Evaluation of Techniques for the Operational, Single Station, Short-Term Forecasting of Rainfall at a Midlatitude Station (Melbourne)

K. FRAEDRICH* AND L. M. LESLIE

Bureau of Meteorology Research Centre, Melbourne, Victoria, 3001, Australia

(Manuscript received 22 September 1986, in final form 21 January 1987)

ABSTRACT

Probability of precipitation (POP) for the 12 hours 0600 to 1800 local time was predicted for Melbourne each day for the three months (winter) period June–August 1986 using six different techniques. These were: a Markov chain model based on 20 years of three-hourly observations; the Australian region limited-area numerical weather prediction (NWP) model; a weighted linear combination of Markov and NWP models; a model output statistics scheme based on the NWP model; an analogue statistics procedure in which a set of the "best" analogues of the NWP forecast were selected; and the manual "official" Bureau of Meteorology forecast for Melbourne, issued by the duty forecaster.

The six techniques were evaluated and compared in terms of Brier scores and also compared with predictions based on climatology. The results of this operational trial indicate that the skill of the combined Markov–NWP model forecasts considerably exceeds the other techniques. The Markov model was next, followed by the other methods which were close together in skill. Some care was taken in the interpretation of these findings as there were differences in lead times associated with the NWP model predictions and the MOS and analogue schemes which were dependent upon it, owing to the operational schedule at Melbourne.

1. Introduction

Probabilistic weather forecasting in Australia has a long history. Some eighty years ago, the world's first probabilistic weather forecasts were issued in Western Australia and the results were published in the *Monthly Weather Review*, Cooke (1906). During the year 1905, daily weather forecasts for two districts in Western Australia were issued with quantitative weights of the forecasters' confidence in their predictions. The predictions of rainfall and of all other weather states were given one of the following five weights: 1) not likely at all; 2) just possible but not likely; 3) more likely right than wrong (wrong four times out of ten); 4) tolerable certainty (wrong once out of ten); and 5) almost absolute certainty. As a matter of interest, it is possible to evaluate the forecasts issued in 1905 in terms of reliability and skill, by assigning probabilities of 0%, 10%, 60%, 90% and 100% to Cooke's weights. Reliability diagrams and (half-) Brier scores can then be deduced and are given in Fig. 1. A tendency to underestimate the probability of future weather states is very apparent, particularly for low weights which are rarely used. The most frequently predicted weights are high and appear to be very reliable. It is therefore not surprising that the Brier scores attain high values.

Since Cooke's time at the turn of the century the automatic recording of long records of station data, coupled with the application of sophisticated statistical techniques and the arrival of powerful computing devices, has encouraged a return to the issuance of probabilistic weather forecasts (Murphy et al., 1985; Fraedrich and Müller, 1986; Williams, 1986). Of course, the major difference between the probabilistic forecasts issued now from those issued in 1905 is that the current forecasts are much more precise. Rather than just being concerned with a general forecast of the weather they attempt to forecast quantities such as rainfall amount or maximum and minimum temperature with high precision.

There has been a proliferation of techniques for short-term prediction of weather states. Broadly speaking, these methods fall into three categories. At the two extremes are the purely statistical methods and the numerical weather prediction (NWP) models. The third category is the statistical–dynamical hybrid models such as the model output statistics (MOS) technique and the analogue statistics approach.

Statistical methods predate the other methods because they require only modest computing resources. The fitting of Markov chain models to climatological records of station surface data has found considerable favor as a method of obtaining forecasts of weather elements such as precipitation (e.g., see Miller, 1981; Fraedrich and Müller, 1983). Markov models have two distinct advantages: forecasts are available immediately

---

* On leave from Institut für Meteorologie, Freie Universität Berlin, Berlin, Federal Republic of Germany.

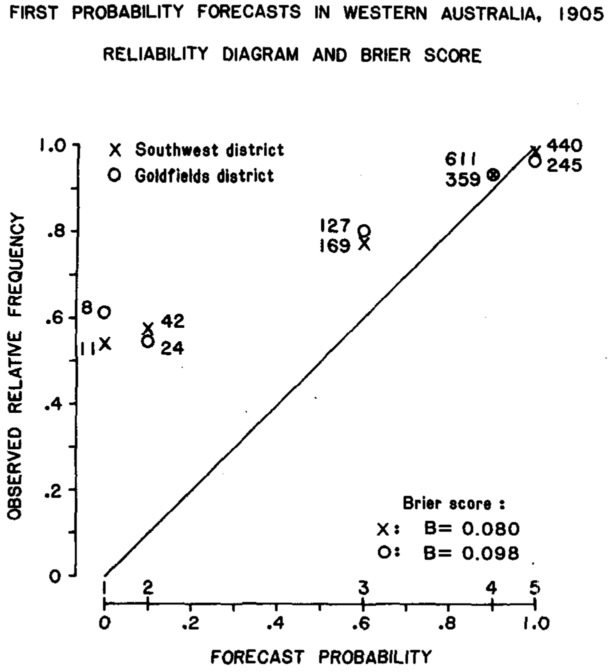

FIG. 1. Reliability diagram of the first probability weather forecasts for two districts near Perth (Western Australia) during 1905. Note that the first two of the five confidence weights have been quantified a posteriori in probabilistic terms. The number of predictions (with weight 1 to 5) also is indicated. The half-Brier score B is included at the lower right.

after the observations have been made because they use only the local surface weather readings as predictors, and they require minimal computation after the climatological records have been processed. In the Australian context, Miller and Leslie (1984) carried out a pilot study in which Markov chains were fitted to the records of three-hourly surface data at six major Australian cities. Probabilistic forecasts of precipitation obtained in this study exhibited sufficient skill to encourage further work and in a subsequent study (Miller and Leslie, 1985) the transition probabilities were extended to be functions not only of the previous state, but also of other surface data including pressure, change in pressure, dewpoint depression, wind speed and wind direction.

