Real-Time Short-Term Forecasting of Precipitation at an Australian Tropical Station

K. FRAEDRICH* AND L. M. LESLIE

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

(Manuscript received 24 June 1987, in final form 17 December 1987)

ABSTRACT

The results of a major real-time trial of techniques for the short-term (12 h ahead) prediction of precipitation for the Australian tropical city of Darwin are described. The trial compared current operational manual forecasting procedures with a range of alternative techniques including statistical methods, numerical weather prediction (NWP), and model output statistics (MOS) which were developed specifically for the trial.

The only technique of those tested which exhibited skill, i.e., consistent superiority over climatology and/or persistence, was a Markov chain model used either individually or in linear combination with other methods. Of particular significance was the relatively poor showing in the tropics of the numerical weather prediction model products, which were worse than both persistence and climatology. The results of this real-time trial should be treated with care because of the small sample size involved.

1. Introduction

Short-term forecasting (that is, up to 12 h ahead) in the tropics has long been recognized as one of the most difficult prediction problems in meteorology. The Australian tropics are no exception, with the skill of the methods currently available for operational short-term forecasting of synoptic scale features, tropical cyclones, and precipitation all having an accuracy below that of climatology and/or persistence (Leslie et al. 1981; Keenan 1982; Holland et al. 1987).

The Australian Bureau of Meteorology has a commitment through its research center to enhance the quality of operational forecasting in the Australian tropics, and several important initiatives have taken place, notably the Australian Monsoon Experiment (AMEX) of December 1986 to February 1987 (Holland et al. 1986), the aim of which is to improve the present level of understanding of the tropical atmosphere of the region embracing northern Australia, Papua New Guinea, and Indonesia during the (Southern Hemisphere) summer months of December to February. The summer monsoon that occurs annually at this time is crucial to the economy of northern Australia as it is responsible for most of the rainfall received each year.

As another part of this commitment to the improvement of operational prediction in the Australian tropics, a real-time trial of existing operational and research methods for the short-term forecasting of precipitation in the Australian tropics was carried out for the city of Darwin during the 1986/87 monsoon, or "wet," season.

The two basic aims of this real-time trial were, first, to evaluate and compare all available techniques to see if there are methods superior to the present operational methods and, second, to see if these methods exhibited gains in accuracy over climatology and/or persistence.

The methods for the short-term prediction of precipitation fall into four main groups: subjective, or manual, forecasts issued by duty forecasters on the basis of their experience and usually assisted by certain objective forecasting tools; statistical forecasts resting on observation-based relationships of various kinds between predictands and predictors (such as Markov chain models); deterministic predictions from schemes such as numerical weather prediction models; and hybrid systems based on combinations of the three categories above. For example, model output statistics (MOS) schemes are a combination of statistical and numerical weather prediction techniques.

An earlier real-time study was made of the accuracy and reliability of all the available techniques for predicting short-term winter precipitation in the midlatitude city of Melbourne (Fraedrich and Leslie 1987a). Six different forecasting techniques, drawn from the four groups mentioned above, were compared and it was found that the best scheme was an error-minimizing linear combination of the Markov chain model of Miller and Leslie (1985) and the Australian region operational numerical weather prediction model of Leslie et al. (1985). The skill of this technique was so high that its use as an operational model was urged strongly by Fraedrich and Leslie (1987a), and it was hoped that

---

* On leave from Institut für Meteorologie, Freie Universität Berlin, Berlin 41, West Germany.

Corresponding author address: Dr. Klaus Fraedrich, Bureau of Meteorology Research Centre, GPO Box 1289K, Melbourne, Victoria 3001, Australia.

a similar level of skill might be maintained when it was applied to the tropics.

In the present study, six techniques for the short-term prediction of precipitation at Darwin (including the operational manual forecasts) were compared in a 90-day trial occupying the "core" of the wet season from December 1986 through February 1987. These techniques included the operational (manual) forecasts, a Markov chain model, the Australian region primitive equations (ARPE) numerical weather prediction model, a model output statistics (MOS) scheme, and two error-minimizing linear combination schemes. All schemes were compared with climatology and persistence forecasts, which can be regarded as defining levels of zero skill. Major differences from the Melbourne study were, first, Darwin rainfall is much more intense than Melbourne winter rainfall so two categories of rainfall, light and heavy (⩾5 mm/6 h), were predicted and, second, Darwin's daytime rainfall has pronounced maxima in intensity and frequency in the afternoon so morning and afternoon predictions were made. A total of three predictions, therefore, was made for each daytime 12-h period: a forecast of any rainfall occurrence in the period 0600 to 1800 LST, and separate forecasts of the occurrence of heavy precipitation in the morning (0600 to 1200 LST) and afternoon (1200 to 1800 LST) intervals.

2. Climatology of Darwin precipitation

The climate of Darwin (12°S, 131°E) is tropical with a monsoon season lasting from mid-December to late March; transition periods from October to mid-December and the month of April; and a dry season occupying the months of May to September. The average annual precipitation is 1536 mm, with the months of December, January and February receiving 224, 409 and 355 mm, respectively.

The core of the wet season, December to February, was chosen for the operational trial. Most of the rain days result from monsoonal weather interspersed with periods of sunny, hot, and humid days, often with a thunderstorm in the afternoon or at night.

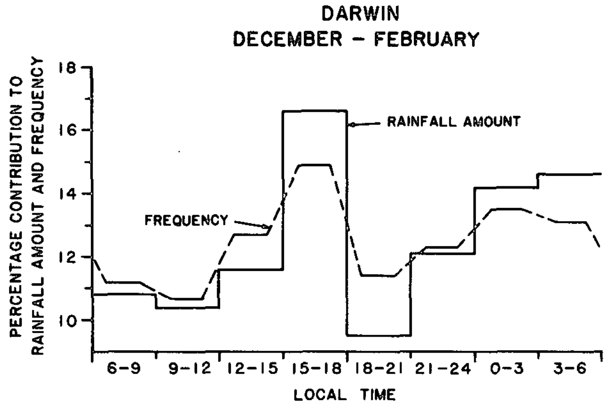

FIG. 1. Diurnal variation of rainfall amount and frequency at Darwin during the wet season December through February (1962–81).

