

Multi-model ensemble projections of future extreme temperature change using a statistical downscaling method in eastern Australia

Bin Wang^{a,b}, De Li Liu^b, Ian Macadam^{c,d}, Lisa V. Alexander^c, Gab Abramowitz^c and Qiang Yu^a

^a School of Life Sciences, Faculty of Science, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia

^b NSW Department of Industry, Skills and Regional Development, Wagga Wagga Agricultural Institute, NSW 2650, Australia

^c Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, NSW 2052, Australia

^d Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

E: bin.a.wang@dpi.nsw.gov.au

Abstract: Projections of changes in temperature extremes are critical to assess the potential impacts of climate change on agricultural and ecological systems. Statistical downscaling can be used to efficiently downscale output from a large number of general circulation models (GCMs) to a fine temporal and spatial scale, which now provides the opportunity for future projections of extreme temperature events. This paper presents an analysis of extreme temperature in data downscaled from ensembles of 13 selected GCMs, out of 28 GCMs, contributing to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR5) under two Representative Concentration Pathways (RCP4.5 and RCP8.5) in eastern Australia. The statistical downscaling procedure begins with spatial interpolation of the monthly gridded data to specific locations of interest using an inverse distance-weighted method, followed by a bias correction towards historical observed climate. Daily climate data for each location are then generated by a modified version of the WGEN stochastic weather generator. The extremes of temperature are described by eleven indices, namely, the annual maximum daily Tmax (TXx), the annual maximum daily Tmin (TNx), the annual minimum daily Tmax (TXn), the annual minimum daily Tmin (TNn), the number of hot days (HD) and frost days (FD), warm days (TX90p) and nights (TN90p), cold days (TX10p) and nights (TN10p) and extreme temperature range (ETR). The results show that downscaled data from most of the GCMs reproduced the correct sign of recent trends in all the extreme temperature indices (except TN10p) although there was much more variation between the individual model runs. An independence weighted mean method was used to calculate uncertainty estimates, which verified that multi-model ensemble projections produced a good consensus compared to the observations in magnitude of the trend in TXx, TN90p, HD, TNn, ETR for the period 1961-2000 when averaged across eastern Australia. In the 21st century the frost days, cold days and nights decrease while more frequent warm days and nights and hot days are projected in the New South Wales (NSW) wheat belt. The changes in temperature extremes under RCP8.5 are more pronounced than that under RCP4.5. Greater warming occurs in the east and northeast of the NSW wheat belt by the end of the 21st century and increases the risk of exposure to hot days around wheat flowering date, which results in farmers needing to reconsider wheat varieties suited to maintain yield. This analysis provides a first overview of projected changes in climate extremes from the ensemble of 13 CMIP5 models with statistical downscaling data in eastern Australia, and supplies important information to mitigate the adverse effects of climate extremes on NSW wheat belt and improve the regional strategy for agricultural systems.

Keywords: Warm indices, cold indices, GCMs, observations, independence weighted mean

1. INTRODUCTION

Future changes in extreme temperature have been projected by different GCMs (Tebaldi *et al.*, 2006). However, GCMs output data are at spatial resolutions of several hundreds of kilometres, and so are too coarse to be used directly in regional scale analysis or resolve small-scale climate processes. This mismatch is a major obstacle for assessing the site-specific impacts of climate change on agricultural systems. The statistical downscaling method of Liu and Zuo (2012) applies statistical transfer functions to GCM output to estimate daily point-scale meteorological series. This method is easily transferable to different regions and provides site-specific estimations under a range of greenhouse gas emission scenarios so long as reliable observations of the climate variable of interest are available (Tolika *et al.*, 2008).