In contrast with the statistical methods which are based firmly in the climatological records, the short-term NWP models are deterministic initial-boundary value predictions that make little reference to the climatology of the forecast domain. They also provide areal average rather than station or point forecasts, although some progress has been made in converting NWP model forecasts of convective precipitation into point forecasts by calculating the lifetime, size and speed of precipitating elements and deriving the probability of precipitation at a given point from these values (Hammerstrand, 1980). Despite their limitations in not being able to forecast subgridscale rain processes such as those associated with convective elements or complex orography, NWP models have become a major tool at weather centers around the world for forecasting precipitation amount and duration at selected stations.

Since the early 1970s much effort has been directed towards combining the raw NWP model with data records from observations at the stations for which precipitation forecasts are required. This conjunction of NWP model output and climatology defines the third class of methods for the prediction of precipitation at specific stations. The regressions provide the link with climatology that is absent from the pure NWP model forecasts. As was mentioned before, there are several kinds of hybrid NWP–statistical regression methods. Perhaps the most well-known of these is the MOS procedure (Glahn and Lowry, 1972) in which regression equations are developed between the required weather element and various quantities forecast by the NWP model. Analogue statistics methods (Stern, 1980) are another, quite different, procedure for relating NWP model output to observational data. In this case, a set of analogues to the NWP model prediction is selected from the analysis records and regressions are applied to this set of analogues to obtain forecasts of precipitation probability and amount for the forecast period.

In common with most weather centers around the world, there is a need for improved accuracy in forecasting the likelihood of precipitation for a city and its surrounding metropolitan area. Melbourne is a midlatitude (38°S) station which has cool, rainy winters with precipitation recorded on an average of about 50 percent of days during the months of June, July and August. Currently there are five precipitation forecasting techniques available at Melbourne: the official Bureau of Meteorology forecast, which is a manual forecast issued by the duty forecaster; the Markov chain statistical technique (Miller and Leslie, 1985); the limited-area NWP model (Leslie et al., 1985); a MOS procedure which regresses the limited area NWP output against observations (Mills and Tapp, 1984; Tapp et al., 1986); and an analogue statistics method (Stern, 1980), which utilizes the best 25 analogues of the NWP model forecast.

As an attempt to improve the quality of precipitation forecasts for Melbourne, and subsequently for other Australian stations, an operational trial was carried out in the winter (June, July and August) of 1986. With the cooperation of the Regional Office in Melbourne, the five methods described above for predicting rainfall in Melbourne were evaluated and compared. The results of the trial are given in section 4.

2. The climatological setting

a. Melbourne

The Melbourne winter season (June, July and August) was chosen for the operational trial because it has

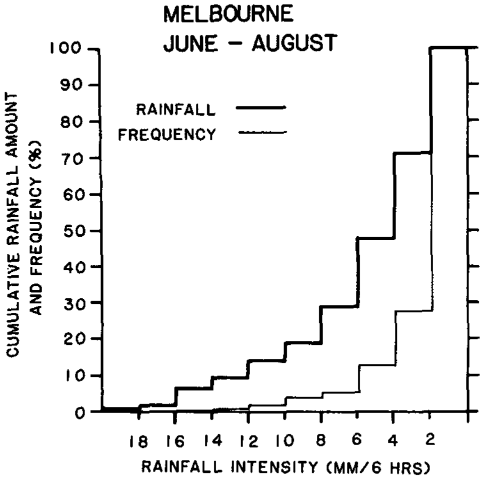

FIG. 2. Cumulative distribution of rainfall amount and frequency versus 6-hourly rainfall intensities for Melbourne (1962–81).

almost an equal number of wet and dry days. Most of the rain days result from the passage of cold fronts, with an average seasonal total over the past 120 years of 157 mm occurring on 48 of the 92 days.

An analysis of the three-hourly observations for the 0600 to 1800 local time period reveals that Melbourne records rain on an average of 35 of the 92 winter half-days. The intensity of the precipitation usually is light, with precipitation rates below 4 mm/6 h accounting for about 50 percent of the total rainfall (see Fig. 2). Because the rainfall is generally light it was decided not to subdivide the rainfall predictions into categories for this study. In a tropical station such as Darwin, precipitation rates below 4 mm/6 h account for only 10 percent of the rainfall and it is desirable for such stations to predict probability of precipitation for at least the categories of light and heavy.

b. Single station versus areal precipitation

An important question for a city as large as Melbourne (approximately 50 × 50 km) is how typical is a single station for the metropolitan area. In fact, there is a large west-to-east gradient in annual rainfall that is basically related to orography (see Fig. 3). To illustrate the representativeness of the Melbourne station, half-daily (0600–1800 local time) precipitation cumulants are calculated from 9 years of pluviograph data for the years 1965–73 at eight suburban stations. These eight stations are combined into two sets of four each from the western and eastern suburbs. Conditional frequency distributions of rainfall occurrence are shown in Fig. 4 to indicate the relative number of stations

