050. Sci Total Environ, 595: 191-200.

Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2015. VSURF: an R package for variable selection using random forests. The R Journal, 7(2): 19-33.

Gessner, U., Naeimi, V., Klein, I., Kuenzer, C., Klein, D., Dech, S., 2013. The relationship between precipitation anomalies and satellite-derived vegetation activity in Central Asia. Global Planet Change, 110: 74-87.

Gu, Y., Brown, J.F., Verdin, J.P., Wardlow, B., 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. Geophysical Research Letters, 34(6).

Guzmán, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A.E., Pote, J.W., 2018. An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels. Agricultural Systems, 159, pp.248-259.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1): 100-108.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning New York. NY: Springer.

Hazaymeh, K., Hassan, Q.K., 2016. Remote sensing of agricultural drought monitoring: a state of art review.

He, B., Wu, J., Lü, A., Cui, X., Zhou, L., Liu, M. and Zhao, L., 2013. Quantitative assessment and spatial characteristic analysis of agricultural drought risk in China. Natural hazards, 66(2), pp.155-166.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ, 83(1-2): 195-213.

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ Modell Softw, 16(4): 309-330.

Kogan, F.N., 1995. Application of vegetation index and brightness temperature for drought detection. Advances in space research, 15(11): 91-100.

Lai, C., Zhong, R., Wang, Z., Wu, X., Chen, X., Wang, P. and Lian, Y., 2019. Monitoring hydrological drought using long-term satellite-based precipitation data. Sci Total Environ, 649: 1198-1208.

Lymburner, L., Tan, P., Mueller, N., Thackway, R., Lewis, A., Thankappan, M., Randall, L., Islam, A. and Senarath, U., 2010. 250 metre dynamic land cover dataset of Australia. Geoscience Australia, Canberra.

McKee, T.B., Doesken, N.J., Kleist, J., 1993, January. The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology (Vol. 17, No. 22, pp. 179-183). Boston, MA: American Meteorological Society.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on geoscience and remote sensing, 42(8): 1778-1790.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. Journal of Hydrology, 391(1–2): 202-216.

Mizzell, E.H.P., 2008. Improving drought detection in the Carolinas: evaluation of local, state, and federal drought indicators, University of South Carolina.

Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. International Journal of Climatology, 27(15): 2103-2111.

Mottaleb, K.A., Gumma, M.K., Mishra, A.K., Mohanty, S., 2015. Quantifying production losses due to drought and submergence of rainfed rice at the household level using remotely sensed MODIS data. Agricultural Systems, 137: 227-235.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3): 247-259.

Nicholls, N., 1997. Increased Australian wheat yield due to recent climate trends. Nature, 387(6632): 484-485.

Niemeyer, S., 2008. New drought indices. Options Méditerranéennes. Série A: Séminaires Méditerranéens, 80: 267-274.

Palmer, W.C., 1965. Meteorological drought. Research Paper No. 45. Washington, DC: US Department of Commerce. Weather Bureau: 59.

Park, S., Im, J., Jang, E., Rhee, J., 2016. Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. Agricultural & Forest Meteorology, 216: 157-169.

Park, S., Im, J., Park, S., Rhee, J., 2017. Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. Agr Forest Meteorol, 237: 257-269.

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., Friedlingstein, P., 2010. The impacts of climate change on water resources and agriculture in China. Nature, 467(7311): 43.

Piao, S., Fang, J., Zhou, L., Guo, Q., Henderson, M., Ji, W., Li, Y., Tao, S., 2003. Interannual variations of monthly and seasonal normalized difference vegetation index (NDVI) in China from 1982 to 1999. Journal of Geophysical Research: Atmospheres, 108(D14).

Quiring, S.M., Ganesh, S., 2010. Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas. Agr Forest Meteorol, 150(3): 330-339.

Quiring, S.M., Papakryiakou, T.N., 2003. An evaluation of agricultural drought indices for the Canadian prairies. Agr Forest Meteorol, 118(1-2): 49-62.

Rhee, J., Im, J., Carbone, G.J., 2010. Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. Remote Sens Environ, 114(12): 2875-2887.

Rouse Jr, J.W., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS.

Song, J., 2015. Bias corrections for Random Forest in regression using residual rotation. Journal of the Korean Statistical Society, 44(2): 321-326.

Swain, S., Wardlow, B.D., Narumalani, S., Tadesse, T., Callahan, K., 2011. Assessment of vegetation response to drought in Nebraska using Terra-MODIS land surface temperature and normalized difference vegetation index. GIScience & Remote Sensing, 48(3): 432-455.

Thornthwaite, C.W., 1948. An Approach toward a Rational Classification of Climate. Geographical Review, 38(1): 55-94.

Touma, D., Ashfaq, M., Nayak, M.A., Kao, S.C., Diffenbaugh, N.S., 2015. A multi-model and multi-index evaluation of drought characteristics in the 21st century. Journal of Hydrology, 526: 196-207.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index - SPEI. Journal of Climate, 23(7): 1696-1718.

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I. and Sides, T., 2018. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecol Indic, 88: 425-438.

Wang, L., Qu, J.J., 2007. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. Geophysical Research Letters, 34(20).

Wang, Q., Shi, P., Lei, T., Geng, G., Liu, J., Mo, X., Li, X., Zhou, H., Wu, J., 2015. The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. International Journal of Climatology, 35(13): 3760-3769.

Wang, Z., Li, J., Lai, C., Wang, R.Y., Chen, X., Lian, Y., 2018. Drying tendency dominating the global grain production area. Global food security, 16, pp.138-149.

Wardlow, B.D., Anderson, M.C., Verdin, J.P., 2012. Remote sensing of drought: Innovative monitoring approaches. CRC Press.

Were, K., Bui, D.T., Dick, O.B., Singh, B.R., 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. Ecol Indic, 52: 394-403.

Wilhite, D.A., 2005. Drought and water crises: science, technology, and management issues. CRC Press.

Wong, G., Lambert, M., Leonard, M., Metcalfe, A., 2009. Drought analysis using trivariate copulas conditional on climatic states. Journal of Hydrologic Engineering, 15(2): 129-141.

Wu, X., Wang, P., Huo, Z., Wu, D., Yang, J., 2018. Crop Drought Identification Index for winter wheat based on evapotranspiration in the Huang-Huai-Hai Plain, China. Agriculture, Ecosystems & Environment, 263: 18-30.

Yagci, A.L., Di, L., Deng, M., 2015. The effect of corn–soybean rotation on the NDVI-based drought indicators: A case study in Iowa, USA, using vegetation condition index. GIScience & Remote Sensing, 52(3): 290-314.

Zarch, M.A.A., Sivakumar, B., Sharma, A., 2015. Droughts in a warming climate: A global assessment of Standardized precipitation index (SPI) and Reconnaissance drought index (RDI). Journal of Hydrology, 526: 183-195.

Zell, A., Mache, N., Huebner, R., Mamier, G., Vogt, M., Schmalzl, M., Herrmann, K.U., 1994. SNNS (stuttgart neural network simulator), Neural Network Simulation Environments. Springer, pp. 165-186.

Zhang, A., Jia, G., 2013. Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. Remote Sens Environ, 134: 12-23.

Zhang, G., Lu, Y., 2012. Bias-corrected random forests in regression. Journal of Applied Statistics, 39(1): 151-160.

Zhang, X., Chen, N., Li, J., Chen, Z., Niyogi, D., 2017. Multi-sensor integrated framework and index for agricultural drought monitoring. Remote Sens Environ, 188: 141-163.

Chapter 4. Dynamic wheat yield forecasts are improved by developing a hybrid approach using biophysical model and machine learning technique

This chapter is based on the following manuscript:

Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., and Yu, Q., 2019. " Dynamic wheat yield forecasts are improved by developing a hybrid approach using biophysical model and machine learning technique." Agricultural and Forest Meteorology. (Second round review)

Highlights

• A hybrid approach was developed to forecast wheat yield for Australia.

• The APSIM crop model and random forest algorithm was integrated.

• Various kinds of growth stage-specific information were considered.

• Satisfactory yield forecasts occurred at 1~2 months prior to harvest.

• Drought was identified as the most influential factor affecting final wheat yields.

Abstract

Early and reliable seasonal crop yield forecasts are crucial for both farmers and decision-makers. Commonly-used methods for seasonal yield forecasting are based on process-based crop models or statistical regression-based models. Both have limitations, particularly in regard to accounting for growth stage-specific climate extremes (such as drought, heat, and frost). In this study, we firstly developed a hybrid yield forecasting approach by blending of multiple growth stage-specific indicators, i.e. APSIM (a process-based crop model)-simulated biomass, and climate extremes, NDVI (Normalized Difference Vegetation Index), and SPEI (Standardized Precipitation and Evapotranspiration Index) before forecasting dates, using a regression model (random forest or multiple linear regression). Plot-scale wheat yield (2008-2017) in the southeastern Australian wheat belt was dynamically forecasted at the end of several targeted growth stages as the growing season progressed to harvest. Results showed that the forecasting accuracy increased significantly for both systems as forecast time approached harvest time. The forecasting system based on random forest outperformed the forecasting system based on multiple linear regression at each forecasting event. Satisfactory yield forecasts occurred at one month (~35 days) prior to harvest (r=0.85, LCCC=0.81, MAPE=17.6%, RMSE=0.70 t ha^{-1}, and ROC score=0.90), and at two months before harvest (r=0.62, LCCC=0.53, MAPE=27.1%, RMSE=1.01 t ha^{-1}, and ROC score=0.88). In addition, drought events throughout the growing season were

identified as the main factor causing yield losses in the wheat belt during the past decade. With the increasing availability of farming-related data, we expect that the yield forecasting system proposed in our study may be widely extended to other comparable cropping regions to produce sufficiently accurate wheat yield forecasts for stakeholders to develop strategic decisions in their respective roles.

Keywords: wheat yield forecast; extreme climate events; remote sensing; APSIM; random forest

4.1 Introduction

Seasonal forecasting of crop yield is becoming increasingly important in both developed and developing countries (Basso and Liu, 2018). This is mainly due to the growing demand for maximizing profits in terms of both farm-level outputs and commodities trading. Early and reliable warning information regarding weather and management impacts on crop yield is crucial for stakeholders to make strategic decisions in their respective roles. For crop producers, once crop yield is site-specifically predicted, appropriate farm management practices and security precaution measures (e.g., grain storage) can be determined. For government policy makers and grain marketing agencies, yield forecasting can provide invaluable information for regulating agricultural markets and determining trading strategies. Therefore, many crop yield forecasting approaches have been developed across the world to provide crop yield outlooks (Cai et al., 2019; Chipanshi et al., 2015; Pagani et al., 2017).

At present, commonly-used yield forecasting approaches can be divided into three categories: (1) field surveys, (2) dynamic process-based crop simulation models, and (3) statistical regression-based models. The field survey method is still in use in many operational yield forecasting systems. It is conducted based on the within-season evaluation of crop growth by experienced farmers or farm managers (Nandram et al., 2014). Their evaluation of yield prospects represents farmers' opinions about the effects of environmental and human factors on the final yield. Numerous crop yield forecasts from farmers can be collected through interviews (e.g., phone interviews) to give a synthetic assessment of yield outlooks for a specified region. However, the field survey method is usually time- and labor-consuming and provides relatively short lead times to inform decisions. Due to these limitations, a large amount of effort has been undertaken to obtain timely and reliable yield forecasts from the other two methods. Crop simulation models are capable of describing key physical and physiological processes by capturing the effects of the complex interactions between crop, soil, weather, and management practices. Thus, they can usually provide satisfactory end-of-season yield forecasts once required input data and parameters are provided. When using crop models for in-season yield forecasting, a major limitation is related to the unknown weather from the forecasting date to the maturity date (Basso and Liu, 2018). Previous studies have used various methods to fill this gap, including historical weather scenarios,

seasonal weather forecasts, and climate model outputs, to run crop simulations to maturity date. However, in regions with large inter-annual climate variability, actual weather conditions can vary significantly from the projected weather conditions (Murphy and Timbal, 2008). Thus, great uncertainty can be introduced into the yield forecasting results. Moreover, crop simulation models usually have limited ability to simulate the effects of extreme climate events (ECEs, such as heat wave, frost, and drought). Some of the limitations are related to uncertainties in parameterization and vague descriptions or over-simplifications of certain processes, which can result in inaccurate yield estimations (Eitzinger et al., 2013). For example, the impacts of heat stress in particular are poorly captured in crop models (Barlow et al., 2015). Most crop models simulate the impacts of extreme temperatures on stem carbohydrate accumulation and distribution or leaf senescence, instead of directly modelling damage to reproductive processes or organs (Feng et al., 2019b). These limitations also raise uncertainty regarding the ability of crop models to properly forecast the end-of-season yield.

Statistical regression-based models relate crop yields to a number of selected predictors, such as meteorological factors and/or vegetation indices derived from remote sensing data, based on historical observations from a given region. Regression equations are then employed as a function of the inputs from other years in this region, or similar regions, to generate yield forecasts. Generally, statistical regression-based models are usually simple and easy to understand and need fewer parameter settings, thereby making them widely-used around the world. As the observed data are increasing in both quantity and quality in recent years, statistical regression-based models usually present satisfactory performance (Lobell and Asseng, 2017; Mathieu and Aires, 2018), especially under conditions characterized by large year-to-year fluctuations in yields, driven by several factors. However, statistical regression-based models are also not free from problems. Most current statistical regression-based models are based on linear regression models, such as multiple linear regression, which cannot capture the nonlinear relations between the dependent and independent variables. Given that crop yields often show nonlinear response to ECEs (Li et al., 2019; Feng et al., 2019b), linear regression models are likely to perform poorly under conditions with frequent climate extremes. Moreover, different meteorological factors and/or vegetation conditions occurring during each growth stage can have different impacts on crop yield. For example, heat or drought events that occur during the flowering stage are likely to cause greater yield losses than those occurring during vegetative stages (Nielsen et al., 2010; Stratonovitch and Semenov, 2015). As statistical regression-based models are usually unable to consider dynamic growth stage changes, they are limited in their ability to disentangle the effects of stage-specific factors.

Given that both crop simulation models and statistical regression-based models have limitations, researchers are attempting to integrate the two types of models in order to achieve complementary

advantages. For example, the Crop Growth Modelling System (CGMS) incorporated crop simulation results and linear regression to provide policy makers with in-season regional yield forecasts of the main food crops in Europe (Kogan et al., 2013; Vossen and Rijks, 1995). Pagani et al. (2017) improved CGMS by incorporating agro-meteorological indices (accounting for drought and cold/heat stress) into the linear regression equation and increased the forecasting reliability by 94% in several European countries. The Integrated Canadian Crop Yield Forecaster (ICCYF) is another regional yield forecasting system for producers, grain traders, and policy makers, which integrates remotely sensed vegetation indices, climate, soil, and crop information through a crop simulation model and linear regression algorithms (Chipanshi et al., 2015). As these systems are still based on crop simulation models and traditional linear regression models, they may be limited in modelling yield losses caused by ECEs. Everingham et al. (2016) developed a random forest model (a machine learning method) with climate indicators and APSIM (a biophysical process-based crop model)-simulated biomass as predictors to forecast regional sugarcane (*Saccharum officinarum* L.) yield and obtained an R^2 of 0.67. However, they did not include remotely sensed indices as predictors, which might also help increase prediction accuracy. In addition, most yield forecasting studies did not consider growth stage-specific indices (Cai et al., 2019; Kern et al., 2018). Therefore, with the increasing availability of farming-related data, a more comprehensive yield forecasting system that incorporates growth stage-specific climate, remote sensing, soil, and crop information through crop simulation models and advanced regression methods is urgently needed for more accurate yield predictions.

Australia is among the most developed agricultural countries and one of the biggest wheat producers and exporters in the world. Early and reliable wheat yield forecasts for Australia become a critical element in providing guidance to farmers and policy makers. As in the areas of the Oceania region, atmospheric circulation patterns play an influential role on Australia's weather, resulting in large inter-annual variability. Frequent extreme climate events, such as drought, heat, and frost have caused severe wheat yield losses during the last decades. As such, previous studies in Australia have focused on the use of agrometeorological models in yield forecasting and have rarely considered remotely sensed information. In this study, we intended to use growth stage-specific information from multiple data sources to make in-season yield forecasts based on the APSIM crop model and the random forest algorithm. The major objectives were to 1) develop a hybrid approach to forecast yield using a biophysical model and machine learning technique at a plot scale, 2) identify the optimum lead time with acceptable accuracy of yield forecasting, and 3) quantify the relative contribution of growth stage-specific predictors for determining end-of-season wheat yield.

4.2 Materials and methods

4.2.1 Study area

The study area was the New South Wales (NSW) wheat belt located in southeastern Australia (Figure 4-1). The wheat belt is located between the Great Dividing Range and the arid interior of Australia. Topography and climatic conditions vary across the extent of the study area, showing a west-east gradient in both altitude and temperature/precipitation. The western areas consist of plains and the eastern areas are mainly mountains with elevations up to 1100 m. Average (1961-2000) growing season (for winter wheat, May-November) precipitation ranges from 171 mm in the southwest to 763 mm in the southeast, and average growing season temperature ranges from 8.3 °C in the southeast to 17.1 °C in the northwest. The NSW wheat belt accounts for about 28% of the total wheat production area in Australia (www.abares.gov.au, 2013-14), thus is crucial for both domestic and international food supply (https://www.dpi.nsw.gov.au/about-us/publications/pdi/2018/wheat). In addition, wheat plants are mainly grown under rainfed conditions in the wheat belt.

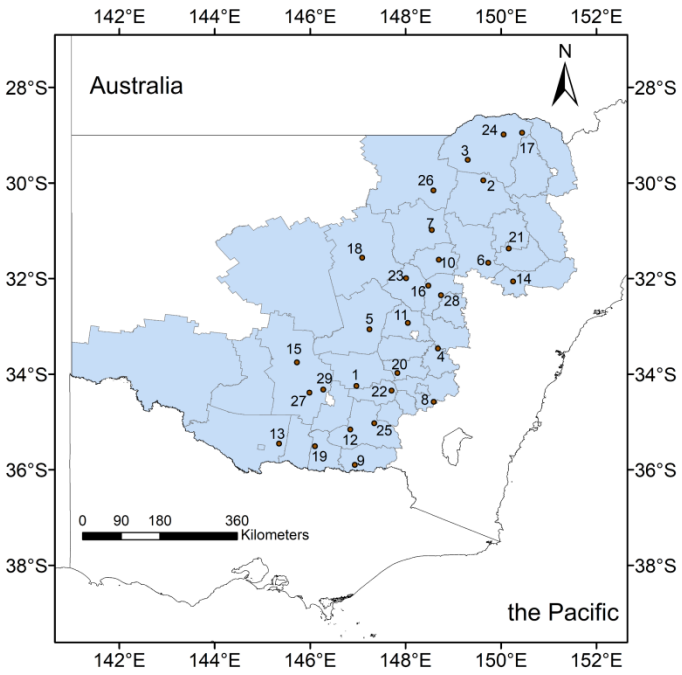


Figure 4-1 Locations of the 29 study sites in the New South Wales wheat belt in southeastern Australia. Detailed information for each site is provided in Table 4-1.

We used 29 sites distributed across most of the wheat belt (Figure 4-1). As with our previous study (Feng et al., 2019a), cropping areas were mostly located in the eastern parts of the wheat belt. Thus, these 29 sites can represent various agro-climatic regions within the belt. These sites are listed in the Grains Research and Development Corporation National Variety Trials (GRDC-

NVT, http://www.nvtonline.com.au/). Long-term wheat variety trials are conducted in these sites and detailed experimental records could be used to develop our model. Table 4-1 presents basic information for the 29 study sites, including location, soil information, climate conditions, and wheat yields.

Table 4-1 Basic information of the 29 study sites in New South Wales, Australia, including location, soil name (details at http://www.asris.csiro.au/), growing season (May-November) rainfall (GSR, mm), growing season temperature (GST, °C), number of years of yield data available (NY), and wheat yield (t ha^{-1}) range for recorded years (2008-2017).

ID	Site	Longitude	Latitude	Soil name	GSR	GST	NY	Yield range
1	Beckom	147.0	-34.2	Kandosol	298	12.9	9	1.1-4.8
2	Bellata	149.6	-29.9	Vertosol	307	15.9	9	1.4-5.3
3	Bullarah	149.3	-29.5	Vertosol	252	17.1	7	1.7-5.4
4	Canowindra	148.7	-33.5	Kandosol	342	12.6	10	1.3-6.0
5	Condobolin	147.2	-33.1	Sodosol	290	12.3	7	1.1-4.6
6	Coolah	149.7	-31.7	Vertosol	507	10.9	9	3.9-4.9
7	Coonamble	148.6	-31.0	Sodosol	267	15.4	10	1.4-5.5
8	Galong	148.6	-34.6	Kandosol	392	10.9	9	2.1-5.5
9	Gerogery	147.0	-35.9	Sodosol	386	12.2	10	3.1-6.9
10	Gilgandra	148.7	-31.6	Sodosol	313	14.3	10	1.7-4.8
11	Goonumbla	148.0	-32.9	Sodosol	331	13.6	9	1.3-5.6
12	Lockhart	146.8	-35.2	Sodosol	312	12.9	9	1.0-5.9
13	Mayrung	145.3	-35.5	Sodosol	251	13.1	8	4.4-7.4
14	Merriwa	150.3	-32.1	Vertosol	308	13.3	10	2.0-5.2
15	Merriwagga	145.7	-33.8	Kandosol	233	14.2	7	1.6-4.9
16	Narromine	148.5	-32.1	Sodosol	289	14.5	1	5.2
17	North Star	150.4	-28.9	Vertosol	270	16.3	9	2.6-5.4
18	Nyngan	147.1	-31.6	Kandosol	253	15.5	7	1.0-4.7
19	Oaklands	146.1	-35.5	Sodosol	306	12.8	10	1.4-5.0
20	Quandialla	147.8	-34.0	Sodosol	317	13.1	9	1.7-5.0
21	Spring Ridge	150.2	-31.4	Vertosol	314	14.4	10	2.7-6.0
22	Temora	147.7	-34.3	Chromosol	305	12.2	8	2.0-5.3
23	Trangie	148.0	-32.0	Sodosol	273	14.7	9	1.2-5.5
24	Tulloona	150.1	-29.0	Vertosol	263	16.7	8	2.3-4.4
25	Wagga wagga	147.3	-35.0	Kandosol	364	12.1	7	2.2-5.2
26	Walgett	148.6	-30.2	Vertosol	239	16.3	6	0.9-5.2
27	Willbriggie	146.0	-34.4	Sodosol	249	13.7	3	4.5-6.1
28	Wongarbon	148.7	-32.3	Kandosol	347	13.5	9	1.3-5.1
29	Yenda	146.3	-34.3	Sodosol	265	13.8	2	3.7-4.4

4.2.2. Data

4.2.2.1. In-situ climate data

Observed daily climate data (2008-2017) for the 29 sites, including solar radiation, precipitation, and minimum and maximum air temperature, were acquired from the Scientific Information for Land Owners patched point dataset (SILO-PPD, https://legacy.longpaddock.qld.gov.au/silo/).

4.2.2.2. Remote sensing data

Many factors can influence wheat growth, including soil type and condition, temperature and precipitation during the growing season, management practices, and diseases. Meteorological information is usually readily available, but impacts of other factors can be difficult to quantify. Remote sensing vegetation indices are the metrics that are able to measure crop canopy conditions directly. Therefore, a number of remote sensing vegetation indices have been developed and used extensively in crop yield estimation and forecasting around the world (Bolton and Friedl, 2013; Huang et al., 2015). The Normalized Difference Vegetation Index (NDVI), has become the most frequently used index in vegetation monitoring (Johnson et al., 2016). In particular, NDVI information is considered to be most useful in dry environments or where the vegetation condition shows high inter-annual variations (Balaghi et al., 2008). Thus, NDVI is able to provide useful information about environmental conditions for a given year. So, in this study, we included NDVI as a predictor to build our wheat yield forecasting system. Daily NDVI data (2008-2017, 500 m spatial resolution) for the study area were obtained from the MODIS/MOD09GA surface reflectance composites hosted by the Google Earth Engine (GEE, https://earthengine.google.com) platform.

4.2.2.3. Wheat trial data

Wheat variety experiment data (2008-2017) for the 29 sites (Figure 4-1) were derived from the GRDC-NVT. The GRDC-NVT is a national project of crop variety testing that assists farmers in making variety decisions. These variety experiments were conducted at plot scale (~1.5 m × ~10 m) at selected trial sites. Three plots at each site were used to determine one observation. Wheat was harvested using well-maintained harvesting equipment at the earliest opportunity after physiological maturity of the plots to minimize grain losses through wind, rain, or pest damage. Other management practices, e.g. sowing, were performed in accordance with local farmers' practices for a certain location. GRDC-NVT data included wheat yield and management practice information (sowing date, fertilization, etc.) for a variety of cultivars. Soil nutrient condition (including organic carbon, phosphorous, total nitrogen, conductivity, and pH) were also available. We chose the wheat cultivar, *Sunvale*, to conduct our study, as it was the most widely used cultivar across the study area and also has been parameterized in the APSIM model. The GRDC-NVT trial data were not available for all 10 years at each site, and as such we ultimately were able to

use 231 sets of wheat trial data. These relatively recent data from experimental trials exhibited no significant technological trends. Therefore, no de-trending approach was implemented to exclude the effects of factors such as pesticide application, fertilizer practices, and varietal improvement, which were not reproduced by modelling.

4.2.2.4. Soil hydraulic properties

Detailed soil hydraulic properties for the 29 sites were acquired from the APSoil database (http://www.asris.csiro.au/) (Dalgliesh et al., 2006). There are more than 800 soil profiles available for Australian agricultural areas in this database. The majority of those soils are parameterized for wheat modelling. Soils that were identified to be geographically closest to our study sites were ultimately selected (Table 4-1). Using close APSoil soil profiles as APSIM input is a common practice used in many other wheat modelling studies in Australia (Innes et al., 2015; Western et al., 2018). The same soil was selected for Coolah (site 6) and Merriwa (site 14).

4.2.3. Modelling methodology

4.2.3.1. APSIM simulations

In the present study, the APSIM (Agricultural Production System sIMulator, http://www.apsim.info/) crop model version 7.7 was implemented to simulate the dynamic changes of wheat phenology and biomass. As the APSIM crop model was developed by Australian institutes, it has been well calibrated in many wheat production areas throughout the Australian wheat belt. The APSIM simulations were set up strictly based on GRDC-NVT trials data (sowing date, variety, fertilization practice, hydraulic properties, and soil nutrient status) at the 29 sites. As stated previously, the *Sunvale* wheat variety is readily available in the APSIM variety database. The dynamic output of biomass and phenology information were then used to feed the statistical models.

4.2.3.2. Regression models

We used two regression models, multiple linear regression (MLR) and random forest (RF), to compare their performance in forecasting wheat yield. MLR is a widely used approach to explore the linear relation between the dependent and independent variables. Compared to ordinary least-squares method, MLR can be viewed as an extension that includes more than one predictor variable. It is easy to understand and use, but usually limited in ability to disentangle the nonlinear relations between the dependent and independent variables.

RF is a popular machine learning method for various regression and classification purposes. In brief, RF is a nonparametric approach that builds numerous independent decision trees and assembles them together to gain a more accurate and stable prediction (Breiman, 2001). RF is capable of modelling nonlinear and hierarchical relationships between the response and the

predictor variables. Our previous studies demonstrated the better performance of RF for exploring the nonlinear effects of ECEs on crop yield in comparison with MLR (Feng et al., 2018; Feng et al., 2019b). In this study, we evaluated the performance of RF for in-season wheat yield forecasting by blending of growth stage-specific APSIM simulated biomass, ECEs, and remote sensing information, with MLR used as a benchmark.

4.2.3.3. Indicators of climate extremes

Wheat cultivation can be divided into 11 Zadok growth stages in the APSIM wheat module: sowing, germination, seedling growth, tillering, stem elongation, booting, awn emergence, flowering (anthesis), milk development, dough development, and ripening (http://apsrunet.apsim.info/svn/apsim/trunk/Documentation/Model,CropandSoil/CropModuleDocumentation/Wheat.html). These stages are dynamic and continuous processes and wheat growth enters a subsequent stage once it finishes a current stage. We intended to forecast final yield once wheat growth finished a stage. Thus, we might need to forecast final yield for 11 times at 11 different dates. However, some growth stages, such as germination and flowering, last for only a few days in APSIM simulations, and therefore it is probably of little value to forecast yield for two times over such a short period. Thus, we grouped these 11 stages into six categories, i.e. S (sowing date), SG (from sowing to end of seedling growth, around 33-43 days), T (from end of seedling growth to end of tillering, around 49-60 days), SE (from end of tillering to end of stem elongation, around 23-27 days), BAF (from end of stem elongation to end of flowering, around 27-30 days), and M (from end of flowering to end of milk development, around 18-20 days). Then we forecasted final yield at the end of each category. The indices for frost and heat were the number of days with minimum/maximum temperatures below/above fixed threshold (<2 °C for frost and >27 °C for heat) (Tashiro and Wardlaw, 1989; Zheng et al., 2012).

The effect of growth stage-specific drought was assessed based on the Agricultural Reference Index for Drought (Woli et al., 2012). It is a general, simple, and daily scale drought index which usually has better performance in agricultural drought assessment compared with many other drought indices (Woli et al., 2013).

$$\mathrm{ARID}_i = 1 - \frac{T_i}{\mathrm{ET}_{o,i}} \tag{4-1}$$

where i denotes ith day, T_i (mm d^{-1}) represents the transpiration on the ith day, and $ET_{o,i}$ (mm d^{-1}) indicates reference evapotranspiration during the ith day. T_i is calculated using a macroscopic modelling method based on soil moisture. $ET_{o,i}$ can be calculated based on the Priestley and Taylor (1972) method because it is assumed to be potential evapotranspiration. Detailed characteristics and calculation processes of the ARID index have been described by Woli et al.

(2012). The values of ARID range from 0 to 1. As values greater than 0.6 are considered to indicate severe water deficit, we selected 0.6 as the threshold to assess drought conditions.

We also included a metrological drought index, Standardized Precipitation and Evapotranspiration Index (SPEI) as a predictor in order to consider the impact of long-term drought on wheat yield. It is usually used to assess drought at monthly, seasonal, or longer time scales, rather than daily as the ARID index does. SPEI characterizes drought through standardizing the difference between precipitation and potential evapotranspiration (Vicente-Serrano et al., 2010). Thus, both precipitation and temperature, the two most relevant factors associated with drought, are considered in the calculation of SPEI. As a standardized index, values around 0 denote normal conditions, while values <-1 and >1 indicate dry and wet conditions, respectively. In addition, the SPEI is developed in a way that can monitor drought at timescales from short (1 month) to long (24 months) periods. 1- to 6-month timescale SPEI values are usually used for meteorological and agricultural drought monitoring, while longer timescale SPEI values are more suitable for hydrological drought monitoring. Thus, in our study, we chose 1-, 3-, and 6-month timescale SPEI as explanatory predictors in the regression models to analyze drought impacts on wheat yield. SPEI is usually calculated based on calendar month. Forecasting events in our study were triggered at the end of each growth category, which might fall on various dates. SPEI values calculated based on calendar month were therefore not likely to exactly cover the previous periods. Therefore, we defined 30 days, 90 days, and 180 days backward from the forecasting date as 1-, 3-, and 6-month timescales respectively, which enabled the SPEI calculation to consider all precipitation and temperature information over a particular timescale. Average NDVI, number of days with heat, frost, or ARID, and SPEI were calculated or counted for a specific wheat growth stage generated from the APSIM simulations.

4.2.3.4. Modelling framework

The overall framework presented in Figure 4-2 shows the order of procedures used in this study for in-season wheat yield forecasting model development and evaluation. APSIM simulated biomass and stage-specific climate and remote sensing information were used as predictors for the MLR and RF models to provide in-season yield forecasts. In detail, we first set up APSIM simulations based on information from GRDC-NVT trials (soil, variety, and management) and ran simulations dynamically according to known climate data. Once APSIM-simulated wheat phenology completed a targeted forecasting stage (S, SG, T, SE, BAF, and M in our study), a forecasting event would be triggered. APSIM simulated biomass, SPEI, stage-specific ECEs, and NDVI, would be extracted or calculated from APSIM output, climate, and remote sensing information, except at S in which only SPEI was available. These factors were used as input for the forecasting models (RF and MLR) to forecast end-of-season yield.