GCMs are a standard tool for deriving projections for future climate change. However, the real climate system is highly complex and it is impossible to adequately describe its processes with an individual climate model. Therefore, authors of model evaluation studies have stated that no single model can be considered ‘best’ and recommend using results from a range of models (Tebaldi and Knutti, 2007). Using the multi-model mean (weighted equally) is the most common and widely used approach. There is some evidence that the “mean model” result, obtained by averaging over the ensemble of models, provides an overall best comparison to observations for climatological mean fields. However, the multi-model mean approach is clearly not appropriate for all types of evaluation, such as variability estimates or extremes, and has significantly less spatial and temporal variance than datasets from individual models or from observations. (Bishop and Abramowitz, 2013). The use of multi-model averages to estimate climatological means works best when each model provides an independent estimate. However, modelling groups share literature, data sets, parameterisations and even sections of model code, so there are several reasons to suspect a lack of statistical independence of every climate prediction in the suite of models we use. Bishop and Abramowitz (2013) derived weights that explicitly account for model dependence defined using covariance of model errors, and we use this approach in this study to synthesize extreme temperature events projected from different GCMs. The optimal weights were estimated to find the linear combination of an ensemble of model simulations that minimizes mean square difference (MSD) between GCM simulated and the observations with respect to the in-sample period.

This study focuses on the analyses of two 40-year simulations: the first examines the time period 1961-2000 (referred to as ‘present’), the second the period 2061-2100 (referred to as ‘future’). The analysis examines temperature extremes in statistically downscaling GCM data and demonstrates the application of the independence weighted mean approach to extreme temperature data from 13 skill-selected GCMs. We investigate extreme temperatures under two RCP scenarios (RCP4.5 and RCP8.5) over eastern Australia so that better use of simulation results can be made in impacts studies, which is beneficial for policymakers and stakeholders in local climate resource and eco-environment management.

2. MATERIALS AND METHODS

2.1. Study area and observed climate data

The study area is located in the NSW wheat belt in eastern Australia. A spatial distribution of 894 meteorological stations used to calculate extreme temperature events is also shown in Fig. 1. The homogeneities of the climate data (annual mean maximum and minimum temperature) from these 894 stations were tested by applying a revision of the two-phase regression model (Lund and Reeves, 2002). Daily temperature data at these stations in eastern Australia for 1889-2010 were extracted from the SILO patched point dataset (PPD,

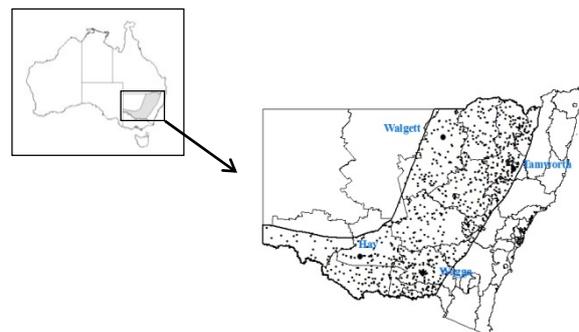


Fig. 1. The NSW wheat belt in eastern Australia and location of 894 weather stations used in this study.

<http://www.longpaddock.qld.gov.au/silo/ppd/index.php>).

2.2. GCM selection

We assessed the output of 28 GCMs from the Coupled Model Intercomparison Project 5 (CMIP5). Details on these models can be found at http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_modeling_groups.pdf. The GCM outputs used here consist of the monthly surface air temperature. Seasonal analyses were performed for winter (June-July-August, JJA), spring (September-October-November, SON), summer (December-January-February, DJF) and autumn (March-April-May, MAM). To identify the relatively satisfactory GCMs in simulating the climatological temperature to downscale over eastern Australia, Taylor's method was used (Taylor, 2001; Wang *et al.*, 2015). The spatial correlation coefficient and standard deviation of each simulation were calculated with respect to 894 station observations for the 50-year average of the reference period 1961-2010 (data not shown). We used 0.4 as criteria of skill score to select 13 eligible GCMs for maximum and minimum temperature in the subsequent analysis. The equation of skill score and detailed selection procedure can be seen Wang *et al.* (2015).

