

# Consistent assessment of global climate model simulations for hydrological application in south eastern Australia

Dewi G.C. Kirono<sup>a</sup>, D.M. Kent<sup>a</sup>, F.H.S. Chiew<sup>b</sup> and J. Teng<sup>b</sup>

<sup>a</sup> CSIRO Water for a Healthy Country National Research Flagship/Centre for Australian Weather and Climate Research, a partnership between CSIRO and the Bureau of Meteorology, Australia

Email: [dewi.kirono@csiro.au](mailto:dewi.kirono@csiro.au)

<sup>b</sup> CSIRO Water for a Healthy Country National Research Flagship/CSIRO Land and Water, Australia

**Abstract:** For a number of reasons, it is necessary to assess global climate models (GCMs) that will be used for estimating the climate change impact on runoff characteristics. Previous studies have used different sets of GCMs and considered many different temporal periods, had different geographical foci or vastly different evaluation approaches. As a result, there is a lack of comparability among studies. This study assesses the GCM's ability to reproduce observed climate variables and the climate drivers affecting rainfall over the South Eastern Australian region using a consistent evaluation skill scores. Such a consistent approach enables one to compare the results from a variety of measures in order to develop a robust framework for GCM selection.

The climate variables being assessed include rainfall, potential evapotranspiration, temperature and mean sea level pressure, while the climate drivers being examined include the Sub-Tropical Ridge (STR), the El Nino Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD) and the Southern Annular Mode (SAM). The study consistently applies some skill scores (including the *M*-Statistic of Watterson, 2008;  $\Delta_{ss}$  of Pierce et al., 2009; and coefficient correlation) to measure the GCMs' reliability on a given field (e.g. mean annual rainfall). However, only the *M*-Statistic is shown here. The preliminary results show that (i) the GCMs generally reproduce the observed mean and coefficient of variation (Cv) of annual climate variables, but not the trend. There is no ground for assuming that performance on long term average is an indicative of performance on long term trend or vice versa. There is also generally very little difference between the *M*-Statistic of different GCMs, particularly for the mean and Cv; (ii) Overall, the GCM representations for the spatial pattern of observed IOD and SAM indices are poorer than their respective ENSO index. None of the models shows consistently good performance on all examined climate drivers, but there are models that indicate consistently poor performance. Therefore, model selection may be best conducted by rejecting models which are consistently poor rather than identifying models that are consistently good across examined climate drivers; and (iii) Consideration of model performance (as quantified by the *M*-Statistic) in projecting future mean annual rainfall over the MDB has relatively little impact. In this regard, the range of future climate scenarios that are used in hydrological impact studies may probably be best constructed from the majority of available archived GCMs simulations.

**Keywords:** *climate change impact, global climate models, rainfall, GCM assessment, GCM selection*

## 1. INTRODUCTION

There are two main steps involved in estimating climate change impacts on runoff characteristics and water availability. The first step is to use global climate model (GCM) simulations and downscaling techniques to obtain catchment-scale projected rainfall time-series. These are then used to drive hydrological models. A set of 25 GCMs is currently available in the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset. Due to limitations on resources it is not always possible to undertake both steps for all 25 simulations in order to develop an ensemble of climate change impacts on runoff characteristics. Therefore, it is necessary to assess the GCMs and select or weight the better GCMs for specific hydrological applications.

There are many studies with different views and approaches on GCM evaluation and selection. Some (e.g. Kirono and Kent, 2011; Vaze *et al.* 2011) argue that one needs to only consider evaluation and selection criteria based on climate variable of interest (for most hydrological applications this are rainfall and potential evapotranspiration), while others (e.g. Smith and Chandler, 2009) suggest the use of a broader suite of variables, including the ability of GCM to represent atmospheric-oceanic climate drivers such as the Southern Annular Mode (SAM) and the El Nino Southern Oscillation (ENSO). This is because climate variables (e.g. rainfall) are associated with the dynamic of climate drivers, and a successful simulation and prediction over a wide range of these drivers increase confidence in the GCM used for climate projections in the future (Randall *et al.* 2007).

