This article has been accepted for publication in the Journal of Southern Hemisphere Earth Systems Science and undergone full peer review. It has not been through the copyediting, typesetting and pagination, which may lead to differences between this version and the final version.

DOI: 10.22499/3.6701.004
Latitudinal variations in the accuracy of model-generated forecasts of precipitation over Australia and south-east Asia

Richard A. Dare and Elizabeth E. Ebert

Australian Bureau of Meteorology

Revised February 2017

Revised version for: Journal of Southern Hemisphere Earth Systems Science

Corresponding author:

Richard A. Dare

R.Dare@bom.gov.au
Abstract

Forecasts of precipitation produced by global and regional versions of the Bureau of Meteorology’s Australian Community Climate and Earth System Simulator (ACCESS) numerical weather prediction models are compared with Tropical Rainfall Measuring Mission (TRMM) observations for the period January 2011 to March 2014. The area considered covers longitudes 110° to 160°E within 40° latitude of the equator, and includes the Australian continent and part of south-east Asia. Forecast accuracy is assessed using objective measures: equitable threat score (ETS), frequency bias (FB), and the ratio of predicted to observed rain volume (RVR). For the assessment, the TRMM and model datasets are both interpolated to consistent grids defined by spacings of 1° longitude and 1° latitude. Assessments based on 3-month seasons show that, in general, the volume of rain predicted by the models is too high, and rainfall is predicted to occur at more locations than observed. Latitudinal variations in values of the ETS reveal marked declines over the equatorial tropics, with minimum values near the equator from December to May and near 10°N from June to November. Differences in the ETS between 24-hour model and persistence forecasts show that the models produce useful predictions of rainfall away from the tropics, poleward of latitudes 5°-15° S and 15°-30° N, depending on the season and model. The 48-hour model predictions are more useful than the 24-hour forecasts, with improvements relative to persistence over most latitudes, although differences are close to zero at low latitudes. Varying the rain threshold used to compute skill metrics shows that the models produce excessive rainfall at low rain rates, generally less than approximately 15 mm day⁻¹, while not enough precipitation is forecast at higher rain rates. Values of the ETS are generally highest at lower rain rates, coinciding with the excessive values of RVR and FB.
1. Introduction

The Australian Bureau of Meteorology (ABOM) operates a suite of numerical weather prediction (NWP) models, which is named the Australian Community Climate and Earth System Simulator (ACCESS) (ABOM 2010, 2012, 2013). This is based closely on the United Kingdom Meteorological Office’s (UKMO) Unified Model (Cullen 1993; Davies et al. 2005; Rawlins et al. 2007). At the ABOM, the global ACCESS-G and regional ACCESS-R models provide numerical forecasts several times each day (ABOM 2012, 2013). Output from these models supports the ABOM’s operational weather forecasting and provides input to other predictions including stream flow and dispersion of volcanic ash, smoke, and other constituents.

The accuracy of forecasts produced by models operated by the ABOM is assessed routinely by comparing them objectively with observed data (Lee 2007; ABOM 2012, 2013; Wu 2013). Numerical predictions of precipitation produced by these models are generally evaluated over the Australian continent (McBride and Ebert 2000; ABOM 2012, 2013). In addition to routine evaluations of forecasts of precipitation over Australia, assessments of ACCESS model predictions of precipitation over other areas have been conducted by, for example, Brown et al. (2010), Catto et al. (2013), Nguyen et al. (2015), and Dare (2016). Precipitation over Australia and surrounding ocean areas was studied by Brown et al. (2010) using a global version of the ACCESS model. They found that light precipitation (1 - 10 mm day⁻¹), was simulated too frequently and that heavy precipitation was of insufficient intensity. Errors in seasonal average precipitation were largest in the tropics, with too much precipitation in the tropical west Pacific Ocean. In studying fronts and precipitation in a
global version of the ACCESS model, Catto et al. (2013) found that the model produced too many fronts over South-East Asia and over Australia. They also noted that their results were consistent with previous studies in which climate models overestimated the frequency of light precipitation. By comparing model predictions of tropical precipitation around Darwin with radar observations, Nguyen et al. (2015) found that a regional version of the ACCESS model overestimated low rain rates, underestimated high rain rates, and overestimated total rainfall, problems that they attributed to the convection scheme. A recent study by Dare (2016) examined the accuracy of the ACCESS-G model predictions of precipitation over the global tropics, by comparing predictions with Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation estimates (Huffman et al. 2007). Consistent with the findings of the other studies noted here, Dare (2016) found that the amount and spatial extent of tropical rainfall was over-predicted, manifest mainly as excessive light rain.

