Australian Meteorological and Oceanographic Journal 63 (2013) 205–212

The relative performance of Australian CMIP5 models based on rainfall and ENSO metrics

Ian Smith[1], Jozef Syktus[2], Leon Rotstayn[1] and Stephen Jeffrey[2]
[1]Centre for Australian Weather and Climate Research (CAWCR), a partnership between CSIRO and the Bureau of Meteorology, CSIRO Marine and Atmospheric Research, Australia
[2] Queensland Department of Science, Information Technology, Innovation and the Arts, Australia

(Manuscript received July 2012; revised December 2012)

We assess the performance of 30 CMIP5 and two CMIP3 models using metrics based on an all-Australia average rainfall and NINO3.4 sea surface temperatures (SSTs). The assessment provides an insight into the relative performance of the models at simulating long-term average monthly mean values, interannual variability and the seasonal cycles. It also includes a measure of the ability to capture observed rainfall–NINO3.4 SST correlations. In general, the rainfall features are reasonably simulated and there is relatively little difference amongst the models but the NINO3.4 SST features appear more difficult to simulate as evidenced by the greater range in metric scores. We find little evidence of consistency in the sense that a relatively good metric score for one feature does not imply a relatively good score for another related (but independent) feature. The assessment indicates that more recent models perform slightly better than their predecessors, especially with regard to the NINO3.4 metrics. We also focus on the ability of models to reproduce the observed seasonal cycle of rainfall–SST correlations since this is a direct indicator of a model's potential utility for seasonal forecasting over Australia. This indicates some relatively good models (CNRM, HadGEM2-ESM, MPI-ESM-LR and MPI-ESM-MR) and some relatively poor models (CSIRO-Mk3.5, FGOALS, GISS-E2-HP1 and INMCM4). We find that the ACCESS1.3 and CSIRO-Mk3.6 models rank as near median performers on this metric and represent improvements over their predecessors (ACCESS1.0, CSIRO-Mk3.0 and CSIRO-Mk3.5).

Introduction

Australia (mainly via the CSIRO and the Bureau of Meteorology) has had a long and successful track record in climate model development and contributions to international modelling endeavours (Smith 2007) and this has continued with contributions to the latest International Climate Model Intercomparison Project (CMIP5). Analyses of CMIP5 results will underpin the next (fifth) Intergovernmental Panel on Climate Change (IPCC) Assessment Report which is due for release in late 2013. Compared to its predecessor (CMIP3), CMIP5 has involved a much more coordinated approach to climate modelling experiments resulting in uniform inputs (atmospheric greenhouse gas, aerosol and ozone concentrations etc.), standardised outputs and a better systematic storage of the results. Details of the experiments and access to the datasets are provided by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) website[1].

One major function of model assessment is to ensure a degree of quality control for coordinated projects such as CMIP3 and CMIP5 since it is quite possible that some model results can suffer from undetected errors (or 'bugs') that degrade the results while it is also possible that errors can creep in during the archiving of the results (e.g. the use of incorrect units). Next, assessments can provide a basis for analysing climate change projections, although there is not a great deal of evidence that model skill affects these (e.g. Santer et al. 2009). Similarly, they can also inform studies which may rely on specific models to generate results such as downscaling studies which rely on plausible host models. However, one of the most compelling reasons is to document the improvement (or lack of) that accompanies various model developments. Intercomparison can therefore serve to identify which parameterisation schemes, numerical

Corresponding author address: Ian Smith,CSIRO Marine and Atmospheric Research, Private Bag 1, Aspendale, 3195, Australia. Email: ian.smith@csiro.au

[1]http://cmip-pcmdi.llnl.gov/cmip5/index.html

techniques etc. are most appropriate for making progress.

Reichler and Kim (2008) performed one of the first systematic assessments of climate models and were able to demonstrate a steady improvement over time. However, as noted by Knutti (2010), 'metrics and criteria for model evaluation must be demonstrated to relate to the projection'. If we consider just the Australian region, previous model assessment studies include: Hope (2006a, b), who pointed to problems with simulations of the winter trough over the southwest of the continent; Suppiah et al. (2007), who assessed the performance of models with respect to how well they reproduced patterns of seasonal average temperature, mean sea level pressure (MSLP) and rainfall; Watterson (1996, 2008), who used a statistic determined from simulated and observed patterns of seasonal average temperature, mean sea level pressure (MSLP) and rainfall over the continent; Perkins et al. (2007) and Maximo et al. (2007), who considered the ability to simulate daily rainfall and daily minimum and maximum temperatures for different regions; Charles (2007), who focussed on the ability to simulate both daily MSLP patterns and the seasonal cycle of monthly average MSLP; Colman et al. (2011), who noted large model biases in simulating features of the monsoon; and Smith and Chandler (2010), who focussed on the Murray Darling Basin of southeastern Australia and assessed models in terms of their ability to simulate key features of both rainfall and the El Niño Southern Oscillation (ENSO).

In choosing which features of the models to address in this study, we can begin by asking what are the key features that would indicate a model is capable of providing skilful seasonal rainfall predictions for Australia. This question is highly relevant since there are strong arguments for developing a unified or 'seamless' approach to the problem of weather forecasts, seasonal predictions and climate change projections (Brown et al. 2012). We are not in a position to test the models with regard to weather forecasts and, since we cannot test the skill of any climate change projections, it is logical to focus the assessments on features which are relevant to Australian rainfall and the ENSO phenomenon. While the Indian Ocean Dipole has also been implicated in Australian rainfall fluctuations, Smith and Timbal (2012) showed that this is relatively minor compared to the role of Pacific Ocean SSTs.

