Di Virgilio, G.; Evans, J.P.; Di Luca, A.; Olsen, R.; Argueso, D.; Kala, J.; Andrys, J.; Hoffmann, P.; Katzfey, J.J.; Rockel, B.:

Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors.

In: Climate Dynamics. Vol. 53 (2019) 5-6, 2985 - 3005.
First published online by Springer: 26.02.2019

DOI: 10.1007/s00382-019-04672-w
https://dx.doi.org/10.1007/s00382-019-04672-w

Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors

Di Virgilio, Giovanni[1], Evans, Jason P.[1, 2], Di Luca, Alejandro[1,2], Olson, Roman [3,4,5], Argüeso, Daniel[6], Kala, Jatin[7], Andrys, Julia[7], Hoffmann, Peter[8,9], Katzfey, Jack J.[8], Rockel, Burkhardt[10]

[1]*Climate Change Research Centre, University of New South Wales, Sydney, Australia*
[2]*Australian Research Council Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, Australia*
[3]*Department of Atmospheric Sciences, Yonsei University, Republic of Korea*
[4]*Center for Climate Physics, Institute for Basic Science, Busan, Republic of Korea*
[5]*Pusan National University, Busan, Republic of Korea*
[6]*Department of Physics, University of Balearic Islands, Palma de Mallorca, Spain*
[7]*School of Veterinary and Life Sciences, Environmental and Conservation Sciences, Murdoch University, Western Australia, Australia*
[8]*Climate Science Centre - CSIRO Oceans and Atmosphere, Aspendale, Victoria, Australia*
[9]*Climate Service Center Germany (GERICS), Helmholtz-Zentrum Geesthacht, Hamburg, Germany*
[10]*Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany*

Correspondence: Giovanni Di Virgilio (giovanni@unsw.edu.au)

Abstract. The ability of regional climate models (RCMs) to accurately simulate current and future climate is increasingly important for impact assessment. This is the first evaluation of all reanalysis-driven RCMs within the CORDEX Australasia framework (four configurations of the Weather Forecasting and Research (WRF) model, and single configurations of COSMO-CLM (CCLM) and the Conformal-Cubic Atmospheric Model (CCAM) to simulate the historical climate of Australia (1981–2010) at 50 km resolution. Simulations of near-surface maximum and minimum temperature and precipitation were compared with gridded observations at annual, seasonal, and daily time scales. The spatial extent, sign, and statistical significance of biases varied markedly between the RCMs. However, all RCMs showed widespread, statistically significant cold biases in maximum temperature which were the largest during winter. This bias exceeded -5 K for some WRF configurations, and was the lowest for CCLM at ±2 K. Most WRF configurations and CCAM simulated minimum temperatures more accurately than maximum temperatures, with biases in the range of ±1.5 K. RCMs overestimated precipitation, especially over Australia's populous eastern seaboard. Strong negative correlations between mean monthly biases in precipitation and maximum temperature suggest that the maximum temperature cold bias is linked to precipitation overestimation. This analysis shows that the CORDEX Australasia ensemble is a valuable dataset for future impact studies, but improving the representation of land surface processes, and subsequently of surface temperatures, will improve RCM performance. The varying RCM capabilities identified here serve as a foundation for the development of future regional climate projections and impact assessments for Australia.

Keywords: CORDEX-Australasia; dynamical downscaling; model bias; precipitation; temperature

1. Introduction

Climate change is a global phenomenon with impacts that manifest at regional and local scales (IPCC 2013). Assessing how these changes will impact physical, ecological, and socio-economic systems and planning response strategies requires robust, high-resolution regional climate projections (IPCC 2012; Rummukainen 2016; Xue et al. 2014). Global climate models (GCMs) provide a basis for this information, however, their coarse resolution lacks the fine-scale details required by the assessment and adaptation planning community (Fowler et al. 2007; Hattermann et al. 2011; Maraun et al. 2010). An effective approach for producing high-resolution climate projections at regional scales is to use regional climate models (RCMs) to dynamically downscale coarse-resolution outputs from GCMs or reanalyses (Giorgi 2006; Laprise 2008; Wang et al. 2004). RCMs use these outputs as initial and lateral boundary conditions to generate projections that better resolve the complex surface characteristics and mesoscale atmospheric processes that are important drivers of regional climate (Di Luca et al. 2012; Giorgi and Bates 1989; Torma et al. 2015). With increased spatial resolution, RCMs can also better resolve convective phenomena and thus improve the simulation of extreme events, such as sub-daily precipitation extremes (Olsson et al. 2015). Accurate simulation of climate extremes by RCMs is increasingly important for climate impact assessment (Halmstad et al. 2013; Sunyer et al. 2017).

The Coordinated Regional Downscaling Experiment (CORDEX) is an initiative of the World Climate Research Programme (WCRP) that aims to improve both the generation and evaluation of downscaled regional climate information (Giorgi et al. 2009). Under the CORDEX framework, regional climate projections based on CMIP5 (Coupled Model Intercomparison Project Phase 5) GCM projections have been produced for fourteen regions worldwide. An important stage in RCM development and the production of future regional climate projections is the evaluation of the models' skill in simulating present-day climatological conditions (Di Luca et al. 2016; Diaconescu et al. 2015; Garcia-Diez et al. 2015). In this capacity, an essential component of CORDEX is the evaluation of multiple RCMs over recent decades using lateral boundary conditions from re-analysis products such as ERA-Interim (Dee et al. 2011).

Evaluations of historical CORDEX RCM simulations forced by ERA-Interim reanalysis have been completed for several regions. These assessments generally show that RCMs capture the main climatological features of the target domain; however, deficiencies are present which vary depending on the model, sub-region, and season. For example, when simulating observed precipitation in Africa, Nikulin et al. (2012) found that RCMs showed marked regional variation, and displayed shortcomings in arid and semi-arid regions. Furthermore, Panitz et al. (2014) reported a dry bias in regions affected by the passage of the West African Monsoon, warm biases in arid regions, and a cold bias over Guinea. RCMs showed reasonably high model accuracy over most of the Middle East and North African domain at annual timescales (Bucchignani et al. 2016). However, a warm summertime bias over North Africa and Saudi Arabia, and a cold bias over the majority of the domain during the boreal winter were also apparent. Evaluations of the EURO-CORDEX domain showed that RCMs simulated the basic spatiotemporal patterns of the European climate. However, model deficiencies included cold and wet biases during most seasons over the majority of Europe and warm and dry summer biases over southern and south-eastern Europe (Kotlarski et al. 2014). Although the general climatological features of South America were reproduced by RCMs, marked wet and cold biases were evident over several regions (Solman et al. 2013).

To date, no evaluation of CORDEX-Australasia has been performed and there is limited information available regarding the capability of ERA-Interim driven RCMs in simulating the Australian climate. While several studies have used RCMs driven with various reanalyses to produce regional climate hindcasts for different regions of the Australian continent (e.g., Evans et al. 2012; Andrys et al. 2015), no intercomparison study has evaluated the relative performance of different RCMs in simulating the Australian climate. Consequently, this paper has three main aims: 1) to evaluate the ability of the CORDEX-Australasia ensemble to simulate the historical temperature and precipitation characteristics of Australia, identifying regions where model biases are common and statistically significant; 2) to assess the relative strengths and weaknesses of individual RCMs; and 3) to assess the possible reasons for deficiencies in model performance. Model evaluation focuses on the entire CORDEX-Australasia ensemble which consists of four configurations of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008), the COSMO-CLM (CCLM) model (Rockel et al. 2008), and the Conformal-Cubic Atmospheric Model (CCAM; McGregor and Dix 2008). We evaluate the ability of this RCM ensemble to simulate near-surface maximum and minimum air temperature and precipitation at annual, seasonal, and daily time scales over Australia. These variables were chosen because they are often used for impact studies and are well-represented in high-quality gridded observational data sets for the Australian continent (King et al. 2013).

