

**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.

Production Note:

Signature of student: 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

Certificate of Original Authorship .....	i
Acknowledgement .....	ii
Publications arising from this thesis.....	iv
Contents .....	v
List of Figures.....	viii
List of Tables .....	xii
Glossary .....	xiii
Abstract.....	xiv
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1 Research background .....	1
1.1.1 Climate change.....	1
1.1.2 Climate variability.....	2
1.1.3 Extreme climate and weather events.....	2
1.1.4 Impacts of climate and weather extremes on crops .....	5
1.1.5 Global climate models .....	8
1.1.6 Modelling methods .....	9
1.2 Research questions and objectives .....	11
1.3 Statement of significance .....	13
1.4 Reference .....	14
<b>Chapter 2. Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia .....</b>	<b>20</b>
Abstract.....	20
2.1 Introduction.....	20
2.2 Materials and methods .....	22
2.2.1 Study area.....	22
2.2.2 Climate and yield data.....	23
2.2.3 Rainfall extreme indices.....	23
2.2.4 De-trending method .....	25
2.2.5 Models.....	25
2.3 Results.....	27
2.3.1 Descriptive statistics of wheat yield.....	27
2.3.2 Model performance .....	27
2.3.3 Relative importance of rainfall extreme indices .....	28
2.3.4 Periodic variation of SPI indices.....	30
2.4 Discussion.....	31
2.5 Conclusion .....	33
2.6 Supporting information.....	34
2.7 References.....	36
<b>Chapter 3. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia .....</b>	<b>41</b>
Abstract.....	41
3.1 Introduction.....	42
3.2 Materials and methods .....	45
3.2.1 Study area.....	45
3.2.2 Data .....	46
3.2.3 Modelling methodology.....	49
3.3 Results.....	54
3.3.1 Model performance assessment .....	54
3.3.2 Relative importance of selected remotely-sensed drought factors.....	57
3.3.3 Comparison with SILO drought maps .....	57
3.3.4 Monitoring agricultural drought.....	60

3.4 Discussion .....	62
3.5 Conclusions .....	64
3.6 Supporting information .....	65
3.7 Reference .....	67
<b>Chapter 4. Dynamic wheat yield forecasts are improved by developing a hybrid approach using biophysical model and machine learning technique .....</b>	<b>71</b>
Abstract .....	71
4.1 Introduction .....	72
4.2 Materials and methods .....	75
4.2.1 Study area .....	75
4.2.2. Data .....	77
4.2.3. Modelling methodology .....	78
4.3 Results .....	84
4.3.1 Model performance .....	84
4.3.2 Optimum forecasting event analysis .....	87
4.3.3 Relative importance of growth selected predictors .....	88
4.4 Discussion .....	89
4.5 Conclusions .....	92
4.6 Supporting information .....	92
4.7 Reference .....	96
<b>Chapter 5. Projected changes in drought across the wheat belt of southeastern Australia using a downscaled climate ensemble.....</b>	<b>101</b>
Abstract .....	101
5.1 Introduction .....	101
5.2 Materials and method .....	104
5.2.1 Study domain description .....	104
5.2.2 Climate data .....	105
5.2.3 Relative SPEI (rSPEI) .....	106
5.2.4 Evaluation of drought characteristics .....	107
5.3 Results .....	109
5.3.1 Temporal changes in drought .....	109
5.3.2 Spatial changes in drought .....	111
5.3.3 Major drivers of drought trends .....	113
5.4 Discussion .....	115
5.5 Conclusions .....	117
5.6 Supporting information .....	118
5.7 Reference .....	120
<b>Chapter 6. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia.....</b>	<b>125</b>
Abstract .....	125
6.1 Introduction .....	126
6.2 Materials and method .....	128
6.2.1 Study sites .....	128
6.2.2 Climate data .....	130
6.2.3 In-situ trial data .....	132
6.2.4 APSIM descriptions .....	133
6.2.5 APSIM simulations .....	133
6.2.6 Climate extremes indicators .....	134
6.2.7 Statistical models .....	135
6.2.8 Hybrid-modelling approach .....	136
6.3 Results .....	137
6.3.1 Model performance .....	137
6.3.2 Effects and projected changes of ECEs .....	140

