Aust. Met. Mag. 34 (1986) 43-56 551.509.314

A diagnostic study of the first year of operational Model Output Statistics forecasts in Australia

R. G. Tapp, G. A. Mills and **G. F. McNamara**
Bureau of Meteorology Research Centre, Melbourne, Australia

(Manuscript received February 1986; revised May 1986)

Operational Model Output Statistics (MOS) forecasts of maximum and minimum temperature and precipitation quantity and probability of occurrence for seven Australian capital cities have been prepared daily since 1 October 1983. The levels of accuracy of the first twelve months of these forecasts are described, and an attempt is made to partition the sources of error in these forecasts into errors in day to day prognoses, deficiencies in the development data set, and errors inherent in the MOS approach. By means of a one-month test it is demonstrated that the MOS forecasts are sensitive to changes in the error characteristics of the numerical weather prediction model used to provide the predictors. It is concluded that the factors which would have the greatest positive impact on the accuracy of MOS forecasts are improved accuracy of the day to day prognoses, and stratification of the development data by season to provide separate equations in each season for each predicted element.

Introduction

The Model Output Statistics (MOS) method for the prediction of individual weather elements based on output from numerical weather prediction models has been widely used since the early 1970s, particularly in the United States of America. Both the technique and its applications have been extensively reviewed by Glahn (1982).

MOS equations for the operational prediction of maximum and minimum temperature, precipitation amount and the probability of precipitation occurrence for seven capital cities in Australia have been developed by Mills and Tapp (1984, hereafter referred to as MT). Forecasts based on these equations were prepared daily from 1 October 1983 in a test mode using operational data sets, and have been issued operationally to Regional Forecast Centres (RFCs) from 16 January 1984. It is the purpose of this paper to document the level of accuracy of the MOS forecasts during the first year of their operational use, considered here to commence on 1 October 1983. Some of the causes of error in the MOS forecasts are determined and partitioned into three categories: those attributable to errors in individual prognoses; those associated with inadequacies in the developmental data set; and those due to the assumptions inherent in the MOS technique.

The following section of this paper gives a brief description of the operational MOS forecast equations; in subsequent sections the approach used to diagnose causes of errors in the MOS forecasts is described, the error statistics of the operational forecasts are discussed and the contributions to the operational errors from various sources are quantified. Finally, the implications of these results for future applications of MOS in Australia are considered.

The operational MOS system

The operational MOS forecasts are based on the 0000 GMT (approximately 0900 local time in the eastern States) 24-hour prognosis from the Australian Region Primitive Equations (ARPE) model, which was described by McGregor et al. (1978) and Leslie et al. (1981). In its operational application, neither the surface drag nor the rainfall parameterisations are activated. Instead, the forecast rainfall is calculated from the model output using the SLYH technique (Younkin et al. 1965), and the forecast humidity fields are adjusted during this stage.

The MOS forecasts are issued to the RFCs at approximately 1300 local time. The temperatures forecast are the minimum expected overnight and the maximum during the following day; both rainfall forecasts apply to the 24-hour period commencing at 0900 local time the following day. Thus the validity time of the ARPE prognosis approximately corresponds with the start of the precipitation forecast period.

The forecast equations were developed from a data set of analysis and prognosis fields from the years 1978-1981 inclusive. The prognoses were a combination of the ARPE archive from 1978-1980 developed by Woodcock (1984), and the operational prognoses archived by the National Meteorological Centre (NMC) of the Bureau of Meteorology commencing in 1981. The potential predictors are essentially analysis and prognosis grid-point values of geopotential height, temperature, relative humidity and southward and eastward wind components at 1000, 850, 700 and 500 mb at the station location, with various degrees of spatial smoothing, together with surface observations taken at the station at the time of the analysis. The predictand for temperature was the value observed at the capital city RFC, while for rainfall it was the average 24-hour rainfall over a network of stations in the metropolitan area in each city. The selected predictors and the rainfall verification networks are listed by MT. The development data set was not long enough to allow seasonal stratification of the data. However, in order to reduce the effect of seasonal variations, the predictors and predictands were all expressed in the form of departures from their monthly means, and the sine and cosine of the serial day of the year were also offered as potential predictors (Woodcock 1984).

The quantitative precipitation forecasts have been verified in ranges, from Category 0 (no rain) to Category 7 (in excess of approximately 80 mm in 24 hours). These ranges are those used by the Bureau of Meteorology for verification of operational (but unissued) quantitative forecasts (Tapp et al. 1986).

Sources of error in MOS forecasts and their identification

Sources of error

The MOS approach to generating forecast equations assumes that elements predicted by a broadscale NWP model at the model grid-points can be related statistically to an actual weather element which occurs at a specific location. Three main types of factors can be identified which contribute to errors in day to day MOS forecasts of a weather element.

(a) Day to day errors in prognoses. The errors in the MOS forecasts which are most readily identified are those due to an error in the NWP model forecast of a MOS predictor on a given day. There are many possible sources of errors of this type, such as the initial conditions offered to the model, lack of model resolution either vertically or horizontally, and lack of incorporation of physical processes in the model. If the prognosis is improved then errors from this source are reduced.

(b) Errors due to assumptions of the MOS approach. Errors in this category may be attributed to a failure by the selected prediction equation to satisfactorily describe the particular situation, and may be associated either with the MOS approach itself, or with the way it is applied.

The MOS approach is a statistical one and an equation derived using MOS will underforecast variance when applied to independent data, so that extreme events would be poorly predicted. Further, the predictive equations will be heavily weighted towards the factors which most frequently influence the particular weather element, so that less common situations are likely to be poorly represented.

In the present application the maximum temperature and rainfall forecasts are particularly affected by having NWP prognoses available only at the beginning of the forecast period. Thus if the normal diurnal temperature cycle is disrupted by a frontal passage or steady rainfall shortly after prognosis validity time, then the MOS temperature forecast is adversely affected. Similarly, there may well be little indication in the prognosis at the forecast location of a rain-producing system which arrives during the forecast period.

The potential predictors offered to the screening regression may not include the most suitable parameters. Only scalar variables at station locations were offered as potential predictors, and it could well be argued on physical grounds that other relevant predictors were not offered to the regression package. It has also been assumed that the relationships between the predictand and all potential predictors is linear, in spite of the non-linearity of atmospheric processes (Lorenz 1956).

No seasonal stratification of the development data was made. It has therefore been assumed that a deviation from the mean of a given predictor has the same impact on the deviation from the mean of the predictand in all seasons.

(c) Errors inherent in the prediction equations arising during development. Practical considerations preclude the acquisition of a 'perfect' development data set, which would ideally be of infinite length, contain error-free prognoses, and cover all possible synoptic situations. In practice, however, it is regarded as satisfactory if the data set is of sufficient length to enable stable equations to be developed, and it is accepted that it will contain an error distribution. It is a fundamental assumption of the MOS approach that systematic errors will be compensated for by the statistical approach. However, only general biases can be accounted for, not systematic errors associated with particular synoptic situations, unless an analogue stratification approach (Lorenz 1956; Paegle 1973) is employed.

