Aust. Met. Mag. 57 (2008) 95-108

Comparison of official and operational consensus forecasts of daily extreme temperatures in 2006

Frank Woodcock[1], Grant Elliott[2] and Milos Setek[3]
[1]Centre for Australian Weather and Climate Research, Melbourne, Australia
[2]Western Australian Regional Office, Bureau of Meteorology, Perth
[3]Bureau of Meteorology, Melbourne, Australia

(Manuscript received November 2007; revised January 2008)

Operational Consensus Forecasts (OCF) and official forecasts of daily maximum and minimum temperature in 2006 are verified for 107 locations where both forecasts are available. At eight capital cities the forecasts extend from one to six days ahead but for the remaining 99 locations the official forecasts usually extend only one day ahead. Over all sites combined, the next day's forecast from OCF is better than official forecasts for both maxima (61 per cent of sites; mean absolute error (MAE) 1.28°C compared with 1.31°C) and minima (75 per cent; MAE 1.48°C compared with 1.53°C). The comparative verification results revealed detailed systematic differences between the two forecasting schemes. For example, the official forecasts of capital city maxima were usually more accurate than OCF whereas for forecasts of capital minima OCF were more accurate.

Forecasters at the three capital cities where official forecasts of maximum temperatures were clearly better than OCF provided summaries of their perceived weaknesses in OCF that enables them to improve on OCF guidance. The relative weaknesses of OCF and official forecasts are discussed with a view to improving future performance.

The main suggestions arising from this study are that OCF may be improved by: (a) basing OCF bias-correction on wind analogues or developing an OCF model output statistics scheme; and (b) rationalising the number of numerical models used in OCF to reduce the over-representation of the Australian regional numerical model in its one to two days ahead forecasts and increasing the number of numerical models that contribute beyond two days ahead. Official forecasts should benefit from documentation of the meteorological situations where OCF performs both well and poorly then using that information when formulating official forecasts. Official forecasts of minimum temperature generally and maximum temperature at most locations would be more accurate if more weight were given to the OCF guidance.

Introduction

Our purpose was to compare the performance in 2006 of Operational Consensus Forecasts (OCF; Woodcock and Engel 2005) of daily temperature extremes with the subsequently issued official forecasts, to diagnose their strengths and weaknesses, and thereby to suggest ways to improve either or both.

The comparison is between the OCF maxima and minima guidance forecasts issued on day+0 for day+1 to day+6 inclusive for eight capital cities (Adelaide, Brisbane, Canberra, Darwin, Hobart, Melbourne, Perth

Corresponding author address: Frank Woodcock, Centre for Australian Weather and Climate Research, GPO Box 1289, Melbourne, Vic. 3001, Australia.
Email: f.woodcock@bom.gov.au

and Sydney) and for day+1 only for 99 regional centres. The partition between capital cities and regional centres was necessary because official forecasts at most regional centres are only issued for day+1. Much of the comparison focuses on relative performance at the capital cities where longer range forecasts could be compared.

OCF and official forecasts are not independent forecasting systems because OCF is scheduled to provide guidance to forecasters to assist in formulating official forecasts. Hence, the official forecasts should outperform OCF. However, our main concern is to identify where the official forecast accuracy systematically differed from OCF. Forecasters at those capitals where official forecasts convincingly outperformed OCF have provided some insight into the methodologies they employed. Their response together with the results from additional experiments both here and overseas provides a basis for our suggestions that may lead to improvements in both OCF and official forecasts in the future.

In the 'Forecasts and verification' section below, the theory and operational production of OCF is outlined and an overview of official forecasts at Regional Forecast Centres is discussed in terms of procedures, deadlines and the information available. The verifying observations and methodology are then discussed. In the Results section the performance of the two forecasting schemes is compared with a view to identifying systematic differences. Discussing the results, including operational forecaster summaries, it is evident that the most successful centres have become aware of the local meteorological conditions where OCF does not perform well. They also note the loss of models available to OCF for forecasts beyond day+2. As part of the discussion, we look at ways to improve OCF and how the official forecasters can make better use of OCF. We conclude by suggesting ways this improvement can be achieved.

Forecasts and verification

Discussion

Comparing forecast quality can often be misleading because despite ensuring that the verification is done over exactly matching events, there are many factors which assist one scheme relative to the other and over which the verifier has little control. In this section some of these factors are discussed. The OCF and official forecast processes are discussed in more detail.

A significant factor affecting the verification results is that OCF issue is scheduled to precede the formulation of official forecasts. This is done to give the official forecasters the option of adopting, modifying or discounting OCF in the formulation of the official forecasts.

OCF is issued about 0400 UTC each day after the required products from the 0000 UTC run of the Australian regional numerical prediction model (LAPS 375) have been generated. This means that the eastern Australian States (Queensland: Qld, New South Wales: NSW, Victoria: Vic. and Tasmania: Tas.) in winter receive OCF guidance forecasts about 1400 local time. The subsequent official forecasts used in this study are issued at 1700 but usually formulated between 1500 and 1600. Thus OCF leads the official forecast formulation by about one to two hours. In South Australia (SA) and the Northern Territory (NT) OCF leads the official forecast formulation by an extra half-hour and in Western Australia (WA) the lead is about four to five hours. In 2006, lead times were one hour shorter in summer in NSW, Vic., Tas. and SA due to daylight saving.

OCF

OCF commenced real-time operations in the Bureau's National Meteorological and Oceanographic Centre (NMOC) in 2005 (NMOC 2005) and has superseded Model Output Statistics (MOS; Glahn and Lowry 1972; Woodcock 1984; Mills and Tapp 1984), which had been the mainstay of NMOC's daily objective guidance to forecasters since 1984. OCF is provided to forecasters via an internal web and to external customers via File Transfer Protocol (FTP). Forecasts are generated for day of issue (day+0) out to seven days ahead (day+7) for over 750 locations ranging from Mawson (67°S, 63°E) to Nauru (0.5°S, 167°E). They are for specific locations and consist of a subset of maximum and minimum air temperature, minimum ground temperature, evaporation, hours of sunshine, precipitation and probability of precipitation according to the availability of the last n days of observations received from the location, where $14 < n < 31$.

OCF forecasts are derived from both local and international numerical model fields available to NMOC using both direct model output (DMO), based on bilinear and temporal interpolation from model grid-point values to locations where a forecast is required, and MOS forecasts from the Australian models. Essentially, OCF is generated by removing the bias from each newly available component forecast (DMO or MOS) then combining them, giving smaller weights to the poorer performing (see Woodcock and Engel (2005) for more details).

Factors that contribute to DMO errors are mostly related to differences between a model's approximate representation of the earth's surface and the true values, and the model's numerical approximations required to parametrise the physics of complex real-world processes. Key factors include:

(a) differences between modelled and real terrain heights;
(b) net radiation received at the ground;
(c) the ratio of sensible to latent heat flux, etc. (see Stensrud and Yussouf (2005) for a more detailed description).