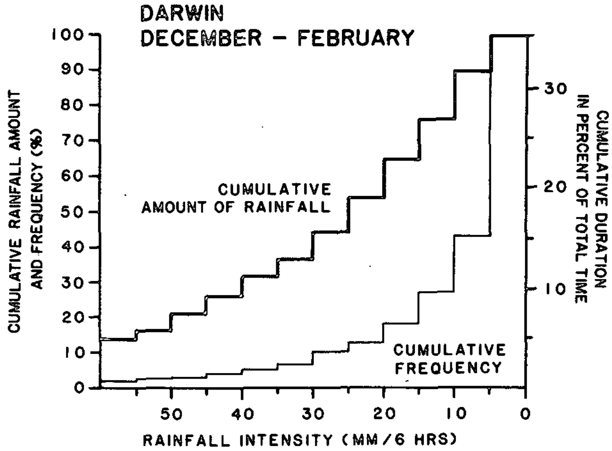

FIG. 2. Cumulative distribution of rainfall amount and frequency vs 6-h rainfall intensities for Darwin (1962–81). For example, this means precipitation occurs in 35 percent of all 6-h periods, and 15 percent of the same total are rainfalls of 5 mm or more.

An analysis of observations at 3-h intervals reveals two maxima in the frequency and amount of precipitation, one in the late afternoon and the other during the night between 0000 and 0600 LST (Fig. 1). The 12-h daytime period 0600 to 1800 LST was chosen for the forecast trial mainly because of its greater significance to the public.

On the average, about 49 percent of the daytime half-days are wet, i.e., 37, 53 and 56 percent in December, January and February, respectively. These frequencies are for rainfall greater than zero measured at Darwin Aerodrome and denoted by RR in the international weather code book. They increase to an average of 73 percent (67, 76 and 76 percent in December, January and February) if the definition of a wet half-day includes precipitation not only at the station but also evidence of precipitation in the neighborhood (denoted by the international weather codes WW = 13–17, 20, 21, 25, 27, 50–69, 80–99).

a. *Classification of precipitation into the categories of light and heavy*

The half-day interval 0600 to 1800 LST was subdivided further into the 6-h morning (0600 to 1200 LST) and afternoon (1200 to 1800 LST) periods. This was suggested in Fig. 1 which shows a clearly defined afternoon maximum in rainfall amount and frequency. An analysis of 6-h rainfall intensities (Fig. 2) shows that precipitation rates greater than or equal to 5 mm/6 h account for 90 percent of the total rainfall but occur in less than 50 percent of the wet 6-h time intervals and occupy less than 15 percent of the total time. Owing to this large drop in frequency, it was decided to choose

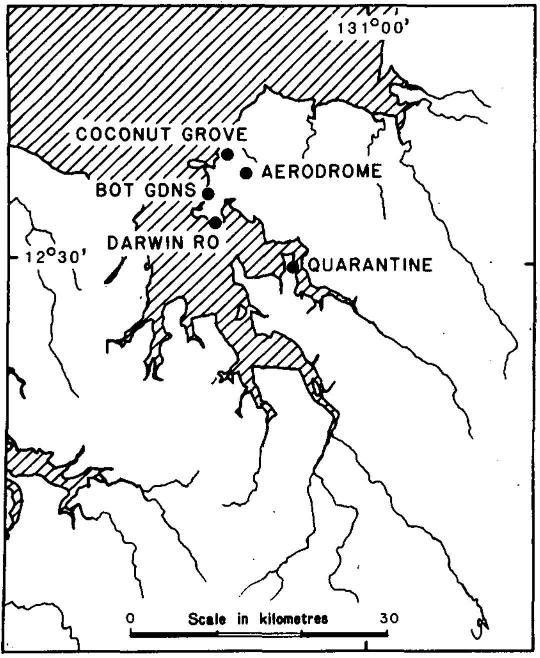

FIG. 3. Location map for rainfall stations in the Darwin area used in this study.

the rate of 5 mm/6 h as a climatologically meaningful threshold intensity for the heavy rainfall category. The actual climatological values of the occurrence of heavy rainfall were 0.12 for the 0600 to 1200 LST period and 0.16 for the 1200 to 1800 LST period. Thus, the likely occurrences of heavy rainfall on a given day in the monsoon period are about 12 percent in the morning and 16 percent in the afternoon. These are small but still significant probabilities. For purposes of comparison, the corresponding climatological values for the midlatitude station of Melbourne are only about 4 percent in the rainy winter season.

b. Representativeness of single station predictions

A basic question for the city of Darwin, like any other metropolitan area, is how typical the single recording station precipitation is for the entire city. To illustrate the representativeness of the Darwin Aerodrome station, daily precipitation cumulants are calculated from seven years (1967–73) of pluviograph data for four neighboring stations whose locations are given in Fig. 3. Conditional frequency distributions of rainfall occurrence are shown in Fig. 4 and indicate the relative number of stations reporting precipitation if it has rained/not rained at Darwin Aerodrome during a 24-h period.

