Aust. Met. Mag. 31 (1983) 57-62 551.524.3.

The use of linear regression to improve official temperature forecasts

F. Woodcock and B. Southern, Head Office, Bureau of Meteorology, Melbourne
(Manuscript received June 1982; revised October 1982)

Model output statistics (MOS) forecasts of daily maximum and minimum temperature are combined with Regional Forecast Centre (RFC) forecasts and 3 pm local observations to produce Local Forecast Equations (LFE) for daily maximum and minimum temperatures for Adelaide, Brisbane, Canberra, Hobart, Melbourne, Perth and Sydney. Testing of the LFE, using independent data, indicates that a substantial improvement on current operational accuracy can be achieved. Specifically, the number of errors exceeding 3°C in the operational forecasts of maximum temperatures can be reduced by about 30 per cent and of minimum temperatures by about 50 per cent. The improvement in the forecasts is mostly due to the inclusion of MOS forecasts in the LFE and partly due to the statistical correction of RFC forecasts. A change to current operational forecast strategy is suggested.

Introduction

It has long been known that it is possible to reduce the error variance in forecasts by the linear combination of two or more independent forecasts of the same event, and Thompson (1977) has suggested that reductions of up to 20 per cent in error variance are theoretically possible using such a scheme. Currently the only forecasts of maximum and minimum temperature available in real time for the capital cities Adelaide, Brisbane, Canberra, Hobart, Melbourne, Perth and Sydney are those issued by Regional Forecast Centres (RFC) in those cities near 5 pm local time for minima and maxima which usually occur near 6 am and 3 pm respectively the following day.

An analysis of official forecast errors for these cities for the mid-season months (January, April, July and October) of 1980 is shown for maximum temperatures in Table 1 and for minimum temperatures in Table 2. In these Tables the Priestley (1945) skill score (P) is defined as:

$$P = 1 - E^2/\sigma_0^2 \qquad \ldots 1$$

where E^2 is the mean square error of the forecast (°C^2) and σ_0^2 is the variance of the observed values. If the mean value of the observations were used as a forecast every day then E^2 equals σ_0^2 and P equals zero indicating no skill; if each daily forecast was accurate, E^2 would equal zero and P would be one. Hence, a series of skilful forecasts would score values of P between zero and one.

Experimental Model Output Statistics forecasts (MOS) of daily maxima and minima for these cities were derived and described by Woodcock (1982). Briefly, these MOS equations were derived using a forward stepwise regression screening of historical records from December 1977 to June 1981 with the mid-season months of 1980 (as in Tables 1 and 2) being excluded from the screening and used as an independent test data set. These MOS equations use Australian region grid-point analysis (Seaman et al. 1977) information valid at approximately 9 am local time, capital city observations at the same time and Australian region grid-point prognoses (McGregor et al. 1978) initialised from the analyses and projected 24 hours. The MOS forecasts would be available as guidance material in the RFC near 2 pm local time. Thus the RFC forecasts can use noon and some 3 pm synoptic information unavailable to the MOS forecasts. The error analysis of the MOS forecasts for maxima and minima for the capital cities during the mid-season months of 1980 is shown in Tables 1 and 2.

There are several operational problems posed by the results shown in these tables. For example, should the RFC forecasts be discontinued at Brisbane, Canberra and Hobart and the MOS forecasts issued instead? Or, should Adelaide RFC ignore the less accurate MOS guidance? More generally, it could be asked, how can the guidance material best be exploited to improve the operational temperature forecasts?

A further problem also arises since the MOS forecasts do not use the latest available synoptic information and cannot do so in the current operational configuration as there is insufficient time to receive the later six hours of synoptic data in the National Meteorological Analysis Centre (NMAC), Melbourne, run MOS equations and transmit the results to the RFC to meet the deadline for forecast issue.

