

Seasonal forecasts of Indian summer monsoon rainfall using local polynomial based non-parametric regression model

ANIL KUMAR ROHILLA, M. RAJEEVAN and D. S. PAI

National Climate Center, India Meteorological Department, Pune – 411 005, India

(Received 10 January 2006, Modified 11 June 2007)

e mails : anilkumar@imd pune.gov.in, anilrohilla@gmail.com

सार – इस शोध-पत्र में भारत में दक्षिण पश्चिमी मानसून वर्षा (जून से सितंबर) और उत्तर पश्चिमी भारत में ग्रीष्म मानसून वर्षा का पूर्वानुमान करने के लिए नए सांख्यिकीय निदर्शों के विवरणों की चर्चा की गई है। ये निदर्श स्थानीय बहुपदीय के आधार पर अप्राचलिक समाश्रयण पद्धति के आधार पर तैयार किए गए हैं। प्रागुक्तियों के दो समुहों (क्रमः 4 और 5 प्रागुक्तियों वाले एस. ई. टी. - I और एस. ई. टी. - II) को क्रमः अप्रैल और जून के उत्तरार्द्ध में पूर्वानुमान देने के लिए दो विभिन्न निदर्शों को विकसित करने के लिए चुना गया है। उत्तर पश्चिमी भारत में मानसून वर्षा के लिए निदर्श विकसित करने हेतु अन्य प्रागुक्त एस. ई. टी. (एस. ई. टी. - III) का चयन किया गया है। प्रागुक्त आँकड़ा समुहों के प्रमुख घटक का विप्लेषण किया गया तथा निदर्श के विकास के लिए पहले दो प्रमुख घटकों को चुना गया। निदर्श को विकसित करने के लिए 1977 – 2005 की अवधि के आँकड़ों का उपयोग किया गया है तथा निदर्श के कौशल का मूल्यांकन करने के लिए जैकनाइफ पद्धति को उपयोग किया गया। आई. एस. एम. आर. का पूर्वानुमान करने के लिए दोनों निदर्शों में उपयोगी कौशल पाए गए हैं तथा विषुद्धतः जल विज्ञान पर आधारित निदर्शों की अपेक्षा ये निदर्श अच्छे पाए गए हैं। अप्रैल और जून निदर्शों की विचाराधीन अवधि में तीन प्रकार के पूर्वानुमानों के लिए हीट स्कोर क्रमः 0.65 और 0.66 है। केवल जलवायु विज्ञान पर आधारित निदर्श के एल. पी. ए. से 10.0 प्रतिषत की तुलना में दीर्घ अवधि औसत से विचाराधीन अवधि में इन निदर्शों की मूल माध्य वर्ग त्रुटि (आर.एम.एस.ई.) क्रमः 5.99 और 6.0 प्रतिषत है। केवल जलवायु विज्ञान पर आधारित निदर्श के एल.पी.ए. से 18.5 प्रतिषत की तुलना में एल. पी. ए. से स्वतंत्र अवधि के दौरान उत्तर पश्चिमी भारत निदर्श का आर. एम. एस. ई. 11.5 प्रतिषत है। घटनाक्रम की इस अवधि में उत्तर पश्चिमी भारत में तीन श्रेणी पूर्वानुमानों का हीट स्कोर 0.55 है।

ABSTRACT. In this paper, details of new statistical models for forecasting southwest monsoon (June-September) rainfall over India (ISMR) and for northwest India summer monsoon rainfall (NWISMR) are discussed. These models are based on the local polynomial based non-parametric regression method. Two predictor sets (SET-I & SET-II consisting of 4 and 5 predictors respectively) were selected for developing two separate models for making predictions in April and late June respectively. Another predictor set (SET-III) was selected for developing model for monsoon rainfall over NW India (NWISMR). Principle Component Analysis (PCA) of predictor data set was done and the first two principal components were selected for model development. Data for the period 1977-2005 have been used for developing the model and the Jackknife method was used to assess the skill of the model. Both the models showed useful skill in predicting ISMR and showed better performance than the model based on pure climatology. The Hit scores for the three category forecasts during the verification period by April and June models are 0.65 and 0.66 respectively. Root Mean Square Error (RMSE) of these models during the verification period is 5.99 and 6.0% respectively from the Long Period Average (LPA) as against 10.0% from the LPA of the model based on climatology alone. RMSE of the Northwest India model during the independent period is 11.5% from LPA as against 18.5% from the LPA of the model based on the climatology alone. Hit score for the three category forecast for NW India during the verification period is 0.55.

Key words – Southwest monsoon, Long range forecasting, Monsoon rainfall, Non-parametric method, Regression, Cross validation.