2. Data and methods

2.1 Model configurations

The RCMs were driven by ERA-Interim boundary conditions with a spatial resolution of approximately 80 km for a 29-year period from January 1981 to January 2010. The WRF RCM configurations used the Advanced Research WRF (ARW) solver which uses a fully compressible, Eulerian and non-hydrostatic equation set. It uses terrain-following, hydrostatic-pressure for the vertical coordinate, which has constant pressure surface at the top of the model. The horizontal grid uses Arakawa C-grid staggering. Its time integration scheme uses the third-order Runge-Kutta scheme, with a smaller time step for acoustic and gravity-wave modes. Further information on WRF can be found in Skamarock et al. (2008). All WRF configurations used a domain with quasi-regular grid spacing of approximately 50 km (0.44° x 0.44° on a rotated coordinate system) covering the CORDEX-Australasia region. Model performance was evaluated for Australia only (Fig. 1). The four configurations of the WRF RCM (UNSW-WRF360J, UNSW-WRF360K, UNSW-WRF360L, and MU-WRF330) used different parameterisations for planetary boundary layer physics, surface physics, cumulus physics, and radiation (Table 1). The UNSW-WRF360J, UNSW-WRF360K, and UNSW-WRF360L configurations were selected from a larger ensemble of WRF RCMs that accurately simulated the south-eastern Australian climate, whilst retaining as much independent information as possible (Evans et al. 2012; Evans et al. 2014; Ji et al. 2014). Parameterisations selected for MU-WRF330 were based on results from a prior sensitivity analysis of WRF to different physics and input data over southwest Western Australia (Kala et al. 2015). The MU-WRF330 simulation (Andrys et al. 2015) was conducted using WRF version 3.3, whereas the three other WRF simulations were conducted using version 3.6.0.

CCAM is a non-hydrostatic, variable-resolution global atmospheric model that includes a number of distinctive features. It uses two-time level, semi-implicit time differencing and semi-Lagrangian horizontal advection with bi-cubic horizontal interpolation. It also incorporates total-variation-diminishing (TVD) vertical advection (McGregor 1993) and reversible staggering (McGregor and Dix 2008). CCAM (version 1209) was run with a global uniform grid configuration at 50 km resolution and used the setup shown in Table 1. When forced with ERA-Interim data, the model setup was similar to the setups described in Katzfey et al. (2016) and Thevakaran et al. (2016), except that a scale-selective filter (i.e., spectral nudging, Thatcher and McGregor, 2009) with a scale of 9000 km was used every six hours for temperature, winds above approximately 900 hPa, and surface pressure. In addition, CCAM used ERA-Interim sea surface temperatures (SST) rather than the bias and variance corrected SSTs developed for CCAM by Hoffmann et al. (2016).

The COSMO model in CLimateMode ('CCLM') is a non-hydrostatic RCM developed from the Local Model (LM) of the German Weather Service. It solves the thermo-hydrodynamic equations for compressible flow in a moist atmosphere on an Arakawa-C grid which is defined on a rotated coordinate system. The vertical grid uses a hybrid coordinate that is terrain-following near the surface and flat near the top of the model. The standard land surface model (LSM) used by CCLM is TERRA-ML (Schrodin and Heise 2001). Further information on the dynamics and physical parametrisations in COSMO-CLM can be found in Doms and Baldauf (2015). For the present simulations, CCLM used a domain with quasi-regular grid spacing of approximately 50 km (0.44° x 0.44° on a rotated coordinate system) covering the CORDEX-Australasia region. Initial 'trial' simulations using the standard version of CCLM (CCLM4.8_clm17) were conducted using a number of different model configurations. These initial simulations showed large temperature overestimates over Australia in comparison to observed near-surface temperature from the CRU TS 3.10 data set (Harris et al. 2014). Subsequent simulations conducted using CCLM coupled to the community land model version 3.5 (CLM3.5, Dickinson et al. 2006) showed a substantial reduction in temperature overestimation. We therefore ran the simulations using the coupled model CCLM4.8_clm17-CLM3.5 (CCLM4-8-17-CLM3-5 in the CORDEX archive nomenclature). The model parameterisations used for CCLM are shown in Table 1.

The namelists used for all simulations evaluated by this study are provided in Online Resource 1. All RCM data were interpolated from the models' native grid to a common regular 0.5° grid for comparison and analysis using a nearest-neighbour algorithm.

2.2 Observations

Australian Gridded Climate Data (AGCD; Jones et al. 2009) were used to evaluate RCM performance. This daily gridded maximum and minimum temperature and precipitation data set has a spatial resolution of 0.05°, and is obtained from an interpolation of station observations across the Australian continent (Jones et al 2009). Observations include temperature minima and maxima only; hence, the ability of RCMs to reproduce mean temperature was not assessed. The majority of these stations are located in the more heavily populated coastal areas with a sparser representation inland, and there are more precipitation stations than temperature stations (refer to Figure 2 of Jones et al. 2009). Cross-validated root mean squared errors (RMSEs) for monthly maximum and minimum temperatures over Australia for 2001–2007 are typically between 0.5 to 1 °C, and 10 to 25 mm mo^{-1} for monthly precipitation (Jones et al. 2009). In order to compare models with slightly different spatial resolutions with gridded observations of a higher resolution, two different approaches can be adopted. One is that model output can be interpolated to match the higher resolution of the gridded observations such that the latter remain unchanged (see for example Vautard et al. 2013 and Zollo et al. 2016). However, in our case, the resolution of the observations is approximately 10 times higher than that of the models (5 by 5 km as compared to approximately 50 by 50 km). A major issue with using the native resolution of the observations as the common grid when evaluating lower resolution model output is that statistics with a strong dependence on the spatial scale (particularly extremes) will not be well evaluated. That is, a perfect model at 50 km would disagree with the observations at 5 km resolution, e.g. due to missing small-scale features. Moreover, interpolating the model output to the much higher resolution of the observational grid provides no additional information than the models' original 50 km grid. Of course, when interpolating the observations to a lower resolution the spatial scale mismatch has also to be taken into account. Here, this is handled by using a conservative re-gridding approach. The AGCD data were therefore re-gridded to correspond with the RCM data on a common 0.5° regular grid using the conservative area-weighted re-gridding scheme of the *Iris version 2.1* library (Met Office 2017) for the *Python version 3.6* programming language. Given AGCD observations are terrestrial data with no coverage over the ocean, only land points were evaluated.

2.3 Evaluation methods

We calculated annual and seasonal means for maximum and minimum temperature and precipitation using monthly averages for each variable. Mean diurnal ranges and 5th and 95th percentiles were calculated for maximum temperature using daily values. The performance of the RCMs in reproducing the observations over these timescales was assessed by calculating the model bias, defined as model outputs minus AGCD observations. The statistical significance of mean annual and seasonal biases compared to the AGCD observations was calculated for each grid cell using t-tests for maximum and minimum temperature ($\alpha = 0.05$) assuming equal variance. The Mann–Whitney U test was used for precipitation given its non-normality. Results on ensemble mean statistical significance were separated into three classes following Tebaldi et al. (2011). Specifically, statistically insignificant areas are shown in colour, denoting that fewer than half of the models are significantly biased. In these areas model bias is generally small; the most desired outcome. In areas of significant agreement (stippled), at least half of RCMs are significantly biased and at least 66% of the RCMs that show a significant difference agree on the direction of bias. In these regions, ensemble bias tends to be in one direction; an undesirable outcome. Areas of significant disagreement are shown in white, where at least half of the models are significantly biased and fewer than 66% of significant models agree on the bias direction. The 66% threshold was selected because it allowed for a single model to disagree with the consensus.

Model performance against observations was also assessed using the RMSE of simulated fields relative to observations. To evaluate the spatial agreement between RCM outputs and observations, we calculated the pattern correlation between simulated and observed fields (Walsh and McGregor 1997). The RMSE and pattern correlation were calculated for each RCM using the annual and seasonal means for each variable of interest.

We also examined the ability of the RCMs to simulate observed temperature and precipitation at daily time scales by comparing the probability density functions (PDFs) for AGCD daily mean observations versus those of the RCMs. PDFs were calculated for the whole study domain and for each natural resource management (NRM) climate region shown in Figure 1. For the PDFs only, all daily values of precipitation below 0.1 mm were omitted from the RCM output, as rates below this amount fall below the detection limit of the stations used to produce the AGCD data. Additionally, the daily rainfall observational network used to produce the AGCD has large gaps in several areas of central Australia; hence, RCM output was masked over these areas. Daily PDFs were compared by calculating the Perkins Skill Score (PSS; Perkins et al. 2007), which measures the common area between two PDFs whereby a PSS value of 1 indicates that the distributions overlap perfectly.

