# Statistical interpretation of general circulation model: A prospect for automation of medium range local weather forecast in India

# ASHOK KUMAR and PARVINDER MAINI

Department of Science & Technology (NCMRWF), New Delhi (Received 10 August 1994, Modified 16 June 1995)

, खार — यद्यपि साधारण परिसंचरण निदर्श (जी. सी. एम.) पर्याप्त रूप से अच्छा मध्यावधि मौसम प्रवानुमान उपलब्ध कराने में समर्थ डोते हैं तथापि विशिष्ट अवस्थिति मौसम प्रवानमान प्रस्तुत करने में वे अपेक्षाकृत कम सक्षम हैं। इसका मुख्य कारण स्थानीय स्थलाकृति और अन्य लक्षणों का हन निदर्शों द्वारा घटिया धस्ततीकरण है। विशिष्ट अवस्थिति मध्यावधि स्थानीय मौसम प्रवनिमान को बेहतर बनाने के लिए .<br>जी. सी. एम. का आंकड़ा संबंधी विश्लेषण (एस. आई.) परमावश्यक है। इस तरह के वस्तुपरक पूर्वानुमान के लिए राष्ट्रीय मध्यावधि मौसम पूर्वानुमान केन्द्र (एन. सी. एम. आर. डब्ल्यू. एफ.), नई दिल्ली में प्रयास किया गया है। इसलिए विशिष्ट अवस्थिति एस. आई. निदर्श तैयार .<br>किए गए हैं और दोषरहित पृवनिमान प्राप्त किया गया है। इसको प्राप्त करने की एक तकनीक परफेक्ट प्रोग विधि (पी. पी. एम.) है। भारत में दस केन्द्रों के मानसन ऋतु (जून से अगस्त) के लिए वर्षण (मात्रात्मक, सम्भव सकारात्मक/नकारात्मक) तथा अधिकतम/न्यूनतम तापमान का पता लगाने के लिए पी. पी. एम. निदर्श विकसित किए गए हैं। एन. सी. एम. आर. डब्ल्यू. एफ. के जी. सी. एम. (आर-40) के उत्पाद तथा हुन पी. पी. एम. निदर्शों का उपयोग इस प्रकार से एस. आई. प्रवनिमान प्राप्त करने के लिए किया जाता है। इस शोष पत्र में दोषरहित प्रवनिमान प्राप्त करने के लिए एस. आई. पूर्वानमान तथा पिछली एक या वो ऋतुओं के प्रेक्षित मानों पर आधारित एक अप्रत्यक्ष विधि की व्याख्या की गई है। मानसन ऋत 1993 के दौरान एन, सी, एम, आर, डब्ल्यू, एफ, डारा 10 कृषि मौसम क्षेत्रीय इकाईयों (ए. एम, एफ, यू.) के लिए निपुणता पर्वक जारी किए गए एस. आई. तथा प्रोक्षेत अन्तिम दोषरहित प्रवनिमान के तुलनात्मक अध्ययन से पता चला है कि एस. आई. प्रवनिमान की सहायता से मध्यावधि स्थानीय मौसम प्रवानुमान स्वचल रूप से प्राप्त किया जा सकता है।

ABSTRACT. The General Circulation Models (GCM), though able to provide reasonably good medium range weather forecast, have comparatively less skill in forecasting location-specific weather. This is mainly due to the poor representation of local topography and other features in these models. Statistical interpretation (SI) of GCM is very essential in order to improve the location-specific medium range local weather forecast. An attempt has been made at the National Centre for Medium Range Weather Forecasting (NCMRWF), New Delhi to do this type of objective forecasting. Hence location-specific SI models are developed and a bias free forecast is obtained. One of the techniques for accomplishing this, is the Perfect Prog. Method (PPM). PPM models for precipitation (quantitative, probability, yes/no) and maximum/ minimum temperature are developed for monsoon season (June to August) for 10 stations in India. These PPM models and the output from the GCM (R-40) operational at NCMRWF, are then used to obtain the SI forecast. An indirect method, based upon SI forecast and observed values of previous one or two seasons, for getting bias fige forecast is explained. A comparative study of skill of bias free SI and final forecast, with the observed, issued from NCMRWF to 10 Agromet Field Units (AMFU) during monsoon season 1993, has indicated that automation of medium range local weather forecast can be achieved with the help of SI forecast.

