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 6, July 2011

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

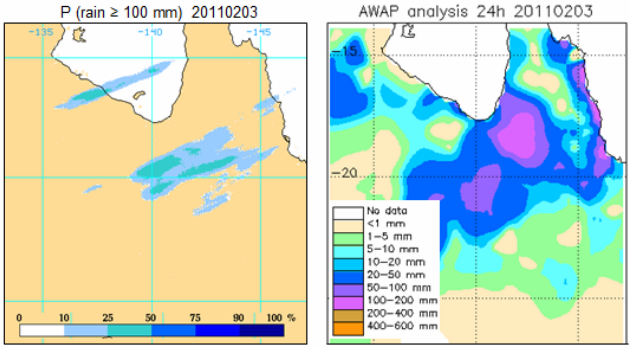
www.cawcr.gov.au

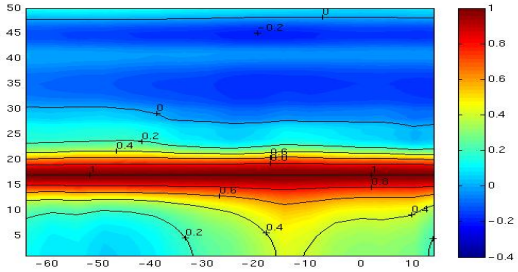
ISSN: 1836-5949

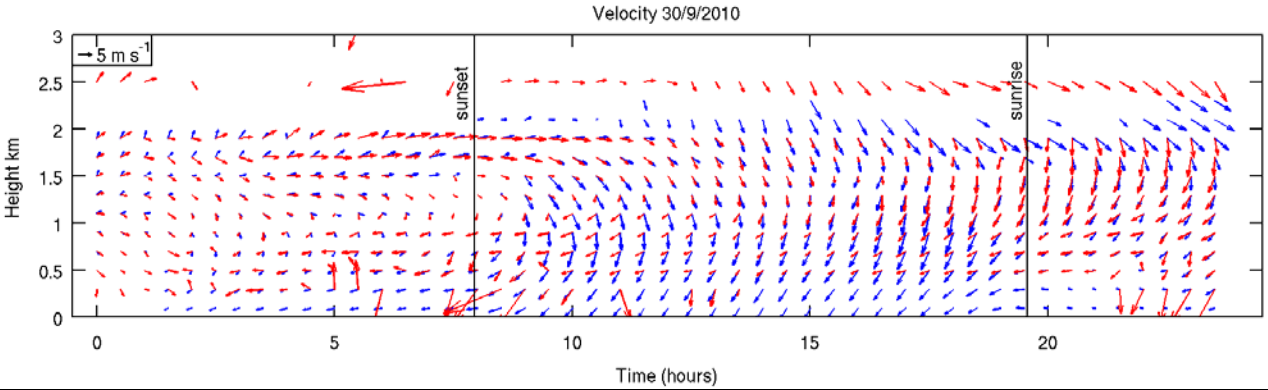
Series: Research Letters (The Centre for Australian Weather and Climate Research); Issue 6.

Copyright and Disclaimer

Contents







Editors: P.A. Sandery, T. Leeuwenburg, G. Wang, A.J. Hollis, 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.

Radius of reliability: A distance metric for interpreting and verifying spatial probabilistic warnings

Elizabeth E. Ebert

The Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

e.ebert@bom.gov.au

Introduction

An important function of national meteorological centres like the Bureau of Meteorology is to provide warnings of the impending danger of high impact weather including heavy rainfall, severe thunderstorms, and tropical cyclones. A difficulty in warning for rare events very far in advance is that their precise location and/or timing may be highly uncertain. To address this predictability issue meteorological centres are increasingly relying on ensemble prediction, which lends itself well to the generation of probabilistic forecasts. Probability forecasts provide users with quantitative uncertainty information that they can use to make more informed decisions, such as whether to take preventative action against storm damage, or even to evacuate.

More than a decade ago, the American Meteorological Society issued a statement promoting the value of probabilistic forecasts (AMS2002, updated in 2008). A well-formed probability forecast defines the event being predicted, its likelihood, as well as the location and period of time over which it is valid (for example, 30% chance of rain accumulation exceeding 10 mm between noon and 6 pm at Melbourne airport).

Some examples of probabilistic weather forecasts now issued by the Bureau include Terminal Aerodrome Forecasts of fog probability, and daily probability of precipitation exceeding various thresholds. In addition, the Bureau will soon begin issuing very short-range probabilistic forecasts of heavy rainfall from the Short-Term Ensemble Prediction System (STEPS; Bowler et al. 2006) and probabilistic nowcasts of thunderstorm impact using the Thunderstorm Environment Strike Probability Algorithm (THESPA; Dance et al. 2009).

The difficulty in pinpointing the location and timing of some kinds of high impact weather means that point probabilities are actually rather low, especially for longer lead-time forecasts. An issue with low probability forecasts of severe weather is that they may go unheeded by the public, even though the risk is non-negligible. To better convey the risk associated with severe thunderstorms the US Storm Prediction Center issues forecasts for the probability of severe thunderstorms occurring within a 40 km radius of each point in the warning area (Kain et al. 2006). This has the effect of increasing the forecast probabilities to a level where the public takes notice.

Convection-permitting ensembles are now being developed to predict high impact weather at short ranges (e.g., Gebhardt et al. 2008; Migliorini et al. 2011). In order to issue forecast probabilities that are likely to result in adequate response to weather risks, post-processing to apply the forecast to a wider area, and therefore inflate the forecast probability, is being tested (S. Theis, personal communication).

On longer time scales, major centres including ECMWF and the Met Office issue ensemble-based forecasts of tropical cyclone strike probability, which predict the likelihood of the tropical cyclone passing within 120 km of a given point in the next five days (van der Grijn, 2002).

Radius of Reliability

The applications mentioned above forecast the

probability of an event occurring *within some specified distance* of the point of interest. By considering a wider set of observations (e.g., at least one event occurring within 40 km), thus increasing the observed frequency, reliability of the inflated probability forecasts is maintained. A reliable forecast is one in which the observed frequency of an event is the same as its forecast probability over a large number of cases and for the range of forecast probabilities.

In some cases the forecast probability at a point may be systematically too high to begin with. This can occur, for example, in the early range of some ensemble forecasts, due to insufficient spread. To restore reliability a post-processing calibration step is often performed whereby excessive probabilities are reassigned to lower values, based on historical verification results. This reduces the frequency of forecasts with high probabilities, and may even eliminate high forecast probabilities altogether when the forecast skill is poor.

Ebert et al. (2011) investigated probabilistic forecasts of heavy rainfall in landfalling Atlantic hurricanes using a new satellite-based product called ensemble tropical rainfall potential (eTRaP). They found that the predictions were over-confident, and that this was due to insufficient spread caused by errors in the predicted location and pattern (rather than the amount) of the heavy rainfall. Calibration of the forecasts to achieve reliability at grid scale would have resulted in some heavy rain probabilities never exceeding 20%. Direct calibration was considered problematic because users might misinterpret the low grid-scale probabilities to mean that heavy rain is unlikely anywhere in the cyclone, when in fact the heavy rain may be very likely but its location is uncertain. In the case of tropical cyclone heavy rainfall a poor risk management decision could lead to dangerous outcomes if appropriate responses to the risks of flooding and landslides were not taken.

Ebert et al. (2011) asked the question: how far must one look to find observed heavy rain with the same frequency as the forecast probability at a point? Building on the "cone of uncertainty" concept used in tropical cyclone track prediction, where an area with 70% likelihood of enclosing the observed track is displayed on the warning chart (Broad et al. 2007), they defined a *radius of uncertainty* for interpreting probability forecasts at a point in space and time. Here we rename this quantity the *radius of reliability* (ROR), to emphasize the importance of reliability in its determination. For a given forecast probability P, ROR can be interpreted as the search radius around the forecast within which at least one observed event can be expected to occur P percent of the time. The ROR can be used to convey the spatial precision of probabilistic forecasts. It can also be viewed as a forecast quality metric, with lower values of ROR indicating more skilful forecasts.

A dense observation network or gridded analysis is needed in order to provide good spatial coverage of the event being predicted. For example, radar data is often used as the reference dataset for verifying rainfall and thunderstorm forecasts.

The iterative methodology for computing the ROR from spatial probability forecasts and observations is given below. Forecasts and observations should be mapped onto a common grid.

1. Create a binary observation grid (i.e., with observed "yes" and "no" values). Observations for continuous variables such as rain accumulation are assigned values of 1 or 0 depending on whether or not they meet the threshold of interest.

2. Spatially extend the binary observations to have a value of 1 within a circle of radius r. This new binary mask defines all grid boxes within distance r of a "yes" observation.

3. Matching the forecast probabilities with the extended binary mask from step 2, construct a reliability table by computing, for each probability value P, the frequency f of a forecast of P being matched with an observed value of 1. Binning of the probability values is necessary if the forecast gives continuous rather than discrete probabilities.

4. Increase r and repeat steps 2 and 3, saving the reliability table for each new value of r. Stop when $f \geq P$ for all values of P, or when r exceeds a search distance beyond which forecasts are no longer considered useful.

5. For each probability P, the ROR is the value of r for which $f = P$.

The determination of ROR as a function of forecast probability is illustrated in Figure 1, where the observed frequency is plotted against the forecast probability for a number of possible values of *r*, using data from Ebert et al. (2011). The diagonal line represents perfect reliability; the intersection of the curves of observed frequency with the diagonal determine the ROR for the associated probability *P*.

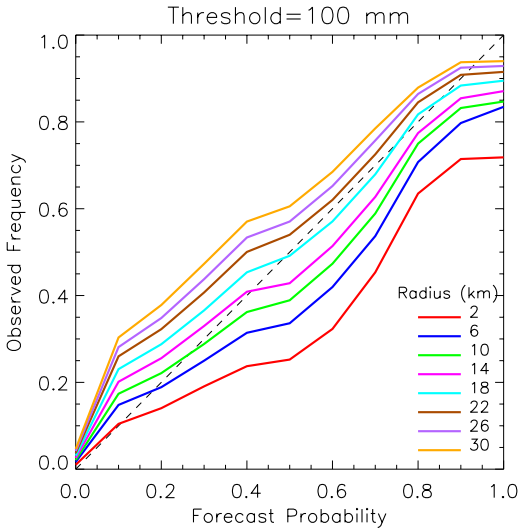


Figure 1 *Reliability diagram for forecast probability P of an observation occurring within radius r of the forecast value, for different values of r. Forecasts are for the probability of 24h rainfall exceeding 100 mm in land-falling Atlantic hurricanes between 2004-2008 (from Ebert et al. 2011).*

Table 1 *Radius of reliability as a function of forecast probability for the example shown in Figure 1.*

Forecast probability *P*	**Radius of reliability ROR (km)**
0.1	2
0.2	7
0.3	11
0.4	13
0.5	18
0.6	20
0.7	20
0.8	16
0.9	21

The corresponding values of ROR are given in Table 1. For example, for *P*=0.5 the curve corresponding to *r*=18 km crosses the diagonal. If the curve does not cross the diagonal, i.e., if the forecast is under-confident for all probability values, then the ROR is undefined since it is not possible to "shrink" the forecast area below grid scale.

While one can compute the ROR for a single gridded probability forecast, it is preferable to aggregate statistics over a large number of samples and then compute the ROR. This is because the forecasts in a single grid are not mutually independent. A large independent sample is required to obtain robust statistical results and to reduce sampling noise in the reliability diagram. Even for the five years of Atlantic hurricanes verified in Figure 1 and Table 1, the sample is not large enough to produce smooth curves and strictly monotonic behaviour.

This spatial approach to evaluating probability forecasts is similar in philosophy to neighborhood verification methods that are now being used to evaluate high resolution deterministic forecasts (Ebert 2009; Mittermaier and Roberts 2010). These methods evaluate forecasts in the neighborhood of a point against the observation at that point, or to the corresponding neighborhood of observations. By verifying forecasts for a range of neighborhood sizes it is possible to determine the scale for which forecasts attain a certain level of skill. In the case of ROR, the neighborhood applies to the observations rather than to the forecast.

Application to Probability of Heavy Rain in Australian Tropical Cyclones

We extend the study of Ebert et al. (2011) to investigate the reliability of eTRaP forecasts for 24h rainfall accumulation in landfalling tropical cyclones over Australia. Reference data comes from the Australian Water Available Project (AWAP) daily rain gauge analyses, which interpolate 24h rain gauge observations made at 9 am local time onto a 0.05° grid over Australia. A successive-corrections method is used to interpolate relative rainfall anomalies, which are then multiplied by the topography-representing climatological field to get optimal daily values (for details see Jones et al. 2009)

ETRaP forecasts were routinely issued by NOAA/NESDIS for all named tropical storms and cyclones around the globe, starting in 2009. Nine eTRaP forecasts for Australia valid at 00 UTC were available to be verified between

December 2009 and February 2011, as listed in Table 2. Note that 00 UTC is within one hour of 9 am LST over all of Australia during summer. Verification was performed at the 4 km scale of the eTRaP product.

Table 2 *24h eTRaP forecasts verified over Australia during 2009-2011.*

Tropical cyclone	Dates
Laurence	15, 20, 21 December 2009
Olga	23, 27 January 2010
Anthony	30 January 2010
Yasi	2, 3 February 2011
Carlos	16 February 2011

An example of a 24h eTRaP forecast for the probability of rain exceeding 100 mm in TC Yasi starting 00 UTC 3 February 2011 is shown in Figure 2. An area with 25-50% probability of exceeding 100 mm was forecast to the southeast of the Gulf of Carpentaria. The heaviest rain was observed about 100 km to the north of where it was predicted to be most likely. This location error is in line with the values found by Ebert et al. (2011), and is attributed to errors in the predicted cyclone track and the evolution of the rain pattern.

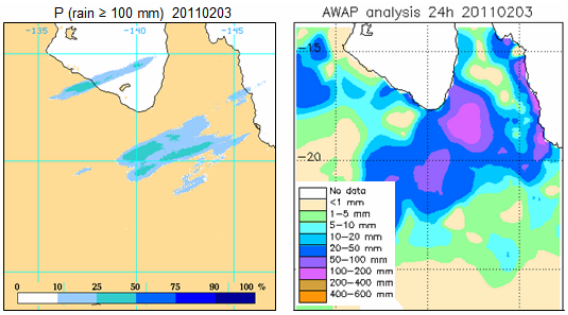


Figure 2 *ETRaP forecast for the probability of 24h rain in TC Yasi of at least 100 mm on 3 February 2011 (left), and the AWAP analysis for the same period (right).*

The radius of reliability was computed for the nine Australian cyclone cases in Table 2, for four 24h rain accumulation thresholds: 50, 100, 150, and 200 mm. Radii out to 160 km were tested; a search distance beyond 160 km was considered no longer useful in a predictive sense. The results are given in Table 3.

An ROR of 2 km means that the forecast is reliable at grid scale (i.e., within 2 km from the centre of the 4 km grid box). The ROR is greater than 2 km for all but the lowest probabilities and thresholds. In general the ROR increases for increasing probability and increasing rain threshold, as a wider search becomes necessary to enclose heavier (and therefore rarer) rain observations. For high probabilities and large thresholds no ROR was found, as it would have exceeded the useful limit of 160 km.

Table 3 *Radius of reliability as a function of forecast probability for 24h eTRaP forecasts over Australia during 2009-2011, verified against AWAP daily gauge analysis, for four rainfall thresholds.*

P	ROR (km)			
	50 mm	**100 mm**	**150 mm**	**200 mm**
0.1	2	2	25	87
0.2	9	9	68	66
0.3	16	60	101	---
0.4	34	48	62	---
0.5	52	81	---	---
0.6	46	---	---	---
0.7	62	---	---	---
0.8	---	---	---	---
0.9	---	---	---	---

Although the small sample size means that these results are subject to considerable uncertainty (as reflected in the non-monotonic increase of ROR with forecast probability), the performance of eTRaP forecasts for Australian tropical cyclones appears to be poorer than for Atlantic hurricanes (Table 1). It is unlikely that the cyclone track errors differ enormously between basins, or that the predicted rain patterns would be much worse in the Australian region. A more probable explanation relates to the differences in the verification data. The US study used gauge-corrected radar observations that contain much greater spatial detail than the fairly smooth gauge analyses available in Australia.

The AWAP analysis provides a spatially complete field but since it is an interpolation of sparse gauge data it underestimates the frequency of the heaviest rain (Beesley et al. 2009). The estimated RORs are therefore likely to be excessive, and caution must be taken when interpreting these results. More robust spatial verification may become possible with the

introduction of rainfall analyses that blend gauge data with remotely sensed radar and satellite precipitation estimates (Seed 2011; Renzullo et al. 2011).

An attempt was made to verify the eTRaP forecasts against observations from the rain gauge network in tropical Australia. However, the network is not dense enough to confidently detect the occurrence of rain within a reasonable radius of a forecast grid box and the ROR values were nonsensical.

Discussion

The radius of reliability (ROR) gives the radius around a forecast point within which an observed event is expected to be found with a frequency equal to the forecast probability. The ROR converts a probabilistic forecast that is over-confident at point scale into a reliable one within a spatial area. It can be thought of as the spatial precision of probabilistic forecasts.

As a forecast quality metric, ROR reflects the forecast uncertainty due to errors in location as well as intensity and spatial structure of an event. Lower values of ROR indicate better quality, more reliable forecasts. It may be possible to decompose the error in forecast probability into contributions from location, intensity, and pattern errors, similar to the contiguous rain area (CRA) approach for deterministic forecasts described by Ebert and McBride (2000). This will be pursued in a future study.

As a calibration tool, ROR can be used to suggest an appropriate spatial scale for issuing probabilistic forecasts. It is normally preferable to calibrate biased probabilistic forecasts to be reliable at the scale at which they are given, i.e., at point or grid scale. However, in the case of dangerous weather the calibration may have unintended and undesirable consequences if the calibrated probabilities are too low to spur protective actions by users of the forecasts. When heeding a warning is important, a spatial interpretation of probability forecasts can lead to higher probabilities that are more likely to cause people to take notice. Moreover, a probability forecast valid for an area is related to the likelihood of disruption at points within that area that may not experience the event directly (e.g., local flooding cutting off roads, power lines down, etc.).

The decision to act on a forecast of dangerous conditions depends on many factors including

- vulnerability (related to location)
- expected loss (related to magnitude)
- range of plausible actions (related to the lead time of the forecast)
- the probability of the event occurring

(Suarez and Tall 2010). The rational decision to act based on a forecast probability is informed by the cost-loss ratio. If the losses incurred by the failure to act are sufficiently high relative to the cost to protect, action may be taken based on quite low forecast probability. However, subjective decisions are not always made in a rational manner – they also depend on a variety of psychological factors, including the numbing effect of false alarms (e.g., Roulston and Smith 2004). The Met Office uses a 60% "chance of disruption" criterion to issue early warnings (Mylne and Legg 2002), reasoning that the public will accept a false alarm ratio of 40%.

Given a probability for which the public is likely to "take notice", then the ROR (search distance) that provides reliable forecasts for that forecast probability can guide the choice for the radius over which a warning applies. For example, if a probability of at least 60% was needed before most people would respond to a forecast of daily rainfall exceeding 100 mm in a land-falling Atlantic hurricane, then warning for heavy rain expected somewhere within a radius of 20 km (from Table 1) might be a good strategy.

