Meteorol. Appl. 8, 11–21 (2001)

Forecasting precipitation over Delhi during the south-west monsoon season

U C Mohanty, N Ravi, and O P Madan, *Centre for Atmospheric Sciences, Indian Institute of Technology, Hauz Khas, New Delhi - 110 016, India*

The south-west monsoon (June–September) is the major rainy season over India. Information about the occurrence of precipitation and the expected quantity at a specific place is important in many sectors of human activity. In this study, objective methods are developed to forecast the probability of precipitation (POP) and provide the quantity of precipitation forecast (QPF) over Delhi. As the onset of the monsoon at Delhi is around 30 June, the models are developed for the months of July, August and September (JAS) using surface and upper-air data for the period 1985–90 and tested with data from JAS for 1994 and 1995. A multiple linear regression equation is developed to forecast the POP and multiple discriminant analysis is used to produce the QPF in terms of one of four groups (0.1–1.0; 1.1–10.0; 10.1–30.0; and ≥ 30.1 mm). The QPF model is used only if precipitation is expected to occur (the POP forecast is turned into a categorical forecast). The categorical forecasts based on the POP exhibit positive skill scores consistently with both the development and independent data sets. The model for QPF also performed satisfactorily

1. Introduction

The south-west monsoon (June–September) is the rainy season which contributes about 80% of the annual rainfall over India. Precipitation is an important weather element that influences various activities. Information about the probability of precipitation (POP) over a specific location/region is important in many areas of human activity such as agriculture, aviation and construction. Knowledge of the expected quantity of precipitation helps in flood and water management. The existing approach for predicting the occurrence of precipitation at a location/region in India is mainly based on a subjective–synoptic approach. In the USA and many European countries, mainly statistical–dynamical methods are used for this purpose. Some of the related work can be found in the following:

- Glahn & Lowry (1969, 1972) used outputs from numerical weather prediction models and developed regression models to forecast POP over different parts of the USA.
- Paegle (1974) compared the forecasts of POP over different parts of the USA, derived from the equations stratified with respect to the synoptic weather pattern and the equations that were not stratified. It was found that the forecasts given by equations using the stratified method are more accurate.
- Kripalani & Singh (1986) developed composite charts of the distribution of probabilities of 24-hour rainfall amounts in two ranges, namely ≥2.5 mm and ≥65 mm, when a monsoon depression is present over the Indian region, at various locations along its track. It is a synoptic–climatological technique and provided reliable forecasts of rainfall probabilities when a monsoon depression is present.
- Upadhyay *et al.* (1986) developed a method to forecast precipitation by considering the fact that the precipitation rates are directly proportional to the large-scale vertical velocity. Using this method, precipitation rates were computed for specific monsoon depression situations over central parts of India.
- Kruzinga (1989) compared the forecasting of POP over the Netherlands using an analogue technique and logistic regression. It was found that for lead times of 1 to 3 days the regression method performed better than the analogue technique.
- Carter *et al.* (1989) discussed the performance of the statistical forecasts that are routinely used in the National Weather Service of USA, for contiguous USA and Alaska.
- Kumar & Ram (1995) developed a technique for quantitative precipitation forecasting over the Rapti catchment region in Uttar Pradesh, India. This is a synoptic-analogue method in which synoptic systems are classified according to the observed rainfall rates in the ranges 11–25 mm, 26–50 mm and >50 mm. Similar studies can be found in Lal *et al.* (1983), Sridharan *et al.* (1994), Dubey & Rajeevan (1994) and Singh *et al.* (1995) for different river catchments over India.
- Raj *et al.* (1996) developed a methodology for forecasting the amount of daily rainfall over

Thiruvananthapuram and Madras regions in the southern parts of India.

- Kumar & Maini (1996) discussed the development and performance of forecasts using the Perfect Prog Method for POP and quantity of precipitation, over various parts of India.

These studies indicate that the use of objective techniques to forecast POP for a specific location in India is very limited, unlike in countries such as the USA where the POP forecasts for specific locations are routinely generated for the entire country.

2. The study

In the present study, objective methods have been developed to forecast the probability of precipitation (POP) and produce a quantity of precipitation forecast (QPF) at Delhi. The normal date of onset of the monsoon over Delhi is 30 June. Hence the models have been developed for the months of July, August and September.

The POP model was initiated at 0830 IST giving the forecast of the POP for the following 24 hours. The QPF model was initiated at the same time only if the POP model indicated that precipitation might occur once the POP is turned into a categorical forecast. It may be noted that the QPF model gives the forecast of 24-hour accumulated precipitation in one of four groups. It is found that a separate model for forecasting the occurrence of precipitation gave consistently better results than including it as a category in the discriminant procedure of the QPF model. In the development of the QPF model, it was observed that rainfall quantity was highly variable for similar atmospheric as well as synoptic conditions. Therefore, the developed model may not be suitable for the entire Delhi region but it is applicable to the place for which it has been developed.

For the development of the model, a meteorological station situated in the north-east sector of Delhi at a distance of about 15 km from central Delhi (Figure 1) was selected (henceforth referred to as Delhi). A long and continuous record of data is available at this site.