Clearly, the first factor is constant unless the model resolution changes. Under quiescent synoptic situations, factors (b) and (c) will vary slowly from the same time one day to the next, but under changing weather situations they may vary quickly.

In OCF, constant and slowly varying errors are removed from component forecasts using a running best easy systematic estimator of error bias (BES; Wonnacott and Wonnacott (1972), section 7.3), where

$$BES = (Q_1 + 2Q_2 + Q_3)/4 \qquad ...1$$

and Q_1, Q_2 and Q_3 are first, second and third quartiles of forecast errors over the n-day sample. BES is more robust than the average or median. Rapid or sudden changes in factors (b) and (c), as may occur from changes in cloudiness, rain, or frontal passage, will not be fully captured by the BES bias-correction.

The last n days of interpolated model forecasts for a site are then bias-corrected; that is, BES is deducted from each component forecast, and the mean absolute error (MAE) of the bias-corrected value is computed via

$$MAE = (\sum_{i=1}^{n} |f_i - o_i|)/n \qquad ...2$$

f_i is the ith forecast and o_i its verifying observation. The BES and MAE are then stored and used to adjust and combine the latest available DMOs and MOS when OCF is run.

To generate OCF the latest available component forecasts are collected and bias-corrected using their stored BES values. The final stage combines the bias-corrected component forecasts for the same site using a weighting for each based on the inverse of its stored MAE. The weight (w_i) for the ith component forecast is

$$w_i = (MAE_i)^{-1} \left(\sum_{i=1}^{n} (MAE_i)^{-1} \right)^{-1} \qquad ...3$$

The weighting increases according to the accuracy over the last n days. The combination process both causes cancellation of random errors in the unbiased component forecasts and a reduction of their extreme errors (Clemen 1989).

The full set of numerical models available to OCF in 2006 is listed in Table 1. Under operational conditions the international models can arrive too late for inclusion in OCF or not at all. The 'Missing days' column of Table 1 lists the number of international models in 2006 that never arrived.

Table 1. Forecast schemes used in 0400 OCF in 2006.

Forecast scheme	*Issue UTC*	*Missing days*	*Approximate arrival UTC*	*Resolution km*	*Resolution h*	*First h*	*Validity Last day Min_T*	*Validity Last day Max_T*
LAPS 050	0000	0	0310	5.0	1	0	1	1
LAPS 125	0000	0	0410	12.5	3	0	2	1
LAPS 375	0000	0	0230	37.5	3	0	3	2
UKGCM	1200	8	1745	140.0	6	0	1	1
USGFS	1200	8	1800	140.0	6	0	1	1
JMA	1200	12	1915	270.0	24	24	2	1
GASP	1200	0	2025	270.0	3	0	4	5
ECMWF	1200	24	2045	40.0	24	24	>7	>7
LAPS MOF	0000	0	0240	I	6	3	2	2
GASP MOF	1200	0	2030	I	12	24	5	6

Note: Missing days are the number of days out of 365 that the model was unavailable but does not include being too late for use in OCF. Arrival times are approximate and 1200 issue time refers to the day prior to OCF issue. Last day Min_T (Max_T) is the number of days after OCF issue that the normal time Eastern Standard Time (EST) occurrence of minimum (maximum) temperature falls within the available model's days-ahead forecast.
LAPS = Australian Limited Area Prediction System.
UKGCM = United Kingdom Global Circulation Model.
USGFS = United States Global Forecast System.
JMA = Japan Meteorological Agency global model.
GASP = Australian Global Analysis and Prediction.
ECMWF = European Centre for Medium-range Weather Forecasts global model.
MOF = Australian Model Output Statistics based on MOS (Glahn and Lowry 1972) and uses multiple linear regression equations to generate site forecasts from model fields.
I = Spatiotemporal interpolation from model grid to city.

Table 1 also shows that the arrival time of global models (UKGCM, USGFS, JMA, GASP and ECMWF; see table for definitions) lags five to eight hours after their nominal issue time. This means that they were based on information 16 hours prior to the OCF run. The later issue of the global models is at 0000 UTC but they arrive in NMOC between 0500 and 0800 UTC and are too late for inclusion in the 0400 UTC issue of OCF.

Official forecasts

Official forecasts for the capital cities are the Bureau's most important and comprehensive public forecasts in terms of detail, frequency of issue, days ahead, and population coverage. They are formulated by professional meteorologists for their home capital cities using a range of objective guidance that includes local and international numerical model predictions, OCF, locally developed objective algorithms, satellite imagery sequences, and an extensive knowledge of local mesoscale weather. There is substantial feedback from peer forecasters, management and the press covering forecast failures: a strong incentive for excellence. At regional centres, outside the capitals, local knowledge is less extensive, site-specific algorithms are rarer, and performance monitoring and feedback are weaker. Hence, the expectation is that official forecasts would be better at the capital cities than at the other regional centres.

Table 2 lists details of the afternoon Regional Forecast Centre issue times in winter. They are issued in time to accompany the 1800 (local time) radio and television news; hence issue times need to be adjusted to accommodate daylight saving in summer.

The official forecasts retrieved from the forecast database for this study were lodged by 1200 UTC. They would usually correspond to the afternoon issue time in Table 2 but could include revised forecasts issued later in the evening but before 1200 UTC.

All official forecasts were formulated with NMOC 0400 UTC OCF guidance available together with the underlying OCF component model forecasts (both raw and bias-corrected) and a flag indicating the best-performing bias-corrected model for that site and weather element over the last 30 days.

Additionally, the forecasters in the Eastern Standard Time (EST) time zone could access 0000 UTC runs of the JMA, UKGCM and USGFS global models that arrive after the OCF issue but before the issue of official forecasts. Similarly the UKGCM and USGFS model guidance beyond day+2 is available for most official forecasts in winter. Occasionally Darwin and Adelaide can access ECMWF 0000 UTC forecasts. Perth would normally receive all the 0000 UTC model guidance that is available after OCF is issued.

Official forecasters were also advantaged by access to USGFS maximum and minimum temperatures out to seven days that, for operational reasons, were not permitted in OCF.

In April 2006 the Bureau increased the number of official forecast days ahead from day+4 to day+6 at all capital cities except Darwin. Hence the verification data for day+1 to day+4 in this study covers 365 days but the day+5 and day+6 data coverage was about 265 days.

Verifying observations

Observed daily maximum and minimum temperatures were retrieved from the National Climate Centre's database. They are quality controlled within the National Climate Centre (e.g. by checking against observer field book entries and nearest neighbour observations). No errors were detected in this study.

