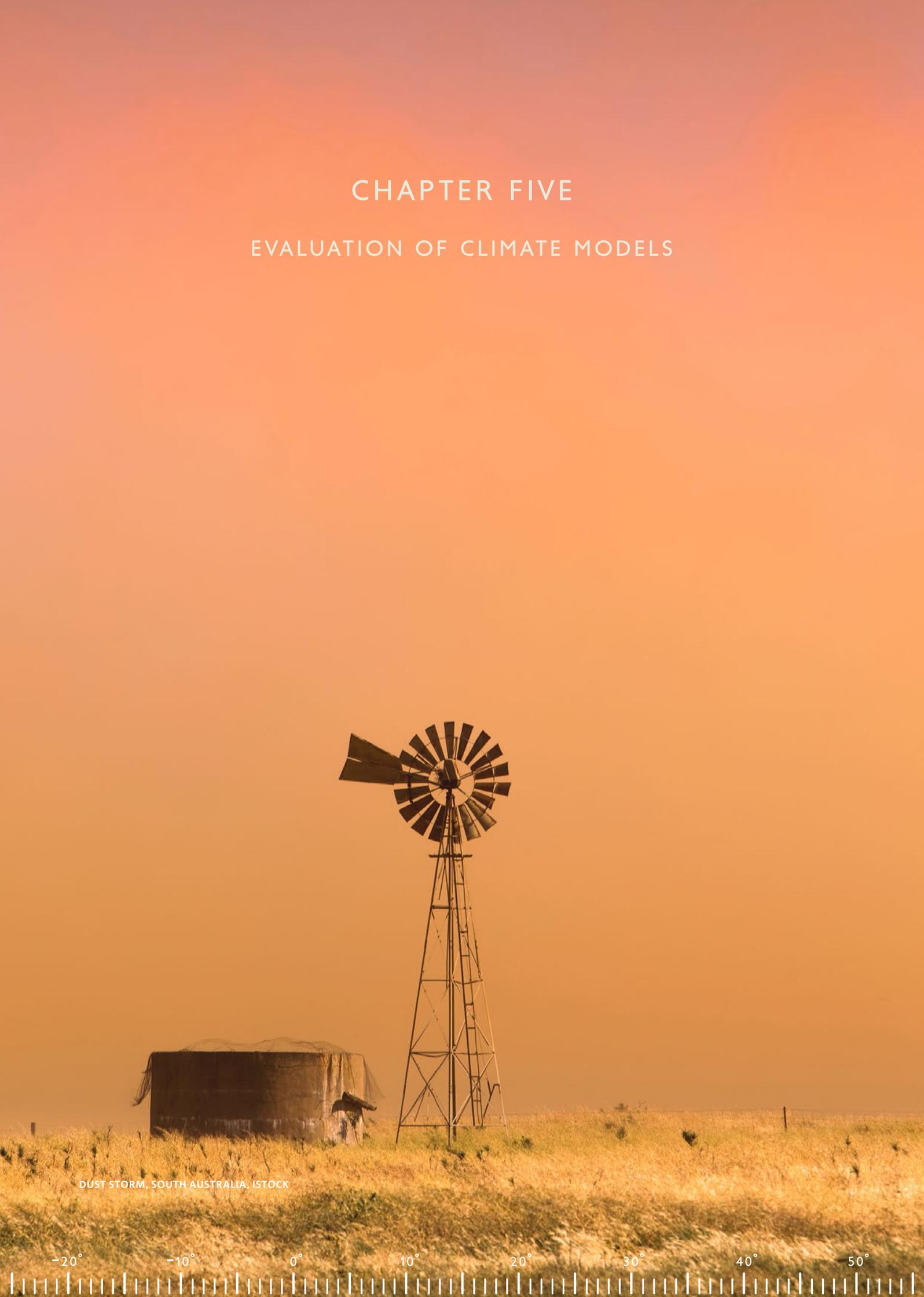


CHAPTER FIVE

EVALUATION OF CLIMATE MODELS



DUST STORM, SOUTH AUSTRALIA, ISTOCK

-20° -10° 0° 10° 20° 30° 40° 50°

CHAPTER 5 EVALUATION OF CLIMATE MODELS

5.1 INTRODUCTION

In this Chapter 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 our assessment of confidence in model-simulated future climate changes (see Section 6.4) 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 CMIP5 models 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 (see also Section 4.1.2 for details on the processes and features) in order to capture the complexity of the climate system as well as satisfy the interests of stakeholders. While this makes it difficult to identify a group of best performing models, 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 are choosing a subset of models for application in impact assessment (*e.g.* through the Climate Futures approach – see Chapter 9). Similarly, the results from the model evaluation are very important when choosing host models for dynamical downscaling (see Section 6.3.3).

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.* 2009a, Raupach *et al.* 2009, 2012) is used for the evaluation 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 datasets used in this chapter are described in Table 5.2.1 and the global climate models (from CMIP3 and CMIP5) are described in Table 3.3.1, including the model labels used throughout this Report.

Features of global climate models, such as resolution and representation of physical processes, and details of the CMIP5 suite of experiments are discussed in Chapter 3.

Most analysis in this chapter is carried out with respect to the historical experiments described in Section 3.3.1. The performance of global climate models with respect

to climatological characteristics, features and processes is evaluated in Section 5.2. The simulation of observed regional climate trends is evaluated in Section 5.3. The simulation of climatic extremes is evaluated in Section 5.4 and downscaling simulations are discussed in Section 5.5, before we conclude in Section 5.6.

5.2 EVALUATION BASED ON CLIMATOLOGICAL CHARACTERISTICS

Global climate models are designed to simulate large-scale processes well. 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 of features and processes these models may simulate realistically. Table 3.3.1 shows the CMIP5 model ensemble and the resolution of both the atmospheric and ocean components of the models. CMIP5 is overall an improved set of global climate models compared to CMIP3 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 are 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. Following is an overview of an assessment for the Australian region.

5.2.1 ASSESSMENT OF HISTORICAL MEAN CLIMATOLOGIES: TEMPERATURE, RAINFALL AND MEAN SEA LEVEL PRESSURE

Figures 5.2.1 and 5.2.2 show a comparison of annual and seasonal climatologies (long-term averages) of temperature and rainfall for Australia. The left column in both figures shows the reference observational data set (AWAP, Jones *et al.* 2009a, Raupach *et al.* 2009; 2012) while the middle column shows an average of a selection of other observational data sets and reanalyses (see Table 5.2.1 for an overview of these) and the right column displays the CMIP5 ensemble mean.

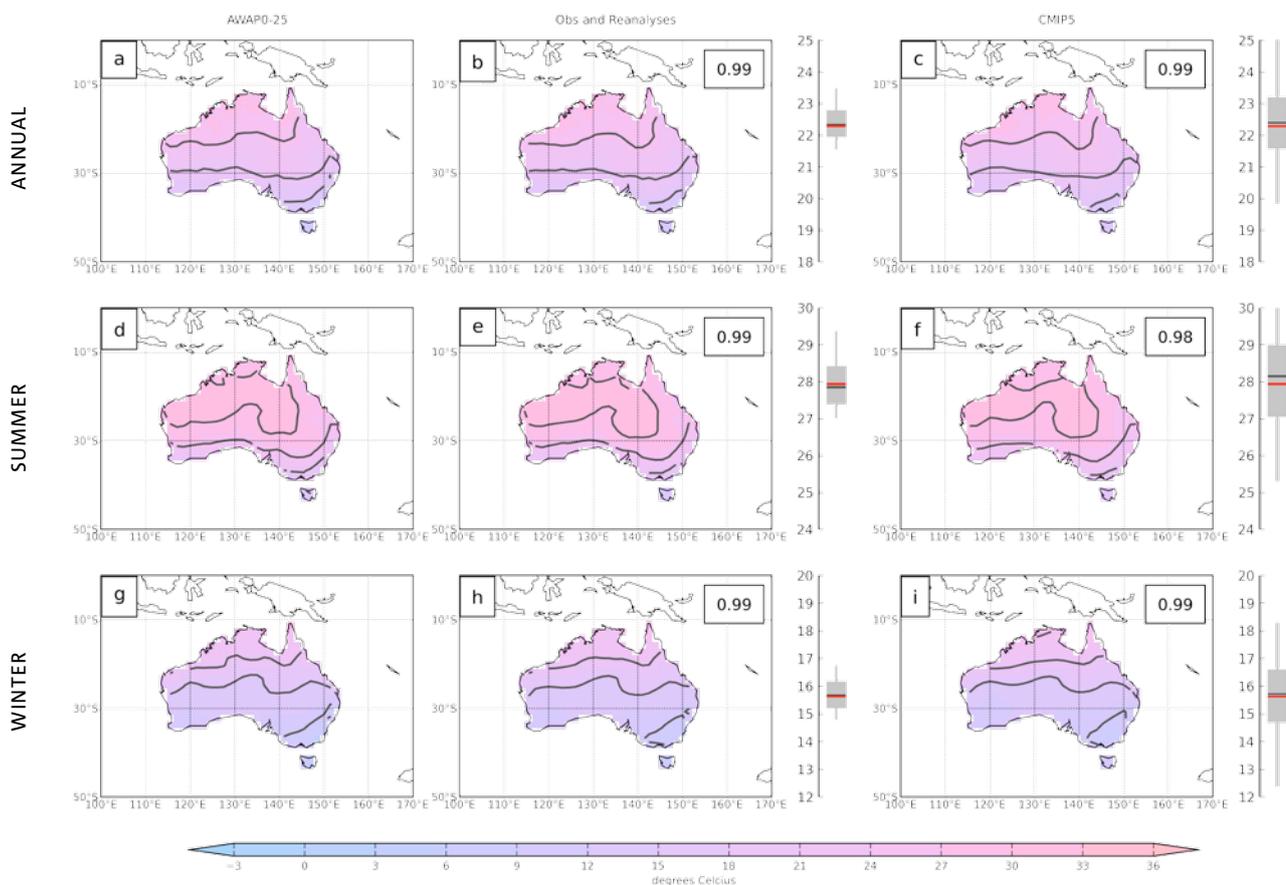


FIGURE 5.2.1: CLIMATOLOGICAL MEAN SURFACE AIR TEMPERATURE FROM AWAP (A, D, G, THE REFERENCE DATA SET), THE AVERAGE OF A SELECTION OF OTHER OBSERVATIONAL DATA SETS (B, E, H, SEE TABLE 5.2.1) AND THE CMIP5 MEAN MODEL (C, F, I) FOR ANNUAL (TOP ROW), SUMMER (DEC-FEB, MIDDLE ROW) AND WINTER (JUN-AUG, BOTTOM 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.