Table 1. Definitions of 11 extreme temperature indices used in this study

ID	Index	Definition	Units
<i>TXx</i>	Warmest daily Tmax	Annual maximum value of daily maximum temperature	°C
<i>TNx</i>	Warmest daily Tmin	Annual maximum value of daily minimum temperature	°C
<i>TXn</i>	Coldest daily Tmax	Annual minimum value of daily maximum temperature	°C
<i>TNn</i>	Coldest daily Tmin	Annual minimum value of daily minimum temperature	°C
<i>HD</i>	Number of Hot days	Frequency of daily maximum temperature >30 °C	days
<i>FD</i>	Number of Frost days	Frequency of daily minimum temperature <2 °C	days
<i>TX90p</i>	Warm days	Number of days in a year with daily maximum temperature >90th percentile	days
<i>TN90p</i>	Warm nights	Number of days in a year with daily minimum temperature >90th percentile	days
<i>TX10p</i>	Cold days	Number of days in a year with daily maximum temperature <10th percentile	days
<i>TN10p</i>	Cold nights	Number of days in a year with daily minimum temperature <10th percentile	days
<i>ETR</i>	Extreme temperature range	Difference between highest daily maximum and lowest daily minimum temperature during the year	°C

2.3 Climate extremes indices

Several measures of extreme temperatures are used to assess extremes by deriving indices from daily temperatures as recommended by the CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) (http://etccdi.pacificclimate.org/indices_def.shtml). A set of 11 temperature indices was selected for this study and their description are given in Table 1. These indices are calculated from daily maximum, minimum temperature and have been developed to assess changes in intensity, duration and frequency of extreme temperature events.

2.4 Statistical downscaling

The monthly gridded data simulated by the GCMs are statistically downscaled to the meteorological observation sites in Fig. 1 using the method described by Liu and Zuo (2012). The downscaling procedure begins with spatial interpolation of the monthly gridded data to specific locations of interest using an inverse distance-weighted method, followed by a bias correction towards historical observed climate (1961-2000). Daily climate data for each location are then generated by a modified version of the WGEN stochastic weather generator (Richardson and Wright, 1984). The parameters required to drive WGEN are derived from the monthly GCM data using relationships derived from daily historical climate data, as described in Liu and Zuo (2012).

2.5 Multi-model ensembles and model dependence

Here we use independence weighted mean (IWM) method accounting for model independence to represent an ensemble of multi-model. The independence measure technique was developed by Bishop and Abramowitz (2013) which applied the covariance in model errors as the basis for a definition of model dependence, specifically

independence coefficients were derived from the error covariance matrix of the bias corrected models (see Bishop and Abramowitz (2013) for details). These coefficients can subsequently be used to optimally weight models for a combination of dependence and mean square error. That is, this measure combines models in independence and performance.

Table 2. Observed and GCMs simulated trends for 11 extreme temperature indices calculated over 1961-2000. Boldface signifies trends that are significant at 5% level. Units as Table 1 (per decade). IWM represents independence weighted mean and AM represents arithmetic mean.

GCMs	<i>TXx</i>	<i>TNx</i>	<i>TX90p</i>	<i>TN90p</i>	<i>HD</i>	<i>TXn</i>	<i>TNn</i>	<i>TX10p</i>	<i>TN10p</i>	<i>FD</i>	<i>ETR</i>
BCC- CSM1.1(m)	0.21	0.14	2.90	3.54	3.26	0.06	0.07	-1.02	-2.71	-2.07	-0.15
CCSM4	-0.13	0.08	1.57	2.96	-0.66	0.18	0.13	-1.15	-2.75	-2.96	-0.25
CESM1(BGC)	0.08	0.22	3.16	3.24	1.51	0.17	0.18	-2.07	-2.76	-3.78	-0.10
CMCC-CM	0.11	0.05	3.98	3.01	2.70	0.21	0.19	-2.49	-3.36	-3.51	-0.06
CMCC-CMS	0.18	0.20	3.89	4.69	2.02	0.04	0.17	-0.90	-4.05	-3.89	0.01
CSIRO- Mk3.6.0	-0.01	0.02	0.22	3.72	0.02	-0.03	0.23	1.36	-3.08	-4.64	-0.22
EC-EARTH	-0.11	0.00	3.18	3.50	0.72	0.19	0.09	-1.29	-2.37	-2.66	-0.18
GISS-E2-H-CC	0.26	0.29	4.71	4.60	3.60	0.12	0.08	-2.65	-3.94	-2.72	0.17
GISS-E2-R	0.03	-0.03	1.65	1.69	-0.73	0.04	0.07	1.46	0.87	-0.89	-0.04
INM-CM4	-0.01	0.11	2.37	2.62	1.83	0.09	-0.03	-1.04	0.18	-0.05	0.02
MPI-ESM-LR	0.21	0.19	4.74	4.22	2.36	0.35	0.09	-3.14	-3.42	-2.68	0.13
MPI-ESM-MR	-0.03	0.15	1.62	3.22	-0.45	-0.01	0.10	0.15	-2.00	-3.25	-0.12
MRI-CGCM3	0.02	0.13	0.42	2.42	-0.79	0.04	0.05	-0.09	-0.38	-0.97	-0.03
AM	0.06	0.12	2.65	3.34	1.18	0.11	0.11	-0.99	-2.29	-2.62	-0.04
IWM	0.06	0.11	2.61	3.33	1.06	0.11	0.11	-0.94	-1.94	-2.57	-0.04
Obs	0.05	0.01	1.70	3.42	0.67	0.07	0.10	-0.42	0.29	-1.83	-0.06

