Australian Government
Bureau of Meteorology

The Centre for Australian Weather and Climate Research
A partnership between CSIRO and the Bureau of Meteorology

CSIRO

CAWCR Research Letters

Issue 9, December 2012

P. A. Sandery, T. Leeuwenburg, G. Wang, K. A. Day (editors)

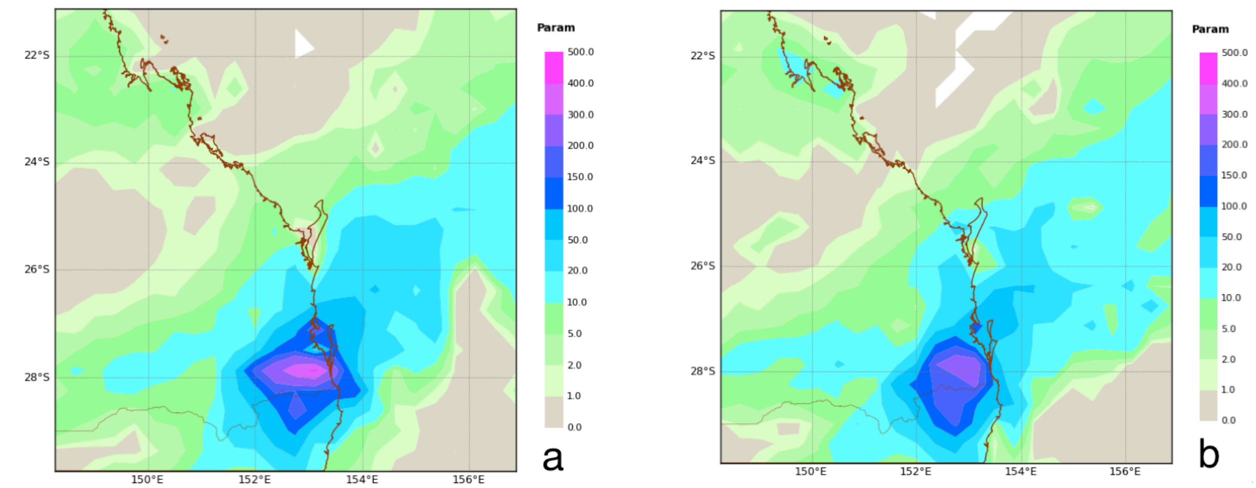
www.cawcr.gov.au

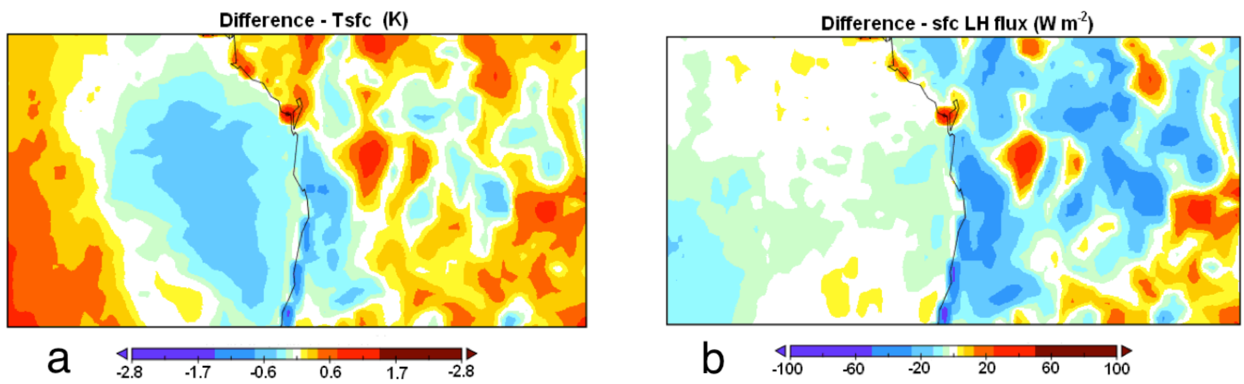
ISSN: 1836-5949

Series: Research Letters (The Centre for Australian Weather and Climate Research); Issue 9.

Copyright and Disclaimer

Contents





Editors: P.A. Sandery, T. Leeuwenburg, G. Wang, K.A. Day
Enquiries: Dr Paul Sandery p.sandery@bom.gov.au
CAWCR Research Letters
The Centre for Australian Weather and Climate Research
Bureau of Meteorology
GPO Box 1298 Melbourne VICTORIA 3001

CAWCR Research Letters is an internal serial online publication aimed at communication of research carried out by CAWCR staff and their colleagues. It follows on from its predecessor *BMRC Research Letters.* Articles in *CAWCR Research Letters* are peer reviewed. For more information visit the *CAWCR* website.

Forecasting the Brisbane flooding event using Ensemble Bred Vector SST initialization and ocean coupling in ACCESS NWP

Vaughan Barras and Paul Sandery

Centre for Australian Weather and Climate Research

700 Collins Street, Docklands, VICTORIA, 3008

V.Barras@bom.gov.au

Introduction

The provision of accurate guidance regarding severe weather events to forecasters is of key importance to the operations of the Bureau of Meteorology ACCESS Numerical Weather Prediction (NWP) system. The predictability of the flooding of southeastern Queensland in early January 2011 is a prime example of this. Forecasts from the ACCESS-G (global) model provided timely warnings that a significant rainfall event was to impact the region at a 5-day lead-time. As it came within the forecast range of the higher resolution 12 km ACCESS-R limited area model, precipitation forecasts were for rainfall totals that exceeded previous daily records at some locations within the Brisbane River catchment. In assessing the overall performance of 12 km ACCESS-R, it is of interest to (i) verify how well the model predicted rainfall for this event, and (ii) given the magnitude of the forecast precipitation amount understand how sensitive these predictions from 12 km ACCESS-R might be to the surface boundary conditions, in particular, that of sea surface temperature (SST). Currently, the operational NWP models use relatively coarse resolution representations of SST to determine fluxes of temperature and moisture into the boundary layer. In a coupled NWP framework it is possible to create an improved SST initial condition as well as incorporate the effects of evolving ocean surface temperatures. Here, we assess the performance of 12 km ACCESS-R for the Brisbane flooding event in both coupled and un-coupled NWP configurations and assess the sensitivity of precipitation forecasts to an alternative representation of SST.

Brisbane flooding event January 2011

Queensland's wettest December on record

The months leading up to January 2011 were extremely wet throughout eastern Australia. This was largely due to the influence of a strong La Nina (Southern Oscillation Index (SOI) = +27.1). Queensland recorded its wettest December on record (National Climate Centre, 2011) with persistent moist easterly airflow associated with anomalously high sea surface temperatures off the northern Australian coastline (Evans and Boyer-Souchet 2012). The landfall of Tropical Cyclone Tasha to the south of Cairns early on Christmas Day brought further heavy rains with daily rainfall totals along the central coast in excess of 200 mm. After its wettest recorded spring season, persistent heavy rains over eastern Queensland resulted in near-saturation of its major water catchments and elevated the risk of major flooding (Figure 1).

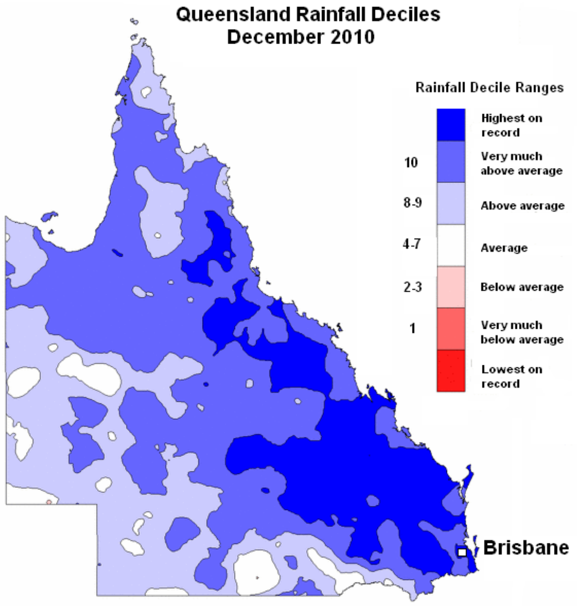


Figure 1: *Queensland rainfall deciles for December 2010. Much of the south and east of the state recorded its highest monthly rainfall on record.*

Flooding of SE Queensland

Early January witnessed a period of very heavy rainfall over southeast Queensland and northern New South Wales under the influence of a well-developed upper level trough. The most intense rains fell over the southern Queensland coast from January 8–12. In the region bounded by Gympie, Brisbane and Toowoomba (covering much of the Brisbane River catchment) the total accumulated rainfall was well in excess of 200 mm. Highest daily rainfall totals for January were recorded on the 10th (Figure 2) at Peachester (298.0 mm) and Lindfield (257.0 mm[1]). In the days following this period there

[1] This was a record for all months at this location. Previous maximum daily rainfall at Lindfield was 237.8 mm recorded on February 9, 1999 (NCC).

was major flooding of the Lockyer and Bremer river catchments as well as the region around Toowoomba resulting in severe property damage and the tragic loss of life. On January 13 the Brisbane River city gauge peaked at 4.46 m, its highest mark since the flood of January 1974[2].

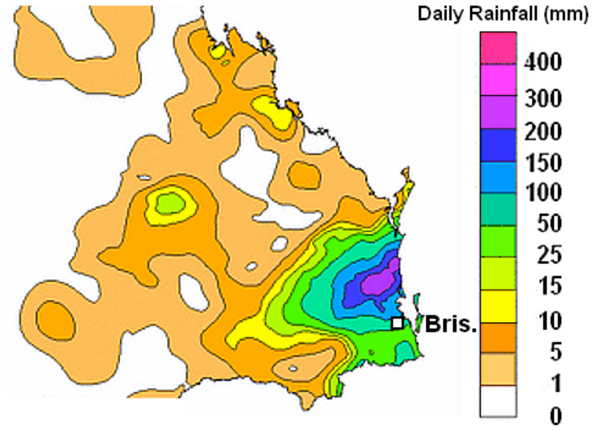


Figure 2: *SE Queensland accumulated rainfall totals for 10th January 2011.*

Validation of 12 km ACCESS-R precipitation

Six days prior to the event, the ACCESS global NWP model (ACCESS-G), predicted rainfall across the coastal region of southeast Queensland. The next day the model predicted that this would be a significant rain event. At three days lead-time, the event fell within the forecast range of the limited area models (12 km ACCESS-R was running in a research trial in parallel with the operational 12km Australian domain ACCESS-A model at the time). From the January 8 basedate, the 12 km ACCESS-R 48-72 hour accumulated precipitation forecast was for maxima in excess of 400 mm over Brisbane and surrounds. As indicated in section 2.2, this was well in excess of record daily rainfall totals measured for the region. When compared to observations taken during the event, the forecast location of maximum rainfall was predicted well by 12km ACCESS-R but was slightly too far south. Whilst area average rainfall amounts were also quite well predicted at this lead-time, the peak accumulated rainfall amounts over the Brisbane River catchment were over-predicted considerably by between 100-200 mm. In terms of location, the forecast was better than ACCESS-G, which tended to hold precipitation off-shore. Other high resolution models (12 km ACCESS-A, and the 5 km model for the Brisbane region, ACCESS-BN) did not forecast out to 72 hours. At their maximum lead-times (48 hours and 36 hours respectively) they forecast the location of the rainfall more accurately to that of 12 km ACCESS-R, however both models still overestimated accumulated rainfall amounts.