While the studies mentioned above concern ACCESS models, forecasts of precipitation produced by other models have shown similar deficiencies. Iorio et al. (2004) noted that there is a well-known tendency at coarse resolution for too much weak precipitation and not enough intense precipitation. They found that this problem was partially eliminated in higher-resolution simulations. Sun et al. (2006) analysed performances of eighteen global models over land and found that these models overestimated the frequency of light precipitation and underestimated the intensity of heavier precipitation. Wilcox and Donner (2007) found that the NOAA/GFDL global model that employed a relaxed version of the Arakawa-Schubert convective parameterisation, produced too much light precipitation and too few heavy rain events. Their experiments with alternative formulations to represent convection showed that global models can be made to produce heavy rain events. Stephens et al. (2010) assessed both NWP and climate models, finding that they produced precipitation that was too frequent and
too light, compared with observations. Some improvements were found when using a cloud resolving model, rather than traditional schemes that parameterised convection. Haiden et al. (2012) examined the performance of several global models, all employing convective parameterisation schemes, and found that predictions of rainfall were less accurate in the tropics compared with the extra-tropics. They also noted that the models over-predicted drizzle and had difficulty in predicting heavy precipitation. Online evaluations of NWP models operated by the European Centre for Medium-Range Weather Forecasts (ECMWF), the UKMO, and National Centers for Environmental Prediction show that 24 and 48-hour forecasts of precipitation from all of these models are least accurate in the tropics when compared with several other domains including Asia, Australia-New Zealand, Europe, North America, Northern Hemisphere and Southern Hemisphere (ECMWF 2017).

The evidence presented above shows that, in general, ACCESS and other numerical models produce poor forecasts of precipitation in the tropics, with too much light rainfall. The current study investigates whether this behaviour in the ACCESS model is characteristic only of the tropics or also reflects characteristics of precipitation in sub-tropical and mid-latitude zones. The domain in which the accuracy of forecasts of precipitation is evaluated includes the Australian continent, the equator, and parts of south-east Asia. Understanding how the quality of rainfall forecasts varies latitudinally will help forecasters and other forecast users to make the most appropriate use of the model guidance. It will also guide improvements to future versions of the model. Forecasts of precipitation produced by both the ACCESS-G and ACCESS-R models over a range of rainfall thresholds are evaluated in this study. Following Dare (2016), model-generated forecasts of precipitation are evaluated here by comparison with Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation estimates (Huffman et al. 2007) over latitudes 40°S to 40°N for the period 18 April 2013 to 28 February 2016.
Details of the numerical models, numerical forecast data and observational data used in the evaluation, along with the areas considered, are discussed in Section 2. Skill metrics used to evaluate the precipitation forecasts are discussed in Section 3. Forecasts of precipitation produced by the model are compared with observations in Sections 4 and 5. Variations in skill metrics over a range of thresholds are considered in Section 6. Section 7 summarises the results and presents conclusions.

2. Models, forecast data, and observations

2.1 Models and forecast data

The version of the ACCESS-G model and ACCESS-R model employed here is referred to as the Australian Parallel Suite 1 (APS1), with details provided by ABOM (2010, 2012, 2013). While the ACCESS-G model provides forecasts for the entire domain shown in Figure 1, the ACCESS-R model produces forecasts for the area shown by the dashed line.