3. RESULTS

3.1 Comparison between observed and modelled temperature extremes during 1961-2000

The observed and downscaled values for each index from 1961 to 2000 averaged across the NSW wheat belt are shown in Fig. 2. There is a trend of increasing observed warm indices (*TXx*, *TNx*, *TX90p*, *TN90p*, *HD*) while decreasing trend can be found in cold extremes (*TX10p*, *FD*) during the historical period 1961-2000. Trends for the observations and each GCM projected for the eleven indices are given in Table 2. It is noteworthy that, of the 11 indices, only *TN90p* (warm nights) presents a significant increase ($p < 0.05$) in the observations and 11 of the 13 models exhibit a statistically significant increase in *TN90p* (Table 2). Although, observed trends of other indices are not significant, the majority of the models obtain the correct sign of trend for each extreme temperature index except *TN10p* (cold nights) when averaged across eastern Australia. More than 10 of 13 models produce trends in seven indices (*TNx*, *TX90p*, *TN90p*, *TXn*, *TNn*, *TX10p*, *FD*) of the same sign as the observed trends. 8 of 13 models agree with the observations in terms of sign of trend in *TXx* and *ETR* and 9 of 13 modes show a consistent trend with the observations in *HD*. While the sign of the extreme temperature trends in the downscaled data generally matches the trends for the observations, the magnitude of change differs between different GCMs and is generally different from the observed values. No one GCM is consistently 'best' at reproducing the observed magnitude of trends across all extreme temperature indicators and some GCMs overestimate or underestimate the observed trends in some of the temperature extremes indices during 1961-2000 (Table 2). Individual GCMs and the observed extreme temperature indices show large inter-annual

variability over historical period 1961-2000 (Fig. 2). However, the IWM projections and the arithmetic mean (AM) from 13 GCMs have less variability in the baseline time series than individual model and the observation due to its combining results from multiple models, which leads to a significant changing trend in 9 of 11 extreme temperature indices. Moreover, the observed and the IWM projected have a good consensus in magnitude of the trend in TXx, TN90p, TXn, TNn, ETR (Table 2). Meanwhile, the RMSE of IWM simulations with observations for all 11 eleven temperature indices is lower than that of AM (data not shown).

3.2 Multi-model ensemble projections of temperature extremes in the 21st century

Fig. 3 shows spatial changes in multi-model ensemble simulated extreme temperature indices between 1961-2000 and 2061-2100 for RCP8.5 based on 894 sites. The results show that temperature extremes increase most in the east and northeast of the NSW wheat belt. In addition, the temporal evolution in temperature indices from 1900 to 2100 under RCP4.5 and RCP8.5 for the IWM results is shown in Fig. 4.