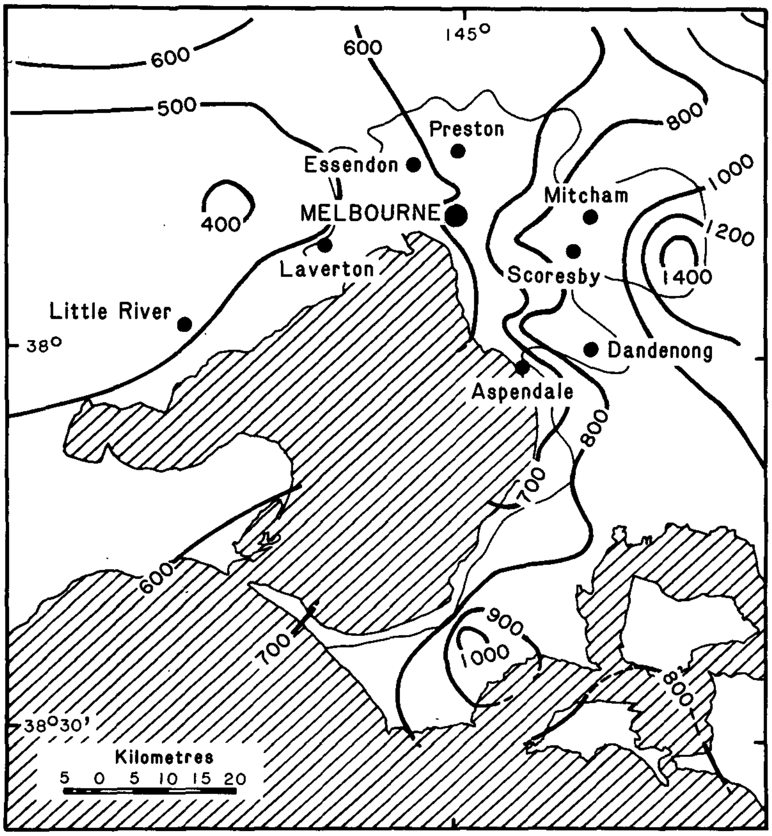

FIG. 3. Annual rainfall distribution (mm yr^{-1}) over the Melbourne metropolitan area.

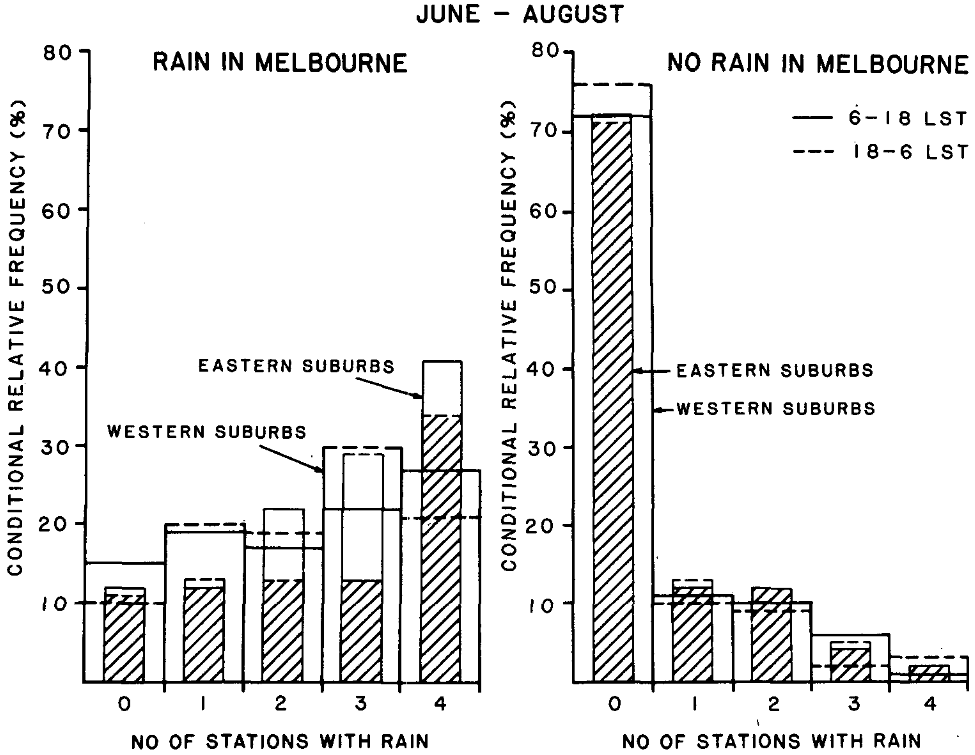

FIG. 4. Relative frequency distribution of the number of suburban stations with rain at the same 12-hourly time intervals for which the Melbourne city station reports rain (left) or no rain (right). Eastern and southeastern (cross hatched) suburbs are Aspendale, Dandenong, Mitcham and Scoresby; the western and northwestern suburbs are Essendon, Laverton, Little River and Preston (see also Fig. 3). The statistics cover 9 years of homogeneous pluviograph records.

reporting precipitation if it has rained/not rained in Melbourne during the 0600–1800 h period.

The observed probability distributions of Fig. 4 for the winter season show that if rain is observed in Melbourne during a 12-h time period there is approximately a 90 percent chance that at least one of the four stations in both the western and eastern suburbs will report rain. In the eastern suburbs the chance that all four stations report rain is about 35 percent. For the western suburbs this value drops to about 25 percent. Conversely, if it has not rained in Melbourne during a 12-h period, there is a 75 percent chance that it has not rained in any of the four stations in both the western and eastern suburbs.

From the climatological distributions just discussed, it therefore can be claimed that predictions of rain/no rain for the station of Melbourne during a 12-h period in winter are quite representative of the entire metropolitan area, but could be refined further by a careful use of Fig. 3.