[2] The peak height measured at the Brisbane River city gauge was 5.45m on January 28-29, 1974.

Revised initialization of SST

Use of Foundation SST in NWP

The over-prediction of rainfall by 12km ACCESS-R for this event suggests either an error in the model physics, or a bias in the external forcing or the initial conditions. With regard to the latter, it is known that NWP uses a single 'foundation' SST field that is retained throughout the forecast period. The Group for High-Resolution Sea Surface Temperature (GHRSST) defines foundation SST as the temperature at the first time of day when the heat gain from solar absorption exceeds that of heat loss at the sea surface (Minnett, 2011). This temperature is, in effect, the nocturnal baseline value from which the diurnal temperature variability of the upper ocean layers begins. Typically it represents the water temperature at a depth of 10 metres.

Ensemble bred vector SST initialization

ACCESS-RC, the coupled limited area model (CLAM) version of 12 km ACCESS-R initializes within the OceanMAPS system (Brassington *et al.*, 2007) and uses an ensemble bred vector (EBV) initialization approach for its initial state (Sandery *et al.*, 2012). This involves the cyclic 'breeding' of an ensemble of the fastest growing coupled atmosphere-ocean modes which can be used as a non-linear filter and to correct the model attractor state (Toth and Kalnay 1997). The cyclic bred vectors generated by the ensemble approach are non-linear generalisations of finite-time normal modes (Frederiksen *et al.* 2010). In our approach we rescale the ocean only in the coupled system so that ensemble variance in the atmosphere measures the sensitivity of the atmospheric model to perturbations of SST. The process begins with the addition of an initial random perturbation of the full ocean temperature field to a control forecast. After 24 hours the difference between this perturbed model run and the control are rescaled to match the RMS amplitude of the initial perturbation. The norm used for rescaling is the RMS anomaly of thermocline temperatures, as in O'Kane *et al.* (2011). The rescaled perturbations are added to the control initial condition and the forecast is run again, and so on. The EBV initial condition for the final forecast is the ensemble mean of this series of rescaled 24 hour bred vectors added to the control initial state.

Validation of EBV fields

Initial conditions for the foundation and EBV SSTs were evaluated against 'super-observations' taken from the Bluelink Ocean Data Assimilation System (BODAS) based on AMSR-E microwave SST (Oke *et al.*, 2008). The term super-observation refers to up-scaled observations designed to reduce representation error with respect to the model horizontal grid resolution (Oke and Sakov, 2008).

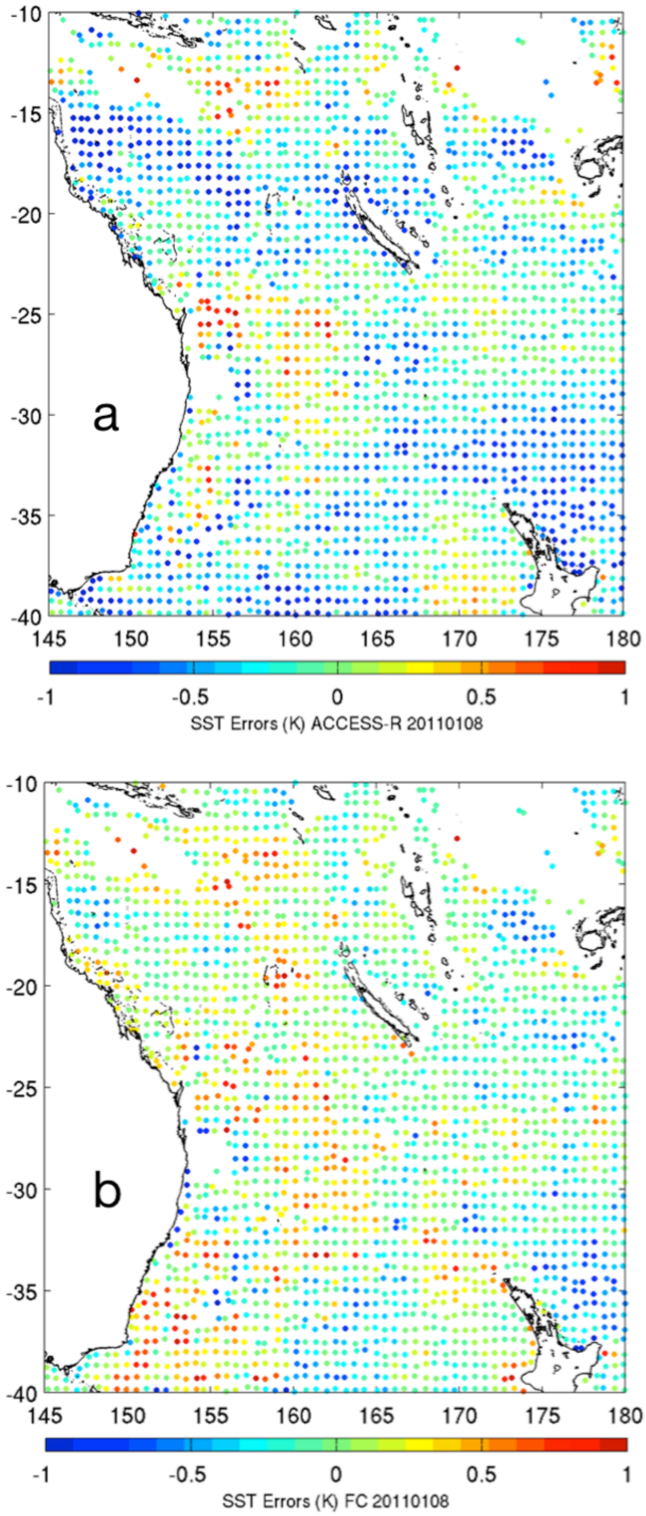


Figure 3: *(a) Difference between foundation SST for 12 km ACCESS-R and AMSR-E SST super-observations from BODAS 24 hour time-average surface temperature for the 48-72h period of the 00 forecast with basedate 8th January 2011. (b) Same as for (a), but for ACCESS-RC using EBV SST initialization.*

Figures 3a and 3b show that during the 48-72 hour forecast period, the mean error for SST is lower in ACCESS-RC when compared to the 12 km ACCESS-R. The standard deviation of these errors is also lower (0.34 and 0.41 respectively). Figure 4 shows a selection of time-averaged ensemble mean bred vectors for the first 24 hours of the ACCESS-RC forecast made on the 6th of January 2011 (portraying the differences between bred perturbations and a control run) for surface temperature, latent heat flux, mean sea level pressure (MSLP), and accumulated precipitation. Each panel illustrates how uncertainty in SST projects into each particular field in the atmospheric model in terms of a dynamic tendency. For example, the mean bred vector for SST is generally within ±0.5 K, and correlates with perturbations in surface latent heat flux, which are generally within ±15 Wm^{-2}. This indicates that the fastest growing modes of uncertainty in the initial state tend to cool the ocean surface temperatures off the southeast Queensland coast in comparison to the control. Similarly, the dynamical tendency of these mean bred vectors is consistent with a small positive MSLP perturbation over the region. Patterns in the mean EBV for accumulated precipitation are less clear given the complexities of the response to perturbations in latent heat flux and other interactions within the model parameterizations. The plots in Figure 4 also show that coastal SSTs near Brisbane tended to be cooler in ACCESS-RC leading to a localized reduction of latent heat flux, stabilization of the atmosphere and an attenuation of the easterly surface winds (not shown). This resulted in an overall reduction in the moisture flux to the atmospheric boundary layer. Comparisons of modelled screen height specific humidity indicated that ACCESS-RC had a drier PBL.

Forecasting the flooding event

Impact of coupled SST upon forecast rainfall

Noting the reduced bias of the EBV-derived SST initial condition and the dynamical tendencies indicated by the ensemble mean bred vectors, it was of interest to test the model sensitivity to these changes to the lower boundary condition. ACCESS-RC was run first in control mode to replicate the 12 km ACCESS-R forecasts. The model ran with foundation SSTs as well as physics settings and initial conditions the same as that of the operational forecast model. An identical model run was then performed with the only changes being the use of the EBV SST initial conditions and a time evolving ocean surface coupled to OceanMAPS every time-step.

A number of forecasts were run for the event using initializations at lead-times increasing by 24 hourly increments. We compared the difference in surface conditions between the two model configurations and analysed the impact of the change upon the 72-hour forecasts of accumulated precipitation.

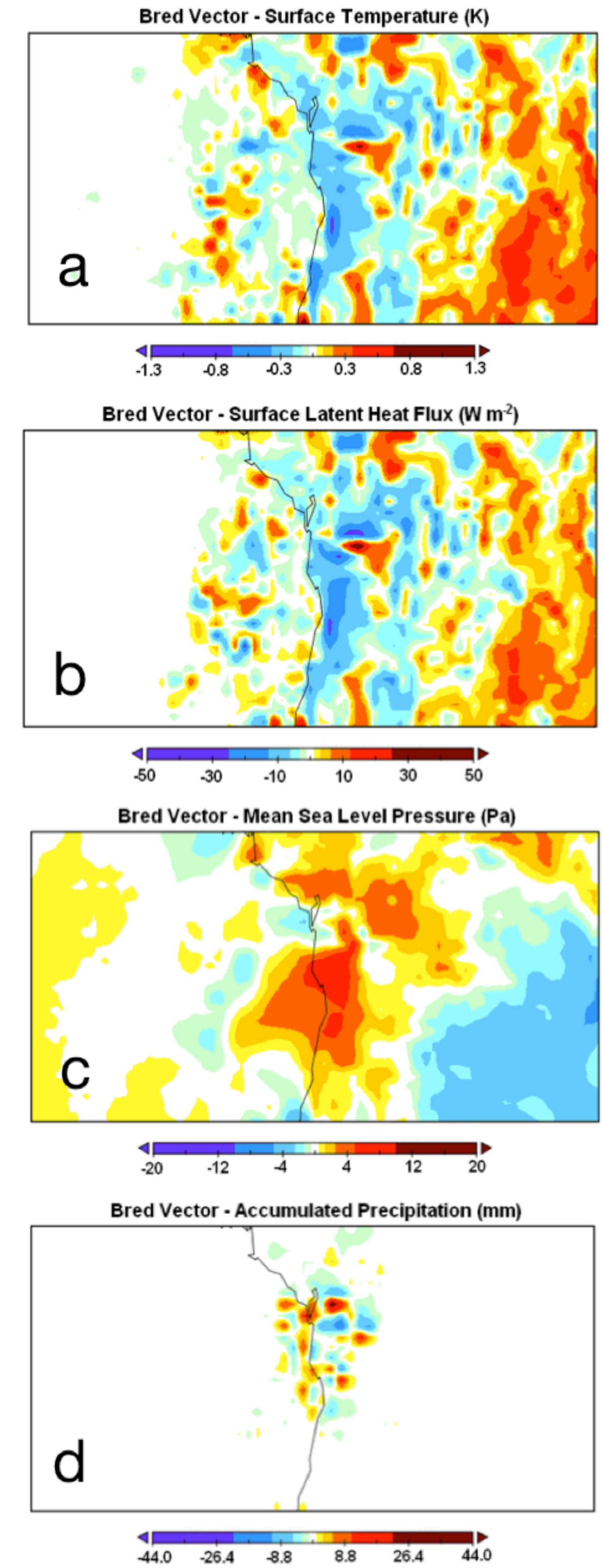


Figure 4: *Mean bred vectors of surface temperature, latent heat flux, MSLP and accumulated precipitation from ACCESS-RC for the forecast made on 6th January 2011.*

The use of EBV initialization for the forecasted SSTs in the ACCESS-RC forecasts resulted in a relative cooling across SE Queensland by between 0.5–2.0 K when compared to 12 km ACCESS-R (Figure 5a). This had a stabilizing effect upon the forecasts for the area and was associated with a subtle increase in surface pressure and decreased near-surface wind speeds. These conditions also resulted in a decrease in surface latent heat flux (Figure 5b) that reduced the supply of moisture from the surface to the developing storm system. The control 12 km ACCESS-R 72 hour forecast of accumulated precipitation for January 10th showed a sharp, intense peak just to the south of Brisbane.The maximum accumulation for this run was 408.9 mm, which was well in excess of even the greatest single gauge observation (Peachester - 298.0 mm). The greatest rainfall accumulation in this case was generally confined to a small number of grid boxes along an east-west orientation (Figure 6a). The ACCESS-RC 72 hour forecast of accumulated precipitation was located in a similar position to that of the operational run.