3. Results

3.1 Maximum temperature

All RCMs overestimate the frequency of lower than average temperatures, as shown by the PDFs of mean daily maximum temperatures across Australia, and underestimate the observed peaks (Fig. 2). The RCMs differ in their simulation of the frequency of warmer than average events, with the four configurations of the WRF RCM underestimating higher temperatures, whereas CCAM and CCLM overestimate occurrences of maximum temperatures higher than 312 K and 314 K, respectively. Overall, MU-WRF330 and CCLM show the best agreement with observations (see PSS scores in Table 2), while the performance of UNSW-WRF360L is comparatively poor. This is generally consistent for the seven NRM climate regions, although the magnitude of the error varies between regions (Fig. 1 and Online Resource 2: Figs. S1-S7).

Ensemble annual mean maximum temperature shows a statistically significant cold bias over most of Australia, which is most intense over the eastern regions (Fig. 3b). Mean bias shows few areas of significant disagreement (white) across Australia, with the majority occurring along portions of the northern and south-eastern coastlines. Additionally, the ensemble mean shows a significant warm bias along sections of the north-western coastline. In terms of individual RCMs, the statistically significant cold bias is the largest for UNSW-WRF360L, which exceeds -5 K over south-eastern Australia (Fig. 3e). UNSW-WRF360L is exceptional in this regard because other WRF configurations display a substantially smaller cold bias. CCAM shows a significant warm bias over a larger area as compared to the other RCMs, being 0.5 to 2.0 K warmer than observations in the semi-arid areas of central and northern Australia. Overall, CCLM has the lowest bias.

Cold biases are reflected in the spatial variation of RMSEs for simulated maximum surface temperatures (Online Resource 2: Fig. S8). For example, UNSW-WRF360L shows a large area of RMSEs > 5 K over south-eastern Australia, whilst RMSEs are lower for CCLM and

MU-WRF330 over the most of the continent. Mean pattern correlations and RMSEs are also consistent with these results, with CCLM having the lowest RMSE (0.97 K, versus the ensemble mean of 1.63 K; Table 3) and MU-WRF330 having the highest mean spatial agreement between observed and simulated fields.

At seasonal time-scales, the cold bias tends to be lower in intensity and spatial extent during summer (DJF, Fig. 4) relative to during winter (JJA, Fig. 5). This change is the most apparent for UNSW-WRF360L, which shows a large cold bias over south-eastern Australia on an annual time-scale that is greatly reduced during DJF (Fig. 4e). Areas of closer agreement between simulated and observed temperatures are also evident across several other regions during DJF, particularly for the WRF RCM configurations (Fig. 4c–f). In contrast, most RCMs display larger and more widespread statistically significant cold biases during the cooler months. This is most apparent during JJA (Fig. 5); however, CCLM and to a lesser extent MU-WRF330, do not follow this pattern. The poor annual performance of UNSW-WRF360L can be attributed to errors during MAM and JJA because RMSEs for the model are markedly higher as compared to other RCMs during these seasons (Table 3).

Figure 6 shows the biases of the 5^{th} and 95^{th} percentiles of daily maximum temperature. CCLM shows the closest agreement with observed 5^{th} percentile temperatures. Whereas the RCMs clearly differ in terms of their representation of annual and seasonal mean maximum temperatures, some similarities are apparent in their simulation of 95^{th} percentile maximum temperatures. Spatial patterns of 95^{th} percentile temperature bias are remarkably similar among the four WRF configurations (Fig. 6i–l), and CCAM and CCLM also share very similar patterns of bias (Fig. 6m–n). MU-WRF330 shows the lowest bias of all WRF RCMs in simulating the 95^{th} percentile across the heavily populated south-eastern coastline. Performance improves slightly for the WRF RCM configurations when simulating 95^{th} percentile maximum temperatures relative to annual mean maximum temperatures (i.e. mean RMSEs are 1.32 K and 1.85 K respectively; Tables 3–4).

3.2 Minimum temperature

Daily minimum temperature PDFs for UNSW-WRF360J and WRF360K match observations more closely as compared to the other simulations (Fig. 7) and produce the highest PSS scores (both scoring 0.98; Table 2). As compared to maximum temperatures, these two RCMs show a reduced tendency to over (under) estimate the occurrence of temperatures at the lower (upper) ends of the distribution. MU-WRF330, CCAM, and CCLM underestimate the frequency of colder than average events and overestimate the occurrence of warmer than average temperatures. Results over specific regions can differ substantially as compared to those over the whole of Australia (Online Resource 2: Figs. S11–17). For example, in contrast to the Australia-wide distribution, both UNSW-WRF360J and WRF360K show larger overestimates of the observed peak over the East Coast region as compared to the other RCMs.

The ensemble annual mean minimum temperature shows a statistically significant warm bias for several central and eastern regions (Fig. 8b). In contrast to the simulation of maximum temperature, all RCMs display significant warm bias over larger areas of the topographically complex eastern coastline. However, there were some prominent areas of significant disagreement over sections of western and northern Australia (Fig. 8b). This can be attributed to MU-WRF330, CCAM, and CCLM having significant warm biases across most of Australia (Fig. 8f–h), while UNSW-WRF360J-K-L show significant cold biases over Western Australia, and several northern and eastern regions (Fig. 8c–e). Notably, UNSW-WRF360J and WRF360K show closer agreement with observed minimum temperatures as compared to the other RCMs, with biases typically in the range of ± 1.5 K (Fig. 8c–d), and their performance is considerably improved relative to maximum temperatures. These two RCMs have the lowest mean RMSEs and low RMSEs across the domain (Table 3 and Fig. S18).

Seasonally, the spatial variation of the signs and magnitudes of the biases for each RCM are fairly similar to their corresponding performance at the annual time-scale (Figs. S19–22). We note that while UNSW-WRF360J and UNSW-WRF360K are fairly consistent across seasons in terms of mean RMSEs (Table 3), RMSE magnitudes are much higher during MAM and JJA for the remaining models and in most cases start increasing in March (Online Resource 2 Fig. S23). Similar to maximum temperatures, the poor annual performance of UNSW-WRF360L can be attributed to difficulties in simulating temperatures during MAM and JJA (Table 3).

3.3 Diurnal temperature range

All RCMs show relatively poor skill in simulating the observed distribution of mean diurnal ranges (Fig. 9). Models overestimate the frequency of smaller temperature ranges and underestimate the observed peak and occurrence of larger diurnal ranges. UNSW-WRF360L and MU-WRF330 perform marginally better than the other RCMs, whereas CCLM has the poorest performance (Table 2).

The ensemble mean diurnal range bias shows widespread areas of significant agreement (Fig. 10b); however, simulated ranges are generally smaller as compared to observed ranges (Fig. 10c-h). The magnitude of this negative bias is the largest over eastern Australia; however, bias decreases in a westerly direction and in some cases its sign is reversed. The ensemble bias shows the largest disagreement over southwest Western Australia. Similar to seasonal maximum and minimum temperatures, most RCMs tend to simulate diurnal ranges more accurately during DJF–SON as compared to during MAM–JJA (Figs. S24–27).

3.4 Precipitation

The PDFs for mean daily precipitation show that UNSW-WRF360J and MU-WRF330 simulate the occurrence of light rainfall events up to 0.5 mm day^{-1} fairly accurately (Fig. 11). UNSW-WRF360J, MU-WRF330, and CCLM simulate the frequency of precipitation events of ≥ 3 mm day^{-1} more accurately than the other models. However, the PSS for these models are only marginally higher as compared to the other RCMs with the exception of UNSW-WRF360K (Table 2). There are some interesting differences in RCM performance between regions (Figs. S28–34). For example, light rainfall events (up to 0.5 mm day^{-1}) are overestimated by several RCMs over the East Coast, while they are simulated more accurately over the Murray Darling Basin, which is adjacent to the East Coast and further inland.

