Point probabilistic prediction of precipitation and quantitative precipitation forecast in Western Himalayas

A. P. DIMRI, U. C. MOHANTY* and L. S. RATHORE**

Snow and Avalanche Study Establishment, Chandigarh – 160 036, India

*Indian Institute of Technology, New Delhi – 110 016, India

**National Center for Medium Range Weather Forecasting, NOIDA, U. P.

सार – भारत के उत्तरी पश्चिमी भू–भाग में हिमालय की विभिन्न पर्वत श्रृंखलाएँ हैं। इस पूरे भू–भाग में ये पर्वत श्रृंखलाएँ प्रकृति के अनुरूप विभिन्न ऊँचाइयों की हैं। इस क्षेत्र में बनकर पश्चिम की ओर बढ़ने वाले निम्न दाब क्षेत्रों और सिनॉप्टिक मौसम तंत्रों जिन्हें पश्चिमी विक्षोभ कहा जाता है, के कारण इस क्षेत्र में शीत ऋतु के दौरान भारी वर्षा होती है। इन तंत्रों के मार्ग की संरचना में भिन्नता के कारण विभिन्न प्रकार के भूभाग निम्न स्तरीय परिसंचरण होता है। तापमान, दाब और सापेक्षिक आर्द्रता जैसे धरातलीय मौसम घटक स्थानीय स्थलाकृतियों पर अत्यंत निर्भर करते हैं। मौसम का प्रायोजित खाका तैयार करने के लिए उपरी स्तर परिसंचरण और धरातलीय मौसम संबंधों के बीच की अनिश्चितताओं को औपचारिक रूप से आँकड़ों की मात्रा में अभिव्यक्त करने का प्रयास किया गया है। वर्षा के संभावित पूर्वानुमान के साथ–साथ मात्रात्मक वर्षण पूर्वानुमान निदर्श देने के लिए सही पूर्वानुमान पद्धति (पी. पी. एम.) का उपयोग किया गया है। इसका उद्देश्य, बृहत हिमालय श्रृंखला के अंतर्गत आने वाले विशेष स्थलों में से एक स्थान सोनमार्ग के संबंध में आगामी 24 घंटों का मौसम पूर्वानुमान देना है। राष्ट्रीय पर्यावरण पूर्वानुमान केन्द्र (एन. सी. ई. पी.) संयुक्त राज्य अमरीका से प्राप्त विश्लेषण आँकड़ों और भारत के भारत मौसम विज्ञान विभाग (आई. एम. डी.) के तीन केन्द्रों के स्थानीय आँकड़ों का उपयोग निदर्श को विकसित करने के लिए किया गया है। विकासात्मक निदर्श के लिए 12 वर्षों (1984–96) की अवधि के दिसंबर, जनवरी, फरवरी और मार्च के महीनों के आँकड़ों को लिया गया है। जबकि भारत मौसम विज्ञान विभाग के अंतर्गत (*i*) एन. सी. ई. पी. विश्लेषण, (*ii*) राष्ट्रीय मध्यावधि मौसम पूर्वानुमान केंद्र विश्लेषण और (*iii*) एन. सी. एम आर. डब्ल्यू. एफ. के वर्ष 1996–97 टी. 80 दिन के दिसंबर, जनवरी, फरवरी और मार्च के महीना के आँकड़ें पूर्वानुमान की जाँच के कार्य के लिए, लिए गए हैं। इनमें प्राप्त हुए परिणामों से यह पता चलता है कि विकासात्मक समूह के लिए वर्षा की संभावना (पी. ओ. पी.) वाला निदर्श 90.4% परिशुद्धता के साथ पूर्वानुमान दे सकता है जबकि इससे जाँच के मामलों में सबसे अच्छा पूर्वानुमान 86.8% तक की परिशुद्धता के साथ दिया जा सकता है। विकासात्मक समूह में क्यू. पी. एफ. निदर्श द्वारा सही पूर्वानुमान 45.0% तक दिया गया है जबकि जाँच के नमूनों में अधिकतम 54.2 % तक की परिशुद्धता निष्पादित की गई।