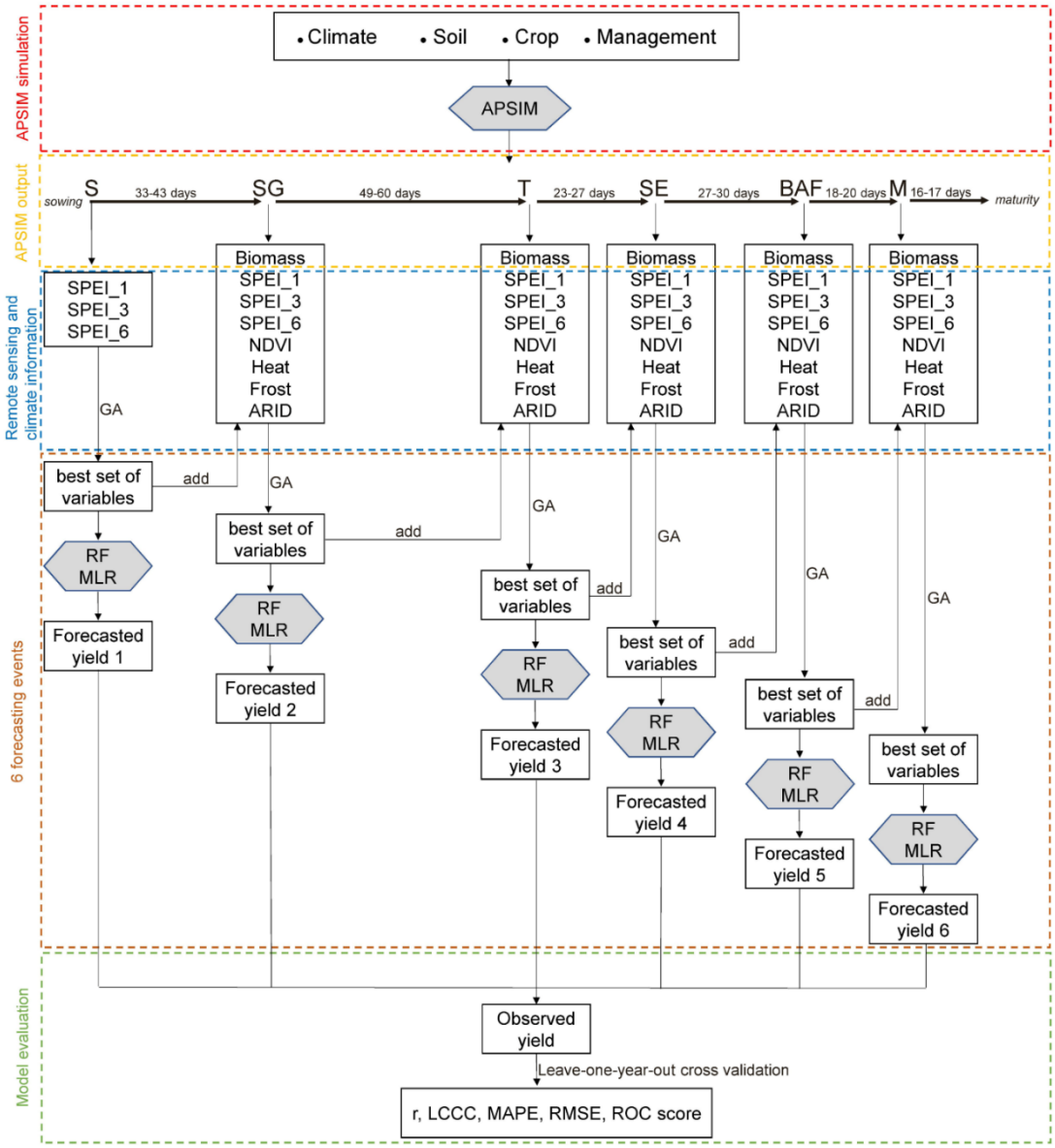


Figure 4-2 Framework for the procedures used in this study, where S: sowing, SG: seedling growth, T: tillering, SE: stem elongation, BAF: booting, awn emergence, and flowering, M: milk development, SPEI_1, SPEI_3, and SPEI_6 represent 1-, 3-, and 6-month timescale Standardized Precipitation and Evapotranspiration Index, ARID: Agricultural Reference Index for Drought, GA: genetic algorithm, MLR: multiple linear regression, RF: random forest, r: Pearson correlation coefficient, LCCC: Lin's concordance correlation coefficient, MAPE: mean absolute percentage error, RMSE: root mean square error, ROC score: receiver operating characteristic score.

As wheat phenology progressed, more and more factors would be gradually added to the forecasting models, which could result in increased computation times and over-fitted model due to correlated factors and the "curse of dimensionality" (Rodriguez-Galiano et al., 2012). Thus, we conducted a feature selection procedure before feeding the forecasting models. In the present study, a nonlinear method, genetic algorithm (GA) (Mitchell, 1998), was used to select the most

relevant factors. GA is a search heuristic for function optimization based on Charles Darwin's theory of natural evolution. For feature selection, a number of subsets from input predictor variables were first sampled as candidate solutions. Their corresponding fitness values (root mean square error, RMSE) were calculated to evaluate the quality of a subset. The subsets with the lowest RMSE were randomly combined to generate offsprings. In this process, two major genetic operators (mutation and crossover) were used, which could substantially affect the fitness value. Mutation takes effect through randomly removing or adding features in a subset. Crossover operates by producing a new subset through combining different features from a pair of subsets. Offsprings replace the old generation based on the criterion that the new generation was better than the old one. This evolutionary process was repeated again and again until the termination of the search procedure. Many generations were then generated, which was likely to create better and better subsets. In the present study, we applied the GA method through the 'caret' package (Kuhn, 2008) coded in R software. Three main parameters were set according to Wang et al. (2018), i.e. mutation rate of 0.1, crossover rate of 0.8, and population size of 50. Selected predictors for each forecasting event based on the GA method are shown in Table 4-2.

Table 4-2 Selected predictors as determined from the genetic algorithm for each forecasting event. S: sowing, SG: seedling growth, T: tillering, SE: stem elongation, BAF: booting, awn emergence, and flowering, M: milk development, SPEI_1, SPEI_3, and SPEI_6: 1-, 3-, and 6-month timescale Standardized Precipitation and Evapotranspiration Index, ARID: Agricultural Reference Index for Drought.

No.	S	SG	T	SE	BAF	M
1	S_SPEI_1	SG_biomass	T_biomass	SE_biomass	BAF_biomass	M_biomass
2	S_SPEI_3	S_SPEI_1	S_SPEI_1	T_SPEI_1	SE_SPEI_6	SE_SPEI_6
3	S_SPEI_6	S_SPEI_3	S_SPEI_6	SE_SPEI_1	BAF_SPEI_1	BAF_SPEI_1
4		S_SPEI_6	SG_SPEI_3	SE_SPEI_3	SG_ ARID	M_SPEI_1
5		SG_SPEI_3	T_SPEI_1	SE_SPEI_6	BAF_ ARID	SG_ ARID
6		SG_ ARID	SG_ ARID	SG_ ARID	BAF_Heat	BAF_ ARID
7		SG_Frost	T_ ARID	T_ ARID	SG_Frost	M_ ARID
8		SG_NDVI	SG_Frost	SE_ ARID	BAF_Frost	BAF_Heat
9				SG_Frost	BAF_NDVI	SG_Frost
10						BAF_Frost

Blending of climate and remote sensing indicators using statistical methods is a common practice in yield forecasting studies. However, most studies calculate indicators at monthly or longer time scale (Cai et al., 2019; Kern et al., 2018), which might be limited in considering stage-specific effects of some indicators. We firstly developed a wheat yield forecasting system which dynamically incorporated growth stage-specific ECEs, SPEI, NDVI, and APSIM-simulated biomass into the RF model. The MLR model was used as a benchmark. Given that crop yields often show nonlinear response to ECEs (Li et al., 2019), RF is expected to show good performance

in exploring nonlinear relationships. The RF model was implemented with the 'randomForest' package (Liaw and Wiener, 2002) based in the R software. The parameters in the RF model were set with $\boldsymbol{m_{try}}$ (the number of randomly selected predictor variables at each node) = the number of predictor variables divided by 3 (rounded down) and $\boldsymbol{n_{tree}}$ (the number of trees to grow in the forest) = 500. The relative importance of variables was estimated using the "%IncMSE" metric in the RF model.

3.2.3.5. Model performance assessment

The 10-year observed wheat yield data from the GRDC-NVT trials were used for comparisons with modeled yields. In doing so, a leave-one-year-out (a whole year of data for all sites was left out) cross validation method was applied to the 231 data sets. Both deterministic and probabilistic measurements were adopted to evaluate model performance. The deterministic measurements included Pearson's correlation coefficient (r), Lin's concordance correlation coefficient (LCCC) (Lin, 1989), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These measurements were defined as follows:

$$r = \frac{\sum_{i=1}^{n}\left(O_i - \bar{O}\right)\left(F_i - \bar{F}\right)}{\sqrt{\sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2}\sqrt{\sum_{i=1}^{n}\left(F_i - \bar{F}\right)^2}} \tag{4-2}$$

$$LCCC = \frac{2r\sigma_O\sigma_F}{\sigma_O^2 + \sigma_F^2 + \left(\bar{O} - \bar{F}\right)^2} \tag{4-3}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{F_i - O_i}{O_i}\right| \tag{4-4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(O_i - F_i\right)^2}{n}} \tag{4-5}$$

where F_i and O_i represent forecasted and observed values, $\bar{F}$ and $\bar{O}$ denote the mean forecasted and observed values, σ_O and σ_F are the variances of observed and forecasted values, n indicates the number of the samples. RMSE and MAPE indicate the average magnitude of the errors in a set of forecasts. r measures the strength of the linear relationship between observations and forecasts. LCCC denotes the degree to which forecasted and observed values follow the 1:1 line through the origin. Model forecasts become increasingly accurate as RMSE and MAPE approach 0 and r and LCCC approach 1.

We also applied a probabilistic forecasting performance measurement: receiver operating characteristic (ROC) score (Fawcett, 2006). The ROC curve was generated by plotting the true positive rate against the false positive rate across various cut-off settings. The ROC score is

defined as the area under the curve, which ranges from 0 to 1. Generally, forecasting models with higher scores are considered more skillful. The ROC analysis is not sensitive to forecasting biases, which is said to be helpful for finding potential skill without considering biases similar to the correlation coefficient (Jha et al., 2019). A detailed description of the ROC analysis can be found in Fawcett (2006). ROC score is a useful measurement of model performance for classification tasks. In the present study, observed yields were first categorized into three terciles: below normal (0% to 32%), normal (33% to 66%), and above normal (67% to 100%). Thus, yield forecast probabilities for each category could be calculated. ROC score was calculated for forecasts of the RF and MLR models for each forecasting event using the R package 'pROC' (Robin et al., 2011).

4.3 Results

4.3.1 Model performance

In this study, APSIM dynamic output, SPEI, stage-specific ECEs, and NDVI were included as predictors into MLR and RF models to make pre-harvest yield forecasts. The performance of both regression models was evaluated by a leave-one-year-out cross validation procedure. Figure 4-3 presents the observed and forecasted yields at the 29 study sites from 2008 to 2017 in the NSW wheat belt. Generally, observed wheat yield varied greatly from year to year in the study area, with a low of 2.5 t ha^{-1} in 2009 and a high of 5 t ha^{-1} in 2016. The temporal variations of the observed wheat yield were successfully captured by both the MLR and RF models, and the accuracy usually improved as growth stage progressed. The RF model tended to better predict observed yields than the MLR model, especially in years with atypical yields, such as 2009 and 2016. Moreover, forecasted yields for each year from the RF model were less variable than from the MLR model over the course of the six growth stage events, meaning that the RF model could provide better forecasts even at earlier growth stages.

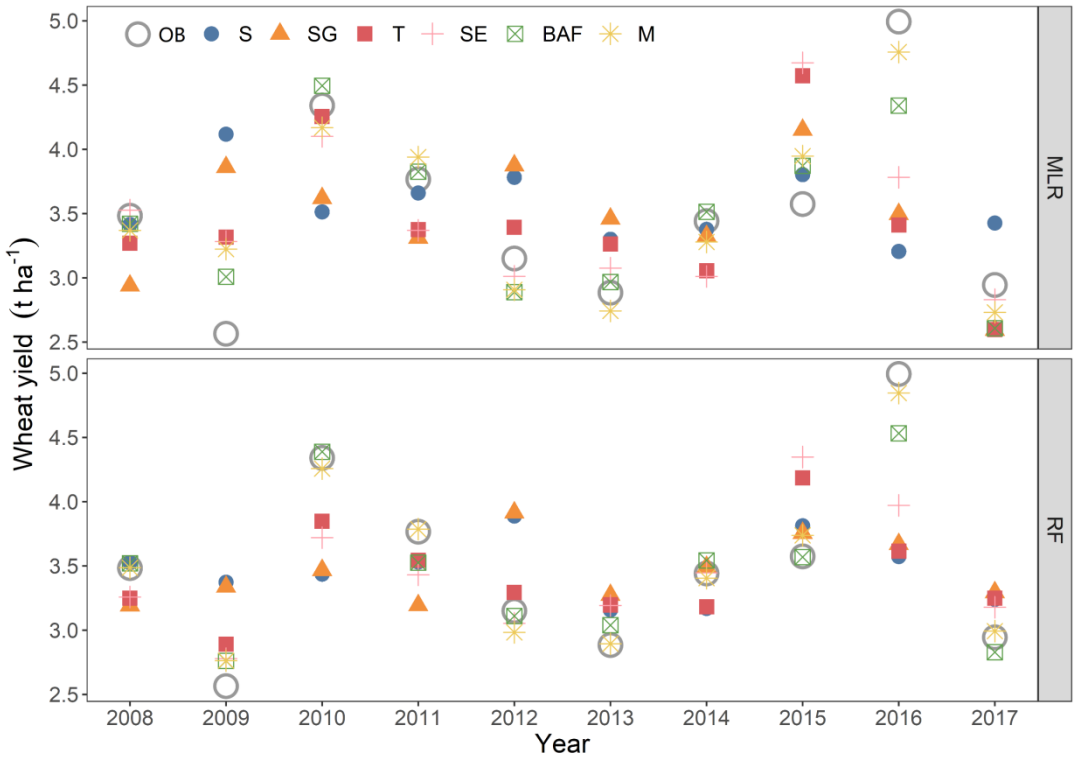


Figure 4-3 Time series of observed and model-forecasted wheat yields based on the six forecasting events from 2008 to 2017. Wheat yields for each year were averaged across the 29 study sites (results for each site can be found in Figure S1 in the supplementary material). Data were generated from the leave-one-year-out cross validation procedure from the two regression models, MLR: multiple linear regression and RF: random forest. Observed and six forecasted wheat yields are shown as gray circles and colored shapes, respectively. OB: observed, S: sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development.

Evaluations of the accuracy of the forecasted yields produced by the two models using the four deterministic metrics (r, LCCC, MAPE, and RMSE) are shown in Figure 4-4. In general, the yield forecast accuracy for both models increased with the progress of wheat growth, as demonstrated by the gradual increase in r and LCCC and decrease in MAPE and RMSE. At the first two forecasting events, S and SG, the two models both showed poor performance in forecasting end-of-season yield, with MAPE values around 40% and RMSE above 1.30 t ha^{-1}. From SG to BAF, the forecasting accuracy increased significantly for both models. For the MLR model, r increased from 0.15 to 0.74, LCCC increased from 0.11 to 0.71, MAPE decreased from 38.4% to 21.3%, and RMSE decreased from 1.33 t ha^{-1} to 0.86 t ha^{-1}. Similarly, for the RF model, r increased from 0.13 to 0.85, LCCC increased from 0.08 to 0.81, MAPE decreased from 38.8% to 17.6%, and RMSE decreased from1.30 t ha^{-1} to 0.70 t ha^{-1}. It should be noted that at each growth stage forecasting event, the RF model always had greater accuracy than the MLR model as denoted by

the four metrics. In addition, forecasting accuracy increased little for both regression models from BAF to M.

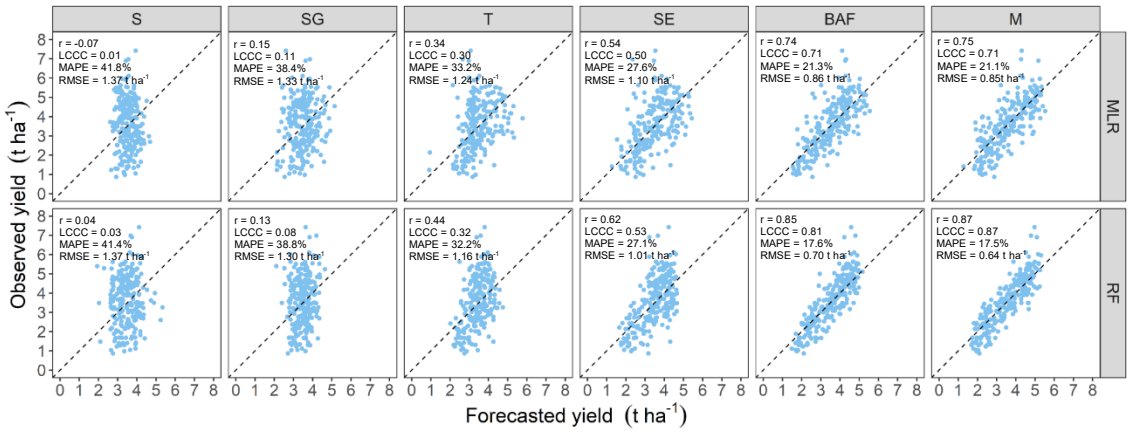


Figure 4-4 Comparison of observed and model-forecasted wheat yields at the six forecasting events (S: sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development) from the two regression models, MLR (multiple linear regression) and RF (random forest). Results of four deterministic metrics of model performance are given in the figure. Dashed lines represent the 1:1 lines.

The evaluation of model performance using the probabilistic measurement, ROC score, is shown in Figure 4-5. The ROC scores of forecasts at the six different forecasting events are consistent with the previously described results from the deterministic measurements, i.e., the ROC scores of forecasts improved as wheat developed. At S and SG, the ROC scores were marginally greater than 0.5, reflecting poor forecasting performance by both models. However, from T onwards, both models had satisfactory forecasting performance, as denoted by ROC scores > 0.6, with ROC scores increasing rapidly from SG to BAF. As with the deterministic measures of model performance, the rate of increase in ROC scores from BAF to M was much slower that at previous growth stages. In general, the RF model had larger ROC scores at all forecasting events. The highest ROC score was 0.9, achieved by the RF model at M.

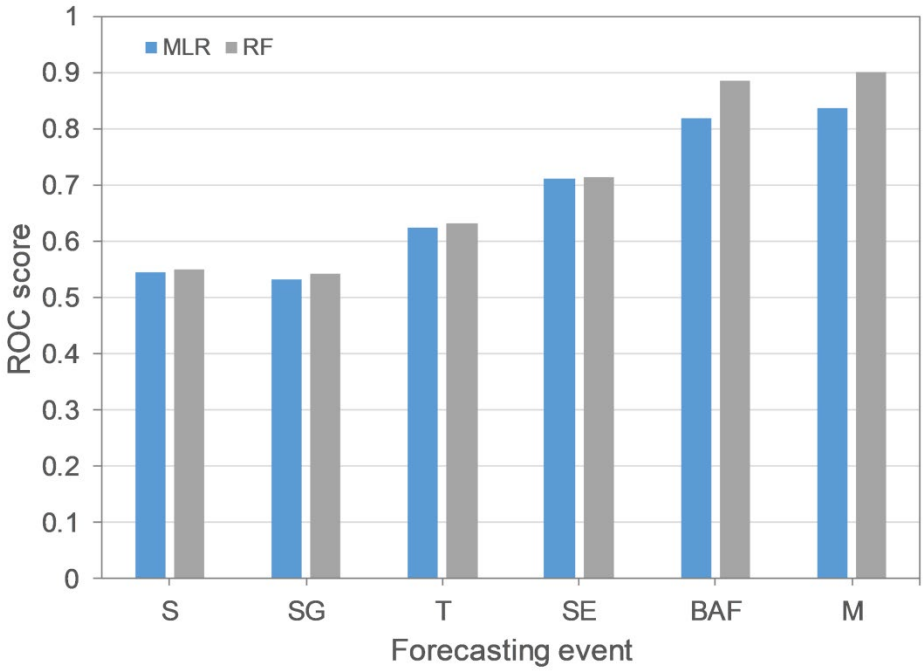


Figure 4-5 Receiver operating characteristic (ROC) scores of observed and model-forecasted wheat yields at the six forecasting events using two regression models, MLR: multiple linear regression and RF: random forecast. S: sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development.

4.3.2 Optimum forecasting event analysis

According to the model performance results, the RF-based forecasting system showed better performance than the MLR-based model for our study area. In general, stakeholders prefer to obtain an accurate yield forecast as early as possible. However, there is a tradeoff between greater accuracy and longer lead time in any yield forecasting system (Basso and Liu, 2018). In our study, model accuracy gradually increased as the growing season progressed towards harvest, but slowed down from BAF to M (Figures 4-4 and 4-5). To determine the optimum forecasting event with respect to the stakeholders' objective, we analyzed the relative changes in model accuracy as growing season progressed. This was done by normalizing the four model performance measurements from 0 to 100% by the following equation

$$x_{new} = \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) * 100\% \qquad (4\text{-}6)$$

where x represents either r, LCCC, MAPE, RMSE, or ROC. This normalization procedure allowed a comparison of the magnitude of change for each performance measurement at each of the six forecasting events. As shown in Figure 6, r_n, $LCCC_n$, and ROC_n score continued to increase from S to M (except for ROC score from S to SG), and $MAPE_n$ and $RMSE_n$ continued to decrease from S to M. The greatest increase in model accuracy occurred between SE and BAF, in which r_n

increased by 24.4%, $LCCC_n$ increased by 33.3%, ROC_n score increased by 47.6%, $MAPE_n$ decreased by 30.4%, and $RMSE_n$ decreased by 46.2%. However, from BAF to M, model accuracy increased only slightly. Thus, BAF can be viewed as the appropriate event providing the most satisfactory forecast, with about a 35-day lead time (Figure 4-2). Forecasting yield at SE or T would improve the lead time to 2 or 3 months, but the model accuracy would greatly decrease.

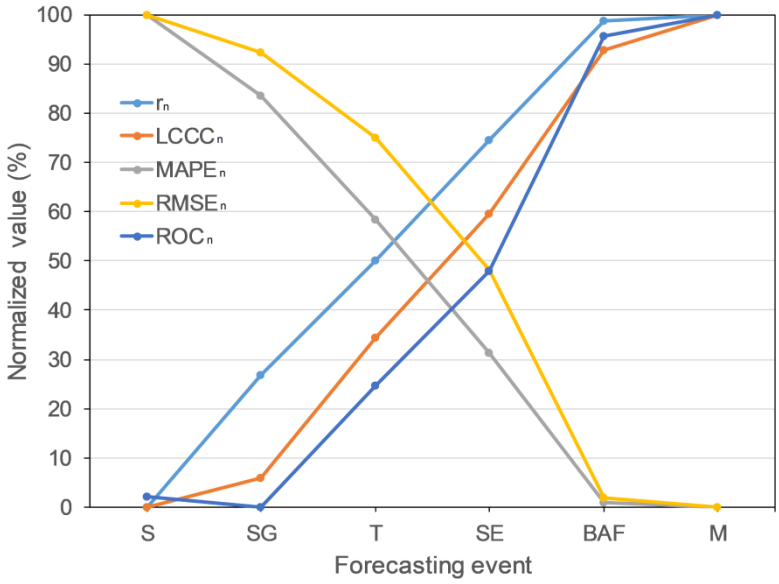


Figure 4-6 Normalized values of the four model performance measurements (r, LCCC, MAPE, RMSE, and ROC score) at the six growth stage forecasting events using the RF-based forecasting model to predict wheat yield at 29 study sites in New South Wales, Australia (2008-2017). S: sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development.

4.3.3 Relative importance of growth selected predictors

The RF model can provide a list of the relative importance of predictors based on each predictor's impact on model accuracy (Were et al., 2015). According to the ranking results (Figure 4-7) for each forecasting event (except for S), there was an increasing trend of the importance values of APSIM-simulated biomass from SG to BAF. APSIM-simulated biomass ranked first in the last three forecasting events. This was expected as biomass is cumulated with crop growth and development. While for other indices, drought was the most influential factor affecting wheat yields in the study area, as the first three indices were all drought related indices at each forecasting event. At the first four forecasting events, SPEI ranked first, but at BAF and M, stage-specific drought (indicated by the ARID index) exceeded SPEI. This may because wheat yield was overly sensitive to daily time scale drought events at BAF and M. Nevertheless, SPEI is a potential index to reflect drought-induced yield losses, as SPEI was involved in all forecasting events and usually ranked high (Figure 4-7). ARID and frost events that occurred during SG were

two indices that were consistently selected at the various forecasting events, meaning that ARID and frost from sowing to the end of seedling growth are likely to greatly affect final wheat yields. Heat and frost events during BAF also had great impacts on final yield and were both selected at the BAF and M forecasting events. In contrast, NDVI was relatively unimportant, only being selected at the SG and BAF forecasting events and was not highly ranked at either time.

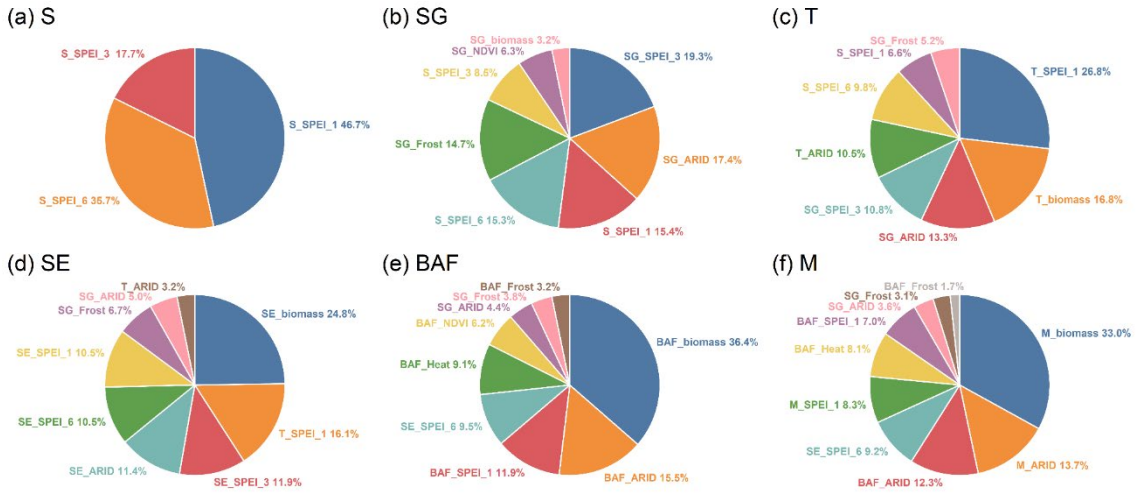


Figure 4-7 Relative importance of selected predictors as determined from the RF (random forest) model for each forecasting event. The results are normalized to sum to 100% and shown in decreasing order in the clockwise direction. S: sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development.

4.4 Discussion

In this study, we combined a crop simulation model (APSIM) with statistical regression-based models to dynamically forecast wheat yield at several points during the growing season based on growth stage-specific climate and remote sensing indices. The APSIM+RF hybrid model obtained satisfactory results in yield prediction. This was primarily because we succeeded in exploiting the merits of each model. Our models not only took advantage of biophysical processes among crop, soil, management, and climate information, but also made use of machine learning technique to account for climate extremes and remote sensing information. In addition, the advanced machine learning technique used in the study showed an overall advantage over traditional regression methods (Figures 4-4 and 4-5) in exploring the relationships between crop yield and environmental factors. Given the increasing availability of farming-related, climate, and remote sensing data, Keating and Thorburn (2018) introduced the blending of advanced statistical and mechanistic models within crop-environment research fields. The yield forecasting system proposed in our study can be regarded as a feasible approach capable of being extended to other wheat cropping areas in order to gain new insights that will guide agricultural practice and grain marketing.

Another advantage of our RF-based wheat yield forecasting system is that it accounts for wheat growth stage-specific ECEs. Generally, crop growth for a given season is subjected to two types of climatic conditions, i.e., mean climate conditions and climate extremes (Challinor et al., 2007). More mean climate conditions tend to result in high harvestable yield, while climate extremes generally lead to yield loss. Most crop models adequately simulate the effects of mean climate conditions but encounter limitations when estimating yield losses due to climate extremes (Barlow et al., 2015). While statistical regression-based models are able to determine relationships between yield and variables quantifying climate extremes, the variables selected are usually vague and not stage-specific. The majority of previous research using regression models has extracted variables based on long time periods, typically covering the entire growing season (Pinke and Lövei, 2017; Wang et al., 2015). This approach may result in inaccurate estimations of yield losses as different crop growth stages can have different tolerances for the same extreme event. For example, Baigorria et al. (2007) demonstrated that the timing and duration of dry periods had different impacts on final maize (*Zea mays* L.) yield. Kern et al. (2018) emphasized the importance of shorter timescales for variable calculation and prepared predictors at monthly resolution, which explained 67% of the variation in winter wheat yields. However, they still did not associate the predictors with concrete growth stages such as anthesis or grain filling. In our study, we acquired crop phenology information by dynamically running APSIM simulations that triggered forecasting events at the end of specific growth stages. Indices of climate extremes were then calculated according to past growth stages. The APSIM-simulated biomass can be viewed as representative of the mean climate condition. Thus, our model considered the effects of mean climate conditions and stage-specific climate extremes simultaneously, thereby resulting in satisfactory yield forecasting results.

In our study, the crop physiological interpretation was easily understood as the yield predictors were associated with concrete growth stages. For these growth stage-specific indices, our results identified biomass, which integrated drought effects, as the most important factor determining yield in the study area. This result was expected as drought is a recurring feature of Australia's climate (Ummenhofer et al., 2009), and has caused severe yield losses during past decades (Feng et al., 2018). In the present study, we found that dry periods during seedling growth are likely to result in carryover effects on wheat growth, while dry periods during anthesis and grain filling are major factors determining final yield. This is because insufficient water supply can restrain leaf expansion and root growth during vegetative growth stages (Chaves et al., 2002), and can reduce mobilization of carbohydrates from vegetative organs to grain during reproductive growth stages (Royo et al., 2006). Meanwhile, SPEI was identified as a potential drought index reflecting the effects of water deficiency on crop yield in this study area as it was always chosen by the RF model as a variable having high importance. Frost and heat events during reproductive growth

stages and frost events during seedling growth also had important impacts on final yield. This is because these stages are more sensitive to temperature anomalies (Hlaváčová et al., 2018). In contrast, the remotely sensed NDVI did not significantly improve model accuracy and was only selected as important by the RF model at the SG and BAF forecasting events, and even at those two stages, NDVI was not ranked high as an important variable influencing wheat yield (Figure 4-7). This might be due to the plot-scale data used in the study. NDVI values with 500 m spatial resolution were too large to reflect the vegetation conditions of specific experimental plots. When applying this system for yield forecasts at larger scale, remote sensing information is likely to contribute more to model accuracy.