The model equations are developed using surface and upper-air data from Delhi for the months of July, August and September (JAS) for the six-year period 1985–1990. Data from stations in the vicinity of Delhi (see Figure 1) are also considered so that advection effects and movement of synoptic systems can be taken into account. In order to develop a multiple regression equation, a total of 375 potential predictors (shown in Table 1) consisting of surface and upper-air observations plus derived parameters are utilised. During the process of selecting the potential predictors, care has been taken to ensure that there are no data gaps. Also, quality control checks are utilised: checks are made on the space, time and synoptic condition consistency. The individual data gaps are filled up by using linear interpolation between previous and subsequent meteorological observations. The developed model was tested with independent data sets from JAS for the period 1994 and 1995.

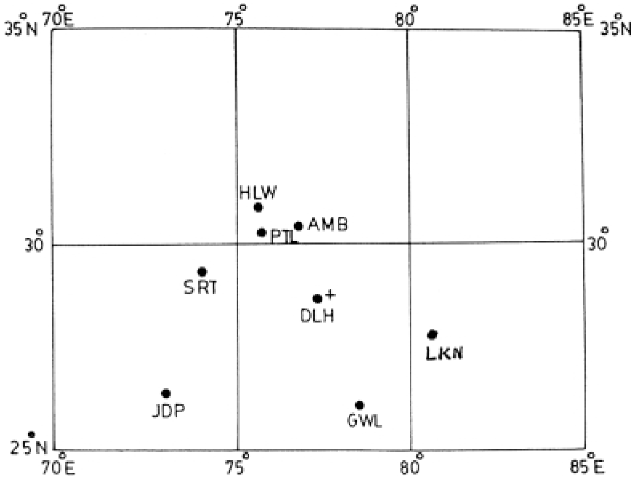

Figure 1. *The location of meteorological stations from which data have been used in this study. AMB: Ambala, DLH: Delhi, GWL: Gwalior, HLW: Halwara, JDP: Jodhpur, LKN: Lucknow, PTL: Patiala, SRT: Suratgarh. The selected place of study is indicated by +.*

3. Characteristic features of rainfall events at Delhi

Prior to the development of the model to forecast the POP and QPF, the characteristic features of rainfall events at Delhi were studied. Hourly data about the occurrence of precipitation events in the months of JAS were compiled for the six-year period (1985–90) covering all types of precipitation (i.e. drizzle, rain, shower and thundershowers). The hourly current weather observations report precipitation only if it occurs at the time of observation but not the quantity. Therefore, information about the occurrence of precipitation is extracted from hourly weather observations and the 24-hour rainfall is extracted from the 0830 IST synoptic observations. A day is categorised as a day of rainfall occurrence if precipitation is reported at any time in the 24-hours commencing from midnight (period 0001 to 2400 hour IST). July and August experienced precipitation on 15 and 14 days respectively, and the month of September had precipitation on 9 days. The decrease in the number of precipitation days in September is due to the withdrawal of the monsoon in the second to third week of the month.

The hourly distribution of occurrence of precipitation is presented in Figure 2. Although rainfall may occur at any time, it has one maximum in the forenoon around 1000 IST and another in the afternoon around 1600 IST. The forenoon maximum is generally due to the formation of stratus in the morning or forenoon and the afternoon maximum is due to increased convective activity. The important synoptic features which aid the occurrence of precipitation are the location of the axis of the monsoon trough (AMT) and its oscillation in the neighbourhood of the station (Rao, 1976), the existence of a slow-moving monsoon low/depression or the interaction between the monsoon and the mid-latitude systems such as western disturbances.

The 24-hour rainfall amounts are found to be highly variable even for similar synoptic conditions. Therefore, these amounts are classified into four groups and the number of cases in each group are presented in Table 2. Rainfall reported as "Trace" is taken as 0.1 mm for this purpose. The percentage frequency of rainfall amount in each group is presented in Figure 3. Table 2 and Figure 3 indicate that out of the 243 cases, nearly 66% are within the first two groups. The highest and the lowest rainfall amounts (in 24 hours) in the development data are 146.6 mm and 0.1 mm respectively.

It is important to point out that the classification of the observed rainfall amount (shown in Table 2) does not follow the India Meteorological Department (IMD) specification. The IMD classification is applicable for the entire country whereas the classification used in this study is applicable only to the selected location.

4. Methodology for forecasting probability of precipitation

In this study a multiple regression equation using a stepwise screening technique (Draper & Smith, 1966) was developed for forecasting the POP over a 24-hour period. For this purpose, the potential predictors (given in Table 1) were subjected to screening. Nine significant predictors were selected in the formulation of the equation for POP. An equation of the following type is assumed:

$$Y = b_o + b_1x_1 + b_2x_2 + ... + b_nx_n \qquad (1)$$

where Y is the value of the predictand obtained by a linear combination of various predictors x_1, $x_2.....x_n$, and b_1, $b_2.....b_n$ are regression coefficients and b_o is the regression constant. The stepwise procedure requires a stopping criterion for the selection of predictors. With such a criterion, the selection of the predictors continues until all significant predictors are included in the regression equation. In the present case, selection of the predictors is stopped when none of the remaining predictors reduces the variance by less than 0.5%.

