A new method of forecasting south-west monsoon rainfall

K. Ranga Rao\textsuperscript{1*}, Naveen Kumar Boiroju\textsuperscript{2} and M. Krishna Reddy\textsuperscript{2}

\textsuperscript{1}Sri Venkateshwara Veterinary University, Dairy Technology Programme, Kamareddy-503 111, A.P., India
\textsuperscript{2}Department of Statistics, Osmania University, Hyderabad – 500 007, A. P., India
*E-mail: rrkollu@yahoo.com

ABSTRACT

In this paper, a new method of forecasting of monthly – seasonal rainfall in India is discussed. This methodology uses only the historical data. The parameters of the model are estimated using least squares principle. The performance of the proposed model has been evaluated on the basis of mean absolute error (MSE), root mean square error (RMSE) and mean absolute percent error (MAPE).

KEY WORDS: Forecast, Least square method, rainfall, time series

INTRODUCTION

The unknown future is a source of anxiety, giving rise to a strong human need to predict it in order to reduce, or ideally eliminate, its inherent uncertainty. Forecasting is a serious professional and scientific endeavor with a certain purpose, namely to provide predictions to be used in formulating decisions and taking actions.

The forecast translates into a decision, and, accordingly, the uncertainty attached to the forecast i.e. error, needs to be endogenous to the decision itself. In other words the use of the forecast needs to be determined- or modified- based on the estimated accuracy about what we should or should not forecast- as some forecasts can be harmful to decision makers. (Box \textit{et al.}, 1994; Salas \textit{et al.}, 1980). The main objective of this paper is to forecast the future monsoon rainfall using the proposed algorithm with the help of previous monsoon rainfall data.

METHODOLOGY

The monsoon rainfall can be predicted using multiple regression analysis and autoregressive models. Prediction performance of the models is measured using the mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage errors (MAPE). (Salas \textit{et al.}, 1980).

Data Source:

The data of south-west monsoon monthly seasonal (June – September) rainfall (in mm) in India for the years 1990 – 2005 is collected from Indian Meteorological Department, Pune, India. (www.tropmet.res.in). The data on south-west monsoon rainfall (in mm) for the period 1990 – 2000 (44 observations) is utilized for model fitting and data for subsequent periods 2001 – 2005 (20 observations) is used for validation. The data on south-west monsoon rainfall is given in Table 1.
Table 1: South-West Monsoon Rainfall (in mm) in India during 1990-2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>191.7</td>
<td>280.4</td>
<td>294.5</td>
<td>205.8</td>
</tr>
<tr>
<td>1991</td>
<td>181.0</td>
<td>276.2</td>
<td>247.2</td>
<td>123.9</td>
</tr>
<tr>
<td>1992</td>
<td>126.3</td>
<td>243.1</td>
<td>289.1</td>
<td>173.2</td>
</tr>
<tr>
<td>1993</td>
<td>172.4</td>
<td>294.9</td>
<td>198.8</td>
<td>239.6</td>
</tr>
<tr>
<td>1994</td>
<td>208.8</td>
<td>355.9</td>
<td>289.8</td>
<td>146.7</td>
</tr>
<tr>
<td>1995</td>
<td>135.1</td>
<td>313.0</td>
<td>259.8</td>
<td>192.4</td>
</tr>
<tr>
<td>1996</td>
<td>184.8</td>
<td>286.6</td>
<td>308.5</td>
<td>155.2</td>
</tr>
<tr>
<td>1997</td>
<td>180.2</td>
<td>296.6</td>
<td>282.4</td>
<td>168.1</td>
</tr>
<tr>
<td>1998</td>
<td>167.8</td>
<td>296.5</td>
<td>261.4</td>
<td>217.3</td>
</tr>
<tr>
<td>1999</td>
<td>175.2</td>
<td>273.5</td>
<td>223.7</td>
<td>190.8</td>
</tr>
<tr>
<td>2000</td>
<td>190.4</td>
<td>273.6</td>
<td>229.7</td>
<td>139.9</td>
</tr>
<tr>
<td>2001</td>
<td>219.0</td>
<td>279.5</td>
<td>209.2</td>
<td>114.2</td>
</tr>
<tr>
<td>2002</td>
<td>180.1</td>
<td>146.2</td>
<td>259.8</td>
<td>151.2</td>
</tr>
<tr>
<td>2003</td>
<td>179.9</td>
<td>317.8</td>
<td>252.3</td>
<td>169.5</td>
</tr>
<tr>
<td>2004</td>
<td>158.7</td>
<td>242.1</td>
<td>248.7</td>
<td>124.6</td>
</tr>
<tr>
<td>2005</td>
<td>143.2</td>
<td>334.1</td>
<td>190.1</td>
<td>206.9</td>
</tr>
</tbody>
</table>

Since given data is part or partial time series data, that is rainfall for the months of June to September, therefore the ARIMA technique does not applicable to the given data and it leads to formulate some new techniques to forecast the south-west monsoon rainfall based on the given historical data. A new method proposed in this paper to forecast the monsoon rainfall in India. (Box et al., 1994).