In order to tease out this issue in regard to eTRaP forecasts in particular, the NOAA NESDIS Satellite Applications Branch surveyed a number of eTRaP users worldwide to ask their preference for how probabilistic heavy rain forecasts should be presented (M. Turk, personal communication, 2011). The choices were:

(a) Give the probability of exceeding a certain rainfall accumulation somewhere within, say, 40 km of the grid box of interest. This approach would be less precise as far as location, but would give a better picture of the overall risk of heavy rainfall in a general area.

(b) Give the probability of exceeding a certain rainfall accumulation for each specific grid box. This approach would be more precise as far as

the location of the highest probability of heavy rainfall, but the probability values will never be all that high because it's so difficult to predict the exact location of heavy rainfall.

Of the 25 responses received, 12 preferred the first, area-based, option and 13 preferred the second, point-specific, option. One user selected both options, pointing out (correctly) that they are complimentary. SAB now plans to issue calibrated eTRaP forecasts both at grid scale and within a 40 km radius of a point (Stan Kidder, personal communication, 2011).

Conclusions

The concept of the radius of reliability (ROR) has been introduced and demonstrated as a verification approach that takes into account location error in the forecast. Smaller values of ROR indicate more reliable, more accurate forecasts. ROR is also a potential calibration tool for interpreting over-confident forecasts. However, the necessity of associating different ROR values with different forecast probabilities and rain thresholds makes it difficult to use in practice.

Therefore, we recommend that ROR be used mainly for verification of spatial probability forecasts in cases where a dense network of observations is available. In Australia, this technique could be used to verify high resolution probability of precipitation forecasts in the southeast and eastern parts of the country where gauge-calibrated and blended radar rainfall mosaics are now produced routinely (Seed 2011).

Acknowledgements

Mike Turk at NOAA NESDIS kindly provided the eTRaP survey results. Mike Turk and Clay Davenport provided the gridded eTRaP forecasts over Australia. Jeff Kepert and Michael Foley in CAWCR provided many comments that helped improve this paper.

References

American Meteorological Society, 2002: Enhancing weather information with probability forecasts. *Bull. Amer. Meteorol. Soc.*, **83**.

American Meteorological Society, 2008: Enhancing weather information with probability forecasts. *Bull. Amer. Meteorol. Soc.*, **89**.

Beesley, C.A., A.J. Frost, and J. Zajaczkowski, 2009. A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets. *18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009.*

Bowler, N. E., C.E. Pierce, and A.W. ***Seed***, 2006:***STEPS***: A probabilistic ***precipitation forecasting*** scheme which merges an extrapolation nowcast with downscaled NWP. *Q.J.R.Meteorol. Soc.,* **132**, 2127-2155.

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the ‘‘cone of uncertainty’’ in Florida during the 2004 hurricane season. *Bull. Amer. Meteor. Soc.*, **88**,651–667.

Dance, S., E. Ebert and D. Scurrah, 2010: Thunderstorm strike probability nowcasting. *J. Atmos. Oceanic. Tech.*, **27,** 79-93.

Ebert, E.E., 2009: Neighborhood verification of high resolution precipitation forecasts. *Wea. Forecasting*, **24**, 1498-1510.

Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Ebert, E.E., M. Turk, S.J. Kusselson, J. Yang, M. Seybold, P.R. Keehn, R.J. Kuligowski, 2011: Ensemble tropical rainfall potential (eTRaP) forecasts. *Wea. Forecasting*, **26**, 213-224.

Gebhardt, C., S. Theis, P. Krahe, and V. Renner, 2008: Experimental ensemble forecasts of precipitation based on a convection-resolving model. *Atmos. Sci. Lett.*, **9**, 67–72.

Jones, D.A., W. Wang and R. Fawcett, 2009: High quality spatial climate data-sets for Australia. *Austral. Meteorol. Oceanogr. J.*, **58**, 233-248.

Kain, John S., S. J. Weiss, J. J. Levit, M. E. Baldwin, D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181.

Migliorini, S., M. Dixon, R. Bannister and S. Ballard, 2011: Ensemble prediction for nowcasting with a convection-permitting model—I: description of the system and the impact of radar-derived surface precipitation rates. *Tellus*, in press.

Mittermaier, M. and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354.

Mylne, K. R., and T. P. Legg, 2002: Early warnings of severe weather from the ECMWF ensemble prediction system. *Preprints,16th Conf. on Probabi-*

lity and Statistics in the Atmospheric Sciences, *Orlando, FL*, Amer. Meteor. Soc., J254.

Renzullo, L.J., A. Chappell, T. Raupach, P. Dyce, L. Ming and Q. Shao, 2011: An assessment of statistically blended satellite-gauge precipitation data for daily rainfall analysis in Australia. *Proc. 34th Intl. Symp. Remote Sensing of Environment, Sydney, Australia, 10-15 April, 2011.*

Seed, A., 2011: Short Term Ensemble Prediction System (STEPS): modelling the uncertainty in rainfall radar estimation and prediction. *WIRADA Science Symposium, 1–5 August 2011, Melbourne, Australia.*

Roulston, M.S. and L.A. Smith, 2004: The boy who cried wolf revisited: The impact of false alarm intolerance on cost–loss scenarios. *Wea. Forecasting*, **19**, 391-397.

Suarez, P. and A. Tall, 2010: Towards forecast-based humanitarian decisions: Climate science to get from early warning to early action. http://www.humanitarianfutures.org/sites/default/files/Towards%20forecast-based%20humanitarian%20decisions%20-%20Discussion%20Paper.pdf

van der Grijn, G., 2002: Tropical cyclone forecasting at ECMWF: New products and validation. *ECMWF Tech. Memo.386*,ECMWF, Reading, UK.

GENESIS: forming an ACCESS-SCM testbed from NWP analyses

Vaughan J. I. Barras and Yuelong Miao

The Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

V.Barras@bom.gov.au

1. Introduction

Single Column Models (SCMs) are useful tools in the development and testing of the various components of General Circulation Models (GCMs). An SCM operates as a single grid column running independently from its parent GCM providing an efficient framework to test its parameterization schemes whilst retaining full access to the model physics. An SCM is driven by supplying it with forcing profiles representative of the large scale. These may be applied to the model by revelation (direct calculation from observations), advective forcing (modifying predicted profile with vertical advection), or relaxation to model profiles over a set timescale (*Randall and Cripe, 1999*). The total tendency of a forcing quantity (ϕ) is described using summation notation as

$$\frac{\partial \phi}{\partial t} = \left(\frac{\partial \phi}{\partial t}\right)_P - \left(\varepsilon_{mn3}\mathbf{V}_m \bullet \nabla\phi + \delta_{m3}\mathbf{V}_m \frac{\partial \phi}{\partial p}\right) + \left(\frac{\phi_B - \phi}{\tau}\right) \quad (1)$$

On the right hand side the first term represents the change due to the model physics, the second term is the change due to the large scale horizontal and vertical advection tendency and the third term is a relaxation to a prescribed background profile (ϕ_B) over a set timescale (τ). By providing the model with forcing conditions representative of the large scale, an SCM may be used to simulate a particular scenario of interest such as a clear sky diurnal cycle or the development of boundary layer stratocumulus. A number of international model intercomparison projects have made use of SCM modelling in this way (*Duynkerke et al., 2004; Lenderink et al., 2004; Cuxart et al., 2006, Svensson and Holtslag, 2007; Wyant et al., 2007*.)

As an enhancement to the ACCESS-SCM, a software tool entitled GENESIS was developed to derive large scale forcings for the model from Numerical Weather Prediction (NWP) output. GENESIS enables a model developer to create an SCM testbed simulation at a desired latitude and longitude based upon a real-world scenario. Typically, SCM studies have been driven by forcings derived from Intensive Observation Periods (IOPs) designed to investigate a particular atmospheric process. The strength of this approach has been the high resolution and relative precision of the observations provided to the model. Despite this, as *Neggers et al. (2010)* point out, the selective nature of the case studies chosen for intense observation may not necessarily reflect a typical climatology of a particular process nor might they adequately target the most significant weaknesses in the model.

The key then is to identify a number of scenarios where models are known to have difficulty and use these as a basis for a process-based evaluation of the relevant model parameterizations. Whilst IOPs provide good coverage and quality data for a limited range of scenarios, the use of GENESIS enables the analysis of a vast number of targeted scenarios with the limitation being upon the resolution and quality of the input data. The question then is: can NWP forcing data sufficiently represent the detail of its parent 3d model simulation, and how can this method be used to enhance process-based model development in ACCESS? We consider here these questions in an investigation of an un-forecasted fog event at Perth Airport during the spring of 2010.

2. Data and study region

2.1 GENESIS

From standard NWP output on pressure levels, GENESIS read fields of surface level pressure (SLP), geopotential height (z), zonal wind (u), meridional wind (v), temperature (T), specific humidity (q) and vertical velocity (ω). GENESIS then used linear interpolation from the surrounding grid points nearest to the user-specified latitude and longitude (Perth Airport, Western Australia (WA), 31.9S 116.0E). This profile data was then interpolated in the vertical onto a Charney-Phillips grid and used as a large scale forcing in the SCM. In the event that the lowest NWP input level (typically 1000hPa) lay well above the lowest SCM levels moisture was kept constant, temperature was adjusted toward the surface according to a dry adiabatic lapse rate and momentum was scaled according to Monin-Obukhov similarity under the assumption of neutral stability conditions. Although substantially idealized, this approach was deemed sufficiently accurate for the purposes of this investigation.

2.2 Case study – Perth Airport, 24th September 2010

Fog was observed at Perth Airport (AP) during the early hours of September 25. During the peak of the event, visibility was reduced to 100 metres causing a significant disruption to local flight operations. The fog formed shortly after the passage of a weak cold front, drawing cool, moist air across southwestern WA. Wind speeds weakened overnight, gradually backing to the southeast and then fell to become light and variable under clear skies, conditions favourable for enhanced near-surface radiative cooling. A layer of fog then formed between 0237 and 0449 local time (1937 and 2149 UTC, 24/9/2010).

The combined infra-red (IR) and near infra-red (NIR) image from MTSAT-1R taken soon after the onset of the event shows that fog was widespread across the southwest of WA at the time (Figure 1). Forecasters found this event difficult to predict given that the prevailing winds were southeasterly. Typically under these conditions fog is unlikely to form given that the airport is located in the lee of the eastern escarpment. As a result, fog was not forecast until very shortly before its onset.

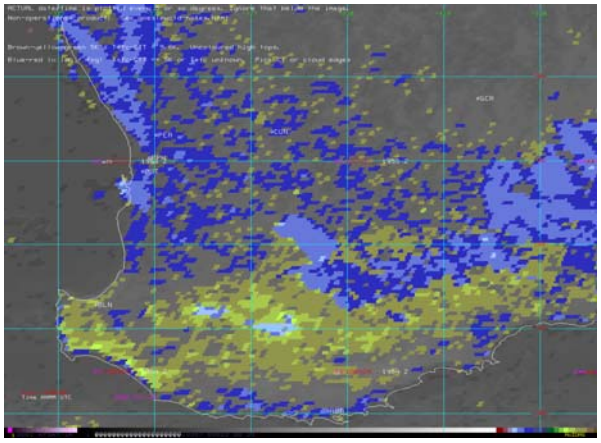


Figure 1 *Satellite fog analysis at 1953 UTC on 24 Sep 2010. The analysis was based on the brightness temperature difference of the infrared (IR) and near infrared (NIR) chancels of Japan's MTSAT-1R geostationary satellite. Spots or speckles shown on the imagery are considered doubtful signals. Different colour refers to the different confidence level of fog detection with the confidence level increasing from yellow to green, green to blue, and blue to light blue.*

2.3 ACCESS-A fog forecast

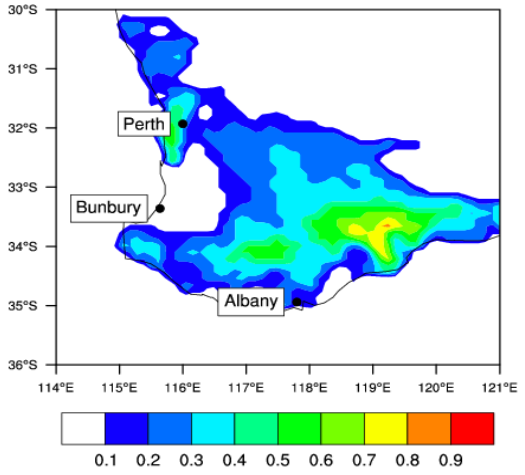


Figure 2 *The 20-hour prediction of fog fraction by the 0000UTC ACCESS-A model on 24 Sep 2010 for the southwest WA region. The fog fraction is equivalent to fog probability ranging from 0 to 1.*

Visibility in ACCESS is diagnosed from relative humidity, liquid water content and a fixed aerosol content. Subgrid-scale variability of visibility is predicted through a stochastic approach resulting in a probability product for a given visibility threshold. Hence fog fraction or fog probability is the areal coverage in percentage terms within a grid box that visibility is less than 1 km. Likewise the probability of mist is the areal coverage of visibility greater than 1km but less than 5 km. Clark et al. (2008) provided a comprehensive description of how visibility and fog fraction are predicted in the operational Met Office Unified Model that ACCESS-A is based on. Although unavailable to forecasters at the time, the experimental field of probability of fog formation in ACCESS-A predicted a moderate to high probability of fog across an extensive region of southwestern WA (Figure 2) occurring at 2000 UTC with a locally high probability of formation over Perth. ACCESS-A appears to have simulated well this event and may provide useful guidance in the future. For the purposes here however, it is of interest to test the extent that the SCM can capture the detail of the event driven by ACCESS-A analyses.

3. Results

3.1 ACCESS-SCM testbed simulation

The SCM (version 7.5) was run with 38 vertical levels for 24 hours at 2 minute timesteps using GENESIS forcings derived from 6-hourly ACCESS-A analysis profiles of zonal and meridional wind speed, temperature, specific humidity, omega and geopotential height. The SCM was run in idealized mode with radiation deactivated. Surface temperatures were therefore prescribed from the regional model to provide the surface forcing. The background geostrophic winds were set to zero.

The results show that even with the idealized nature of the large scale forcings, the SCM was able to reproduce well the basic features of the fog event. As indicated by the probability of fog formation diagnostic, conditions became favourable for fog formation from 1200 UTC. However, while ACCESS-A took time to develop to a maximum at 2100 UTC, the SCM reached its peak probability much more rapidly (Figure 3a). The fog probability curves were comparable in magnitude.

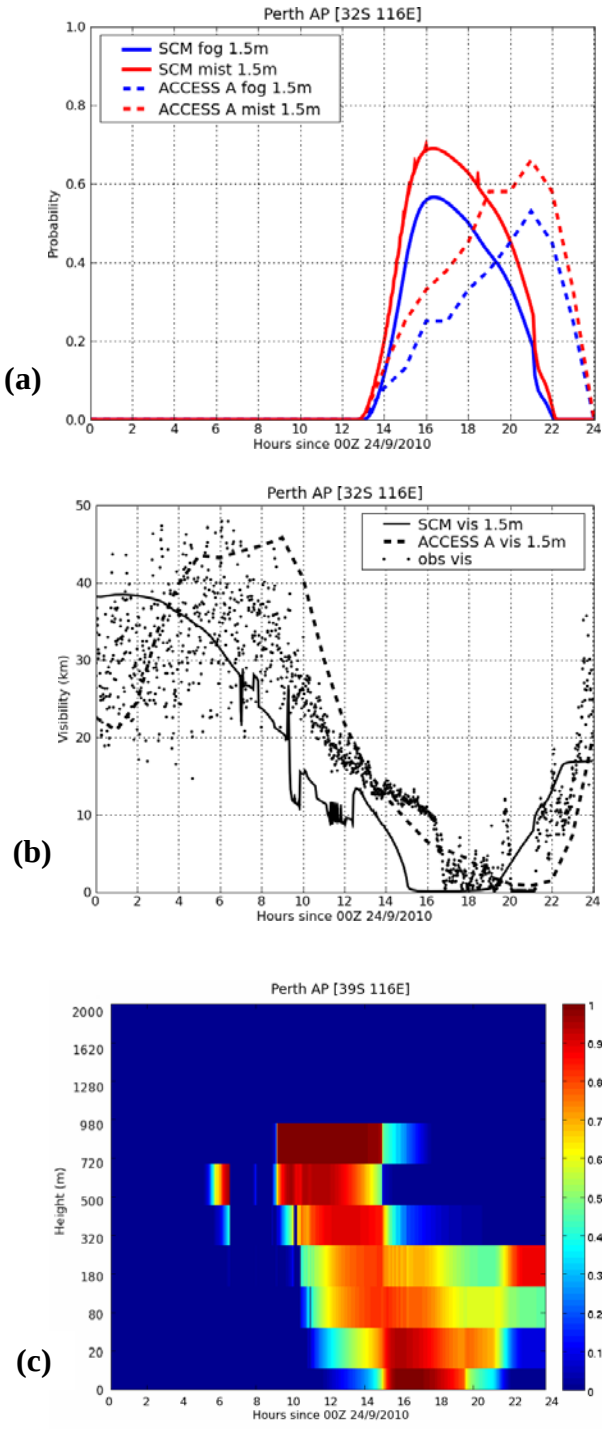


Figure 3 *Comparison between ACCESS-A forecast and ACCESS-SCM simulations of fog event at Perth AP, 24/9/2010. (a) Probability of fog and mist at 1.5m (b) Visibility (km) (c) Bulk low cloud fraction (SCM only). Note Y-axis is model theta levels and not a linear scale.*

In terms of visibility, the SCM once again appeared to develop fog and low cloud earlier than ACCESS-A. Visibility in the SCM fell consistently below 10 km from about 1400 UTC (Figure 3b), falling to zero between 1500 and 1900 UTC. ACCESS-A visibility fell below 10 km an hour later than the SCM, reaching a minimum of 817 metres at 2100 UTC. Corresponding visiometer observations recorded visibility below 10 km just after 1600 UTC and fell to zero briefly at 1800 UTC, and again from 2000 until just after 2100 UTC. A feature of the SCM simulation is that it formed cloud at low levels quite early. The cloud base then gradually lowered to the surface until fog was eventually formed (Figure 3c). Observations indicated that the formation mechanism of fog for this event was not due to a lowering of the cloud base as sometimes occurs overnight, but from clear sky radiative cooling. Celiometer measurements of cloud base recorded clear skies right up until the time of fog formation (Figure 4a). There is an indication here that although the model may be forming fog at the right time, the process of formation may not be correct.