Equation (1) gives the forecast in probabilistic terms. In the development of the equations, the value of the predictand, Y, is taken as 1 if precipitation occurs and 0 if it does not. The value of Y varies from 0 (0%) to 1 (100%). Using the developed equations, the values of Y are re-calculated for all observations from the developmental sample. If the re-estimated value of Y is greater than 1, it is made equal to 1 and if it is less than 0, it is made equal to 0. The re-calculated values of Y are grouped into intervals of 0.1. For each group the observed probability of occurrence and non-occurrence of precipitation are evaluated. The results are presented in Figure 4 for the POP model along with the best-fitting curve. The purpose is to find the critical value of Y for deciding the cut-off for converting the probability forecasts into a categorical forecast of occurrence or non-occurrence of precipitation. The cut-off of $Y = 0.45$ is chosen because for a smaller value the probability of non-occurrence is more than that of occurrence. Thus, based on this critical value, the probability forecast is converted into a categorical Yes/No forecast.

- If the value of Y is less than 0.45, non-occurrence of precipitation is forecast (No).
- If Y is greater than or equal to 0.45, occurrence of precipitation is forecast (Yes).

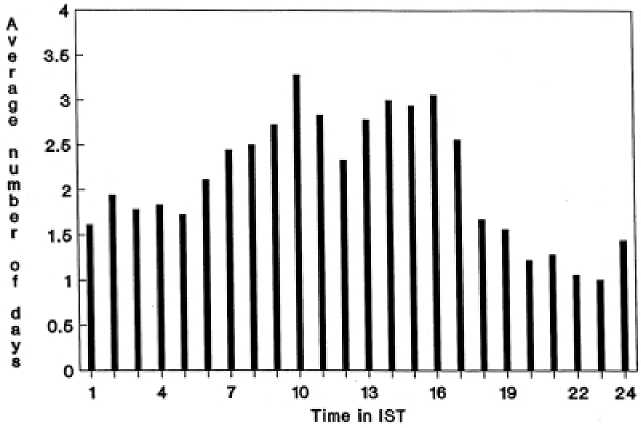

Figure 2. *Hourly distribution of precipitation events at Delhi.*

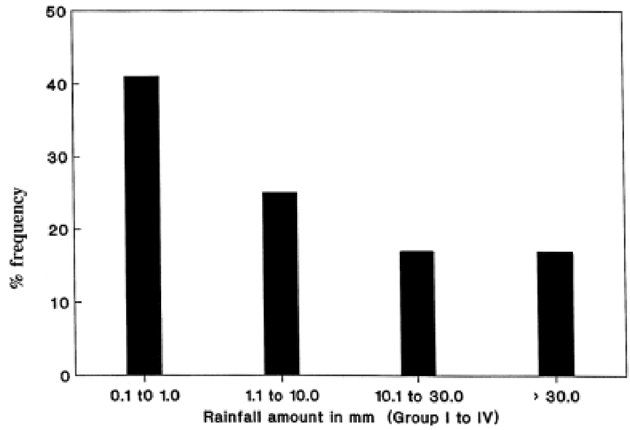

Figure 3. *Percentage frequency of rainfall amount in four groups, JAS 1985–90.*

Table 1. *List of potential predictors and their notations used in the present study. The total number of predictors is 375*

	Predictors and their notations	Stations	Total
Surface	Dry bulb (TT) and dew point (TD) temperatures, maximum (Tmax) and minimum (Tmin) temperatures. and their 24-hour changes (ΔTT, ΔTD, Δtmax, ΔTmin), dew point depression (DPD), relative humidity (RH), total cloud amount (TTLN), zonal (U) and meridional (V) components of wind.	Delhi (DLH) Ambala (AMB) Jodhpur (JDP) Gwalior(GWL)	144
Upper-air	Dry bulb (TT) and dew point (TD) temperatures, mixing ratio (w), zonal (U) and meridional (V) components of wind, potential, wet bulb and equivalent potential temperatures (? ,?w and ?e) at standard pressure levels (850, 700, 500, 400, and 300 hPa), thermal advection, wind shear, temperature differences between different levels from surface to 300 hPa, lapse rate of temperature potential equivalent.	Delhi, Jodhpur	208
Persistence	Occurrence of precipitation previous day.	Delhi	1
Stability indices	Showalter index , Rackliff index, Jefferson's modified index, Convective index of REAP, George index, Vertical totals index, Cross totals index, Total Totals index, Modified vertical totals total index, Modified cross totals index, Modified total totals index, Lifted index, Potential instability index, Severe weather threat index.	Delhi	15
Derived dynamical parameters	Divergence, vorticity and vertical velocity at different levels.	Delhi, Gwalior Jodhpur	7

Table 2. *Classification of observed rainfall amounts in the developmental data (JAS 1985–1990)*

Label	Group	Rainfall amount (mm)	Number of cases	Percentage number of cases
Very light	I	0.1 to 1.0	99	41
Light	II	1.1 to 10.0	60	25
Moderate	III	10.1to 30.0	42	17
Heavy	IV	≥ 30.1	42	17

The selected predictors that form the equation are mostly surface and upper-air parameters at Delhi. It may be noted that, in the process of screening, the data of neighbouring stations were also included but their contribution to overall reduction of variance is not significant and hence were not selected. The importance of the selected predictors and their physical relationship with the predictand are discussed below.

The predictors that are selected in the POP equation and the variance explained by them are given in Table 3.

