

Evaluation of CMIP3 and CMIP5 Models over the Australian Region to Inform Confidence in Projections

A. Moise¹, L. Wilson², M. Grose², P. Whetton², I. Watterson², J. Bhend², J. Bathols², L. Hanson¹, T. Erwin², T. Bedin², C. Heady² and T. Rafter²

¹ Bureau of Meteorology, Australia

² CSIRO Oceans and Atmosphere Flagship, Australia

(Manuscript received November 2014; accepted April 2015)

Model evaluation is an important tool to help rate confidence in climate model simulations. This can add to the overall confidence assessment for future projections of the Australian climate. Additionally it can highlight significant model deficiencies that may affect the selection of a subset of models for use in impact assessment.

Here we present results from an extensive model evaluation undertaken as part of the Natural Resource Management (NRM) Project in order to inform the newest set of climate change projections for Australia.

The assessment covers mean climate skill over Australia as well as variability measures and teleconnections from up to 47 CMIP5 models and 23 CMIP3 models (for comparison where appropriate). Additionally, the skill in representing important climate features such as MJO, SAM, blocking and cut-off lows are also reviewed. Selected extremes are evaluated as well as simulations of two different types of downscaling simulations used within the NRM project. Finally, an attempt is made to synthesise this information in order to highlight a small group of CMIP5 models which show consistent deficiencies in representing the Australian climate and its features.

Introduction

Climate models are our primary tools for investigating the response of the Earth's climate system to forcings such as greenhouse gases, and for making projections of the future climate. It is crucial to evaluate the individual models and ensembles of models used in climate studies (Flato et al. 2013). The evaluation of models in simulating the current climate and recent climate trends is a ground to accept or reject models for use in a particular application, and is an important component in assessing the confidence in future projections from the ensemble used.

In this paper climate models are evaluated by using measures of agreement between model simulations and observations of the present climate of the Australian region. The results of this model evaluation contribute to the assessment of confidence in model-simulated future climate changes (CSIRO and Bureau of Meteorology 2015) and also to the assessment of the adequacy of any model, or models in general, for particular applications. Recent IPCC Assessment Reports also use model evaluation to guide confidence in projections of future climate (IPCC 2007, 2013).

The ability of individual models in the Coupled Model Intercomparison Project phase 5 (CMIP5) archive (Taylor et al. 2012) to simulate the Australian climate can vary depending on which aspect of a model simulation is considered. There is a wide range of climate features that have been included in this evaluation, in order to capture the complexity of the climate system. These features include the mean and seasonal cycle of variables such as surface temperature, rainfall and wind, but also the tele-connection of Australian rainfall to the main drivers of variability such as the El Niño Southern Oscillation (see also Section 4.1.2 in CSIRO and Bureau of Meteorology 2015). It is not possible to produce an

exhaustive evaluation of every aspect of the climate system relevant to regional climate change, however the assessment of the mean state of surface variables and a few important aspects driving climate variability is useful evidence for informing confidence in projections. Obtaining a wide range of assessment metrics, some of which are not independent, makes it difficult to identify a group of best performing models, however it is possible to identify a small subset of models that perform consistently poorly across many aspects of the climate, or that perform poorly on critical aspects of the climate. Such information on poorly performing individual models is relevant when users of climate change projection information are choosing a subset of models for application in impact assessment, such as through the Climate Futures approach – see Whetton et al. (2012). Similarly, the results from the model evaluation are one very important consideration when choosing host models for further downscaling approaches (Evans et al 2013).

Global climate models are designed to simulate large-scale processes. On a smaller regional scale, the spatial and temporal details of these processes are simulated with much more varying capacity. On even smaller scales, processes might not be directly simulated by global climate models at all (i.e. tropical cyclones). Climate model resolution will give a rough indication of the spatial extent as to what features and processes these models may simulate realistically.

CMIP5 is overall an improved set of global climate models compared to the previous CMIP3 (Meehl et al. 2007) in terms of model formulation. The improvements arise from the increase in horizontal and vertical resolution; an improved representation of processes within the climate system (i.e. aerosol-cloud interactions, and the carbon cycle in the subset that are Earth-System-Models, ESMs) and also the availability of a larger number of ensemble members improves statistics overall (Chapter 9 in IPCC, 2013). On a continental and global scale, this has also led to an improved ability to simulate historical climate. Some examples of this ability are reported by IPCC (2013) and include the representation of:

- Global mean surface temperature, including trends over the recent decades
- Long-term global mean large-scale rainfall patterns (but less well than temperature)
- Regional mean surface temperature (sub-continental scales)
- Annual cycle of Arctic sea ice extent (and recent trends)
- Trends in ocean heat content
- ENSO simulation
- Extreme events, especially temperature related ones
- Recent ozone trends

Beside these improvements, certain areas have not improved since the previous IPCC Assessment in 2007. This includes important systematic errors and biases such as the "cold tongue" bias (e.g. the sea surface temperature difference between East and West equatorial Pacific is too large in models, leading to a cold bias in the Western Pacific), problems in simulating the diurnal cycle of rainfall, the Madden-Julian Oscillation, and more. In many cases, there is a large inter-model spread leading to enhanced uncertainty, however amongst the models that do not include carbon cycle these are reduced compared to CMIP3 (IPCC, 2013).

Apart from spatial resolution, models also employ different physical schemes representing atmospheric and oceanic processes (such as clouds and convection schemes). One of the main aims of model evaluation is to assess the skill of these models through standardised inter-comparisons. The CMIP5 experiments allow for such a comparison. This paper will focus on the assessment of climate models for the Australian region.

Data and Methods

At the core of every model evaluation is a set of high quality observations to which model simulations can be compared. The high quality data set from the Australian Water Availability Project (AWAP, Jones et al. 2009; Raupach et al. 2009, 2012) is used for the evaluation of the spatial distribution of rainfall and temperature over the Australian continent. These provide an excellent indicator of mean climate across Australia. For the assessment of trends in temperature a recently updated high-quality reference station data set is used (ACORN-SAT – Trewin 2013). For several climate fields (including rainfall and temperature) there are multiple global data sets available, which allows for an extension of the evaluation over a wider region including ocean regions surrounding Australia, and an estimate of the uncertainty in observations when multiple data sets are used for the same climate field. The various observational and reanalysis datasets used in this paper are described in Table 1 and the global climate models (from CMIP3 and CMIP5) are described in Table 2, including the model labels used throughout this paper. For the analysis of ENSO observational data were derived from the Hadley Centre Sea Ice and SST dataset (HadISST; Rayner et al. 2003).

Most analysis in this paper is carried out using the Run 1 simulation from the *historical experiments* of the models taking part in the Coupled Model Inter-comparison Phase 5 (CMIP5) experiment coordinated by the World Climate Research program (WCRP) and described in Taylor et al. (2012). For the historical experiments, a comprehensive set of historical anthropogenic emissions and land-use and land-cover change data have been assembled in order to produce a relatively homogeneous ensemble of historical simulations with common time-series of forcing

agents (e.g. a prescribed set of concentrations). For most evaluation tasks, we compare climate over the time slice 1986–2005 as a representative recent climate period consistent with the ‘baseline’ period used in CSIRO and Bureau of Meteorology (2015) and IPCC (2013). When evaluating climate variability (and if the observational data set is available for longer periods) the period 1950–2005 is used. It will be stated clearly which period was used.

Table 1 List of global gridded observational (white) and reanalysis data sets (grey), their climate fields used, time coverage, origin and reference. The reference data sets for surface air temperature and rainfall over Australia are AWAP and ACORN-SAT. The abbreviation pr refers to precipitation; tas: surface air temperature; mslp: mean sea level pressure; sst: sea surface temperature.

<i>Gridded data set</i>				
Name	Fields	Period	Origin	References
AWAP	pr	1900–2012	Australian Water Availability Project, Bureau of Meteorology and CSIRO	Jones et al. 2009a; Raupach et al., 2009 and 2012
ACORN-SAT	tas	1910–2012	Australian Climate Observations Reference Network – Surface Air Temperature, Bureau of Meteorology	Trewin 2013
CMAP	pr	1979–2008	Climate Prediction Centre Merged Analysis of Precipitation	Xie & Arkin, 1997
GPCC	pr	1901–2010	Global Precipitation Climatology Centre (GPCC)	Rudolf et al., 2005; Beck et al., 2005
GPCP	pr	1979–2008	Global Precipitation Climatology Project 2	Huffman et al., 2009; Adler et al., 2003
CRU	tas	1901–2006	Climate Research Unit temperature database	Harris et al., 2013
GISS	tas	1850–2006	NASA Goddard Institute for Space Sciences (GISS) Surface Temperature Analysis	GISTEMP; Hansen et al., 2010
HADCRU	tas, pr	1901–2008	Met Office Hadley Centre and Climate Research Unit	HadCRUT3; Brohan et al., 2006
COREv2	pr, tas, mslp	1958–2006	CLIVAR Working Group on Ocean Model Development (WGOMD) Coordinated	Large & Yeager, 2009 & 2004
HOAPS	pr, fluxes	1987–2005	Hamburg Ocean Atmosphere parameters and fluxes satellite	Fennig et al., 2012; Andersson et al., 2010
HadISST	sst	1870–2010	Hadley Centre Sea Ice and Sea Surface Temperature dataset	HadISST2; Rayner et al., 2003
HadSLP2	mslp	1850–2004	Hadley Centre Sea Level Pressure dataset	Allan and Ansell 2006
CFSR	pr, winds, mslp, tas	1979–2009	NCEP Climate Forecast System Reanalysis	Saha et al., 2010
Merra	pr, winds, mslp, tas	1979–2011	Modern Era Retrospective-analysis for Research and Applications	Rienecker et al., 2011
ERA40	pr, winds, mslp, tas	1958–2002	European 40-year reanalysis	Uppala et al., 2005
ERA_INT	pr, winds, mslp, tas	1979–2011	ERA-interim	Dee et al., 2011
NCEP	pr, winds, mslp, tas	1948–2011	NCEP/NCAR reanalysis 1	Kalnay et al., 1996
NCEP2	pr, winds, mslp, tas	1979–2011	NCEP/DOE reanalysis 2	Kanamitsu et al., 2002
JRA25anl	pr, winds, mslp, tas	1979–2010	Japanese 25-year reanalysis	Onogi et al., 2007

All atmospheric (oceanic) data were assessed by interpolating to the grid spacing of the observed data set. When calculating the multi-model mean (MMM), data were interpolated to a common 1.5° latitude / longitude grid prior to analysis. Region-specific information on model evaluation will be focussed on climatic regions that have been identified by Natural Resource Management (NRM) authorities within Australia as important regional clusters. Figure 1 shows a map of the Australian topography as well as these eight clusters including the naming convention for each of the NRM cluster regions. Along with the eight clusters, regional information is presented for four ‘super-clusters’: Southern Australia (SA) made up of the Southern Slopes, Murray Basin and South and South-western Flatlands clusters; Eastern Australia (EA), made up of the Central Slopes and East Australia clusters; Northern Australia (NA), made up of the Monsoonal North and Tropical North clusters; and the Rangelands (R), containing only the Rangelands cluster. Some results are also presented for ‘sub-clusters’ that are divisions of clusters: ECN East Coast North and ECS East Coast South dividing the EC cluster along the Queensland–NSW border; MNE Monsoonal North East and MNW Monsoonal North West dividing MN along the Queensland–Northern Territory border; RN Rangelands North and RS Rangelands South dividing Rangelands roughly through the middle; four sub-clusters for Southern Slopes dividing it into quarters (SSTE Southern Slopes Tasmania East, SSTW Southern Slopes Tasmania West, SSVW Southern Slopes Victoria West and SSVE Southern Slopes Victoria East); and SSWFE South and South Western Flatlands East and SSWFW South and South Western Flatlands West, dividing up SSWF into the Western Australia and South Australia portions. See CCiA (2015) for more information about regionalisation.

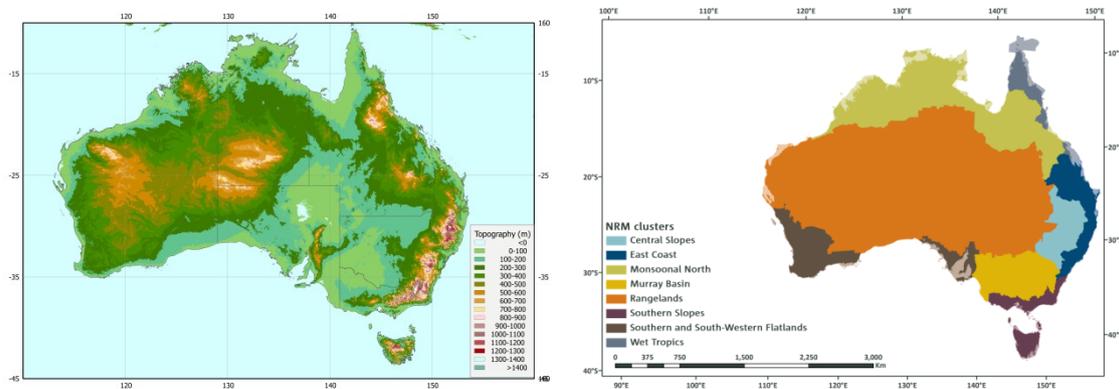
Table 2 List of CMIP5 and CMIP3 ocean-atmosphere general circulation models including the grid resolution for the ocean and atmosphere components (in degrees) and the size of a single atmosphere grid cell (in km).

<i>CMIP5 Model ID</i>	<i>Institute and Country of Origin</i>	<i>Ocean horizontal resolution (°lat x °lon)</i>	<i>Atmosphere horizontal resolution (°lat x °lon)</i>	<i>Atmosphere Eq. resolution Latitude (Km)</i>	<i>Longitude (Km)</i>
ACCESS-1.0	CSIRO-BOM, Australia	1.0×1.0	1.9×1.2	210	130
ACCESS-1.3	CSIRO-BOM, Australia	1.0×1.0	1.9×1.2	210	130
BCC-CSM1-1	BCC, CMA, China	1.0×1.0	2.8×2.8	310	310
BCC-CSM1-1-M	BCC, CMA, China	1.0×1.0	1.1×1.1	120	120
BNU-ESM	BNU, China	0.9×1.0	2.8×2.8	310	310
CanCM4	CCCMA, Canada	1.4×0.9	2.8×2.8	310	310
CanESM2	CCCMA, Canada	1.4×0.9	2.8×2.8	310	310
CCSM4	NCAR, USA	1.1×0.6	1.2×0.9	130	100
CESM1-BGC	NSF-DOE-NCAR, USA	1.1×0.6	1.2×0.9	130	100
CESM1-CAM5	NSF-DOE-NCAR, USA	1.1×0.6	1.2×0.9	130	100
CESM1-FASTCHEM	NSF-DOE-NCAR, USA	1.1×0.6	1.2×0.9	130	100
CESM1-WACCM	NSF-DOE-NCAR, USA	1.1×0.6	2.5×1.9	275	210
CMCC-CESM	CMCC, Italy	2.0×1.9	3.7×3.7	410	410
CMCC-CM	CMCC, Italy	2.0×1.9	0.7×0.7	78	78
CMCC-CMS	CMCC, Italy	2.0×2.0	1.9×1.9	210	210
CNRM-CM5	CNRM-CERFACS, France	1.0×0.8	1.4×1.4	155	155
CNRM-CM5-2	CNRM-CERFACS, France	1.0×0.8	1.4×1.4	155	155
CSIRO-Mk3-6-0	CSIRO-QCCCE, Australia	1.9×0.9	1.9×1.9	210	210
EC-EARTH	EC-EARTH, Europe	1.0×0.8	1.1×1.1	120	120
FIO-ESM	FIO, SOA, China	1.1×0.6	2.8×2.8	310	310
GFDL-CM2p1	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
GFDL-CM3	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
GFDL-ESM2G	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
GFDL-ESM2M	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
GISS-E2-H	NASA/GISS, NY, USA	2.5×2.0	2.5×2.0	275	220
GISS-E2-H-CC	NASA/GISS, NY, USA	1.0×1.0	1.0×1.0	110	110
GISS-E2-R	NASA/GISS, NY, USA	2.5×2.0	2.5×2.0	275	220
GISS-E2-R-CC	NASA/GISS, NY, USA	1.0×1.0	1.0×1.0	110	110
HadCM3	MOHC, UK	1.2×1.2	3.7×2.5	410	280
HadGEM2-AO	NIMR-KMA, Korea	1.0×1.0	1.9×1.2	210	130
HadGEM2-CC	MOHC, UK	1.0×1.0	1.9×1.2	210	130
HadGEM2-ES	MOHC, UK	1.0×1.0	1.9×1.2	210	130
INMCM4	INM, Russia	0.8×0.4	2.0×1.5	220	165
IPSL-CM5A-LR	IPSL, France	2.0×1.9	3.7×1.9	410	210
IPSL-CM5A-MR	IPSL, France	1.6×1.4	2.5×1.3	275	145
IPSL-CM5B-LR	IPSL, France	2.0×1.9	3.7×1.9	410	210
MIROC4h	JAMSTEC, Japan	0.3×0.2	0.56×0.56	60	60
MIROC5	JAMSTEC, Japan	1.6×1.4	1.4×1.4	155	155
MIROC-ESM	JAMSTEC, Japan	1.4×0.9	2.8×2.8	310	310
MIROC-ESM-CHEM	JAMSTEC, Japan	1.4×0.9	2.8×2.8	310	310
MPI-ESM-LR	MPI-N, Germany	1.5×1.5	1.9×1.9	210	210
MPI-ESM-MR	MPI-N, Germany	0.4×0.4	1.9×1.9	210	210

<i>CMIP5 Model ID</i>	<i>Institute and Country of Origin</i>	<i>Ocean horizontal resolution (°lat x °lon)</i>	<i>Atmosphere horizontal resolution (°lat x °lon)</i>	<i>Atmosphere Latitude (Km)</i>	<i>Eq. resolution Longitude (Km)</i>
MPI-ESM-P	MPI-N, Germany	1.5×1.5	1.9×1.9	210	210
MRI-CGCM3	MRI, Japan	1.0×0.5	1.1×1.1	120	120
MRI-ESM1	MRI, Japan	1.0×0.5	1.1×1.1	120	120
NorESM1-M	NCC, Norway	1.1×0.6	2.5×1.9	275	210
NorESM1-ME	NCC, Norway	1.1×0.6	2.5×1.9	275	210
AVERAGE				228	190
Highest resolution				60	60
Lowest resolution				410	410

<i>CMIP5 Model ID</i>	<i>Institute and Country of Origin</i>	<i>Ocean horizontal resolution (°lat x °lon)</i>	<i>Atmosphere horizontal resolution (°lat x °lon)</i>	<i>Atmosphere Latitude (Km)</i>	<i>Eq. resolution Longitude (Km)</i>
bccr-bcm2-0	BCCR, Norway	1.0×1.0	2.8×2.8	310	310
cccma-cgcm3-1	CCCMA, Canada	1.9×1.9	3.7×3.7	410	410
cccma-cgcm3-1-t63	CCCMA, Canada	1.4×0.9	2.8×2.8	310	310
cnrm-cm3	CNRM, France	2.0×1.0	2.8×2.8	310	310
csiro-mk3-0	CSIRO, Australia	1.9×0.9	1.9×1.9	210	210
csiro-mk3-5	CSIRO, Australia	1.9×0.9	1.9×1.9	210	210
gfdl-cm2-0	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
gfdl-cm2-1	NOAA, GFDL, USA	1.0×1.0	2.5×2.0	275	220
giss-aom	NASA/GISS, USA	4.0×3.0	4.0×3.0	440	330
giss-model-e-h	NASA/GISS, USA	1.0×1.0	5.0×4.0	550	440
giss-model-e-r	NASA/GISS, USA	5.0×4.0	5.0×4.0	550	440
iap-fgoals1-0-g	IAP, China	1.0×1.0	2.8×2.8	310	310
ingv-echam4	INGV, Italy	1.0×1.0	1.1×1.1	120	120
inmcm3-0	INM, Russia	2.5×2.0	5.0×4.0	550	440
ipsl-cm4	IPSL, France	2.0×1.0	3.7×2.5	410	280
miroc3-2-hires	CCSR, Japan	1.2×0.6	1.1×1.1	120	120
miroc3-2-medres	CCSR, Japan	1.4×0.9	2.8×2.8	310	310
miub-echo-g	MIUB, Germany/Korea	2.8×2.3	3.7×3.7	410	410
mpi-echam5	MPI-M, Germany	1.0×1.0	1.9×1.9	210	210
mri-cgcm2-3-2a	MRI, Japan	2.5×2.0	2.8×2.8	310	310
ncar-ccsm3-0	NCAR, CO, USA	1.1×0.5	1.4×1.4	155	155
ncar-pcm1	NCAR, CO, USA	1.0×1.0	2.8×2.8	310	310
ukmo-hadcm3	MOHC, UK	1.2×1.2	3.8×2.5	420	280
AVERAGE				325	290
Highest resolution				120	120
Lowest resolution				550	440

Figure 1 Map of Australia showing the (a) state boundaries and topography; and (b) main eight NRM cluster regions Central Slopes (CS), East Coast (EC), Monsoonal North (MN), Murray Basin (MB), Rangelands (R), Southern Slopes (SS), Southern and South-Western Flatlands (SSWF) and Wet Tropics (WT). (source of Figure 1b: CCiA 2015)



This paper is structured as follows: the performance of global climate models with respect to climatological characteristics, features and processes is evaluated in the following sections. Then the skill of the models in reproducing important climate features is described. Following that, there is an overview of how the recent observed trends in rainfall and temperature are captured by the models. The simulation of climatic extremes is evaluated in next and downscaling simulations are discussed as well. Finally, a discussion and conclusion is provided.