TABLE 5.2.1: LIST OF GLOBAL GRIDDED OBSERVATIONAL (BLUE) AND REANALYSIS DATA SETS (GREEN), THEIR CLIMATE FIELDS USED, TIME COVERAGE, ORIGIN AND REFERENCE. THE AWAP AND ACORN-SAT REFERENCE DATA SETS FOR SURFACE AIR TEMPERATURE AND RAINFALL OVER AUSTRALIA ARE SHOWN IN ORANGE. THE ABBREVIATION PR REFERS TO PRECIPITATION; TAS: SURFACE AIR TEMPERATURE; MSLP: MEAN SEA LEVEL PRESSURE.

GRIDDED DATA SET		PERIOD	ORIGIN	REFERENCES
NAME	FIELDS			
AWAP	PR	1900-2012	Australian Water Availability Project, Bureau of Meteorology and CSIRO	Jones <i>et al.</i> 2009a; Raupach <i>et al.</i> 2009 and 2012
ACORN-SAT	TAS	1910-2012	Australian Climate Observations Reference Network – Surface Air Temperature, Bureau of Meteorology	Trewin 2013
CMPA	PR	1979-2008	Climate Prediction Centre Merged Analysis of Precipitation	Xie & Arkin, 1997
GPPC	PR	1901-2010	Global Precipitation Climatology Centre (GPPC)	Rudolf <i>et al.</i> 2005; Beck <i>et al.</i> 2005
GPCP	PR	1979-2008	Global Precipitation Climatology Project 2	Huffman <i>et al.</i> 2009; Adler <i>et al.</i> 2003
CRU	TAS	1901-2006	Climate Research Unit temperature database	Harris <i>et al.</i> 2013
GISS	TAS	1850-2006	NASA Goddard Institute for Space Sciences (GISS) Surface Temperature Analysis	GISTEMP; Hansen <i>et al.</i> 2010
HADCRU	TAS, PR	1901-2008	Met Office Hadley Centre and Climate Research Unit	HadCRUT3; Brohan <i>et al.</i> 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 <i>et al.</i> , 2012; Andersson <i>et al.</i> 2010
HadISST	SST	1870-2010	Hadley Centre Sea Ice and Sea Surface Temperature dataset	HadISST2; Rayner <i>et al.</i> 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 <i>et al.</i> 2010
Merra	PR, WINDS, MSLP, TAS	1979-2011	Modern Era Retrospective-analysis for Research and Applications	Rienecker <i>et al.</i> 2011
ERA40	PR, WINDS, MSLP, TAS	1958-2002	European 40-year reanalysis	Uppala <i>et al.</i> 2005
ERA_INT	PR, WINDS, MSLP, TAS	1979-2011	ERA-interim	Dee <i>et al.</i> 2011
NCEP	PR, WINDS, MSLP, TAS	1948-2011	NCEP/NCAR reanalysis 1	Kalnay <i>et al.</i> 1996
NCEP2	PR, WINDS, MSLP, TAS	1979-2011	NCEP/DOE reanalysis 2	Kanamitsu <i>et al.</i> 2002
JRA25anl	PR, WINDS, MSLP, TAS	1979-2010	Japanese 25-year reanalysis	Onogi <i>et al.</i> 2007

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 5.2.1) although the ensemble mean model climate is cooler than AWAP across northern parts over Western Australia. During winter (Jun-Aug) the model ensemble mean model is 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 distribution of temperature.

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 5.2.1. While 50 per cent of the models are within $\pm 1^\circ\text{C}$ 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 5.2.1 additionally indicate that there is some discrepancy amongst the other observational data sets and reanalysis data sets with respect to temperature across Australia. However, 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. Mean rainfall differences are shown in Figure 5.2.2. There is a general tendency for the models to simulate 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.

The winter rainfall regime (across southern coastal regions of Australia) on the other hand is generally too dry as simulated by the climate models, especially in Tasmania. This could be caused by insufficient global model resolution of two kinds: (a) some model resolutions are too coarse to simulate the correct land-sea contrast over Tasmania; (b) even if global models have information over Tasmania it is usually not enough to correctly simulate 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.

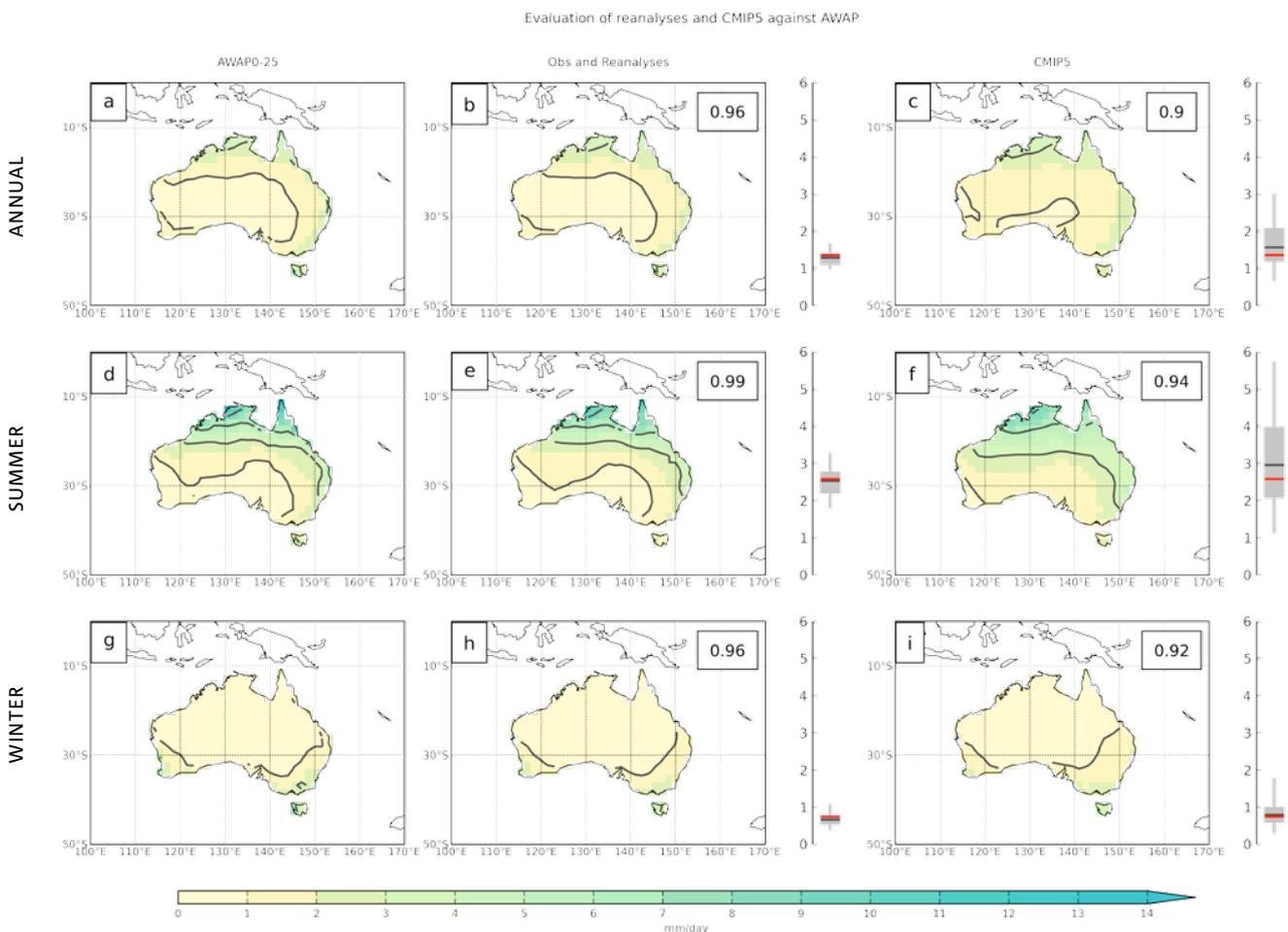


FIGURE 5.2: CLIMATOLOGICAL MEAN RAINFALL FROM AWAP (A, D, G, THE REFERENCE DATA SET), THE AVERAGE OF A SELECTION OF OTHER OBSERVATIONAL DATA SETS (B, E, H, SEE TABLE 5.2.1) AND THE CMIP5 MEAN MODEL (C, F, I) FOR ANNUAL (TOP ROW), SUMMER (DEC-FEB, MIDDLE ROW) AND WINTER (JUN-AUG, BOTTOM ROW) RAINFALL. 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.

Figures 5.2.3 and 5.2.4 show the cluster-based assessment of the CMIP5 model biases in seasonal surface air temperature and rainfall climatologies. In general, the CMIP5 models are able to capture seasonal temperatures much better than rainfall.

During summer, the model simulated median temperatures (Figure 5.2.3) are very close to AWAP reference values, particularly for the warmer regions across northern and central Australia (Monsoonal North and Rangelands 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.

During winter, the majority of climate models have warm biases over some regions of south-eastern Australia (Southern Slopes cluster). 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.