- The first predictor is the relative humidity at 850 hPa level, which contributes positively to the occurrence of precipitation. This indicates that the higher values of humidity are favourable for the formation of clouds and hence precipitation.
- The second predictor is the zonal component of wind at the surface which contributes negatively to the predictand. The zonal component of wind becomes negative with easterlies and favours the occurrence of precipitation. The easterly wind at the surface also indicates that the axis of the monsoon trough lies south of Delhi. Further, the easterly wind during this season is usually associated with higher moisture content.
- The third predictor is the wind speed at the 700 hPa level. This predictor is negatively correlated with the predictand such that the lower values of this predictor are favourable for the occurrence of precipitation. In the monsoon season, the 700 hPa level is the transition level between the lower and middle troposphere. Hence, a stronger wind speed at this level may inhibit the growth of clouds and hence may not favour precipitation.
- The fourth predictor is the thickness of the 500 and 700 hPa layer, and this predictor positively con-

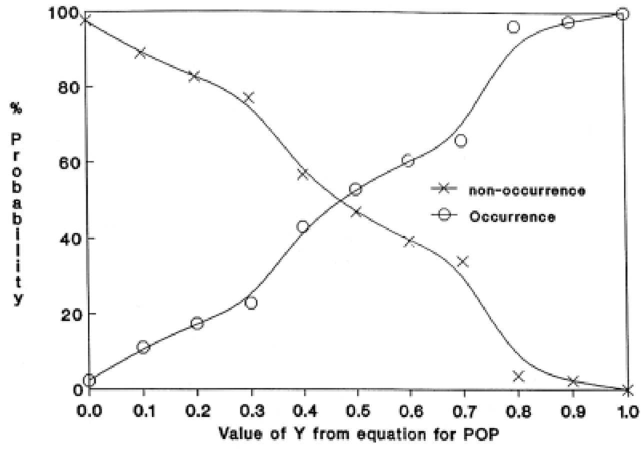

Figure 4. *Probability of occurrence/non-occurrence of precipitation events at Delhi using regression method.*

tributes to the predictand. Higher thickness values indicate warmer temperature in the layer and hence are favourable for convection and thus indicative of precipitation.

- The fifth and seventh predictors are the 24-hour changes in dewpoint depression (DPD) at 0530 and 1430 IST at surface level, and both contribute negatively to the predictand. Lower positive values or negative values of these predictors are favourable for the occurrence of precipitation. Negative values indicate an increase in the moisture content at surface level and hence are favourable for formation of clouds and precipitation.
- The sixth predictor is the wind direction at the 300 hPa level. It contributes negatively to the predictand. Therefore lower values are favourable for the occurrence of precipitation. The lower values of direction indicate the presence of north-easterly to easterly winds.
- The eighth predictor is the dry bulb temperature at 0830 IST which contributes positively to the predictand.
- The final predictor is the dewpoint temperature at 0530 IST at the surface. It also contributes positively to the predictand. The higher values of this parameter indicate higher moisture content and hence are favourable for the occurrence of precipitation.

The conditions that favour the formation of clouds and hence are indicative of the occurrence of precipitation events are as follows: high values of warm and moist air close to the ground as indicated by the predictors at the surface level, high values of moisture as indicated by the relative humidity at 850 hPa level, and warm air in the layer between 500 and 700 hPa.

It is seen that all the predictors selected in the formulation of the POP model are significant and physically related to the precipitation process.

5. Methodology for forecasting quantity of precipitation

The quantity of precipitation at a location is highly variable and difficult to predict because it depends on the type of synoptic system and the associated clouds. The precipitation at Delhi during the monsoon season is mainly convective in nature and hence the quantity of precipitation would depend upon the distribution of convective clouds around the station as well as the type of synoptic system that is causing the precipitation. When the monsoon trough is located close to the station, the precipitation is mainly in the form of thundershowers or showers (Rao, 1976). Therefore, precipitation is rather localised and patchy, and the amount may differ significantly between nearby locations. Due to the high spatial and temporal variability of the

Table 3. *Equation, predictors, variance explained and correlation coefficient for forecasting POP*

POP= – 4.787500 + 0.00591*A1 – 0.00863*A2 – 0.00910*A3 + 0.00177*A4 – 0.01434*A5 – 0.00052*A6 – 0.01528*A7 + 0.01230*A8 + 0.00867*A9

	Name of predictor	Level (hPa)	Time (IST)	Variance explained	Cumulative variance explained	Correlation coefficient
A1	Relative humidity	850	0530	17.5	17.5	+0.42
A2	Zonal component of wind	Surface	0530	2.3	19.8	–0.28
A3	Wind speed	700	0530	2.3	22.1	–0.19
A4	Thickness	500–700	0530	1.5	23.6	+0.23
A5	24-hr change in DPD	Surface	1430	1.4	25.0	–0.20
A6	Wind direction	300	0530	0.8	25.8	–0.27
A7	24-hr change in DPD	Surface	0530	0.7	26.5	–0.12
A8	Drybulb temperature	Surface	0830	0.5	27.0	+0.26
A9	Dewpoint temperature	Surface	0530	0.5	27.5	+0.29

Multiple correlation coefficient = 0.52

monsoon rainfall, a four-group classification of rainfall amount was used (see Table 2). The classification has been defined after carefully examining the distribution of rainfall amount at Delhi.

In the development of the dynamical–statistical model for forecasting the quantity of precipitation, multiple discriminant analysis (MDA) was used. The MDA procedure yields (G–1) discriminant functions for the G groups which are used to classify an event. The details can be found in Miller (1962). Some of the studies where MDA has been used for forecasting the precipitation amount can be found in Klein (1978) and Wilson (1982).

