MAUSAM, 50, 3 (July 1999), 299-310

Letters to the Editor

551.509.33 : 551.577

AN ARMA MODEL FOR LONG-RANGE FORECASTING OF RAINFALL OVER COIMBATORE

1. Agriculture is rightly described as a gambling with monsoon. In this game, human efforts to understand the nature and to adopt his cultivation practices to it is a legend. The problem of uncertainty of weather, especially rainfall and man's efforts to understand the process of change and to adopt himself to it are particularly important in dry land agriculture. The importance of information on amount of rainfall sequences and lengths of wet and dry spells for crop planning and to carry out agricultural practices has been discussed by Pandarinath (1991). Due to larger variation in the weather system, a variety of rainfall patterns are observed which are erratic and quite undependable.

Several methods have been tried to model rainfall amounts. They can be broadly classified as distribution models, regression models and progression models. The distribution models are not useful for forecasting future rainfall. Agricultural management practices could be modified towards increased and sustainable productivity only when future rainfall can be predicted well in advance. Hence, regression and progression models have been tried to model rainfall amounts.

The variations in Indian monsoon rainfall are due to some natural causes. This line of thinking paved the way for investigating the causes for the variability of monsoon rainfall in India. El-Nino and Southern Oscillation have been identified in recent years as promising signals of monsoon variability.

Many authors worked on these lines. The important works were by Thapliyal (1982, 1990, 1993), Gowarikar *et al.* (1989, 1991) and Thapliyal and Kulshrestha (1992).

The regression models used so far for long range forecasting of Indian monsoon rainfall are general models to represent the entire country and for the entire South-West monsoon season (June-September). Such general models will not be applicable for smaller regions. They are also not valid for shorter time periods such as weeks.

Moreover, they involve numerous climatic parameters for which data collection is not easy and economical. A great deal of efforts and resources are necessary to collect relevant data. Hence, there is a need to find out a model that is useful for long range forecasting of rainfall for smaller regions and shorter periods and which utilizes relatively easily available data. Application of time series technique may be a solution for the above problems

The auto regression moving average (ARMA) models are powerful tools to forecast time series. The application of ARMA schemes to time series problems was popularised by Box and Jenkins (1970). Their approach is commonly known as Box-Jenkins methodology.

No much work has been done on application of time series models to forecast rainfall. Rangaswamy and Kulandaivelu (1980) used pure auto regressive (AR) model to forecast annual rainfall for Coimbatore, Tamil Nadu, India. Thapliyal (1982, 1991, 1993) has shown that the lead indicator ARIMA models can be used for forecasting rainfall. Borah and Bora (1995) used seasonal ARIMA model to predict monthly rainfall around Guwahati, Assam, India.

Although a variety of tools were attempted for forecasting rainfall, there is no universally acceptable and reliable tool. This study is an attempt in that direction.

2. Rainfall data for Coimbatore city in Tamil Nadu State, India have been used for this study since daily, weekly, seasonal and annual rainfall data are available for 91 years continuously from 1907-97.

The commonly used criterion to choose the most reliable forecast model is the minimum mean square error (MSE). The selected model might fit very well the data from which the unknown parameters are estimated. However, the agreement between the forecasts and future data that are not used for estimation need not be as good. Hence, comparison of forecasts with actual future observations can be additional useful tool for model evaluation and selection (Box and

TABLE 1

Rainfall forecast for South west and North east monsoon- Coimbatore 1982-97

(Rainfall in mm)

Year	South west monsoon			North east monsoon		
	Observed	Forecast 1	Forecast 2	Observed	Forecast 1	Forecast 2
1982	156	89	155	353	191	290
1983	158	110	162	425	254	310
1984	175	220	183	417	436	342
1985	226	112	164	221	294	322
1986	195	121	167	216	237	311
1987	91	132	168	509	64	279
1988	257	236	179	98	111	285
1989	142	165	171	299	155	287
1990	125	213	178	238	174	289
1991	228	177	172	109	372	335
1992	295	108	165	400	427	349
1993	106	216	177	432	512	374
1994	177	221	179	430	443	354
1995	122	164	172	210	517	376
1996	193	197	176	376	238	303
1997	180	175	179	605	345	410
Mean square error		5693	3015		33875	16649
Root mean square error		75.4	54.9		184.1	129.0

Forecast 1 : Without shrinkage constant, Forecast 2: With shrinkage constant

Tiao, 1975). In practical situation one can not expect many future observations. Therefore, in a series the initial part may be used for model building and the remaining part for forecast evaluation and comparison.