Precipitation forecasts produced by the ACCESS models are evaluated from 18 April 2013 to 28 February 2016, corresponding to the period when ACCESS-G, ACCESS-R, and observational data were all available. During this period, the ACCESS-G model operated with a resolution of N320, a horizontal grid spacing of approximately 40 km (ABOM 2012), while the ACCESS-R model operated with a horizontal grid spacing of approximately 12 km (ABOM 2013). Precipitation forecasts for the 0-24 and 24-48 hour forecast period are assessed here, each beginning and ending at 00 UTC on successive days.
2.2 Observations

TRMM data (Huffman et al. 2007), available since 1997, allow assessment of tropical rainfall over wide areas including over remote ocean locations where there are no conventional sources of observations, such as rain gauges. There are several products based on TRMM data. The TRMM dataset used here to verify the ACCESS rainfall forecasts is known as “3B42”, but is referred to here as “TRMM”. Estimates of precipitation in this dataset use TRMM observations as a calibrator for microwave estimates from a number of satellites, which are then combined and corrected using monthly gauge observations where available. Chen et al. (2013a, b) assessed the accuracy of TRMM daily estimates of rainfall produced by tropical cyclones over the Pacific Ocean and over Australia. They found that the TRMM estimates are most accurate over the ocean, with some limitations identified when tropical cyclone rainfall is enhanced by orography. They also noted that although TRMM data depend on indirect observations of rainfall, rain gauges also have limitations, such as interference from wind and evaporation. Over land, Buarque et al. (2011) and Fleming et al. (2011) found that TRMM data are comparable with rain gauge data over monthly to annual periods of time. Due to the poor detection of non-convective rain (Ebert et al. 2007), satellite estimates of precipitation perform best during summer (Gottschalck et al. 2005). Reasons why this is the case are explained in detail by Ebert et al. (2007). Multiple studies have found that TRMM data provide the most accurate satellite-based estimates of daily precipitation (e.g., Liu et al. 2015 and references therein). Therefore, while acknowledging some limitations, the TRMM data are used here to allow assessment of the performance of the ACCESS-G and ACCESS-R models. Daily observations of rainfall available in the TRMM 3B42 product cover 24-hour periods from 00 UTC, in phase with the 0-24 and 24-48 hour periods of precipitation predicted by the model.
2.3 Areas evaluated

As TRMM data are used to evaluate the model forecasts, the analysis is limited to latitudes within 40° of the equator (EQ). The total area considered covers longitudes 110° to 160°E and latitudes 40°S to 40°N. Variations in the ability of the models to forecast precipitation over Australia and northward into the tropics and south-east Asia are assessed by employing eight individual areas covering successive intervals of 10° latitude. The eight areas considered, four on each side of the equator, are shown in Figure 1, each covering 50° longitude by 10° latitude (greater than 5000 x 1000 km²). Model predictions of rainfall are evaluated against TRMM data within each of the eight areas, within which a grid is defined with intervals between grid points of 1.0° longitude and 1.0° latitude. To carry out the verification, both model and observed data are interpolated to this common horizontal grid within each of the eight boxes. Experiments in which grid intervals less than 1.0° were employed showed that the choice of grid interval did not have a large impact on the results of the objective evaluation. The skill metrics described in Section 3 are used to compare observed and predicted rainfall for each day of the 34 month period. The area covered by the regional forecast model (dashed line) shows that the regional model does not cover all eight areas. As a consequence, areas labelled 2N, 3N, and 4N contain forecasts from the global model only, while the other five boxes contain forecasts from both the regional and global models.

3. Skill metrics
The three metrics used to assess the accuracy of each rainfall prediction are equitable threat score (ETS), frequency bias (FB), and rain volume ratio (RVR). ETS and FB are described by Wilks (1995) and Ebert and McBride (1997). These metrics depend on the construction of a contingency table, the cells of which contain aggregates of the number of locations where predicted and observed rain amounts are above or below a selected threshold of rainfall. Consequently, the table contains four cells, as shown by Table 1. The formulae used to compute ETS and FB are shown here in Equations 1 and 2, respectively. Although a range of thresholds is considered later, the initial assessment of the rainfall is based on a threshold of 1 mm day\(^{-1}\), following Ebert et al. (2007), which means that points with values of rainfall below this level are considered to be free of rain, in the respective predicted and observed fields. These two metrics do not assess rain intensities within areas of rain, but essentially assess the occurrence of rain in the predicted and observed fields.