One approach has been the provision of objective guidance based on MOS forecasts and later local observations (Glahn 1980). However, in view of the improvement in accuracy theoretically obtainable by combining forecasts, it is considered desirable to

Table 1. Error analysis of official MOS (in parenthesis) daily maximum forecasts for the mid-season months of 1980.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.43 (2.33)	1.86 (1.68)	2.32 (2.31)	1.83 (1.79)	2.09 (2.70)	1.87 (2.01)	1.52 (1.89)
Mean square error (°C²)	4.18 (9.32)	6.04 (4.71)	9.58 (8.02)	5.67 (4.82)	7.31 (11.62)	6.66 (7.27)	5.02 (6.89)
Priestley skill score	0.61 (0.00)	0.22 (0.32)	0.25 (0.30)	0.29 (0.42)	0.41 (0.19)	0.55 (0.44)	0.38 (0.15)
Correlation between forecast and observed	0.81 (0.62)	0.48 (0.63)	0.52 (0.64)	0.58 (0.67)	0.64 (0.51)	0.78 (0.73)	0.64 (0.61)

Table 2. Error analysis of official and MOS (in parenthesis) daily minimum forecasts for the mid-season months of 1980.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.30 (1.62)	1.14 (1.12)	2.04 (2.00)	1.54 (1.32)	1.50 (1.37)	1.52 (1.53)	1.30 (1.07)
Mean square error (°C²)	3.62 (3.99)	2.79 (2.02)	6.44 (6.08)	4.04 (2.63)	3.90 (3.28)	4.10 (3.51)	2.51 (1.75)
Priestley skill score	0.60 (0.54)	0.11 (0.29)	0.33 (0.49)	0.13 (0.51)	0.49 (0.51)	0.28 (0.38)	0.08 (0.51)
Correlation between forecast and observed	0.70 (0.77)	0.53 (0.65)	0.62 (0.72)	0.52 (0.74)	0.67 (0.76)	0.69 (0.69)	0.47 (0.76)

include the official forecasts in the statistical processing. In this paper, therefore, linear regression screening is used to combine the RFC forecasts, the MOS forecasts and the capital city 3 pm observations to produce new predictive equations, the Local Forecast Equations (LFEs) for use in the RFC for the following day's minimum and maximum temperature forecasts. It is assumed that the RFC forecasters would insert the appropriate 3 pm observations and their own forecasts into the LFE and obtain revised forecasts. One fundamental requirement of this approach is that the RFC forecasters do not use the MOS guidance in formulating their forecasts. Whilst this procedure is a radical departure from current Australian professional practice, the results of this study show such a convincing reduction of forecast errors that the proposed change warrants serious consideration.

This paper briefly describes the data and method used and discusses the results obtained.

Data and screening

The period of data for this study is governed by the period of prognosis archives available, that is, from December 1977 to December 1981 inclusive.

MOS forecasts for both maximum and minimum temperatures were derived by applying Woodcock's (1982) equations (based on the December 1977 to June 1981 data but excluding the mid-season months of 1980) to the archived real time Australian region analysis and prognosis grid-point fields and to the 9 am archived capital city observations for the December 1977 to December 1981 period. RFC forecasts for this period were obtained from information supplied by the RFC to the Bureau's routine verification programs whilst capital city 3 pm observations were obtained from the Bureau's Information Services Sub-section. For any one RFC the potential predictor data set vector used to derive the LFE contained the elements shown in Table 3.

Table 3. Potential predictors vector for local linear regression equation.

1	RFC forecast.
2	3 pm station level pressure.
3	3 pm surface temperature.
4	3 pm surface dew-point.
5	3 pm eastward wind component.
6	3 pm northward wind component.
7	3 pm visibility.
8	3 pm total cloud cover.
9	3 pm low cloud height.
10	3 pm total low cloud cover.
11	Total rainfall in 24 hours to 9 am.
12	MOS forecast.

Each element of the daily vector was expressed as a deviation from its monthly mean value computed for the December 1977 to December 1981 period. Finally, for each RFC the data for the mid-season months of 1980 were withheld from the regression screening and used for an independent test of the derived equations. Approximately 1400 data vectors were screened and 120 used for testing for each capital city and for each predictand.

The screening algorithms have been described by Woodcock (1982). Output from these algorithms contained those potential predictors which in combination maximised the explained variance of the predictand; the number of predictors to be selected being prescribed by the user.

In this study seven LFE were specified for each city and each predictand. The simplest LFE contained two elements, one predictor and a constant, the simplest of the remaining LFE contained two predictors and a constant and so on to the eight element LFE, containing seven predictors and a constant.