The QPF model was initiated at 0830 IST only when the POP model predicted the occurrence of precipitation. The QPF model gives the forecast of the most likely group to which the 24-hour accumulated rainfall belongs. For the development of the QPF model, the predictors considered are the same as those selected in the POP model. Since there are four groups in the present study, the MDA procedure yielded three discriminant functions of the form:

$$z_g = w_1x_1 + w_2x_2 + \dots\dots\dots + w_mx_m \qquad (2)$$

where z_g are the discriminant scores (functions), w_i are the discriminant weights (coefficients) and x_i are the independent variables. The mean values of the predictors are given in Table 4 and the discriminant functions are given in Table 5. The interpretation of the discriminant weights involve examination of the sign and the magnitude of the weights. Independent variables with relatively larger weights contribute more to the discriminating power of the function than the smaller ones. Thus, when the sign is ignored, each weight represents the relative contribution of its associated variable to that discriminant function. The sign merely denotes that the variables make either a positive or a negative contribution.

The models are evaluated using developmental data as well as independent data sets.

6. Performance of the POP model

The performance of the regression model for forecasting POP over Delhi has been evaluated using developmental data from the months of JAS for the six-year period 1985–90 and independent data from JAS for the years of 1994 and 1995. The probabilistic forecast derived from the prediction equation was converted into categorical occurrence/non-occurrence forecast using the cut-off value of 0.45. For the purpose of verification of the categorical forecasts, a 2×2 contingency table was prepared and the verification parameters and skill scores evaluated as defined in Appendix A.

6.1. Performance of the POP model with developmental data

The verification of the POP model with the developmental data indicated that the model could predict the occurrence and non-occurrence cases satisfactorily.

The performance of the model in July 1987 produced interesting results. In this year, the onset of the monsoon was considerably delayed and the monthly rainfall frequency and amount for July 1987 was below normal. Precipitation events were observed on only 7 days against the normal of 15 days. Further, the precipitation events were generally associated with thundery activity, mostly during the period 0200–0900 IST or 1200–1800 IST. The POP model could predict 5 out of 7 occurrence events and 21 out of 24 days of non-occurrence events.

On very few occasions, with the developmental data set, the model failed to forecast the observed event. Such cases were analysed to try to establish the causes of the failure. For example, on 11 August 1989, when the forecast was incorrect (forecast non-occurrence, observed occurrence), the value of relative humidity at 850 hPa was found to be lower than on previous and subsequent days. The values of relative humidity were 72.9, 48.2 and 63.4% on 10, 11 and 12 August 1989 respectively. The large decrease in the value of this parameter alone resulted in the wrong forecast on that day.

The skill scores and other verification measures of the POP model with the developmental data are presented in Table 6. The table shows that the model could predict 74 % of the occurrence events (POD = 0.74) and 72% of the non-occurrence events accurately ($C\text{-}NON$ = 0.72). The bias is 1.14, which is very close to the perfect bias of 1.0. The regression model for POP yielded skill scores varying between 0.46 and 0.53 (CSI = 0.53, TSS = 0.47 and HSS = 0.46). The overall percentage of correct forecasts is 73%. The correlation coefficient between observed and predicted events is 0.50.

6.2. Performance of the POP model with independent data

The performance of the model for forecasting probability of precipitation is verified with data from JAS 1994 and JAS 1995. It is seen from the results that the model could predict the observed events very well. The individual as well as continuous occurrences, such as 3 July 1994, 5 August 1994, 12–31 July 1994, 28–31 August 1994 and 17–19 July 1995, are predicted accurately. Similarly, the model also predicts the observed non-occurrence events correctly. For example, 4 August 1994, 16–19 and 21–30 September 1994, and 19–30 September 1995 are the days when the model

Table 4. *Mean values of the predictors in the four groups of the QPF model*

Predictor	Level (hPa)	Time (IST)	Group I	Group II	Group III	Group IV
Relative humidity	850	0530	79.20	78.05	82.87	83.93
Zonal component of wind	Surface	0530	–0.676	–0.214	0.000	0.289
Wind speed	700	0530	11.00	10.50	9.24	10.67
Thickness	500–700	0530	2510.44	2515.14	2514.97	2520.67
24-hr change in DPD	Surface	1430	–0.33	–0.68	–1.94	–1.54
Wind direction	300	0530	169.84	162.85	167.40	182.96
24-hr change in DPD	Surface	0530	–0.35	–0.27	–0.48	–0.83
Drybulb temperature	Surface	0830	24.87	25.20	25.13	25.67
Dewpoint temperature	Surface	0530	23.12	22.33	22.68	23.68

Table 5. *Coefficients of discriminant functions in the QPF model*

Predictor	Level (hPa)	Time (IST)	Function 1	Function 2	Function 3
Relative humidity	850	0530	+0.04402	+0.17708	–0.17907
Zonal component of wind	Surface	0530	–0.49947	–0.33482	+0.30658
Wind speed	700	0530	–0.33748	–0.34833	+0.21352
Thickness	500–700	0530	–0.00123	+0.00273	–0.00724
24-hr change in DPD	Surface	1430	+0.71590	+0.83971	–0.86981
Wind direction	300	0530	+0.00084	–0.00351	+0.00806
24-hr change in DPD	Surface	0530	–0.07268	–0.16725	+0.25498
Drybulb temperature	Surface	0830	–0.16256	–0.04528	+0.03822
Dewpoint temperature	Surface	0530	+0.30079	+0.00652	+0.07203