We firstly compare how well models reproduce the present day mean, interannual variability and seasonal cycle of monthly all-Australian average rainfall and monthly NINO3.4 sea surface temperatures (SSTs). All-Australian average rainfall provides a convenient indicator since it tends to represent the northern tropics during the relatively wet summer months and the southern regions during the winter months and (as will be shown) its variability is dominated by ENSO events. The NINO3.4 index is the standard operational index used by NOAA's Climate Prediction Center to monitor ENSO (McPhaden 2012).

Secondly, we compare how well the models can reproduce the seasonal cycle of correlations between the monthly rain and monthly SSTs. This study therefore focuses on the ability of models to capture the variability of rainfall and SSTs at both the seasonal and interannual timescales. For assessments focussing on Australian models, spatial patterns, or other variables, we refer the reader to other papers in this issue (e.g. Bi et al. 2013, Dix et al. 2013, Irving et al. 2012, Rashid et al. 2013, and Watterson et al. 2013).

Models

The Australian models that contributed to CMIP3 are referred to as the CSIRO-Mk3.0 and 3.5 models, while the contribution to CMIP5 includes the CSIRO-Mk3.6 model and the ACCESS1.0 and 1.3 models. The CSIRO-Mk3.6 model differs from CSIRO-Mk3.5 model (Gordon et al. 2010) by the inclusion of an interactive aerosol scheme, an updated radiation scheme and other changes to the atmospheric physics package. The model also includes dynamic sea ice and a soil–canopy scheme with prescribed vegetation properties, but no carbon cycle. For further details of the model see Rotstayn et al. (2011, 2012) and, for preliminary assessments of its performance see Rotstayn et al. (2010) and Syktus et al. (2011).

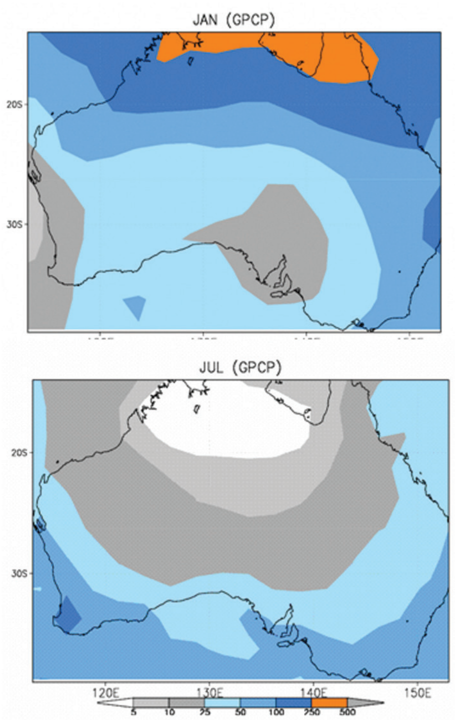
The two ACCESS models were developed after a decision was made in 2004 to combine the operational and climate modelling requirements of both CSIRO and the Bureau of Meteorology. It has also involved collaboration with the UK Meteorological Office. A major feature of ACCESS model development is that it takes a so-called 'unified model' approach in which the core model is designed for adaptation to weather forecasting, seasonal forecasting and climate variability and climate change simulations. The atmospheric component for ACCESS1.0 is based on the Met Office Hadley Centre model HadGEM2 (version r1.1) while the component for ACCESS1.3 is based on the Met Office's subsequent Global Atmosphere (GA) 1.0. For further details see Bi et al. (2013), Dix et al. (2013), Rashid et al. (2013), and Watterson et al. (2013).

We assess the results from a total of 32 models comprising 30 CMIP5 models and the two CMIP3 models (CSIRO-Mk3.0 and CSIRO-Mk3.5) which are included in order to estimate evidence of improved performance over time. Similarly, the performance of the ACCESS1.3 model relative to the ACCESS1.0 model also provides an indication of the effect of model improvements. (Note that the 30 sets of CMIP5 model results represent a large sample of the almost 60 sets of results that had been, and were still being generated, on 1 March 2013). Each assessment is based on the results from just one ensemble member from each model.

Data and methodology

The GPCP dataset (version 2.2) is based on data from over 6 000 rain gauge stations and satellite observation which have been merged to estimate monthly rainfall on a 2.5°

Fig. 1. Average (1979 to 2010) Australia region rainfall totals (mm) for (a) January and (b) July.



global grid from 1979 to the present (Adler et al. 2003). A separate dataset known as GPCC (version 6.0) comprises gridded, monthly land surface rainfall values based on rain gauge observations beginning in 1901 (Schneider et al. 2011). All-Australian average (1901 to 2000) monthly average values have been calculated using GPCC values averaged over the box region defined by 39°S to 14°S and 113°E to 153°E. Similarly, monthly averages for the NINO3.4 SST values have been calculated using observed grid point values for the box region 5°S to 5°N and 170°W to 130°W from the HadISST dataset[2].

