CSIRO PUBLISHING

Journal of Southern Hemisphere Earth Systems Science, 2020, **70**, 393
https://doi.org/10.1071/ES17009_CO

Corrigendum

ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system

Debra Hudson, Oscar Alves, Harry H. Hendon, Eun-Pa Lim, Guoqiang Liu, Jing-Jia Luo, Craig MacLachlan, Andrew G. Marshall, Li Shi, Guomin Wang, Robin Wedd, Griffith Young, Mei Zhao and Xiaobing Zhou

Journal of Southern Hemisphere Earth Systems Science, 2017, **67**(3), 132–159.
doi:10.1071/ES17009

The authors wish to advise that the original Figure 2 was incorrect. The nature of the error was that *instantaneous* model output was compared to *averaged* reanalysis data for daily 500 hPa geopotential height anomalies over the Southern Hemisphere extratropics for the calculation of the normalised root-mean-square error (RMSE) curves. The corrected figure is provided below, which uses instantaneous model and reanalysis data. Forecasts from ACCESS-S1 and GloSea5 were initialised on 1 May for the period 1993–2012 and are based on five-member ensembles. The verification data are from the ERA-Interim reanalysis (Dee *et al.* 2011).

The error does not affect the conclusions drawn from Figure 2 in the text of the paper, namely that the difference between the error and spread curves is much smaller in ACCESS-S1 (red curves) than in GloSea5 (black curves). GloSea5 is significantly under-dispersed out to the limit of predictability, whereas the ACCESS-S1 ensemble is more reliable.

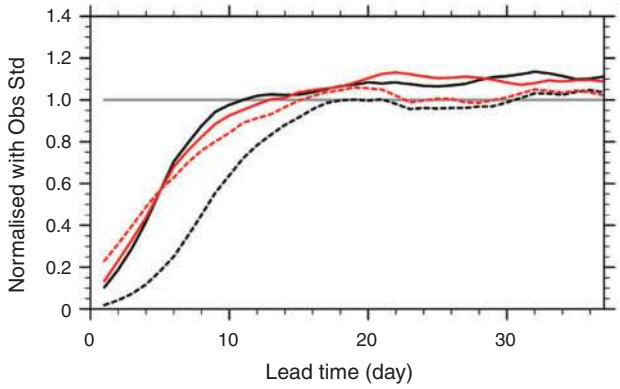


Fig. 2. Normalised RMSE (solid) and ensemble spread (dashed) for daily 500 hPa geopotential height anomalies over the Southern Hemisphere extratropics (20°S–60°S) from the Met Office GloSea5-GC2 system (black) and from ACCESS-S1 (red). Forecasts were initialised on 1 May for the period 1993–2012 and are based on five-member ensembles. The verification data are from the ERA-Interim reanalysis (Dee *et al.* 2011).

Reference

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteorol. Soc.* **137**, 553–597. doi:10.1002/QJ.828

ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system

Debra Hudson,[1] Oscar Alves,[1] Harry H. Hendon,[1] Eun-Pa Lim,[1] Guoqiang Liu,[1] Jing-Jia Luo,[1] Craig MacLachlan,[2] Andrew G. Marshall,[1] Li Shi,[1] Guomin Wang,[1] Robin Wedd,[1] Griffith Young,[1] Mei Zhao,[1]and Xiaobing Zhou[1]

[1] Bureau of Meteorology, Melbourne, Australia
[2]Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received July 2017; accepted December 2017)

ACCESS-S1 will be the next version of the Australian Bureau of Meteorology's seasonal prediction system, due to become operational in early 2018. The multi-week and seasonal performance of ACCESS-S1 has been evaluated based on a 23-year hindcast set and compared to the current operational system, POAMA. The system has considerable enhancements compared to POAMA, including higher vertical and horizontal resolution of the component models and state-of-the-art physics parameterisation schemes. ACCESS-S1 is based on the UK Met Office GloSea5-GC2 seasonal prediction system, but has enhancements to the ensemble generation strategy to make it appropriate for multi-week forecasting, and a larger ensemble size.

ACCESS-S1 has markedly reduced biases in the mean state of the climate, both globally and over Australia, compared to POAMA. ACCESS-S1 also better predicts the early stages of the development of the El Niño Southern Oscillation (through the predictability barrier) and the Indian Ocean Dipole, as well as multi-week variations of the Southern Annular Mode and the Madden-Julian Oscillation — all important drivers of Australian climate variability. There is an overall improvement in the skill of the forecasts of rainfall, maximum temperature (Tmax) and minimum temperature (Tmin) over Australia on multi-week timescales compared to POAMA. On seasonal timescales the differences between the two systems are generally less marked. ACCESS-S1 has improved seasonal forecasts over Australia for the austral spring season compared to POAMA, with particularly good forecast reliability for rainfall and Tmax. However, forecasts of seasonal mean Tmax are noticeably less skilful over eastern Australia for forecasts of late autumn and winter compared to POAMA.

The study has identified scope for improvement of ACCESS-S in the future, particularly 1) reducing rainfall errors in the Indian Ocean and Maritime Continent regions, and 2) initialising the land surface with realistic soil moisture rather than climatology. The latter impacts negatively on the skill of the temperature forecasts over eastern Australia and is being addressed in the next version of the system, ACCESS-S2.

Corresponding author: Debra Hudson, Bureau of Meteorology, GPO Box 1289, Melbourne, VIC, Australia 3001
Email: debra.hudson@bom.gov.au

1 Introduction

The Bureau of Meteorology (BoM) has routinely made coupled model seasonal forecasts with POAMA (Predictive Ocean Atmosphere Model for Australia) since 2002. From May 2013 the official BoM seasonal outlook for Australian climate has been based on the POAMA-2.4 system (Hudson et al. 2013). Substantial improvements in forecast accuracy, reliability and products from the POAMA system have been made since its initial inception in 2002. These improvements stem from upgrades to the ocean assimilation (Yin et al. 2011), atmosphere-land initialisation (Hudson et al. 2011) and to the ensemble approaches to sample both observational (Hudson et al. 2013) and model uncertainty (Wang et al. 2011; Langford and Hendon 2013). However, no change has been made since 2002 to the basic coupled atmosphere-ocean-land surface model in the forecast system. As a result, POAMA has become outdated both in terms of parameterised physics that were based on model developments in the 1990's, and horizontal and vertical resolution that was limited by available computing resources in the early 2000's.

To address these shortcomings, BoM is making a major upgrade to the seasonal forecasting capability by implementing a new coupled-model seasonal prediction system, called ACCESS-S1 (the Australian Community Climate and Earth-System Simulator-Seasonal prediction system version 1), which is planned to replace POAMA in early 2018. The ACCESS-S1 system is based on the UK Met Office's global coupled model seasonal forecast system GloSea5-GC2 (Global Seasonal forecast system version 5 using the Global Coupled model configuration 2; MacLachlan et al. 2015), but includes some BoM enhancements to the ensemble generation strategy. The GloSea5-GC2 system is made available to BoM through the Unified Model Partnership, of which BoM is a core partner.

The GC2 model (Williams et al. 2015) in the GloSea5-GC2 system includes a number of state-of-the-art features compared to POAMA, especially substantially higher horizontal and vertical resolution in both atmosphere and ocean model components, improved atmospheric physics parameterization (e.g. Brown et al. 2012), a multi-level land surface model (Best et al. 2011), and an interactive multi-level sea-ice model (Hunke and Lipscomb 2010; Rae et al. 2015). GC2 has ~60 km horizontal resolution (N216) in the mid-latitudes, which resolves details of orography and topography such as the Dividing Range in eastern Australia and the island state of Tasmania. It has 85 vertical levels with a fully resolved stratosphere. In comparison, POAMA has ~250 km horizontal resolution and seventeen vertical levels in the atmosphere, with only five levels in the stratosphere (above the 200 hPa level).

The ocean model in GC2 (Madec 2008; Megann et al. 2014) is eddy-permitting, with a 0.25° horizontal resolution globally and 75 vertical levels that provides 1 metre resolution in the top 10 metres and 2 metre resolution between 10-20 metres. In comparison, the POAMA ocean model has a 0.5° latitude by 2° longitude resolution in the tropics reducing to 1.5° by 2° at the poles and with only 25 vertical levels (with ~15 metre resolution near the surface). The increased oceanic resolution in GC2 means that tropical instability waves are fully resolved, thus leading to improved depiction of the mean state and variability in the tropical eastern Pacific (e.g., Graham 2014). Substantial improvements to the depiction of mixed layer processes, coastal currents and upwelling zones, and other sub-surface processes associated with the El Niño Southern Oscillation (ENSO) are also achieved. GC2 includes climate forcing of greenhouse gases with realistic time-variation, whereas greenhouse gas forcing is held constant in POAMA.

The final version of the ACCESS-S1 system has been configured, tested and implemented, with hindcasts produced for the period 1990-2012. In this study we describe the multi-week (sub-seasonal/intra-seasonal) and seasonal forecast skill of ACCESS-S1 based on the hindcast set and make comparisons to POAMA, the system that it will replace. The real-time design of the ACCESS-S1 system is also described.

2 ACCESS-S1 forecast system

The key components of the ACCESS-S1 forecast system are summarised in Table 1. The coupled model version and data assimilation for ACCESS-S1 are the same as the Met Office GloSea5-GC2 system (MacLachlan et al. 2015). The key differences between ACCESS-S1 and GloSea5-GC2 are the ensemble size, method of ensemble generation, the hindcast size, and the real-time configuration, which are detailed below.

Table 1 The ACCESS-S1 forecast system

Atmospheric model	Global Atmosphere 6.0: The Unified Model (UM; Walters et al. 2017, Williams et al. 2015). Horizontal resolution: N216 (~60 km in the mid-latitudes) Vertical resolution: 85 levels
Land surface model	Global Land 6.0: Joint UK Land Environment Simulator (JULES; Best et al. 2011; Walters et al. 2017) with 4 soil levels
Ocean model	Global Ocean 5.0: Nucleus for European Modelling of the Ocean (NEMO ORCA25 ; Madec 2008, Megann et al. 2014). Horizontal resolution: 0.25° Vertical resolution: 75 levels. Level thicknesses range from 1 m near the surface to about 200 m near the bottom (6000 m depth).
Sea ice model	Global Sea Ice 6.0: Los Alamos sea ice model (CICE; Hunke and Lipscomb 2010; Rae et al. 2015).
Coupler	Ocean Atmosphere Sea Ice Soil coupler (OASIS3, Valcke, 2013)
Atmosphere initial conditions	ERA-Interim (Dee et al. 2011) for the hindcasts Operational Global NWP analysis for the real-time forecasts
Ocean and ice initial conditions	Met Office Forecast Ocean Assimilation Model (FOAM) which uses the NEMO 3-dimensional variational ocean data assimilation (NEMOVAR; Mogensen et al. 2009, 2012; Waters et al. 2015)
Land surface initial conditions	Climatological soil moisture (MacLachlan et al. 2015)
Ensemble generation	Initial condition uncertainty: BoM perturbation scheme combined with lagged initial conditions Model uncertainty/unresolved processes: Stochastic Kinetic Energy Backscatter version 2 (SKEB2; Bowler et al. 2009).

2.1 The coupled model and data assimilation system

The global coupled model GC2 is comprised of the UM Global Atmosphere 6.0 (GA6.0), Global Ocean 5.0 (GO5.0), Global Land 6.0 (GL6.0) and Global Sea Ice 6.0 (GSI6.0) (Williams et al. 2015). The UM GA6.0 atmospheric model is run at a N216 horizontal resolution, which equates to about 60 km in the mid-latitudes, and there are 85 levels in the vertical with a well-resolved stratosphere (35 levels are above 18 km; Walters et al. 2017). The land surface model (GL6.0) is

the JULES (Joint UK Land Environment Simulator) model (Best et al. 2011; Walters et al. 2017) and is tightly coupled to the atmosphere model on the same horizontal grid. It has four soil levels, with a sophisticated representation of soil and surface hydrology, and fluxes of heat and moisture within the soil and to the atmosphere. Land cover heterogeneity is represented. The ocean model (GO5.0) is based on the NEMO3.4 (Nucleus for European Modelling of the Ocean) ORCA25 version model and has a 0.25° horizontal resolution using a tripolar grid with 75 levels in the vertical (with a 1 m resolution in top 10 m) (Madec 2008, Megann et al. 2014). The sea-ice model (GSI6.0) is the CICE4.1 model (Rae et al. 2015) and has five thickness categories. It is tightly coupled to the ocean model on the same horizontal grid. The atmosphere/land and the ocean/sea ice are coupled every three hours using the Ocean Atmosphere Sea Ice Soil coupler (OASIS3, Valcke 2013).

Climate forcings for greenhouse gases (e.g. CO2 and methane) are set to observed values up to the year 2005 and after this the emissions follow the Intergovernmental Panel on Climate Change (IPCC) RCP4.5 scenario (MacLachlan et al. 2015). For ozone, the observational seasonally-varying climatology is used.

The atmospheric initial conditions for the real-time predictions are taken from the global numerical weather prediction (NWP) 4D-Var data assimilation analyses using the ACCESS-G model at BoM, which is very similar to the assimilation run at the Met Office (e.g., Rawlins et al. 2007). The hindcasts are initialised from ERA-Interim reanalyses (Dee et al. 2011), which are interpolated onto the N216 grid, following the approach of GloSea5 (MacLachlan et al. 2015). The ocean and sea ice are initialised using the FOAM (Forecast Ocean Assimilation Model) analyses (Blockley et al. 2014) in both the real-time forecasts and the hindcasts, as per GloSea5. FOAM uses the same ocean-sea ice models as GC2 and uses the NEMO 3-dimensional variational ocean data assimilation (NEMOVAR; Mogensen et al. 2009, 2012; Waters et al. 2015) to produce an analysis of currents, surface height, temperatures, salinity and sea ice based on satellite and in situ observations of sea surface temperature, sea level anomaly satellite data, sub-surface temperature and salinity profiles, and satellite observations of sea-ice concentration. Soil moisture for both the hindcasts and the real-time forecasts are initialised with monthly climatologies to ensure they are initialised consistently. An inconsistency between the forecast and hindcast soil moisture initialisation can result in spurious near surface temperature biases in the calibrated forecast. The lack of a consistent land surface reanalysis covering the historical period up to real-time has forced the use of climatology. The soil temperature and snow cover are initialised with time-varying initial conditions from ERA-Interim (Dee et al. 2011) for the hindcasts and from the NWP analysis for the real-time forecasts.