3. The rainfall prediction models

a. The Melbourne rainfall Markov chain

The basis for the stochastic single station forecast model is the Markov chain model described by Miller and Leslie (1985). It consists of four mutually exclusive states ($j = 1, \cdots 4$) including three cloud states (0–2 octas, 3–5 octas, 6–8 octas) and one rain state. The rain state is defined by the station having received measurable precipitation during the previous three hours, or with present/past weather indicating rain at or near the station (that is, international weather code WW: 13–17, 20–27, 29, 50–99). The probability of precipitation (POP) is then given by

$$\mathrm{POP}(j,t,h) = a(j,t,h) + \sum_{k=1}^{3} b_k(j,t,h)X_k$$

where t is the current time of day, h is the number of hours ahead and the covariates X_k ($k = 1, 2, 3$) are the surface pressure, the dewpoint depression, and the east–west component of the wind. Other covariates were found to be of less significance. Note that different intercepts, a, are used for each month and common slopes, b, are fitted for all months. The model has been fitted by ordinary least-squares to the data.

Equation (1) defines the POP as the chance of a rainfall event to occur at least once during the period h. With increasing h, the POP rises to its limiting value of 1, or 100 percent. For the present operational trial, a 12-h period was chosen as a convenient time interval, and this is sufficiently long for persistence to be inadequate, but not so long that the Markov prediction approaches climatology.

The Markov chain model POP forecasts can be displayed conveniently in graphical form, as shown in Fig. 5 in a probability of precipitation versus surface pressure diagram for each of the four basic weather states, and with adjustments for dewpoint depression. Due to the definition of the Markov chain rainfall state (which also includes nearby precipitation processes by using the present/past weather observations), a dewpoint depression may exist concurrently with rainfall (Fig. 5). Variations in the POP values as functions of wind are of less significance and are not shown in Fig. 5. Note that there is no lead time for Markov chain predictions.

b. Numerical weather prediction model

The NWP model used to provide the forecasts of precipitation is the Australian Bureau of Meteorology's new operational limited-area model. It is a 12-level, 150-km horizontal resolution, primitive equations model and it provides twice-daily 36-h forecasts from database times of 0900 and 2100 local time. Full details of the model are given by Leslie et al. (1985).

Rainfall predictions for Melbourne during the period 0600–1800 local time are calculated from the 9-h–21-h subset of the 36-h forecast based on the initial data at 2100 hours the evening before. Thus, the NWP model has a forecast lead time of 9 hours. The 21-h lead time predictions (issued at 0900 local time) are also analyzed to obtain a description of the skill varying with lead time. The Melbourne rainfall is obtained by interpolating rainfall predictions from the four nearest gridpoints.

The NWP model provides a wide range of forecast guidance. For the purposes of the POP trial, the relevant output displays the NWP forecasts of rainfall as well as mean sea level pressure, cloud cover and screen temperature at three-hourly intervals.

c. Combined Markov-NWP scheme

The two independent Markov and NWP schemes can be combined linearly to achieve a level of skill that

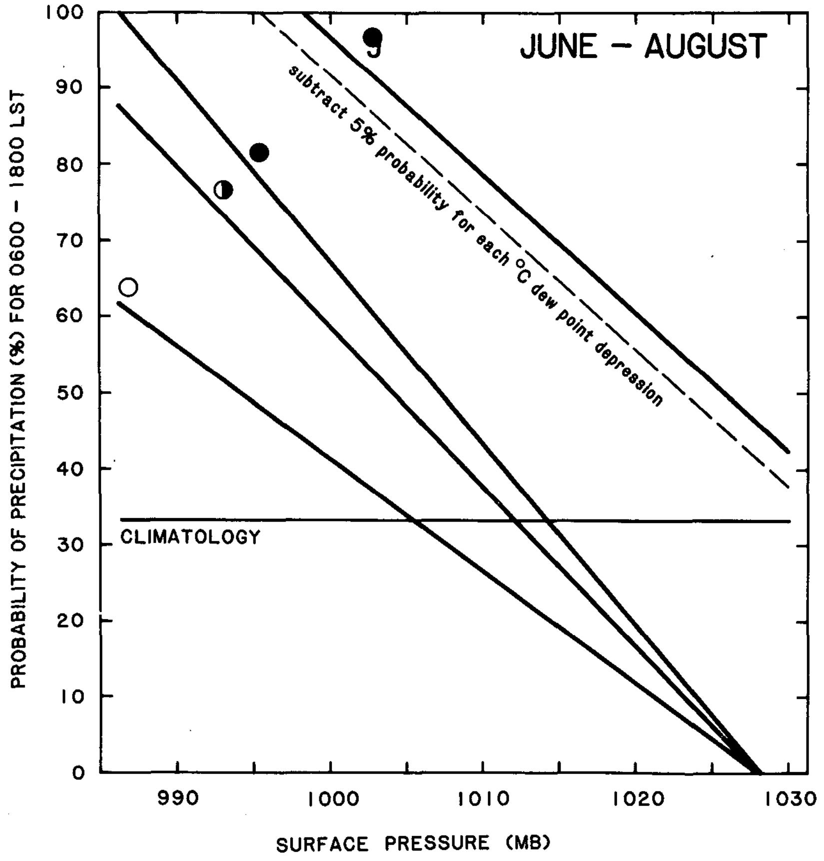

FIG. 5. Melbourne single-station probability of precipitation (0600–1800 LST) changing with surface pressure, cloud cover and rainfall states observed during the preceding hour (O, ◑, ●, ❟ are the 0–2, 3–5, 6–8 octas, rain states). Note that for the rainfall state POPs decrease with decreasing dewpoint depression, which is due to the definition of the rainfall state (section 3a). Climatology is also included.

exceeds the skill of each scheme individually. Fraedrich and Leslie (1986) have shown that the predictive scheme

$$\mathrm{POP}^* = a\,\mathrm{POP}_{\mathrm{MARKOV}} + (1-a)\mathrm{NWP}, \qquad (2)$$

where NWP = 0 or 1 for no rain and rain NWP forecasts, and $\mathrm{POP}_{\mathrm{MARKOV}}$ is given by Eq. (1). From the hindcast experiments the optimum value of a was calculated to be 0.7.

d. Model output statistics

Model output statistics (MOS) forecast equations were developed by Mills and Tapp (1984) for the operational prediction of local weather at the seven major Australian cities, including Melbourne. Meteorological quantities such as maximum/minimum temperature, rainfall and POP were the predictands in a regression scheme that uses parameters forecast by the NWP model described in section 3b as predictors.