Results of an evaluation based on climatological characteristics

Here we compare the climatology of surface temperature, rainfall and wind in the CMIP3 and CMIP5 models to observations. For reference, Table 2 shows spatial resolution of both the atmospheric and ocean components of the CMIP5 models.

Assessment of historical mean climatologies: temperature, rainfall and mean sea level pressure

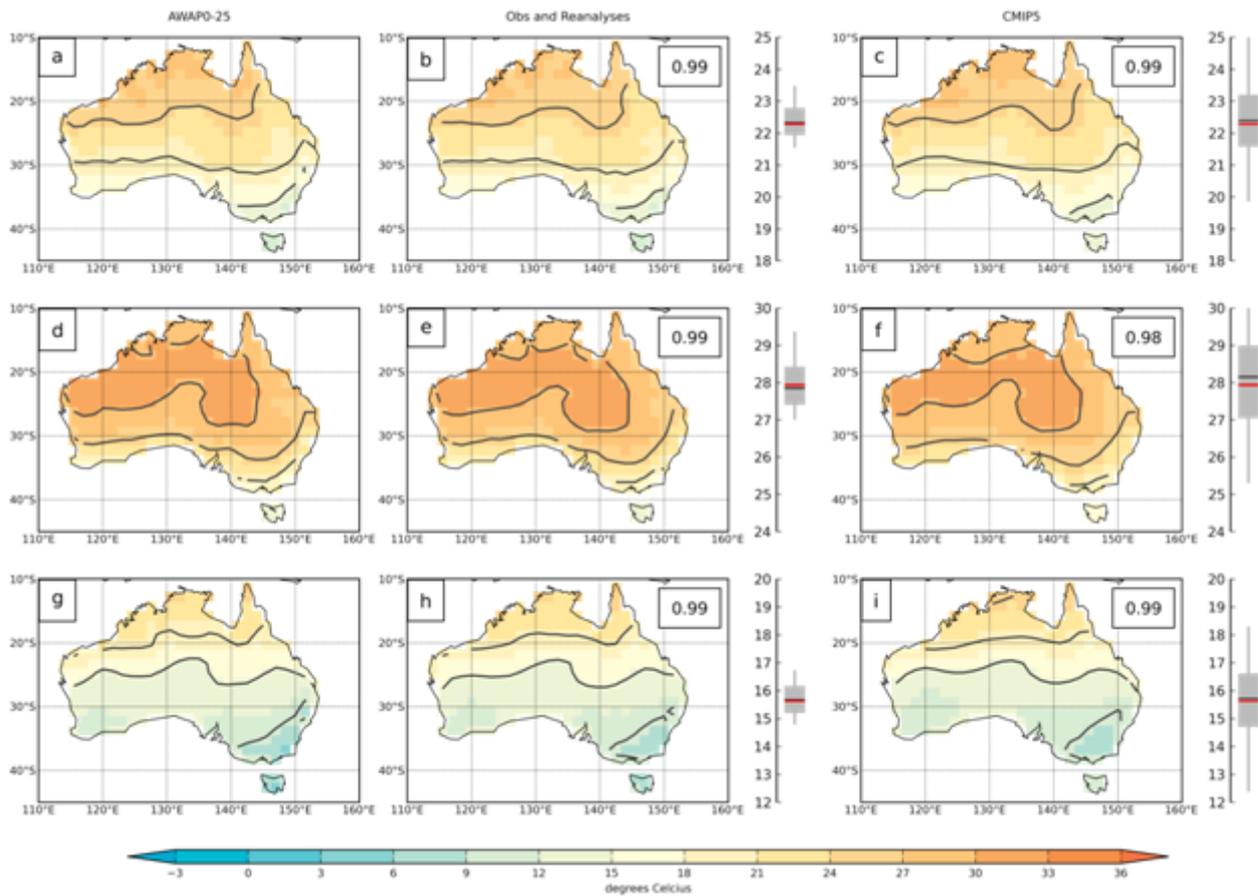
Figures 2 and 3 show a comparison of annual and seasonal climatologies (long-term averages) of temperature and rainfall for Australia for the period (1986–2005). The left column in both figures shows the reference observational data set (AWAP, Jones et al. 2009; Raupach et al. 2009, 2012) while the middle column shows an average of a selection of other observational data sets and reanalyses (7 for temperature and 8 for rainfall; see Table 1 for an overview of these) and the right column displays the CMIP5 ensemble mean.

On average, the CMIP5 models capture the climatological temperature distribution across the continent very well. The north-south gradient in temperature is correctly simulated as well as the coastal versus inland differences during summer (middle row in Figure 2) with cooler areas reaching a bit too far inland over northern parts of Western Australia. During winter (Jun–Aug) the model ensemble mean model seems slightly too warm over northern Australia as well as coastal regions in the south east and Tasmania. Pattern correlations are generally very high for the mean model.

There is a substantial spread in the Australia-averaged temperature amongst the CMIP5 models as indicated by the spread in the box-whiskers in Figure 2. While 50 per cent of the models are within ± 1 °C of the AWAP reference data, some of the models are several degrees warmer or colder. The box-whiskers belonging to the middle column in Figure 2 additionally indicate that there is some discrepancy amongst the other observational data sets and reanalysis data sets with respect to temperature across Australia. But this discrepancy is generally less than half of the spread seen in the CMIP5 models.

Some of the model differences in temperature are driven by their differences in the simulation of the hydrological cycle and Figure 3 shows their skill in simulating rainfall across Australia. There is a general tendency to have too much rainfall across north-western Australia and reaching too far into the interior of the continent (summer and annual case). North-eastern regions show somewhat less summer rainfall in the models compared to AWAP.

Figure 2 Climatological mean surface air temperature from AWAP (left column, the reference data set), the average of a selection of other observational data sets (middle column, see Table 1) and the CMIP5 mean model (right column, see Table 2) for annual (top row), summer (Dec-Feb, middle row) and winter (Jun-Aug, bottom row) surface air temperature. The averaging period is 1986-2005 and the units are degrees Celsius (°C). The contours highlight the 9, 15, 21, 27 and 33 °C thresholds for better comparison. The number in the top right corner indicates the spatial correlation between the corresponding data and AWAP. The spread in the data sets is indicated by the box-whisker to the right of each subplot: each shows the Australia-averaged surface air temperature where the grey box refers to the middle 50 % of the data and the whiskers show the spread from minimum to maximum. The thick black line is the median of the underlying data and the red line is AWAP.

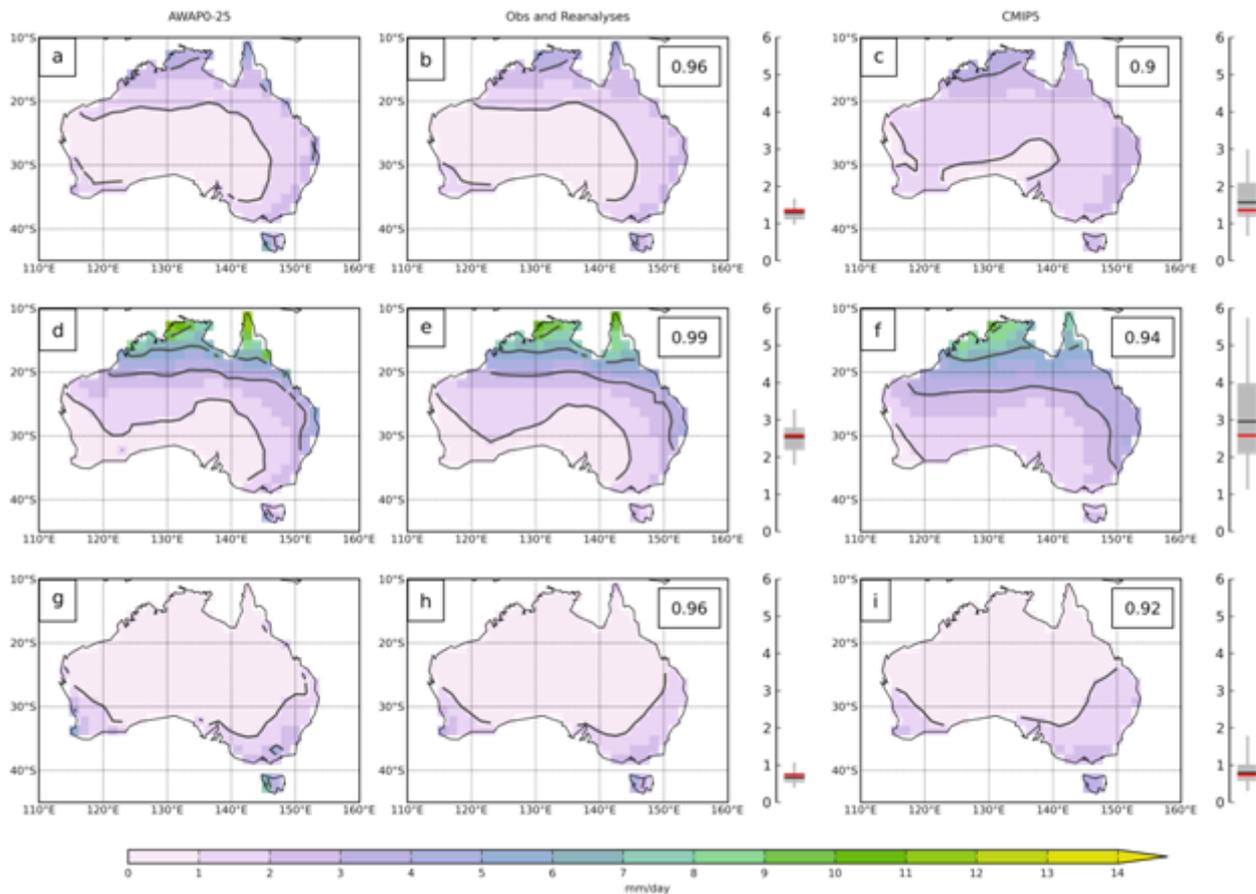


The winter rainfall regime (across southern coastal regions of Australia) on the other hand is generally too dry, especially in Tasmania. This could be caused by two aspects of insufficient resolution in global climate models: (a) some models’ resolution is too coarse to represent the land mass of Tasmania at all; (b) even if global models include Tasmania they do not sufficiently resolve the topographically driven high rainfall regimes particularly over western regions of Tasmania. Therefore the pattern correlations are lower for rainfall compared to temperature and the model spread for summer rainfall is very large.

Figures 4 and 5 show the assessment of the CMIP5 model biases in seasonal surface air temperature and rainfall climatologies averaged over the clusters, super-clusters and sub-clusters. In general, the CMIP5 models are able to capture seasonal temperatures much better than rainfall which is well known and has been reported widely (IPCC 2007 and 2013).

During summer, the model simulated median seasonal temperatures (Figure 4) are very close to AWAP reference values, particularly for the warmer regions across northern and central Australia (MN, R clusters for example). While the temperature range within the model ensemble can be as large as 3 °C (with some models showing an even larger cold bias in southern regions and Tasmania), the majority of the models are within +/-1 °C of the observed values.

Figure 3 Climatological mean rainfall from AWAP (left column, the reference data set), the average of a selection of other observational data sets (middle column, see Table 1) and the CMIP5 mean model (right column, see Table 2) for annual (top row), summer (Dec-Feb, middle row) and winter (Jun-Aug, bottom row) rainfall. The averaging period is 1986-2005 and the units are mm per day. The contours highlight the 1, 3, 6, and 9 mm/day thresholds. The number in the top right corner indicates the spatial correlation between the corresponding data and AWAP. The spread in the data sets is indicated by the box-whisker to the right of each subplot: each shows the Australia-averaged rainfall where the grey box refers to the middle 50 % of the data and the whiskers show the spread from minimum to maximum (for CMIP5 data only). The thick black line is the median of the underlying data and the red line is AWAP.



During winter, the majority of climate models have warm biases over some regions of south-eastern Australia (Southern Slopes cluster, see Figure 1b for details on clusters.). Most other cluster regions are very well simulated with the median temperatures often within 1 °C of the AWAP values. Noteworthy is the large overall spread between the models, which can reach more than 4 °C between the warmest and coldest model for a particular cluster.

Overall, the biases in temperature point towards a deficiency in some models in capturing the north-south temperature gradient across Australia in either the summer or winter season.

Rainfall biases averaged over clusters, super-clusters and sub-clusters are shown in Figure 5 for summer and winter. The skill of models in simulating climatological rainfall varies strongly across Australia: for example during summer, models capture rainfall amounts over regions with moderate to high seasonal rainfall totals such as the monsoon regions (except the Wet Tropics) and along the East Coast but show more variable skill elsewhere. To illustrate the spread in skill in simulating summer rainfall across tropical regions of Australia, Figure 6 shows the December to February rainfall climatology (in mm/day) across Northern Australia from 72 models (CMIP5 and CMIP3). Shown are also the observed climatology and the ensemble mean models for CMIP3 and CMIP5. In order to better compare the results between CMIP3 and CMIP5 models, the climatology has been calculated using the same period (1980-1999). While there is a fairly large model spread (particularly over the monsoon affected regions), the median rainfall is close to the AWAP data in summer. Some models fail to produce a monsoon-related rainfall climatology over northern Australia while others show much too strong rainfall amounts that extend too far south into the continent.

Figure 4 Cluster averaged mean surface air temperature (units: °C) for summer (Dec-Feb, top) and winter (Jun-Aug, bottom) from all CMIP5 models (represented by box-whisker bars), AWAP (red circle) and several other observations and re-analysis data sets (coloured dots). The box-whiskers display the middle 50 % of the CMIP5 models (box, including the median of the CMIP5 models as thick black line) and the range (whiskers) while outlier models are shown as black circles (i.e. they are more than 1.5 times the box width from the median away). The time period used for rainfall averages is 1986-2005. Regions are marked on the x-axis: see section 2 for definitions.

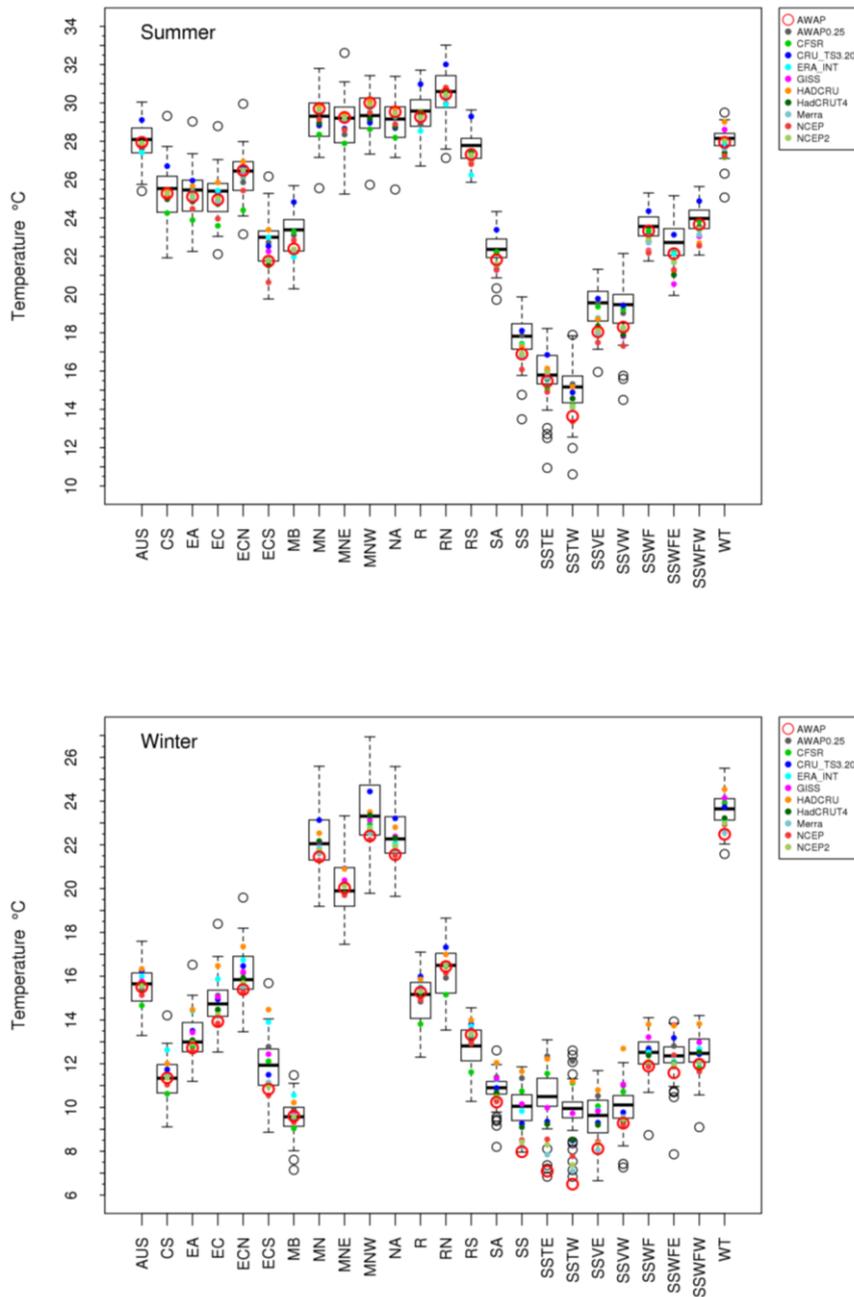


Figure 5 Cluster averaged rainfall (units: mm per day) for summer (Dec-Feb, top) and winter (Jun-Aug, bottom) from all CMIP5 models (represented by box-whisker bars), AWAP (red circle) and several other observations and re-analysis data sets (coloured dots). The box-whiskers display the middle 50 % of the CMIP5 models (box, including the median of the CMIP5 models as thick black line) and the range (whiskers) while outlier models are shown as black circles (i.e. they are more than 1.5 times the box width from the median away). The time period used for rainfall averages is 1986-2005. Regions are marked on the x-axis: see section 2 for definitions.