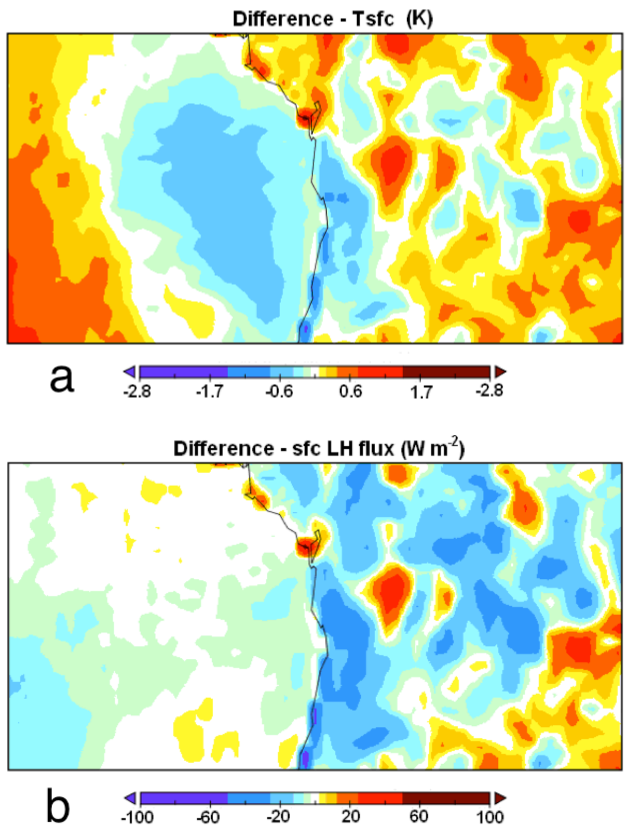


Figure 5: *Mean differences in (a) surface temperature, (b) latent heat flux between ACCESS-RC and ACCESS-R for forecasts made for the 6-13th of January 2011.*

However, as shown in Figure 6b, the maximum accumulation of precipitation was substantially reduced (253.0 mm). Similarly, whilst the total area of rainfall was relatively unchanged the distribution of rain across the Brisbane region was slightly more spread out. This pattern in the accumulated precipitation indicates a 'smoothing out' of the precipitation distribution across the region.

Time-series of rainfall amounts from the model for the Brisbane region were validated against observations of basin average rainfall for the Brisbane River catchment (Figure 7). There is a strong similarity between the rates of precipitation accumulation with time early in the forecast period.

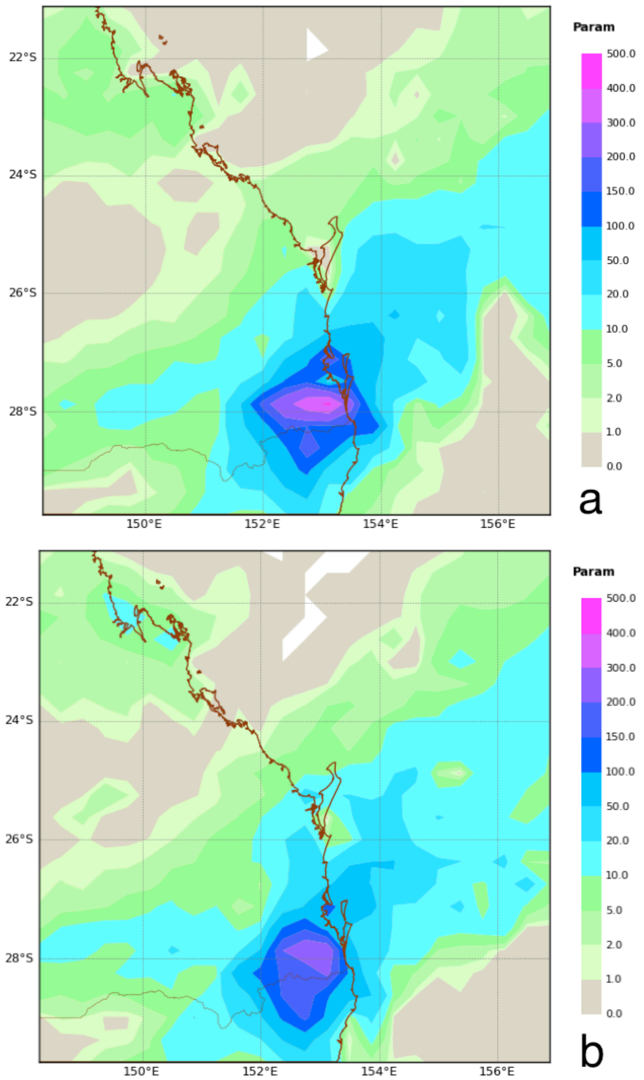


Figure 6: *48-72h forecast accumulated rainfall for Jan 10, 2011 (a) 12 km ACCESS-R (max 409 mm) (b) ACCESS-RC (max 253 mm).*

This is followed by a rapid divergence between the models during January 10 2011. It should be noted that the difference between the models was not between the rates of precipitation during this period (almost identical) but in duration (12 km ACCESS-R: ~17 hours; ACCESS-RC: ~10 hours). There may be a few reasons for this. Most apparent is the increase in local atmospheric stability arising from the surface-cooling tendency. Combined with the local decrease in moisture flux to the atmosphere, both would act to reduce the accumulated precipitation. Less clear is the response of the model parameterization to the change in SST and what the precise nature of the observed response in rainfall may be. Furthermore, the ACCESS-RC result was the product of changes not only to the SST initial condition, but also to the time evolution of SST via coupling to OceanMAPS. It would be of interest to verify the model sensitivity to each of these changes individually. Addressing these questions will form a future phase of this work.

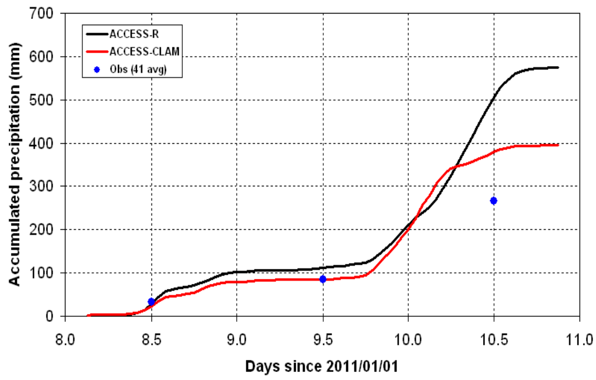


Figure 7: *Comparisons of model forecasts of accumulated rainfall for 8-10 January, 2011 for Brisbane region. Observations are basin area averages for the Brisbane River catchment.*

Whilst the coupling of SST did bring about a decrease in predicted rainfall (~ 33%) bringing it much closer to observations, there remained a substantial over-prediction of rainfall for this event. This may reflect a high sensitivity of the convection parameterization with the increase in grid resolution, but it is of note how strongly the model responds to what is a relatively small change in the lower boundary condition. A further test of 12 km ACCESS-R against other limited area model configurations (such as ACCESS-A) would be valuable in this regard.

Conclusion

ACCESS-R 12 km 72 hour forecasts of accumulated precipitation were evaluated for the Brisbane flooding event of January 2011. Whilst the model proved to represent well the timing and location of the event at long lead-times, there was a tendency for the model to overpredict the total rainfall. The sensitivity of 12 km ACCESS-R to perturbations in SST was investigated by running identical forecasts using a coupled atmosphere-ocean NWP model, ACCESS-RC. Using an EBV initialization approach, the coupled version of the model ran with less biased initial ocean temperatures and evolved its SST via coupling to OceanMAPS. This change in SSTs resulted in a ~33% reduction in precipitation accumulation in ACCESS-RC, bringing forecasts much closer to observations. This improvement was in part due to local surface cooling and the reduction in moisture flux to the boundary layer due to the coupling of SST. However, questions still remain regarding the sensitivity of the model response, such as to what extent was the rainfall sensitive to different SST initial conditions or to the coupling to the ocean model. Also what is the precise nature of the rainfall response to perturbations in SST and is this a linear or non-linear process. ACCESS-R 12 km responded quite strongly to a small change in the SST boundary condition, therefore it is of interest to test whether this sensitivity exists in other configurations of the limited area model.

Although this study is one isolated example, it demonstrates model sensitivity to small changes to the SST lower boundary condition, particularly under the strong dynamic forcing associated with a severe rainfall event. From these experimental forecasts there is evidence that a more representative SST lower boundary condition can be generated by initialization with an Ensemble Bred Vector initialization approach and coupling to an evolving ocean model SST. In contrast to the current use of single value foundation SSTs to derive fluxes from the ocean, coupled NWP allows for more frequent updates to the ocean fluxes being supplied to the boundary layer. As demonstrated here, small changes in this lower boundary condition can have a substantial impact upon model forecasts of extreme events.

Acknowledgments

Thanks to Kevin Tory, Hongyan Zhu, Mike Naughton and Kamal Puri for their helpful comments on these results, and to Wenming Lu for help in preparing some of the figures.

References

Brassington, G.B., T. Pugh, C. Spillman, E. Schulz, H. Beggs, A. Schiller and P.R. Oke, 2007. BLUElink> Development of operational oceanography and servicing in Australia. Journal of Research and Practice in Information Technology, 39, pp. 151-164.

Evans, J., and I. Boyer-Souchet, 2012: Local sea surface temperatures add to extreme precipitation in northeast Australia during La Niña, Geophys. Res. Lett., doi:10.1029/2012GL052014, in press.

Frederiksen J.S., C.S. Frederiksen and S.L. Osbrough, 2010. Seasonal ensemble prediction with a coupled ocean-atmosphere model. Australian Meterological and Oceanographic Journal, 59, 53-66.

Minnett, P.J., 2011: Clarifications on SST definitions, Proceedings of the GHRSST XII Science Team Meeting, Edinburgh, Scotland, June 27 – July 1, 2011, 310 pp. See also: https://www.ghrsst.org/ghrsst-science/sst-definitions.

National Climate Centre, 2011: Frequent heavy rain events in late 2010/early 2011 lead to widespread flooding across eastern Australia, Special Climate Statement 24, Bureau of Meteorology, 28 pp. Online: http://www.bom.gov.au/climate/current/statements/scs24c.pdf

O'Kane, T.J., P.R. Oke and P.A. Sandery, 2011. Predicting the East Australian Current. Ocean Modelling, 38(3-4), 251-266.