The MOS POP forecasts have been issued operationally to the Melbourne regional office forecasters since January 1984, following an operational trial (Tapp et al., 1986) in which the MOS predictions of rainfall amount and probability were shown to be of sufficient skill to warrant routine usage.

The operational MOS forecasts are issued once daily based on the 24-h forecast from the NWP model valid at 2300 UTC. In order to include the MOS POPs in the present comparative study, it was necessary for the MOS equations to be generated for the 12-h period from 0600–1800 local time using the 24-h NWP forecast run from base time 2300 UTC the morning before. Thus, the MOS forecasts for the 12-h period involved a lead time of 21 hours, which is a considerable restriction on the skill of the method. However, no alternative was available as the MOS equations have not yet been calculated for 1100 UTC.

e. Analogue technique

A fourth system that is used to predict probability of precipitation for Melbourne is the analogue technique devised by Stern (1985). The original analogue method developed by Stern was a single analogue technique, but it performed poorly during a trial of its skill in October 1985 and has been replaced by an analogue statistics procedure in which a set of analogues is selected and a regression analysis is applied to this set to obtain a forecast of the required weather elements. The analogue statistics model was subjected to a trial in the spring of 1978 and its performance was only marginally worse than the duty forecaster's predictions.

In the present trial the statistics analogue model has been adapted to provide a POP forecast based on the 25 best analogues of the 24-h forecast valid at 0900 local time. The analogue statistics are then used to calculate a rainfall probability for the 12-h period 0600 to 1800 local time. There are no analogue statistics calculated at 2100 local time as yet, so the analogue technique has a lead time of 24 hours which, as in the case of the MOS forecasts, is a limiting factor on the skill of the method.

f. Duty forecaster

The Australian Bureau of Meteorology has a regional office in Melbourne. This regional office is responsible for issuing worded weather forecasts to the public at various times of the day. No POP forecasts are routinely issued by the regional office, but for the purposes of the operational trial described in this paper, the duty forecasters issued POP forecasts at 0545 local time for the period 0600–1800 local time.

The duty forecasters have available as forecast guidance only the information provided by the NWP system. The remaining four predictions (Markov, the Markov–NWP combination, MOS and analogue) were not prepared in time for the forecasters to use. The duty forecasters were required to enter a probability of precipitation prediction into their work sheets just before 0600 local time each day of the three-month trial.

4. Results of the operational trial

The trial was carried out over a total of 100 half-days (0600–1800 local time) from 22 May to 29 August 1986. The last week of May originally was intended as a "buffer" period during which any problems could be dealt with. However, no difficulties arose, and the data for the period were included in the trial.

From the climatological records for Melbourne the average number of wet half-days for the period covered by the trial is 38. The basic question is how typical was the 1986 season. As it turned out there were 32 half-days on which measurable precipitation (≥ 0.2 mm) was recorded. These were 2 wet half-days in the last 10 days of May, 7 wet half-days in June, 15 in July and 8 in the first 29 days of August. In summary, July was a wet month, while the other months were drier than normal. An unforeseen complication arose with the number of days (5) on which a trace of rain was recorded. This represented a significant percentage of the total wet half-days. It was decided to present the results with trace defined as both dry and wet.

The six predictive schemes were evaluated and compared using a number of assessment techniques and the results are presented below.

a. Brier scores

An objective measure for comparing the skill of the precipitation forecasts is the half-Brier score (Brier, 1950) given by

$$B = \frac{1}{m}\sum_{i=1}^{m}(\delta_i - \mathrm{POP}_i)^2 \qquad (3)$$

where m (=100) is the number of forecasts, POP_i is the probability of precipitation predicted for the ith day, and $\delta_i = 0$ or 1 if the observation was dry or wet, and $\delta_i = 0.5$ if trace was observed. Note that all schemes were verified against the half-daily single-station precipitation measured in Melbourne, which does not include rain near the station as indicated by present/past weather states. The Brier score has a minimum value of 0 if all forecasts were perfect. For categorical (NWP) predictions $\mathrm{POP}_i = 0$ or 1 for dry or wet forecasts. It is noted that for binary forecasts the half-Brier score is identical with the percentage of incorrect predictions.

The half-Brier scores for the six schemes are shown in Table 1. By far the best predictions were made by the combined Markov–NWP scheme, with a half-Brier score of 0.11. Next came the Markov scheme with a score of 0.14, followed by the manual forecasters with 0.17. The MOS method and the NWP model were close together in skill, with 0.182 and 0.181, respectively. Finally, the analogue statistics technique recorded a score of 0.20. For climatology predictions $\mathrm{POP}_i = \mathrm{POP}_c = 0.31$, 0.32 and 0.35 are used in June, July and August.