1. Introduction

Accurate seasonal forecast of Indian summer monsoon (June-September) rainfall (ISMR) is an essential ingredient in the water resource planning, management for reservoir operations, agricultural activities and flood emergency responses. The India Meteorological Department (IMD) has been issuing objective long range forecasts of the south-west monsoon rainfall for more than

80 years using statistical methods. The first objective model was based on statistical correlations, which resulted from the extensive and pioneering work of Sir Gilbert Walker (Walker 1923 & 1924). Since then, IMD's operational long range forecasting system has undergone many changes in its approach and scope over the years. During the period 1988-2002, IMD's operational forecasts for the seasonal ISMR for the country as a whole were based on the 16 parameter power regression and

parametric models (Gowariker *et al.* 1989 & 1991). Following the failure of operational forecast in 2002, a critical evaluation of the 16-parameter power regression and parametric models was made and in 2003, a new two-stage forecasting strategy was adopted with a provision for forecast update in June (Rajeevan *et al.*, 2004). According to this new strategy, forecasts for the seasonal ISMR as a whole are issued in two stages. The first stage forecast is issued in mid April using 8 parameter power regression and probabilistic models and an update or second stage forecast using 10 parameter power regression and probabilistic model is issued by the end of June along with separate forecasts for four homogeneous regions (Northwest India, Central India, South Peninsula, Northeast India) over India.

In view of the failures of operational forecast during the two recent drought years (2002 & 2004), efforts were made to develop the models with improved skill to further support the forecasting strategy (Rajeevan *et al.*, 2006). As a part of these efforts, different statistical techniques were also tested for the development of forecast models. Many new methods of model development and predictor selection were adopted. The work presented in this study is the result of such efforts. Here, we report the results of model development based on a non-parametric method, called local regression. This method was adopted and tested for forecasting ISMR and northwest India summer monsoon rainfall (NWISMR). In the section 2, we discuss the data used for this study and in the section 3 we discuss predictors used for this study. Methodology of the model development is discussed in section 4. In section 5, results are discussed and the conclusions are presented in section 6.

2. Data

The main data set used was the monthly NOAA Extended Reconstructed Global Sea Surface Temperature version 2 (ERSST.v2) data at $2^\circ \times 2^\circ$ latitude \times longitude grid (Smith and Reynolds 2004). This data set was produced based on the latest version of the Comprehensive Ocean Atmosphere Data Set (COADS) release 2 observations (Woodruff *et al.*, 1998). The monthly ERSST data are available from 1854 onwards. In this study, we have used the ERSST.v2 data for the period from January 1958 to May 2005.

The ISMR data and NWISMR data are taken from the data records of the India Meteorological Department, Pune. The ISMR series used was based on the seasonal (June-September) monsoon rainfall data of all the 36 meteorological sub-divisions of India. The seasonal ISMR over the country as a whole was calculated as the area weighted average of seasonal rainfall of all 36 sub-

divisions. The seasonal ISMR was expressed as the percentage departure from the Long Period Average (LPA), which is equal to 88 cm. Coefficient of Variance (CV) for ISMR, is 10%. When the ISMR during a year is $>10\%$ ($<10\%$) of LPA, the year is termed an excess (deficient) monsoon year. All other years are termed as normal monsoon years.

The NWISMR series used was based on the seasonal (June-September) monsoon rainfall data of 9 meteorological sub-divisions of India. For this purpose the NW India is considered to consist of following 9 sub-divisions, Jammu and Kashmir, Himachal Pradesh, Utranchal, West UP, East UP, Haryana, Punjab, East Rajasthan, and West Rajasthan. The seasonal NWISMR was calculated as area weighted average of seasonal rainfall of these 9 sub-divisions. The seasonal NWISMR was expressed as percentage departure from LPA, which is equal to 61 cm. CV for NWISMR, is 18.5%. When NWISMR during a year is $>18.5\%$ ($<18.5\%$) of LPA, the year is termed as excess (deficient) monsoon year.

Another data set used in this study was the monthly reanalysis data of surface sea level pressure and 850 hPa zonal wind of National Centers for Environmental Prediction (NCEP) (<http://www.cdc.noaa.gov/>.) The spatial resolution of the data is $2.5^\circ \times 2.5^\circ$, latitude \times longitude grid. In addition, we have used the monthly land surface air temperature (LST) data from 5 Europe stations obtained from the publication "Monthly Climate Data for the world" published by NCAR. These stations are: Orland, Oslo/Gendermon, Ostursund/Frosen, Karlstad and De Bilt. Monthly mean Nino 3.4 index was obtained from the web site (<http://www.cpc.ncep.noaa.gov/>) of Climate Prediction Centre, NOAA). All the data sets were used for the period 1958-2005. In addition, we have used the monthly mean Warm Water Volume (WWV) data over Pacific (McPhaden 2003, Meinen and McPhaden 2000, Rajeevan and McPhaden 2004) for the period 1958-2005. The WWV data available on real time at <http://www.pmel.noaa.gov/tao/elnino/wwv/> were based on the upper ocean temperature field analysis.

3. Predictors used in the study

As described in the introduction, the main objective of the paper was to develop new prediction models for the long range forecasting of seasonal ISMR. This demands development of two sets of models; one set (useful for the first stage forecast issued by mid April) and the second set (useful for the second stage or update forecast issued by end of June).