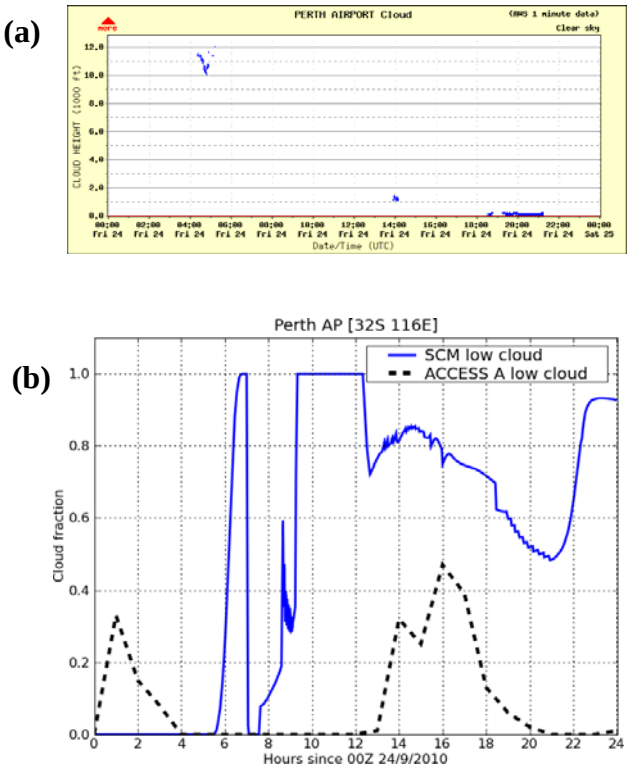


Figure 4 *Low cloud comparison for Perth AP (a) celiometer observations of cloud base height. Note height is in ft (m 0.3). (b) Low cloud fraction comparison for SCM and ACCESS-A.*

3.2 Model intercomparison

For the SCM simulation to be useful as a testbed for parameterization development it is important to check to what extent it replicates the behaviour of the ACCESS-A simulation. As described in the previous section, fog formation in the SCM was due to a lowering of the cloud base overnight. ACCESS-A exhibits similar behaviour forming cloud prior to the formation of fog indicating a common fog formation mechanism in both the idealized and full versions of the model (Figure 4b). Although the simulated mechanism may not be realistic, it is notable that the SCM is capable of replicating the general characteristics and processes of the ACCESS-A simulation with a relatively simple configuration of forcing. A feature common to the SCM is that it responds very strongly to the large scale forcings it is given, particularly with respect to moisture. This should be noted carefully during the preparation of an SCM testbed. The processes of interest should be known to the investigator in order to ensure that the model is forced in an appropriate fashion.

4. Conclusion

A software tool known as GENESIS was developed to derive large scale forcings for the ACCESS-SCM using NWP output. Using this approach, a large number of specifically chosen scenarios may be examined that specifically target key processes in the model. A fog event occurring at Perth Airport during the early hours of September 25, 2010 formed a basis to test whether the SCM forced in this way could sufficiently represent the detail of its parent 3D model (ACCESS-A) and to examine how this method may be used to enhance process-based model development in ACCESS.

It was found that even when supplied with a very basic set of large scale forcings, the SCM was capable of reproducing the general feature of interest seen in ACCESS-A. Although not identical to the full model, the timing and magnitude of the event were also comparable. Importantly, the SCM appeared to respond in terms of fog formation in the same way as ACCESS-A making it a useful framework in

which to establish a testbed for process evaluation and testing. From this point, several options are available to the developer. Firstly, similar analyses may be performed for a range of relevant case studies to verify a common shortcoming in the model parameterization. Secondly, once the testbed has been established, the model developer may then use this as a basis for testing of alternate model settings or parameterization schemes. Naturally, development of this kind can be done without needing to run the SCM. However the greatest benefit of GENESIS is the versatility it brings to the use of the model as well as the degree of customization of the event it puts in the hands of the developer.

Acknowledgements

We would like to acknowledge Peter Hurley for his support and contributions during the early development of GENESIS. We would also like to thank Greg Roff for his ongoing development of the code.

References

Clark, P. A., et al., 2008: Prediction of visibility and aerosol within the operational Met Office Unified Model. I: Model foremulation and variational assimilation, Q. J. R. Meteorol. Soc., 134, 1801-1816.

Cuxart, J., et al., 2006: Single Column Model intercomparison for a stably stratified atmospheric boundary layer, Boundary-Layer Meteorol., 118, 273-303.

Duynkerke, P., et al., 2004: Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case, Q. J. R. Meteorol. Soc., 130, 3269–3296.

Lenderink, G., et al., 2004: The diurnal cycle of shallow cumulus clouds over land: A single-column model intercomparison study, Q. J. R. Meteorol. Soc., 130, 3339-3364.

Randall, D.A., and D. G. Cripe, 1999: Alternative methods for specification of observed forcing in single-column models and cloud system models, Journ. Geophys. Res., 104, 24527-24545.

Svensson, G., and B. Holtslag, 2007: The diurnal cycle – GABLS second intercomparison project, GEWEX News, 17 (1), 9-10.

Wyant, M. C., et al., 2007: A single-column model intercomparison of a heavily drizzling stratocumulus-topped boundary layer, Journ. Geophys. Res., 112, D24204, doi:10.1029/2007JD008536.

Applications of latitude dependent covariance for ACCESS-R regional data assimilation

Xudong Sun and Peter Steinle

The Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

x.sun@bom.gov.au

1. Introduction

The Australian Bureau of Meteorology (ABM) has recently introduced a new numerical weather prediction (NWP) system based on the Australian Community Climate Earth System Simulator (ACCESS, see Puri et al., 2010). Its data assimilation system is based on the UK Met Office 4dVAR (Rawlins et al. 2007). One of the issues faced by any operational NWP system is the latitudinal extent of the regional domain – from 65°S, near the Antarctic coast to 17°N. Whereas the UK Met-Office regional assimilation system only supports constant background error covariances (Version 23.4 and earlier). Such a constraint was not considered to be appropriate for the Australian regional domain. To reflect the statistical differences in short range NWP accuracy for tropical, continental and polar regions, regional latitude dependent background covariances are derived and applied to the data assimilation system. This background error covariance formulation has been integrated in the UK Met-Office VAR system (Version 24.3 and later) and now is now used operationally within the ABM. This paper will briefly discuss these developments and present some validation results.

2. Discussion

The derivation of ACCESS specific background error covariance statistics is similar to that used for UK global 4dVAR system. The procedure involves using the NMC method (Parrish and Derber, 1992) to calculate vertical covariances at every 5 degrees of latitude. This covariance data set comes from 24 versus 48h forecast differences. A variable matrix transformation can then be derived to include the latitudinal variation in the 4dVAR data assimilation scheme.

In a controlled variable transform environment, matrix analysis such as eigenvector decomposition and inner-product multiplication and vector rotation is applied to the covariance $\mathbf{B}_i^{\nu}$, (where i represents the number of latitude bands in the formulation and ν depicts vertical covariance), so that the covariance corresponds to transformed variables and it can readily be inversed. The matrix factor $\mathbf{M}_i$ is:

$$\mathbf{M_i} = (\mathbf{Z}\mathbf{B}_i^{\nu}\mathbf{Z^t}) \qquad (1)$$

Where $\mathbf{Z}$ represents the above mentioned matrix analysis process. In the data assimilation stage, $\mathbf{M}_i$ is then applied to the transformed variables so that latitude dependency can be accommodated. Further information on modeling of covariances within the UK 4dVAR system and the application of latitude dependant covariance to variational analysis can be found in above mentioned documentation and Lorenc et al. (2000), Bannister (2008).

Figure 1 shows a diagram of 5 different covariance statistical methods/options that can be used in our regional NWP suite (ACCESS-R). These methods can be explained as follows: Method 1 involves direct estimation from the average global covariance statistics without modification. Method 2 follows step 1, but the vertical modes (i.e.: eigenvector and rotation) are re-calculated over the regional domain. Clearly, $\mathbf{M}_i$ is not regarded as latitude-dependent. Method 2 is currently used in UK Met Office regional data assimilation. Because the covariance is averaged over the whole regional assimilation domain, Methods 1 and 2 are more suitable for smaller regional and city model scale data assimilation. Methods 3 and 4 are the initial steps of the ACCESS latitude-dependent covariance development. These options allow the latitudinal variation from the global covariance to be used in the regional data assimilation, where the corresponding latitude band of the covariance is directly copied from global covariance. Following Methods 3 and 4, method 5 is further developed by re-calculating the vertical modes using variable rotation and eigenvector decomposition. The

reason for re-calculating the modes is that for global covariance, $\mathbf{M}_i$ is calculated using the global domain. It is necessary that for the Australian regional model, $\mathbf{M}_i$ is derived from using the statistics within this domain.

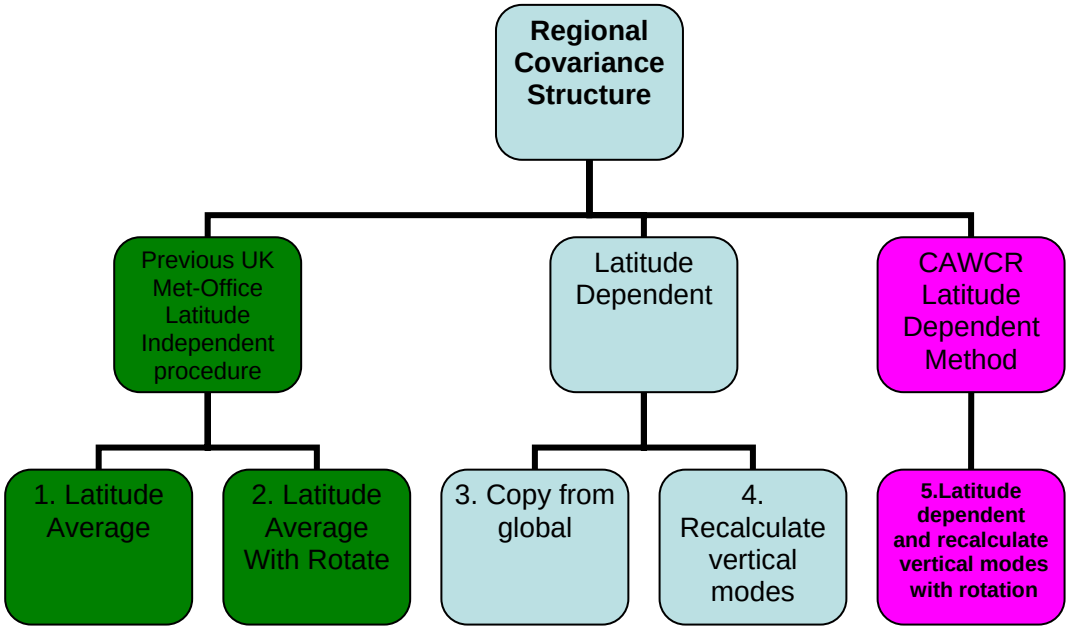
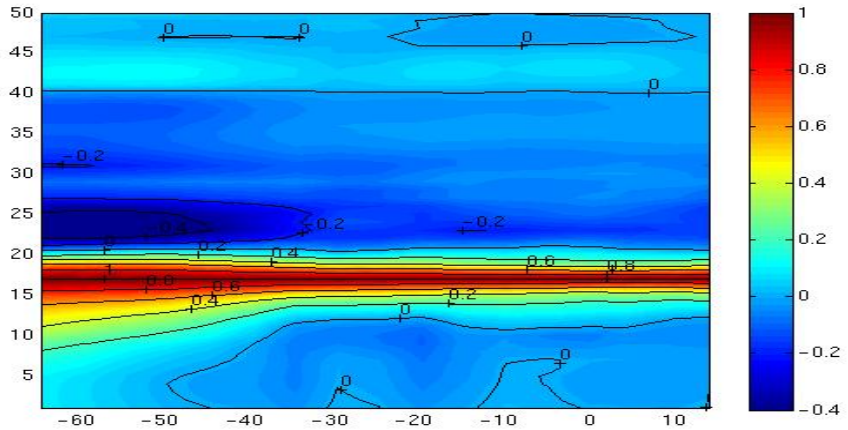


Figure 1 *Diagram showing 5 different methods/options used in ACCESS-R (Note that 5 is our main focus)*

Figure 2 gives vertical correlations of 500 hPa (corresponding to level 17) temperature and pressure background errors with each of its 50 levels from the ACCESS latitude dependent covariance file, derived from Oct. 2008 data. The latitudinal variation of the vertical structures is clearly shown, in particular the broadening of the temperature structures south of 40°S, with a negative lobes in lower stratosphere. Ingleby (2001) studied vertical correlation structure using Jan., 1999 global Met-Office 3dVAR model data. It shows similar pattern when compared to Figure 2(a). Similar results can also be found in Rabier et al. (1998). For tropical vertical forecast error covariance estimates, little detailed research is available. Some of the difficulties involve the strong seasonal variations of the statistics and the lack of "truth" fields. Ingleby (2001) has also shown that in the tropics, the vertical correlation length scale is generally small comparing to mid-latitude for temperature and wind; while tropical length scale is larger for mass (pressure). This is consistent with our results in Figure 2. This can be partly explained by the differences in geostrophic balance over the tropical and mid-latitude area.

(a)



(b)

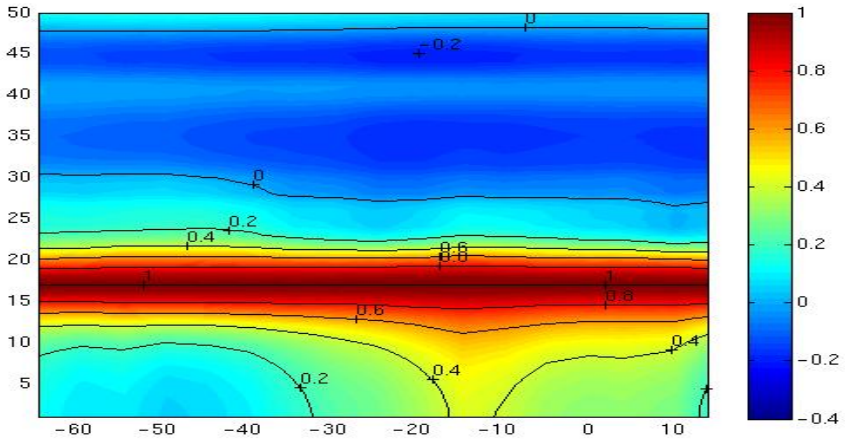


Figure 2 *500hPa temperature(a) and pressure (b) vertical correlations with each of its 50 levels from our covariance files. In the Figure, Vertical axis is model level, horizontal axis is latitudes from -65°S to 18°N. The scale bar represents correlation values.*

This section has discussed the nature of the latitudinal variation in background error covariances and how this can be incorporated into the ACCESS variational assimilation scheme. The justification of using latitude dependent covariance is further discussed in Rabier (1998). The next section investigates the impact of these covariances on NWP forecast skill.

3. Testing and Verification

The sensitivity of the ACCESS 4dVAR system to latitudinally varying background error covariances can be shown by using two artificial observations located at (146.5°E, 43°S) and (146.5°E, 22.5°S). Both assumed temperature increase of 3°C at 850 hPa and observation time coincided with the analysis time. The analysis increments from assimilating these observations are shown in Figure 3 with latitude independent covariance (a) and latitude dependent covariance (b). The figures show the corresponding wind increments for a cross-section along 146°E. In near-tropical areas the observations give similar increments for both (a) and (b), while higher correlation at 43°S gives higher wind magnitude and larger impact area from the latitude dependent covariances shown in Figure 3(b) relative to latitude independent correlations shown in Figure 3(a). This is consistent with the contours in Figure 2(a) although not at the same level.

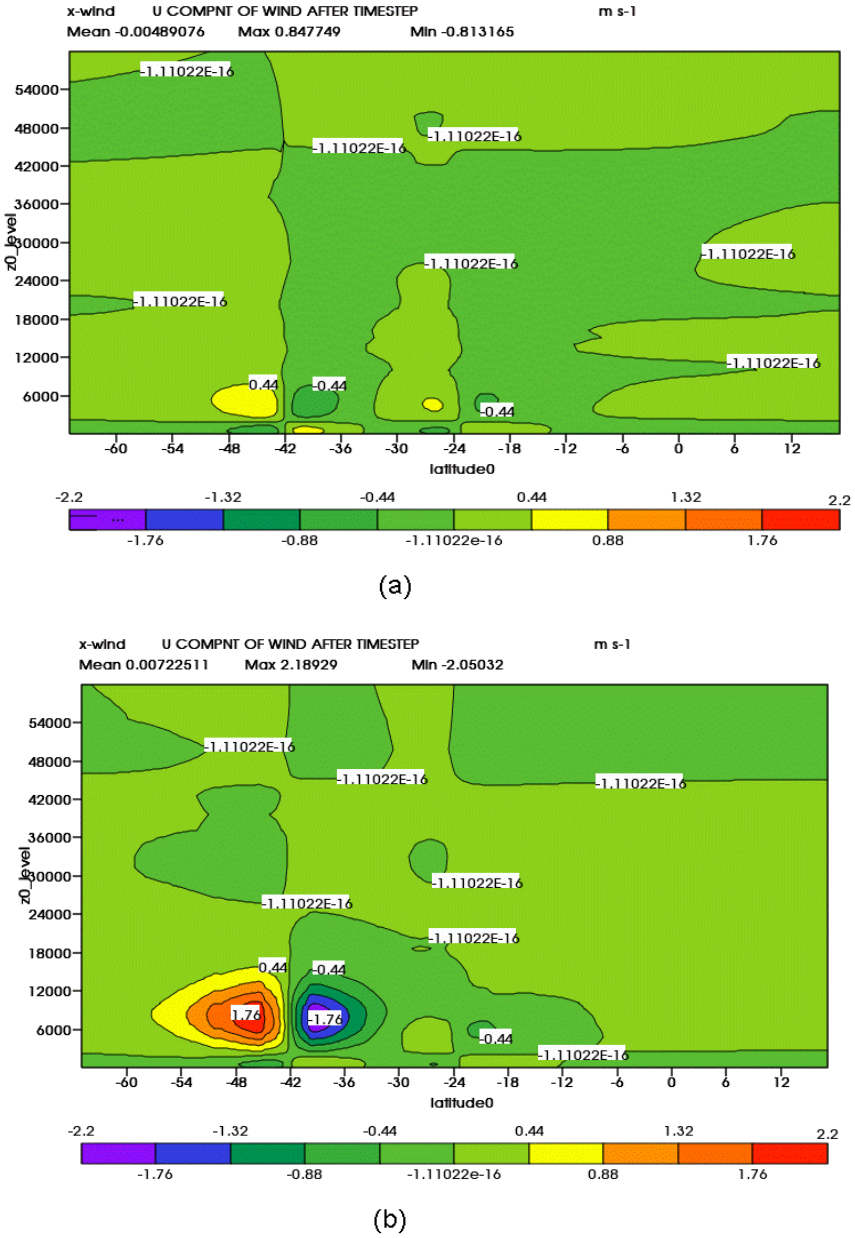


Figure 3 *Cross-section of U-wind increment with (a) Latitude constant covariance and (b) Latitude dependent covariance.*

Finally, to demonstrate the overall performance of using new latitude dependent background covariance statistics to the regional forecast, verification method is used to assess the NWP performance: the model forecast is validated by the radio-sonde observations. Here, we compare the results from methods 2 and 5 (Figure 1). The specific domain of our regional NWP suite (ACCESS-R) is 65.0°S to 17.125°N, 65.0°E to 184.625°E.

The resolution of the background is 0.375° and 50 vertical levels with a 6 hourly 4dVAR window centered on 00, 06, 12 and 18UTC. Figure 4 shows the vertical average geo-potential and wind Root Mean Square Error (RMSE). These results show some improvement for the RMSE; particularly in the upper atmosphere. At this level, 65-70% of the days in this period show a positive impact. However, there is little improvement below the 500 hPa level. Further research may be necessary for determining the causes and its inconsistency. Notice the verification results show no improvement at 2-day or longer forecasting time

scales. This is probably due to the lateral boundary conditions becoming the dominant influence on forecast error.