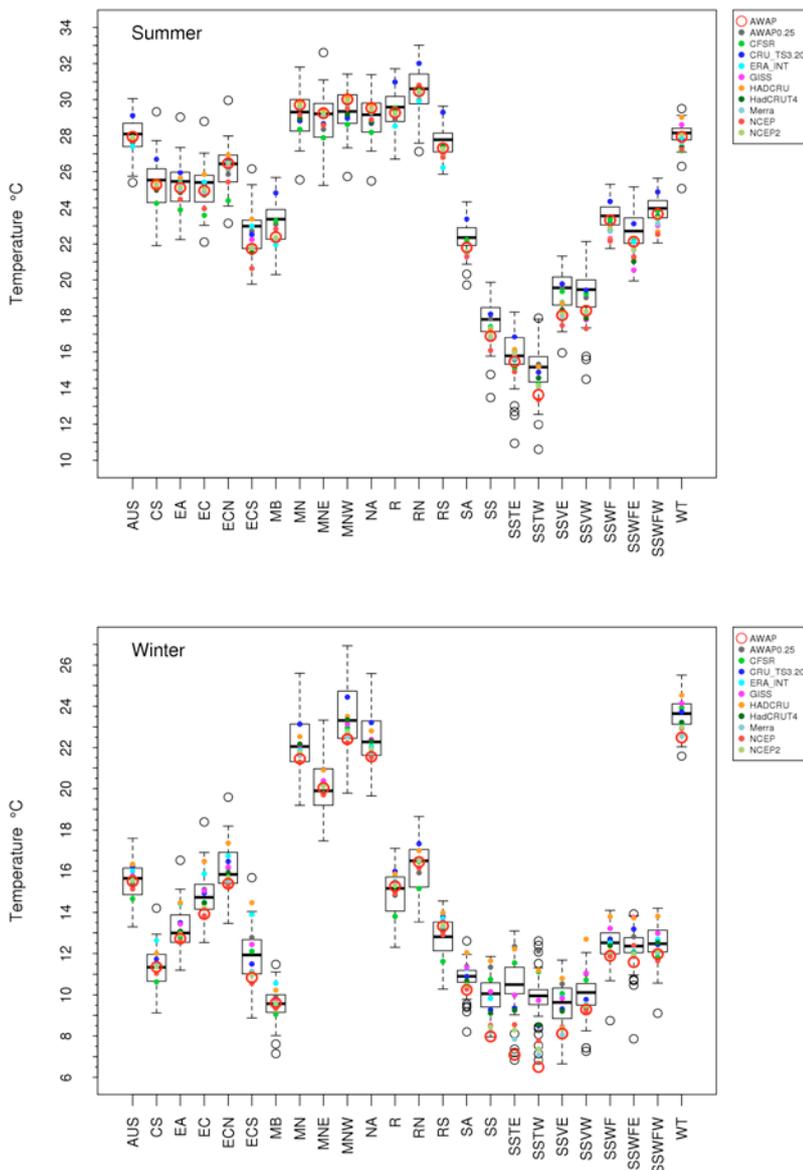


FIGURE 5.2.3: CLUSTER AVERAGED MEAN SURFACE AIR TEMPERATURE (UNITS: °C) FOR SUMMER (DEC-FEB, TOP) AND WINTER (JUN-AUG, BOTTOM) FROM ALL CMIP5 MODELS (REPRESENTED BY BOX-WHISKER BARS), AWAP (RED CIRCLE) AND SEVERAL OTHER OBSERVATIONS AND RE-ANALYSIS DATA SETS (COLOURED DOTS). THE BOX-WHISKERS DISPLAY 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 AWAY FROM THE MEDIAN). THE TIME PERIOD USED IS 1986–2005.

Cluster-based rainfall biases are shown in Figure 5.2.4 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 (except the East Coast South sub-cluster), but show more variable skill elsewhere.

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. Along the tropical east coast (Wet Tropics cluster), 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

clusters) with wet biases of up to 20 mm/month. 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 (Southern Slopes and Murray Basin clusters), except for the East Coast cluster where the GCM ensemble median rainfall is a good match to the observed rainfall. 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 well along the tropical regions (Wet Tropics and Monsoonal North clusters) and also the higher winter rainfall of the East Coast cluster. Dry biases are common

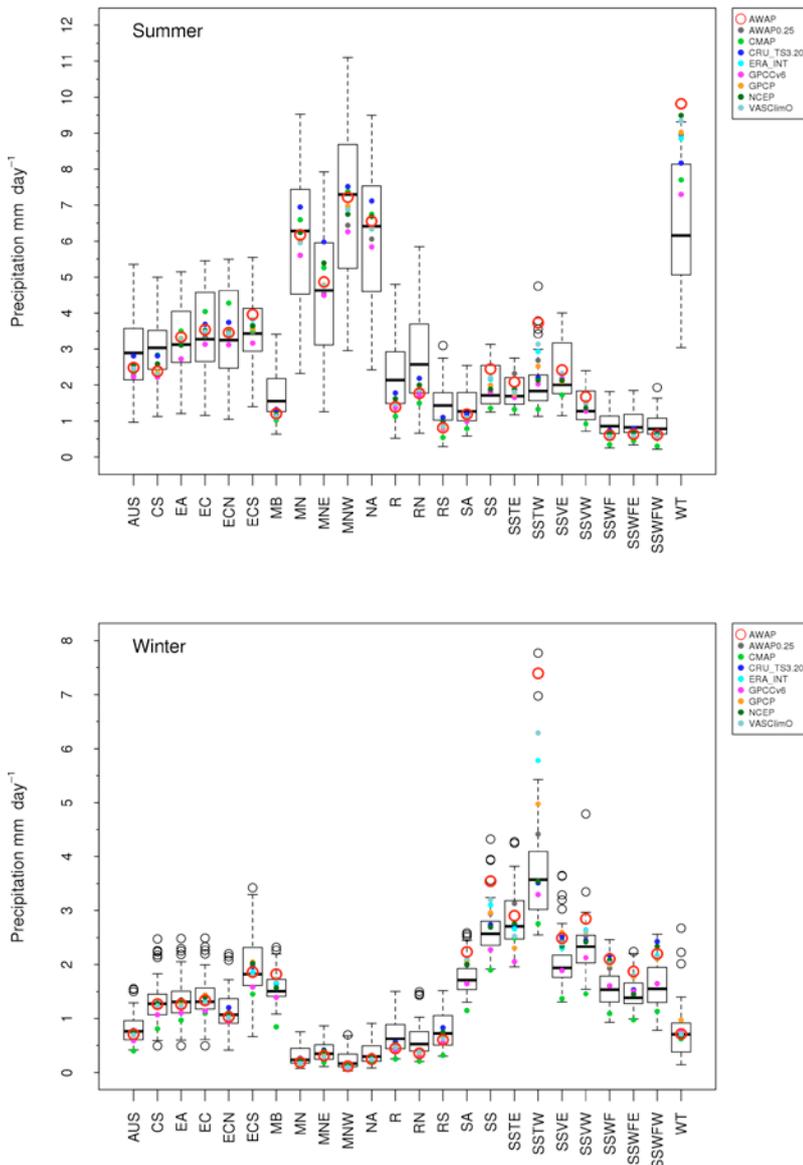


FIGURE 5.2.4: CLUSTER AVERAGED RAINFALL (UNITS: MM PER DAY) FOR SUMMER (DEC-FEB, TOP) AND WINTER (JUN-AUG, BOTTOM) FROM ALL CMIP5 MODELS (REPRESENTED BY BOX-WHISKER BARS), AWAP (RED CIRCLE) AND SEVERAL OTHER OBSERVATIONS AND RE-ANALYSIS DATA SETS (COLOURED DOTS). THE BOX-WHISKERS DISPLAY THE MIDDLE 50% 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 AWAY FROM THE MEDIAN). THE TIME PERIOD USED IS 1986–2005.



in mountainous regions, which is likely due to model resolution being insufficient to simulate local orographic enhancement of rainfall.

Figure 5.2.5 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.

5.2.2 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 (see Box 5.1; Watterson, 1996) which has also been used for the previous *Climate Change in Australia* projections (CSIRO and BOM, 2007).

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 each model) and Watterson *et al.* (2013b)