To estimate the effect of different lead times, the half-Brier scores of (at least) some of the predictions schemes were evaluated for 9- and 21-h lead time forecasts: the NWP model, the climate–NWP combination ($\mathrm{POP}^* = a\ \mathrm{POP}_c + (1 - a)$ NWP with the optimal weight $a = 0.5$, see Fraedrich and Leslie, 1986), and the Markov–NWP combination (2) keeping the same weight $a = 0.7$ for both 9- and 21-h lead time. The results (Table 2) show the expected decrease in skill with increasing lead time for all schemes. Furthermore, in a broad sense the MOS scheme can be seen as combining NWP and climatology. This hypothesis is supported by the 21-h lead time skill of the climate–NWP combination (0.19), which is only about 5% worse than the score for MOS (0.18). Moreover, assuming this hypothesis to hold also for the 9-h predictions, both MOS and the climate–NWP combination forecasts would attain skills which are comparable with the Markov model; the performance of the Markov–NWP combination, however, would remain best.

b. Reliability

If the POPs are perfectly reliable then precipitation should occur on the percentage of days, say $X\%$, for which the POP forecast was $X\%$. Figure 6 shows the realized relative frequency of the occurrence of precipitation (that is, the relative number of 12-hourly time intervals from 0600 to 1800 local time on which rain fell in Melbourne) grouped in 20% POP forecast ranges. Because a significant fraction of the total number of rain half-days is categorized as trace rainfall, the results are shown for the cases where a trace is counted as either "rain" or "no rain."

TABLE 1. Half-Brier scores for the six predictive methods and climate.

	Predictive Method						
	Markov–NWP	Markov	Forecaster	MOS	NWP	Analogue	Climate
Half-Brier score	0.11	0.14	0.17	0.18	0.18	0.20	0.22

TABLE 2. Half-Brier scores for 9 and 21 h lead time (see text).

Predictive scheme	Lead time	
	9 h	21 h
NWP	0.18	0.26
Climate–NWP	0.15	0.19
Markov–NWP	0.11	0.14
MOS		0.18

The diagonal in each diagram of Fig. 6 represents perfect reliability which, of course is unachievable; even in a coin tossing experiment one has no guarantee to obtain exactly 50% heads or tails. The following features of the predictive schemes should then be noted:

1) All the objective models and the manual forecasts show a tendency to overpredict, that is, to forecast higher probabilities of rain than observed, except for the analogue scheme, which shows a tendency to underforecast. It should be noted that the Markov chain has been calibrated for a rainfall state which includes rain and trace observed at the station plus those situations for which single station present/past weather observations indicate rain nearby (see section 3a). This explains the tendency for the Markov chain (and the NWP–Markov combination) to overpredict rainfall when being verified against raingage measurements. This may also be the reason for the greater reliability of Markov and Markov–NWP predictions in the extremes of the probability range.

2) The manual forecasts are very reliable for dry days, but the skill tapers off in the range between 0.4 and 0.8. If the trace rainfall counts as a wet day, the forecasters regain reliability in the 0.8 to 1.0 range. Thus, the manual forecasts are reliable for highly predictable wet or dry situations but are inadequate in more uncertain situations.

3) The number of forecasts by the Markov–NWP combined and the Markov predictive schemes is fairly equally distributed over the probability ranges. The manual forecasts use POP = 0 or 1 very frequently, on about 60% of occasions. If trace rainfall counts as a wet half-day then they are reliable, indicating that they tend to forecast rain with safety in mind. (This also is reflected in the forecasters' bias, which will be discussed in section 4c.) In contrast with the manual forecasts, MOS uses the middle range 0.4 to 0.6 most frequently (about 40% of all forecasts) but only forecasts POP > 0.8 on three occasions. This is not surprising as regression predictions tend increasingly towards the mean and inflation adjustments, or other transfor-