The ensemble bias for annual mean precipitation shows significant agreement across the eastern, southern, western, and central regions of Australia (Fig. 12b), with areas of significant disagreement occurring mainly over northern Australia and a narrow strip along the eastern coastline. With the exception of MU-WRF330, RCMs show wet biases across large areas of the eastern, central, and southern regions. Some dry biases are also apparent; for example, UNSW-WRF360K, CCAM, and CCLM underestimate rainfall over the monsoonal north, whereas the remaining RCMs display a wet bias in this region. RMSEs are also comparatively high along the northern coastline for all RCMs (Fig. S35). MU-WRF330 displays a wet bias along the eastern coastline, and a dry bias over the lowlands to the west of the Great Dividing Range (Fig. 1) and across the southern half of Australia. Furthermore, MU-WRF330 overestimates rainfall over much of the northern half of Australia and as such, the spatial variation of its bias is an approximate mirror-image to that of CCAM. CCLM has the lowest annual mean RMSE of 15.58 mm mo^{-1} as compared to the ensemble mean of 20.62 mm mo^{-1} (Table 3).

Seasonally, many RCMs remain significantly wet-biased over much of eastern Australia, albeit with some regional variations in the sign of the bias. For example, several RCMs show a dry bias over northern regions during DJF, which subsequently switches to a wet bias during MAM, JJA, and SON (Figs. S36–39). The majority of RCMs are better able to capture the spatial pattern of precipitation during DJF, as compared to other seasons or annually, as evidenced by the mean pattern correlations (Table 3). Conversely, when RMSEs are considered, RCMs are most inaccurate during DJF, while accuracy is highest during JJA (Table 3). The strong seasonality of RCM skill is summarised by the RMSE annual cycles in Fig. S40.

4. Discussion

In summary, RCMs were generally cold-biased for maximum temperature, warm-biased for minimum temperature, and overestimated precipitation. However, model performance varied considerably between seasons and the different RCMs and RCM configurations. The following sections discuss potential mechanisms for these differences.

4.1 WRF

Cold biases were more widespread and typically larger for the four WRF configurations as compared to CCAM and CCLM. The unified Noah LSM used by all the WRF configurations is a potential source of this bias. Previous studies have demonstrated that use of this LSM can result in cold biases over European snow-covered regions during winter and overestimations of soil moisture and evaporation during summer (Garcia-Diez et al. 2015). While snow occupies a small proportion of the land surface in south-eastern Australia during cooler months, an excess of soil moisture is a potential explanation for the simulated cold bias. To investigate this hypothesis, the temporal correlation of the 29-year time series between monthly biases in precipitation and monthly biases in maximum temperature was calculated (Fig. 13). A strong negative correlation between mean monthly precipitation biases and mean monthly maximum temperature biases was apparent over most of Australia. Pearson's *r* values averaged across Australia for the four WRF configurations ranged from -0.44 to -

0.18. These associations also displayed strong seasonal variability; negative correlations between biases were larger and more widespread during DJF as compared to during JJA (e.g. for UNSW-WRF360J mean r = -0.60 versus r = -0.18, respectively; see Online Resource 2: Figs. S41–S42). These findings support the hypothesis that precipitation overestimation is a likely cause of the large maximum temperature cold bias in the WRF simulations. This is consistent with previous studies which have identified Australia as a soil moisture-atmosphere coupling "hot spot" for maximum temperature (Hirsch et al. 2014). Importantly, this negative correlation was reversed for biases in minimum temperature and precipitation (Fig. S43). Moreover, the more accurate simulation of 95th percentile maximum temperatures than annual mean maximum temperatures by the WRF RCM configurations may also be linked to this precipitation bias. Hot extremes in Australia often occur during dry conditions and are hence less affected by the mean precipitation overestimate. Future studies will investigate the drivers of the maximum temperature cold bias using soil moisture observations. Furthermore, since soil moisture is influenced by the LSM, it would also be informative to trial several LSMs with WRF with the aim of improving the representation of land surface processes, and subsequently, the simulation of near-surface temperatures.

The cold bias was more intense for UNSW-WRF360L as compared to other WRF configurations. UNSW-WRF360L was the only configuration to use CAM3 radiation schemes, suggesting that the strong cold bias can be partially attributed to the radiative scheme. This is supported by Katragkou et al. (2015) who also found that using CAM3 resulted in large cold biases.

The WRF configurations showed significant warm biases along portions of the north-western coastline, which were consistent with dry biases over this region. The spatial patterns of 95th percentile maximum temperature bias were also remarkably similar over this region for the four WRF RCM configurations. This consistent north-western bias must be viewed in the context of the relative sparseness of meteorological stations in this region, and the fact that many stations are located near the coastline where temperatures are lower than further inland. These issues increase the uncertainty of the AGCD observations relative to areas with denser station coverage. The strong relationship between station density and AGCD errors over the north-west and the western interior was noted by Jones et al. (2009), with these regions showing much larger cross-validated RMSEs than elsewhere (see their Figures 2 and 5). Given that other physical settings varied between the different WRF RCMs, it is difficult to identify a specific physical parameterisation that underlies this bias. However, it could also be partially inherited from the ERA-Interim lateral boundary conditions (Moalafhi et al. 2016).

UNSW-WRF360J and WRF360K both showed close agreement with regards to observed minimum temperatures with fairly small biases. This may partially stem from their use of the Mellor-Yamada-Janjic local PBL scheme, which was found to contribute to an accurate simulation of minimum temperature over Southern Spain (Argueso et al. 2011). These two RCM configurations differed only in terms of the cumulus scheme used (UNSW-WRF360J - Kain-Fritsch; UNSW-WRF360K - Betts-Miller-Janjic). Previous sensitivity studies for eastern Australia found that in WRF, these cumulus schemes do not have a large influence on minimum temperature (Evans et al. 2012).

In terms of precipitation biases, similarities between the WRF configurations included dry biases over parts of Western Australia and wet biases over the topographically complex terrain of south-eastern Australia. This south-eastern wet bias changed to a dry bias during winter, which coincides with a substantial improvement in model performance. Rainfall over south-eastern Australia is typically more frequent during the cooler months due to cold fronts moving across southern Australia. These wet biases may be partially inherited from the ERA-Interim lateral boundary conditions, which has a positive precipitation bias over eastern Australia as compared to the Global Precipitation Climatology Centre version 7 observed precipitation (Tuinenburg and de Vries 2017). Most of the model wet biases observed in the present evaluation were largest over eastern Australia. However, despite the fact that the RCMs assessed were driven by ERA-Interim, in many respects they showed quite different patterns of precipitation biases, suggesting that other factors also contributed to this bias. For example, precipitation biases demonstrated by ERA-Interim-forced WRF models over Germany were linked to the models' cumulus scheme not being tuned to European conditions (Warrach-Sagi et al. 2013). While Australia and Germany are very different regions, the cumulus scheme employed by Warrach-Sagi et al. (2013; Kain Fritsch) was used in three of the WRF configurations in the present study. As was the case in Germany, this cumulus scheme was not tuned for Australian conditions. Future work should assess whether using a higher resolution, such as the 20 km resolution selected for CORDEX2, together with more recent cumulus physics schemes, such as Grell-Freitas (Grell and Freitas 2014) and multiscale Kain-Fritsch (Zheng et al. 2016), will yield precipitation simulations over Australia that are more accurate than the current results.

4.2 CCLM

CCLM simulations have been performed over several CORDEX domains (e.g. Africa – Panitz et al. 2014, the Middle East North Africa - Bucchignani et al. 2016, and Europe - Kotlarski et al. 2014). Given that CCLM is based on the COSMO weather forecast model, it has been developed to provide good results for the European domain. For other CORDEX domains, the optimal setup differs from that of the European domain, and also between the various domains. A comparison of results between regions should therefore be performed with caution. The CCLM setup for CORDEX Australasia was based on CORDEX Africa simulations with two major differences. Firstly, the Bechtold et al. (2008) convection scheme was used instead of the Tiedtke (1989) scheme. The former was chosen due to the findings of Lange et al. (2015) who compared both schemes over South America and found that the Bechtold scheme resulted in an improved representation of precipitation. Tests during the setup phase of the present CCLM simulation confirmed that these findings also applied to Australia. Secondly, as described above in section 2.1 Model configurations, the standard LSM, TERRA-ML (Schrodin and Heise 2001), was replaced by CLM3.5 (Dickinson et al. 2006) in order to obtain a better representation of land surface processes.