ABSTRACT. Northwest India is comprised of various Himalayan mountain ranges. These ranges are having different altitude and orientations all along this region. During winter season enormous amount of precipitation is received in this region due to westward moving low pressure synoptic weather systems called Western Disturbances (WD). Variable terrain gives rise to low level circulation during the passage of these systems. Surface weather elements like temperature, pressure and relative humidity are highly dependent on local topography. To draw projected weather, uncertainties involved in the relationship between upper level circulation and surface weather is tried to be formally expressed in statistical terms. Perfect Prognostic Method (PPM) is used to forecast Probability of Precipitation (PoP) occurrence, followed by Quantitative Precipitation Forecast (QPF) model. The objective is to give projected weather in lead time of 24 hour at one of the specific sites, Sonamarg, situated in Great Himalayan range. Analysis data from the National Center for Environmental Prediction (NCEP), US and station data of three stations from India Meteorological Department (IMD), India is used for development of model. Data of December, January, February and March (DJFM) months for 12 year (1984-96) is taken for developmental mode. Whereas IMD data with (*i*) NCEP analysis, (*ii*) NCMRWF analysis and (*iii*) NCMRWF's T80 day 1 forecast for DJFM months for 1996-97 is considered for the verification purpose. Result shows that PoP model could predict with 90.4% accuracy for developmental set, whereas in verification cases best prediction is made with accuracy of 86.8%. In case of QPF model percentage correct forecast is made with 45.0% in developmental set, whereas maximum 54.2% accuracy is achieved in verification sample.

Key words – Prediction, Circulation, Precipitation, Skill.

1. Introduction

Extratropical cyclones (WDs) are westward moving low pressure synoptic weather systems, which originate somewhere over mid Atlantic and travel towards east over Iran, Afghanistan, Pakistan and north India. These weather systems take southern most tracks during winter season and pass over northwest India. These WDs yield

enormous amount of precipitation during winter months, *viz.*, December, January, February and March (DJFM), in the form of snow, over northwest India. Northwest India is comprised of complex mountain ranges and hence having variable terrain and complex orography. Surface weather elements like precipitation, minimum and maximum temperatures are highly dependent upon local topography and local atmospheric circulations. Though low level circulations are purely dependent upon above factors, but upper level circulations are not so dependent upon them. Moreover, uncertainties involved in the relationship between the circulation and surface weather can be formally expressed by a suitable statistical relation. Hence, a statistical relation developed between upper air circulation around the location of interest and observed values of the surface weather element at the location, by using chosen predictors and suitable statistical technique which will account for the effect of these local conditions. This indicates that statistical forecast so obtained will have improved forecast skill using numerical outputs. Basically two methods are used for statistical based forecast. These are the Perfect Prognostic Method (PPM) (Klein *et al*. 1959) and the Model Output Statistics (MOS) (Glahn and Lowry, 1972). Precipitation is one of the important weather elements that influence various activities. Prediction of Probability of Precipitation (PoP) at specific site and time is important in many areas of human and natural hazard activity. Subsequently, knowledge of expected quantity not only helps in assessing avalanche threat perception in mountainous terrain of the region, but also helps in flood, water and forest management in the region.

The site and time specific prediction of the occurrence of precipitation in northwest India is studied to estimate the threat perception due to existing avalanche danger along the highways. The existing approach is mainly based on PPM and uses numerical outputs of meteorological fields for prediction purposes. In the present study, statistical dynamical model is developed to forecast the PoP and produce a QPF. Since northwest India remains infested with moving WDs during winter season therefore DJFM months are considered to develop the deterministic models.

In the present paper, section 2 describes the data and experimental set up used in the study. Section 3 and 4 illustrates the model formulation of PoP and QPF respectively. Results pertaining to these models are described in section 5. Broad conclusions are presented in section 6.

