Vol. 67: 99–115, 2016
doi: 10.3354/cr01366

CLIMATE RESEARCH
Clim Res

Published February 24

Evaluation of long-term precipitation and temperature Weather Research and Forecasting simulations for southeast Australia

Fei Ji[1,*], Jason P. Evans[2], Jin Teng[3], Yvonne Scorgie[1], Daniel Argüeso[2], Alejandro Di Luca[2]

[1]Office of Environment and Heritage, Department of Planning and Environment, Queanbeyan, New South Wales 2620, Australia
[2]Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales 2052, Australia
[3]CSIRO Land and Water Flagship, Canberra, ACT 2600, Australia

ABSTRACT: The New South Wales (NSW)/Australian Capital Territory (ACT) Regional Climate Modelling (NARCliM) project aims to deliver robust climate change projections for southeast Australia at a scale relevant for decision-making. In the first phase of the project, the Weather Research and Forecasting (WRF) model with 3 physics scheme combinations, driven by NCEP/NCAR reanalysis dataset as 'perfect' boundary conditions, was run for a 60 yr period from 1950–2009 to assess the model's ability to simulate regional climate for southeast Australia. In this study, model results for daily precipitation and maximum and minimum temperatures were compared to gridded observations from the Australian Water Availability Project (AWAP) to evaluate model performance at varying time scales using a number of statistical metrics. Results show that all simulations have good representation of daily, monthly, seasonal, annual, multiannual and decadal variation in precipitation and temperature. However, there is a bias in precipitation in the northwest part of the domain (25–100%) and along the Great Dividing Range (75–150%). The temperatures are systematically underestimated across the domain (2–3°C for maximum temperature and 1–2°C for minimum temperature), suggesting the need for bias correction. The evaluation results indicate that the cumulus scheme is critical to precipitation simulation, and planetary boundary layer and radiation schemes are more important in temperature simulations. The findings from this study give us confidence in the WRF model for long-term regional climate modelling for southeast Australia. They also provide guidance in the parameterisation of the WRF model in providing more reliable precipitation and/or temperature projections.

KEY WORDS: NARCliM · Weather Research and Forecasting · WRF · Precipitation · Maximum and minimum temperatures

1. INTRODUCTION

Global climate models (GCMs) are the most valuable tools to project future climate conditions, but their coarse resolution (about 100–300 km) cannot provide regional and local climate details for impact studies (Vaze et al. 2011). Many downscaling methods have therefore been developed to transform climate information from the coarse resolutions of GCMs to regionally relevant resolutions (10–50 km). Dynamical downscaling (e.g. Frei et al. 2003, 2006, Déqué et al. 2005, Giorgi & Lionello 2008) with regional climate models (RCMs) is one of the commonly used downscaling methods. Many previous projects including RMIP (Fu et al. 2005), PRUDENCE (Christensen et al. 2007), ENSEMBLES (van der Lin-

*Corresponding author: fei.ji@environment.nsw.gov.au

den & Mitchell 2009), NARCCAP (Mearns et al. 2012), and CLARIS-LPB (Solman et al. 2013) have produced regional climate projections using RCMs. Globally coordinated projects such as the Coordinated Regional Climate Downscaling Experiment (CORDEX) (Giorgi et al. 2009) collate and compare the RCM simulations. Likewise, the New South Wales (NSW)/Australian Capital Territory (ACT) Regional Climate Modelling (NARCliM) is a project that endeavours to use RCMs to deliver robust climate change projections for NSW and the ACT at a scale relevant for use in biophysical impact assessments and local-scale decision-making (Evans et al. 2014).

In the NARCliM project, 4 GCMs and 3 RCMs were objectively selected to undertake future projections (Evans & Ji 2012a, Evans et al. 2014). In the first phase of the project, the 3 RCMs were used to downscale the NCEP/NCAR Reanalysis Project 1 data (Kalnay et al. 1996) from 1950–2009 to assess the RCM's ability to simulate the observed regional climate. The 3 RCMs are physics scheme combinations of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008) which were objectively selected from 36 physics scheme combinations (Evans & Ji 2012b, Evans et al. 2014).

Previous studies have shown that the WRF model performs well in simulating the regional climate for southeast Australia (Evans & McCabe 2010, 2013, Ji et al. 2011, 2014, Evans & Westra 2012, Evans et al. 2012, 2013b). Evans & McCabe (2010) undertook an evaluation of regional climate simulated over the Murray-Darling Basin for the period 1985–2008. They found that WRF was able to capture daily, seasonal and inter-annual variability of precipitation and mean temperature as well as recent extreme events and a dominant inter-annual variation mechanism (El Niño–Southern Oscillation, ENSO). Similar model evaluations were done for other regions. Argüeso et al. (2012) evaluated WRF-simulated mean and extreme precipitation in Spain for 1970–1999. They found that the WRF model was largely capable of capturing the various precipitation regimes, although substantial errors were still observed in monthly precipitation, especially during spring. Andrys et al. (2015) undertook multi-decadal (1981–2010) evaluation of WRF capabilities in simulating precipitation and temperature extremes for Western Australia, and they concluded that the WRF model was able to simulate daily, seasonal and annual variation in temperature and precipitation reasonably well, including extreme events. Recently, Katragkou et al. (2015) evaluated an ERA-Interim reanalysis-driven WRF multi-physics ensemble for the time period 1990–2008 over the EURO-CORDEX domain. They found that the WRF model can represent the present climate with a reasonable degree of fidelity. However, temperatures tend to be underestimated, and the largest temperature spread and biases are seen in winter over north-eastern Europe. Precipitation is overestimated in both winter and summer but with a larger magnitude in summer.

From the common findings of the above studies, it is clear that the WRF model is able to represent regional climate, although systematic biases can be seen in the simulations. As different combinations of physics parameterisations were used in these studies, it seems that biases for temperature and precipitation are linked to the boundary conditions, different physical mechanisms and geographic location of the model domain.

The main purpose of this study was to evaluate three 60 yr simulations driven by NCEP/NCAR Reanalysis Project 1 (NNRP) reanalysis data to assess the RCM performance for southeast Australia. The impact of physics scheme options on precipitation and maximum and minimum temperatures in long-term simulations is also discussed. The results will provide guidance when using NARCliM future projections for impact studies.

2. DATA

2.1. NARCliM data

The selected RCMs are 3 physics scheme combinations of the Advanced Research WRF (ARW) Version 3.3, which are referred to as R1, R2 and R3 (Table 1). The 3 RCMs were chosen from 36 physics scheme combinations (2 planetary boundary layer, 2 cumulus, 3 microphysics and 3 radiation schemes) based on 2 criteria: (1) adequate performance when simulating historic climate. The full set of RCMs is evaluated over the domain of interest in order to remove from the set any models that are not able to adequately simulate the climate; (2) the selected RCMs are the most independent from each other. From the set of RCMs that perform well, a subset is chosen such that each chosen RCM is as independent as possible from the other RCMs (Evans et al. 2013a, 2014). The model domain covers the CORDEX Australasia region (Giorgi et al. 2009) with 50 km resolution and the NARCliM domain with 10 km resolution (Fig. 1). Within the NARCliM domain, the Great Dividing Range (the most substantial mountain range in

Table 1. Configuration of the 3 regional climate models in NARCliM. The planetary boundary layer physics/surface layer physics include Yonsei University (YSU)/Similarity theory (MM5) and Mellor-Yamada-Janjic (MYJ)/Similarity theory (Eta). The cumulus physics are Kain-Fritsch (KF) and Betts-Miller-Janjic (BMJ) schemes. The radiation physics include Dudhia/RRTM and CAM/CAM shortwave/long wave schemes. Other physics include microphysics (WDM 5 class) and land surface physics (Noah LSM)

NARCliM ensemble member	Planetary boundary layer physics / surface layer physics	Cumulus physics	Micro-physics	Land surface physics	Shortwave / longwave radiation physics
R1	MYJ/Eta similarity	KF	WDM 5 class	Noah LSM	Dudhia/RRTM
R2	MYJ/Eta similarity	BMJ	WDM 5 class	Noah LSM	Dudhia/RRTM
R3	YSU/MM5 similarity	KF	WDM 5 class	Noah LSM	CAM/CAM

Australia) stretches thousands of kilometres from Queensland, running the entire length of the eastern coastline through NSW, and then into Victoria. The mountain range runs almost north–south, intercepting the prevailing westerly winds in winter and the easterly winds in summer, resulting in more precipitation to the west of the mountain range in winter and to the east in summer, and acting as a major climatic barrier separating southeast Australia into distinct climate zones.

Both domains used 30 vertical levels spaced closer together in the planetary boundary layer. Simulations were run from 1950–2009, and outputs (daily accumulated precipitation and maximum and minimum temperature) from the NARCliM domain were evaluated in this study.

The model simulations used 6 hourly boundary conditions from the NNRP, which was the only available reanalysis to provide a record long enough for this purpose. The NNRP is a combination of observations and a global atmosphere model output (Kalnay et al. 1996). By using as many observations as possible, the NNRP produces an estimate of the state of the atmosphere that is as close to reality as possible. Recently, a number of new reanalyses have been produced that have generally improved on earlier reanalyses like NNRP (e.g. Peña-Arancibia et al. 2013); however, these reanalyses are not long enough to allow examination of inter-decadal variability, and hence are not used here. A number of reanalyses cover longer periods (e.g. ERA-20C), but they were not available when the experiments started.

2.2. Australian Water Availability Project (AWAP) data

The gridded daily precipitation totals, and maximum and minimum temperatures from the AWAP (www.bom.gov.au/jsp/awap; Jones et al. 2009) are used as observations in this study. This dataset is at a resolution of 0.05° by 0.05° (approximately 5 km by 5 km) and is obtained by interpolating data from a network of stations.