The particular potential predictors selected for inclusion in each LFE were not specified, the best predictors being chosen from the complete predictor vector by the screening algorithm.

Results and interpretation

Although LFE containing up to seven predictors were derived for both predictands it was found that little improvement was obtained in maximum forecasts by the inclusion of more than two predictors and that there was only a small improvement in the minimum forecasts gained by including additional predictors. Consequently only three-element equations are considered in the following discussion on maximum forecasts while both three and six-element equations are included for minima.

Maxima

Since the RFC and MOS forecasts were always included in the three-element LFE, coefficients of these two predictors are shown in Table 4, which is derived from the dependent data. Comparison of this table with Table 1 (independent data) reveals some consistencies. At Adelaide, where the MOS maximum forecasts were inferior to the RFC forecasts (Table 1), the RFC component in the LFE (Table 4) has a larger coefficient than for the MOS component. Similar consistencies occur for Brisbane, Canberra, Melbourne and Perth. Discrepancies between the two tables for Hobart and Sydney suggest that for the mid-season months of 1980 the relative skill of the RFC and MOS forecasts was reversed from that in the dependent sample.

Table 4. Coefficients of RFC and MOS forecasts in the three-element LFE.

	Maxima		
	RFC	MOS	Constant
Adelaide	0.80	0.23	0.02
Brisbane	0.45	0.66	-0.01
Canberra	0.44	0.56	-0.02
Hobart	0.58	0.43	0.02
Melbourne	0.67	0.37	0.01
Perth	0.74	0.21	0.02
Sydney	0.40	0.63	-0.03
	Minima		
	RFC	MOS	Constant
Adelaide	0.61	0.39	-0.00
Brisbane	0.28	0.72	-0.02
Canberra	0.40	0.62	-0.00
Hobart	0.51	0.47	-0.02
Melbourne	0.43	0.60	0.00
Perth	0.50	0.48	0.01
Sydney	0.40	0.61	-0.01

An analysis of maximum temperature forecast errors for the three-element LFE is shown in Table 5, together with the percentage improvement over the RFC errors. It is possible to test the statistical significance of the improvement of the LFE forecasts over the RFC forecasts for all cities combined using a one-tail, matched-pair t-test. This test was applied to each of the four verification parameters separately as follows. The mean of a verification parameter for each city over the four test months was obtained for both Bureau and LFE forecasts and their difference (d) calculated. The mean ($\bar{d}$) and standard deviation (σ_d) of these differences over the (n=7) capital cities was obtained and the t-value computed from

$$t = \bar{d}\, n^{1/2} / \sigma_d \qquad \ldots 2$$

From the percentage points of a t_{n-1} distribution, the significance of the difference of $\bar{d}$ with respect to zero is derived. The assumption required for validity of this t-test is that the values of d are independent and approximately normally distributed; neither were tested but are probably satisfied.

Although the verification parameters measure different features of the forecast skill, they cannot be considered as independent. For example, two sets of forecasts with a significant difference in absolute error are also likely to show a significant difference in the mean square error gained. Nevertheless using the one-tail, matched-pair t-test, the LFE forecasts were significantly better for all cities combined at the 5 per cent level for each of the verification parameters in Table 5, despite the fact that they were worse at Adelaide and showed little difference at Sydney.

It does not immediately follow that the inclusion of MOS guidance material has improved the RFC forecasts' accuracy since the impact of regressing the RFC forecasts against the observed maxima may have improved the RFC forecasts without MOS forecasts being used at all. This factor was examined by deriving a two-element LFE (containing the RFC forecast and a constant) and testing the equation on the same independent 1980 data. This was done by rerunning the screening algorithms on the original potential predictor vectors with the MOS forecasts removed. The results are shown in Table 6. Comparison of the improvements over the RFC forecasts (in parenthesis) of Tables 5 and 6 shows that, whilst some improvement (which was not statistically significant overall) was due to the impact of the RFC regression, the major impact was due to the inclusion of MOS guidance material. The one-tail, matched-pair t-test showed that overall the LFE forecasts were significantly better than the statistically corrected RFC forecasts at the 5 per cent level for each of the verification parameters in Tables 5 and 6.