Although generally cold biased, CCLM resulted in the most accurate representation of maximum temperatures in terms of mean annual and seasonal RMSEs. CCLM showed a maximum temperature bias that was also low, i.e. ±2 K across most of Australia. The reasonable results for annual and seasonal mean maximum temperature are partially due to the change of the LSM as described above, which is consistent with previous results for CCLM simulations (e.g. Panitz et al. 2014). Furthermore, we compared the surface solar radiation intensity simulated by CCLM with Surface Radiation Budget (SRB) data (SRB Science Team 2012). This revealed that CCLM simulated lower global radiation (i.e. direct + diffuse solar radiation) and lower net radiation as compared to the SRB data values, a tendency that would lead to lower simulated maximum surface temperatures. However, attribution of the radiation bias shown by CCLM to an overestimation of cloud cover and/or aerosols has not been established. This is because a comparison of observed and modelled cloud cover is not straightforward and requires a tool such as the International Satellite Cloud Climatology

Project (ISCCP) data simulator. Hence, an analysis of cloud cover using satellite measurements of this type merits future investigation. Furthermore, Zubler et al. (2011) and Kothe et al. (2014) found major deficiencies (over Europe and Africa, respectively) when using the aerosol climatology of Tanré et al. (1984) which is the default aerosol climatology used in CCLM. However, both of these studies changed the CCLM program code to accommodate alternative aerosol climatologies to that of Tanré et al., and therefore used unofficial CCLM versions. The Tanré aerosol climatology is the only aerosol scheme implemented in the official released CCLM version 4.18_clm17 used in the CORDEX-Australasia simulations. Therefore, it is not currently possible to conduct sensitivity tests to assess the relationships between different aerosol climatologies and uncertainties in the radiation components. However, in the most recent official version of CCLM (version 5.0), an alternative aerosol climatology can be selected via a namelist setting. An analysis of the influence of aerosol climatology on radiation bias over Australia will therefore be possible for future simulations.

CCLM overestimated the occurrence of warmer than average mean daily minimum temperatures, and overestimated annual mean minimum temperatures by approximately 3 to 4 K over most of Australia. A comparison of the simulated terrestrial radiation budget to SRB data (SRB Science Team 2012) showed that CCLM overestimated nighttime downward fluxes and also net fluxes, both factors which would contribute to an overestimation of minimum surface temperatures. The combined underestimation of maximum temperatures together with an overestimation of minimum temperatures is one explanation for CCLM's estimates of small diurnal temperature ranges.

CCLM showed fairly close agreement with observed rainfall across the semi-arid inland regions of Australia, whereas it underestimated precipitation across northern Australia and along most of the coastline. This dry bias over coastal areas and tropical Northern regions is consistent with findings by Panitz et al. (2014). The precipitation intensity simulated by CCLM shows a steep gradient between the northern Australian peninsulas and the adjacent ocean areas (not shown). Panitz et al. (2014) stated that "CCLM seems unable to fully transport inland the moisture from the ocean". This may not only affect the water vapor transport, but also the transport of cloud and precipitable water. More recently, Li et al. (2018) observed that precipitation biases shown by CCLM over the CORDEX-East Asian domain were closely linked to biases of water vapor transport. Although the model versions and domains of these studies are different to those of our study, inaccuracy in simulating water vapor transport processes is a possible reason for the precipitation biases observed over some Australian regions. Further investigation is required to understand the causes of the precipitation biases shown by CCLM over Australia, and in particular to test whether they are related to biases in water vapor transport.

4.3 CCAM

In contrast to the other models, the CCAM simulation was conducted on a global even/uniform grid and spectrally nudged towards the ERA-Interim data using a scale-selective filter. Hence, the parameterisations were selected to perform well globally and not for a particular region or resolution. In addition, the filter settings used to force the ERA-Interim data were not restrictive (i.e. mainly forcing features with scales larger than 9000 km). Furthermore, CCAM was not constrained by lateral boundary data.

CCAM overestimated occurrences of maximum temperatures at both the lower and upper ends of the observed distribution and was similar to CCLM in this regard. CCAM overestimated maximum temperatures across large regions of northern and central Australia at an annual timescale and during most seasons. Conversely, it was generally cold-biased over the southern half of the country, particularly over the temperate regions of south-western and eastern Australia. Similar to the WRF results, the regions of maximum temperature bias correspond strongly with those of precipitation bias, which suggests that maximum temperature underestimation is related to excessive soil moisture and evaporation and vice versa.

CCAM simulated minimum temperatures more accurately than maximum temperatures. In their evaluation of the current climate of Vietnam, Katzfey et al. (2016) found that CCAM simulated maximum temperatures less accurately than minimum temperatures, which is consistent with our findings. Notably, these results are consistent across very different domains. Although more detailed analysis is required, the CABLE LSM used by CCAM may have some inaccuracies related to the simulation of prescribed soil surface albedo and parameterised vegetation albedo (Wang et al. 2011), issues which would primarily affect the simulation of maximum temperatures.

CCAM's diurnal temperature range PDF, like the observed PDF, has only one major peak, though this peak is shifted slightly towards the lower values. In contrast, the PDFs of the other models show bimodal peaks. The seasonal biases in diurnal temperature are also smaller than those of the other models, except possibly during JJA. Consequently, the CCAM results show a general temperature offset, but a fairly accurate simulation of the diurnal cycle, which could be informative for impact modelling and assessment studies in fields such as agriculture (e.g. Lobell 2007) and human health (e.g. Lambrechts et al. 2011).

CCAM was generally dry-biased over northern regions and wet-biased over the southern half of Australia. However, this northern dry bias was only associated with the wetter seasons (DJF and MAM) because it was reduced during JJA and switched to a wet bias during SON. The CCAM version used by the present study (version 1209) also underestimated precipitation during the Vietnamese wet season (summer) and overestimated precipitation during the dry season (winter) (Katzfey et al. 2016). Similar to the results reflected in the daily precipitation PDFs of the present study, CCAM also accurately simulated daily observed light rainfall events over Vietnam for a threshold rate of 1 mm day^{-1} (Nguyen et al. 2014). Initial experiments that tested different convection scheme settings showed that simulated rainfall over tropical regions was sensitive to the profiles and rates of entrainment and detrainment, which are configured by various settings in the *kuonml* namelist options (see Online Resource 1). As described below, experiments that have used updated convection scheme settings have substantially improved the simulation of rainfall as compared to the results noted here.

The CCAM code evaluated by the present study used a new prognostic aerosol scheme which overestimated the concentration of SO_2. This overestimation of SO_2 concentrations would affect CCAM's cloud microphysics (indirect effects), shortwave radiation (direct effects) and rainfall (via the number of condensation nuclei). Subsequent refinements to the CCAM code (version 3355) have alleviated the SO_2 overestimation issue. Furthermore, additional refinements have been made to the convective parameterisation and explicit cumulus scheme, as well as to the CABLE LSM. More recent simulations that incorporate these refinements show substantial improvements in the simulation of maximum and minimum temperatures and precipitation over Australia (i.e. the magnitudes of biases are substantially reduced). These model refinements and new results will be discussed in a future paper.

5. Conclusions

This study evaluated the ability of six reanalysis-driven RCMs/RCM configurations within the CORDEX Australasia framework to simulate maximum and minimum temperature and precipitation over Australia at daily, seasonal, and annual time scales. In doing so, we address an important knowledge gap because no such RCM evaluations currently exist for Australia. RCMs were generally cold-biased when simulating maximum temperatures over Australia, behaviour that was particularly characteristic of the WRF RCM configurations. Negative correlations were observed between mean monthly biases in precipitation and maximum temperature which supports the general conclusion that RCM cold bias is associated with precipitation overestimation. The configurations of CCAM and CCLM were quite different to those of the WRF models. Taking this into account, CCAM and CCLM performed quite well and offer useful complements to the WRF configurations assessed. Future refinements to model configurations in the CORDEX Australasia ensemble that reduce overestimation of precipitation, and subsequently soil moisture and evaporation, would improve model performance for this region. Since soil moisture is influenced by the LSM, it would also be beneficial to test different LSMs with the aim of improving the representation of land surface processes, and subsequently of surface temperatures. Overall, the CORDEX Australasia ensemble is valuable for use in further studies. The RCM configurations assessed here are currently being used to perform future climate change projections for Australia, forced by GCM outputs from CMIP5. Our assessment of the abilities of these RCMs/RCM configurations to simulate Australian temperature and precipitation, particularly over heavily populated regions, can thus help inform decision-making by the adaptation community. Furthermore, the varying model capabilities reported here can also help guide experiment design and model configuration for climate change impact studies over Australia.