For the first stage forecast model, a predictor set requiring data up to March (SET-I) was used and for the

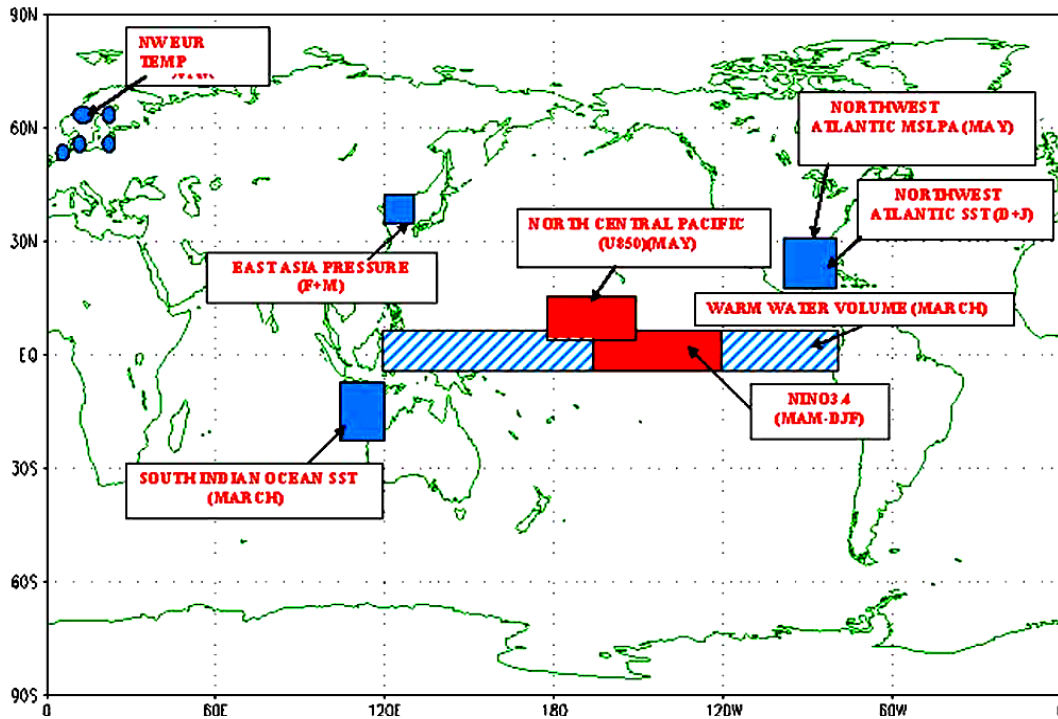


Fig. 1. Geographical locations of 8 predictors used for ISMR

second stage forecast model, another predictor set requiring data upto May (SET-II) was used. Fig. 1 shows the geographical regions in which these predictors are defined. The predictors in SET-I and SET-II are listed in Tables 1 and 2 respectively. SET-I and II contain 4 and 5 predictors respectively. One predictor (East Asia Surface Pressure Anomaly) is common for both the sets. The SST predictors were derived as the simple arithmetic average of the monthly ERSST.v2 anomalies over the respective geographical region. The time periods used for the averaging are given in Tables 1 and 2. The pressure and wind predictors were similarly derived from the NCEP reanalysis data. The LST anomaly over the northwest Europe was computed as the average of LST anomalies of 5 land stations from Europe (Orland, Oslo/Gendermon, Ostursund/Frosen, Karlstad and De Bilt). The seasonal tendency in the NINO3.4 anomaly index was computed by subtracting monthly anomalies averaged over the winter season (DJF) from that averaged over the spring season MAM (March to May). All anomalies were computed using the climatological base period of 1971-2000. One of the SST predictors (equatorial southeast Indian Ocean) in the predictor set II showed significant warming trend during the data period. Hence, time series of this predictor was de-trended by removing the linear trend fitted for the period 1951-2000 from the time series.

More details of all the predictors in set I and set II can be found in Rajeevan *et al.* (2006). One more predictor set (SET-III) was identified, which was used for forecasting of NWISMR. Table 3 shows the details of the predictors used for forecasting NWISMR. The relationship of the predictors defined in Set I & II with ISMR and Set III with NWISMR was found to be stable (C.C. near or above 5% significant level) during most of the analysis period.

4. Methodology

4.1.1. Local regression

Local regression was applied in a variety of fields in late 19th and early 20th centuries (Henderson, 1916). The current popularity of local regression as a statistical procedure is largely due to the Lowess/Loess procedure (Cleveland 1979, Cleveland and Devlin 1988).