\[
ETS = \frac{ZH - FM}{(F + M)N + (ZH - FM)} \tag{1}
\]

\[
FB = \frac{H + F}{H + M} \tag{2}
\]

In Equations 1 and 2, H is the number of Hits or correct forecasts of occurrence of rain, F is the number of False Alarms, M is the number of Missed forecasts, and Z is the number of correct predictions of no rainfall, or rainfall less than the threshold. N is the total number of forecasts made.

Values of the ETS range from a minimum of \(-1/3\) to a maximum of 1 with a value of 1 indicating that the rain was forecast at the correct location with no false alarms. The FB
metric assesses the over- and under-prediction of rainfall occurrence, where the ideal value is unity.

To provide some quality information on the amount of rain produced, the rain volume (RV) is computed for each forecast period within each area, for both predicted and observed fields. The ratio of model predicted RV to observed RV provides the RVR that indicates whether the model produced too much rainfall (RVR > 1) or too little (RVR < 1).

When computing seasonal values of the ETS and FB, the values of H, F, M, and Z for each day in each box are accumulated in the contingency table over the entire season. These accumulated values are used to compute ETS and FB, rather than computing values of these metrics for each day and subsequently computing mean daily values for the season. Similarly, for the RVR, the observed RV and forecast RV are each accumulated over a season, from which the RVR is computed, rather than computing daily values of RVR from which a mean is computed.

Finally, a reference forecast of persistence (i.e., yesterday's observations are the forecast for today) is also verified to help put the model verification results in context. Forecasts that cannot outperform persistence are of limited value.

4. Maps of observed and predicted precipitation
Before discussing results based on computed skill metrics, it is informative to consider maps of precipitation over the domain of interest. Two examples are presented here. The first is the annual mean daily observed and predicted precipitation (Figure 2). Comparison of Figures 2a and 2b shows that the model has some skill in predicting the general locations of precipitation throughout the entire area. For example, note the light precipitation over western and central Australia, the much heavier precipitation within 10º of the equator with local minima to the northeast of Borneo, and the swath of relatively high values extending from Taiwan to southern Japan and further to the east. However, despite these general qualitative agreements between observed and predicted patterns of precipitation, there are many locations where the predicted values are higher than those observed. For example, within 10º of the equator, the area covered by values of model-predicted precipitation above 8 mm day\(^{-1}\) is larger than that covered by the corresponding observations.

The second example is the observed and predicted precipitation for the 24-hour period beginning at 00 UTC 12 January 2014 (Figure 3). This day was selected because the skill metrics within box 1S, covering the area from the equator to 10ºS, are similar to the mean December-January-February (DJF) values of the metrics for this box presented later. A comparison of Figures 3a and 3b shows that there is general agreement between observed and predicted fields regarding the locations of the highest amounts of precipitation, above 20 mm day\(^{-1}\). However, the model has performed less well in predicting the highest amounts of precipitation, above 100 mm day\(^{-1}\). The areas of light precipitation (<10 mm day\(^{-1}\)) predicted by the model are excessive when compared with observations. This feature is possibly the most outstanding difference between the observed and predicted fields. In box 1S, although the model has failed to predict the heaviest precipitation, the RV is actually
greater than that observed due to the model prediction of excessive widespread light precipitation.