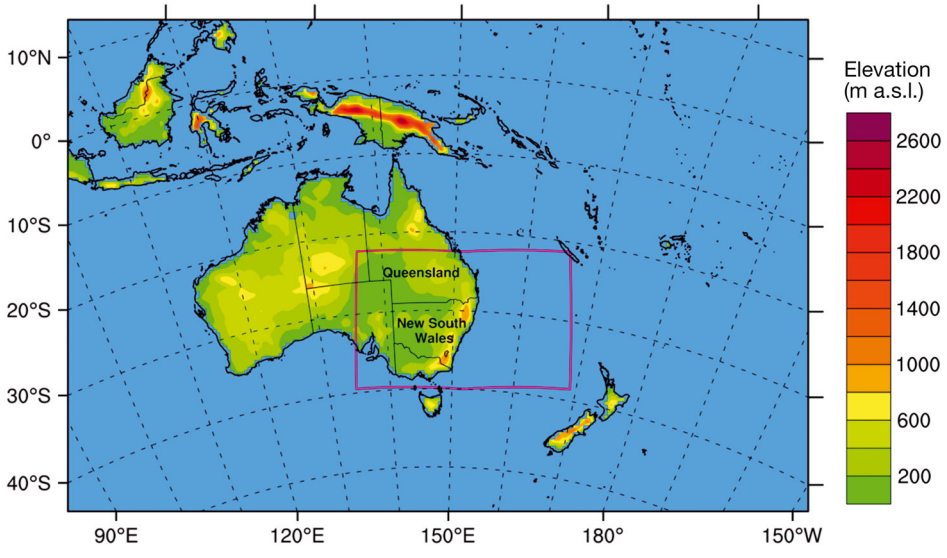


Fig. 1. Weather Research and Forecasting (WRF) model domains with grid spacing of about 50 km (outer CORDEX domain shown as map extent) and 10 km (inner NARCliM domain shown with red outline). All evaluations are conducted for the inner NARCliM domain

The climate of southeast Australia varies from desert climate in the west, to grassland climate in the centre, to temperate climate on the southeast coast and subtropical climate on the northeast coast. The climate in this region is influenced by a number of larger-scale climate drivers such as ENSO, the Indian Ocean Dipole (IOD) and the Southern Annular Mode (SAM). These climate drivers interact over south-east Australia and produce a highly variable climate from year to year. The mean seasonal precipitation, and maximum and minimum temperatures for 1950–2009 are shown in Fig. 2. There is a clear west–east gradient in precipitation for southeast Australia. Higher precipitation is observed on the eastern seaboard, with more than 2000 mm annual precipitation, and less than 200 mm annual precipitation in the northwest part of the domain. The seasonality of precipitation is obvious for this area with summer (DJF) dominant precipitation in the north, winter (JJA) dominant precipitation in the south and the Snowy Mountains, and uniformly distributed seasonal precipitation in between. A northwest–southeast gradient is present in the observed maximum and minimum temperatures across the region. The temperature is highest in the northwest corner (above

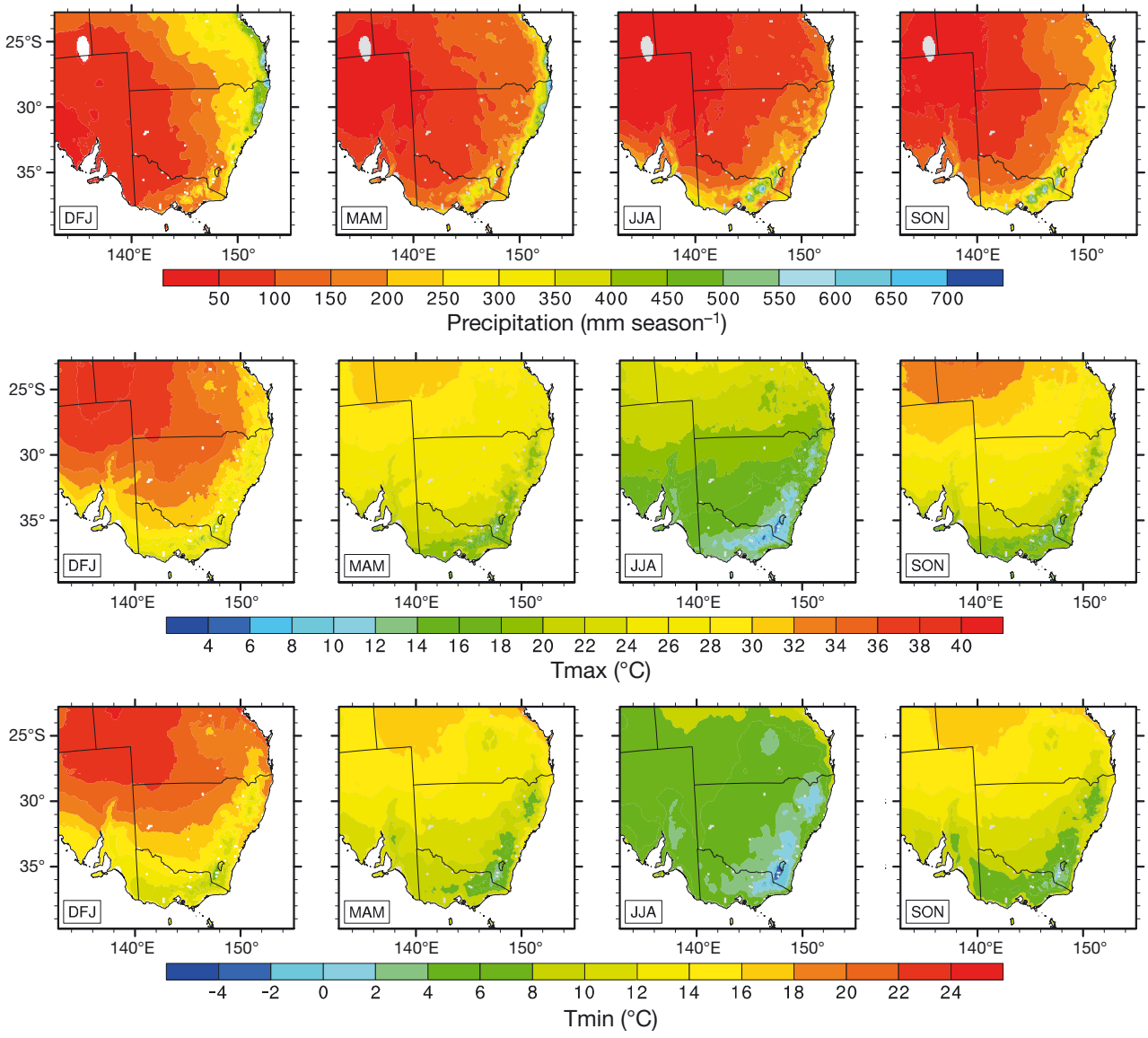


Fig. 2. Australian Water Availability Project (AWAP) observations of precipitation (top row), maximum temperature (central row) and minimum temperature (bottom row) averaged for each season: summer (DJF), autumn (MAM), winter (JJA) and spring (SON). The grey spot on the northwest corner (top row) indicates a shortage of data

40°C in summer) and lowest near the high-altitude Snowy Mountains (< –4°C in winter).

Due to complex topography and large precipitation and temperature variability in this region, the domain was divided into 11 climate zones (Fig. 3) based on the characteristics of observed precipitation and temperature, by using a regionalization method based on Argüeso et al. (2011), but modified to incorporate both temperature and precipitation in a single set of climate divisions. Within the 11 climate zones, Zones 5, 8 and 11 are zones with uniform precipitation across seasons. Those to the north of the uniform precipitation zones (Zones 1, 2, 3, 4, 9, 10) are summer-dominant precipitation zones, and the remainder (Zones 6 and 7) are winter-dominant precipitation zones. The monthly distribution and annual anomaly for precipitation and maximum and minimum temperature were analysed for each of the 11 climate zones.

3. EVALUATION METHODOLOGY

The gridded data from AWAP were re-gridded to the grid specifications of the WRF outputs, using inverse distance weighting. The evaluation was undertaken on each WRF grid cell and on all WRF grid cells within each climate zone.

In this study, simulations were assessed by quantifying bias in magnitude and similarity in temporal and spatial variability. On a grid cell basis, the simulations were evaluated using the spatial and temporal correlation (R), mean bias, mean absolute error (MAE) and root mean square error (RMSE). For each climate zone, the zone-averaged rainfall and temperature time series were used to analyse monthly distributions and annual anomalies for precipitation and maximum and minimum temperatures.

The probability density functions (PDFs) of daily precipitation and maximum and minimum temperatures were compared using the skill score defined by Perkins et al. (2007). This skill score (Sscore) is simply a measure of the common area between 2 PDFs, expressed as the empirical bins used to create the PDF:

$$Sscore = 100 \times \sum_{i=1}^{N} \min(Z_m, Z_o) \quad (1)$$

where N is the number of bins used to calculate the PDF, Z_m is the frequency of values in a given bin from the model, and Z_o is the frequency of values in a given bin from the observations. A perfect simulated PDF would lead to an Sscore value of 100.

Taylor diagrams (Taylor 2001) provide a concise way of graphically summarizing how well patterns of simulations match observations in terms of their spatial correlation, their root mean square (RMS) difference and the ratio of their variations (represented by their standard deviations). These diagrams are especially useful in evaluating multiple aspects of complex models or in gauging the relative skill of many different models.

Finally, a number of Expert Team (ET) on Climate Change Detection and Indices, ETCCDI (http://etccdi.pacificclimate.org/list_27_indices.shtml) were also calculated. These indices are the number of days that exceed 10 mm (R10), the number of frost nights (minimum temperature <0°C) and the number of hot days (maximum temperature >35°C). As 35°C is a typical temperature for summer days in southeast Australia, we used the threshold of 35°C for hot days instead of 25°C for summer days in this study.