The measuring period for observed maximum and minimum temperatures is provided in Table 2 and corresponds exactly to the OCF forecast period but differs from the official forecast period. The difference may slightly handicap the official forecast per-

Table 2. Details of official wintertime forecast issue times, OCF issue time and verifying observation period.

Forecast	*Location*	*Issue time UTC*	*Forecasts Days ahead*	*Measuring period**	
				Maximum	*Minimum*
official	WA	0830	1 – 6	6 am to midnight	6 pm to 9 am
	NT	0700	1 – 4	6 am to midnight	6 pm to 9 am
	SA	0700	1 – 6	6 am to midnight	6 pm to 9 am
	Qld	0700	1 – 6	6 am to midnight	6 pm to 9 am
	NSW	0700	1 – 6	6 am to midnight	6 pm to 9 am
	ACT	0700	1 – 6	6 am to midnight	6 pm to 9 am
	Vic.	0700	1 – 6	6 am to midnight	6 pm to 9 am
	Tas.	0700	1 – 6	6 am to midnight	6 pm to 9 am
OCF	All	0400	1 – 7	9 am to 9 am	9 am to 9 am
observation	All			9 am to 9 am	9 am to 9 am

Note: Issue times are subject to minor operational fluctuations. Official issue times are earlier in summer when daylight saving is enforced. Days ahead is for the capital cities after April 2006: for regional centres one day ahead is usual.
* denotes local clock time.

formance when the observed maximum occurs between midnight and 0600 or the minimum occurs after 0900 but before 1800 (local time).

Method

Both OCF and official forecasts were extracted from the Bureau's site forecast database. No error checking was undertaken other than against observations when large differences occurred or when there were large differences between OCF and the official forecast. One official forecast in the forecast database had the tens digit missing and was corrected.

For this study, OCF temperatures were rounded to the nearest integer prior to verification in order to match the precision of official forecasts.

Verification was undertaken only when both OCF and official forecasts were available. The verification parameters included:

(a) bias, where

$$bias = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) \qquad ...4$$

where n is the number of verifying observations;

(b) MAE (see Eqn 2);

(c) the explained variance (V%) of the observations by the forecasts as the square of the correlation coefficient between forecasts and observations, where

$$V\% = 100\left(\sum_{i=1}^{n}(o_i-\bar{o})(f_i-\bar{f})/\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}\sqrt{\sum_{i=1}^{n}(f_i-\bar{f})^2}\right)^2 \qquad ...5$$

(d) the percentage of absolute errors greater than 2.5°C , and those greater than 4.5°C, and;

(e) the Priestley skill score (P; see Mason 1982) and used here as

$$P = 1 - (\text{MSE} / \sigma_o^2) \qquad ...6$$

where MSE is the mean square error of forecasts

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2 \qquad ...7$$

and σ_o^2 is the variance of the verifying observations.

In addition to consolidated and individual site verification, we follow Baars and Mass (2005) and compare the performance of official forecasts and OCF on day+1 after stratifying according to persistence error; i.e. observed day+0 (the persistence forecast) minus observed day+1 (the verifying observation). OCF was expected to perform worse than official forecasts when large persistence errors occurred because OCF, in using a 15 to 30-day calibration, would not respond sufficiently quickly to accommodate a changed regime when, for example, the biases and weightings learned over a hot dry period were applied after a sudden cold change. This expectation is supported by results from Cheng and Steenburgh (2007), where forecasts based on a seven-day running mean bias-correction of a single model DMO perform better under quiescent conditions than under dramatic, large-scale synoptic pattern changes.

Again following Baars and Mass (2005), we also investigate how the accuracy of forecasts varies as the verifying temperatures depart from the daily climate normal: i.e. climate day+1 – observed day+1. OCF was expected to perform relatively poorly in extreme departures from normal because the 30-day bias-correction learning period used by OCF would be dominated by near-normal events. This comparison is undertaken for day+1 alone and for day+1 through day+6 combined.

Results

Capital city maximum temperature forecasts

Summary verification results for the capital city maximum temperature forecasts are provided in Table 3. The bias of both sets of forecasts was negligible. Figure 1 displays the corresponding MAEs. Overall, official forecasts of maxima were better than OCF maxima from day+1 to day+6 and over all verification parameters. The improvement of official forecasts over OCF was largest at day+3, then at day+4 and next at day+1. Although worse than the official forecasts, the better relative results for OCF occurred for verifications day+2, day+5 and day+6 while the poorer verification results were on day+3 and day+4.

Table 4 displays the number of capitals where official or OCF has the lower MAE. Similar to Table 3, the official forecasts were clearly better than OCF for day+1, day+3 and day+4 forecasts. More detailed analysis (not shown) shows only three capital official forecasts convincingly outperform OCF. As Table 4 shows, OCF MAEs are lower than official forecast MAEs at two capitals on day+1, three on day+2, and four out of seven capitals on day+6. What is noteworthy here is that despite the convincing overall result in Table 3 that official forecasts were more accurate than OCF, some capital official forecasts performed relatively poorly. The Perth and Hobart results were particularly good compared to OCF.

Following Baars and Mass (2005), the performance of OCF and official forecasts on day+1, stratified according to the change in maxima between day+0 and day+1 (day+0 – day+1), is investigated. For this type of 'distributions-orientated verification' (Brooks and Doswell 1996) OCF is expected to perform relatively poorly when large daily tempera-

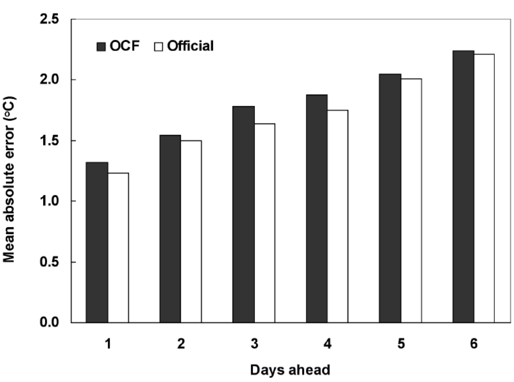
Table 3. Summary of OCF and official (OFF) maximum temperature forecast verification over all capital cities combined. Best results are in bold font.