Oke, P.R. and P. Sakov, 2008. Representation Error of Oceanic Observations for Data Assimilation. Journal of Atmospheric and Oceanic Technology, 25, 1004-1017.

Oke, P.R., G.B. Brassington, D.A. Griffin and A. Schiller, 2008. The Bluelink ocean data assimilation system (BODAS). Ocean Modelling, 21, pp. 46-70.

Sandery P.A, T.J. OKane, G.B Brassington and J Freeman, 2012. Coupled Initialisation in an ocean-atmosphere tropical cyclone prediction system, submitted to QJRMS.

Toth, Z., and E. Kalnay, 1997. Ensemble forecasting at NCEP and the breeding method, Mon. Weath. Rev., 125, 3297-3319.

Evaluation of Medium Range Weather Forecasts Based on Short-term Forecasts

Deryn Griffiths and Jiwon Park

New South Wales Regional Office,

Bureau of Meteorology, Australia

D.Griffiths@bom.gov.au

Introduction

Official weather forecasts in Australia are produced manually and presented as text, and more recently in gridded format. Judging the skill of many aspects of these forecasts is typically a time consuming process and not undertaken routinely. There has been a history of verifying maximum and minimum temperature forecasts at point locations. However, the difficulty of verifying other aspects of forecasts is compounded by the difficulty of decoding text forecasts, inadequate observation datasets, and the large data sets required to verify probabilistic forecasts using standard indices and techniques for assessing probabilistic forecasts.

In this paper a new method of assessing medium range forecasts is described. The key to the method is to use the short-term (24 hour) official forecast to judge the skill of the medium-term forecast produced 4 to 6 days earlier. The use of 24-hour forecasts instead of verifying observations overcomes many of the difficulties which otherwise impede the assessments of the forecasts. We discuss why any error to do with using 24-hour forecasts in place of observations is considered acceptable.

This paper presents an example of the application of the new method to assess the relative skill of automated and official medium-term forecasts. The forecasts included descriptive text for points and areas, and gridded deterministic and gridded probabilistic forecasts for New South Wales and the Australian Capital Territory. Although only a limited set of forecasts were analyzed, the results changed the practice of the forecasters to rely more on the automated forecasts than they did previously.

Methodology

During the last quarter of 2011, the NSW Office of the Australian Bureau of Meteorology conducted a subjective assessment of official forecasts for 21 distinct days. The forecasts had lead-times of 5 to 7 days. Automated forecasts were also created and collected. The days surveyed were based on convenience of collection of the data rather than targeting any particular weather situations.

Official Forecasts: Official weather forecasts for NSW and the ACT were produced manually, using a Graphical Forecast Editor (GFE). Forecasts were created up to seven days in advance, for every point in NSW on a 6km by 6km grid, some forecasts being deterministic and some probabilistic. The GFE allows forecasters to manipulate objective guidance, or the previous forecast, to match their policy formulated based on objective guidance inside and outside the GFE, climatology and experience.

Screen grabs were collected of three gridded forecasts for each day within the survey. The forecasts were (i) a 3pm local time step of Wind, (ii) the chance of exceeding 0.2mm rainfall in a 24-hour period (referred to as DailyPoP) and (iii) the amount of rain expected to fall with at least a 50% confidence (referred to as DailyPrecip50Pct).

The GFE has an automatic text formatter, which creates text forecasts from the gridded forecasts. Twenty text forecasts created this way, based on Official Forecast grids, were collected each day. These comprised six "Metro" forecasts for areas varying from 7 to 130 grid cells, four "Town" forecasts for one grid cell, and 10 "Précis" forecasts which provide a summary forecast, of no more than 30 characters, for one grid cell. The Précis points chosen were selected primarily from within the Metro areas or Towns assessed.

Table 1: *Text forecasts assessed each day.*

Forecast Type	Forecast Locations
Metro (area)	Sydney, Canberra, Central Coast, Newcastle, Alpine, Wollongong
Town (point)	Coffs Harbour, Katoomba, Lismore, Orange
Précis (point)	Sydney, Thredbo, Canberra, Gosford, Newcastle, Wollongong, Coffs Harbour, Katoomba, Orange, Lismore*, Goulburn*

**Lismore and Goulburn were assessed most, but not all, days.*

Automated Forecasts: Acceptable standard editing techniques for days 5 to 7 gridded forecasts have been developed in the NSW office. These rely heavily on guidance, with some restricting of values, such as removing any gale force strength winds.

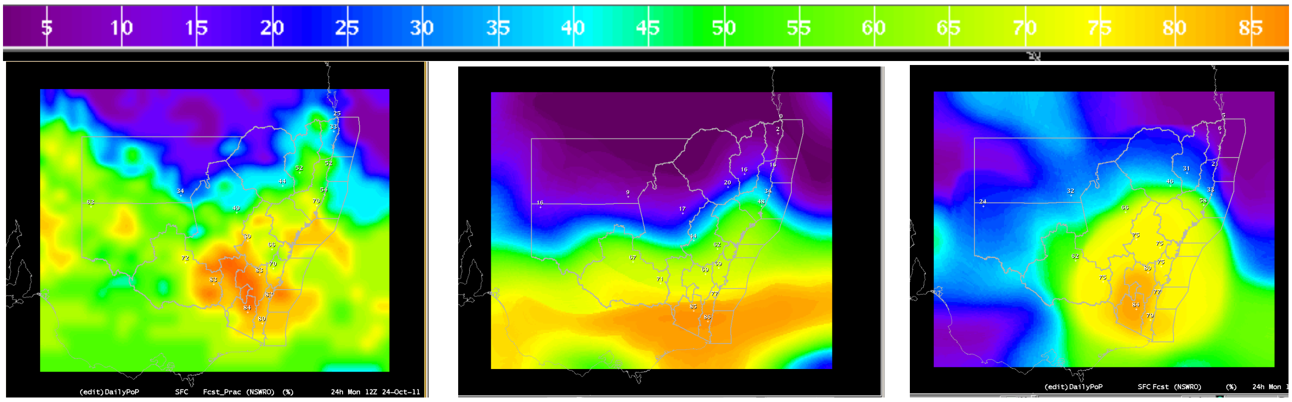


Figure 1: *An example of a set of gridded forecasts of the 24-hour Chance of Exceeding 0.2mm (DailyPoP) for New South Wales, all with the same valid period. Left: Automated (5-day lead-time) Middle: Official (5 day lead-time) Right: Official (1 day lead-time)*

The restriction of values is based on what is climatologically likely so as not to forecast an extreme event at that lead-time based on only one model run. The standard editing techniques also apply some consistency checks, and use only the word "Showers" rather than trying to distinguish between "Showers", "Rain", "Drizzle" and "Thunderstorms."

The guidance used is Optimal Consensus Forecasts (OCF) for Minimum and Maximum Temperature, Cloud Cover, Chance of Rain and Amount of Rain, and direct numerical model output for Wind. The components of OCF are bias corrected for the temperature forecasts (Engel *et al.*, 2007). The Chance of Rain is calibrated (Bureau of Meteorology, 2011b). Multi-model averages are used for Cloud Cover and Amount of Rain.

What is called "automated forecasts" in this paper, were created manually using the standard editing techniques.

The automated forecasts collected were for the same time period as the official forecasts and were based on the guidance available at the time the official forecasts were being prepared.

Verifying Information: Day 1 official forecasts, (issued around 4pm for the following day), were collected corresponding to the medium-term forecasts to be assessed. These Day 1 forecasts are referred to as "verifying forecasts." Historical point-based verification of forecasts against observations has shown the increase in skill of Day 1 forecasts compared to 5 to 7 day forecasts. For example, forecast rainfall probability for Sydney in 2008-09 had Brier Skill Scores referenced by climatology of 0.06 and 0.21 at lead-times of 7 days and 5 days respectively, improving to 0.44 at a 1 day lead-time (Bureau of Meteorology, 2009). Similarly, 2008 to 2011 Canberra Maximum Temperature forecasts had root mean square errors that improved from 2.2 and 2.7 at Days 5 and 7 to only 1.5 at Day 1 (Bureau of Meteorology, 2011a). The significant increase in skill gave reassurance in using Day 1 forecasts as a reference for assessing the longer-term forecasts. The Day 1 official forecasts were considered appropriate as "verification" in place of observations as they represented the best possible forecast according to the forecaster on duty. This is reasonably assumed to be significantly more accurate than a longer term forecast, as supported by the point-based verification quoted above. In addition, the verifying forecasts were of the same format as the forecasts assessed, making comparison relatively easy. There were no textual descriptions based on observations, or gridded analyses readily available for cloud cover or rainfall. What observational data was available was in a different format to the forecasts being assessed. The verifying forecasts were a suitable and practical solution to assessing the longer-term forecasts.

Figure 1 shows an Automated and Official Forecast of the 24-hour chance of exceeding 0.2 mm (DailyPoP) at a 5-day lead-time, and the verifying forecast, which is the official forecast for the same valid time, but with only a 1-day lead-time. Table 2 shows an example of an automated and official text forecast for the alpine area. It is for a 24-hour period and an area of approximately 250 km^2.

Table 2: *An example of a set of text forecasts for a 24-hour period and the alpine area, an area of approximately 250* km^2 *in southern New South Wales.*

Automated forecast (5-day lead-time)	Partly cloudy. Isolated showers. Light winds.
Official forecast (5 day lead-time)	Cloudy. Areas of rain. Winds east to northeasterly averaging up to 20km/h tending northeasterly up to 30km/h around dawn.
Official verifying forecast (1 day lead-time)	Cloudy. Heavy showers developing around dawn, easing to scattered showers around midday. Winds east to southeasterly averaging up to 25km/h.

Also shown in Table 2 is the verifying forecast, which is the official forecast for the same area and with the same valid time, but with only a 1-day lead-time.

Table 3: *Rating Scale used to assess forecasts.*

Score	Description
1	Automated Forecast better than Official Forecast; Official Forecast unacceptable
2	Automated Forecast better than Official Forecast; both forecasts acceptable
3	Both forecasts equally good
4	Official Forecast better than Automated Forecast; both forecasts acceptable
5	Official Forecast better than Automated Forecast; Automated Forecast unacceptable

Rating Scale: The verifying forecast was used to determine which of the medium range forecasts, Automated or Official, was better, if either. Three days were assessed by three people to check the level of consensus in the forecasting team. Following this, one forecaster (the second author) assessed all 21 days of forecasts. The assessment was kept simple with a rating scale of 1 to 5 used as shown in Table 3. By "unacceptable" we meant that the forecast might cause embarrassment to the Bureau of Meteorology.

Results

Assessments: Approximately 420 text forecasts were assessed, and the ratings are shown in Table 4, both grouped together and separated according to lead-time. The ratings for 63 gridded forecasts are shown in Table 5 and arranged as for Table 4. The results show that both the Automated and Official forecasts were almost always acceptable. The Automated forecast was as good as, or better than, the Official forecast about 80% of the time, although for text forecasts with a lead-time of 5 days, the analysis suggests that the Official Forecasts may be slightly better than the Automated Forecasts.