6.3.3 Differences between APSIM projected and the hybrid model projected future wheat yields .....	142
6.4 Discussion .....	143
6.5 Conclusions .....	146
6.6 Supporting information .....	146
6.7 Reference .....	148
<b>Chapter 7. Final conclusions and future research .....</b>	<b>155</b>
7.1 Final conclusions .....	155
7.2 Limitations and future research.....	158



# List of Figures

## Chapter 1

- Figure 1-1 Impacts of shifts in temperature distribution on temperature extremes (IPCC, 2012). Three possible change patterns are plotted. .... 3

## Chapter 2

- 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. .... 23
- 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..... 28
- 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..... 29
- 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. .... 31

## Chapter 3

- 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. .... 45
- 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. .... 50
- 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. .... 52
- 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 neural 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. .... 55

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 neural network (MLP) models at two stations in New South Wales, Australia.....	56
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.....	57
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.....	58
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.....	59
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.....	60
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...	62
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.....	62
<i>Chapter 4</i>	
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.....	75
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.....	81
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.....	85
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.....	86

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. .... 87

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. .... 88

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. .... 89

**Chapter 5**

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. .... 104

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. .... 104

Figure 5-3 Framework of the procedures used in this study. .... 108

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. .... 110

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 ( $|Z_S| > 1.96$ ). .... 111

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. .... 112

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. .... 113

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 ( $\Delta DF = a * \Delta P (\%) + b * \Delta T (^\circ C)$ ) was then built in both zones for each season and each period. All the regression coefficients ( $R^2$ ) shown were significant ( $P < 0.05$ ). ..... 114

*Chapter 6*

Figure 6-1 Locations of the 29 study sites in the New South Wales wheat belt in south-eastern Australia. .... 129

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..... 137

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. .... 138

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. .... 139

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. .... 140

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 10<sup>th</sup> and 90<sup>th</sup> 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. .... 141

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. .... 143

## List of Tables

### Chapter 2

Table 2-1 List of 11 rainfall extremes indices.....	24
Table 2-2 Regression coefficients of first-difference wheat yield ( $\Delta$ Yield, $\text{kg}\cdot\text{ha}^{-1}$ ) with the 11 rainfall extreme indices in the multiple linear regression (MLR) model ( $\Delta$ Yield = $a\Delta$ SPI <sub>MAM</sub> + $b\Delta$ SPI <sub>JJA</sub> + $c\Delta$ SPI <sub>SON</sub> + $d\Delta$ SPI <sub>DJF</sub> + $e\Delta$ 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. ....	29

### Chapter 3

Table 3-1 Remote sensing indices used in the prediction of agricultural drought in the New South Wales wheat belt, Australia. ....	46
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).....	47
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.....	48
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. ....	50
Table 3-5 Drought predictor variables selected by VSURF for two data clusters in New South Wales, Australia. Predictor variable definitions are given in Table 1. The “1” and “3” designations appended to the end of the variable names designate 1-month and 3-month time scale means. ....	55

### Chapter 4

Table 4-1 Basic information of the 29 study sites in New South Wales, Australia, including location, soil name (details at <a href="http://www.asris.csiro.au/">http://www.asris.csiro.au/</a> ), growing season (May-November) rainfall (GSR, mm), growing season temperature (GST, °C), number of years of yield data available (NY), and wheat yield ( $\text{t ha}^{-1}$ ) range for recorded years (2008-2017). ....	76
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.....	82

### Chapter 5

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. ....	105
---	-----

### Chapter 6

Table 6-1 A brief description of the 29 study sites used in the study, including location, Soil No.(details at <a href="http://www.asris.csiro.au/">http://www.asris.csiro.au/</a> ), GSR (mm), GST (°C), HY, HDR and AMWY ( $\text{t ha}^{-1}$ ). ....	130
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 <a href="https://cmip.llnl.gov/cmip5/availability.html">https://cmip.llnl.gov/cmip5/availability.html</a> . ....	131
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. ....	135

## Glossary

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

## 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 21<sup>st</sup> 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