2.2 Ensemble generation

2.2.1 Model uncertainty

Model uncertainty in ACCESS-S1 (and GloSea5-GC2) is simulated using the Stochastic Kinetic Energy Backscatter (SKEB2) scheme (Bowler et al. 2009) in the atmospheric model. SKEB2 is designed to represent unresolved processes and to provide grid-scale perturbations during the model integration (MacLachlan et al. 2015). Each ensemble member evolves differently due to these grid-scale perturbations.

2.2.2 Initial condition uncertainty

The representation of initial condition uncertainty for ACCESS-S1 differs from that of the Met Office GloSea5-GC2 system. The GloSea5-GC2 system incorporates initial condition uncertainty using a lagged ensemble (combining forecasts from earlier start times). There are no perturbed initial conditions. Rather, for a single start time, differences between ensemble members develop solely as a result of the stochastic physics (SKEB2). This approach is appropriate for seasonal prediction, which is the focus of the GloSea5-GC2 system, but is not optimal for multi-week forecasting as discussed below and shown in our previous research with POAMA (Hudson et al. 2013). Furthermore, the design of the hindcast set as produced at the Met Office, with four forecast starts per month that are roughly eight days apart (see Section 2.3), means that a viable lagged ensemble cannot be generated or assessed for multi-week predictions using the hindcasts, because the time between forecast starts is too great which will negatively impact skill. Ultimately, for the next version of the system (ACCESS-S2) perturbed ensembles will be produced using the BoM's assimilation scheme (e.g., Hudson et al. 2013; Yin et al. 2011). However, in the meantime, a viable yet practical approach has been developed to produce perturbed ensemble forecasts from ACCESS-S1 that are suitable for multi-week prediction in real-time and in the hindcasts.

Based on previous experience with the POAMA system and as reported in Magnusson et al. (2009), we perturb the initial condition by adding appropriately scaled, randomly chosen seven-day differences of reanalysed atmospheric states from

the period 1981-2010 from the ERA-interim reanalysis (Dee et al. 2011). This is a revision of the random field perturbation described by Magnusson et al. (2009). Random field perturbations are derived by scaling the difference between analyses on two different days that are randomly chosen. Our refinement of this approach is to calculate the difference between days seven days apart. The idea in using seven-day differences is that typical patterns of error growth that are relevant to multi-week timescales will be selected. We made small trials of using one-day and three-day differences and settled on using seven-day differences because the patterns of difference were of larger scale and so will project more heavily onto the error structures at longer lead times (Magnusson et al. 2009), which may be relevant to the multi-week timescales that we are targeting. For this version of ACCESS-S1, no perturbations are added to the land surface, ocean or sea-ice initial states, assuming that most of the error growth in the first two-three weeks of the forecast will be in the atmosphere.

Perturbations for a forecast start time in a given month are created by randomly sampling seven-day differences from reanalyses in the same month from the period 1981-2010. The differences for the three dimensional fields of u-wind, v-wind, temperature (T) and specific humidity (q), as well as for surface pressure, are multiplied by a scaling factor so that the perturbations have magnitude equal to analysis uncertainty. The scaled perturbations are then added and subtracted from the unperturbed initial atmospheric state each day so that pairs of perturbed members are created.

The scaling factor is based on the magnitude of the surface pressure perturbations. Let the randomly chosen seven-day difference in a surface pressure analysis be:

$$\gamma(x,y,t) = Surfp(x,y,t+7) - Surfp(x,y,t), \tag{1}$$

where Surfp is the surface pressure analysis from the ERA-interim reanalysis (Dee et al. 2011) at time t and t+7 days. The magnitude of the perturbations is defined instantaneously at each latitude and for the zonal mean as:

$$'\Upsilon(y,t) = [\gamma(x,y,t)^2]^{1/2}, \tag{2}$$

where [] is the zonal mean. The mean magnitude of all possible perturbations from the period 1981-2010 based on seven-day differences in a given month is simply formed by time averaging the instantaneous amplitudes:

$$\bar{\Upsilon}(y) = \{\Upsilon(y,t)\}, \tag{3}$$

where {} indicates a time average. The mean magnitude of these raw perturbations over 1981-2010 is displayed in Figure 1a (green) for the month of May (as an example). As expected, the amplitude is greatest in the extratropics and a minimum in the tropics.

To get an idea of the relative magnitude of these perturbations, we compare to the daily standard deviation of the analyses similarly defined:

$$\bar{A}(y) = \{[Surfp(x,y,t)'^2]^{1/2}\}, \tag{4}$$

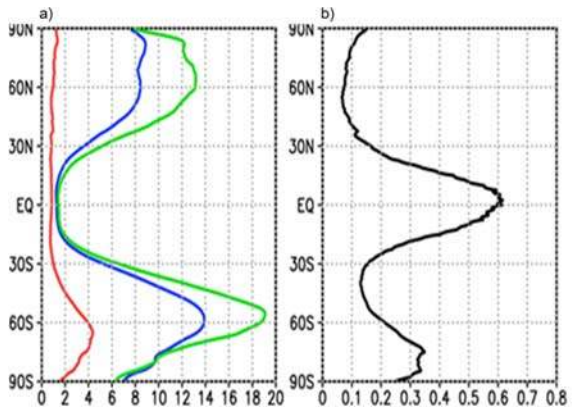
where the prime indicates an anomaly from the daily climatological seasonal cycle. This zonal mean standard deviation of the analyses $\bar{A}(y)$ for the month of May is shown in Figure 1a (blue). It is seen to have similar latitudinal shape as for the mean of the seven-day differences, $\bar{\Upsilon}(y)$, but of slightly smaller amplitude in the extratropics. Importantly, however, in the tropics, the daily standard deviation of the analyses is of near equal magnitude to the amplitude of the seven-day differences (the implication of this is discussed below).

The magnitude of a seven-day difference perturbation, $'\Upsilon(y,t)$, is then scaled to have amplitude equal to analysis uncertainty. We define analysis uncertainty as the time mean, root mean squared instantaneous difference of two state-of-the-art reanalyses for surface pressure. In this case we compare ERA-Interim reanalyses (Dee et al. 2011) to the NCEP-NCAR reanalyses (Kalnay et al. 1996):

$$\overline{A}(y) = \{[(SurfP_{ERA}(x,y,t) - SurfP_{NN}(x,y,t))^2]^{1/2}\}. \tag{5}$$

The mean analysis uncertainty $\overline{\Delta}(y)$ for the month of May is displayed in Figure 1a (red). In contrast to the daily standard deviation, the analysis uncertainty is relatively constant in latitude, but with a peak in the higher latitudes of the Southern Hemisphere, which probably reflects less in situ data in order to constrain the analyses.

Figure 1 Examples of the perturbation calculation for 1st May 00Z data: (a) the mean magnitude of raw perturbations based on all randomly selected seven-day differences of surface pressure ($\overline{Y}$; green line) from 1981-2010, daily standard deviation of ERA-Interim reanalysis surface pressure ($\bar{A}$; blue line) and the mean magnitude of observational uncertainty ($\overline{\Delta}$; red line) and (b) the mean raw scaling factor (prior to limiting to maximum value 0.2).



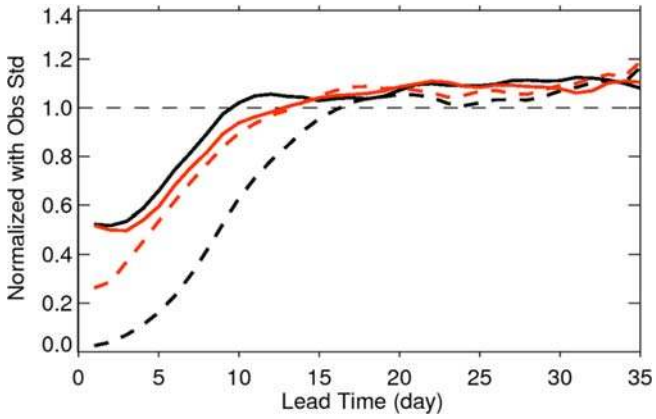
The scaling factor for each seven-day difference perturbation (applied to the full 3 dimensional fields of u v, T, and q and surface pressure) is then derived as the ratio of mean analysis uncertainty to the instantaneous zonal mean amplitude of the seven-day difference: $\overline{\Delta}(y)/\,'\Upsilon(y,t)$. The mean amplitude of the scaling factor, $\overline{\Delta}(y)/\,\{'\Upsilon(y,t)\}$, is displayed in Figure 1b for the month of May 1981-2010. In the extratropics, the scaling factor is generally less than about 0.2. However, in the tropics the mean scaling factor exceeds 0.5, reflecting that analysis uncertainty is relatively large in the tropics. Hence, scaling the seven-day difference perturbations using this full scaling factor would result in adding perturbations in the tropics that have similar magnitude as the observed standard deviation, which could have a deleterious effect on the forecast quality. We thus chose to limit the maximum scaling factor to always be < 0.2. This decision was ad hoc, but appears to have minimal impact in the extratropics, while in the tropics the resulting ensemble is slightly under dispersed but appears to be an optimum compromise between error growth and reliability. In order to optimise the efficiency of the code, we created pairs of perturbed initial conditions by adding in the positive and negative of each perturbation.

As an example, to create a perturbed initial condition for a hindcast initialized on 1 November 1998, we randomly select another date in November from the period 1981-2010, excluding 1998. Assume we select 9 November 2007. We then generate the difference fields for surface pressure, u, v, T, and q from the ERA-interim reanalyses on 16 November 2007 and 9 November 2007. We then scale these difference fields by a multiplier that is only a function of latitude and is applied equally to all variables at all levels. The multiplier forces the root zonal mean square amplitude of the seven-day surface pressure differences to have amplitude equal to the root time mean, zonal mean square difference in surface pressure between NCEP-NCAR and ERA-Interim reanalyses, with the provision that the scale factor cannot exceed 0.2. The scaled perturbations for surface pressure, u, v, T, and q are then added into the unperturbed initial conditions. This procedure is repeated four times by randomly selecting four other dates. Another five perturbed initial conditions are formed by subtracting, rather than adding, these five sets of perturbations from the unperturbed initial condition.

The efficacy of this method of generating perturbed initial conditions is demonstrated in Figure 2, which shows the root-mean-square error (RMSE) and ensemble spread for daily 500 hPa geopotential height anomalies over the Southern Hemisphere extratropics (20°S-60°S) from the Met Office GloSea5-GC2 system (the control) and from a sensitivity experiment using ACCESS-S1 designed to test the impact of the new method of ensemble generation. Both use a five-member ensemble initialised with ERA-Interim reanalyses (Dee et al. 2011) on the 1st May for 1990-2012 (note: although there are only five ensemble members, the results shown in Figure 2 are computed over more than 30,000 grid boxes). For the control (black curves), the spread is produced solely through the stochastic physics scheme. The sensitivity experiment (red curves) uses an unperturbed central member plus two pairs of perturbed members, developed as described above. Stochastic physics is also included in the perturbed members. The verification data are the National Centers for Environmental

Prediction/National Center for Atmospheric Research (NCEP-NCAR) global reanalysis 1 dataset (Kalnay et al. 1996). The skill scores (RMSE and spread) are normalised by the standard deviation of the NCEP-NCAR data. For a climatological forecast (zero anomaly), the normalised RMSE (NRMSE) would be one, and thus forecasts are typically deemed skilful for NRMSE < 1 (i.e., forecasts are skilful when the error is smaller than that from forecasting climatology). The perturbed ensemble (red) is skilful to about lead time twelve days, whereas the control is limited to about nine days (black; Figure 2). The biggest difference between the control and the experiment is the much improved relationship between the error and ensemble spread in the sensitivity experiment. In a perfect ensemble system, over a large sample of forecasts, the ensemble spread will be equal to the RMSE of the ensemble mean for all lead times (Palmer et al. 2006; Weisheimer et al. 2011). The difference between the error and spread curves is much smaller in the sensitivity experiment compared to the control (Figure 2). The control is significantly under-dispersed out to the limit of predictability (i.e. ~day twelve), which will result in unreliable probabilistic forecasts on multi-week timescales.

Figure 2 Normalised RMSE (solid) and ensemble spread (dashed) for daily 500 hPa geopotential height anomalies over the Southern Hemisphere extratropics (20°S-60°S) from the Met Office GloSea5-GC2 system (black) and from a sensitivity experiment using ACCESS-S1 (red). Forecasts were initialised on 1st May for the period 1990–2012 and are based on five-member ensemble means. The verification data are from the NCEP-NCAR reanalysis (Kalnay et al. 1996).



Similar spread-error relationships as seen in Figure 2 for the Southern Hemisphere are obtained for the Northern Hemisphere extratropics (not shown). In the tropics (not shown), the new perturbation method results in initially slow growth of forecast spread (but still with far more spread than from stochastic physics alone), suggesting that the new method is far from optimal for the tropics where diabatic heating anomalies are playing a primary role, for instance resulting from the Madden-Julian Oscillation.

The new method is a workable first approach to extend the capability of the S1 seasonal system to multi-week lead times, but will be improved with subsequent development of a coupled assimilation-initialisation method in ACCESS-S2. Forecast reliability in the real-time system will additionally be improved by including time-lagged forecasts initialised one day apart. However, as mentioned previously, the long lag in the hindcasts (~eight days) will negatively impact the skill and will not be indicative of what could be obtained with one-day lags as will be available in the real-time system. For the real-time system (see below), the size of the ensemble (i.e. the number of one-day lags) will be determined by the number of successively earlier forecasts starts that can be included without compromising the skill.