2.   Data and experimental set up

The PoP model is initiated at 0300 UTC giving the forecast of the PoP for the following 24 hours. The QPF model is initiated at the same time only if the PoP model indicates 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 the four groups.

For the development of the model, a station in northwest India, Sonamarg (Lat. 75° 17′ 57″, Long. 34° 18′ 11″ and altitude 2745 m) was selected. The model equations are developed using surface and upper air data of DJFM months for the 12 year period (1984-96). In order to develop a multiple regression equation, a total 2454 potential predictors, consisting of surface and upper air observations plus derived parameters are utilized. During the development process data quality checks were enforced. The developed model was tested with independent data sets from DJFM for the period 1996- 97.

PPM model is developed by using analysis data obtained from the National Center for Environmental Prediction (NCEP), US and upper air data of Patiala, Jodhpur and Delhi from the India Meteorological Department (IMD), India. The NCEP analysis is global data with resolution of 2.5° Lat. × 2.5° Long. grid and at 12 vertical pressure levels, whereas IMD data is station data. Further, as being a very coarse grid, from existing complex topographic and terrain conditions' point of view, NCEP analysis data are interpolated at station points on a five concentric circles, Sonamarg as a center, with increasing radius from 0.5° to 2.5° with a 0.5° interval. These six station points, at each circle around Sonamarg are selected by starting anticlockwise from east direction with 60° intervals, as shown in Fig. 1. Due to interpolation of data, maximum atmospheric circulation at and around Sonamarg will be taken care off.

Precipitation can be treated either as a continuous or a binary predictand. If measurable precipitation is observed, the binary predictand value is set to 1; if no measurable precipitation is observed, the predictand value is set to 0. The threshold value of precipitation is taken as 0.1 cm at Sonamarg, which is the least measurable precipitation as snow. Precipitation reported on a particular day is the accumulated snow depth in 24 hour ending at reporting time, *i.e.*, 0300 UTC. Snowfall depths are classified into four groups: 0.1-12.0 cm, 12.1-24.0 cm, 24.1-48.0 cm and ≥48.1 cm. This classification is used for avalanche forecasting in India. But for rest of the work snow depths are converted into the corresponding water equivalent and then compared with the model's

TABLE 1

Equation, predictors, variance explained and correlation coefficient for forecasting PoP

$$PoP = 10.6262 - 0.4232* E3(w)_{850\,hPa}^{0000\,UTC} + 0.0402* P(TT)_{Surface}^{0000\,UTC} - 0.0379* SW2(TT)_{850\,hPa}^{+1200\,UTC} + 0.1639* E3(q)_{500\,hPa}^{0000\,UTC}$$

$$+0.0027* P(u)_{500\,hPa\sim850\,hPa}^{0000\,UTC} - 0.0005* D(dd)_{Surface}^{-1200\,UTC} +0.0152* SW2(u)_{700\,hPa}^{+1200\,UTC} - 0.0166* D(\theta)_{900\,hPa\sim Surface}^{0000\,UTC} +0.0067* P(TD)_{850\,hPa}^{-1200\,UTC}$$

Predictor	Type of data	Variance explained	Cumulative variance explained	Correlation coefficient
$E3(w)_{850\,hPa}^{0000\,UTC}$	NANA	23.9	23.9	-0.49
$P(TT)_{Surface}^{0000\,UTC}$	STN	4.5	28.4	+0.39
$SW2(TT)_{850\,hPa}^{+1200\,UTC}$	NANA	1.4	29.8	+0.02
$E3(q)_{500\,hPa}^{0000\,UTC}$	NANA	0.8	30.6	+0.38
$P(u)_{500\,hPa\sim850\,hPa}^{0000\,UTC}$	STN	0.8	31.4	+0.24
$D(dd)_{Surface}^{-1200\,UTC}$	STN	0.5	31.9	-0.24
$SW2(u)_{700\,hPa}^{+1200\,UTC}$	NANA	0.5	32.4	+0.17
$D(\theta)_{900\,hPa\sim Surface}^{0000\,UTC}$	STN	0.5	32.9	-0.08
$P(TD)_{850\,hPa}^{-1200\,UTC}$	STN	0.4	33.3	+0.23

Multiple correlation coefficient = 0.58

NANA : NCEP reanalysis data
STN : Station data

precipitation fields. It may be noted that while converting the snow depth into the water equivalent, snow density is taken into consideration by computing the standard volume, density and mass relation.