4. Conclusions

We have extended the background error covariance formulation that is used in the 4dVAR assimilation for the regional NWP (ACCESS-R) system by allowing for latitude dependence in background error covariances. The covariance statistics are calculated every 5^o of latitude. The validation results have produced more realistic forecasts compared to that of average covariance, particularly from 500 hPa to the top of tropopause.

Instead of generating regional latitude dependent covariance from global covariance, we have also derived regional covariance from regional model itself. In this way, the covariance statistics is better considering regional model has much high resolution than that of the global model. This research is continuing.

Acknowledgments

The authors like to thank the staff within the ACCESS team, in particular, Dr Yi Xiao, Dr Chris Tingwell and Jin Lee for their support and help in this project. Assistance from UK Met Office scientists are also greatly appreciated

References

Bannister, R. N., 2008: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. Quarterly Journal of the Royal Meteorological Society, **134**, 1971-1996.

Ingleby, N.B., 2001: The statistical structure of forecasterrors and its representation in The Met-Officeglobal 3-D Variational Data Assimilation Scheme.*Quart. J. R. Meteorol. Soc*., **127**, 209-231.

Lorenc, A. C.; Ballard, S. P.; Bell, R. S.; Ingleby, N. B.; Andrews, P. L. F.; Barker, D. M.; Bray, J. R.; Clayton, A. M.; Dalby, T.; Li, D., 2000: The Met. Office global three-dimensional variational data assimilation scheme, *Quarterly Journal of the Royal Met. Soc.*, Vol **126**, pp. 2991-3012.

Parrish, D. F., and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747–1763.

Puri, K. et al., 2010: Preliminary results from Numerical Weather Prediction implementation of ACCESS. *CAWCR Research Letters*. No. 5, 15-22.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, F. Bouttier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Journal of the Royal Met. Soc.,* Vol. 124, pp1783–2179

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 Met. Soc.,* Vol **133**, pp. 347-362.

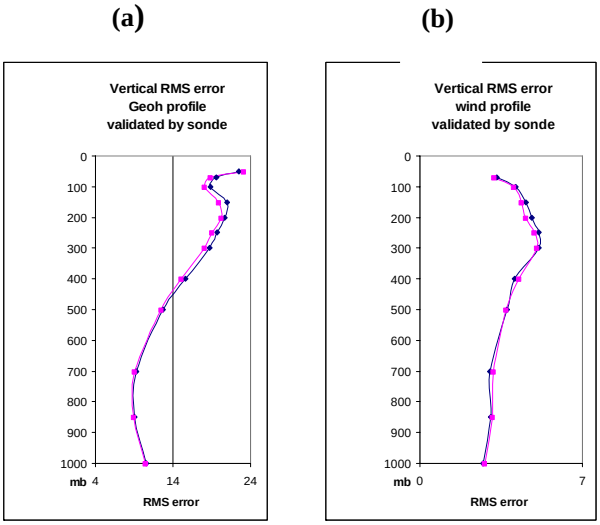


Figure 4 *ACCESS-R RMS Validation of 36 hour forecasting. (a) Geopotential height. (b) Wind velocity. (Data from Sept. 25 to Oct. 10. Red line is from latitude dependent scheme)*

Capping friction velocity over high topography in ACCESS

Vaughan J. I. Barras[A], Adrian Lock[B]

[A]The Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

[B]Met Office, Exeter, UK

v.barras@bom.gov.au

Introduction

The accurate representation of the effect of flow over topography upon surface fluxes of momentum, heat and moisture is an important issue in Numerical Weather Prediction (NWP). Owing to the relatively coarse resolution of global NWP models with respect to the horizontal scales of even major mountain ranges such as the Himalaya, parameterisations are required to approximate the effects of topography upon near-surface wind speed and areal average surface fluxes. These effects are commonly described in terms of a 'form drag' in which a region of locally high pressure exists on the windward side of a mountain range acting to enhance mixing relative to an equivalent homogeneous surface. A simple yet effective approach has been the representation of aerodynamic drag in terms of an adjustment to the roughness length, z_0 (Mason, 1985) commonly referred to as the 'effective roughness length'. Observation has found that even over mountainous terrain a logarithmic profile wind is applicable near the surface (Grant and Mason, 1990). As a result, a prediction of velocity at a particular height over complex terrain can be made by determining the value of the effective roughness length (Wood and Mason, 1993). We describe here the sensitivity of the present formulation of form drag in UM 6.4 over the high topography of the Himalaya (Figure 1).

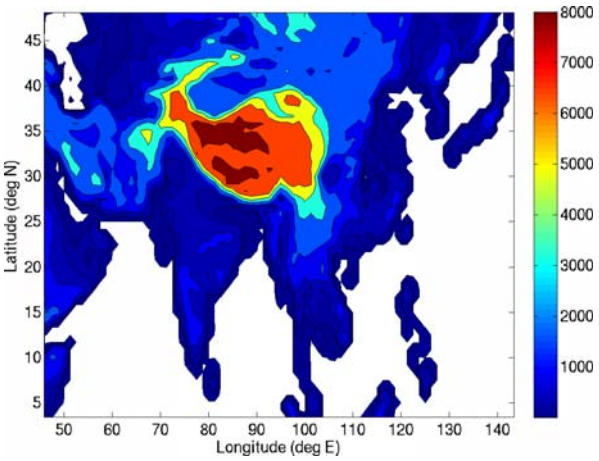


Figure 1 *UM topography of Central Asia. Contour spacing is 1, 50, 100, 200, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000 metres above mean sea level.*

Form drag in UM 6.4

The parameterisation of form drag in UM 6.4 makes use of an effective roughness length (z_0^{eff}). For momentum the formulation is based upon that of Wood and Mason (1993) and Milton and Wilson (1996). Scalar quantities follow the parameterisation of Hewer and Wood (1998). However, for the purposes of this study, only the parameterisation of form drag for momentum will be examined. Wood and Mason (1993) examined the parameterisation of form drag under neutral conditions only, describing the effects of topography in terms of z_0^{eff}. Milton and Wilson (1996) expanded this

parameterisation to incorporate a stability dependence such that:

$$\frac{\kappa^2}{\left[\ln\left(\frac{z_h}{z_0^{eff}}\right)\right]^2} = 0.5 C_D^{orog} \frac{A}{S} f_D(Ri_B) + \ldots$$

$$\ldots \frac{\kappa^2}{\left[\ln\left(\frac{z_h}{z_{0m}}\right)\right]} \qquad (1),$$

where z_0 is the homogeneous roughness length, z_0^{eff} is the effective roughness length, z_h is equal to $\sqrt{2}\sigma_h$ (standard deviation of the orographic height), κ is the Von-Kármán constant, C_D^{orog} is the orographic drag coefficient (= 0.3), A/S is the ratio of the large scale orographic variance against the silhouette area of the orography and $f_D(Ri_B)$ is a stability parameter which decreases from unity to zero for Bulk Richardson numbers ranging between $0 \leq Ri_B \leq Ri_{crit}$. For the parameterisation of form drag in UM 6.4, the relatively simple effective roughness length approach has been shown to be remarkably robust even under increasingly stable conditions (Belcher and Wood, 1996; Brown and Wood, 2003). However, the underlying assumption of the consistent structure of the boundary layer for homogeneous and inhomogeneous terrain is perhaps an oversimplification which can lead to inconsistencies in the predicted wind profile (Wood et al., 2001; Brown and Wood, 2003; Beljaars et al., 2004). Similarly, the stability dependence of z_0^{eff} is set according to a predefined critical value of $Ri_{crit} = 0.5$.

U_{star} diagnostic testing

The above formulation of form drag in ACCESS defines a stability dependence upon z_0^{eff}. Once the boundary layer stability increases beyond a critical threshold, the effective roughness length is equated with the roughness length of homogeneous terrain (z_0).

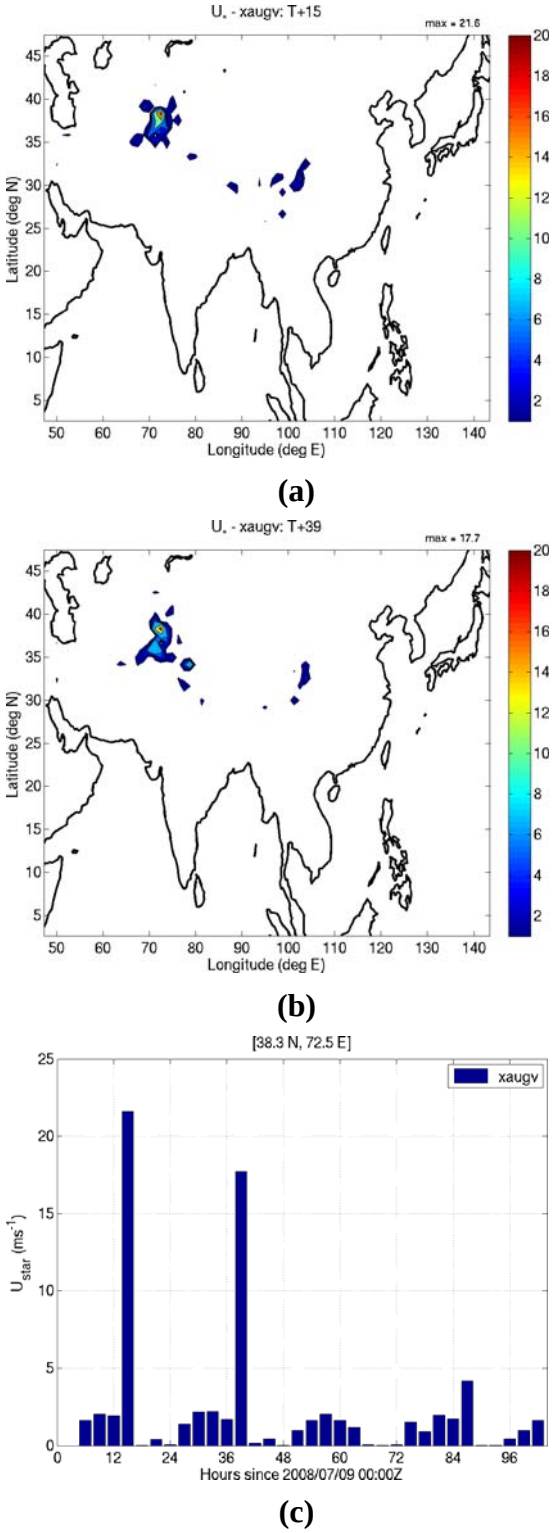


Figure 2 *U_* over continental Asia at timestep (a) T+15 and (b) T+39. Contour interval is 1, 2, 3, 4, 6, 10, 14, 20 ms^{-1}. (c) Timeseries of u_* at western Himalaya grid point.*

Physically, this represents the suppression of vertical motion by stratification such that any deepening of the boundary layer due to orographic form drag will become negligible.

This deactivation of the effective roughness length with increasing stability results in a generally acceptable behaviour of surface drag in the model. However, difficulties may arise in the region of very high topography during periods of transition in boundary layer stability. At these locations, z_0^{eff} can become particularly large (of the order ~50m) compared with a grassland surface roughness of ~0.02m. As a result, considerable changes can occur to the surface drag coefficient which affects the surface exchange coefficient (c_D) and the explicit calculation of the friction velocity (u_*) during the transition. In the UM, this explicit value is used for the u_* diagnostic such that

$$u_* = \sqrt{FC_D|\Delta v|^2} \qquad (2),$$

where F is the land tile fraction (= 1 in this case), C_D is the surface conductance ($c_D V$) where V is the effective wind speed for surface turbulent exchanges and $|\Delta v|$ is the magnitude of the wind shear between the surface and the lowest model layer. Figures 2a, and b illustrate the problem of large z_0^{eff} during an NWP 48 hour forecast initialised at 00Z on 09/07/2008. At T+15 and T+36 the boundary layer stability transitions from a well-mixed to a stable condition. During both periods a sharp peak in the value of u_* occurs over the western edge of the Himalaya where the topography rises sharply (see Figure 1). During the transition to stable conditions, the surface roughness decreases markedly with the disengagement of the effective surface roughness resulting in excessive wind shear and subsequent values of u_* (Figure 2c).

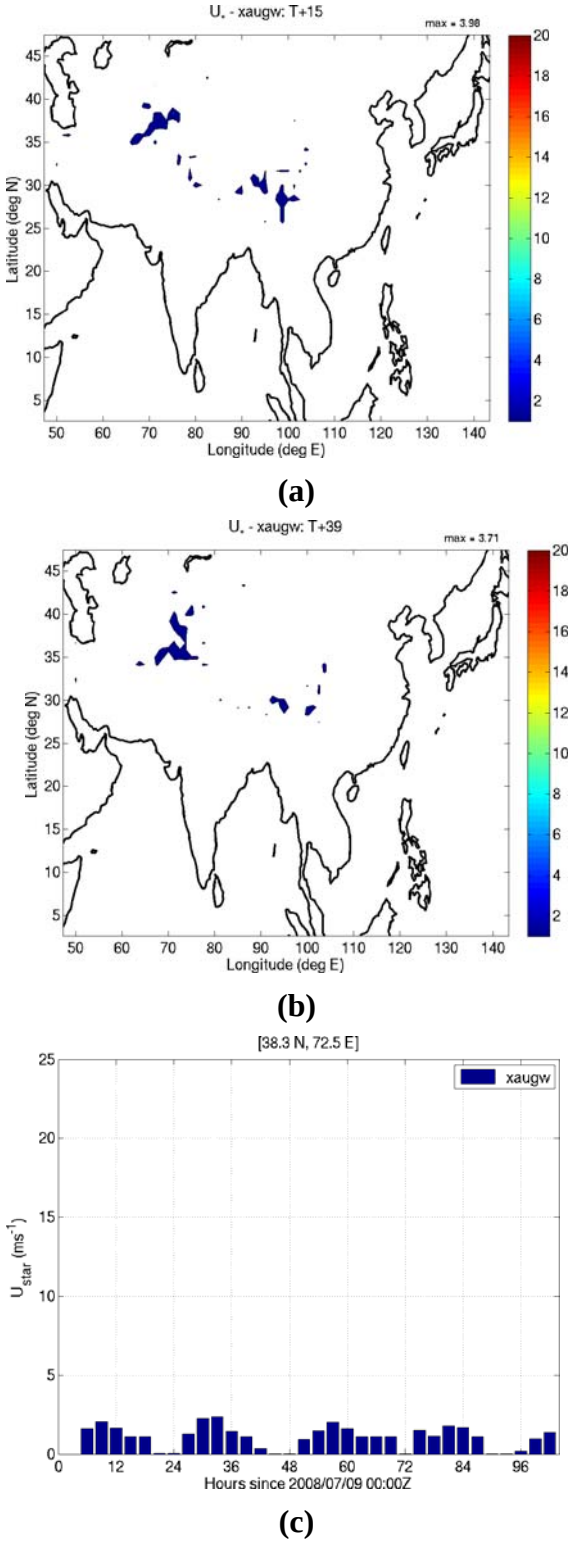


Figure 3 *Capped U_* over continental Asia at timestep (a) T+15 and (b) T+39. Contour interval as for Figure 2a, b. (c) Timeseries of capped u_* at western Himalaya grid point.*

Modification and testing of U_{star} diagnostic

In orderto overcome the problem of unrealistic values of u_* arising during stability transitions in the boundary layer, a capping threshold was implemented. Following the calculation of u_*, its value was compared against a maximum allowable surface stress value

$$u_* = \sqrt{z_h \tau'_{\max}} \qquad (3),$$

where z_h is the height above the surface of the boundary layer top, and $\tau'_{\max} = 0.05$ is an imposed maximum threshold stress gradient across the boundary layer. Maintaining reasonable values of u_* is important as it is also used in the calculation of the velocity scale for momentum in the non-local diffusion coefficient profile (Holtslag and Boville, 1993):

$$w_m = \left(u_*^3 + 0.6 w_*^3 \right)^{1/3} \qquad (4a),$$

where

$$w_* = \left(\left(\frac{g}{\theta_{v0}} \right) \left(\overline{w'\theta'_v} \right)_0 z_h \right) \qquad (4b),$$

such that $\left(\overline{w'\theta'_v} \right)_0$ is the surface buoyancy flux, θ_{v0} is the virtual potential temperature at the surface and g is the acceleration due to gravity. With the capping of u_* applied to an otherwise identical NWP forecast a substantial difference can be seen in the peak values over the western Himalaya (Figure 3a, and b) decreasing u_* to values below 4ms^{-1} (Figure 3c). The effect upon u_* at other timesteps is minimal.

Evaluation of implicit U_{star} calculation

Surface stress for PBL wind increments in the UM is given by calculating u_* such that

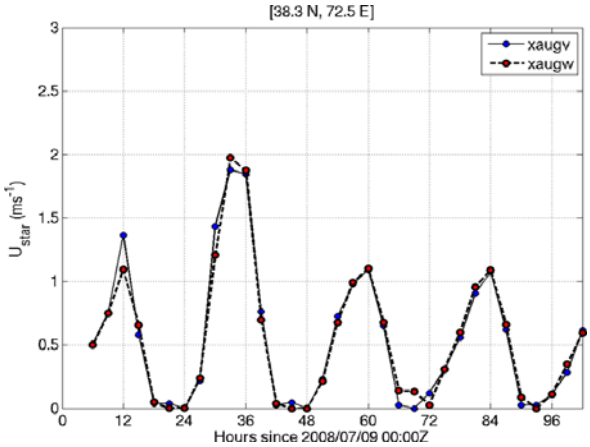


Figure 4 *Timeseries of implicit u_* values at western Himalaya grid point. Identifiers xaugv and xaugw denote standard and capped u_* forecast runs respectively. Note the scale on the y-axis.*

$$u_* = \left[\left(\frac{\tau_x}{\rho_0} \right)^2 + \left(\frac{\tau_y}{\rho_0} \right)^2 \right]^{1/4} \qquad (5),$$

where τ_x and τ_y are the respective zonal and meridional components of the surface stress and ρ_0 is the surface air density. The implicit calculation of the stress at the surface is:

$$\vec{\tau}_i = \rho_0 c_D \left(\gamma \, \vec{U}_{n+1} + (1-\gamma) \, \vec{U}_n \right) \left| \, \vec{U}_n \right| \qquad (6),$$

where ρ_0 is the surface air density and c_D is the surface exchange coefficient. Note the difference between this and the surface conductance C_D:

$$C_D = \frac{\kappa \, u_*}{\Phi_m} \qquad (7a)$$

$$c_D = \frac{\kappa^2}{\Phi_m^2} \qquad (7b)$$

It can be seen in (5) and (6) that by determining u_* using the surface stress (τ_i) directly, this calculation is less sensitive to changes arising from the (de)activation of z_0^{eff} across the stability threshold. Figure 4 shows some sensitivity of u_* about the stability thresholds, but nothing like the sharp changes seen in the explicit calculation (Figure 2c). Similarly, there is very little

change to the implicitly calculated u_* with the application of the capping threshold.