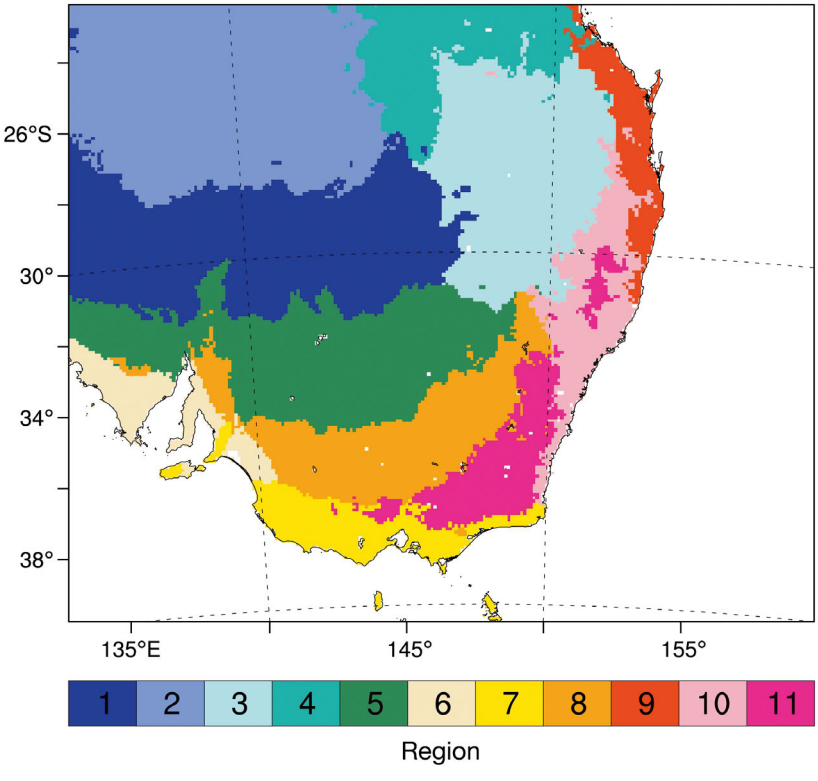


Fig. 3. Climate zones (n = 11) in the NARCliM domain. See Section 2.2. in the main text for details on zone characteristics

4. RESULTS

The results of the 3 WRF simulations were compared against the AWAP observations of precipitation and maximum and minimum temperatures, and the evaluation results are presented here. First, the ability of the RCM simulations to reproduce the observed characteristics at a daily time scale was examined using the Sscore. The distribution of monthly precipitation and temperature was then compared to observations for each of the 11 climate zones. Next, the RCMs were evaluated against observations at annual and seasonal time scales, and their ability to reproduce the inter-annual and decadal variability was assessed. Finally, the capability of WRF in simulating a few extreme indices was also evaluated.

4.1. Daily time scale

The accurate description of the precipitation and temperature PDFs is essential for characterizing precipitation and temperature regimes. Here, daily PDFs of the variables of interest are calculated and compared, quantifying the similarity using the Sscore of Perkins et al. (2007). All daily values of precipitation below 0.2 mm d^{-1} were omitted, as rates below this amount were not recorded in the observations. Perkins Sscores for precipitation and maximum and minimum temperatures are shown in Fig. 4.

For the majority of the domain, WRF is able to capture the daily precipitation PDF, with Perkins Sscores >90. The Sscores for the area west of the Great Dividing Range is >96, which is much higher than Sscores of various GCMs found by Perkins et al. (2007) and higher than other previous RCM studies (Evans & McCabe 2010, Argüeso et al. 2012). Lower scores can be seen along the Great Dividing Range, which highlights the intrinsic difficulty of modelling precipitation processes in topographically complex regions, along with the higher uncertainty associated with observations in these regions. It should be noted that, even in these regions, the Sscore remains above 80, indicating that most of the PDFs are simulated well.

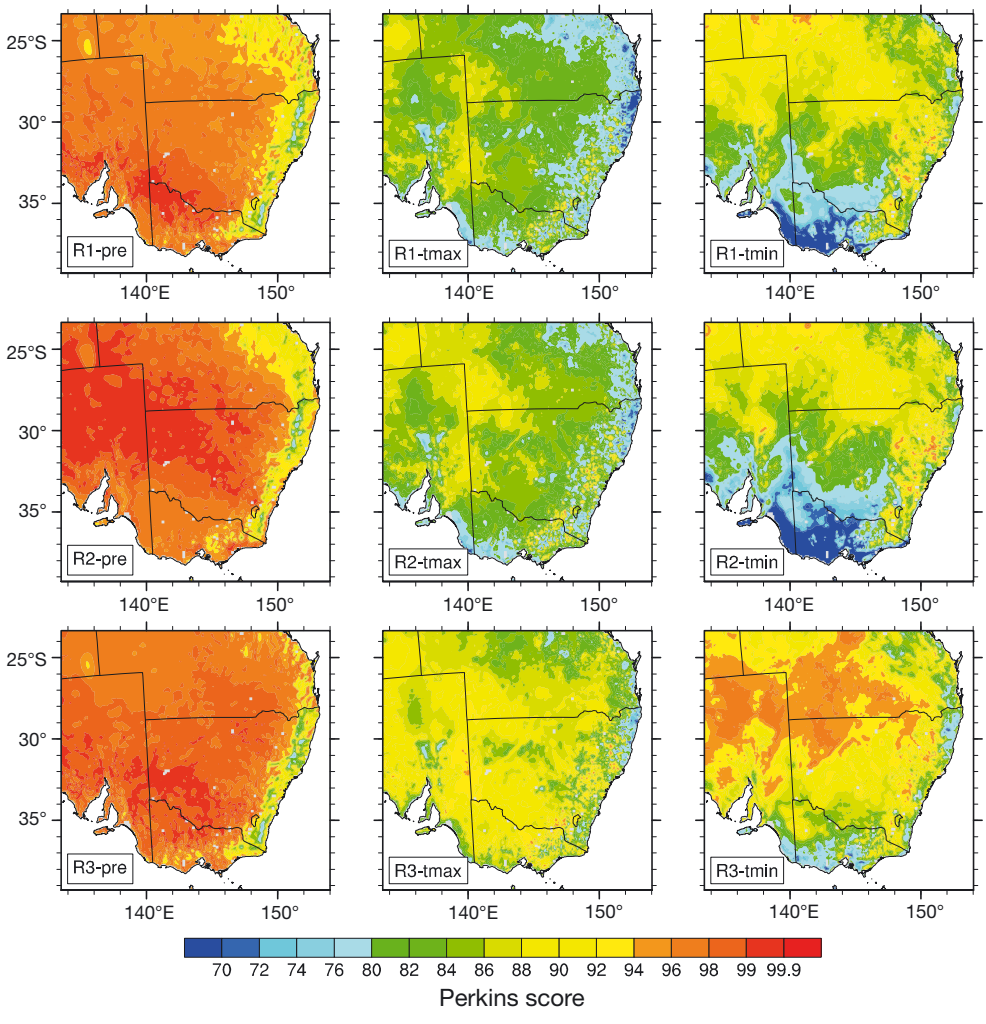


Fig. 4. Perkins skill scores calculated for precipitation (left column), maximum temperature (central column) and minimum temperature (right column) from R1, R2 and R3 simulations. A value of 100 represents perfect skill

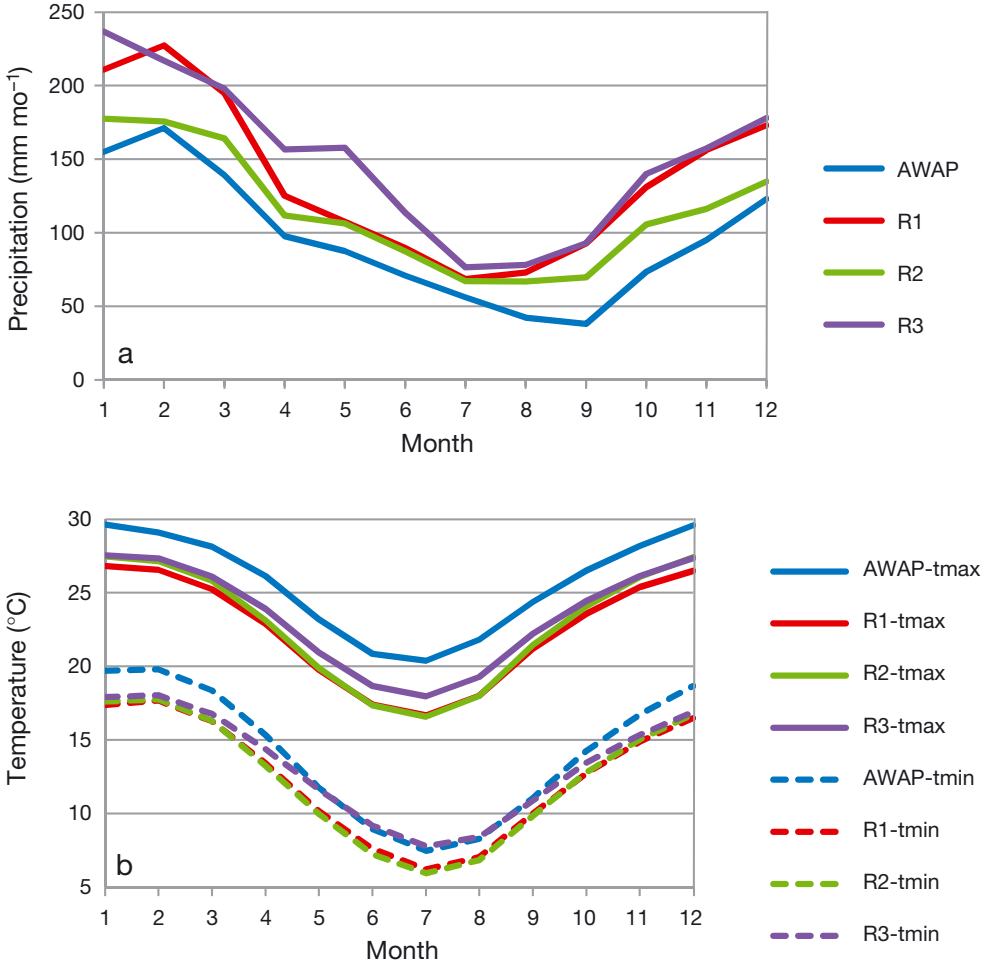


Fig. 5. Monthly distribution of (a) precipitation and (b) maximum (Tmax, solid lines) and minimum (Tmin, dashed lines) temperatures for Zone 9 (see Fig. 3 for zones). The 4 colours represent 1 observation (AWAP) and 3 simulations (R1, R2, R3)

Within the 3 simulations, R2 and R3 are better than R1, and R2 is slightly better than R3 for the southeast coast.