With regard to climate variables, existing studies usually assess the GCMs ability to simulate the long term mean of observed climate variables such as rainfall, mean sea-level pressure and temperature (e.g. Suppiah *et al.*, 2007). Regarding climate drivers, the studies commonly assess the GCMs performance in simulating climate drivers such as ENSO and the Indian Ocean Dipole (IOD) and their teleconnection with rainfall (e.g. Cai *et al.* 2009a). These studies often use different sets of GCMs, consider different temporal outputs and geographical areas or use different evaluation approaches and metrics in their assessments. This makes it difficult to synthesize and to apply their results in selecting GCMs that can be considered for developing rainfall projections to be used in hydrological modelling.

This study uses a consistent approach to evaluate and select the GCMs for subsequent use in hydrological impact modelling over the South Eastern Australian (SEA) region. Specifically, this study assesses the GCMs' ability to reproduce (i) observed climate variables; (ii) main climate drivers of rainfall in SEA; and (iii) teleconnection between those climate drivers and SEA rainfall. Furthermore, the study examines implications of model performance considerations on rainfall projections.

## 2. ASSESSMENT BASED ON CLIMATE VARIABLES

Rainfall (P) and potential evapotranspiration (PET) are key input variables to catchment water balance assessments. Other important variables include mean-sea level pressure (MSLP) and air temperature (T) which are both commonly used as predictors in statistical downscaling techniques. Therefore, this study begins by assessing the ability of the GCMs to simulate P, PET, T and MSLP.

MSLP is assessed over the Australian region, while the other variables (P, T and PET) are each assessed over the Murray Darling Basin (MDB) and over each of the three climate-regions within the SEA region as defined in Timbal and Fernandez (2009), i.e. the northern part of the MDB (NMDB), the eastern sea-board (ESB) and the south-west of Eastern Australia (SWEA). The ability of each GCM to represent the spatial pattern of observed monthly, seasonal and annual mean, coefficient of variation (Cv) and long-term trend are quantified using a number of skill score metrics (including the *M*-Statistic of Watterson (2008), the delta skill score ( $\Delta_{ss}$ ) of Pierce *et al.* (2009) and root mean square error). However, due to a limited space in the current paper, only the *M*-Statistic, as used in the CSIRO and BoM (2007), is shown here. The *M*-Statistic is a non-dimensional skill score that can be used to quantify the similarity of the gridded model to the gridded observed fields as follows (Watterson, 2008):

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

With *mse* being the mean square error between the model field *X* and observed field *Y*, *V* is the variance and *G* is the mean of the subscripted field. All statistics are calculated over the set of spatial points within the region.

The modelled MSLP data are assessed against the MSLP reanalysis database of the National Center for Environmental Prediction (NCEP) (Kalnay *et al.*, 1996). The observed rainfall and temperature data come from the Australian Bureau of Meteorology’s (BoM) high-quality data set (Jones and Beard, 1998). PET is estimated from climate variables from the SILO gridded database (Jeffrey *et al.*, 2001) using Morton’s areal evapotranspiration algorithm (Morton, 1983; Chiew and McMahon, 1991; BoM, 2001). Both the GCM and the observed data are regridded to a 1° grid size prior to the evaluation.