3.2.1 Warm extremes (TXx, TNx, TX90p, TN90p, HD)

Fig. 4a and b show that TXx and TNx are projected to increase, and before the year 2050 there is no distinguishable differences between the two scenarios. In the last 40 years of the 21st century, TXx increases by around 2.8 °C for RCP4.5 and 4.6 °C for RCP8.5 compared to 1961-2000. The increase of TNx is by approximately 2.4 °C and 4.2 °C for RCP4.5 and RCP8.5, respectively. Warm days (TX90p) and warm nights (TN90p) also exhibit increases. It is observed from Fig. 4c and d that TX90p and TN90p show consistent increase over the 21st century for both scenarios RCP4.5 and RCP8.5. At the end of the 21st century, TX90p increases dramatically by around 178.4% for RCP4.5 and 336.2% for RCP8.5 compared to the historical period. TN90p rises by approximately 193.5% and 396.6% for RCP4.5 and RCP8.5, respectively. For the number of hot days (HD) (Fig. 4e), the temporal evolution of HD shows consistent increase under RCP8.5, which is similar to other warm extreme indices. The rise of HD is by around 40.0% for RCP4.5 and 68.3% for RCP8.5 compared to the 1961-2000 at the end of the 21st century.

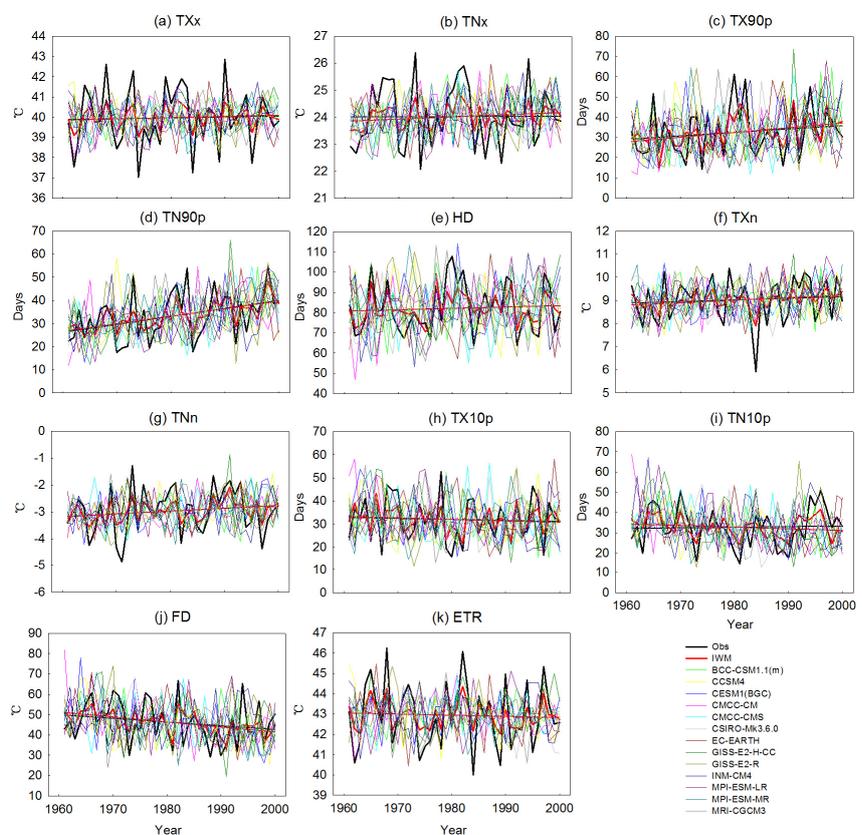


Fig. 2. Observed (black line) and simulated including multi-model independence weighted mean (IWM, red line) time series of extreme temperature indices from 1961 to 2000 averaged across the NSW wheat belt.

3.2.2 Cold extremes (TXn, TNn, TX10p, TN10p, FD)

There is a projected positive trend for TXn of around 1.7 °C for RCP4.5 and 3.0 °C for RCP8.5 by 2100 compared to 1961-2000 (Fig. 4f). The increase for TNn is lower than that in TXn, increasing by about 1.2 °C for RCP4.5 and 2.3 °C for RCP8.5 (Fig. 4g). In contrast, cold days (TX10p) and cold nights (TN10p) are projected to decrease in the future. The magnitude of the decrease by the end of 21st century is 62.4% for RCP4.5 and 79.7% for RCP8.5 (Fig. 4h). It can be seen that TN10p is also projected to decline, by 72.2% for RCP4.5 and 89.1% for RCP8.5 (Fig. 4i). As shown in Fig. 4j, the temporal evolution of FD shows a consistent drop under two scenarios in the 21st century with around 52.6% for RCP4.5 and 73.0% for RCP8.5 compared to the baseline period.