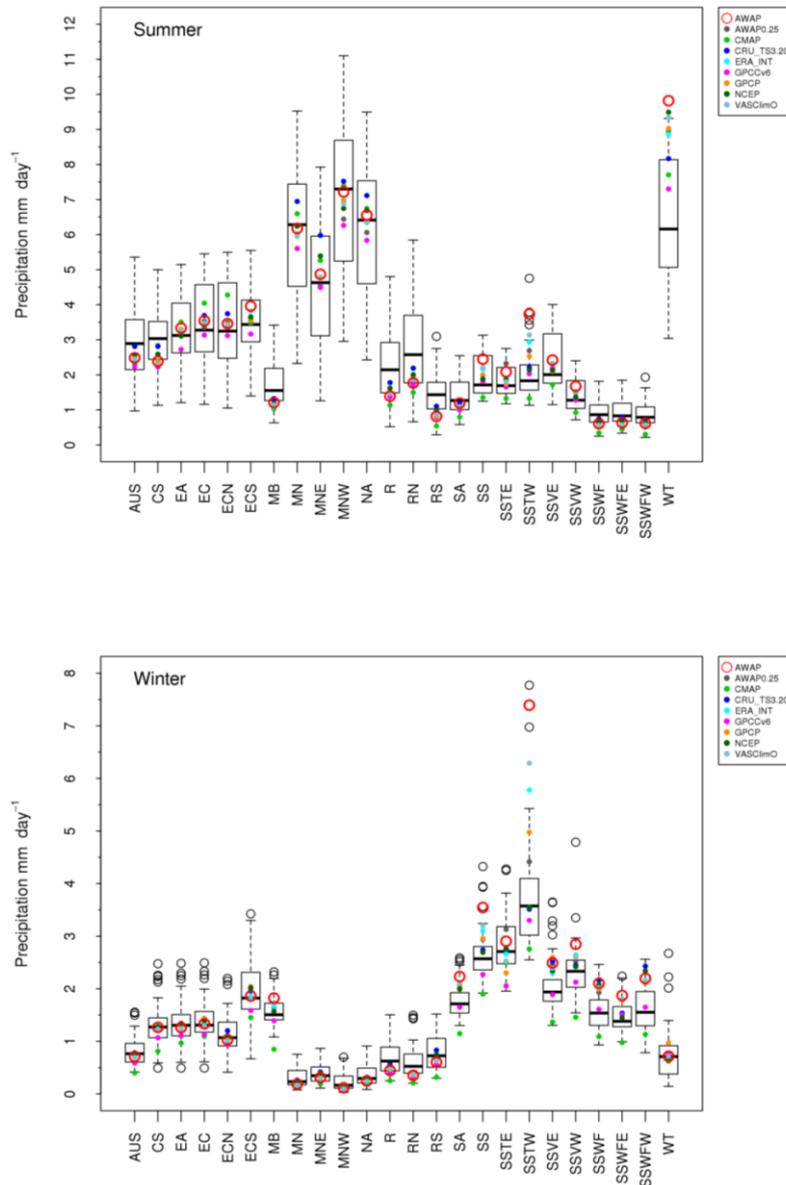
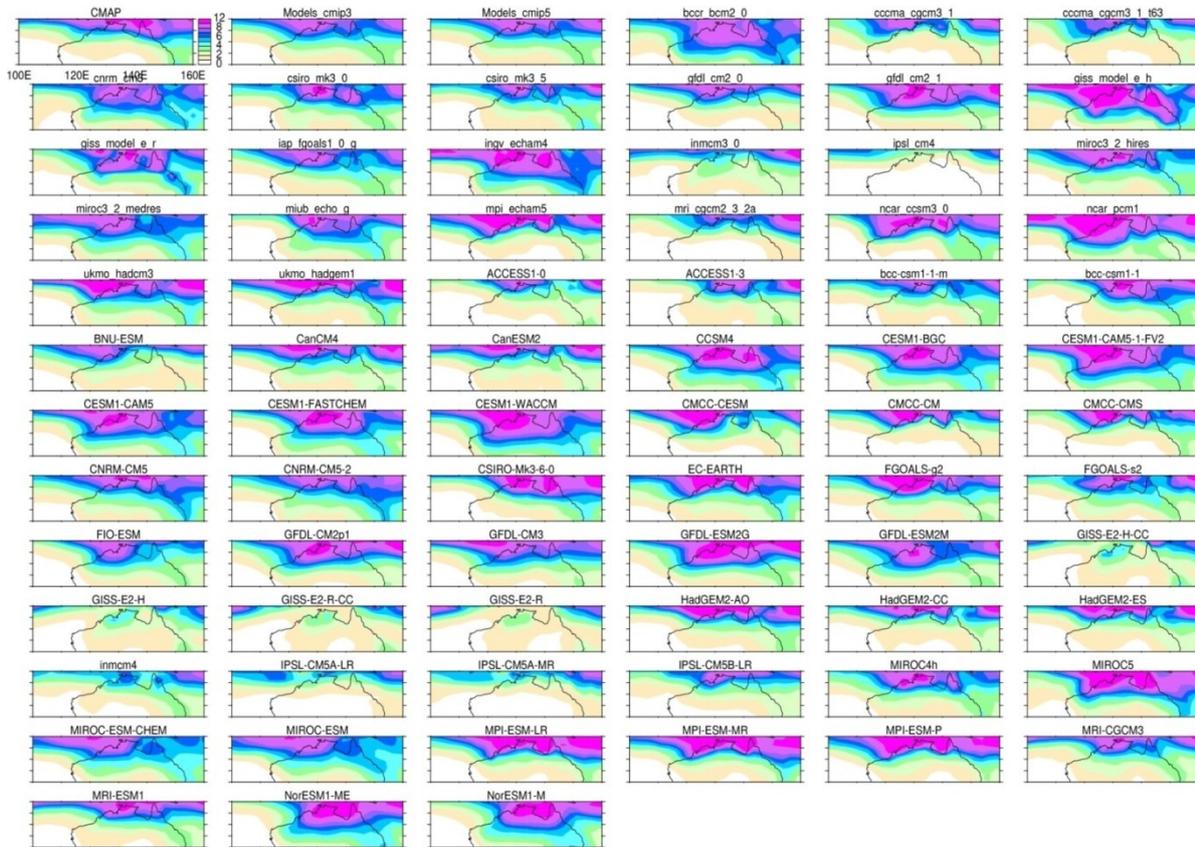


Figure 6 Dec-Feb rainfall climatology (in mm/day) across Northern Australia from 72 models (CMIP5 and CMIP3). Shown are also the observed climatology (AWAP; top left) and the ensemble mean models for CMIP3 (second top left) and CMIP5 (third top left). The climatological means are taken over the same time period (1980-1999) for both CMIP3 and CMIP5 models.

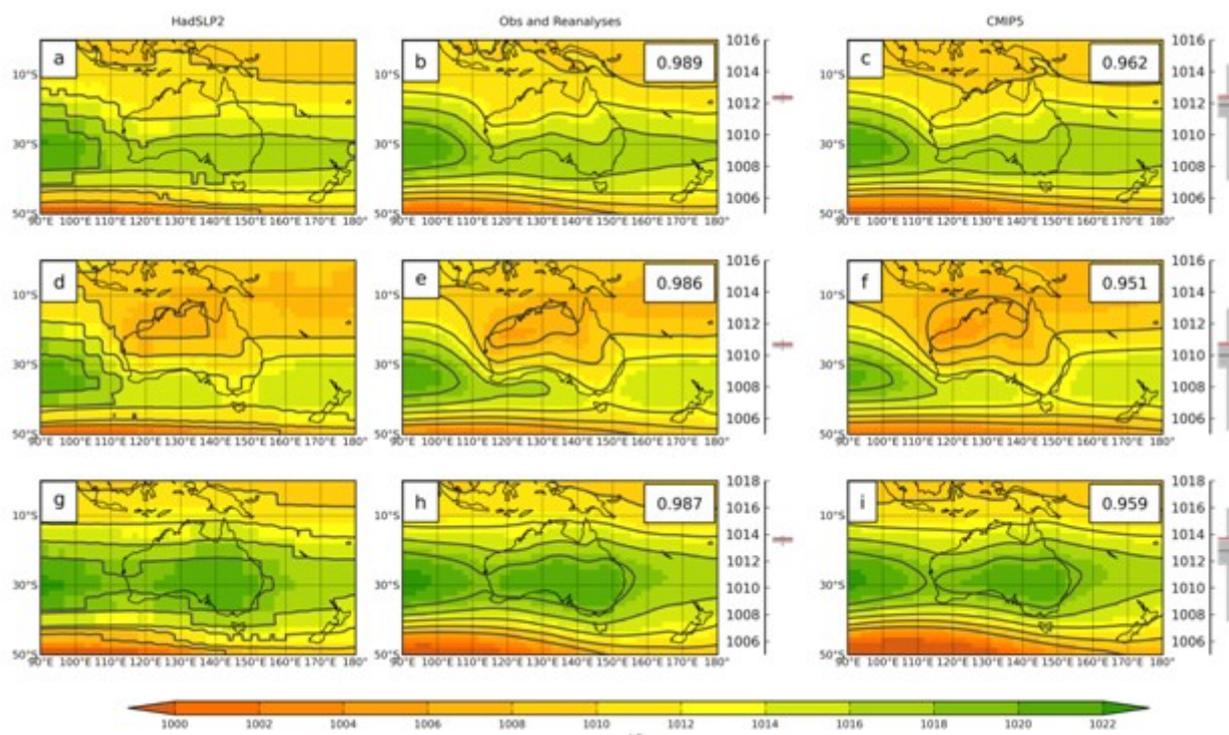


Along the tropical east coast, models show a substantial dry bias. Further south and inland, there is a general tendency for models to overestimate summer rainfall (i.e. Rangelands, Southern Slopes, Murray Darling Basin, Central Slopes) with wet biases of up to 20 mm/month (see Figure 5). Further south (Tasmania), the model biases are reversed with strong dry biases of around 20 mm/month for the entire region.

During winter (e.g. the main rainfall period for southern clusters), the model ensemble shows a dry bias over most of the higher rainfall regions (all Southern Slopes and Murray Basin clusters), except for the East Coast cluster where the median model rainfall is a good match to the observed AWAP rainfall (Figure 5). The dry bias is particularly large in Tasmania where almost all models underestimate winter rainfall. For the large Rangelands cluster, winter rainfall is slightly overestimated. The models capture lower rainfall totals along the tropical regions (Wet Tropics and Monsoonal North) well and also the higher winter rainfall with the East Coast cluster. Dry biases are common in small mountainous regions (such as the Flinders Ranges in South Australia), which is likely due to model resolution being insufficient to simulate local orographic enhancement of rainfall (Figure 5).

Figure 7 shows the comparison of annual and seasonal climatologies of mean sea level pressure for the wider region around Australia from observations and the ensemble mean. The middle and bottom rows display the shift between summer and winter pressure climatologies. During summer, the monsoonal low over north-west Western Australia dominates with high pressure systems pushed south of the continent. During winter, the high pressure system over the continent dominates. On average, the CMIP5 models capture these patterns very well (high spatial correlations) but the heat low during summer across the ‘top end’ is too deep and broad. The model spread is several hectopascals (hPa) either side of the mean sea level pressure.

Figure 7 Climatological mean sea level pressure from HadSLP2 (left column, used as the reference data set), the average of a selection of 7 other observational data sets (middle column, see Table 1) and the CMIP5 mean model (right column, see Table 2) for annual (top row), summer (Dec-Feb, middle row) and winter (Jun-Aug, bottom row) mean sea level pressure. The averaging period is 1986–2005 and the units are hectopascals (hPa). The contours highlight several thresholds for better comparison. The number in the top right corner indicates the spatial correlation between the corresponding data and ERA-Interim (with values closer to "1" indicating a better correlation). The spread in the data sets is indicated by the box-whisker to the right of each subplot: each shows the Australia-averaged mean sea level pressure where the grey box refers to the middle 50 % of the data and the whiskers show the spread from minimum to maximum. The thick black line is the median of the underlying data and the red line is HadSLP2.



Assessment of spatial structure of historical mean climatologies: M-scores for rainfall and temperature

The correct representation of climatological seasonal rainfall is a very important test for climate models. Questions such as how well the models capture the southward extent of the monsoon are a typical example addressing this issue. Similarly important and somewhat related is the representation of temperature distribution across Australia. There are several methods that can be used to evaluate spatial characteristics from climate models. Here we applied the M-Statistic (Watterson, 1996) which has also been used for the previous *Climate Change in Australia* projections (CSIRO and BoM, 2007). The *M* statistic or skill score is used as a metric for agreement between simulated and observed climatological fields over a particular region.

Two recent studies have made use of skill scores based on the M statistic for seasonal climatologies of selected climatic variables. Watterson et al. (2013a) used a simple test for overall skill in basic surface climate (calculating M-scores for a combination of temperature, rainfall and mean sea level pressure for each model) and Watterson et al. (2013b) applied tests of various features of climate (such as the sub-tropical jet).

The calculations were done for the super-cluster regions as well as the entire continent and the overall average of the M-scores (for three variables and four seasons) for each region and each model are given in Table 3. The score is out of a maximum of 1000. All CMIP5 models show an M-Score of over 500/1000 for the Australian domain, but the scores tend to be lower in smaller regions that have less spatial variation. The top scoring model for the full Australian region (AUS) is ACCESS1.0, but others do best for other regions.

Given the continuing use and validity of CMIP3 results, there is interest in how the two ensembles compare. Here we compare the results for 24 CMIP3 models given in Watterson et al. (2013a). We can see that the top results are a little lower in CMIP3 than in CMIP5, with differences from 14 to 111 points (Table 3). The means show a consistent, and larger, improvement for CMIP5 compared to CMIP3, by 57 points for AUS. In fact, several CMIP3 models have poor scores, lowering the CMIP3 mean considerably.

The best performing CMIP5 models on these scores are: ACCESS1-0, bcc-csm1-1-m, EC-EARTH, HadGEM2-ES, MPI-ESM-LR and MPI-ESM-MR. The worst performing models are BNU-ESM, CESM1-WACCM, CMCC-CESM, GISS-E2-H, GISS-E2-H-CC, MIROC-ESM and MIROC-ESM-CHEM.

Table 3: Overall skill scores for 40 CMIP5 models over five Australian domains. The values are the average M score, times 1000, for temperature, rainfall and mean sea level pressure, and the four seasons. The top values are highlighted in red and lowest three values in blue. Also shown are the overall averages and top model score for the CMIP5 ensemble as well as for CMIP3 for comparison.

<i>Model</i>	<i>AUS</i>	<i>SA</i>	<i>EA</i>	<i>NA</i>	<i>R</i>
ACCESS1-0	727	575	514	540	677
ACCESS1-3	691	492	463	532	583
bcc-csm1-1	684	464	447	513	604
bcc-csm1-1-m	711	573	490	525	611
BNU-ESM	564	388	260	400	462
CanESM2	706	542	447	544	616
CCSM4	642	519	429	492	533
CESM1-BGC	653	518	471	488	543
CESM1-CAM5	659	589	640	475	511
CESM1-WACCM	555	360	429	410	442
CMCC-CESM	549	355	240	283	479
CMCC-CM	663	583	416	532	554
CMCC-CMS	672	471	408	553	568
CNRM-CM5	706	587	450	537	584
CSIRO-Mk3-6-0	613	431	362	467	500
EC-EARTH	711	636	569	499	587
FGOALS-g2	653	518	398	417	551
FIO-ESM	641	480	347	451	546
GFDL-CM3	676	546	571	465	542
GFDL-ESM2G	638	467	499	389	527
GFDL-ESM2M	607	383	396	393	515
GISS-E2-H	586	458	426	358	432
GISS-E2-H-CC	581	473	405	344	430
GISS-E2-R	575	516	406	350	445
GISS-E2-R-CC	614	543	459	394	477
HadGEM2-AO	711	499	499	541	634
HadGEM2-CC	698	533	472	538	628
HadGEM2-ES	720	556	506	554	674
inmcm4	657	455	423	434	569
IPSL-CM5A-LR	581	395	299	512	532
IPSL-CM5A-MR	612	477	360	556	527
IPSL-CM5B-LR	625	424	307	498	569
MIROC5	644	488	431	499	521
MIROC-ESM	549	434	321	379	451
MIROC-ESM-CHEM	561	450	344	386	456
MPI-ESM-LR	720	542	520	567	650
MPI-ESM-MR	705	513	513	587	625
MRI-CGCM3	659	511	482	434	559
NorESM1-M	604	480	471	368	505
NorESM1-ME	594	475	455	362	488
CMIP5 Average	643	492	434	464	543
CMIP3 Average	586	442	383	407	473
CMIP5 Top model	727	636	640	587	677
CMIP3 Top model	706	587	529	573	625

Figure 8 Average annual cycles of surface air temperature for Australia (top left) and selected regions (North Australia - NA, Rangelands - R, Southern Slopes – SS, Southern Australia - SA, and East Australia - EA) from CMIP5 models. Each grey line represents a model simulation, the black line being the ensemble mean and observations (AWAP) shown as a brown line. The averaging period is (1986-2005)