Figure 1 shows the *M*-Statistic summarising the ability of each of the GCMs to reproduce the mean, Cv and trend of annual P, T and PET over the MDB, as well as MSLP over Australia for 1975-2004. An *M*-Statistic of one indicates a perfect match between the modelled and observed values and a value of  $\leq 0$  indicates a very poor match. Figure 1 indicates that the GCMs generally reproduce the observed mean and Cv of the annual climate variables, but not the trend. There is no clear association among the performances on mean and performances on Cv and/or trend. For example, the correlation coefficient between the *M*-Statistic of mean and trend of P is 0.14. There is also generally very little difference between the *M*-Statistic of different models, particularly on the mean and Cv fields. For example, all GCMs have an *M*-Statistic of 0.99 for the mean T and MSLP fields. Meanwhile, the average, maximum and minimum *M*-Statistic for the mean P is 0.77, 0.86 and 0.56, respectively. This makes it difficult to identify the better or poorer GCMs based on the *M*-Statistic skill scores for the mean and Cv of climate variables assessed here.

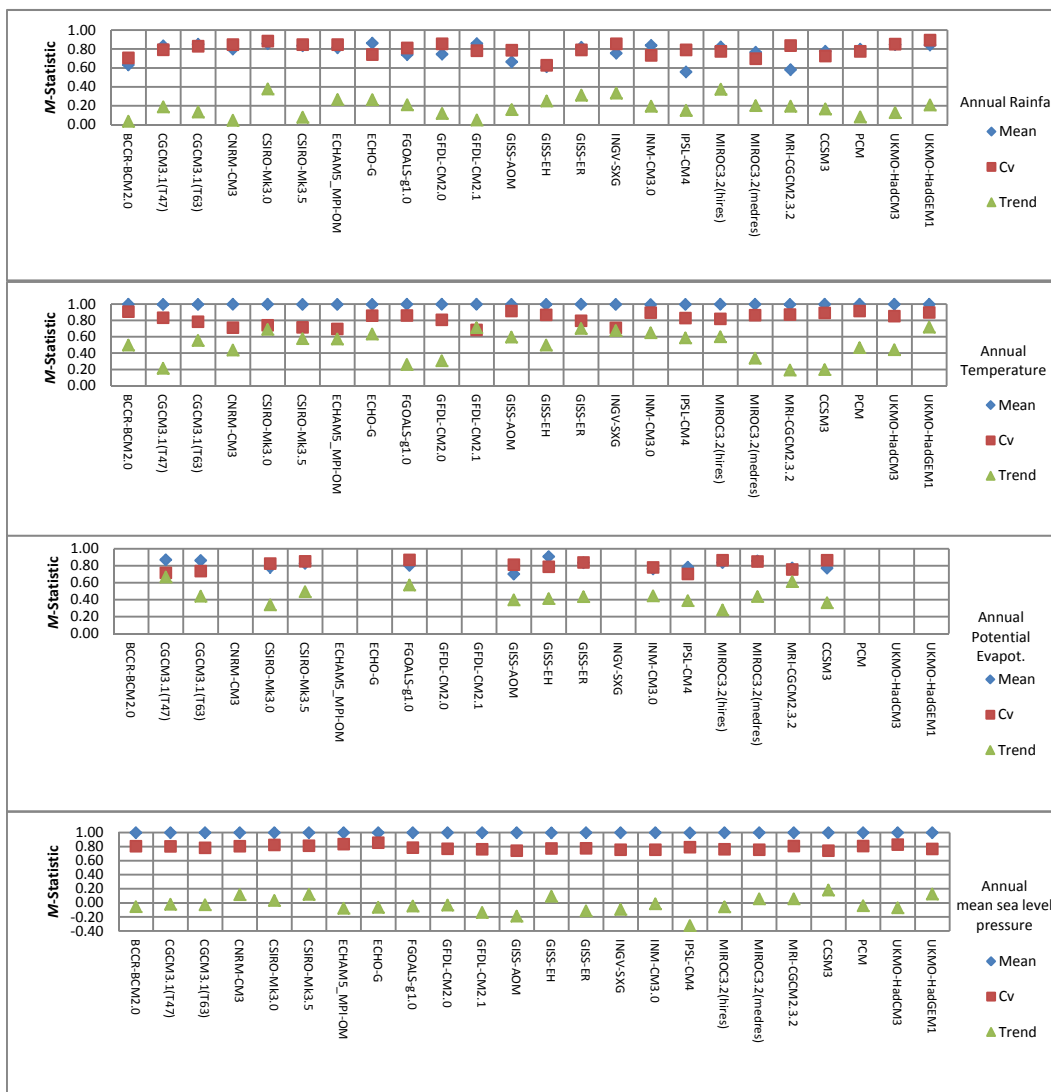


Figure 1. GCM’s ability to simulate observed rainfall (P), temperature (T), potential evapotranspiration (PET) (mean annual value, Cv of annual values and trend) across the MDB and mean-sea level pressure across Australia as measured by the *M*-Statistic.

### 3. ASSESSMENT BASED ON CLIMATE DRIVERS AND THEIR TELECONNECTION WITH RAINFALL

The climate drivers being investigated include the ENSO, the IOD, the SAM and the sub-tropical ridge (STR). In this study the ENSO and IOD modes of variability are diagnosed as the first Empirical Orthogonal Function (EOF) of sea surface temperature over the Pacific and Indian oceans respectively while the SAM is the first EOF mode of mean sea level pressure over the Southern Ocean (Cai *et al.*, 2009b). The STR derivation follows that of Drosowsky (2005). The modelled climate driver indices from the GCMs are compared to the observations over the period 1950-1999. The observation data for the ENSO and IOD are obtained from the HadISST sea surface temperature database, and the SAM data are obtained from the NCEP-reanalysis MSLP data. Two datasets are used to compare the GCM simulations of STR: the BoM dataset (Drosowsky, 2005) which is based on station data; and the dataset derived from MSLP