The maps of the Perkins Sscores for maximum and minimum temperatures illustrate that across the domain, WRF is able to capture the daily PDF for maximum and minimum temperatures well. Poorer scores are observed for both maximum and minimum temperatures in the southern part of the domain and the northeast coast, especially for minimum temperature. These scores are comparable to the findings of previous studies (Evans & McCabe 2010, Andrys et al. 2015). Among the 3 simulations, R3 gives better results than the R1 and R2 simulations for temperature.

4.2. Monthly time scale

Monthly distributions of precipitation and maximum and minimum temperatures are analysed to assess the ability of WRF to represent the annual cycle. The zone-averaged monthly variables of interest and their MAEs are calculated and compared for each of the 11 climate zones. The MAE values of precipitation and temperature for Zone 9 are close to the domain-averaged MAE values; therefore, the results for this zone are shown in Fig. 5 as a typical example.

Zone 9 is a summer-dominant precipitation zone with ~150 mm mo^{-1} of precipitation in summer (DJF) and ~50 mm mo^{-1} in winter (JJA). The 3 simulations capture the seasonality of precipitation well; however, they all overestimate monthly precipitation, especially in warm months. Among the 3 simulations, R2 performs the best; the MAE is ~20.3 % of observed precipitation, which is smaller than 43.6 % for R1 and 56.8 % for R3.

For temperature, all simulations have a 2–3°C constant cold bias for maximum temperature across all seasons. The R3 simulation, with an MAE of 2.2°C, performs slightly better than R1 (3.2°C) and R2 (2.8°C). A 1–2°C cold bias in minimum temperature is observed in R1 and R2 simulations for all seasons. R3 with an MAE of 0.9°C performs better than R1

(1.7°C) and R2 (1.8°C), despite underestimating minimum temperature in warm seasons and overestimating minimum temperature in cold seasons.

The results for other climate zones are provided in the Supplement, available at www.int-res.com/articles/suppl/c067p099_supp.pdf. All 3 simulations are able to capture the seasonality of precipitation (Fig. S1 in the Supplement). However, R1 and R3 simulations are prone to exaggerate the seasonality by over-estimating precipitation in warm months, especially for summer-dominant precipitation zones. R2 is better than R1 and R3 in matching precipitation seasonality. The bias and MAE for precipitation are larger in warm months for summer-dominant and uniform precipitation zones, and in cold months for winter-dominant precipitation zones. All simulations can capture the seasonality of maximum and minimum temperatures (Fig. S2). However, there is a 2–3°C constant cold bias for maximum temperature, and 1–2°C cold bias for minimum temperature, except for R3, which overestimates minimum temperature in cold months and underestimates minimum temperature in warm months. R3 has the least bias in temperature when compared with R1 and R2.

Table 2. Mean absolute error (MAE) for monthly precipitation, maximum temperature (Tmax) and minimum temperature (Tmin) in 3 simulations (R1, R2, R3) for 11 climate zones (see Fig. 3 for zones). The smallest MAE for each climate zone is highlighted in **bold**

Zone	MAE								
	Precipitation (%)			Tmax (°C)			Tmin (°C)		
	R1	R2	R3	R1	R2	R3	R1	R2	R3
1	52.4	**17.6**	52.0	3.0	2.5	**2.1**	2.1	2.4	**0.6**
2	84.7	**22.9**	101.8	3.1	2.5	**2.4**	1.8	2.1	**0.9**
3	58.9	**24.8**	45.8	3.2	2.7	**2.0**	1.5	1.6	**0.3**
4	70.7	**26.1**	70.3	3.2	2.7	**2.3**	1.7	1.8	**0.5**
5	38.6	**11.6**	29.7	3.2	2.8	**2.2**	2.4	2.8	**0.5**
6	24.1	**12.5**	22.7	2.5	2.4	**1.8**	2.4	2.8	**0.7**
7	19.7	18.5	**14.6**	2.3	2.3	**1.6**	2.0	2.5	**1.3**
8	30.1	**12.8**	22.3	3.3	2.9	**2.1**	2.3	2.7	**0.8**
9	43.6	**20.3**	56.8	3.2	2.8	**2.2**	1.7	1.8	**0.9**
10	51.5	**26.2**	65.5	3.2	2.7	**2.1**	1.3	1.4	**0.9**
11	53.9	**33.4**	61.6	3.1	2.6	**1.8**	1.0	1.1	**0.4**
Mean	48.0	**20.6**	49.4	3.0	2.6	**2.1**	1.8	2.1	**0.7**

The MAE for monthly precipitation and temperature for 11 climate zones are summarized in Table 2. For precipitation, the domain-averaged MAE for the R2 simulation is about 20%, which is less than half of the MAE for R1 and R3 simulations. The MAE for R2 is about one-third to a half of the MAE for R1 and R3 for all climate zones, except for Zone 7, where the MAE for R2 is similar to that for R1 and R3. The domain-averaged MAEs for maximum and minimum temperature are about 3, 2.6 and 2.1°C, and 1.8, 2.1 and 0.7°C for R1, R2 and R3, respectively. The MAE for maximum temperature is consistently smaller for R3 compared to R1 and R2 for all climate zones. The MAE for minimum temperature is much smaller (less than one-third of the maximum MAE for most climate zones) for R3 than R1 and R2.

4.3. Seasonal and annual time scales

Fig. 6 shows the biases of precipitation and maximum and minimum temperatures at annual time scales (results at seasonal scales are shown in Figs. S3–S5). The relative biases demonstrate how differently the 3 simulations behave.

The R2 simulation produces reasonably good results over most of the domain and throughout the year, although some positive biases are observed along the Great Dividing Range, particularly during the autumn (MAM). In addition, the model also overestimates precipitation in the northwest of the domain, especially during the winter (JJA) and spring (SON). The R2 performance is especially good for summer (DJF) where the bias is below 25% over the majority of the region. As the northern part of the domain is a summer-dominant precipitation zone, the low bias in summer precipitation leads to low bias in annual precipitation. R1 and R3 are similar in bias for annual and seasonal precipitation. They generally demonstrate good performance in precipitation for most of the domain with a bias below 50%, but substantially overestimate precipitation along the Great Dividing Range and the northwest of the domain for all seasons, especially in autumn (MAM) and spring (SON), leading to larger annual deviations.

All simulations have a negative bias in annual and seasonal maximum temperature across the domain. The cold biases for the R3 simulation are generally <3°C for annual and seasonal maximum temperature, which is smaller than those for R1 and R2. The largest biases can be seen in winter (JJA) and the smallest biases in summer (DJF), which are similar to the results from Andrys et al. (2015).

R1 and R2 have cold biases for annual and seasonal minimum temperature for most of the domain, with larger biases (>4°C) in winter and spring relative to