Table 5. Analysis of three-element LFE maximum forecast errors for the mid-season months of 1980. The percentage improvement over RFC forcecasts is shown in parenthesis.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.55 (-8)	1.56 (16)	1.95 (16)	1.64 (10)	2.03 (3)	1.70 (9)	1.57 (-3)
Mean square error (°C²)	4.47 (-7)	4.05 (33)	6.81 (29)	4.03 (29)	6.72 (8)	4.57 (31)	4.87 (3)
Priestly skill score	0.64 (5)	0.41 (86)	0.44 (76)	0.48 (66)	0.48 (17)	0.65 (18)	0.36 (-5)
Correlation between forecast and observed	0.82 (0)	0.64 (33)	0.66 (23)	0.69 (19)	0.68 (6)	0.82 (5)	0.68 (6)

Table 6. Analysis of maximum forecast errors of the regressed RFC forecasts for the mid-season months of 1980. Percentage improvements over the RFC errors are shown in parenthesis.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.49 (-4)	1.73 (7)	2.24 (3)	1.80 (2)	2.01 (4)	1.82 (3)	1.56 (-3)
Mean square error (°C²)	4.31 (-3)	5.14 (15)	9.11 (5)	4.96 (13)	6.83 (7)	6.25 (6)	4.90 (2)
Priestley skill score	0.63 (3)	0.22 (0)	0.29 (16)	0.36 (24)	0.46 (12)	0.57 (4)	0.35 (-8)
Correlation between forecasts and observed	0.81 (0)	0.47 (-2)	0.54 (4)	0.61 (5)	0.66 (3)	0.77 (-1)	0.64 (0)

Table 7. Number of maximum forecast errors exceeding T °C for RFC and LFE (in parenthesis) for the mid-season months of 1980.

T	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
3	7 (13)	18 (12)	23 (15)	20 (11)	19 (19)	14 (9)	12 (12)
4	3 (6)	12 (3)	18 (8)	6 (0)	13 (10)	10 (5)	9 (8)
5	2 (2)	6 (2)	11 (7)	2	7 (7)	4 (3)	5 (6)
6	1 (1)		5 (4)	1	4 (2)	4 (0)	3 (2)
7			5 (2)		1 (0)		
8			3 (1)				
9			1 (0)				
10							

Another aspect of forecast quality not highlighted in the previous analyses is the distribution of large forecast errors. Since these errors attract most public criticism of the Bureau, an analysis of these is important. It is essential that a new technique improves the forecasts where the existing methods fail rather than improve an already acceptable product. A comparison of RFC and the three-element (RFC and MOS components plus a constant) LFE is shown in Table 7 and the combined city results in Fig. 1. These results show that the LFE forecasts are a substantial improvement over the operational RFC maximum forecasts. The number of large errors has been reduced by approximately 30 per cent. Even at Adelaide, where the RFC forecasts are most accurate, the LFE forecasts make the same number of errors greater than 5°C as the RFC forecasts.

Minima

Comparing minimum forecasts for the RFC and the three-element LFE equations incorporating the

Fig. 1 Distribution of large errors of maximum temperature forecasts for RFC (x) and LFE (·) for all cities for the mid-season months of 1980.

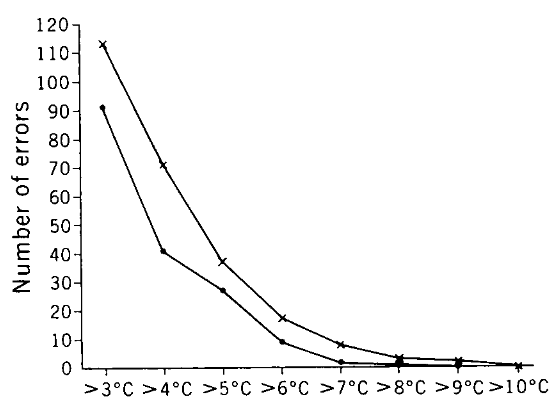

Table 8. Analysis of LFE minimum forecast errors for the mid-season months of 1980. The percentage improvement over the RFC forecast errors are shown in parenthesis.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.26 (3)	1.04 (9)	1.90 (7)	1.17 (24)	1.21 (19)	1.40 (8)	1.01 (22)
Mean square error (°C²)	2.73 (25)	1.68 (40)	5.35 (17)	2.08 (49)	2.42 (38)	2.87 (30)	1.51 (40)
Priestly skill score	0.69 (15)	0.38 (245)	0.55 (67)	0.58 (346)	0.62 (27)	0.47 (68)	0.58 (488)
Correlation between forecast and observed	0.84 (6)	0.66 (25)	0.74 (19)	0.78 (50)	0.81 (21)	0.74 (7)	0.79 (68)