5. Accuracy of daily precipitation forecasts during each season

5.1 Rain volume ratio

Values of parameters H, F, M, Z, and RV from daily forecasts are accumulated over four 3-month seasons for each of the eight boxes to provide an indication of the ability of each model to accurately predict daily precipitation. Seasonal values of RVR based on the 0-24 hour forecast period, for DJF, March-April-May (MAM), June-July-August (JJA), and September-October-November (SON), show that ACCESS-G generally produces too much rainfall (RVR > 1) over most latitudes (Figure 4a). Values are similar but a little lower during the 24-48 hour forecast period (Figure 4b). The highest values for each season are mostly located in the Northern Hemisphere. For the longitudes of interest here, the RVR maxima present in the Northern Hemisphere correspond to areas where relatively high amounts of precipitation are observed due to the Asian monsoon (Lin et al. 2010) and tropical cyclones (Jiang et al. 2011). In comparison, the Australian region experiences a less active TC season (Jiang et al. 2011), a monsoon that covers a smaller area (Lin et al. 2010), and relatively light precipitation over the Australian continent, as shown by Figure 2a. Seasonal and latitudinal variations are associated with the seasonal shifts in the location of the intertropical convergence zone, the monsoon (Lin et al. 2010) and tropical cyclones (Jiang et al. 2011). As the precipitation associated with both the monsoon and tropical cyclones is largely of a convective nature, the RVR maxima in the Northern Hemisphere may be due to misrepresentation of convective precipitation, an issue that has been identified previously (Wilcox and Donner 2007, Stephens et al. 2010, Nguyen et al. 2015).
Seasonal RVRs for ACCESS-R (Figure 4c and d) are generally lower than those produced by ACCESS-G, particularly during DJF, JJA and SON where values are less than unity over some latitudes. As ACCESS-G and ACCESS-R use the same methods to forecast precipitation, these differences in RVRs are likely due to the different grid resolutions employed by the two models. Forecasts based on persistence (Figure 4e and f) are clearly superior to those of both models, with RVRs consistently closer to unity across all latitudes during all four seasons, as would be expected. However, for the 24-48 hour forecast period, ACCESS-R compares well with persistence during DJF, MAM, and SON over latitudes south of 20ºS and during JJA near the equator. Across the multi-month time scales considered here, it would be expected that persistence-based forecasts of RV would be superior to predictions produced by numerical models because while models produce new forecasts for each day, persistence-based forecasts use past daily observations. When accumulated over a season, the persistence-based data are approximately equal to the observations on which they are based, only offset by 24 or 48 hours. Therefore, the persistence-based RVR computed for a season would be expected to be equal to, or close to, unity. Here, there are variations from unity by approximately 5% (Figure 4f). These variations are due to the imbalance between the persistence-based forecast rain volume on the first two days of the period and the observed rain volume during the final two days of the period.

5.2 Frequency bias

Although there are variations between seasons and across latitudes, seasonal values of FB produced by ACCESS-G are generally in the range from 1.5 to 2.1, showing that this model
is predicting precipitation at too many locations, relative to observations (Figures 5a and b). Values of FB based on ACCESS-R forecasts (1.4 to 1.8, Figures 5c and d) are slightly less than those produced by the global model, but remain above unity. Over Australian latitudes, particularly south of 20ºS, both the global and regional models exhibit an increase in FB from DJF to MAM to JJA, followed by a decrease into SON. This may be associated with the increase in the occurrence of frontal systems producing precipitation over southern Australia during JJA relative to other seasons, and the ability of the ACCESS model to simulate fronts and their associated precipitation as discussed by Catto et al. (2013). The precipitation from these systems may be represented within the NWP model by the microphysics scheme on the grid-scale, or by activation of the convection scheme. Determining which component of the model may be responsible for the enhanced FB during winter is a matter for future investigation. An additional factor to consider is that TRMM data may underestimate precipitation during winter (Ebert et al. 2007). In terms of FB, persistence forecasts (Figures 5e and f) are superior to those of both ACCESS-G and ACCESS-R, with values close to unity across all latitudes during each season for both the 0-24 and 24-48 hour forecast periods.

5.3 Equitable threat score

Latitudinal variations in values of the ETS based on ACCESS-G forecasts (Figure 6a and b) reveal marked declines from approximately 10ºS to 20ºN during DJF and MAM, and from approximately 5ºS to 30ºN during JJA and SON. In general, this diagram shows that the global model has remarkably less skill in forecasting daily precipitation over low latitudes, compared with the relatively high values of ETS recorded at latitudes further from the equator. The regional model ACCESS-R also exhibits a reduction in ETS for areas closer to the equator, with values of ETS for each season and latitude very similar to those of the
global model (Figure 6c and d). The relatively poor performance of the numerical models in predicting precipitation over low latitudes is likely to be associated with the parameterisation of convection (Stephens et al. 2010, Nguyen et al. 2015).