Given that there will be an error distribution in the development data set it is important that the development data be as homogeneous as possible, and that the error characteristics of the prognoses do not change. Glahn (1982) has reviewed the possible consequences arising when the operational prognosis model undergoes changes which alter its error characteristics from those of the prognoses in the development data sets. He argues that where systematic biases have been accommodated by the MOS equations their alteration through modification of the model would degrade the resultant MOS forecasts. There must therefore be a large improvement in overall prognosis skill accompanying the change of model if this degradation is to be offset. In the Australian context there has been a change in options used in the Australian version of the ARPE model from those used in generating the bulk of the development data set, and some of the consequences of this will be addressed later in this paper.

Quantification of contributions to errors

It is possible to identify some of the sources of error in the MOS forecasts, and to quantify these errors to some extent. The errors due to day to day prognosis error can be identified by replacing those MOS predictors from the prognosis with values extracted from the objective analyses valid at prognosis verifying time and recomputing the MOS forecasts. That is, the forecast derived is that which would have been obtained had the numerical prognosis perfectly described the analysed conditions. (Note that this is not the classical 'perfect prognosis' approach (see Glahn 1982).) These forecasts are referred to as the QPP forecasts.

For the temperature forecasts perfect prognosis equations were also developed by MT. That is, analyses at prognosis verification time were used as the development sample. If used in the forecast mode, that is, with predictors from a prognosis, then the forecast skill declines, chiefly due to the error characteristics of the prognosis model not being accounted for in the screening regression process. However, if these perfect prognosis equations are applied to analyses valid at prognosis verification time they provide in hindsight a useful further diagnosis of MOS forecast error, since these forecasts eliminate the error contribution which arose because the error distribution of the prognosis data in the development data set caused less than optimum selection of predictors or a reduction in the weights given to the predictors selected. This approach was used in a slightly different context by Woodcock and Nicholls (1982). The error contribution of these APP forecasts contains the effects of the limitations of the MOS approach itself and the effects of deficiencies in the way in which MOS has been applied here.

Accuracy of operational MOS forecasts

Maximum temperature forecasts

Figures 1 (a) to (g) show the root mean square (RMS) forecast error for each month for the MOS, QPP and APP forecasts, and the monthly standard deviations for the same forecasts together with the observed standard deviation. On the RMS plot the RMS errors obtained by MT for January, April, July and October independent data withheld during equation development are shown by the barred horizontal lines.

The first conclusion to be drawn is that there are no obvious systematic and significant differences between the skill of the forecasts obtained during development and the operational (MOS) forecasts, given that there is considerable variation from year to year in the skill of these forecasts.

Secondly if, as has been proposed, the effect of the individual prognosis errors can be assessed by the difference between the MOS and the QPP forecasts, it can be seen that the skill of the MOS forecasts would have been considerably improved during the warmer months at Adelaide, Canberra, Melbourne, Hobart and Perth if the prognoses had been more accurate.

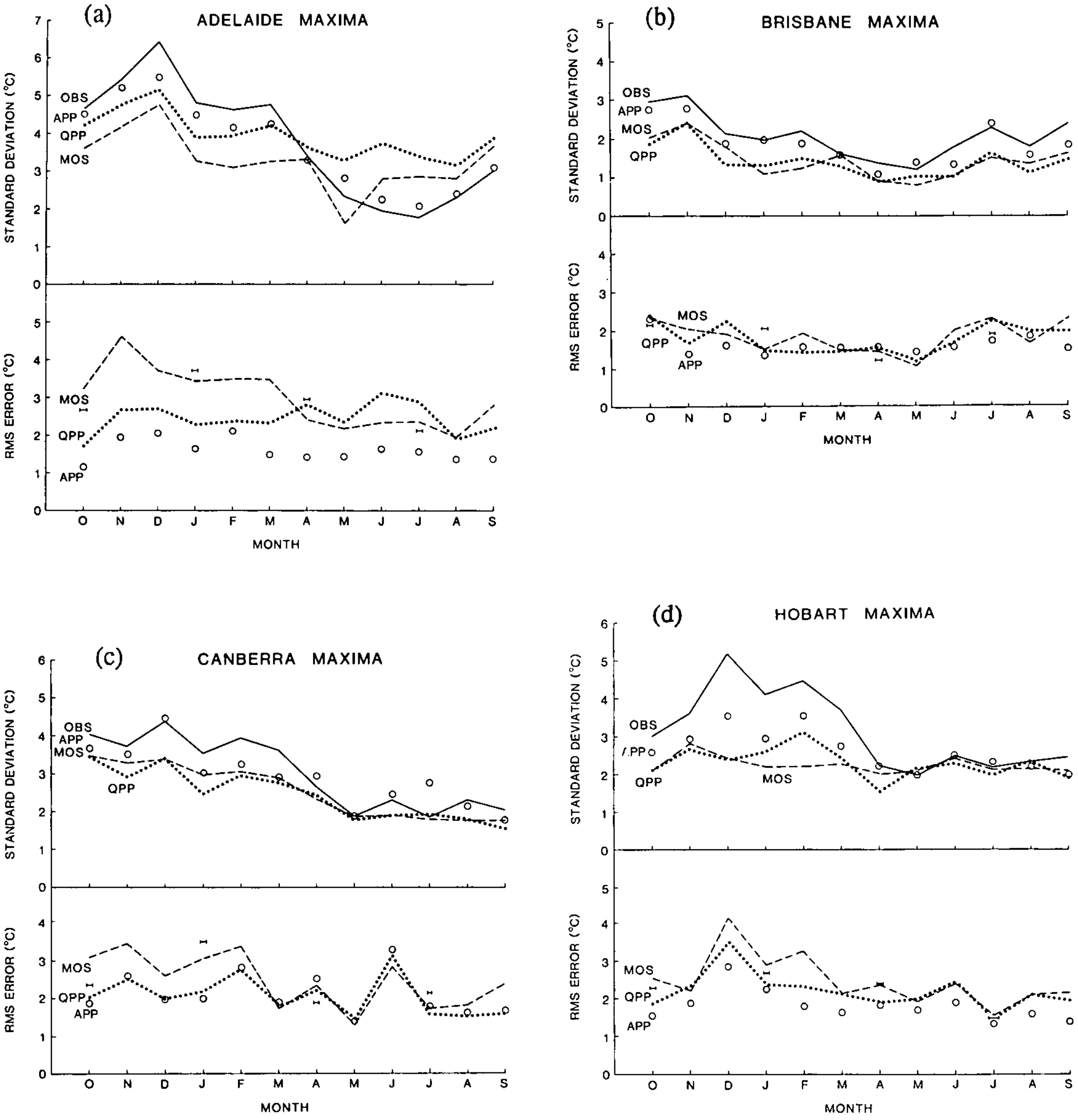
However, the skill of the QPP forecasts became less than that of the MOS forecasts during the winter months at Adelaide, Melbourne and Perth, and to a lesser extent at Canberra and Hobart. Part of the reason for this change becomes apparent when the monthly standard deviations of the MOS, QPP and APP forecasts are compared with the observed standard deviations. To use Adelaide as an example, since the effect is perhaps most marked there, it can be seen from Fig. 1 (a) that during the warmer months the standard deviations for the MOS forecasts were well below those observed, as is to be expected from a statistical scheme. Those of the QPP forecasts lay between the two, and the standard deviations of the APP forecasts were even closer to the observed values. However, while the APP forecasts followed reasonably closely the seasonal decline exhibited by the observed standard deviations, although with a somewhat damped amplitude, neither the MOS nor the QPP forecasts showed the same variability, and, indeed, the QPP forecasts had a much greater standard deviation during the winter months than was observed. Thus, the use of a single equation for the full year underestimates variance during the months when a high variance is observed and overestimates it when a low variance is observed. The error characteristics of the prognosis data set are also seen to affect the generated equations. The different characteristics of the data in the analyses give a different forecast error structure (Glahn 1982).