Author contributions

JE, AD, RO and DA designed and ran the UNSW WRF experiments. JK and JA ran the MU WRF experiments. PH and JJK ran the CCAM experiment. GD and JE conceived the research aims. GD designed and performed the analyses. GD prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We thank the NCAR Mesoscale and Microscale Meteorology Division for developing and maintaining WRF. We thank Marcus Thatcher and John McGregor at CSIRO Oceans and Atmosphere for developing CCAM, for help with the post-processing software to produce the CORDEX output, and for helpful discussions regarding CCAM. Logistical support was provided by the Climate Change Research Centre at the University of New South Wales, by the National Computing Infrastructure National Facility at Australian National University and by the Pawsey Supercomputing Centre. This project is supported through funding from the Earth Systems and Climate Change Hub of the Australian Government's National Environmental Science Programme and the NSW government Office of Environment and Heritage. JK is supported by an Australian Research Council (ARC) Discovery Early Career Researcher Grant (DE170100102). AD is also supported by ARC grant (DE170101191). RO was supported by the Basic Science Research Program through National Research Foundation of Korea (NRF-2017K1A3A7A03087790), and through the Institute for Basic Science (project code IBS-R028-D1). DA received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.743547. We thank two anonymous reviewers for their constructive feedback on this manuscript.

References

Andrys J, Lyons TJ, Kala J (2015) Multidecadal Evaluation of WRF Downscaling Capabilities over Western Australia in Simulating Rainfall and Temperature Extremes J Appl Meteorol Climatol 54:370-394 doi:10.1175/jamc-d-14-0212.1

Argueso D, Hidalgo-Munoz JM, Gamiz-Fortis SR, Esteban-Parra MJ, Dudhia J, Castro-Diez Y (2011) Evaluation of WRF Parameterizations for Climate Studies over Southern Spain Using a Multistep Regionalization J Clim 24:5633-5651 doi:10.1175/jcli-d-11-00073.1

Bechtold P et al. (2008) Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales Quarterly Journal of the Royal Meteorological Society 134:1337-1351 doi:10.1002/qj.289

Bucchignani E, Mercogliano P, Rianna G, Panitz HJ (2016) Analysis of ERA-Interim-driven COSMO-CLM simulations over Middle East - North Africa domain at different spatial resolutions International Journal of Climatology 36:3346-3369 doi:10.1002/joc.4559

Dee DP et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system Quarterly Journal of the Royal Meteorological Society 137:553-597 doi:10.1002/qj.828

Di Luca A, Argueso D, Evans JP, de Elia R, Laprise R (2016) Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales Journal of Geophysical Research-Atmospheres 121:1575-1590 doi:10.1002/2015jd024009

Di Luca A, de Elia R, Laprise R (2012) Potential for added value in precipitation simulated by high-resolution nested Regional Climate Models and observations Clim Dyn 38:1229-1247 doi:10.1007/s00382-011-1068-3

Diaconescu EP, Gachon P, Scinocca J, Laprise R (2015) Evaluation of daily precipitation statistics and monsoon onset/retreat over western Sahel in multiple data sets Clim Dyn 45:1325-1354 doi:10.1007/s00382-014-2383-2

Dickinson RE et al. (2006) The Community Land Model and its climate statistics as a component of the Community Climate System Model J Clim 19:2302-2324 doi:10.1175/jcli3742.1

Doms G, Baldauf M (2015) A Description of the Nonhydrostatic Regional COSMO-Model Part I: Dynamics and Numerics. DWD, Offenbach, Germany, pp 164

Evans JP, Ekström M, Ji F (2012) Evaluating the performance of a WRF physics ensemble over South-East Australia Clim Dyn 39:1241-1258 doi:10.1007/s00382-011-1244-5

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 doi:10.5194/gmd-7-621-2014

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling International Journal of Climatology 27:1547-1578 doi:10.1002/joc.1556

Freidenreich SM, Ramaswamy V (1999) A new multiple-band solar radiative parameterization for general circulation models Journal of Geophysical Research: Atmospheres 104:31389-31409 doi:10.1029/1999JD900456

Garcia-Diez M, Fernandez J, Vautard R (2015) An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation Clim Dyn 45:3141-3156 doi:10.1007/s00382-015-2529-x

Giorgi F (2006) Regional climate modeling: Status and perspectives J Phys IV 139:101-118 doi:10.1051/jp4:2006139008

Giorgi F, Bates GT (1989) The Climatological Skill of a Regional Model over Complex Terrain Monthly Weather Review 117:2325-2347 doi:10.1175/1520-0493(1989)117<2325:tcsoar>2.0.co;2

Giorgi F, Jones C, Asrar G (2009) Addressing climate information needs at the regional level: The CORDEX framework WMO Bulletin 53:175–183

Grell GA, Freitas SR (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling Atmos Chem Phys 14:5233-5250 doi:10.5194/acp-14-5233-2014

Halmstad A, Najafi MR, Moradkhani H (2013) Analysis of precipitation extremes with the assessment of regional climate models over the Willamette River Basin, USA Hydrol Process 27:2579-2590 doi:10.1002/hyp.9376

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset International Journal of Climatology 34:623-642 doi:10.1002/joc.3711

Hattermann FF, Weiland M, Huang SC, Krysanova V, Kundzewicz ZW (2011) Model-Supported Impact Assessment for the Water Sector in Central Germany Under Climate Change-A Case Study Water Resour Manag 25:3113-3134 doi:10.1007/s11269-011-9848-4

Hirsch AL, Pitman AJ, Seneviratne SI, Evans JP, Haverd V (2014) Summertime maximum and minimum temperature coupling asymmetry over Australia determined using WRF Geophys Res Lett 41:1546-1552 doi:doi:10.1002/2013GL059055

Hoffmann P, Katzfey JJ, McGregor JL, Thatcher M (2016) Bias and variance correction of sea surface temperatures used for dynamical downscaling Journal of Geophysical Research-Atmospheres 121:12877-12890 doi:10.1002/2016jd025383

IPCC (2012) Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PM (eds). Cambridge, UK and New York, NY

IPCC (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. doi:10.1017/CBO9781107415324

Ji F, Ekström M, Evans JP, Teng J (2014) Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model Theoretical and Applied Climatology 115:297-304 doi:10.1007/s00704-013-0904-2

Jones DA, Wang W, Fawcett R (2009) High-quality spatial climate data-sets for Australia Aust Meteorol Oceanogr J 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 doi:10.1007/s00382-014-2160-2

Katragkou E et al. (2015) Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble Geoscientific Model Development 8:603-618 doi:10.5194/gmd-8-603-2015

Katzfey J et al. (2016) High-resolution simulations for Vietnam - methodology and evaluation of current climate Asia-Pac J Atmos Sci 52:91-106 doi:10.1007/s13143-016-0011-2

King AD, Alexander LV, Donat MG (2013) The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia International Journal of Climatology 33:2376-2387 doi:10.1002/joc.3588

Kothe S, Panitz HJ, Ahrens B (2014) Analysis of the radiation budget in regional climate simulations with COSMO-CLM for Africa Meteorol Z 23:123-141 doi:10.1127/0941-2948/2014/0527

Kotlarski S et al. (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble Geoscientific Model Development 7:1297-1333 doi:10.5194/gmd-7-1297-2014

Kowalczyk E, Wang Y, M Law R, L Davies H, L Mcgregor J, Abramowitz G (2006) The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model vol 1615.

Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, Thomas MB, Scott TW (2011) Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti Proc Natl Acad Sci U S A 108:7460-7465 doi:10.1073/pnas.1101377108

Lange S, Rockel B, Volkholz J, Bookhagen B (2015) Regional climate model sensitivities to parametrizations of convection and non-precipitating subgrid-scale clouds over South America Clim Dyn 44:2839-2857 doi:10.1007/s00382-014-2199-0

Laprise R (2008) Regional climate modelling J Comput Phys 227:3641-3666 doi:10.1016/j.jcp.2006.10.024

Li DL et al. (2018) Present Climate Evaluation and Added Value Analysis of Dynamically Downscaled Simulations of CORDEX-East Asia J Appl Meteorol Climatol 57:2317-2341 doi:10.1175/jamc-d-18-0008.1

Lobell DB (2007) Changes in diurnal temperature range and national cereal yields Agricultural and Forest Meteorology 145:229-238

Maraun D et al. (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user Rev Geophys 48:34 doi:10.1029/2009rg000314

McGregor JL (1993) The CSIRO 9-level atmospheric general circulation model. Melbourne, CSIRO Australia

McGregor JL (2003) A new convection scheme using a simple closure. BMRC research report 93. Melbourne, Australia

McGregor JL, Dix MR (2008) An updated description of the Conformal-Cubic atmospheric model. High Resolution Numerical Modelling of the Atmosphere and Ocean. Springer, New York. doi:10.1007/978-0-387-49791-4_4

Met Office (2018) Iris: A Python library for analysing and visualising meteorological and oceanographic data sets version 2.1. Exeter, Devon, UK

Moalafhi DB, Evans JP, Sharma A (2016) Evaluating global reanalysis datasets for provision of boundary conditions in regional climate modelling Clim Dyn 47:2727-2745 doi:10.1007/s00382-016-2994-x

Nguyen KC, Katzfey JJ, McGregor JL (2014) Downscaling over Vietnam using the stretched-grid CCAM: verification of the mean and interannual variability of rainfall Clim Dyn 43:861-879 doi:10.1007/s00382-013-1976-5

Nikulin G et al. (2012) Precipitation Climatology in an Ensemble of CORDEX-Africa Regional Climate Simulations J Clim 25:6057-6078 doi:10.1175/jcli-d-11-00375.1

Olsson J, Berg P, Kawamura A (2015) Impact of RCM Spatial Resolution on the Reproduction of Local, Subdaily Precipitation J Hydrometeorol 16:534-547 doi:10.1175/jhm-d-14-0007.1

Panitz H-J, Dosio A, Büchner M, Lüthi D, Keuler K (2014) COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44° and 0.22° resolution Clim Dyn 42:3015-3038 doi:10.1007/s00382-013-1834-5

Perkins SE, Pitman AJ, Holbrook NJ, 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 doi:10.1175/jcli4253.1

Raschendorfer M (2001) The new turbulence parameterization of LM. COSMO-Newsletter, No. 1, Feb 2001, 89-97. vol 1.

Ritter B, Geleyn J-F (1992) A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations Monthly Weather Review 120:303-325

Rockel B, Will A, Hense A (2008) The Regional Climate Model COSMO-CLM(CCLM) Meteorol Z 17:347-348 doi:10.1127/0941-2948/2008/0309

Rotstayn LD (1997) A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes Quarterly Journal of the Royal Meteorological Society 123:1227-1282 doi:10.1002/qj.49712354106

Rummukainen M (2016) Added value in regional climate modeling Wiley Interdiscip Rev-Clim Chang 7:145-159 doi:10.1002/wcc.378

Schrodin E, Heise E (2001) The Multi-Layer Version of the DWD Soil Model TERRA LM. COSMO Technical Report No.2, pp 16, Sep 2001, DWD, Offenbach, Germany. .

Seifert A, Beheng KD (2001) A double-moment parameterization for simulating autoconversion, accretion and selfcollection Atmospheric Research 59:265-281 doi:10.1016/s0169-8095(01)00126-0

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2008) A description of the Advanced Research WRF Version 3. NCAR Tech Note NCAR/TN-475+STR. NCAR. Boulder, CO

Solman SA et al. (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 doi:10.1007/s00382-013-1667-2

Sunyer MA, Luchner J, Onof C, Madsen H, Arnbjerg-Nielsen K (2017) Assessing the importance of spatio-temporal RCM resolution when estimating sub-daily extreme precipitation under current and future climate conditions International Journal of Climatology 37:688-705 doi:doi:10.1002/joc.4733

Tanré D, Geleyn J-F, Slingo J (1984) First results of the introduction of an advanced aerosol-radiation interaction in the ECMWF low resolution global model. In: Gerber HE, Deepak A (eds) Aerosols and Their Climatic Effects. Hampton, Va, p 133

SRB Science Team (2012) SRB Data. Hampton, VA, USA. doi:10.5067/SRB/REL3.1_LW_3HRLY_NC_L2

Tebaldi C, Arblaster JM, Knutti R (2011) Mapping model agreement on future climate projections Geophys Res Lett 38 doi:doi:10.1029/2011GL049863

Thatcher M, McGregor JL (2009) Using a Scale-Selective Filter for Dynamical Downscaling with the Conformal Cubic Atmospheric Model Monthly Weather Review 137:1742-1752 doi:10.1175/2008mwr2599.1

Thevakaran A, McGregor JL, Katzfey J, Hoffmann P, Suppiah R, Sonnadara DUJ (2016) An assessment of CSIRO Conformal Cubic Atmospheric Model simulations over Sri Lanka Clim Dyn 46:1861-1875 doi:10.1007/s00382-015-2680-4

Tiedtke M (1989) A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models Monthly Weather Review 117:1779-1800 doi:10.1175/1520-0493(1989)117<1779:acmfsf>2.0.co;2

Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps Journal of Geophysical Research: Atmospheres 120:3957-3972 doi:10.1002/2014JD022781

Tuinenburg OA, de Vries JPR (2017) Irrigation Patterns Resemble ERA-Interim Reanalysis Soil Moisture Additions Geophys Res Lett 44:10341-10348 doi:10.1002/2017gl074884

Vautard R et al. (2013) The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project Clim Dyn 41:2555-2575 doi:10.1007/s00382-013-1714-z

Walsh K, McGregor J (1997) An assessment of simulations of climate variability over Australia with a limited area model International Journal of Climatology 17:201-223 doi:10.1002/(sici)1097-0088(199702)17:2<201::aid-joc118>3.3.co;2-r

Wang YP et al. (2011) Diagnosing errors in a land surface model (CABLE) in the time and frequency domains J Geophys Res-Biogeosci 116:18 doi:10.1029/2010jg001385

Wang YQ, Leung LR, McGregor JL, Lee DK, Wang WC, Ding YH, Kimura F (2004) Regional climate modeling: Progress, challenges, and prospects J Meteorol Soc Jpn 82:1599-1628 doi:10.2151/jmsj.82.1599

Warrach-Sagi K, Schwitalla T, Wulfmeyer V, Bauer H-S (2013) Evaluation of a climate simulation in Europe based on the WRF–NOAH model system: precipitation in Germany Clim Dyn 41:755-774 doi:10.1007/s00382-013-1727-7

Xue YK, Janjic Z, Dudhia J, Vasic R, De Sales F (2014) A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability Atmospheric Research 147:68-85 doi:10.1016/j.atmosres.2014.05.001

Zheng Y, Alapaty K, Herwehe JA, Genio ADD, Niyogi D (2016) Improving High-Resolution Weather Forecasts Using the Weather Research and Forecasting (WRF) Model with an Updated Kain–Fritsch Scheme Monthly Weather Review 144:833-860 doi:10.1175/mwr-d-15-0005.1

Zollo AL, Rillo V, Bucchignani E, Montesarchio M, Mercogliano P (2016) Extreme temperature and precipitation events over Italy: assessment of high-resolution simulations with COSMO-CLM and future scenarios International Journal of Climatology 36:987-1004 doi:doi:10.1002/joc.4401

Zubler EM, Folini D, Lohmann U, Luthi D, Schar C, Wild M (2011) Simulation of dimming and brightening in Europe from 1958 to 2001 using a regional climate model Journal of Geophysical Research-Atmospheres 116:13 doi:10.1029/2010jd015396

Tables

Table 1. List of CORDEX RCMs analysed and their configurations.

Model / Version	Responsible institution	Planetary boundary layer physics / surface layer physics	Cumulus physics	Microphysics	Shortwave and longwave radiation physics	Land surface	Vertical levels
UNSW-WRF360J	University of New South Wales (UNSW)	Mellor-Yamada-Janjic/ETA Similarity	Kain-Fritsch	WRF Double-Moment 5	Dudhia/RRTM	Noah LSM	30
UNSW-WRF360K		Mellor-Yamada-Janjic/ETA Similarity	Betts-Miller-Janjic	WRF Double-Moment 5	Dudhia/RRTM	Noah LSM	30
UNSW-WRF360L		Yonsei University/MM5 Similarity	Kain-Fritsch	WRF Double-Moment 5	CAM3/CAM3	Noah LSM	30
MU-WRF330	Murdoch University	Yonsei University/MM5 Similarity	Kain-Fritsch	WRF Single-Moment 5	Dudhia/RRTM	Noah LSM	30
CCAM	CSIRO	Monin-Obukhov Similarity Theory stability-dependent boundary-layer scheme (McGregor 1993)	Mass-flux closure (McGregor 2003)	Liquid and ice-water scheme (Rotstayn 1997)	GFDL (Freidenreich and Ramaswamy 1999)	CABLE (Kowalczyk et al. 2006)	27
CCLM4-8-17-CLM3-5	Climate Limited-area Modelling Community	Prognostic turbulent kinetic energy (Raschendorfer 2001)	Bechtold et al. (2008)	Seifert and Beheng (2001), reduced to one moment scheme	Ritter and Geleyn (1992)	CLM; (Dickinson et al. 2006)	35

Table 2. Perkins skill scores (PSS) for the six RCMs for daily minimum and maximum temperature, diurnal temperature, and daily precipitation. Bold values indicate the RCM with the highest PSS.