3. Forecast formulation of Probability of Precipitation (PoP)

The PoP forecasts are constructed; following multiple regression equation with a stepwise regression technique (Draper and Smith 1966) with stopping criteria is used for forecast over a 24 hour period. Nine significant predictors were selected and then subjected to the development of PPM models. An equation of following type is assumed ;

$$Y_{24} = a_0 + \sum_{i=1}^{n} a_i x_i \qquad (1)$$

where a_i's are the regression coefficient and a_0 is the regression constant, Y is the predictand value obtained by a linear combination of selected prediction x_i's.

PoP forecast for next 24 hour (Y_{24}) is taken from Eqn. (1). While developmental procedure, the value of predictand, Y, is taken as 1 if precipitation occurs and 0 if it does not.

(*i*) If the value of Y is less than 0.55, non-occurrence of precipitation is forecast (no)

(*ii*) If the value of Y is greater than or equal to 0.55, occurrence of precipitation is forecast (yes)

The predictors that are selected in the PoP model and variance explained by them are given in Table 1. Letters and numbers (prefix to the notation of the selected predictor) in the notation of the predictor represent the geographical direction toward which that station is located
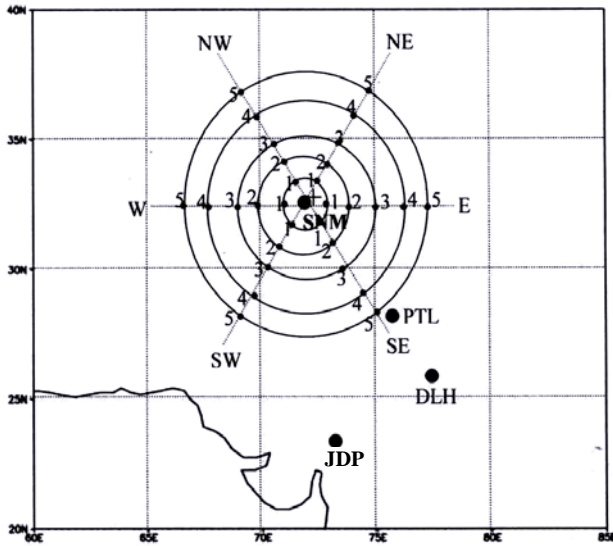
Fig. 1. The location of meteorological stations from which data have been used in this study. PTL : Patiala, DLH : Delhi, JDP : Jodhpur. The SNM : Sonamarg is the selected place of study and is indicated by +. NCEP analysis data is interpolated at locations numbered from 1 to 5 along various geographical directions and are marked by •

and the number of the circle in which that predictor belongs, respectively (Fig. 1). The superscript represents the time at which that candidate predictor is observed and the subscript represents the level at which that candidate predictor is observed or levels between which the mean/difference of that candidate predictor is computed. These are interpolated from the NCEP reanalysis data. For example, $\mathrm{E3}(w)_{850\,\mathrm{hPa}}^{0000\,\mathrm{UTC}}$ indicates the vertical component of wind (w) from the NCEP reanalysis interpolated at circle point 3 towards the east. The superscript shows that this predictor comes from 0000 UTC, and the subscript shows that it is for 850 hPa. Predictors only having letters prefixed to their notations are station data from IMD and that letter is first letter of that station name. For example, $\mathrm{P}(TT)_{\mathrm{Surface}}^{0000\,\mathrm{UTC}}$ represents the dry bulb temperature (TT) at the surface which is observed at 0000 UTC at Patiala. The cumulative variance explained, correlation coefficients and multiple correlation coefficient explained by the selected predictors at Sonamarg is also presented in Table 1.