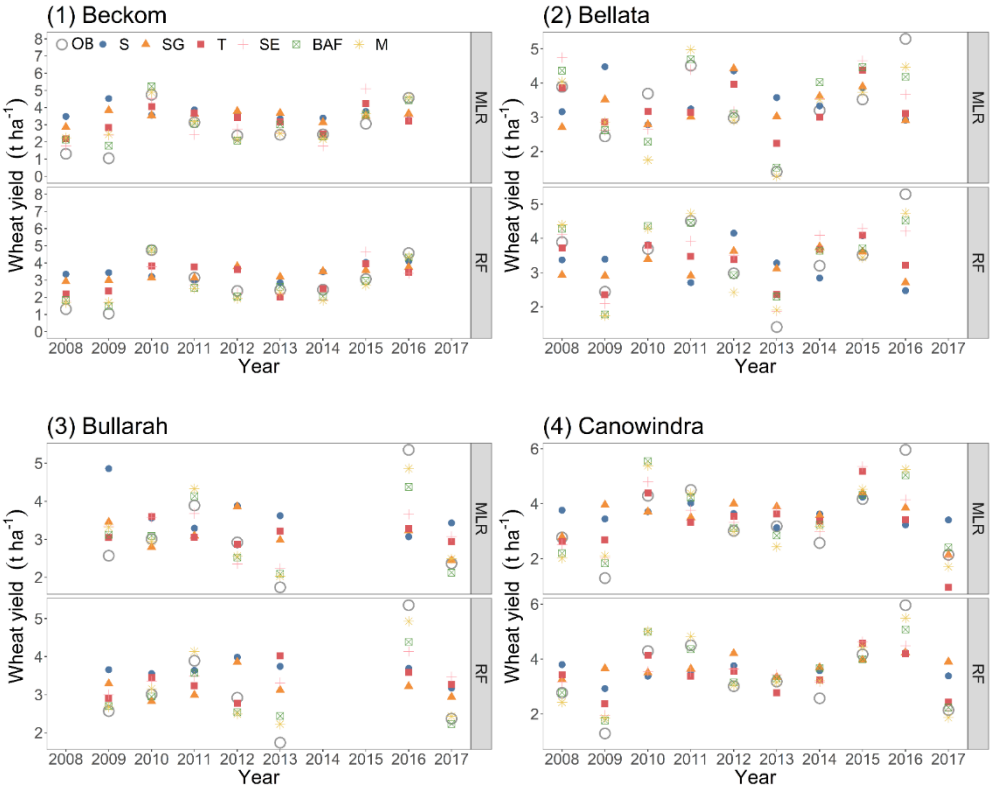
Our proposed yield forecasting system has great potential for practical agricultural production systems. In the present study, we achieved satisfactory results for plot-level wheat yield forecasting with a 35-day lead time (r=0.85, LCCC=0.81, MAPE=17.6%, RMSE=0.70 t ha^{-1}, and ROC score=0.90) and with a two-month lead time (r=0.62, LCCC=0.53, MAPE=27.1%, RMSE=1.01 t ha^{-1}, and ROC score=0.88). In comparison with other yield forecasting studies conducted in Australia, Cai et al. (2019) obtained a result of R^2=0.73 with a 2-month lead time for wheat yield forecasting in Australia. However, region-level wheat yield records (rather than plot-level wheat yields as used in our study) were used in their study. Wheat yields at a smaller scale are more difficult to forecast due to the variable conditions even within the same region, which tends to require more kinds of data sources with finer resolution. Further development of data acquisition techniques will allow the acquisition of more detailed farming-related data (Filipi et al. 2019), such as agronomic information, soil moisture conditions (Peng et al., 2017), and high spatiotemporal remote sensing images for a growing season (Zambrano et al., 2018). Our proposed method can also be extended to and validated for larger areas to determine crop yield outlooks.

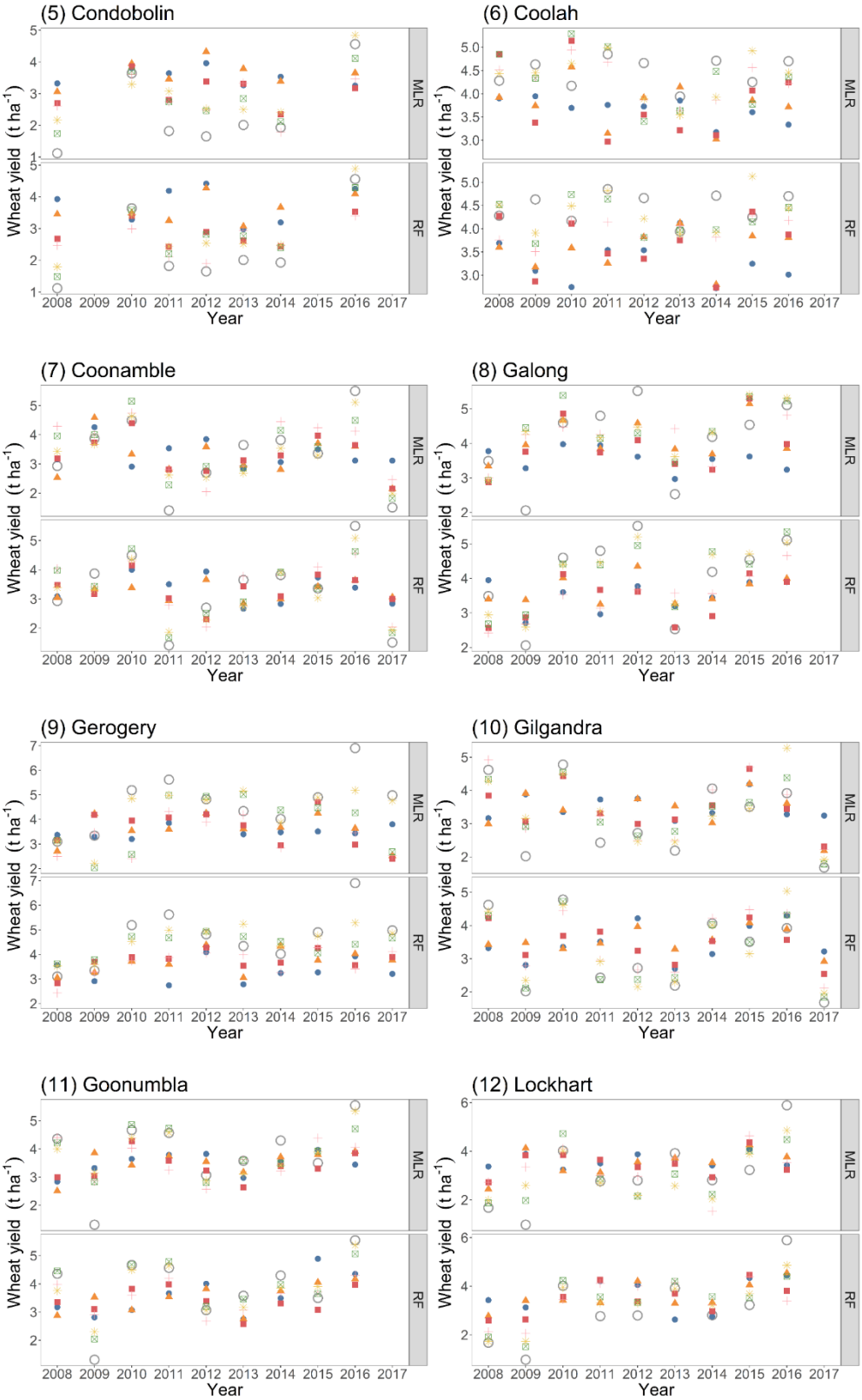
Our forecasting system satisfactorily predicted yield in the NSW wheat belt. However, this method required a large amount of data from different sources, including soil, climate, crop, management, and remote sensing information. Thus, it might not be suitable for data-poor areas. Nevertheless, with societal and economic developments, more and more areas will have sufficient data available to implement this yield forecasting method. In addition, El Nino Southern Oscillation (ENSO)-related indices are also frequently used indicators for yield forecasting in Australia, but were not considered in our study. Future studies using a similar modelling approach to ours may use additional information (such as ENSO-related indices) and potentially achieve even greater forecasting accuracy.

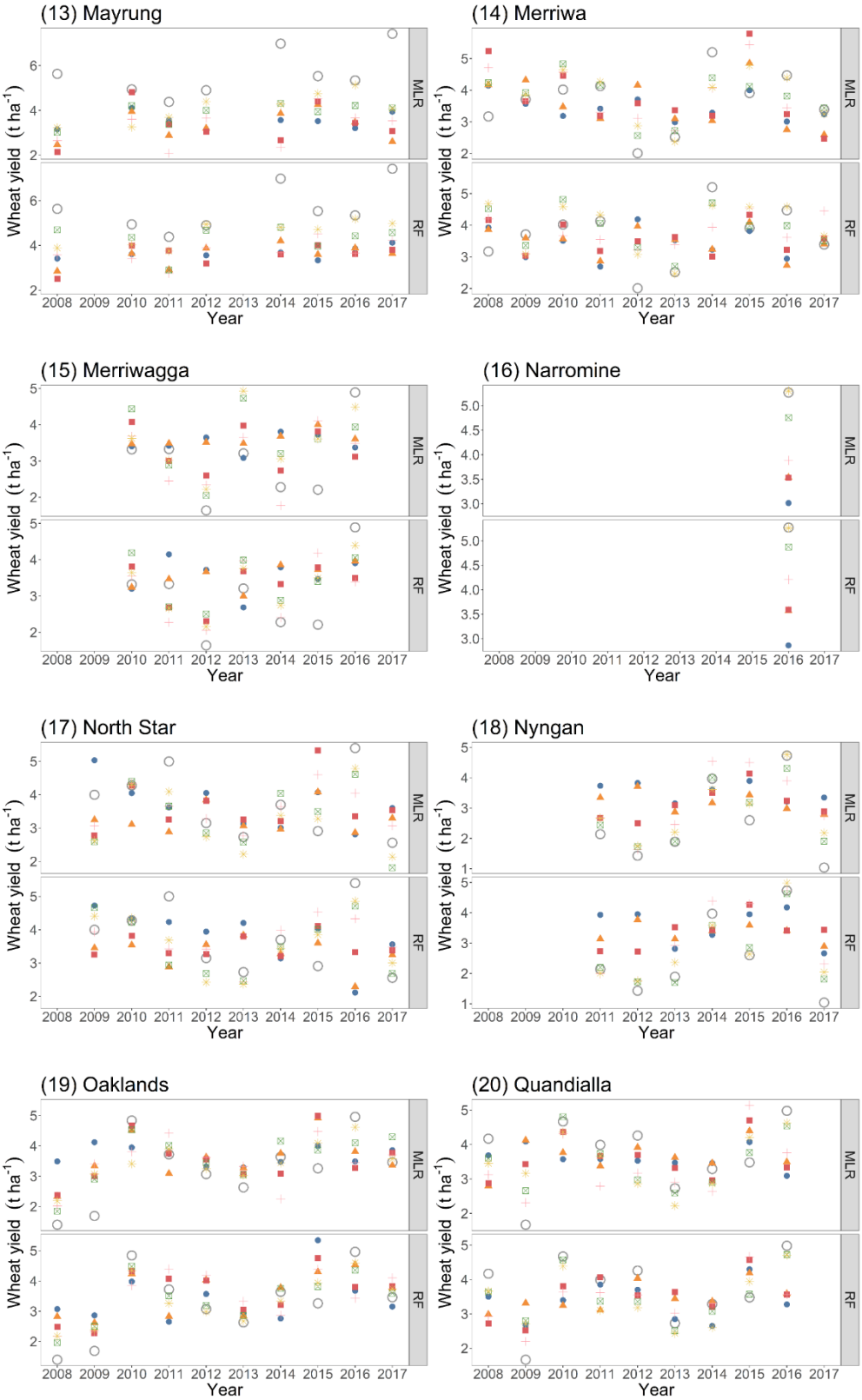
4.5 Conclusions

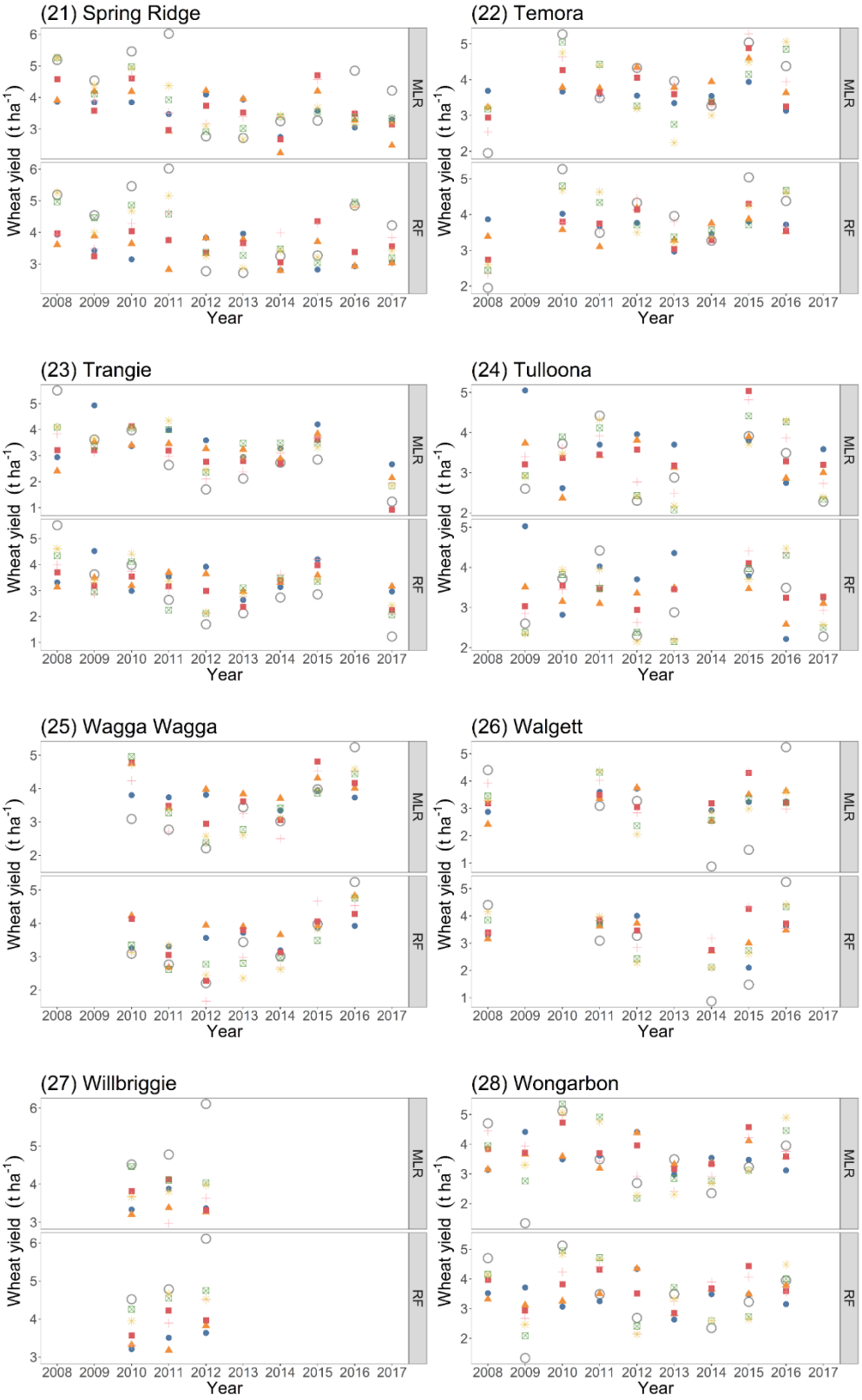
In the present study, we succeeded in developing a wheat yield forecasting system by incorporating multiple data sources such as crop model output, indices of extreme climate, and remote sensing information into two statistical regression-based models. Plot-level wheat yields were dynamically forecasted at the end of targeted growth stages during the growing season progressing to harvest. Stage-specific extreme climate events were fully considered in the system. We found that the machine learning-based model produced more accurate forecasts of wheat yield than the traditional multiple linear regression model. The optimum forecasting events that produced sufficiently accurate yield predictions were those providing one- and two-month lead times. We expect that this forecasting system using crop simulation modelling and a machine learning method specifically addressing the effects of stage-specific climate extremes on crop yield can be used for operational forecasting purposes in Australia and potentially other similar dryland cropping systems around the world. With the further development of information technology and remote sensing technology, our proposed system can be directly extended to region- and country-scale forecasts. This yield-forecasting method will become increasingly important in providing information to mitigate the detrimental effects of climate change on global food supply.

4.6 Supporting information




(5) Condobolin
(6) Coolah
(7) Coonamble
(8) Galong
(9) Gerogery
(10) Gilgandra
(11) Goonumbla
(12) Lockhart
Wheat yield (t ha-1)
Year
MLR
RF
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017


(13) Mayrung
(14) Merriwa
(15) Merriwagga
(16) Narromine
(17) North Star
(18) Nyngan
(19) Oaklands
(20) Quandialla
Wheat yield (t ha-1)
Year
MLR
RF
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017


(21) Spring Ridge
(22) Temora
(23) Trangie
(24) Tulloona
(25) Wagga Wagga
(26) Walgett
(27) Willbriggie
(28) Wongarbon
Wheat yield (t ha-1)
Year
MLR
RF
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017

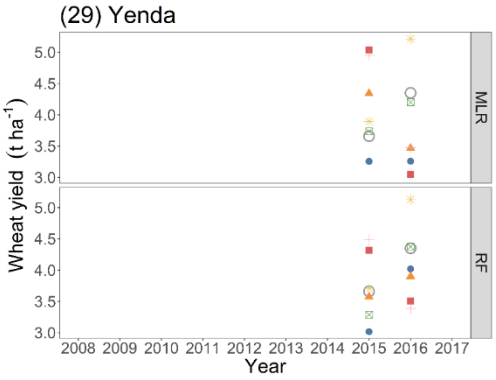


Figure 4-S1 Time series of observed and model-forecasted wheat yields based on the six forecasting events from 2008 to 2017 for all study sites. Data were generated from the leave-one-year-out cross validation procedure from the two regression models, MLR: multiple linear regression and RF: random forest. Observed and six forecasted wheat yields are shown as grey circles and colorful shapes, where OB: observed, S : sowing, SG: end of seedling growth, T: end of tillering, SE: end of stem elongation, BAF: end of flowering, and M: end of milk development.

4.7 Reference

Baigorria, G.A., Jones, J.W., Shin, D.-W., Mishra, A. and O'Brien, J.J., 2007. Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs. Climate Research, 34(3): 211-222.

Balaghi, R., Tychon, B., Eerens, H. and Jlibene, M., 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. International Journal of Applied Earth Observation and Geoinformation, 10(4): 438-452.

Barlow, K.M., Christy, B.P., O'Leary, G.J., Riffkin, P.A. and Nuttall, J.G., 2015. Simulating the impact of extreme heat and frost events on wheat crop production: A review. Field Crops Research, 171: 109-119.

Basso, B. and Liu, L., 2018. Seasonal crop yield forecast: Methods, applications, and accuracies. Advances in Agronomy, Chapter Four, pp.201-255.

Bolton, D.K. and Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agr Forest Meteorol, 173: 74-84.

Breiman, L., 2001. Random Forest. Machine Learning, 45: 5-32.

Brown, H., Huth, N. and Holzworth, D., 2018. Crop model improvement in APSIM: using wheat as a case study. European Journal of Agronomy, 100: 141-150.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L. and Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology, 274, pp.144-159.

Challinor, A., Wheeler, T., Craufurd, P., Ferro, C. and Stephenson, D., 2007. Adaptation of crops to climate change through genotypic responses to mean and extreme temperatures. Agriculture, ecosystems & environment, 119(1-2): 190-204.

Chaves, M.M., Pereira, J.S., Maroco, J., Rodrigues, M.L., Ricardo, C.P.P., Osório, M.L., Carvalho, I., Faria, T. and Pinheiro, C., 2002. How plants cope with water stress in the field? Photosynthesis and growth. Annals of botany, 89(7), pp.907-916.

Chipanshi, A., Zhang, Y., Kouadio, L., Newlands, N., Davidson, A., Hill, H., Warren, R., Qian, B., Daneshfar, B., Bedard, F. and Reichert, G., 2015. Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape. Agricultural and Forest Meteorology, 206, pp.137-150.

Dalgliesh, N., Wockner, G. and Peake, A., 2006. Delivering soil water information to growers and consultants, Proceedings of the 13th Australian Agronomy Conference. Australian Society of Agronomy Perth, Western Australia, pp. 10-14.

Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rötter, R., Kersebaum, K.C. and Olesen, J.E., 2013. Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. The Journal of Agricultural Science, 151(6), pp.813-835.

Everingham, Y., Sexton, J., Skocaj, D. and Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron Sustain Dev, 36(2).

Fawcett, T., 2006. An introduction to ROC analysis. Pattern recognition letters, 27(8): 861-874.

Feng, P., Wang, B., D.L. Liu, Xing, H., Ji, F., Macadam, I., Ruan, H. and Yu, Q., 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic change, 147(3-4), pp.555-569.

Feng, P., Wang, B., D.L. Liu, and Yu, Q., 2019a. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, pp.303-316.

Feng, P., Wang, B., D.L. Liu, Waters, C. and Yu, Q., 2019b. 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, pp.100-113.

Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D., Pozza, L.E., Ugbaje, S.U., Jephcott, T.G., Paterson, S.E., Whelan, B.M. and Bishop, T.F., 2019. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture, pp.1-15.

Hlaváčová, M., Klem, K., Rapantová, B., Novotná, K., Urban, O., Hlavinka, P., Smutná, P., Horáková, V., Škarpa, P., Pohanková, E. and Wimmerová, M., 2018. Interactive effects of high temperature and drought stress during stem elongation, anthesis and early grain filling on the yield formation and photosynthesis of winter wheat. Field crops research, 221, pp.182-195.

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D. and Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. Agricultural and Forest Meteorology, 204, pp.106-121.

Innes, P.J., Tan, D.K.Y., Van Ogtrop, F. and Amthor, J.S., 2015. Effects of high-temperature episodes on wheat yields in New South Wales, Australia. Agricultural and Forest Meteorology, 208, pp.95-107.

Jha, P.K., Athanasiadis, P., Gualdi, S., Trabucco, A., Mereu, V., Shelia, V. and Hoogenboom, G., 2019. Using daily data from seasonal forecasts in dynamic crop models for yield prediction: A case study for rice in Nepal's Terai. Agricultural and Forest Meteorology, 265, pp.349-358.

Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A. and Bédard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agr Forest Meteorol, 218: 74-84.

Keating, B.A. and Thorburn, P.J., 2018. Modelling crops and cropping systems—evolving purpose, practice and prospects. European journal of agronomy, 100, pp.163-176.

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P. and Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agricultural and forest meteorology, 260, pp.300-320.

Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O. and Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. International Journal of Applied Earth Observation and Geoinformation, 23, pp.192-203.

Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of statistical software, 28(5): 1-26.

Kussul, N., Lavreniuk, M., Skakun, S. and Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 14(5), pp.778-782.

Lawrence, I. and Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, pp.255-268.

Li, Y., Guan, K., Yu, A., Peng, B., Zhao, L., Li, B. and Peng, J., 2019. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US. Field Crops Research, 234, pp.55-65.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3): 18-22.

Lobell, D.B. and Asseng, S., 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environmental Research Letters, 12(1): 015001.

Mathieu, J.A. and Aires, F., 2018. Assessment of the agro-climatic indices to improve crop yield forecasting. Agricultural and forest meteorology, 253, pp.15-30.

Mitchell, M., 1998. An introduction to genetic algorithms. MIT press.

Murphy, B.F. and Timbal, B., 2008. A review of recent climate variability and climate change in southeastern Australia. International Journal of Climatology, 28(7): 859-879.

Nandram, B., Berg, E. and Barboza, W., 2014. A hierarchical Bayesian model for forecasting state-level corn yield. Environmental and ecological statistics, 21(3), pp.507-530.

Nielsen, D.C., Halvorson, A.D. and Vigil, M.F., 2010. Critical precipitation period for dryland maize production. Field crops research, 118(3): 259-263.

Pagani, V., Guarneri, T., Fumagalli, D., Movedi, E., Testi, L., Klein, T., Calanca, P., Villalobos, F., Lopez-Bernal, A., Niemeyer, S. and Bellocchi, G., 2017. Improving cereal yield forecasts in Europe–The impact of weather extremes. European journal of agronomy, 89, pp.97-106.

Peng, J., Loew, A., Merlin, O. and Verhoest, N.E., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Reviews of Geophysics, 55(2), pp.341-366.

Pinke, Z. and Lövei, G.L., 2017. Increasing temperature cuts back crop yields in Hungary over the last 90 years. Global change biology, 23(12): 5426-5435.

Priestley, C. and Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon Weather Rev, 100(2): 81-92.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics, 12(1), p.77.

Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M. and Jeganathan, C., 2012. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens Environ, 121: 93-107.

Royo, C., Villegas, D., Rharrabti, Y., Blanco, R., Martos, V. and García del Moral, L., 2006. Grain growth and yield formation of durum wheat grown at contrasting latitudes and water regimes in a Mediterranean environment. Cereal Research Communications, 34(2-3), pp.1021-1028.

Stratonovitch, P. and Semenov, M.A., 2015. Heat tolerance around flowering in wheat identified as a key trait for increased yield potential in Europe under climate change. Journal of experimental botany, 66(12): 3599-3609.

Tashiro, T. and Wardlaw, I.F., 1989. A comparison of the effect of high temperature on grain development in wheat and rice. Ann Bot-London, 64(1): 59-65.

Ummenhofer, C.C., England, M.H., McIntosh, P.C., Meyers, G.A., Pook, M.J., Risbey, J.S., Gupta, A.S. and Taschetto, A.S., 2009. What causes southeast Australia's worst droughts?. Geophysical Research Letters, 36(4).

Vicente-Serrano, S.M., Beguería, S. and López-Moreno, J.I., 2010. A multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index - SPEI. Journal of Climate, 23(7): 1696-1718.

Vossen, P., Rijks D., 1995. Early crop yield assessment of the E.C. countries: the system implemented by the Joint Research Centre. EUR Publication of the Office for Official Publications of the E.C. Luxembourg, pp.120.

Wang, B., Chen, C., D.L. Liu, Asseng, S., Yu, Q. and Yang, X., 2015. Effects of climate trends and variability on wheat yield variability in eastern Australia. Climate Research, 64(2), pp.173-186.

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., D.L. Liu, Simpson, M., McGowen, I. and Sides, T., 2018. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecological indicators, 88, pp.425-438.

Were, K., Bui, D.T., Dick, O.B. and Singh, B.R., 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. Ecol Indic, 52: 394-403.

Western, A.W., Dassanayake, K.B., Perera, K.C., Argent, R.M., Alves, O., Young, G. and Ryu, D., 2018. An evaluation of a methodology for seasonal soil water forecasting for Australian dry land cropping systems. Agricultural and Forest Meteorology, 253, pp.161-175.

Woli, P., Jones, J.W. and Ingram, K.T., 2013. Assessing the Agricultural Reference Index for Drought (ARID) using uncertainty and sensitivity analyses. Agronomy Journal, 105(1): 150-160.

Woli, P., Jones, J.W., Ingram, K.T. and Fraisse, C.W., 2012. Agricultural reference index for drought (ARID). Agronomy journal, 104(2): 287-300.

Zambrano, F., Vrieling, A., Nelson, A., Meroni, M. and Tadesse, T., 2018. Prediction of drought-induced reduction of agricultural productivity in Chile from MODIS, rainfall estimates, and climate oscillation indices. Remote Sens Environ, 219: 15-30.

Zheng, B., Chenu, K., Doherty, A. and Chapman, S., 2014. The APSIM-Wheat Module (7.5 R3008), http://www.apsim.info/Portals/0/Documentation/Crops/WheatDocumentation.pdf (accessed May 2018).

Zheng, B., Chenu, K., Fernanda, D.M. and Chapman, S.C., 2012. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivium) varieties? Global Change Biology, 18(9): 2899-914.

Chapter 5. Projected changes in drought across the wheat belt of southeastern Australia using a downscaled climate ensemble

This chapter is based on the following manuscript:

Feng, P., Liu, D.L., Wang, B., Waters, C., Zhang, M. and Yu, Q., 2019. Projected changes in drought across the wheat belt of southeastern Australia using a downscaled climate ensemble. International Journal of Climatology, 39(2), pp.1041-1053.

Abstract

Drought is viewed as a naturally recurring phenomenon in many Australian agricultural systems. Identifying regional changes in frequency and severity of drought induced by climate change is required to develop regionally specific adaptation strategies. In this study, we provided a first look at the impacts of climate change on 21st century drought characteristics over the New South Wales wheat belt of south-eastern Australia. These impacts were assessed from an ensemble of 28 statistical downscaled global climate models under Representative Concentration Pathway 8.5. A modified relative standardized precipitation and evapotranspiration index (rSPEI) at the seasonal scale (3 months) was used to analyse temporal and spatial changes in drought. Results indicated that there was a tendency towards more frequent and severe winter-spring droughts over the study area. Moreover, winter-spring drought prone areas were expected to expand from west to east. Until the end of the 21st century, more than half the wheat belt would be vulnerable to winter-spring drought. The combined effects of reduced precipitation and increased temperature during future winter and spring seasons were the main reasons causing these changes of drought. In addition, summer and autumn droughts would have both slight temporal and spatial changes across the study region. This study also revealed that traditionally dry areas would likely experience an increased frequency of drought compared to wetter areas when subjected to a same increase in temperature or decrease in precipitation. Furthermore, the western part of the wheat belt might be unsuitable for winter crops in the future, or at least exposed to an increased risk of variable yield and would require a gradual transformation which might include summer crops or pastures. Investments in cropping land should be focused on the east part of the wheat belt to achieve more consistent financial returns.

Key words: Drought; Climate change; rSPEI; Spatiotemporal variations; South-eastern Australia

5.1 Introduction

Drought is a temporal and recurrent phenomenon, which originates from prolonged absence, shortage or unusual distribution of precipitation compared to the normal pattern. The occurrences

of above-average temperature which lead to increased evaporation can also inevitably aggravate drought occurrence. Given a high level of confidence that climate change will lead to increased temperature and changed precipitation pattern (Field, 2012), drought conditions are likely to change greatly in many parts of the world (Ahmadalipour et al., 2016; Spinoni et al., 2017). However, as drought is often a period- and region-specific disaster (Wilhite, 1993), it is difficult to identify the occurrence and the severity of a drought event based on simple and fixed standards of precipitation or temperature anomalies in a particular region (Morid et al., 2007; Trinh et al., 2017). Therefore, there is an urgent requirement to develop appropriate methods to identify expected changes of drought conditions at regional scales, which is critical for land managers to develop mitigation and adaptation strategies.

The Coupled Model Intercomparison Project phase 5 (CMIP5, https://cmip.llnl.gov/cmip5/) is a powerful tool to analyse the projections of 21st century climate change. A number of physical-based global climate models (GCMs) are available for obtaining future drought projections based on the assumptions of the future economic development or associated greenhouse gas (GHG) emissions. As the real climate system is immensely complex, no single model is capable of describing its overall process adequately even in a particular region (Tebaldi and Knutti, 2007). Recent studies tend to use multiple GCMs to assess future drought conditions. For example, Dai (2012) managed to project worldwide drought conditions until the end of 21st century based on 14 GCMs under representative concentration pathways 4.5 (RCP4.5). The results of Dai (2012) indicated severe droughts in the next few decades over many mid-latitude areas such as the eastern USA, Europe and Australia, because of either increased evaporation and/or decreased precipitation. Ahmadalipour et al. (2016) used 21 CMIP5 GCMs to assess drought projections and revealed a significant increase in frequency and intensity of future summer droughts across the United States under RCP8.5. Kirono et al. (2011) demonstrated that there is a likely risk (more than 66% probability) of at least doubling drought frequency and an increase in drought affected areas in south-eastern Australia by 2070 based on projections from 14 CMIP3 GCMs under SRES-A1B and A2 emission scenario. The use of multiple GCMs is considered to reduce model uncertainties and provide more reliable future projections (Mpelasoka et al., 2018).