2.3 Real-time and hindcast system configuration

The ACCESS-S1 operational real-time system includes a perturbed ensemble of eleven members (five pairs of perturbed members and an unperturbed central member) generated every day and running out to a full six calendar months. An additional twenty-two members per day are run out to six weeks lead time to support the multi-week products. Forecast products are based on a 99-member lagged ensemble (i.e., using nine successive days of forecasts for seasonal products and three successive days of forecasts for multi-week products).

A hindcast set is required to support the real-time system. These hindcasts (retrospective forecasts) are used to provide statistical calibration for the real-time system (e.g. define mean climatologies, thresholds, deciles, etc.) and to estimate forecast quality and evaluate model performance, which informs the operational service and can guide future improve-

ments. In addition, the hindcasts provide a dataset for developing products for BoM services, such as the Climate Outlooks and Seasonal Streamflow Forecasts and for interfacing to applications modelling, such as crop modelling. ACCESS-S1 uses the same ocean, sea-ice, atmosphere and land initial conditions as the Met Office GloSea5-GC2 system. It is not possible to run hindcasts daily so as to be directly compatible with real-time system, both because the daily initial conditions are not available from Met Office and the computing cost would be prohibitive. Initial conditions for the hindcasts are only available on the 1st, 9th, 17th, and 25th of the month. Hence, the hindcast set for ACCESS-S1 consists of an eleven-member ensemble from each of the 1st, 9th, 17th, and 25th of every month for the period 1990-2012.

The hindcasts generated on the following start-dates are used in this paper to estimate the performance of ACCESS-S1 over the twenty three-year period: 25th of January, April, July, October; and 1st of February, May, August, November. For the verification of thee-month mean seasonal forecasts we have created a time-lagged twenty-two-member ensemble by combining the forecasts initialised on the 1st of the month with those from the 25th of the month prior (e.g., eleven members from 25th January combined with eleven members from 1st February). These dates allow analysis of the performance of forecasts of the four standard seasons at a one-month lead time i.e., December-February (DJF), March-May (MAM), June-August (JJA) and September-November (SON).

A time-lagged ensemble is not used for verification of the multi-week forecasts (i.e. the eleven-member ensemble from the 1st and 25th of the months are used as independent forecasts). For forecasts on the multi-week timescale, the skill of forecasting fortnight one (weeks one and two), fortnight two (weeks two and three) and fortnight three (weeks three and four) of the forecasts is assessed.

A sufficiently long hindcast set is necessary for including an adequate number of cases of the low frequency influences on Australian climate, such as ENSO, and for determining how skill may vary based on the state of these climate drivers. A twenty three-year hindcast set is likely too short to be able to stratify seasonal timescale skill based on different climate drivers and their associated phases. In addition, a sample size of twenty three is considered small for obtaining statistically robust results. For the next version of the system, ACCESS-S2, the hindcast period will be increased to 30+ years, made possible with the implementation of the BoM's own data assimilation system, hence not being reliant on the Met Office initial conditions, as it is for ACCESS-S1.

3 The POAMA-2 system: A baseline for comparison

In this study, we compare the performance of the ACCESS-S1 system to the POAMA system which is currently used to make operational seasonal forecasts at BoM (forecasts available at http://www.bom.gov.au/climate/ahead/). Note that POAMA remains experimental for forecasts on the multi-week timescale. For full details of the model, data assimilation and ensemble generation refer to Hudson et al. (2013; note, the current system is referred to in their paper as P2-M). A brief description of POAMA is provided here.

POAMA is a fully coupled ocean-atmosphere model and data assimilation system. The atmospheric model component is the BoM's Atmospheric Model version 3 (BAM3; Colman et al. 2005) and has a T47 horizontal resolution (~250 km) and seventeen levels in the vertical. This low horizontal resolution means that Tasmania is not resolved as land. The land surface component is a simple bucket model for soil moisture (Manabe and Holloway, 1975) and has three soil levels for temperature. The ocean model is the Australian Community Ocean Model version 2 (ACOM2) (Schiller et al., 1997; 2002), which is based on the Geophysical Fluid Dynamics Laboratory Modular Ocean Model (MOM version 2). The ocean grid resolution is 2° in the zonal direction and 0.5° in the meridional direction at the Equator, which gradually increases to 1.5° near the poles. The ocean model has 25 vertical levels. The atmosphere and ocean models are coupled using the Ocean Atmosphere Sea Ice Soil (OASIS) coupling software (Valcke et al., 2000). POAMA has no sea ice model and there are no time-varying greenhouse gases. The CO2 level is fixed to 345 ppm and ozone and sea ice are climatological.

Forecasts are initialised from observed atmospheric and oceanic states. POAMA obtains ocean initial conditions from the POAMA Ensemble Ocean Data Assimilation System (PEODAS; Yin et al. 2011) and atmosphere and land initial conditions from the Atmosphere-Land Initialisation scheme (ALI; Hudson et al. 2011). The ALI assimilation consists of a nudging of the model atmospheric fields (u, v, T, and q) towards an existing analysis every six hours. For the period 1980-2002, the ERA-40 reanalyses (Uppala et al. 2005) are used, and for the period 2003-onward the analyses from the BoM's operational NWP system are used. Through this nudging process the land surface is realistically initialised in response to the surface fluxes generated in the atmosphere.

PEODAS, the ocean assimilation, is an approximate form of an Ensemble Kalman Filter and generates an ensemble of ocean states each day including a central unperturbed ocean analysis (Yin et al 2011). PEODAS is based on the multivariate ensemble optimum interpolation system of Oke et al. (2005), but uses covariances derived from the time evolving model ensemble. In situ temperature and salinity observations are assimilated into the central unperturbed analysis, and corrections to currents are generated based on the ensemble cross-covariances with temperature and salinity. The ensemble of ocean states is generated by perturbing the surface fluxes within a possible range of observational errors.

Perturbed initial atmosphere-ocean-land states are produced by a breeding scheme by integrating an ensemble of ten members of coupled model integrations that are nudged back to the central member (from the assimilation) once per day (Hudson et al. 2013). To address model uncertainty, POAMA has adopted a multi-model ensemble strategy using three different configurations of the atmospheric model. Hence, a 33-member ensemble is generated that comprises an eleven-member ensemble for each model version.

The POAMA hindcast spans 1981-2014, with a 33-member ensemble generated every five days (on the 1st, 6th, 11th, 16th, 21st, 26th of every month). In order to make a fair comparison with ACCESS-S1, the analysis in this paper is performed over the twenty three common hindcast years (1990-2012), using the closest matching start dates (i.e. the 1st and 26th from POAMA) and only eleven ensemble members per start date from POAMA. In order to sample model uncertainty fairly from POAMA, we use ensemble members from each of the three versions, as follows: four members from version 1, three members from version 2, and four members from version 3. As described for ACCESS-S1 (Section 2.3), a time-lagged twenty-two member-ensemble is used for the seasonal forecasts.

4 Forecast system evaluation

4.1 Mean state bias

The mean state bias is considerably reduced in ACCESS-S1 compared to POAMA. Figure 3 shows the bias (for all the standard seasons combined) for sea surface temperature (SST; top) and rainfall (bottom). POAMA exhibits a cold bias over much of the global oceans (Figure 3b), a typical feature of many climate models. The cold bias in ACCESS-S1 is markedly smaller than in POAMA, as are the warm biases along the west coasts of North and South America and southern Africa (Figure 3a). The latter may be due to more realistic upwelling in the higher resolution ACCESS-S ocean and atmosphere models.

ACCESS-S1 does, however, exhibit pronounced cold biases in the equatorial east Pacific, the equatorial Atlantic and the eastern Indian Ocean in the Maritime Continent region (Figure 3a). These biases develop rapidly and generally intensify with lead time (Zhou et al. 2015; Lim et al. 2016a). The cold bias in the Maritime Continent region is most pronounced for forecasts of winter (JJA) and spring (SON) (Lim et al. 2016a).The equatorial cold biases in ACCESS-S1 SST are related to dry rainfall biases in most seasons (not shown), thus clearly seen in the mean bias across the seasons (Figure 3c), although the dry bias (like the cold bias) is prominent in winter and spring over the Maritime Continent (Lim et al. 2016a). There is a wet bias over the central-western Indian Ocean (Figure 3c), which occurs primarily in spring (SON) and summer (DJF) (Lim et al. 2016a). It is likely that the SST and rainfall biases over the Indian Ocean have a negative impact on the prediction skill of the Indian Ocean Dipole (IOD) (Section 4.2.2) and the teleconnection between the Indian Ocean and Australian climate (Lim et al. 2016a).

The higher resolution of ACCESS-S1 results in a much improved representation of the climate of Australia. For example, Figure 4 shows the seasonal mean climatology for spring (SON) for rainfall, maximum temperature (Tmax) and minimum temperature (Tmin) from ACCESS-S1 and POAMA compared to observations. The observations used are the 5 km resolution Australian Water Availability Project (AWAP) gridded datasets (Jones et al. 2009). Tasmania is not resolved as land in POAMA (not shown on the POAMA maps in Figure 4), whereas it is resolved in ACCESS-S1 and the model is clearly able to differentiate between the very different rainfall climatology of western and eastern Tasmania (Figure 4). ACCESS-S1 better captures the high rainfall associated with the Great Dividing Range over south-eastern Australia. POAMA has a strong dry bias over the whole continent in spring, which is greatly reduced in ACCESS-S1. ACCESS-S1 also has a better representation of south-western Australian rainfall, where POAMA is again too dry (Figure 4). In addition, ACCESS-S1 has a much improved climatology for temperature (Figure 4). There is an improved representation of the coldest temperatures (Tmax and Tmin) along the east coast and the Great Dividing Range. Although the errors are considerably less in ACCESS-S1, there are still notable biases, with Tmax being too cold over much of Australia in spring (Figure 4, but dif-

ference maps not shown), as well as in the other seasons (not shown). For Tmin in spring, ACCESS-S1 has a cold bias over most of central and northern Australia, but there is a warm bias over eastern and southern regions. In the other three seasons the warm bias is more pronounced and also dominates northern Australia (not shown). The largest bias for rainfall in ACCESS-S1 occurs in summer (DJF), where there is a distinct wet bias over most of Australia (not shown).

Figure 3 Bias in the ensemble mean climatology of ACCESS-S1 (left) and POAMA (right) for three-month mean SST (°C; top) and rainfall (mm/day; bottom) at one-month lead time from all four start dates (based on a twenty-two-member ensemble) for the period 1990-2012. Bias is calculated relative to Reynolds OI v2 SST analyses (Reynolds et al. 2002) for SST and ERA-Interim (Dee et al. 2011) for rainfall. The root-mean-square deviation (RMSd) is indicated for each panel.

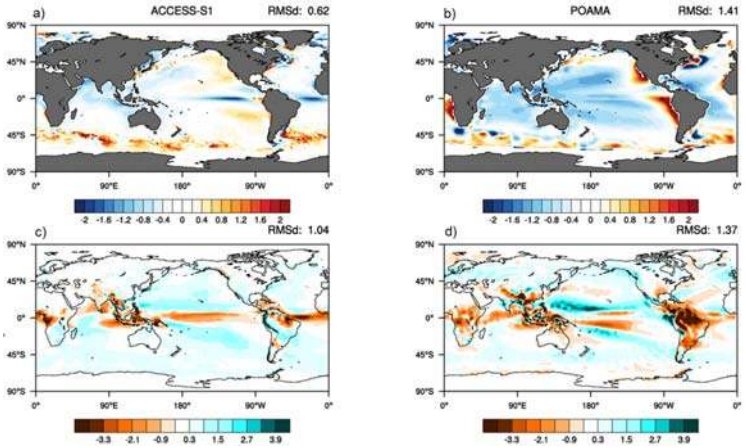
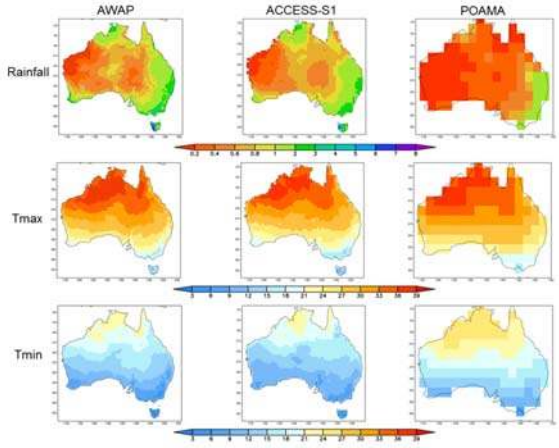


Figure 4 Rainfall (mm/day; top row), Tmax (°C; middle row) and Tmin (°C; bottom row) climatology for the spring season (SON) from AWAP observations (Jones et al. 2009) (left column), ACCESS-S1 (middle column) and POAMA (right column). The results are shown on the respective native grids of the observations (5 km) and models (N216 and T47 respectively). The climatology for ACCESS-S1 and POAMA are at a one-month lead time.



Having a good climatology, as ACCESS-S1 generally does for Australian climate (and much improved compared to POAMA), does not necessarily ensure that forecasts of departures from normal will be skilful. The skill of the forecasts is addressed in the following sections.

4.2 Climate drivers

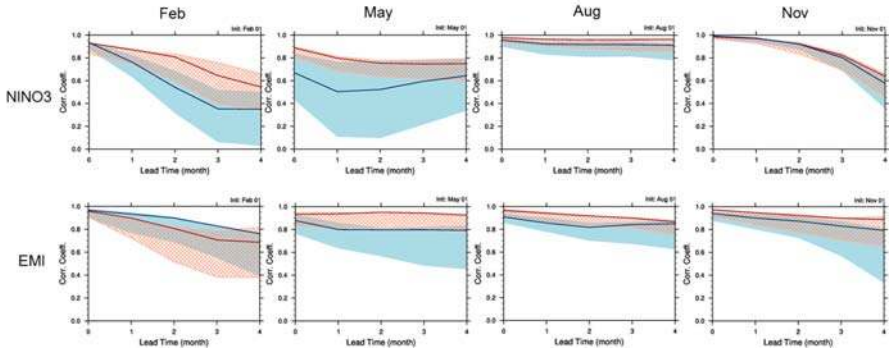
In this section we evaluate the ability of ACCESS-S1 to predict some of the key large-scale climate drivers responsible for predictable multi-week to seasonal variations of Australian climate. Skilful regional climate prediction also requires that

the model can simulate the teleconnection of the drivers to regional climate. Examining the teleconnections is beyond the scope of this paper, but Lim et al. (2016a) have done a preliminary investigation, focussing on teleconnections relevant to south-eastern Australia. They found that the teleconnection strength of ENSO and the IOD to rainfall over south-eastern Australia is substantially weaker in ACCESS-S1 than observed. Furthermore, the teleconnection of the SAM to south-eastern Australian rainfall is deficient in late autumn to winter (Lim et al. 2016a).

