On Single-Station Forecasting: Probability of Precipitation in Berlin

Klaus Fraedrich* and Klaus Müller

Institut für Meteorologie, Freie Universität Berlin, Dietrich-Schäfer-Weg 6–10, D-1000 Berlin 41, West Germany

(Manuscript received 07.02.1986, in revised form 13.03.1986)

Abstract:

Probability of precipitation (PoP) forecasts quantify the inherent uncertainty of the weather process and its prediction. Since 1983 a PoP-forecast experiment is conducted with the University Weather Service providing subjective short-term PoP-forecasts for Berlin which cover a 12-hour time interval without lead time. These forecasts are evaluated and compared with independent single station stochastic model predictions and observations at Berlin-Dahlem. Both subjective and model forecasts show a tendency to underestimate PoP's with subjective forecasts being less reliable in the intermediate PoP-range; the Brier scores of subjective and model predictions are comparable. Interpreting all predicted PoP's > 0.3–0.4 as precipitation (and vica versa), about 79 % of the subjective and 77 % of the model predictions were correct compared with 73 % correct point-forecasts by numerical weather predictions.

Zusammenfassung: Kurzfristvorhersage der Niederschlagswahrscheinlichkeit an einer Station (Berlin):

Vorhersagen der Niederschlagswahrscheinlichkeit (PoP) quantifizieren die Unsicherheiten, die zum Wetterablauf und seiner Vorhersage gehören. Seit 1983 wird ein Vorhersageexperiment für Niederschlagswahrscheinlichkeiten mit dem Universitätswetterdienst der Freien Universität Berlin durchgeführt, bei dem subjektive Kurzfristprognosen für ein 12-Stunden Zeitintervall erstellt werden. Diese Vorhersagen werden ausgewertet und verglichen mit unabhängigen Prognosen eines stochastischen Modells sowie den Beobachtungen an einer Station. Subjektive und Modellvorhersagen unterschätzen die Niederschlagswahrscheinlichkeit während des Experiments, wobei die subjektiven Prognosen weniger zuverlässig sind im mittleren Bereich der Wahrscheinlichkeiten; die Brier-scores von subjektiven und Modell-Prognosen sind von vergleichbarer Größe. Interpretiert man alle Niederschlagswahrscheinlichkeiten PoP > 0.3–0.4 als Niederschlag (und umgekehrt), so sind 79 % der subjektiven und 77 % der Modellvorhersagen für Berlin-Dahlem korrekt. Für denselben Zeitraum erhält man 73 % korrekte Punkt-Prognosen mit einem numerischen Wettervorhersagemodell.

Résumé: Prévisions à partir d'une seule station: probabilitè de précipitations à Berlin

Les prévisions de la probabilité de précipitations (PoP) quantifient l'incertitude inhérente aux processus atmosphériques et à leur prévison. Depuis 1983, une expérience de PoP est conduite avec le service des prévisions de l'université fournissant des PoP subjectives à courte échéance pour Berlin et couvrant une période de 12 heures. Ces prévisions sont contrôlées et comparées avec celles d'un modèle stochastique indépendant établi à partir d'une seule station. Les observations utilisées sont celles de Berlin – Dahlem. Les deux prévisions, subjective et à partir du modèle, montrent une tendance à sous-estimer la PoP, les prévisions subjectives étant moins fiables dans le domaine d'échéance intermédiaire. Les scores de Brier pour les deux types de prévisions sont comparables. En interprétant toutes les PoP > 0,3 – 0,4 comme des précipitations (et vice versa), 79 % des prévisions subjectives et 77 % des prévisions du modèle furent correctes. Ces valeurs sont à comparer aux 73 % de réussite des prévisions de caractère local établies à partir de modèles numériques.

* On leave at the Bureau of Meteorology Research Centre, Melbourne, Australia

0005-8173/86/03 0427–08 $ 03.00/0

1 Introduction

Short-term prediction of future weather states and its improvement are a current task of weather services. Although comprehensive models of mesoscale phenomena (and/or model output statistics applied to NWP models) are most commonly used as guidance for the forecaster, such models follow only one road to short-term prediction or local nowcasting. The other road is the application of simple stochastic models based on single station observations. These models offer means of providing rapid forecasts of future weather states (e.g. cloud amount or sunshine duration and precipitation) using single station observations which are immediately available at the starting time of the forecast. Stochastic models are fitted to the local climatology of a station for which the forecaster has to formulate his prediction. Furthermore, probability measures can be included in the forecast to quantify the inherent randomness or uncertainty of the local weather evolution leaving the risk evaluation of the forecast and the subsequent decision making process to the customer.