The importance of applying appropriate indices for drought assessment has been addressed in a number of studies (Heim, 2002; Mishra and Singh, 2010). Evaluating drought characteristics systematically and comprehensively at regional scales may be problematic using a series of values. In the past few decades, researchers have managed to develop numerous drought indices by integrating climate factors including precipitation, temperature, and evapotranspiration into a single value. The most widely used drought indices include the palmer drought severity index (PDSI), the self-calibrating PDSI (Wells et al., 2004), the moisture anomaly index (Z-index) (Palmer, 1965), and the standardized precipitation index (SPI) (Thomas et al., 1993). Generally,

these indices improve the identification of the onset of a drought event as well as the measurement of drought severity, which then allows an assessment of spatial and temporal features of drought in various areas. However, most drought indices have a fixed temporal range. For example, PDSI can only capture droughts on time scales of more than 9 months (Guttman, 1998; Lloyd-Hughes and Saunders, 2002), and it cannot be used to detect dry periods at shorter time scales. The SPI is designed in a way that can identify droughts at time scales from small (to one month) up large (to 72 month) periods. However, the SPI is calculated using precipitation alone, and it fails to take into account the important contribution of temperature via evaporation (Nicholls, 2004). Another drought index, the standardized precipitation evapotranspiration index (SPEI) (Vicente-Serrano et al., 2010), has been developed to overcome this shortcoming of the SPI. The SPEI is calculated based on the difference between precipitation and potential evapotranspiration (PET), thereby accounting for energy balance and temperature changes as well as precipitation, but also retaining positive traits of the SPI. While the SPEI is a recently developed drought index, it has been widely used and received considerable attention in numerous studies to analyse drought condition (e.g. Gao et al., 2017; Wang et al., 2015). However, the SPEI still has a shortcoming. For example, a given amount of precipitation and PET at a wet station that produces a negative SPEI may produce a positive SPEI at a dry station. In another word, dryness and wetness are relative to the local historical average rather than the absolute difference between precipitation and PET at a certain station. Therefore, the SPEI has limitations for spatial comparison. In this study, we introduced a relative SPEI (rSPEI) which is based on regional average rather than local conditions, to improve the performance of the original SPEI.

Australia is the driest inhabited continent in the world and drought is an expected feature of the Australian climate (Ummenhofer et al., 2009). Drought causes large agricultural losses in Australia. For example, in south-eastern Australia, drought reduced the agricultural Gross National Product by around 30% in 1994, 2002 and 2006 (Kirono et al., 2011). It is likely that climate change will further exaggerate drought impacts in this region (BOM and CSIRO, 2016). The main objectives of this study are to (1) use rSPEI as an indicator of drought to examine the spatial and temporal characteristics of future drought occurrence across major farming regions of south-eastern Australia until the end of 21st century from an ensemble of 28 CMIP5 GCMs; (2) identify the major climatic factors which contribute to the change of drought frequency under future climate change; (3) identify suitable adaptation measures to mitigate the negative impacts of drought in the study area.

5.2 Materials and method

5.2.1 Study domain description

The domain of the study is the New South Wales (NSW) wheat belt (141.0°-152.0°E, 28.5°-36.1°S) of south-eastern Australia, which covers an area of 360 000 km^2 (Figure 5-1) (Liu et al., 2014). There is an east-west gradient in both elevation and precipitation/temperature across the study area. The eastern part of the wheat belt consists of mountains with an elevation up to 1100 m and the western areas are mainly plains. Average temperature ranges from 11 °C in the south-east to 20 °C in the north-west and average annual precipitation ranges from 1000 mm in the south-east to 200 mm in the south-west (Figure 5-2). Overall, the eastern part of the wheat belt is wet and cold, whilst the western part is dry and warm. In NSW wheat belt, winter crops commonly include wheat and canola, while summer crops are mainly sorghum and maize. In particular, wheat is the most important commodity which contributes to 15% of the gross value of agricultural production in NSW (http://www.abs.gov.au/Agriculture).

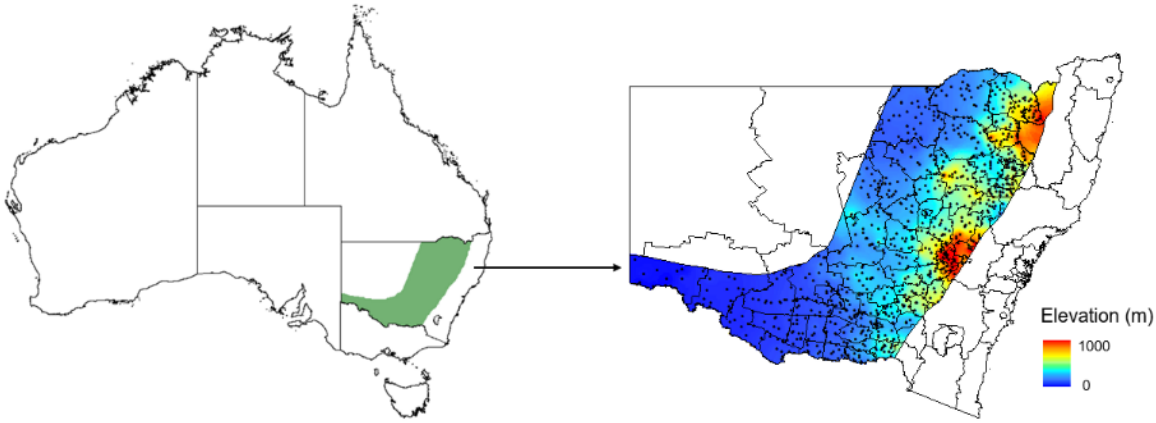


Figure 5-1 The study area is located in the New South Wales wheat belt of south-eastern Australia. Black points are the locations of 931 weather stations used in the study.

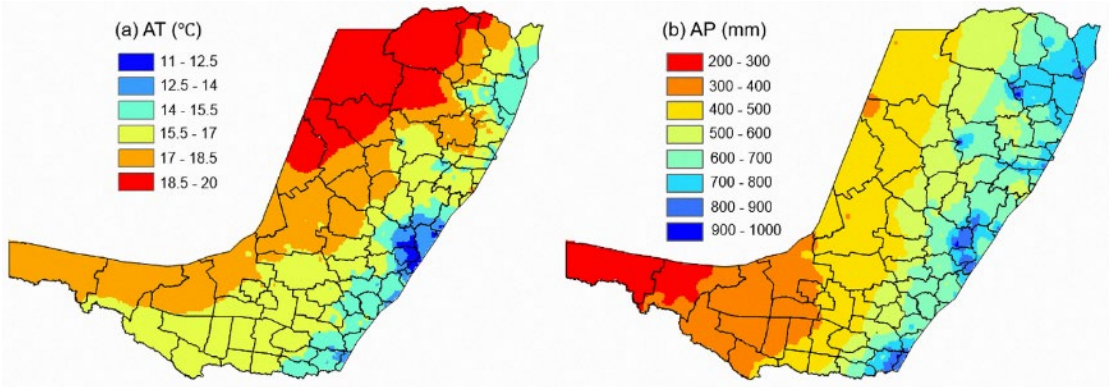


Figure 5-2 Spatial distributions of (a) annual mean temperature (AT) and (b) mean annual precipitation (AP) over the New South Wales wheat belt of south-eastern Australia during 1961-1990.

5.2.2 Climate data

In this study, the monthly gridded data of 28 GCMs (Table 5-1) were acquired from CMIP5. These GCMs are from different climate modeling institutions all over the world. Detailed descriptions of these GCMs can be found at https://cmip.llnl.gov/cmip5/.

Raw GCMs are unable to produce regional scale projections, because they are normally at coarse spatial resolutions (100-300 km grid spacing). Thus, we downscaled raw GCMs data to weather observation stations using a weather-generator based statistical downscaling approach which was developed by NSW Department of Primary Industries at Wagga Wagga Agricultural Institute (NWAI-WG) and has been described by Liu and Zuo (2012). This approach has been frequently used in recent climate change research (e.g. Anwar et al., 2015; He et al., 2017; Wang et al., 2016) which allows daily data of meteorological factors from monthly gridded GCMs to be derived. It can also correct biases in the raw GCMs. Briefly, the first step of the downscaling procedure was to interpolate the monthly gridded data for each weather station (931 stations in the NSW wheat belt, Figure 1) using Inverse Distance Weighted (IDW) method (Bartier and Keller, 1996). A bias correction method was applied in this step to enable the resulting monthly station data to match with observed data (downloaded from Scientific Information for Land Owners patched point dataset, http://www.longpaddock.qld.gov.au/silo/ppd/index.php) using quantile-quantile (QQ) mapping technique (Zhang, 2005; Zhang, 2007). Secondly, the bias-corrected monthly values were disaggregated to daily data through the modified WGEN weather generator (Richardson and Wright, 1984). Further details of procedures and results of this method can be found in Liu and Zuo (2012).

Table 5-1 List of 28 GCMs under RCP8.5 future climate scenarios used in this study for statistical downscaling outputs of 931 stations over the New South Wales wheat belt of south-eastern Australia.

Model ID	Name of GCM	Abbr. of	Institute ID	Country
01	BCC-CSM1.1	BC1	BCC	China
02	BCC-CSM1.1(m)	BC2	BCC	China
03	BNU-ESM	BNU	GCESS	China
04	CanESM2	CaE	CCCMA	Canada
05	CCSM4	CCS	NCAR	USA
06	CESM1(BGC)	CE1	NSF-DOE-NCAR	USA
07	CMCC-CM	CM2	CMCC	Europe
08	CMCC-CMS	CM3	CMCC	Europe
09	CSIRO-Mk3.6.0	CSI	CSIRO-QCCCE	Australia
10	EC-EARTH	ECE	EC-EARTH	Europe
11	FIO-ESM	FIO	FIO	China

12	GISS-E2-H-CC	GE2	NASA GISS	USA
13	GISS-E2-R	GE3	NASA GISS	USA
14	GFDL-CM3	GF2	NOAA GFDL	USA
15	GFDL-ESM2G	GF3	NOAA GFDL	USA
16	GFDL-ESM2M	GF4	NOAA GFDL	USA
17	HadGEM2-AO	Ha5	NIMR/KMA	Korea
18	INM-CM4	INC	INM	Russia
19	IPSL-CM5A-MR	IP2	IPSL	France
20	IPSL-CM5B-LR	IP3	IPSL	France
21	MIROC5	MI2	MIROC	Japan
22	MIROC-ESM	MI3	MIROC	Japan
23	MIROC-ESM-CHEM	MI4	MIROC	Japan
24	MPI-ESM-LR	MP1	MPI-M	Germany
25	MPI-ESM-MR	MP2	MPI-M	Germany
26	MRI-CGCM3	MR3	MRI	Japan
27	NorESM1-M	NE1	NCC	Norway
28	NorESM1-ME	NE2	NCC	Norway

In GCMs dataset, four representative concentration pathways (RCP2.6, RCP4.5, RCP6.0, and RCP8.5) are available, which describe four possible future climates based on future concentrations of greenhouse gases. RCP8.5 represents the most serious condition with continuous rising emissions throughout the 21st century. Studies have shown that trends in greenhouse gases concentrations since 2000 agree better with those projected by RCP8.5 than any other scenarios (Diffenbaugh and Field, 2013; Peters et al., 2011). Therefore, RCP8.5 was utilized in this study as it has projections which are most likely to be achieved in the future (Ribeiro et al., 2016).

5.2.3 Relative SPEI (rSPEI)

The original SPEI is based on climatic water balance and allows for the contribution of temperature in drought assessment. It uses the difference between precipitation (P) and evapotranspiration (PET) as parameter to characterize drought. It is a standardized index for which a value of 0 represents the median P-PET (i.e., normal conditions), while dry conditions are denoted by negative values (i.e., -2 for extremely dry) and wet conditions are denoted by positive values (i.e., 2 for extremely wet). Generally, a value of < -1 is viewed as drought condition. In addition, the SPEI can be calculated on different timescales according to specific aims.

Conceptually, one SPEI value indicates the deviation of P-PET at a given station for a given period from ‘normal condition’. This raises a problem. For a wet station and a dry station within a region, a given amount of P-PET that produces negative SPEI (say -1) at the wet station might

supposedly produce positive SPEI at the dry station. So, the relative aridity condition remains uncertain between stations when using the original SPEI. Dubrovsky et al. (2008) ever used a relative SPI to make comparisons of absolute drought conditions. Here, we introduced the rSPEI which improved the calculation of the original SPEI and can be applied to spatial analysis.

The process of calculating the original SPEI mainly consists of two steps. 1) fit the P-PET series into a log-logistic distribution to acquire parameters; 2) convert the distribution into a normal distribution to determine SPEI values. In this case, the same P-PET series is used in both steps. For the calculation of rSPEI, we created a reference P-PET series by aggregating all monthly P-PET totals of the 931 stations.

$$\overline{D}_j = \overline{P}_j - \overline{PET_j} \tag{5-1}$$

where $\overline{P}_j$, $\overline{PET_j}$, and $\overline{D}_j$ are the averaged total precipitation, the accumulated PET and the deficit of *j-th* month at the 931 weather stations. Then, the averaged accumulated P-PET at *k*-month scale is calculated by

$$\begin{cases} \overline{X}_{i,j}^{k} = \sum_{l=13-k+j}^{12} \overline{D}_{i-1,l} + \sum_{l=1}^{j} \overline{D}_{i,l} & if\ j < k \\ \overline{X}_{i,j}^{k} = \sum_{l=j-k+1}^{j} \overline{D}_{i,l} & if\ j \geq k \end{cases} \tag{5-2}$$

where $\overline{X}_{i,j}^{k}$ is the accumulated P-PET at *k*-month scale in *j-th* month of *i-th* year; $\overline{D}_{i,l}$ is the monthly P-PET in *l-th* month of *i-th* year. The parameters for the log-logistic distribution were acquired according to this reference series. Then, the values of the rSPEI relative to the reference distribution function were acquired for each station. In this study, the rSPEI was calculated for a 3-month time period in order to investigate seasonal drought attributes. This above process enabled us to compare the P-PET deviation for each location using the distribution function that indicated the climate optimum of the given region rather than that of the individual location. A comparison between the rSPEI and the SPEI values is shown in Figure 5-S1 in the supplementary material.

5.2.4 Evaluation of drought characteristics

Figure 5-3 illustrates the whole process of this study. We assessed future drought characteristics from three aspects: temporal changes, spatial changes, and major drivers of these changes.

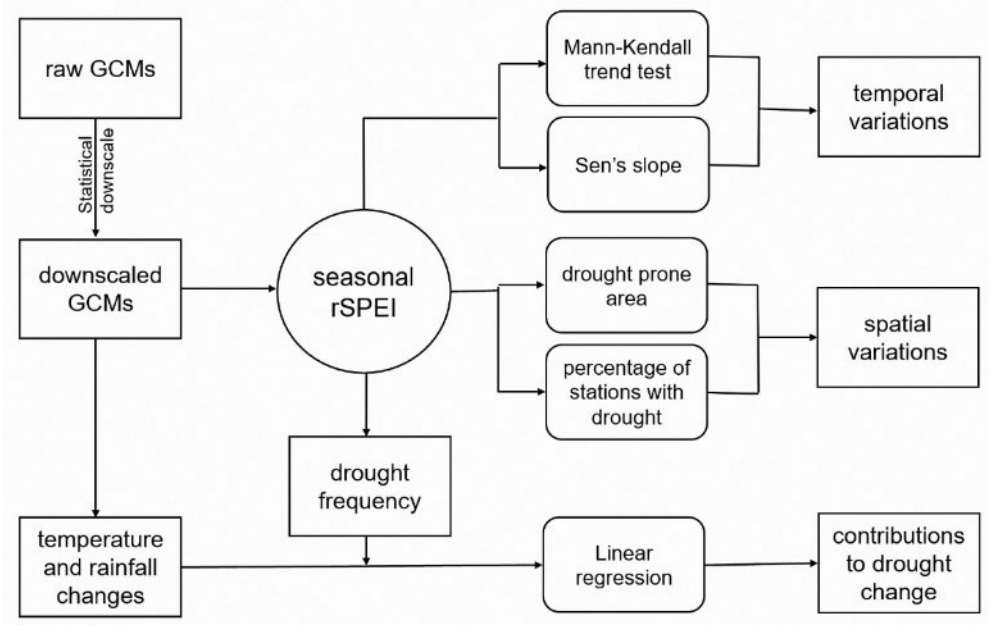


Figure 5-3 Framework of the procedures used in this study.

5.2.4.1 Temporal changes

We assessed the likely changes of drought severity across the wheat belt. As the rSPEI values represent the severity of drought, analysing the time series of the rSPEI values for each station can reveal temporal changes in drought. Two methods, i.e. the Mann-Kendall trend test and the Sen's slope, were used. The Mann-Kendall trend test, which is often used to assess trends in hydroclimatological time series (Hamed and Ramachandra Rao, 1998; Lutz et al., 2016; Serrano-Notivoli et al., 2018), was applied to test the significance of trends on the time series of the rSPEI for each station. On the other hand, the trend of a climate variable may not be assessed to be statistically significant while it might be of practical interest (Shahid, 2010; Sheikhy et al., 2017). Therefore, in our study, we also applied linear trend analysis on rSPEI time series using the Sen's slope (Sen, 1968) which could provide a robust estimation of trend. In addition, these two methods were both performed using the R package 'trend' (Thorsten, 2018).

5.2.4.2 Spatial changes

Different zones of the study area are characterized with rather different climate conditions (Figure 5-2). Thus, some zones might be more vulnerable to drought compared to others. These zones are defined as drought prone zones. A zone is identified as drought prone zone where there is a high drought frequency value. Drought frequency (DF, %) indicates the number of drought events occurring for a given period (Spinoni et al., 2013). Specified thresholds of DF are usually needed to identify the drought prone zone for different areas. Many studies (Patel et al., 2007; Sonmez et al., 2005; Wilhelmi and Wilhite, 2002) have defined the thresholds to filter drought prone areas, mainly ranging from 20% to 30%. In this study, we chose the upper value of the threshold (30%)

because Australia is typically a dry continent. Seasonal drought events (rSPEI < -1) of each weather station were firstly counted for four 30-year periods (1961-1990, 2011-2040, 2041-2070, and 2071-2100) and then the DFs were calculated. The DFs were then interpolated for each grid cell (~3 km) using IDW method. Areas with more than 30% of seasonal drought events were then identified for each period and each season.

5.2.4.3 Major drivers of the changes of drought

A descriptive statistical analysis was conducted using a least-squared multiple linear regression model between the changes in DF (%, Figure 5-S2) and the changes in and precipitation (%, Figure 5-S3) and temperature (°C, Figure 5-S4). The regression allowed for a better evaluation of drought attributes, thereby providing useful insights into the nature and strength of the relationships.

5.3 Results

5.3.1 Temporal changes in drought

Temporal variation of seasonal rSPEIs in all the 931 weather stations over the NSW wheat belt for the period of 1961-2100 are given in Figure 5-4. For each year, the rSPEI values projected by the 28 GCMs for the 931 stations were presented as a distribution. Distribution for each year was a complete presentation of seasonal drought conditions of the whole wheat belt based on the 28 GCMs. For example, in the spring of 1961, the concentrated position of the distribution was around 0.3. This meant that most weather stations had an rSPEI value of ~0.3, thus most of the wheat belt was in near normal climatic conditions. Therefore, changes in the distribution could then illustrate the overall change in drought condition for the entire wheat belt over the study period. Since the peak positions of every distribution could to some degree represent the concentrated distribution positions, we linked the peak positions using black lines to indicate trend. The distributions showed a significant ($P<0.001$) decreasing trend in both spring and winter over the study period. The spring and winter periods decreased from 0.3 to nearly -1, indicating the majority of the wheat belt might experience future moderate drought conditions for these seasons. However, autumn and summer had relatively small trends and the rSPEI values for the majority of the wheat belt consistently fluctuated around 0. Therefore, in general, the wheat belt was expected to experience drier conditions in spring and winter but had little change in summer and autumn.

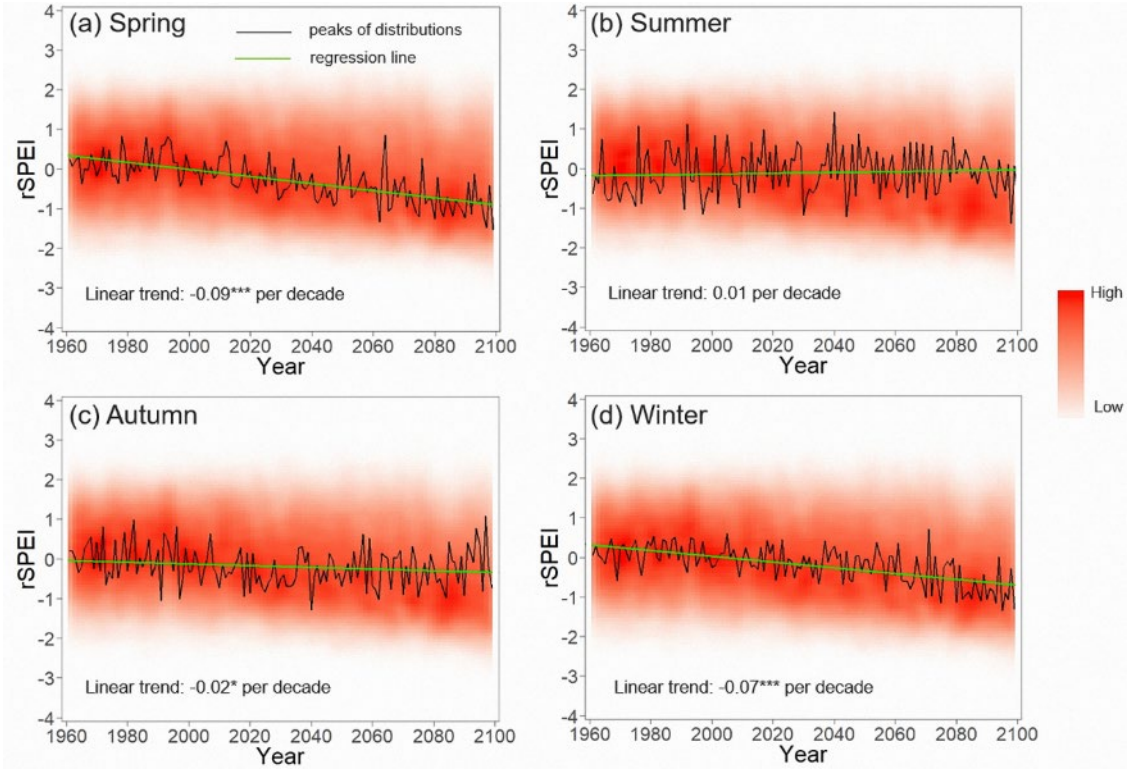


Figure 5-4 Changes in rSPEI values of the 931 weather stations in the New South Wales wheat belt of south-eastern Australia during 1961-2100. Seasonal rSPEI values were first calculated for the 931 weather stations based on 28 GCMs. Then, for each year, seasonal rSPEI values from all the weather stations and GCMs were shown in a distribution. The red shaded area in the figure indicates the distribution of the rSPEI values for each year. The deeper the red colour, the more concentrated the distribution of the rSPEI values. Each distribution has a peak that indicates the most concentrated position. The black line captures the peaks of distributions of the rSPEI values for each year, so its change can to some extent represent the change in the distributions of the rSPEI values from the 931 weather stations based on 28 GCMs. The green line shows the linear trend of the rSPEI peaks. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$.

Figure 5-5 shows the changes of drought severity over the wheat belt in the future using rSPEI as a drought indicator. Trends of rSPEI were calculated through Sen's slope. The black triangles and circles indicate weather stations with a significant changing trend according to Mann-Kendall trend test. Generally, trends were consistent with the results indicated in Figure 5-4 and significant decreasing trends were projected in spring and winter for the entire wheat belt. Given that the rSPEI thresholds of -1, -1.5, and -2 indicate moderate, severe, and extreme drought condition, respectively, most of the wheat belt was likely to suffer a higher-level drought in spring and winter by the end of this century. While for summer and autumn, areas with slightly increased drought intensity were mainly located in the south-western and northern areas, respectively. However, these trends were slight compared to those in spring and winter. In addition, north-eastern areas

during summers and south-western areas during autumns were expected to experience decreased drought intensity.

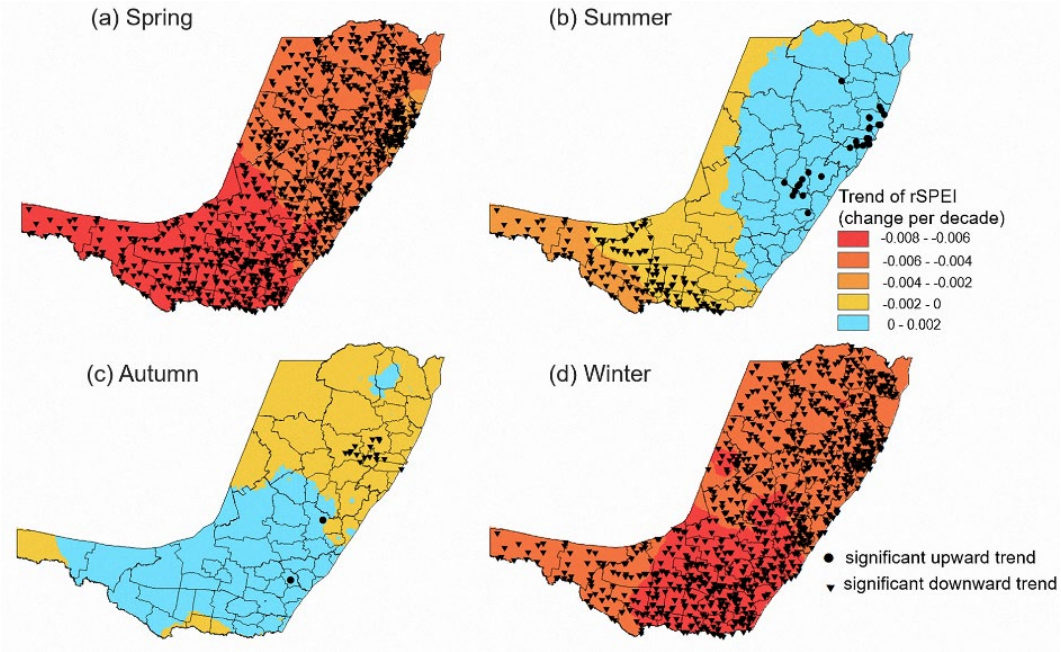


Figure 5-5 The trends of long-term seasonal drought across the New South Wales wheat belt of south-eastern Australia in 2011-2100. The rSPEI is calculated for each of 28 GCMs and the average trend per decade (based on Sen's slope) from all downscaled GCMs is calculated for the 931 weather stations. We interpolated the trend of rSPEI using IDW method as the resolution of ~3 km. The black triangles and circles indicate stations with a significant changing trend based on the Mann-Kendall trend test ($|Z_S| > 1.96$).

5.3.2 Spatial changes in drought

As the weather stations used in this study are relatively evenly distributed over the study area, the percentage of drought stations (rSPEI < -1) could be used to estimating the area size experiencing drought. Figure 5-6 illustrates the percentage of drought stations according to the rSPEI values of the 931 weather stations over the NSW wheat belt for the period of 1961-2100. Despite the inter-annual variability and GCM uncertainty, increasing trends were apparent in both winter and spring. About 40% of stations in the wheat belt could be expected to suffer continuous spring and winter drought by the end of the 21st century. However, in summer and autumn, the percentages of drought stations fluctuated around 20% over time, with no obvious trend.

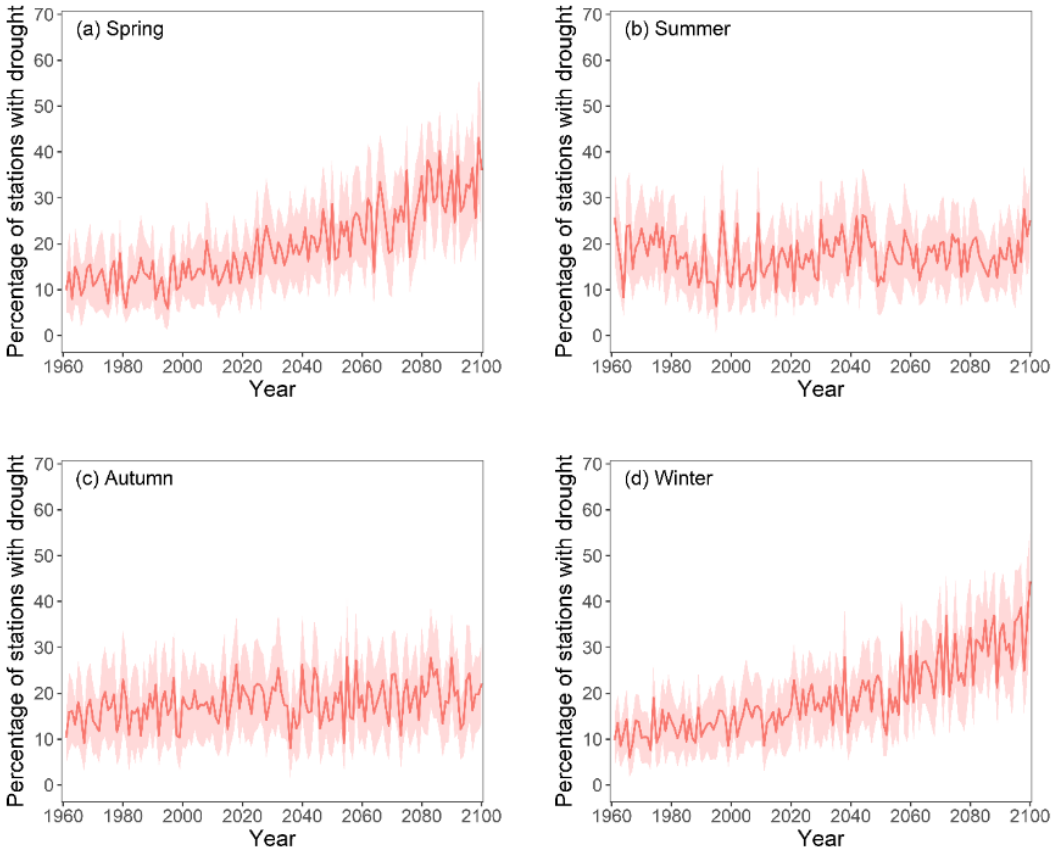


Figure 5-6 Percentage of stations with drought (rSPEI < -1) in the 931 weather stations across the New South Wales wheat belt of south-eastern Australia for 4 seasons during 1961-2100. For each year, the percentage of stations with drought was calculated for each of 28 downscaled GCMs and multi-model ensemble mean values (red lines) were then plotted. The shading denotes the 95% confidence intervals for the 28 GCMs.

Drought prone areas were primarily in the western part of the wheat belt (Figure 5-7). This was particularly noticeable for spring and winter drought (Figure 5-6). For example, only a small area (14.6%, about 5.27×10^4 km^2) in the north-west of the wheat belt was drought vulnerable during 1961-1990 spring period. However, over time, the eastern limit expanded eastward, and by the end of the 21st century over half (58.8%) of the wheat belt was at a high risk of experiencing spring drought. The winter drought prone area followed a similar expansion rate, but the area mainly expanded in the northern part of the wheat belt. Summer and autumn drought prone areas had slight changes over time.

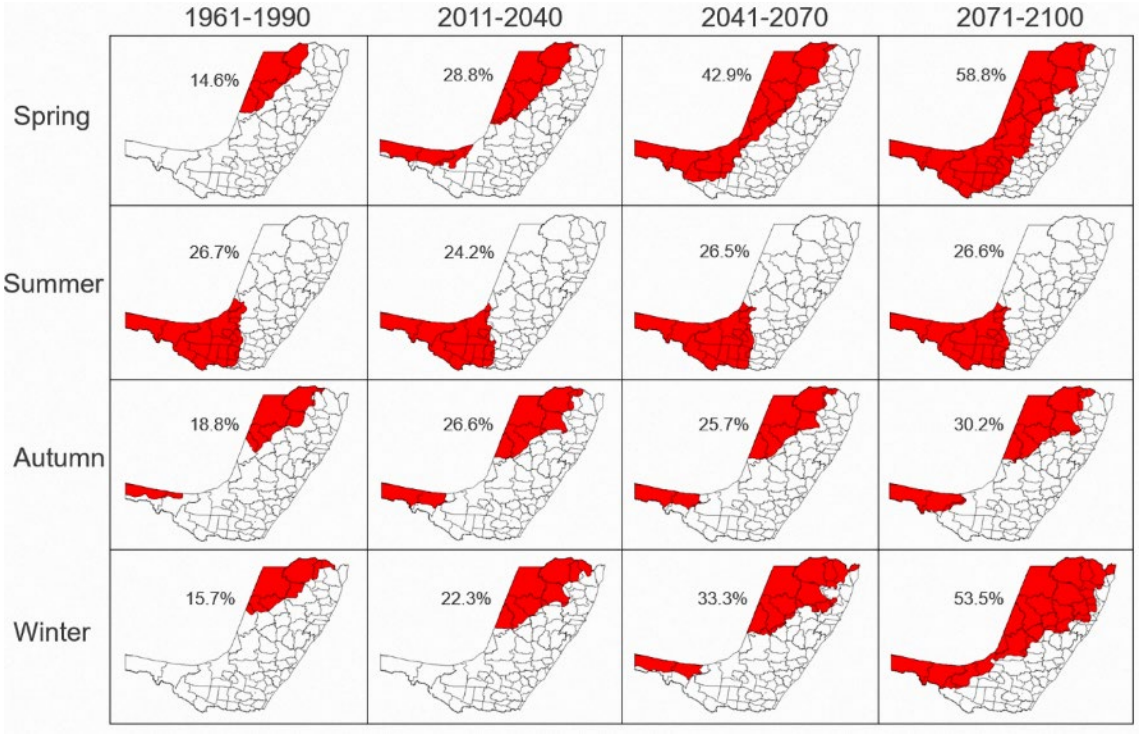


Figure 5-7 Changes in seasonal drought prone areas in the New South Wales wheat belt of south-eastern Australia. Drought prone area is defined as an area with more than 30% of seasonal drought events (rSPEI < -1) for four 30-year periods (1961-1990, 2011-2040, 2041-2070 and 2071-2100). For each year and each season, multi-model ensemble means of rSPEI values were calculated for all stations based on the 28 downscaled GCMs and then percentages of < -1 values were calculated for each period. Then, the percentages were interpolated for each grid cell (~3 km). Areas with more than 30% of seasonal drought events are shown in red parts. The number in each plot denotes the proportion of the red area.