fields of the NCEP-reanalysis. The *M*-Statistic, the  $\Delta_{ss}$  and coefficient correlation (R) are used to quantify the ability of each of the GCMs in reproducing the spatial pattern of each of the indices. The same skill-scores are also used to quantify the GCMs modelled teleconnection, the linear correlation between each of the indices and the SEA rainfall (see also Kent *et al.* (2011) for a more detailed assessment of the STR simulations).

Figure 2 shows the ability of the GCMs to reproduce the spatial pattern of the observed ENSO index. Overall, most GCMs represent the historical spatial mean of the ENSO index reasonably well, with *M*-Statistic values mostly larger than 0.5. The GISS-AOM is the poorest model (with an *M*-Statistic of less than 0.4), which is consistent with the results of Irving *et al.* (2011) who rank this GCM in the lowest bottom. The results for the IOD and SAM are not shown here, but in general the GCM representation of the IOD and SAM are poorer than their respective ENSO representation. For example, the average *M*-Statistics of all models for the ENSO, IOD and SAM indices are 0.59, 0.34 and 0.04, respectively. The GISS-AOM's *M*-Statistic for IOD and SAM are 0.46 and 0.05.

Figure 3 summarises the *M*-Statistic that quantifies the GCMs' ability in representing the teleconnection between the annual STR-Intensity and SEA rainfall. Apart from a few GCMs with very high *M*-Statistic values (i.e. MIROC-medres and IPSL-CM4), many GCMs have relatively low *M*-Statistic ( $\leq 0.4$ ). The preliminary results for the teleconnection between ENSO, IOD and/or SAM with the SEA rainfall are relatively similar to this (not shown here).

Overall, there seems no single model that is consistently good across all the examined climate drivers and their teleconnection with rainfall. For example, the MIROC3.2(hires)'s *M*-Statistic for ENSO, IOD, SAM and STR-rainfall teleconnection are 0.71, 0.43, 0.1 and 0.66, respectively while the CSIRO-MK3.5's are 0.65, 0.32, 0.26 and 0.6, respectively. However, there are models that consistently poor across all the examined climate drivers. For example, the GISS-AOM has an *M*-Statistic of 0.20, 0.46, 0.05 and 0.23 for the ENSO, IOD, SAM, and STR-rainfall teleconnection, respectively. This suggests that identifying/rejecting models which are consistently poor is more feasible than those which are consistently good across examined variables.