Local regression means fitting a regression equation locally. Suppose x is a predictor and y is predictand and there are n pairs of data $(x_1, y_1), \dots, (x_n, y_n)$. In simple linear regression an equation of the form $y = ax + b$ is fitted and the coefficients a & b are estimated based on

TABLE 1

Details of predictors used for first stage forecast (SET-I)

No.	Parameter	Period	Spatial domain	C.C. with ISMR (1977-2005)
A1	North Atlantic SST Anomaly	Dec + Jan	20° N - 30° N, 100° W - 80° W	-0.48**
A2	East Asia Surface Pressure Anomaly	Feb + Mar	35° N - 45° N, 120° E - 140° E	0.64**
A3	Europe Land Surface Air Temperature Anomaly	Jan	5 Stations	0.50**
A4	Warm Water Volume	Feb + Mar	5° S - 5° N, 120° E - 80° W	-0.33

* and ** indicate statistical significant at 5% and 1% level respectively

TABLE 2

Details of predictors used for second stage forecast (SET-II)

No.	Parameter	Period	Spatial domain	C.C. with ISMR (1977-2005)
J1	North Atlantic MSLP Anomaly	May	20° N - 30° N, 100° W - 80° W	-0.44*
J2	Equatorial SE Indian Ocean SST Anomaly	Feb + Mar	20° S - 10° S, 100° E - 120° E	0.49**
J3	East Asia Surface Pressure Anomaly	Feb + Mar	35° N - 45° N, 120° E - 140° E	0.64**
J4	Nino 3.4 SST Anomaly Tendency	MAM (0) – DIF (0)	5° S - 5° N, 170° W - 120° W	-0.47*
J5	North Central Pacific Zonal Wind Anomaly at 850 hPa	May	5° N - 15° N, 180° E - 150° W	-0.37*

* and ** indicate statistical significant at 5% and 1% level respectively

TABLE 3

Details of predictors used for northwest India model

No.	Parameter	Period	Spatial domain	C.C with NWISM (1977-2005)
N1	South Atlantic MSLP Anomaly	Jan	35° N - 45° N, 60° - 50° W	-0.59**
N2	North Atlantic MSLP Anomaly	May	17.5 - 27.5, 55 - 42.5° W	0.41*
N3	South Pacific MSLP Anomaly	May	12.5 - 22.5, 142.5 - 150° E	-0.49**
N4	North Atlantic SST Anomaly	Jan	22 - 28° N, 86 - 78° W	-0.51**
N5	East Asia Surface Pressure Anomaly	Feb + Mar	35 - 45° N, 120 - 130° E	0.57**

* and ** indicate statistical significant at 5% and 1% level respectively

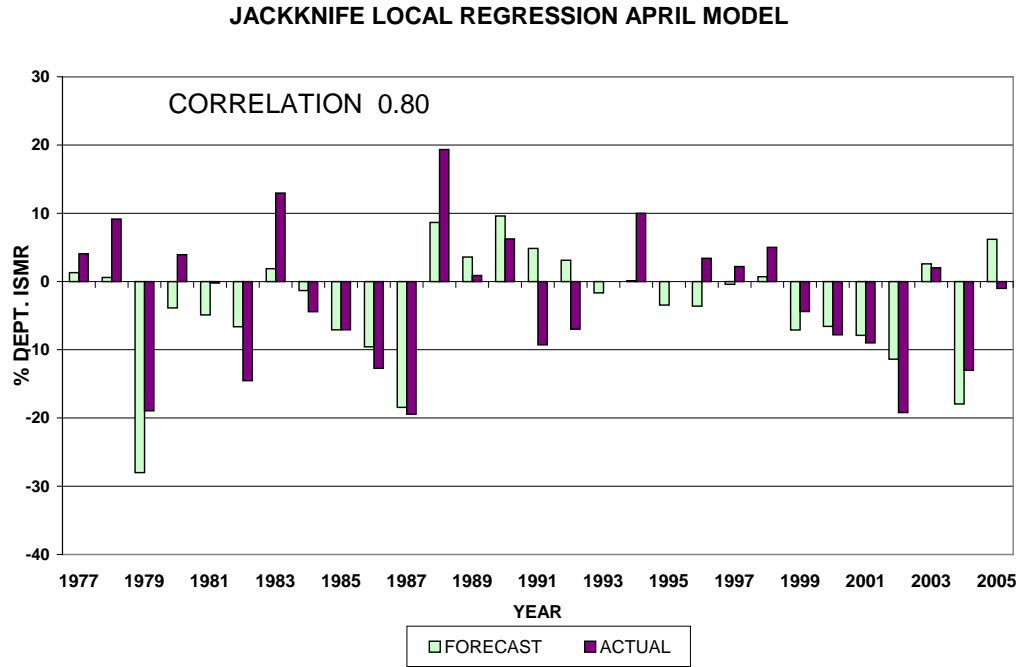


Fig. 2. Performance of local regression model using Jackknife technique for April forecasts, which used predictors from SET-I

method of least squares. The value of a & b are derived based on the whole data set and on all the n observations.