In the present study, 75 years data were used for model construction. The minimum required sample size suggested in the time series literature is 50. Thus, with the data of first 75 years starting from 1907 the models were constructed and the one- step-ahead forecast was made for the year 1982. Similarly, taking the data of previous 75 years for model construction the forecasts were made for the year 1983-97.

Since the crops are usually grown during South-West(SW) and North-East(NE) monsoon seasons, the models were constructed for these series only.

3. The ARMA model of order p and q is represented as

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$$

where, X's are deviations of actual values from the mean μ, t is the time period, ϕ's are the auto regression parameters, e's are the errors and θ's are the moving average parameters. The parameters are estimated using Box-Jenkins methodology.

In the usual ARMA models the parameters are constants. Makridakis and Wheelwright (1978) have developed a modelling approach with time-varying parameters. Their approach is known as adaptive filtering. It consists of a heuristic recursive algorithm that revises the parameter estimates as each new observation becomes available. The heuristic optimization procedure does not require any a *priori* knowledge about the time series in question. The parameters of the ARMA model are estimated through a non-linear least squares approach by using the method of steepest descent. The formula for adapting the parameters of the auto regressive (AR) model according to the method of steepest descent is

$$\phi_{it}' = \phi_{it-1} + 2\, k\, e_t * X_{t-i*}\ ,\ i=1,2,\ldots,p \qquad (1)$$

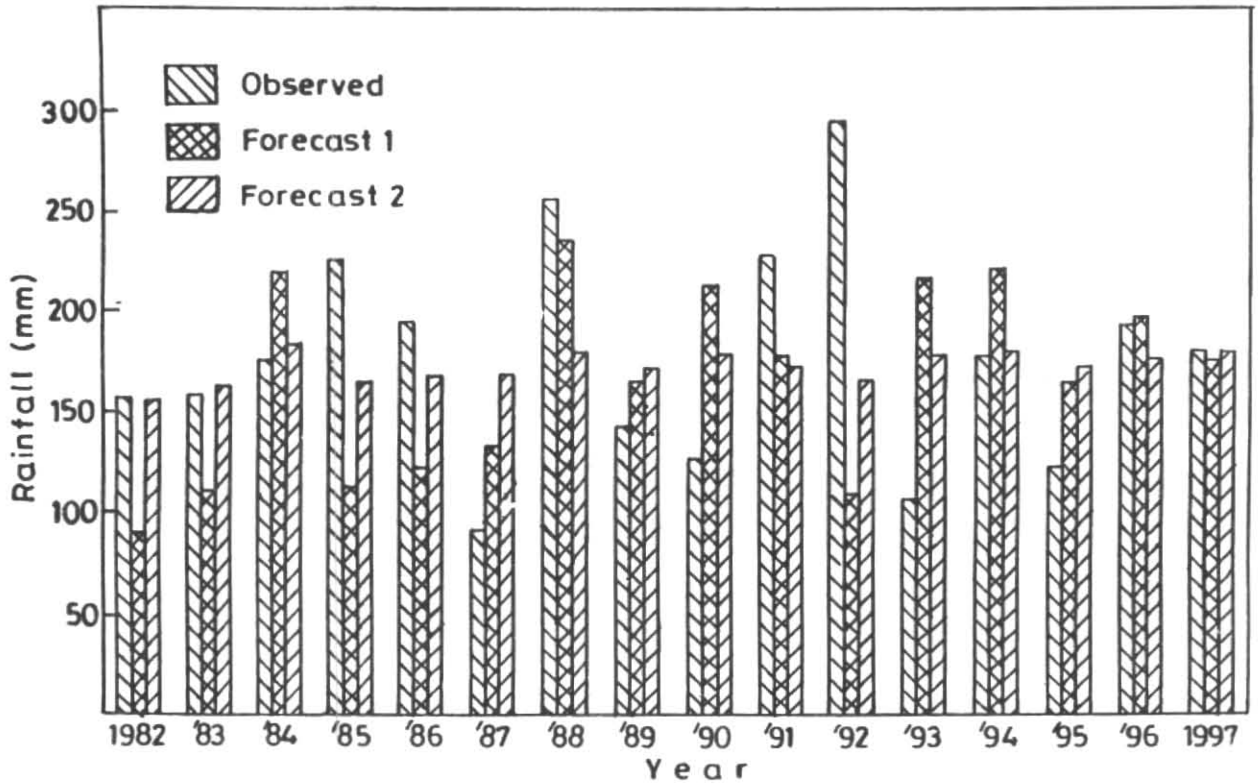

Fig. 1. Rainfall forecast for southwest monsoon Coimbatore (1982-97). Forecast 1 - without the index of shrinkage, Forecast 2 - with the index of shrinkage