Conclusion

The existing formulation of orographic form drag in UM 6.4 has been investigated to test the impact of imposing a cap upon the value of u_*. The stability dependence of the effective surface roughness as outlined by Milton and Wilson (1996) has been shown to give unrealistically large values of u_* during stability transition of the boundary layer between well-mixed and stable conditions in regions of high topography. To address this, a maximum stress gradient across the depth of the boundary layer was imposed which greatly reduced the magnitude of the explicitly calculated value of u_* during these transition periods whilst having a minimal effect upon the values of u_* for the rest of the simulation. However, implicit calculations of u_* were shown to be generally unaffected by the problem and to the imposition of a capping threashold upon z_0^{eff}.

Whilst delivering a satisfactory outcome, the capping of u_* in the way described here is limited. Brown and Wood (2003) point out that although in general the area-averaged structure of boundary layer structure is largely indistinguishable between flat and undulating terrain (Grant and Wood, 1990; Wood and Mason, 1993) there remain important differences between the two, particularly in relation to the vertical wind profile.

Wood et al. (2001) and Beljaars et al. (2004) advocate a different approach to the parameterization of orographic drag by the specification of stress profiles rather than sole reliance upon effective surface roughness. Wood et al. (2001) approximate the surface drag as

$$\left(u_*^{eff}\right)^2 \approx u_{*0}^2 + \frac{F_{p\_x}}{S_d} e^{-z/l} \qquad (8),$$

where u_{*0} is the surface friction velocity (the non-orographic surface stress) and F_{p_x}/S_d is a parameterization of the pressure force exerted by the topography (Wood and Mason, 1993). Wood et al. (2001) add an exponential decay parameter where z is the height above the mean orographic height and l is a decay scale (see reference for details). As further work, this more explicit approach may also be tested as an alternative representation of orographic form drag, particularly with respect to the structure of the boundary layer under stable conditions. However, being a more explicit approach, it may be more susceptible to numerical problems with the relatively long timesteps in the NWP model.

Acknowledgements

The author gratefully acknowledges Ivor Blockley of the National Meteorological and Oceanographic Centre for the identification of the model error and the advice given by Les Logan, Asri Sulaiman and Zhian Sun of the Centre for Australian Weather and Climate Research during the setup and configuration of the model for this investigation.

References

Belcher, S.E., and N. Wood, 1996: Form and wave drag due to stably stratified turbulent flow over low ridges, *Quart. J. Roy. Meteor. Soc.*,**122**, 863-902.

Beljaars, A.C.M., A.R. Brown, and N. Wood, 2004: A new parameterization of turbulent orographic form drag, *Quart. J. Roy. Meteor. Soc.,* **130**, 1327-1347.

Brown, A.R., and N. Wood, 2003: Properties and parameterization of the stable boundary layer over moderate topography, *Journ. Atm. Sci.*, **60**, 2797-2808.

Grant, A.L.M, and P.J. Mason, 1990: Observations of boundary-layer structure over complex terrain, *Quart. J. Roy. Meteor. Soc.,* **116**, 159-186.

Hewer, F.E., and N. Wood, 1998: The effective roughness length for scalar transfer in neutral conditions over hilly terrain, *Quart. J. Roy. Meteor. Soc.,* **124**, 659-685.

Mason, P.J., 1985: On the parameterization of orographic drag, in *ECMWF Seminar on physical parameterization for numerical models of the*

atmosphere, pp.139-165. ECMWF, Shinfield Park, Reading, UK.

Milton, S.F., and C.A. Wilson, 1996: The impact of parameterized subgrid-scale orographic forcing on systematic errors in a global NWP model, *Mon. Weather Rev.*, **124**, 2023-2045.

Wood, N., and P.J. Mason, 1993: The pressure force induced by neutral, turbulent flow over hills. *Quart. J. Roy. Meteor. Soc.*,**119**, 1233-1267.

Wood, N., A.R. Brown, and F.E. Hewer, 2001: Parameterizing the effects of orography on the boundary layer: an alternative to effective roughness lengths, *Quart. J. Roy. Meteor. Soc.*, **127**, 759-777.

A Comparison between Wind Profiles from the Sydney (Terrey Hills) Doppler Radar and the Sydney Airport Wind Profiler

Susan Rennie[A], Daniel McIntosh[B], Alan Seed[A], Peter Steinle[A], Matt Tully[B]

[A]The Centre for Australian Weather and Climate Research

[B]Observations and Engineering Branch

Bureau of Meteorology, Melbourne, Victoria

s.rennie@bom.gov.au

Summary

Wind observations from the Sydney (Terrey Hills) Doppler weather radar were compared with those from a nearby wind profiler at Sydney Airport, in order to evaluate biases between the two instruments. The Australian Doppler radars are relatively new and their usage is being developed. Therefore evaluation against other instruments is informative, and can establish how these instruments may be used for NWP assimilation and verification

Data were collected for eleven days encompassing various weather types. To facilitate comparison, the Doppler radial winds were converted to vertical profiles using the Velocity Azimuth Display (VAD) technique. Signal from both precipitation and clear air (mostly insects) were used. The profiles from each instrument were then analyzed for biases and dependencies. Some biases were anticipated from the inherent characteristics of the observation techniques, i.e. Full Correlation Analysis (FCA) for the wind profiler versus Doppler VAD technique. For example, the FCA technique typically leads to a 10–15% wind speed underestimation[1], and produces less accurate direction at low speed. The VAD technique performs poorly at low wind speeds and is most accurate with linear wind fields and broad observation coverage. The findings coincided with these expectations; the VAD speed was higher by several m s^{-1}, and the direction agreement was substantially worse at low speeds. The bias due to use of insect echoes versus precipitation was also considered, but was generally not able to be isolated from other biases.

[1] When statistically compared against co-located Radio Sonde observations.

Instruments

The two instruments are located near the New South Wales coast, about 25 km apart (Figure 1). The Doppler radar at Terrey Hills is located to the north of the wind profiler at Sydney Airport. The instruments are very near the coastline, and

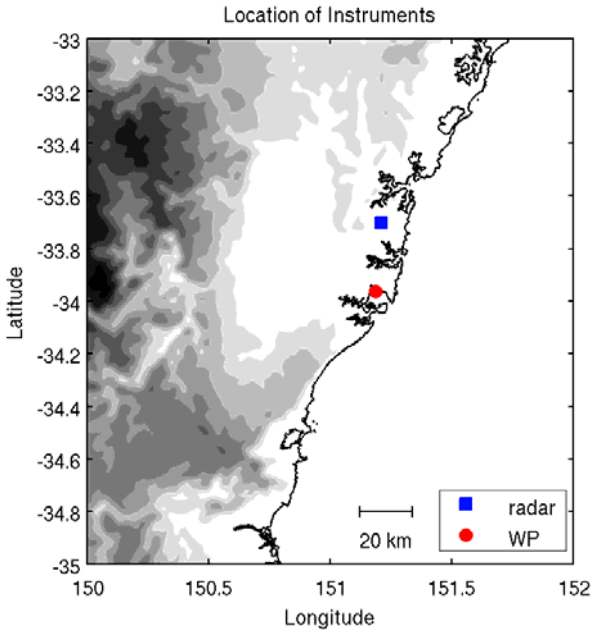


Figure 1 *Location of the Doppler radar at Terrey Hills and wind profiler (WP) at Sydney Airport.*

not far from mountains to the west. The distance between the instruments could allow substantial spatial variability in the wind field. For example a sea breeze may arrive at different times at the two locations. Furthermore, the radar measurements are derived by scanning a large area with radius perhaps up to 100 km at low elevation, whereas the wind profiler observes the atmosphere directly above it.

Wind Profiler

The wind profiler (WP) is a Boundary Layer Tropospheric Radar manufactured by ATRAD and located at Sydney Airport. The radar operates with a spaced antenna at 54.1 MHz (Vincent et al., 1998). The wind was calculated using Full Correlation Analysis (FCA) with the manufacturer-supplied software. Standard output included average values over 15, 30 or 60 minutes. Half-hourly values were used as a compromise between quality and frequency. Wind profiler heights were in 100 m increments to 2000 m, then in 300 m increments from 2200 m.

Wind profilers are expected to yield a speed bias of 10–15% below the true wind speed (Holdsworth and Reid, 2004b, Holdsworth and Reid, 2004a, Kariko et al., 2007). This is due to the "triangle size effect", proportional to the spacing of the antenna array (Holdsworth and Reid, 2004a). At low speeds (<5–8 m s^{-1}) the direction becomes unreliable. Heavy rain can also cause erroneous measurements.

Doppler VAD profiles

The Sydney (Terrey Hills) Doppler radar is an S-band radar with 1º beam width, which scans 14 elevations every 6 minutes. Elevations are 0.5º, 0.9º, 1.3º, 1.8º, 2.4º, 3.1º, 4.2º, 5.6º, 7.4º, 10º, 13.3º, 17.9º, 23.9º, and 32º. The PPI (plan position indicator) scans can be used to estimate the wind at the radar location over a range of heights through the Velocity Azimuth Display (VAD) technique (Andersson, 1992, Browning and Wexler, 1968, Michelson et al., 2000). The algorithm used here followed that of Michelson et al. (2000). VAD profiles were calculated for each elevation, and then all profiles were averaged into vertical bins of 100 m. Note that the height of the volume sampled by the radar beam, determined by the beam's width and trajectory, would be much larger than the 100 m interval (especially at long ranges) so the vertical profile was effectively smoothed.

The VAD algorithm assumed a linear wind field to fit a sinusoid to the radial wind measurements. It required diametrically opposite pairs of radial wind observations. Therefore, more data would be utilized with full azimuthal coverage than partial coverage. For example, scattered showers are likely to produce poor profiles due to few available data pairs. The calculation only proceeded if there were at least 8 data pairs. For quality control, the mean square residuals (R^2) of the best fit of the sinusoid were calculated as an error proxy. To remove gross errors, values were ignored where R^2>50. Then the speed, zonal (**u**) and meridional (**v**) components were averaged in 100 m intervals for all 14 elevations, and direction recalculated from **u** and **v**. If there was only one value in a height bin, or if the standard deviation of the **u** or **v** values was greater than 6, the average for that bin was rejected. This threshold removed very poor estimates but was not stringent. Subsequently, VADs with speed less than 2 m s^{-1} were removed, as the direction estimate was liable to be poor for low speeds, especially with noisy or patchy data. Typically the VAD calculation is limited to observations within a range limit (e.g. 30 km) over which the linear wind field assumption should hold. However, no such limit was applied here. Therefore the uncertainty will increase with height, and may be quite large at high altitudes.

The VAD calculation does not account for target velocity independent from the wind. In the case of precipitation, this is the fall velocity, which can be estimated from echo strength (as a proxy for drop size). For clear-air returns (assumed to be predominantly insects), the insect velocity may include vertical and horizontal contributions from insect air speed. The clear air velocity at low speeds is likely to be dominated by insect flight, which is another reason to remove values with speed less than 2 m s^{-1}. At low elevations, the vertical velocity of targets will have negligible impact. Horizontal velocity was corrected for elevation angle, which has a small effect at high elevations and a negligible effect at low elevations.

The Bureau of Meteorology also routinely calculates a variational VAD version (VVD, http://www.cawcr.gov.au/projects/wep_vvd/), which uses a variational method to fit the

sinusoid to the radial velocity. The variational method does not have the limitation of requiring diametrically opposite pairs, and so may achieve more accuracy by using available data. A comparison of VADs and VVDs from several days was made. There was generally good agreement of **u** and **v** within a few m s^{-1}. The uncertainty of a VAD was typically a few m s^{-1} (from root mean square error) and the uncertainty for the (hourly averaged) VVD was estimated as 1–4 m s^{-1}, based on the spread of values over one hour. There is occasionally a large error because VVDs are not calculated with dealiased data. VVDs also tend toward zero when there are few (noisy) data points, whereas VAD quality control (and reliance on diametrically opposite pairs) produces missing data flags. This last issue is largely due to differences in quality control procedures.

Analysis

The wind profiles and VADs were compared directly at equivalent nominal heights, and the difference (VAD−WP) calculated. Hence a positive speed difference indicates the VAD had the higher speed. The direction difference was normalised to within ±180º and positive indicates VAD observation was more anticlockwise. The differences were also analysed after segregation according to range of conditions, to investigate their effects. Data were collected from the eleven days listed in Table 1, in half-hourly intervals. All times are UTC, and local time (EST) is UTC+10h. An initial analysis included the first six days. The statistics did not change substantially with the additional five days.

Heavy rain was anticipated to impact the quality of wind estimates from the wind profiler.

Table 1 *List of case days and weather types.*

19/2/2010	Early light rain, clear air during day.
11/5/2010	Mostly clear air, light drizzle earlier in the day.
17/6/2010	Heavy rain first 6 hours. Light clear air later in day.
3/8/2010	Clear air/light showers early on.
18/8/2010	Heavy rain.
30/9/2010	Clear air, sea breeze 3–5 UTC.
7/11/2010	Clear air leading up to thunderstorm.
8/11/2010	Severe thunderstorm with hail.
16/12/2010	Thunderstorm.
29/1/2011	Clear air, some shear.
1/2/2011	Clear air, data missing early on.

Incidences when the WP was under heavy rain were extracted, i.e. when from the radar's 14 scans at least 5 gates directly above the wind profiler observed >10 dBZ reflectivity. This condition only occurred on four days, for a total of 20 half-hourly samples. For the first day, the VAD showed a higher speed, with difference at this time generally greater than 5 m s^{-1}. The observed velocity ranged up to 40 m s^{-1} at 6000 m altitude. The second day had only one incident, and did not look different from adjacent times' observations. The third day had a negative speed bias associated with the period. The fourth day showed variable bias, especially at high altitudes, and the VAD had a much higher R^2 error during the heavy rain (a hailstorm passing through caused a highly non-linear wind field). Direction differences were not obviously greater during heavy rain, but large differences were apparently correlated with VAD error and VAD speed.

Mean statistics

The overall mean differences in speed and direction at each height were found by averaging across all days, and all data (Figure 2). The mean difference in speed was almost entirely positive, i.e. VAD on average faster than WP. Below 2 km (large contribution from clear air), the mean difference was less than 2 m s^{-1}. Above 2 km (precipitation, fewer samples available, and larger error likely) the mean difference was often in the range 2–5 m s^{-1}. Cumulatively, 90% of all individual speed differences were within −4.5m s^{-1} and 6.75 m s^{-1}, with RMS difference

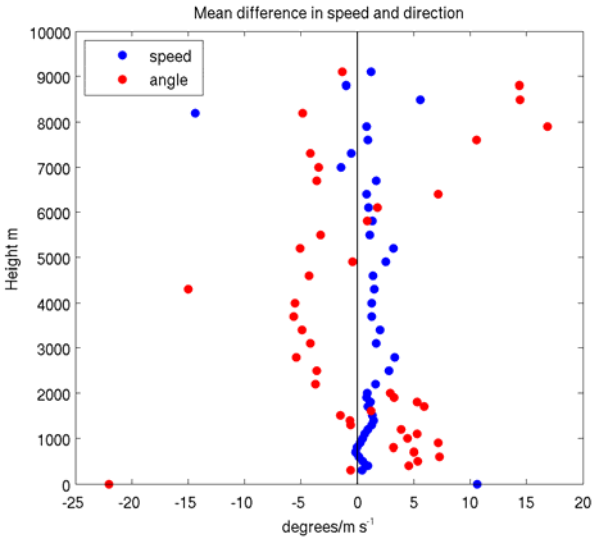


Figure 2 *Mean difference in speed and direction.*

of 3.6 m s^{-1} and mean difference of 1 m s^{-1}.

The mean speed difference normalised to the VAD speeds was also considered. A negative bias will be emphasized if the difference is negative when the VAD speed is low. Below 2–2.5 km (range containing clear air) the overall bias varied between −20% and 8%, but each day's average was localised to a smaller range. Positive and negative bias was observed from both clear air and precipitation. Above 2.5 km the mean difference was usually positive, but ranged between −13% and 20%. This suggests a positive bias is common at higher altitudes, where the wind speed is usually much greater than zero. Echoes at these altitudes are from precipitation only, so the target velocity should not cause overestimation of the VAD speed. Therefore the WP speed is likely underestimated. When the difference was normalised by WP speed, the result ranged between 5% and 35% (generally lower at low levels) but the variation with altitude was less pronounced. Examining the time series of data, speed differences appeared to be systematic rather than random. Differences persisted over several hours, and were often limited in vertical range, suggesting different air masses were associated with particular regions of bias.

The individual comparison of speeds is summarised in a 2D histogram (Figure 3). The distribution of speeds below the 1-to-1 line indicates the tendency of the WP to measure lower speeds than the Doppler radar. The line of best fit is also shown, but the fit is not really meaningful near the VAD speed truncation at 2 m s^{-1}.

The mean difference in direction (Figure 2, red dots)was also height dependent. Below 2 km it was mostly between −2º and 10º. At 2–6 km it was mostly −6º to 0º, and above this, between −5º and 15º; the spread increasing as data were sparser and had higher uncertainty. Ninety percent of individual direction differences ranged between −50º and 58º, with RMS difference 37º. It is marked that, particularly at low levels, the direction bias was not distributed about zero. The VAD wind direction was more anticlockwise below 2 km. The wind was most frequently from the east-southeast.

A time series of direction differences showed a persistence that indicated systematic rather than

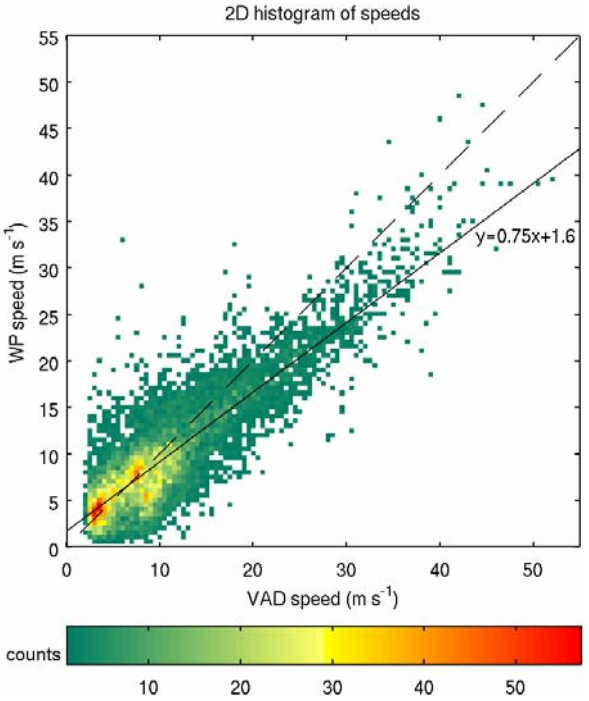


Figure 3 *2D histogram comparing VAD and WP speeds showing 1-to-1 line (dashed) and line of best fit (solid).*

random differences. Systematic differences could result if there were prevailing winds and spatial variation, e.g. sea breeze or topography. The features in the time series of directions did not elucidate a spatial variation or lag, possibly due to the coarseness of temporal sampling. However, some systematic differences were clearly associated with passing gust fronts and wind changes. For example, a sea breeze passing through early on 30/9/2010, or a storm front on 8/11/2010 that resulted in a large direction change. Such changes are not immediately registered by the VAD because the change needs to propagate past the radar before the diametrically opposite radial velocities represent the changed wind. When the change is directly over the radar, the VAD has a very poor fit, due to the nonlinearity of the wind field. Finally, insect air speed could cause a systematic direction difference that persists at all heights. Without auxiliary information it was difficult to account for these effects.