The performance of the model have been evaluated on the basis of MAE, RMSE and MAPE which are given by

\[ MAE = \frac{1}{n} \sum_{t=1}^{n} |Y_t - \hat{Y}_t| \]  

(1)

\[ RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2} \]  

(2)

\[ MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \]  

(3)

where \( Y_t \) is the original rainfall in different years and \( \hat{Y}_t \) is the forecasted rainfall in the corresponding years and \( n \) is the number of years used as forecasting period.
Proposed method of forecasting

The purpose of time series study is to measure chronological variations. The main objective in analyzing time series is to understand, interpret and evaluate change in the phenomena in the hope of more correctly anticipating the course of future events. In a mathematical approach, the objective is to build a model from past observations and use this model for future occurrences. (Box et al., 1994).

One can expect current month rainfall in the monsoon period may depend on its previous month rainfall (rainfall may spread in the current and previous months) and also it may depend on the rainfall in the same month of previous year monsoon (seasonality). Assuming that the monthly rainfall moderately correlated with the previous month rainfall, a statistical recurrence model is developed to forecast the rainfall in June to September in each year by combining the naïve method and regression methods. (Salas et al., 1980).

Model Development:

In this method, the current month figure is expressed in terms same month figure of previous year and previous month figure in the current year. For the given data we have four months namely June, July, August and September in each year. Let $y_{k1}$ represents the rainfall of June for the $k^{th}$ year, $y_{k2}$ represents the rainfall of July for the $k^{th}$ year, $y_{k3}$ represents the rainfall of August for the $k^{th}$ year, and $y_{k4}$ represents the rainfall of September for the $k^{th}$ year, and $k=1990, 1991, \ldots, 2005$.

Let the annual time series $Y_k = [y_{k1}, y_{k2}, y_{k3}, y_{k4}]^T$ where $k=1990, 1991, \ldots, 2005$.

The recurrence forecasting model for monthly-seasonal rainfall is given by

$$\hat{Y}_{k+1} = C.(A.Y_k + B.Y_{k-1}) \quad (4)$$

where $C$ is the yearly increment constant represents the average growth over the years, $A$ and $B$ are the coefficient matrices and $Y_k$ is the vector of time series values for the year $k$.

The yearly increment constant ($C$) is calculated as follows.

1. Calculate the yearly increments

$$I_k = \frac{T_k}{T_{k-1}} \quad \text{where } T_k \text{ is the total rainfall of the year } k.$$

2. Now compute yearly increment constant $C$ as geometric mean of $I_k$ values.

The coefficient matrices $A$ and $B$ are calculated as follows:

$$A = \begin{bmatrix}
\alpha_{11} & 0 & 0 & 0 \\
0 & \alpha_{22} & 0 & 0 \\
0 & 0 & \alpha_{33} & 0 \\
0 & 0 & 0 & \alpha_{44}
\end{bmatrix} \quad B = \begin{bmatrix}
0 & 0 & 0 & 0 \\
\beta_{21} & 0 & 0 & 0 \\
0 & \beta_{32} & 0 & 0 \\
0 & 0 & \beta_{43} & 0
\end{bmatrix}$$
The above coefficients are estimated using least squares method from the given data and are given below.

\[ \hat{\alpha}_{11} = \frac{\sum_k y_{k1} y_{(k-1)1}}{\sum_k y_{(k-1)1}^2} \]  \hspace{1cm} (5)

\[ \hat{\beta}_{21} = \frac{\sum_k y_{k2} y_{(k-1)2} \sum_k y_{(k-1)2} y_{k1} - \sum_k y_{k2} y_{k1} \sum_k y_{(k-1)2}^2}{\left( \sum_k y_{(k-1)2} y_{k1} \right)^2 - \sum_k y_{k1}^2 \sum_k y_{(k-1)2}^2} \]  \hspace{1cm} (6)

\[ \hat{\alpha}_{22} = \frac{\sum_k y_{k2} y_{(k-1)2} - \hat{\beta}_{21} \sum_k y_{(k-1)2} y_{k1}}{\sum_k y_{(k-1)2}^2} \]  \hspace{1cm} (7)