Useful stochastic models for local short-term weather prediction are Markov chains with a finite number of states observed at equally spaced discrete times. Such models have been used in meteorology since 1962 (GABRIEL and NEUMANN, 1962) to simulate daily rainfall in Tel Aviv, and, more extensively, to predict 24 different weather elements with an hourly time interval (MILLER et al. 1981) at stations throughout the United States. FRAEDRICH and MÜLLER (1983) used mathematical results deduced for Markov chains and other stochastic models to make probabilistic predictions; they estimated the time until the next occurrence of an event or the probability that a particular event will occur, say, r times in the next n time periods. Consequently, instead of the commonly used forecast for the chances of a weather state to occur at a prescribed time step n, they predicted the chances of a weather state to occur at least once within a prescribed time interval (say from 0 to n). Such interval forecasts are deduced from Markov chains fitted to various surface weather data of Berlin recorded every three hours. Following this approach of interval prediction, MILLER and LESLIE (1984, 1985) used Australian data to forecast the probability of rain up to twelve hours ahead. Furthermore, the University Weather Service (Berlin) conducted an experimental program; forecasters made their first subjective probability forecasts of precipitation during summer 1983, before probability forecasts were included to the routines for local services (telefon weather service since 1984, radio and television 1986). In this note we analyse the subjective interval forecasts for the probability of precipitation in Berlin issued by the University Weather Service during three summer and three winter seasons since 1983. We apply various verification schemes (reliability diagrams, Brier scores) to assess the subjective forecast quality and to compare it with a stochastic model. Finally, the probability forecasts are evaluated in relation to a categorial (non-probabilistic) interpretation (i.e. rain versus no-rain), which allows consumers to make optimum use of the predicted probability of precipitation by evaluating risk and forecast quality. Local clients of the University Weather Service seem to be satisfied with quantitative probability forecasts. A public response, however, is rare, and if so, mostly positive.

2 Probability of Precipitation (PoP) Forecasting

Single-station short-term forecasts of the probability of precipitation have been made for Berlin since 1983. These forecasts are issued at 6 GMT and express the probability of a precipitation event (PoP) to occur (at least) once within the following 12-hourly time interval from 6 to 18 GMT. Subjective PoP-forecasts prepared by the University Weather Service forecasters are based on all information available at that time (numerical weather predictions, regional weather maps, radar and satellite observations and, as additional guidance, the PoP's predicted by a stochastic model fitted to the single station data of Berlin). The forecasters were briefly acquainted with the related statistical background and, after an experimental phase during summer 83, subjective probability forecasts have been formulated on an operational basis. Thus 690 subjective forecasts are available for the following analysis.

The ideas for subjective PoP-forecasts are based on derivations from stochastic processes which are fitted to single station observations to estimate the probability of precipitation events during a short-term period (FRAEDRICH and MÜLLER, 1983; FRAEDRICH et al., 1984). More precisely, the models predict the chances for a precipitation event to occur at least once in a 12-hourly time interval following the last weather observation, which becomes immediately available at, say, 6 GMT. Therefore, the results of the subjective forecasts of PoP's are compared with results of a selected stochastic model which also serves as guidance for the forecasters. The model is a first order Markov chain whose inputs are four classes of past weather states describing the weather situation of the last three hours (i.e. from 3 to 6 GMT). These classes are three states of cloud cover (cloud cover above, about and below 4/8) and one state of precipitation including showers, etc.; these four states are mutually exclusive. This basic model is improved by allowing transition probabilities to change with the surface pressure; the fitting follows a linear regression estimation (MILLER, 1964). Our stochastic model is fitted to the single station data of Berlin-Dahlem (1970–80) for 10 summer seasons (May through August) and 10 winter seasons (November through February).

Reliability diagram (Figure 1): If the PoP's are well predicted, then precipitation should occur, for example, on x % of the cases for which the PoP-forecast was x %. Figure 1 shows the realized (relative) frequency of the occurrence of precipitation (i.e. the number of 12-hour time intervals from 6 to 18 GMT at Berlin-Dahlem with $RR > 0.0$ mm) versus the PoP's predicted for the same time interval. The diagonal line represents perfect reliability.