5.3.3 Major drivers of drought trends

The rSPEI calculations were based on precipitation and temperature, so changes in these variables would result in changes in the rSPEI. While the results of the rSPEI showed dramatic changes in future drought conditions, it was not clear if this was being driven by an increase in temperature, or by a decrease in precipitation. As different regions had different climates and aridity conditions, the primary factor might differ from location to location. In order to explore this possibility, seasonal changes in temperature and precipitation projected by the 28 downscaled GCMs were extracted for all weather stations in three 30-year periods (2011-2040, 2041-2070 and 2071-2100) compared to the baseline (1961-1990) (Figure 5-S3 and 5-S4). Based on elevation, the study area was divided into two zones (Figure 5-8), the eastern zone which is a mountainous area with an elevation of >300 m and the western zone which is mainly occupied by plains (< 300 m). Climatically, the western zone is warm and dry, while the east is cold and wet (Figure 5-2). Spatial regression analysis was then applied between the changes in drought frequency (Figure 5-S2) and

the changes in temperature and precipitation using least-squares multiple regression (ΔDF = *a**Δprecipitation (%) + *b**Δtemperature (°C)) in both zones for each season and each period.

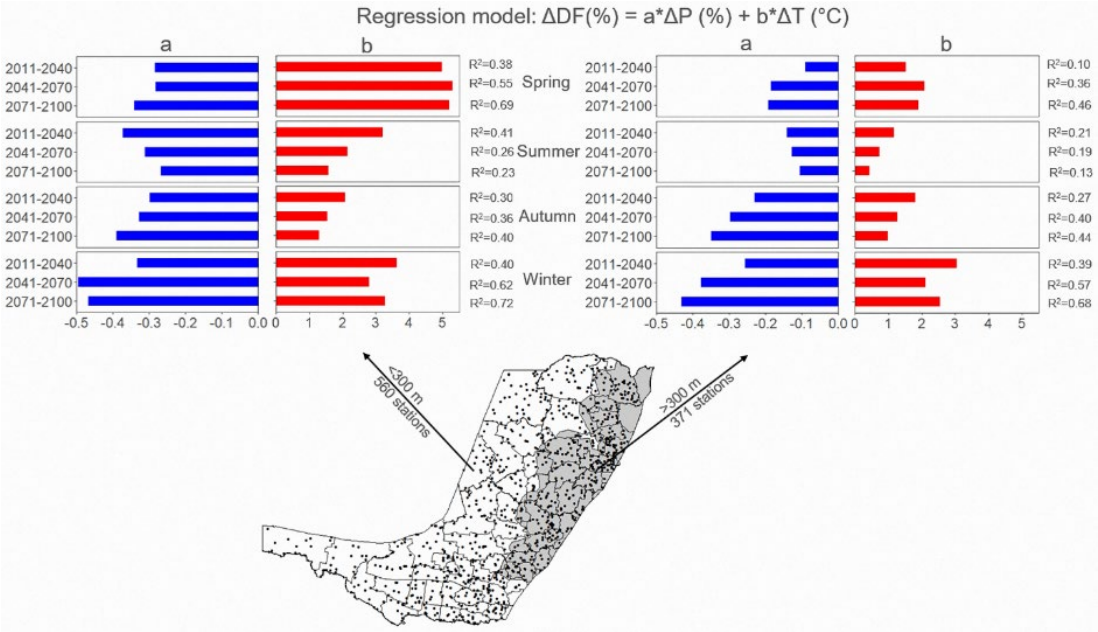


Figure 5-8 Regression analysis of the impacts of temperature (a) and precipitation (b) on drought frequency (DF, %) in the western (left) and the eastern (right) zones of the New South Wales wheat belt of south-eastern Australia. Changes of DF, averaged annual total precipitation (P) and annual mean temperature (T) were firstly calculated for three, 30-year periods (2011-2040, 2041-2070, and 2071-2100) relative to 1961-1990 for each weather station within the western (n=560) and eastern (n=371) zones for all the GCMs. Least-squares multiple regression model (ΔDF = *a**ΔP (%) + *b**ΔT (°C)) was then built in both zones for each season and each period. All the regression coefficients (R^2) shown were significant ($P < 0.05$).

The R^2 (coefficient of determination) for each period and each season was higher in the western zone than the east (Figure 5-8). This meant that changes in seasonal precipitation and temperature were better able to explain DF change in the western zone. In addition, the magnitude of the regression coefficients (*a*, *b*) were also larger in the western zone than those in the east, which meant that the same decrease in precipitation or increase in temperature could result in a greater increase of DF in the western zone compared to the eastern.

Overall, temperature and precipitation changes had varying effects on drought in different seasons over time. For example, in spring and winter, DF increased greatly until 2100 with a small change in *b* but a large increase in *a*. Thus, although precipitation decreased (Figure 5-S3), and temperature increased (Figure 5-S4) over time simultaneously, the decrease in precipitation could be viewed as the major factor causing drought increase in spring and winter. However, in summer, both temperature and precipitation were expected to increase across the entire wheat belt, which leaded to a slight change in drought frequency. Therefore, both *a* and *b* decreased over time in

summer. However, in autumn, a slight increase in precipitation was detected (Figure 5-S3) but could not offset the effects caused by increased temperature. Thus, *a* also increased over time in autumn.

5.4 Discussion

This study investigated drought projections for the wheat belt of south-eastern Australia using the ensemble of 28 statistical downscaled CMIP5 GCMs. The long-term seasonal drought trends (Figure 5-4 and 5-5) showed that the whole wheat belt was expected to suffer more severe spring and winter droughts. These findings are consistent with the results from CCIA (2015), which reported future decreased spring and winter precipitation in southern Australia. For example, more than half of 40 GCMs projected >15% decrease in winter-spring precipitation in our study area until the 2090s under RCP8.5 (CCIA, 2015). It should be emphasized that winter and spring are key growth periods for winter crops such as wheat and canola. Previous studies have shown that in dryland agriculture of south-eastern Australia, precipitation declines can be amplified 1.5-1.7 times in wheat yield losses (Dijk et al., 2013). In our study, the projected increase of spring and winter droughts would inevitably cause decreased yields and increased risks to cropping systems in the study area. In addition, as the Australian climate is highly variable (Potgieter et al., 2012), the increasing trends would likely be accompanied by dramatic fluctuation as inter-annual extremely dry conditions occur more frequently (Alexander and Arblaster, 2017).

The spatial trends in drought vulnerability corresponded with temporal trends in drought. Spring and winter drought vulnerable areas were likely to expand remarkably (Figure 5-7). Historically, the western zone of the wheat belt was typically a drought vulnerable area due to lower precipitation and higher temperature (Figure 5-2). Crop yields in these areas were typically lower compared to the eastern areas (Hochman et al., 2016). Our results showed that drought prone areas were likely to increase, which meant that the areas of low-level crop yield would expand. By 2100, more than half of the wheat belt was expected to have low crop yields or might not be suitable for growing winter cereal crops. On the other hand, summer and autumn drought prone areas were primarily in the south-west and north-west and might only experience slight changes (Figure 5-6 and 5-7). This could be attributed to the increases of projected precipitation during summers and autumns in the late decades of the 21st century (Figure 5-S3).

In general, spatial and temporal increases in drought events were primarily driven by decreased precipitation and also evaporation through increased temperatures (Sheffield and Wood, 2007). It is commonly recognized that climate change will cause changes in temperature and precipitation all over the world (Pachauri and Meyer, 2015). An absolute amount of change in precipitation or temperature may result in distinctly different changes in drought conditions in different climate regions. Our results demonstrated that temperature and precipitation possessed greater ability to

regulate the aridity condition in traditional drought prone areas. A similar decrease in precipitation or increase in temperature would cause a greater increase in drought frequency in dry areas compared to wet areas (Figure 5-8). This creates a challenge for historically dry areas in face of climate change. In the future, even though a slight increase in temperature or decrease in precipitation will inevitably increase the risk of drought, thereby reducing the agricultural production capacity.

Our results also showed that areas of projected increased precipitation were also at risk of increased drought frequency. For example, in autumn, most of the NSW wheat belt was projected to receive more rain after 2041 (Figure 5-S3). However, the area vulnerable to drought in autumn would still increase (Figure 5-7). This was mainly because the slight increase in precipitation might not fully compensate the increasing water demand caused by the increasing temperature (Liu et al., 2017). In addition to this, climate projections have shown that future precipitation is likely to be characterized by low frequency, high intensity and uneven intra-annual precipitation distribution (Bao et al., 2017). More intense precipitation events may result in increased runoff (Trenberth et al., 2014) without replenishment of soil moisture. In this case, while summer precipitation amount might increase significantly (Figure 5-S3), aridity was projected to remain relatively high. Therefore, areas with projected increased precipitation might also experience drought conditions. In addition, current drought indices which only consider the amount of precipitation are of limited utility when evaluating actual drought conditions. A more comprehensive drought index that not only considers precipitation amount but also takes into account of precipitation frequency and intensity is urgently needed for the evaluation of future drought.

The rSPEI used in our study, nevertheless, had proved to be a useful index in assessing regional drought conditions. Generally, studying drought at regional scales can better discern local characteristics of drought, resulting in a more accurate projection of drought conditions (Sheffield et al., 2009). The rSPEI can help achieve this goal by detecting relative drought prone areas within a particular region. The values of the rSPEI are uniquely associated with the set of weather series at all stations rather than at a single station. In this way, the rSPEI provides an objective and effective quantification of the relative intensity of drought events and their frequency with regard to the whole region (Marcos-Garcia et al., 2017; Trnka et al., 2009). Thus, the results can assist stakeholders to develop regionally specific adaptation strategies and disaster response measures. It would be worthwhile to employ the rSPEI presented here to access other drought prone areas to extract more generalized conclusions.

In addition, even though uncertainties exist in GCMs and multi-model ensemble method, results from our work could be regarded as an indication of the very likely future. Mitigation and adaptation strategies should be prepared in advance in order to minimize the adverse effects of

future severe droughts on crop production. Our results showed that the NSW wheat belt was really a climatologically-diverse region, so coping strategies should be specific in different zones. For example, cropping in the western zones of the south-eastern Australian wheat belt would be an increasingly risky enterprise. Therefore, changing enterprise type e.g. incorporating a livestock component, purchasing additional cropping land or moving the cropping enterprise to areas with more reliable precipitation or access to irrigation might prove alternative adaptive responses for the near future. While in the eastern zones, changing sowing dates or crop rotations, stubble management, incorporating shorter growing season varieties and even fallow, might represent adaptive responses to cope with future drought. We hope that drought projections from this work are able to provide useful information for long-term planning for stakeholders.

5.5 Conclusions

Temporal and spatial characteristics of future seasonal droughts in the New South Wales wheat belt of south-eastern Australia until the end of 21st century were analysed in this study based on 28 statistical downscaled global climate models. The relationship between drought frequency and temperature as well as precipitation was also examined. The major conclusions are as follows:

(1) Spring and winter droughts were expected to be more severe over the wheat belt of south-eastern Australia, while summer and autumn drought intensities might change little.

(2) The winter and spring drought prone areas were likely to spread from west to east significantly and more than half of the wheat belt would be vulnerable to winter and spring droughts by 2100. In contrast, the summer and autumn drought prone areas were primarily in the south-west and north-west, respectively, with only slight changes and sometimes even a decrease in the future.

(3) Traditionally dry areas would likely suffer a greater increase in drought frequency compared to wet areas when subjected to a same increase in temperature or decrease in precipitation.

We believe this study provides useful information for local farmers and policy makers with respect to evaluating the impacts of drought change on cropping systems in south-eastern Australia. However, additional studies which for example, examining the effect of rainfall intensity in a drought index, may help to increase our confidence in accurately projecting future drought change.

5.6 Supporting information

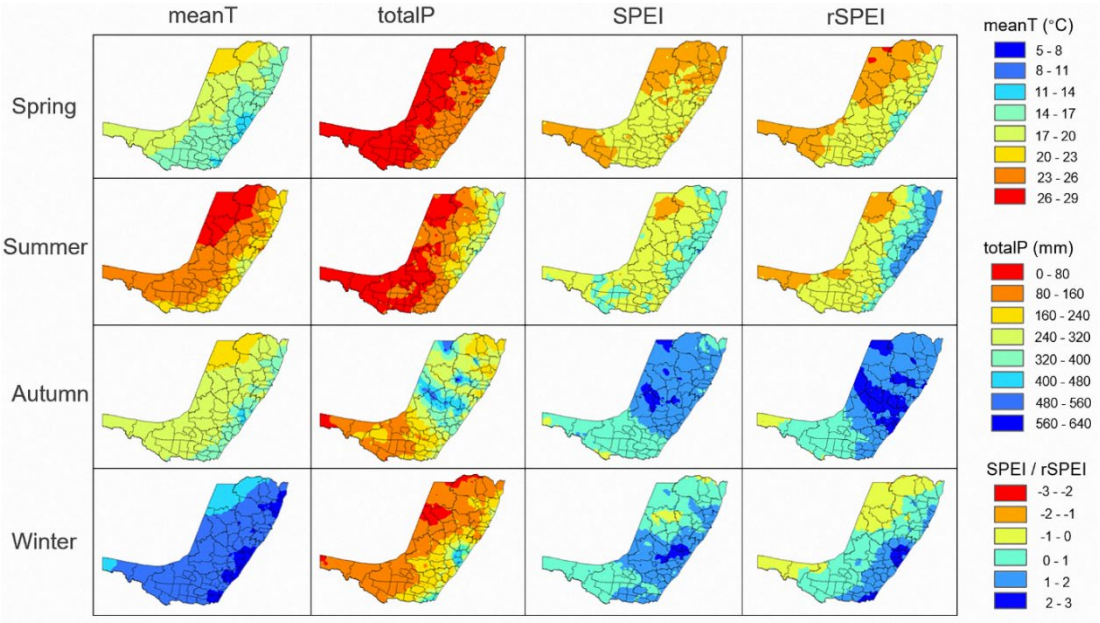


Figure 5-S1 A comparison between the rSPEI and the SPEI values. meanT: mean temperature; totalP: total precipitation. As the SPEI is a standardized index, it cannot be shown in the form of multi-year average. Thus, we took the seasonal SPEI and rSPEI values of 1990 as an example to compare the two indices. Generally, areas with more precipitation and lower temperature are less likely to suffer from drought. As shown in the figure, compared to the seasonal SPEIs, the seasonal rSPEIs are more consistent with the temperature and precipitation conditions. For example, the rSPEI matches better with the precipitation condition for the 1990 winter. Thus, the rSPEI performs better in identifying relatively dry and wet areas and is more suitable for spatial analysis of drought.

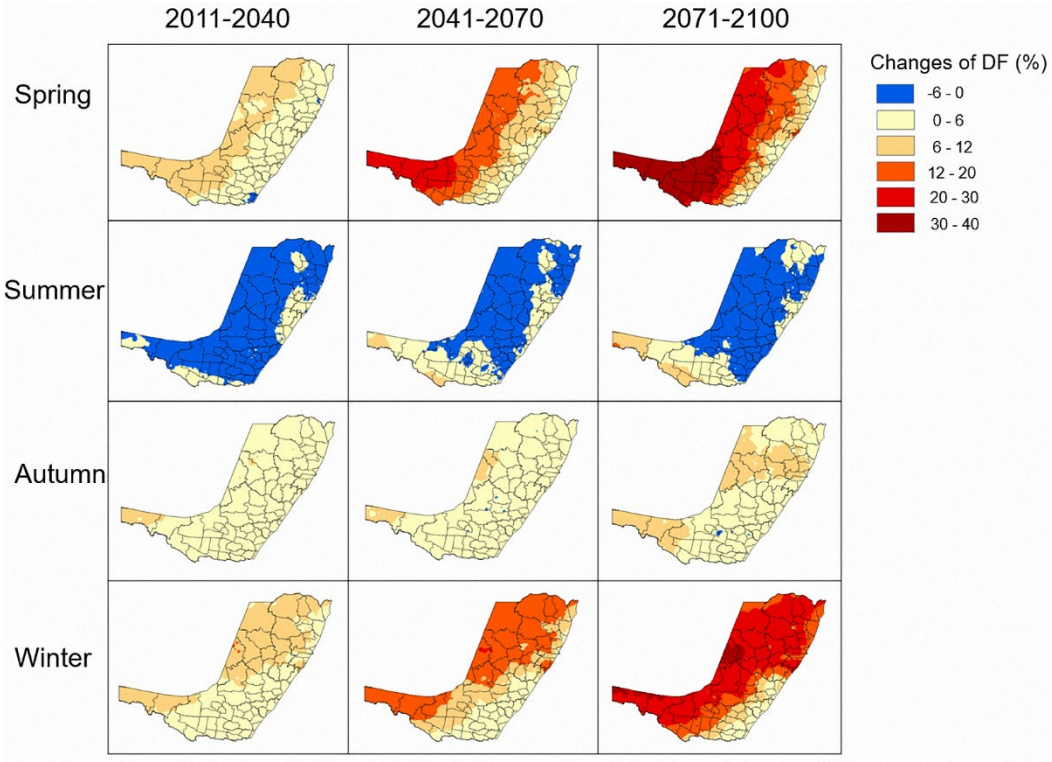


Figure 5-S2 Projected changes in seasonal DF (drought frequency, percentage of < -1 rSPEI values during a 30-year period). Calculations are done for each of the 28 GCMs and the mean changes are plotted here.

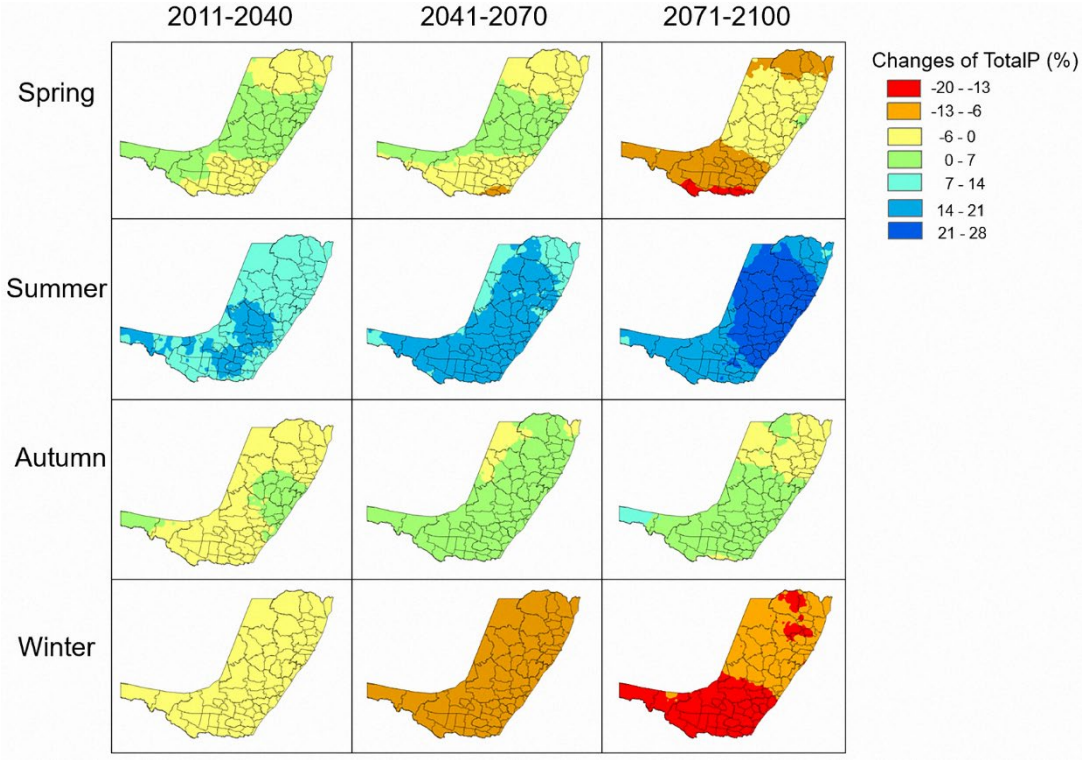


Figure 5-S3 Percentage change in seasonal totalP (total precipitation) projections. Calculations are done for each of the 28 GCMs and the mean changes are plotted here.

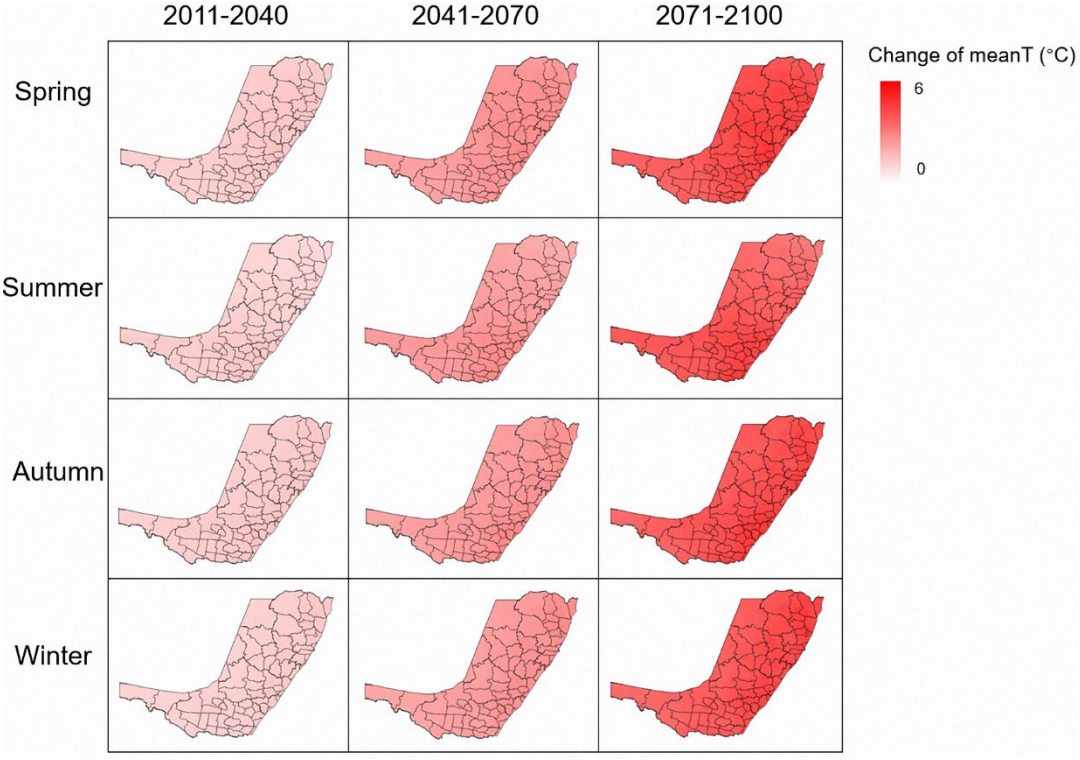


Figure 5-S4 Projected changes in seasonal meanT (mean temperature) in degrees Celsius. Calculations are done for each of the 28 GCMs and the mean changes are plotted here.

5.7 Reference

Ahmadalipour, A., Moradkhani, H. and Svoboda, M., 2017. Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble. International Journal of Climatology, 37(5): 2477-2491.

Alexander, L.V. and Arblaster, J.M., 2017. Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. Weather and climate extremes, 15: 34-56.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome, 300(9): D05109.

Anagnostopoulou, C., 2017. Drought episodes over Greece as simulated by dynamical and statistical downscaling approaches. Theoretical and Applied Climatology, 129(1-2): 587-605.

Anwar, M.R. et al., 2015. Climate change impacts on phenology and yields of five broadacre crops at four climatologically distinct locations in Australia. Agricultural Systems, 132: 133-144.

Bao, J.W., Sherwood, S.C., Alexander, L.V. and Evans, J.P., 2017. Future increases in extreme precipitation exceed observed scaling rates. Nature Climate Change, 7(2): 128-+.

Bartier, P.M. and Keller, C.P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). Computers & Geosciences, 22(7): 795-799.

Beguería, S., Vicente‐Serrano, S.M., Reig, F. and Latorre, B., 2014. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models,

tools, datasets and drought monitoring. International Journal of Climatology, 34(10): 3001-3023.

BOM and CSIRO, 2016. State of the Climate.

Buser, C.M., Kunsch, H.R., Luthi, D., Wild, M. and Schar, C., 2009. Bayesian multi-model projection of climate: bias assumptions and interannual variability. Clim Dynam, 33(6): 849-868.

CCIA, 2015. Climate change in Australia: Projections for Australia's NRM regions. https://www.climatechangeinaustralia.gov.au/en/publications-library/technical-report/ (assessed 23 Jan 2018).

Collins, M., 2007. Ensembles and probabilities: a new era in the prediction of climate change. Philos T R Soc A, 365(1857): 1957-1970.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nature Climate Change, 3(1): 52-58.

Diffenbaugh, N.S. and Field, C.B., 2013. Changes in ecologically critical terrestrial climate conditions. Science, 341(6145): 486-492.

Dijk, A.I. et al., 2013. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resources Research, 49(2): 1040-1057.

Duan, Q.Y. and Phillips, T.J., 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. J Geophys Res-Atmos, 115.

Dubrovsky, M., Svoboda, M.D., Trnka, M. and Hayes, M.J., 2009. Application of relative drought indices in assessing climate-change impacts on drought conditions in Czechia. Theoretical and Applied Climatology, 96(1): 155-171.

Gao, X.R. et al., 2017. Temporal and spatial evolution of the standardized precipitation evapotranspiration index (SPEI) in the Loess Plateau under climate change from 2001 to 2050. Sci Total Environ, 595: 191-200.

Guttman, N.B., 1998. Comparing the Palmer Drought Index and the standardized precipitation index. J Am Water Resour As, 34(1): 113-121.

Hamed, K.H. and Rao, A.R., 1998. A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology, 204(1-4): 182-196.

Hargreaves, G.H. and Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Applied engineering in agriculture, 1(2): 96-99.

He, L. et al., 2017. Multi-model ensemble projections of future extreme heat stress on rice across southern China. Theoretical and Applied Climatology: 1-12.

Heim, R.R., 2002. A Review of Twentieth-Century Drought Indices Used in the United States. Bulletin of the American Meteorological Society, 83(8): 1149-1165.

Hochman, Z., Gobbett, D., Horan, H. and Garcia, J.N., 2016. Data rich yield gap analysis of wheat in Australia. Field Crops Research, 197: 97-106.

IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental Panel on Climate Change. Journal of Clinical Endocrinology & Metabolism, 18(6): 586-599.

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Ishida, K., Gorguner, M., Ercan, A., Trinh, T. and Kavvas, M.L., 2017. Trend analysis of watershed-scale precipitation over Northern California by means of dynamically-downscaled CMIP5 future climate projections. Sci Total Environ, 592: 12-24.

Kirono, D.G.C., Kent, D.M., Hennessy, K.J. and Mpelasoka, F., 2011. Characteristics of Australian droughts under enhanced greenhouse conditions: Results from 14 global climate models. Journal of Arid Environments, 75(6): 566-575.

Lesk, C., Rowhani, P. and Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production. Nature, 529(7584): 84-87.

Liu, D.L., Anwar, M.R., O'Leary, G. and Conyers, M.K., 2014. Managing wheat stubble as an effective approach to sequester soil carbon in a semi-arid environment: Spatial modelling. Geoderma, 214: 50-61.

Liu, D.L. and Zuo, H., 2012. Statistical downscaling of daily climate variables for climate change impact assessment over New South Wales, Australia. Climatic Change, 115(3-4): 629-666.

Liu, M., Xu, X., Sun, A.Y. and Wang, K., 2017. Decreasing spatial variability of drought in southwest China during 1959–2013. International Journal of Climatology.

Lloyd-Hughes, B. and Saunders, M.A., 2002. A drought climatology for Europe. International Journal of Climatology, 22(13): 1571-1592.

Lutz, S.R. et al., 2016. Hydroclimatic and water quality trends across three Mediterranean river basins. Sci Total Environ, 571: 1392-1406.

Marcos-Garcia, P. et al., 2017. Combined use of relative drought indices to analyze climate change impact on meteorological and hydrological droughts in a Mediterranean basin. Journal of Hydrology, 554.

Masud, M.B., Khaliq, M.N. and Wheater, H.S., 2017. Future changes to drought characteristics over the Canadian Prairie Provinces based on NARCCAP multi-RCM ensemble. Clim Dynam, 48(7-8): 2685-2705.

Maurer, E.P., Roby, N., Stewart-Frey, I.T. and Bacon, C.M., 2017. Projected twenty-first-century changes in the Central American mid-summer drought using statistically downscaled climate projections. Regional Environmental Change: 1-12.

Mckee, T.B., Doesken, N.J. and Kleist, J., 1993. The Relationship of Drought Frequency and Duration to Time Scales.

Mishra, A.K. and Singh, V.P., 2010. A review of drought concepts. Journal of Hydrology, 391(1–2): 202-216.

Morid, S., Smakhtin, V. and Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. International Journal of Climatology, 27(15): 2103-2111.

Nicholls, N., 2004. The changing nature of Australian droughts. Climatic Change, 63(3): 323-336.

Nijp, J.J. et al., 2017. Including hydrological self-regulating processes in peatland models: Effects on peatmoss drought projections. Sci Total Environ, 580: 1389-1400.

Palmer, W.C., 1965. Meteorological Drought. U.S.department of Commerce Weather Bureau Research Paper.

Patel, N.R., Chopra, P. and Dadhwal, V.K., 2007. Analyzing spatial patterns of meteorological drought using standardized precipitation index. Meteorological Applications, 14(4): 329-336.

Peters, G.P. et al., 2012. Rapid growth in CO 2 emissions after the 2008–2009 global financial crisis. Nature Climate Change, 2(1): 2.

Potgieter, A. et al., 2013. Spatial impact of projected changes in rainfall and temperature on wheat yields in Australia. Climatic Change, 117(1-2): 163-179.

Reifen, C. and Toumi, R., 2009. Climate projections: Past performance no guarantee of future skill? Geophysical Research Letters, 36(13).

Ribeiro, B.R., Sales, L.P., De Marco Jr, P. and Loyola, R., 2016. Assessing mammal exposure to climate change in the Brazilian Amazon. PloS one, 11(11): e0165073.

Richardson, C.W. and Wright, D.A., 1984. WGEN: A Model for Generating Daily Weather Variables. U.S. Dept. of Agriculture, Agriculture Research Service.

Sen, P.K., 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. Journal of the American Statistical Association, 63(324): 1379-1389.

Shahid, S., 2011. Trends in extreme rainfall events of Bangladesh. Theoretical & Applied Climatology, 104(3-4): 489-499.

Sheffield, J., Andreadis, K.M., Wood, E.F. and Lettenmaier, D.P., 2009. Global and Continental Drought in the Second Half of the Twentieth Century: Severity-Area-Duration Analysis and Temporal Variability of Large-Scale Events. Journal of Climate, 22(8): 1962-1981.

Sheffield, J. and Wood, E.F., 2008. Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. Clim Dynam, 31(1): 79-105.