where, ϕ_{it}' are the new parameters

ϕ_{it-1} are the old parameters

k is the learning constant

X_{t-i*} are standardized values at time t-i

e_{t*} is the standardized error term at time t

$e_t = X_t - \hat{X}_t$

The learning constant k determines the speed of adoption. It will be in the range of $0<k<1/p$ where p is the order of the AR model. The standardization is done by dividing X_i and e_i values by $[\sum_{i:1}^{p} X_{t-i}^2]^{1/2}$. The standardization coefficient is used so that the X_i and e_i values fall between 0 and 1.

Their procedure starts with the determination of order of the model and initial estimates of the parameters as in the other methods. The appropriate formula for adapting the parameters is used till all the N observations are used. One complete pass through the series $X_{p+1}, X_{p+2}, ..., X_N$ is called an iteration. Usually several iterations are required to get optimum parameters. If the relative change in the MSE of one-step ahead forecasts from one iteration to the next is smaller than a predetermined constant, the iterations are stopped. The estimate of parameters from the last iteration is used in the forecast of future observations.

According to Makridakis and Wheelwright, the simplicity, economy and self-adapting procedure of the adaptive filtering are the three main advantages of it. There are no real practical problems in applying the method as long as the learning constant k is set as stated already. The optimization will converge when k is properly specified even if the series is non-stationary or the wrong order of the model is specified.

The adaptive filtering has been questioned on both theoretical and empirical grounds by several authors (Ekern, 1976; Golder and Settle, 1976; Montgomery and Contreras, 1977). A major criticism of the adaptive filtering technique has been the fact that it does not make explicit reference to a model for the parameters changes and that it does not discuss the assumptions that are necessary to yield the revision values in equation (1). Makridakis and Wheelwright (1978) admit that the method of adaptive filtering needs a better theoretical basis. Since the parameters are not fixed, it is difficult to use the classical statistical concepts