Under the theory of Loess, suppose x_0 is an observation of x , we consider the interval $(x_0 - \alpha, x_0 + \alpha)$ where α is suitably chosen so as to include the points of x that lie in the local neighborhood of x_0 . Suppose there are m such points, now we build up a linear/polynomial regression model based on these m points using method of least squares. However a weighted function is defined corresponding to each observation x , the weight varies with the distance of x from x_0 such that more weight is assigned to closer points. The regression equation is then developed. When we input $x = x_0$ in the regression equation, the estimated value of y say y^{\wedge}_0 can be estimated. If polynomial regression form is used only a second order polynomial is employed. Thus for each point x_0 , a different regression equation should be worked out. This technique takes into consideration the non-linear nature of relation between x and y . Over all model for local regression is expressed as $y_t = f(x_t)$ as different from $y = ax + b$ which is expression for conventional linear regression .

If there is more than one independent variable the technique of multiple regression or multiple polynomial regression could be used.

The underlying model for local regression is

$$Y_t = f(x_t) + e_t \tag{1}$$

where $x_t = (x_{1t}, x_{2t}, x_{3t}, \dots, x_{pt})$, $t = 1, 2, \dots, N$. This is similar to the linear regression model, but the function f could be linear or nonlinear, and the errors e_t are assumed to be normally distributed with zero mean and variance σ^2 . The key difference from linear regression is that the function f is fit 'locally' to estimate the value of Y . The value of the function at any point x_i is obtained by

- (i) Identifying a small number $K = \alpha N$, where $\alpha \in (0, 1)$ of neighbours to x_i
- (ii) Fitting a polynomial of order p to the neighbours, identified from the observations that are closest to x_i in terms of the Euclidian distance or another such metric (Mahalanobis distance; Yates *et al.* 2003).
- (iii) The fitted polynomial is then used to estimate the mean value of the dependent variable. The coefficients of the polynomial are estimated using a weighted least-squares approach.

The theoretical background of the local polynomial method is described in detail in Loader (1999), who refers

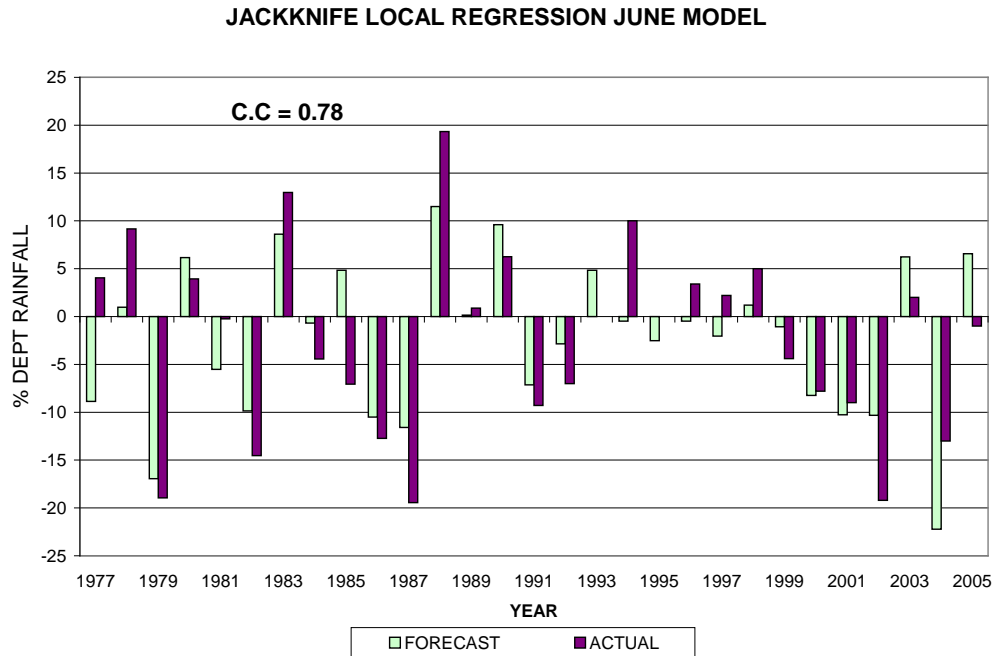


Fig. 3. Performance of local regression model using Jackknife technique for June forecasts, which used predictors from SET-II

to it as LOCFIT. We have used the same terminology in this paper.

The step for generating the forecast is as follows:

(i) For a new value of the predictor set, the mean value Y new is estimated using the LOCFIT approach as described above

The key parameters to be estimated in the LOCFIT model are the size of the neighbourhood (K or α) and the order of the polynomial p . These parameters are obtained using objective criteria such as the generalized cross-validation (GCV) function or likelihood function:

$$\text{GCV}(\alpha, p) = \frac{\sum_{i=1}^N \frac{e_i^2}{N}}{\left(1 - \frac{m}{N}\right)^2} \quad (2)$$

Where e_i is the error (*i.e.*, difference between the model estimate and observed), N is the number of data points and m is the number of parameters (predictors). For a suite of α and p values the GCV function is computed from Equation (2) and the combination that gives the least GCV value is selected. For stability purposes, the minimum neighbourhood size should be twice the number

of parameters to be estimated in the model. If a first-order (*i.e.*, linear) polynomial is selected, and if the neighbourhood includes all the observations (*i.e.*, $K = N$ or $\alpha = 1$), then results in the traditional linear regression. Thus, LOCFIT can be viewed as a superset. For developing the model, we used the software LOCFIT developed by Loader (1999), which is available on-line at (<http://cm.bell-labs.com/cm/ms/departments/sia/project/locfit/index.html>).