Unacceptable Assessments: One graphical and two text forecasts were assessed as unacceptable. The particular forecasts were as follows. (i) One official graphical forecast showed an extreme (0%) probability forecast, which was re-assessed as near 50% in the verifying forecast. (ii) One automated text forecast was inconsistent within itself, with showers and mainly sunny. This forecast was most likely due to the standard editing techniques to create the automated forecast not being followed correctly. (iii) One automated point forecast indicated showers even though that point was within a district for which the district forecast did not indicate showers. This sort of inconsistency between point and spatial forecast is a known limitation of the way the text is created in the GFE that can affect an official or automated forecast when showers are only expected in a very small proportion of a district.

Gridded Forecast Assessments: An attempt was made to understand why the automated gridded forecasts were superior to the Official ones. The wind grids of the automated and official forecasts were almost identical, suggesting the standard editing techniques were being used to create the official forecast. The official forecasts of DailyPoP (the chance of rain exceeding 0.2 mm in a 24-hour period) and of DailyPrecip50Pct (the amount of rain expected to fall with at least a 50% confidence) showed that forecasters had a common practice of limiting the DailyPoP to no more than 50% at a 5-7 day lead-time. This practice followed an earlier iteration of the standard editing procedures relevant when the OCF guidance for DailyPoP was uncalibrated. The results of the assessment confirmed that the more up-to-date procedures, relying more on the guidance, were appropriate, and gave better forecasts than the procedures used by many forecasters to create the official forecasts.

Table 4: *Text Forecast Analysis*

Score	1 (Automated better)	2	3	4 (Official better)	5
+5, 6 and 7 day forecasts	0%	23%	53%	24%	< 1%
+7 day forecasts	0%	24%	57%	18%	< 1%
+6 day forecasts	0%	33%	46%	21%	0%
+5 day forecasts	0%	12%	55%	32%	< 1%

Table 5: *Graphical Forecast Analysis*

Score	1 (Automated better)	2	3	4 (Official better)	5
+5, 6 and 7 day forecasts	2%	33%	52%	13%	0%
+7 day forecasts	0%	38%	62%	0%	0%
+6 day forecasts	0%	29%	57%	14%	0%
+5 day forecasts	5%	33%	38%	24%	0%

Response of Forecasters to the Results: For forecasters in the NSW Regional Office, the analysis confirmed that a forecast based on consensus guidance is suitable at 5 to 7 day lead-times. The results were convincing enough to most forecasters, for at least the day 6 and 7 lead-times, to change their practice so as to rely heavily on the latest guidance and standard editing techniques. By doing so, they have been able to free up time to spend on improving their shorter term forecasts and on contributing to office projects to improve forecasts in a strategic manner. Forecasters have requested that the standard editing techniques be coded to enable them to be run in a truly automated way, reducing the risk of operator errors and reducing the time spent by forecasters on Days 5 to 7. This was done and introduced into the NSW Regional Office in a preliminary manner in September 2012. The results of the analysis highlight the need for a routine verification scheme to assist forecasters make the best use of the available guidance.

Comparisons to Other Studies: The results of this work are consistent with those of the Project Phoenix study in Canada. That study used a different analysis technique and found that forecasters add little if any value to automated forecasts beyond a 48-hour lead-time (McCarthy *et al*., 2007).

Further Work

Objective Assessment of Probabilistic Forecasts: For medium-term probabilistic forecasts, the assessment used above could be extended to an objective assessment. For example, scores such as the Brier Score could be applied using a Day 1 forecast in place of the usual observed probability of 0 or 1 according to whether the event was observed to occur or not. In fact, in many situations, particularly when trying to analyze rainfall in data-sparse regions, and even in regions with radar coverage, it is difficult to be confident of whether the event occurred, and it may be more honest to allow the "observed" probability to take a value between 0 and 1 according to our confidence of it having occurred. For the assessment described here, only screen grabs were captured. The automated forecasts were not saved in a gridded format. It would have been interesting to compare an objective assessment of the probability forecasts to the subjective assessments made.

Summary

Comparing medium-term forecasts to short-term forecasts of the same style provides an effective way to assess the longer lead-time forecasts. A limited assessment period provided sufficient information to change the practice of forecasters. The index used in the assessment was a simple but effective way of comparing the alternative medium-term forecasts. The technique of using short-term forecasts as a reference allowed assessment of probability forecasts, and allowed assessment of forecasts for which there were no corresponding observations or analyses available.

Acknowledgments

Thanks to Beth Ebert for encouragement to publish the results.

References

Bureau of Meteorology 2009, *New South Wales and Australian Capital Territory Annual Report 2008-2009.*

Bureau of Meteorology 2011a, *New South Wales and Australian Capital Territory Annual Report 2010-2011.*

Bureau of Meteorology 2011b, *NMOC Operations Bulletin No. 87 - Upgrades to the Operational Gridded OCF and PME Systems.*

Engel, C., Ebert, E., & Lane, T. P. (2007). Mesoscale gridded operational consensus forecasts. Proceedings of the 22nd Conference on Weather Analysis and Forecasting and the 18th Conference on Numerical Weather Prediction.

McCarthy P.J., Ball D., Purcell W. 2007. Project Phoenix – Optimizing the Machine-Person Mix in High Impact Weather Forecasting *Proceedings of the 22nd Conference on Weather Analysis and Forecasting and the 18th Conference on Numerical Weather Prediction* American Meteorological Society.

Initializing ACCESS from ERA-Interim data

Greg Roff[A], Tom Green[B], Martin Dix[A] and Scott Wales[C]

[A]Centre for Australian Weather and Climate Research,

700 Collins Street, Docklands, Victoria, 3008, Australia

[B]Met. Office, Fitzroy Road, Exeter, EX1 3PB, United Kingdom

[C]Computational Modelling Support, School of Earth Sciences,

University of Melbourne, Australia 3010

g.roff@bom.gov.au

Introduction

The Australian Community Climate and Earth System Simulator (ACCESS) is a coupled model using an Ocean Atmosphere Sea Ice Soil (OASIS) coupler (Valcke, 2006) to link the atmosphere with land surface, ocean and sea ice components (Puri, 2005). ACCESS versions 1.0 and 1.3 are being used for the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5), Coupled Model Intercomparison Project Phase 5 (CMIP5), and Atmospheric Model Intercomparison Project (AMIP) simulations. These two ACCESS versions are based on the United Kingdom Met Office (UKMO) Unified Model (UM) 7.3 HadGEM2 (Collins *et al*., 2008) and HadGEM3 (Hewitt *et al*., 2010), respectively, with a major difference being that the former uses the UKMO Met Office Surface Exchange Scheme (MOSES) land surface scheme (Cox *et al*., 1999) while the latter the Australian developed CSIRO Atmosphere Biosphere Land Exchange (CABLE) scheme (Kowalczyk *et al*., 2006).

The atmospheric and land surface components of the coupled model can be run with ancillary files providing the information which would normally come from the other components of the coupled model (sea surface temperature, sea ice thickness etc). This version of the coupled model, which is what ACCESS will refer to from now on, is much faster to run and can be used for Numerical Weather Prediction (NWP) short-term/seasonal forecasting purposes. However, in NWP experiments, such as running hindcasts for seasonal prediction or case studies on bushfire, flooding or tropical cyclone situations, the initial conditions are very important and so a standard source of initial conditions is desirable.

This paper introduces a procedure to create initial conditions suitable for ACCESS from the European Centre for Medium Range Weather Forecasting (ECMWF) Re-analysis (ERA) dataset ERA-Interim (Dee *et al*., 2011). This dataset covers the years 1979-present, but the same procedure can also be used for ERA-40 (Uppala *et al*., 2005), which covers years 1957-2002. Difficulties encountered when initializing low and high-top versions of ACCESS are also examined.

ACCESS operation

The ACCESS 1.0 and 1.3 UMUI (UM User Interface) build and N96L38 run job identifiers on both the Bureau's in-house supercomputer *solar* and its equivalent at NCI *vayu* can be seen in Appendix Table 1. These jobs have been compared and give identical results on both computing systems using the same ancillary files and can now be used by the wider community to run experiments based on the ACCESS AR5/CMIP5 AMIP simulations.

These jobs are limited in that they require UM formatted initial conditions. For climate simulations, where initial condition specifics are not essential, there are suitable initial conditions, or 're-start' files, available from previous ACCESS coupled model experiments. However for shorter range NWP or seasonal forecast situations, where initial conditions may be critical, such data may not be readily available. This prompted the development of a procedure to enable ACCESS to be initialized from ERA-Interim data. The Build and Run jobs on *solar* for this reconfiguration are also seen in Appendix Table 1 and note that there are two reconfiguration run jobs corresponding to a low-top model with 38 vertical levels and a high-top model with 85 vertical levels.

A short description of this procedure and the relevant UMUI jobs on *solar* are discussed in the following sections. The Appendix has more detailed instructions centred on the UMUI jobs in Appendix Table 1 and also indicates the location of a help directory on *solar* where software code and example datasets can be found. The corresponding *vayu* reconfiguration jobs and help directory are under construction.

Reconfiguration with ERA-Interim data

Though the UMUI has the capacity to run from ECMWF gridded binary (grib) formatted data this has not been possible in Australia (due to copyright restrictions) until recently when Tom Green created a

UM vn7.9 patch. This patch requires the ECMWF GRIB API, which is an application program interface accessible from C and FORTRAN programs developed for encoding and decoding World Meteorological Organization (WMO) FM-92 GRIB edition 1 and edition 2 messages. Note that the patch will not be incorporated into the UM trunk until at least UM vn8.4 is released (private communication, Thomas Green). On *solar*, this patch has been applied to the UMUI UM vn7.9 reconfiguration build job and the resultant reconfiguration executable has been tested in the UM vn7.9 N96L38[3] reconfiguration run job. A copy of the patch (patch.diff) and instructions on how to apply it can be found in the help directory along with copies of a sample ERA-Interim grib file and the resultant UM formatted reconfiguration file. ERA-Interim daily data at 00Z, 06Z, 12Z and 18Z from 1979 to the present can be obtained freely from http://data-portal.ecmwf.int/.

While several institutions have various copies of various fields from the ERA-Interim dataset, none have the complete dataset. As the above procedure will enable the Bureau of Meteorology, CSIRO and the Universities to initialize ACCESS from ERA-Interim data, and to consolidate all these datasets, a site has been created on the Data Centric Compute (DCC) resource at the National Computing Infrastructure National Facility (NCI-NF) which enables easy access to this dataset and which is updated monthly. Details on how to access this dataset at NCI on the DCC are at http://climate-cms.unsw.wikispaces.net/ERA+INTERIM.

The patch works with full (0.75°x0.75°) or low (1.5°x1.5°) horizontal resolution ERA-Interim grib data on ECMWF model or pressure levels in the vertical as seen in Figure 1. The various height ranges available are: high resolution data on the actual 91 vertical ECMWF model levels extending to 0.01 hPa; high resolution also on a reduced 60 vertical model levels extending to 0.1 hPa; and, low resolution on 37 pressure levels extending to 1 hPa. Note, however, that the first is only available for the Year of Tropical Convection (YOTC) period from 2008-05-01 to 2010-04-30 (see http://www.ucar.edu/yotc for YOTC details).