Table 9. Analysis of minimum forecast errors of the regressed RFC forecasts for mid-season months of 1980. Percentage improvements on the RFC forecast errors are shown in parenthesis.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.31 (-1)	1.19 (-4)	2.15 (-5)	1.27 (18)	1.39 (7)	1.45 (5)	1.18 (9)
Mean square error (°C²)	3.22 (11)	2.31 (17)	6.46 (0)	2.47 (39)	3.26 (16)	3.23 (20)	2.15 (14)
Priestley skill score	0.63 (5)	0.05 (-54)	0.45 (36)	0.49 (277)	0.51 (4)	0.43 (54)	0.42 (425)
Correlation between forecast and observed	0.80 (1)	0.54 (2)	0.67 (8)	0.73 (40)	0.75 (12)	0.72 (4)	0.68 (45)

RFC and MOS components indicates some consistency between Tables 2 and 4. Thus, the relative magnitude of the coefficients of MOS and RFC components (Table 4) reflects the accuracy of the two forecasts in Table 2. At Hobart and Perth there are discrepancies possibly due to a reversal of relative skill of the two forecasts in the developmental and 1980 test periods.

Table 8 shows the analysis of errors of the three-element LFE for the mid-season months of 1980 together with the percentage improvements of these forecasts over those of the RFC (from Table 2). The LFE forecasts were superior in every respect to the RFC forecasts at the 0.5 per cent significance level of a one-tail, matched-pair t-test of station means for all cities combined.

Again it is necessary to consider the improvement due to regression screening of the RFC forecasts before the impact of MOS forecasts is ascertained. This analysis is shown in Table 9. Comparing Tables 9 and 8, percentage improvement over the RFC errors shows that the inclusion of the MOS guidance forecasts was an important factor causing the overall improvements brought about by the three-element LFE. Overall, the LFE forecasts were significantly better at the 5 per cent level than the statistically corrected RFC forecasts for each of the verification parameters in Tables 8 and 9.

The distribution of large errors in the minimum forecasts is shown in Table 10 for each city and the combined result for all cities is shown in Fig. 2. Approximately 50 per cent of the large errors in RFC forecasts can be eliminated by using the LFE strategy.

Since the usual time of occurrence of the minima (near 6 am) falls between the 3 pm observational data and the validity time of the prognosis (9 am), it is likely that the observational data would have more impact on the accuracy of the LFE minimum forecasts than they did for maxima. The particular elements of the observational data included in the LFE by the screening algorithm varied with the city, the number of elements required to give the best forecasts over the test period similarly varied.

For brevity, the final discussion concerns the six-element LFE which was chosen for each capital city by the screening algorithms. This was not necessarily the best LFE for an individual city, but performed best overall on the test data.

The most common 3 pm observational elements chosen were surface temperature and total or low cloud cover. Table 11 shows the error analysis of the

Fig. 2 Distribution of large errors of minimum temperature forecasts for RFC (x) and LFE (·) for all cities for mid-season months of 1980.

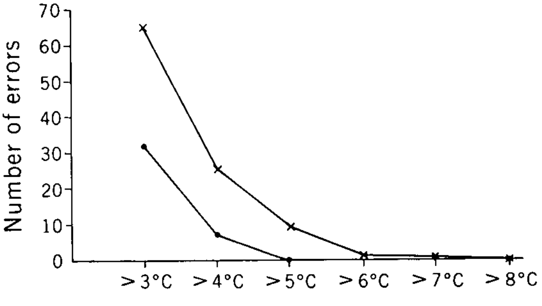

Table 10. Number of minimum forecast errors exceeding T °C for RFC and LFE (in parenthesis) for mid-season months of 1980.