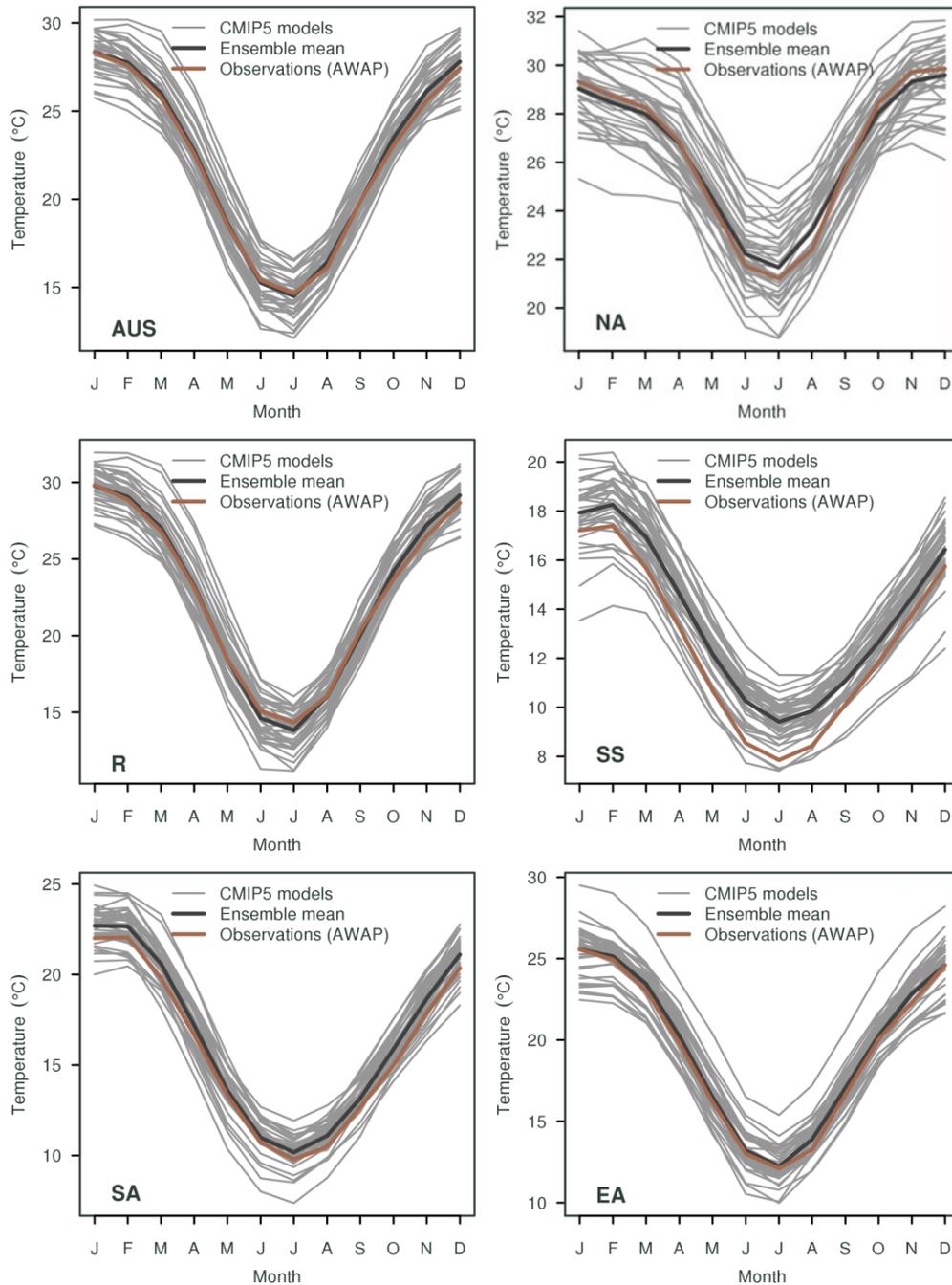
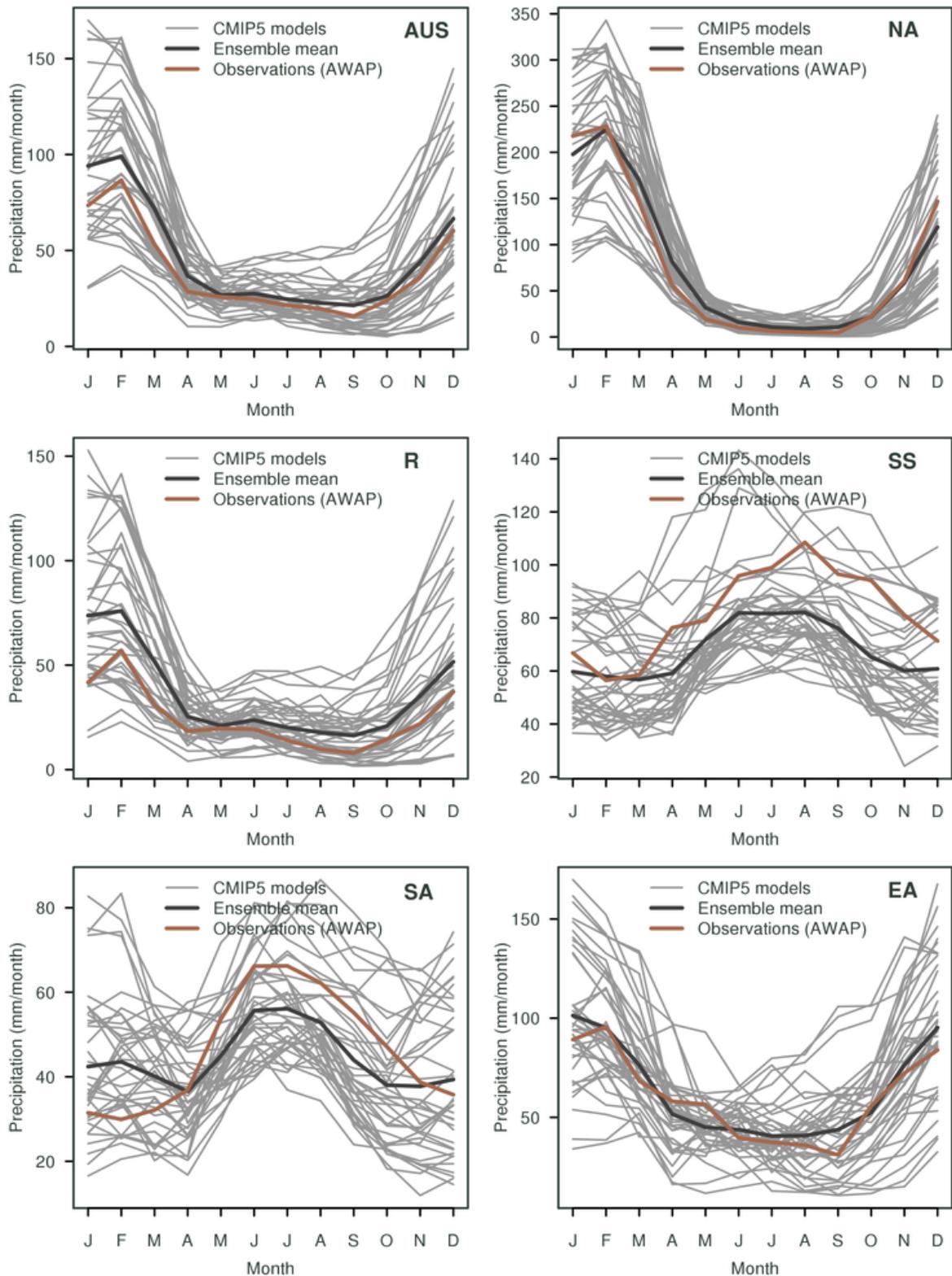


Figure 9 Average annual cycles of rainfall for Australia (top left) and selected regions (North Australia - NA, Rangelands - R, Southern Slopes – SS, Southern Australia - SA, and East Australia - EA) from CMIP5 models. Each grey line represents a model simulation, the black line being the ensemble model mean and observations (AWAP) shown as a brown line. The averaging period is (1986-2005).



Assessment of annual cycles of historical climate: rainfall and temperature

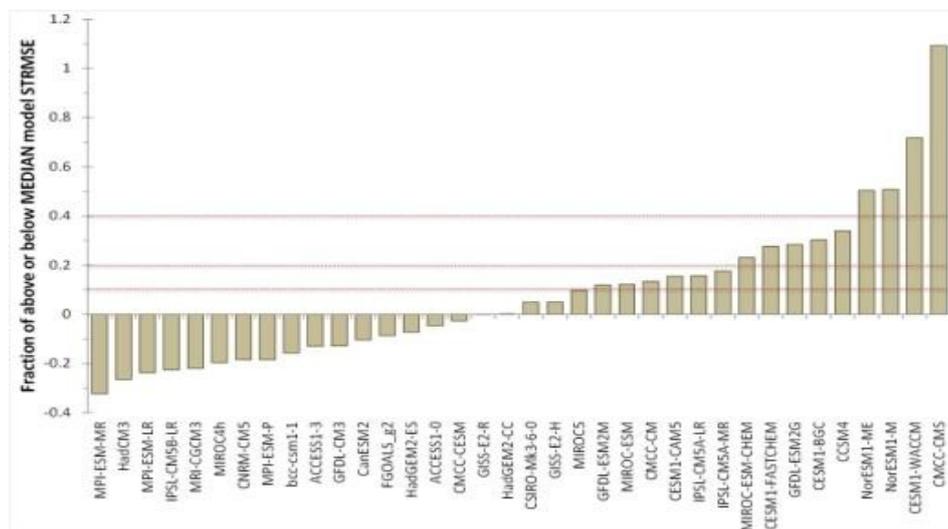
The annual cycle is one of the main climate features which, particularly for temperature and rainfall, is a crucial element to simulate correctly. Figure 8 shows the surface air temperature annual cycles for Australia and the super-cluster regions. In general there is very good agreement, but with some model spread around the mean. For most regions (and months) the multi-model average is within half a degree of the AWAP value. One exception is the Southern Slopes cluster (Figure 8 bottom left) where models are too warm year round, but most of the differences there are related to biases over Tasmania (which is often poorly resolved) and not over the Victorian region within Southern Slopes.

Figure 9 shows the corresponding rainfall annual cycles for the same regions. As mentioned earlier, because of the small-scale processes involved in rainfall simulation, it is more difficult to correctly simulate rainfall, particularly across small regions such as some of the clusters.

Regions with a pronounced annual rainfall cycle, such as monsoon dominated Northern Australia, show good model skill with the multi-model average matching the AWAP cycle – albeit with large inter-model spread. Other regions show more varying model skill and while the average might still be close to AWAP there is significant departure by some models. In the example of a fairly "flat" annual rainfall cycle (Southern Slopes, Figure 9 bottom left), some models show even a reversed annual cycle (for example the NorESM1-ME model).

The spatial-temporal root-mean-square-error (STRMSE) is used as a skill measure for the 1986–2005 annual-average rainfall cycle (following Gleckler et al. 2008). It combines spatial deviations from observed patterns for each month, thereby reflecting also the skill of simulating the annual cycle. This error measure is portrayed in Figure 10 as a relative error by normalizing the result by the median error of all model results. For example, a value of 0.20 indicates that a model's STRMSE is 20% larger than the median CMIP5 error for that variable, whereas a value of -0.20 means the error is 20% smaller than the median error. For Australia, the median STRMSE for the CMIP5 models is close to 1 mm/day. The group of models that show significantly lower STRMSE values for rainfall are: MPI-ESM-MR, HadCM3, MPI-ESM-LR, IPSL-CM5B-LR, MRI-CGCM3, MIROC4h, CNRM-CMS, CNRM-CMS, MPI-ESM-P, bcc-csm1-1, ACCESS1-3, GFDL-CM3, CanESM2, FGOALS_g2, HadGEM2-ES, ACCESS1-0, CMCC-CESM, GISS-E2-R, HadGEM2-CC, CSIRO-Mk3-6-0, GISS-E2-H, MIROC5, GFDL-ESM2M, MIROC-ESM, MIROC-CM, CMCC-CM, CESM1-CAM5, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC-ESM-CHEM, CESM1-FASTCHEM, GFDL-ESM2G, CESM1-BGC, CCSM4, NorESM1-ME, NorESM1-M, CESM1-WACCM and CMCC-CMS.

Figure 10 Space-time root mean square error of CMIP5 models rainfall across Australia. The error is scaled to show the fractional error higher or lower than the median model error. Values below zero indicate better than median error and values above zero higher than median error. The red dotted lines display thresholds of 10%, 20% and 40% above the median error (1.13 mm/day).



Assessment of additional climate features and associated skill scores

Here we present a review of previous studies that have examined the simulation of relevant climate features that impact Australia, and present a limited amount of targeted new analysis that complements this review. For a more detailed description of the climate features, see for example Chapter 4 in CSIRO and Bureau of Meteorology (2015).

As part of the UK Met Office's CAPTIVATE project (Scaife et al. 2011; stands for Climate Processes, Variability and Teleconnections), an evaluation of simulated Australian climate features was used. The tests were initially applied to the three Australian CMIP5 models (ACCESS 1.0, ACCESS 1.3 and CSIRO Mk3.6), with the results described in Watterson et al. (2013b). Here we consider only the tests for climatological features, which have been somewhat modified to suit the available data and NRM interests.

Again, the M statistic is used to quantify the agreement between each model and the observations, in each season. The variable and domains depend on the test, as outlined in Table 4. The variables surface air temperature (tas) and precipitation (pr), without being averaged as was done in Table 3, are tested and the domain is that of the Australian land area. The domain for the variable sea level pressure (psl) is over the larger region as described in Table 4 in order to capture the pressure systems extending past the continent. Also tested over Australian land are incoming solar radiation (rsds) and the diurnal temperature range (DTR; using maximum and minimum temperature). Data for DTR are missing for 3 models (CMCC-CM, MPI-SEM-MR and NorESM1-ME).

The diurnal temperature range is an important indicator for models' representation of extreme cold and warm temperatures, and therefore contributes to the skill in representing temperature extremes. Table 5 shows a very large spread along M-skill scores for DTR in CMIP5 models (from 94 to 496) which indicates that (a) the simulation of DTR is the least skilful of all features listed in Table 4 and (b) the spread of skill is largest amongst the CMIP5 models for DTR compared to the other features.

The ERA-Interim data set is again used as representing the observations for rsds (although given some doubt about its representation of cloud cover, the rsds fields, and the resulting scores may not be reliable – see Naud et al. 2014). For DTR, the AWAP data set is used.

The four other tests use wind data in zonal (east-west) and meridional (north-south) direction, which are available from 20 of the 40 models. The tests for wind at 850 hPa and 200 hPa are over the larger region (see Table 4) and are representative measures for skill in capturing the larger atmospheric circulation both closer to the surface (850 hPa) and further aloft (200 hPa). The tests for 'Subtropical Jet' and 'Monsoon Onset' are very simplified tests of winds over smaller rectangular regions (see Table 4 for domain details). Wind data from ERA-Interim are used as representing the observations.

Table 4 Overview of tests for nine features of Australian climate, with variables and domain given. All tests are done across the four seasons, except for 'Monsoon Onset', which is over SON and DJF only.

<i>Feature</i>	<i>Fields - CMIP5 name</i>	<i>Domain</i>
1.5m Temperature	tas	Australia – land only
Rainfall	pr	Australia – land only
Solar radiation	rsds	Australia – land only
Diurnal Temperature Range	DTR	Australia – land only
Zonal and meridional winds at 850 hPa height	ua, va 850hPa	Region (longitude: 105 °E-165 °E; latitude: 0 °S-50 °S)
Zonal and meridional winds at 200 hPa height	ua, va 200hPa	Region (longitude: 105 °E-165 °E; latitude: 0 °S-50 °S)
Sea Level Pressure	psl	Region (longitude: 105 °E-165 °E; latitude: 0 °S-40 °S)
Sub-tropical Jet	ua, 850, 500, 200 hPa	East Australia (longitude: 140 °E-150 °E; latitude: 15 °S-40 °S)
Monsoon Onset.	ua, va 1000 hPa ua 850hPa	North Australia (longitude: 120 °E-150 °E; latitude: 10 °S-20 °S)

The scores for the nine tests are given in Table 5. Quantities with larger spatial variation tend to have smaller scores, in particular DTR. Even the best score for DTR, from ACCESS1.0, is only 496.

The top six models (averaged over the nine tests) are: ACCESS-1.0, CMCC-CM, CNRM-CM5, HadGEM2-CC, HadGEM2-ES, and MPI-ESM-MR. Note that three out of these six have the same atmosphere model. The worst performing models are: BNU-ESM, CESM1-WACCM, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC, IPSL-CM5A-LR, MIROC-ESM and MIROC-ESM-CHEM.

The main purpose of the scores is to support the NRM project by providing information about the quality of the models being used for projections. Naturally, these tests are only for the climate of the past decades, and the link between such skills and the reliability of climate changes is not well established. Nevertheless, skill in simulating the features of climate through the four seasons can add confidence in a model's ability to simulate changes that follow from global warming (Whetton et al. 2007). This confidence is part of the overall assessment of projected changes in Australia's climate. The scores for the nine features add to the information available for assessment. It is important to note that both versions of ACCESS are well ranked in most of these tests.

Table 5 Skill scores for 40 CMIP5 models for nine features of Australian climate. The values are the average M score, times 1000. The top values are highlighted in red and lowest values in blue. Where there was some missing data, the score couldn't be calculated and are indicated by 'xxx'. Variables are: TAS mean surface temperature, PR precipitation, RSDS solar radiation, DTR diurnal temperature range, W850 wind at the 850 hPa level, W200 wind at the 200 hPa level, psl mean sea level pressure, STJ subtropical jet, MO monsoon onset.