The observed probability distributions of Fig. 4 for the wet season show that if rain has fallen at Darwin Aerodrome during a 24-h time period, there is more than a 95 percent chance that at least one of the other four stations will report rain. Conversely, if it has not rained at Darwin Aerodrome, there is a 65 percent

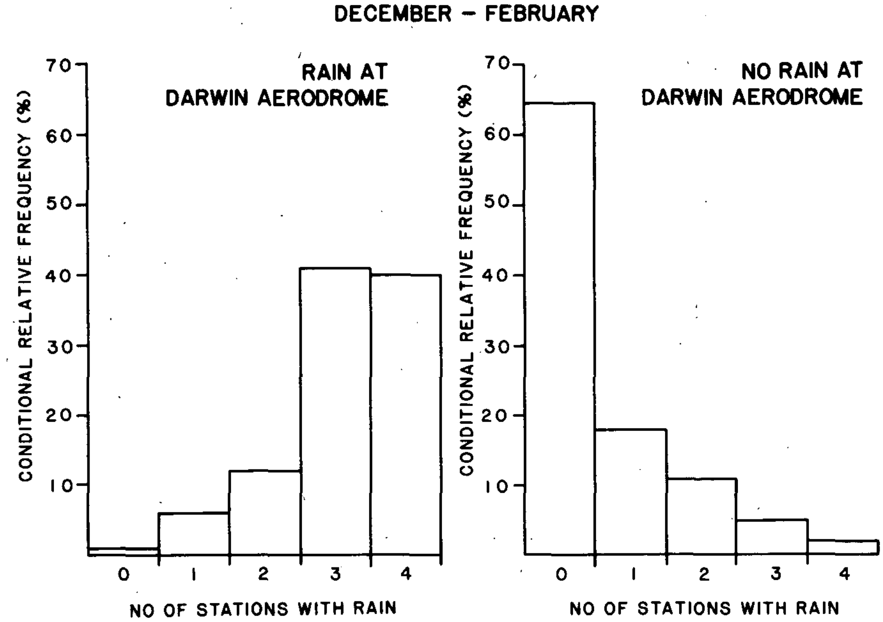

FIG. 4. Relative frequency distribution of the number of suburban stations with rain on the same day for which Darwin Aerodrome station reports rain (left) or no rain (right). The locations of the four stations Coconut Grove, Botanical Gardens, Darwin Regional Office, and Quarantine Station are shown in Fig. 3. The statistics cover eight years of homogeneous pluviograph records during the summer monsoon.

chance that it has not rained at any of the other four stations.

The climatological distribution, therefore, indicates that the rainfall predictions for Darwin Aerodrome are representative of the entire metropolitan area, but could possibly be refined further by a careful use of the data examined for Fig. 4.

3. The rainfall forecasting techniques

a. Manual forecasts

Manual forecasts were issued at 0600 LST each day and also at 1800 LST the previous evening, thereby enabling an evaluation of the influence of a 12-h lead time on the skill of the predictions. The actual predictions made by the forecasters were probability of precipitation (PoP) forecasts and categorical (wet/dry) forecasts. These forecasts took the form of predictions for the entire 0600 to 1800 LST half-day and for the occurrence of heavy rainfall (⩾5 mm/6 h) in the morning (0600 to 1200 LST) and afternoon (1200 to 1800 LST) intervals.

The forecasters had the 0600 LST observations available, plus all climatological data and the NWP and MOS forecasts, but not the Markov chain forecasts. In addition they were introduced to PoP forecasting by a three-week preliminary trial prior to the real-time experiment, but had no feedback during the experiment.

b. Statistical forecasts

1) MARKOV CHAIN MODEL

The basis of the purely statistical forecasts in this study was the Markov chain model developed originally by Fraedrich and Müller (1983) and applied by Miller and Leslie (1985) to a number of major Australian cities. This model was fitted to the archived 3-h surface data for Darwin and used to predict the probability of precipitation for the half-day period 0600 to 1800 LST and to predict the probability of heavy rain (⩾5 mm/6 h) in the morning and afternoon periods.

For the half-day forecasts in which the only prediction is the probability of any precipitation, four mutually exclusive states are used. These are three cloud states (0–2 oktas, 3–5 oktas, and 6–8 oktas, all with no rain) and one rain state, which is the same as that mentioned in section 2, and referred to as WW-rain. For the Markov chain model of Miller and Leslie (1985) the probability of precipitation for a given state j (=1, · · ·, 4), at current time t, and for h hours ahead is:

$$\mathrm{PoP}(j, t, h) = a(j, t, h) + \sum_{k=1}^{4} b_k(j, t, h)X_k.$$

The covariates X_k ($k = 1, \cdots, 4$) are respectively the surface pressure, the diurnally corrected 3-h pressure change, the dewpoint depression, and the east–west component of the wind. Other covariates, such as the north–south wind component, were found to be of less significance. The intercepts a are different for each month but common slopes b are fitted for all months.

The Darwin Markov chain model is summarized in Fig. 5 where the PoP (percent) is plotted as a function of the diurnally corrected 3-h pressure change for each of the three cloud states and the rain state. The range of PoP values for the individual months of December (D), January (J) and February (F) also is indicated. The effects of dewpoint depression are included and climatological values are indicated for reference purposes.

An assessment has been made of the effect of the covariates on the accuracy of the Markov chain forecasts. Inclusion of the covariates affects the probability of precipitation forecast values on a given day by as much as 20 percent. However, the half-Brier scores for the Markov chain forecasts made with and without the covariates on the dependent dataset (1962–81) were almost identical, being 0.143 and 0.145 respectively. This result suggests that the Markov chain could be used in its original form (Fraedrich and Müller 1983) with the cloud cover and rain states as predictors, with little change in overall accuracy. For this study, it was decided to include the covariates because they gave improved probability forecasts at the ends of the range, i.e., in the 0 to 10 and 90 to 100 percent PoP intervals.

Furthermore, because most of the accuracy of the Markov chain forecasts is contributed by the basic cloud cover and rain states, the model may be applied to future seasons with considerable confidence. In the dependent data sample, the climatological probabilities of precipitation for the four states are distinctly different, the seasonal mean values for 12-h forecasts being 0.366, 0.643, 0.733 and 0.916 respectively for the three cloud cover and the rainfall states.