Key words - Numerical weather prediction (NWP), General circulation model (GCM), Statistical interpretation (SI), Perfect Prog. Method (PPM), Model Output Statistics (MOS), Quantitative Precipitation (QP), Probability of Precipitation (PoP), MAX (Maximum), A. MIN (Minimum).

### 1. Introduction

Objective forecasts of local weather elements can best be obtained by using statistical methods to complement the raw output of Numerical Weather Prediction (NWP) models. When the objective forecasting is inherently statistical in nature and depends for its input on NWP models, it is said to be an interpretative system and is called as statistical interpretation (SI) of NWP model output.

Surface weather elements, like maximum/ minimum temperature and rainfall, are of great importance to the user and are highly dependent on local topographic and environmental conditions. In the NWP models it is difficult to include all the local topographic and environmental conditions of each and every point considered in the model. Problem becomes more grave in the case of General Circulation Models (GCM) because of coarse resolution.



Fig. i. Two methods of combining numerical and statistical weather forecasting in schematic form

Surface weather is manifestation of interaction between local topographic conditions and upper air circulation. Hence a statistical technique, which develops concurrent relationship between the upper air circulation and the surface weather parameter, will have inbuilt accounting capability for these local conditions. This relation can be developed for a particular surface weather element for a specific season over previous five to six years.

# 2. Methodology

### 2.1. Two methods of statistical interpretation

There are two methods of obtaining the SI forecast. The first is the Perfect Prog Method (PPM), (Klein et al. 1959) which utilizes observed historical data on observed circulation to specify local weather elements from concurrent weighted combinations of meteorological parameters. To use the derived equations for making a forecast, we apply them to the output of numerical prognostic models shown by dashed arrow in Fig. 1, (Klien and Glahn 1974). Although errors in the numerical prediction will produce corresponding errors in the statistical forecast, the latter will improve each time the former is improved. A major advantage of this method is that stable forecasting relations can be derived for individual locations and seasons from a long period of record. It can be applied even if numerical model undergoes a major change, i.e., same relation will still hold good. One of its disadvantages is that it takes no account of errors and uncertainties in the numerical model and the other disadvantage is that relation at the time of derivation may not hold good at the time of forecast. This may be due to a major change in the local environmental conditions of the station or due to the adoption of a new analysis and forecast system. Under such a situation the PPM equations will have to be developed again.

Second statistical technique is known as model output statistics (MOS), (Glahn and Lowry 1972). The predictor sample in MOS usually consists of a relatively short period of prognostic data produced by numerical models. Thus the MOS method involves archiving the output from the numerical models and matching it with observations of local weather as in Fig. 1 (Klien and Glahn 1974). Forecast equations are then derived using statistical techniques. This way the systematic errors in NWP products are removed. The local topographic and environmental conditions of a location are also automatically accounted for in the forecast system. A drawback of this technique is that a sufficient sample of model output is required in order to derive a stable relation. Hence, it cannot be applied immediately when a new model is made operational. Also if the model undergoes a major change the MOS relations will have to be developed again.

On examining the advantages and disadvantages of both the techniques, it can be concluded that although MOS is superior to PPM, it cannot be applied at NCMRWF for two reasons; (i) lack of a sufficient sample of NWP output and (ii) rrequent changes in the models, due to which the MOS equations have to be redeveloped everytime the model undergoes a major change. So, under the present circumstances PPM based SI forecast seems more appropriate.

# 2.2. Development of PPM model

For developing PPM models, atleast data for three seasons (same) of six months is required (Carter 1989). As sufficient analysis/forecast data from the operational analysis and forecast system at NCMRWF was not available, the development of PPM models for monsoon season (June-August) based upon six years (1985-90) of (TOGA)

#### TABLE 1

Meteorological parameters considered as possible predictors



analysis  $(2.5 \times 2.5)$  from the European Centre for Medium Range Weather Forecasting (ECMWF) was taken up, although it has its own limitations over the Indian region.

NCMRWF prepares three-day location specific forecast for 28 agroclimatic zones on bi-weekly basis. As SI forecast is also used for the preparation of final forecast to be disseminated to these agromet field units (AMFU), hence SI models were developed only for those stations for which past observed data (1985-90) was readily available.