Figure 1 shows maps of the distribution of the GPCP estimates for Australian mean January and July rainfall and highlights the relatively wet northern regions in summer and the relatively wet southern regions in winter while Fig. 2 shows the long-term (1901 to 2000) monthly average values (and associated standard deviations) for both the all-Australian rainfall and the NINO3.4 SSTs. The seasonal cycle for rainfall is characterised by a maximum in February, corresponding to the middle of the tropical wet season, and a minimum in August, corresponding to the middle of the dry season for much of the continent. The standard deviations indicate that interannual variability is closely related to the mean rainfall. The seasonal cycle for NINO3.4 SSTs is characterised by maxima in April/May and minima in November/December/January but, in contrast to the rainfall, the standard deviations are higher during the cooler months and lower during the warmer months.

[2]http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_hadisst

Fig. 2. Observed monthly values (bars) and associated standard deviations (grey bars) for (a) all-Australian average rainfall (mm per day) and (b) NINO3.4 sea surface temperatures (°C).

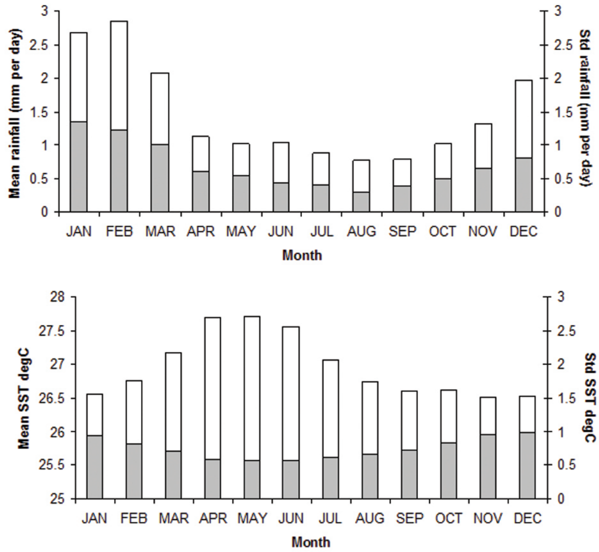


Fig. 3. Observed correlations between all-Australian monthly rainfall and NINO3.4 SSTs.

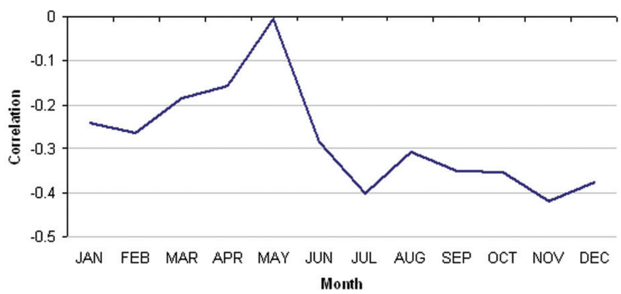


Figure 3 shows the correlations (r) between monthly rainfall and monthly NINO3.4 SSTs. The average value is –0.28 and it can be seen that a significant ($p > .05$, $|r| > .2$) relationship exists for all months except March, April and May. These minima corresponds to the so-called 'predictability barrier' that characterises ENSO events (Barnston et al. 2012).

In assessing climate models there are obviously a large number of metrics that can be used depending on the feature of interest or the intended purpose for the models. For example, Watterson et al. (2013) describes the assessment of models using metrics related to the spatial features, the variability and teleconnection patterns for several variables. Perkins et al. (2012) and Rashid et al. (2013) use metrics based on ENSO-related features such as frequency, the spatial SST anomaly patterns and links between ENSO and other climate variables. Here we focus on metrics which provide some insight into how well the models simulate key features of Australian rainfall and ENSO as revealed in Figs 2 and 3. We compare simulated and observed bias, variability, and