For the predictions of the 6-h probability of heavy precipitation, two rain states are used. These are precipitation less than 5 mm, and precipitation greater than or equal to 5 mm in the past 6 h, giving a total of five mutually exclusive states for the Markov chain.

2) PERSISTENCE AND CLIMATOLOGY

For reference purposes forecasts based on persistence and climatology are included. These forecasts provide a kind of "zero-skill" base from which to evaluate the quality of the forecasts obtained by the other techniques described in section 3.

c. Numerical weather prediction model

The operational Australian region NWP model is a 12-level, 150 km horizontal resolution model that provides twice-daily 36-h forecasts from base times of 1200 UTC and 0000 UTC, i.e., 2130 and 0930 Darwin LST. Full details of the model are given by Leslie et al. (1985).

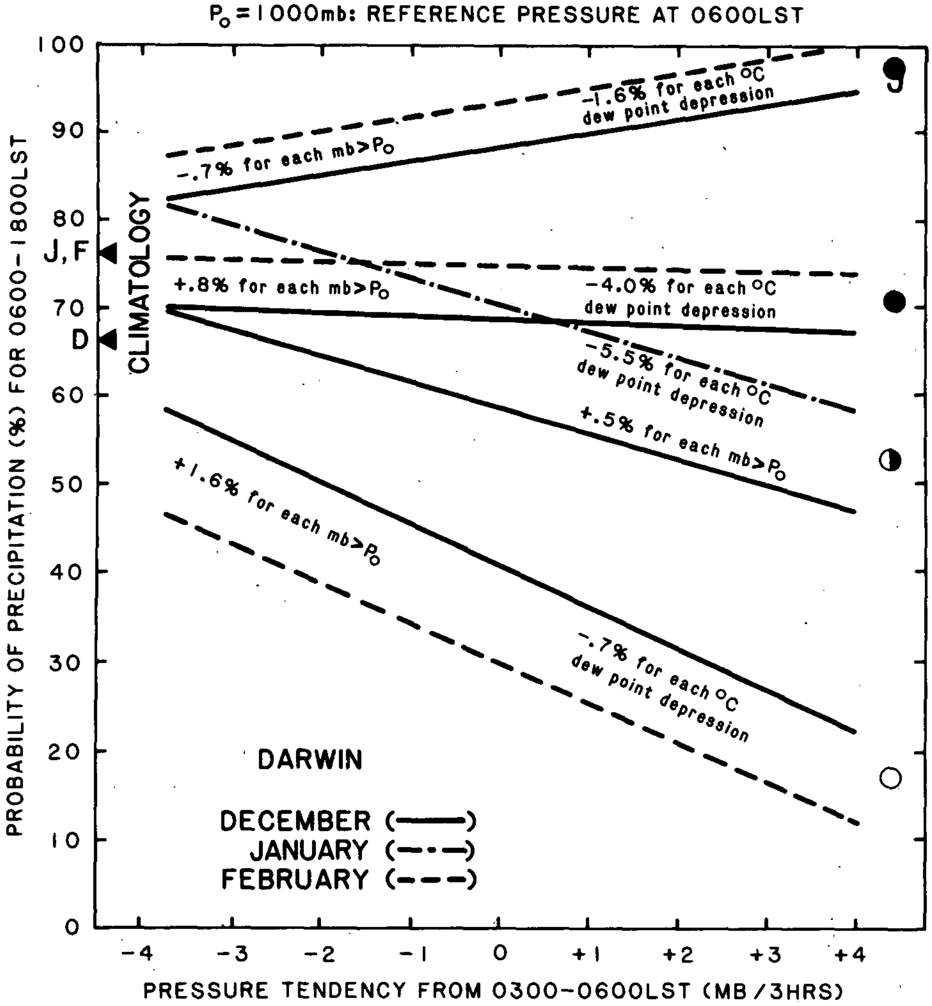

FIG. 5. Darwin Aerodrome single-station Markov chain probability of precipitation (0600–1800 LST) changing with pressure tendency, cloud cover, and rainfall states observed during the preceding hour at fixed surface pressure 1000 mb. ○, ◑, ●, ❟ represent the 0–2, 3–5, 6–8 octas and rain states, respectively. Note that the PoP change also with dewpoint depression, surface pressure, and the east–west surface wind (not indicated here). Climatology is included. D, J, F = December, January, February. Parallel lines (dashed, full, or dash-dotted) denote largest monthly differences for each cloud cover or rainfall state.

Rainfall forecasts for the period 0600 to 1800 LST in Darwin are provided from both the 2130 LST and 0930 LST base time forecasts. Thus there are lead times of 8½ and 20½ h in the Darwin rainfall predictions.

d. Hybrid models

These models take the form of combinations of other models. Two kinds of hybrid models were used in this study, the model output statistics (MOS) scheme and an error-minimizing linear combination of various schemes.

1) MOS SCHEME

Model output statistics forecast equations were derived for Darwin, based on the regressions from the NWP forecasts of section 3c from 0000 UTC (0930 LST) the evening before. Thus there is a lead time of 20½ h for the MOS scheme. No MOS regression equations are available at other times of the day. The predictions were rainfall and PoP for the entire half-day and for the morning (0600 to 1200 LST) and afternoon (1200 to 1800 LST) periods (Tapp et al. 1986).

2) LINEAR COMBINATIONS OF FORECASTING TECHNIQUES

Several independent forecasting methods were combined in an error-minimizing linear fashion:

$$\text{PoP}_{\text{combination}} = a\text{PoP}_{\text{scheme 1}} + (1 - a)\text{PoP}_{\text{scheme 2}}$$

where a is the optimal weighting factor calculated from the 20-yr dataset 1962–81. The value of a depends on the particular schemes that are combined (for more details, see Fraedrich and Leslie 1987b). Schemes which were combined included Markov and persistence, forecasters (categorical) and climatology, NWP model and climatology, and persistence and climatology.