Finally, if the APP forecast skill indicates the level of skill which could be attained with this application of MOS if the development data set was error free, it can be seen that further considerable gains could have been achieved in Adelaide, Melbourne, Perth and Hobart. Expressed another way, a considerable component of the MOS forecast error for these cities was due to sub-optimal equations having been selected due to the errors in the prognoses in the development data set.

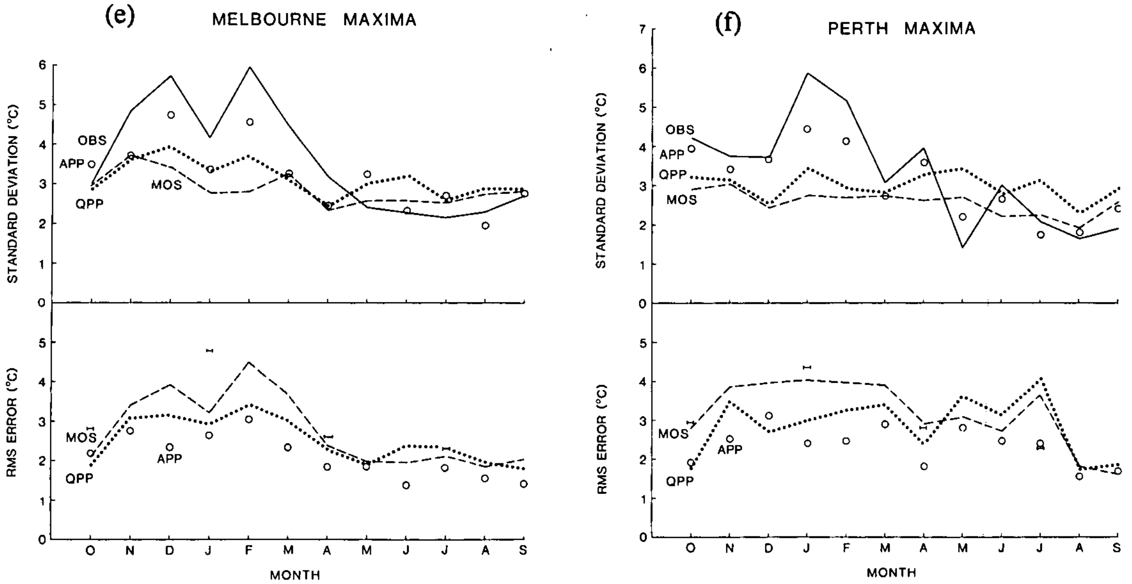
Minimum temperature forecasts

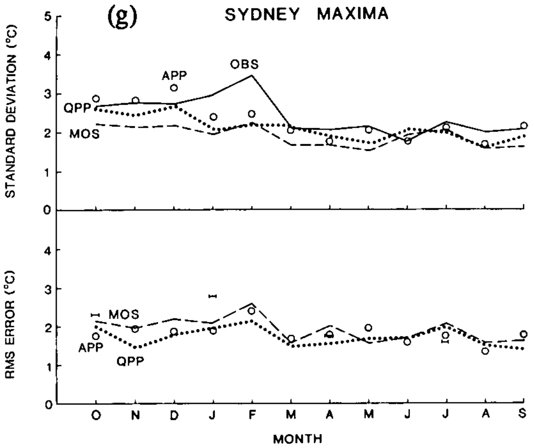
The same set of RMS error and standard deviation

Fig. 1 Root mean square errors and standard deviations of forecasts of maximum temperature for daily forecasts between October 1983 and September 1984, inclusive. Curves are presented for MOS, QPP and APP forecasts (see text for explanations). The standard deviations of the observed maxima are also shown, together with the skill achieved on independent data for January, April, July and October during equation development (horizontal bars).



statistics were prepared for the MOS, QPP and APP minimum temperature forecasts. The differences in skill between the three forecast types were less pronounced than for the maximum temperatures, so the verification statistics are presented as seasonal values in Table 1 rather than in graph form. The chief conclusions to be drawn from these results are that there was some contribution to the MOS forecast errors from day to day prognosis errors in Canberra in summer, autumn and spring, in Hobart in summer, autumn and winter, and to a lesser extent in Brisbane in summer and Melbourne in winter. There was a further contribution to the MOS errors from errors in the developmental data in Adelaide





during the whole year, in Melbourne in spring, summer and autumn, in Hobart in spring and summer, and a large effect in Perth during summer.

It is noticeable that there is much less contrast between the MOS, QPP and APP forecasts of minimum temperature than there was for the maximum temperatures. Part of the reason for this may be the lower RMS errors for these forecasts than for the maximum temperature forecasts, but an additional factor is undoubtedly the persistence component of each minimum temperature forecast equation, which reduced the overall sensitivity of the forecasts to the prognosis predictors. The accuracy of the minimum temperature forecasts is also likely to be assisted by having the prognosis predictors valid shortly *after* the typical time of occurrence of the predictand rather than *before* it as is the case with the maximum temperature, since there is a greater likelihood in the former situation that the prognosis will be representative of the conditions which influence the predictand.

Rainfall forecasts

Forecast skill. The skill of the forecasts of probability of precipitation (PoP) overall is most clearly apparent when the data are presented graphically in the form of relative operating characteristics (ROCs) where, for a range of decision criteria, the frequency of correct forecasts of the occurrence of a particular event, the 'hit rate', is plotted against the frequency with which that event is predicted but does not occur, the 'false alarm rate'. The applicability of this technique to meteorological data has been discussed in detail by Mason (1982).