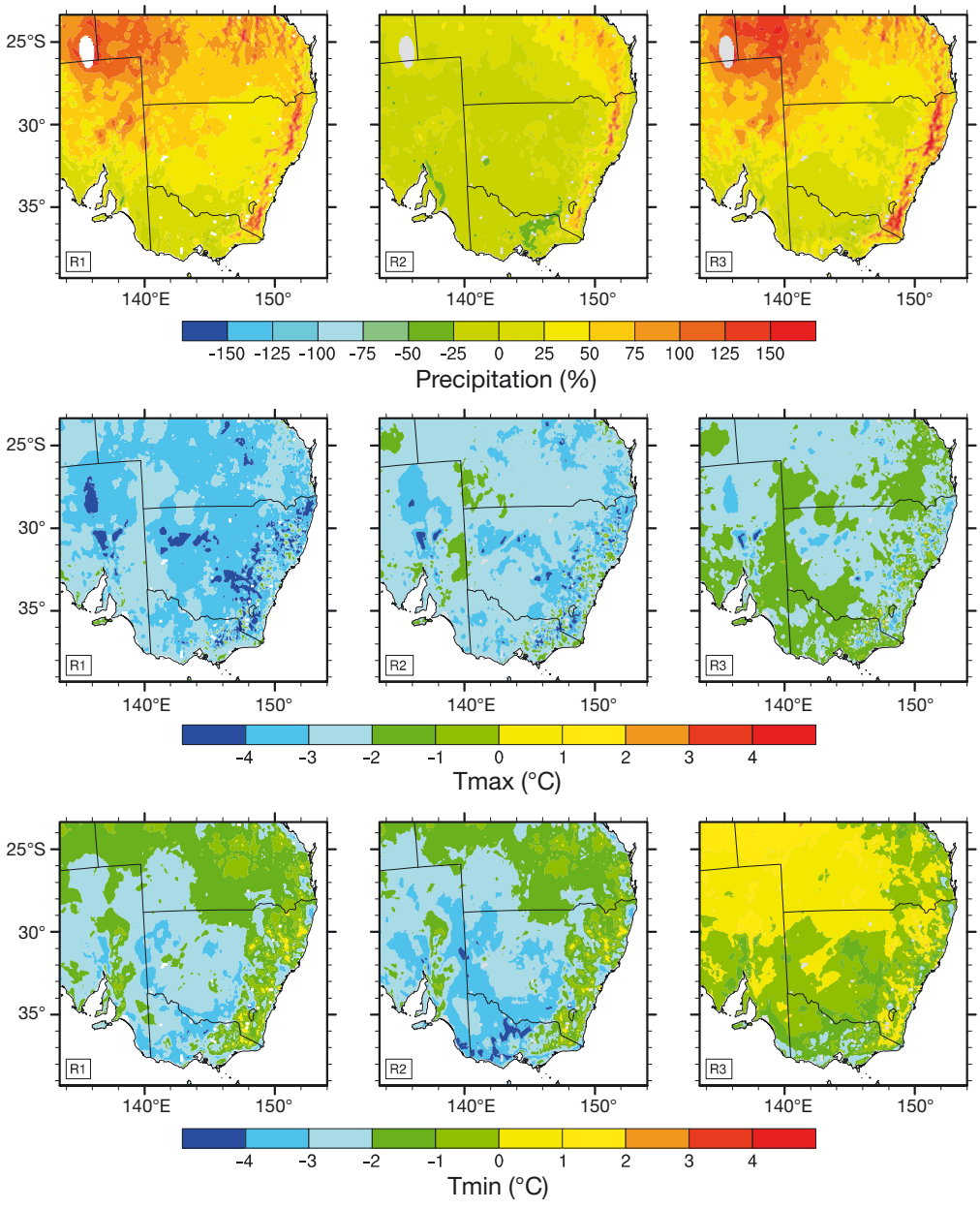
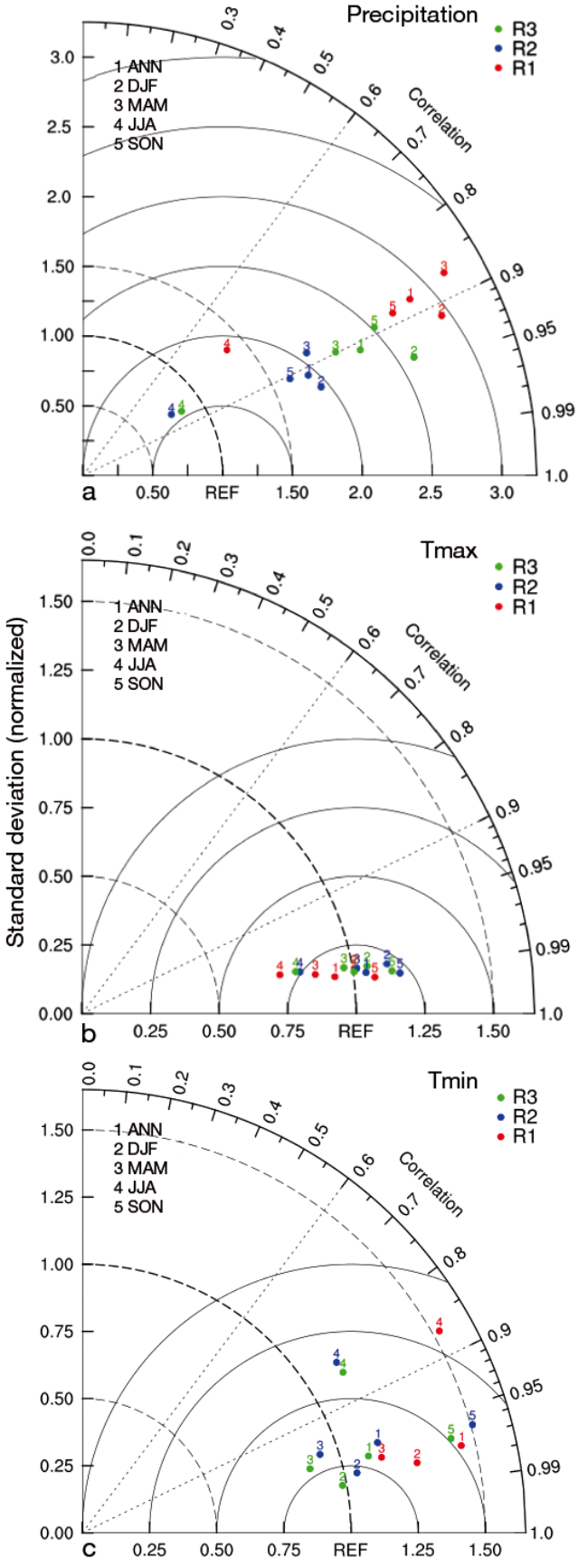


Fig. 6. Biases in annual precipitation (top row), maximum temperature (central row) and minimum temperature (bottom row) for R1, R2 and R3 simulations

summer and autumn (2–3°C), especially in the south. The bias for R3 is slightly different from R1 and R2; it has a positive bias in the north to northwest of the domain and negative bias elsewhere. Positive biases occur across seasons, with larger values in spring (SON) which cause the positive annual deviations in those areas. The magnitude of biases for R3 is under 2°C, which is similar to the range reported by Andrys et al. (2015). The magnitude of biases for R1 and R2 is slightly larger than that for R3, although it is still smaller than in some of the simulations by Katragkou et al. (2015).

The Taylor diagram provides a different angle to show a comparison between simulations and observations. Fig. 7 allows a succinct comparison of models using metrics of spatial correlation (R), RMS difference (the RMS differences between the dots and 'REF' field are proportional to the their distance apart, in the same units as standard deviation. See Taylor 2001) and the variance ratio calculated for annual and seasonal precipitation and maximum and minimum temperatures. Better model performance is indicated by values that are close to the *x*-axis and near the line that is labelled 'REF', which represents a 1:1 variance ratio.

Overall, model performance indicates a very high spatial agreement with observations (0.85 for precipitation and 0.95 for temperature,) with the highest



correlation in summer (DJF) and lowest correlation in winter (JJA) for precipitation and minimum temperature, and no substantial difference between seasons for maximum temperature. The amplitude of variability is generally higher than the observations for most seasonal precipitation and minimum temperature, with smallest variability in winter for precipitation and largest variability in winter for minimum temperature. Seasonal differences for maximum temperature are not substantial. When comparing the 3 simulations, R2 performs better in reproducing precipitation and R3 for minimum temperature. There is not much difference among the 3 simulations with regards to how well they reproduce maximum temperature.

4.4. Inter-annual time scale

Plots of the precipitation, maximum and minimum temperature annual anomalies for Zone 8 are shown in Fig. 8 as an example. All simulations are able to reproduce the observed precipitation annual anomalies, especially for wet periods in the late 1950s, mid-1970s and 1990s and dry periods in the late 1960s, 1980s and 2000s. The temporal correlations of annual precipitation anomalies are 0.47, 0.62 and 0.44 for R1, R2 and R3, respectively. For temperature, all simulations capture the observed maximum and minimum temperature annual anomalies better for warm periods in the 1980s and 2000s and cold periods in the late 1950s and late 1970s. The temporal correlations for temperature are higher than those for precipitation, which are 0.68, 0.74 and 0.70, and 0.78, 0.76 and 0.77 for R1, R2 and R3 simulated maximum temperature and minimum temperature, respectively. There is a clear rising trend in observed temperatures with about 0.02 and 0.01°C yr^{-1} rate of increase for maximum and minimum temperature, respectively, which is well represented in all 3 simulations. However, simulations tend to exaggerate the rate of increase in minimum temperature.

The performance of the 3 simulations is not easily judged from Fig. 8 alone. However, the R2 simulation stands out with the least bias in long-term means and the highest temporal anomaly correlation in precipi-

Fig. 7. Taylor diagrams for (a) precipitation, (b) maximum temperature and (c) minimum temperature. The 3 simulations (R1, R2, R3) are shown in different colours, and annual means are labelled with a 1 and seasons are numbered from 2 to 5 for DJF, MAM, JJA and SON, respectively

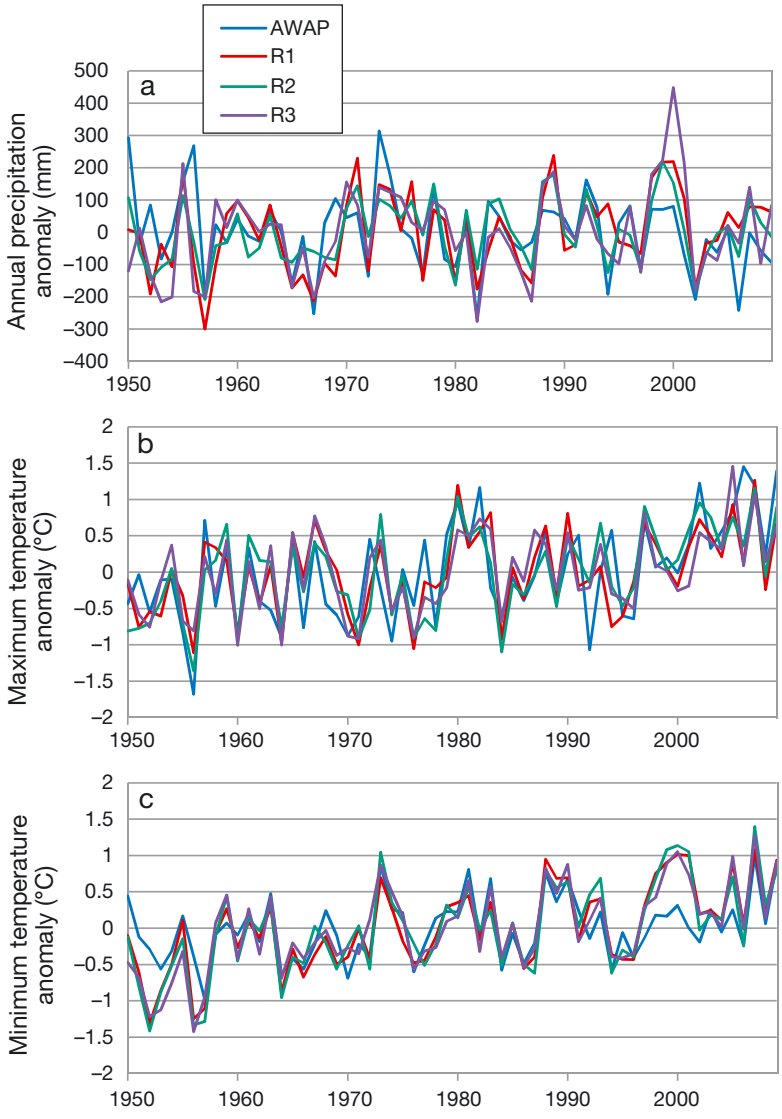


Fig. 8. Anomaly time series for (a) annual precipitation, (b) maximum temperature and (c) minimum temperature for Zone 8 (see Fig. 3 for zones) from 1950–2009. The 4 different colours represent 1 observation (AWAP) and 3 simulations (R1, R2, R3)