4. Results of the real-time trial

a. The 1986/87 climatology

The real-time trial was carried out over a period of three months from December 1986 through February 1987. A basic question is how typical the 1986/87 wet season was in Darwin. As shown in the observations part of Table 1, the 1986/87 season was very close to climatology. There were 49 percent of half-days (0600 to 1800 LST) with measurable rain (RR). In December there were 7 rain days plus 3 trace days, in January 19 rain days, and in February 19 rain days plus 1 trace day. Climatological averages are 11, 18 and 16, respectively. The 1986/87 season was drier than average in December but wetter in January and February, because the monsoon started later and was wetter when it did arrive. If nearby rainfall was included, using the WW observations, a much wetter picture emerges, with 79 percent rainy half-days (WW): 19 in December, 27 in January, and 25 in February, compared with climatological averages of 18, 23 and 21, respectively.

Owing to the large differences between wet half-days defined by Darwin Aerodrome measurable rain (RR) and weather code rain (WW) it was decided to include both results and to verify for both the RR and WW wet half-days. In broad terms this might be thought of as a distinction between point and areal rainfall. In addition, trace rainfall (that is, less than 0.2 mm in the forecast period) was observed on four occasions and was verified differently, as indicated below.

b. Brier scores

An objective measure for comparing the accuracy of the precipitation forecasts is the half–Brier score (Brier 1950) or mean square error of probabilistic forecasts, given by

$$B = \frac{1}{m}\sum_{i=1}^{m}(\delta_i - \mathrm{PoP}_i)^2,$$

where m (=90 for this trial) is the number of forecasts, PoP_i is the predicted probability of precipitation for the ith day, and $\delta_i = 0$ or 1 if the observation was dry or wet at Darwin Aerodrome for WW and RR verifications. Rain trace is taken as $\delta_i = 0$ for RR verifications. The half–Brier score has a minimum value of 0 if all forecasts were perfect. For categorical predictions, such as those from the NWP model or from persistence, $\mathrm{PoP}_i = 0$ or 1 for dry or wet forecasts. For these binary forecasts the half–Brier score is identical with the relative number of incorrect predictions (Murphy 1986).

The half–Brier scores for the predictive schemes are shown for both the WW and RR verifications (RR is in parentheses) and, wherever possible, for two different lead times (Table 2). A number of conclusions may be drawn from Table 2:

(i) The best predictions for both WW and RR rainfall are generated by the Markov chain, even though it was calibrated only for the WW-rain forecasts. Climatology is next, followed by the forecasters (both PoP and categorical forecasts), MOS, and persistence. The NWP model performs worst of all with only about 50 percent correct forecasts. In fact the NWP model accuracy is worse for the shorter lead time, as is the forecasters', a feature also noted by Holland et al. (1987) for Darwin predictions.

(ii) The forecasters' PoP predictions are better when compared with WW data indicating a tendency to overforecast RR rain. The forecasters' categorical predictions of RR are slightly worse than climatology but superior to persistence.

(iii) Unfortunately there were five missing days of manual forecasts. Assuming the extreme case that all five forecasts are correct, i.e., $(\delta_i - \mathrm{PoP}_i)^2 = 0$, or incorrect, $(\delta_i - \mathrm{PoP}_i)^2 = 1$, the half–Brier scores, B, decrease [or increase] to $85B/90$ [to $(85B + 5)/90$]. This amounts to a 6 percent reduction or 22 percent increase

TABLE 1. Relative frequencies of Darwin half-days (0600–1800 LST) with rain observed and predicted during the 1986/87 December–February wet season. Observations are taken from the Darwin Aerodrome station international weather code (WW) and raingauge (RR). Climate refers to 1962–81 means.

	Rainfall climatology of									
	observations		predictions							
Dec–Feb	Climate	1986/87 season	Forecasters		NWP		Persistence	MOS PoP > 0.5	Markov PoP > 0.5	Markov − persistence
Lead time (h)	—	—	0	12	8½	20½	12	20½	0	0
Percent rain WW (RR)	.73 (.49)	.79 (.49)	.78	.79	.54 (.46)	.47 (.40)	.79 (.49)	.66	.77	.78 (.60)

TABLE 2. Half-Brier scores for Darwin half-day (0600–1800 LST) probability of precipitation forecasts. For lead times of various schemes, see Table 1. Note that forecasters also give categorical predictions (i.e., rain/no rain). The predictions are verified by Darwin Aerodrome international weather code WW and raingauge RR (in parentheses) observations.

	Predictive schemes										
	Probabilistic				Categorical			Combination			
Lead times	Forecaster	MOS	Climate	Markov	Forecaster categorical	NWP	Persistence	Forecaster categorical – climate	NWP – climate	Persistence – climate	Markov – persistence
0h/8½ h WW (RR)	.195 (.219)	/	.153 (.216)	.147 (.173)	.200 (.365)	.511 (.544)	.222 (.356)	.144 (.231)	.263 (.305)	.151 (.217)	.127 (.167)
12h/20½h WW (RR)	.187 (.226)	.221 (.223)	.153 (.216)	/	.226 (.381)	.461 (.539)	/	.166 (.242)	.231 (.289)	/	/

of the half–Brier score, respectively, which means that the forecaster's accuracy would not have been seriously degraded by the missing days.