Generally, the worst direction agreement occurred when the direction uncertainty was high (sparse data, low speeds), or in strong vertical shear. There were several cases where the wind changed direction over a short height range, in all cases between 1000 and 2000 m. The direction

Table 2 *Vector correlation coefficient (r^2) for data separated by WP speed, and number of samples (n).*

Data range	r^2	n
All data	0.79	9399
>8 m s^{-1}	0.82	4517
4–8 m s^{-1}	0.62	3379
<4 m $^{-1}$	0.29	1503

difference sometimes reached 180º in this region. Besides real differences in atmospheric profiles, the instruments themselves may produce incongruous observations of shear regions. The VAD height assumed a standard atmosphere for beam propagation, so the height may have been misallocated. The measurement volume of the radar beam encompassed a greater height than the VAD bin, which would have caused vertical smoothing.

Comparison of speed and direction difference separated by WP speed

The wind profiler was predicted to have worse direction estimates at low wind speeds. (This was true for VADs also, but very bad values were usually excluded.) The differences in speed and direction were split according to wind profiler speeds of <4 m s^{-1}, 4–8 m s^{-1} and >8 m s^{-1}. On most days, data fell within one or two of these categories only. The (absolute) direction difference was typically a few tens of degrees, and the average absolute direction difference increased at low speeds (Figure 4). Vector correlation between (**u**,**v**) observations from each instrument were calculated following the method of Crosby et al. (1993), and normalised to between 0 and 1. The correlation coefficients are shown in Table 2. Correlation decreased with decreasing WP wind speed, using the categories defined above. At speeds below 4 m s^{-1}, the correlation was very low.

Incidents of largest disagreement were partly due to regions of shear at 1–2 km, and likely also partly due to raw radar data having spatial patchiness and a nonlinear wind field. Additionally, the VAD direction may have reduced accuracy at low speeds; firstly for clear air where insect air speed will have greater influence, and secondly because the VAD fit becomes less accurate if the maximal radial

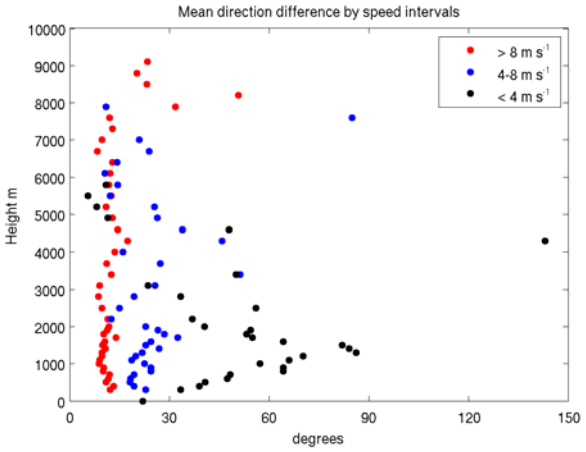


Figure 4 *Mean absolute difference in direction, separated by speed observed by the wind profiler.*

velocities are not distinct. The result shown in Figure 4 is probably attributable to all of the above, which were in effect at different times.

Clear air or precipitation

Variations in speed difference mostly coincided with weather features, and transitions between precipitation and clear air. There was no appreciable difference between day and night for most cases. An exception was 30/9/2010 (Figure 5) which was clear air with a few distant showers. The VAD speed ranged from <5 m s^{-1} during the daytime (before sunset and after sunrise as the figure shows one day UTC, or 10 am to 10 am EST) to 5–12 m s^{-1} overnight. In this case the speed difference was around zero during the daytime, when the speeds were low, and around 5 m s^{-1} overnight. This may have been due to nocturnal migration of moths that can fly at several metres per second. Note that it was too early in the season for locusts. The direction did not show a similar bias, but did appear to vary with VAD speed. Scatterers were sparser and patchier during the day, and denser during the night. Another variation was observed on 29/1/2011, where a marked speed difference was observed in an upper (~2 km) nocturnal jet, while the lower (~400 m) jet that persisted from afternoon through to morning appeared similar with both instruments. The jets corresponded with two layers in the reflectivity, and different directions (45ºT and −120ºT), suggesting separate migration layers. Bird migrants are also possible at 2 km.

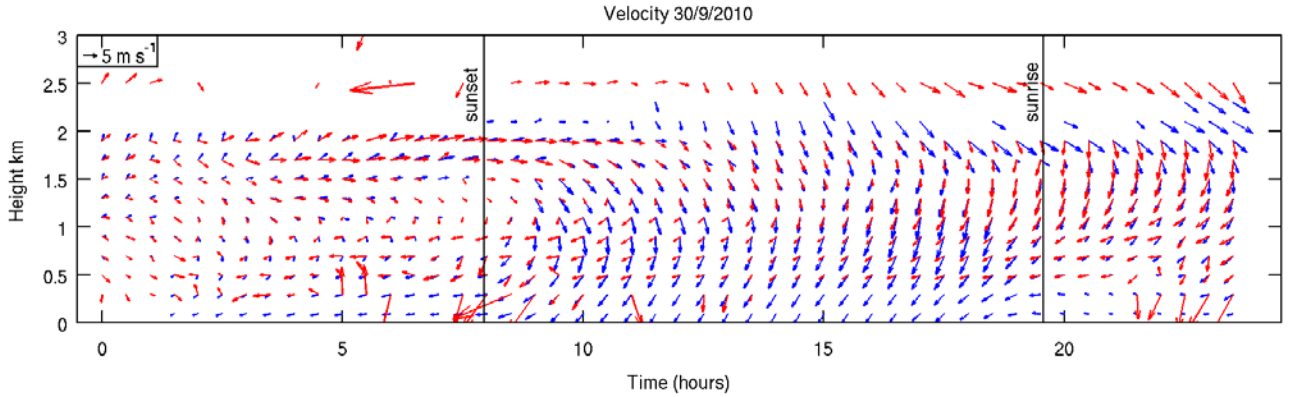


Figure 5 *Velocity vectors for Doppler radar (blue) and wind profiler (red) for 30/9/2010. The velocities are from clear air. There appears to be a difference in observed wind velocity between daytime and nighttime. Vectors represent* ***u*** *and* ***v*** *and are displayed with equal horizontal and vertical scales.*

There exists some evidence of Australian insects controlling their migration: the observation of common orientation with both entomological radar (Drake, 1983) and scanning weather radar (Author's unpublished data (in prep.)). Using weather radar to detect a 'dumbbell' pattern in the reflectivity, insect orientation at an acute angle to the wind was observed in rural NSW, particularly early in the day or night. However, it is difficult to observe this at a coastal location such as Sydney because insects avoid migrating over the ocean, preventing the symmetrical reflectivity pattern.

To distinguish clear air and precipitation velocities, the data were reanalysed by first dividing the raw Doppler data into high reflectivity (>2 dBZ) precipitation and low reflectivity (<5 dBZ) light rain and insects. Note that typical values of reflectivity for precipitation and clear air overlap in the interval 0–20 dBZ. For high reflectivity, agreement was good when there were sufficient data for a good VAD estimate. The WP slow bias was apparent, particularly at high wind speeds, as shown in the example in Figure 6.

For the low reflectivity data, often sufficient precipitation remained to produce VADs when clear air was minimal. Therefore, periods that were known to be without clear air echoes were ignored. For the remaining clear air periods, there may have been larger differences in the direction. However, these times also include very low wind speeds. It is difficult to separate the sources of disagreement or error at very low wind speeds, as discussed above. Error due to spatial variation in the wind might be reflected in the VAD error. However, VAD error is also proportional to wind speed, hence these two effects may cancel each other out. Ultimately, it is not obvious if there is an insect bias; the small sample size may be a factor.

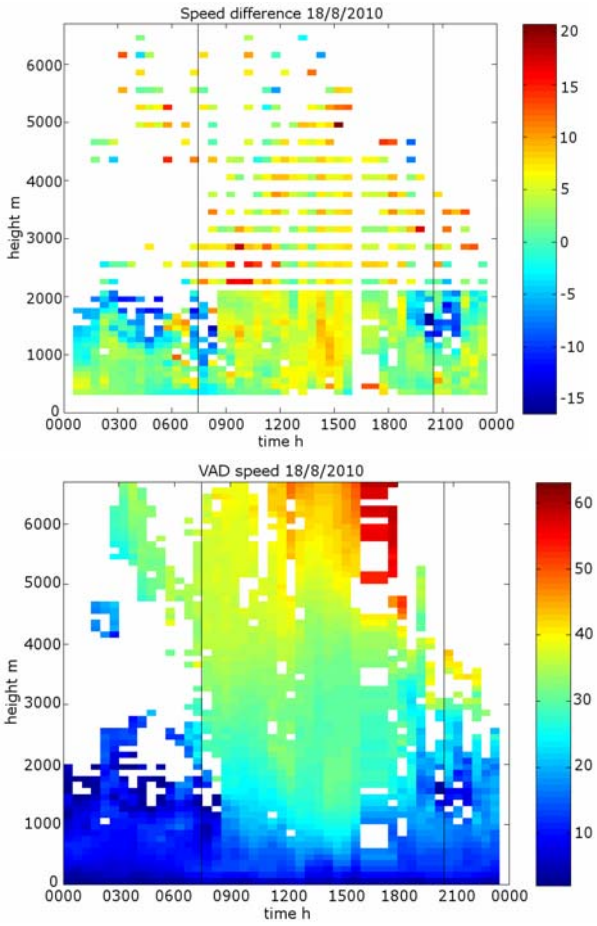


Figure 6 *18/8/2010, a rainy night, with some clear air during the previous afternoon. Top panel, difference in speed using high reflectivity observations. Bottom panel, VAD speed. Vertical lines denote sunset and sunrise.*

Conclusions

The results presented here explore the wind profiles measured by two instruments, a Doppler S-band radar and a wind profiler, and attempt to quantify and account for the differences. Data obtained on eleven days were analysed, the Sydney Airport wind profiler, and VADs derived from the Terrey Hills Doppler weather radar, showed fair agreement in speed and direction. The RMS speed difference was 3.6m s^{-1} and RMS direction difference was ~37º. This suggests that these instruments may be used to validate each other against gross errors, and may be informative in conjunction with numerical weather prediction. Further comparison with other instruments such as GPS sondes, which provide positional information, may also better identify discrepancies between the instruments.

There were some systematic differences. The VADs were on average faster by 1m s^{-1}. The direction agreed well at high speeds, but showed some systematic biases which seemed to relate to wind speed. At low wind speeds both the wind profiler and the VADs were expected to measure direction less accurately. Furthermore, other factors including spatial variations, vertical shear and insect air speed could also cause biases or errors that are more pronounced at low speeds.

Using precipitation, the VAD's limit to accuracy depended on spatial variability and coverage, i.e. smooth stratiform rain is best for VAD applications. Clear air echo usually provided broad coverage and often spatially uniform velocity observations, but with an air speed bias (which may be quantified if the target type is known), and a direction bias, which is more difficult to determine.

References

Andersson, T. (1992) A method for estimating the wind profile and vertical speed of targets from a single Doppler radar. In *WMO Technical conference on Instruments and Methods of Observation (TECO-92)*, Vol. 49 World Meteorological Organization, Vienna, Austria, pp. 380-384.

Browning, K. A. and Wexler, R. (1968) The determination of kinematic properties of a wind field using Doppler radar. *Journal of Applied Meteorology,* **7,** 105-113.

Crosby, D., Breaker, L. and Gemmill, W. (1993) A proposed definition for Vector Correlation in Geophysics: Theory and Application. *Journal of Atmospheric and Oceanic Technology,* **10,** 355-267.

Drake, V. A. (1983) Collective orientation by nocturnally migrating Australian plague locusts, *Chortoicetes terminifera* (Walker) (Orthoptera: Acrididae): a radar study. *Bulletin of Entomological Research,* **73,** 679-692.

Holdsworth, D. and Reid, I. (2004a) The Buckland Park MF radar: routine observation scheme and velocity comparisons. *Annales Geophysicae,* **22**.

Holdsworth, D. and Reid, I. (2004b) Comparisons of full correlation analysis (FCA) and imaging Doppler interferometry (IDI) winds using the Buckland Park MF radar. *Annales Geophysicae,* **22,** 3829-3842.

Kariko, A., Lockett, D. and Atkinson, R. (2007) Australian Bureau of Meteorology Upper Wind and Radar Profiler Network. In *AMS 33rd Conference On Radar Meteorology*Cairns, Queensland, Australia, pp. P13B8.

Michelson, D. B., Andersson, T., Koistinen, J., Collier, C. G., Riedl, J., Szturc, J., Gjertsen, U., Nielsen, A. and Overgaard, S. (2000) BALTEX Radar Data Centre Products and their Methodologies. In *SMHI Reports. Meteorology and Climatology*.Swedish Meteorological and Hydrological Institute, Norrkoping.

Vincent, R., Dullaway, S., MacKinnon, A., Reid, I., Zink, F., May, P. T. and Johnson, B. (1998) A VHF boundary layer radar: First results. *Radio Science,* **33,** 845-860.

Real-time predictions of coral bleaching risk for the Great Barrier Reef: Summer 2010/2011

C. M. Spillman

The Centre for Australian Weather and Climate Research

Bureau of Meteorology, Melbourne, Victoria

c.spillman@bom.gov.au

Mass coral bleaching events are primarily due to anomalously high ocean temperatures (Hoegh-Guldberg 1999) and have occurred on the Great Barrier Reef (GBR) in 1998, 2002 and 2006, causing widespread damage (Berkelmans et al 2004, Weeks et al. 2008). Under climate change such events are predicted to increase in both frequency and severity (Donner et al. 2005), which underscores the importance of developing appropriate management plans to minimise reef damage during such events.

The Great Barrier Reef Marine Park Authority (GBRMPA) has developed a comprehensive Coral Bleaching Response Plan, comprising of three components: 1) an early warning system, 2) assessment and monitoring, and 3) communication and public education (Maynard et al. 2009). Operational real-time seasonal forecasts of coral bleaching risk, developed at the Australian Bureau of Meteorology, form an important part of the early warning system and are used to identify future conditions conducive to mass bleaching. Forecasts are generated daily using the Predictive Ocean Atmosphere Model for Australia (POAMA), a coupled global ocean-atmosphere model and data assimilation ensemble forecast system developed by the Bureau of Meteorology and CSIRO.

The real-time POAMA forecast system (v1.5) operationally produces a nine month forecast each day, with the outlooks based on an ensemble of the 30 most recent daily forecasts. The variability of the results among the forecasts (i.e. ensemble members) gives an indication of the uncertainty in the future evolution of the climate system and provides information as to the probability distribution of future conditions. Forecast products are updated daily and hosted online by the Bureau of Meteorology at http://www.bom.gov.au/oceanography. For further model and real-time forecast generation information see Spillman and Alves (2009) and Spillman (2011).

Predictions on a seasonal time-scale are the most practical for reef managers, as advance warning of potential bleaching events allows the implementation of management strategies prior to bleaching onset to minimise reef damage (Marshall and Schuttenberg 2006). POAMA forecasts are used to brief government ministers, environmental managers, tourist operators and the public as to the risk of bleaching in the GBR. Forecasts are also used to plan research surveys and inform monitoring programs such as Bleachwatch, a community based coral reef monitoring initiative by the GBRMPA.

This paper summarizes observed ocean conditions in the GBR region for the summer of 2010/2011 and the skill of seasonal real-time predictions for this period.

Observed conditions

The annual sea surface temperature anomaly (SSTA) in the Australian region during 2010 was +0.54 C (referenced to the 1961 to 1990 average), the warmest value on record (Bureau of Meteorology 2011a). High individual monthly sea surface temperature records were also set in March, April, June, September, October, November and December 2010. In winter and spring of 2010 SST values were very warm over the GBR. These warm temperatures however were heavily mitigated in the following months by a series of extreme weather events including a

vigorous monsoon, heavy rainfall and cyclonic activity.

A moderate to strong La Niña event developed in 2010, transitioning from El Niño conditions established in 2009. NINO3.4 is an index used to describe ENSO and is defined as the areal average of monthly SST anomalies in the tropical Pacific Ocean (5 S–5 N, 170–120 W). Monthly NINO3.4 values for the summer of 2010/2011 were -1.3°C in November, -1.4 C in December, -1.5 C in January, -1.0 C in February and -0.6°C in March (Bureau of Meteorology 2011b). These values indicate a La Niña event which resulted in the early onset of the Australian monsoon in mid December, which was vigorous and sustained. Heavy rainfall occurred in coastal Queensland during December 2010, the wettest December on record for the state, and January 2011, causing extensive flooding (Bureau of Meteorology 2011c).

La Niña conditions are also often associated with higher numbers of tropical cyclones in the Australian region than usual (November-April) (Bureau of Meteorology 2011d). The most severe cyclone in the 2010/2011 summer was Tropical Cyclone Yasi which crossed the Queensland coast on 2 February 2011 (Figure 1) (Bureau of Meteorology 2011e). TC Yasi was a Category 5 tropical cyclone and one of the most powerful to have affected Queensland since records commenced. Significant wind damage was reported between Innisfail and Townsville (Bureau of Meteorology 2011f).

All of these factors minimised the build up of heat stress across the GBR over summer. Of a total of 1206 reef health reports received July 2010 to March 2011, 75% of reports recorded only minor bleaching and less than 1% indicated high levels of bleaching (Great Barrier Reef Marine Park Authority 2011). Damage sustained by the reef during TC Yasi and flooding is still being assessed.

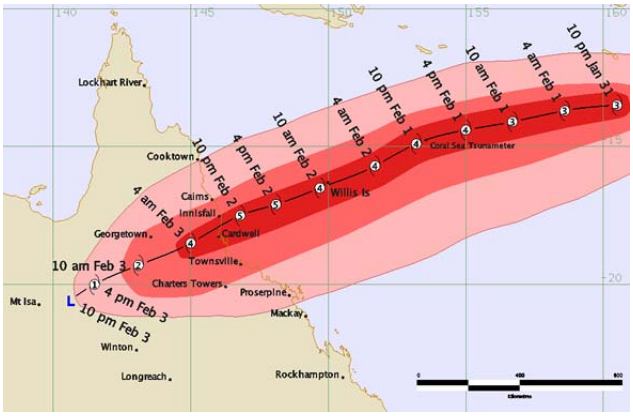


Figure 1 *Track and intensity information for Severe Tropical Cyclone Yasi.*[6]

POAMA SST Anomaly Forecasts

POAMA forecasts for December 2010issued 1 December, 1 November and 1 October are shown in Figure 2a, together with observed conditions. Observed average conditions for December were very warm, with anomalies exceeding 1.3 C in parts of the northern GBR region. POAMA forecasts indicated warmer than normal conditions for all forecast issue dates shown, though did not capture the magnitude of observed anomalies. The regionally averaged SST anomaly (SSTA) values (GBR Index; Spillman and Alves 2009) were 0.86 C for observed and 0.49 C, 0.69 C and 0.63 C for predictions issued 1 December, 1 November and 1 October respectively. The GBR Index forecast issued on 1 December 2010 is compared with observed values in Figure 3, which highlights that although the predicted ensemble mean for December is too low, the observed is captured within the ensemble spread. Both observed and the ensemble mean values also fall within the upper model tercile. The underestimate of ocean temperatures could be due to POAMA anticipating an excessively vigorous monsoon during December or local conditions not captured by the model.