4.2.1 ENSO

ENSO is the primary driver of seasonal to interannual variations of Australian rainfall, particularly over eastern regions during winter (JJA) and spring (SON) (e.g., Risbey et al. 2009; Langford et al. 2011). Figure 5 (top row) shows the skill in forecasting the NINO3 index for both POAMA (blue) and ACCESS-S1 (red). The NINO3 index measures the SST anomaly in the eastern Pacific (150ºW-90ºW, 5ºN-5ºS), a key region of strong SST variability. Skill is calculated using the temporal correlation coefficient between predicted and observed indices and is based on three-month running mean data. The skill of the ensemble mean (based on a twenty-two-member ensemble) as well as the range in skill from the individual ensemble members (cross-hatched/shaded regions) is shown. Strong persistence of ENSO from austral winter to spring and into summer accounts for the good prediction skill of NINO3 that is seen for both POAMA and ACCESS-S1 in forecasts that start in August and November (Figure 5, top row). Both POAMA and ACCESS-S1 show reduced skill in predicting NINO3 anomalies for forecasts initialised in February and May (Figure 5, top row), highlighting the well-known difficulty in predicting ENSO through the so-called boreal spring predictability barrier. However, it is for these forecast start dates that ACCESS-S1 produces significantly more skilful forecasts for NINO3 compared to POAMA, providing improved predictions of the development of ENSO throughout autumn (MAM) and winter (JJA) (Figure 5).

Figure 5 Correlation skill for predictions of the NINO3 index (top row) and the El Niño Modoki index (EMI; bottom row) as a function of lead time from ACCESS-S1 (red and cross-hatching) and POAMA (blue and shaded) forecasts initialised on the 1st of February, May, August and November respectively. The forecasts are based on the twenty-two-member ensemble and the skill of the ensemble mean (solid line) and the range in skill of the individual ensemble members (cross-hatching/shading) is shown. The Reynolds OI v2 SST analyses (Reynolds et al. 2002) are used as the observations.



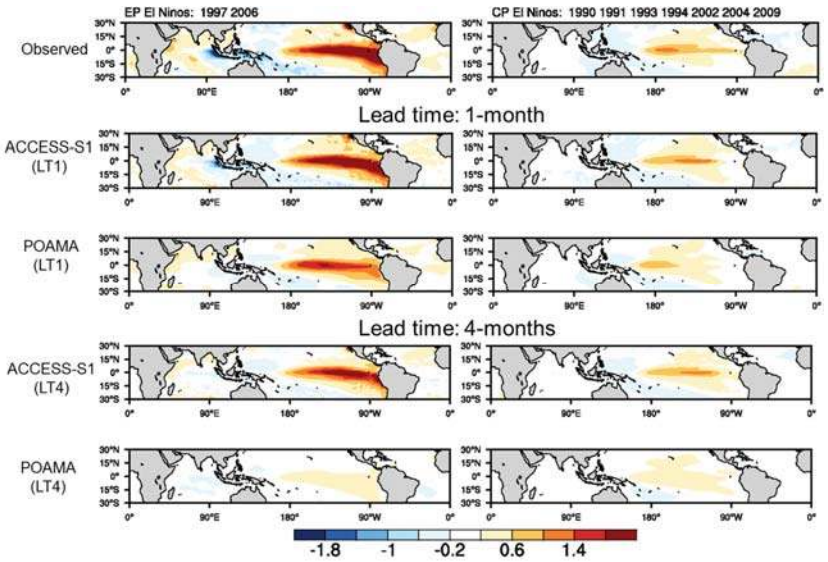
Australian rainfall is sensitive to east-west variations of ENSO, such that below-average rainfall is more strongly related to El Niño events that have their maximum warming located in the central Pacific rather than the eastern Pacific (Wang and Hendon 2007; Hendon et al. 2009; Lim et al. 2009). The El Niño Modoki index (EMI; Ashok et al. 2007) is a measure of these central Pacific ENSO events, and is calculated as

$$\mathrm{EMI} = \overline{SST}_{(10°\mathrm{S}\text{-}10°\mathrm{N},\ 165°\text{-}220°\mathrm{E})} - 0.5*\overline{SST}_{(15°\mathrm{S}\text{-}5°\mathrm{N},\ 250°\text{-}290°\mathrm{E})} - 0.5*\overline{SST}_{(10°\mathrm{S}\text{-}20°\mathrm{N},\ 125°\text{-}145°\mathrm{E})}, \quad (6)$$

where the overbar indicates the spatial average of the SST anomaly. The EMI has one of the strongest relationships with eastern Australian rainfall from autumn through to spring (Langford et al. 2011; Lim et al. 2012) compared to other ENSO indices. ACCESS-S1 produces notably more skilful EMI forecasts than POAMA for forecasts initialised in May and August, i.e., for predictions of austral winter and spring El Niño Modoki conditions (Figure 5; note that while POAMA shows improved prediction of the EMI from February compared to ACCESS-S1, there is considerable overlap between the two systems in the range of skill from the ensemble members).

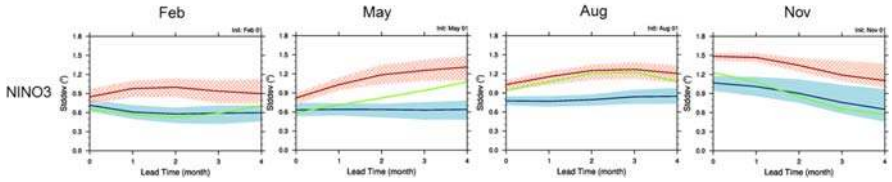
In addition, ACCESS-S1 is better able to distinguish between the patterns of eastern Pacific (EP) and central Pacific (CP) El Niño events, particularly as lead time increases. Figure 6 shows SST composites for the spring (SON) season for EP (1997, 2006) and CP (1990, 1991, 1993, 1994, 2002, 2004, 2009) El Niño events from observations and from forecasts at 1-month and 4-month lead times from ACCESS-S1 and POAMA. Years are classified as EP if the standardised NINO3 index is > 1 and the standardised EMI index is < 0.7; and as CP if the standardised EMI is ≥ 0.7. This analysis is somewhat limited by the small number of EP El Niño events in the hindcast period. As discussed in Sullivan et al. (2016) and Zhao et al. (2016), CP El Niño events have been more frequent in recent decades and EP El Niños less frequent. Nonetheless, it is clear that the POAMA SST anomaly patterns of the EP and CP El Niños become increasingly indistinguishable as lead time increases (see also Hendon et al. 2009), whereas ACCESS-S1 is able to maintain the distinction between the two types of El Niño (Figure 6). ACCESS-S1 is also able to maintain realistic amplitude of the anomalies with lead time for both EP and CP El Niños, where POAMA's SST anomalies become too small (Figure 6). Lim et al. (2016a) also found that ACCESS-S1 better captures the teleconnection between central Pacific ENSO and south-eastern Australian rainfall when forecasts are initialised in August compared to POAMA.

Figure 6 SST anomaly (°C) composites for two eastern Pacific (EP) El Niño events (left) and seven central Pacific (CP) El Nino events (right) during spring (SON). Observations (top row) and ensemble-mean forecasts (based on twenty-two-member ensembles) at 1-month (initialised on 1st August) and at four-month (initialised on 1st May) lead times from ACCESS-S1 and POAMA are shown.



The ACCESS-S1 predicted amplitude of the EMI index (given by the standard deviation of the magnitude of the EMI index) is generally realistic for all start dates and lead times, but is underestimated by POAMA (not shown, but see Lim et al. 2016a). However, ACCESS-S1 tends to predict ENSO with too strong amplitudes when defined using the NINO3 (Figure 7) and NINO34 (not shown) indices. Figure 7 shows that ACCESS-S1's amplitude of ENSO (NINO3) is generally realistic for forecasts initialised in August for the subsequent spring and summer, but at other times it is too large.

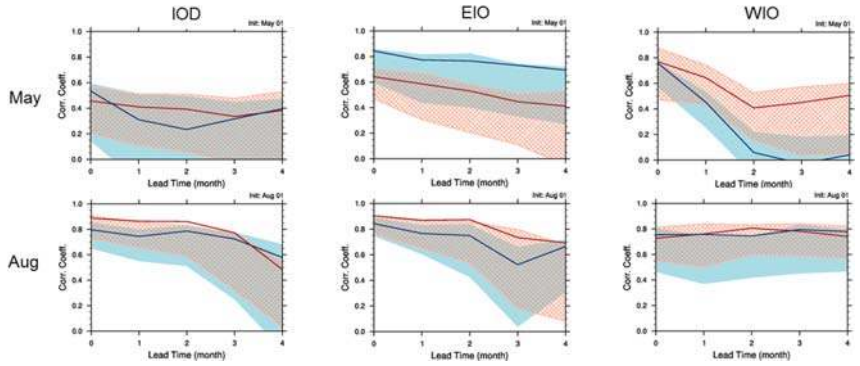
Figure 7 As for Figure 5, but showing the standard deviation of the magnitude of the predicted NINO3 index. The green line indicates the observed standard deviation of NINO3.



4.2.2 Indian Ocean Dipole (IOD)

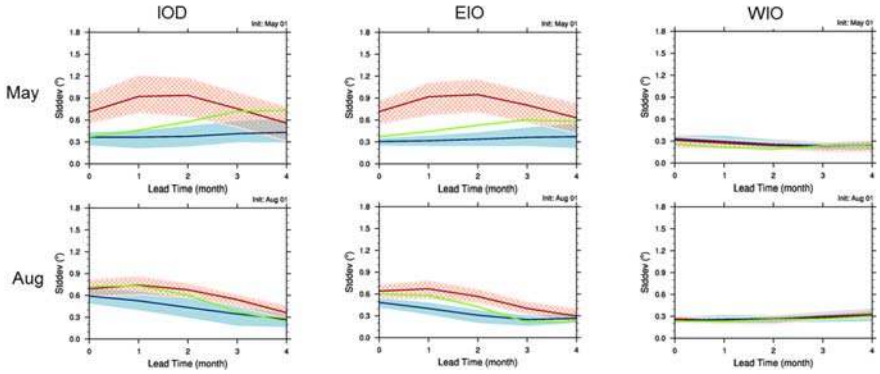
The IOD is another important driver of Australian rainfall, particularly over southern regions in winter and spring (e.g., Risbey et al. 2009; Langford et al. 2011). The IOD is monitored with the Dipole Mode Index (DMI), which is the SST anomaly difference between the western (50–70°E; 10°S–10°N) and eastern (90–110°E; 10°S–0°) tropical Indian Ocean (Saji et al. 1999). The IOD amplitude peaks in spring (SON) and is at a minimum in January-April (e.g. Zhao and Hendon 2009). Figure 8 shows the skill of forecasts of the IOD, and the eastern and western component poles, from May and August forecast starts (i.e., covering periods when the IOD is most active). There is a large degree of overlap between the range in ensemble-member skill from ACCESS-S1 and POAMA. However, as shown by the ensemble mean, the winter-to-early spring IOD is better predicted by ACCESS-S1 for both start times (Figure 8). For forecasts from May, this appears to stem from the much improved skill in predicting the western pole. Improved skill in the western Indian Ocean is presumably derived from improved prediction of eastern Pacific ENSO (Figure 5), as tropical western Indian Ocean SST variations are largely a remote response to El Niño in the Pacific. However, even with this improvement made by ACCESS-S1, the skill for predicting the IOD from May initial conditions is not high, likely due to inherent limits of predictability of the IOD (Shi et al. 2012) and the reduced predictability of the IOD post-2000 (Lim et al. 2016b).

Figure 8 Correlation skill for predictions of the IOD (left), eastern pole of the IOD (EIO; middle) and western pole of the IOD (WIO; right) as a function of lead time from ACCESS-S1 (red and cross-hatching) and POAMA (blue and shaded) forecasts initialised on the 1st of May (top) and August (bottom row). The forecasts are based on the twenty-two-member ensemble and the result for the ensemble mean (solid line) and the range from the individual ensemble members (cross-hatching/shading) is shown. The Reynolds OI v2 SST analyses (Reynolds et al. 2002) are used as the observations.



As with ENSO, the IOD is too strong in ACCESS-S1 for forecasts initialised in May for the subsequent winter (JJA) and spring (SON) seasons (Figure 9). This stems from the larger than observed amplitude of the eastern pole of the IOD (Figure 9). In contrast, IOD activity in POAMA is somewhat damped (Figure 9).

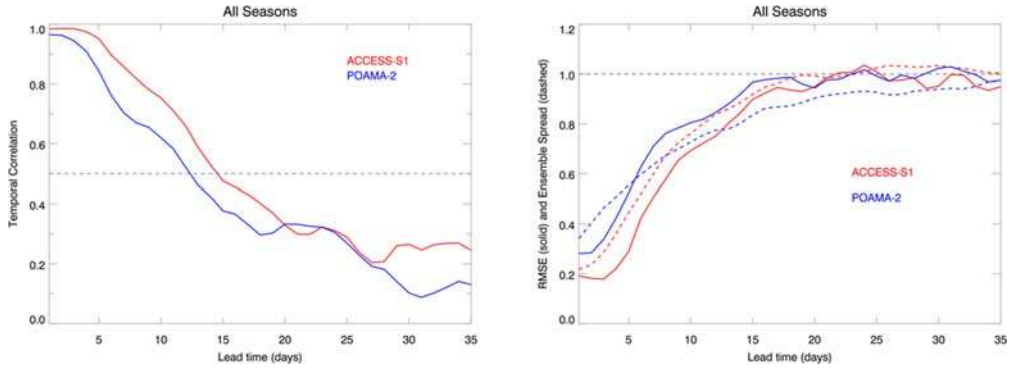
Figure 9 As for Figure 8, but showing the standard deviation of the magnitude of the predicted IOD, EIO and WIO index. The green line indicates the observed standard deviation of the IOD, EIO and WIO respectively.