Table 3. Temporal correlations (R) between simulations (R1, R2, R3) and observations for anomalies in annual precipitation and maximum and minimum temperatures for 11 climate zones (see Fig. 3 for zones). The largest R value for each climate zone is highlighted in **bold**

Zone	R								
	Precipitation			Tmax			Tmin		
	R1	R2	R3	R1	R2	R3	R1	R2	R3
1	0.40	**0.66**	0.44	0.66	**0.67**	0.61	**0.83**	0.76	**0.83**
2	0.34	**0.71**	0.47	**0.63**	0.62	0.59	0.76	0.69	**0.80**
3	0.41	**0.51**	0.41	0.55	**0.65**	0.60	0.75	0.78	**0.81**
4	0.47	**0.51**	0.48	0.52	**0.58**	**0.58**	0.70	0.74	**0.80**
5	0.45	**0.68**	0.46	0.68	**0.75**	0.67	**0.80**	0.76	0.78
6	0.42	0.46	**0.59**	0.70	**0.75**	0.73	0.79	**0.80**	0.78
7	0.61	0.56	**0.65**	0.79	0.82	**0.83**	**0.86**	**0.86**	**0.86**
8	0.47	**0.62**	0.44	0.68	**0.74**	0.70	**0.78**	0.76	0.77
9	0.29	**0.38**	0.33	**0.64**	0.60	**0.64**	0.75	**0.78**	0.77
10	0.44	**0.56**	0.45	0.68	**0.72**	0.68	0.76	**0.77**	0.75
11	0.42	**0.51**	0.41	0.68	**0.73**	0.68	**0.79**	0.77	**0.79**
Mean	0.43	**0.56**	0.47	0.65	**0.69**	0.66	0.78	0.77	**0.79**

tation, and the R3 simulation is superior with the least bias in the long-term temperature mean.

The results for other zones are provided in the Supplement (Figs. S6–S8). The temporal correlations for other zones are summarized in Table 3. The R2 simulation is better correlated to observed precipitation temporally and shows less bias in simulating long-term precipitation means for most of the zones, except Zones 6 and 7 that are winter-dominant precipitation zones. The R3 simulation shows the least bias in long-term temperature means and highest temporal correlation to observed minimum-temperature anomalies. However, the R3 temporal correlations to observed maximum-temperature anomalies are not the best among the 3 simulations. The temporal correlations for temperature are higher than those for precipitation, which suggests that WRF performs better in capturing long-term variability of temperature than precipitation.

4.5. Decadal time scale

The biases at decadal time scales are presented in Figs. S9–S11. For precipitation, the simulations perform well, with <50% in bias for most of the regions across all decades, with the exception of the northwest part of the domain and along the Great Dividing Range, which are overestimated. The 1950s and 1970s are wet decades for southeast Australia; during these periods, the biases are smaller than other time periods. Within the 3 simulations, R2 out-performs R1 and R3 by having the least bias in the northwest of the domain and along the Great Dividing Range.

There is a larger bias in maximum temperature for all decades. The R1 simulation has the largest bias (3–4°C cooler

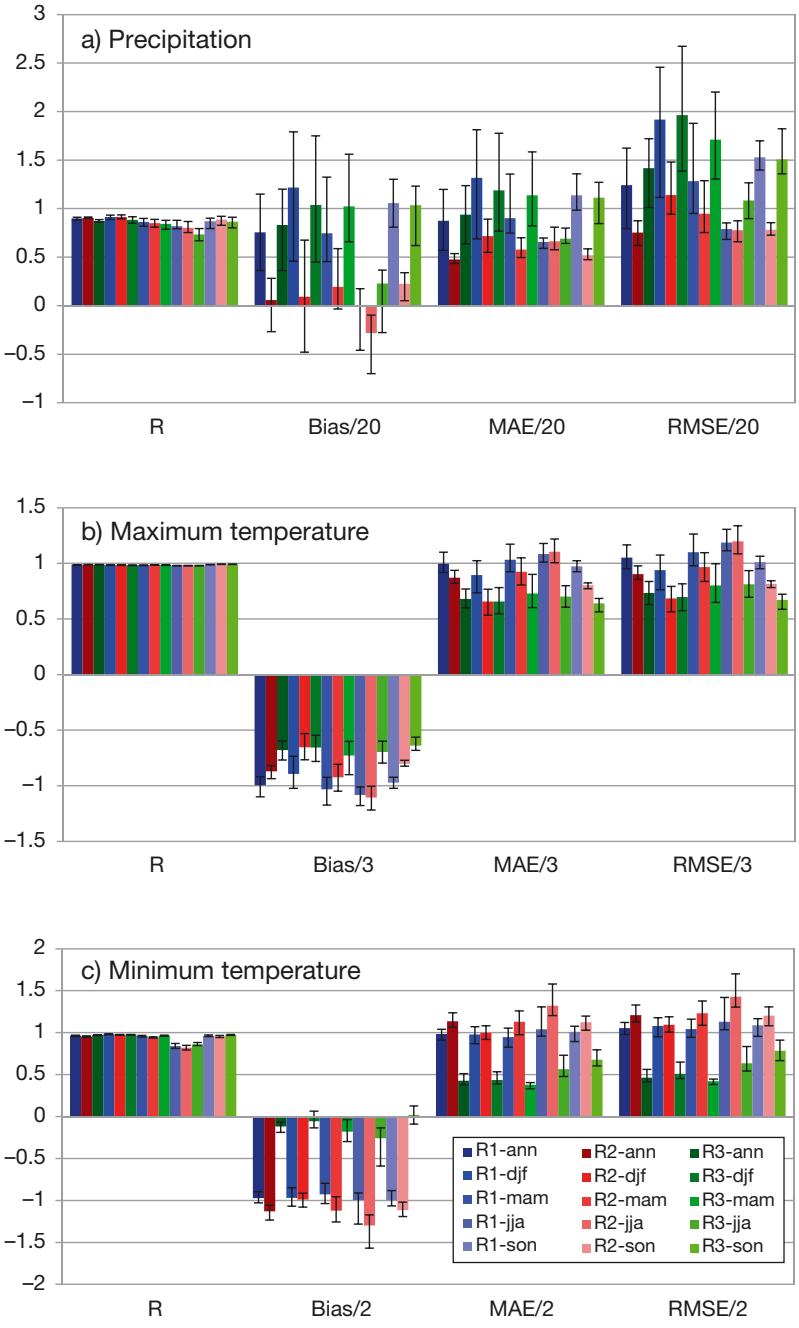


Fig. 9. Statistical results for decadal (a) mean precipitation, (b) maximum temperature and (c) minimum temperature based on annual and 4-season data. The block columns represent results for 60 yr means and error bars represent the inter-decade range (minimum to maximum). Blue, red and green are for R1, R2 and R3 simulations, respectively. Bias, mean absolute error (MAE) and root mean square error (RMSE) were divided by 20, 3 and 2, respectively, for precipitation, maximum temperature and minimum temperature, to allow them to be plotted in the same graph

than the observations), followed by R2 (2–3°C bias). R3 has the least negative bias (1–2°C).

The bias in minimum temperature shows interesting patterns. It is generally smaller in the north of the domain and coastal areas than elsewhere. R1 and R2 have a negative bias while R3 has a positive bias in the north and a negative bias in the south of the domain. The magnitude of bias is smaller for R3 than for R1 and R2.

The decadal statistical results for annual and seasonal precipitation and maximum and minimum temperatures are presented in Fig. 9. The columns represent the results for the 60 yr means of precipitation and temperature, and error bars show the inter-decade range (minimum to maximum).

The results show good spatial correlations (>0.9) for long-term mean precipitation with some seasonal variation (smaller value in winter: 0.75). R2 is slightly better than R1 and R3 for most seasons. All simulations have a positive bias in annual and seasonal precipitation, except for R2 simulated winter precipitation. The magnitude of bias is much smaller for R2 than R1 and R3 for most seasons, with no clear seasonal change. However, a larger bias of ca. 40 mm mo^{-1} can be seen in summer (DJF) and spring (SON) for R1 and R3. The same seasonality for the MAE and RMSE is also observed for R1 and R3. The magnitudes of the MAE and RSME are smaller for R2 than R1 and R3 for annual and seasonal precipitation. There is not much difference in inter-decade range of R for the 3 simulations, but the range of the bias, MAE and RMSE for R2 is consistently smaller than R1 and R3 for annual and seasonal precipitation.

The spatial correlations of maximum and minimum temperatures are better than that of precipitation, even if there is a similar seasonal change in R for minimum temperature with low values in winter. There is a consistent negative bias in maximum and minimum temperatures, with much smaller bias for R3 than R1 and R2, which have a 2–3°C cold bias for maximum temperature and a 1–2°C for minimum temperature. The MAE and RMSE values indicate a range similar to that for bias. Among the 3 simulations, R3 has much

smaller systematic bias than R1 and R2 for annual and seasonal temperatures.

4.6. ETCCDI

A number of ETCCDI are examined, including: number of days with daily precipitation >10 mm (R10), number of hot days and number of frost days (FD). The results are shown in Fig. 10.