There are some non-parametric approaches for estimating the function locally. But we have adopted the LOCFIT technique, which is easy to implement. LOCFIT has been used for several hydroclimate applications (Lall, 1995), for spatial interpolation of precipitation (Rajagopalan and Lall, 1998), flood frequency estimation (Apipattanavis *et al.*, 2005) and Seasonal forecasting of Thailand summer monsoon rainfall (Nkrintra *et al.*, 2005).

For the Local Regression method, owing to the small sample size we used polynomial of order 1 (*i.e.*, local linear fit). Local neighbourhood size (α) for this study is chosen as 0.40 for all the models.

4.1.2. Model development period

Various components of the Indian monsoon exhibit significant inter-decadal variability. Modulation of inter-annual variability by inter-decadal variability influences

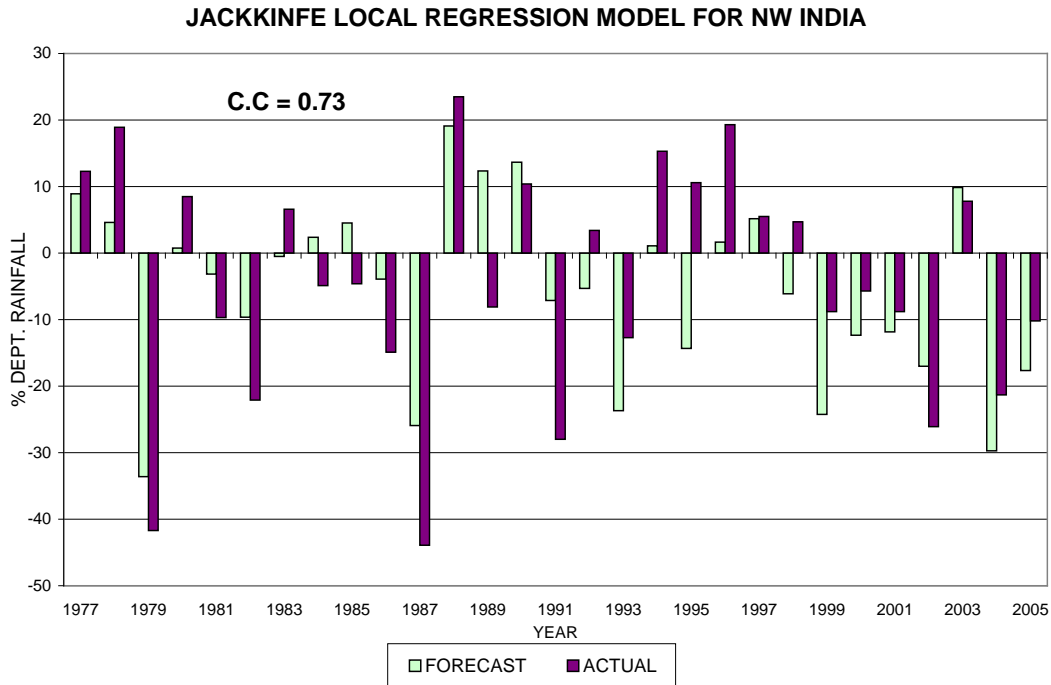


Fig. 4. Performance of Local regression model using Jackknife technique for northwest India summer monsoon rainfall, which used predictors from SET-III

predictability of seasonal mean monsoon (Goswami 2005). A recent study by Goswami (2004) revealed that potential predictability of monthly mean summer monsoon climate has decreased by almost a factor of two during the recent decades (1980s and 1990s) compared to decades of 1950s and 1960s associated with the major inter-decadal transition of climate in mid 1970s.

Therefore, for the model development, we have used data from 1977-2005 in this study. Owing to a small sample size, the Jackknife technique (Crask and Parreault 1977; Tukey 1958) which is most suitable for checking the model performance when data period is small was adopted. In accordance with the Jackknife method, prediction for each of the years (say i^{th} year) within the given data period of k years was done using the remaining $k-1$ years. For example for predictions of year 1977 model was developed for the period 1978 to 2005 (28 years). For 1978 prediction model was developed using year 1977 and 1979-2005 (28 years) data.

To reduce the dimension of the predictor data set, we have first performed a Principal Component Analysis (PCA) and the few (First 2) PCs (principle components) having highest correlation coefficient (CC) with the

predictand (ISMR, NWISMR) is selected. The selected PCs are then used as predictors in the regression analysis.