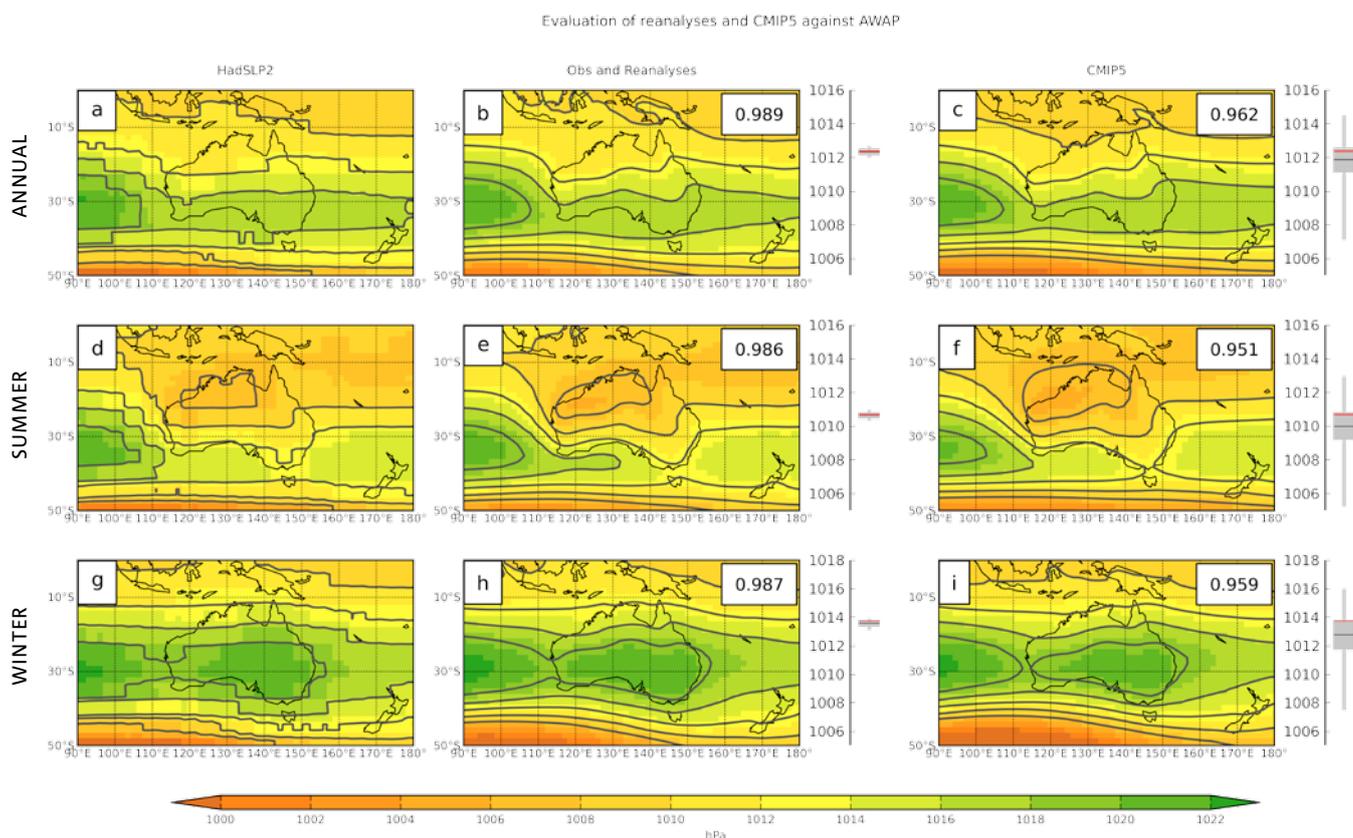


FIGURE 5.2.5: CLIMATOLOGICAL MEAN SEA LEVEL PRESSURE FROM HADSLP2 (A, D, G, USED AS THE REFERENCE DATA SET, SEE TABLE 5.1), THE AVERAGE OF A SELECTION OF OTHER OBSERVATIONAL DATA SETS (B, E, H, SEE TABLE 5.2.1) AND THE CMIP5 ENSEMBLE MEAN OF THE MODELS (C, F, I) FOR ANNUAL (TOP ROW), SUMMER (DEC-FEB, 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 X1, X2, AND X3 CONTOURS ARE HIGHLIGHTED FOR BETTER COMPARISON. THE NUMBER IN THE TOP RIGHT CORNER INDICATES THE SPATIAL CORRELATION BETWEEN THE CORRESPONDING DATA AND HADSLP2 (WITH VALUES CLOSER TO ONE 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.

-20° -10° 0° 10° 20° 30° 40° 50°

applied tests of various features of climate (such as the subtropical jet).

The calculations were done for the super-cluster regions (see Figure 2.3) used in this assessment: Southern Australia (SA), Eastern Australia (EA), Northern Australia (NA), and the Rangelands (R). The overall average of the M-scores (for three variables and four seasons) for each region and each model are given in Table 5.2.2. The score is out of a maximum of 1000. Most CMIP5 models show considerable skill in each region. 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. Watterson *et al.* (2013a) gave results for 24 models in CMIP3. The top results are a little lower than for CMIP5, with differences from 14 to 111 points, as can be seen in Table 5.2.2. The means show a consistent, and larger, improvement for CMIP5, by 57 points for AUS. In fact, several CMIP3 models have poor scores, lowering the CMIP3 mean considerably.

The best performing 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.

BOX 5.1: M STATISTIC FOR AGREEMENT

The *M* statistic or skill score is used as a metric for agreement between simulated and observed climatological fields over a particular region (Watterson, 1996). For the model field *X* and observed field *Y*, the *M* score is given by

$$M = (2/\pi) \arcsin[1 - mse / (V_x + V_y + (GX - G_y)^2)],$$

with *mse* the mean square error between *X* and *Y*, and *V* and *G* the spatial variance and regional mean, respectively, of the fields (as subscripted).

Like the correlation coefficient, *M* is non-dimensional and varies between +1 and -1, with 0 signifying no agreement. However, with *M*, 1 is reached only for identical fields (*mse* = 0), and *M* approaches 1 more slowly (for instance, global temperature fields typically correlate at *r* = 0.99, while *M* is 0.90). Tabulated values are multiplied by 1000. Averages of *M*-scores for the four seasons, for individual variables over the Australian region, are given in Table 5.2.2. Note that the *M* statistic is x1000. The average of scores for the three standard variables temperature, precipitation and pressure gives the score for overall skill in simulating the present climate for each model in Table 5.2.2. This overall score was used as a weight for each model in calculating projections by the 2007 method (see Section 6.2). Watterson *et al.* (2013a) further compares scores from CMIP5 and CMIP3 models over all continents.

TABLE 5.2.2: 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 VALUES IN BLUE. ALSO SHOWN ARE THE OVERALL AVERAGES AND TOP MODEL SCORE FOR THE CMIP5 ENSEMBLE AS WELL AS FOR CMIP3 FOR COMPARISON.

MODEL	AUS	SA	EA	NA	R
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

5.2.3 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 5.2.6 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 5.2.6 bottom right) 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.

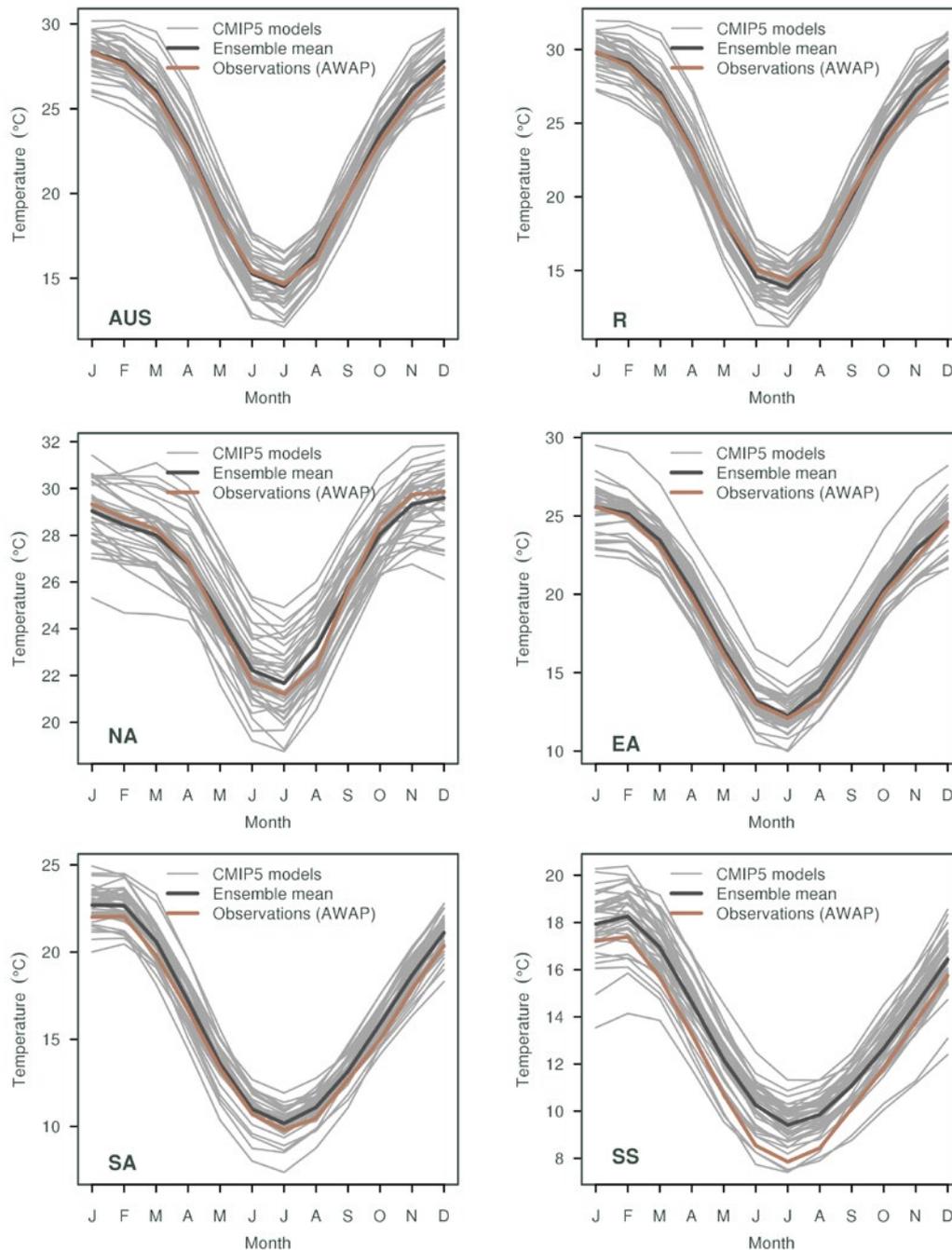


FIGURE 5.2.6: AVERAGE ANNUAL CYCLES OF SURFACE AIR TEMPERATURE FOR AUSTRALIA (TOP LEFT) AND SELECTED REGIONS (EAST AUSTRALIA - EA, NORTH AUSTRALIA - NA, RANGELANDS - R, SOUTHERN AUSTRALIA - SA, AND SOUTHERN SLOPES - SS) FROM CMIP5 MODELS. EACH GREY LINE REPRESENTS A MODEL SIMULATION, THE BLACK LINE BEING THE ENSEMBLE MEAN AND OBSERVATIONS (AWAP) SHOWN AS A BROWN LINE. THE AVERAGING PERIOD IS 1986–2005.

Figure 5.2.7 shows the corresponding rainfall annual cycles for the same regions. 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 cluster, Figure 5.2.7 bottom right), some models show even a reversed annual cycle (also see Figure 7.2.7).