Table 6. *Verification measures for the POP model*

Measure	Dependent data (JAS 1985–90)	Independent data (JAS 1994)	Independent data (JAS 1995)
Probability of Detection (POD)	0.74	0.82	0.77
False Alarm Rate (FAR)	0.35	0.26	0.36
Miss Rate (MR)	0.26	0.18	0.23
Correct Non-occurrence (C-NON)	0.72	0.58	0.56
Critical Success Index (CSI)	0.53	0.58	0.54
True Skill Score (TSS)	0.47	0.40	0.32
Heidke Skill Score (HSS)	0.46	0.42	0.36
Bias (BIAS)	1.14	1.13	1.19
Percent Correct (PC)	74	73	67
Correlation Coefficient (CC)	0.46	0.43	0.32

accurately predicted the observed non-occurrence events.

The observed event of continuous occurrence of precipitation between 12 and 31 July 1994 is accurately predicted by the model, except on two days, 13 and 21 July 1994, when the model incorrectly forecast non-occurrence.

However, it was noticed during the analysis that there were a few occasions when the forecasts were incorrect: for example, 13 July 1994 (forecast non-occurrence, observed occurrence) and 24 July 1995 (forecast occurrence, observed non-occurrence). On 13 July 1994, the relative humidity at 850 hPa (52.4%), which is the first predictor to enter the model, was lower than on 12 July 1994 (70%) and 14 July 1994 (65%). The large reduction of this parameter alone caused the deviation in the forecast on 13 July 1994.

The skill scores and the other verification measures of the POP model with JAS 1994 and JAS 1995 data sets can be seen in Table 6. Most of the verification measures in the dependent and independent data sets are consistent and the performance of the model is good. However, the correct non-occurrence (C-NON) score in the independent data set is about 0.56 against the dependent sample value of 0.72.

The analysis further indicated that the few incorrect forecasts could mainly be attributed to the large increase/decrease in the values of the DPD at the surface level or the meridional component of wind at the 850 hPa level. Such departures in the values of the predictors were found to be mainly due to the occurrence of thunderstorms or continuous precipitation.

From the analysis of the skill scores it is apparent that the POP model is able to predict the occurrence/non-occurrence of precipitation events satisfactorily. However, the non-occurrence of precipitation events are slightly under-predicted, particularly in JAS 1994. This could possibly be due to the fact that the months of July and August 1994 had 23 and 21 days of precipitation respectively compared with the normal 15 and 13 days respectively. Moreover, the precipitation events occurred continuously between 12 and 31 July 1994 under favourable synoptic conditions.

7. Performance of the QPF model

The model for forecasting quantity of precipitation was developed using multiple discriminant analysis (MDA). The three discriminant functions were evaluated with the developmental as well as the independent data sets. The skill scores and the other verification measures were calculated using a 4×4 contingency table (see Appendix B). A set of observations (i.e. the nine predictors) was assigned to one of the four groups using the sum of squared distance principle. That is, an observation y is assigned to group g if:

$$\sum_{m=1}^{M}[d_m(y-\bar{x}_g)]^2 \le \sum_{m=1}^{M}[d_m(y-\bar{x}_h)]^2 \text{ for all } h \ne g \quad (3)$$

where d_m are the discriminant functions (in our case $M = 3$), y is the set of observations of the predictors (x_1 .... x_9) and $\bar{x}_g$ is the vector of mean values of the predictor variables (Table 4) in the four groups.

7.1. Performance of the QPF Model with developmental data

The analysis of the individual cases of amount of precipitation in the four groups indicated that the QPF model is able to classify the observed rainfall amount into the correct groups satisfactorily.

The skill scores and other verification parameters of the model with the developmental data are presented in Table 7. The table shows that about 47% of the rainfall amounts are correctly grouped. The bias exhibited by the model in the four groups is 0.79, 0.80, 1.19 and 1.60 respectively; these results indicate that the events in Group IV are over-predicted by the model, whereas the Group I and Group II events are under-predicted. The model yielded a Heidke Skill Score of 0.28. The Critical Success Index (CSI) varied between 0.21 and 0.39 for the groups.

Table 7. *Contingency table and skill scores of the QPF model with the developmental data set (JAS 1985–90)*

Observed	Forecast				
	I	II	III	IV	Total
I	50	13	18	18	99
II	12	25	9	14	60
III	7	8	16	11	42
IV	9	2	7	23	41
Total	78	48	50	66	242

Measure	Group			
	I	II	III	IV
Bias	0.79	0.80	1.19	1.60
Critical Success Index	0.39	0.30	0.21	0.27
Percentage Correct = 47%		Heidke Skill Score = 0.282		

7.2. Performance of the QPF Model with independent data

The frequency of cases of 24-hour rainfall amounts as observed in the independent data sets in the four groups (as defined in section 3) for each of the groups is presented in Figure 5. This shows that 61% of the rainfall amounts in JAS 1994 and 56% in JAS 1995 is