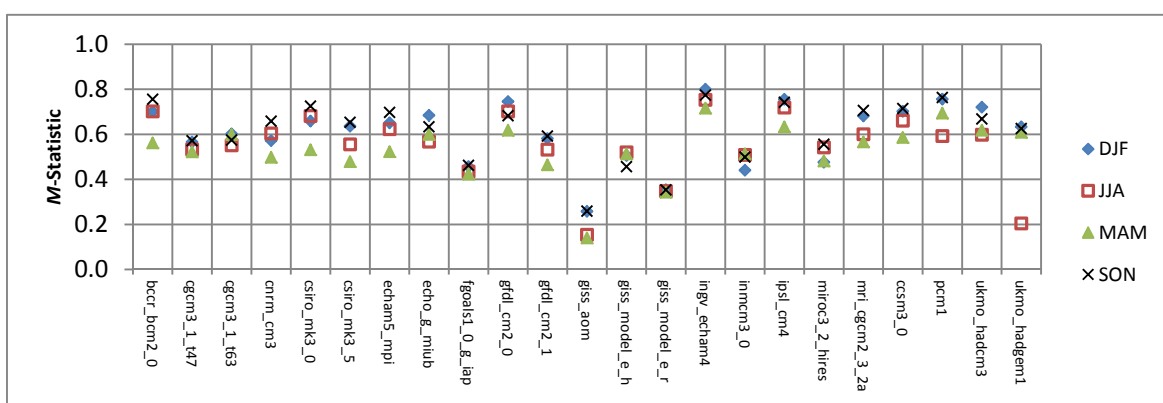


Figure 2. *M*-Statistic summarising the GCMs' ability to simulate the observed spatial pattern of mean seasonal ENSO index for 1951-2007.

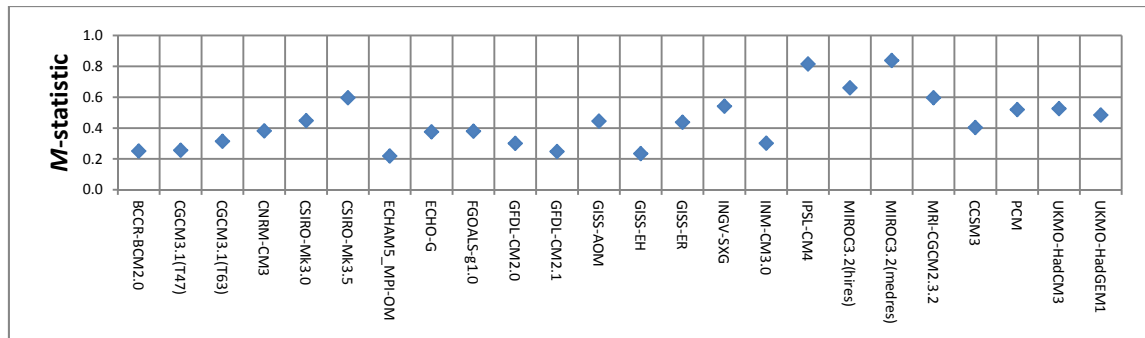


Figure 3. *M*-Statistics summarising the GCMs' ability to simulate the observed teleconnection between the STR-Intensity and SEA rainfall during 1948-2002.

#### 4. IMPLICATIONS ON RAINFALL PROJECTIONS

There are three common practices in developing climate projections based on the output of multiple models: (i) use all the available models, with an assumption that they are all equally good; (ii) use all the models, but weight each of the models based on a given performance; and (iii) select/reject models according to a certain criteria. However, there is no a standard method as to how assigning the weights and/or to select/reject models. Arguably, different approaches and considerations will results in different projections.