Forecast	*Days ahead*	*Cases*	*Bias (°C)*	*MAE (°C)*	*Exp var(%)*	*%>2.5 (°C)*	*%>4.5 (°C)*	*Priestley*
OCF	1	2920	0.06	1.30	85	15	3	0.85
OFF			-0.06	**1.21**	**87**	**12**	**2**	**0.87**
OCF	2	2920	0.04	1.54	81	20	4	0.81
OFF			-0.10	**1.50**	**82**	**18**	**3**	**0.82**
OCF	3	2920	0.07	1.78	76	25	7	0.75
OFF			-0.11	**1.64**	**79**	**22**	**5**	**0.79**
OCF	4	2615	0.08	1.88	70	27	8	0.69
OFF			-0.09	**1.75**	**74**	**24**	**6**	**0.73**
OCF	5	1857	0.04	2.05	67	30	10	0.65
OFF			-0.08	**2.01**	**69**	**30**	**9**	**0.67**
OCF	6	1829	-0.09	2.24	62	34	12	0.60
OFF			-0.14	**2.21**	**63**	**34**	**12**	**0.62**

Table 4. Number of capitals where the mean absolute error (°C) was lower. There are eight capitals for 1 to 4 inclusive days ahead and seven thereafter because Darwin does not have official forecasts beyond four days ahead. Ties have been discounted and better results are in bold font. OFF = official forecast. For example for day+1 forecasts of maximum temperature, the official mean absolute error was lower than the OCF mean absolute error at six capital cities, and the OCF mean absolute error was lower at two capital cities.

Days ahead	*1*		*2*		*3*		*4*		*5*		*6*	
Forecast	OFF	OCF	OFF	OCF	OFF	OCF	OFF	OCF	OFF	OCF	OFF	OCF
Maximum temp.	**6**	2	**4**	3	**5**	1	**5**	2	**4**	3	**3**	4
Minimum temp.	2	**5**	0	**8**	0	**6**	2	**6**	1	**4**	2	**4**

Fig. 1 Mean absolute error of maximum temperature forecasts over the Australian capital cities 2006 (1829 to 2920 events).



ture changes occur. As Fig. 2 shows, there is no strong signal until the change in temperature exceeds 10°C, but the sample sizes are so small that the results are not significant. Both official and OCF forecasts make the worst error of +8°C on an occasion when the daily temperature dropped 18°C from the previous day. There was a slight tendency for OCF to underestimate the maximum temperature more than the official forecasters do when large warm changes occur (see comments from Hobart forecasters in the Discussion section below), but performances after sharp cool changes are similar. The comparative verification relative to persistence was not extended beyond the day+1 forecast.

In a second 'distributions-orientated verification' (again following Baars and Mass 2005), the variation in forecast errors as the verifying observed temperature deviates from climatological normal is investigated. The one-day-ahead result is shown in Fig. 3. Usually the OCF mean error for maximum tempera-

Fig. 2 Mean error of OCF and official day+1 maximum temperature forecasts against corresponding persistence error for all capitals combined.

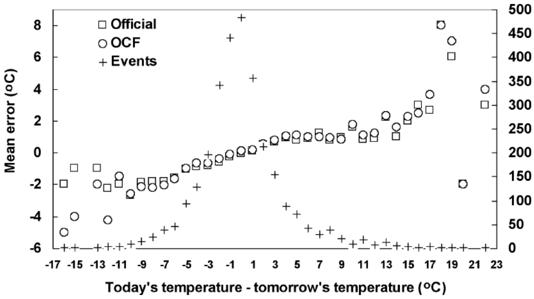
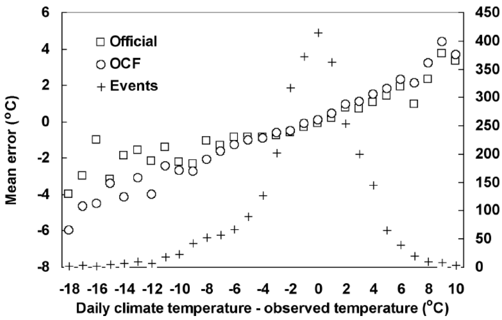


Fig. 3 Mean error of OCF and official maximum temperature forecasts for one day ahead against departures of verifying observations from daily climate temperatures.



ture is larger than the corresponding official forecast mean error and the difference increases as the verifying temperature departure from climatology becomes more extreme. However, the difference between them is negligible until the departure exceeds about 6°C. The larger discrepancies occur once the departure from normal exceeds 10°C but sample sizes drop quickly as departures increase.

The results for one to six days ahead combined are shown in Fig. 4. Again it is evident that the OCF forecasts of maximum temperature deteriorate more rapidly than official forecasts as the verifying temperature becomes more extreme. The separation between OCF and official forecast errors is less distinct when the longer-range forecasts are included than it is for the day+1 forecast.

The percentage of days when one forecast was better than the other is also used as a measure of quality. The results are provided in the 'Maximum' column of Table 5. Here, OCF is marginally superior at day+2

Fig. 4 As Fig. 3 but for day+1 to day+6 combined.

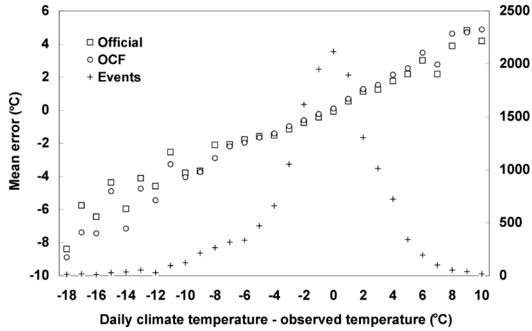


Table 5. Percentage of days with the more accurate forecasts (better results in bold font). OFF = official forecast.

		Maximum		*Minimum*	
Days ahead	*Days*	*OCF*	*OFF*	*OCF*	*OFF*
1	2920	25	**30**	**27**	25
2	2920	**31**	30	**36**	26
3	2920	29	**35**	**35**	29
4	2615	29	**33**	**33**	30
5	1857	30	**30**	**30**	26
6	1829	30	**31**	**29**	28
All	15061	29	**32**	**32**	27

Table 6. Number of capitals with the higher percentage of lower mean absolute errors (ties discounted and better results in bold font). OFF = official forecast.

		Maximum		*Minimum*	
Days ahead	*Sites*	*OCF*	*OFF*	*OCF*	*OFF*
1	8	2	**6**	4	4
2	8	**5**	3	**8**	0
3	8	2	**6**	**6**	2
4	8	3	**5**	**5**	3
5	7	**4**	3	**5**	2
6	7	**4**	3	2	**4**
All	46	20	**26**	**30**	15

ahead while the official forecasts are clearly superior at day+1, day+3 and day+6. The 'Maximum' column of Table 6 lists the number of capitals with the highest percentage of more accurate forecasts and highlights a sharp deterioration in OCF relative to official forecasts from day+2 to day+3.

Capital city minimum temperature forecasts

Summary results for minimum temperature forecast verification are shown in Table 7. OCF bias is generally smaller than the corresponding official forecast bias. OCF biases are negligible and positive whereas the official forecast biases are larger but negative. All the verification parameters in Table 7 rank OCF better than official forecasts for day+1 to day+6 inclusive. The MAE chart (Fig. 5) shows that the largest differences between OCF and official forecast accuracy occurred for day+2 and day+3 with only minor improvements of OCF over official forecasts for day+1 and day+6.