Results for the R10 index indicate that the general gradient from east to west and 2 centres on the northeast NSW coast and Snowy Mountains are accurately captured by all simulations. However, all simulations overestimate along the Great Dividing Range, especially on the northeast NSW coast. R1 and R3 simulations also overestimate days with daily precipitation >10 mm in the northern part of the domain, which implies that the rainfall bias is mainly caused by larger events. The bias from the R2 simulation is far smaller than the other two simulations.

The cold bias in the temperature simulations is reflected in the hot day results. The spatial area with <10 hot days is slightly overestimated in the simula-

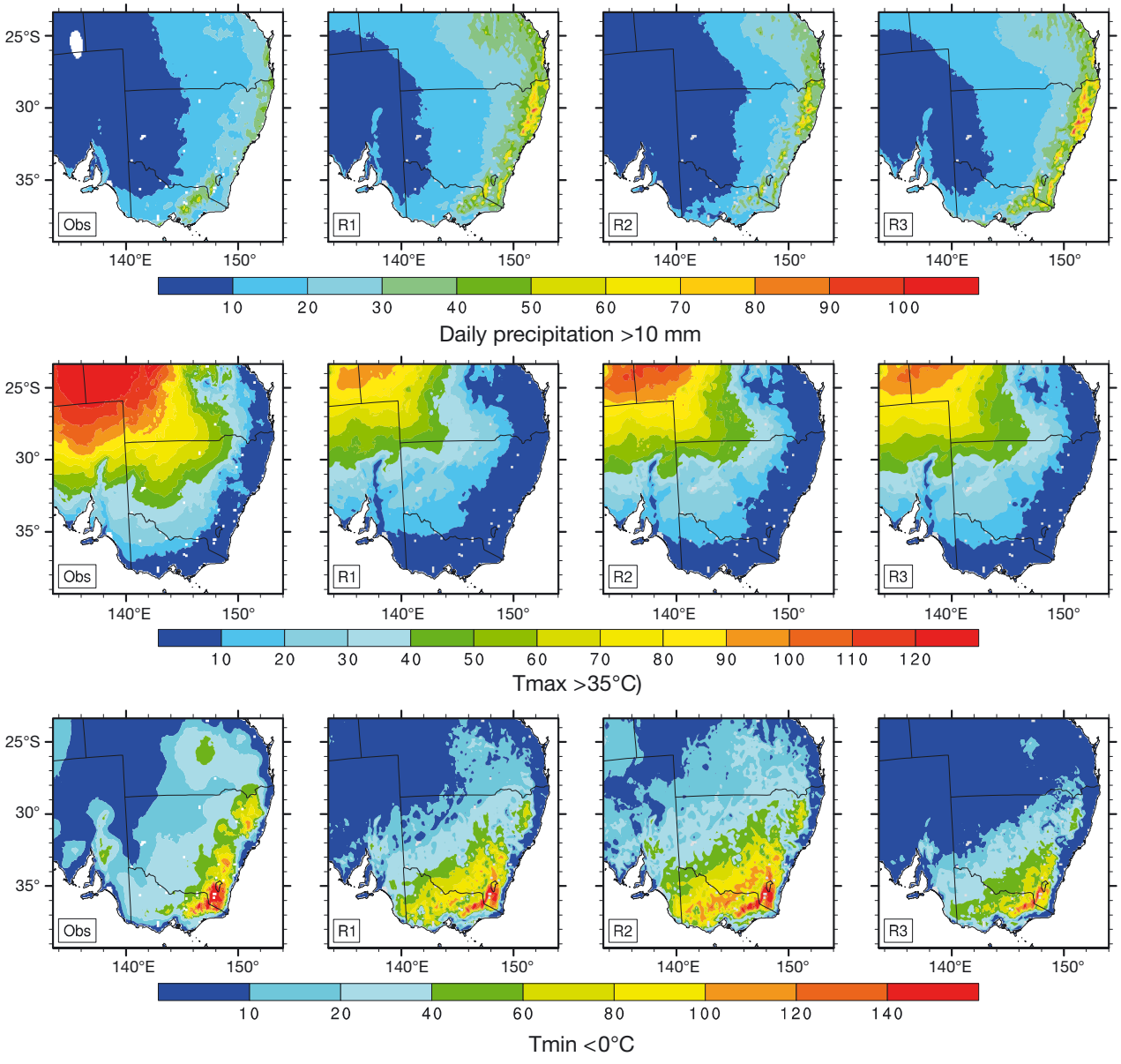


Fig. 10. Days with daily precipitation >10 mm (top row), maximum temperature >35°C (centre row) and minimum temperature <0°C (bottom row) calculated from observations (obs) and 3 simulations (R1, R2, R3) from 1950–2009

tions. However, the northwest to southeast gradient is nearly perfectly represented by all simulations, although the simulations underestimate the number of hot days in the northwest of the domain. R2 and R3 simulations give better results than the R1 simulation.

The FD index results show that there are >40 frost days a year along the Great Dividing Range, with the longest period in the Snowy Mountains where the frost period exceeds 5 mo. This distinctive feature is fully characterized in all 3 simulations. Nevertheless, the simulations, especially R1 and R2, overestimate the frost period in the southern part of the domain. All 3 simulations tend to slightly underestimate the frost period in the northern part of the domain, which is particularly true for R3.

5. DISCUSSION

All 3 simulations reproduce precipitation and temperature reasonably well at different time scales (daily, monthly, seasonal, inter-annual and decadal) for the domain of interest. However, there is a positive bias in precipitation and a negative bias in maximum and minimum temperatures.

As the precipitation in the study area is dominated by convective rainfall during warm months, it was expected that the simulated precipitation is sensitive to physics scheme combinations (Flaounas et al. 2011, Crétat et al. 2012). As the main focus of this study was to assess the capability of WRF to simulate regional climate for southeast Australia, the design of the experiments made it difficult to analyse the impact of the physics scheme on modelled precipitation. However, the influence of the physics scheme selection can still be observed in the modelling results. R2 has the same physics scheme options as R1 except for the cumulus scheme (CU). The Betts-Miller-Janjic (BMJ) scheme is used in R2, while the Kain–Fritsch (KF) scheme is used in R1. The difference between R1 and R2 simulated precipitation shows that the BMJ scheme is better than the KF scheme in terms of monthly biases and absolute errors and also regarding the daily PDF. This suggests that the KF scheme is less suitable for long-term climate simulations for southeast Australia. This result is surprising, as it contradicts the findings from previous studies (Cohen 2002, Kerkhoven et al. 2006, Ji et al. 2014, Pennelly et al. 2014, Gilmore et al. in press) that considered the KF scheme as a good option for capturing major precipitation centres in short-term weather simulations. These studies compared different CU schemes in simulating convective rainfall and concluded that the KF scheme is well suited for mesoscale model runs, as its assumption about the consumption of convective available potential energy is appropriate for mesoscale time and space scales. The scheme has a realistic convective trigger, and it considers entrainment and detrainment more realistically than other schemes. However, some of the studies have also pointed out that although the KF scheme simulates heavy and moderate phases of rainfall well, it has a tendency to over-estimate light rainfall, which has little effect on short-term simulations, but can be problematic for long-term climatic applications (Kerkhoven et al. 2006). In our study, the 2 RCMs that used the KF cumulus scheme (R1 and R3) both overestimated light rainfall, which contributed to the overall overestimation of annual rainfall for the region, especially in the northwest domain where the annual rainfall is low (<200 mm yr^{-1}). Alapaty et al. (2012) also reported the excessive precipitation simulated by the WRF model, and suggested that by including subgrid-scale cloud-radiation interactions in the KF scheme, the overestimated precipitation can be largely reduced in both weather (short-term) and climatic (long-term) simulations.

The simulations generally show a consistent cold bias, especially for maximum temperature; the cold bias is much larger in magnitude than the difference among the 3 simulations. This indicates that the main source of the bias is not the choice of planetary boundary layer (PBL), CU, microphysics (MP) and radiation (RA). Evans & McCabe (2010) found that NNRP has a cold bias for southeast Australia. Kala et al. (2015) also noted that NNRP was not as accurate as ERA-I data, with a larger bias in temperature. It is reasonable to deduce that, at least partially, the systematic bias in temperature simulations might be inherited from the driving data, although further work is required to confirm this hypothesis.

Zeng et al. (2015) compared 4 land surface schemes (LSS) in the WRF model to simulate high-temperature events. They found that the surface air temperature is sensitive to LSS and lead time. Their analysis showed that Noah LSS generally underestimates the surface air temperature, and the systematic simulation error increases with lead time. García-Díez et al. (2015) suggested that the Noah scheme overestimates total soil moisture content. The excess soil moisture influences the energy partitioning, shifting it towards too much evaporation, and making it difficult for the occurrence of a 'dry regime' where evapotranspiration is limited by soil moisture instead of incoming energy.

Therefore Noah LSS, which is used in all 3 of our simulations, is another source of the cold bias shown in the temperature.

The large difference in modelled temperature between R1 and R3 simulations suggests that the modelled temperature is sensitive to the choice of PBL and RA schemes since they use the same CU and MP schemes but different PBL and RA schemes. This is consistent to the findings from Evans et al. (2012) and Kala et al. (2015).

ETCCDI indices are good indictors for assessing the capability of WRF to simulate regional climate variability. We only presented a few indices here to demonstrate the capability of WRF. The results do show that WRF has skill in simulating regional climate variability and moderate extremes.

The results from this study provide end-users ground for confidence in the WRF model for long-term regional climate simulation in southeast Australia. The NARCliM project used 4 GCMs to drive 3 RCMs (total of 12 ensemble members) for 3 time periods (1990–2009, 2020–2039, 2060–2079). The results from the GCM-driven R2 simulations should be more trustworthy in terms of precipitation relative to other simulations in the ensemble. Similarly, minimum temperatures for R3 simulations are likely to be better than those from other simulations.

6. CONCLUSIONS

This paper presents an evaluation of 3 RCM simulations over southeast Australia for a 60 yr period (1950–2009) to examine the ability of the WRF model to accurately represent the climatology of the region.