Models are developed for rainfall and temperature. In these models the predictands chosen are: quantitative precipitation (QP), Probability of Precipitation (PoP) and Maximum/Minimum (MAX/ MIN) temperatures. As rainfall is highly variable, hence model is developed using cube root of quantitative precipitation. The analysis data fields at 1000, 850, 700 and 500 hPa levels for forty seven parameters are chosen for inclusion in the set of predictors (Table 1). This includes mean sea level pressure (MSLP), 1000-500 hPa precipitable water (PPW), saturation deficit (SD), thickness and rate of change of moist static energy (RMSE). These were carefully selected to include all available factors which might contribute to the surface temperatures and rainfall, laying emphasis on their individual contribution to the percentage of variance explained. Climatic value of predictand, SIN and COS of the day are also tried as predictors, but not retained as they do not contribute much to the percentage of variance.

Although for the development of PPM models three reference times, viz., 0000 and 1200 UTC of previous day and 0000 UTC of the same day were considered, but after a scrutiny of the results hence obtained, reference time at which the value of the predictor is to be considered (Fig. 2) for developing the model equations, is chosen as follows:

- $(i)$  0000 UTC of the same day for MIN temperature, on which it is attained:
- (ii) 1200 UTC of the same day for MAX temperature, on which it is attained and
- (iii) 0000 UTC of the same day, 1200 UTC of the previous day and average of the two, for 24hr accumulated rainfall (OP and PoP) on which it is reported.

This implies that 47 predictors each, are required for development of maximum/minimum temperature models, whereas for rainfall the number of predictors is thrice as much, i.e., 141  $(47 \times 3)$  predictors.

In order to get the value of a particular predictor, representative of the station, its values at nine grid points around the station were considered (Fig. 3).



Fig. 2. Reference time for predictors

(Woodcock 1984). This is achieved by employing canonical correlation (Rousseau 1982). In this technique the first canonical variate, which is the best linear combination of the values of a predictor at the nine grid points that has maximum correlation with the predictand, is taken as the value of a particular predictor at the station. These canonical variates are found for each of the predictors to obtain a new predictor set.

#### 2.3. PPM model equations

Using the stepwise selection procedure, only those predictors which explain most of the variance are selected as the final predictors. Selection of variables as predictors is terminated if the variable just selected contributes less than a critical value to the total percentage of variance explained by the predictors already selected (Woodcock 1984). In order to get a significant percentage of variance explained by the predictors selected, this critical value is taken as 1.0% for MAX/MIN temperature and PoP but for QP it is relaxed to 0.5%. These selected predictors are then used for developing the PPM model equations.

Two to three predictors for maximum temperature and three to seven predictors for minimum temperature are selected. 1000-500 hPa saturation deficit and 850 hPa temperature are important in predicting maximum temperature whereas for prediction of minimum temperature 850 hPa temperature, 500 hPa temperature and 850-500 hPa thickness play an important role.



Fig. 3. The grids considered around a station

Five to eight variables for PoP and six to twelve variables for QP are selected as predictors. Mean relative humidity and 850 hPa meridional wind component are generally selected for predicting both PoP and QP, but in the case of QP the vorticityat 850 hPa is also one of the important predictors chosen. The estimates obtained from these model equations for PoP are called the regression estimates of event probabilities (Glahn 1982).

Station

Delhi

**Section** 

#### TABLE 2

Threshold values (mm) for quantitative precipitation



The equation relating one predictand to several predictors is a simple linear form and can be written as.

$$
Y = a_0 + a_1 X_0 + a_2 X_1 + \ldots + a_n X_n \qquad (1)
$$

where  $a_i$ 's are the multiple regression coefficients.

#### 2.4. SI forecast from R-40 model output

SI forecast is obtained everyday from R-40 model output by using PPM models developed. For getting 24, 48, 72 and 96-hour forecast for MAX/ MIN temperature, QP and PoP, values of the predictors from the R-40 model output are put in the corresponding PPM equations and value of the predictand is obtained during monsoon 1993. Large bias has been found in this forecast in the sense that the predictand thus obtained may be away from the expected value, that is, it may be either overpredicted or under-predicted. Method for removing bias is developed using SI forecast from R-40 model and observed data of 1991 and 1992 monsoon season (June-August). Thus, a bias-free SI forecast is got along with the operational run of R-40 model in 1993 monsoon.

The bias in the temperature predictions is removed by using an indirect method (Glahn et al.