The percentage of days with more accurate forecasts (Table 5) suggests that OCF forecasts of daily minima convincingly outperform official forecasts at day+2 to day+5 days inclusive and are of similar accuracy to official forecasts for day+1 and day+6. The number of capitals where the MAE is lower (Table 6) shows that no capital city official forecast MAE is lower than corresponding OCF at day+2.

Figure 6 compares the performance of official and OCF next day forecasts, stratified according to the difference in minima between forecast issue day and the next day. Similarly to maximum temperature, official forecasts are expected to outperform OCF when large day-to-day temperature changes occur. However, as Fig. 6 shows, there is no evidence that OCF deteriorates relative to official forecasts as the change from one day to the next increases in magnitude.

Similarly, when considering extreme departures from climatology (Fig. 7 for day+1 and Fig. 8 for day+1 to day+6 combined) there is no evidence that OCF minima deteriorate relative to official forecasts as the verifying temperature becomes more extreme.

Fig. 5 As Fig. 1 but for minimum temperatures.

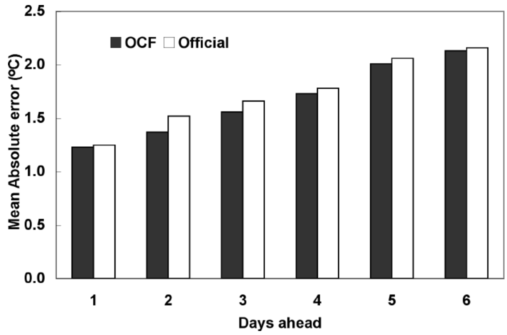
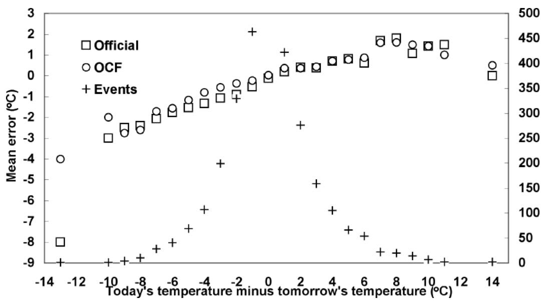


Fig. 6 As Fig. 2 but for minimum temperatures.



Regional centre day+1 forecasts of temperature maxima and minima

The verification over regional centres for the next day forecasts is summarised in less detail than the capital city forecasts in Table 8.

Table 7. As Table 3 but for minimum temperature forecasts.

Forecast	*Days ahead*	*Cases*	*Bias (°C)*	*MAE (°C)*	*Exp var (%)*	*%>2.5 (°C)*	*%>4.5 (°C)*	*Priestley*
OCF	1	2920	**0.04**	**1.23**	**88**	**11**	**1**	**0.89**
OFF			-0.20	1.27	88	12	1	0.88
OCF	2	2919	**0.04**	**1.37**	**86**	**16**	**2**	**0.86**
OFF			-0.17	1.52	83	19	3	0.82
OCF	3	2919	**0.06**	**1.56**	**82**	**21**	**3**	**0.82**
OFF			-0.15	1.66	80	23	5	0.80
OCF	4	2614	**0.05**	**1.73**	**75**	**25**	**5**	**0.75**
OFF			-0.13	1.78	73	27	6	0.73
OCF	5	1857	**0.01**	**2.01**	**62**	33	**8**	**0.61**
OFF			-0.13	2.06	60	**33**	9	0.59
OCF	6	1829	**0.01**	**2.13**	**58**	35	**10**	**0.57**
OFF			-0.14	2.16	56	**38**	11	0.54

Fig. 7 As Fig. 3 but for minimum temperatures.

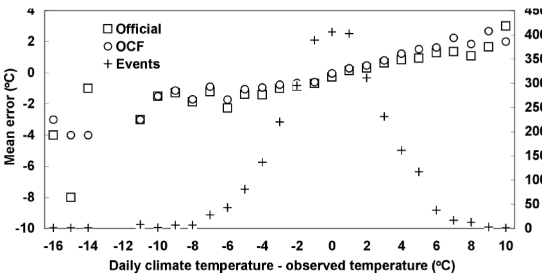


Fig. 8 As Fig. 4 but for minimum temperatures.

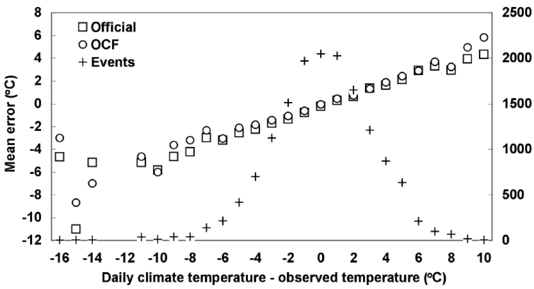


Table 8. Mean absolute error (MAE) of OCF and official forecasts and the percentage of sites with a lower MAE for the 99 regional sites, each with over 300 verifications. Better results are in bold font. OFF = official forecast.

		Maxima		*Minima*	
Region	*Sites*	*OCF*	*OFF*	*OCF*	*OFF*
		Mean absolute error (MAE) °C			
WA	22	1.39	**1.37**	**1.41**	1.48
SA	13	1.42	**1.33**	**1.50**	1.58
NT	8	**0.93**	1.07	**1.12**	1.19
Qld	17	**0.96**	1.00	**1.26**	1.27
NSW	15	**1.42**	1.47	**1.71**	1.73
Vic.	19	**1.40**	1.50	**1.70**	1.82
Tas.	5	**1.21**	1.22	1.47	1.47
ALL	99	**1.28**	1.31	**1.48**	1.53
		% sites with a lower MAE			
WA	22	36	**64**	**82**	14
SA	13	15	**85**	**77**	23
NT	8	**87**	13	50	50
Qld	17	**71**	18	**59**	41
NSW	15	**87**	13	**60**	40
Vic.	19	**84**	16	**95**	5
Tas.	5	40	**60**	**60**	40
ALL	99	**63**	37	**73**	27

Because OCF is an objective system, there is no reason to expect its accuracy at regional centres to differ substantially from its accuracy at capital cities. Here, however, it is evident that overall for day+1 maximum temperatures forecasts, OCF is more accurate than the official forecasts. Only the SA regional centre official forecasts substantially outperformed OCF. This result is in sharp contrast to the capital city results for day+1 in Table 3. The difference is mainly because the official forecast MAE over the capital cities is 7.6% lower than over the regional centres while the OCF capital city MAE is 1.5% larger than over the regional centres.