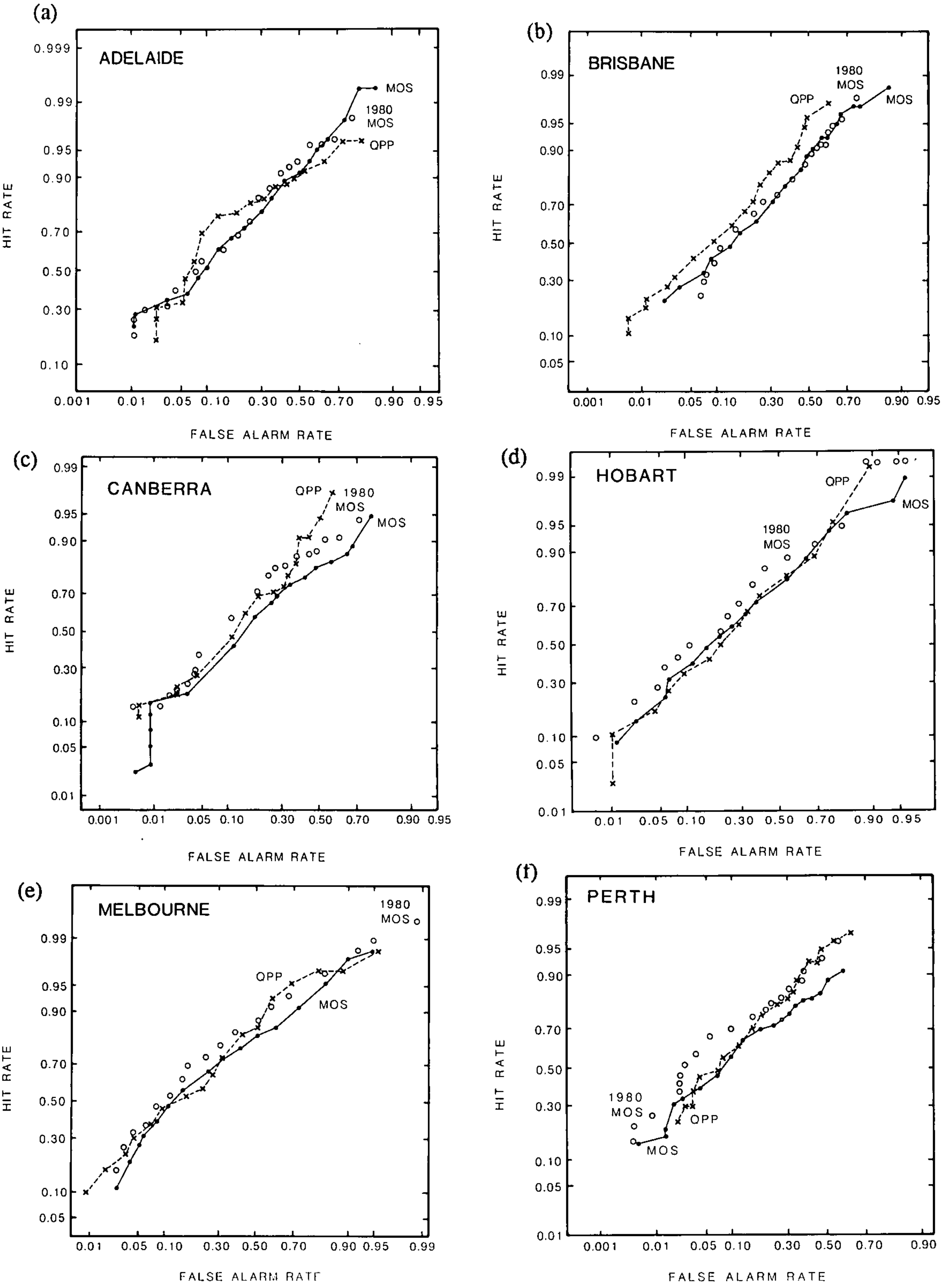
Only at Brisbane was the skill of the PoP forecasts through the period of this investigation comparable with that obtained when the developed equations were tested previously on eight months of independent data from 1980. At all other cities a drop in skill occurred, although at Adelaide the decrease was only slight (Figs 2 (a)-(g)).*

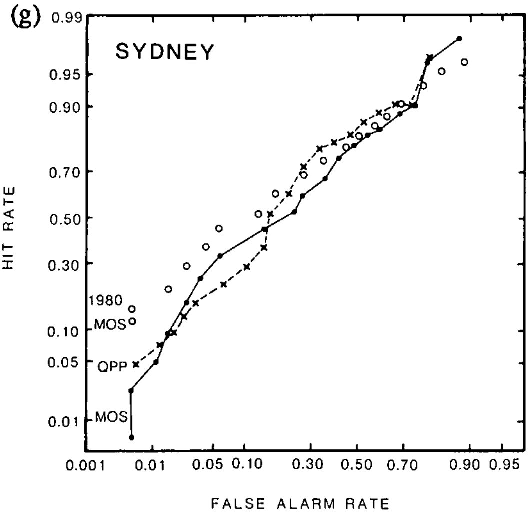
The skill which could have been achieved during 1983-84 had the day by day prognoses been error free

---

* In Fig. 2 skill increases from the lower right corner towards the upper left corner of each graph; when the hit rate and false alarm rate are equal the forecasts show no skill over predictions based on the sample climatology.

Fig. 2 Relative operating characteristics for daily MOS and QPP forecasts of probability of precipitation occurrence from October 1983 to September 1984, inclusive. Also shown is the skill achieved during equation development with forecasts for eight months of independent data from 1980 (1980 MOS).





can be seen from the additional data plotted in Figs 2 (a) to (g) for QPPs. At all cities except Hobart an overall improvement in the skill of the forecasts would have been achieved had the prognoses been correct, although only a small improvement would have occurred at Melbourne. Considering the forecasts from all cities together, this loss of skill attributable to day to day errors in the prognoses can be seen (Fig. 3) in all months to varying degrees when the levels of skill of the MOS and QPP PoP forecasts are compared in terms of Brier scores (Lowry and Glahn 1976). The dependence of the Brier score on climatology has been removed by making the assessments of skill relative to that of forecasts based on the long-term climatological probabilities of rain occurrence in each month (Glahn and Jorgensen 1970). Categorical predictions of rain occurrence generated from both the MOS and QPP forecasts using a probability of 50 per cent to delineate

Table 1. Root mean square errors (RMSE) and standard deviations (SD), in °C, of MOS, QPP and APP forecasts of minimum temperature, together with the standard deviations of the observed minima (OBS), for seven Australian cities. The data from October 1983 to September 1984, inclusive, are subdivided by season.

		Summer		Autumn		Winter		Spring	
		RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
Adelaide	MOS	2.05	2.89	2.33	1.91	2.13	1.36	1.97	2.06
	QPP	2.02	2.53	2.33	1.70	1.82	1.15	1.78	1.79
	APP	1.58	3.12	1.99	2.31	1.69	1.55	1.83	2.35
	OBS		4.14		3.30		2.28		2.95
Brisbane	MOS	1.20	1.02	1.25	1.28	1.40	1.51	1.42	1.50
	QPP	1.02	1.01	1.17	1.30	1.32	1.37	1.38	1.60
	APP	1.03	1.14	1.07	1.57	1.28	1.65	1.39	1.69
	OBS		1.35		1.74		1.90		2.00
Canberra	MOS	3.18	2.15	2.93	2.49	3.02	2.20	2.75	2.17
	QPP	2.81	2.14	2.58	2.48	3.04	1.76	2.23	2.28
	APP	2.65	2.27	2.64	2.67	2.95	1.93	2.28	2.38
	OBS		3.23		3.54		3.09		3.14
Hobart	MOS	2.13	1.60	2.02	1.48	1.84	1.56	1.71	1.61
	QPP	1.80	1.67	1.58	1.20	1.63	1.32	1.73	1.37
	APP	1.55	2.32	1.51	1.82	1.62	1.97	1.37	2.07
	OBS		2.48		1.98		2.31		2.45
Melbourne	MOS	2.05	1.87	1.99	1.63	2.02	1.64	1.94	1.72
	QPP	2.03	1.98	2.01	1.70	1.80	1.62	2.00	1.72
	APP	1.73	2.18	1.72	1.59	1.82	1.49	1.51	1.91
	OBS		2.29		2.39		2.37		2.55
Perth	MOS	1.75	2.32	2.06	1.60	2.00	1.50	1.68	1.63
	QPP	1.67	2.18	1.92	1.22	2.09	1.38	1.44	1.58
	APP	1.14	2.99	1.80	2.02	1.88	2.31	1.26	2.09
	OBS		2.92		2.46		2.39		2.19
Sydney	MOS	1.29	1.74	1.57	1.29	1.51	1.19	1.46	1.85
	QPP	1.23	1.50	1.50	1.44	1.48	1.16	1.39	1.67
	APP	1.20	1.72	1.31	1.86	1.38	1.40	1.30	1.82
	OBS		1.75		2.16		1.83		2.23

Fig. 3 The skill of daily MOS and QPP PoP forecasts for seven Australian cities combined, between October 1983 and September 1984, inclusive. The forecasts are assessed using Brier scores, expressed as skill relative to forecasts based on climatology, and Hanssen and Kuipers' skill scores for categorical predictions of rain occurrence. The bias of the categorical forecasts of rain occurrence is also shown. The skill and bias achieved during equation development for categorical forecasts from eight months of independent data from 1980 are also indicated (-·-).