To examine the impact of performance considerations (PR, T, MSLP, ENSO, etc) — for selecting a set of best models for rainfall projections, we start by ranking the models according to their *M*-Statistic and select the top five models for a particular variable (e.g. ENSO). Since no model is consistently good across examined variables, the top five models can be slightly different for variables. For instance, based on the *M*-Statistic of MSLP, the top five models are CNRM-CM3, CSIRO-MK3.5, MRI-CGCM3.2, NCAR-CCSM and UKMO-Hadgem1 whereas those based on the STR-teleconnection' skill score are MIROC3.2(medres), IPSL-CM4, MIROC3.2(hires), MRI-CGCM3.2 and CSIRO-MK3.5. Secondly, we calculated the range (10<sup>th</sup>, median and 90<sup>th</sup> percentile) of percent change in future mean annual rainfall over the MDB based on all GCMs and based on the top five models (selected based on each variables performance) (as shown in Figure 4). The figure indicates: (i) with regard to the median, all considerations suggest a decrease in future rainfall of around 4%, with the ENSO- and IOD-based selections indicating a lesser degree of drying (i.e. around 1%); (ii) regarding the ranges of projection, there is no distinct difference among different considerations. The range of uncertainty (i.e. the 90<sup>th</sup> minus 10<sup>th</sup> values) for most of them is around 10%, except that of the MSLP-based selection (i.e. around 6%); and (iii) while others imply some possibility of increase in rainfall, the MSLP- and SAM-based selections do not. The same experiment, but with the set of top 10 instead of top 5 models, also indicates similar results (not shown here). The exception being the MSLP-based selection gives relatively similar ranges of uncertainty compared to others (i.e. about 10%).

In order to understand the implication of different weighting approaches (i.e. equally-weighted or unweighted versus performance-based-weighted approaches), we weight the models according to their *M*-Statistic for each of the variables (shown as a crossed symbol in Figure 4). For example, the cross symbol in the PR-column shows the projections based on all models—but weighted with the *M*-Statistic for PR whereas the cross symbol in the All-GCMs-column represent the projections based on all models-equally weighted. The figure suggests that the weighted-based approach gives relatively similar results in comparison to that of the equally-weighted (or non-weighted) approach. For instance, both the equally-weighted and performance-based-weighted approaches suggest a decrease of around 4% in future rainfall. The exception being the SAM-performance-weighted approach which suggest a 7% decrease. Such a similar projection is a reflection of the very little difference between the *M*-Statistic of different models (as mentioned previously).

All of these results suggest that application of different approaches (equally-weighted, performance-based-weighted, and performance-based-selection) and different considerations (performance on PR, T, MSLP, ENSO, etc) will results in dissimilar ranges of rainfall projections over the MDB. However, for most cases, the dissimilarities are relatively small (more or less 3%). This suggest that, based on the *M*-Statistic skill score, model performance consideration in projecting future mean annual rainfall over the MDB has relatively little implication.