3.2.3 Extreme temperature range (ETR)

The time series of ETR shows a consistent increase but with large fluctuations and a small difference between the two scenarios in the 21st century (Fig. 4k). The amount of ETR increase in the future is small, increasing by approximately 1.5 °C for RCP4.5 and 2.2 °C for RCP8.5 compared to the 1961-2000.

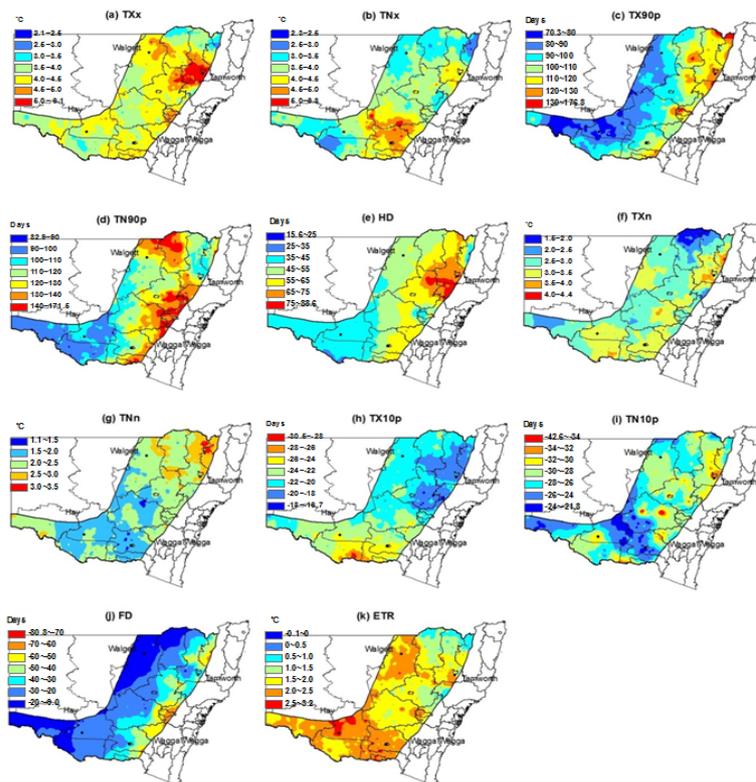


Fig. 3. Changes in multi-model ensemble (independence weighted mean, IWM) simulated extreme temperature indices during the period 2061-2100 compared to 1961-2000 under RCP8.5.

4. DISCUSSION AND CONCLUSION

The trends of some extreme temperature indices in the downscaled data for some individual GCMs were opposite to observed values (Table 2), which is possibly linked to temporal downscaling procedure and/or unforced variability. Like many other statistical downscaling methods, this method relies on empirical relationships between observational and GCM simulated data (Liu and Zuo, 2012). Although WGEN parameters derived from observations are improved in this method, extreme events may not be realistically represented in the output. Each GCM simulated extreme climate indices still had larger differences with the observations (Fig. 2). This may be enhanced by improving the WGEN parameters with fine scale outputs of dynamically downscaled results, rather than using coarse scale GCM data, in combination with historical records. However, considering the simulation of extreme climate we do not yet expect each downscaled model to accurately reproduce observed absolute quantities or rates of change because of large interannual variability. The outcome of a forecast can be improved by combining results from multiple models, which is based on the fundamental assumption that errors tend to be cancelled if the models are independent.

In the NSW wheat belt, wheat crops are projected to experience more severe heat stress but decreases in frost risk in the eastern part of the NSW wheat belt. Grain growers try to choose suitable varieties and sow after the autumn break, which can allow flowering to occur during the period with minimum stress of frost and heat. However, as frost days are projected to decrease in frequency and hot days are projected to increase, this will lead

to a shift of the low-risk flowering window compared to current climate. As a consequence, grain growers need to choose suitable varieties and sow after autumn break, which can allow flowering to occur during the period with minimum stress of frost and heat.

Estimates of future changes in temperature extremes are essential information for stakeholders and policymakers.