4.2.3 Southern Annular Mode (SAM)

In previous work we have assessed the ability of POAMA to predict the SAM (Marshall et al. 2012; Hudson et al. 2013), an important driver of multi-week climate variability and extremes over Australia (Marshall et al. 2012; Marshall et al. 2014; Lim et al. 2016c). The daily SAM index is calculated for observations by projecting observed daily mean sea level pressure (MSLP) anomalies (NCEP–NCAR reanalysis; Kalnay et al. 1996) onto the leading empirical orthogonal function (EOF) of observed zonally averaged monthly mean MSLPs between 25°S and 75°S. For POAMA and ACCESS-S1, the daily mean predicted MSLP anomalies from each model ensemble member, respectively, are projected onto the observed EOF to calculate their respective indices (see Marshall et al. 2012 for details). The temporal correlation, root-mean-square error (RMSE) and ensemble spread of the SAM index are shown in Figure 10. ACCESS-S1 shows improved skill to predict the SAM index compared to POAMA (Figure 10). Correlation skills above 0.5 are achieved out to about 15 days for ACCESS-S1 and 12 days for POAMA. Based on the RMSE statistic, ACCESS-S1 produces skilful forecasts of the SAM out to about 21 days and POAMA out to 17 days (based on the period over which the RMSE is less than that of a climatological forecast) and the ACCESS-S1 error is less than that of POAMA over the entire first two weeks of the forecast. ACCESS-S1 also has an ensemble with an improved error-spread relationship, i.e. there is a closer match between the ensemble mean error and the ensemble spread. POAMA forecasts are over-dispersed (spread exceeds error) in the first 5 days and then under-dispersed (spread is less than the error) thereafter. ACCESS-S1 generally has a well dispersed ensemble, although it is slightly over-dispersed.

Figure 10 SAM index forecast skill calculated from all eight forecast start dates (based on eleven-member ensembles; see section 2.3) in the 23-year hindcast period. Correlation of the ensemble mean with observed (left) and RMSE (solid) and ensemble spread (dashed) (right) out to 35 days lead time is shown for ACCESS-S1 (red) and POAMA (blue). The dashed horizontal line represents the RMSE of a climatological forecast.



Although SAM predictability has traditionally been regarded as low beyond about two weeks, recent work has shown that there is the potential to make seasonal predictions of the SAM due to its association with ENSO (Lim et al. 2013) and due to persistent stratospheric circulation anomalies that couple to the troposphere in austral spring (Seviour et al. 2014). ACCESS-S1 produces an improved prediction of the seasonal SAM (based on three-month mean data) at both 0-month and one-month lead times (at 0-month lead the correlation with observed is r=0.50 for POAMA and r=0.57 for ACCESS-S1 for all four forecast start dates combined and using a twenty-two-member ensemble, and at one-month lead r=0.36 for POAMA and r=0.47 for ACCESS-S1).

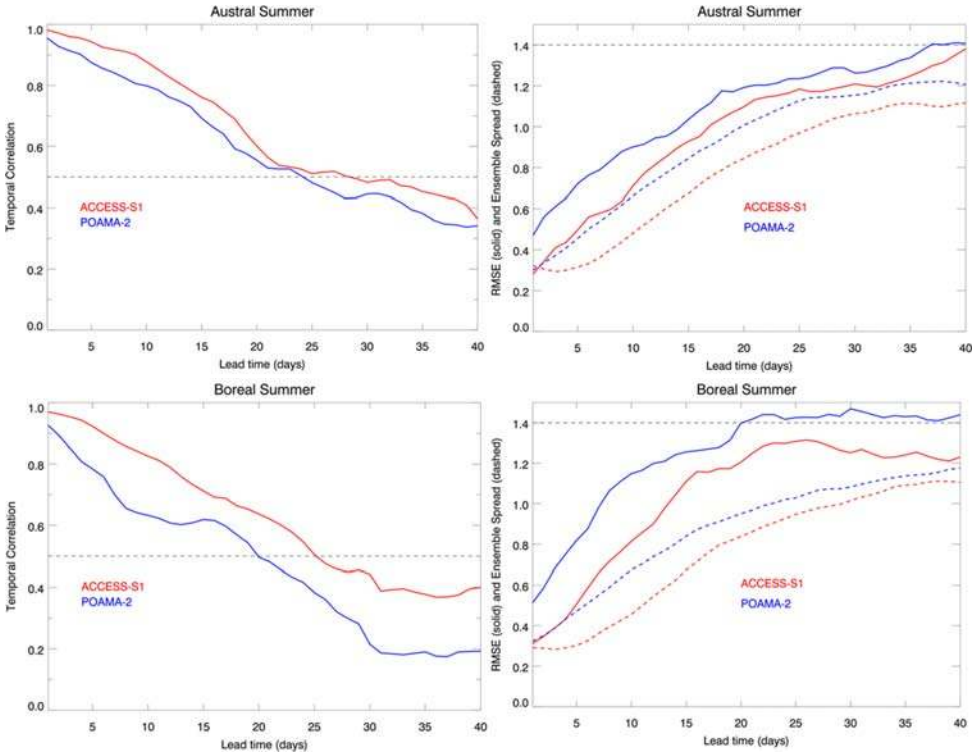
4.2.4 Madden-Julian Oscillation (MJO)

The MJO (Madden and Julian 1971) is the dominant mode of intra-seasonal variability in the tropics and recognised as an important driver of Australian weather and climate (e.g., Wheeler et al., 2009; Marshall et al. 2011; Marshall et al. 2014). The MJO is depicted using the Real-Time Multivariate MJO (RMM) index (Wheeler and Hendon 2004), which captures the eastward propagation of its large-scale structure in convection and zonal wind along the equator. Calculation of the MJO index is described in Marshall et al. (2011; 2014). The NCEP–NCAR reanalysis (Kalnay et al. 1996) is used for the observations.

Figure 11 shows the skill in predicting the MJO for forecasts initialised in the austral summer (start dates: 25 October, 1 November, 25 January, 1 February) (top row) and boreal summer (start dates: 25 April, 1 May, 25 July, 1 August) (bottom

row) halves of the year respectively. In each season ACCESS-S1 provides a four-five day improvement in skill compared to POAMA — with skill (correlation > 0.5) out to about 28 days (compared to 24 days) in the austral summer and about 25 days (compared to 20 days) in the boreal summer (Figure 11). ACCESS-S1 has a lower RMSE than POAMA over the entire 40-day lead time shown for both seasons. ACCESS-S1 shows more improvement over POAMA in the boreal summer prediction of the MJO than the austral summer.

Figure 11 Performance of forecasts of the MJO from ACCESS-S1 (red) and POAMA (blue) initialised in the austral summer half of the year (25 October, 1 November, 25 January, 1 February) (top) and boreal summer half of the year (25 April, 1 May, 25 July, 1 August) (bottom) respectively. The bivariate correlation for the RMM index using the ensemble mean is shown in the left panels. Forecasts are considered skilful for correlations exceeding 0.5. The RMSE of the ensemble mean (solid curves) and the ensemble spread (dashed curves) for the bivariate RMM index is shown on the right panels (the dashed line at 1.4 indicates the RMSE of a climatological forecast).

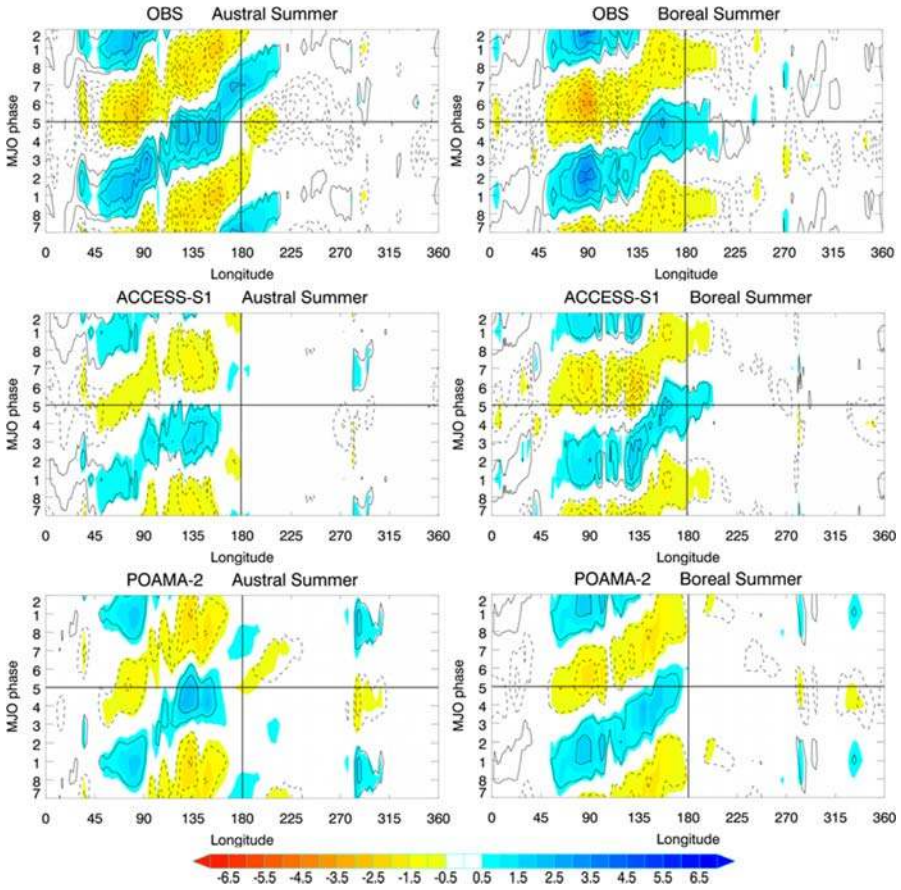


Both POAMA and ACCESS-S1 forecasts are under-dispersed (spread is less than the error) (Figure 11). ACCESS-S1 initially has a well dispersed ensemble (spread matches the RMSE at day one), but the spread does not grow for the next 5 days. In contrast, the ensemble spread from POAMA grows steadily from the start of the forecast. This may reflect the more optimal atmospheric initial condition perturbations in the POAMA system, which may be more effective at emphasising the fast-growing errors associated with large-scale tropical convection. The breeding method perturbs those modes that grow fastest during the analysis cycle and are thus the modes most likely to dominate the analysis errors (Toth and Kalnay 1993; Hudson et al. 2013). The next version of ACCESS-S (version 2) will incorporate this breeding method to generate the initial condition perturbations, so it will be interesting to see if this will improve on the spread-error characteristics of the ensemble compared to ACCESS-S1.

A common problem in models is propagation of the MJO and associated convection across the Maritime Continent region and into the western Pacific. Figure 12 shows composites of rainfall anomalies (averaged 10°S-10°N), for each MJO phase as a function of longitude from forecasts at two-four weeks lead time. For forecasts initialised in the boreal summer half of the year (Figure 12 right panels), ACCESS-S1 shows realistic eastward propagation of the MJO, with rainfall anomalies extending past the Dateline, unlike that with POAMA. For the austral summer half of the year (Figure 12 left panels), eastward propagation into the western Pacific is inhibited in both models, but perhaps slightly more so in ACCESS-S1.

Overall, the depiction of MJO-rainfall in ACCESS-S1 is similar, if not a modest improvement on, MJO-rainfall in POAMA.

Figure 12 Composites of rainfall anomalies (mm/day) (shaded; the standardised rainfall amplitude is shown in contours) averaged 10°S-10°N for each MJO phase as a function of longitude, from observations (top row; Climate Prediction Centre Merged Analysis of Precipitation; Xie and Arkin 1997), and ACCESS-S1 (middle row) and POAMA (bottom row) forecasts at two-four weeks lead time. Results are shown for the austral summer half of the year (forecasts initialised on 25 October, 1 November, 25 January, 1 February) (left) and boreal summer half of the year (forecasts initialised on 25 April, 1 May, 25 July, 1 August) (right).

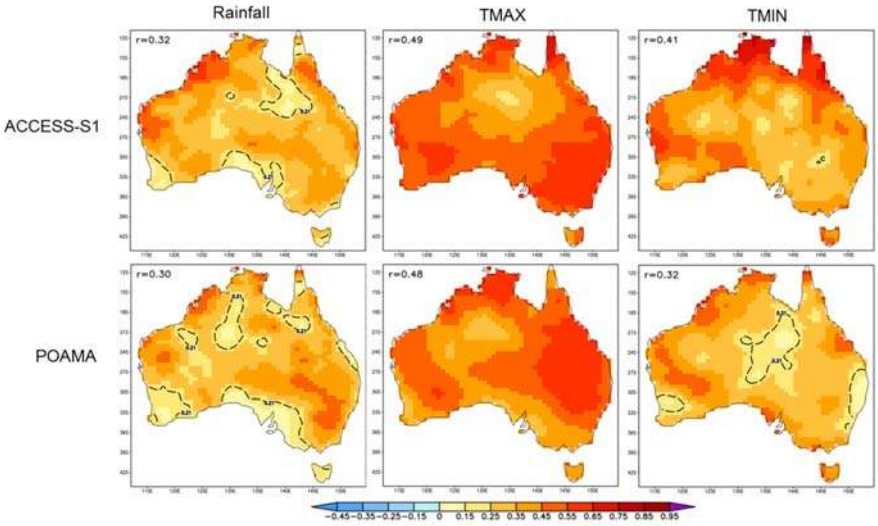


4.3 Skill for forecasts of Australian climate

4.3.1 Seasonal

Overall, ACCESS-S1 has a similar level of performance to POAMA for seasonal skill over Australia for predictions of temperature and rainfall. This is clear from Figure 13 which shows the correlation skill for seasonal forecasts of rainfall, Tmax and Tmin at 0-month lead time for all seasons combined. ACCESS-S1 forecasts of Tmax are, however, more skilful over southern Australia, but less skilful over north-eastern Australia (Figure 13). ACCESS-S1 also generally has more skilful forecasts of Tmin compared to POAMA, particularly over northern Australia.