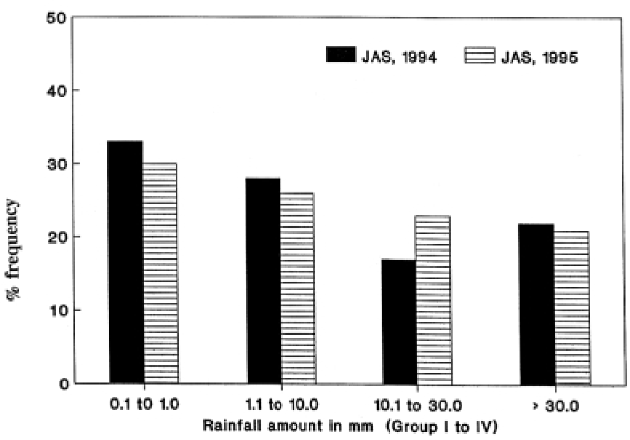

Figure 5. *Percentage frequency of rainfall amount in four groups, JAS 1994 and JAS 1995.*

within the first two groups. The QPF model was verified for its performance with the data sets for the period JAS 1994 and JAS 1995.

The analysis of the results indicated that the model is able to classify the observed rainfall amounts into the respective groups with reasonable accuracy.

The skill score for the JAS 1994 and JAS 1995 data sets is presented in Table 8. The overall percentage of correct forecasts was 44.5%. The bias of the model in the four groups is 1.59, 0.30, 1.05 and 1.04 respectively. This indicates that Group II events are under-predicted while Group I events are over-predicted. Thus the events belonging to Groups I and II show greater misclassification. One of the possible reasons is that JAS 1994 had a significantly higher number of days of precipitation. The Heidke Skill Score is 0.24 and the Threat Score varies between 0.17 to 0.38.

The verification of the QPF model with the developmental and independent data sets indicated that the model could classify the observed events satisfactorily. The skill scores indicated that there is a positive skill in the prediction of the model.

8. Conclusions

From the results of this study the following broad conclusions can be drawn.

- The equation for POP provides satisfactory results in forecasting categorically occurrence/non-occurrence events of precipitation during the next 24 hours. The model exhibited positive skill scores consistently with both the developmental as well as independent data sets.
- The model for forecasting the quantity of precipitation by classification into groups performed satisfactorily. The Group I events are slightly over-predicted while Group II events are under-predicted.
- Based on the percentage of correct forecasts, the prediction of occurrence/non-occurrence of precipitation events by the POP model is considerably higher than the prediction of quantity by the QPF model.
- The POP and QPF models are developed for a place 15 km to the north-east of Delhi. Due to the highly convective nature of precipitation with large spatial and temporal variability, the models may need minor modification for application at different locations even within Delhi. The quantitative precipitation is highly variable in space and time and the associated atmospheric conditions change very rapidly. It is possible that the model output statistics (MOS) approach to the problem may improve the QPF when sufficient data from a numerical model become available.

Table 8. *Same as Table 7 but for independent data set for JAS 1994 and JAS 1995*

Observed	Forecast				
	I	II	III	IV	Total
I	23	0	4	5	32
II	9	8	7	3	27
III	9	0	6	5	20
IV	10	0	4	8	22
Total	51	8	21	21	101

Measure	Group			
	I	II	III	IV
Bias	1.59	0.30	1.05	1.04
Critical Success Index	0.38	0.30	0.17	0.23
Percentage Correct = 44.5%		Heidke Skill Score = 0.24		

Acknowledgement

The authors wish to thank the Directorate of Meteorology, Indian Air Force, and in particular its Director, Air Vice Marshal K. C. Varma, for providing the necessary data and many stimulating discussions. The authors also acknowledge the help of the India Meteorological Department in providing valuable data sets for accomplishing this study. The work was partly supported by the Aeronautical Research & Development Board, the Defence Research and Development Organisation and the Department of Science and Technology, Govt. of India at the Centre for Atmospheric Sciences, IIT, Delhi.

Appendix A. Verification measures used for forecast evaluation

Observed	Forecast	
	Yes	No
Yes	*A*	*B*
No	*C*	*D*

The values in the contingency table are defined as follows.

- When an event is predicted to occur (forecast occurrence) and in reality it does occur (observed occurrence) then it is classified as *A*, otherwise (observed non-occurrence) it is classified as C.
- When an event is predicted not to occur (forecast non-occurrence) and in reality it does occur (observed occurrence) then it is classified as *B*, otherwise (observed non-occurrence) it is classified as *D*.

The total number of forecasts generated is $A+B+C+D$.

Probability of Detection (POD)

$$POD = \frac{A}{A+B}$$

False Alarm Rate (FAR)

$$FAR = \frac{C}{C+A}$$

Miss Rate (MR)

$$MR = \frac{B}{B+A}$$

Correct Non-occurrence (C-NON)

$$C-NON = \frac{D}{D+C}$$

Critical Success Index (CSI)

$$CSI = \frac{A}{A+B+C}$$

True Skill Score (TSS)

$$TSS = \frac{A}{A+B} + \frac{D}{D+C} - 1$$

Heidke Skill Score (HSS)

$$HSS = \frac{2(AD - BC)}{B^2 + C^2 + 2AD + (B+C)(A+D)}$$

Bias (BIAS)

$$BIAS = \frac{A+C}{A+B}$$

Percentage Correct (PC)

$$PC = \frac{A+D}{A+B+C+D} 100\%$$

Appendix B. Categorical verification of forecasts (four category events)