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 5.2.8 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 and CNRM-CM5. Those that have at least 30 % higher STRMSE than the median error are: CESM1-BGC, CCSM4, NorESM1-ME, NorESM1-M, CESM1-WACCM and CMCC-CMS.

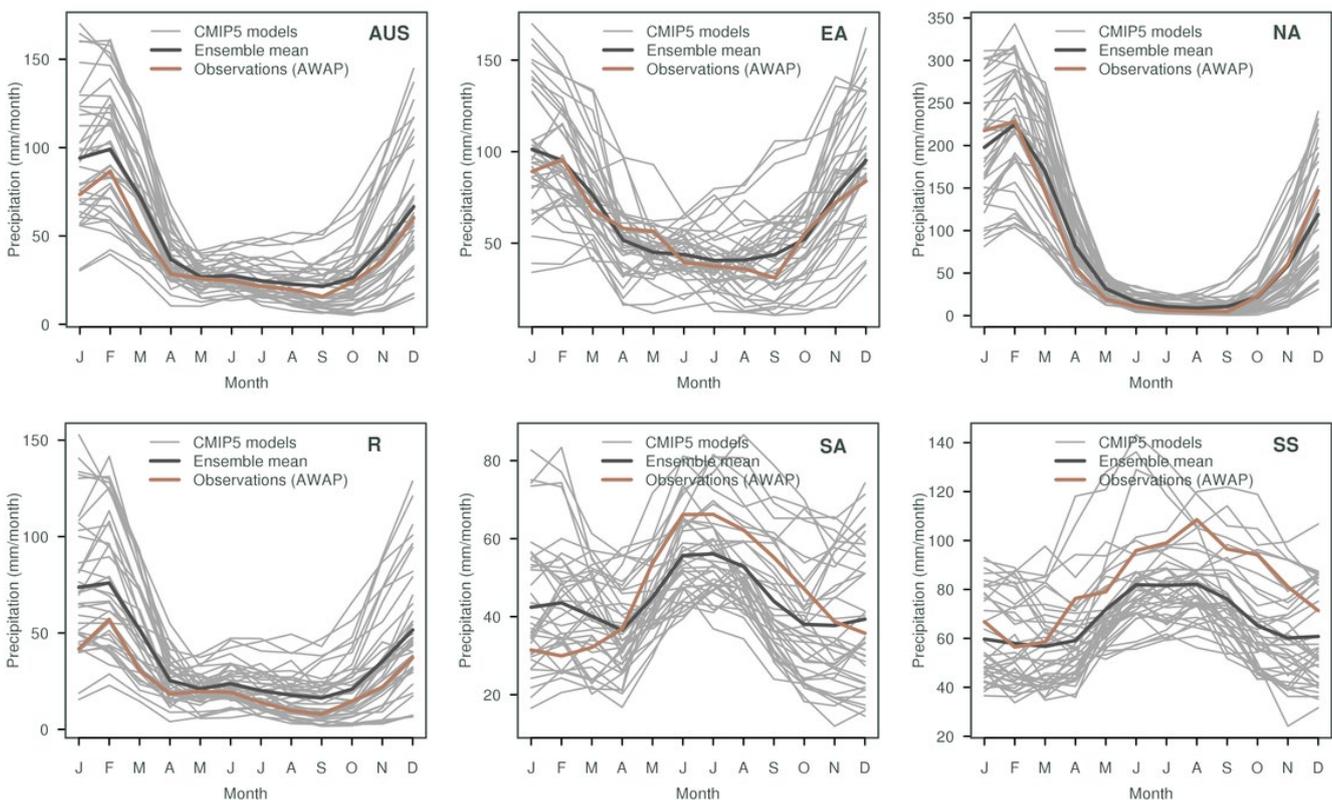


FIGURE 5.2.7: AVERAGE ANNUAL CYCLES OF RAINFALL FOR AUSTRALIA (TOP LEFT) AND SELECTED REGIONS (EAST AUSTRALIA - EA, NORTH AUSTRALIA - NA, RANGELANDS - R, SOUTHERN AUSTRALIA - SA, AND SOUTHERN SLOPES - SS) FROM CMIP5 MODELS. EACH GREY LINE REPRESENTS A MODEL SIMULATION, THE BLACK LINE BEING THE ENSEMBLE MODEL MEAN AND OBSERVATIONS (AWAP) SHOWN AS A BROWN LINE. THE AVERAGING PERIOD IS 1986–2005.

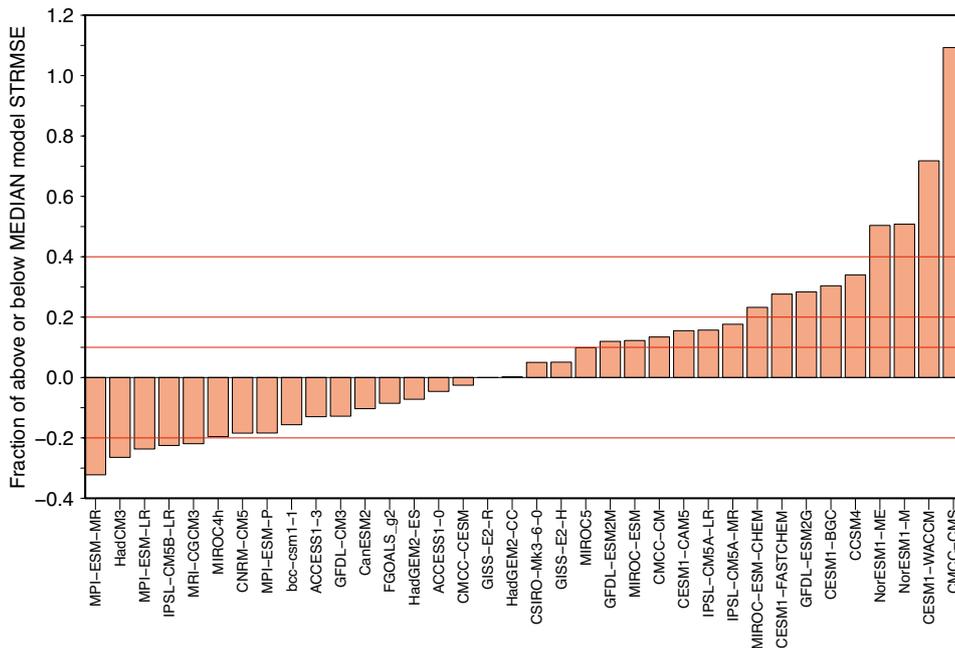


FIGURE 5.2.8: SPACE-TIME ROOT MEAN SQUARE ERROR OF CMIP5 MODELS RAINFALL ACROSS AUSTRALIA. THE ERROR IS SCALED TO SHOW THE FRACTIONAL ERROR HIGHER OR LOWER THAN THE MEDIAN MODEL ERROR. VALUES BELOW ZERO INDICATE BETTER THAN MEDIAN ERROR AND VALUES ABOVE ZERO HIGHER THAN MEDIAN ERROR. THE RED DOTTED LINES DISPLAY THRESHOLDS OF 10, 20 AND 40 % ABOVE THE MEDIAN ERROR (1.13 MM/DAY).

5.2.4 ASSESSMENT OF ADDITIONAL CLIMATE FEATURES AND ASSOCIATED SKILL SCORES

INTRODUCTION

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 5.2.3. The variables surface air temperature (tas) and precipitation (pr), without being averaged as in Table 5.2.2, 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 5.2.3 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).

TABLE 5.2.3: 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.

FEATURE	FIELDS- CMIP5 NAME	DOMAIN
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)
Subtropical 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 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 5.2.4 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 skillful of all features listed in Table 5.2.3 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 corresponding scores from the rsds field may not be reliable). 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 5.2.3) 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 5.2.3 for domain details). Wind data from ERA-Interim are used as representing the observations.

The scores for the nine tests are given in Table 5.2.4. Quantities with smaller 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. This confidence is part of the overall assessment of projected changes in Australia's climate (see Section 6.4). 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.

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, see also Section 4.1). 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.* 2012, 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 5.2.9 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 skillful 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).



TABLE 5.2.4: 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'.