Table 1. CMIP5 model scores for ten metrics (see text for details). In column 1, the Australian models are highlighted in colour: ACCESS1.0 (light green), ACCESS1.3 (dark green), CSIRO-Mk3.0 (orange), CSIRO-Mk3.5 (dark orange), CSIRO-Mk3.6 (red). Within the main body of the table, the colour coding indicates relatively 'good' (blue), relatively 'mediocre' (white) and relatively 'poor' (yellow) scores.
In column A, 'good' scores are those within ten per cent of the observed mean (1.7 mm per day) while the 'poor' scores are those in excess of +/– 50 per cent.
In column B, the 'good' scores lie between 90 per cent and 110 per cent while the 'poor' scores are either less than 80 per cent or greater than 120 per cent.
In columns C, D, G and H the 'good' scores are greater than 0.90 while the 'poor' scores are less than 0.80.
In column E the 'good' scores lie within 0.5 °C of the observed value while the 'poor' scores differ by more than 1.0 °C.
In column F the 'good' scores lie between 90 per cent and 110 per cent while the 'poor' scores are either less than 70 per cent or greater than 130 per cent
In column I, the 'good' scores lie between –0.18 and –0.28, while the 'poor' scores are greater than zero.
In column J, the 'good' scores are greater than +0.50 while the 'poor' scores are less than zero.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
	Rainfall	*Rainfall*	*r: rainfall*	*r: rainfall*	*NINO34*	*NINO34*	*r: NINO34*	*r: NINO34*	*rainfall (NINO34)*	*rainfall (NINO34)*
	bias	*std ratio*	*seasonal cycle*	*std seasonal cycle*	*bias*	*std ratio*	*seasonal cycle*	*std seasonal cycle*	*mean*	*seasonal cycle*
	mm day–1	*%*			*degC*	*%*				
ACCESS1.0	–0.01	92	0.89	0.86	–0.81	78	0.75	0.97	–0.17	–0.16
ACCESS1.3	0.36	123	0.91	0.91	–0.79	87	0.23	0.54	–0.23	0.30
BCC-CSM1-1	0.69	122	0.95	0.92	–0.39	171	0.58	0.76	0.04	–0.07
BNU-ESM	–0.11	101	0.76	0.96	–1.38	173	0.67	0.78	–0.37	0.47
CanESM2	0.00	93	0.99	0.89	–1.09	128	0.89	0.94	–0.02	0.38
CCSM4	1.03	111	0.98	0.93	–0.88	121	0.77	0.90	–0.30	0.34
CESM1-BGC	0.95	113	0.98	0.93	–0.81	105	0.67	0.71	–0.29	0.20
CESM1-CAM5	0.73	102	0.98	0.94	–1.95	106	0.87	0.96	–0.35	–0.02
CNRM	0.83	105	0.97	0.95	–1.35	121	0.86	0.85	–0.32	0.57
CSIRO-Mk3.0	0.23	110	0.97	0.97	–2.14	185	0.34	0.06	–0.02	–0.33
CSIRO-Mk3.5	0.07	111	0.98	0.95	–0.32	143	0.66	–0.47	–0.27	–0.21
CSIRO-Mk3.6	0.16	107	0.97	0.97	–3.23	109	0.96	–0.40	–0.34	0.27
EC-EARTH	0.51	101	0.97	0.96	–2.81	53	0.84	0.76	0.02	–0.13
FGOALS	0.40	106	0.98	0.92	–1.59	85	0.85	0.90	0.03	–0.56
FIO-ESM	0.21	103	0.98	0.93	–0.52	129	0.54	0.56	–0.25	–0.19
GFDLCM3	0.26	107	0.95	0.94	–1.82	136	0.06	0.73	–0.27	0.58
GISS-E2-HP1	0.22	95	0.92	0.92	0.26	68	0.73	0.86	–0.12	–0.37
GISS-E2-RP1	0.00	72	0.93	0.89	0.48	73	0.88	0.86	0.00	0.20
HadGEM2-AO	0.14	99	0.96	0.95	–1.32	108	0.56	0.57	–0.17	0.02
HadGEM2-CC	–0.16	91	0.85	0.94	–1.95	109	0.54	0.53	–0.21	0.09
HadGEM2-ESM	–0.03	93	0.92	0.96	–1.84	103	0.52	0.51	–0.19	0.68
INMCM4	0.33	94	0.97	0.95	–2.67	67	0.29	0.30	–0.18	–0.30
IPSL-CM5A-LR	–0.60	62	0.82	0.85	–2.57	97	0.74	–0.27	–0.20	–0.25
IPSL-CM5A-MR	–0.61	61	0.86	0.91	–2.08	107	0.56	–0.04	–0.20	0.36
IPSL-CM5B-LR	–0.21	86	0.88	0.88	–1.12	83	0.03	0.43	–0.07	–0.03
MIROC5	1.10	132	0.98	0.91	–1.26	134	0.94	0.94	–0.29	0.43
MIROC-ESM	1.20	112	0.79	0.92	–2.98	63	0.90	0.86	–0.23	–0.01
MPI-ESM-LR	0.19	103	0.99	0.92	–2.60	112	0.87	0.57	–0.35	0.72
MPI-ESM-MR	0.26	109	0.98	0.98	–2.38	97	0.92	0.45	–0.24	0.57
MRI-CGCM3	0.22	95	0.92	0.92	–1.94	87	-0.35	0.46	–0.13	0.21
NorESM1-M	1.23	118	0.98	0.93	–1.33	96	0.72	0.90	–0.33	0.32
NorESM1-ME	1.24	123	0.98	0.95	–1.52	98	0.85	0.87	–0.35	0.33

the seasonal cycle for both the all-Australian average rainfall and the NINO3.4 SST and also calculate and compare the monthly correlations between these variables.

The ten metrics are evaluated in this study are:

- (A): the annual rainfall bias (mm per day);
- (B): the ratio (per cent) of the 12-month average interannual standard deviation of rainfall relative to the observed average;
- (C) and (D): the correlation between the 12 long-term average monthly values and standard deviations for rainfall and observed values (i.e. a measure of how well the seasonal cycle is captured);
- (E) the annual bias for the NINO3.4 SST (ºC);
- (F), (G) and (H): as for (B), (C) and (D) except for the NINO3.4 SST;
- (I): The average correlation between the 12 average monthly rainfall totals and the NINO3.4 SSTs (the observed average = –0.28);
- (J): the correlation between the 12 monthly correlations and observed correlations (i.e. a measure of how well the relationships evident in Fig. 3 are simulated).

Results

The results for each of the 32 models across the ten metrics are listed in Table 1 with the Australian models highlighted and the colour coding chosen to indicate relatively 'good' and relatively 'poor' performers. The definition of relatively 'good' or relatively 'poor' is subjective, but serves to provide a visual impression of overall model performance.