Figure 13 Correlation skill for 0-month lead time predictions of seasonal mean rainfall (left), Tmax (middle) and Tmin (right) using all four forecast start dates (twenty-two-member ensemble) from ACCESS-S1 (top) and POAMA (bottom). Statistically significant correlations exceed 0.2 (95 per cent confidence level; n=92; dashed line). If there is no dashed line shown on a map, then all the correlations are significant. Observations are the AWAP (Jones et al. 2009) analyses. The average correlation (r) over Australia is shown in each panel.



The skill maps for individual seasons at 0-month and 1-month lead time are shown in Lim et al. (2016a) and Hudson et al. (2017) and summarised Australia-wide in Figures 14-16. The latter figures show the accuracy (proportion correct), reliability error and resolution scores for probabilistic seasonal forecasts of above/below median for rainfall, Tmax and Tmin respectively. Proportion Correct is the number of correct forecasts for the occurrence and non-occurrence of the event divided by the total number of forecasts i.e., it determines the fraction of correct forecasts (a perfect score is 1). To calculate the score, the probabilistic forecast is converted to a category forecast. If the forecast probability of above median rainfall, for example, is greater than 50 per cent, then is categorised as a 'yes' for an above-median event and this is then compared to the observed outcome. The reliability error and resolution scores are terms in the Brier score and are based on the probabilistic forecasts (Wilks 2006). The reliability error measures the conditional bias of the probabilistic forecasts, such that perfect reliability will give a reliability error of 0. Reliability error is calculated as follows:

$$\text{Reliability error} = \frac{1}{n}\sum_{i=1}^{I} N_i(Y_i - \bar{O}_i)^2 \quad , \tag{7}$$

where n is the total number of forecasts, N is the number of forecasts in probability category i, Yi is the probability of i-th category and $\bar{O}_i$ is the mean of relative observed frequency when forecasts in the i-th category are issued (see Murphy 1973, Wilks 2006 and Hudson et al 2013 for more details). A forecast system has good resolution and a high value for the resolution score if the forecast probabilities can discern periods with different observed relative frequencies of the event (i.e., frequencies that differ from the climatological base rate; Wilks 2006). If the climatological probability is always forecast, then the resolution is zero. The score is calculated as follows:

$$\text{Resolution score} = \frac{1}{n}\sum_{i=1}^{I} N_i(\bar{O}_i - \bar{O})^2, \text{ where } \bar{O} \text{ is the mean of } \bar{O}_i. \tag{8}$$

Figure 14 Proportion correct (left), reliability error (middle) and resolution (right) for seasonal rainfall forecasts of above/below median for each season as indicated (using twenty-two-member ensembles) at a 0-month (top) and 1-month (bottom) lead time from ACCESS-S1 (red) and POAMA (blue). Observations are the AWAP (Jones et al. 2009) analyses.

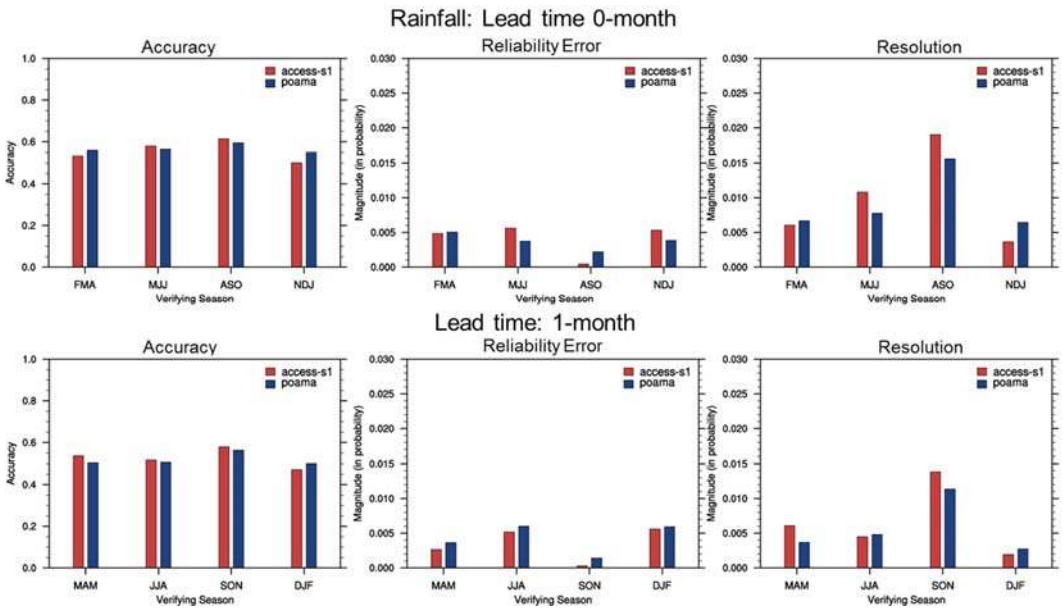


Figure 15 As for Figure 14, but for Tmax.

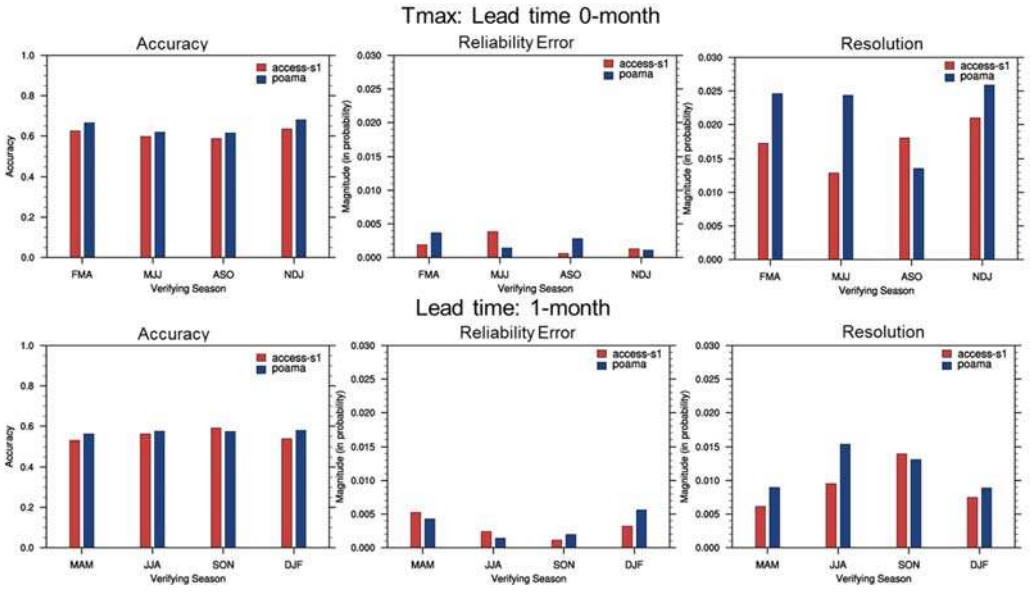
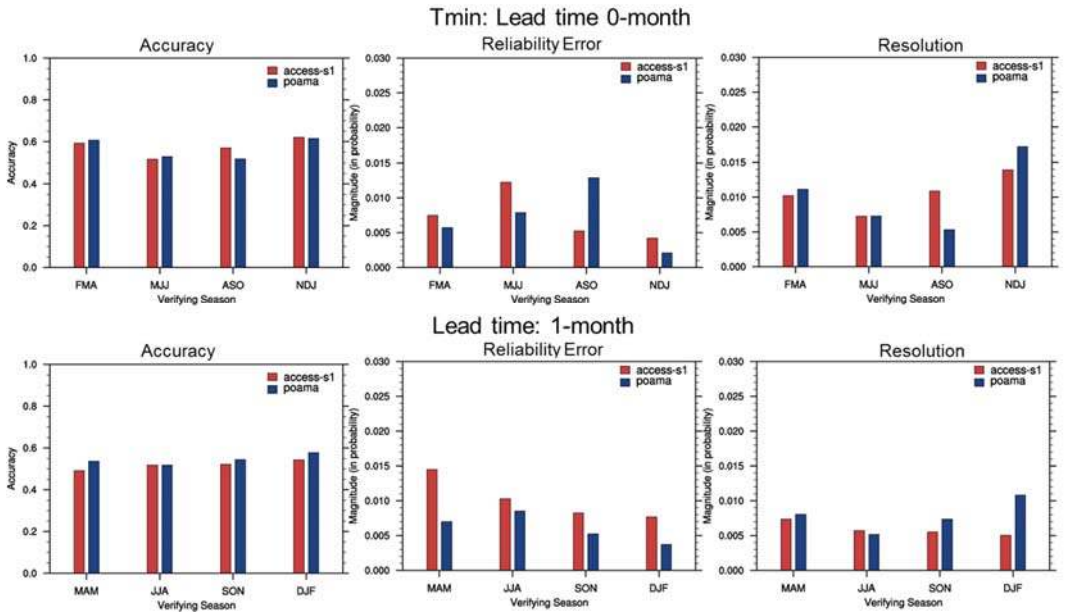


Figure 14 again shows that the skill of forecasting seasonal rainfall is similar between the systems. ACCESS-S1 appears to be an improvement over POAMA for forecasts of late winter (August-October; ASO) and spring (SON), particularly in terms of reliability and resolution in an Australia-wide sense, with reliability being near perfect (Figure 14). On a regional basis, ACCESS-S1 shows improved accuracy over south-eastern, eastern-coastal and south-western Australia in late winter (Figure 17 top row and Lim et al. 2016a). Lim et al. (2016a) also show that ACCESS-S1 has improved and skilful predictions of late autumn (May-July; MJJ) and late winter (ASO) rainfall at a three-month lead time (i.e. forecasts initialised in February and May respectively) over south-eastern Australia.

Figure 16 As for Figure 14, but for Tmin.



One of the key differences between the two prediction systems is in the prediction of temperature, particularly Tmax, over eastern Australia in late autumn (MJJ) (Figure 17 bottom row) and winter (JJA) (not shown). ACCESS-S1 shows markedly degraded performance compared to POAMA. This manifests in the Australia-wide results in the reliability error and resolution scores where ACCESS-S1 is worse than POAMA (Figure 15, 16). This poor performance is largely related to using climatological land surface initial conditions (soil moisture) to initialise the forecasts, rather than realistic initial conditions. This was demonstrated by running a sensitivity experiment to determine the impact on forecast skill of initialising using time varying (realistic) soil moisture (Zhao et al. 2017). As described in Zhao et al. (2017), the stand-alone version of JULES (the same as that used in the coupled ACCESS-S1 model) was run forced by bias corrected three-hourly ERA-interim data (Dee et al. 2011) (2m temperature, 2m specific humidity, downwards longwave radiation, downwards shortwave radiation, surface pressure, total precipitation, 10m wind speed) to obtain a land reanalysis (1990-2014, with a ten-year spin-up period starting in 1980). The time varying fields of soil moisture from this land reanalysis were used to initialise the forecasts of the sensitivity experiment, which included an eleven-member ensemble on the 1st May for the period 1990-2012. This experiment was compared to the ACCESS-S1 forecasts for the same period and start dates. Results showed that when the land surface is initialised with realistic initial conditions, rather than climatology, there are significant improvements in the prediction skill on multi-week to seasonal timescales over Australia of Tmax, with moderate improvements to Tmin and precipitation (Zhao et al. 2017). For example, Figure 18 shows that the skill for Tmax forecasts for May (the first month of the forecast) are significantly improved, and better than POAMA, when the land surface is initialised with realistic initial conditions and that the largest impact is seen over north-eastern Australia. For the next version of ACCESS-S (version 2), the land surface will be initialised with realistic initial conditions rather than climatology using the BoM's data assimilation scheme. The experiment described above suggests that this will significantly improve the skill of Tmax forecasts in the winter half of the year, as well as Tmin and precipitation but to a lesser degree (Zhao et al. 2017).

Figure 17 Proportion correct of 0-month lead time seasonal forecasts of above/below median for (top) rainfall in ASO (August-October) and (bottom) Tmax in MJJ (May-July) from ACCESS-S1 (left) and POAMA (right). Proportion correct of the climatological forecast is 0.5, thus accuracy greater than 0.55, shown with colour shading, is considered to be skilful. Observations are the AWAP (Jones et al. 2009) analyses.