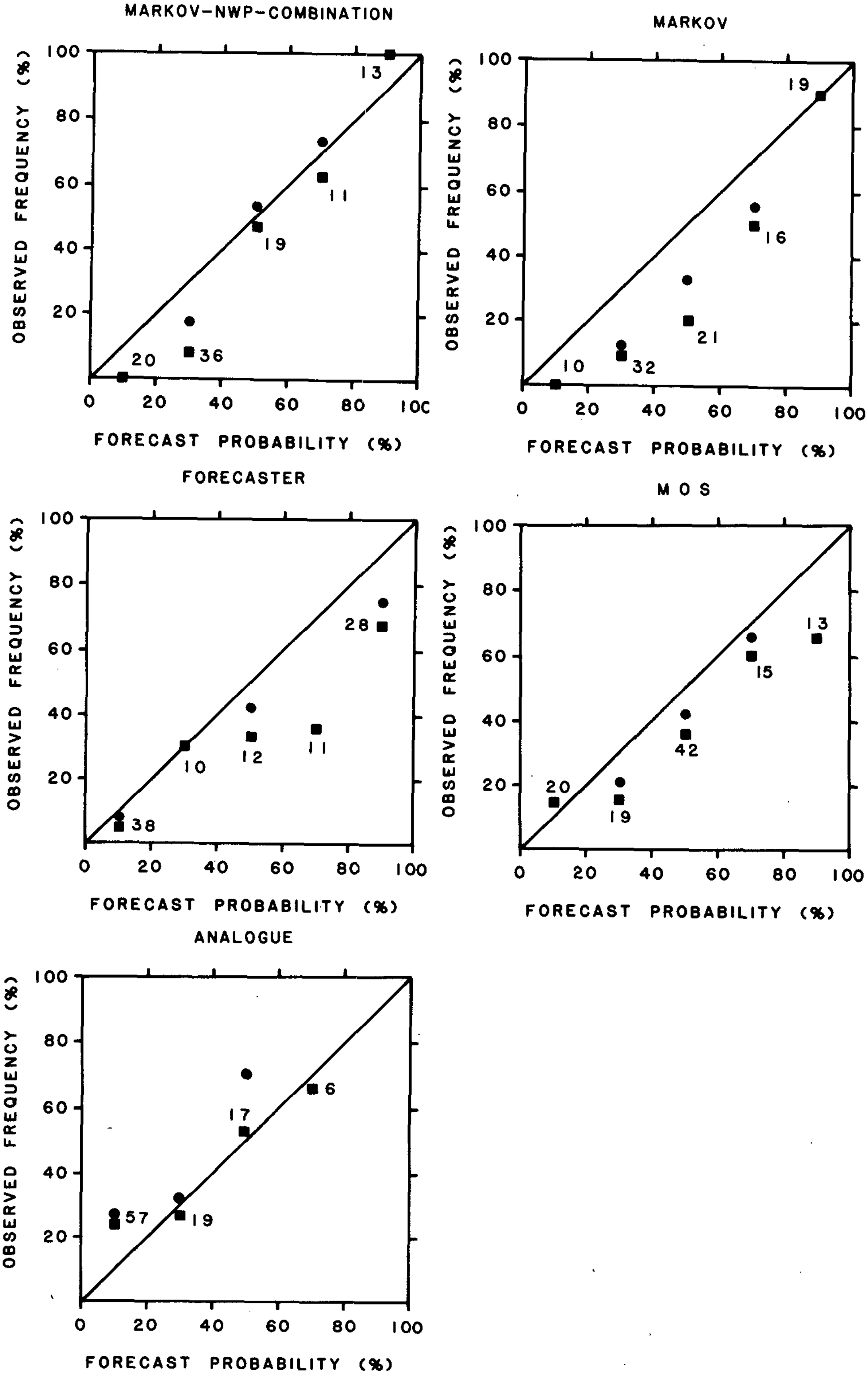

FIG. 6. Reliability diagrams for five probability of precipitation forecast methods: Markov–NWP Combination, Markov, Forecaster, MOS and Analogue-Predictions. The number of cases in each 20% range is indicated. Trace rainfall can be defined as "no rain" (■) or "rain" (●).

mations have not been used for the MOS forecasts (Tapp et al., 1986). The analogue scheme predicts POP $\leqslant 0.2$ for about 60% of the cases and no POP > 0.8.

4) The Markov chain model overpredicts in the middle POP range but is almost perfect at high and low probabilities. This is in contrast with the MOS method.

5) The Markov–NWP combination also is excellent in the high and low POP ranges. For all 20 Markov–NWP POP forecasts less than 0.2 no rain was recorded and for all 13 POP forecasts greater than 0.8 rain was observed. Furthermore, apart from the 0.2 to 0.4 range the reliability in all other categories is almost perfect (with and without considering trace as rainfall).

c. Evaluating POP forecasts in categorical (binary) terms

Although categorical predictions are not optimal, they are still common practice in weather forecasting. Therefore, we generate categorical forecasts of rainfall occurrence from probabilistic (POP) predictions by interpreting $0.0 \leqslant \text{POP} \leqslant 0.5$ as dry and $0.5 < \text{POP} \leqslant 1.0$ as wet. This 50% threshold value was chosen from Markov chain hindcast evaluation at which value the categorical, or binary, interpretation produces the minimum number of incorrect rainfall state predictions if verified against the present/past weather states (defined in section 3a).

Given the 50% threshold, 2×2 contingency tables may be constructed of the form

	Predicted	
Observed	Wet	Dry
Wet	A	B
Dry	C	D

for all the probability schemes and the NWP scheme. The contingency table is given in Table 3 for the six predictive schemes and for climatology. Values calculated with trace defined as wet are given in brackets.

The contingency table may be used to calculate three further measures of forecast skill:

1) The overall skill of the POP forecasts expressed in binary terms may be assessed using the discriminant of Hanssen and Kuipers, which is independent of the climatological frequency of the occurrence of rain (Woodcock, 1976). This score is defined by

$$V = \frac{AD - BC}{(A+B)(C+D)}, \tag{4}$$

and the values of V for all schemes and climatology (i.e., no rain predicted for every half-day: $V = 0$) are given in the first row of Table 4. Not unexpectedly, the Markov–NWP scheme is much superior to the other methods, despite the relatively poor NWP results.

2) The relative number of correct wet and dry predictions,

$$\text{Percent} - \text{correct} = \frac{A+D}{(A+B+C+D)}, \tag{5}$$

is given in the second row of Table 4. Once again the Markov–NWP combination is by far the best predictor with only 12 out of 100 incorrect predictions compared with 20 to 25% of incorrect predictions for the other schemes, with MOS being the worst performing of all schemes with about 30% incorrect predictions. This is due to the large amount of $0.4 < \text{POP} \leqslant 0.6$ MOS-predictions in the neighborhood of the 50% threshold value and therefore should not be regarded as a significant failure of MOS. Note, however, that both MOS and analogue percent-correct forecasts perform as well as climatology (predicting no rain for every half-day).

3) The *bias* is a measure of the predictive model climate compared with the observed climatology, for rain,

$$\text{Bias} = \frac{A+C}{A+B}.$$

The three schemes, Markov–NWP combination, Markov, and MOS represent the Melbourne 1986 winter precipitation climate reasonably well (with bias ~ 1), in particular if trace is considered as wet. All predictive schemes (except the Markov–NWP combination) show a tendency to overforecast rain by their bias larger than one. In particular, forecasters tend to be on the "safer" side with their POP predictions (i.e., they overforecast). Of course, the climatology categorical prediction of no rain every half-day reveals zero bias. Due to the tendency of the Markov chain to overpredict rainfall (see section 4b) the bias is expected to be larger than 1; otherwise, a bias near 1 would have been observed. Note that bias of all categorical predictions may be

TABLE 3. Contingency table for the six predictive methods (plus climatology) based on a threshold of 50% for dry/wet distinction. Figures in brackets are for trace taken as "wet."