At the daily time scale, the daily precipitation PDF is reproduced with good skill throughout the domain, with slightly poorer skill along the Great Dividing Range. In terms of maximum and minimum temperatures, the overall shape of the temperature PDF is good throughout the domain, with some poorer performance found in southern Victoria. Among the 3 simulations, R2 is the best in capturing daily precipitation, and R3 is the most skilled at simulating minimum temperature.

At the monthly time scale, the simulations tend to overestimate precipitation in warm months and systematically underestimate temperature. However, the simulations capture the annual cycles rather well. R2 and R3 simulations performed best in simulating precipitation and temperature, respectively.

At seasonal time scales, the simulations generally perform well but tend to overestimate precipitation along the Great Dividing Range and in the northwest of the domain for all seasons. In general, better precipitation results are generated for winter compared to other seasons. The simulations capture the spatial pattern of maximum and minimum temperatures relatively well despite a 2–3°C cold bias for maximum temperature and 1–2°C cold bias for minimum temperature. The performance of the 3 simulations does not differ much when simulating maximum temperatures. However, R2 and R3 are better in performance when simulating precipitation and minimum temperature individually.

The simulations are able to reproduce the interannual variability accurately, especially for the major wet/dry and hot/cold periods. R2 has the highest temporal correlations for annual precipitation anomalies and the least bias in long-term means for precipitation, and R3 has the smallest bias in long-term means for temperature.

The ability of WRF to capture the variability in precipitation and temperature at decadal scales is also evaluated. Generally, WRF performs well in reproducing the spatial patterns of precipitation and temperature, but bias in the amount of precipitation and temperature is substantial across the domain. R2 and R3 simulations are superior in reproducing decadal variability in precipitation and temperature separately.

The differences in precipitation and temperature between the simulations indicated that the CU scheme is critical to precipitation simulations, and PBL and RA schemes are more important in temperature simulations.

Overall, the simulations plausibly represent the spatial and temporal patterns of observations. However, it is evident that there is overestimation of precipitation along the Great Dividing Range and in the northwest of the domain, and systematic underestimation of temperature across the domain. The R2 simulation is the most skilled at reproducing precipitation and R3 at reproducing temperature at varying time scales.

Acknowledgements. This work was made possible by funding from the NSW Environmental Trust for the Eastern Seaboard Climate Change Initiative–East Coast Low Project, the NSW Office of Environment and Heritage backed NSW/ACT Regional Climate Modelling Project (NARCliM) and the Australian Research Council as part of the Future Fellowship FT110100576 and Linkage Project LP120200777. The modelling work was undertaken on the NCI high performance computers in Canberra, Australia, supported by the Australian Commonwealth Government.

LITERATURE CITED

Alapaty K, Herwehe JA, Otte TL, Nolte CG and others (2012) Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling. Geophys Res Lett 39:L24809, doi:10.1029/2012GL054031

Andrys J, Lyons TJ, Kala J (2015) Multi-decadal evaluation of WRF downscaling capabilities over Western Australia in simulating precipitation and temperature extremes. J Appl Meteorol Climatol 54:370–394

Argüeso D, Hidalgo-Muñoz JM, Gámiz-Fortis SR, Esteban-Parra MJ, Dudhia J, Castro-Díez Y (2011) Evaluation of WRF parameterizations for climate studies over southern Spain using a multistep regionalization. J Clim 24: 5633–5651

Argüeso D, Hidalgo-Muñoz JM, Gámiz-Fortis SR, Esteban-Parra MJ, Castro-Díez Y (2012) Evaluation of WRF mean and extreme precipitation over Spain: present climate (1970–1999). J Clim 25:4883–4897

Christensen J, Carter T, Rummukainen M, Amanatidis G (2007) Evaluating the performance and utility of regional climate models: the PRUDENCE project. Clim Change 81:1–6

Cohen C (2002) A comparison of cumulus parameterizations in idealized sea-breeze simulations. Mon Weather Rev 130:2554–2571

Crétat J, Pohl B, Richard Y, Drobinski P (2012) Uncertainties in simulating regional climate of southern Africa: sensitivity to physical parameterizations using WRF. Clim Dyn 38:613–634

Déqué M, Jones RG, Wild M, Giorgi F and others (2005) Global high resolution versus limited area model climate change projections over Europe: quantifying confidence level from PRUDENCE results. Clim Dyn 25:653–670

Evans JP, Ji F (2012a) Choosing GCMs. NARCliM Technical Note 1. NARCliM Consortium, Sydney

Evans JP, Ji F (2012b) Choosing the RCMs to perform the downscaling. NARCliM Technical Note 2. NARCliM Consortium, Sydney

Evans JP, McCabe MF (2010) Regional climate simulation over Australia's Murray Darling basin: a multitemporal assessment. J Geophys Res 115:D14114, doi:10.1029/2010JD013816

Evans JP, McCabe MF (2013) Effect of model resolution on a regional climate model simulation over southeast Australia. Clim Res 56:131–145

Evans JP, Westra S (2012) Investigating the mechanisms of diurnal precipitation variability using a regional climate model. J Clim 25:7232–7247

Evans JP, Ekstrom M, Ji F (2012) Evaluating the performance of a WRF physics ensemble over South-East Australia. Clim Dyn 39:1241–1258

Evans JP, Ji F, Abramowitz G, Ekström M (2013a) Optimally choosing small ensemble members to produce robust climate simulations. Environ Res Lett 8:044050

Evans JP, Fita L, Argüeso D, Liu Y (2013b) Initial NARCliM evaluation. In: Piantadosi J, Anderssen RS, Boland J (eds) MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Adelaide, p 2765–2771

Evans JP, Ji F, Lee C, Smith P, Argüeso D, Fita L (2014) Design of a regional climate modelling projection ensemble experiment – NARCliM. Geosci Model Dev 7: 621–629

Flaounas E, Bastin S, Janicot S (2011) Regional climate modelling of the 2006 West African monsoon: sensitivity to convection and planetary boundary layer parameterisation using WRF. Clim Dyn 36:1083–1105

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL (2003) Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. J Geophys Res 108:4124, doi:10.1029/2005JD005965

Frei C, Schöll R, Fukutome S, Schmidli J, Vidale PL (2006) Future change of precipitation extremes in Europe: intercomparison of scenarios from regional climate models. J Geophys Res 111:D06105, doi:10.1029/2005JD005965

Fu C, Wang S, Xiong Z, Gutowski WJ and others (2005) Regional Climate Model Intercomparison Project for Asia. Bull Am Meteorol Soc 86:257–266

García-Díez M, Fernández J, Vautard R (2015) An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. Clim Dyn 45:1–16

Gilmore JB, Evans JP, Sherwood SC, Ekström M, Ji F (in press) Extreme precipitation in WRF during the Newcastle East Coast Low of 2007. Theor Appl Climatol, doi:10.1007/s00704-015-1551-6

Giorgi F, Lionello P (2008) Climate change projections for the Mediterranean region. Global Planet Change 63: 90–104

Giorgi F, Jones C, Asrar GR (2009) Addressing climate information needs at the regional level: the CORDEX framework. WMO Bull 58:175–183

Ji F, Evans JP, Ekstrom M (2011) Using dynamical downscaling to simulate rainfall for East Coast Low events. In: Chan F, Marinova D, Anderssen RS (eds) MODSIM2011, 19th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, Perth, p 2733–2739

Ji F, Ekstrom M, Evans JP, Teng J (2014) Evaluating precipitation patterns using physics scheme ensembles from a regional atmospheric model. Theor Appl Climatol 115: 297–304

Jones D, Wang W, Fawcett R (2009) High-quality spatial climate datasets for Australia. Aust Meteorol Mag 58: 233–248

Kala J, Andrys J, Lyons TJ, Foster IJ, Evans BJ (2015) Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. Clim Dyn 44:633–659

Kalnay E, Kanamitsu M, Kistler R, Collins W and others (1996) The NCEP/NCAR 40 year reanalysis project. Bull Am Meteorol Soc 77:437–471

Katragkou E, García-Díez M, Vautard R, Sobolowski S and others (2015) Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. Geosci Model Dev 8:603–618

Kerkhoven E, Gan TY, Shiiba M, Reuter G, Tanaka K (2006) A comparison of cumulus parameterization schemes in a numerical weather prediction model for a monsoon rainfall event. Hydrol Process 20:1961–1978

Mearns LO, Arritt R, Biner S, Bukovsky MS and others (2012) The north American regional climate change assessment program: overview of phase I results. Bull Am Meteorol Soc 93:1337–1362

Peña-Arancibia JL, van Dijk AIJM, Renzullo LJ, Mulligan M (2013) Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method

for regions in Australia and South and East Asia. J Hydrometeorol 14:1323–1333

Pennelly C, Reuter G, Flesch T (2014) Verification of the WRF model for simulating heavy precipitation in Alberta. Atmos Res 135–136:172–192

Perkins SE, Pitman AJ, Holbrook N, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376

Skamarock WC, Klemp JB, Dudhia J, Gill DO and others (2008) A description of the Advanced Research WRF Version 3. NCAR Technical Note. NCAR, Boulder, CO

Solman SA, Sanchez E, Samuelsson P, da Rocha RP and others (2013) Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: model performance and uncertainties. Clim Dyn 41:1139–1157

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106: 7183–7192

Van der Linden P, Mitchell JFB (eds) (2009) ENSEMBLES: climate change and its impacts. Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Exeter

Vaze J, Teng J, Chiew FHS (2011) Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. Hydrol Process 25:1486–1497

Zeng XM, Wang N, Wang Y, Zheng Y and others (2015) WRF-simulated sensitivity to land surface schemes in short and medium ranges for a high temperature event in East China: a comparative study. J Adv Model Earth Syst 7:1305–1325

Editorial responsibility: Filippo Giorgi, Trieste, Italy

Submitted: August 4, 2015; Accepted: December 14, 2015
Proofs received from author(s): February 1, 2016