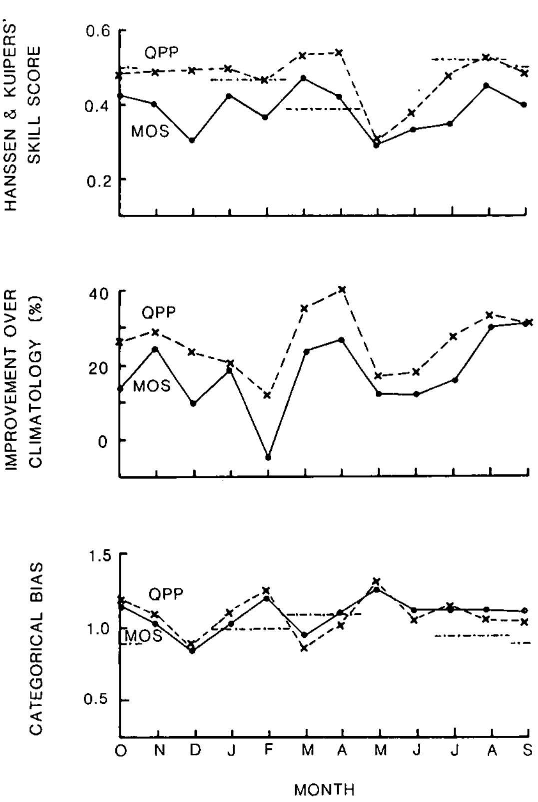
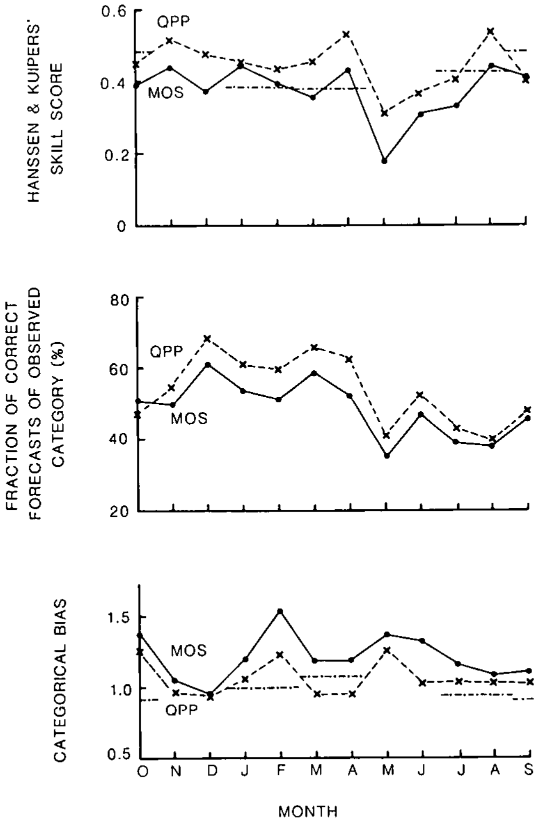


Fig. 4 The skill of daily MOS and QPP QPFs for seven Australian cities combined, between October 1983 and September 1984, inclusive. Skill is assessed in terms of the fraction of correct forecasts of observed rain category (%), and using Hanssen and Kuipers' skill scores for categorical forecasts of rain occurrence. The bias of the categorical prediction of rain occurrence is also shown. The skill and bias achieved during equation development for categorical forecasts from eight months of independent data from 1980 are also indicated (-·-).



between predictions of rain and no rain (Tapp et al. 1986) show similar comparative skill (Fig. 3), assessed in terms of Hanssen and Kuipers' skill scores (Hanssen and Kuipers 1965; Woodcock 1981). Only at Canberra and Hobart did the categorical predictions developed from the PoP forecasts show skill comparable with that previously obtained using independent data from 1980.

Categorical predictions of precipitation occurrence made using the quantitative precipitation forecasts (QPFs) would also have been better had the prognoses been perfect, and greater precision in the prediction of the rain amount would also have resulted (Fig. 4). This is further seen in the skill of predictions of larger rain totals, although the small sample sizes in individual months distort the month by month variations to some extent (Fig. 5). The 1983-84 QPFs showed slightly less skill overall than did those for 1980, but the loss of skill was less than that which occurred for the PoP forecasts. Whereas the development data and the 1980 independent data had shown that the MOS PoP predictions anticipated precipitation occurrence more skilfully than did the MOS QPFs, that superiority was considerably reduced in the 1983-84 data. This loss of skill is further revealed by the reduced reliability of the MOS PoP forecasts during the latter period (Fig. 6 cf. Fig. 1 of MT).

A decrease in skill was apparent in the prediction of all rainfall categories. The skill of the QPFs as

Fig. 5 The skill of daily MOS and QPP forecasts of 24-hour rain totals of Category 3 or greater for seven Australian cities combined, between October 1983 and September 1984, inclusive, assessed in terms of Hanssen and Kuipers' skill scores for prediction of event occurrence and the number of forecasts of Category 3 or greater correct within one category. The bias of the categorical forecasts is also presented, together with the skill and bias achieved during equation development for categorical forecasts from eight months of independent data from 1980 (---).

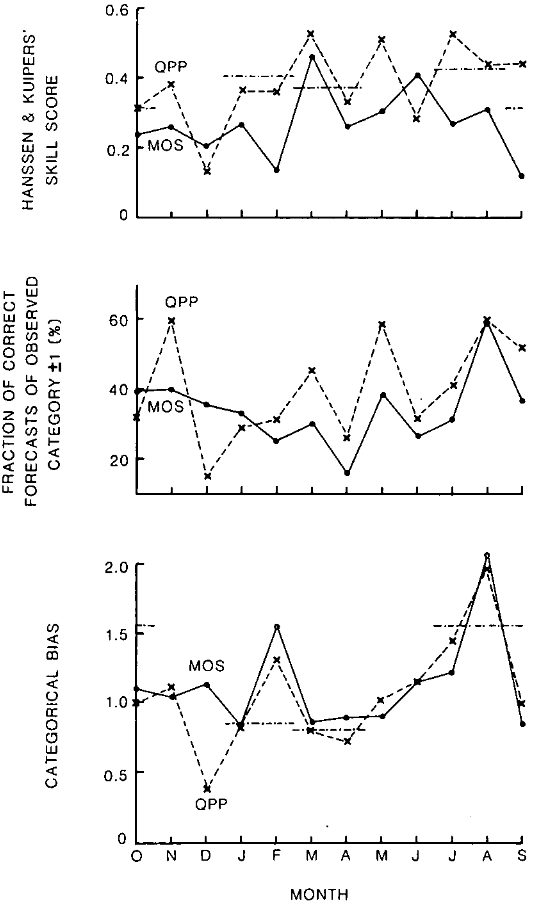
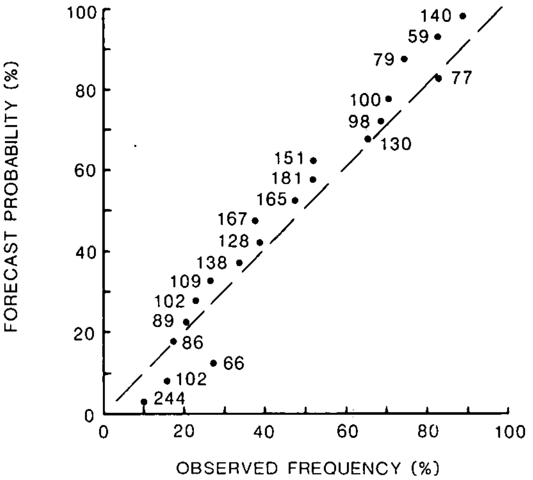