<i>Model</i>	<i>TAS</i>	<i>PR</i>	<i>RSDS</i>	<i>DTR</i>	<i>W850</i>	<i>W200</i>	<i>psl</i>	<i>STJ</i>	<i>MO</i>
ACCESS1-0	832	552	604	496	760	750	834	798	645
ACCESS1-3	792	544	606	198	690	678	798	738	496
bcc-csm1-1	780	499	699	295	657	684	716	687	543
bcc-csm1-1-m	766	525	744	365	xxx	xxx	811	xxx	Xxx
BNU-ESM	755	451	534	120	xxx	xxx	615	xxx	Xxx
CanESM2	824	492	705	426	717	718	812	712	500
CCSM4	816	379	602	172	720	758	802	744	611
CESM1-BGC	824	400	645	184	xxx	xxx	801	xxx	Xxx
CESM1-CAM5	806	493	544	188	xxx	xxx	815	xxx	Xxx
CESM1-WACCM	743	281	337	94	xxx	xxx	673	xxx	Xxx
CMCC-CESM	641	479	644	289	xxx	xxx	481	xxx	Xxx
CMCC-CM	794	486	698	xxx	xxx	xxx	757	xxx	Xxx
CMCC-CMS	729	564	725	358	xxx	xxx	673	xxx	Xxx
CNRM-CM5	742	602	770	485	xxx	xxx	863	xxx	Xxx
CSIRO-Mk3-6-0	744	482	601	400	691	666	657	647	658
EC-EARTH	687	701	xxx	315	xxx	xxx	765	xxx	Xxx
FGOALS-g2	755	535	725	235	667	737	586	625	624
FIO-ESM	817	424	705	141	xxx	xxx	636	xxx	Xxx
GFDL-CM3	781	564	790	172	741	724	731	623	653
GFDL-ESM2G	716	472	617	122	712	724	798	771	593
GFDL-ESM2M	728	469	630	118	745	740	731	726	589
GISS-E2-H	661	490	271	228	662	647	738	748	486
GISS-E2-H-CC	610	501	269	181	xxx	xxx	769	xxx	Xxx
GISS-E2-R	651	461	286	272	xxx	xxx	760	xxx	Xxx
GISS-E2-R-CC	731	472	279	265	xxx	xxx	779	xxx	Xxx
HadGEM2-AO	808	600	644	496	xxx	xxx	797	xxx	Xxx
HadGEM2-CC	800	541	723	474	737	718	782	781	638
HadGEM2-ES	807	561	715	457	730	735	801	744	602
inmcm4	681	524	730	290	657	683	815	635	439
IPSL-CM5A-LR	796	403	414	118	622	659	507	473	390
IPSL-CM5A-MR	825	404	406	100	674	688	612	531	446
IPSL-CM5B-LR	760	596	519	128	xxx	xxx	559	xxx	Xxx
MIROC5	793	432	805	338	xxx	xxx	778	xxx	Xxx
MIROC-ESM	790	342	710	271	519	561	488	552	319
MIROC-ESM-CHEM	790	333	695	265	517	574	516	560	300
MPI-ESM-LR	830	593	812	232	xxx	xxx	743	xxx	Xxx
MPI-ESM-MR	808	640	799	xxx	xxx	xxx	704	xxx	Xxx
MRI-CGCM3	726	599	652	350	xxx	xxx	743	xxx	Xxx
NorESM1-M	730	347	558	162	699	699	779	774	627
NorESM1-ME	724	343	559	xxx	676	699	752	785	623

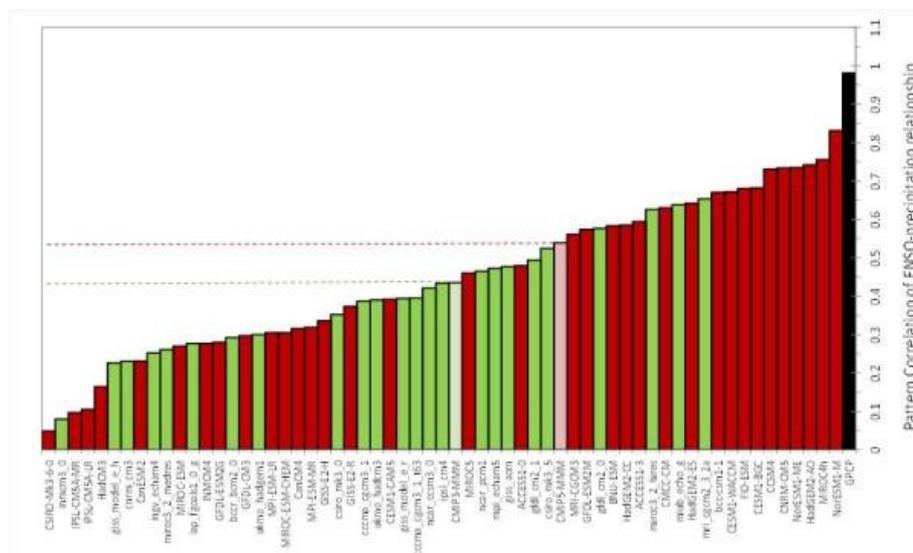
The El Niño-Southern Oscillation

The El Niño-Southern Oscillation (ENSO) phenomenon is the dominant driver of climate variability on seasonal to interannual time scales for Australia (see for example Risbey et al. 2009b; Wang et al. 2004a; Wang & Hendon 2007; Cai et al. 2011). While there has been an improvement in the simulation of ENSO in climate models from CMIP3 to CMIP5 (see for example Guilyardi et al. 2009; Chapter 14 in IPCC, 2013), some systematic errors remain and impact to some extent on the simulation of the relationship between ENSO and Australian rainfall (Watanabe et al. 2011; Weller and Cai 2013a). However, there are improvements in the multi-model mean which is mostly due to a reduced number of poor-performing models (Flato et al. 2013).

The ENSO-rainfall teleconnection involves mechanisms similar to those related to the rainfall response to global warming (Neelin et al. 2003) and therefore provides a valuable insight into each model's rainfall response. While CMIP5 models display a slightly better skill in Australian rainfall reductions associated with El Niño (Neelin 2007; Cai et al. 2009; Coelho and Goddard 2009; Langenbrunner and Neelin 2013), there is not much additional improvement over CMIP3. There is also little change in their abilities to represent the correlations between the equatorial Pacific sea surface temperatures (Niño 3.4 region) and north Australian sea surface temperatures (Catto et al. 2012a, 2012b) with models failing to adequately capture the strength of the negative correlations during the second half of the year. In general, the evolution of sea surface temperatures in the north Australian region during El Niño and La Niña is still problematic for models to simulate.

The teleconnection patterns from ENSO to rainfall over Australia are reasonably well simulated in the key September-November season (Cai et al. 2009; Weller and Cai 2013b) in the CMIP3 and CMIP5 multi-model mean. Figure 11 shows the ranked list of the skill of this relationship in both CMIP3 and CMIP5 models. While there is clearly a majority of CMIP5 models towards the more skilful end of the list, there are a few CMIP5 models showing very little correlation (CSIRO-Mk3-6-0, IPSL-CM5A-MR, IPSL-CM5A-LR, HadCM3) or only small correlation (CanESM2, MIROC-ESM, INMCM4, and GFDL-ESM2G). These results are similar to those found previously for CMIP3 (Cai et al. 2009) and more recently for CMIP5 (Jourdain et al. 2013).

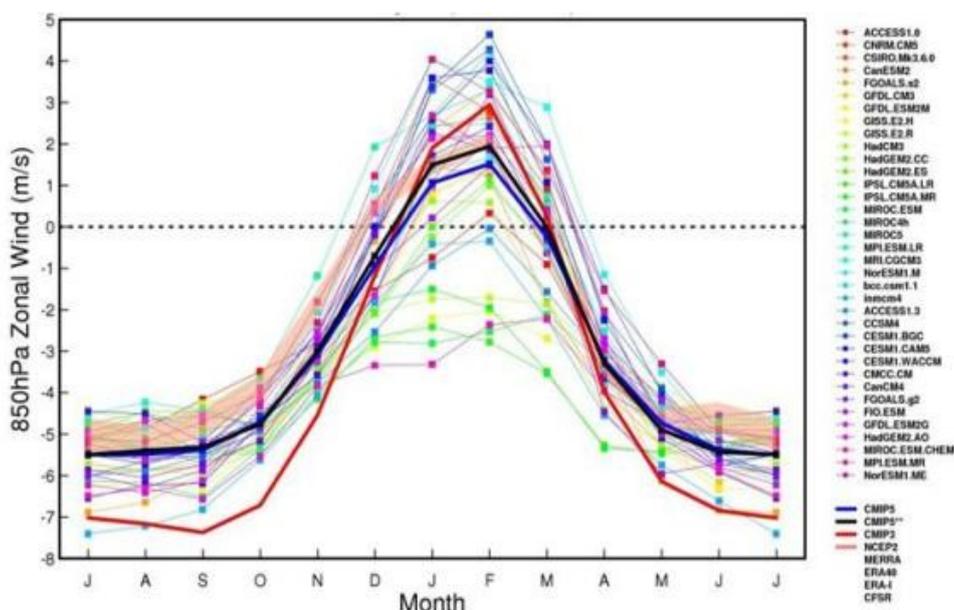
Figure 11 Pattern correlations (against the CMAP-HadISST reference pattern) of the ENSO-Australian-rainfall teleconnection pattern for each CMIP5 (red) and CMIP3 model (green). A second observed data set (GPCP-HadISST, black bar) is shown on the right while the models are ordered in increasing skill towards the right. The ensemble mean of the models are shown in pink (for CMIP5) and light green (for CMIP3).



The Australian monsoon

The Australian monsoon is the main driver of annual variation in the tropical regions (Trenberth et al. 2000; Wang and Ding 2008; Moise et al. 2012) and therefore is an important feature for climate models to correctly simulate. This will also enhance confidence in future projections of mean changes and associated impacts (Colman et al. 2011).

Figure 12 Monthly seasonal climatology of 850 hPa zonal wind (1986-2005), averaged over 120-150°E, 10-20°S land only for 37 CMIP5 models, 5 reanalysis products (CFSR, MERRA, NCEP2, ERA40 and ERA-INT; forming the pink shaded band) and the ensemble mean of the models of CMIP3 (red) and CMIP5 (blue). The thick black line represents the ensemble mean of CMIP5 excluding models not simulating monsoon westerlies over this region.



The Australian monsoon is characterised by an annual reversal of the low level winds and well defined dry and wet seasons (Moise et al. 2012; Wang and Ding 2008), and its variability is primarily connected to the Madden-Julian Oscillation (MJO) and ENSO. Most CMIP3 models poorly represent the characteristics of the global monsoons and monsoon teleconnections (Randall et al. 2007), with some improvement in CMIP5 with respect to the mean climate, seasonal cycle, intraseasonal, and interannual variability (Sperber et al. 2013; also see Figure 9 for Northern Australia). Figure 12 shows the annual cycle of low level zonal winds for CMIP5 models and several reanalysis data sets for the Australian monsoon. On average the models reversal to westerlies starts later than in the reanalysis (December) but has a similar timing in the switch to easterlies in March. Several models fail to simulate monsoon westerlies over northern Australia altogether: GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC-ESM, INMCM4, ACCESS1-3 and MIROC-ESM-CHEM.

While the entire annual rainfall cycle has been assessed earlier using the spatial-temporal root-mean-square error (STRMSE, see Figure 10) here we focus on the wet season only and assess the spatial distribution of wet season rainfall from the models over the tropical Australia domain. Figure 13 shows the ranked list of the skill of Australian tropical rainfall distribution in both CMIP3 and CMIP5 models. While there is clearly a majority of CMIP5 models towards the more skillful end of the list, there are a few CMIP5 models showing very little skill (MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P) or only small skill (GFDL-ESM2G, MIROC5, and HadCM3).

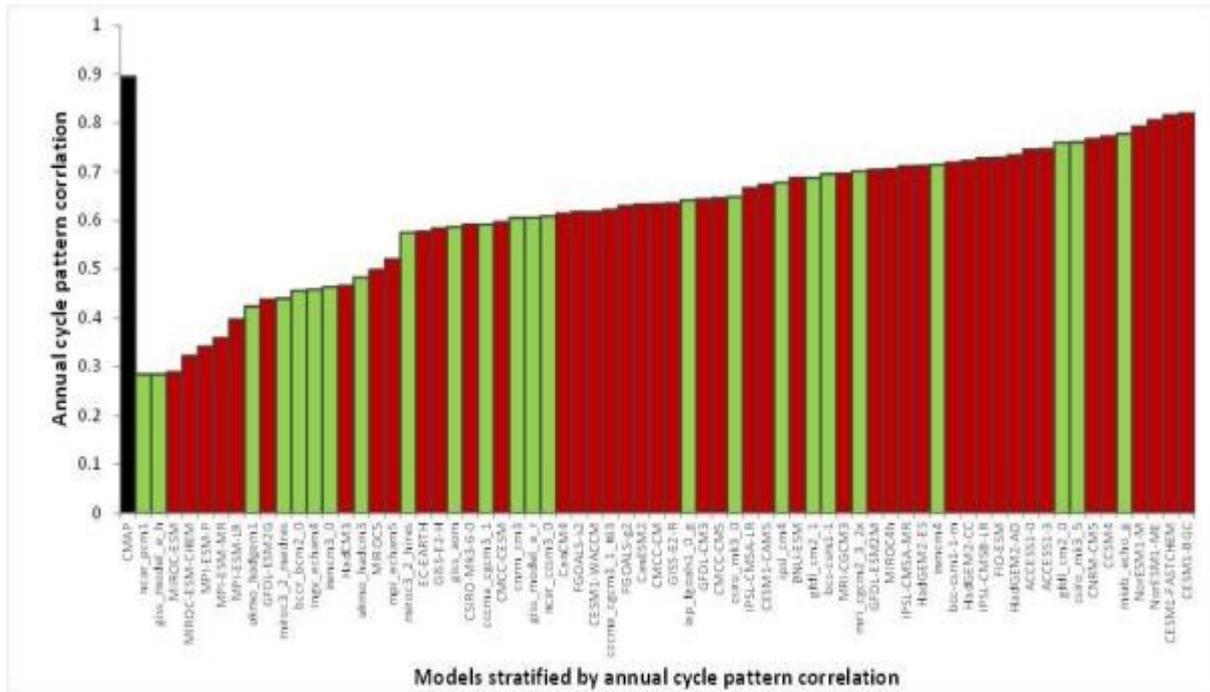
With respect to the onset of the Australian monsoon, Table 5 also includes the M-score for skill in monsoon onset for CMIP5 models. While better models reach a score above 600, several models score below 400 and both MIROC-ESM and MIROC-ESM-CHEM are close to 300.

Atmospheric blocking

The climate along the Australian mid-latitudes is predominantly affected by weather regimes such as west-east moving pressure systems or east coast lows, and blocking weather regimes are often associated with extreme rainfall events (Risbey et al. 2009a). During blocking, the prevailing mid-latitude westerly winds and storm systems are interrupted by a local reversal of the zonal flow resulting in enhanced rainfall events. The strongest correlation between blocking (using a blocking index) and Australian rainfall is during autumn but also in winter. It affects mainly south-eastern regions of the continent (Risbey et al. 2009a).

Climate models in the past have universally underestimated the occurrence of blocking. As in CMIP3, most of the CMIP5 models globally still significantly underestimate blocking (Dunn-Sigouin and Son, 2013). Increasing model resolution is expected to improve model representation of blocking significantly (IPCC, 2013, Chapter 14).

Figure 13 Wet season (Nov-Feb) rainfall pattern correlations (against GPCP reference data set) for CMIP5 (red) and CMIP3 models (green) over tropical Australia. A second observed data set (CMAP, black bar) is shown on the left while the models are ordered in increasing skill towards the right.



During atmospheric blocking the upper tropospheric westerly air stream typically splits into two sections. The strength of this split can be assessed through a combination of the upper air zonal wind field (at 500hPa) at different latitudes integrated into a simple *Blocking Index* (BI, Pook and Gibson 1999 see also Risbey et al. 2009a; Grose et al. 2012):

$$BI = 0.5(ua_{25} + ua_{30} - ua_{40} - 2ua_{45} - ua_{50} + ua_{55} + ua_{60})$$

Where ua_x is the zonal wind at 500hPa at latitude x (degrees south). The BI is calculated here at longitude 140°E which represents the region over Australia where blocking is typically observed. The CMIP5 models were evaluated with respect to the seasonal correlations of the Blocking Index in autumn and winter to rainfall across relevant south-eastern cluster regions (Central Slopes, East Coast South, Murray Basin, Southern Slopes) (results not shown). Almost half of the models assessed showed reasonable correlations across several clusters. Models that showed very low skill in reproducing this relationship include ACCESS1-3, CanCM4, GFDL-ESM2G, GISS-E2-R and GISS-E2-H.

Southern Annular Mode

The Southern Annular Mode (SAM) is the most dominant driver for large-scale climate variability in the mid- and high-latitudes of the Southern Hemisphere (Thompson and Solomon, 2002) – describing the alternation of atmospheric mass between high- and mid-latitudes. This alternation affects pressure and wind patterns across southern parts of Australia and therefore also impacts on rainfall in these regions (for more detail, see Hendon et al. 2007; Risbey et al. 2009b). When SAM is in its high phase there are higher pressures over southern Australia, wind anomalies are predominantly easterly and rainfall is reduced on west-facing coastlines but enhanced on east-facing regions.

CMIP3 and CMIP5 models are able to produce a clear Southern Annular Mode (Raphael and Holland, 2006; Zheng et al. 2013; Barnes and Polvani 2013) but there are relatively large differences between models in terms of the exact shape and orientation of this pattern.

The Indian Ocean Dipole

Similar to ENSO, the Indian Ocean dipole mode (IOD) is an ocean-atmosphere phenomenon located in the tropical Indian Ocean. The main period of impact on Australian rainfall is spring (Sep-Nov) and depending on the phase of the IOD, the ENSO impact can be enhanced over

Australia. If the IOD is in its positive phase, El Niños can result in stronger reduction of rainfall and if the IOD is in its negative phase, La Niñas show further enhanced rainfall (Risbey et al. 2009b).

Most CMIP3 and CMIP5 models are able to reproduce the general features of the IOD but show a large spread in the strength of the IOD (Saji et al. 2006; Liu et al. 2011; Cai and Cowan 2013). Most models also show a location bias in the westward extension of the IOD. No substantial improvement is seen in CMIP5 compared to CMIP3 (Weller and Cai 2013a).

A majority of CMIP3 and CMIP5 models also simulate the observed correlation between IOD and ENSO. The magnitude of this correlation varies substantially between models, but seems independent of each model’s simulation of ENSO (Saji et al. 2006; Jourdain et al. 2013).

The teleconnection patterns from both ENSO and IOD to precipitation over Australia are reasonably well simulated in the key September–November season (Cai et al. 2009; Weller and Cai 2013b) in the CMIP3 and CMIP5 multi-model mean.

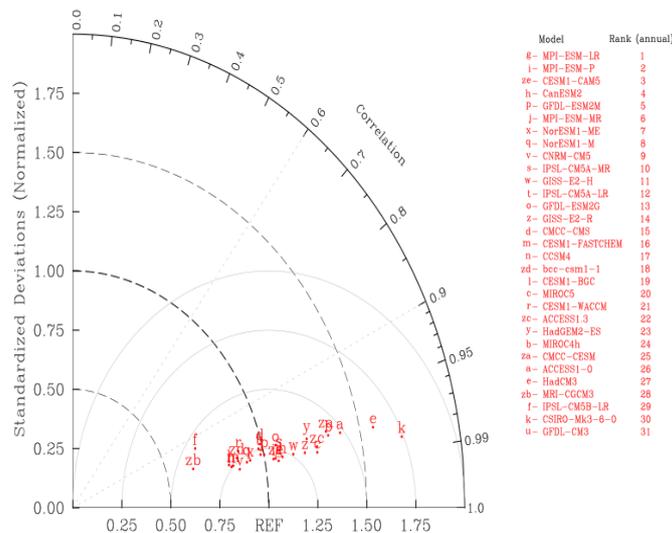
An additional way to assess the spatial structure of the IOD is by computing the Taylor statistics (Taylor, 2001) of the tropical Indian Ocean sea surface temperatures where the IOD occurs as shown in Figure 14. These statistics (spatial correlation; spatial root-mean-square error and spatial standard deviation) can highlight non-temporal deficiencies in the simulation of this feature. Most CMIP5 models simulated very high spatial correlations. Combining the statistics into a skill score as proposed by Taylor (2001) we find that while most CMIP5 models show very high spatial correlations (above 0.95), the main difference between more skilful and less skilful models lies in their simulation of the spatial variability of sea surface temperatures (horizontal spread of letters in Figure 14). In particular, MRI-CGCM3 and CSIRO-MK3-6-0 have a much reduced variability and GFDL-CM3, IPSL-CM5B-LR, ACCESS1-0 and HadCM3 show a far too strong variability (furthest right from the reference dashed line). This is similar to previous results from other studies, such as Jourdain et al. (2013).