To examine the skill of the forecast models Hit score for the three category (deficit, normal, excess) forecast during the verification period was evaluated. Hit Score is proportion of correct forecasts and is expressed as the ratio of forecasts in correct category to the total number of forecasts. In addition, root mean square error (RMSE) and CC between predicted and actual rainfall also have been calculated.

5. Results and discussion

The performance of the models for ISMR using the Jackknife method is shown in Table 4. As seen in the Table 4, RMSE for the period 1977-2005 of the April model is 5.99%. RMSE for the June model is 6.06%. RMSE of predictions based on climatology alone was 10% of LPA. Thus performance of both the models was far better than that of the model based on climatology.

Fig. 2 shows the actual and predicted ISMR for the April model based on local regression. Fig. 3 shows the same for the June model. As seen in Fig. 2, the extreme

TABLE 4
Performance of statistical models using Jackknife technique (Period 1977-2005)

Skill score	LOCFIT April model	MR April model	LOCFIT June model	MR June model	NW India LOCFIT model	Northwest India MR model
Hit Score	0.65	0.51	0.66	0.67	0.55	0.51
RMSE % of LPA	5.99	6.25	6.0	6.0	11.5	11.73
Correlation coefficient	0.80	0.76	0.78	0.77	0.73	0.71
Bias	-0.85%	-0.03%	-0.06%	-0.08%	-0.77%	-0.31%

years like 1979, 1982, 1986, 1987, 2002 and 2004 are well predicted by the model. However in some years, the performance was not good. In 1983, an excess monsoon year, the model predicted a slight excess. The model did not predict large deficiency in 1991 and 1992. However, April model were able to predict the sign of ISMR every year after 1996. Two recent drought years (2002 & 2004) were also correctly predicted by this model. CC between actual and predicted ISMR for this model is 0.80, which is statistically significant.

The June model, as seen in Fig. 3 predicted all the extreme years (1979, 1982, 1983, 1986, 1987, 1988, 2002 and 2004) well. But for the years like 1977, 1985, 1994 (normal monsoon years) actual and predicted ISMR are of opposite sign. Compared to the April model, June model showed improved performance in 1983, 1991 and 1992. From, 1998 onwards, this model was able to predict correctly the sign of ISMR every year. Further, this model was able to predict the recent two drought years (2002 & 2004) correctly. CC between actual and predicted ISMR for the June model is 0.78, which is statistically significant.

Fig. 4 shows the performance of local regression model for NWISMR. The extreme years like 1979, 1987, 1988, 2002 and 2004 are well explained by the model. Model was not able to explain the years like (1989, 1991, 1995 and 1996). From, 1999 onwards, this model was able to predict correctly the sign of NWISMR every year. Further, this model was able to predict two recent drought years (2002 and 2004) correctly. From the Table 4 it is clear that CC between actual and predicted NWIR for this model is 0.74, which is statistically significant. RMSE for this model is 11.5%. RMSE of predictions based on

climatology was 18.5 % of LPA. Thus performance of this model is far better than model based on pure climatology.

Locfit model was compared with a simple Multiple Regression (MR) model using the same predictors. The jackknife method was used to assess the skill of the models. For the MR model also, first two PCAs of the predictor set was selected. Table 4 also shows the results of MR models used for comparison. It is clear from Table 4 that Hit scores for LOCFIT models are greater than MR models and RMSE of the model based on LOCFIT method is lower than MR model, thus suggesting that LOCFIT models are performing better as compared to the MR models.

6. Conclusions

Two statistical models were developed using local regression method, to support the IMD's two-stage long range forecasting strategy for ISMR. The April model was based on a predictor set which used data upto the month of March to support the first stage forecast. June model was based on a predictor set which used data upto the month of May to support the second stage forecast. For assessing the performance of both the models during the period 1977-2005, the Jackknife technique was applied. Both the models showed good skill in forecasting the ISMR during most of the years considered in this study. Particularly, during most of the extreme ISMR years, the predicted ISMR was close to the actual value. Both models were also able to correctly predict the recent two drought monsoon years (2002 & 2004). RMSE during the independent forecasting period for both the models (used for both the first and second stage forecast) was relatively smaller than that for model based on climatology. Hit score of the models are 0.65 and 0.66 for April and June model respectively. CC between actual and predicted ISMR during verification period for April and

June model are 0.8 and 0.78, which are statistical significant. Northwest India model also showed good skill in forecasting NWISMR during the verification period. Hit score of the model is 0.55. This model is able to capture most of the extreme years. From 1998 onward this model is able to capture the sign of NWISMR each year. RMSE for this model is 11.5 % of LPA as against 18.5 % of LPA model based on pure climatology alone.

Acknowledgments

We are thankful to Dr. (Mrs) N. Jayanthi, ADGM(R) for giving us encouragement and support. We also appreciate the sincere efforts made by the staff members of the National Climate Center. We also thank the anonymous referee for their valuable comments and suggestions, which helped in improving the quality of the paper.