RCM	Temp. max.	Temp. min.	Diurnal range	Precipitation
UNSW-WRF360J	0.94	**0.98**	0.56	0.76
UNSW-WRF360K	0.94	0.98	0.57	0.69
UNSW-WRF360L	0.88	0.91	0.64	0.72
MU-WRF330	**0.95**	0.91	**0.68**	0.76
CCAM	0.90	0.94	0.62	0.76
CCLM	0.95	0.90	0.17	**0.78**

Table 3. Diagnostics for six RCMs for annual and seasonal mean minimum and maximum temperature and precipitation for the period January 1981 to January 2010 with Australian Gridded Climate Data as reference data. Bold values indicate the RCM with the best diagnostic score.

		Pearson's *r*							**RMSE**						
	Period	UNSW-WRF360J	UNSW-WRF360K	UNSW-WRF360L	MU-WRF330	CCAM	CCLM	Ensemble Mean	UNSW-WRF360J	UNSW-WRF360K	UNSW-WRF360L	MU-WRF330	CCAM	CCLM	Ensemble Mean
	Annual	0.895	0.899	0.869	**0.908**	0.904	0.903	0.90	1.73	1.55	2.85	1.28	1.37	**0.97**	1.63
	DJF	0.837	0.839	0.856	**0.858**	0.845	0.841	0.85	1.90	1.66	1.70	1.66	1.77	**1.28**	1.66
Temp. Max. (K)	MAM	0.894	0.898	0.858	0.904	0.897	**0.906**	0.89	2.10	1.95	3.36	2.02	1.86	**1.27**	2.09
	JJA	0.917	0.919	0.817	0.922	0.919	**0.925**	0.90	2.43	2.23	5.87	1.67	2.18	**1.32**	2.62
	SON	0.906	0.909	0.901	**0.915**	0.908	0.904	0.91	1.47	1.45	1.77	1.09	1.70	**1.04**	1.42
	Annual	**0.902**	0.897	0.896	0.900	0.899	0.889	0.90	**0.84**	0.87	1.57	1.83	1.25	2.33	1.45
	DJF	0.908	0.901	0.904	0.909	**0.912**	0.901	0.91	**1.09**	1.11	1.19	2.00	1.10	1.84	1.39
Temp. Min. (K)	MAM	**0.896**	0.891	0.876	0.894	0.888	0.876	0.89	**1.18**	1.21	2.02	1.79	1.56	2.62	1.73
	JJA	0.855	0.852	0.826	**0.856**	0.852	0.844	0.85	1.19	**1.14**	2.95	1.89	2.15	2.86	2.03
	SON	**0.915**	0.909	0.906	0.907	0.915	0.907	0.91	**1.03**	1.15	1.39	2.29	1.43	2.23	1.59
	Annual	0.730	0.630	**0.775**	0.766	0.712	0.681	0.72	28.00	20.31	18.63	21.64	19.59	**15.58**	20.62
	DJF	0.818	0.753	0.818	**0.836**	0.789	0.796	0.80	60.93	48.99	51.90	58.89	50.80	**37.06**	51.43
Prec. (mm mo^{-1})	MAM	0.630	0.547	**0.682**	0.660	0.611	0.471	0.60	41.65	35.68	35.19	40.10	36.36	**26.08**	35.84
	JJA	0.720	0.715	0.771	0.775	0.788	**0.794**	0.76	19.89	18.31	15.28	15.72	21.24	**11.40**	16.97
	SON	0.741	0.739	0.803	0.756	**0.803**	0.752	0.77	30.08	20.82	19.39	21.74	25.01	**13.02**	21.68

Table 4. Summary diagnostics for six RCMs when simulating extreme (5th and 95th percentile) maximum and minimum temperature for 1981 to 2010 using Australian Gridded Climate Data as reference data. Bold values indicate the RCM with the best diagnostic score.

		Pearson's r							RMSE						
	Percentile	UNSW-WRF360J	UNSW-WRF360K	UNSW-WRF360L	MU-WRF330	CCAM	CCLM	Ensemble Mean	UNSW-WRF360J	UNSW-WRF360K	UNSW-WRF360L	MU-WRF330	CCAM	CCLM	Ensemble Mean
Temp. Max. (K)	**5th**	0.93	0.93	0.80	0.93	0.94	**0.94**	0.91	2.42	2.21	7.87	1.66	2.24	**1.17**	2.93
	95th	0.87	0.88	**0.88**	0.87	0.80	0.79	0.85	1.63	1.35	1.26	**1.03**	1.66	1.38	1.38
Temp. Min. (K)	**5th**	0.88	0.88	0.84	**0.89**	0.88	0.87	0.87	**1.03**	1.07	2.85	2.18	1.72	3.14	2.00
	95th	0.90	0.90	0.89	0.90	**0.91**	0.89	0.90	**0.92**	0.95	1.04	2.54	1.08	2.19	1.45

Figure captions

Fig. 1 Topographic variation across the study domain, Australia. Approximate location of the Great Dividing Range is delineated in white. NT=Northern Territory; QLD=Queensland; NSW=New South Wales; ACT = Australian Capital Territory; TAS = Tasmania; VIC = Victoria; SA = South Australia; WA = Western Australia. Inset **a** shows natural resource management (NRM) climate regions (MDB = Murray Darling Basin). Inset **b** shows the CORDEX Australasia domain

Fig. 2 Probability density functions of mean daily maximum near-surface air temperatures (K) across Australia. Panels a-f show the PDF of a specific RCM/RCM configuration relative to that of Australian Gridded Climate Data (AGCD) observations

Fig. 3 Annual mean near-surface atmospheric maximum temperature bias with respect to Australian Gridded Climate Data (AGCD) observations for the RCMs. Stippled areas indicate locations where an RCM shows statistically significant bias ($P < 0.05$). **b** Significance stippling for the ensemble mean bias follows Tebaldi et al. (2011). Statistically insignificant areas are shown in colour, denoting that less than half of the models are significantly biased. In areas of significant agreement (stippled), at least half of RCMs are significantly biased, and at least 66% of the significant RCMs agree on the direction of the bias. Areas of significant disagreement are shown in white, which are where at least half of the models are significantly biased and less than 66% significant models agree on the bias direction - see main text for additional detail on the stippling regime

Fig. 4 Summer (DJF) maximum temperature bias with respect to AGCD observations with stippling as per Fig. 3

Fig. 5 Winter (JJA) maximum temperature bias with respect to AGCD observations with stippling as per Fig. 3

Fig. 6 Biases in 5^{th} percentile (panels a-g) and 95^{th} percentile (panels h-n) mean maximum temperatures simulated by the RCMs, relative to AGCD with stippling ($P < 0.05$)

Fig. 7 Probability density functions of mean daily minimum near-surface air temperatures across Australia

Fig. 8 Annual mean minimum temperature bias (K) with respect to AGCD observations for the RCMs with stippling as per Fig. 3

Fig. 9 Probability density functions of mean diurnal ranges across Australia

Fig. 10 Bias in the mean diurnal ranges simulated by RCMs relative to observed mean diurnal ranges

Fig. 11 Probability density functions of mean daily precipitation

Fig. 12 Annual mean precipitation bias of the RCMs with stippling as per Fig. 3

Fig. 13 a Temporal correlations between observed mean monthly maximum temperature (tasmax) and precipitation (pr), **b-c** Biases in modelled versus observed tasmax and pr, **d-i** Temporal correlations between mean monthly biases in maximum temperature and precipitation