MODEL	TAS	PR	RSDS	DTR	WIND850	WIND200	PSL	ST JET	MONS ONS
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

-20° -10° 0° 10° 20° 30° 40° 50°

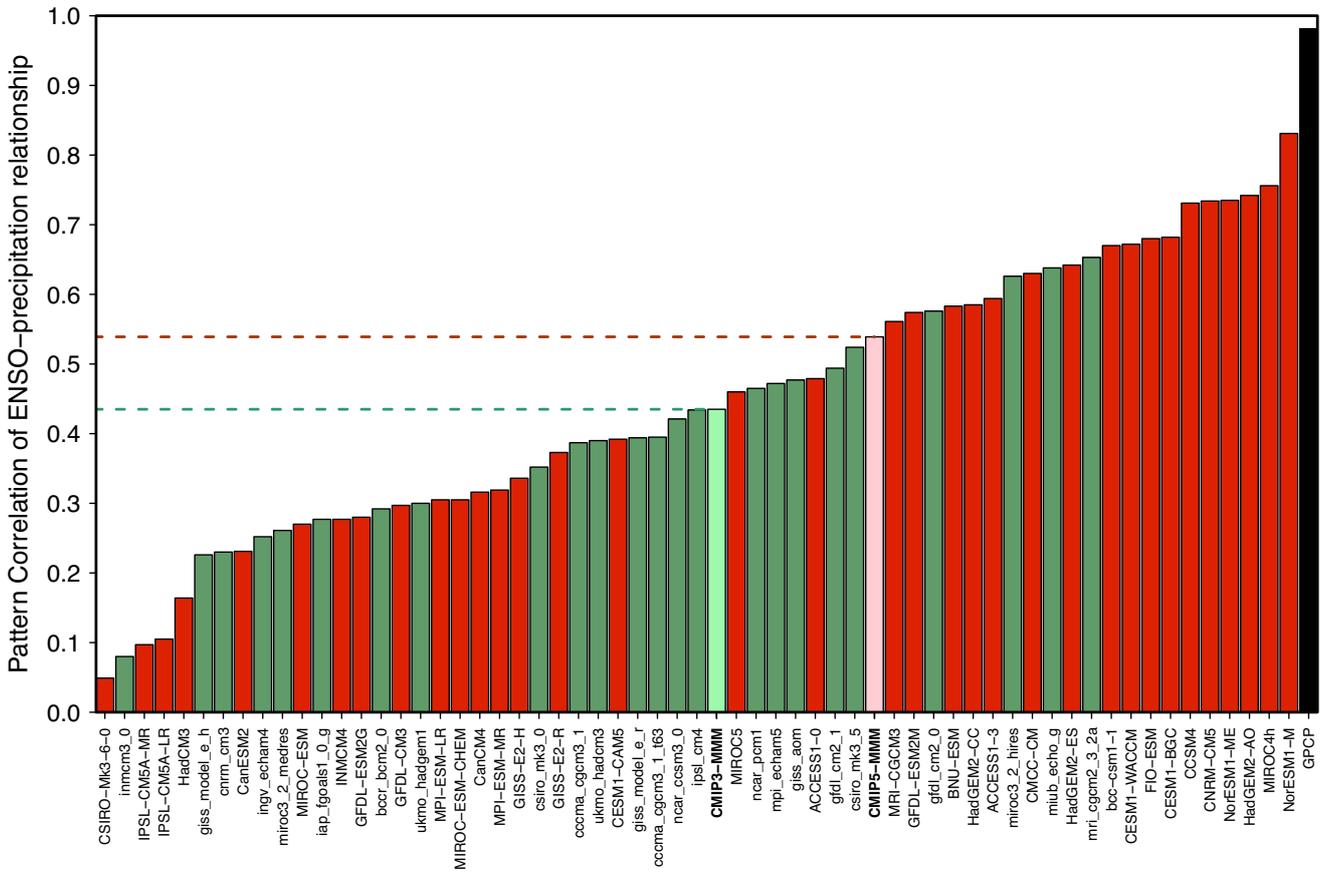


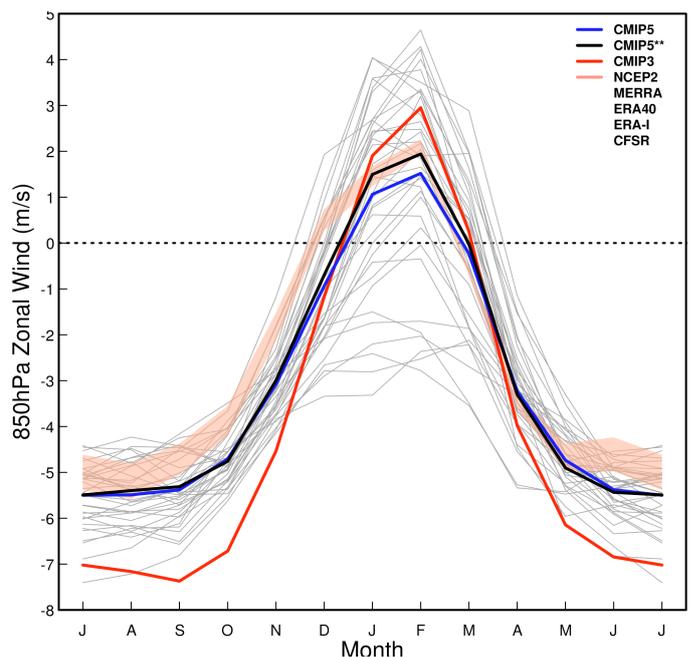
FIGURE 5.2.9: PATTERN CORRELATIONS (AGAINST THE CMAP-HADISST REFERENCE PATTERN) OF THE ENSO-AUSTRALIAN-RAINFALL TELECONNECTION PATTERN FOR EACH CMIP5 (RED) AND CMIP3 MODEL (GREEN). A SECOND OBSERVED DATA SET (GPCP-HADISST, BLACK BAR) IS SHOWN ON THE RIGHT WHILE THE MODELS ARE ORDERED IN INCREASING SKILL 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).

The 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

FIGURE 5.2.10: 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.



poorly represent the characteristics of the monsoon 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 5.2.7 top right for Northern Australia). Figure 5.2.10 shows the annual cycle of low level zonal winds for CMIP5 models and several reanalysis data sets. On average the models reversal to westerlies starts later than in the reanalysis, 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 5.2.8) 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 5.2.11 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 skilful 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 5.2.4 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 east-west moving pressure systems or East Coast Lows, and blocking weather regimes are often associated with extreme rainfall events (see Section 4.1.2 and also 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

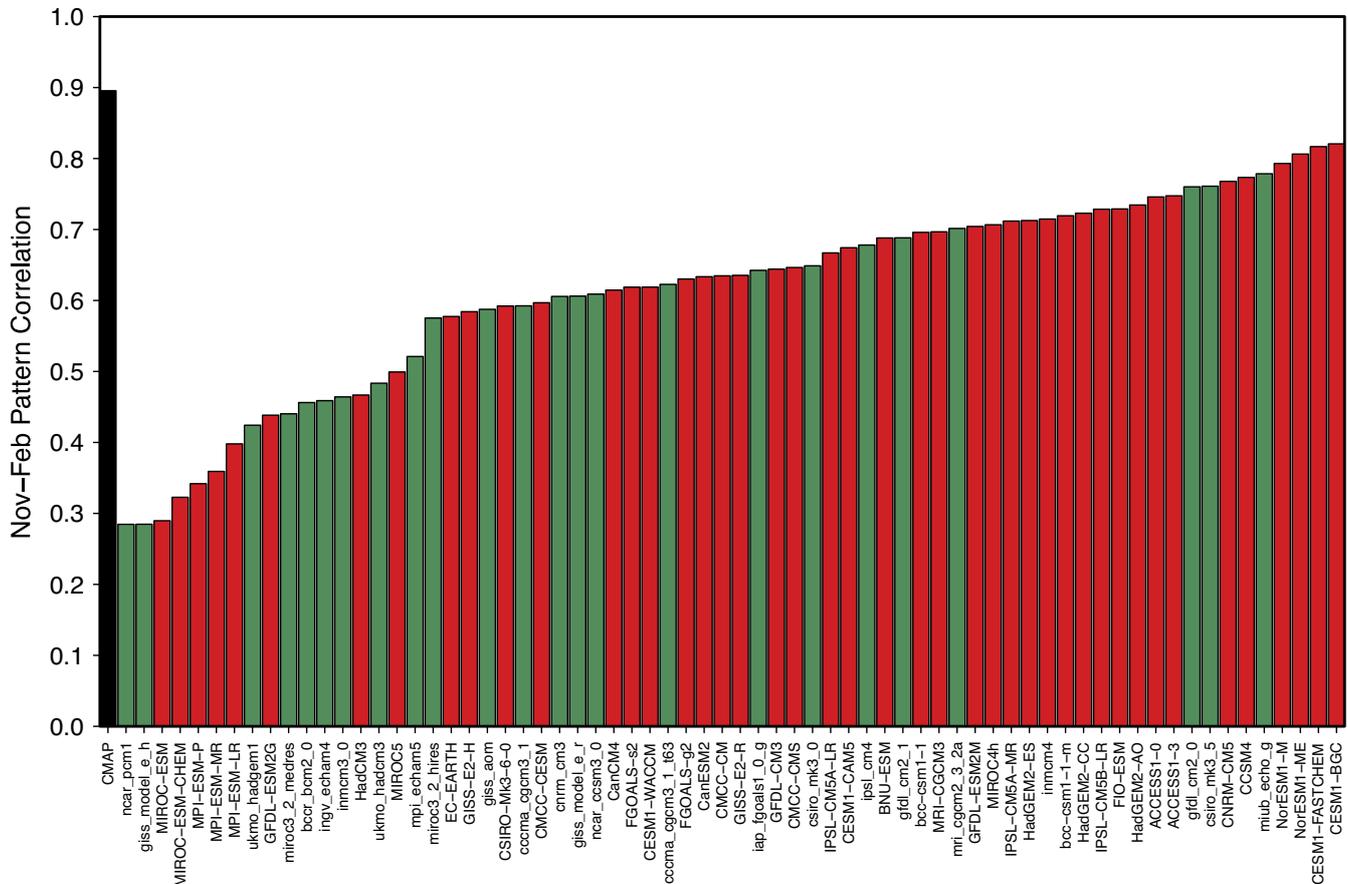


FIGURE 5.2.11: WET SEASON (NOV-FEB) RAINFALL PATTERN CORRELATIONS (AGAINST GPCP REFERENCE DATA SET) FOR CMIP5 (RED) AND CMIP3 MODELS (GREEN) OVER TROPICAL AUSTRALIA. A SECOND OBSERVED DATA SET (CMAP, BLACK BAR) IS SHOWN ON THE LEFT WHILE THE MODELS ARE ORDERED IN INCREASING SKILL TOWARDS THE RIGHT.

-20° -10° 0° 10° 20° 30° 40° 50°

models 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).

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):

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

Where u_{ax} 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 Section 4.1.2 and also 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) (see Section 4.1.2) 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.

One 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 5.2.12. 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 5.2.12). 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).

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; see Section 4.1).

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).

Sperber and Kim (2012) provided a simplified metric synthesising the skill of CMIP3 and CMIP5 model results. 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 east 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 subtropical high pressure belt which separates the mid-latitude westerly winds to the south and the south-east trade winds to the north (see Section 4.1.2 for more information on subtropical ridge). 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 (see Figure 4.1.1).

The evaluation 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 5.2.1 for overview of reanalysis data sets). The annual cycle in the pressure and latitude of the subtropical high pressure belt known as the subtropical 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). Preliminary results indicate that results are similar in CMIP5 models (not shown).