The choice of which ERA-Interim dataset to select should depend on the ACCESS model configuration being used, since ACCESS vertical model levels that are not covered by the ERA-Interim data will be kept constant at the highest vertical values available or linearly extrapolated (depending on UMUI reconfiguration settings) – either of which will not be realistic and may lead to problems with the simulation, as shown later. Several ACCESS configurations for different simulations, with their number of vertical levels indicated by L, are also shown in Figure 1. These range from climate and NWP through to chemistry and middle-atmosphere cases: the standard climate L38 ACCESS 1.0/1.3; the present operational Australian Parallel Suite 0 L50 NWP APS0; the ACCESS chemistry model L60 UKCA; the proposed new operational L70 NWP APS1; a high-top test version of ACCESS 1.3 L85 WIRADA. A low vertical resolution simulation, such as L38 ACCESS, can use any of the ERA-Interim datasets as its full height range is always covered by the ERA-Interim data. However, if for example the L70 NWP configuration is used then this will not be true. This is not a problem provided the extrapolated layer aloft is not too large as then the upper levels have been found to quickly adjust to realistic flows (as was seen in Roff *et al.*, 2010).

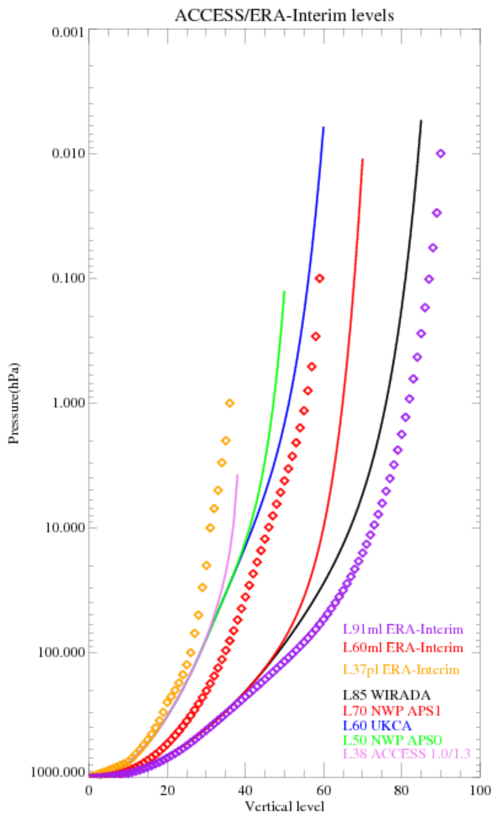


Figure 1: *Vertical level number against pressure height for various model configurations (solid lines) and datasets (diamonds): L38=ACCESS 1.0/1.3; L50=NWP APS0; L60=UKCA; L70=NWP APS1; L85=WIRADA; ERA-Interim model level L91 and L60, and pressure level L37. Here L indicates the number of levels and ml and pl indicate if ERA-Interim are model levels or pressure levels, respectively.*

The reconfiguration requires four 3D fields (U, V, T and Q) and four surface fields (surface pressure, skin temperature, geopotential and land-sea mask – with the latter two being invariant) from ERA-Interim and placed in one grib file. Note that all the grib fields must

[3] N96L38 indicates the resolution of the ACCESS run with: N96 indicating 2x96=192 longitudinal grid points and 1.5x96-1=145 latitudinal grid points; L38 indicating 38 model levels in the vertical.

be on the same latitude/longitude grid. This is then supplied as the start "dump", or initial conditions, for the reconfiguration run job to produce a N96L38 um-formatted initial condition file. The help directory has examples of these while Figure 2(a) shows zonal mean zonal wind field from the grib file and Figure 2(b) shows this field from the reconfigured file.

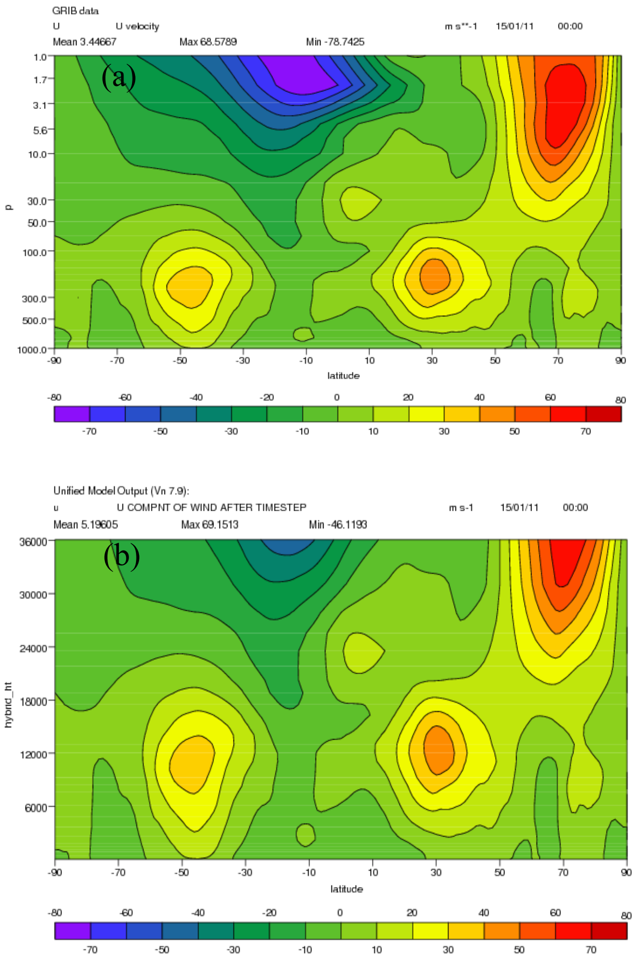


Figure 2: (a) Zonal mean zonal wind contour plot on latitude/pressure level (hPa) axes from ERA-Interim low-resolution L37 grib data. (b) The corresponding field after reconfiguration to N96L38 um-format on latitude/height (m) axes.

Note that in the reconfiguration step the horizontal interpolation can be either bilinear or area averaged and the vertical interpolation linear with or without extrapolation. The best results, as shown here, were found when area averaged horizontal interpolation and linear interpolations without extrapolation in the vertical are used in the reconfiguration job.

The ERA-Interim reconfigured file produced can now be used as initial conditions for ACCESS 1.0. Unfortunately ACCESS 1.3 cannot use the initial condition file yet. This is because the reconfigured file is configured to run with the four soil levels and nine surface types of the MOSES land-surface scheme – which is used in ACCESS 1.0 – and not the six soil levels and seventeen tiles used in the CABLE land-surface scheme run in ACCESS 1.3.

In order to create ACCESS 1.3 initial conditions python scripts are used to copy the atmospheric fields from the ACCESS 1.0 reconfigured file created above onto a basic ACCESS 1.3 AMIP dump file which has suitable soil moisture levels and tiles. Instructions on how to do this are in the help directory along with the python scripts in a tar file as well as examples of an AMIP file and the final ACCESS 1.3 dump file produced.

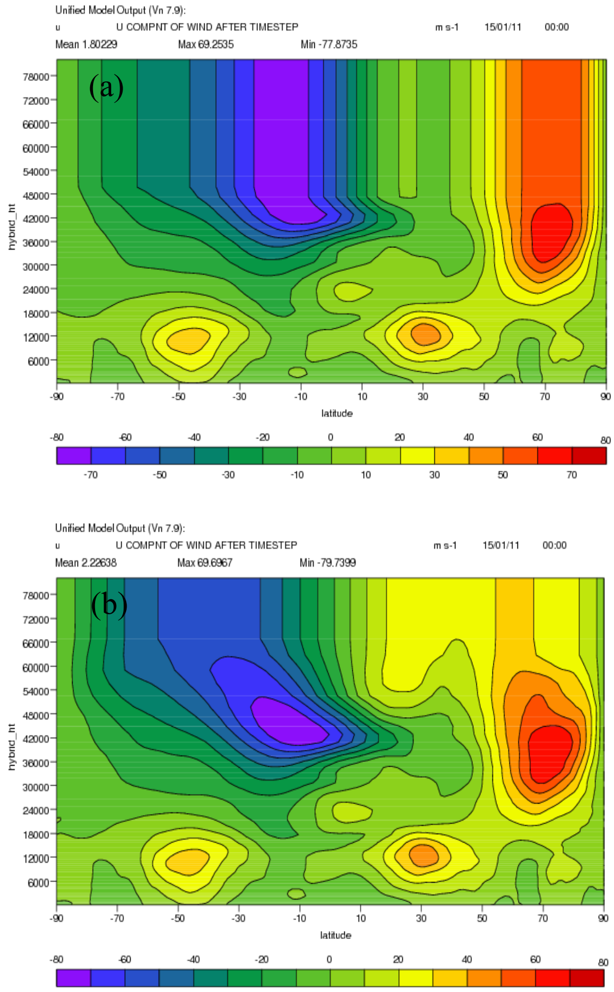


Figure 3: A zonal wind zonal section through: (a) a simple reconfiguration to N96L85 (~82 km top) of the N96L38 (~36 km top) ERA-Interim L37 pressure level reconfiguration dump file seen in Figure 2(b); (b) N96L85 ERA-Interim L60 model level reconfiguration dump file.

The impact of low-resolution pressure level and full-resolution model level ERA-Interim data on N96L38 ACCESS 1.3 initial conditions are small, as both extend beyond the height of the L38 model and are due to the increased vertical resolution in the full-resolution ERA-Interim data.

Initializing a high-top ACCESS model

The above procedures have been used to successfully run N96L38 ACCESS 1.0 and 1.3 AMIP simulations initialized from ERA-Interim L37 pressure level data, as discussed, and L60 model level data (L91 model level YOTC data was not used because of the need to start the simulation in 1981). Figure 2(b) shows the zonal mean zonal wind field from such a dump file created from the L37 pressure level data, however, when an L85 high-top version of ACCESS 1.3 was attempted it was first initialized with a standard reconfiguration of the L38 ACCESS 1.3 dump file created above from 38 to 85 levels, with model tops near ~36 km and ~82 km, respectively. The AMIP run thus initialized failed due to the L85 model top being too far above the L38 top (see Figure 1). The simple reconfiguration could only apply constant, or even worse, linearly extrapolated values at these levels taken from the top of the L38 dump file, as seen in Figure 3(a). These unrealistic constant values over such a large vertical range (5-0.01 hPa) were unstable and led to CFL failure due to very large zonal winds forming in the polar night jet.

The solution was to again use python scripts to copy corresponding L60 model level ERA-Interim atmospheric fields, which had been reconfigured to L85 via the *solar* UM vn7.9 N96L85 reconfiguration run job, onto the L85 dump file, resulting in initial conditions seen in Figure 3(b). The AMIP simulation then ran successfully, even though the very top of L85 (0.1-0.01 hPa) did have extrapolated values as the model quickly adjusted them to realistic values.

Conclusions

Standard jobs indicated in Appendix Table 1 have been created on *solar* and *vayu* which enable ACCESS 1.0 and 1.3 to be run and produce identical results. These will enable researchers from the Universities, CSIRO and the Bureau of Meteorology to carry out experiments and compare them to the ACCESS AR5/CMIP5 AMIP runs.

These standard ACCESS jobs have been enhanced by the creation of reconfiguration jobs on *solar* which enable ERA-Interim data to be used as initial conditions. This enables the ACCESS model, whether in climate mode or as used in NWP and seasonal forecasting, to be suitably initialized so it can be used to examine extreme weather events (floods, bushfires, and tropical cyclones) and to be used in hind-casts for seasonal prediction studies. Similar reconfiguration jobs are being created for *vayu*. The expansion of ACCESS into these areas has further been supported by the creation of a continuously updated ERA-Interim repository at NCI which the Universities, CSIRO and the Bureau of Meteorology can access. These three advances should enable more ACCESS development and experimentation.