At latitudes away from the equator, values of the ETS based on persistence forecasts (Figure 6e) are clearly lower than values based on model forecasts. Over the equatorial tropics, ETSs based on persistence forecasts are approximately 0.2, while model-based ETSs are below 0.2 at this location. ETSs based on the 24-48 hour persistence forecasts (Figure 6f) are clearly lower than during the 0-24 hour period (Figure 6e). In comparison, ETSs based on the ACCESS model forecasts (Figures 6a-d) show only small changes between the two forecast periods. Consequently, the 48 hour persistence forecasts provide ETSs across nearly all latitudes (Figure 6f) that are less than those produced by the corresponding ACCESS forecasts (Figures 6b and d). Differences between ETSs based on model and persistence forecasts are presented in Figure 7. ACCESS-G 0-24 hour forecasts produce higher values of the ETS than those produced by persistence forecasting poleward of approximately 15°S and 15°N during DJF and MAM (Figure 7a). During JJA and SON, the global model produces superior values of ETS over latitudes south of approximately 10°S and north of approximately 25°N. For the 24-48 hour period, ACCESS-G produces forecasts that are superior to those based on persistence except from approximately 10°S to 20°N where differences are close to zero (Figure 7b). Differences in ETS between ACCESS-R and persistence forecasts (Figures 7c and d) are very similar to the differences based on the global model. Both exhibit an approximately linear decline in ETS difference from 35°S towards the equator. The relatively good performance of the ACCESS model forecasts during the 24-48 hour period is due largely to the decrease in the accuracy of persistence forecasts from the 0-24 to 24-48 hour periods. These results showing that predictions of precipitation are less
accurate in the tropics compared with the extra-tropics are consistent with past findings of Haiden et al. (2012) and with the ECMWF (2017) online evaluations.

6. Skill metric variations with precipitation threshold

Up to this point, skill metrics have been based on values of precipitation over a 24-hour period above a threshold of 1 mm day\(^{-1}\). In this section, skill metrics are recomputed for a range of precipitation thresholds. Figures 8 and 9 contain a range of results for the eight boxes, the two models, and three metrics (RVR, FB, and ETS). Results presented here are limited to the 0-24 hour forecast period because they do not differ greatly from those found for the 24-48 hour forecast period. Although the results vary between locations and seasons, the model forecasts produce RVRs closest to unity for rain thresholds below approximately 15 mm day\(^{-1}\) (Figures 8a, 8b, 9a, 9b). Above this precipitation threshold, RVRs are generally less than unity. In general, the ACCESS models examined here produce too much precipitation at low rain rates but not enough at high rain rates.

Values of FB show that the models predict precipitation that is too widespread for thresholds less than approximately 15 mm day\(^{-1}\) (Figures 8c, 8d, 9c, 9d). The FB is closest to unity for rain thresholds ranging from approximately 5 to 20 mm day\(^{-1}\), with differences between boxes being larger than differences between the two models. These distributions of FB also show that the precipitation amounts greater than approximately 15 mm day\(^{-1}\) are under-forecast. The results discussed here concerning FB are in general agreement with one of the findings by Nguyen et al. (2015), who compared radar observations with a higher resolution version of the ACCESS model over Darwin, and noted that the model underestimated
intensity and overestimated frequency of precipitation. Many other models have also been found to overestimate the frequency of light precipitation (Dai 2006, Sun et al. 2006), an issue associated with the over-activation of the convective parameterisation (Dai 2006, Nguyen et al. 2015).