Observed	Forecast				
	I	II	III	IV	Total
I	*a*	*b*	*c*	*d*	*J*
II	*e*	*f*	*g*	*h*	*K*
III	*i*	*j*	*k*	*l*	*L*
IV	*m*	*n*	*o*	*p*	*M*
Total	*N*	*O*	*P*	*Q*	*T*

The total number of observed events in category I is:

$$J = a + b + c + d$$

The total number of forecast events in category I is:

$$N = a + e + i + m$$

In the similar way O, K, P, L, Q and M are computed. Then, the total number of events are given by:

$$T = J + K + L + M = N + O + P + Q$$

The Percentage Correct (PC) and Critical Success Index (CSI) are given by:

$$PC = \frac{a+f+k+p}{T}$$

$$CSI = \frac{a}{J+N-a}, \frac{f}{K+O-f}, \frac{k}{L+P-k}, \frac{p}{M+Q-p}$$

The CSI is lowered if one tries to over-forecast a rare category to 'catch' the occurrence of this category.

The Bias provides a check on whether the categories forecast with the correct frequency. Perfect bias is 1.

$$Bias = \frac{N}{J}, \frac{O}{K}, \frac{P}{L}, \frac{Q}{M}$$

The Heidke Skill Score (HSS) is given by:

$$HSS = \frac{(a+f+k+p) - \frac{JN+KO+LP+MQ}{T}}{T - \frac{JN+KO+LP+MQ}{T}}$$

References

Carter, G. M., Dallavalle, P. J. & Glahn, H. R. (1989). Statistical forecasts based on the national meteorological centre's weather prediction system. *Wea. Forecasting*, **4:** 401–412.

De, U. S., Desai, D. S. & Bhandari, S. G. (1995). Compilation of Weather in India, June–September, 1994. *Mausam*, **46:** 343–366.

Draper, N. R. & Smith, H. (1966). *Applied Regression Analysis*. Wiley & Sons, New York, 407 pp.

Dubey, D. P. & Rajeevan, M. (1994). Application of statistical synoptic quantitative precipitation forecast for Narmada and Tapti Basin. *Mausam*, **45:** 87–89.

Glahn, H. R. & Lowry, D. A. (1969). An operational method for objectively forecasting probability of precipitation. *ESSA Tech. Memo. WBTM 27*, 24 pp.

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

Klein, W. (1978). Statistical forecast of local weather by means of Model output statistics. In *ECMWF Seminar on the Interpretation and Use of Large-Scale Numerical Weather Products*, ECMWF, reading, 186–220.

Kripalani, R. H. & Singh, S. V. (1986). Rainfall probabilities and amounts associated with monsoon depressions over India. *Mausam*, **37:** 111–116.

Kruzinga, S. (1989). Statistical interpretation of ECMWF products in Dutch weather service. In *ECMWF Seminar/Workshop on Interpretation of NWP Products*, ECMWF, Reading, 360–365.

Kumar, A. & Maini Parvinder (1996). Statistical interpretation of general circulation model: A prospect for automation of local weather forecasts in India. *Mausam*, **47:** 229–234.

Kumar, A. & Ram, L. C. (1995). Semi-quantitative forecasts for Rapti catchment by synoptic analogue method. *Mausam*, **46:** 97–99.

Lal, J., Dey, J. S. & Kapoor, K. L. (1983). Semi-quantitative precipitation forecasts for Gomti catchment by synoptic analogue method. *Mausam*, **34:** 309–312.

Miller, R. G. (1962). Statistical prediction by discriminant analysis, *Meteorological Monograph*, **25**, American Meteorological Society, 3–14.

Paegle, J. N. (1974). Prediction of precipitation probability based on 500mb flow types. *J. Appl. Meteor*ol, **13:** 213–220.

Raj, Y. E. A., Jayanta Sarkar & Ramakrishnan, B. (1996). Daily rainfalls through a quantitative precipitation forecast model over Thiruvananthapuram and Madras areas for the monsoon of 1992. *Mausam*, **47:** 349–354.

Rao, Y. P. (1976). Southwest monsoon. *Meteorological Monograph. Synoptic Meteorology*, India Meteorological Department, 367 pp.

Singh, K. M., Prasad, M. C. & Prasad, G. (1995). Semi-quantitative forecasts for river Pun-Pun by synoptic analogue method. *Mausam*, **46:** 149–154.

Sridharan, S., Vizaya Bhaskar, V. & Murty, C. N. (1994). Precipitable water vapour and rainfall over Vishakhapatnam. *Mausam*, **45:** 372–376.

Upadhyay, D. S., Apte, N. Y., Surinder Kaur & Singh S. P. (1986). A dynamical approach to quantitative precipitation forecast. *Mausam*, **37:** 369–372.

Vislocky, Robert, L. & Young, G.S. (1989). The use of perfect prog forecasts to improve model output statistics forecasts of precipitation probability. *Wea. Forecasting*, **4:** 203–209.

Wilson, L. J. (1982). Weather element prediction by discriminant analysis. In *ECMWF Seminar/Workshop on Interpretation of Numerical Weather Prediction Products*, ECMWF, Reading, 311–346.

Wilson, L. J. & Sarrazin, R. (1989). A classical-REEP short-range forecast procedure. *Wea. Forecasting*, **4:** 502–516.