In probabilistic forecasts for December issued 1 November and 1 October, the model indicates that the likelihood of SSTA exceeding 0.6 C was up to 80% in the central and northern GBR in forecasts (Figure 4a). The threshold of 0.6 C is a rule of thumb used by reef managers in mid-summer to indicate ocean conditions that may lead to coral bleaching. However in the forecast issued 1

December, probabilities did not indicate a high likelihood of threshold exceedance over most of the reef, despite values above this threshold being observed.

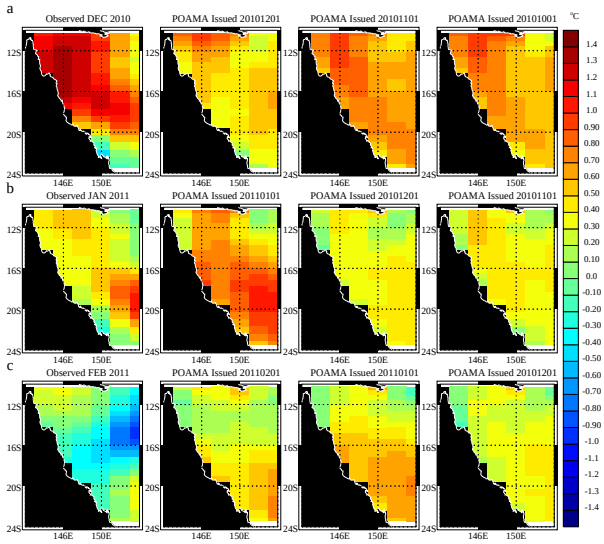


Figure 2 *Observed and POAMA SST anomalies (ensemble mean) in the GBR region for target months (a) December 2010, (b) January 2011 and (c) February 2011 from forecasts issued at the start of each month, 1 month prior and 2 months prior. Derived from outlooks issued 1 October2010 to 1 February 2011.*

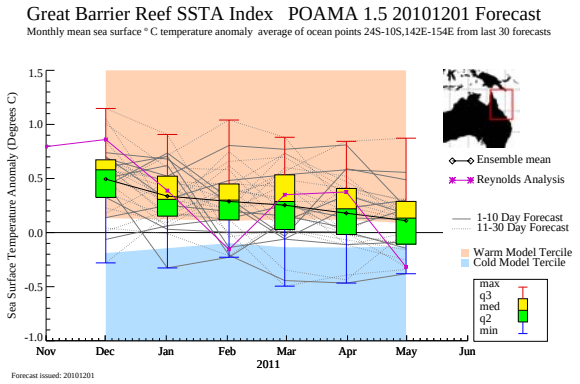


Figure 3 *POAMA monthly GBR Index values for December 2010 to May 2011in the official outlook issued on 1 December 2010, with the distribution by quartiles of the ensemble composed of the last 30 forecasts (grey lines). Overlaid is the ensemble mean (black) and the observed Reynolds GBR index (pink). The shading indicates upper and lower climatological terciles from the POAMA v1.5 hindcasts.*

Observed conditions for January 2011 were considerably cooler than those recorded in December, particularly in the northern GBR (Figure 2b). This can be attributed to the vigorous monsoon and heavy rainfall experienced over north eastern Australia in January. In general SST values were 0.3-0.6 C above climatology, with even warmer conditions noted offshore in the southern GBR. The region that was the exception however, was the southern coast where below average temperatures were evident. The POAMA SSTA forecast issued for January 2011 on 1 January (Figure 2b) represented the observed north-south temperature gradient quite well, though values were 0.3-0.5 C warmer than observed at the coast and in the northern GBR.

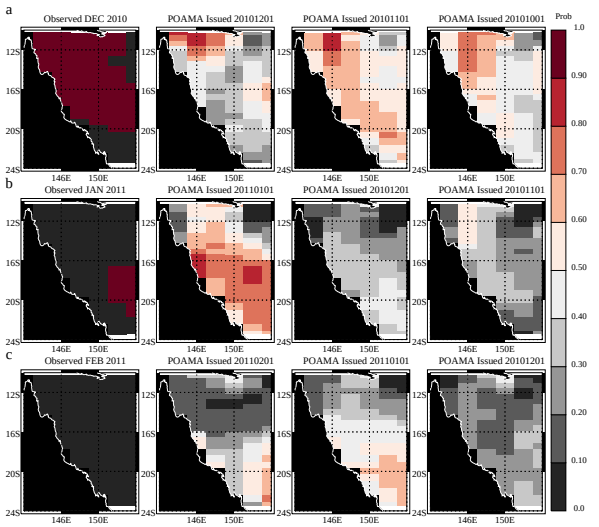


Figure 4 *Observed SST anomalies and probabilities of POAMA SST anomalies exceeding 0.6 °C in the GBR region for target months (a) December 2010, (b) January 2011 and (c) February 2011 fromforecasts issued at the start of each month, 1 month prior and 2 months prior. Derived from outlooks issued 1 October2010 to 1 February 2011.*

The predicted GBR Index was also higher than observed, i.e., 0.65°C compared to an observed value of 0.39°C (Figure 5). However forecasts issued 1 December (Figure 3) and 1 November captured average January observed conditions well, with GBR Index values 0.34 C and 0.36 C. The corresponding probabilistic forecasts indicate low probabilities (< 50% or less than 15 out of 30 forecasts) of exceeding 0.6 C throughout most of the region (Figure 4b). However in the forecast issued 1 January, the model indicates up to 90% chance of exceeding this threshold in the southern GBR, coinciding with observed values.

In February, observed conditions were again significantly cooler than those seen in December and January, with much of the region cooler than

climatology. In the area north of 16 S and east of 150 E, temperatures were up to 1 C cooler than the long term average. This is most likely due to the impacts of TC Yasi, the path of which roughly corresponds with this particularly cool region (Figure 1), combined with the continuation of a strong monsoon and heavy rainfall. POAMA predictions for February indicated warmer conditions than observed for all forecast issue dates shown, particularly in the southern GBR (Figure 2c). Conditions were predicted to be cooler in the northern GBR than the south, as was observed, but values were still overpredicted by the model. The ensemble spread captured the observed GBR Index value of -0.16 C in each of the outlooks issued 1 December (Figure 3), 1 January (Figure 5) and 1 February (Figure 6), though ensemble mean values were significantly higher than was observed i.e. 0.33 C, 0.42 C and 0.29 C. However the three probabilistic forecasts all indicate low probabilities (< 50%) of SST anomalies exceeding 0.6 C over most of the GBR region (Figure 4c).

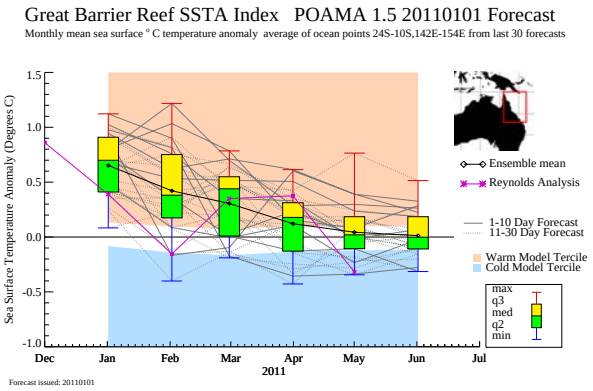


Figure 5 *POAMA GBR Index values for January-June 2011 in the outlook issued 1 January 2011, as per Figure 3.*

POAMA Accumulated Thermal Stress Forecasts

Degree Heating Months (DHM) give an indication of the persistence of thermal stress at a location and are calculated as the sum of positive monthly anomalies, referenced to the long term mean temperature of the warmest summer month, over a rolling 3 month time period (Spillman 2011, Spillman et al. 2010; 2011a). Model DHM forecasts were generated by summing over forecasts for the first three months from the outlook issue date i.e. for a forecast issue date of 1 December 2010, anomaly forecasts for December 2011, January 2011 and February 2011 are accumulated (Spillman et al. 2011a). This product is based on similar products developed by the NOAA Coral Reef Watch program, and is still under development.

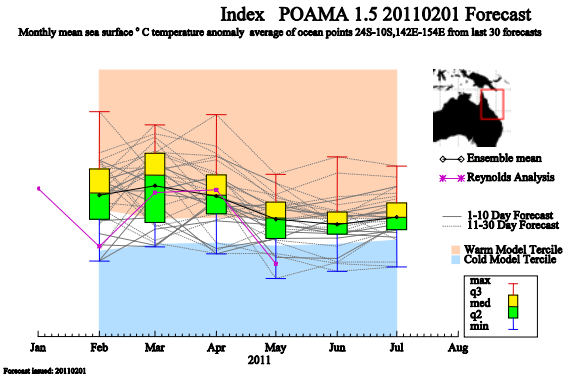


Figure 6 *POAMA GBR Index values for February-March 2011 in the outlook issued 1 February 2011, as per Figure 3.*

Model forecasts of DHM values in the tropical oceans generated on 1 December, 1 January and 1 February compared reasonably well to observed values for the corresponding seasons (Figure 7). DHM values ≥ 1 were observed in the 2010-2011 summer season around Papua New Guinea, north eastern Australia, central Western Australia and south of 24 S across the Pacific. Unlike the previous summer of 2009/2010 which saw high DHM values in the central equatorial Pacific due to a persistent El Niño signal (Spillman et al. 2010), there were no DHM values registered in this region in the summer of 2010/2011. This pattern was a result of the La Niña event, with DHM values highlighting the warm anomalies typical of such an event around eastern Australia.

The model replicates this spatial pattern well though under-predicts peak values. Along the north-east coast of Australia in the GBR region, the model predicts DHM values ≤ 2 and reproduces the general spatial pattern of the observed DHM values (Figure 7). However it doesn't capture the amplitude of observed values in the northern GBR in December-January-February (DJF). Predicted values however still indicate an increased risk of coral bleaching, and would give cause for concern for reef management, though not at the scale of past significant mass bleaching events, such as that of

2001/2002 (DHM values were ≥ 3.0; Spillman et al. 2011b). In February-March-April, POAMA overestimated thermal stress values. This is likely due to cooling effect of TC Yasi on ocean temperatures, as well as the influence of local atmospheric processes, which cannot be resolved by the model.

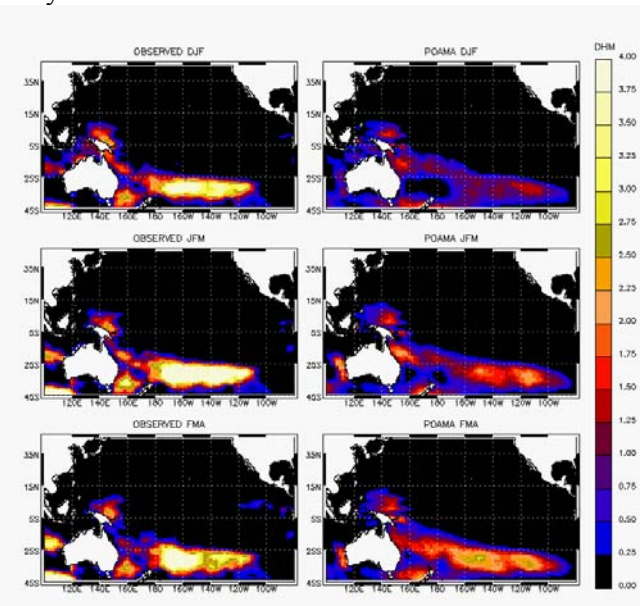


Figure 7 *Observed DHM values (first column) and POAMA ensemble mean DHM forecasts (second column) for outlooks issued on 1 December 2010, 1 January 2011 and 1 February 2011.*

Summary

The summer of 2010/2011 in the Great Barrier Reef region was characterised by a number of extreme weather events including a strong and vigorous monsoon, heavy flooding and Severe Tropical Cyclone Yasi. POAMA forecasts correctly predicted warmer than average conditions for December up to two months ahead, slightly overestimated January anomalies and indicated cooler temperatures by comparison for February. Observed values of the GBR Index were captured within the ensemble spread for all months, though the model ensemble mean differed from the observed more in some months than others. Forecasts of DHM values showed the model predicted the general patterns of thermal stress around the east coast of Australia, consistent with the typical pattern associated with a La Niña event. DHM values in the GBR region were elevated in December, indicating an increased risk of coral bleaching, though these decreased as summer progressed and only minor bleaching was reported during the summer.

References

Berkelmans R, De'ath G, Kininmonth S, Skirving W (2004) A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns and predictions. Coral Reefs 23:74-83.

Bureau of Meteorology (2011a). Annual Australian Climate Statement for 2010: http://www.bom.gov.au/announcements/media_release/climate/change/20110105.shtml

Bureau of Meteorology (2011b). ENSO wrap up December 2010-March 2011, http://www.bom.gov.au/climate/search/enso-wrap-up.shtml?bookmark=no-rm

Bureau of Meteorology (2011c) Bureau of Meteorology. Seasonal Climate Summary for Queensland – March 2011, http://www.bom.gov.au/climate/current/season/qld/archive/201102.summary.shtml

Bureau of Meteorology (2011d). ENSO Feature - The 2010-11 La Niña, http://www.bom.gov.au/climate/enso/feature/ENSO-feature.shtml

Bureau of Meteorology (2011e). Severe Tropical Cyclone Yasi report, http://www.bom.gov.au/cyclone/history/yasi.shtml

Bureau of Meteorology (2011f) Monthly Weather Review for Queensland, February 2011, http://www.bom.gov.au/climate/mwr/

Donner SD, Skirving WJ, Little CM, Oppenheimer M, Hoegh-Guldberg O (2005) Global assessment of coral bleaching and required rates of adaptation under climate change. Glob Change Biol 11:2251-2265.

Great Barrier Reef Marine Park Authority (2011). Coral bleaching forecast and status for the Great Barrier Reef (accessed June 2011), http://www.gbrmpa.gov.au/corp_site/key_issues/climate_change/management_responses/coral_bleaching_status

Hoegh-Guldberg O (1999) Coral bleaching, climate change and the future of the world's coral reefs. Review. Marine and Freshwater Research 50:839–866.

Marshall PA, Schuttenberg HZ (2006) A Reef Manager's Guide to Coral Bleaching. Great Barrier Reef Marine Park Authority, Australia pp167 (ISBN 1-876945-40-0).

Maynard JA, Johnson JE, Marshall PA, Eakin CM,

Goby G, Schuttenberg H, Spillman CM (2009) A Strategic Framework for Responding to Coral Bleaching Events in a Changing Climate. Environmental Management 44:1–11.

Spillman CM (2011) Operational real-time seasonal forecasts for coral reef management. Journal of Operational Oceanography, 4(1):13-22.

Spillman CM, Alves O (2009) Dynamical seasonal prediction of summer sea surface temperatures in the Great Barrier Reef. Coral Reefs 28:197-206.

Spillman CM, Alves O, Hudson DA (2011a) Seasonal prediction of thermal stress accumulation for coral bleaching in the tropical oceans. Monthly Weather Review, 139:317-331.

Spillman CM, Alves O, Hudson DA (2011b) Predicting thermal stress for coral bleaching in the Great Barrier Reef using a coupled ocean-atmosphere seasonal forecast model. International Journal of Climatology, *in press*.

Spillman CM, Hudson DA, Alves O (2010) Real-time seasonal SST predictions for the Great Barrier Reef during the summer of 2009/2010. CAWCR Research Letters, 4:11-19.

Weeks SJ, Anthony KRN, Bakun A, Feldman GC (2008) Improved predictions of coral bleaching using seasonal baselines and higher spatial resolution. Limnology and Oceanography 53:1369-1375.

POAMA-2 SST Skill Assessment and Beyond

Guomin Wang, Debra Hudson, Yonghong Yin, Oscar Alves, Harry Hendon,
Sally Langford, Guo Liu, and Faina Tseitkin
Centre for Australian Weather and Climate Research,
Bureau of Meteorology, Melbourne, Victoria
g.wang@bom.gov.au

Introduction

POAMA (Predictive Ocean Atmosphere Model for Australia) is an intra-seasonal to inter-annual climate prediction system based on coupled ocean and atmosphere general circulation models (Alves et al, 2003). POAMA-1.5 was implemented operationally in the Bureau in July 2007 (Wang et al, 2008). POAMA-1.5 was an update from the original version POAMA-1, which was developed jointly between the former Bureau of Meteorology Research Centre (BMRC), the division of CSIRO Marine Research (CMR) and the Managing Climate Variability (MCV) program.

Although POAMA is based on coupled climate models that represent and predict global climate variability, the main focus for POAMA-1 was the prediction of Sea Surface Temperature (SST) anomalies associated with the El Niño/Southern Oscillation (ENSO). It was initialized from a univariate system that assimilates only subsurface ocean temperature. In line with the original focus on predicting SST anomalies associated with ENSO, the initial atmospheric conditions for POAMA-1 forecast were not based on observed atmospheric conditions, rather they were derived from an AMIP-type simulation subject to the observed SST. Furthermore, hindcasts had only one member.

The focus of POAMA-1.5 expanded to include the direct prediction of regional climate variability over Australia. To do so, a new atmospheric/land initialization (ALI: Hudson et al 2011a) scheme was implemented and hindcasts were produced with 10 member ensembles in order to better sample the full range of possible forecast outcomes. The POAMA-1.5 system shows reasonable skill, at least at short lead time, for prediction of seasonal rainfall and surface temperature across Australia. (Lim et al, 2011). Although not specifically designed for intra-seasonal time-scales, the POAMA-1.5 system also demonstrates reasonable skill for intra-seasonal prediction of the Madden-Julian Oscillation (Rashid et al, 2010) and regional rainfall and temperatures (Hudson et al 2011b) to a lead time of 3-4 weeks.

A new version, POAMA-2, was implemented in Bureau operations in early 2011. Influenced by research progress made in seasonal prediction science and experience in operational practice during the past few years, POAMA-2 has undergone substantial upgrades both in system design and operational implementation strategies. This paper summarises the POAMA-2 system. Forecast performance is compared to that from POAMA-1.5 over a common hindcast period with a focus on prediction skill of tropical SST and Australian rainfall.. The status of the development of a version of POAMA-2 more suitable for intra-seasonal timescales is briefly discussed.

POAMA-2 System

The main model and coupler modules in POAMA-2 are similar to those used in POAMA-1.5. They includes the ocean model ACOM2 (Australian Community Ocean Model version 2), the atmospheric model BAM3 (the Bureau of Meteorology Research Centre Atmospheric Model version 3) and the OASIS3 (Ocean Atmosphere Sea Ice Soil version 3) coupler. The major upgrades of POAMA-2 over POAMA-1.5 are focused on the forecast system rather on the model modules, and include:

(1) an advanced ocean data assimilation scheme PEODAS,
(2) an ocean ensemble generation strategy which also generates ocean perturbations,
(3) a multi-model ensemble (MME) approach,
(4) a larger ensemble over a longer hindcast period,
(5) a new real-time forecast strategy that is consistent with the hindcast strategy (i.e., consistent climatologies)

The ocean analysis system for POAMA-1.5 was based on a univariate optimum interpolation system that assimilates only in-situ temperature observations. Due to the lack of appropriate multivariate formulations, this approach has a detrimental effect on the salinity and velocity fields of the analysed ocean state. To rectify these limitations a new ocean analysis system PEODAS (POAMA Ensemble Ocean Data Assimilation System; Yin et al., 2011) has been developed. PEODAS is an approximate form of the ensemble Kalman filter. It is based on the multivariate ensemble optimum interpolation system of Oke et al (2005), but uses covariances from a time evolving model ensemble. The construction of the ensemble in PEODAS explicitly represents errors in surface forcing, and the ocean model error is accounted for by introducing ocean perturbations through a method of additive inflation. Both in-situ temperature and salinity observations are assimilated, and current corrections are generated based on the ensemble cross-covariances with temperature and salinity. PEODAS has demonstrated a significant quantitative improvement in skill over its predecessor (Yin et al. 2011), especially in relation to salinity. Ensembles of analysed ocean states from PEODAS provide ocean initial conditions used to perturb the POAMA-2 ensemble hindcasts.