(iv) The categorical schemes all performed poorly but could provide skillful and reliable forecasts when combined linearly with climatology:

$$\mathrm{PoP}_{\mathrm{combination}} = 0.5\mathrm{PoP}_{\mathrm{categorical}} + 0.5\mathrm{PoP}_{\mathrm{climatology}}.$$

The weighting factor 0.5 holds exactly for unbiased predictions producing the same climatology as observed (Fraedrich and Leslie 1987b), thereby minimizing the half–Brier score. The combined forecasts produce slight to significant improvements over the individual schemes. In particular, the combined forecasters' (categorical)–climate, and the persistence–climate combinations show considerable quality improvements, coming close to the accuracy of the Markov forecasts when verified against WW observations. For RR-rain this improvement is still large but smaller because of the difference between the 1986/87 observations and the 20-yr climatology. Although the number of RR events for the 20-yr climatology and 1986/87 season are similar, the actual distribution of RR events varies during the seasons. Also, the NWP–climate combination achieves a similar level of accuracy as the MOS scheme for the WW verification, thereby confirming an earlier result from a midlatitude realtime trial (Fraedrich and Leslie 1987a) and demonstrating its relevance to the tropics.

(v) All of the purely statistical methods—Markov chain, climatology, and persistence—have low Brier scores. This suggests that combining the methods may lead to an even greater accuracy. As one of the reviewers pointed out, since B is a quadratic scoring rule the score of an average forecast will be better than the average of the individual scores and this may contribute to the improvement obtained. Markov (M) and persistence (P) may be combined in an error-minimizing linear manner (Fraedrich and Leslie 1987b):

$$\mathrm{PoP}_{\mathrm{MP}} = a\mathrm{PoP}_{\mathrm{M}} + (1 - a)\mathrm{PoP}_{\mathrm{P}}$$

with the optimal weights $a = 0.66$ for WW-rain and $a = 0.90$ for RR-rain. As seen in Table 2 these forecasts show by far the best results. There may be a physical basis for the great accuracy of the Markov–persistence combination in this study. Persistence, which is on a time scale of a day, can reasonably be expected to provide good synoptic scale information in weather regimes where the events to be predicted occur more than about 50 percent of the time in extended bursts, such as monsoonal periods and breaks. The Markov chain, which operates on 3-h data, then acts like a mesoscale correction to the larger scale prediction provided by persistence.

c. *Reliability*

Despite the relatively small sample, reliability diagrams are useful in displaying both the reliability and

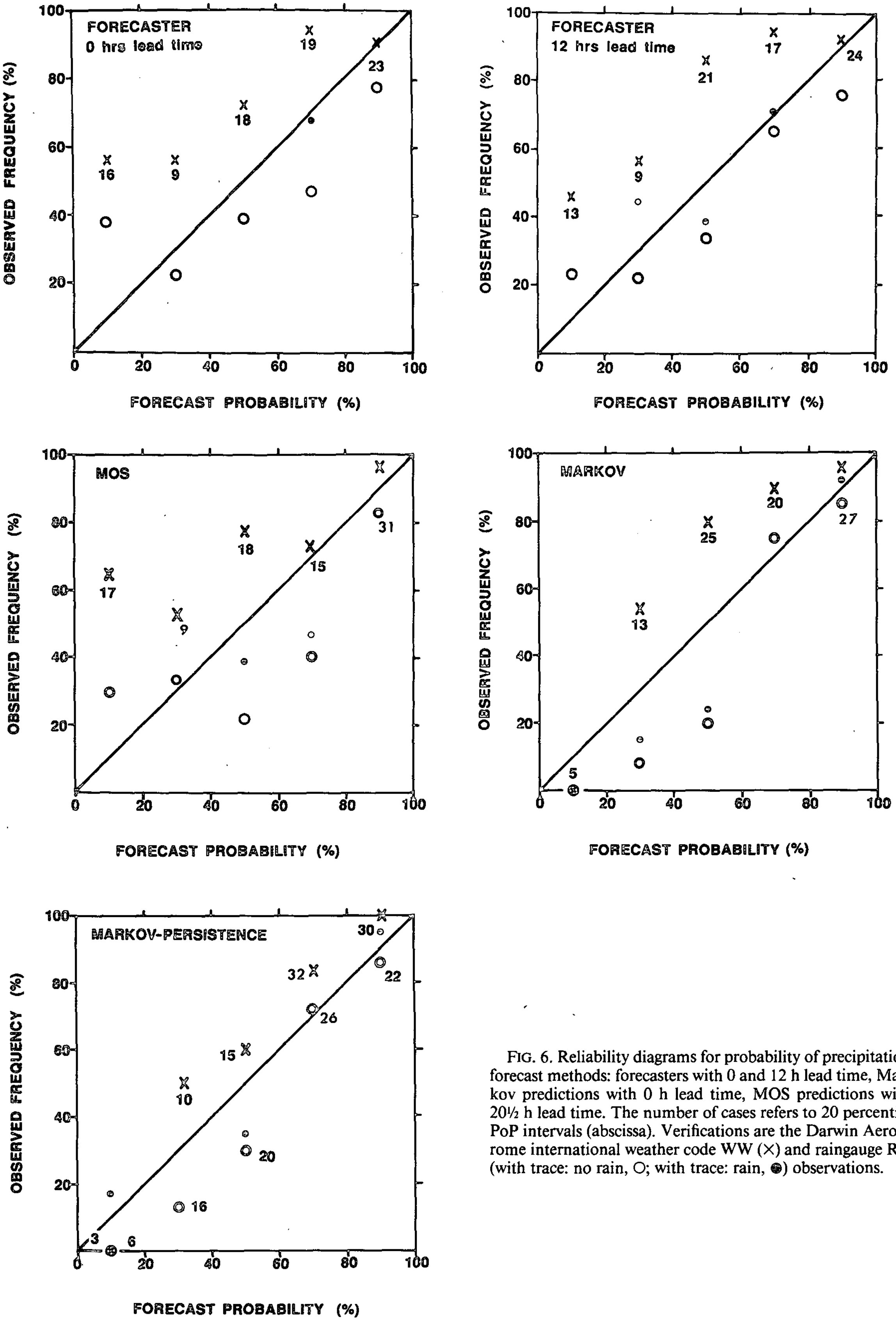

FIG. 6. Reliability diagrams for probability of precipitation forecast methods: forecasters with 0 and 12 h lead time, Markov predictions with 0 h lead time, MOS predictions with 20½ h lead time. The number of cases refers to 20 percentile PoP intervals (abscissa). Verifications are the Darwin Aerodrome international weather code WW (×) and raingauge RR (with trace: no rain, ○; with trace: rain, ●) observations.

distribution of the forecasts for the various PoP intervals. If the PoP predictions are perfectly reliable, then precipitation should occur on the percentage of days, say X percent, on which the PoP forecast was X percent. Figure 6 shows the realized relative frequency of precipitation, that is, the relative number of 12-h daytime intervals on which WW- and RR-rain (× and ○) were observed at Darwin Aerodrome. Trace rainfall (●) also is included. The diagonal in each diagram represents the theoretically achievable perfect reliability line. The following features are noteworthy:

(i) All schemes show a tendency to overpredict RR-rain and to underestimate WW-rain.