Sheikhy, N.T., Aris, A.Z., Sefie, A. and Keesstra, S., 2017. Detecting and predicting the impact of land use changes on groundwater quality, a case study in Northern Kelantan, Malaysia. Sci Total Environ, s 599–600: 844-853.

Shrestha, M.K., Recknagel, F., Frizenschaf, J. and Meyer, W., 2017. Future climate and land uses effects on flow and nutrient loads of a Mediterranean catchment in South Australia. Sci Total Environ, 590-591: 186-193.

Sonmez, F.K., Komuscu, A.U., Erkan, A. and Turgu, E., 2005. An analysis of spatial and temporal dimension of drought vulnerability in Turkey using the standardized precipitation index. Natural Hazards, 35(2): 243-264.

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P. and Vogt, J., 2014. World drought frequency, duration, and severity for 1951–2010. International Journal of Climatology, 34(8): 2792–2804.

Strzepek, K., Yohe, G., Neumann, J. and Boehlert, B., 2010. Characterizing changes in drought risk for the United States from climate change. Environmental Research Letters, 5(4).

Tebaldi, C. and Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions, 365(1857): 2053.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical review, 38(1): 55-94.

Trenberth, K.E. et al., 2014. Global warming and changes in drought. Nature Climate Change, 4(1): 17-22.

Trinh, T., Ishida, K., Kavvas, M.L., Ercan, A. and Carr, K., 2017. Assessment of 21st century drought conditions at Shasta Dam based on dynamically projected water supply conditions by a regional climate model coupled with a physically-based hydrology model. Sci Total Environ, 586: 197.

Trnka, M. et al., 2009. Developing a regional drought climatology for the Czech Republic. International Journal of Climatology, 29(6): 863-883.

Ummenhofer, C.C. et al., 2009. What causes southeast Australia's worst droughts? Geophysical Research Letters, 36(4): 1-5.

Venkataraman, K., Tummuri, S., Medina, A. and Perry, J., 2016. 21st century drought outlook for major climate divisions of Texas based on CMIP5 multimodel ensemble: Implications for water resource management. Journal of Hydrology, 534: 300-316.

Vicente-Serrano, S.M., Beguería, S. and López-Moreno, J.I., 2010. A multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index - SPEI. Journal of Climate, 23(7): 1696-1718.

Wang, B. et al., 2016a. Multi-model ensemble projections of future extreme temperature change using a statistical downscaling method in south eastern Australia. Climatic Change, 138(1-2): 85-98.

Wang, G.L., Yu, M. and Xue, Y.K., 2016b. Modeling the potential contribution of land cover changes to the late twentieth century Sahel drought using a regional climate model: impact of lateral boundary conditions. Clim Dynam, 47(11): 3457-3477.

Wang, Q.F. et al., 2015. The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. International Journal of Climatology, 35(13): 3760-3769.

Wells, N., Goddard, S. and Hayes, M.J., 2004. A Self-Calibrating Palmer Drought Severity Index. Journal of Climate, 17(12): 2335-2351.

Wilhelmi, O.V. and Wilhite, D.A., 2002. Assessing Vulnerability to Agricultural Drought: A Nebraska Case Study. Natural Hazards, 25(1): 37-58.

Wilhite, D.A., 1993. Drought Assessment, Management, and Planning: Theory and Case Studies. Springer US.

Yu, M., Li, Q., Hayes, M.J., Svoboda, M.D. and Heim, R.R., 2013. Are droughts becoming more frequent or severe in China based on the Standardized Precipitation Evapotranspiration Index: 1951–2010? International Journal of Climatology, 34(34): 545-558.

Zhang, X.C., 2005. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. Agricultural & Forest Meteorology, 135(1–4): 215-229.

Zhang, X.C., 2007. A comparison of explicit and implicit spatial downscaling of GCM output for soil erosion and crop production assessments. Climatic Change, 84(3-4): 337.

Chapter 6. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia

This chapter is based on the following manuscript:

Feng, P., Wang, B., Liu, D.L., Waters, C. and Yu, Q., 2019. 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, pp.100-113.

Highlights

• A hybrid model was developed by integrating the APSIM model and the RF model.

• The hybrid model outperformed the APSIM model in predicting observed wheat yield.

• The APSIM model might underestimate future yield losses caused by climate extremes.

• Increasing heat events were identified to be the major factor causing future yield losses.

Abstract

Accurately assessing the impacts of extreme climate events (ECEs) on crop yield can help develop effective agronomic practices to deal with climate change impacts. Process-based crop models are useful tools to evaluate climate change impacts on crop productivity but are usually limited in modelling the effects of ECEs due to over-simplification or vague description of certain process and uncertainties in parameterization. In this study, we firstly developed a hybrid model by incorporating the APSIM model outputs and growth stage-specific ECEs indicators (i.e. frost, drought and heat stress) into the Random Forest (RF) model, with the multiple linear regression (MLR) model as a benchmark. The results showed that the APSIM+RF hybrid model could explain 81% of the observed yield variations in the New South Wales wheat belt of south-eastern Australia, which had a 33% improvement in modelling accuracy compared to the APSIM model alone and 19% improvement compared to the APSIM+MLR hybrid model. Drought events during the grain-filling and vegetative stages and heat events immediately prior to anthesis were identified as the three most serious ECEs causing yield losses. We then compared the APSIM+RF hybrid model with the APSIM model to estimate the effects of future climate change on wheat yield. It was interesting to find that future yield projected from single APSIM model might have a 1-10% overestimation compared to the APSIM+RF hybrid model. The APSIM+RF hybrid model indicated that we were underestimating the effects of climate change and future yield might be lower than predicted using single APSIM informed modelling due to lack of adequately accounting for ECEs-induced yield losses. Increasing heat events around anthesis and grain-

filling periods were identified to be major factors causing yield losses in the future. Therefore, we conclude that including the effects of ECEs on crop yield is necessary to accurately assess climate change impacts. We expect our proposed hybrid-modelling approach can be applied to other regions and crops and offer new insights of the effects of ECEs on crop yield.

Keywords: extreme climate events; wheat yield; APSIM; random forest; hybrid model

6.1 Introduction

As the global population and living standards increase, demand for stable foods such as wheat is expected to increase by 60% towards the middle of 21st century (Alexandratos and Bruinsma, 2012; Godfray and Toulmin, 2010). Sustainably improving crop production is urgently needed to meet this demand. However, the ongoing impacts of climate change will increase the risk of meeting this demand for crop production (Howden et al., 2007; IPCC, 2014). In particular, climate change-induced increases of extreme climatic events (ECEs) are recognised as the major threat to crop production (Trnka et al., 2014; Watson et al., 2017; Wheeler and Von Braun, 2013). In recent decades, ECEs have resulted in increased yield losses around the world (Lesk et al., 2016). For example, in south-eastern Australia, drought and co-occurring heat stress reduced the agricultural Gross National Product by around 30% in 1994, 2002 and 2006 (Kirono et al., 2011). Therefore, accurately estimating current and future ECEs-induced yield losses is urgently needed to assess the sustainability of our agricultural production.

ECEs are usually defined as atypical precipitation, temperature, and other weather factors compared to their historical distributions (IPCC, 2012). During a crop growth cycle, various ECEs are likely to occur and cause varying degrees of yield losses. In this study, we focused on three agro-climatic extremes, i.e. drought, heat stress, and frost, which are widely considered to have great impacts on crop growth and yield (Guarin et al., 2018). Drought is currently the main constraint to crop yield in rainfed systems. Drought-induced insufficient water supply can adversely impact crop growth in all growing stages, by restraining root growth and leaf expansion during the vegetative growth stage, and limiting photosynthesis, carbon allocation and yield formation during reproductive stages (Chaves et al., 2002). The relationship between heat stress (high temperatures above specified thresholds) and yield losses have also been identified in numerous studies (Innes et al., 2015; Pagani et al., 2017a; Semenov and Shewry, 2011). This relationship could be potentially explained by a number of mechanisms, including decreased net photosynthesis (Rezaei et al., 2015), increased maintenance respiration rates (Innes et al., 2015), and accelerated plant development (Stratonovitch and Semenov, 2015). Crops are most vulnerable to heat stress at reproductive stages, especially from anthesis to maturity. A single, isolated extreme heat event at anthesis can considerably reduce grain yield, but a continuous period of heat stress can lead to almost total yield loss (Porter and Semenov, 2005). As with

drought and heat, frost (low temperatures below specified thresholds) can also reduce crop growth in all growing periods from the seedling stage to harvest. For example, leaves of wheat seedlings are vulnerable to extreme cold conditions and may wither (Fuller et al., 2007). When the wheat inflorescence is forming but prior to flowering, frost events can result in sterile flowers, decreasing grain number (Barlow et al., 2015).

Two distinct methods have been widely used to examine climate-yield relationships, i.e. process-based crop models and statistical models. Process-based crop models have been developed to account for the complex interactions between the local environment, the crop genotype, and management practices (Chenu et al., 2017). In recent years, crop models have been widely used to characterise the effects of historical and future ECEs on crop yield in multiple regions around the world (Cammarano and Tian, 2018; Harrison et al., 2014; Jin et al., 2017; Lobell et al., 2015). While process-based crop models can provide a comprehensive understanding of the timing, frequency and intensity of ECEs on crop growth (Watson et al., 2017), they have limitations. Some of the limitations relate to over-simplification or vague description of certain process and uncertainties in parameterisation, which can lead to inaccurate results (Eitzinger et al., 2013). These limitations are especially obvious in simulating ECEs (Barlow et al., 2015). For example, heat stress impacts are particularly poorly captured in crop models (Fischer, 2011; White and Hoogenboom, 2010). For example, most crop models simulate the effects of high temperature on leaf senescence and stem carbohydrate accumulation and distribution, rather than directly model damage to reproductive organs and processes. This raises uncertainty over the application of crop models to properly account for yield losses resulting from ECEs and the validity in assessing long-term impacts of ECEs under climate change (Schauberger et al., 2017).

Statistical models use various regression methods to link historical yields to historical climate data which are then used to make predictions about yields under altered climate conditions (Schlenker and Roberts, 2009). They are easy to handle and relatively easy to compute. With the increasing availability and improved quality of observed data, statistical models usually have a high level of accuracy (Folberth et al., 2019; Innes et al., 2015). Moreover, newly emerging machine learning algorithms may improve the ability of statistical models to explore climate-yield relationships (Chlingaryan et al., 2018). Machine learning algorithms are capable of disentangling the effects of co-linear climate variables and analysing hierarchical and nonlinear relationships between the predictors and the response variable through an ensemble learning approach (Shalev-Shwartz and Ben-David, 2014), which usually result in better performance compared to conventional linear regression models (Everingham et al., 2015; Feng et al., 2018; Jeong et al., 2016). However, a major limitation of statistical models is that they usually only provide a simple evaluation of impacts, rather than provide a deeper understanding of physiological constraints required to inform adaptation strategies (Roberts et al., 2017). Thus,

results from statistical models might sometimes be vague and ambiguous in aiding targeted development of adaptive practices.

In recent years, the value of combining both process-based crop models and statistical models is gaining recognition. Pagani et al. (2017b) incorporated outputs from the sugarcane model Canegro (Inman-Bamber, 1991) and agro-climatic indicators into multiple linear regressions to reproduce recorded yield. Their results showed that the combined model increased prediction accuracy by about 20% compared to each individual model. Everingham et al. (2016) obtained similar higher levels of accuracy by combining APSIM model and random forest (RF) algorithm. Guzmán et al. (2017) combined DSSAT model (Jones et al., 2003) and support vector regression model for a comprehensive assessment of groundwater variability to demonstrate that the hybrid model performed better in characterizing groundwater variability. In evaluating the effects of ECEs on wheat yield, the existing state-of-art studies are still based on either crop models (Cammarano and Tian, 2018) or statistical models (García-León et al., 2019). Our study will firstly use the combination of the two kinds of models, i.e. crop models and machine learning techniques, to explore new insights of the effects of ECEs on wheat yield.

Australian wheat production is crucial to global food security, because Australia is one of the world's major grain exporters (Grundy et al., 2016). The New South Wales (NSW) wheat belt is the main wheat production area of south-eastern Australia, accounting for 27% of the national production (www.abares.gov.au, 2013-14). However, inter-annual wheat yields in this area are highly variable. Compared to the long-term mean, up to 1 $t \cdot ha^{-1}$ yield loss has occurred frequently over the past three decades (http://www.abs.gov.au/Agriculture). There is a very definite possibility that recurrent extreme events drive the inter-annual variability in wheat yields (Hughes et al., 2015). Moreover, drought and heat stress are projected to increase due to the changing climate (BOM and CSIRO, 2016). In this study, we combined the APSIM model (Holzworth et al., 2014) and RF algorithm to build a hybrid model to evaluate impacts of ECEs on wheat yields. The main objectives were to 1) develop a hybrid model to reproduce historical observed wheat yields in the NSW wheat belt, 2) quantify the relative importance of growth stage-specific drought, heat, and frost events in determining wheat yields, and 3) compare the yield differences projected by the APSIM alone and the hybrid model under future climate change.

6.2 Materials and method

6.2.1 Study sites

The NSW wheat belt (Figure 6-1) is located in south-eastern Australia, with its western border bounded by the semi-arid interior. It accounts for nearly 30% of the total areas planted to wheat in Australia (www.abares.gov.au, 2013-14), making it important in terms of both domestic and

international food security (Ray et al., 2015). Generally, wheat is mainly grown under rainfed conditions and the typical growing season is May to November (Gomez-Macpherson and Richards, 1995).

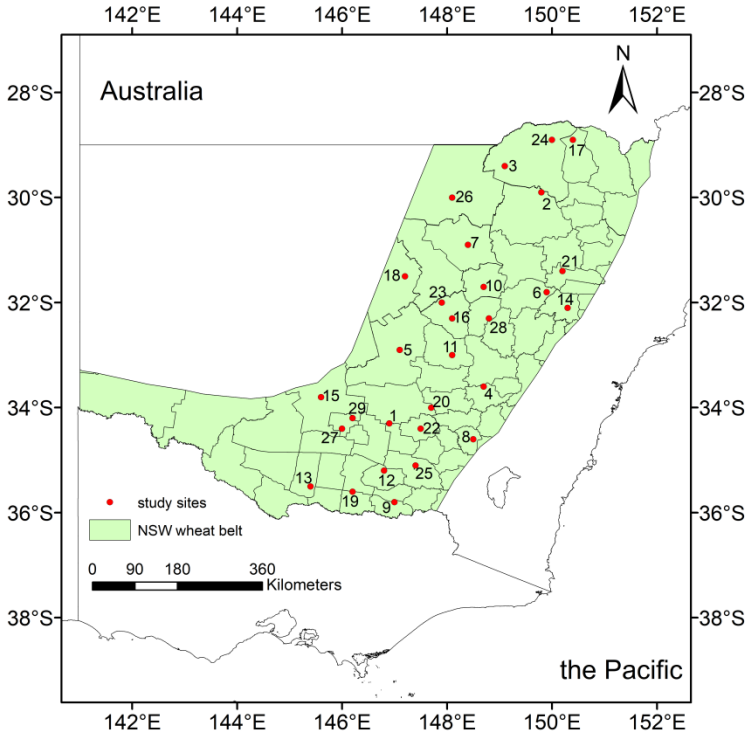


Figure 6-1 Locations of the 29 study sites in the New South Wales wheat belt in south-eastern Australia.

The NSW wheat belt is characterized by variable topography and climatic conditions. There is an east-west gradient in both elevation and precipitation/temperature. The eastern part of the wheat belt consists of mountains with elevation up to 1100 m and the western areas are mainly plains. Average growing season temperature ranges from 8.3 °C in the south-east to 17.1 °C in the north-west and the average growing season precipitation ranges from 171 mm in the south-west to 763 mm in the south-east in 1961-2000 (Wang et al., 2017a). In addition, the climate is characterized by large inter-annual variability mainly due to El Niño Southern Oscillation (Murphy and Timbal, 2008; Power et al., 1998). We used 29 sites that are listed in the Grains Research and Development Corporation National Variety Trials (GRDC-NVT, http://www.nvtonline.com.au/) and also located in the wheat belt. These sites are scattered throughout most of the wheat belt (Figure 6-1) to represent the range of agro-climatic zones across this area. We used NVT datasets because these trials were conducted in recent years and had detailed experimental records to calibrate and validate our model. A brief description of the 29 study sites, including climate and annual mean wheat yield, is shown in Table 6-1.

Table 6-1 A brief description of the 29 study sites used in the study, including location, Soil No.(details at http://www.asris.csiro.au/), GSR (mm), GST (°C), HY, HDR and AMWY (t ha^{-1}).

ID	Site	Soil	GSR	GST	HY	HDR	AMWY
1	Beckom	543	298	12.9	2009-2016	-	3.12
2	Bellata	83	307	15.9	2009-2011,2013-2016	-	3.77
3	Bullarah	126	252	17.1	2009,2010,2013,2016	-	3.48
4	Canowindra	703	342	12.6	2008,2011,2013,2014,2017	-	3.73
5	Condobolin	688	290	12.3	2008,2010-2014,2016	2016	2.66
6	Coolah	868	507	10.9	2008-2012,2014-2016	2008-	4.82
7	Coonamble	247	267	15.4	2008-2014,2016,2017	-	3.44
8	Galong	545	392	10.9	2008,2010,2011,2013-	2016	4.52
9	Gerogery	176	386	12.2	2008-	2017	5.15
10	Gilgandra	249	313	14.3	2008-2013,2015-2017	-	3.35
11	Goonumbla	193	331	13.6	2008-2013,2015,2016	-	4.08
12	Lockhart	539	312	12.9	2008,2009,2011-	2016	3.53
13	Mayrung	538	251	13.1	2010,2016	-	5.64
14	Merriwa	868	308	13.3	2008-2016	-	3.91
15	Merriwagga	696	233	14.2	2010,2012,2015,2016	-	3.13
16	Narromine	686	289	14.5	2016	-	5.36
17	North Star	237	270	16.3	2010,2011,2016	-	4.16
18	Nyngan	246	253	15.5	2011-2017	-	2.80
19	Oaklands	186	306	12.8	2008-2017	-	3.60
20	Quandialla	693	317	13.1	2008,2010-2016	-	3.91
21	Spring	127	314	14.4	2008,2012,2013,2015,2016	2009-	4.47
22	Temora	913	305	12.2	2010,2011,2013-2016	-	4.16
23	Trangie	683	273	14.7	2010-2013,2015,2017	2008	2.97
24	Tulloona	865	263	16.7	2009-2013,2015,2016	-	3.41
25	Wagga	498	364	12.1	2010-2016	2010-	3.52
26	Walgett	1016	239	16.3	2014-2016	-	3.17
27	Willbriggie	697	249	13.7	2010	-	5.66
28	Wongarbon	685	347	13.5	2008-2011,2016	2008	3.54
29	Yenda	697	265	13.8	2015,2016	-	4.22

*Note: GSR: growing season rainfall; GST: growing season temperature; HY: harvest year; HDR: heading date record year; AMWY: and annual mean wheat yield.

6.2.2 Climate data

Historical (2008-2017) daily climate data (rainfall, maximum and minimum air temperature, and solar radiation) for the 29 study sites were downloaded from Scientific Information for Land Owners patched point dataset (SILO-PPD, https://silo.longpaddock.qld.gov.au/) (Jeffrey et al., 2001).

Future (2020-2100) climate data were obtained for each of the 29 sites from 34 different global climate models (GCMs, Table 6-2). The monthly climate data from these GCMs are provided by different climate modeling institutions all over the world. Detailed descriptions of these GCMs can be found at the Coupled Model Inter-comparison Project phase 5 (CMIP5, https://cmip.llnl.gov/cmip5/). Generally, raw GCMs are at coarse temporal (monthly) and spatial (100-300 km grid spacing) resolutions and therefore cannot be directly used to feed site-based crop models. Here we used a statistical downscaling method (NWAI-WG, (Liu and Zuo, 2012)), to downscale the monthly gridded data simulated by raw GCMs to daily climate data for each of the 29 sites. This approach has been frequently used in recent climate change research (Liu et al., 2017; Wang et al., 2015; Wang et al., 2017b). In addition, two representative concentration pathways (RCP4.5 and RCP8.5) are available in the GCMs dataset and were utilized in this study.

Table 6-2 List of 34 GCMs under RCP4.5 and RCP8.5 future climate scenarios used in this study for statistical downscaling outputs of the 29 sites over the New South Wales wheat belt in south-eastern Australia. Details of the 34 GCMs can be found at https://cmip.llnl.gov/cmip5/availability.html.

Model ID	Name of GCM	Abbr. of	Institute ID	Country
1	ACCESS1-0	AC1	CSIRO and BoM	Australia
2	ACCESS1-3	AC2	CSIRO and BoM	Australia
3	BCC-CSM1-1	BC1	BCC	China
4	BCC-CSM1-1-m	BC2	BCC	China
5	BNU-ESM	BNU	GCESS	China
6	CanESM2	CaE	CCCMA	Canada
7	CCSM4	CCS	NCAR	USA
8	CESM1-BGC	CE1	NSF-DOE-NCAR	USA
9	CESM1-CAM5	CE2	NSF-DOE-NCAR	USA
10	CESM1-WACCM	CE5	NSF-DOE-NCAR	USA
11	CMCC-CM	CM2	CMCC	Europe
12	CMCC-CMS	CM3	CMCC	Europe
13	CNRM-CM5	CN1	CNRM-GAME	France
14	CSIRO-Mk3-6-0	CSI	CSIRO-QCCCE	Australia
15	EC-EARTH	ECE	EC-EARTH	Europe
16	FIO-ESM	FIO	FIO	China
17	GISS-E2-H	GE1	NASA GISS	USA
18	GISS-E2-H-CC	GE2	NASA GISS	USA
19	GISS-E2-R	GE3	NASA GISS	USA
20	GFDL-CM3	GF2	NOAA GFDL	USA
21	GFDL-ESM2G	GF3	NOAA GFDL	USA
22	GFDL-ESM2M	GF4	NOAA GFDL	USA

23	HadGEM2-AO	Ha5	NIMR/KMA	Korea
24	INM-CM4	INC	INM	Russia
25	IPSL-CM5A-LR	IP1	IPSL	France
26	IPSL-CM5A-MR	IP2	IPSL	France
27	IPSL-CM5B-LR	IP3	IPSL	France
28	MIROC5	MI2	MIROC	Japan
29	MIROC-ESM	MI3	MIROC	Japan
30	MIROC-ESM-CHEM	MI4	MIROC	Japan
31	MPI-ESM-LR	MP1	MPI-M	Germany
32	MRI-CGCM3	MR3	MRI	Japan
33	NorESM1-M	NE1	NCC	Norway
34	NorESM1-ME	NE2	NCC	Norway

6.2.3 In-situ trial data

The GRDC-NVT is a national program of comparative crop variety testing with standardized trial management, data generation, collection and dissemination. Crop variety trial data from the GRDC-NVT have been frequently used by the scientific community in recent years (Dreccer et al., 2018; Shen et al., 2018; Zheng et al., 2012). In the present study, we used in-situ wheat trial data (2008-2017) for the 29 sites across the wheat belt. These data include sowing date, heading date (only available at several sites and years, Table 6-1), and yield for dozens of wheat varieties. In addition, soil nutrient status (including total nitrogen, phosphorous, organic carbon, pH, and conductivity) and fertilization practice (including date and fertilizer type) were also available. We chose four varieties, i.e. *Suntop*, *Sunvale*, *Ventura*, and *Wallup*, which are widely cultivated across the wheat belt and have also been well parameterized on the vernalization and photoperiod response in the APSIM model. The parameters for vernalization sensitivity were 2.0, 2.8, 1.5, and 1.5 for each variety respectively; and for photoperiod sensitivity, 3.5, 3.0, 3.0, and 3.5 respectively. In the APSIM-Wheat module, photoperiod and vernalization are two important factors that determine wheat phenology (Keating et al., 2003). As a result, the four selected varieties tended to have different dates of growing stages which would result in differences in responses to growth stage-specific ECEs. Thus, using the four varieties will enable a comprehensive evaluation of impacts of various ECEs on wheat yields. In addition, as each variety was only available in parts of years, we eventually collected 516 trials data. The yield data did not show obvious skewed distribution (Figure 6-S1) and can be used for regression analysis. In addition, as these data were experimental data and very recent, no significant technological trend was detected after examination. Thus, no de-trending method was applied to remove factors (e.g., changes in management practices, pesticide application) not reproduced by the modelling solution (Pagani et al., 2017b).

6.2.4 APSIM descriptions

APSIM (Agricultural Production System sIMulator) version 7.7 (http://www.apsim.info/) (Holzworth et al., 2014) was used to simulate historical and future wheat phenology, final biomass and yield at the 29 sites. The APSIM wheat module was developed in Australia and has been applied to numerous studies across the Australian wheat belt (Asseng et al., 2011; Chenu et al., 2013; Lobell et al., 2015). In particular, the wheat module has been shown to adequately simulate a number of processes at a daily time step, including phenological development, biomass accumulation, and yield formation, for multiple cultivars, soil moisture and nutrient status, weather conditions, and farming management practices (Jin et al., 2017). Phenological development is determined by temperature and cultivar features as mentioned above. Biomass accumulation is determined by both radiation interception and soil water limitation, while yield formation is calculated based on a simple assimilate partitioning rule (Tao et al., 2017).

In the APSIM wheat module, several simplified descriptions of certain processes have been defined to regulate the effects of ECEs on crop growth (Barlow et al., 2015). Frost and heat events are incorporated by stress functions which can lead to leaf senescence, while drought (water stress) events are defined as functions that can restrain leaf expansion and biomass accumulation. Detailed descriptions of these functions have been described by Zheng et al. (2014). It should be noted that these functions are mostly simple and linear. Moreover, many other damages of ECEs on crop growth are not considered in APSIM, such as frost and heat-induced sterility around anthesis. In other words, APSIM simplifies several processes and ignores some limitations on yields, which may result in poorly modelling ECEs (Barlow et al., 2015).

6.2.5 APSIM simulations

The four varieties used in this study are available in the APSIM variety bank. In addition, there are more than 800 soil profiles in the APSoil database (Dalgliesh et al., 2006) available for Australian agricultural areas. Most of these soils have already been parameterized for modelling wheat. We finally selected 28 soil profiles (Table 6-1) that are geographically closest to study sites. A same soil was used in Coolah and Merriwa sites. Detailed hydraulic properties of all these soils can be found at http://www.asris.csiro.au/.

For the calibration dataset (2008-2017), we set up the APSIM simulations strictly according to the NVT trial data (variety, sowing date, soil nutrient status and fertilization practice). The output yield and phenology data were then directly used to compare with observed data. For the climate change impacts (1961-2100), APSIM simulations were driven by the 34 downscaled GCMs for each of the 29 sites. The four varieties were simulated for each site and for each GCM. For the management options, a sowing window starting on the 1st of May and ending on the 30th of June

was used with the option "must sow". The fertilizer at sowing was 100 kg/ha of urea (equivalent to 46 kg/ha of N) under each future climate scenario. All other options were left as the defaults.

Effects of elevated CO_2 concentration were considered in simulations of future scenarios. In APSIM, CO_2 influences plant growth through regulating transpiration efficiency, radiation use efficiency, and critical leaf nitrogen concentration. However, APSIM has no function to generate time-varying values of CO_2 concentration. Thus, we added a function to APSIM so that it could calculate yearly atmospheric CO_2 concentrations through empirical relations of calendar year (Liu et al., 2014; Liu et al., 2017). For RCP4.5 scenario, the atmospheric CO_2 concentration was calculated by:

$$[CO_2]_{year} = 650.18 + \frac{0.000075326 \times y - 0.16276}{0.00022299 - \frac{727.97}{y^2}} - 0.00018747 \times (y - 2045)^3 \quad (6\text{-}1)$$

For RCP8.5, it was fitted by:

$$[CO_2]_{year} = 1034.3 + \frac{267.78 - 1.6188 \times y}{4.0143 + \frac{53.342}{y^{5.2822}}} + 21.746 \times \left(\frac{y - 2010}{100}\right)^3 + 100.65 \times \left(\frac{y - 1911}{100}\right)^3 \quad (6\text{-}2)$$

where y is the calendar year from 1900 to 2100 (i.e. y = 1900, 1901, …, 2100).

6.2.6 Climate extremes indicators

In APSIM, wheat cultivation is divided into 11 stages, i.e. sowing, germination, emergence, end of juvenile (EJ), floral initiation (FI), flowering (F), start of grain filling (SGF), end of grain filling, maturity, harvest rips, and end crop. In this study, we took EJ, FI, F, and SGF into consideration, as they represent the four main growing stages. In this study, we intended to evaluate impacts of three kinds of ECEs (Table 6-3) at the four main growing stages on wheat yields. The indicators for heat and frost are simple counts of days with maximum/minimum temperatures above/below fixed thresholds (Tashiro and Wardlaw, 1989; Zheng et al., 2012). The impact of water deficit was assessed using the ARID - Agricultural Reference Index for Drought (Woli et al., 2012). This drought index is a simple, general, soil-plant-atmosphere metric. It usually performs better than many other drought indices in agricultural drought evaluations (Woli et al., 2013).

$$\mathrm{ARID}_i = 1 - \frac{T_i}{\mathrm{ET}_{\mathrm{o},i}} \quad (6\text{-}3)$$

where i represents the ith day, T_i is the transpiration during the ith day (mm d^{-1}), and $ET_{o,i}$ is the reference evapotranspiration on the ith day (mm d^{-1}). When calculating ARID, $ET_{o,i}$ is assumed to be equal to potential evapotranspiration and can be estimated using the Priestley and Taylor

(1972) method. T_i is estimated through a macroscopic modeling approach which is based on the water content. Detailed descriptions and calculation processes can be found in Woli et al. (2012). ARID values fall between 0 and 1. Values higher than 0.6 are usually recognized as high water stress, thus we chose 0.6 as the threshold to evaluate daily drought condition.

To calculate these indicators, we first ran APSIM simulations and obtained wheat phenology information, including dates and duration of wheat growing stage. Then, according to the phenology information, stage-specific ECEs indicators were calculated or counted out. After calculation, we found that heat events rarely happened at the EJ stage and frost events rarely happened at the F and SGF stages during the historical period (2008-2017). Thus, we eventually selected 9 extreme events at the four growing stages (Table 6-3).

Table 6-3 List of extreme climate events used in this study. Heat events were calculated at FI, F, and SGF stages. Frost events were calculated at EJ and FI stages. Drought events were calculated at EJ, FI, F, and SGF stages. Thus, totally 9 weather extreme indicators were used in this study.

Extreme event	Description	Growth stage
Heat	Number of days with daily maximum temperature >27°C	FI, F, SGF
Frost	Number of days with daily minimum temperature <0°C	EJ, FI
Drought	Number of days with ARID >0.6	EJ, FI, F, SGF

*Note: EJ: end of juvenile; FI: floral initiation; F: flowering; SGF: start of grain filling; ARID: Agricultural Reference Index for Drought.

6.2.7 Statistical models

RF (random forest) is a nonparametric and ensemble learning algorithm originated from classification and regression trees (Breiman, 2001). It is a nonparametric technique which builds multiple decision trees and combines them together to obtain a prediction. Thus, the RF usually presents good accuracy in spite of the presence of missing values and outliers (Elavarasan et al., 2018). Moreover, the RF can approximate functions with both linear and non-linear relations and can also identify the relationship between the response and a variable, which is conditional upon other variables (Hoffman et al., 2018). Given that the effects of ECEs on crop yields are often nonlinear (Lobell et al., 2011; Schlenker and Roberts, 2009), RF is expected to perform well in assessing the nonlinear relationship. Our previous studies (Feng et al., 2018; Feng et al., 2019; Wang et al., 2018) have demonstrated that the RF model usually performed better than many other machine learning techniques, in agricultural-based applications. In addition, RF mode is able to provide the relative importance of each predictor in determining response variable. Therefore, in the present study, we intended to take advantage of RF to enhance the ability of APSIM in simulating the effects of ECEs on wheat yields.

We also used the MLR to build the hybrid model with the APSIM model. MLR is a commonly used regression method to model the linear relationship between the independent variables and

the dependent variable. It is considered to be the extension of ordinary least-squares that involves more than one explanatory variable. It is easy to understand and implement, but usually limited in disentangling the nonlinear relationships between the predictors and the response.