In the case of rainfall bias (column A), nine models have values within ten per cent of the observed value (1.68 mm per day) while six models have values in excess of 20 per cent. In the case of interannual variability, column B indicates that most model values fall within ten per cent of the observed value while six have values in excess of 20 per cent. There is a tendency for the biases to be similar in both cases. The models do reasonably well at simulating the seasonal cycle of rainfall (column C) with the poorest performers being BNU-ESM ($r = 0.76$) and MIROC-ESM ($r = 0.79$). The seasonal cycle of variability of rainfall is also very well simulated by the models with the lowest correlation ($r = 0.85$) associated with the HadGEM2-CC model. All the models do well at capturing the seasonal cycle of variability in monthly rainfall (column D), with values ranging from between 0.85 to 0.98.

Figure 4 shows the time series of seasonal rainfall for the Australian models compared to several other models. The GISS-E2-RP1 model has a near zero bias but has a relatively weak seasonal cycle. The IPSL-CM5A-MR model has a negative bias, particularly during the spring and summer months. The NORESM1-ME model overestimates rainfall, but has a realistic seasonal cycle. Figure 5 shows that the EC-EARTH model closely matches the observed monthly variation in rainfall variability, the CSIRO-Mk3.6 model tends to underestimate variability through the winter and spring months, while the IPSL-CM5a-MR model underestimates it during spring and summer. The MIROC5 model tends to overestimate variability in all months. In general, there is not a great deal of difference between the models when it comes to simulating the basic features for rainfall.

Fig. 4. A comparison between simulated and observed monthly average all-Australian rainfall values from selected models.

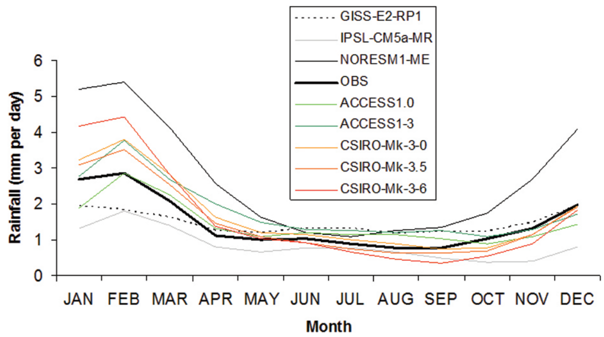
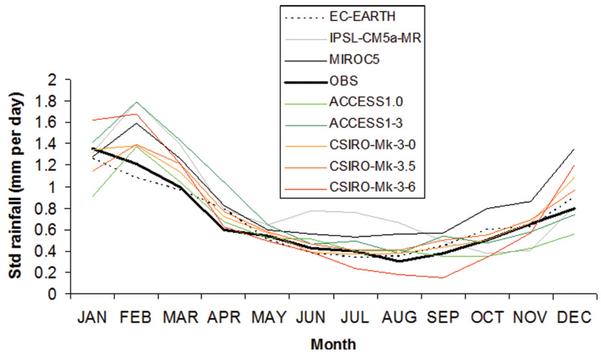


Fig. 5. A comparison between simulated and observed inter-annual standard deviation of monthly all-Australian rainfall values from selected models.



In the case of NINO3.4 biases (column E), all models bar two (GISS-E2-HP1, GISS-E2-RP1) are too cool and reflect the fact that most models still suffer (to varying degrees) from the cold tongue bias problem which afflicted the CMIP3 models (Vanniere et al. 2012). The GISS-E2-HP1 model yields the least biased (+0.26 °C) result while the CSIRO-Mk3.6 model yields the most biased (–3.23 °C) result.

Unlike rainfall, the degree to which the models can capture interannual variability (column F) ranges from as low as 53 per cent (EC-EARTH) to as high as 185 per cent (CSIRO-Mk3.0). Eleven models capture the observed variability to within about ten per cent. Not many models do very well at capturing the seasonal cycle of NINO3.4 SSTs (column G). The GFDLCM3 (0.06), IPSL-CM5-LR (0.03) and MRI-CGMCM3 (–0.35) models are particularly 'poor' while the CSIRO-Mk3.6 (0.96), MIROC5 (0.94) and MPI-ESM-Mr (0.92) models are particularly 'good'. The ability to simulate the seasonal cycle of NINO3.4 variability (column H) is quite variable. The values vary from as low as –0.47 for the CSIRO-MK3.5 model to as high as +0.97 for the ACCESS1.0 model.

Figure 6 compares simulated and observed seasonal cycles for NINO3.4 SST and shows that, despite possessing a relatively large cool bias the CSIRO-Mk3.6 model does extremely well at capturing the seasonal cycle. The GISS-E2 -HP1 and –RP1 models are slightly too warm while the MRI-CGCM3 model performs very poorly and is almost completely out of phase with the observations.

Figure 7 compares the monthly values for interannual variability and indicates that one of the better performing models according to this metric is the NORESM1-ME model, while the EC-EARTH model both underestimates the observations and shows very little seasonal variation. Unlike the two ACCESS models, all three CSIRO models tend to both overestimate variability and fail to capture the observed seasonal cycle.

Finally, the ability to simulate the rainfall–SST relationship (columns I and J, Fig. 8) also varies considerably. The MPI-ESM-LR model does best (r = +0.72) at capturing the seasonal cycle while the FGOALS model (r = –0.56, not shown) does worst. The BCC-CSM1-1 model (r = –0.07) does poorly, mainly because it simulates near zero or positive correlations in all months. The ACCESS1.0 (r = +0.30) and CSIRO-Mk3.6 (r = +0.23) models perform moderately well whereas their predecessors perform relatively poorly on this metric.