The corresponding results for minimum temperature forecasts over regional centres are consistent with the capital city minima results, with OCF clearly providing the better forecasts overall. It is interesting to note that, except for Canberra, both the official forecast and OCF are much more accurate at capital cities than for the regional centres, thereby indicating that most capital city minima forecasts are relatively easy, presumably resulting from heat island, maritime and low altitude influences restricting large variations in minimum temperature.

The overall improvement of OCF over official forecasts (in terms of MAE) for maxima is 2% and for minima is 3%; both values are lower than the corresponding 9% and 13% cited in Woodcock and Engel (2005). OCF produces lower MAEs than official forecasts for 60% of sites for maxima and 75% for minima compared to 67% and 76% in Woodcock and Engel (2005).

Discussion

Forecasters' comments

The main purpose of this study is to identify ways to improve the objective forecasts (OCF) so that in the future it provides better guidance for the formulation of official forecasts. In order to gain further insight to the more successful official forecasting results in 2006, the views of the forecasters were sought.

Official capital city maximum temperature forecasts at Hobart, Darwin and Perth were superior to OCF for most days ahead. Regional Office forecasters at these capitals supplied the summary information below concerning how their forecasts were able to outperform OCF.

At Hobart there is no single methodology employed, but several approaches may have contributed to the improvement of official forecasts over OCF for maximum temperature.

(a) In the warmer months, sunny days with no sea-breeze are often much warmer than the sea-breeze affected days beforehand. OCF forecasts are too cold on these no sea-breeze days.

(b) There is a significant cooling after any cool change and OCF is too warm on these days.
(c) In winter, under anticyclonic conditions, Hobart is affected by cold air drainage down the Derwent Valley; this cold air is often not mixed out during the day and suppresses the maximum temperature and OCF tends to be too warm. These situations are recognised by official forecasters.
(d) Forecasters consider LAPS 050 provides better guidance than OCF for extreme warm days when there is little cloud and the lower levels of the atmosphere can become well mixed.
(e) The local fire weather model, which generates temperatures for fire weather locations from a climatological database using thickness, total cloud, rainfall, and wind velocity, provides useful guidance. It is used as a cross-check against model and OCF guidance by the forecasters.
(f) The number of models used in the longer term OCF is low; beyond two days ahead only GASP and ECMWF are available (see Table 1). Forecasters usually have least confidence in the GASP and thus tend to use the ECMWF component forecasts as guidance. However, for the longer range guidance, forecasters also have access to USGFS and UKGCM and can select those models which are considered the better performers.

Points (a), (b) and (c) above provide examples where local knowledge has been used to improve on OCF, while (d) and (e) show where local comparative verification may have played a similar role. Point (f) highlights the weakness in OCF evident in Table 1 and is addressed in the 'Possible improvements to OCF' subsection below.

At Darwin, both official forecast and OCF maximum temperature MAEs were the lowest of all capital cities. Nevertheless a summary of the Northern Territory Official Regional Office forecaster's assessment of OCF for maximum temperature is interesting as points (a) and (b) below reinforce other documented assessments of OCF within the Bureau of Meteorology.

The forecasters consider that OCF performs worst when:
(a) wet and cloudy conditions occur during the wet season, such as when the monsoon onset is developing or when easterly waves are approaching;
(b) evaporative cooling occurs from dry season rain clouds, which produces cooler temperatures than OCF predicts.

For day+1 maximum temperatures, Darwin forecasters often predict the time of onset of the day+1 sea-breeze by modifying the observed day+0 time of onset according to expected changes in synoptic-scale meteorological conditions (especially low-level wind velocity and stability). Sea-breeze onset effectively nullifies further temperature rises and hence governs the maximum temperature.

Perth official forecasts outperformed OCF convincingly at all projections. Possible reasons are the long lead-time between OCF issue and official forecast issue so that WA receives all 0000 UTC model guidance that is unavailable to OCF and also receives USGFS and UKGCM model predictions to day+6. However, the MAE for official maximum temperature forecast for day+1 for all 22 WA sites combined (excluding Perth) was 1.37°C and for OCF was 1.39°C. For Perth alone, the respective values were 1.13°C and 1.40°C. The Perth official next day forecast MAE was the third lowest of the State's 23 sites whereas the OCF MAE was close to the median. These results suggest:

1. Perth's official next-day maximum temperature forecasts for day+1 (at least) were anomalously accurate.
2. The impact of later numerical model data and a shorter lead time was of minor benefit to WA.

The anomalous accuracy of official forecasts at Perth is partly due to the well documented performance of OCF there. An approach for predicting Perth maximum temperatures built on that documentation was developed by one of us (G.E.) based on the observations that OCF bias varies with wind regime and time of year (similar to the summary comments (a), (b), and (c) above made by the Hobart forecasters). It differs mainly from NMOC's OCF guidance in that the Perth bias-correction of each DMO is developed from the best set of 28 closest matching pressure gradient (a surrogate for wind velocity) analogues near the corresponding day of the year over the last two years instead of using the last 30 days' bias-correction as used in NMOC.

In detail, the pressure gradient matching compares the current model forecast pressure gradient between Perth and each of three sites located about 150 km to the north, east and south with corresponding model forecast pressure gradients archived over the past two years and within 28 days of the corresponding day of the year. The matching is undertaken separately for both 0000 and 1200 UTC model runs of GASP, ECMWF, JMA, LAPS 050, LAPS 125, LAPS 375, UKGCM and USGFS to generate 140 analogues for each model (14 for the current year plus 28 for each of the two previous years multiplied by two, for the number of model run times/day). The 20 per cent best historical gradient-matched analogues are then chosen and, from these, the corresponding model biases and MAEs are computed.

The final forecast is generated as the average of the appropriately bias-corrected latest available component model forecasts ('Gradient-matched all' in Table 9) but one variation has been to include only those models whose historical MAE was less than the average MAE for all models ('Gradient-matched

50%' in Table 9). The results of a 10-month comparative verification of this experimental forecast scheme at Perth against a simplified OCF ('Raw OCF') are shown in Table 9. The simplification employed a simple mean rather than the inverse MAE weighted bias-correction described in the OCF subsection earlier.

Table 9 shows that both 'Gradient-matched all' and 'Gradient-matched 50%' outperformed the 'Raw OCF' (i.e. using latest 28-day bias-correction) for both maxima and minima. All results in Table 9 use the same model runs and so there were no lead-time advantages for Perth.

The improvement of 'Gradient-matched OCF' over 'Raw OCF' was greater for maxima in summer and minima in winter. The improvement is probably due to Perth's coastal location where different model biases for onshore and offshore wind regimes are likely.

Possible improvements to OCF

Pressure-gradient and day-of-year analogues. The use of pressure-gradient and day-of-year analogues, as employed in Perth, provided a marked improvement over a simplified OCF. The improvement is believed to be due to Perth's coastal location where different model biases for onshore and offshore winds are likely. If this assumption is generally applicable, the methodology may improve OCF for many coastal sites (including Adelaide, Brisbane, Hobart, Melbourne and Sydney) where a marked contrast in temperature between onshore and offshore wind regimes can prevail.