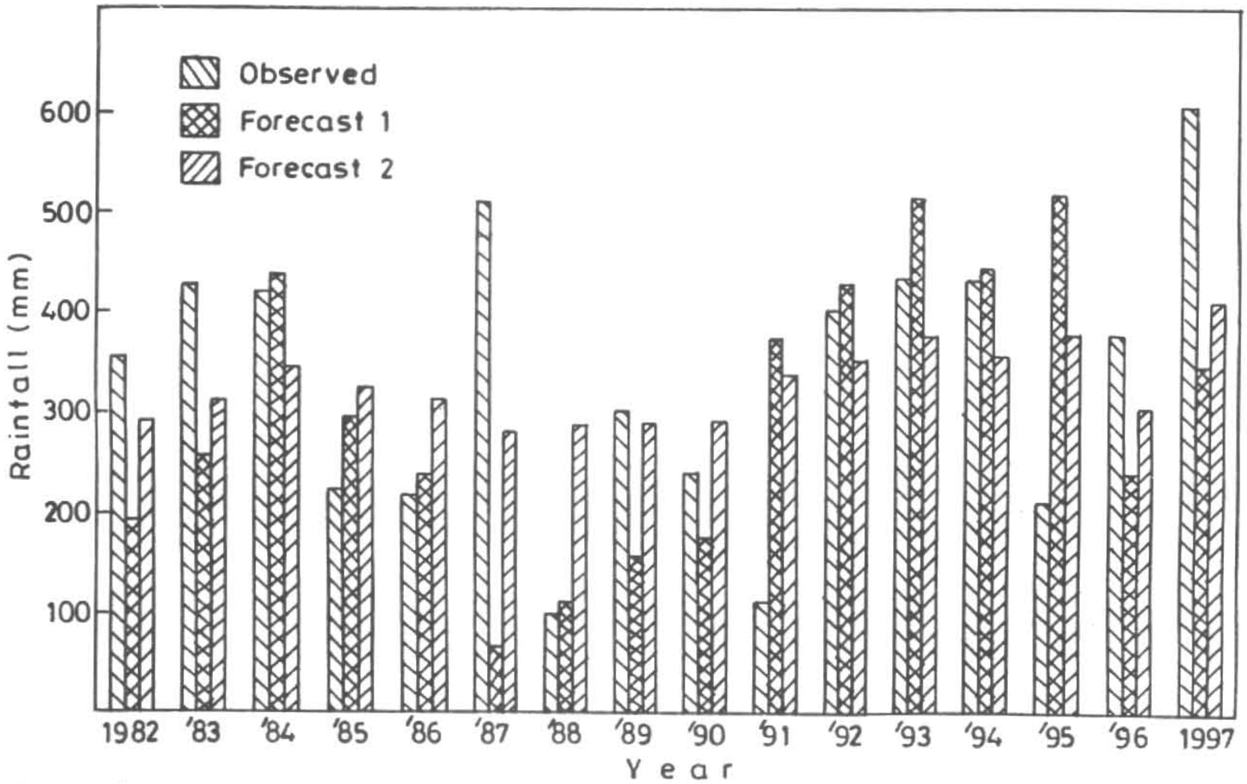

Fig. 2. Rainfall forecast for northeast monsoon - Coimbatore 1982-97. Forecast 1 - without the index of shrinkage, Forecast 2 - with the index of shrinkage

used in the time series analysis.

4. The auto correlations for the seasonal and annual rainfall data of Coimbatore reveal that the rainfall series is random. In such situations the usual Box-Jenkins methodology which depends on auto correlation functions may not be used to fit the time series model. Instead, the Makridakis - Wheelwright procedure of adaptive filtering may be tried.

Following the stepwise autoregression procedure described by Newbold and Granger (1974) and employing adaptive filtering technique with $N = 75$ and $k = 0.197$ it was found that AR(50) model fitted well to the SW and NE monsoon rainfall data of Coimbatore. The MSE approached zero as the number of iterations increased to 25. However, the MSE for the post sample forecasts increased as the number of interations increased. It was observed that with two iterations the MSE of the post sample forecasts was minimum in almost all the cases.

Using AR (50) models and adapting adaptive filtering with two interations the forecasts were made for SW as well as NE monsoon seasons for 16 years from 1982 to 1997. The results are presented in Table 1.

In both SW and NE monsoon rainfall series the models fitted well in the fitting phase. However, they did not forecast well in the forecasting phase which can be observed from the results in Table 1. Chatfield (1975) observed that with large number of parameters close to the number of observations the model may fit the series well, but, the forecasts may be highly erratic with such models.

Similar problem arises in predictions using usual regression models. According to Copas (1983) the fit of a regression predictor to new data is nearly always worse than its fit to the original data. To overcome this problem he suggested the use of shrinkage constant which will give a uniformly lower prediction MSE than least squares. In the present study the shrinkage constant was used to reduce the MSE of the forecast using AR model. The index of shrinkage is given by

$$K = \frac{1 - 2\delta^2}{1 + (P - 4)\delta^2}$$

where $\delta^2 = \sigma^2 / n\beta^T V\beta$

$V = n^{-1}X^TX$

$X = (n\text{-}p) \times p$ matrix of observed rainfall

$= X_1, X_2, X_3, ...X_p$

$X_2, X_3, X_4, ... X_{p+1},$

$X_3, X_4, X_5, ...X_{p+2}$

$X_{n-p}, X_{n-(p-1)}, X_{n-(p-2)}, ...X_{n-1}$

β = vector of parameters

p = number of parameters

n = number of observations

The results are given in Table 1. In case of SW monsoon rainfall the MSE for the post sample forecasts was 5693. When shrinkage constant was used, the MSE reduced to 3015. Similarly in case of NE monsoon rainfall the forecast MSE was 33875 without using shrinkage constant while it was 16649 using shrinkage constant.

For Coimbatore the long term average rainfall is 174 mm for the SW monsoon and 321 mm for the NE monsoon. The variances for the SW and NE monsoon rainfall are 82.5 and 139.9 respectively. The root mean square error for the sample forecasts was 43.0 in case of SW monsoon and 80 in case of NE monsoon, which are far less than the standard deviations of the actual observations.