An objective measure of reliability

$$\text{rel} = |\overline{\text{PoP}_i} - \text{PoP}_{\text{obs}}|$$

is the absolute value of the difference between the average of the predicted PoP's ($\overline{\text{PoP}_i}$) and the observed relative frequency of precipitation (PoP_{obs}); rel = 0 is obtained for predictions free of bias. It appears that the stochastic model (rel = + 0.056) shows better reliability than the forecaster (rel = + 0.106) for the period under investigation.

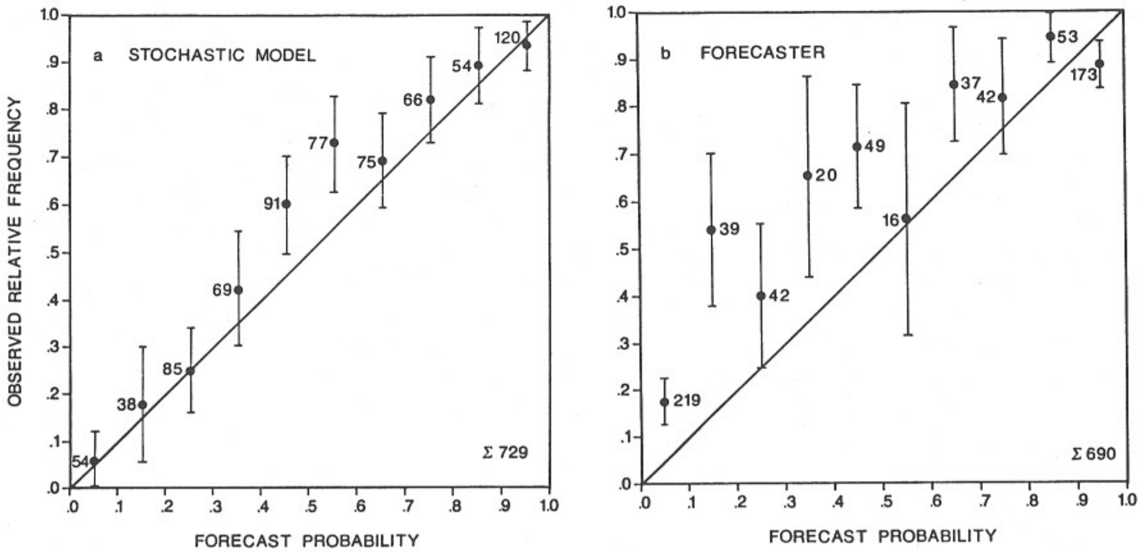
For each PoP-range the sample size (number of predictions) and the 95 % confidence intervals are added to the observed relative frequencies of precipitation (vertical axis). The length of the confidence intervals is approximately inversely proportional to the square root of the sample size. The PoP-forecasts are grouped in 10 % ranges for both stochastic model (Figure 1a) and subjective forecasts (Figure 1b). The following results should be noted:

(i) The relative frequency of precipitation occurrences rises when the predicted PoP increases; i.e. both model and forecasters are able to quantify the uncertainty inherent in forecasting precipitation.

(ii) Independent predictions of the stochastic model (Figure 1a) seem to be more reliable than the subjective forecasts (Figure 1b) which show larger deviations from the diagonal of perfect reliability in the domain $0.1 < \text{PoP} \leqslant 0.9$. One notices more PoP-intervals of subjective forecasts, whose confidence limits do not include the diagonal of perfect reliability, than PoP-intervals of stochastic model predictions.

(iii) Subjective predictions show a strong tendency to underestimate the PoP's in the sense that – using an extreme case – 35 (or 71 %) of 49 subjectively predicted PoP's in the range $0.4 < \text{PoP} < 0.5$ realize precipitation (Figure 1b).