Table 9. Mean absolute error (MAE, °C) of a 10-month trial of OCF modifications at Perth. Raw is similar to the NMOC OCF but excludes performance weighting and uses bias-correction learnt over the last 28 rather than 30 days. 'Gradient-matched all' uses bias-correction learnt from 20% of the closest pressure gradient analogues for each component model. 'Gradient-matched 50%' is similar to 'Gradient-matched all' but uses only those component models with a historical MAE less than the mean of all MAEs in the learning period.

	Raw OCF	*Gradient-matched all*	*Gradient-matched 50%*
Maximum	1.46	1.20	**1.15**
Minimum	1.42	**1.28**	1.30

OCF MOS. The conclusion of Cheng and Steenburgh (2007), that MOS responds more quickly than a running seven-day bias-correction to changing synoptic conditions, suggests a promising approach that may be more generally implemented than using analogues. Developing a MOS forecast based on OCF daily and hourly fields (Engel 2005) with only a small number of potential predictors (to make frequent updating viable) may offer a model-independent MOS system as a more generally applicable alternative to analogue matching.

Rationalising the models available to OCF. Woodcock and Engel (2005) and the Hobart forecasters have noted the small number of component forecasts in OCF beyond day+2. Many studies have shown that forecasts can be improved by combining unbiased estimates (e.g. Winkler et al. 1977) and have shown that the incremental improvement in accuracy as the number of contributing forecasts increases declines asymptotically. The impact of adding component forecasts to OCF was undertaken in a recent study applying OCF to forecasts of sea-state parameters from ten international and relatively independent numerical models. RMSE for all successive combinations of n (ranging from 2 to 9) out of ten models was derived for both significant wave height forecasts and mean wind speed measured by moored buoys. An equation of the form

$$RMSE = AB^{n} + C \qquad ...8$$

where RMSE is the average root mean square error of the n combinations and A, B (<1.0) and C are constants, provides a close to perfect fit to the experimental results for both significant wave height and mean wind speed. As Eqn 8 shows, RMSE asymptotically approaches C as the number of models increases and, importantly here, if the number of OCF components decreases, then the RMSE of OCF forecasts increases.

Table 10 lists the OCF skill relative to official forecasts at the capital cities, based on MAEs, for both maxima and minima together with the number of component models available to OCF. Skill (S) is defined as

$$S = 100\,(MAE_{OFF} - MAE_{OCF})\,/\,MAE_{OFF} \qquad ...9$$

For maximum temperature, Table 10 shows OCF improves in accuracy relative to the official forecasts from day+1 to day+2 then deteriorates markedly from day+2 to day+3 before improving steadily from day+3 onwards. Likewise for minima forecasts, OCF skill relative to official forecasts of minimum temperatures increases from day+1 to day+2 ahead then declines quickly over day+3 and day+4. Beyond that the two schemes converge slowly. The marked relative deterioration in OCF corresponds to the marked decline in the number of component forecasts available to OCF as Eqn 7 suggests.

This trend in objective relative to official forecast accuracy differs from the USA results (Dallavalle and Dagostaro 2004; Baars and Mass 2005) where official forecast errors grow faster than model-based forecast errors. The official forecasts are often superior for day+1 but model guidance forecasts become increasingly superior as the forecast lead-time increases. Hence, the evidence suggests that OCF's loss of skill relative to official forecasts is due to its loss of contributing models. If so, it is easily remedied.

Despite this argument for increasing the numerical models for the forecasts beyond day+2, there may be some improvement in OCF if the number of models at day+1 and day+2 were actually decreased. This is because it is important that the contributing DMOs are as independent as possible (Gupta and Wilton 1987). This means they should exhibit acceptably low error correlations. Hence, a potential risk in the current OCF configuration in Table 1 is of a very high error correlation across slightly different configurations of LAPS (LAPS 375, LAPS 125 and LAP 050) resulting in too much weight being given to essentially the same model and correspondingly insufficient weight to the more independent models. Woodcock and Greenslade (2007) addressed a similar problem with Australian and international models used in OCF to predict significant wave heights. They found that replacing the DMOs from the three Australian models by their average improved the OCF forecasts. A similar result may apply here.

Table 10. The skill of OCF relative to official forecasts. Negative skill indicates OCF is worse than official forecasts. Skill increases as OCF improves relative to official forecasts.

	Maximum temperature		*Minimum temperature*	
Days ahead	*Components*	*OCF skill*	*Components*	*OCF skill*
1	10	-7	10	3
2	5	-3	6	10
3	3	-9	4	6
4	3	-7	3	3
5	3	-2	2	2
6	2	-1	1	1

Possible improvements to official forecasts

Documentation of OCF performance. Summaries from official forecasters suggest that it is important to document synoptic situations where OCF performs well and where it performs badly. Then, forecasters can use the documentation to modify OCF when formulating their official forecast as the need arises. Most Regional Forecast Centre forecasters are aware of the synoptic conditions where OCF performs poorly but the extent to which it is formally documented and included in local forecast formulation policy is unknown. The normal expectation would be that under quiescent conditions OCF produces forecasts that need little or no modification.

Give more weight to OCF in official forecasts. The result that OCF minima were more accurate than official forecasts at 74 per cent of sites for day+1 and even at the capital cities at day+1 to day+6 indicates that at many sites the official minimum temperature forecasts would have been more accurate if OCF were used unchanged. Similarly, at 63 per cent of regional centres, the official maximum temperature MAEs would have been lower if the OCF were used instead. To a lesser degree official maximum temperature forecasts would have improved also.

Combine OCF and official forecasts. One method of giving more weight to OCF within the formulation of the afternoon official forecasts is provided here. Woodcock and Southern (1983) showed that it was possible to improve official forecasts using predetermined linear regression equations to combine the existing official forecasts with MOS guidance to form a new forecast. The combined forecasts reduce large errors in official maximum (minimum) temperature forecasts by about 30 per cent (50 per cent) on independent data. Since that study, frequent upgrades to LAPS 375 have resulted in a deterioration of its MOS guidance and the combination of MOS and official forecasts is more problematic. However even a simple combination of OCF and an official forecast is worth investigating.

Table 11 shows the result of taking the average of the 0000 UTC OCF and the morning official forecast (issued about 11 am in eastern States). Both the morning official forecast and 0400 UTC OCF are available before the afternoon official forecast is issued. The results for minima forecasts show that the combination of already existing official forecasts and the available guidance OCF is more accurate than the afternoon official forecast. This is perhaps not too surprising since the 0000 UTC OCFs are mostly better than the later official forecasts. What is more surprising is that the combined forecasts at day+2, day+5 and day+6 have lower MAEs than the later afternoon

Table 11. Summary of combined OCF 0000 UTC and official 0000 UTC maxima and minima temperature forecast verification over all capital cities combined. Results in bold font are where the verification results for the combined forecasts were better than the subsequent 1200 UTC official forecast results.