Acknowledgments

Many people have been involved in this work and the authors would like to thank: Mike Rezny, Paola Petrelli, Ben Evans, Gareth Berry, Christine Chung, Asri Sulaiman, Maggie Zhou, Christian Jakob and Nathan Bindoff.

References

Collins, W.J., N. Bellouin, M. Doutriaux-Boucher, N. Gedney, T. Hinton, C. D. Jones, S. Liddicoat, G. Martin, F. O'Connor, J. Rae, C. Senior, I. Totterdell, S. Woodward, T. Reichler, J. Kim, 2008: Evaluation of the HadGEM2 model. Met Office Hadley Centre Technical Note no. HCTN 74, available from Met Office, Fitzroy Road, Exeter EX1 3PB, http://www.metoffice.gov.uk/publications/HCTN/index.html

Cox, P.M., R.A. Betts, C.B. Bunton, R.L.H. Essery, P.R.Rowntree and J. Smith, 1999: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, *Clim. Dyn*., **15**, 183-203.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, H´olm EV, Isaksen L, K°allberg P, K¨ohler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Th´epaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc*. **137**: 553–597. DOI:10.1002/qj.828.

Hewitt, H.T., Copsey, D., Culverwell, I.D., Harris, C.M., Hill, R.S.R., Keen, A.B., McLaren, A.J., and Hunke, E.C., 2010: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, Geosci. Model Dev. Discuss., 3, 1861–1937, 2010, www.geosci-model-dev-discuss.net/3/1861/2010/doi:10.5194/gmdd-3-1861-2010.

Kowalczyk, E.A., Y.P. Wang, R.M. Law, H.L. Davies, J.L. McGregor and G. Abramowitz, 2006: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, http://www.cmar.csiro.au/eprint/open/kowalczykea_2006a.pdf.

Puri, K, 2005: Project plan for ACCESS, 2005, available at http://www.accessimulator.org.au/report/index.html.

Roff, G., D. W. J. Thompson, and H. Hendon (2011), Does increasing model stratospheric resolution improve extended range forecast skill? , *Geophys. Res. Lett.* , **38** , L05809, doi:10.1029/2010GL0465.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., 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., van de Berg, L., 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. *Quart. J. R. Meteorol. Soc*., **131**, 2961 3012.doi:10.1256/qj.04.176.

Valcke, S.: OASIS3 User Guide (prism 2-5). PRISM Support Initiative No 3, 68 pp., 2006.

Appendix

Appendix Table 1 below lists the ACCESS 1.0 and 1.3 UMUI build and N96L38 run job identifiers on both the Bureau's in-house supercomputer *solar* and its equivalent at NCI *vayu*. Also in this Table are the Build and Run jobs on *solar* for the reconfiguration of ERA-Interim data, with the latter for both a low-top model with 38 vertical levels and a high-top model with 85 vertical levels.

Note: In the following we assume the *solar*/*dcc* user accounts are glr/glr548 and the umui jobs are as listed. If you do not need to change the source code then you do not need to build new executables, and can just follow the run instructions listed.

APPENDIX Table 1: *ACCESS 1.0/1.3 build and run jobs on solar and vayu and their original sources on the CSIRO server cherax[4], as well as the ERA-Interim Reconfiguration jobs on solar.*

ACCESS 1.0	*solar*	*vayu*	*cherax*
Build	waaac	saaqa	xajbf
Run N96L38	waaad	saaqa	xajbp
ACCESS 1.3	***solar***	***vayu***	***cherax***
Build	waaaf	saaqb	uaakc
Run N96L38	waaag	saaqb	uaakg
Reconfiguration jobs on *solar*			
Build		xbfii	
Run N96L38 and N96L85		xbfik, xbfij	

There is a help directory on *solar* (*~glr/UMic_from_ERA*) that details how to use the above reconfiguration build and run jobs to create initial conditions for the ACCESS 1.0 and 1.3 build and run jobs listed in the Table. Copies of sample data files are held in a sub-directory (*Files*) and the instructions cover the following steps:

1. Getting ERA-Interim pressure level grib data from ECMWF or DCC and combining them into a single grib file *ei2011011500pl.grib*;
2. Installing the patch (*path.diff*) and then running the UMUI UM vn7.9 **xbfii** job to build the reconfiguration executable. Then running this executable via UMUI UM vn7.9 **xbfik** job using the ERA-Interim grib file as input to create viable ACCESS 1.0 initial conditions *recei2011011500pl*;
3. Running the UMUI UM vn7.3 **waaac** job to build the ACCESS 1.0 executable. Then running this executable via UMUI UM vn7.3 **waaad** job using the *recei2011011500pl* initial conditions to carry-out a N96L38 ACCESS 1.0 simulation (a sample output file *waaada.pcl12e0* is provided);
4. Create ACCESS 1.3 initial conditions by merging the ERA-Interim *recei2011011500pl* initial conditions with an ACCESS 1.3 dump file from a previous coupled or AMIP run which has appropriate CABLE fields for the date in question *dzsjc_amip20110111_orig*. Merging is done via python scripts held in the tar file *py_era_recon.tar* to create viable ACCESS 1.3 initial conditions *ac1.3_2011011500*;
5. Running the UMUI UM vn7.3 **waaaf** job to build the ACCESS 1.3 executable. Then running this executable via UMUI UM vn7.3 **waaag** job using the *ac1.3_2011011500* initial conditions to carry-out a N96L38 ACCESS 1.3 simulation (a sample output file *waaaga_pcb500* is provided).

[4] Note: *Solar* jobs are under user glr; *vayu* jobs under saw562; and *cherax* ACCESS 1.0/1.3 jobs under ras029 and yan06j, respectively.

Application performance improvement by use of partial nodes to reduce memory contention

Ilia Bermous[A], Jörg Henrichs[B] and Michael Naughton[A]

[A]Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

[B] Oracle Corporation Australia Pty Limited, Advanced Customer Support, Melbourne

I.Bermous@bom.gov.au

Introduction

The number of computational cores is increasing rapidly on modern supercomputers, which now include from tens to hundreds of thousands of relatively powerful cores. At the same time, processor performance since 1980 has been increasing at a much faster rate than memory bandwidth (Graham *et al*. 2005, pp.106-107). The authors made a conclusion that the memory bandwidth bottleneck will become a serious problem in the future affecting performance scaling. This situation has been reached on current systems: memory bandwidth is one of the constraining factors for execution speed for HPC applications of interest. Basically, the processes cannot run at full speed due to memory bandwidth limitation, which is aggravated by the fact that more and more cores compete for the same memory. This can be viewed as memory contention between the several cores on a node.

This paper examines the effect of reducing the number of used cores per node for a real-world application, the UK Met Office Unified Model (Davies *et al*., 2005). Steenman-Clark and Cole (2010) also presented results using similar ideas at the 2010 NCAS Workshop. In our investigation, the first point is to demonstrate memory contention by showing that the same model configuration runs faster on less than the full number of cores per node, spread across a larger number of compute nodes. Then the question is whether there are practical situations where it may be beneficial to use less than fully-committed nodes.

The impetus for this study came from the development of the next version of the Bureau of Meteorology's operational Australian Region 3-day numerical weather prediction system, i.e. the 12-km APS1 ACCESS-R system, with 1088 x 746 x 70 grid (see Puri *et al*., 2010, for ACCESS details). In order to fit within operational deadlines, this model system needs to be run in a 120 min time window from the observational cut-off time to delivery of products to Bureau forecasters. The initial version was taking 20-25 min too long; several scheduling and computational performance improvements were made, reducing this time below 100 minutes; the usage of 6 instead of the maximum available 8 cores per node for the major model computational steps was responsible for 12-15 min of this reduction.

Memory bandwidth for Nehalem processors

Results of this paper were produced on the National Computational Infrastructure (NCI) National Facility (NF) Oracle/Sun Constellation Cluster. The system has 1492 nodes containing two quad-core 2.93 GHZ Intel Nehalem X5570 CPUs with hyper-threading disabled. Each CPU contains 8 MB shared L3 cache, and the maximum memory bandwidth per CPU is 32GB/s (Intel Nehalem product sheet, 2009).

The stream benchmark (McCalpin *et al*., 1995) was used to measure the overall bandwidth achieved as a function of number of cores used (Figure 1).

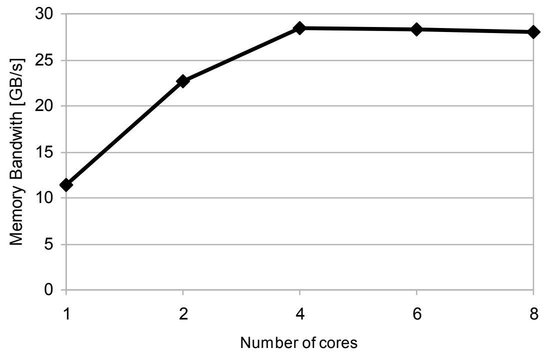


***Figure 1:** Memory bandwidth using stream benchmark (copy operation).*

The setting KMP_AFFINITY=scatter was used to give an even distribution of threads across sockets. The maximum measured memory bandwidth on a single core was 11.4 GB/s. With 2 cores this figure is doubled as both sockets are utilised, resulting in a memory bandwidth of 22.7 GB/s. The memory bandwidth increases by 20% with the use of 4 cores to 28 GB/s, but thereafter remains the same for 6 and 8 cores. These figures show that the memory bandwidth per core decreases for 4 or more cores (Figure 2).

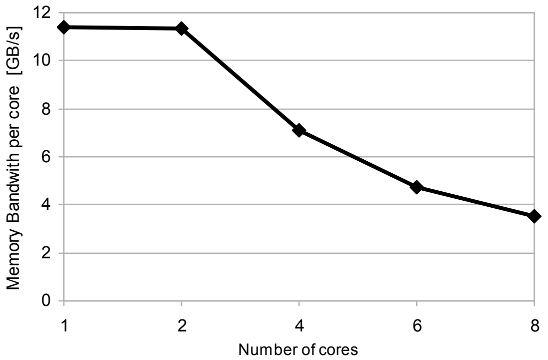


Figure 2 *Memory bandwidth per core using stream benchmark (copy operation).*

The question is: is it possible to decrease the overall runtime of an application by not using all cores on a socket? Leaving some cores idle provides several advantages:

1. It results in a higher memory bandwidth per core. This means that for memory bound applications the core appears to be faster.
2. The L3 cache is shared between all cores on a socket. Since we are distributing processes (which do not share memory) to each core, each core only has a fraction of the overall L3 cache available. If fewer cores are used then a larger fraction of L3 cache is available for each process.
3. By running fewer processes, there is also potentially more free memory available on a node, which the system can use for buffering IO.
4. It makes the application less susceptible to operating system jitter: the operating system needs a core to handle interrupts and other running processes (load sensors for the scheduler, probes to detect any hardware problems a node might have, etc.). If all cores of a node are used by the application, the application needs to be interrupted. Especially when using large number of nodes, those interrupts can have a significant impact on the overall performance. If cores are available for handling operating system tasks, there will be less interruption of the actual application.