Taking the square root of MSE for sample forecasts as allowable error, the closeness of the forecast may be judged. On this basis the forecast for SW monsoon may be termed as satisfactory if it is 43 mm above or below the actual rainfall. Similarly, if the forecast for NE monsoon is 80 mm above or below the observed rainfall it may be taken as satisfactory. Based on this criterion it was observed that the forecasts for SW monsoon were close to the actual rainfall in 10 out of 16 years when shrinkage constant was used. But they were close only in 6 out of 16 years when the shrinkage constant was not used. In case of NE monsoon the forecasts were close to the actual rainfall in 8 out of 16 years, with or without using shrinkage constant.

Both in terms of MSE and closeness of forecasts, the usage of shrinkage constant improves the forecasting ability of the model. Further improvements in forecast may be possible with some other modifications similar to shrinkage constant. The validity of the usefulness of shrinkage constant in forecasting models has to be checked by applying it to rainfall data of more number of places.

References

Borah, D.K. and Bora, P.K., 1995, "Predicting the monthly rainfall around Guwahati using a seasonal ARIMA model", *Ind. Soc. Ag. Statistics*, **47**, 3, 278-287.

Box, G.E.P. and Jenkins, G.M., 1972, "Time series analysis - Forecasting and control", Holden - Day, Inc., San Francisco.

Box, G.E.P., and Tiao, G.C., 1975, "Intervention analysis with applications to economic and environmental applications", *J. Amer. Statist. Assoc.*, **70**, 70-79.

Chatfield, C., 1975, "The analysis of time series : Theory and Practice", Chapman and Hall, London.

Copas, J.B., 1983, "Regression, prediction and shrinkage", *J. R. Statist. Soc. B.*, **45**, 311-354.

Ekern, S., 1976, "Forecasting with adaptive filtering", *Opl. Res. Q.*, **27**, 705-715.

Golder, E.R. and Settle, J.G., 1976, "On adaptive filtering", *Opl. Res. Q.*, **27**, 857-868.

Gowariker, V., Thapliyal, V., Sarker, R.P., Mandal, G.S. and Sikka, D.R., 1989, "Parametric and power regression models: New approach to long range forecasting of monsoon rainfall in India", *Mausam*, **40**, 115-122.

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 south west monsoon rainfall over India", *Mausam*, **42**, 125-130.

Makridakis, S. and Wheelwright, S.C., 1978, "Forecasting : Methods and applications", John Wiley and Sons, New York.

Montgomery, D.C. and Contreras, L.E., 1977, "A note on forecasting with adaptive filtering", *Opl. Res. Q.*,**28**, 87-91.

Newbold, P. and Granger, C.W.J., 1974, "Experience with forecasting univariate time series and the combination of forecasts", *J. R. Statist. Soc. A.*, **137**, 131-165.

Pandarinath, N., 1991, "Markov chain model probability of dry and wet weeks during monsoon period over Andhra Pradesh", *Mausam*, **42**, 393-400.

Rangaswamy, R. and Kulandaivelu, R., 1980, "A model for forecasting rainfall, paper presented in the National Seminar on Agroclimatology, Tamil Nadu Agricultural University, Coimbatore, India, July 26-29, 1980, 43-46.

Thapliyal, V., 1982, "Stochastic dynamic model for long range prediction of monsoon rainfall in Peninsular India", *Mausam*, **33**, 399-404.

Thapliyal, V., 1990, "Long range prediction of summer monsoon rainfall over India: Evaluation and development of new models", *Mausam*, **41**, 339-346.

Thapliyal, V., 1991, "Issue of tropical long-range forecast -status and development of new forecast models, WMO/TD No 395, 267-272.

Thapliyal, V., 1993, "Long-range forecasting of summer monsoon rainfall and date of onset over India", in "Advances in Tropical Meteorology: Monsoon variability, satellite applications and modelling", Keshavamurthy, R.N. and Toshi, P.C., (Eds), Tata McGrow -Hill Publishing Co., New Delhi, 3-21.

Thapliyal, V. and Kulshrestha, S.M., 1992, "Recent models for long range forecasting of southwest monsoon rainfall in India", *Mausam*, **43**, 239-248.

Walker, G.T., 1910, "Correlation in seasonal variation of Weather II", *Mem. India Met. Dept.*, **21**, 22.

R. RANGASWAMY

Tamil Nadu Agricultural University
Coimbatore- 641003, India

24 December 1997, Modified 4 March 1999