1991). In this method, regression equation is developed between observed and predicted values of  $Y$  starting from day one to day  $n$ , where  $n$  is fixed. The corrected value of the predictand  $(n+1)$  is hence obtained using the coefficients thus developed (say, b) and the predicted value of the  $(n+1)$ <sup>th</sup> day, *i.e.* 

$$
Y_{(n+1)\text{ corr}} = a + b + Y_{(n+1)\text{ pred}}
$$

Again the regression is developed between the observed and predicted values of Y starting from day 2 to day  $(n+1)$ , which is again used for getting the corrected value from the predicted value for  $(n+2)$ <sup>th</sup> day and so on. In this way bias-free forecast is obtained for all the days. The fixed number of days, i.e., n used for development of equation each time may be different for different cases. Optimal  $n$ for bias removal for minimum temperature is found to be 20 and for maximum temperature it is 10. Same  $n$  is used for getting operational SI forecast during 1993 monsoon.

For QP, threshold value for a rainy day (Table 2) and for PoP, value of the constant factor (Table 3) for adding to the forecast probability, have been obtained by trial and error method so that skill scores get maximised during 1991 and 1992 monsoon. YES/NO forecast from PoP is derived by using the criterion that if PoP is less than 0.5, then

TABLE 3

Constants to be added for probability of precipitation

 $48-hr$ 

 $0.0$ 

 $\sim$ 

 $24-hr$ 

 $0.0$ 

 $\sim$ 

Forecast

 $72-hr$ 

 $0.0$ 

 $\sim$ 

 $96-hr$ 

 $0<sub>0</sub>$ 

 $\sim$ 

# TABLE 4

Skill scores for rainfall

<b>Station</b>	<b>SI/FNL</b> Forecast	Skill score					
		Ratio (%)	HK	<b>RMSE</b>	Correct (%)	Usable (96)	Unusable (% )
Delhi	<b>SI</b>	76.9	0.48	33.1	73.3	3.3	23.3
	<b>FNL</b>	74.4	0.35	31.2	79.3	0.0	20.7
Ludhiana	`SI	82.1	0.56	13.8	<b>78.1</b>	6.3	15.6
	<b>FNL</b>	79.5	0.33	13.4	87.1	6.5	6.5
Hisar	SI	76.9	0.42	41.4	76.7	13.3	10.0
	<b>FNL</b>	61.5	0.05	41.2	87.5	0.0	12.5
Pantnagar	SI	61.5	0.20	17.0	66.7	4.2	29.2
	<b>FNL</b>	60.0	0.19	16.5	60.9	8.7	30.4
Udaipur	SI	74.4	0.58	15.8	65.5	10.3	24.1
	<b>FNL</b>	76.9	0.53	15.9	70.0	3.3	26.7
Anand	SI	84.6	0.57	28.2	75.8	3.0	21.2
	<b>FNL</b>	69.2	0.43	27.6	59.3	3.7	37.0
Hyderabad	SI	59.0	0.17	8.1	69.6	8.7	21.7
	<b>FNL</b>	45.5	$-0.01$	10.2	40.0	6.7	53.3
Jabalpur	SI	76.9	0.54	33.8	46.7	3.3	50.0
	<b>FNL</b>	76.9	0.51	32.5	36.7	10.0	53.3
Parbhani	SI	56.4	0.10	26.8	68.2	4.6	27.3
	<b>FNL</b>	51.5	0.06	29.4	52.9	5.9	41.2
Raipur	SI	71.8	0.43	24.2	42.9	3.6	53.6
	<b>FNL</b>	71.8	0.42	22.2	35.7	7.1	57.1

SI - Statistical interpretation forecast

FNL - Final forecast prepared by forecasters' panel

it is considered as no rain case, otherwise it is considered as rain case (Mason 1979). Rainfall forecast is obtained as a hybrid of QP and PoP forecasts after combining the two in the following ways (Tapp et al. 1985):

- (1) If  $PoP < 0.5$  and  $QP = m$ , then give  $rain = 0.0$
- (2) If PoP  $\geqslant 0.5$  and QP = m. then give  $rain = m$
- (3) If  $PoP \ge 0.5$  and  $QP = 0.0$ , then give  $rain = 0.1$

where  $m$  is the forecast value of rainfall from  $QP$ equations. In case (3) rain is taken as 0.1 mm as this is the minimum possible rainfall that can be measured.

# 3. Discussion of results

After a bias-free SI forecast is obtained, a verification study is conducted for monsoon 1993. This is done by comparing the skill of SI forecast and that of final forecast issued from NCMRWF to the 10 AMFUs during monsoon 1993.