Forecast	*Days ahead*	*Cases*	*Bias (°C)*	*MAE (CO)*	*Exp Var (%)*	*%>2.5 (°C)*	*%>4.5 (°C)*
Maxima	1	2920	0.02	1.31	86	14	2
	2	2920	0.01	**1.47**	**83**	18	4
	3	2920	0.02	1.67	79	23	5
	4	2615	0.05	1.78	73	24	6
	5	1857	0.04	**1.97**	**70**	**29**	**9**
	6	1829	-0.07	**2.15**	**65**	**32**	**12**
Minima	1	2920	-0.06	**1.24**	**89**	**12**	**1**
	2	2919	-0.05	**1.41**	**85**	**17**	**2**
	3	2919	-0.01	**1.54**	**76**	**25**	**5**
	4	2614	0.01	**1.70**	**76**	**25**	**5**
	5	1857	-0.05	**1.95**	**64**	**31**	**8**
	6	1829	-0.05	**2.09**	**60**	**34**	**9**

capital city official forecasts. At day+3 and day+4 the combined forecast MAEs (1.67°C and 1.78°C, respectively) were only marginally worse than for the official forecasts (1.64°C and 1.75°C).

Concluding remarks

Comparative verification of OCF and official forecasts has helped identify their relative strengths and weaknesses. As a result it has been possible to identify promising areas to improve both. These are:

OCF may be improved by:

- basing OCF bias-correction on pressure-gradient-matched (or wind velocity) and day-of-year analogues and/or building an OCF MOS system with a small number of potential predictors;
- including more component models that extend beyond day+2 in the OCF ensemble and replacing the LAPS models in OCF by their average.

The comparative verification results offer some clues to future improvements to official forecasts or at least where resources could be released when necessary.

At the capital cities where official forecasts performed considerably better than OCF there has been substantial investment in documenting where and when OCF performs relatively poorly and this information is used to improve on their use of OCF. This approach seems to be both important and successful.

OCF minima forecasts appear to be more accurate than official forecasts at most sites including capital cities, suggesting that official forecast policy should put more weight on the OCF guidance. To a lesser extent the same suggestion applies to official maximum forecasts for regional centres.

Official forecast policy may also benefit from the knowledge that OCF forecasts at day+2 are more accurate than day+1 relative to official forecasts.

At capital cities, the afternoon minimum temperature forecast MAEs for the next day onwards rarely improve on the MAEs for the average of the official morning forecast and the 0400 UTC OCF while the afternoon maximum temperature MAEs at day+1 are the only forecasts that are substantially better than the average of the morning official forecast and the 0400 UTC OCF.

Acknowledgments

Many thanks are due to Andrew Amad-Corson for modifying the forecast database record structure to facilitate this study, and to John Flook for generating the capital city daily climatology and the official forecast and OCF verification statistics for the regional centres. Tom Durrant kindly provided the analysis of the impact on OCF accuracy resulting from a change to the number of its component models.

Malcolm Downing (Tasmania/Antarctica Regional Office, Hobart) and Andrew Tupper (Northern Territory Regional Office, Darwin) coordinated the summarising comments from their colleagues giving their reasons for the official forecasts outperforming OCF. David Williams (NSW and ACT Regional Office) documented the performance of OCF under cloudy and rainy conditions over inland Australia. Beth Ebert, Michael Foley and Tony Bannister provided valuable and constructive reviews of the original manuscript. Two anonymous reviewers provided additional helpful suggestions.

References

Baars, J.A. and Mass, C.F. 2005. Performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *Weath. forecasting, 20*, 1034-47.

Brooks, H.E. and Doswell C.A. 1996. A comparison of measures-orientated and distribution-orientated approaches to forecast verification. *Weath. forecasting, 11*, 288-303.

Cheng, W.Y.Y. and Steenburgh, W.J. 2007. Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Weath. forecasting, 22*, 1304-18.

Clemen, R.T. 1989. Combining forecasts: A review and annotated bibliography. *Int. J. Forecasting, 5*, 559-83.

Dallavalle, J.P. and Dagostaro, V.J. 2004. Objective interpretation of numerical weather prediction model output — A perspective based on verification of temperature and precipitation guidance. *Preprints, Symp. On the 50th Anniversary of Operational Numerical Weather Prediction*, College Park, MD, Am. Met. Soc. CD-ROM, 6.1.

Engel, C. 2005. Hourly Operational Concensus Forecasts (OCF), *BMRC Research Report No. 115*, Bur. Met., Australia.

Glahn, H.R. and Lowry, D.A. 1972. The use of model output statistics (MOS) in objective weather forecasting. *Jnl appl. Met., 11,* 1203-11.

Gupta, S. and Wilton, P.C. 1987. Combination of forecasts: An extension. *Manage. Sci., 33*, 356–72.

Mason, I. 1982. A model for assessment of weather forecasts. *Aust. Met. Mag., 30*, 291-303.

Mills, G.A. and Tapp, R.G. 1984. The Australian operational MOS forecast system. *Tech. Rep. 56*, Bur. Met., Australia, 50 pp.

NMOC 2005. Operational Consensus Forecasts (OCF) (31 March 2005). http://web.bom.gov.au/nob/nmoc/stan/opsbull/apob60/apob60.shtml.

Stensrud, D.J. and Yussouf, N. 2005. Bias-corrected short-range ensemble forecasts of near surface variable. *Met. Appl., 12*, 213-30.

Winkler, R.L., Murphy, A.H. and Katz, R.W. 1977. The consensus of subjective probability forecasts: Are two, three, …, heads better than one? *Preprint. 5th Conference on Probability and Statistics, No. 15-18*, Las Vegas, Nevada, USA, Am. Met. Soc., Boston, MA, 57-62.

Wonnacott, T.H. and Wonnacott, R.J. 1972. *Introductory Statistics*. Wiley, NY, 510 pp.

Woodcock, F. and Southern, B. 1983. Use of linear regression to improve official temperature forecasts. *Aust. Met. Mag., 31*, 57-62.

Woodcock, F. 1984. Australian experimental model output statistics forecasts of daily maximum and minimum temperature. *Mon. Weath. Rev., 112*, 2112-21.

Woodcock, F. and Engel, C. 2005. Operational Consensus Forecasts. *Weath. forecasting, 20*, 101-11.

Woodcock, F. and Greenslade, D.J.M. 2007. Consensus of numerical model forecasts of significant wave heights. *Weath. forecasting, 22,* 792-803.