In the four southern boxes, the highest ETSs are produced for precipitation thresholds less than approximately 5 mm day\(^{-1}\) (Figures 8e and 8f). In the north, highest ETSs occur for precipitation thresholds close to 10 mm day\(^{-1}\) (Figure 9e and 9f). The reductions in ETS evident in the northern boxes as precipitation thresholds approach zero are much less evident at low precipitation thresholds in the southern boxes. This might be due to differences in weather regimes between the Northern and Southern Hemispheres over the longitudes of interest, such as differences in precipitation associated with monsoon activity (Lin et al. 2010) and tropical cyclones (Jiang et al. 2011). Another influence might be differences in the TRMM estimation of rainfall over oceanic versus continental surfaces, with better estimates over oceanic surfaces (Chen et al. 2013a).

There is a clear separation in values of ETS corresponding to boxes located next to the equator (boxes 1S and 1N) compared with boxes further poleward (Figures 8e, 8f, 9e). This separation occurs across all rates of precipitation. This feature demonstrates again the relatively poor performance of the model precipitation forecasts near the equator. Also notice in the RVR and FB results (Figures 8a-d and 9a, c) that values in boxes 1S and 1N are among the highest of all boxes at lowest rain rates and among the lowest of all boxes at highest rain rates.
7. Summary and conclusions

Metrics based on seasonal accumulations of parameters used to assess daily precipitation forecasts produced by the ACCESS-G and ACCESS-R models show that, in general, the volume of rain predicted by the models is too high, and rainfall is predicted to occur at more locations than observed. These results are consistent with the findings of Stephens et al. (2010) and Nguyen et al. (2015).

Assessing skill metrics within eight separate areas, from 40ºS to 40ºN, provides a clear picture of the latitudinal variations in the ability of the ACCESS models to accurately predict precipitation. The ability of the models to accurately forecast daily precipitation is lowest over low latitudes. ACCESS-G forecasts produce latitudinal variations in values of the ETS with marked declines between approximately 10ºS and 20ºN during DJF and MAM, and from approximately 5ºS to 30ºN during JJA and SON. The ACCESS-R model also produces relatively low values of the ETS near the equator, with values of ETS for each season and latitude very similar to those of the ACCESS-G model. ACCESS model forecasts of precipitation are superior to persistence away from the tropics and during the 24-48 hour period.

Varying the precipitation threshold used to compute skill metrics shows that the ACCESS models produce too much precipitation at rates generally less than approximately 15 mm day$^{-1}$, while not enough precipitation is forecast at higher rain rates. This result is in general agreement with the findings of previous studies, such as Brown et al. (2010) and
Nguyen et al. (2015). While this general pattern occurs within boxes at different latitudes, there is a clear separation in values of the ETS across all precipitation thresholds, with lower values of the ETS corresponding to boxes adjacent to the equator (boxes 1S and 1N) compared with boxes further poleward.

Although evidence presented here identifies a deficiency in the ability of the ACCESS models to accurately predict tropical precipitation, it also provides an opportunity to investigate components of the ACCESS models that are responsible for poor forecasts of precipitation. Modifications made to physics components of the ACCESS models with the aim of improving the prediction of precipitation in the equatorial tropics also have the potential to impact positively on the simulation of heat and moisture profiles and wind fields in the tropical atmosphere, which in turn may positively affect vital forecasting challenges including, for example, the prediction of the dispersion of volcanic ash in the tropical atmosphere and the development and movement of tropical cyclones. Additional focus on verifying forecasts of tropical rainfall from the model during the model development phases would be helpful.

The operational version of the ACCESS-G model was recently upgraded from APS1, the subject of the current work, to APS2. Details of this upgrade are provided by ABOM (2016). Improvements in forecasts of precipitation over Australia include small reductions in the bias and small increases in the ETS (ABOM 2016). Further work is required to examine whether these improvements are also found over tropical areas to the north of the Australian continent.
Acknowledgements

The authors acknowledge the contributions made by members of the Earth System Modelling team and the Bureau National Operations Centre who support the development and implementation of the ACCESS models. We thank Lawrie Rikus for helpful discussions on this work, and Charmaine Franklin and David Smith for providing constructive comments on the manuscript. We thank Ying Jun Chen for assistance in obtaining the TRMM 3B42 data. We also thank the two anonymous reviewers and associate editor Steve Siems for their constructive comments.