Data, only for three days during a week for which final forecast is issued, is considered for this verification study. Maximum and minimum temperature trends are considered for the threeday forecast, which is obtained by adding the

#### TABLE 5

Skill scores for maximum temperature



Skill scores for minimum temperature



**SI** - Statistical interpretation forecast

FNL - Final forecast prepared by forecasters' panel

trend to the temperature of the day on which forecast is issued, both for SI and final forecast

A man-machine-mix approach is followed for preparation of local weather forecast. It is an integrated scheme in which forecaster's panel uses all types of information, viz., direct model output (DMO), statistical interpretation (SI), conventional interpretation of graphical output and synoptic, and then gives the final forecast.

The ratio scores and the HK scores of rainfall forecast (Table 4) obtained from SI models are higher for most of the stations and are equal in one or two stations. The percentage of correct forecast is also higher in the case of SI forecast for most of the stations.

3A-36 IMD/96

 $Un$ usable  $(%)$ 

 $27.8$ 25.6 27.8 46.2  $36.1$ 38.5  $278$ 53.9  $22.2$ 30.8  $28$ 18.0

5.6 9.1 5.6  $23.1$ 56  $24.2$ 16.7 38.5

SI - Statistical interpretation forecast

FNL - Final forecast prepared by forecasters' panel

#### TABLE 7

# Critical values used for error structure



The correlations obtained from SI forecast in the case of MAX/MIN temperatures (Tables 5 and 6) are more than that obtained by final forecast for

majority of the stations. The root mean square error (RMSE) of MAX/MIN temperatures are also lower in the case of SI forecasts than in the case of final forecasts for almost all the stations. Critical values for error structure are given in Table 7.

Verification study shows that SI forecast is having good skill and is comparable to that of final forecast issued from NCMRWF. Hence SI forecast has the scope of being used as the automatic medium range local weather forecast to be issued from NCMRWF.

#### Acknowledgements

The authors are extremely grateful to Dr. R. K. Datta (Retd.), Project Co-ordinator, NCMRWF for his valuable suggestion and guidance. Authors express their special thanks to Dr. S. K. Mishra, Director General, NCMRWF, Dr. Jagdish Bahadur and Dr. L. S. Rathore. Directors DST for their overall help and guidance. Authors are also extremely thankful to Dr. U.C. Mohanty, former Jt. Director (NCMRWF), for allowing to use the ECMWF analysis data available with him and to Dr. A.K. Bohra, Director. DST and Shri G.R. Ivenger, SSO-I. NCMRWF for their help in getting the R-40 global model output and to Shri D.N. Ram for his help in graphical verification.

#### References

Carter, G. M., 1989. Statistical forecasts based on the National Meteorological Center's numerical weather prediction system, Weather and Forecasting, 401-426.

- Glahn. H. R., 1982, Statistical Weather Forecasting, Statistical Interpretation of NWP Products, Seminar Workshop, ECMWF, 263-310.
- Glahn, H. R., Murphy, A. H., Wilson, L. J. and Jensenius, J.S., 1991, Programme on short and medium range weather prediction research (PSMP) WMO, PSMP No. 34, WMO/TD No. 421.
- Glahn, H. R. and Lowry, D. A., 1972, "The use of model output statistics (MOS) in objective weather forecasting", J. Appl. Meteor., 1203-1211.
- Klien. W. H. and Glahn, H. R., 1974, "Forecasting local weather by means of model output statistics", Bull. Amer. Meteor. Soc., 55, 10, 1217-1227.
- Klien. W. H., Lewis, B. M. and Enger, I., 1959, "Objective prediction of five-day mean temperature during winter", J. Meteor., 16, 672-682.
- Mason, L. 1979. "On reducing probability forecasts to YES/ NO forecasts". Mon. Weath. Rev., 207-211.
- Rousseau, D., 1982, Work on the statistical adaptation for local forecasts in France, Statistical Interpretation of NWP Products, Seminar/Workshop, ECMWF, 347-364.
- Tapp. R. G., Woodcock, F. and Mills, G. A., 1985, "The application of model output statistics to precipitation prediction in Australia", Mon. Weath. Rev., 114,  $50 - 61.$
- Woodcock, F., 1984, "Australian experimental model output statistics forecasts of daily maximum and minimum temperature", Mon. Weath. Rev., 112, 2112-2121.