How well do the Australian models perform? Irving et al. (2012) found a slight improvement in the performance of CMIP5 models at simulating spatial features of Australian rainfall and noted that both CMIP3 and CMIP5 models experienced difficulties in capturing features of the seasonal cycle at regional scales. Here we note a slight improvement in the rainfall metric scores for ACCESS1.3 compared to the earlier version ACCESS1.0, but no similar improvement is evident with the CSIRO models,

Both sets of models have cool biases with respect to the NINO3.4 SST but while the ACCESS models tend to underestimate SST variability, as also noted by Watterson et al. (2013), the CSIRO models tend to overestimate. The CSIRO-Mk3.6 model performs best of all five models at capturing both the magnitude of the SST variability and the seasonal cycle, but the ACCESS models do much better at capturing the seasonal variation of SST anomalies.

Despite having the coldest bias (–3.23 °C), the CSIRO-Mk3.6 model does well at simulating the magnitude of NINO3.4 interannual variability (109 per cent) and does best (of all the models) at capturing the seasonal cycle (r = +0.96) and certainly represents a significant improvement over that of its CMIP3 predecessors. The ACCESS.1.0 model also captures the seasonal cycle relatively well (r = +0.75) whereas the ACCESS1.3 model performs relatively poorly (r = 0.23), a finding also reported by Rashid et al. (2013). These differences are consistent with the findings of Guilyardi et al. (2012) who found that CMIP5 models exhibited a clear improvement over CMIP3 models in simulating key ENSO features. Rashid et al. (2013) also note that the ACCESS models simulate key ENSO better than most of the previous generation models.

Fig. 6. A comparison between simulated and observed monthly average NINO3.4 SST values from selected models.

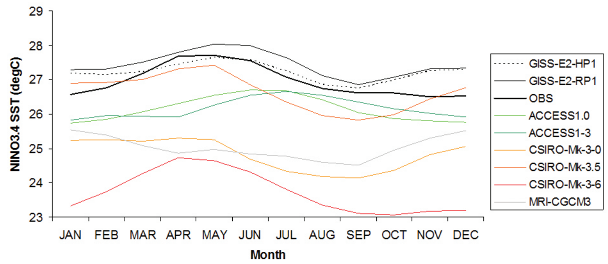


Fig. 7. A comparison between simulated and observed NINO34 interannual standard deviation of monthly NINO3.4 SST values from selected models.

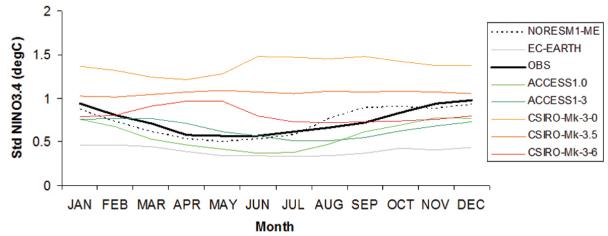
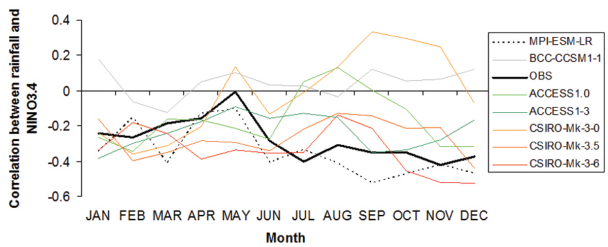


Fig. 8. A comparison between simulated and observed monthly all-Australian rainfall – NINO3.4 SST correlations from selected models.



Finally, in terms of capturing the seasonal cycle of the rainfall–SST relationship, the early versions perform relatively poorly, with correlations less than zero. However, the more recent versions (ACCESS1.3 and CSIRO-Mk3.6) perform moderately well with positive correlations (+0.30 and +0.27) that place them close to the median (+0.20) of all values.

Can we rank the models based on these metrics? This is a difficult question to answer since it depends very much on the relevance of each metric to any particular problem that is being addressed. Almost all models provide a reasonable estimate of the seasonal cycle of rainfall, so this does not provide much discrimination. On the other hand, the NINO3.4 SST results show little evidence of consistency in that it is difficult to identify models which either perform well or perform poorly across all metrics. As a consequence, we conclude that it is difficult to attempt a ranking of the models based solely on any one of the first eight metrics.

If we are interested in the potential for producing seasonal rainfall predictions, then a focus on capturing the observed relationships between the NINO3.4 SSTs and rainfall could be expected to be relevant. On this basis, we can use the results to suggest that the CNRM, HadGEM2-ESM, MPI-ESM-LR and MPI-ESM-MR models can be ranked high, and the CSIRO-Mk3.5, FGOALS, GISS-E2-HP1 and INMCM4 models ranked low. By the same criterion, the ACCESS1.0 and CSIRO-Mk3.5 models can also be ranked relatively low, while the ACCESS1.3 and CSIRO-Mk3.6 can be ranked as near-median performers.