References


<table>
<thead>
<tr>
<th></th>
<th>Observed: Yes</th>
<th>Observed: No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted: Yes</td>
<td>H</td>
<td>F</td>
</tr>
<tr>
<td>Predicted: No</td>
<td>M</td>
<td>Z</td>
</tr>
</tbody>
</table>

**Table 1:** Contingency table to assess accuracy of rainfall forecasts, showing the number of Hits (H), False Alarms (F), Misses (M), and correct predictions of rainfall less than the threshold (Z).
Figure captions

Figure 1: Positions of the eight 50° longitude by 10° areas, numbered 1-4 in each hemisphere, and labelled north (N) and south (S) of the equator. The ACCESS-R domain is outlined by the dashed line.

Figure 2: Annual daily mean precipitation (mm day$^{-1}$) data from (a) TRMM observations, and (b) ACCESS-G predictions, for the two-year period 20130418 to 20150417.

Figure 3: (a) Daily precipitation (mm day$^{-1}$) data for the 24-hour period beginning at 00 UTC 12 January 2014, from (a) TRMM observations, and (b) ACCESS-G predictions.

Figure 4: The distribution of RVR across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence.

Figure 5: The distribution of FB across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence.

Figure 6: The distribution of ETS across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence. For reference, the ETS=0.2 line is plotted, corresponding approximately to the 24-hour persistence skill in the tropics.

Figure 7: The distribution of ETS differences between model and persistence forecasts across the latitudes covered by the eight boxes, for each of the four 3-month periods, for 0-24 and 24-48 hour forecasts, (a, b) ACCESS-G, (c, d) ACCESS-R.
Figure 8: Values of skill metrics RVR, FB, ETS over a range of precipitation thresholds, for four areas (1S to 4S) based on ACCESS-G (left column) and ACCESS-R (right column) 24-hour forecasts.

Figure 9: Values of skill metrics RVR, FB, ETS over a range of precipitation thresholds, for four areas (1N to 4N) based on ACCESS-G (left column) and ACCESS-R (right column) 24-hour forecasts. Note that areas 2N to 4N do not fall within the ACCESS-R domain.
Figure 1: Positions of the eight 50° longitude by 10° areas, numbered 1-4 in each hemisphere, and labelled north (N) and south (S) of the equator. The ACCESS-R domain is outlined by the dashed line.
Figure 2: Annual daily mean precipitation (mm day$^{-1}$) data from (a) TRMM observations, and (b) ACCESS-G predictions, for the two-year period 20130418 to 20150417.
Figure 3: (a) Daily precipitation (mm day$^{-1}$) data for the 24-hour period beginning at 00 UTC 12 January 2014, from (a) TRMM observations, and (b) ACCESS-G predictions.
Figure 4: The distribution of RVR across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence.
Figure 5: The distribution of FB across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence.
Figure 6: The distribution of ETS across the latitudes covered by the eight boxes, for each of the four 3-month periods, based on 0-24 and 24-48 hour forecasts produced by (a, b) ACCESS-G, (c, d) ACCESS-R, (e, f) persistence. For reference, the ETS=0.2 line is plotted, corresponding approximately to the 24-hour persistence skill in the tropics.
Figure 7: The distribution of ETS differences between model and persistence forecasts across the latitudes covered by the eight boxes, for each of the four 3-month periods, for 0-24 and 24-48 hour forecasts, (a, b) ACCESS-G, (c, d) ACCESS-R.
Figure 8: Values of skill metrics RVR, FB, ETS over a range of precipitation thresholds, for four areas (1S to 4S) based on ACCESS-G (left column) and ACCESS-R (right column) 24-hour forecasts.
Figure 9: Values of skill metrics RVR, FB, ETS over a range of precipitation thresholds, for four areas (1N to 4N) based on ACCESS-G (left column) and ACCESS-R (right column) 24-hour forecasts. Note that areas 2N to 4N do not fall within the ACCESS-R domain.