The Madden-Julian Oscillation

During summer the eastward propagating feature of enhanced and diminished convection from the Indian Ocean into the western Pacific known as the Madden-Julian Oscillation (MJO; (Madden and Julian, 1972, 1994)) mainly affects the tropics north of 15 °S. It is one of the dominating features of intra-seasonal variability (60-90 days) and plays a major role in the onset of the Australian monsoon (Wheeler et al. 2009).

Various diagnostics have been used to assess the skill of simulating the MJO in climate models (Waliser et al. 2009; Xavier 2012). The main model errors in representing the MJO relate to the skill in the model convection schemes and their mean state biases (Kim et al. 2012; Mizuta et al. 2012; Inness et al. 2003).

Figure 14 Taylor plot of spatial statistic of sea surface temperatures from CMIP5 models over the tropical Eastern Indian Ocean against HadISST observed sea surface temperatures (REF point at horizontal axis). Each letter represents one CMIP5 model's simulation averaged over the period (1986-2005).



Sperber and Kim (2012) provided a simplified metric synthesising the skill of CMIP3 and CMIP5 model results. These metrics are based on lag correlation analysis of principal component time series of daily outgoing long-wave radiation. Some of the more skilful models are those with

higher resolution (CNRM-CM5, CMCC-CM) while several models showed very low coherence in the propagation of the convection: MIROC-ESM-CHEM, INM-CM4, IPSL-CM5A-MR, IPSL-CM5A-LR, MIROC-ESM, GFDL-ESM2G and HadGEM2-ES.

While Sperber and Kim (2012) show that the simulation of the MJO is still a challenge for climate models (see also Lin et al. 2006; Kim et al. 2009; Xavier et al. 2010), there has been some improvement in CMIP5 in simulating the eastward propagation of the summer MJO convection (Hung et al. 2013). Further improvements have been reported for the MJO characteristics in the Pacific (Jiang et al. 2013). In general, CMIP5 models have improved compared to previous generations of climate models with respect to the MJO (Waliser et al. 2003; Lin et al. 2006; Sperber and Annamalai 2008).

Winds and atmospheric circulation

Wind fields across Australia are associated with large scale circulation patterns and their seasonal movement. Across the southern half of Australia, average wind conditions are influenced by the seasonal movement of the sub-tropical high pressure belt (called the Sub-tropical Ridge) which separates the mid-latitude westerly winds to the south and the south-east trade winds to the north. Across the north of Australia, from about November to March the Asian-Australian monsoon interrupts the trade winds bringing a north-westerly flow across northern Australia.

The evaluation in this study of winds in climate models therefore mainly focuses on these two large scale seasonal changes: the north-south shift across the southern half of Australia and the east-west reversal of winds across tropical Australia. Due to the sparseness of long-term, high quality wind measurements from terrestrial anemometers, a high quality gridded data set for wind is not available over Australia (Jakob, 2010). Therefore 10 m winds from reanalysis products are commonly used as a baseline against which climate model winds are compared (see Table 1 for overview of reanalysis data sets). The annual cycle in the pressure and latitude of the sub-tropical high pressure belt known as the Sub-tropical Ridge (STR) is fairly well represented in the CMIP3 mean, but each model has some biases in position and intensity (Kent et al. 2013). This means there are typically some biases in the northern boundary of the westerly circulation. Also, the relationship between the STR and rainfall variability is poorly simulated in some models and trends in the pressure of the ridge are underestimated by all CMIP3 models (Kent et al. 2013; Timbal and Drosowsky 2013), and the results are similar in CMIP5 models (Grose et al. 2015).

The path of westerly weather system generally to the south of the subtropical ridge is known as the 'storm track' and is a crucial feature of rainfall variability in southern Australia. The representation of the storm track, and its connection to processes such as ENSO, has improved from CMIP3 to CMIP5 but certain models still show poor performance (Grainger et al. 2014).

Regarding the wind reversal over tropical Australia during the monsoon season, Figure 12 shows the annual cycle of low level zonal winds for CMIP5 models and several reanalysis data sets. As mentioned earlier, on average the models' reversal to westerlies starts later compared to the reanalysis but has a similar timing in the switch back to easterlies in March. As noted, several models fail to simulate monsoon westerlies over northern Australia altogether.

Evaluation of simulated rainfall and temperature trends

In addition to the long-term climatology and the annual cycle (see previous section), climate models are also evaluated with respect to how well they are able to reproduce observed climate change. Aspects of climate change have been extensively evaluated at global to continental scales and the simulated warming is found to agree well with observations (Stone et al. 2009). Changes in global precipitation, on the other hand, are less well reproduced in simulations (Zhang et al. 2007). Recently, global climate models have also been evaluated against observed regional climate change (van Oldenborgh et al. 2009, van Haren et al. 2012, van Oldenborgh et al. 2013, Bhend and Whetton, 2013).

Recent regional trends in seasonal mean daily maximum and minimum temperature and rainfall have been evaluated (Bhend and Whetton, 2013). Simulated trends in the historical experiment from the CMIP5 ensemble are compared to the observed trends. Climate models used here have been run with a comprehensive set of observed and reconstructed boundary conditions including the changing atmospheric concentrations of greenhouse gases, aerosols, and ozone as well as solar irradiance changes. The models thus produce a realistic - within model limitations - representation of recent climate change. It is important to note, however, that a portion of the observed and simulated recent change is due to natural internal variability in the climate system. This part of climate change differs between observations and simulations, as the simulations are not constrained to exhibit internal variability that is in phase with the observed internal variability. The remainder of the change – the signal – is due to changes in external forcing mechanisms and therefore in principle reproducible in long-term simulations. Only this deterministic, forced component of climate change can be used for evaluation of climate models. Therefore, being able to separate signal from noise (internal variability) is crucial when evaluating transient behaviour in climate models and a multitude of methods to achieve this exists (Bindoff et al. 2013). For simplicity, we assumed here that the regional signal in both temperature and rainfall over the period from 1956 to 2005 is approximately linear. Simulated seasonal rainfall and daily maximum and minimum temperature trends from 42 global climate models in the CMIP5 ensemble are compared with observed trends in the station-based gridded datasets. ACORN-SAT (Trewin, 2013) and CRU TS3.20

(Harris et al. 2013) were used for temperature, and AWAP (Jones et al. 2009, Raupach et al. 2009; 2012) and CRU TS3.20 for precipitation (listed in Table 1 under CRU). We compute linear trends from 1956 to 2005 using ordinary least squares regression.

The observed trends in seasonal mean daily maximum temperature from 1956 to 2005 show significant warming in eastern and southern Australia and widespread cooling (some of which is statistically significant) in the summer half-year in north-western Australia (Figure 15 a-d). The ensemble median simulated trends for the same period show consistent warming and less than 10 per cent of the simulations reproduce the cooling in spring (SON) and summer (DJF) in north-western Australia (Figure 15 e-h). While the trend biases are locally significant (at 90 per cent level based on a simple estimate of internal variability in observations and model time series) in the majority of the climate models, the area where significant differences are found is generally not larger than what one would expect due to internal variability alone. Results for trends in seasonal mean daily minimum temperatures are qualitatively similar (not shown).

Figure 15 Observed trend in seasonal mean daily maximum temperature from 1956 to 2005 (a-d) and median of simulated trends from 42 CMIP5 models (e-h). Stippling in a-d denotes areas where the observed trend is significantly different from zero at the 10% level. Crosses (dashes) in e-h denote areas where less than 10 per cent of the simulated trends are as large (small) as the observed trend.

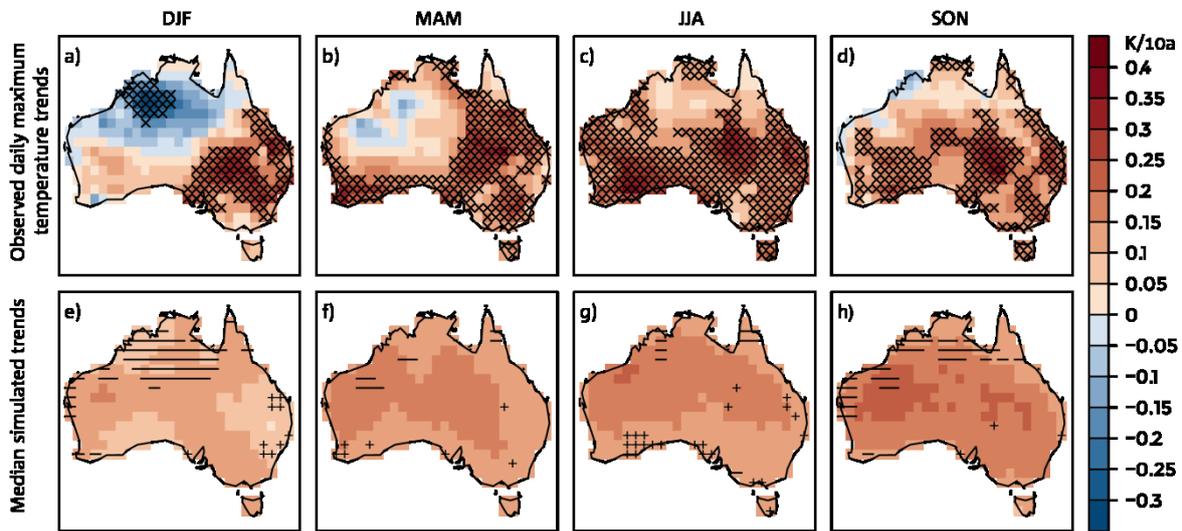
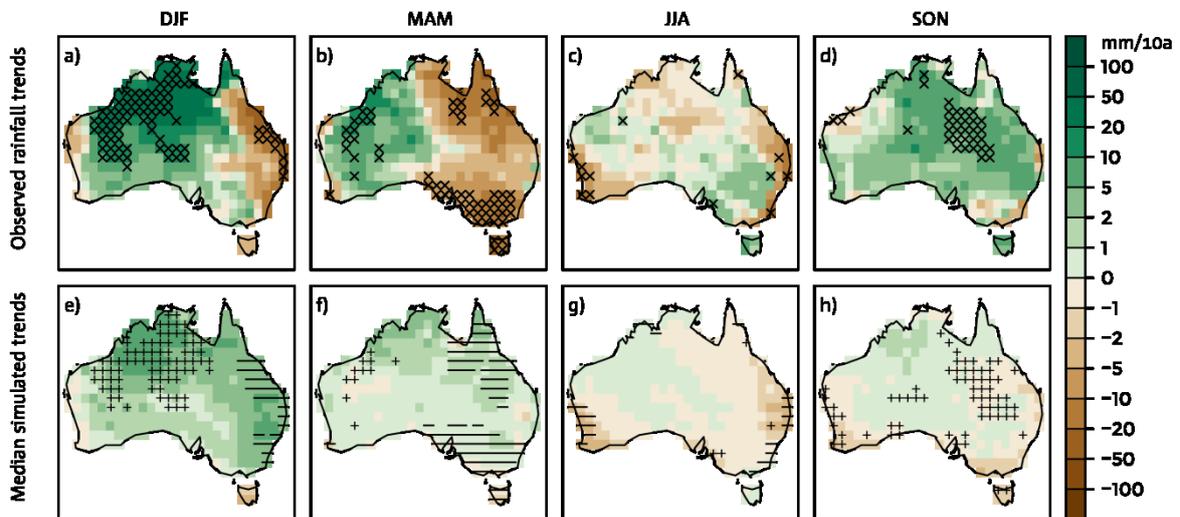


Figure 16 As in Figure 15 but for seasonal rainfall in mm per decade



The picture is similar for rainfall. The area where significant differences are found between observed and simulated rainfall trends is generally not larger than what one expects due to internal variability alone (Figure 16 a-d). Less than 10 per cent of the models reproduce the significant wetting in north-western Australia in summer (DJF), the drying in south-eastern Australia in autumn (MAM) and the wetting in north-eastern Australia in spring (SON). Additional analyses reveal that the majority of the models significantly (at the 10 per cent level) underestimate the observed wetting in north-western Australia in summer and the observed drying in south-eastern Australia in spring (not shown.) (i.e. due to random chance).

In conclusion, areas where the CMIP5 ensemble fails to reproduce observed trends from 1956-2005 in seasonal mean daily maximum and minimum temperature and seasonal rainfall are evident. The extent of the areas for which these discrepancies exist, however, is generally not larger than expected due to the pronounced variability on inter-annual to decadal scales. Therefore, there is no conclusive evidence that CMIP5 models fail to reproduce recent observed trends in daily maximum and minimum temperature and rainfall. Nevertheless, confidence in rainfall projections is inevitably reduced where consistency is low, particularly north-western Australia in summer and south-eastern Australia in autumn.

Evaluation of extremes in climate models

Extreme events refer to weather and climate events near the ‘tail’ of the probability distribution. They are in general difficult to realistically represent in climate models. The 2007 IPCC AR4 concluded that models showed some considerable skill in simulating the statistics of extreme events (especially for temperature extremes) despite their coarse resolution (Randall et al. 2007). In a separate report, the IPCC has conducted an assessment of extreme events in the context of climate change: the Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX) (IPCC, 2012). Although climate model evaluation with respect to extreme events was not done in a consistent manner in SREX, model performance was taken into account when projections uncertainty was assessed.

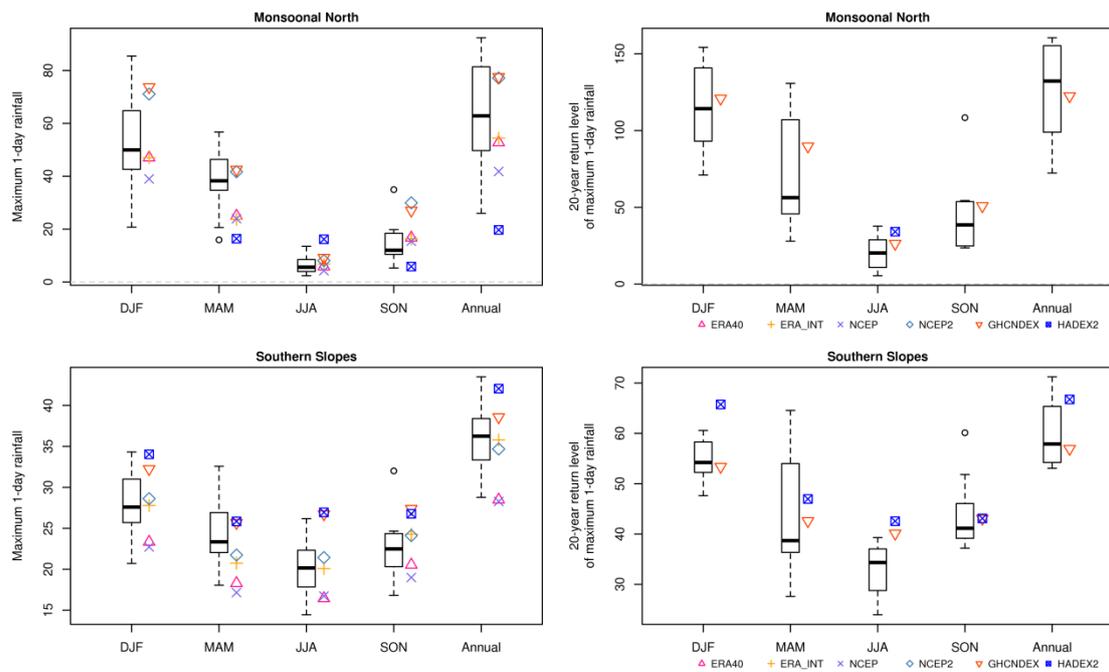
The evaluation of the simulation of extremes in climate models is important because the impacts of climate change will be experienced more profoundly in terms of the frequency, intensity or duration of extreme events (e.g., heat waves, droughts, extreme rainfall events).

The recently published IPCC AR5 WG1 report (IPCC, 2013) summarised that the global distribution of temperature extremes are represented well by CMIP5 models. Furthermore, it reported that CMIP5 models tend to simulate more intense and thus more realistic precipitation extremes than CMIP3, which could be partly due to generally higher horizontal resolution. Related to this is the statement that CMIP5 models are also able to better simulate aspects of large-scale drought.

Specifically for Australia, we have assessed the bias in three of the extreme indices from the CMIP5 model ensemble: annual and seasonal maximum of daily maximum temperature (Txx); annual and seasonal minimum of daily minimum temperature (Tnn) and the annual and seasonal maximum 1-day rainfall event (rx1day). Additionally, the 20-year return value of these quantities has been compared to observed values. There are currently two global observation-based data sets available to assess climate extreme indices: the GHCNDEX (Donat et al. 2013a; Fischer and Knutti 2014) and the HadEX2 (Donat et al. 2013b) data set, with the latter having less spatial coverage, in particular across northern Australia.

A comparison between CMIP5 model daily maximum rainfall and observations is shown in Figure 17 for seasons and annually for two example clusters; Monsoonal North (MN) and Southern Slopes (SS). With less data coverage in tropical Australia for the HadEX2 data set, for the Monsoonal North we focus on how the models are placed compared to the GHCNDEX data points (red downward triangles in Figure 17). Overall the observed daily maximum rainfall amounts are mostly captured by the Interquartile Range (middle 50 % of CMIP5 model simulations) of the model ensemble. This is true for both the maximum daily rainfall event within a year (Figure 17a) as well as the maximum daily rainfall event over a 20 year period (Figure 17b). For summer and averaged over the entire Monsoonal North cluster, the latter event is around 120 mm in the observations and very close to the ensemble median. The spread is fairly substantial – particular towards the lower end with some models showing less than half the observed rainfall during the maximum event. The reason for this could be the lower skill in the representation of both tropical cyclones. For the Southern Slopes cluster, the CMIP5 models have a tendency to underestimate the maximum 1-day rainfall event during a year (Figure 17c) but are still within range. The 20-year event (Figure 17d) is somewhat better captured. Noteworthy is the fact that despite on average receiving more rainfall during winter (JJA), the maximum one-day rainfall events are stronger in the summer months. This is the contribution of intense convective events during the summer season.

Figure 17 Daily extreme rainfall (left column: for daily maximum rainfall per year; ((a) for cluster MN and (c) for cluster SS): for daily maximum rainfall in 20 years ((b) for cluster MN and (d) for cluster SS) across seasons and annually (units are mm/day). CMIP5 models are represented by the box-whisker while coloured symbols represent reanalysis products (see Table 5.2.1 of the CCiA Technical Report) and two gridded observational data products (see text).



Other clusters show very similar results quantitatively: fairly large model spread around median maximum rainfall values that are not too far from that observed.

However, it should be noted that this assessment is for rain events averaged at large spatial scales, whereas many extreme rainfall events in the real world occur at a far smaller spatial scale. These events are included not as single small-scale events but aggregated over each larger grid cell.

Annual and 20 year daily maximum and minimum temperatures show similar biases to mean temperature: a slight cold bias for maximum and slight warm bias for minimum temperatures (not shown). As with rainfall, the model spread is fairly large (up to 10 degrees for some seasons for both daily maximum and minimum temperature).

In summary, the CMIP5 models are able to capture the annual maximum 1-day rainfall event reasonably well. Additionally, they are able to simulate both annual and seasonal daily maximum and minimum temperatures with some skill (not shown).