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 5.2.10 shows the annual cycle of low level zonal winds for CMIP5 models and several reanalysis data sets. As mentioned above, 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 earlier, several models fail to simulate monsoon westerlies over northern Australia altogether.

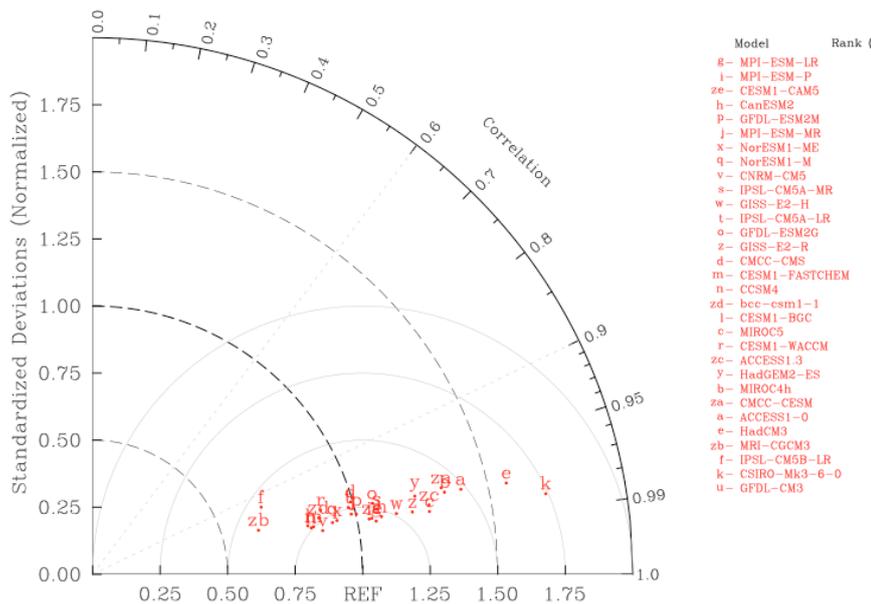


FIGURE 5.2.12: TAYLOR PLOT OF SPATIAL STATISTIC OF SEA SURFACE TEMPERATURES FROM CMIP5 MODELS OVER THE TROPICAL EASTERN INDIAN OCEAN AGAINST HADISST OBSERVED SEA SURFACE TEMPERATURES (REF POINT AT HORIZONTAL AXIS). EACH LETTER REPRESENTS ONE CMIP5 MODEL’S SIMULATION AVERAGED OVER THE PERIOD (1986-2005).



5.3 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 (see Section 3.1 for further discussion). 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 assume 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 (see Table 3.3.1) 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.* 2009a, Raupach *et al.* 2009; 2012) and CRU TS3.20 for precipitation (listed in Table 5.2.1 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 5.3.1a-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 5.3.1e-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).

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 5.3.2a-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.



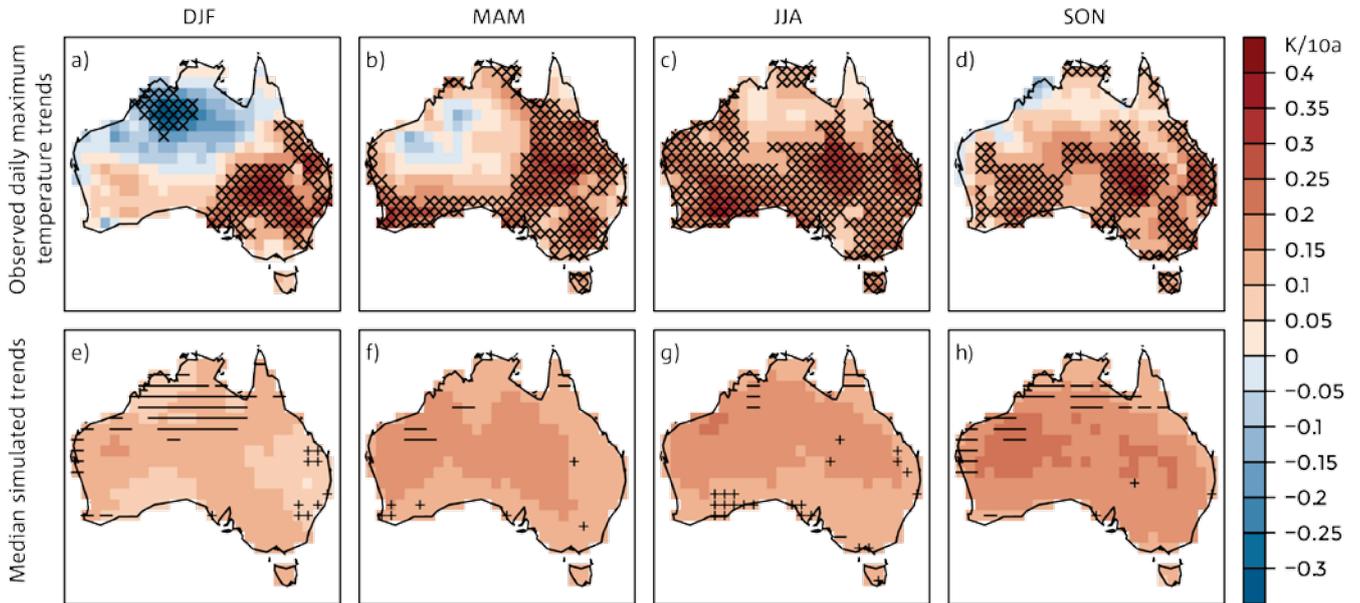


FIGURE 5.3.1: OBSERVED TREND IN SEASONAL MEAN DAILY MAXIMUM TEMPERATURE FROM 1956 TO 2005 (A-D) AND MEDIAN OF SIMULATED TRENDS FROM 42 CMIP5 MODELS (E-H). CROSSES IN A-D DENOTES AREAS WHERE THE OBSERVED TREND IS SIGNIFICANTLY DIFFERENT FROM ZERO AT THE 10 % LEVEL. CROSSES (DASHES) IN E-H DENOTE AREAS WHERE LESS THAN 10 % OF THE SIMULATED TRENDS ARE AS LARGE (SMALL) AS THE OBSERVED TREND.

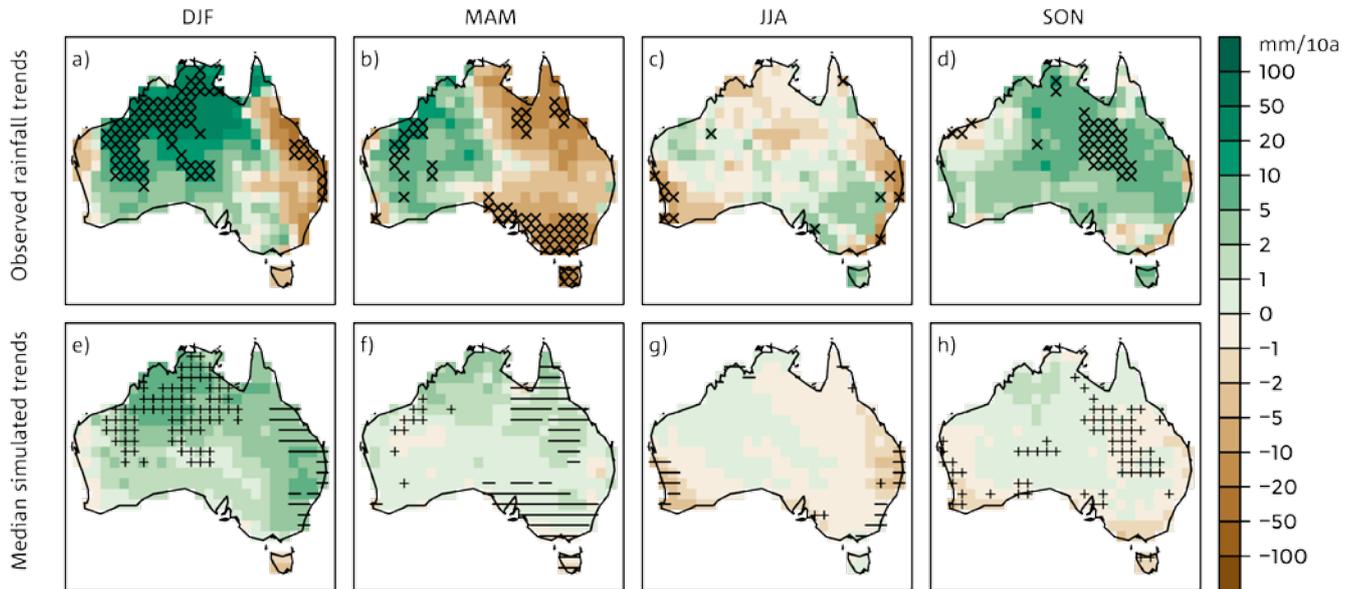


FIGURE 5.3.2: AS IN FIGURE 5.3.1 BUT FOR SEASONAL RAINFALL IN MM PER DECADE.

-20° -10° 0° 10° 20° 30° 40° 50°

5.4 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 *Fourth Assessment Report* concluded that models showed some considerable skill in simulating the statistics of extreme events (especially for temperature extremes) given 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 *Fifth Assessment Report* (Working Group 1) (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. Also, CMIP5 models are 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 (T_{xx}); annual and seasonal minimum of daily minimum temperature (T_{nn}) 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, particular across northern Australia.