References

- Apipattanavis, S., Rajagopalan, B. and Lall, U., 2005, "Local polynomial technique for flood frequency analysis", *Journal of Hydrology*, submitted for publication.
- Cleveland, W. S., 1979, "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*, **74**, 829-836.
- Cleveland, W. S. and Devlin, S. J., 1988, "Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association*, **83**, 596-610.
- Crask, M. R. and Perreault, W. D., Jr., 1997, "Validation of discriminant analysis in marketing research", *Journal of Marketing Research*, **14**, 60-68.
- Goswami, B. N., 2004, "Interdecadal changes in potential predictability of the Indian summer monsoon", *Geophys. Res. Letters*, **31**, L16208, doi: 10.1029/2004GL020337.
- Goswami, B. N., 2005, "The Asian Monsoon: Interdecadal Variability, in Global Monsoon System: Research and Forecast", Edited by C. P. Chang, Bin Wang and N. C. Lau, Chapter **26**, p455.
- Gowariker, V., Thapliyal, V., Kulshrestha, S. M., Mandal, G. S., Sen Roy, N. and Sikka, D. R., 1991, "A power regression model for long range forecast of southwest monsoon rainfall over India", *Mausam*, **42**, 125-130.
- Gowariker, V., Thapliyal, V., Sarkar, R. P., Mandal, G. S., Sikka, D. R., 1989, "Parametric and power regression models: New approach to long range forecasting of monsoon rainfall in India", *Mausam*, **40**, 115-122.
- Henderson, R., 1916, "Note on graduation by adjusted average", *Actuarial Soc. Amer.*, **17**, 43-48.
- Kung, E. C. and Sharif, T. A., 1982, "Long range forecasting of the Indian Summer monsoon onset and rainfall with upper air parameters and SST", *J. Met. Soc. Japan*, **60**, 672-681.
- Lall, U., 1995, "Recent advances in non-parametric function estimation", *Reviews of Geophysics*, **33**, 1093-1102.
- Loader, C., 1999, "Local Regression Likelihood", Springer: New York.
- McPhaden, M. J., 2003, "Tropical Pacific Ocean heat content variations and ENSO persistence barriers", *Geophys. Res. Letter*, **30**, 1480, doi: 10.1029/2003GL016872.
- Meinen, C. S. and McPhaden, M. J., 2000, "Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña", *J. Clim.*, **13**, 3551-3559.
- Nkrintra, Singhrattna, Rajagopalan, Balaji, Clark, Martyn and Krishna Kumar, K., 2005, "Seasonal forecasting of Thailand summer monsoon rainfall", *Int. J. Climatol.*, **25**, 649-664.
- Rajagopalan, B. and Lall, U., 1998, "Locally weighted polynomial estimation of spatial precipitation", *J. Geographic Information and Decision Analysis*, **2**, 44-51.
- Rajeevan, M. and McPhaden, M. J., 2004, "Tropical Pacific upper ocean heat content variations and Indian summer monsoon rainfall", *Geophys. Res. Letters*, **31**, L18203, doi:10.1029/2004GL020631.
- Rajeevan M., Pai, D. S., Dikshit, S. K. and Kelkar, R. R., 2004, "IMD's New Operational Models for Long Range Forecast of South-west Monsoon Rainfall over India and their Verification for 2003", *Current Science*, **86**, 422-431.
- Rajeevan, M., Pai, D. S. and Rohilla, Anil Kumar, 2005, "New statistical models for long range forecasting of south west monsoon rainfall over India", National Climate Center research report No. 1.
- Rajeevan, M., Pai, D. S., Rohilla, Anil Kumar and Lal, B., 2006, "New statistical models for long range forecasting of southwest monsoon rainfall over India", *Climate Dynamics*, DOI 10.1007/s00382-006-0197-6.
- Smith, T. M. and Reynolds, R. W., 2004, "Improved Extended Reconstruction of SST (1854-1997)", *Journal of Climate*, **17**, 2466-2477.
- Tukey, J. W., 1958, "Bias and confidence in non-quite large samples", *Annals of Mathematical Statistics*, **29**, p614.
- Walker, G. T., 1923, "Correlation in seasonal variations of weather, VIII, A preliminary study of world weather", *Mem. India Met. Deptt.*, **24**, 75-131.
- Walker, G. T., 1924, "Correlation in seasonal variations of weather, IV, A further study of world weather", *Mem. India Met. Deptt.*, **24**, 275-332.

Woodruff, S. D., Diaz, H. F., Elms, and Worely, S. J., 1998, "COADS release 2 data and metadata enhancement for improvements of marine surface flux fields", *Phys. Chem. Earth*, **23**, 517-527.

Yates, D., Gangopadhyay, S., Rajagopalan, B. and Strzepek, K., 2003, "A nearest neighbour bootstrap technique for generating regional climate scenarios for integrated assessments", *Water Resources Research*, **39**, 7, 1199, SWC 7-1-7-15.