4. Forecast formulation of Quantity Precipitation Forecast (QPF)

The precipitation at these selected sites is mainly due to the WDs. The amount of precipitation is extensively modulated due to the existing orography. Mesoscale circulation contributes immensely to define the type and amount of precipitation. Due to the high spatial and

temporal variability of the precipitation, a four group classification of snow depth is used: 0.1 cm to 12.0 cm, 12.1 cm to 24.0 cm, 24.1 cm to 48.0 cm and ≥ 48.1 cm.

Probabilistic QPF models are developed using multiple discriminant analysis (MDA). The MDA procedure yields (G-1) discriminant functions for the G groups, which are used to classify an event (Miller 1962). Wilson (1982) has used MDA for forecasting precipitation amounts.

The QPF model was initiated at 0300 UTC only if PoP model forecasts the occurrence of precipitation as yes *i.e.*, 1. This model gives probabilistic forecast of the most likely group to which the 24 hour precipitation belongs to. While developing QPF model, the predictors considered are same as selected in the PoP model. Since there are four groups in the present study, the MDA procedure yielded three discriminant functions of the form;

$$z_g = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots\ldots + w_m x_m \qquad (2)$$

where z_g are discriminant scores (functions), w_1 are the discriminant weights (coefficient) and x_1 are the independent variables.

The model for forecasting QPF was developed using MDA. The three discriminant functions were evaluated with the developmental as well as the independent data sets. A set of observations (*e.g.*, 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

$$\sum_{m=1}^{M} \left[d_m\left(y - \overline{x}_g\right)\right]^2 \leq \sum_{m=1}^{M} \left[d_m\left(y - \overline{x}_h\right)\right]^2 \text{ for all } h \neq \mathrm{g} \qquad (3)$$

where d_m are the discriminant functions (in our case $m = 3$), y is the set of observations of the predictors ($x_1,\ldots.x_9$) and x_g is the vector of mean values of the predictor variables in the four groups.

5. Result and discussion

The results of the PoP and QPF models determined with dependent/developmental data sets of DJFM 1984-96 and independent data sets of DJFM 1996-97 are presented. The performances of PoP and QPF models are evaluated by computing various statistical skill scores. Verification of categorical forecasts and the percentage of correct forecasts are also computed. Comprehensive analyses are carried out to assess model skills of the three experiments.

TABLE 2

Verification measures for the PoP model

Measure	Dependent data (DJFM 1984-96)	IMD NCEP Analysis	IMD NCMRWF Analysis	IMD NCMRWF day 1 forecast
		Independent data (DJFM 1996-97)		
Probability of detection (POD)	0.84	0.71	0.63	0.74
False alarm rate (FAR)	0.07	0.21	0.08	0.08
Miss rate (MR)	0.16	0.29	0.37	0.26
Correct non-occurrence (C-NON)	0.95	0.92	0.98	0.96
Critical success index (CSI)	0.79	0.60	0.60	0.70
True skill score (TSS)	0.79	0.63	0.61	0.70
Heidke skill score (HSS)	0.80	0.64	0.66	0.72
Bias (BIAS)	0.91	0.89	0.68	0.80
Përcentcorrect(PC)	90.36	85.12	86.78	86.78

TABLE 3

Contingency table and skill scores of the QPF model with the developmental data set of IMDINCEP (DJFM 1984-96)

Observed	I	II	III	IV	Total
		Forecast			
No rain	31	7	1	1	40
I	140	44	40	17	241
II	45	28	21	31	125
III	25	11	32	39	107
IV	8	5	10	39	62
Total	218	88	103	126	535

Measure	I	II	III	IV
		Group		
Bias	0.91	0.70	0.96	2.03
Critical success index	0.44	0.15	0.18	0.26
Percentage correct = 45.0%		Heidke skill score = 0.22		

In this section, emphasis has been given to the performance of PoP and QPF models based on the PPM concept using independent data sets for DJFM 1996-97. However, for the sake of completeness, the performance of these models with development data sets DJFM 1984-96 is also presented.