Conclusions

An assessment of 30 CMIP5 and two CMIP3 models has been conducted based on metrics related to all-Australian region average rainfall and ENSO. The main findings are that:

- most models simulate both the mean and the interannual variability of rainfall reasonably well, but with a tendency to overestimate;
- all models capture the seasonal cycle of rainfall (the minimum correlation being +0.76);
- almost all models underestimate the annual mean value for the NINO3.4 SST due to a persistent cool tongue bias;
- simulated values for the interannual variability of NINO3.4 SSTs vary widely (the ratio of model to observed values varies from a minimum of only 53 per cent to a maximum of 185 per cent); and
- most models can capture the seasonal cycle of the SSTs, but several yield very poor correlations.

We find little consistency across the metrics which makes it difficult to combine the different scores in any meaningful way. For example, BNU-ESM model performs relatively poorly at representing the seasonal cycle of rainfall ($r = 0.76$), but captures the rainfall–SST relationship reasonably well ($r = 0.47$). The CSIRO-Mk3.6 model has a severe cool bias (–3.23 °C) in representing the NINO3.4 SSTs, but represents the seasonal cycle quite well ($r = +0.96$). However, if we choose the ability to capture the seasonal cycle of the rainfall–NINO3.4 SST relationship as a potential indicator of the skill of the models at performing seasonal predictions, then it is possible to identify relatively strong and weak performers. We can rank four models (CNRM, HadGEM2-ESM, MPI-ESM-LR and MPI-ESM-MR) high and four models (CSIRO-Mk3.5, FGOALS, GISS-E2-HP1 and INMCM4) low. Amongst the Australian models, while the ACCESS1.3 and CSIRO-Mk3.6 can only be ranked as near median performers, we note that they both represent improvements over their predecessors.

It is important to note that these findings relate to the chosen metrics only, and do not necessarily imply anything about the performance of models in other regards. They do, at least, provide indications of where some models may need to be closely scrutinised before they are used for seasonal predictions or climate change studies.

Acknowledgments

The authors acknowledge the support provided by the Queensland Government Department of Science, Information Technology, Innovation and the Arts who provided high-performance computing facilities for the CSIRO-Mk3.6 model simulations. It also acknowledges the support provided under the Australian Climate Change Science Program (an Australian Government Initiative) for development of ACCESS and CSIRO-Mk3.6, and the assistance of the National Computing Infrastructure (NCI) National Facility, where the ACCESS CMIP5 simulations were performed. We acknowledge the World Climate Research Programme Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (whose models are listed in Table 3 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Most of the analyses described in this paper were conducted using the KNMI Climate Explorer (climexp.knmi.nl; van Oldenborgh and Burgers 2005). We also thank Tony Hirst and the anonymous reviewers who provided valuable feedback on an early draft of this paper.

References

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin E. 2003. The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *J. Hydrometeor., 4,* 1147–1167.

Barnston, A.G., Tippett, M.K., L'Heureux, M.L., Li, S. and DeWitt, D.G. 2012. Skill of Real-Time Seasonal ENSO Model Predictions during 2002-11: Is Our Capability Increasing?. *Bull. Am. Meteorol. Soc., 93,* 631–651.

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Rashid, H.A., Uotila, P., Hirst, A.C., Kowalczyk, E.A., Golebiewski, M., Sullivan, A., Yan, H., Hanna, N., Franklin, C., Sun, Z., Vohralik, P., Watterson, I., Zhou, X., Fiedler, R., Collier, M., Ma, Y., Noonan, J., Stevens, L., Uhe, P., Zhu, H., Hill, R., Harris, C., Griffies, S. and Puri, K. 2013. The ACCESS Coupled Model: Description, Control Climate and Preliminary Validation, *Aust. Met. Oceanogr. J.,* 63, 41–64.

Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J, & Shelly, A. 2012. Unified modeling and prediction of weather and climate: a 25-year journey. *Bull. Am. Meteorol. Soc., 93,* 1865–1877.

Charles, S.P., Bari, M., Kitsios, A. and Bates, B.C. 2007. Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine Catchment, Western Australia. *Int. J. Climatol., 27*, 1673–1690.

Colman, R.A., Moise A.F. and Hanson L.I. 2011. Tropical Australian climate and the Australian monsoon as simulated by 23 CMIP3 models. *J. Geophys. Res., 116,* D10116, doi:10.1029/2010JD015149.

Dix, M., Vohralik, P., Bi, D., Rashid, H.A., Marsland, S., O'Farrell, S., Uotila, P., Hirst, A.C., Kowalczyk, E.A., Sullivan, A., Yan, H., Franklin, C., Sun, Z., Watterson, Collier, M., Noonan, J., Rotstayn, L., Stevens, L., Uhe, P. and Puri, K. 2013. The ACCESS Coupled Model: Documentation of core CMIP5 simulations and initial results. *Aust. Met. Oceanogr. J.,* 63, 83–99.

Gordon, H., O'Farrell, S., Collier, M., Dix, M., Rotstayn, L., Kowalczyk, E., Hirst, A. and Watterson, I. 2010. The CSIRO Mk3.5 Climate model. The Centre for Australian Weather and Climate Research: *CAWCR Technical Report No. 021.*

Guilyardi, E., Bellenger, H., Collins, M., Ferrett, S., Cai W. and Wittenberg A. 2012. A first look at ENSO in CMIP5. *Clivar Exch 17,* 29–32.

Hope, P.K. 2006a. Shifts in synoptic systems influencing southwest Western Australia. *Clim. Dyn., 26,* 751–64.