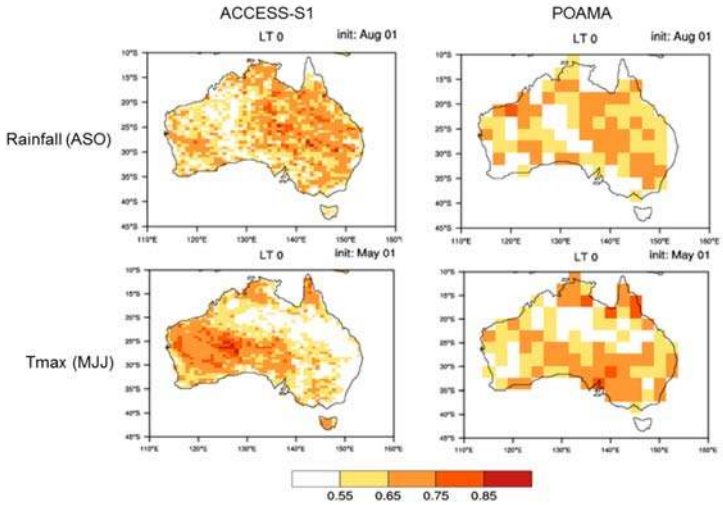
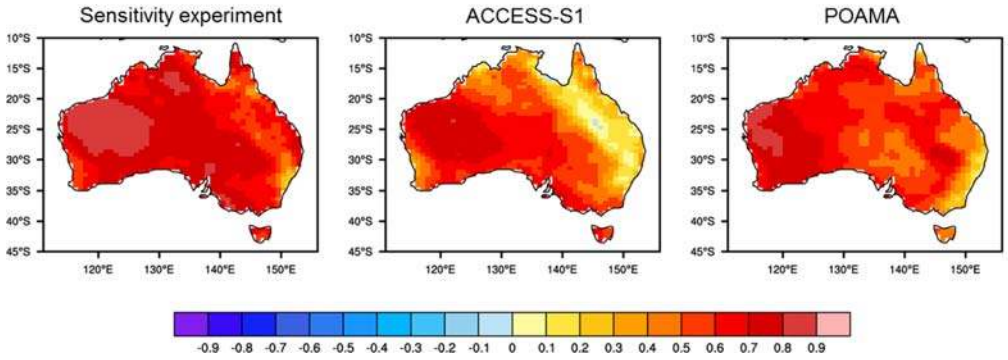


Figure 18 Correlation skill for Tmax for 0-month lead forecasts of the month of May (1990-2012) from the sensitivity experiment where the forecasts are initialised with realistic soil moisture conditions (left), from ACCESS-S1 where the forecasts are initialised using climatological soil moisture (middle), and from POAMA (right).



4.3.2 *Multi-week*

The correlation skill for forecasts of fortnightly mean rainfall, Tmax and Tmin anomalies for fortnight one (week one and two), fortnight two (weeks two and three) and fortnight three (weeks three and four) of the forecast are shown in Figures 19-21, using all available forecast start dates. For ACCESS-S1 there are statistically significant positive correlations Australia-wide for Tmax and Tmin at all three lead times; and for rainfall for the first two fortnights (except for isolated small regions in fortnight two). The correlation skill for rainfall for fortnight three (weeks three and four) of the forecast is not statistically significant over large parts of eastern Australia (Figure 19). Multi-week skill for Tmax and Tmin is generally higher than that for rainfall at all lead times. There is a clear reduction in skill with forecast lead time, i.e. going from fortnight one to fortnight three.

ACCESS-S1 is a clear improvement over POAMA for all variables for fortnights one and two (Figures 19-21). It is also better over most regions for fortnight three for Tmax and Tmin. Superior atmospheric initial conditions (together with improved physics and dynamics) in ACCESS-S1 are likely playing a large role in the improved multi-week forecast skill. ACCESS-S1 uses ERA-Interim atmospheric initial conditions (Dee et al. 2011) whereas POAMA uses degraded ERA-40 (Uppala et al. 2005) initial conditions (degraded due to the nudging). The better atmospheric initial conditions in AC-

CESS-S1 compared to POAMA outweighs the negative effect of using climatological soil moisture initial conditions (Zhao et al. 2017). However, the land surface experiments described previously indicate that the multi-week skill of Tmax and Tmin can be expected to increase even further in ACCESS-S, particularly from week four onwards, if realistic land surface initial conditions are used (Zhao et al. 2017).

Figure 19 Correlation skill for rainfall for fortnight one (weeks one and two) (left column), fortnight three (weeks two and three) (middle column) and fortnight three (weeks three and four) (right column) from ACCESS-S1 (top) and POAMA (bottom) using all of the eight forecast start dates noted in section 2.3 (eleven-member ensemble). The dashed line indicates the correlation value above which the correlations are statistically significant (i.e. r>0.14 is significant, 95 per cent confidence level, n=184). If there is no dashed line shown on a map, then all the correlations are significant.

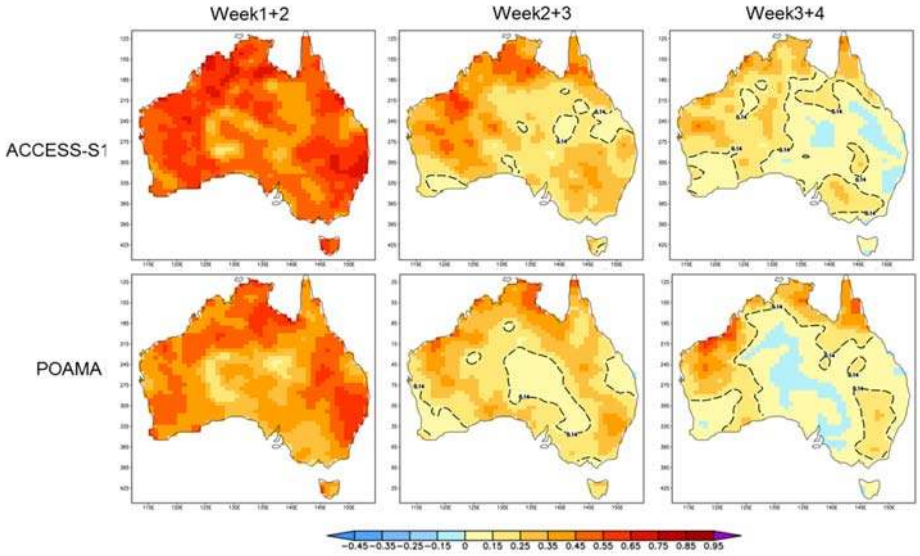


Figure 20 As for Figure 19, but for Tmax.

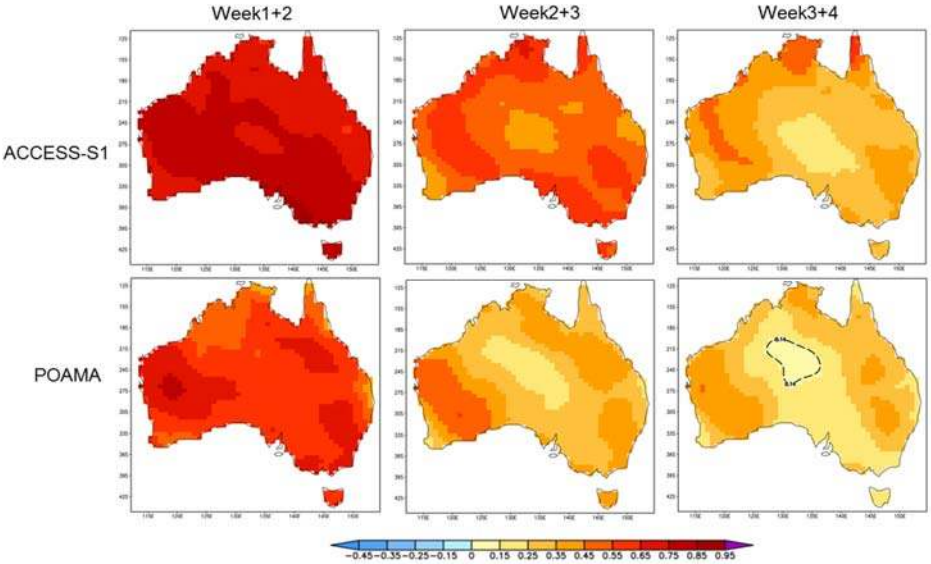
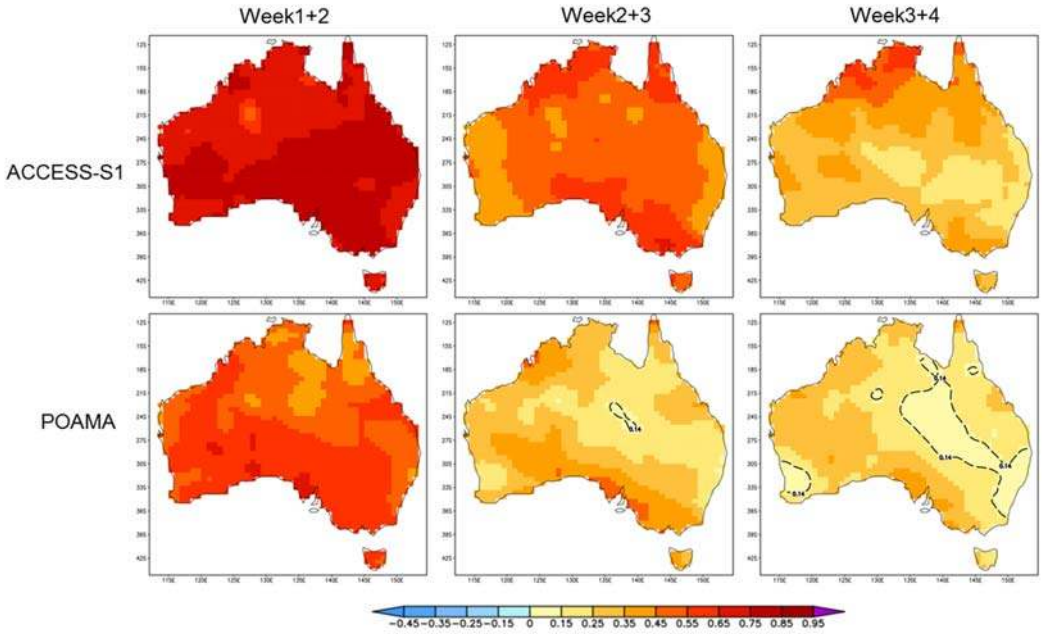


Figure 21 As for Figure 19, but for Tmin.



5 Conclusions and future developments

ACCESS-S1 will be the next version of the Bureau of Meteorology's seasonal prediction system, due to become operational in early 2018. In this paper we have evaluated the performance of ACCESS-S1 and compared it to the current operational system, POAMA. ACCESS-S1 will have considerable enhancements compared to POAMA, including higher resolution of the component models and state-of-the-art physics parameterisation schemes. ACCESS-S1 is based on the Met Office GloSea5-GC2 seasonal prediction system, but has local enhancements to the ensemble generation strategy to make it appropriate for multi-week forecasting and a larger ensemble size.

The evaluation is based on eleven-member ensemble forecasts for 1990-2012. Seasonal forecast performance is based on a twenty-two-member time-lag ensemble, created by combining forecasts initialised on the 1st of the month with the 25th of the month prior. Keeping in mind some degree of uncertainty stemming from the relatively small ensemble size and the short hindcast period for the comparison of the two systems, evaluation of the ACCESS-S1 forecasts and comparison of its skill to that from POAMA is summarised by:

- reduced mean climate bias;
- improved prediction of the early stages of the development of ENSO (through the boreal spring predictability barrier) and the IOD;
- improved distinction between the SST patterns of eastern and central Pacific El Niños;
- improved prediction of the daily SAM index, by about four days;
- improved prediction of the 0- and 1-month lead seasonal SAM;
- improved prediction of the MJO index, by about five days, especially in boreal summer;
- improved eastward propagation of the MJO-related rainfall anomalies in the boreal summer;
- improved spatial detail in seasonal and multi-week forecasts for Australia;
- improved seasonal forecasts over Australia for late winter (August to October at 0-month lead) and spring (September to November at one-month lead), with particularly good forecast reliability for rainfall and Tmax. Forecast accuracy for rainfall is improved over south-eastern, eastern-coastal and south-western Australia; and
- overall improved prediction of multi-week Australian climate, particularly for fortnights one (weeks one and two) and two (weeks two and three).

However, there is still a large scope for improvement of ACCESS-S1. Some of the shortfalls of ACCESS-S1 diagnosed in this study, in Lim et al. (2016a) and Hudson et al. (2017) include:

- rainfall and SST biases in the Indian Ocean and Maritime Continent region;
- overestimation of strengths of the ENSO and the IOD;
- indications that the ensemble is still under-dispersed at short lead times, particularly in the tropics, as diagnosed with the MJO index;
- inhibited eastward propagation of MJO-related rainfall anomalies into the western Pacific in the austral summer half of the year;
- weak/deficient teleconnections between Australian rainfall over eastern and south-eastern Australia and ENSO, the IOD and the SAM respectively, particularly in the winter half of the year; and
- initialisation of soil moisture with climatology, rather than realistic soil initial conditions that vary year-to-year. Initialising using climatology impacts negatively on the skill of the multi-week (from week four) and seasonal forecasts, particularly for Tmax over eastern Australia and to a lesser degree on Tmin over south-eastern Australia for forecasts in late autumn and winter.

A caveat of these results is the relatively small sample size in hindcast years (n=23) used for the evaluation of forecast performance. A period of twenty three years is typically too short to obtain statistically reliable results, particularly for aspects of the prediction system that are inherently noisy, such as for forecasts of relatively small regions of Australian climate (in contrast to less noisy, more large-scale aspects like ENSO). In addition, a sufficiently long hindcast set is necessary for including an adequate number of cases of the low frequency influences on Australian climate, like the different phases of ENSO, and for knowing how the skill may vary based on the state of these climate drivers. The size of the hindcast for ACCESS-S1 is limited due to the availability and reliance on the Met Office's initial conditions. However, for ACCESS-S2 we plan to create a longer hindcast set, spanning at least 30 years.

Development of the next version of ACCESS-S (version 2) is already underway. ACCESS-S2 will include the BoM's ensemble generation and data assimilation system (e.g., Hudson et al. 2013) and will not be reliant on the Met Office initialisation strategy. The land surface will be initialised with realistic initial conditions and optimal perturbations will be applied to the atmosphere and ocean.

Acknowledgments

Support for this work was provided by the Australian Government through the Agricultural Competitiveness White Paper, which established a roadmap of practical actions to grow the agriculture sector in Australia, as well as by the Managing Climate Variability Program (MCV00041) and the Victorian Climate Initiative. The work was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. Claire Spillman, Felicity Gamble and two anonymous reviewers are thanked for their useful comments in the preparation of this manuscript.

References

Ashok, K., Behera, S.K., Rao, S.A. and Weng, H. 2007. El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research*, 112, C1107, doi:10.1029/2006JC003 798.

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B. and Harding, R.J. 2011. The Joint UK Land Environment Simulator (JULES), model description–Part 1: Energy and water fluxes. *Geoscientific Model Development,* 4, 677–699, doi: 10.5194/gmd-4-677-2011.