A comparison between CMIP5 model daily maximum rainfall and observations is shown in Figure 5.4.1 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 5.4.1). Overall the observed daily maximum rainfall amounts here are mostly captured

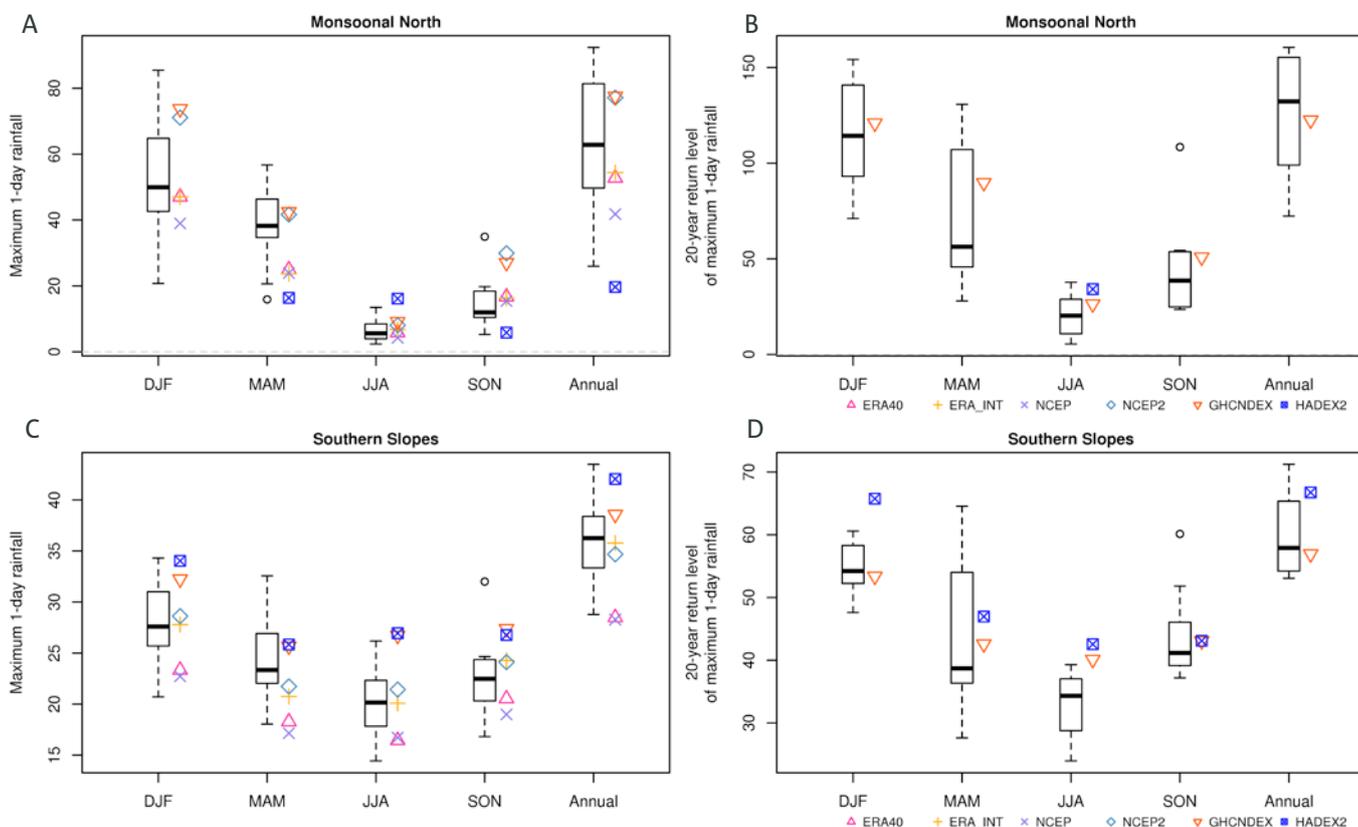


FIGURE 5.4.1: DAILY EXTREME RAINFALL (LEFT COLUMN: FOR DAILY MAXIMUM RAINFALL PER YEAR; (A) FOR CLUSTER MN AND (C) FOR CLUSTER SS); FOR DAILY MAXIMUM RAINFALL IN 20 YEARS ((B) FOR CLUSTER MN AND (D) FOR CLUSTER SS) ACROSS SEASONS AND ANNUALLY (UNITS ARE MM/DAY). CMIP5 MODELS ARE REPRESENTED BY THE BOX-WHISKER WHILE COLOURED SYMBOLS REPRESENT REANALYSIS PRODUCTS (SEE TABLE 5.2.1) AND TWO GRIDDED OBSERVATIONAL DATA PRODUCTS (SEE TEXT).

-20° -10° 0° 10° 20° 30° 40° 50°

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 5.4.1a) as well as the maximum daily rainfall event over a 20 year period (Figure 5.4.1b). 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.

For the Southern Slopes cluster, the CMIP5 models have a tendency to underestimate the maximum 1-day rainfall event during a year (Figure 5.4.1c), but are still within range. The 20-year event (Figure 5.4.1d) 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.

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. 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.

5.5 EVALUATION OF DOWNSCALING SIMULATIONS

This Report includes regional climate change projections information from two downscaling methods (described in detail in later in Section 6.3). Therefore it is necessary to also provide some information about how well these simulations perform over the historical period.

Each of the two methods used here 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).

We also assessed 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 5.5.1 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.

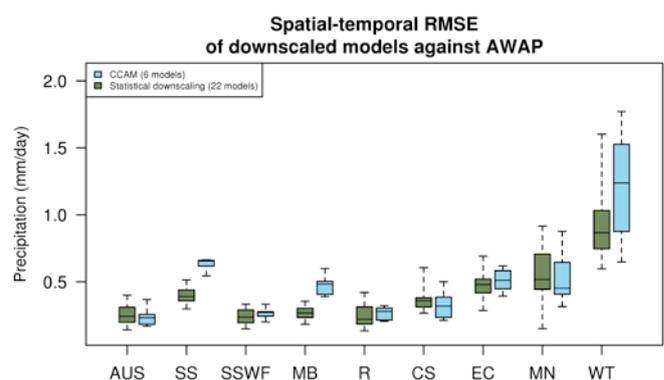


FIGURE 5.5.1: BOX-WHISKER PLOT OF THE SPATIAL-TEMPORAL ROOT MEAN SQUARE ERROR (STRMSE; LARGER VALUES INDICATE LARGER ERRORS COMPARED TO OBSERVATIONS) FOR RAINFALL FROM TWO DOWNSCALED ENSEMBLES AGAINST AWAP RAINFALL FOR AUSTRALIA AND THE EIGHT CLUSTER REGIONS. THE DOWNSCALED ENSEMBLES ARE THE BOM-SDM 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 5.5.2 shows a comparison of the two downscaling 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 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 5.2.8) to around 0.25 mm/day in downscaled simulations here) that may then reveal regional detail in the climate change signal at finer scale than GCMs can (see further discussion of these points in Sections 6.3 and 7.2).

5.6 SYNTHESIS OF MODEL EVALUATION

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.

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 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.

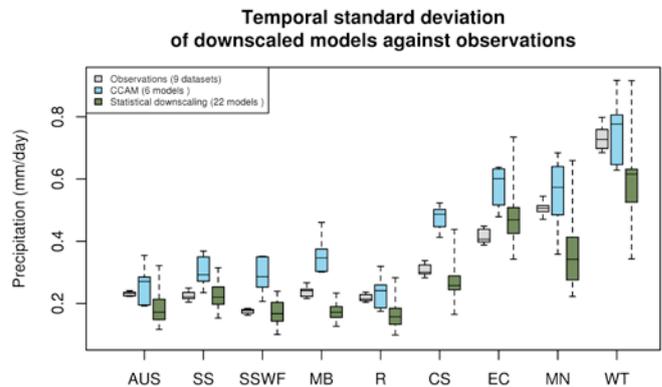


FIGURE 5.5.2: BOX-WHISKER PLOT OF THE TEMPORAL STANDARD DEVIATION OF ANNUAL RAINFALL (1986–2005) FROM TWO DOWNSCALED ENSEMBLES AND GRIDDED OBSERVATIONAL RAINFALL FOR AUSTRALIA AND THE EIGHT CLUSTER REGIONS. THE DOWNSCALED ENSEMBLES ARE THE BOM-SDM STATISTICAL DOWNSCALED ENSEMBLE (GREEN) AND THE CCAM DYNAMICAL DOWNSCALED ENSEMBLE (BLUE). THE OBSERVATIONAL DATA INCLUDES AWAP.

The GISS-E2 models (GISS-E2-H, GISS-E2-H-CC and GISS-E2-R) and MIROC-ESM models (MIROC-ESM MIROC-ESM-CC) 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 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. 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 5.2.2). 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, subtropical jet, circulation – see Table 5.2.4 and Section 5.2.4) 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

There is no conclusive evidence that CMIP5 models fail to reproduce 1956-2005 observed trends in daily maximum and minimum temperature and rainfall. The extent of the areas where discrepancies exist are generally not larger than expected due to the pronounced variability on inter-annual to decadal scales. Nevertheless, confidence in rainfall projections is inevitably reduced where consistency is low, particularly 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 applied here, 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. As will be seen in Section 6.2, in forming ranges of projected change using CMIP5, factoring in model performance (by different methods weighting or model elimination) does not have a strong affect, and is not done routinely in the ranges of projected change presented in Chapter 7 (although there are some exceptions noted in that chapter). 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 (see Section 6.4 and Chapter 7). Secondly, poor performing models are flagged in the Climate Futures tool (Chapter 9), to guard against these models being selected when forming a small set of models for use in impact assessment.

From the results of the analysis presented in the individual sections of this chapter, the following models were identified as poor performing models, for the reasons outlined (see also Table 5.6.1). All of these models should be used with caution in any projection work in 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). 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.
- 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.



TABLE 5.6.1: 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.

MODEL	Low M-Score for PR and TAS	Low STRMSE	Low score for climate features	Low ENSO – rainfall tele connection	No Monsoon westerlies	Wet season rainfall not good spatially	Rainfall relationship to Blocking not good	IOD spatial variability too low or too high	Simple MJO skill not good	
ACCESS1-0								X		1
ACCESS1-3					X		X			2
BNU-ESM	X		X			X				3
CanCM4							X			1
CanESM2				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		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