Hope, P.K. 2006b. Future changes in synoptic systems influencing southwest Western Australia. *Clim. Dyn., 26,* 765–80.

Irving, D., Whetton, P. and Moise, A. 2012. Climate projections for Australia: a first glance at CMIP5. *Aust. Met. Oceanogr. J.,* 62, 211–225.

Knutti, R. 2010. The end of model democracy? *Editorial for Climatic Change, 102,* 395–404, doi:10.1007/s10584-010-9800-2.

Maximo, C.C., McAvaney, B.J., Pitman, A.J. and Perkins, S.E. 2007. Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation. *Int. J. Climatol.,* in press. DOI: 10.1002/joc.1612

McPhaden, M.J. 2012. A 21st century shift in the relationship between ENSO SST and warm water volume anomalies. *Geophys. Res. Lett., 39,* L09706.

Perkins, S.E., Pitman, A.J., Holbrook, N.J. and McAveney, 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.

Perkins, S.E., Irving, D.B., Brown, J.R., Power, S.B., Moise, A.F., Colman R.A. and Smith I. 2012. CMIP3 ensemble climate projections over the western tropical Pacific based on model skill. *Clim. Res., 51,* 35–58.

Rashid, H.A., Sullivan, A., Hirst, A.C., Bi, D., Zhou, X. and Marsland, S.J. 2013. Evaluation of simulated El Nino-Southern Oscillation in ACCESS coupled model simulations for CMIP5. *Aust. Met. Oceanogr. J.,* 63, 161–80.

Reichler, T. and Kim, J. 2008. How well do coupled models simulate today's climate? *Bull. Am. Meteorol. Soc., 89*, 303–311.

Rotstayn, L.D., Collier, M.A., Dix, M.R., Feng, Y., Gordon, H.B., O'Farrell, S.P., Smith, I.N. and Syktus, J.I. 2010. Improved Simulation of Australian Climate and ENSO-related rainfall variability in a global climate model with an interactive aerosol treatment. *Int. J. Climatol. 30,* 1067–1088. doi: 10.1002/joc.1952

Rotstayn, L.D., Collier, M.A., Mitchell, R.M., Qin, Y. and Campbell, S.K. 2011. Simulated enhancement of ENSO-related rainfall variability due to Australian dust. *Atmos. Chem. Phys., 11,* 6575–6592, doi:10.5194/acp-11-6575-2011.

Rotstayn, L.D., Jeffrey, S.J., Collier, M.A., Dravitzki, S.M., Hirst, A.C., Syktus, J.I. and Wong, K. 2012. Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations, *Atmos. Chem. Phys., 12,* 6377–6404, doi:10.5194/acp-12-6377-2012.

Santer, B.D., Taylor, K.E., Gleckler, P.J., Bonfils, C., Barnett, T.P., Pierce, D.W., Wigley, T.M.L., Mears, C., Wentz, F.J., Bruggemann, W., Gillett, N.P., Klein, S.A., Solomon, S., Stott, P.A. and Wehner, M.F. 2009. Incorporating model quality information in climate change detection and attribution studies. *Proc. Nat. Acad. Sci.* doi: 10.1073/pnas.0901736106.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B. and Ziese, M. 2011. GPCC Full Data Reanalysis Version 6.0 at 2.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD_GPCC/FD_M_V6_250

Smith, I.N. 2007. Global climate modelling within CSIRO: 1981 to 2006. *Aust. Met. Oceanogr. J., 56,* 153–166.

Smith, I.N. and Chandler, E. 2010. Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance. *Climatic Change, 102,* 377–393, DOI 10.1007/s10584-009-9757-1.

Smith, I.N. and Timbal, B. 2012. Links between tropical indices and southern Australian rainfall. *Int. J. Climatol.,* 32, 33–40.

Suppiah, R., Hennessy, K.J., Whetton, P.H., McInnes, K., Macadam, I., Bathols, J., Ricketts J. and Page C.M. 2007. Australian climate change projections derived from simulations performed for the IPCC 4th Assessment Report. *Aust. Met. Mag., 56,* 131–152.

Syktus, J., Jeffrey, S., Rotstayn , L., Wong, K., Toombs, N., Dravitzki, S., Collier, M., Hamalainen C. and Moeseneder C. 2011. The CSIRO-QCCCE contribution to CMIP5 using the CSIRO Mk3.6 climate model. *19th International Congress on Modelling and Simulation*, Perth, Australia, 12–16 December 2011.

Vannière, B., Guilyardi, E., Madec, G., Doblas-Reyes, F.J. and Woolnough, S. 2012. Using seasonal hindcasts to understand the origin of the equatorial cold tongue bias in CGCMs and its impact on ENSO. *Clim. Dyn.*, 1–19.

van Oldenborgh, G.J. and Burgers, G. 2005. Searching for decadal variations in ENSO precipitation teleconnections, *Geophys. Res. Lett., 32,* L15701, doi:10.1029/2005GL023110, 2005.

Watterson, I.G. 1996, Non-dimensional measures of climate model performance, *Int. J. Climatol., 16,* 379–391.

Watterson, I. G. 2008. Calculation of probability density functions for temperature and precipitation. *J. Geophys. Res., 113* (D12).

Watterson, I.G., Hirst A.C. and Rotstayn L.D. 2013. A skill-score based evaluation of simulated Australian climate. *Aust. Met. Oceanogr. J.,* 63, 181–90.