(iv) The stochastic model also underestimates the daytime PoP's, because it has been calibrated by the complete data set including day and night time hours. However, the independent model PoP-predictions analysed in this note are made only for the day time hours 6 to 18 GMT; i.e. the influence of the diurnal



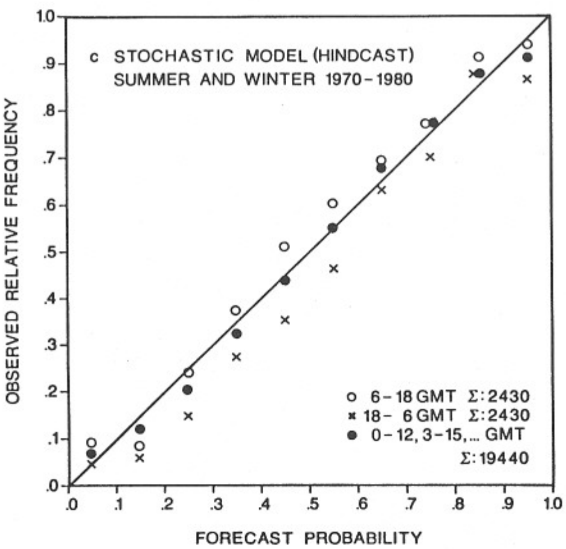


Figure 1

Reliability diagram: Relative frequency of occurrence of precipitation in each 10 % range of probability of precipitation (PoP) predicted for 12-hour time intervals. The number of forecasts and 95 % confidence limits are included for each PoP-forecasts interval; left (a): independent forecasts by stochastic single station model; right (b): forecaster of the University Weather Service (Berlin); (c) stochastic model PoP-hindcasts for day and/or night time intervals.

cycle leads to PoP under-(over) estimates during day-(night) time. This is shown in Figure 1c based on model hindcasts.

(v) Forecasters used the extreme classes PoP $\leqslant 0.1$ and > 0.9 most frequently i.e. for 57 % of their subjective predictions. PoP's $\leqslant 0.1$ reveal a small underestimation of the precipitation; it rains during 18 % of the cases. Whereas PoP's > 0.9 tend to overestimate precipitation; 88 % of the time intervals are wet. As these over- and underestimates are not too serious, one should expect a reasonably good skill score (next subsection). The intermediate PoP-range ($0.1 < \text{PoP} \leqslant 0.9$), however, is less often predicted by forecasters in contrast to stochastic model predictions.

(vi) Model PoP-predictions (Figure 1a) are more equally distributed over the complete PoP-range. In the intermediate PoP-range forecasters should use the guidance of the model.

■ **Table 1 Brier scores** of subjective (forecaster) and stochastic model predictions for all individual seasons. Climatology and persistence forecasts are added for comparison.

	Brier score of 12-hour time interval precition 6–18 GMT						
	Su 83	Wi 83/84	Su 84	Wi 84/85	Su 85	Wi 85/86	all seasons
forecaster	.116	.130	.218	.168	.173	.234	.173
model	.154	.184	.198	.137	.173	.160	.168
climatology	.244	.222	.260	.233	.255	.217	.238
persistence	.285	.408	.382	.250	.325	.289	.323

Brier score (Table 1): An objective measure to compare the quality of subjective forecasts and model performance is the half Brier score (BRIER, 1950):

$$B = m^{-1} \sum_{i=1}^{m} (PoP_i - \delta_i)^2$$

where m is the number of forecasts, PoP_i is the probability of precipitation predicted for the i-th time interval, and $\delta_i = 1$ or 0, if the time interval was wet or dry. The Brier score has a minimum value $B = 0$ for perfect forecasting. Table 1 shows Brier scores for subjective and model forecasts for individual summer and winter seasons and for all predictions combined. Furthermore, Brier scores based on climatology and persistence PoP-predictions are also included; climatology is defined by a fixed $PoP_i = 46.2\,\%$ for summer (60.8 % in winter); this is the mean relative frequency of wet ($RR > 0.0$ mm) 12-hour time intervals at Berlin-Dahlem. Persistence means $PoP_i = 1$ or 0 depending on whether the preceding (3–6 GMT) past weather state was wet or dry. The following results can be deduced from Table 1:

(i) Subjective and stochastic model PoP-forecasts show a considerably better score than climatology and persistence predictions.

(ii) The stochastic model obtains a slightly better Brier score than the subjective forecaster, if one discards the exceptionally dry summer 1983 and the winter 1983/84. For this summer the persistence forecast (of no-precipitation) performs particularly well in comparison to the other summer seasons.