Evaluation of downscaling simulations

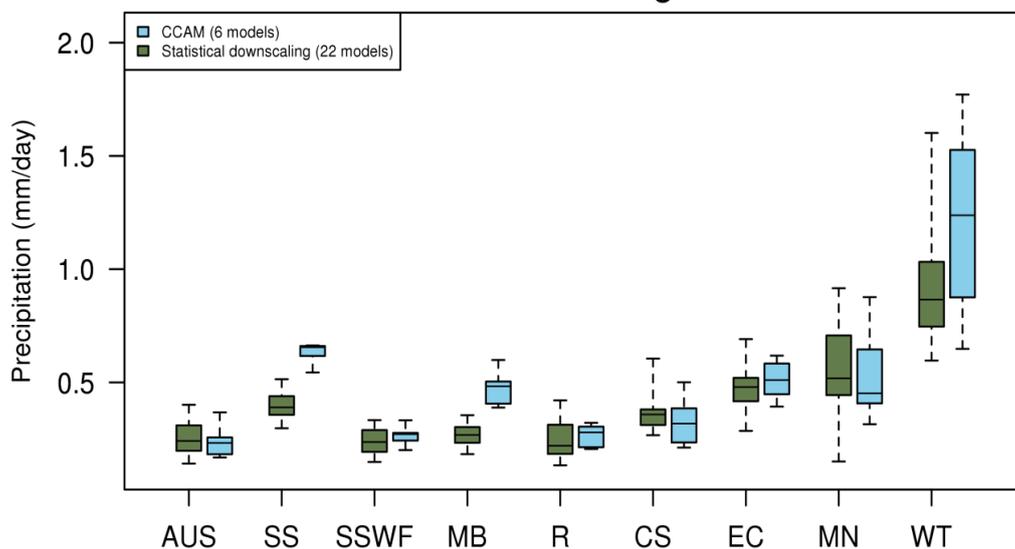
The new Climate Change projections for Australia (CSIRO and Bureau of Meteorology, 2015) includes regional climate change projections information from two downscaling methods (one dynamical downscaling – CCAM; and one statistical downscaling – BOM-SDM). Since model evaluation forms one of the lines of evidence used to construct confidence levels around projected changes of Australian climate, it is necessary to also provide some information about how well the downscaling simulations perform over the historical period.

Each of the two methods used have important aspects that bring them closer to observations. For the statistical downscaling method, observed relationships between local synoptic situations and the large scale climate are used to build the statistical model. This usually leads to a very close representation of the observed climate in the statistical downscaling model, (almost) independent of the choice of host global climate model. A set of 22 global climate models have been used as hosts and the resulting statistical downscaling model simulations are all very similar over the historical period (1986-2005). For the dynamical downscaling method, the monthly sea surface temperature data used as input from each global climate model simulation are initially adjusted to match the observed mean climate before being used to build the dynamical

downscaling simulation. This means the resulting dynamically downscaled simulations are again fairly similar to each other and to the observations over the historical period (1986–2005). Not surprisingly then, for the mean climate we find that all performance metrics are very high for temperature and rainfall, as well as for mean sea level pressure in the dynamical downscaling simulations (not shown).

Additionally assessed are two measures of temporal variability for rainfall: the annual cycle (through the spatial-temporal root mean square error, STRMSE following Gleckler et al. (2008) as above) and the inter-annual variability of rainfall – both at cluster level. Figure 18 shows the comparison of the STRMSE for both ensembles across the cluster regions and the entire continent. Even though the dynamical downscaling ensemble only has 6 members (compared to 22 for the statistical downscaling ensemble), the spread in performance is quite similar for both. Apart from the Southern Slopes (SS) and Murray Basin (MB) clusters, the size of the error is comparable between the two ensembles as well. The dynamical downscaling shows larger STRMSE than the statistical downscaling for SS and MB clusters. For all other non-tropical clusters, the median STRMSE is mostly below 0.5 mm/day. The larger inter-model spread is seen for the tropical cluster regions (Wet Tropics and Monsoonal North) where climatological rainfall is very high and seasonal differences are also very pronounced.

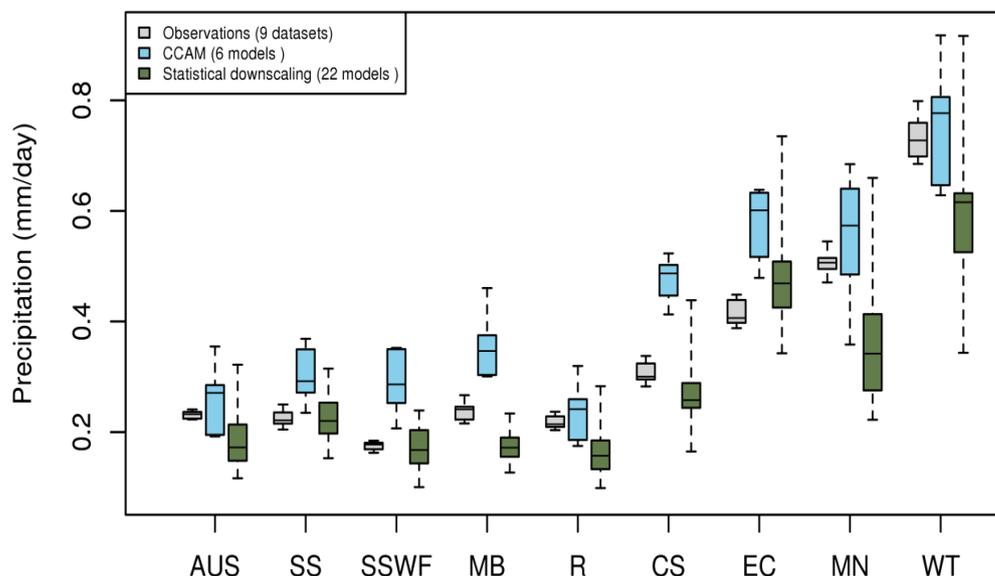
Figure 18 Box-Whisker plot of the spatial-temporal root mean square error (STRMSE; larger values indicate larger errors compared to observations) for rainfall from two downscaled ensembles against AWAP rainfall for Australia and the eight cluster regions. The downscaled ensembles are the BOM-SDM statistical downscaled ensemble (green) and the CCAM dynamical downscaled ensemble (blue).



The year-to-year variability of rainfall is an important feature of the climate within each of the clusters and Figure 19 shows a comparison of the two downsampling ensembles against various observational data sets (including AWAP) for the period 1986–2005. In the observations, the inter-annual variability is fairly modest except along the East Coast and over tropical Australia where the impact of ENSO and monsoonal rainfall is strong. The statistical downscaling ensemble is able to capture the extra-tropical inter-annual rainfall variability well, whereas the dynamical downscaling ensemble shows especially good skill over the tropical clusters and the Rangeland cluster.

It should be noted that whereas downscaling generally involves processes that bring the simulation of the current climate further in line with observations, the downscaling simulations inherit much of the climate change ‘signal’ from the host model. Therefore the set of excellent evaluation metrics shown above does not lead to a proportional increase in the confidence in the projections from downscaling compared to GCMs. They do however show that both downscaling methods achieved their aim: to produce higher resolution outputs with smaller biases than GCMs (compare 1.13 mm/day median STRMSE across Australia in GCMs (Figure 10) to around 0.25 mm/day in downscaled simulations) that may then reveal regional detail in the climate change signal at finer scale than GCMs.

Figure 19 Box-Whisker plot of the temporal standard deviation of annual rainfall (1986-2005) from two downscaled ensembles and gridded observational rainfall for Australia and the eight cluster regions. The downscaled ensembles are the BOM-SDM statistical downscaled ensemble (green) and the CCAM dynamical downscaled ensemble (blue). The observational data includes AWAP.



Discussion and Conclusion

Model evaluation is an important tool to help rate confidence in climate model simulations. This can add to the overall confidence assessment for future projections of the Australian climate. Additionally it can highlight significant model deficiencies that may affect the selection of a subset of models for use in impact assessment. Following is a synthesis and discussion of the main findings of the model evaluation for Australian climate, drawing on the original work and also the literature review presented above.

Atmospheric variables

The CMIP5 models are able to capture the broad-scale characteristics of the 1986-2005 average surface air temperature, rainfall and surface wind climatology. However they display some important deficiencies in simulating the finer details, especially for rainfall. Sometimes model skill can be impacted by large scale biases in the models. For example in some models the so-called "cold-tongue" bias in the central Pacific Ocean influences Australian mean rainfall directly as well as through the ENSO teleconnection to Australian rainfall and therefore results in an additional bias in the annual rainfall cycle. There are also biases in the representation of the seasonal wind reversal across tropical Australia around the onset of the monsoon.

The GISS-E2 models (GISS-E2-H, GISS-E2-H-CC and GISS-E2-R) and MIROC-ESM models (MIROC-ESM and MIROC-ESM-CHEM) provide consistently poorer simulations of the average climate across all atmospheric variables examined. Additionally, IPSL-CM5A-LR shows deficient simulations for several fields and both NorESM1-M models are particularly deficient for mean rainfall across Australia.

Regions and clusters

Some regions and clusters are more difficult to simulate than others (for temperature and rainfall). This is typically the case when (a) the region or cluster is quite small and therefore only a few grid cells contribute to the statistics; and (b) where topography and coastlines play a major role. For example, the skill of simulation of rainfall is acutely linked to surface fields such as topography, coastlines and land surface cover. This is one of the reasons why rainfall varies strongly at regional scales. Therefore higher resolution models can potentially better resolve these processes. The Wet Tropics region is a good example for both being a small cluster region and having significant topography. Others are the Southern Slopes sub-clusters in Tasmania and the East Coast cluster.

For rainfall, the two models CESM1-WACCM and CMCC-CESM show particularly poor simulations across regions in Australia and GISS models GISS-E2-H, GISS-E2-H-CC and GISS-E2-R are similarly deficient mainly over the Wet Tropics and Rangelands regions (Table 3). A few other models showing deficiencies only over some regions include BNU-ESM (for Southern and Eastern Australia); GFDL-ESM2M (for

Southern Australia); IPSL-CM5A-LR and IPSL-CM5B-LR (for Eastern Australia); and MIROC-ESM (scoring the lowest for the entire continent).

Climate features and patterns of variability

Most of the CMIP5 global climate models are able to reproduce the major climate features (SAM, monsoon, pressure systems, sub-tropical jet, circulation – see Table 5) and modes of variability (seasonal cycle, ENSO, Indian Ocean Dipole). Three models (IPSL-CM5A-MR, IPSL-CM5A-LR and CSIRO-MK3-6-0) show unusually low skill with respect to the ENSO-rainfall teleconnection. This is partly due to their bias in the equatorial sea surface temperatures. The following models do not simulate the reversal to monsoon westerlies across tropical Australia during the monsoon season: GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC-ESM, INMCM4, ACCESS1-3 and MIROC-ESM-CHEM.

Recent observed trends

The trend analysis did not provide conclusive evidence that CMIP5 models fail to reproduce 1956-2005 observed trends in daily maximum and minimum temperature and rainfall. However, lack of consistency in simulated recent rainfall trends would warrant reduced confidence in projected changes – this is particularly the case for north-western Australia in summer and south-eastern Australia in autumn.

Extremes

CMIP5 models are able to capture the annual maximum 1-day rainfall events across different clusters reasonably well. Additionally, they are able to simulate both annual and seasonal daily maximum and minimum temperatures with some skill.

Downscaling simulations

Because of the inherent nature of the downscaling methods, the rainfall and temperature climatology is simulated very well. Some differences between the statistical and dynamical method are seen when evaluating climate variability, with the dynamical scheme showing better ability to simulate higher inter-annual variability (in the tropics) while the statistical scheme shows better ability across the southern half of Australia.

CMIP5 model reliability and implication for projections

Despite some models performing poorly across multiple evaluation metrics, the approach adopted for generating climate change projections for Australia has been to equally weight all participating CMIP5 models. In forming ranges of projected change for Australia using CMIP5, factoring in model performance (by different methods weighting or model elimination) does not have a strong effect (CSIRO and Bureau of Meteorology 2015), and is not done routinely in the ranges of projected change presented in this work (CSIRO and Bureau of Meteorology 2015). Nevertheless the model performance results are used in two other important ways. First, they are considered in formulating the confidence rating that is attached to the CMIP5 projections (CCiA 2014). Secondly, poor performing models are flagged in the *Climate Futures* tool (Whetton et al. 2012; Clarke et al. 2011), to guard against these models being selected when forming a small set of models for use in impact assessment.

Finally, from the results of the analysis presented in the individual sections of this paper, the following models were identified as poor performing models, for the reasons outlined (and summarised in Table 6). All of these models should be used with caution in any projection work within Australian regions or for variables, where the noted model deficiencies are likely to be particularly relevant. The models are:

MIROC-ESM and MIROC-ESM-CHEM don't simulate temperature and rainfall over Australia well. They also do not produce monsoon westerlies during the monsoon season and therefore show deficient wet season rainfall (spatial distribution). Both models score low on the simple MJO skill (propagating convection into tropical region), reported by Sperber and Kim (2012). MIROC-ESM additionally shows deficient ENSO-rainfall teleconnection for Australia.

GISS-E2H, GISS-E2H-CC and GISS-E2R show low scores for temperature and rainfall across Australia. They also simulate low scores averaged across various climate features and don't produce monsoon westerlies during the wet season over tropical Australia. Two of the three GISS models do not show a correlation between blocking and rainfall over Australia.

Table 6: Summary of models scoring low on various skill metrics used throughout the model evaluation process. For each evaluation the lowest 6 -8 models are included. The column on the right gives the overall sum of how often a model fell into the lower group.

<i>Model</i>	<i>Low M-Score for PR and TAS</i>	<i>Low STRMSE</i>	<i>Low score for climate features</i>	<i>Low ENSO – rainfall tele connection</i>	<i>No Monsoon westerlies</i>	<i>Wet season rainfall not good spatially</i>	<i>Rainfall relationship to Blocking not good</i>	<i>IOD spatial variability too low or too high</i>	<i>Simple MJO skill not good</i>
ACCESS1-0								X	1
ACCESS1-3					X		X		2
BNU-ESM, CanCM4	X	X				X			3
CanESM2				X			X		1
CCSM4,		X							1
CESM1-BGC,		X							1
CESM1-WACCM,	X	X	X						3
CMCC-CESM,	X								1
CMCC-CMS		X							1
CSIRO-Mk3-6-0				X				X	2
GFDL-CM3								X	1
GFDL-ESM2G				X			X		3
GISS-E2-H,	X	X			X		X		4
GISS-E2-H-CC,	X	X			X				3
GISS-E2-R		X			X		X		3
HadCM3				X				X	2
HadGEM2-ES								X	1
INMCM4				X	X			X	3
IPSL-CM5A-LR				X	X			X	3
IPSL-CM5A-MR				X	X			X	3
IPSL-CM5B-LR							X		1
MIROC-ESM	X			X	X	X		X	5
MIROC-ESM-CHEM	X				X	X		X	4
MPI-ESM-LR						X			1
MPI-ESM-MR						X			1
MPI-ESM-P						X			1
MRI-CGCM3							X		1
NorESM1-M,		X							1
NorESM1-ME,		X							1

IPSL-CM5A-MR and IPSL-CM5A-LR show unusually low skill with respect to the ENSO-rainfall teleconnection over Australia. These two IPSL models also have deficient simulation of larger circulation (no monsoon westerlies) and propagating convection (low MJO related skill) across tropical Australia.

CESM1-WACCM and BNU-ESM are equally low in skill for temperature and rainfall simulations across Australia and averaged over nine climate features important for Australia. Additionally, CESM1-WACCM shows deficiencies in simulating the annual cycle of rainfall while BNU-ESM has lower skill in the spatial representation of wet season rainfall.

Similar to the IPSL models mentioned above, the INMCM4 model has low skill in representing the ENSO-rainfall relationship for Australia and does not produce monsoon westerlies during the wet season over tropical Australia. Additionally, there is low MJO related skill.

GFDL-ESM-2G has low skill in representing the ENSO-rainfall relationship for Australia and does not show a correlation between blocking and rainfall over Australia. Additionally, there is low MJO related skill.

Acknowledgements

This work was supported by the Australian Government (Department of Environment), the Bureau of Meteorology and CSIRO.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's 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.

This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. 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. Hydrometeorology*, 4, 1147-1167.
- Allan, R. and Ansell, T. 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850-2004. *J. Climate*, 19, 5816-5842.
- Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H. and Schulz, J. 2010. The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS-3. *Earth System Science Data*, 2, 215-234.
- Barnes, E. A. and Polvani, L. 2013. Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models. *J. Climate*, 26, 7117-7135.
- Beck, C., Grieser, J. and Rudolph, B. 2005. A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. *Geophysical Research Abstracts*, Vol. 7, 07154.
- Bhend, J. and Whetton, P. 2013. Consistency of simulated and observed regional changes in temperature, sea level pressure and precipitation. *Climatic Change*, 118, 799-810.
- Bindoff, N. L., Stott, P. A., Achutarao, K.M., Allen, M.R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I.I., Overland, J., Perlwitz, J., Sebbari, R. and Zhang, X. 2013: Detection and Attribution of Climate Change: from Global to Regional. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels A., Xia, Y., Bex, V. & Midgley, P.M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. www.climatechange2013.org and www.ipcc.ch.
- Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. and Jones, P. D. 2006. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.*, 111, D12106.
- Cai, W., and Cowan, T. 2013. Why is the amplitude of the Indian Ocean Dipole overly large in CMIP3 and CMIP5 climate models? *Geophys. Res. Lett.*, 40, 1200–1205, doi:10.1002/grl.50208.
- Cai, W., Sullivan, A. and Cowan, T. 2009. Rainfall Teleconnections with Indo-Pacific Variability in the WCRP CMIP3 Models. *J. Climate*, 22, 5046-5071.
- Cai, W., Van Rensch, P., Cowan, T. and Hendon, H.H. 2011. Teleconnection pathways for ENSO and the IOD and the mechanism for impacts on Australian rainfall. *J. Climate*, 24, 3910-3921.
- Cai, W., Zheng, X.-T., Weller, E., Collins, M., Cowan, T., Lengaigne, M., Yu, W. and Yamagata, T. 2013. Projected response of the Indian Ocean Dipole to greenhouse warming. *Nature Geoscience*, 6, 999-1007.
- Cai, W., Borlace, S., Lengaigne, M., Van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., Mcphaden, M., Wu, L., England, M., Wang, G., Guilyardi, E. and Jin, F. 2014. Increasing frequency of extreme El Nino events due to greenhouse warming. *Nature Climate Change*, 4, 111-116.
- Catto, J., Nicholls, N. and Jakob, C. 2012a. North Australian Sea Surface Temperatures and the El Nino-Southern Oscillation in Observations and Models. *J. Climate*, 25, 5011-5029.
- Catto, J., Nicholls, N. and Jakob, C. 2012b. North Australian Sea Surface Temperatures and the El Nino-Southern Oscillation in the CMIP5 Models. *J. Climate*, 25, 6375-6382.
- Clarke, J. M., Whetton, P. H., and Hennessy, K. J. 2011. Providing application-specific Climate Projections datasets. CSIRO's Climate Futures Framework. *International Congress on Modelling and Simulation (MODSIM), Perth, December 2011*.