5.1. *Performance of probability of precipitation (PoP) model*

The regression model for forecasting PoP at these sites is evaluated using the developmental DJFM data for the 12 year period 1984-96 and the independent data for DJFM 1996-97. For the purpose of verification of the categorical forecasts, a 2×2 contingency table is prepared and the verification parameters and skill scores are evaluated as defined in Appendix-I (Wilks, 1995).

The skill scores and other verification measures of the PoP model with the development data is presented in Table 2. With dependent data (DJFM 1984-96), the table shows that the model could predict 84.0% of occurrence events (POD = 0.84) and 95.0% of the non-occurrence events accurately (C-NON = 0.95). The false alarm rate (FAR) is of 0.07, which illustrates that the PoP has a better POD of the non-occurrence (C-NON) of

precipitation. This may be attributed to the fact that the number of no precipitation days is generally higher compared to the number of days with precipitation. The Heidke skill score (HSS) is 0.8 for Sonamarg. The bias is 0.91, which is somewhat close to the perfect bias 1.0. The regression model for PoP yielded skill scores varying between 0.79 and 0.80 (CSI = 0.79, TSS = 0.79 and HSS = 0.80). The overall percentage of correct forecast is 90.4%. All the forecast verification indices estimated from the contingency tables clearly demonstrate that the PoP model provides very satisfactory performance regard to with the development data.

The performance of the model for forecasting PoP is investigated with independent data set (DJFM 1996-97) extensively using different predictors. These data sets of selected predictors are from (*i*) IMD and NCEP analysis, (*ii*) IMD and NCMRWF analysis and (*iii*) IMD and NCMRWF T80 model 1 day forecast output. The comparison indicates that, as with dependent cases, the skill of model in predicting the non-occurrence events more successfully with correct non-occurrence (C-NON) is higher than the probability of detection (POD). The FAR in all the three cases does not exceed 0.21. The Heidke skill score at Sonamarg is in the range of 0.64 to 0.72. The overall performances of all the three experiments as illustrated by percent correct are quite satisfactory with in the range of 85.0% to 86.0%.

5.2. *Performance of quantitative precipitation forecast (QPF) model*

QPF models based on discriminant analysis to predict the categorical quantity of precipitation are evaluated with development and independent data sets. The skill scores and the other verification measures are calculated using a 4 × 4 contingency table (Appendix-II; Wilks, 1995).

The skill score and other verification measure of the model with the developmental data are presented in Table 3. The critical success index (CSI) for the QPF model is higher in group 1 as compared to other groups. This illustrates the fact that QPF model can predict the precipitation amount in lower snowfall category better than in higher categories. The HSS of 0.22 is found in developmental sample of data. It is interesting to note that the overall performance of the QPF model in terms of percent correct (PC) with the developmental sample is about 45.0%. The bias exhibited by the model in the four groups is 0.91, 0.70, 0.96 and 2.03 respectively. All the forecast estimated from the contingency model shows that the QPF model is able to provide reasonable performance at Sonamarg.

TABLE 4

Contingency table and skill scores of the QPF model with the independent data \ set (DJFM 1996-97)