UM speedup results and discussion

The application used here is version 8.0 of the UK Met Office Unified Model, known as the UM, at N512L70 resolution, a resolution which has been commonly used for UM benchmarks and for daily operational global numerical weather prediction. The UM model has been used as the ACCESS model (Puri *et al*., 2010) for operational forecasting in the Australian Bureau of Meteorology since 2009. The runs are 24 hour integrations with a 10 minute time step; the N512L70 resolution global model has grid size 1024x769x70. To avoid any issues with possible I/O contention, all the output of fields from the model was switched off. The model executable was built with the Intel 12.1.8.273 compiler and OpenMPI 1.4.3 library. The OpenMP multithreading implementation available with the UM8.0 source version has not been used in these runs. The MPI decomposition is based on horizontal domain decompositions in the latitude and longitude directions, with the model grid space divided into subdomains, where each subdomain contains a complete set of vertical levels and a rectangular horizontal subsection. The following decompositions from 96 to 3072 cores were used in the runs: 8x12; 12x16; 16x24; 24x24; 24x32; 30x32; 32x36; 32x42; 32x48; 36x48; 40x48; 44x48; 44x52; 48x52; 48x56; 48x60; 48x64. To get uniformly distributed cores on exclusive-use nodes for reduced cores-per-node the following mpirun command options

```
mpirun -bysocket -bind-to-core –npernode N …
```

were specified with the usage of N=6 and N=4 cores per node.

For each configuration, several runs were made on a fairly busy large multi-user system. Variations in elapsed times were mostly, but not always, small; we ignore these variations here, as they are not the focus of this study. In each case, the shortest elapsed times were used for our comparisons, as an estimate of the times which would be obtained on a dedicated system without interference. Some runs were repeated until consistent times were obtained.

Figures 3 and 4 show the speedups relative to the elapsed time achieved with 96 cores on fully committed nodes, as a function of number of used cores (Figure 3) and reserved cores (Figure 4), where the reserved cores include both the actual used cores and the reserved but unused cores in the 4 and 6 cores-per-node cases. Note that the number of reserved cores is increased by factor 4/3 compared to the used cores in the case of 6 cores-per-node, and it is doubled in the case of 4 cores-per-node.

Figure 3 shows that the model runs 11-18% faster on 6 cores-per-node than 8. On 4 cores-per-node a further 11-22% improvement is achieved, with the fastest runs being close to 16 times speedup for 4-core runs with 1920 cores (i.e. 20 times the number of cores for the 96-core run). The shape of the scaling curves is similar for all three curves. In the 6 and 8 cores-per-node runs the scaling limit has effectively been reached at 1920 cores, and there is only a marginal 10% speedup with 60% increase in cores from 1920 to 3072. So while it is not worth adding cores above 1920 when using fully committed nodes, the overall runtime can still be improved by using partial nodes. The 4-core runs were made up to only 1920 core decompositions, since these required greater total core counts to run.

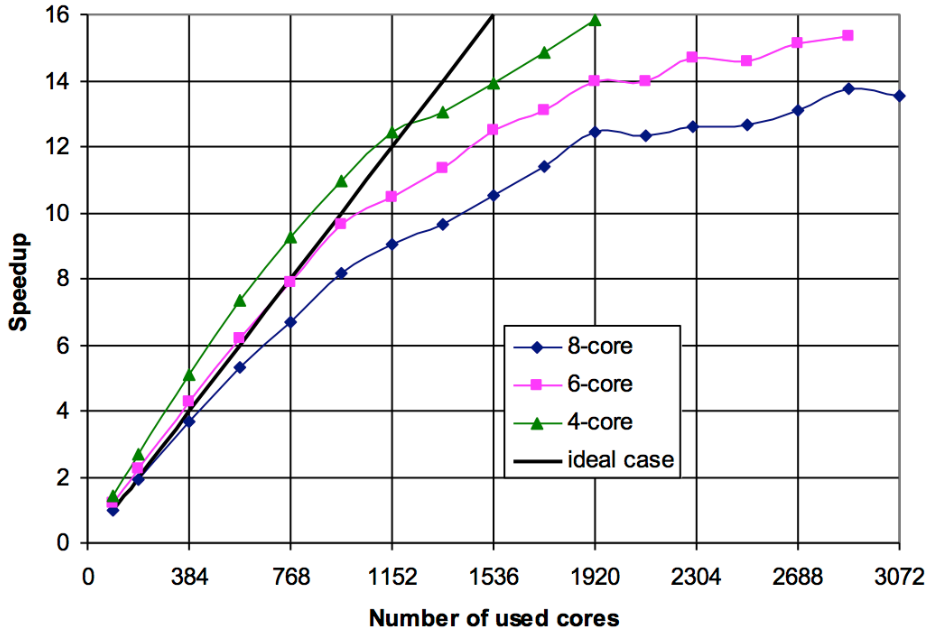


Figure 3: *Elapsed time speedup as a function of number of used cores. Reference point is 96 core run on fully committed nodes with an elapsed time of 3283 seconds.*

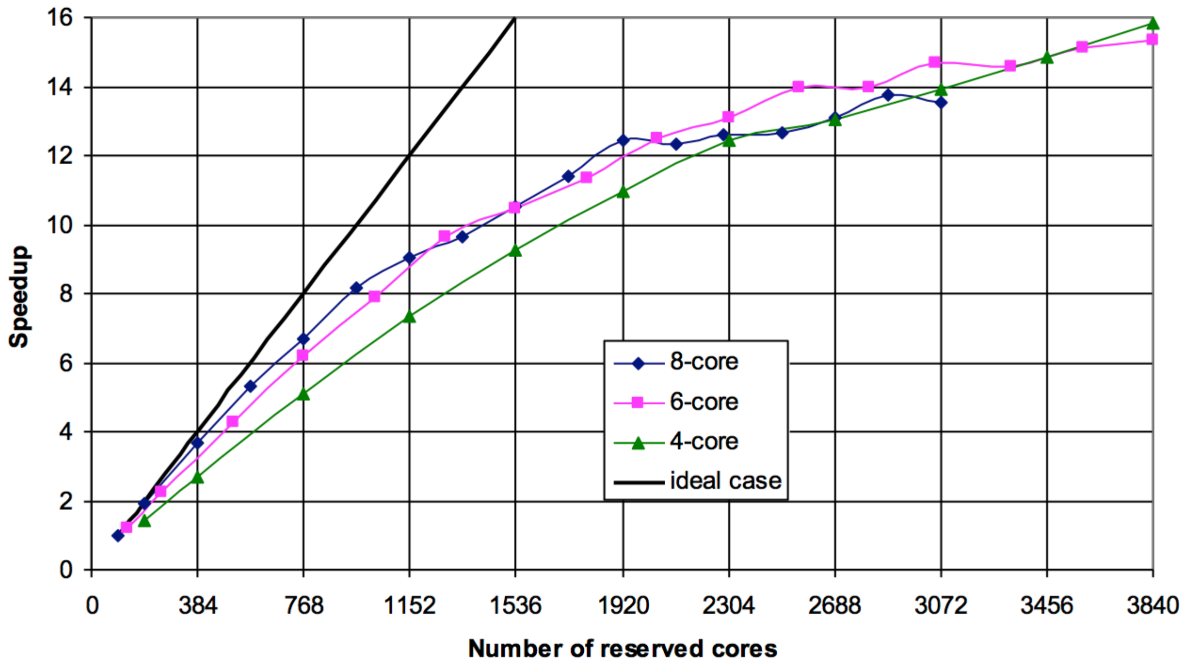


Figure 4: *Elapsed time speedup as a function of number of reserved cores. Reference point is 96 core run on fully committed nodes with an elapsed time of 3283 seconds.*

Figure 4 shows that the efficiency increase of the 6-core runs compared to 8-core is approximately commensurate with the overhead of reserving unused cores, but an overall gain at high core counts is achieved. Even when counting in all reserved cores the application's performance can be improved by using partial nodes. For 4-cores-per-node the speedup is clearly lower than 8- and 6-cores for the same number of nodes reserved through most of the range, until the scaling limit tail-off is reached. It appears that the cores are under-used overall with 4 cores, as the efficiency gained from further memory contention reduction is less than the capacity wasted by leaving cores idle. At the highest core-counts the 4-core run eventually achieves the fastest time, suggesting that memory contention is still present even past the point where the application has almost stopped speeding up with the use of more cores.

We would like to note that even 2 cores-per-node usage (not shown in the figures) also gave similar further improvement relative to 4 cores-per-node, but the improvement was relatively modest compared to the number of cores left idle to achieve it.

Another UM model case to mention is the example of smaller-size UM model applications. In those cases the ultimate speed of the model is limited not by the flattening off of the scaling curve, but by the MPI decomposition limits. One of the reasons for these application limits is that the model was developed in the mid 1990's when there were no requirements to have applications scaling beyond several hundred cores. Two cases are the ACCESS-C 4-km resolution "City-domain" weather forecast models, which have around 300 grid points in latitude and longitude directions, and standard climate resolutions such as N96, which has 192 x 145 grid points. The UM code imposes constraints on the sizes of halos for the semi-Lagrangian dynamics upstream departure point data exchange which limits the decomposition size to around 200-300 cores, which is still in the range where the application is scaling. In these cases there is also potential for application speedup by use of partial nodes, since indications are that memory contention is also present there too.

Conclusions

Running with reduced number of cores per node has been used to demonstrate the effect of memory contention on the computational performance of the UM numerical weather prediction model. This approach can be very useful in practice if the runtime of an application needs to be improved, but adding more cores is not efficient because the limit of scalability is reached, or not possible because of constraints in the application. It is less work than trying to optimise the application (e.g. making better use of cache to reduce necessary memory bandwidth, or overcoming limits in the application). We used this approach to save 12 to 15 minutes for the operational ACCESS-R forecast suite, which was taking too long for operational requirements.

In general, the benefits will depend on the relationship between the speedup per core when using fewer than the full number of cores per node and the scaling of the application. While the balance between processor and memory speed requirements will vary between applications, it is likely that many other applications will fall into this memory-intensive category for which the memory bandwidth will be the limiting factor affecting the application performance.

This approach for memory-bandwidth-intensive applications has been demonstrated to have practical benefits for both running in a fixed time window and maximising efficiency on the system for this application.

Acknowledgments

We would like to thank David Singleton (NCI) for providing useful comments during collection of the performance information on the NCI system, and Robert Bell (CSIRO) for helpful comments at the internal review stage.

References

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M.H., Staniforth, A., White, A. A. and Wood, N. 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **131**, 1759–1782.

Graham, S.L., Snir, M., and Patterson, C.A., 2005: Getting Up To Speed: The Future Of Supercomputing. *National Academies Press*, 289p.

Intel Nehalem Product Sheet, 2009: http://ark.intel.com/products/37111/Intel-Xeon-Processor-X5570-8M-Cache-2_93-GHz-6_40-GTs-Intel-QPI. Retrieved 29/11/2012.

McCalpin, J.D., 1995: Memory Bandwidth and Machine Balance in Current High Performance Computers. *IEEE Computer Society Technical Committee on Computer Architecture (TCCA) Newsletter*, December 1995.

Puri, K. *et al*., 2010: Preliminary results from Numerical Weather Prediction implementation of ACCESS. *CAWCR Research Letters*, **5**, 15-22.

Steenman-Clark, L., and Cole, J.: Coping with multi-core processors. NCAS Annual Meeting, July 2010. http://cms.ncas.ac.uk/index.php/component/docman/doc_download/349-lsc-talk-ncas-annual-meeting. Retrieved 10/12/2012