Fig. 6 The reliability of MOS PoP forecasts for October 1983 to September 1984 for seven Australian cities combined. Figures indicate the number of forecasts in each 5 per cent range of probabilities.



a whole deteriorated at all cities except Hobart and Perth, while the loss of skill in predicting large rain totals was greatest at Canberra and Perth. However, prediction of large totals was in fact better at Sydney for 1983-84 than for the 1980 data.

It will be noted that the QPFs, in particular, consistently showed a positive bias in prediction of rain occurrence (Fig. 4), where the bias $B = R_F / R_O$ is the ratio of the number of rain days forecast to the number observed. On the other hand, the overprediction of large rain totals in individual months was mostly within what might be considered a reasonable tolerance for a relatively small sample of an infrequent event. However, major over-forecasting occurred in August, and to a lesser extent in February and July (Fig. 5). Further, in the twelve months such events were overpredicted at Brisbane by more than a factor of two, and underforecast at Canberra by a similar amount. The overprediction of the occurrence of large rain amounts occurred despite the fact that the equations used were transformed during development to remove this bias and the transformations had operated satisfactorily during testing on independent data (Tapp et al. 1986). A tendency to overpredict (underpredict) rain occurrence during the wet (dry) season was noted for both Brisbane and Perth.

Causes of major errors. An assessment was undertaken of the major MOS forecast failures, in terms of the main contributing factors discussed previously. A major error was considered for the present purpose to be an occasion on which the actual and predicted totals differed by a factor of three or more, and one or both totals exceeded 10.5 mm (that is, Category 4 or greater), or when either the forecast PoP exceeded 95 per cent and no rain was recorded or the predicted PoP was less than 5 per cent but an amount between 0.5 mm and 10.5 mm was registered. The results of this analysis are shown in Table 2.

Of the errors attributable to the failure of the statistical technique several factors can be identified as being particularly important. These are:

- rain-generating systems which arrived late in the

Table 2. The number of major failures of MOS precipitation forecasts for seven Australian cities, October 1983 to September 1984, inclusive. The principal cause of errors is identified as being due to failures of either the prognoses or the statistical technique used, or both.

	Statistical	Prognosis	Both	Total
Adelaide	12	3	4	19
Brisbane	27	17	7	51
Canberra	15	10	8	33
Hobart	11	6	4	21
Melbourne	23	6	1	30
Perth	15	11	7	33
Sydney	20	14	8	42
Total	123	67	39	229

MOS forecast period. Melbourne, Canberra and Sydney were the main places to experience frequent errors attributable to this;

- the selected predictors were inappropriate to the particular situation (included here are rare events);
- the distribution of mesoscale precipitation areas within a system, or topographical influences, were not adequately captured by the statistics. This factor was particularly important at Brisbane, Sydney and Hobart.

Effects of changes in prognosis model error characteristics

As was described by Tapp et al. (1986), the first three years of prognoses used for equation development were prepared using the ARPE model in a research mode (see Woodcock 1984), while the last year (1981) consisted of prognoses performed operationally and archived by the Bureau of Meteorology. Three changes in the way the ARPE model was used accompanied this change of archive. Firstly, the manner in which rainfall was predicted changed from being through a determination within the model of the moist processes, using the precipitation parameterisation schemes described by McGregor et al. (1978), to the use of the SLYH model (Younkin et al. 1965). Secondly, the surface drag parameterisation was switched off in the operational model, but was used for the first three years of the prognosis archive. Thirdly, in April 1981 time-dependent lateral boundary conditions (Leslie et al. 1981), specified by a hemispheric forecast model, were incorporated into the operational system. These changes all affected to some extent the homogeneity of the development data set, but the length of record at that time was insufficient to assess their effects. The equations selected may thus not be optimally matched to the operational prognoses.

A further inadvertent change to the operational system took place in February 1984, when the SLYH prognosis was not generated following the prognosis model run which was used to provide the first guess fields for the subsequent objective analysis. This had the effect of gradually increasing the amount of moisture over the oceanic (data void) areas in the objective analyses, with obvious consequences for any subsequent prognoses. This was corrected on 23 July 1984 but the system only slowly returned to an equilibrium climatology. There was dramatic overprediction of large rain totals in August 1984 at Melbourne and Hobart which caused the peak in the bias evident in Fig. 5 , which can be partly attributed to this problem.

Since the time of the initial development of the forecast equations the archiving of operational prognoses has continued, and it is now possible to investigate whether the error characteristics of the operational prognoses differ from those of the research prognoses. Correlations between elements from the prognoses and their verifying analyses have therefore been prepared for the three years of research prognoses, 1978-1980, and for three years of operational prognoses, 1981-1983, in the manner of Tapp (1983). The correlations in each period for some of the most commonly selected predictors for both temperature and precipitation at the seven cities are shown in Table 3. There has generally been a lowering of skill in the prediction of the low-level wind components and all the relative humidity parameters. (The correlations for 850 mb and 500 mb height forecasts are also included in Table 3 to demonstrate that the forecast skill of all fields did not decline from the first three-year period to the second. Unfortunately these fields are not the ones most used in the MOS equations.)

There are two possible causes of this loss of skill. There may have been a change in the actual climatology of the weather patterns from one three-year period to the next. Alternatively, the way in which the numerical prognosis model was used may have changed, in which case a reduction in the skill of the operational MOS forecasts will probably have resulted (Glahn 1982).

Certainly, during the 1981-83 period all States except Western Australia were affected by one of the worst droughts on record (Gibbs 1984). The statistical characteristics of the data from the

Table 3. **Correlations between forecast parameters frequently selected in MOS equations for temperature and rainfall prediction, and the same parameters taken from the verifying analyses. Correlation coefficients are presented for data from two three-year periods, 1978-1980 and 1981-1983, for seven Australian cities. The parameters referred to are geopotential heights (Z), temperatures (T), relative humidities (RH) and zonal (U) and meridional (V) wind components at 1000, 850, 700 or 500 mb.**