PoP-Forecast evaluation (Table 2): Given a threshold value PoP*, we define all interval predictions PoP_i to fall into the category "wet" (precipitation), if

wet: $1 \geqslant PoP_i > PoP^*$

and into the category "dry" (no-precipitation), if

dry: $0 \leqslant PoP_i < PoP^*$

Now we can construct contingency tables for any threshold value PoP* relating the numbers of correct and incorrect predictions of PoP's to observed wet and dry categories. The number of incorrect predictions for dry or wet intervals increases (decreases) with increasing the probability threshold PoP*, which separates dry from wet categories (Table 2, first and second column). Thus there exists a critical threshold PoP_c which gives the minimum of incorrect predictions for both wet and dry time intervals (Table 2, third column; note that different numbers of wet or dry predictions enter the percentage of the total incorrect forecasts). This critical threshold value PoP_c defines a reasonable boundary for consumers to evaluate the PoP-forecast: the total forecast error obtains a minimum if a predicted

$$PoP_i \gtrless PoP_c \text{ is interpreted } \begin{matrix}\text{wet}\\ \text{dry}\end{matrix}$$

■ **Table 2 Percentage of incorrect PoP-predictions** by stochastic model and forecasters, if PoP's larger (smaller) than threshold value PoP* are interpreted as precipitation (no precipitation): incorrect forecasts are distinguished for wet or/and dry time intervals. One notices the minimum of 21 % (23 %) incorrect subjective (stochastic model) forecasts.

Frequency (in %) of incorrect forecasts of wet or/and dry 12-hour time intervals						
probability threshold wet > PoP* ⩾ dry	stochastic model wet	stochastic model dry	stochastic model total	forecaster wet	forecaster dry	forecaster total
> .1	36	6	34	23	18	21
> .2	33	11	30	21	24	22
> .3	27	18	24	17	26	21
> .4	23	24	23	16	28	22
> .5	18	34	26	14	34	25
> .6	16	41	30	12	35	25
> .7	11	46	34	12	39	29
> .8	9	50	40	10	43	32
> .9	8	53	46	12	48	39

■ **Table 3 Contingency table** of single station observations at Berlin versus predicted grid point precipitation from numerical weather prediction model (BKF). PoP-predictions by forecasters (MvD) and stochastic single station model (Model) are also included; PoP's > 0.3–0.4 are interpreted as wet, PoP's ⩽ 0.3–0.4 as dry categories (see Table 2). One notices about 27 % incorrect forecasts by the BKF-model (21 % by the forecasters, 23 % by stochastic model).

	observed								
predicted	dry BKF	dry MvD	dry Model	wet BKF	wet MvD	wet Model	Σ BKF	Σ MvD	Σ Model
dry	237	222	186	124	78	60	361	300	246
wet	70	65	107	294	325	376	364	390	483
Σ	307	287	293	418	403	436	725	690	729

From Table 2 one can evaluate the PoP-forecasts:

(i) The critical threshold of the minimum of incorrect model predictions lies at $PoP_c \sim 0.4$. Only 23 % of all independent model forecasts were incorrect, if one interpretes PoP's larger (smaller) than $PoP_c \sim 0.4$ as wet (dry) 12-hour time intervals. Stochastic model hindcasts of the summer and winter seasons (not shown) reveal a critical threshold $PoP_c \sim 0.5$ with a minimum of 23 % incorrect predictions.

(ii) Forecasters reduce the critical limit to $PoP_c \sim 0.3$; i.e. in the average they tended to underestimate the chances of precipitation as shown in Figure 1. The minimum of incorrect predictions lies at 21 % taking PoP's > 0.3 as wet (⩽ 0.3 as dry) time interval predictions.

Comparison with numerical predictions (Table 3): Stochastic model predictions (based on single station observations only) and subjective forecasts (based on all information available) can be compared with numerical weather predictions of precipitation provided by the Deutscher Wetterdienst BKF-model for the gridpoint Berlin (13.6 E, 52.2 N). From the 96-hour forecast issued at 00 GMT we used the 6–18 GMT amount of rainfall predictions and the observations at Berlin-Dahlem during the same 12-hour time interval. A contingency table (Table 3) of predicted and observed precipitation (discarding the amount) yields 27 % incorrect predictions of the BKF-model for the period under investigation.

The percent of incorrect predictions is equivalent to the half Brier score (see above) of categorial forecasts, if wet or dry predictions are measured by $PoP_i = 1$ or 0.