(ii) The forecasters and the MOS scheme show an almost equally distributed use of all PoP classes which as a consequence leads to large errors for small PoP values ($0 \leq \text{PoP} \leq 20$ percent). In particular the Markov and Markov–persistence combination performed well. The reliability of the forecasters for all other PoP values is surprisingly good, especially as they received no feedback during the trial.

(iii) The wet forecasts ($80 \text{ percent} \leq \text{PoP} \leq 100$ percent) exhibit almost perfect reliability.

d. Evaluation in categorical terms

Although categorical predictions are not optimal, they are a common practice in tropical rainfall forecasting. Therefore, we generated categorical forecasts of rainfall occurrence from probabilistic predictions (PoP) by interpreting $0 \leq \text{PoP} \leq 0.5$ as dry and $0.5 \leq \text{PoP} \leq 1.0$ as wet. This 50 percentile threshold value was chosen from Markov chain hindcast evaluations for which the categorical interpretation of PoP forecasts produces the minimum number of incorrect wet/dry predictions if verified against WW-rain. In the following this threshold also is used for the other probabilistic prediction schemes. Given the 50 percentile threshold, 2 × 2 contingency tables may be constructed of the form

	Predicted	
Observed	Wet	Dry
Wet	A	B
Dry	C	D

for all predictive schemes (Table 3 for WW- and RR-rain verification). Note that forecasters also issue categorical predictions which will be analyzed here. Unfortunately there were five missing days of manual forecasts, but this should not affect the accuracy greatly. Measures of forecast quality are calculated and are presented in Table 4 and discussed in the following:

(i) The overall skill of the forecasts in categorical terms is assessed by the discriminant of Hanssen and Kuipers, which is independent of the occurrence of rain (Woodcock 1976):

TABLE 3. Contingency table for predictive schemes verified at Darwin Aerodrome. Probability forecasts are categorized by the threshold value of 50 percent for dry/wet distinction. Figures refer to the international weather code WW and raingauge observations RR with trace = dry, in parentheses. Note that there were 5 fewer manual forecasts made than other forecasts (see text).

	Predicted "wet"							Predicted "dry"						
Observed	Forecaster categorical	NWP	Persistence	MOS	Climate	Markov	Markov – persistence	Forecaster categorical	NWP	Persistence	MOS	Climate	Markov	Markov – persistence
"Wet" WW (RR)	58 (37)	37 (18)	61 (28)	52 (37)	71 (37)	59 (42)	62 (41)	9 (5)	34 (26)	10 (16)	19 (7)	0 (7)	12 (2)	9 (3)
"Dry" WW (RR)	8 (29)	12 (23)	10 (16)	8 (23)	19 (22)	5 (22)	8 (13)	10 (14)	7 (23)	9 (30)	11 (23)	0 (24)	14 (24)	11 (33)

TABLE 4. The Hanssen and Kuipers score, percentage of correct forecasts, and bias for the predictive techniques. The numbers are based on the contingency table (Table 3).

Scores WW (RR)	Forecaster categorical	NWP	Persistence	MOS	Climate	Markov	Markov – persistence
Hanssen and Kuipers	.42 (.21)	.11 (.09)	.33 (.29)	.31 (.34)	.0 (.36)	.57 (.48)	.45 (.71)
Percent correct	.80 (.60)	.49 (.46)	.78 (.64)	.70 (.66)	.79 (.68)	.81 (.73)	.81 (.82)
Bias	.99 (1.57)	.69 (.93)	1.0 (1.0)	.84 (1.37)	1.27 (1.34)	.90 (1.34)	.99 (1.23)

$$V = (AD - BC)/(A + B)(C + D).$$

$V = 0$ if rain is predicted every half-day. Not unexpectedly the Markov chain model performs best for both WW- and RR-rain verifications (only exceeded by the Markov–persistence combination when verified against rain (RR). This is followed by MOS, persistence, and forecasters' predictions.

(ii) The relative number of correct wet and dry predictions is defined by

$$\text{percent} - \text{correct} = (A + D)/(A + B + C + D).$$

Here again the Markov chain performs best if verified against WW-rain; the differences among climatology, persistence and the forecasters are negligibly small, while MOS performs worse than (as well as) persistence for WW-rain (for measurable RR-rain).

(iii) Bias is a measure of the predictive schemes' climate versus the observed climate:

$$\text{Bias} = (A + C)/(A + B).$$

Persistence, of course, shows no bias which also holds for the forecasters' categorical predictions and very closely for Markov forecasts if verified against WW-rain. Compared with measurable rain (RR), there is a general tendency of schemes to overpredict rain except for NWP (which, however, did not show a noticeable overall skill).

e. Probability predictions of heavy rainfall

In addition to the PoP predictions of the occurrence of rain during the 12-h daytime period of 0600 to 1800 LST in Darwin, PoP predictions of heavy rainfall ($\geqslant 5$ mm/6 h) were made. These predictions were divided into the 6-h periods before and after noon (1200 LST). The time intervals and the intensity threshold were based on the climatological information described in section 2. It is worthwhile again to note that the sample size is small, as the trial was conducted only for one three-month period.