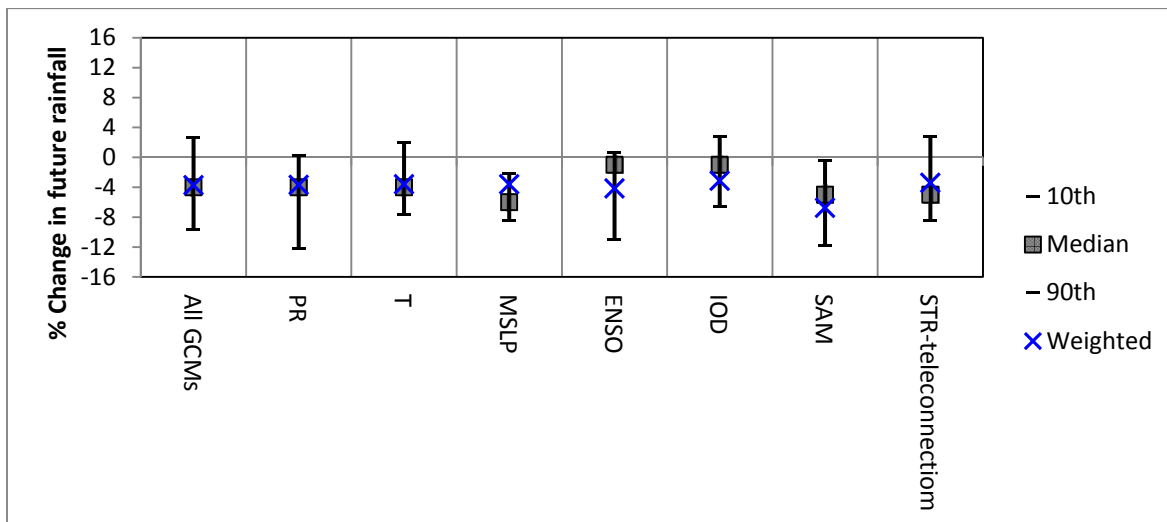


Figure 4. Ranges (10<sup>th</sup>, median and 90<sup>th</sup>) of percent changes in mean annual rainfall per degree global warming calculated from different sets of GCMs (All GCMs, and the top five GCMs selected based on their performance on PR, T, MSLP, ENSO, and so on). The cross symbols indicate projections based on all GCMs but weighted according to a given performance (PR, T, and so on) the cross symbol in the ‘All GCMs’ column shows the equally-weighted projection.

### 5. SUMMARY

Unlike most previous studies, this study assesses GCMs ability to reproduce not only the long-term mean but also Cv and trend of climate variables. This study also quantifies the ability of GCMs in reproducing the spatial pattern of some climate drivers (e.g. ENSO and IOD) and their teleconnection with SEA rainfall, using a consistent evaluation skill scores, so that the results can be comparable. Based on the *M*-Statistic skill scores calculated for a broad suite of climate variables and climate drivers relevant to the South East Australia region, we observed the followings:

- The GCMs reproduce the spatial pattern of observed mean and Cv of the annual climate variables (P, T, PET, MSLP) relatively well, but not the trend. For instance, the average *M*-Statistics for mean, Cv and trend for P are 0.77, 0.8 and 0.19, respectively. Since the coefficient correlation between the *M*-Statistic on mean and trend is very small (e.g. for P it is only 0.14), there is no ground for assuming that performance on long term average is an indicative of performance on long term trend or vice versa. There is also generally very little difference between the *M*-Statistic of different models, particularly on the mean and Cv fields. This makes it difficult to identify the better or poorer GCMs based on the *M*-Statistic skill scores. Furthermore, when the *M*-Statistics are applied to weight the multiple models used in developing rainfall projection, the results are similar to those based on equally-weighted models.
- Overall, the GCM representations of the spatial pattern of observed IOD and SAM indices are poorer than their respective ENSO index. The average *M*-Statistics for the ENSO, IOD and SAM indices are 0.59, 0.34 and 0.04, respectively. None of the models show consistently good performance on all examined climate drivers, but there are models that indicate a consistent poor performance. Therefore, model selection may be best conducted by rejecting models with consistent poor performance rather than identifying models that are consistently good across examined variables.
- Consideration of model performance (as quantified by the *M*-Statistic) in projecting future mean annual rainfall over the MDB has relatively little implication. In this regard, the range of future climate scenarios that are used in hydrological impact studies may probably best constructed from the majority of available archived GCMs simulations.