	Adelaide		Brisbane		Canberra		Hobart		Melbourne		Perth		Sydney	
Parameter	1978 -1980	1981 -1983	1978 -1980	1981 -1983	1978 -1980	1981 -1983	1978 -1980	1981 -1983	1978 -1980	1981 -1983	1978 -1980	1981 -1983	1978 -1980	1981 -1983
1000 T	0.873	0.856	0.851	0.883	0.869	0.867	0.799	0.751	0.864	0.850	0.863	0.741	0.851	0.867
850 U	0.822	0.786	0.872	0.855	0.855	0.805	0.827	0.801	0.836	0.786	0.817	0.839	0.849	0.798
850 V	0.853	0.801	0.807	0.784	0.829	0.778	0.823	0.804	0.842	0.802	0.759	0.708	0.831	0.786
850 RH	0.709	0.691	0.685	0.754	0.740	0.695	0.486	0.430	0.736	0.683	0.675	0.669	0.686	0.687
700 RH	0.727	0.660	0.804	0.779	0.743	0.723	0.436	0.328	0.691	0.621	0.531	0.490	0.759	0.722
500 RH	0.596	0.491	0.789	0.525	0.715	0.493	0.397	0.305	0.598	0.451	0.365	0.344	0.734	0.492
500 U	0.859	0.841	0.932	0.931	0.870	0.869	0.831	0.834	0.850	0.847	0.816	0.860	0.885	0.883
500 V	0.844	0.828	0.882	0.863	0.842	0.830	0.855	0.848	0.853	0.837	0.736	0.754	0.864	0.851
850 Z	0.892	0.905	0.889	0.895	0.902	0.924	0.881	0.901	0.896	0.914	0.818	0.840	0.903	0.923
500 Z	0.944	0.948	0.943	0.950	0.955	0.961	0.926	0.935	0.943	0.948	0.901	0.905	0.956	0.963

analyses for the two samples showed small changes in most parameters at all cities. However, the corresponding changes in the prognoses were mostly larger and not necessarily in the same direction as those in the analyses. Further, even when the analysed parameters showed little change between the two three-year samples the statistical characteristics of the prognoses generally still altered. The low-level wind fields and the moisture fields are those most likely to be affected by changes to the surface drag and moisture parameterisations, respectively. Any effect of changes in the prevailing weather conditions on the ability of the prognoses to represent the analysed conditions thus appears to have been only minor, and the loss of skill of the prognoses can be mostly attributed to the different configurations of the prediction model during each period. It should also be noted that the data used for equation development included that from the FGGE period when the improved data base, particularly over the southern oceans, had a positive impact on the skill of the prognoses in the Australian region (Leslie et al. 1981).

In order to gain some idea of the quantitative effects on MOS forecasts of the changes in prognosis parameters, prognoses were regenerated for each day of January 1984 using the ARPE model in its fixed boundary, research mode. MOS forecasts were then prepared using these prognoses for comparison with the operational MOS forecasts. (January was selected for this investigation because the MOS forecasts of rainfall for Brisbane were surprisingly poor in this month, compared with those obtained during development, and because it was before the change of operational practice in February 1984.)

Table 4 shows the RMS errors and the number of errors greater than or equal to 5 °C for maximum and minimum temperature forecasts based on both the operational and the re-run prognoses for January 1984. The forecasts from the re-run prognoses were slightly worse than the operational MOS forecasts in Perth and Sydney, but were considerably improved in several other cities, notably Adelaide, Canberra and Melbourne for maximum temperatures, and Adelaide, Canberra, Hobart and Melbourne for the minimum temperatures. The most significant result is the improvement in the minimum temperature forecasts for Canberra, with the RMS error being nearly 1 °C less and the number of large errors being reduced from six to two.

A dramatic improvement was also noted in the prediction of precipitation at Canberra (Table 5), which was highlighted by two days on which totals of Categories 5 and 6 were recorded and which were now predicted within one category where previously Category 2 had been forecast for both. Forecasts for Perth, where only one rain day was recorded in January 1984, reflect the difficulty of improving on climatology as forecast guidance at this time of year. Elsewhere, there was a general increase in skill of the MOS precipitation forecasts at all cities except Hobart when the regenerated prognoses were used.

Table 4. Root mean square (RMS) errors and the number of errors of 5°C or more for MOS forecasts generated for seven Australian cities using forecast predictors obtained from the ARPE model run in its operational mode and rerun in its research mode for January 1984 (see text for details).

City	RMS Errors		Number of errors 5 °C or more	
	Operations	Research	Operations	Research
	Maximum temperature forecasts			
Adelaide	3.44	2.99	6	4
Brisbane	1.53	1.49	0	0
Canberra	3.06	2.63	4	2
Hobart	2.92	2.96	4	5
Melbourne	3.21	2.73	4	3
Perth	4.03	4.27	9	8
Sydney	2.10	2.45	2	3
	Minimum temperature forecasts			
Adelaide	2.28	1.92	1	1
Brisbane	0.84	0.92	0	0
Canberra	3.64	2.70	6	2
Hobart	2.06	1.73	2	0
Melbourne	2.23	1.65	1	0
Perth	1.81	1.82	0	0
Sydney	1.47	1.51	1	0

Sydney PoP forecasts and QPFs of large totals were also less skilful, but the categorical predictions at Sydney based on QPFs improved (Table 5). There was also a general improvement in the bias for prediction of rain days, particularly for the QPFs. Although the bias at Brisbane was reduced there was still a strong tendency to overpredict rain occurrence.

The loss of skill at Hobart, particularly in the categorical forecasts of rain occurrence, and the associated increase in the underprediction of rain days, which resulted from the use of prognoses generated with the research form of the ARPE model, came about primarily because several occasions on which Category 1 totals were recorded and for which totals of up to 0.5 mm had been forecast were now predicted as 'fine'. Both the QPFs and the PoP forecasts use the height and the zonal wind component at 1000 mb as predictors (Mills and Tapp 1984), and the change in the forecasts apparently relates to changes in the predictions of these parameters in the Hobart region. The fixed boundary conditions lead to the research version predicting more intense cyclones over the Southern Ocean in this month, resulting in stronger zonal pressure gradients. It appears that strengthening of the zonal wind had a greater effect on the MOS forecasts than did changes to 1000 mb heights.

It thus appears that although the superficial assessment based on Fig. 1 suggested that the operational MOS temperature forecasts were performing with a level of skill similar to that obtained during development, in reality they showed some sensitivity to the changed error characteristics of the ARPE model in its operational guise. In some cities this sensitivity was undesirably large.

The sensitivity of the precipitation forecasts to those changes was greater. They depended much more heavily on the moisture parameters, which are the ones most strongly affected by the alterations to the prognosis model in the present situation and which showed the greatest changes in statistical characteristics between the two three-year samples examined (e.g. Table 2). As with the temperature forecasts, even the PoP forecasts at Hobart, which used no moisture predictors, showed some loss of skill relative to that achieved using data from 1980 (Fig. 2 (d)). The reduced skill of the other precipitation forecasts generated from the operational prognoses is clearly seen in the results presented in Figs 2 to 5.

Conclusions

This examination of twelve months of MOS forecasts prepared daily for seven Australian cities has shown that the level of skill achieved during equation development for maximum and minimum temperature forecasts has been largely maintained. However, the forecasts of rain amount and the probability of precipitation occurrence suffered some loss of skill, with the latter being particularly affected. This loss of skill of the precipitation forecasts arose because the MOS forecasts depend heavily on predictors obtained from the 24-hour ARPE prognoses, and the form of the prediction model with which most of the data used for equation development were generated differed from that used operationally. It was the characteristics of the moisture parameters, in particular, which were altered by this change, and the precipitation forecasts were heavily dependent on those parameters. A close examination of the

Table 5. The skill of MOS precipitation forecasts generated for seven Australian cities using forecast predictors obtained from the ARPE model used in both its operational mode and its research mode for January 1984, and compared with QPP forecasts for the same month (see text for details).