Observed	Forecast				Total
	I	II	III	IV	
(a) IMD/NCEP analysis data set					
No rain	6	1	0	0	7
I	8	4	1	1	14
II	1	2	1	1	5
III	1	2	0	3	6
IV	0	0	1	1	2
Total	10	8	3	6	27
Group					
Measure	I	II	III	IV	
Bias	0.71	1.6	0.5	3.0	
Critical success index	0.5	0.2	0.0	0.14	
Percentage correct = 41.2%			Heidke skill score = 0.17		
(b) IMD/NCMRWF analysis data set					
No rain	2	0	0	0	2
I	10	1	2	1	14
II	1	1	1	1	4
III	0	1	1	2	4
IV	0	0	1	1	2
Total	11	3	5	5	24
Group					
Measure	I	II	III	IV	
Bias	0.78	0.75	1.25	2.5	
Critical success index	0.67	0.17	0.13	0.17	
Percentage correct = 54.2%			Heidke skill score = 0.30		
(c) IMD/NCMRWF day 1 forecast data set					
No rain	1	1	1	0	3
I	5	1	9	4	19
II	2	1	3	1	7
III	0	1	5	3	9
IV	0	0	1	1	2
Total	7	3	18	9	37
Group					
Measure	I	II	III	IV	
Bias	0.37	0.43	2.0	4.5	
Critical success index	0.24	0.11	0.23	0.10	
Percentage correct = 32.4%			Heidke skill score = 0.11		

The performances of QPF model with the independent data from DJFM 1996-97 is evaluated using different type of predictors and are compared and are presented in Table 4. On comparison it is seen that experiment 2 exhibits the best model forecast with 54.2% correct forecast. The model bias in each group varied between 0.37 and 4.5. The high HSS is seen with IMD-NCMRWF analysis. Similarly high CSI values are obtained in this set of data. This illustrates the fact that second set of data produces better prediction of

probability of occurrence of the quantity of precipitation in respective categories than other set of data at Sonamarg. Further, bias close to 1 is presented by second experiment in all the groups as compared to experiment 1 and 3. The overall performance, therefore shows that data set of IMD and NCMRWF analysis gives better results than rest of the two.

6. Conclusion

The equation of PoP provides satisfactory results in forecasting categorical occurrence/non-occurrence events of precipitation during next 24 hour. The model for forecasting the QPF by classification into groups performed satisfactorily.

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 Sonamarg. Due to highly complex terrain/topography precipitation has high spatial and temporal variability, hence the model needs minor modification for application at different location. The QPF 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.

With the advent of various numerical models, to state the projection of future state of weather has become more common. Improvement in these forecasts can be carried out by downscaling associated errors either by dynamical methods or by statistical methods.

References

Draper, N. R. and Smith, H., 1966, "Applied Regression Analysis", Wiley and Sons, New York, p407.

Glahn, H. R. and Lowery, D. A., 1972, "The use of model output statistics (MOS) in objective weather forecasting", *J. Appl. Met.*, **11**, 1203-1211.

Klein, W., Lewis, B. M. and Enger, I., 1959, "Objective prediction of 5-day temperature during winter", *J. Met.*, **16**, 672-682.

Miller, R. G., 1962, "Statistical Prediction by Discriminant Analysis", *Meteor. Monogr. Amer. Meteor. Soc.*, **25**, 3-14.

Wilks, D.S., 1995, "Statistical Methods in the Atmospheric Sciences", Academic Press, p466.

Wilson, L. J., 1982, "Weather element prediction by discriminant analysis", Seminar/Workshop on Interpretation of NWP Products, Reading, U.K., European Centre for Medium Range Weather Forecasting, 311-364.

Appendix I

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,

1. 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.
2. 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.
3. A+B : Total number of cases of occurrence of precipitation as observed.
4. C+D : Total number of cases of non-occurrence of precipitation as observed.
5. A+B+C+D: Total number of forecasts.

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) for occurrence

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

Percentage Correct (PC)

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

For a best/perfect forecast series : B = 0 and C = 0 and hence
POD = 1, FAR = 0, MR = 0, C – NON = 0, Bias = 1, CSI = 1, TSS = 1, HSS = 1, PC = 100%

Appendix II

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

Total number of observed events in category I is:
J = a + b + c + d
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 numbers of events are :
T = J + K + L + M = N + O + P + Q
Percentage Correct (PC)

$$PC = \frac{a+f+k+p}{T} 100\%$$

Critical Success Index (CSI)

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

Heidke skill score (HSS)

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

_____