6.2.8 Hybrid-modelling approach

Figure 6-2 illustrates the processes of combining the APSIM and RF (or MLR) models in our study. First, the APSIM wheat module was run to simulate wheat phenology, biomass and yield based on NVT trial datasets. The outputs of phenology date were then used as references for the calculation of the 9 indicators at the four growth stages. Lastly, APSIM simulated biomass and the 9 indicators were applied as predictors in RF (or MLR) for estimating wheat yield. In this study, we proposed the RF (or MLR) model as an external modification which was expected to help improve the performance of APSIM model in simulating the effects of growth stage-specific ECEs.

We performed the RF model using the R package "randomForest" (Liaw and Wiener, 2002). Two parameters were needed to be determined before the implement of the model, i.e. n_{tree} (the number of trees to grow in the forest) and m_{try} (the number of randomly selected predictor variables at each node). We set the n_{tree} as the default values of 500. While for m_{try}, it could affect the model accuracy. As the dataset was not large, we adopted a trial and error analysis to determine the value of m_{try}. Values of 1 to 10 were tried and 5 was chosen finally as it leaded to a little higher model accuracy. The relative importance of variables was assessed through the "%IncMSE" metric in the RF model. In addition, the MLR model was performed using the R package "Rattle" (Williams, 2011).

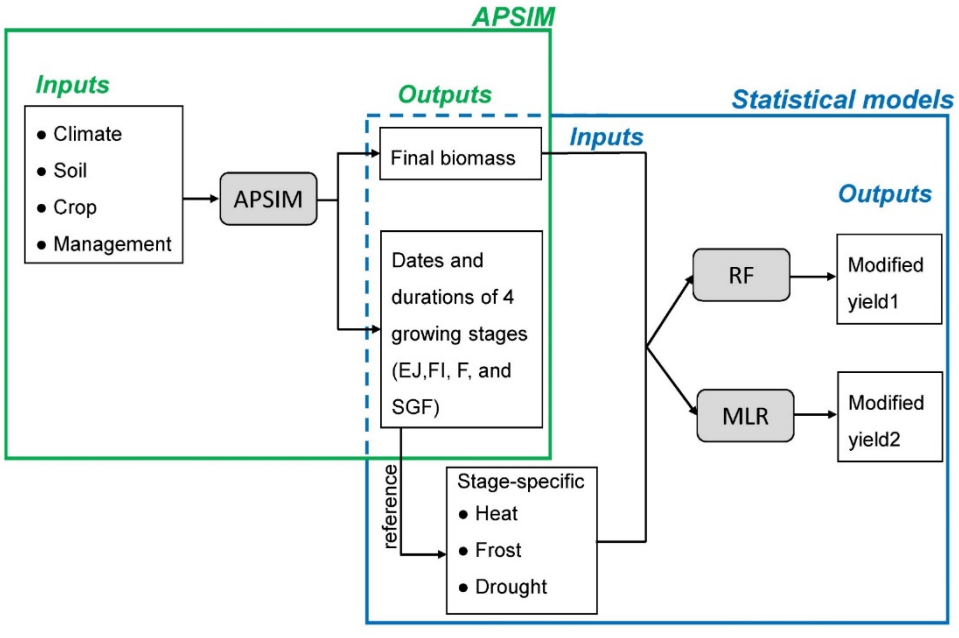


Figure 6-2 Diagram of the input and output per model for the APSIM+RF (or MLR) hybrid model applied in this study. EJ: end of juvenile; FI: floral initiation; F: flowering; SGF: start of grain filling; RF: random forest; MLR: multiple linear regression.

6.2.9 Model performance assessment

The NVT trial data (516 trials, 2008-2017) were used to calibrate the models. The output yields from the APSIM model were directly used to compare with the observed data. While for the RF model, a 10-fold cross validation approach was applied to the 516 data. The coefficient of determination (R^2) and root mean square error (RMSE) were used for model evaluation following:

$$R^2 = \left(\frac{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} \sqrt{\sum_{i=1}^{n} \left(P_i - \overline{P}\right)^2}} \right)^2 \quad (6\text{-}4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(O_i - P_i\right)^2}{n}} \quad (6\text{-}5)$$

Where n is the number of samples, O_i and P_i denote observed and simulated values, and $\overline{O}$ represents the mean of observed values. Generally, the model with higher R^2 and lower RMSE is considered to be the more accurate model.

6.3 Results

6.3.1 Model performance

The performance of the APSIM wheat module was evaluated by comparing simulated and observed wheat grain yield and flowering date (Figure 6-3). Agreement between simulations and observations was described by the root mean square error (RMSE), the coefficient of determination (R^2) and the slope of the regression lines. As shown in Figure 6-3a, the simulated flowering dates were consistent with observed dates, with an RMSE of 5.01 days (R^2=0.82, y=0.72x+76.71, P<0.01), suggesting that the APSIM was able to provide a satisfactory estimation of wheat flowering dates. As the flowering stage was generally viewed as the indicative stage of the entire wheat phenology, it was likely that APSIM could also provide fairly good estimations of the other three growth stages. This laid a foundation for our subsequent calculation of stage-specific ECEs indicators. For wheat yields, the model was able to explain 61% of the variation and the RMSE was 0.86 t ha^{-1}. This was a common and acceptable result for large-scale crop model simulations (Jin et al., 2017), even though the accuracy was not high. In general, inaccurate simulations were due to the absent or rough assumptions around certain factors, including pest, diseases, and weather extremes as we mentioned above. In subsequent analysis, we managed to increase the accuracy through improving the ability of simulating ECEs.

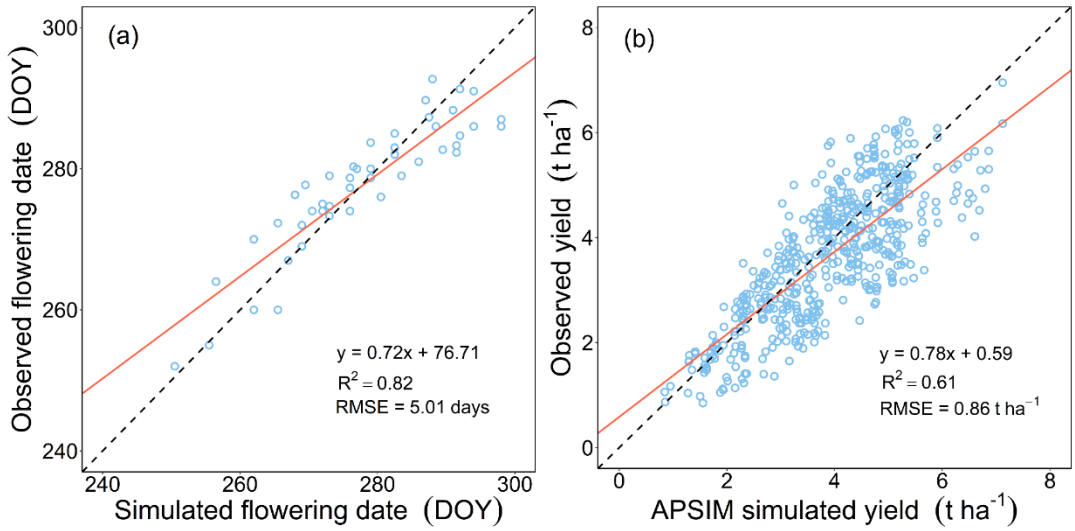


Figure 6-3 Comparison of observed and APSIM simulated values of grain yield and flowering date from 2008 to 2017 at the 29 sites across the New South Wales wheat belt. Totally 516 yield data and 47 flowering date data were collected from National Variety Trials of Australia (see text for more detail about this dataset). Dashed lines are the 1:1 ratio line. Red lines are the linear regression fit.

We used the MLR model and the RF model as external modification on the APSIM model and created two hybrid models to predict wheat yield. Compared to the APSIM model alone, both hybrid models showed higher accuracy in reproducing the observed yields (Figure 6-4). The APSIM+RF hybrid model explained 81% of the variation in observed yields, an increase of 33% compared to the APSIM model. It also reduced the RMSE by 0.32 t ha^{-1}. Moreover, the slope of the regression function was close to 1.0, meaning that the APSIM+RF hybrid model was unbiased in simulation of wheat yields for our study area. While for the APSIM+MLR hybrid model, its model accuracy increased slightly compared to the APSIM model alone and was far below the APSIM+RF hybrid model. Thus, the external modification using the RF model with ECEs indicators could greatly improve the performance of the APSIM model.

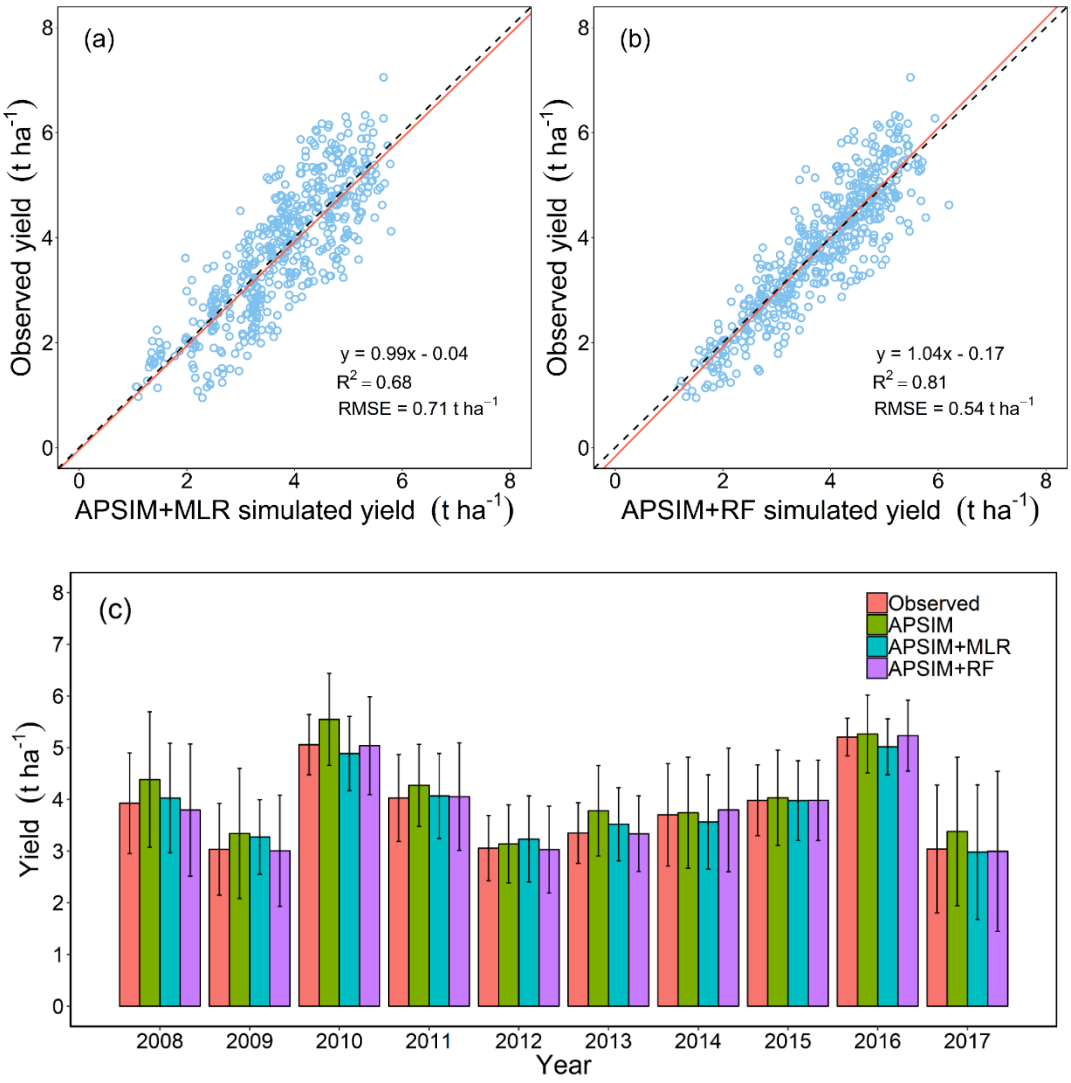


Figure 6-4 Comparison of observed, APSIM simulated, APSIM+MLR simulated, and APSIM+RF simulated wheat yields from 2008 to 2017 at the 29 sites across the New South Wales wheat belt. (a) observed vs. APSIM+MLR hybrid model simulated. (b) observed vs. APSIM+RF hybrid model simulated. (c) time series of the four kinds of yields. In (c), error bars indicate the standard deviation from yields at the 29 sites.

Figure 6-4c shows the time series of the three kinds of wheat yields, i.e. observed, APSIM simulated, and the two hybrid models simulated, from 2008 to 2017. The observed yields ranged from 3.0 t ha^{-1} to 5.2 t ha^{-1}, with the greatest inter-annual variability. In general, all models' simulations successfully captured the temporal pattern of the observed wheat yields. However, the APSIM model tended to slightly overestimate the yields in almost every year. This was particularly evident in 2008, 2010, and 2013, where an overestimate of up to 0.5 t ha^{-1} occurred. These overestimations can be attributed to an underestimate of the effects of ECEs on wheat yields. Through incorporating ECEs indicators, the APSIM+RF hybrid model succeeded in making the simulated yields more consistent with the observed yields.

6.3.2 Effects and projected changes of ECEs

The historical occurrence of ECEs is shown in Figure 6-5. Drought was the most commonly occurring ECE during the historical period. In general, the four growth stages, i.e. EJ, FI, F, and SGF, usually last for 72-88, 37-48, 6-9, and 25-34 days respectively for the four cultivars. Thus, nearly one quarter, half, and half of the EJ, F, and SGF stages respectively experienced drought conditions. While for heat, it commonly occurred around and post anthesis. The SGF stage was most vulnerable to heat stress and up to two-thirds of this stage might be under heat threat. Frost events mainly occurred during the EJ stage and a few frost events also occurred at the FI stage. Figure 6-4b and Figure 6-5 show that there was an obvious and direct relationship between wheat yields and the occurrences of ECEs. Wheat yields were much lower during years with a higher occurrence of ECEs, such as 2012 and 2017.

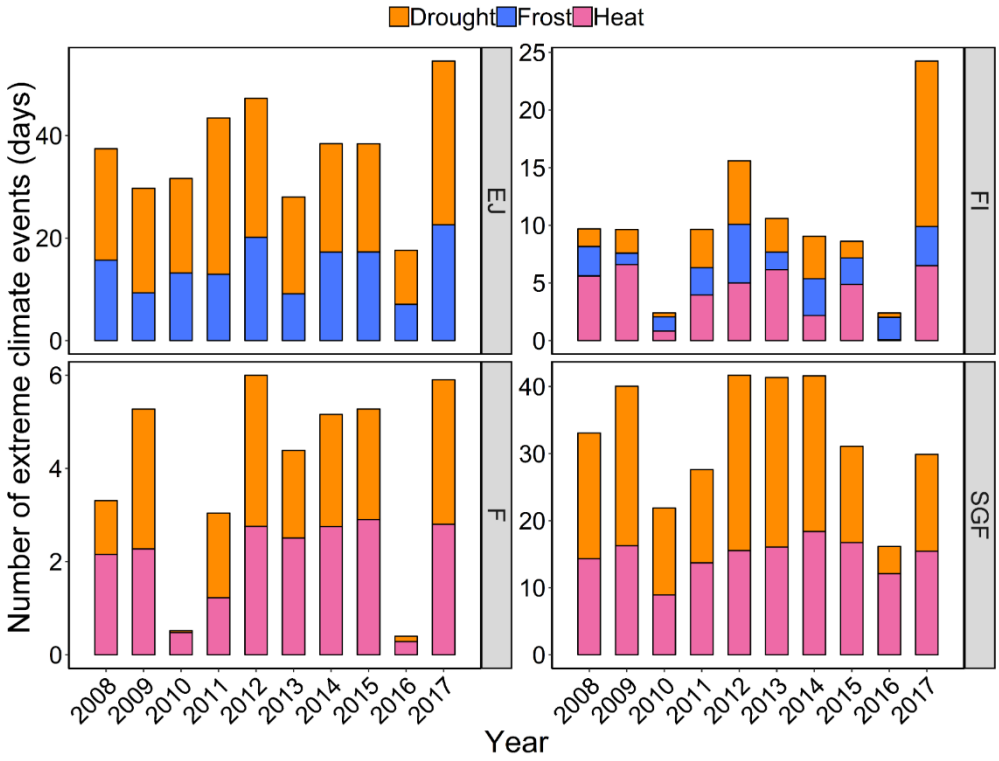


Figure 6-5 Number of extremes climate events occurred from 2008 to 2017. Values for each year were averaged values of the 29 study sites.

According to the above results, the RF model could potentially improve the performance of the APSIM model in simulating the effects of ECEs. We then obtained the relative importance (percentage values in Figure 6-6) and the marginal effect (lines in Figure 6-6) of each predictor from the RF model. The trend of the line, rather than the actual values, describes the nature of the dependence between the response and the predictor variables. All ECEs, except frost during the EJ stage, had negative effects on wheat yield. Drought events occurred at the SGF and EJ stages had high importance values, meaning that they were more harmful to wheat growth. The third

was heat events at the FI stage. Thus, even though heat events were more common during the SGF stage (Figure 6-5), they tended to cause more yield losses if occurred at the FI stage. Frost events at the EJ stage had positive effects on wheat yield. This may be due to that winter wheat plant is capable of withstanding extreme cold before the initiation of flowering (Fowler and Carles, 1979). Moreover, the plant also requires enough exposure to cool temperature for jarovization, which will affect subsequent growth and development (Robertson et al., 1996). In addition, the responses of wheat yield to different ECEs at a same stage were different. For example, the F stage was more vulnerable to heat rather than drought, while the SGF stage was more vulnerable to drought rather than heat.

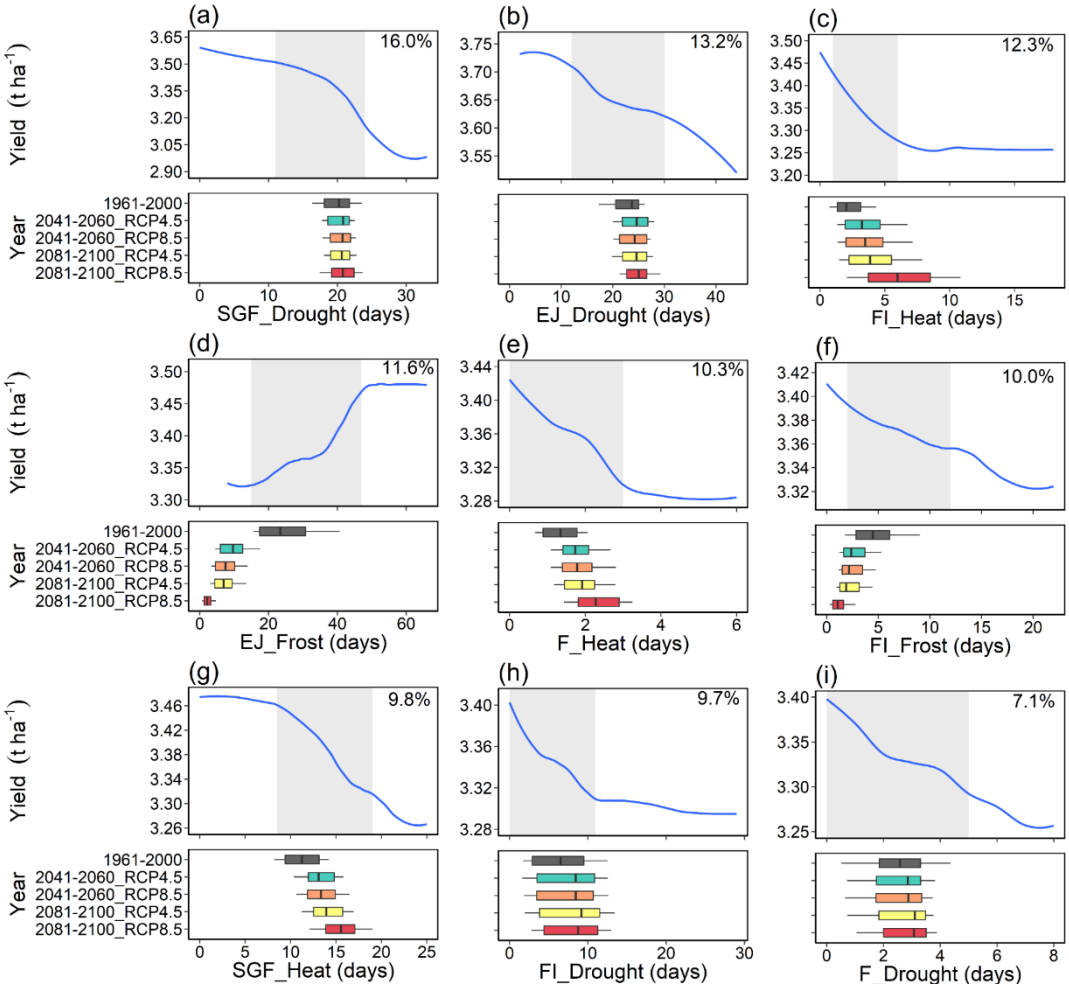


Figure 6-6 Partial dependence of wheat yield change on extreme climate events and projected changes in each event under RCP4.5 and RCP8.5. The random forest model could provide partial dependence of the change in the response (blue lines) for selected predictors, when accounting for the average effect of all other driver predictors. The blue lines are smoothed representations of the response, with fitted values (model predictions) for the calibration data. The trend of the line, rather than the actual values, describes the nature of the dependence between the response and predictors. The shaded area denotes calibration data between the 10th and 90th percentile. The

percentages values denote the relative importance of each predictor generated from the random forest model. The box plots indicate the occurrences of extreme climate events during the baseline (1961-2000) period and two future periods (2041-2060 and 2081-2100) based on the 34 downscaled GCMs. Box boundaries indicate the 25th and 75th percentiles across GCMs, whiskers below and above the box indicate the 10th and 90th percentiles. The black lines within each box indicate the multi-model median. EJ, FI, F, and SGF indicate end of juvenile, floral initiation, flowering, and start of grain filling, respectively.

The shaded area (Figure 6-6) denotes the 10^{th} to 90^{th} percentile of each variable in the calibration dataset. All lines have a sharp change in the shaded area, meaning that a small change of each ECE may have a large impact on wheat yield. As shown in the boxplots in Figure 6-6, drought events were projected to have small increase in occurrence at the four study stages, but heat events were likely to increase significantly. In particular, under RCP8.5, heat events during the end of the 21^{st} century may double compared to the baseline period. Frost events were likely to decrease, by >10 days at the EJ stage and by 1-4 days at the FI stage. However, the decrease of frost events at the EJ stage might also cause yield reductions. Thus, in general, more yield losses were indicated as a result of changes in future ECEs.

6.3.3 Differences between APSIM projected and the hybrid model projected future wheat yields

As the RF model performed better in improving the performance of the APSIM model compared to MLR. We then used the APSIM model and the APSIM+RF model to evaluate the impacts of future climate change on wheat yield in the study area. Projected changes in simulated wheat yield from the APSIM model and the hybrid model for two of the study sites are shown in Figure 6-7 (other sites in Figure 6-S2). We calculated changes of simulated wheat yields for each site and found that trends from APSIM-simulated wheat yield differed in different sites. However, the APIM+RF hybrid model-simulated yield was projected to decrease in all study sites. For example, Figure 6-7 shows the APSIM-simulated yields were projected to decrease by 0.5-3% at Bellata but increase by 1.5-3% at Mayrung. This might be due to different soil conditions and climate projections at the two sites. However, the APSIM+RF hybrid model-simulated wheat yields were projected to decrease at both sites. According to the multi-model ensemble mean values (2041-2060), the APSIM+RF hybrid model-simulated wheat yields were 4% and 3% lower than the baseline levels at Bellata and Mayrung, respectively. Moreover, the yield reductions magnified over time. The differences between the two models were mainly caused by different responses to ECEs. Climate change might result in various trends of the APSIM projections at different sites, but the ECEs changes, especially the increase of heat and drought events (Figure 6-6), were most likely to reduce the yields to lower levels.

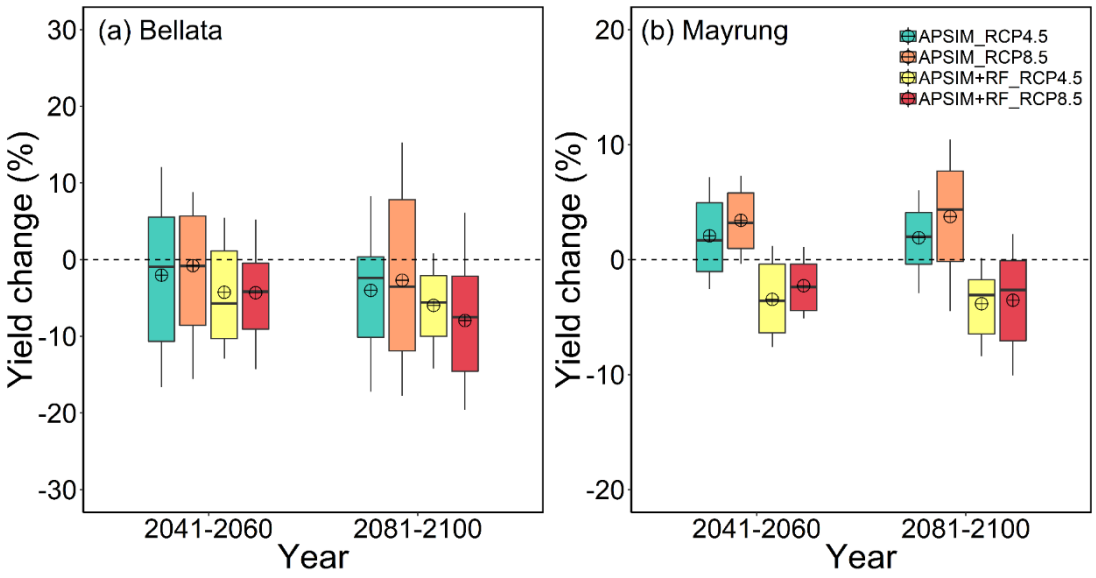


Figure 6-7 Projected changes in simulated wheat yield from the APSIM model and the APSIM+RF hybrid model for two of the study sites. Changes were estimated between two future periods (2041-2060 and 2081-2100) and the baseline period (1961-2000) under RCP4.5 and RCP8.5 based on the 34 downscaled GCMs. Box boundaries indicate the 25th and 75th percentiles across GCMs, whiskers below and above the box indicate the 10th and 90th percentiles. The black lines and crosshairs within each box indicate the multi-model median and mean respectively.

6.4 Discussion

The comparison between the results obtained from the APSIM model and the hybrid model showed that the APSIM+RF hybrid model is better at reproducing historical wheat yields. Using the RF algorithm as an external modification on the APSIM model outputs appears to improve the performance of the individual APSIM model (Figure 6-3 and Figure 6-4). Moreover, the RF model also outperformed the MLR model. Everingham et al. (2016) conducted a similar study through incorporating the biomass simulated by the APSIM model and several climate indices into the RF algorithm to simulate sugarcane yield and also obtained a high R^2 of ~0.8. The most likely explanation is that the external statistical model may help improve the performance of the crop model by simulating the effects of these selected climate indices. Using the crop model outputs as predictors may also improve the ability of the statistical model to consider more agro-climatic processes. Keating and Thorburn (2018) proposed the possible research trend of blending mechanistic and empirical statistical models. The approach outlined in this paper can be viewed a feasible method which may be easily extended to other wheat growing regions to obtain new insights to guide agricultural practices.

In recent years, researchers have been concerned with ECEs because of their remarkable damage to crops (Lesk et al., 2016; Rezaei et al., 2015). It was reported that nearly one-quarter of all damage and losses in the agriculture sector are caused by ECEs in many countries (FAO, 2015). As ECEs may impact during different stages of a crop cycle and cause varying degrees of yield losses, it is necessary to identify the most harmful ECEs in a particular region. In our study, we found that drought events during grain-filling and vegetative stages have the greatest adverse effects on wheat yields across a wide range of agro-climatic zones. Drought events around anthesis may have relatively small effects on wheat yields. One reason is probably that yield formation mainly occurs at the grain filling stage (Royo et al., 2006). Moreover, the number of drought events occurring around anthesis is relatively low. A short-term drought event is not able to severely reduce the final yield (Clarke et al., 1992).

In contrast, the impacts of heat and frost events on wheat yield are usually direct and rapid (Barlow et al., 2015). We found heat events around anthesis (FI_Heat and F_Heat), result in a large negative impact on wheat yields, even though they may only occur over a few days in the growing season , which is consistent with Balla et al. (2009)'s and Hays et al. (2007)'s studies. This is mainly because the sensitivity of the wheat plant to various ECEs varies with different growth stages (Hlaváčová et al., 2018). During anthesis, the awns or spikes start to emerge from the flag leaf as the grain begins to form. However, even a single frost or heat event may cause sterility and aborted grains, thereby reducing the number of grains in the inflorescence. While during the grain-filling period, as grains have formed, the adverse impacts from heat events will be reduced (Barlow et al., 2015).

The ECEs that have occurred during the historical period have already resulted in large yield losses, so their likely changes induced by climate change are needed to be assessed. In our study, we found that drought events are expected to have a slight increase, while heat events around anthesis and grain-filling periods will likely be more common. Semenov and Shewry (2011) obtained similar results in a study in Europe and demonstrated that wheat plants are expected to suffer more heat stress than drought in the future. Given the severe and rapid damage caused by heat stress around anthesis, a single day increase in heat events is likely to cause great wheat yield losses. On the other hand, while fewer frost days are expected because of global warming, we found that frost events during vegetative stages had a positive effect on wheat yield across our study region. Thus, in general, ECEs changes will result in future increased risks for wheat production in the study area.

Given the more unfavorable weather conditions in the future, the possible trend of future wheat yields is frequently discussed among researchers. The most commonly used method to assess climate impacts is a combination of process-based crop models and GCMs. For example, Qian et al. (2016) reported that an average increase in wheat yields of 26-37 % to be expected during

2041-2070 compared to a baseline period (1971-2000) in the Canadian Prairies using the CERES-Wheat model and two GCMs. Wang et al. (2017b) conducted a study in Australia using the APSIM model and 11 GCMs to demonstrate that there would be a decrease in the yield in the south-eastern Australian wheat belt throughout the 21st century. The majority of these previous and similar climate change impacts studies only used one single crop model. However, model comparison studies, especially for climate change studies, have emphasized the limited ability of crop models to account for ECEs. Sánchez et al. (2014) observed that a number of crop models adequately predict mean yields, but are less able to predict extreme low yields, due to their inability to handle ECEs. Hochman et al. (2012) reported that the APSIM model poorly accounted for ECEs such as severe frost and was also overly optimistic about water limited yield impacts in some seasons and locations. Therefore, it is questionable whether crop model-based projections accurately reflect the direction and magnitude of the effects of climate change on yield.

In our study, we used the hybrid model which systematically incorporated the APSIM model output and ECEs indicators through random forest algorithm to predict future wheat yields and found that a single APSIM model may have 1-10% overestimation for future yield projections compared to the APSIM+RF hybrid model (Figure 6-S2). The overestimation was mainly caused by underestimating the ECEs-induced yield losses. This is likely to be a common phenomenon in crop model projections, as the most popular crop models poorly account for impacts of ECEs (Eitzinger et al., 2013). However, as future ECEs are projected to increase in most part of the world (IPCC, 2012), previous projections based on crop models might overestimate the most likely yield level in the future. Thus, whether an increase or a decrease of wheat yield is projected in a particular region using crop models, the most likely achieved yields may be lower due to the underestimation of ECEs-induced yield losses.

Appropriate adaptation strategies can be developed in order to maintain and improve wheat yields in the face of current and future increasing ECEs in the NSW wheat belt. Two kinds of strategies are frequently discussed among researchers, i.e. minimizing and escaping the adverse effects of ECEs. In terms of minimizing the adverse effects of ECEs, the main approach is to increase the resistance of crops through breeding. From our study, the heat tolerance traits will be important. Stratonovitch and Semenov (2015) also reported heat tolerance around flowering in wheat as a key trait for increased yield potential in Europe under climate change. On the other hand, escaping the adverse effects of climate extremes is also a potential approach, which mainly aims to stagger the reproductive stages to avoid suffering ECEs. Adjusting sowing date is currently the most effective farming management practice to avoid negative impacts of ECEs. Optimising sowing dates can lead to suitable duration of the pre-anthesis period for accumulating biomass and suitable flowering and grain-filling windows without frost, heat, and terminal drought (Bell et al., 2014). Shortening the whole growth period may also help avoid suffering terminal heat and

drought events, but yield may also decline because of short growth length. Therefore, more studies that take into account both breeding and selection approaches as well as farming practices are required to maintain and improve wheat yields in face of increasing ECEs.