(a) PoP forecasts, assessed using the improvement in Brier scores, relative to forecasts based on climatology, and categorical forecasts of precipitation occurrence generated from the PoP forecasts, assessed using Hanssen and Kuipers' skill scores. The bias of the categorical precipitation forecasts is also presented.

	Improvement in skill over climatology (%)			Hanssen and Kuipers' skill score			Bias		
	Operations	Research	QPP	Operations	Research	QPP	Operations	Research	QPP
Adelaide	16.3	24.4	18.8	0.2619	0.2024	0.3452	1.0	0.571	0.714
Brisbane	−12.3	6.6	−26.4	0.1333	0.0462	−0.0083	2.0	1.769	1.8
Canberra	26.1	40.7	35.6	0.6048	0.5571	0.7095	0.9	1.0	1.3
Hobart	34.9	33.8	45.1	0.3542	0.2750	0.6167	1.0	0.533	1.133
Melbourne	17.8	20.1	23.9	0.1750	0.1708	0.3000	0.563	0.688	0.688
Perth	−18.6	−55.4	−3.4	0.0	0.0	0.0	0.0	0.0	0.0
Sydney	29.6	16.7	20.9	0.4916	0.4664	0.4202	0.882	1.118	0.941

(b) QPFs, assessed in terms of the fraction of correct forecasts of observed rain category (%), and categorical forecasts of rain occurrence assessed in terms of Hanssen and Kuipers' skill scores. The bias of categorical forecasts is also presented.

	Hanssen and Kuipers' skill score			Fraction of correct forecasts of observed category (%)			Bias		
	Operations	Research	QPP	Operations	Research	QPP	Operations	Research	QPP
Adelaide	0.2619	0.4464	0.4464	64.5	74.2	74.2	1.0	1.0	1.0
Brisbane	0.0564	0.1692	0.1167	14.3	25.0	19.4	1.923	1.462	1.667
Canberra	0.5095	0.5571	0.4048	54.8	61.3	71.0	1.1	1.0	0.7
Hobart	0.4750	0.2042	0.4750	67.7	54.8	67.7	0.733	0.333	0.733
Melbourne	0.3625	0.3000	0.4958	45.2	54.8	64.5	0.750	0.688	0.625
Perth	0.0	0.0	0.0	96.8	96.8	96.8	0.0	0.0	0.0
Sydney	0.2143	0.2521	0.0966	29.0	35.5	29.0	1.647	1.294	1.529

(c) QPFs for rain totals of Category 3 or greater (24-hour totals in excess of 5.5 mm), assessed in terms of Hanssen and Kuipers' skill scores for prediction of event occurrence, and the number of forecasts of Category 3 or greater which were correct within one category. The bias of the categorical forecasts is also presented. No data are presented for Adelaide, Hobart or Perth because no totals of Category 3 or greater were recorded in those cities during January 1984.

	Hanssen and Kuipers' skill score			Fraction of forecasts correct within one category (%)			Bias		
	Operations	Research	QPP	Operations	Research	QPP	Operations	Research	QPP
Brisbane	0.4286	0.6190	0.2619	71.4	42.9	28.6	1.0	1.0	1.0
Canberra	0.0867	0.4200	0.3333	16.7	50.0	33.3	0.500	0.833	0.333
Melbourne	0.0	0.4655	0.0	0.0	50.0	0.0	0.0	1.0	0.0
Sydney	0.1061	0.3224	0.3649	33.3	37.5	29.2	0.833	0.792	0.833

MOS temperature predictions revealed that they were also affected, but to a lesser degree. Considerable care should be taken to assess the effects of any changes in prognosis error characteristics which might result from proposed changes to the prognosis model, when any subsequent MOS equations are developed.

Replacement of forecast predictors with equivalent parameters from the verifying analyses has shown that a considerable overall increase in skill can be achieved in all the elements forecast if the model predictions can more closely describe the observed conditions. Such improvements could be achieved by applying output from an upgraded model to the present prediction equations, even though the gains may be less than optimum due to the factors mentioned above. Further gains of skill would be expected when sufficient data had been collected using the upgraded prediction model to allow redevelopment of the forecast equations.

The twelve months of temperature forecasts examined here indicate that some seasonal factors are not being captured satisfactorily through the use of a single equation applied throughout the year. Forecasts for the southern cities, in particular, should benefit from the use of separate equations in each season once an archive of sufficient length is available for their development. Rain forecasts are also likely to be improved in this manner, particularly at those locations which experience a large seasonal variation in rainfall. Hammons et al. (1976) have demonstrated the advantages of seasonal stratification of data in generating MOS temperature forecast equations.

Acknowledgments

Several people assisted with software development or data processing during this project. In particular, Ms R. Khakham wrote part of the operational MOS software and F. Ohis assisted with routine preparation of MOS forecasts and with compilation of verification data.

References

Gibbs, W. J. 1984. The great Australian drought: 1982-1983. *Disasters, 8,* 89-104.

Glahn, H. R. 1982. Statistical weather forecasting. *ECMWF Seminar/Workshop 1982, Interpretation of Numerical Weather Prediction Products.* ECMWF, Reading, Berkshire, England, 263-310.

Glahn, H. R. and Jorgensen, D. L. 1970. Climatological aspects of the Brier P-score. *Mon. Weath. Rev., 98,* 136-41.

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

Hanssen, A. W. and Kuipers, W. J. A. 1965. On the relationship between the frequency of rain and various meteorological parameters. Koninklijk Nederlands Meteorologisch Institut, *Meded. Verhand, 81,* 2-15.

Leslie, L. M., Mills, G. A. and Gauntlett, D. J. 1981. The impact of FGGE data coverage and improved numerical techniques in numerical weather prediction in the Australian region. *Q. Jl R. met. Soc., 107,* 629-42.

Lorenz, E. N. 1956. Empirical orthogonal functions and statistical weather predictions. *Sci. Rep. No. 1, Contract AF 19 (609) 1566,* Dept of Meteorology, MIT, 49pp.

Lowry, D. A. and Glahn, H. R. 1976. An operational model for forecasting probability of precipitation — PEATMOS PoP. *Mon. Weath. Rev., 104,* 221-32.

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.

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. Report 56,* Bur. Met., Australia, 50 pp.

Peagle, J. N. 1973. Prediction of precipitation probability based on 500 mb flow types. *Jnl appl. Met., 13,* 213-20.

Tapp, R. G. 1983. A statistical comparison of concurrent analysis, prognosis and observational data for seven Australian cities. *Tech. Report 52,* Bur. Met., Australia, 30 pp.

Tapp, R. G., Woodcock, F. and Mills, G. A. 1986. The application of model output statistics to precipitation prediction in Australia. *Mon. Weath. Rev., 114,* 50-61.

Woodcock, F. 1981. Hanssen and Kuipers' discriminant related to the utility of yes/no forecasts. *Mon. Weath. Rev., 109,* 172-3.

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

Woodcock, F. and Nicholls, N. 1982. A methodology for numerical prognosis evaluation using objective local weather retrieval. *Aust. Met. Mag., 30,* 155-62.

Younkin, R. J., Larue, J. A. and Sanders, F. 1965. The objective prediction of clouds and precipitation using vertically integrated moisture and adiabatic vertical motions. *Jnl appl. Met., 4,* 3-17.