T °C	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
3	9 (6)	3 (1)	23 (16)	10 (1)	7 (3)	11 (3)	2 (1)
4	4 (3)	1 (0)	10 (4)	2 (0)	3 (0)	4 (0)	1 (0)
5	2 (0)	1	2 (0)		2	1	1
6		1					
7		1					

Table 11. Error analysis of six-element LFE daily minimum forecasts for the mid-season months of 1980. The percentage improvement over the RFC is shown in parenthesis.

	Adelaide	Brisbane	Canberra	Hobart	Melbourne	Perth	Sydney
Mean absolute error (°C)	1.29 (1)	0.94 (18)	1.86 (9)	1.16 (25)	1.22 (19)	1.30 (14)	1.00 (23)
Mean square error (°C²)	2.83 (22)	1.43 (49)	5.12 (20)	2.00 (50)	2.41 (38)	2.51 (39)	1.49 (41)
Priestly skill score	0.67 (12)	0.49 (345)	0.57 (73)	0.59 (354)	0.62 (27)	0.55 (96)	0.60 (650)
Correlation between forecast and observed	0.83 (5)	0.72 (36)	0.76 (23)	0.79 (52)	0.78 (16)	0.78 (13)	0.78 (66)

six-element LFE and the percentage improvement over the RFC minimum forecast errors (from Table 2). Comparison of the percentage improvement of the six-element LFE minimum forecast errors over the corresponding three-element LFE values (Table 8) shows that a small but consistent improvement is brought about by using the 3 pm observational data. Overall, only for the mean square errors and the Priestley skill score were the six-element LFE forecasts significantly better at the 5 per cent level than the three-element LFE forecasts.

The large error analysis for the six-element LFE is not shown; there were 30 errors greater than 3°C and six errors greater than 4°C, a slight improvement on the three-element LFE.

Conclusions

This study has shown that a substantial reduction of errors in the operational RFC forecasts of daily maxima and minima can be achieved using LFE incorporating the RFC and MOS forecasts for maxima and these two elements plus some 3 pm observations for minima. The major reduction in error is due to the use of MOS forecasts and a minor reduction due to the statistical correction of the RFC forecasts.

Although the effect of using surface observations was small, experience elswhere (Hammons et al. 1976) has been that the importance of these observations varies considerably with season. As further prognosis data become available it will be possible to produce forecasts using seasonally derived equations, thereby reducing errors even further.

Use of this system requires that RFC forecasters make their forecasts without knowledge of the MOS guidance and then include them as a component of the LFE forecasts. This departs from established practice, but such a procedure would be easily implemented when MOS forecasts become available in real time and is well suited for inclusion in an automated system of RFC operations. The impressive improvement shown suggests that serious consideration be given to adopting this proposed strategy.

Acknowledgments

The authors are indebted to Alan Miller, CSIRO Division of Mathematics and Statistics, PO Box 310, South Melbourne, Victoria, for use of his regression software, Neville Nicholls (ANMRC), Peter Price and Tom Keenan (Bureau of Meteorology) for their reviews and comments. Mary Smedley kindly typed the manuscript.

References

Glahn, H. R. 1980. Plans for the development of a local AFOS MOS program (LAMP). *Preprints, Eighth Conference on Weather Forecasting and Analysis, Denver, Colo.* American Meteorological Society, Boston, 302-5.

Hammons, G. A., Dallaville, J. P. and Klein, W. H. 1976. Automated temperature guidance based on three-month seasons. *Mon. Weath. Rev., 104,* 1557-64.

McGregor, J. L., Leslie, L. M. and Gauntlett, D. J. 1978. The ANMRC limited area model: consolidated formulation and operational results. *Mon. Weath. Rev., 106,* 427-38.

Priestley, C. H. B. 1945. Accuracy of upper air forecasts. *Meteorological Research Paper 222,* British Met. Office, UK, 12 pp.

Seaman, R. S., Falconer, R. L. and Brown, J. 1977. Application of a variational blending technique to numerical analysis in the Australian region. *Aust. Met. Mag., 25,* 3-23.

Thompson, P. D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Weath. Rev., 105,* 228-9.

Woodcock, F. 1982. Australian experimental model output statistics forecasts of daily maximum and minimum temperature. *Bull. Am. met. Soc.* (in press).