	Predicted "wet"						Predicted "dry"						
Observed	Markov + NWP	Markov	Forecaster	MOS	NWP	Analogue	Markov + NWP	Markov	Forecaster	MOS	NWP	Analogue	Climate (observed)
"wet"	28 (30)	26 (27)	26 (29)	21 (25)	20 (22)	7 (7)	4 (7)	6 (10)	6 (8)	11 (12)	12 (15)	25 (30)	32 (37)
"dry"	8 (6)	14 (13)	20 (17)	20 (16)	5 (10)	5 (5)	60 (57)	54 (50)	48 (46)	48 (47)	63 (53)	62 (57)	68 (63)

TABLE 4. The Hanssen and Kuipers score, percentage of correct forecasts, and bias for the six predictive techniques. Values for trace taken as wet are in brackets.

	Method						
Scores	Markov–NWP	Markov	Forecaster	MOS	NWP	Analogue	Climate
Hanssen and Kuipers	0.76 (0.72)	0.61 (0.57)	0.52 (0.51)	0.55 (0.42)	0.36 (0.42)	0.14 (0.11)	0 (0)
Percent-correct	88 (87)	80 (77)	74 (75)	69 (73)	83 (75)	69 (64)	68 (63)
Bias	1.13 (0.97)	1.25 (1.08)	1.46 (1.24)	1.78 (0.86)	1.28 (1.11)	0.37 (0.32)	0 (0)

reduced by choosing POP-threshold values which minimize the incorrect forecasts for each scheme individually.

5. Conclusions

The performance of six methods for the short-term predictions of precipitation for the Australian midlatitude city of Melbourne has been assessed in an operational trial carried out over a period of 100 days, during the winter months, 22 May–29 August. The main aims of the operational trial were to compare the various methods available to the Australian Bureau of Meteorology in terms of a range of measures of skill. The measures include the half-Brier score, the reliability and, interpreting the forecasts as categorical, the Hanssen–Kuipers score, the percentage of correct forecasts and the bias.

The linear combination of the independent Markov chain and NWP predictions proved to be by far the best of all the schemes evaluated in the trial, with a Brier score of 0.11 compared with 0.14 for the Markov scheme, 0.17 for the forecasters, 0.18 for MOS and NWP and 0.21 for the analogue statistics method.

When the probabilistic forecasts were interpreted as categorical forecasts of dry or wet days using a 50% threshold distinction between dry and wet days, the Markov–NWP combination retained its high level of skill correctly forecasting the occurrence of wet and dry half-days on 88% of occasions. The skill of the Markov–NWP combination was sufficiently high to suggest that it could be used on a routine basis for 12-h probabilistic and/or categorical forecasting of precipitation for Melbourne.

Further improvements in the Markov–NWP combination can be expected to come from a number of directions. The skill of the NWP model has been improving steadily over the years and can reasonably be anticipated to improve further in the future. The Markov chain method could also be refined further to remove several obvious shortcomings such as the failure to distinguish cloud types and to differentiate between fog and overcast conditions.

Acknowledgments. We would like to thank the Regional Director of the Victorian Regional Office, Tony Powell, for his permission and enthusiastic support for the trial. We are grateful also to Bob Wright and Harvey Stern for continuous interest in the trial.

Other persons we would like to acknowledge include Graham Mills, Roger Tapp and Gerald McNamara who prepared the MOS forecasts; David Pike who tabulated and coordinated the results of the trial; Peter Yew for drafting the figures; Ms. C. Aiello and Ms. F. Nikou for typing; and the referees for their comments.

REFERENCES

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Cooke, W. E., 1906: Forecasts and verifications in Western Australia. *Mon. Wea. Rev.*, **34**, 23–24.

Fraedrich, K., and K. Müller, 1983: On single-station forecasting: Sunshine and rainfall Markov chains. *Beitr. Phys. Atmos.*, **56**, 108–134.

——, and K. Müller, 1986: On single-station forecasting: Probability of precipitation in Berlin. *Beitr. Phys. Atmos.*, **59**, 427–434.

——, and L. M. Leslie, 1987: Combining predictive schemes in short term forecasting. *Mon. Wea. Rev.*, **115**, 1640–1644.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.

Hammarstrand, U., 1980: A model to predict the probability of precipitation. *Mon. Wea. Rev.*, **108**, 793–803.

Leslie, L. M., G. A. Mills, L. W. Logan, D. J. Gauntlett, G. A. Kelly, M. J. Manton, J. L. McGregor and J. M. Sardie, 1985: A high-resolution primitive equations NWP model for operations and research. *Aust. Meteor. Mag.*, **33**, 11–35.

Miller, A. J., and L. M. Leslie, 1984: Short-term, single-station forecasting of precipitation. *Mon. Wea. Rev.*, **112**, 1198–1205.

——, and ——, 1985: Short-term, single-station probability of precipitation forecasting using linear and logistic models. *Beitr. Phys. Atmos.*, **58**, 517–527.

Miller, R. G., 1981: GEM: A statistical weather forecasting procedure. NOAA Tech. Rep. NWS 28, National Weather Service.

Mills, G. A., and R. G. Tapp, 1984: The Australian operational MOS forecast system. Bureau of Meteorology, Tech. Rep. 56, 50 pp. [GPO Box 1289K, Melbourne, Vic. 3001 Australia.]

Murphy, A. H., W. R. Hsu, R. L. Winkler and D. S. Wilks, 1985: The use of probabilities in subjective quantitative precipitation forecasts: Some experimental results. *Mon. Wea. Rev.*, **113**, 2075–2089.

Stern, H., 1980: A system for automated forecasting guidance. *Aust. Meteor. Mag.*, **28**, 141–154.

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

Williams, G. M., 1986: Probability forecast results of the PROFS 1983 summer exercise. *Mon. Wea. Rev.*, **114**, 295–306.

Woodcock, F., 1976: The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Wea. Rev.*, **104**, 1209–1214.