The results of the PoP predictions are presented in Table 5 and the following salient features emerge:

(i) The Brier scores clearly show the superiority of the Markov chain predictions, with a Brier score of 0.069 before noon and 0.094 after noon. Next in accuracy is climatology (0.102 and 0.116), followed by the forecasters (0.117 and 0.191), and the numerical weather prediction model (0.120 and 0.130).

The climatological chances of heavy rain are less than 20 percent in all months, being 0.07, 0.13 and 0.16 for 0600–1200 LST in December, January and February, respectively; and 0.11, 0.20 and 0.16 for 1200–1800 LST. During the period of this real-time trial, the respective observed monthly frequencies were close to the climatological values, being 1, 8, 3 (relative frequencies 0.03, 0.26, 0.11) for 0600 to 1200 LST and 1, 7, 5 (relative frequencies 0.03, 0.23, 0.18) for 1200 to 1800 LST in December, January, and February 1986/87 respectively. Thus, relatively uncommon events are being predicted, and the results should therefore be evaluated with this factor in mind.

(ii) All forecast methods show more accuracy in the morning than the afternoon. This may be due to the lead time and/or to the lower frequency of occurrence of heavy rainfall events in the morning. The greatest

TABLE 5. Half-Brier scores for Darwin (0600–1200 and 1200–1800 LST) probability of precipitation (RR $\geqslant$ 5 mm/6 h).

	Half-Brier score									
	Forecaster		MOS		Markov		Climate		NWP	
	Lead time									
	06–12	12–18	06–12	12–18	06–12	12–18	06–12	12–18	06–12	12–18
0 h/8½ h	.117	.191	/	/	.069	.094	.102	.116	.133	.133
12 h/20½ h	.122	.129	.220	.149	/	/	.102	.116	.111	.122
RR $\geqslant$ 5 mm/6 h	Probabilistic								Categorical	

degradation of accuracy between morning and afternoon forecasts comes from the forecasters who appeared to overestimate the strength of the afternoon maximum.

(iii) The forecasters and the MOS technique also were evaluated with lead times of 12 and 20½ h. The forecasters' accuracy in predicting heavy rain in the afternoon was greatly improved with a 12-h lead time.

5. Conclusions

An evaluation was made of six methods for predicting short-term (12 h ahead) precipitation at the Australian tropical city of Darwin. In a real-time trial during the 1986/87 monsoon season it was found that the only technique showing an overall level of accuracy greater than that of climatology was the Markov chain model. The current manual operational forecasts were inferior to climatology as were the other methods assessed, including the operational Australian region primitive equations NWP model and a model output statistics (MOS) scheme. Caution is needed in interpreting the findings of this real-time trial as it was conducted over only 90 days, a relatively small sample. However, the Darwin Regional Forecasting Centre intends to carry out more real-time trials, using the present findings as the basis for the future trials.

The results of the trial are consistent with findings elsewhere for short-term forecasting in the Australian tropics (see, for example, Leslie et al. 1981; Keenan 1982; Holland et al. 1987). They suggest that in the immediate future gains in accuracy can readily be obtained from a better use of archived data in statistical schemes such as the Markov chain model. Such models have much appeal because they do not rely heavily on computing power and other expensive technology not readily available to the Darwin regional office and other tropical centers around the world. On the other hand, improvements in skill of the deterministic forecasting techniques in the tropics, such as numerical weather prediction models, ultimately will come from a combination of better understanding of physical processes in the tropics, databases of high quality and resolution, and continuing advances in computer technology. However, these improvements are still some time in the future.

Acknowledgments. We would like to thank the staff and the Director of the Northern Territory Regional Office, Geoff Love, for enthusiastic support of the trial. Other persons we would like to acknowledge include Alan Miller, Jim Arthur, Mark Williams, Roger Tapp (who prepared the MOS forecasts and read the manuscript carefully), Peter Yew (who tabulated and coordinated the results of the trial), and Mrs. Fay Stroumos for typing the paper.

In addition, we acknowledge the reviewers and a coeditor of *Weather and Forecasting* whose comments and criticisms helped us greatly in revising this article.

REFERENCES

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

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

——, and L. M. Leslie, 1987a: Evaluation of techniques for the operational, single station, short-term forecasting of rainfall at a midlatitude station (Melbourne). *Mon. Wea. Rev.*, **115**, 1645–1654.

——, and ——, 1987b: Combining independent predictive schemes in short-term forecasting. *Mon. Wea. Rev.*, **115**, 1640–1644.

Holland, G. J., J. L. McBride, R. K. Smith, D. Jasper and T. D. Keenan, 1986: The BMRC Australian monsoon experiment: AMEX. *Bull. Amer. Meteor. Soc.*, **67**, 1466–1472.

——, L. M. Leslie, K. Fraedrich and G. B. Love, 1987: The challenge of very short range prediction in the tropics. *Proc. Symp. Mesoscale Analysis and Forecasting,* Vancouver, ESA SP-282, 287–295.

Keenan, T. D., 1982: A diagnostic study of tropical cyclone forecasting in Australia. *Aust. Meteor. Mag.*, **30**, 69–80.

Leslie, L. M., G. A. Mills and D. J. Gauntlett, 1981: The impact of FGGE data coverage and improved numerical techniques on numerical weather prediction in the Australian region. *Quart. J. Roy. Meteor. Soc.*, **107**, 629–642.

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

Miller, A. J., and L. M. Leslie, 1985: Short-term, single-station probability of precipitation forecasting using linear and logistic models. *Beitr. Phys. Atmos.*, **58**, 517–527.

Murphy, A. H., 1986: Comparative evaluation of categorical and probabilistic forecasts: Two alternatives to the traditional approach. *Mon. Wea. Rev.*, **114**, 245–249.

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.

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