6.5 Conclusions

Impacts of heat, frost, and drought events on current and future wheat yields were analyzed in this study based on a hybrid model which incorporated the APSIM model output and climate extremes indicators into the Random Forest algorithm. The likely changes of the climate extremes were also discussed. The major conclusions are as follows:

(1) Drought and pre-anthesis heat events are the major climate extremes causing current wheat yield losses in the New South Wales wheat belt of south-eastern Australia.

(2) In general, future climate in the New South Wales wheat belt is expected to be more unfavorable. Drought events are projected to remain at historical levels, while heat events are projected to increase in the future. Frost events during the vegetative and pre-anthesis stages will decrease.

(3) Future yield projections from conventional process-based crop models might have a 1-10% overestimation because of the underestimation of climate extremes-induced yield losses.

In addition, the combination of process-based crop models and statistical models showed high performance in modelling extreme climate events and is worthy of consideration for future research. We believe this study would provide some useful information for local farmers and policy makers with respect to development of adaptation strategies in face of increased climate extremes under climate change.

6.6 Supporting information

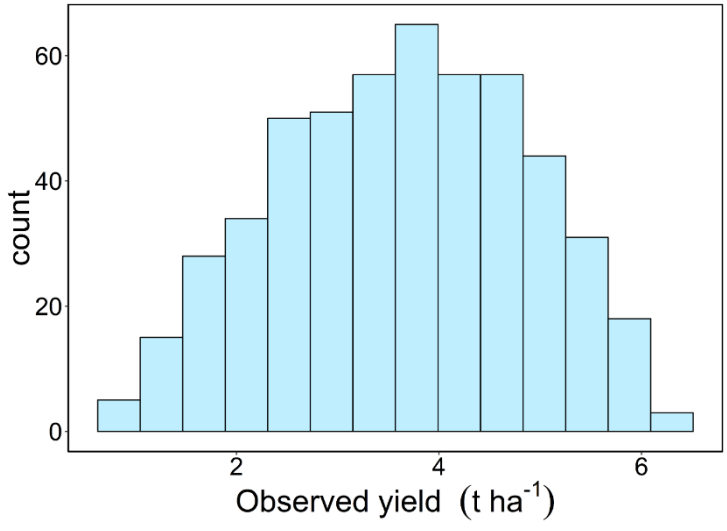


Figure 6-S1 Distribution of observed wheat yields (2008-2017) for the 29 sites across the New South Wales wheat belt.

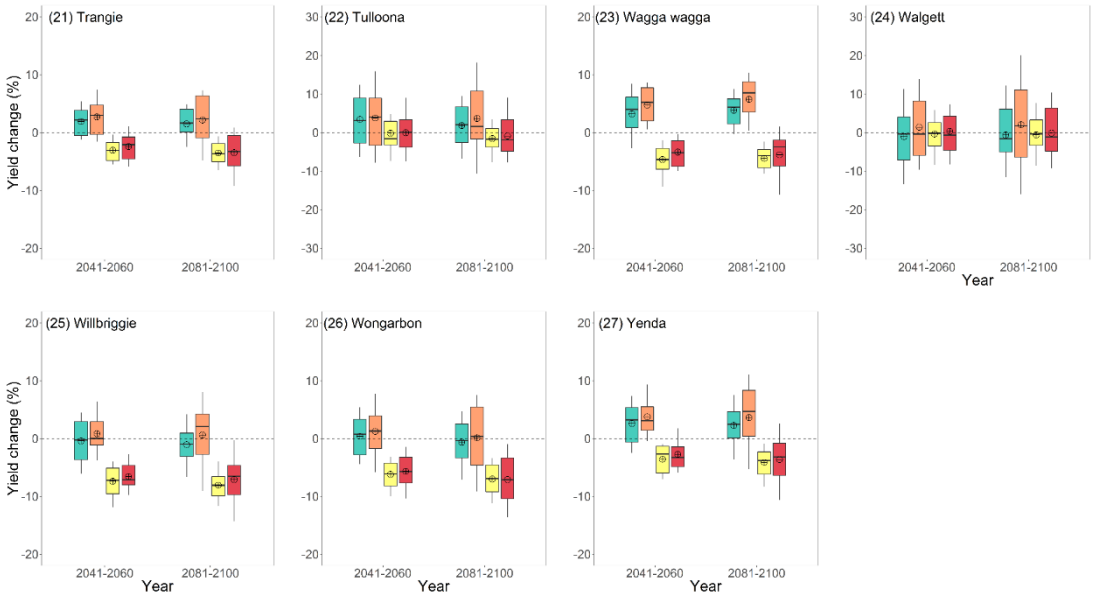


Figure 6-S2 Projected changes in simulated wheat yield per hectare from the APSIM model and the APSIM+RF hybrid model for the study sites. Changes were estimated between two future periods (2041-2060 and 2081-2100) and the baseline period (1961-2000) under RCP4.5 and RCP8.5 based on the 34 downscaled GCMs. Box boundaries indicate the 25th and 75th percentiles across GCMs, whiskers below and above the box indicate the 10th and 90th percentiles. The black lines and crosshairs within each box indicate the multi-model median and mean respectively.

6.7 Reference

Alexandratos, N. and Bruinsma, J., 2012. World agriculture towards 2030/2050: the 2012 revision, ESA Working paper FAO, Rome. No. 12-03. pp. 147.

Asseng, S., Foster, I. and Turner, N.C., 2011. The impact of temperature variability on wheat yields. Global Change Biology, 17(2): 997–1012.

Balla, K., Bencze, S., Janda, T. and Veisz, O., 2009. Analysis of heat stress tolerance in winter wheat. Acta Agronomica Hungarica, 57(4): 437-444.

Barlow, K.M., Christy, B.P., O'Leary, G.J., Riffkin, P.A. and Nuttall, J.G., 2015. Simulating the impact of extreme heat and frost events on wheat crop production: A review. Field Crops Research, 171: 109-119.

Bell, L., Lilley, J., Hunt, J. and Kirkegaard, J., 2014. Optimising grain yield and grazing potential of crops across Australia's high rainfall zone: a simulation analysis. 1. Wheat. Crop and Pasture Science.

BOM and CSIRO, 2016. State of the Climate.

Breiman, L., 2001. Random Forest. Machine Learning, 45: 5-32.

Cammarano, D. and Tian, D., 2018. The effects of projected climate and climate extremes on a winter and summer crop in the southeast USA. Agr Forest Meteorol, 248: 109-118.

Chaves, M.M., Pereira, J.S., Maroco, J., Rodrigues, M.L., Ricardo, C.P.P., Osório, M.L., Carvalho, I., Faria, T. and Pinheiro, C., 2002. How plants cope with water stress in the field? Photosynthesis and growth. Annals of botany, 89(7), pp.907-916.

Chenu, K., Deihimfard, R. and Chapman, S.C., 2013. Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt--spatial and temporal trends. New Phytologist, 198(3): 801-20.

Chenu, K. et al., 2017. Contribution of crop models to adaptation in wheat. Trends in plant science, 22(6): 472-490.

Chlingaryan, A., Sukkarieh, S. and Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture, 151, pp.61-69.

Clarke, J.M., DePauw, R.M. and Townley-Smith, T.F., 1992. Evaluation of methods for quantification of drought tolerance in wheat. Crop Science, 32(3): 723-728.

Dalgliesh, N., Wockner, G. and Peake, A., 2006. Delivering soil water information to growers and consultants, Proceedings of the 13th Australian Agronomy Conference. Australian Society of Agronomy Perth, Western Australia, pp. 10-14.

Dreccer, M.F., Fainges, J., Whish, J., Ogbonnaya, F.C. and Sadras, V.O., 2018. Comparison of sensitive stages of wheat, barley, canola, chickpea and field pea to temperature and water stress across Australia. Agricultural and Forest Meteorology, 248, pp.275-294.

Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rötter, R., Kersebaum, K.C. and Olesen, J.E., 2013. Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. The Journal of Agricultural Science, 151(6), pp.813-835.

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y. and Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: A survey. Computers and Electronics in Agriculture, 155, pp.257-282.

Everingham, Y., Sexton, J. and Robson, A., 2015. A statistical approach for identifying important climatic influences on sugarcane yields. In: Proceedings of the 37th Annual Conference of the Australian Society of Sugar Cane Technologists, 37(37): 8-15.

Everingham, Y., Sexton, J., Skocaj, D. and Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron Sustain Dev, 36(2).

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.

Feng, P., Wang, B., Liu, D.L., Xing, H., Ji, F., Macadam, I., Ruan, H. and Yu, Q., 2018. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. Climatic change, 147(3-4), pp.555-569.

Feng, P., Wang, B., Liu, D.L. and Yu, Q., 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, pp.303-316.

Fischer, R., 2011. Wheat physiology: a review of recent developments. Crop and Pasture Science, 62(2): 95-114.

Folberth, C., Baklanov, A., Balkovič, J., Skalský, R., Khabarov, N. and Obersteiner, M., 2019. Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. Agricultural and Forest Meteorology, 264, pp.1-15.

Fowler, D.B. and Carles, R.J., 1979. Growth, Development, and Cold Tolerence of Fall-acclimated Cereal Grains 1. Crop science, 19(6), pp.915-922.

Fuller, M.P., Fuller, A.M., Kaniouras, S., Christophers, J. and Fredericks, T., 2007. The freezing characteristics of wheat at ear emergence. European Journal of Agronomy, 26(4): 435-441.

García-León, D., Contreras, S. and Hunink, J., 2019. Comparison of meteorological and satellite-based drought indices as yield predictors of Spanish cereals. Agricultural Water Management, 213, pp.388-396.

Godfray, H.C.J. and Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Science, 327(5967): 812-8.

Gomez-Macpherson, H. and Richards, R., 1995. Effect of sowing time on yield and agronomic characteristics of wheat in south-eastern Australia. Australian Journal of Agricultural Research, 46(7): 1381-1399.

Grundy, M.J. et al., 2016. Scenarios for Australian agricultural production and land use to 2050. Agricultural systems, 142: 70-83.

Guarin, J.R., Asseng, S., Martre, P. and Bliznyuk, N., 2018. Testing a crop model with extreme low yields from historical district records. Field Crops Research.

Guzmán, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A.E. and Pote, J.W., 2017. An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels. Agricultural Systems.

Harrison, M.T., Tardieu, F., Dong, Z., Messina, C.D. and Hammer, G.L., 2014. Characterizing drought stress and trait influence on maize yield under current and future conditions. Global change biology, 20(3): 867-878.

Hays, D.B., Do, J.H., Mason, R.E., Morgan, G. and Finlayson, S.A., 2007. Heat stress induced ethylene production in developing wheat grains induces kernel abortion and increased maturation in a susceptible cultivar. Plant Science, 172(6): 1113-1123.

Hlaváčová, M., Klem, K., Rapantová, B., Novotná, K., Urban, O., Hlavinka, P., Smutná, P., Horáková, V., Škarpa, P., Pohanková, E. and Wimmerová, M., 2018. Interactive effects of high temperature and drought stress during stem elongation, anthesis and early grain filling on the yield formation and photosynthesis of winter wheat. Field crops research, 221, pp.182-195.

Hochman, Z., Gobbett, D., Holzworth, D., McClelland, T., van Rees, H., Marinoni, O., Garcia, J.N. and Horan, H., 2013. Reprint of "Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia". Field Crops Research, 143, pp.65-75.

Hoffman, A.L., Kemanian, A.R. and Forest, C.E., 2018. Analysis of climate signals in the crop yield record of sub‐Saharan Africa. Global change biology, 24(1), pp.143-157.

Holzworth, D.P., Huth, N.I., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D. and Brown, H., 2014. APSIM–evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software, 62, pp.327-350.

Howden, S.M., Soussana, J.F., Tubiello, F.N., Chhetri, N., Dunlop, M. and Meinke, H., 2007. Adapting agriculture to climate change. Proceedings of the national academy of sciences, 104(50), pp.19691-19696.

Hughes, L., Steffen, W., Rice, M. and Pearce, A., 2015. Feeding a hungry nation: Climate change, food and farming in Australia. Climate Council of Australia.

Inman-Bamber, N., 1991. A growth model for sugar-cane based on a simple carbon balance and the CERES-Maize water balance. South African Journal of Plant and Soil, 8(2): 93-99.

Innes, P.J., Tan, D.K.Y., Ogtrop, F.V. and Amthor, J.S., 2015. Effects of high-temperature episodes on wheat yields in New South Wales, Australia. Agricultural & Forest Meteorology, 208: 95-107.

IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental Panel on Climate Change. Journal of Clinical Endocrinology & Metabolism, 18(6): 586-599.

IPCC, 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ Modell Softw, 16(4): 309-330.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R. and Kim, S.H., 2016. Random forests for global and regional crop yield predictions. PLoS One, 11(6), p.e0156571.

Jin, Z., Zhuang, Q., Wang, J., Archontoulis, S.V., Zobel, Z. and Kotamarthi, V.R., 2017. The combined and separate impacts of climate extremes on the current and future US rainfed maize and soybean production under elevated CO2. Global change biology, 23(7), pp.2687-2704.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J. and Ritchie, J.T., 2003. The DSSAT cropping system model. European journal of agronomy, 18(3-4), pp.235-265.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N., Meinke, H., Hochman, Z. and McLean, G., 2003. An overview of APSIM, a model designed for farming systems simulation. European journal of agronomy, 18(3-4), pp.267-288.

Keating, B.A. and Thorburn, P.J., 2018. Modelling crops and cropping systems—Evolving purpose, practice and prospects. European Journal of Agronomy.

Kirono, D.G.C., Kent, D.M., Hennessy, K.J. and Mpelasoka, F., 2011. Characteristics of Australian droughts under enhanced greenhouse conditions: Results from 14 global climate models. Journal of Arid Environments, 75(6): 566-575.

Lesk, C., Rowhani, P. and Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production. Nature, 529(7584): 84-87.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3): 18-22.

Liu, D.L., Anwar, M.R., O'Leary, G. and Conyers, M.K., 2014. Managing wheat stubble as an effective approach to sequester soil carbon in a semi-arid environment: Spatial modelling. Geoderma, 214: 50-61.

Liu, D.L., Zeleke, K.T., Wang, B., Macadam, I., Scott, F. and Martin, R.J., 2017. Crop residue incorporation can mitigate negative climate change impacts on crop yield and improve water use efficiency in a semiarid environment. European journal of agronomy, 85, pp.51-68.

Liu, D.L. and Zuo, H., 2012. Statistical downscaling of daily climate variables for climate change impact assessment over New South Wales, Australia. Climatic Change, 115(3-4): 629-666.

Lobell, D.B., Bänziger, M., Magorokosho, C. and Vivek, B., 2011. Nonlinear heat effects on African maize as evidenced by historical yield trials. Nature climate change, 1(1): 42.

Lobell, D.B., Hammer, G.L., Chenu, K., Zheng, B., McLean, G. and Chapman, S.C., 2015. The shifting influence of drought and heat stress for crops in northeast Australia. Global change biology, 21(11), pp.4115-4127.

Murphy, B.F. and Timbal, B., 2008. A review of recent climate variability and climate change in southeastern Australia. International Journal of Climatology, 28(7): 859-879.

Pagani, V., Guarneri, T., Fumagalli, D., Movedi, E., Testi, L., Klein, T., Calanca, P., Villalobos, F., Lopez-Bernal, A., Niemeyer, S. and Bellocchi, G., 2017a. Improving cereal yield forecasts in Europe–The impact of weather extremes. European journal of agronomy, 89, pp.97-106.

Pagani, V., Stella, T., Guarneri, T., Finotto, G., Van den Berg, M., Marin, F.R., Acutis, M. and Confalonieri, R., 2017b. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. Agricultural Systems, 154, pp.45-52.

Porter, J. and Semenov, M., 2005. Crop responses to climatic variation. Philosophical Transactions of the Royal Society B Biological Sciences, 360(360): 2021-35.

Power, S., Tseitkin, F., Torok, S., Lavery, B., Dahni, R. and McAvaney, B., 1998. Australian temperature, Australian rainfall and the Southern Oscillation, 1910-1992: coherent variability and recent changes. Australian Meteorological Magazine, 47(2), pp.85-101.

Priestley, C. and Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon Weather Rev, 100(2): 81-92.

Qian, B., De Jong, R., Huffman, T., Wang, H. and Yang, J., 2016. Projecting yield changes of spring wheat under future climate scenarios on the Canadian Prairies. Theoretical and applied climatology, 123(3-4): 651-669.

Ray, D.K., Gerber, J.S., Macdonald, G.K. and West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nature Communications, 6(5989): 5989-5989.

Rezaei, E.E., Webber, H., Gaiser, T., Naab, J. and Ewert, F., 2015. Heat stress in cereals: mechanisms and modelling. European Journal of Agronomy, 64: 98-113.

Roberts, M.J., Braun, N.O., Sinclair, T.R., Lobell, D.B. and Schlenker, W., 2017. Comparing and combining process-based crop models and statistical models with some implications for climate change. Environmental Research Letters, 12(9): 095010.

Robertson, M.J., Brooking, I.R. and Ritchie, J.T., 1996. Temperature response of vernalization in wheat: modelling the effect on the final number of mainstem leaves. Annals of Botany, 78(3), pp.371-381.

Royo, C., Villegas, D., Rharrabti, Y., Blanco, R., Martos, V. and García del Moral, L., 2006. Grain growth and yield formation of durum wheat grown at contrasting latitudes and water regimes in a Mediterranean environment. Cereal Research Communications, 34(2-3), pp.1021-1028.

Sánchez, B., Rasmussen, A. and Porter, J.R., 2014. Temperatures and the growth and development of maize and rice: a review. Global Change Biology, 20(2): 408–417.

Schauberger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Khabarov, N., Müller, C. and Pugh, T.A., 2017. Consistent negative response of US crops to high temperatures in observations and crop models. Nature communications, 8, p.13931.

Schlenker, W. and Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proceedings of the National Academy of Sciences of the United States of America, 106(37): 15594-15598.

Semenov, M.A. and Shewry, P.R., 2011. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. Scientific reports, 1: 66.

Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press.

Shen, J., Huete, A., Tran, N.N., Devadas, R., Ma, X., Eamus, D. and Yu, Q., 2018. Diverse sensitivity of winter crops over the growing season to climate and land surface temperature

across the rainfed cropland-belt of eastern Australia. Agriculture, ecosystems & environment, 254, pp.99-110.

Stratonovitch, P. and Semenov, M.A., 2015. Heat tolerance around flowering in wheat identified as a key trait for increased yield potential in Europe under climate change. Journal of Experimental Botany, 66(12).

Tao, F., Rötter, R.P., Palosuo, T., Díaz-Ambrona, C.G.H., Mínguez, M.I., Semenov, M.A., Kersebaum, K.C., Nendel, C., Cammarano, D., Hoffmann, H. and Ewert, F., 2017. Designing future barley ideotypes using a crop model ensemble. European journal of agronomy, 82, pp.144-162.

Tashiro, T. and Wardlaw, I.F., 1989. A comparison of the effect of high temperature on grain development in wheat and rice. Ann Bot-London, 64(1): 59-65.

Trnka, M., Rötter, R.P., Ruiz-Ramos, M., Kersebaum, K.C., Olesen, J.E., Žalud, Z. and Semenov, M.A., 2014. Adverse weather conditions for European wheat production will become more frequent with climate change. Nature Climate Change, 4(7), p.637.

Wang, B., Liu, D.L., Asseng, S., Macadam, I., Yang, X. and Yu, Q., 2017a. Spatiotemporal changes in wheat phenology, yield and water use efficiency under the CMIP5 multimodel ensemble projections in eastern Australia. Climate Research, 72(2), pp.83-99.

Wang, B., Liu, D.L., Asseng, S., Macadam, I. and Yu, Q., 2015. Impact of climate change on wheat flowering time in eastern Australia. Agr Forest Meteorol, 209: 11-21.

Wang, B., Liu, D.L., O'leary, G.J., Asseng, S., Macadam, I., Lines‐Kelly, R., Yang, X., Clark, A., Crean, J., Sides, T. and Xing, H., 2017b. Australian wheat production expected to decrease by the late 21st century. Global change biology, 24(6), pp.2403-2415.

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Liu, D.L., Simpson, M., McGowen, I. and Sides, T., 2018. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecological indicators, 88, pp.425-438.

Watson, J., Zheng, B., Chapman, S. and Chenu, K., 2017. Projected impact of future climate on water-stress patterns across the Australian wheatbelt. Journal of experimental botany, 68(21-22): 5907-5921.

Wheeler, T. and Von Braun, J., 2013. Climate change impacts on global food security. Science, 341(6145): 508-513.

White, J.W. and Hoogenboom, G., 2010. Crop response to climate: ecophysiological models, Climate change and food security. Springer, pp. 59-83.

Williams, G., 2011. Data mining with Rattle and R: The art of excavating data for knowledge discovery. Springer Science & Business Media.

Woli, P., Jones, J.W. and Ingram, K.T., 2013. Assessing the Agricultural Reference Index for Drought (ARID) using uncertainty and sensitivity analyses. Agronomy Journal, 105(1): 150-160.

Woli, P., Jones, J.W., Ingram, K.T. and Fraisse, C.W., 2012. Agricultural reference index for drought (ARID). Agronomy journal, 104(2): 287-300.

Zhang, X.C., 2005. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. Agricultural & Forest Meteorology, 135(1–4): 215-229.

Zhang, X.C., 2007. A comparison of explicit and implicit spatial downscaling of GCM output for soil erosion and crop production assessments. Climatic Change, 84(3-4): 337.

Zheng, B., Chenu, K., Doherty, A. and Chapman, S., 2014. The APSIM-Wheat Module (7.5 R3008), http://www.apsim.info/Portals/0/Documentation/Crops/WheatDocumentation.pdf (accessed May 2018).

Zheng, B., Chenu, K., Fernanda, D.M. and Chapman, S.C., 2012. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivium) varieties? Global Change Biology, 18(9): 2899-914.

Chapter 7. Final conclusions and future research

7.1 Final conclusions

This study systematically investigated the impacts of present and future climate extremes on wheat yield and their likely change in the future in the NSW wheat belt. The models developed and analytical results present in this project provided insights into the underlying relationships between wheat yield and climate extremes events.

(1) Rainfall extremes were identified as dominant factors affecting wheat yield variation and could explain more than half yield variability in the NSW wheat belt. In the eastern slopes and southern plains of the belt, growing season rainfall and consecutive dry days were major factors causing yield variation. By contrast, in the northern plains, pre-growing season rainfall was included as one of the most important factors. Overall, wheat yield variability in the study area was mainly caused by frequent water shortage, while extreme wetness within growing season had a small effect as it occurred less frequently.

The comparison between the results obtained from the MLR and RF models showed that the RF model was better at predicting wheat yield variation based on rainfall extremes. The RF model outperformed the MLR model in all three regions based on the four validation measurements used. The MLR model tended to distribute higher importance values to 2-3 indices and largely disregarded the remaining indices. This could be because multicollinearity might still exist, and one index might mask the contribution from another. Since the RF was a nonlinear algorithm, its result was not affected by multicollinearity. So, the MLR results should be interpreted with great caution. However, this did not affect the capacity of the MLR model to obtain a good fit with regression, or the quality of predictions from the regression. Furthermore, compared to the RF, the MLR model could also quantitatively estimate variable contribution. Consequently, both the two models had their own strengths in exploring climate-yield relationships.

(2) Seasonal agricultural drought could be effectively monitored using satellite remote sensing information machine learning technique in the wheat belt. The bias-corrected random forest model successfully produced SPEI drought maps which were consistent with drought maps derived from station-based dataset. This model had several advantages. First, the drought factors used in the present study were based on global remote sensing data (e.g., TRMM and MODIS) which covered from 50°S to 50°N, so the proposed approach could be applied to any vegetated region in the world. Second, this product could generate results comparable to the ground-based database. Third, this product was cost-free as the remotely-sensed data are easily accessible over the internet. In addition, I also found that machine learning-based remote sensing drought monitoring was more suitable for semi-arid and vegetation-sensitive environments. Nevertheless,

the bias-corrected random forest model could still provide satisfactory results in other types of environments. As such, the drought monitoring approach proposed in this study could be extended to any vegetated region where remote sensing data were available, even in areas with limited in-situ data availability, to provide detailed spatial information regarding drought extent and severity.

The proposed drought monitoring model was also very useful for guiding agricultural practices. In rainfed croplands across the world, drought was widely viewed as the major threat causing yield losses. Timely drought warning systems could help farmers develop optimal strategies to ultimately reduce drought damage. In our study, high correlation coefficients were found between observed wheat yields and model-predicted seasonal SPEIs. Thus, the remotely-sensed, fused SPEI could be a good tool to predict wheat yield based on estimated SPEI, ultimately providing invaluable information for grain price estimations. In addition, appropriate drought-response measures could be worked out in advance. For example, if SPEI_Summer reached a low level, a drought-resistant wheat cultivar could be adopted in the NSW wheat belt to mitigate yield losses, or a fallow season could be recommended to limit production costs in a year in which very low yields were anticipated due to severe drought.

The superior estimation results derived from the BRF model in our study were likely attributable to reduced susceptibility to over-fitting and superior ability in dealing with the hierarchical and non-linear relationships that might exist between the ground-based drought index and remotely-sensed drought factors. Furthermore, the bias-corrected method used in the present study also helped improve the performance of the original RF model. The original RF model usually performed poorly in reproducing extreme observations, with large values underestimated and small values overestimated. This might result in bias in drought monitoring as drought was usually defined as abnormally dry conditions compared with normal conditions. The bias-correction method used in our study significantly reduced the prediction bias with smaller RMSE. Thus, the combination of the RF and bias-correction methods outlined in our study was a promising approach for drought monitoring and can be extended to other cropping regions to obtain new insights to guide agricultural practices.

(3) Temporal and spatial characteristics of future seasonal drought in the wheat belt until the end of 21st century were also analysed in this project based on 28 statistical downscaled global climate models. The relationship between drought frequency and temperature as well as precipitation was also examined. I found that spring and winter droughts were expected to be more severe over the wheat belt, while summer and autumn drought intensities might change little. Winter and spring drought prone areas were likely to spread from west to east significantly and more than half of the wheat belt would be vulnerable to winter and spring droughts by 2100. In contrast, the summer and autumn drought prone areas were primarily in the south-west and north-west, respectively, with only slight changes and sometimes even a decrease in the future. Traditionally dry areas

would likely suffer a greater increase in drought frequency compared to wet areas when subjected to a same increase in temperature or decrease in precipitation.

Even though uncertainties exist in GCMs and multi-model ensemble method, results from our work could be regarded as an indication of the very likely future. Mitigation and adaptation strategies should be prepared in advance in order to minimize the adverse effects of future severe droughts on crop production. Our results showed that the NSW wheat belt was really a climatologically diverse region, so coping strategies should be specific in different zones. For example, cropping in the western zones of the southeastern Australian wheat belt would be an increasingly risky enterprise. Therefore, changing enterprise type, for example, incorporating a livestock component, purchasing additional cropping land or moving the cropping enterprise to areas with more reliable precipitation or access to irrigation might prove alternative adaptive responses for the near future. While in the eastern zones, changing sowing dates or crop rotations, stubble management, incorporating shorter growing season varieties and even fallow, might represent adaptive responses to cope with future drought.

(4) The hybrid model developed by incorporating APSIM output, climate extremes indices, and/or remote sensing information into machine learning models enhanced the capability of conventional crop models or statistical models to simulate the impacts of climate and weather extremes. This hybrid model could be used for pre-harvest yield forecasting with fully consideration of the impacts of stage-specific extreme climate events. Satisfactory forecasts of crop yield could be achieved several months before harvest. It could be used for operational forecasting purposes in Australia and potentially other similar dryland cropping systems around the world. Further, with the development of information technology and remote sensing technology, this proposed model could be directly extended to region- and country-scale yield forecasts. In addition, this hybrid model was also used to evaluate the impacts of climate change on wheat yield in the wheat belt. I found that future yield projections from conventional process-based crop models might have a 1-10% overestimation because of the underestimation of climate extremes-induced yield losses. Future climate conditions in the NSW wheat belt was expected to be more unfavorable. Drought events were projected to remain at historical levels, while heat events during wheat reproductive stages were projected to increase significantly in the future. I believe this study would provide some useful information for local farmers and policy makers with respect to development of adaptation strategies in face of increased climate extremes under climate change.

The comparison between the results obtained from the APSIM model and the hybrid model showed that the APSIM + RF hybrid model was better at reproducing historical wheat yields. Using the RF algorithm as an external modification on the APSIM model outputs appeared to improve the performance of the individual APSIM model. Moreover, the RF model also outperformed the MLR model. This was primarily because I succeeded in exploiting the merits

of each model. The external machine learning model could help improve the performance of the crop model by simulating the effects of climate and weather extremes. Using the crop model outputs as predictors could also improve the ability of the machine learning model to account for multiple agro-climatic processes. Another advantage of the APSIM + RF hybrid model was that it could account for wheat growth stage-specific climate and weather extremes. As different crop growth stages could have different tolerances for a same extreme event, the hybrid could achieve higher accuracy on estimating yield losses induced by various climate and weather extremes. I believe the APSIM + RF hybrid model would provide useful information for local farmers and policy makers with respect to development of adaptation strategies in face of increased climate extremes under climate change.

7.2 Limitations and future research

In spite of the overall contributions of this project presented above, there are a number of limitations which require further investigation in the future.

(1) Drought indices. This is regarding the work summarized in Chapter 2-6. Multiple drought indices, e.g. SPI, SPEI, and ARID, are used in this project. All these indices are limited in considering rainfall-induced runoff. Heavy rainfall events can result in large amount of runoff, which cannot be used in areas where rainfall fell. Thus, these indices can lead to inaccurate estimation of drought events. Moreover, heavy rainfall events are expected to increase in most areas of Australia in face of climate change. These indices may not be appropriate for climate change studies. Thus, new drought indices that consider the effect of rainfall intensity are urgently needed for more accurate estimation of drought.

(2) Model uncertainties. This is related to work summarized in Chapter 5 and Chapter 6. Multiple GCMs are used in these two chapters to provide projections of drought and yield to the end of the 21st century. Projected drought or yield under future climate scenarios could be significantly different due to the differences in GCMs, RCPs, and simulation models, indicating large uncertainties in future drought or yield projections. Future research should quantify the uncertainties originating from various sources in order to evaluate the reliability of drought or yield projections.

(3) Machine learning models. This is regarding the work summarized in Chapter 2, 3, 4, and 6. In this project, machine learning models show better performance in drought monitoring and wheat yield variation analysis compared to traditional regression models. However, it should be noted that, compared to traditional regression models, machine learning models generate only the importance order of variables, but cannot be used for quantifying factor contributions, which

limits their use in further analysis. Future research should introduce more advanced machine learning algorithms to explore climate-yield relationships.

(4) The hybrid model. This is related to work summarized in Chapter 4 and Chapter 6. The hybrid yield simulating and forecasting model satisfactorily predicts yield in the NSW wheat belt. However, this model requires a large amount of data from different sources, including soil, climate, crop, management, and remote sensing information. Thus, it may not be suitable for data-poor areas. Nevertheless, with societal and economic developments, more and more areas will have sufficient data available to implement this yield forecasting method. In addition, ENSO-related indices are also frequently used indicators for yield forecasting in Australia but are not considered in this project. Future studies using a similar modelling approach to mine may use additional information (such as ENSO-related indices) and potentially achieve even greater forecasting accuracy.

(5) Remote sensing drought monitoring model. This is related to work summarized in Chapter 3. I acknowledge that the spatial resolution of 500 m used in the present study is somewhat coarse. Drought estimations could be improved with use of remote sensing data of finer spatial resolution (e.g., 30 m). Furthermore, the bias-corrected model still has some limitations in predicting extreme conditions of drought. Future research should be conducted using more advanced fusion models and more drought-related factors from more detailed data sources to achieve improved performance in drought estimation.