### ACKNOWLEDGMENTS

This work is funded by the South Eastern Australian Climate Initiative (SEACI). We wish to thank David Post and two anonymous reviewers for their constructive comments on an earlier version of this paper. We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison



(PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

## REFERENCES

- BoM. (2001). Climatic Atlas of Australia: Maps of Evaporation. Wang Q.J., McConachy F.L.N., Chiew F.H.S, James R., de Hoedt G.C., Wright W.J. Bureau of Meteorology, Melbourne, Australia.
- Cai, W., Sullivan, A. and Cowan, T. (2009a). Methods of assessing the performance of IPCC-AR4 models in simulating Australian rainfall teleconnections with Indo-Pacific climate drivers. In Anderssen, R.S., R.D. Braddock and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 2377-2383. ISBN: 978-0-9758400-7-8. <http://www.mssanz.org.au/modsim09/G1/cai.pdf>.
- Cai, W., Sullivan, A., Cowan, T. (2009b). Rainfall teleconnection with Indo-Pacific Variability in the WCRP CMIP3 models, *Journal of Climate*, 22, 5046-5071.
- Chiew, F.H.S. and McMahon, T.A. (1991). The applicability of Morton and Penman evapotranspiration estimates in rainfall-runoff modelling. *Water Resources Bulletin*, 27, 611-620.
- CSIRO and BoM. (2007). Climate change in Australia, *Technical report*, 148pp. <http://www.climatechangeinaustralia.com.au/resources.php>
- Drosowsky, W. (2005). The latitude of the subtropical ridge over eastern Australia: The L index revisited, *International Journal of Climatology*, 25, 1291–1299, 10.1002/joc.1196.
- Irving, D.B., Perkins, S.E., Brown, J.R., Gupta, A.S., Moise, A.F., Murphy, B.F., Muir, L.C., Colman, R.A., Power, S.B., Delage, F.P. and Brown, J.N. (2010). Evaluating global climate models for climate change projections in the Pacific island region. *Climate Research*. doi:10.3354/cr01028.
- Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software*, 16, 309–330.
- Jones, D.A. and Beard, G. (1998). Verification of Australian monthly district rainfall totals using high resolution gridded analyses. *Australian Meteorological Magazine*, 47, 41-54.
- 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, *Bulletin of the American Meteorological Society*, 77, 437–471.
- Kent, D.M., Kirono, D.G.C., Timbal B. and Chiew, F.H.S. (2011). Representation of the Australian subtropical ridge in the CMIP3 models. *International Journal of Climatology*. Submitted.
- Kirono, D.G.C. and Kent, D.M. (2011). Assessment of rainfall and potential evaporation from global climate models and its implications for Australian regional drought projection. *International Journal of Climatology*, 31, 1295-1308.
- Morton, F.I. 1983. Operational estimates of areal evaporation and their significance to the science and practice of hydrology. *Journal of Hydrology*, 66, 1-76.
- Pierce, D.W., Barnett, T.P., Santer, B.D. and Glecker, P.J. (2009). Selecting global climate models for regional climate change studies. *PNAS*, 106, 8441-8446.
- Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A. and Taylor, K.E. (2007). Climate Models and Their Evaluation. 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 [Solomon SD, D Qin, M Manning, Z Chen, M Marquis, KB Averyt, Tignor M and HL Miller (eds.)] Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Smith, I.N. and Chandler, E. (2009). Refining rainfall projections for the Murray Darling Basin of south-east Australia – the effect of sampling model results based on performance. *Climatic Change*, doi:10.1007/s10584-009-9757-1.
- Suppiah, R., Hennessy, K.J., Whetton, P.H., McInnes, K.L., Macadam, I., Bathols, J.M., Ricketts, J.H. and Page, C.M. (2007). Australian climate change projections derived from simulations performed for the IPCC 4th Assessment Report. *Australian Meteorological Magazine*, 56, 131-152.
- Vase, J. Teng, J. and Chiew, F.H.S. (2011). Assessment of GCM simulations of annual and seasonal rainfall and daily rainfall distribution across south-east Australia. *Hydrological Processes*, 25, 1486-1497.
- Watterson, IG. (2008). Calculation of probability density function for temperature and precipitation change under global warming. *Journal of Geophysical Research*, 113, D12106, doi:10.1029/2007JD009254.