- Coelho, C. and Goddard, L. 2009. El Niño-Induced Tropical Droughts in Climate Change Projections. *Journal of Climate*, 22, 6456-6476.
- Colman, R., Moise, A. and Hanson, L. 2011. Tropical Australian climate and the Australian monsoon as simulated by 23 CMIP3 models. *Journal of Geophysical Research-Atmospheres*, 116.
- CSIRO and Bureau of Meteorology. 2007. *Climate change in Australia: technical report*, CSIRO Marine and Atmospheric Research, Aspendale, Australia.
- CSIRO and Bureau of Meteorology. 2015. *Climate Change in Australia. Information for Australia's Natural Resource Management Regions*. Technical Report, CSIRO and Bureau of Meteorology, Australia
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G. and Bauer, P. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Met. Soc.*, 137, 553-597.
- Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R. and Caesar, J. 2013a. Global Land-Based Datasets for Monitoring Climatic Extremes. *Bull. Amer. Met. Soc.*, 94, 997-1006.
- Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Tank, A., Kruger, A., Marengo, J., Peterson, T., Renom, M., Rojas, C., Rusticucci, M., Salinger, J., Elrayah, A., Sekele, S., Srivastava, A., Trewin, B., Villarreal, C., Vincent, L., Zhai, P., Zhang, X. and Kitching, S. 2013b. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res.: Atmospheres*, 118, 2098-2118.
- Dunn-Sigouin, E. and Son, S. 2013. Northern Hemisphere blocking frequency and duration in the CMIP5 models. *J. Geophys. Res.: Atmospheres*, 118, 1179-1188.
- Evans, J. P., Ji, F., Abramowitz, G. and Ekstrom, M. 2013. Optimally choosing small ensemble members to produce robust climate simulations. *Environmental Research Letters*, 8(4): 044050.
- Fennig, K., Andersson, A., Bakan, S., Klepp, C. and Schroeder, M. 2012. *Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 3.2 - Monthly Means / 6-Hourly Composite*. (electronic publication). Satellite Application Facility on Climate Monitoring.
- Fischer, E. and Knutti, R. 2014. Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, 41, 547-554.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S. and Eyring, V. 2013. Evaluation of climate models. In: T.F. Stocker, D. Qin, G. Plattner et al. (eds.), *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, United Kingdom and New York, USA, Cambridge University Press.
- Gleckler, P., Taylor, K. and Doutriaux, C. 2008. Performance metrics for climate models. *J. Geophys. Res.: Atmospheres*, 113, D06104, doi: 10.1029/2007JD008972.
- Grainger, S., Frederiksen, C. S. and Zheng, X. 2014. Assessment of modes of interannual variability of Southern Hemisphere atmospheric circulation in CMIP5 models. *J. Climate*, 27, 8107-8125.
- Grose, M., Pook, M., McIntosh, P., Risbey, J. and Bindoff, N. 2012. The simulation of cutoff lows in a regional climate model: reliability and future trends. *Clim. Dyn.*, 39, 445-459.
- Grose, M.R., Timbal, B., Wilson, L., Bathols, J., and Kent, D. 2015. The subtropical ridge in CMIP5 models, and implications for projections of rainfall in southeast Australia. *Aust. Met. Oceanogr. J.*, 65, 90–106.
- Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Van Oldenborgh, G. J. and Stockdale, T. 2009. Understanding El Niño in ocean-atmosphere general circulation models: progress and challenges. *Bull. Amer. Met. Soc.*, 90, 325-340.
- Hansen, J., Ruedy, R., Sato, M., and Lo, K. 2010. Global surface temperature change. *Reviews in Geophysics*, 48, RG4004.
- Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H. 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.*, 34, 623–642.
- Hendon, H., Thompson, D. and Wheeler, M. 2007. Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode. *J. Climate*, 20, 2452-2467.
- Huffman, G. J., Adler, R. F., Bolvin, D. T. and Gu, G. 2009. Improving the global precipitation record: GPCP Version 2.1. *Geophys. Res. Lett.*, 36, L17808.
- Hung, M., Lin, J., Wang, W., Kim, D., Shinoda, T. and Weaver, S. 2013. MJO and Convectively Coupled Equatorial Waves Simulated by CMIP5 Climate Models. *J. Climate*, 26, 6185-6214.
- Inness, P., Slingo, J., Guilyardi, E. and Cole, J. 2003. Simulation of the Madden-Julian oscillation in a coupled general circulation model. Part II: The role of the basic state. *J. Climate*, 16, 365-382.
- IPCC. 2007. Summary for Policy Makers. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Eds., Cambridge University Press, Cambridge, United Kingdom.

- IPCC. 2012. Summary for Policymakers. *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, in: C. B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.) (Ed.). pp. 582 pp.
- 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*. In: Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (ed.).
- Jakob, D. 2010. Challenges in developing a high-quality surface wind-speed data-set for Australia. *Aust. Met. Oceanogr. J.*, 60, 227-236.
- Jiang, X., Maloney, E., Li, J. and Waliser, D. 2013. Simulations of the Eastern North Pacific Intraseasonal Variability in CMIP5 GCMs. *J. Climate*, 26, 3489-3510.
- Jones, D. A., Wang, W. and Fawcett, R. 2009. High-quality spatial climate data-sets for Australia. *Aust. Met. Oceanogr. J.*, 58, 233-248.
- Jourdain, N. C., Gupta, A. S., Taschetto, A. S., Ummenhofer, C. C., Moise, A. F. & Ashok, K. 2013. The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations. *Clim. Dyn.*, 41, 3073-3102.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Met. Soc.*, 77, 437-471.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J., Fiorino, M. and Potter, G. 2002. NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Met. Soc.*, 83, 1631-1643.
- Kent, D., Kirono, D., Timbal, B. and Chiew, F. 2013. Representation of the Australian sub-tropical ridge in the CMIP3 models. *Int. J. Climatol.*, 33, 48-57.
- Kim, D., Sperber, K., Stern, W., Waliser, D., Kang, I., Maloney, E., Wang, W., Weickmann, K., Benedict, J., Khairoutdinov, M., Lee, M., Neale, R., Suarez, M., Thayer-Calder, K. and Zhang, G. 2009. Application of MJO Simulation Diagnostics to Climate Models. *J. Climate*, 22, 6413-6436.
- Kim, D., Sobel, A., Del Genio, A., Chen, Y., Camargo, S., Yao, M., Kelley, M. and Nazarenko, L. 2012. The Tropical Subseasonal Variability Simulated in the NASA GISS General Circulation Model. *J. Climate*, 25, 4641-4659.
- Langenbrunner, B. and Neelin, J. 2013. Analyzing ENSO Teleconnections in CMIP Models as a Measure of Model Fidelity in Simulating Precipitation. *J. Climate*, 26, 4431-4446.
- Large, W. G. and Yeager, S. 2009. The global climatology of an interannually varying air-sea flux data set. *Clim. Dyn.*, 33, 341–364.
- Large, W. G. and Yeager, S. 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies. CGD Division of the National Center for Atmospheric Research, *NCAR Technical Note: NCAR/TN-460+STR*.
- Lin, J., Kiladis, G., Mapes, B., Weickmann, K., Sperber, K., Lin, W., Wheeler, M., Schubert, S., Del Genio, A., Donner, L., Emori, S., Gueremy, J., Hourdin, F., Rasch, P., Roeckner, E. and Scinocca, J. 2006. Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Climate*, 19, 2665-2690.
- Liu, L., Yu, W. and Li, T. 2011. Dynamic and Thermodynamic Air-Sea Coupling Associated with the Indian Ocean Dipole Diagnosed from 23 WCRP CMIP3 Models. *J. Climate*, 24, 4941-4958.
- Madden, R. A. and Julian, P. R. 1972. Description of global-scale circulation cells in the tropics with a 40-50 day period. *J. Atmos. Sci.*, 29, 1109-1123.
- Madden, R. A. and Julian, P. R. 1994. Observations of the 40-50-day tropical oscillation - a review. *Mon. Wea. Rev.*, 122, 814-837.
- Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J. & Taylor, K. E. 2007: THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research, *Bull. Amer. Met. Soc.*, 88, 1383-1394, doi: 10.1175/BAMS-88-9-1383.
- Mizuta, R., Yoshimura, H., Murakami, H., Matsueda, M., Endo, H., Ose, T., Kamiguchi, K., Hosaka, M., Sugi, M., Yukimoto, S., Kusunoki, S. and Kitoh, A. 2012. Climate Simulations Using MRI-AGCM3.2 with 20-km Grid. *J. Met. Soc. Japan*, 90A, 233-258.
- Moise, A., Colman, R. and Brown, J. 2012. Behind uncertainties in projections of Australian tropical climate: Analysis of 19 CMIP3 models. *J. Geophys. Res.: Atmospheres*, 117, D10103, doi: 10.1029/2011JD017365.
- Naud, C. M., Booth, J. F., and Del Genio, A. D. 2014. Evaluation of ERA-Interim and MERRA Cloudiness in the Southern Ocean. *J. Climate*, 27, 2109–2124.
- Neelin, J., Chou, C. and Su, H. 2003. Tropical drought regions in global warming and El Nino teleconnections. *Geophys. Res. Lett.*, 30, 2275, doi: 10.1029/2003GL018625.
- Neelin, J. D. 2007. Moist dynamics of tropical convection zones in monsoons, teleconnections, and global warming. In: T. Schneider and A. Sobel (eds.) *Moist dynamics of tropical convection zones in monsoons, teleconnections and global warming*, Princeton, Princeton University Press.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N. and Taira, R. 2007. The JRA-25 Reanalysis. *J. Met. Soc. Japan. Ser. II*, 86, 369-432.

- Pook, M. and Gibson, T. 1999. Atmospheric blocking and storm tracks during SOP-1 of the FROST Project. *Aust. Met. Mag* (special issue), 51-60.
- Raupach, M.R., Briggs, P.R., Haverd, V., King, E.A., Paget, M. and Trudinger, C.M. 2009. Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. *CAWCR Technical Report 13*. Centre for Australian Weather and Climate Research, Melbourne, 67 pp.
- Raupach, M.R., Briggs, P.R., Haverd, V., King, E.A., Paget, M. and Trudinger, C.M. 2012. *Australian Water Availability Project*. CSIRO Marine and Atmospheric Research, Canberra, Australia.
- Randall, D. A. and coauthors. 2007. Climate models and their evaluation. In: Solomon, S. (ed.) *Climate change 2007 - the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*. Cambridge: Cambridge University Press. Pp. 589-662.
- Raphael, M. and Holland, M. 2006. Twentieth century simulation of the southern hemisphere climate in coupled models. Part 1: large scale circulation variability. *Clim. Dyn.*, 26, 217-228.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C. and Kaplan, A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.: Atmospheres*, 108, D1.
- Reinecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M. and Woollen, J. 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, 24, 3624–3648.
- Risbey, J., Pook, M., McIntosh, P., Ummenhofer, C. and Meyers, G. 2009a. Characteristics and variability of synoptic features associated with cool season rainfall in southeastern Australia. *Int. J. Climatol.*, 29, 1595-1613.
- Risbey, J., Pook, M., McIntosh, P., Wheeler, M. and Hendon, H. 2009b. On the Remote Drivers of Rainfall Variability in Australia. *Mon. Wea. Rev.*, 137, 3233-3253.
- Rudolf, B., Beck, C., Grieser, J. & Schneider, U. 2005. *Global Precipitation Analysis Products*. Global Precipitation Climatology Centre (GPCC), DWD, Internet publication, 1-8.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G. and Goldberg, M. 2010. The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Met. Soc.*, 91, 1015–1057.
- Saji, N., Xie, S. and Yamagata, T. 2006. Tropical Indian Ocean variability in the IPCC twentieth-century climate simulations. *J. Climate*, 19, 4397-4417.
- Scaife, A. A., Copesey, D., Kendon, L., Vidale, P.-L., Martin, G. M., Milton, S., Moufouma-Okia, W., Roberts, M. J., Palmer, Walters, D. N. and Williams, K. 2011. *Final Report of the CAPTIVATE model development project*. Deliverable D3.2.7 (Internal Report), Met. Office Hadley Centre.
- Sperber, K. and Kim, D. 2012. Simplified metrics for the identification of the Madden-Julian oscillation in models. *Atmos. Sci. Letters*, 13, 187-193.
- Sperber, K. and Annamalai, H. 2008. Coupled model simulations of boreal summer intraseasonal (30-50 day) variability, Part 1: Systematic errors and caution on use of metrics. *Clim. Dyn.*, 31, 345-372.
- Sperber, K., Annamalai, H., Kang, I., Kitoh, A., Moise, A., Turner, A., Wang, B. and Zhou, T. 2013. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.*, 41, 2711-2744.
- Stone, D. A., Allen, M. R., Stott, P. A., Pall, P., Min, S. K., Nozawa, T. and Yukimoto, S. 2009. The Detection and Attribution of Human Influence on Climate. *Annual Review of Environment and Resources*, 34, 1-16.
- Taylor, K. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.: Atmospheres*, 106, 7183-7192.
- Taylor, K. E., Stouffer, R. J. and Meehl, G. A. 2012. An overview of CMIP5 and the experiment design. *Bull. Amer. Met. Soc.*, 93, 485-498.
- Thompson, D. and Solomon, S. 2002. Interpretation of recent Southern Hemisphere climate change. *Science*, 296, 895-899.
- Timbal, B. and Drosowsky, W. 2013. The relationship between the decline of South Eastern Australia rainfall and the strengthening of the subtropical ridge. *Int. J. Climatol.*, 33, 1021-1034.
- Trenberth, K., Stepaniak, D. and Caron, J. 2000. The global monsoon as seen through the divergent atmospheric circulation. *J. Climate*, 13, 3969-3993.
- Trewin, B. 2013. A daily homogenized temperature data set for Australia. *Int. J. Climatol.*, 33, 1510-1529.
- Uppala, S., Kallberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M., Beljaars, A., Van De Berg, L., Bidlot, J., Bormann, N.,

- Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B., Isaksen, L., Janssen, P., Jenne, R., McNally, A., Mahfouf, J., Morcrette, J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. 2005. The ERA-40 re-analysis. *Quart. J. Roy. Met. Soc.*, 131, 2961-3012.
- Van Haren, R., Van Oldenborgh, G., Lenderink, G., Collins, M. and Hazeleger, W. 2012. SST and circulation trend biases cause an underestimation of European precipitation trends. *Clim. Dyn.*, 40, 1-20.
- Van Oldenborgh, G. J., Drijfhout, S., Van Ulden, A., Haarsma, R., Sterl, A., Severijns, C., Hazeleger, W. and Dijkstra, H. 2009. Western Europe is warming much faster than expected. *Clim. Past*, 5, 1-12.
- Van Oldenborgh, G. J., Reyes, F. D., Drijfhout, S. and Hawkins, E. 2013. Reliability of regional climate model trends. *Environmental Research Letters*, 8, 014055.
- Waliser, D., Jin, K., Kang, I., Stern, W., Schubert, S., Wu, M., Lau, K., Lee, M., Krishnamurthy, V., Kitoh, A., Meehl, G., Galin, V., Satyan, V., Mandke, S., Wu, G., Liu, Y. and Park, C. 2003. AGCM simulations of intraseasonal variability associated with the Asian summer monsoon. *Clim. Dyn.*, 21, 423-446.
- Waliser, D., Sperber, K., Hendon, H., Kim, D., Wheeler, M., Weickmann, K., Zhang, C., Donner, L., Gottschalck, J., Higgins, W., Kang, I., Legler, D., Moncrieff, M., Vitart, F., Wang, B., Wang, W., Woolnough, S., Maloney, E., Schubert, S., Stern, W., Vitart, F., Wang, B., Wang, W. and Woolnough, S. 2009. MJO Simulation Diagnostics. *J. Climate*, 22, 3006-3030, doi: 10.1175/2008JCLI2731.1.
- Wang, C., Picaut, J., Wang, C., Xie, S. and Carton, J. 2004a. Understanding ENSO physics - A review. *Earth's Climate: the Ocean-Atmosphere Interaction*, 147, 21-48.
- Wang, G. and Hendon, H. H. 2007. Sensitivity of Australian rainfall to El-Nino variations. *J. Climate*, 20, 4211-4226.
- Wang, B. and Ding, Q. 2008. Global monsoon: Dominant mode of annual variation in the tropics. *Dynamics of Atmospheres and Oceans*, 44, 165-183.
- Watanabe, M., Chikira, M., Imada, Y., and Kimoto, M. 2011 Convective control of ENSO simulated in MIROC. *J. Climate*, 24, 543-562.
- Watterson, I. 1996. Non-dimensional measures of climate model performance. *Int. J. Climatol.*, 16, 379-391.
- Watterson, I., Bathols, J. and Heady, C. 2013a. What influences the skill of climate models over the continents? *Bull. Amer. Met. Soc.*, 95, 689-700.
- Watterson, I., Hirst, A. and Rotstayn, L. 2013b. A skill score based evaluation of simulated Australian climate. *Aust. Met. Oceanogr. J.*, 63, 181-190.
- Weller, E. and Cai, W. 2013a. Asymmetry in the IOD and ENSO Teleconnection in a CMIP5 Model Ensemble and Its Relevance to Regional Rainfall. *J. Climate*, 26, 5139-5149.
- Weller, E. and Cai, W. 2013b. Realism of the Indian Ocean Dipole in CMIP5 Models: The Implications for Climate Projections. *J. Climate*, 26, 6649-6659.
- Wheeler, M., Hendon, H., Cleland, S., Meinke, H. and Donald, A. 2009. Impacts of the Madden-Julian Oscillation on Australian Rainfall and Circulation. *J. Climate*, 22, 1482-1498.
- Whetton, P., Macadam, I., Bathols, J. and O'Grady, J. 2007. Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophys. Res. Lett.*, 34, 14, doi: 10.1029/2007GL030025.
- Whetton, P., Hennessy, K., Clarke, J., McInnes, K., and Kent, D. 2012. Use of Representative Climate Futures in impact and adaptation assessment. *Climatic Change*, 115, 433-442.
- Xavier, P. 2012. Intraseasonal Convective Moistening in CMIP3 Models. *J. Climate*, 25, 2569-2577.
- Xavier, P., Duvel, J., Braconnot, P. and Doblas-Reyes, F. 2010. An Evaluation Metric for Intraseasonal Variability and its Application to CMIP3 Twentieth-Century Simulations. *J. Climate*, 23, 3497-3508.
- Xie, P. and Arkin, P. A. 1997. Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bull. Amer. Met. Soc.*, 78, 2539-2558.
- Zhang, X. B., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A. and Nozawa, T. 2007. Detection of human influence on twentieth century precipitation trends. *Nature*, 448, 461-465.
- Zheng, F., Li, J., Clark, R. T. and Nnamchi, H. C. 2013. Simulation and projection of the Southern Hemisphere annular mode in CMIP5 models. *J. Climate*, 26, 9860-9879.