Although this analysis contributes to this effort, there are many uncertain factors when assessing changes in extreme indices at the regional scale. More work on generating projections from more advanced, higher resolution GCMs with advanced multi-model ensemble methods, as well as analysing uncertainties related to model structure and internal parameters and improving downscaling and bias correction techniques, are needed for a better understanding of the future extreme changes over NSW wheat belt.

Acknowledgements

The author acknowledges that the Chinese Scholarship Council provided the scholarship and the NSW Department of Primary Industries provided office facilities for conducting this work.

REFERENCES

- Bishop CH, Abramowitz G (2013) Climate model dependence and the replicate Earth paradigm. *Climate dynamics* 41:885-900.
- Liu DL, Zuo H (2012) Statistical downscaling of daily climate variables for climate change impact assessment over New South Wales, Australia. *Climatic change* 115:629-666.
- Lund R, Reeves J (2002) Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate* 15:2547-2554.
- Richardson CW, Wright DA (1984) WGEN: A model for generating daily weather variables. ARS (USA).
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* (1984–2012) 106:7183-7192.
- Tebaldi C, Hayhoe K, Arblaster JM, Meehl GA (2006) Going to the extremes. *Climatic change* 79:185-211.
- Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365:2053-2075.
- Tolika K, Anagnostopoulou C, Maheras P, Vafiadis M (2008) Simulation of future changes in extreme rainfall and temperature conditions over the Greek area: a comparison of two statistical downscaling approaches. *Global and Planetary Change* 63:132-151.
- Wang B, Liu DL, Asseng S, Macadam I, Yu Q (2015) Impact of climate change on wheat flowering time in eastern Australia. *Agricultural and Forest Meteorology* 209:11-21.

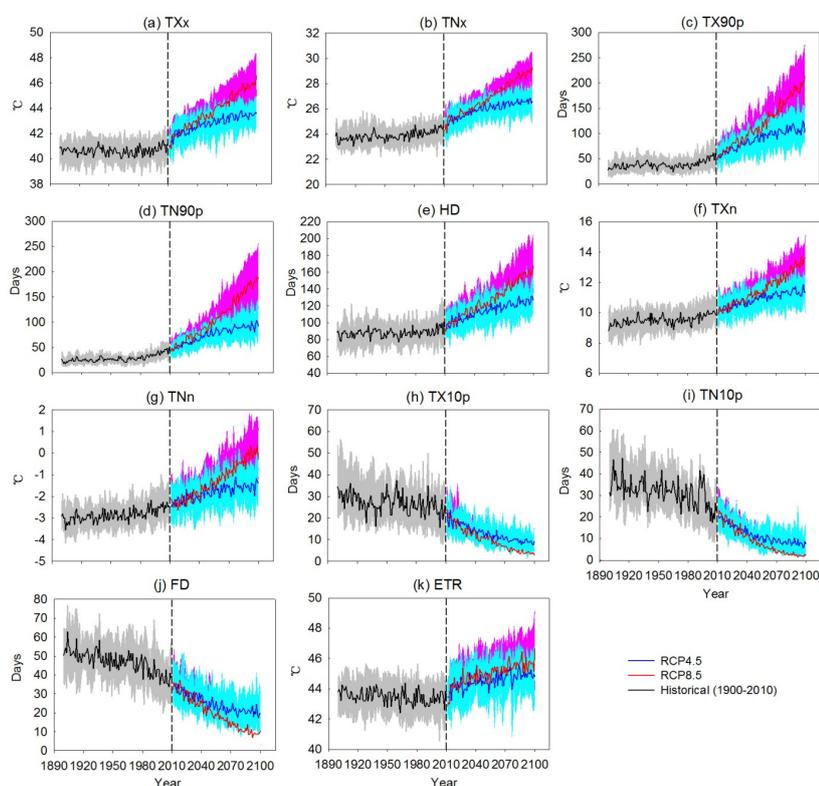


Fig. 4. Time series (1900-2100) of multi-model ensemble (independence weighted mean, IWM) projected each extreme temperature index under RCP4.5 and RCP8.5 averaged across the NSW wheat belt. The top and bottom bounds of shaded area are the 90th and 10th percentile of the annual value from the 13 GCMs simulations.