Blockley, E.W., Martin, M.J., McLaren, A.J., Ryan, A.G., Waters, J., Lea, D.J., Mirouze, I., Peterson, K.A., Sellar, A. and Storkey, D. 2014. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, *Geoscientific Model Development*, 7, 2613–2638, doi:10.5194/gmd-7-2613-2014.

Bowler, N., Arribas, A., Beare, S., Mylne, K.E. and Shutts, G. 2009. The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 135, 767–776

Brown, A.R., Milton, S., Cullen, M., Golding, B., Mitchell, J. and Shelly, A. 2012. Unified modelling and prediction of weather and climate: a 25 year journey, *Bulletin of the American Meteorological Society*, 93, 1865–1877, doi:10.1175/BAMS-D-12-00018.1

Colman, R., Deschamps, L., Naughton, M., Rikus, L., Sulaiman, A., Puri, K., Roff, G., Sun, Z. and Embery, G. 2005. BMRC Atmospheric Model (BAM) version 3.0: Comparison with mean climatology. BMRC Research Report No. 108, Bureau of Meteorology Research Centre, 32 pp.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597.

Graham, T. 2014. The importance of eddy-permitting model resolution for simulation of the heat budget of tropical instability waves. *Ocean Modelling*, 79: 21–32, doi: 10.1016/j.ocemod.2014.04.005.

Hendon, H.H., Lim, E., Wang, G., Alves, O. and Hudson, D. 2009. Prospects for Predicting Two Flavours of El Nino. *Geophysical Research Letters*, doi:10.1029/2009GLO40100.

Hudson, D., Alves, O., Hendon, H.H. and Wang, G., 2011. The impact of atmospheric initialisation on seasonal prediction of tropical Pacific SST. *Climate Dynamics*, 36, 1155-1171.

Hudson, D., Marshall, A.G., Yin, Y., Alves, O. and Hendon, H.H. 2013. Improving intraseasonal prediction with a new ensemble generation strategy. *Monthly Weather Review*, 141, 4429-4449.

Hudson, D., Shi, L., Alves, O., Zhao, M., Hendon, H.H. and Young, G. 2017. Performance of ACCESS-S1 for key horticultural regions. Bureau Research Report No. 20, Bureau of Meteorology, Australia (available from: http://www.bom.gov.au/research/research-reports.shtml).

Hunke, E.C. and Lipscomb, W.H. 2010. CICE: The Los Alamos sea ice model documentation and software users' manual, Version 4.1, LA-CC-06-012, Los Alamos National Laboratory, N.M.

Jones, D.A., Wang, W. and Fawcett R., 2009. High–quality spatial climate data–sets for Australia. *Australian Meteorological and Oceanographic Journal*, 58, 233–248.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society,* 77,437–471.

Langford, S., Hendon, H.H. and Lim. E-P. 2011. Assessment of POAMA's predictions of some climate indices for use as predictors of Australian rainfall. CAWCR Technical Report, No. 031, Bureau of Meteorology, Australia (available from: http://www.cawcr.gov.au/publications/).

Langford, S. and Hendon, H. 2013. Improving Reliability of Coupled Model Forecasts of Australian Seasonal Rainfall. *Monthly Weather Review*, 141, 728–741, doi: 10.1175/MWR-D-11-00333.1.

Lim, E.-P., Hendon, H.H., Hudson, D., Wang, G. and Alves, O. 2009. Dynamical forecasts of inter-El Niño variations of tropical SST and Australian springtime rainfall. *Monthly Weather Review*, 137,3796-3810.

Lim, E-P., Hendon, H.H., Langford, S. and Alves, O. 2012. Improvements in POAMA2 for the prediction of major climate drivers and south eastern Australian rainfall. CAWCR Technical Report No. 051, Bureau of Meteorology, Australia.

Lim, E.-P., Hendon, H.H. and Rashid, H.A. 2013. Seasonal predictability of the Southern Annular Mode due to its association with ENSO. *Journal of Climate*, 26, 8037-8054.

Lim, E.-P, Hendon, H.H., Hudson, D., Zhao, M., Shi, L., Alves, O. and Young G. 2016a. Evaluation of the ACCESS-S1 hindcasts for prediction of Victorian seasonal rainfall. Bureau Research Report No. 19, Bureau of Meteorology Australia (available from: http://www.bom.gov.au/research/research-reports.shtml).

Lim, E.P., Hendon, H.H., Zhao, M. and Yin, Y. 2016b. Inter-decadal variations in the linkages between ENSO, the IOD and south-eastern Australian springtime rainfall in the past 30 years. *Climate Dynamics*, doi: 10.1007/s00382-016-3328-8.

Lim, E.P., Hendon, H.H., Arblaster, J.M., Chung, C., Moise, A.F., Hope, P., Young, G. and Zhao, M. 2016c. Interaction of the recent 50 year SST trend and La Niña 2010: amplification of the Southern Annular Mode and Australian springtime rainfall. *Climate Dynamics*, 47(7-8), 2273-2291.

MacLachlan, C., Arribas, A., Peterson, K.A., Maidens, A., Fereday, D., Scaife, A.A., Gordon, M., Vellinga, M., Williams, A., Comer, R.E., Camp, J., Xavier, P. and Madec, G. 2015. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, *Quarterly Journal of the Royal Meteorological Society*, 141, 1072-1084 doi:10.1002/qj.2396

Madden, R.A. and Julian, P.R. 1971. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, 28, 702–708

Madec, G., and co-authors, 2008. NEMO ocean engine. Technical Report Note du Pole de od´elisation No 27, ISSN No 1288-1619, Institut Pierre-Simon Laplace (IPSL), France.

Magnusson, L., Nycander, J. and Källén, E. 2009. Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus A*, 61: 194–209. doi:10.1111/j.1600-0870.2008.00385.x

Manabe, S. and Holloway, J. 1975: The seasonal variation of the hydrological cycle as simulated by a global model of the atmosphere. *Journal of Geophysical Research*, 80, 1617-1649.

Marshall, A.G., Hudson, D., Wheeler, M.C., Hendon, H.H. and Alves, O. 2011. Assessing the Simulation and Prediction of Rainfall Associated with the MJO in the POAMA Seasonal Forecast System. *Climate Dynamics*, 37, 2129-2141, doi:10.1007/s00382-010-0948-2

Marshall, A.G., Hudson, D., Wheeler, M.C., Hendon, H.H. and Alves, O. 2012. Simulation and prediction of the Southern Annular Mode and its influence on Australian intra-seasonal climate in POAMA. *Climate Dynamics*, 38, 2483-2502, doi:10.1007/s00382-011-1140-z.

Marshall, A.G., Hudson, D., Wheeler, M., Alves,O., Hendon, H.H., Pook, M.J. and Risbey, J.S. 2014. Intra-seasonal drivers of extreme heat over Australia in observations and POAMA-2. *Climate Dynamics*. 43, 1915-1937

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J. and Sinha, B. 2014. GO5.0: the joint NERC–Met Office NEMO global ocean model for use in coupled and forced applications, *Geoscientific Model Development*, 7, 1069–1092, doi:10.5194/gmd-7-1069-2014

Mogensen K., Balmaseda M. and Weaver A.T., Martin M., Vidard A. 2009. NEMOVAR: A variational data assimilation system for the NEMO ocean model. In ECMWF Newsletter, Walter Z. (ed.) 120: 17–21. ECMWF, Reading, UK.

Mogensen K., Balmaseda M.A. and Weaver A.T. 2012. 'The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4', Technical Report TR-CMGC-12-30. CERFACS: Toulouse, France.

Murphy, A.H. 1973. A new vector partition of the probability score. *Journal of Applied Meteorology*, 12, 595–600

Oke, P.R., Schiller, A., Griffin, D.A. and Brassington, G.B. 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quarterly Journal of the Royal Meteorological Society*, 131, 3301-3311.

Palmer, T., Buizza, R., Hagedorn, R., Lawrence, A., Leutbecher, M. and Smith, L. 2006. Ensemble prediction: A pedagogical perspective. ECMWF Newsletter, No. 106, ECMWF, Reading, United Kingdom, 11–17.

Rae, J.G.L., Hewitt, H.T., Keen, A.B., Ridley, J.K., West, A.E., Harris, C.M., Hunke, E.C. and Walters, D.N. 2015. Development of the Global Sea Ice 6.0 CICE configuration for the Met Office Global Coupled Model. *Geoscientific Model Development*, 8, 2221–2230

Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C. and Payne, T.J. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133, 347–362. doi:10.1002/qj.32

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes D.C. and Wang, W. 2002. An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15, 1609-1625.

Risbey J.S., Pook M.J., McIntosh P.C., Wheeler M.C. and Hendon H.H. 2009. On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, 137, 3233–3253.

Saji, N.H., Goswami, B.N., Vinayachandran, P.N., Yamagata, T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* 401, 360–363.

Schiller, A., Godfrey, J.S., McIntosh, P. and Meyers, G. 1997. A global ocean general circulation model climate variability studies. CSIRO Marine Research Report No. 227, 60 pp.

Schiller, A., Godfrey, J.S., McIntosh, P.C., Meyers, G., Smith, N.R., Alves, O., Wang, G. and Fiedler, R. 2002. A New Version of the Australian Community Ocean Model for Seasonal Climate Prediction. CSIRO Marine Research Report No. 240.

Seviour W.J.M, Hardiman S.C., Gray L.J., Butchart N., MacLachlan C. and Scaife A.A. 2014. Seasonal prediction of the Southern Annular Mode and Antarctic ozone. *Journal of Climate*, 27, 7462–7474

Shi, L., Hendon, H.H., Alves, O., Luo, J–J., Balmaseda, M. and Anderson, D. 2012. How predictable is the Indian Ocean Dipole? *Monthly Weather Review*, 140, 3867–3884.

Sullivan, A., Luo, J.J., Hirst, A.C., Bi, D., Cai, W. and He, J. 2016. Robust contribution of decadal anomalies to the frequency of central-Pacific El Niño. *Scientific Reports*, 6, 38540.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at the NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, 74, 2317–2330.

Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Berg, L.V.D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.

P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961-3012.

Valcke, S., Terray, L. and Piacentini, A. 2000. OASIS 2.4 Ocean Atmospheric Sea Ice Soil users guide, Version 2.4. CERFACS Tech. Rep, CERFACS TR/CMGC/00-10, 85pp.

Valcke, S. 2013. The OASIS3 coupler: A European climate modelling community software, *Geoscientific Model Development*, 6, 373–388, doi:10.5194/gmd-6-373-2013

Walters, D., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Vidale, P.L., Wilkinson, J., Willett, M., Woolnough, S. and Xavier, P. 2017. The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, 10, 1487-1520, doi:10.5194/gmd-10-1487-2017.

Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A. and While, J. 2015. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141: 333–349. doi:10.1002/qj.2388

Wang, G. and Hendon, H.H. 2007. Sensitivity of Australian rainfall to inter-El Niño variations. *Journal of Climate*, 20, 4211-4226.

Wang, G., Hudson, D., Yin, Y., Alves, O., Hendon, H., Langford, S., Liu, G. and Tseitkin, F. 2011. POAMA-2 SST skill assessment and beyond. CAWCR Research Letters, No. 6, 40–46, Bureau of Meteorology, Australia.

Weedon, G.P., Gomes, S., Viterbo, P., Shuttleworth, W.J., Blyth, E., Österle, H., Adam, J.C., Bellouin, N., Boucher, O. and Best, M. 2011. Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, 12, 823–848

Weisheimer, A., Palmer, T.N. and Doblas-Reyes, F.J. 2011. Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters*, 38, L16703, doi:10.1029/2011GL048123.

Wheeler, M.C. and Hendon, H.H. 2004. An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, 132, 1917–1932.

Wheeler M.C., Hendon H.H., Cleland S., Meinke H. and Donald A. 2009. Impacts of the Madden–Julian oscillation on Australian rainfall and circulation. *Journal of Climate*, 22, 1482–1498.

Wilks, D. 2006. Statistical Methods in Atmospheric Sciences. 2nd ed. Academic Press, 627 pp.

Williams, K. D., Harris, C.M., Bodas-Salcedo, A., Camp, J., Comer, R.E., Copsey, D., Fereday, D., Graham, T., Hill, R., Hinton, T., Hyder, P., Ineson, S., Masato, G., Milton, S. F., Roberts, M.J., Rowell, D. P., Sanchez, C., Shelly, A., Sinha, B., Walters, D.N., West, A., Woollings, T. and Xavier, P.K. 2015. The Met Office Global Coupled model 2.0 (GC2) configuration, *Geoscientific Model Development*, 8, 1509-1524, doi:10.5194/gmd-8-1509-2015.

Yin, Y., Alves, O. and Oke, P.R. 2011. An ensemble ocean data assimilation system for seasonal prediction. *Monthly Weather Review*, 139, 786-808.

Xie P. and Arkin P.A. 1997. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society,* 78, 2539–2558

Zhao, M. and Hendon, H.H. 2009. Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. *Quarterly Journal of the Royal Meteorological Society,* 135, 337–352. doi:10.1002/qj.370

Zhao, M., Hendon, H.H., Alves, O., Liu, G. and Wang, G. 2016. Weakened Eastern Pacific El Niño Predictability in the Early Twenty-First Century. *Journal of Climate*, DOI: http://dx.doi.org/10.1175/JCLI-D-15-0876.1

Zhao, M., Zhang, H-Q. and Dharssi, I. 2017. Impact of Land-surface Initialization on ACCESS-S1 and Comparison with POAMA. Bureau Research Report, No. 23., Bureau of Meteorology, Australia (available from: http://www.bom.gov.au/research/research-reports.shtml)

Zhou, X., Luo, J., Alves, O. and Hendon, H. 2015. Comparison of GLOSEA5 and POAMA2.4 Hindcasts 1996-2009: Ocean Focus. Bureau Research Report No. 10, Bureau of Meteorology, Australia (available from: http://www.bom.gov.au/research/research-reports.shtml)