It should be noted (a) that the observations used for verification are the same for both stochastic single station or numerical weather prediction model and forecaster, and (b) that one observes a surprisingly good correlation between 12-hourly single station and areal precipitation (FRAEDRICH et al., 1986). Comparing the quality of deterministic and stochastic precipitation predictions one should realize that numerical models forecast the amount of rainfall, but we have used only the information whether there will be precipitation or not. PoP-predictions, however, quantify the uncertainty of the precipitation event.

The contingency table (Table 3) can be expanded by subjective and stochastic model PoP-forecasts, if we use the interpretation (deduced by Table 2) that a PoP-forecast means wet (dry) for PoP's above (below) the critical threshold PoP_c (i.e. PoP > 0.3–0.4 or < 0.3–0.4). One notices 21 % incorrect subjective and 23 % incorrect stochastic model predictions for the 3 summer and winter seasons.

3 Conclusion

The results of the Berlin PoP-forecast experiment described in this note may have an impact on regional short-term weather forecasts. At present, the uncertainty is seldom or not at all expressed in the forecasts issued by the weather services (MURPHY, 1981; DAAN and MURPHY, 1982). However, the experiments show that the uncertainty – in particular for the precipitation process – can be quantified. Furthermore, stochastic single station models seem to offer an excellent guidance for the forecaster. They show considerable reliability and good confidence in estimating uncertainty; another advantage is that they provide rapid forecasts based on single station weather observation, which are immediately available.

A method to evaluate PoP-forecasts has been introduced comparing probabilistic with categorial predictions. Interpreting all predicted PoP's larger (smaller) than $PoP_c \sim 0.3$–0.4 as precipitation (no-precipitation) we obtain the relatively high percentage of 77 % (79 %) correct stochastic model (subjective) forecasts. Not unexpectedly one observes $PoP_c \sim 0.5$ for both day and night time stochastic model predictions.

The single station stochastic models may be improved by including the diurnal cycle (e.g. FRAEDRICH and MÜLLER, 1983) or fitting to monthly (instead of seasonal) data. Although this would improve the forecast skill by a few points the impact on the forecaster may not be very relevant as she/he needs general guidance in the uncertainty estimates but not decide about a few percent of PoP-deviations.

Finally, we presume that the successful predictions of our stochastic single station forecast model are due to the observed cloud cover states, which enter as initial values. They are defined as past weather states of a three hourly period advancing the forecast time interval. These observed cloud cover states seem to define an ideal meso-scale forecast information set. We conjecture that this set of observations can hardly be replaced by related cloud parameterizations of numerical weather prediction models to obtain a similar skill of MOS-based PoP-forecasts.

Acknowledgement

Thanks are due to Mrs. M. SCHOLZ for typing the manuscript, Mr. H. HAUG for drawing.

References

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

DAAN, H. and A. H. MURPHY, 1982: Subjective probability forecasting in the Netherlands: some operational and experimental results. Meteorol. Rdsch. **35**, 99–112.

FRAEDRICH, K. and K. MÜLLER, 1983: On single station forecasting: sunshine and rainfall Markov chains. Beitr. Phys. Atmosph. **56**, 108–134.

FRAEDRICH, K., Th. DÜMMEL and K. MÜLLER, 1984: Stochastische Wettervorhersage an einer Station (Berlin). In "100 Jahre Deutsche Meteorologische Gesellschaft in Berlin 1884–1984", 209–236.

FRAEDRICH, K., R. BACH and G. NAUJOKAT, 1986: A single station climatology of central European fronts: number, time and precipitation statistics. Beitr. Phys. Atmosph. **59**, 54–65.

GABRIEL, K. R. and J. NEUMANN, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. Quart. J. R. Meteor. Soc. **88**, 90–95.

MILLER, R. G., 1964: Regression estimation of event probabilities. Travellers Research Center Techn. Report, 7411–121, Hartford, Coun., 153 pp.

MILLER, R. G., 1981: GEM: a statistical weather forecasting procedure. US Dept. of Commerce, NOAA, Silver Spring, Md., 89 pp.

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

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

MURPHY, A. H., 1981: Subjective quantification of uncertainty in weather forecasts in the United States. Meteorol. Rdsch. **34**, 65–77.