To address model uncertainty, POAMA-2 has also adopted a pseudo multi-model strategy using three different model configurations. The three configurations are:

P2.4c: standard atmospheric physics,
P2.4b: bias corrected version i.e. same model as P2.4a but with ocean-atmosphere fluxes corrected to reduce climatological biases,
P2.4a: modified atmospheric physics configured to use an alternative parameterization of shallow convection, which leads to slightly reduced mean state drift.

Because the physics options in P2.4c are similar to those in POAMA-1.5 (P1.5b), we will compare results from P1.5b and P2.4c in order to highlight impacts of the improved ocean initial conditions from PEODAS. We will also compare the complete ensemble of P24 (thirty members from 24a, 24b and 24c combined) with P15b in order to highlight the full benefit of the new P24 system that is running in operations.

An additional benefit of PEODAS is that ocean initial conditions were generated back to 1960, so that a hindcast set has now been produced with P24 for 50 years (1960-2009). Atmopspheric-land initial conditions were also generated back to 1960 using ALI. Initialized on the 1st day of each calendar month, a 10 member, 9-month forecast has been generated for each of the three model configuration of POAMA-2. Each member's initial conditions differ only in the ocean state (atmosphere-land state is same for all ensemble members) by taking the 10 different realizations of the PEODAS ocean analysis. Therefore there are a total of 30 members for each 9-month hindcast, initialized 12 times a year and for 50 years. In this paper we have limited the comparison of the hindcasts between POAMA-1.5 and POAMA-2 to the common period of 1980-2006.

Our analysis begins by examining the simulated El Niño variability in the different versions of the model. Figure 1 shows the standard deviation of the Nino3.4 SST index (SST averaged 170ºW-120ºW, 5ºN-5ºS) as a function of forecast lead time using all start months 1980-2006. The observational data (dashed line in Fig. 1) is from Hurrell et al (2008).

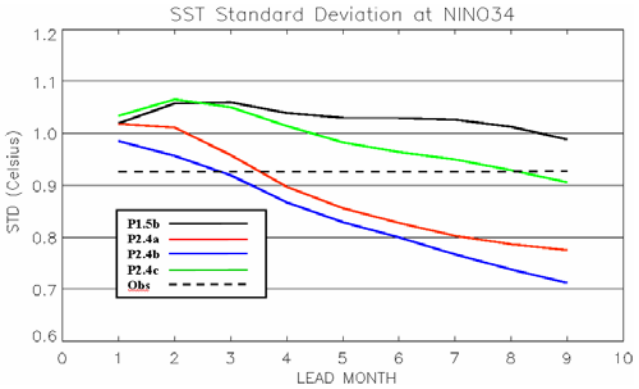


Figure 1 *SST standard deviation for the Nino3.4 region with lead time in months. Data used for forecasts are from 10 members of each experiment initialized on 1st of each calendar month during 1980-2006.*

There are several points worth noting from Fig. 1. First, each model configuration of POAMA-2 has its own characteristic in representing ENSO variability, with P2.4c having the strongest variability, P2.4b (flux corrected version) the weakest, and P2.4a in between. Second, there is a substantial difference in simulated ENSO variability after 3 month lead time between P1.5b and P2.4c, although the model components for these two version are identical. Zhao et al. (2011) conclude that this difference in simulated ENSO variability between P1.5b and P2.4c stems from the mean differences in the ocean initial states (primarily differences in the mean salinity) between PEODAS and the original univariate system use in P1.5b, which results in different simulated mean states that support different levels of ENSO variability. Figure 1 also suggests that ENSO behaves differently in each of the POAMA-2 model configurations, which also stems from different simulated mean states in each of the three model version (Zhao et al. 2011). This different level of simulated ENSO variability will also affect the simulated teleconnections of ENSO to other parts of the world. A potential benefit of this different level of simulated ENSO variability by the three different versions of POAMA-2 is increased ensemble spread, which is required for improved reliability of probabilistic climate forecasts. The results shown below support the idea that the MME approach of POAMA-2 does lead to improved probabilistic forecasts.

Tropical SST Prediction Skill

(a) forecast accuracy

A commonly used measure of forecast accuracy is the anomaly correlation coefficient (ACC) between forecast and observed quantities. Forecast anomalies are formed by subtracting the model climatology from each individual hindcast, which has the effect of removing the systematic mean state error. The model climatology is defined as a function of lead time and initial month. In this paper we define a lead time of one month as monthly mean of the first month of the forecast, therefore the forecasts go out to a total lead time of nine months.

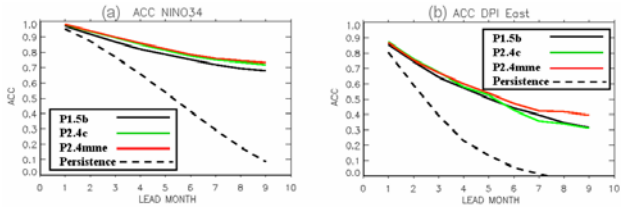


Figure 2 *SST anomaly correlation calculated for each lead time in months for (a) Nino3.4 and (b) east node of Indian Ocean Dipole. Data used for forecasts are from ensemble means of each experiment initialized on 1st of each calendar month of years 1980-2006. Ensemble means are from 10 members for P1.5b and P2.4c, and 30 members for P2.4MME.*

The anomaly correlation for the Nino3.4 SST index and the east pole of the Indian Ocean Dipole are shown in Fig. 2 for the forecasts from P1.5b (10 members), P2.4c (10 members), and POAMA-2 (30 members; P2.4MME). Comparing P1.5b and P2.4c highlights the impact of improved ocean initial conditions, and comparing P1.5b and P2.4MME highlights the impact of all changes introduced in POAMA-2.

The increase in ACC of Nino3.4 SST is most prominent for both P2.4c and P2.4MME suggesting that the new ocean data assimilation PEODAS has had a positive impact on forecast skill of ENSO. POAMA-2 has gained around one lead month for a given level of correlation compared to POAMA-1.5. For SST anomalies over the east Indian Ocean, the skill from all three models is similar for lead times up to ~5 months. For longer lead times, P2.4MME has outperformed both P1.5b and P2.4c. It is

important to keep in mind, however, that the overall level of skill in the Indian Ocean is still much lower than in the Pacific and this is common to all three systems.

(b) forecast reliability

In addition to forecast accuracy, a key indicator of the confidence of a forecast is its reliability, which is defined as how well the forecast probabilities of occurrence match the observed frequency of occurrence. Here, we construct an attributes diagram whereby the relative observed frequency of an event is plotted against the forecast probability. Five bins with an equal probability interval of 0.2 are used. A perfectly reliable forecast will lie along the diagonal 1:1 straight line in the attributes diagram, such that the event is observed to occur with the same frequency as the probability with which it was forecast.

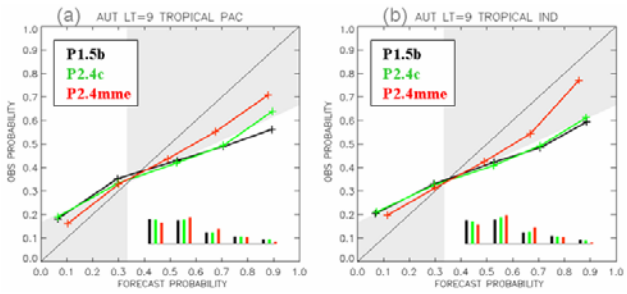


Figure 3 *Attributes diagram of probabilistic SST anomaly forecast of above upper tercile at lead time of 9 month over (a) tropical Pacific (150ºE-100ºW, 25ºS-25ºN), and (b) tropical Indian Ocean (50ºE-110ºE, 45ºN-35ºS). Data used for forecasts are from 10 members for P1.5b and P2.4c, and 30 members for P2.4MME, initialized on 1st of each calendar month during years 1980-2006. Bar charts are relative frequency of event occurrence. Bin width is 0.2. Shaded area represents the region where a forecast will contribute positively to the Brier Skill Score when compared to climatology.*

Construction of the attributes diagram is sensitive to sample size, so rather than use just the Nino34 index, we assess the SST forecasts at each grid point over the tropical Pacific Ocean (150ºE-100ºW, 25ºS-25ºN) and tropical Indian Ocean (50ºE-110ºE, 45ºN-35ºS). Here we consider forecasts for the probability of being in the upper or lower tercile of the SST distribution. Forecast probability is determined by counting the number of ensembles falling in a given category at eache model grid, and then aggregated over an area with weights proportional to grid size. The attributes diagram is based on the aggregated probability forecasts.

The prediction of SST anomalies over the tropical Pacific and Indian Oceans are reliable for lead times up to 6 months for P1.5b, P2.4c and P2.4MME, with P2.4MME being the best (figures not shown). At longer leads P2.4MME is relatively reliable, as shown in Fig. 3 at 9 months lead. Even at 9 months lead, the forecasts from P2.4MME for SST to be in the upper tercile are reliable, whereas forecasts from P1.5b and P2.4c are not. Interestingly forecasts for lower tercile Pacific SST (cold events) have better reliability than forecasts for warm events (figures not shown).

The above results indicate that the strength of MME for increasing forecast reliability. The larger ensemble size over that from P15b helps, but the biggest benefit for increased reliability comes from combining three configurations of POAMA-2 rather than more ensemble members being used.

Australian Rainfall Forecast Skill

a) forecast accuracy

We now assess whether the increased forecast skill from POAMA-2 for tropical SST carries over to prediction of Australian rainfall. We focus on forecast skill for lead times 2-4 months, which is typical of an operational setting to make forecasts for the upcoming season. Forecast accuracy is assessed using a two-category accuracy score for above and below median. The forecast accuracy is defined as the correct number of forecasts for above median plus the correct number of forecasts for being below median, divided by the total number of forecasts. The accuracy score for the four main seasons are shown in Fig. 4. Note that we use 9 of the ten members to produce the probabilistic forecast (so that we avoid the possibility of exactly 50% chance of being above/below median). The 9 member subset is taken randomly (without replacement) from the full 10 member ensemble of each of the model configurations. This is repeated 100 times and averaged to determine the

probabilistic forecasts. From Fig. 4 we see that forecasts in austral spring and autumn are consistently the most predictable seasons and all three models lack skill in summer. There is a shift in the location of the regions of high accuracy between each version of the model, particularly in autumn, but the continent averaged accuracy is consistently slightly higher for P2.4c and P2.4MME compared with P1.5b.

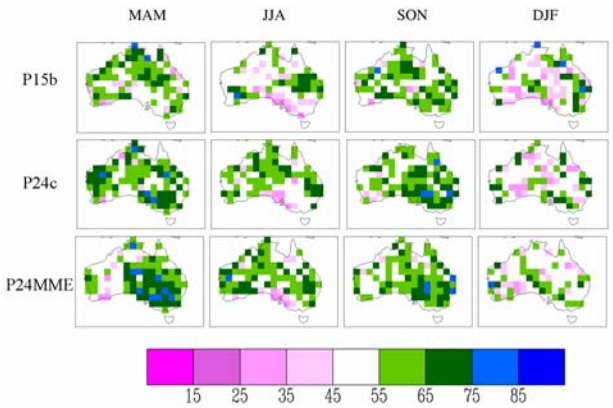


Figure 4 *Accuracy score (as a percentage) for above median rainfall probability forecast for P1.5b, P2.4c and P2.4MME, for the years 1980-2005, averaged over lead 2-4 months. An accuracy score greater than 50 per cent (indicated by green or blue grid points) is considered skilful. Observational verification data is from Jones and Weymouth (1997).*

b) forecast reliability

For a forecast to be valuable, it must be reliable as well as accurate. The attributes diagram for prediction of rainfall in the upper tercile is shown in Fig. 5. Here, all grid points over Australia are considered using all start months. To test the impact of the increased ensemble size for P2.4MME, both the full 30 member and a half size 15 member P2.4MME result are given. In the latter case a 5 member subset is taken randomly (without replacement) from the full 10 member ensemble of each of the three model configurations of POAMA-2 and a probability forecasts is formed based on this subset. This is repeated 10 times, and the average is used as the final probability forecast. Results are similar with other choices of subset size and number of repetition. The similarly good reliability using this subset of the P2.4MME compared to the full P2.4MME, and the minimal improvement in P2.4c compared to P1.5b, confirms that the major contribution to improved reliability of the P2.4MME is from the use of three model versions rather than increased ensemble size.

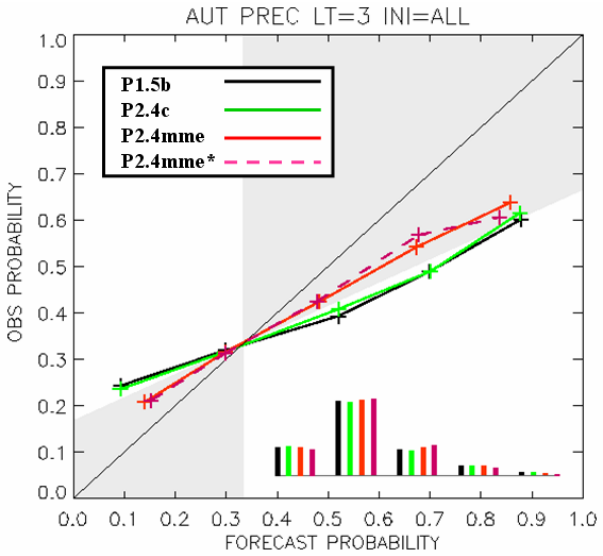


Figure 5 *Reliability diagram of Australian rainfall anomaly probability forecast of above upper tercile averaged over lead 2-4 months. Data used for forecasts are from 10 members for P1.5b (black) and P2.4c (green), 30 members for P2.4MME (red), and 15 members for P2.4MME* (pink dashed), initialized on 1st of each calendar month of years during 1980-2006. Bar charts are relative frequency of event occurrence for each experiment for each probability forecast bin.*

Development of an Intra-seasonal Forecast System

To fill the current prediction capability gap between weather forecasts and seasonal outlooks for Australia, POAMA-2 is being developed with a specific intra-seasonal component. Motivation for development of the intra-seasonal system based on POAMA-2 is provided by some promising results using POAMA-1.5 for prediction of the MJO (Rashid et al. 2010; Marshall et al. 2011) and of regional Australian rainfall and temperatures (Hudson et al. 2011b). The key focus for development of the POAMA-2 system is enhancements to ensemble initialization and generation.

A key limitation of the POAMA-2 seasonal

prediction system for application to intra-seasonal forecast is the use of a single atmospheric initial condition for each forecast. Although ensemble spread is provided through the perturbed ocean initial conditions, the development of this spread is too slow to be useful at intra-seasonal timescales where forecast uncertainty is primarily determined by the atmosphere. To this end, a coupled ensemble initialisation (CEI) system has been developed which produces perturbations to both the ocean and atmosphere at the initial time of the forecasts. Initial assessment of some forecasts produced using this perturbation strategy indicates that good spread is developed in the first weeks of the forecasts. Initial indications are that this initialisation strategy also has benefits on the seasonal timescale. Furthermore, this new initialisation strategy is closely aligned with our long-term goal of fully-coupled data assimilation with POAMA. Discussion of the performance of the POAMA-2 intra-seasonal system will be the topic of a future paper.

Conclusion

The new POAMA-2 seasonal forecast system demonstrates improved forecast accuracy and reliability for SST in the central tropical Pacific, which we interpret to mean that POAMA-2 has increased forecast skill to predict El Niño. The new system demonstrates marginally higher forecast skill in the east tropical Indian Ocean. We attribute the increased skill in the Pacific to the improved ocean analysis system, PEODAS, introduced in POAMA-2 and to the use of a MME approach.

Although forecast accuracy for rainfall over Australia is only marginally improved using POAMA-2, there is a substantial improvement in forecast reliability at short lead times. This improvement in reliability primarily derives from the use of the multi-model combination and not from the use of larger ensembles.

The POAMA-2 intra-seasonal system is currently being developed. Preliminary results suggest it is likely to have increased skill on time scales from one week to one or two months compared to POAMA-1.5. The first season forecast skill from this system is also likely to be higher than that from the POAMA-2 seasonal system.

Acknowledgements

Some of the development of the POAMA-2 system was carried out in joint projects with support from the Managing Climate Variability Program administered through the Australian Grains Research and Development Corporation.

References

Alves, O., G. Wang, A. Zhong, N. Smith, F. Tseitkin, G. Warren, A. Schiller, S. Godfrey and G. Meyers, 2003: POAMA: Bureau of Meteorology operational coupled model seasonal forecast system. Proc. National Drought Forum, Brisbane, Apr 2003, pp. 49-56. Available from DPI Publications, Department of Primary Industries, GPO Box 46, Brisbane, Qld 4001, Australia.

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

Hudson, D., O. Alves, H.H. Hendon, and A.G. Marshall. 2011b: Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Q. J. Royal Meteor. Soc.*, **137**,673-689.

Hurrell, J.W., J.J. Hack, D. Shea, J.M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153.

Jones, D.A., G. Weymouth, 1997: An Australian monthly rainfall dataset. *Technical Report 70*, Bureau of Meteorology.

Lim, E.-P., H.H. Hendon, D.L.T. Anderson, A. Charles, and O. Alves 2011: Dynamical, statistical-dynamical and multi-model ensemble forecasts of Australian spring season rainfall. *Mon. Wea. Rev*. **139**, 958-975.

Marshall A. G., D. Hudson, M.C. Wheeler, H.H. Hendon, and O. Alves, 2011: Assessing the simulation and prediction of rainfall associated with the MJO in the POAMA seasonal forecast system. *Climate Dynamics*. *doi:10.1007/s00382-010-0948-2.*

Oke, P. R., A. Schiller, D.A. Griffin, and G.B. Brassington, 2005: Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quart. J. Roy. Meteor. Soc.*, **131**, 3301-3311.

Rashid H. A., H. H. Hendon, M. C. Wheeler, and O. Alves, 2010: Predictability of the Madden_Julian Oscillation in the POAMA Dynamical seasonal prediction system. *Climate Dynamics*. **36**, 649-661**.**

Wang, G. O. Alves, D. Hudson, H. Hendon, G. Liu and F. Tseitkin 2008: SST skill assessment from the new POAMA-1.5 system, *BMRC Res. Lett*. No. 8, 2-6. www.cawcr.gov.au/publications/BMRC_archive/researchletter/reslett_08.pdf

Yin, Y., O. Alves, and P. R. Oke, 2011: An ensemble ocean data assimilation system for seasonal prediction. *Mon. Wea. Rev.*, **139**, 786-808.

Zhao, M., H. H. Hendon, D. Anderson, Y. Yin, and O. Alves, 2011: Impact of assimilating salinity on the simulated mean state and variability in a coupled seasonal forecast model., to be submitted.