\[ \hat{\beta}_{32} = \frac{\sum_k y_{k3} y_{(k-1)3} \sum_k y_{(k-1)3} y_{k2} - \sum_k y_{k3} y_{k2} \sum_k y_{(k-1)3}^2}{\left( \sum_k y_{(k-1)3} y_{k2} \right)^2 - \sum_k y_{k2}^2 \sum_k y_{(k-1)3}^2} \]  \hspace{1cm} (8)

\[ \hat{\alpha}_{33} = \frac{\sum_k y_{k3} y_{(k-1)3} - \hat{\beta}_{32} \sum_k y_{(k-1)3} y_{k2}}{\sum_k y_{(k-1)3}^2} \]  \hspace{1cm} (9)

\[ \hat{\beta}_{41} = \frac{\sum_k y_{k4} y_{(k-1)4} \sum_k y_{(k-1)4} y_{k3} - \sum_k y_{k4} y_{k3} \sum_k y_{(k-1)4}^2}{\left( \sum_k y_{(k-1)4} y_{k3} \right)^2 - \sum_k y_{k3}^2 \sum_k y_{(k-1)4}^2} \]  \hspace{1cm} (10)

\[ \hat{\alpha}_{44} = \frac{\sum_k y_{k4} y_{(k-1)4} - \hat{\beta}_{43} \sum_k y_{(k-1)4} y_{k3}}{\sum_k y_{(k-1)4}^2} \]  \hspace{1cm} (11)

**Model Validation:**

The given data is divided into two independent sets namely in-sample set and out-of-sample set. In-sample set contains the monsoon rainfall during 1990-2000 is used for the model building and the obtained model is tested on the out-of-sample set consists of the monsoon rainfall during the 2001-2005 years. For the given data, the model parameters are \( C = 0.993 \) and the coefficient matrices.
Therefore the model is

\[
\begin{bmatrix}
\hat{y}_{(k+1)1} \\
\hat{y}_{(k+1)2} \\
\hat{y}_{(k+1)3} \\
\hat{y}_{(k+1)4}
\end{bmatrix} = (0.993)
\begin{bmatrix}
0.97 & 0 & 0 & 0 \\
0 & 0.53 & 0 & 0 \\
0 & 0 & 0.22 & 0 \\
0 & 0 & 0 & 0.27
\end{bmatrix}
\begin{bmatrix}
y_{k1} \\
y_{k2} \\
y_{k3} \\
y_{k4}
\end{bmatrix} + (0.993)
\begin{bmatrix}
0.79 & 0 & 0 & 0 \\
0 & 0.69 & 0 & 0 \\
0 & 0 & 0.47 & 0
\end{bmatrix}
\begin{bmatrix}
\hat{y}_{(k+1)1} \\
\hat{y}_{(k+1)2} \\
\hat{y}_{(k+1)3} \\
\hat{y}_{(k+1)4}
\end{bmatrix}
\]

For example,

\[
\hat{y}_{2001,Jan} = 0.993(0.97 \times 190.4) = 183.395
\]
\[
\hat{y}_{2001,Jul} = 0.993(0.53 \times 273.6 + 0.79 \times 183.395) = 287.861
\]
\[
\hat{y}_{2001,Aug} = 0.993(0.22 \times 229.7 + 0.69 \times 287.861) = 247.414
\]
\[
\hat{y}_{2001,Sep} = 0.993(0.27 \times 139.9 + 0.47 \times 247.414) = 152.979
\]

The obtained MAE, RMSE and MAPE values using this method and the values for the fitted data and validation set are given in Table 2.

### Table 2: Measures of errors for the proposed method

<table>
<thead>
<tr>
<th>Data</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitting</td>
<td>29.530</td>
<td>37.747</td>
<td>14.194</td>
</tr>
<tr>
<td>Validation</td>
<td>29.659</td>
<td>43.742</td>
<td>16.398</td>
</tr>
</tbody>
</table>

MSE : Mean absolute error, RMSE : Root mean square error, MAPE: Mean absolute percent error

Out-of-Sample forecasts using the proposed method are presented in Table 3.
Table 3: Forecasts of South-West Monsoon Rainfall (in mm) in India

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Observed rainfall</th>
<th>Forecasts of rainfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Jun</td>
<td>219.000</td>
<td>183.395</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>279.500</td>
<td>287.861</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>209.200</td>
<td>247.414</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>114.200</td>
<td>152.979</td>
</tr>
<tr>
<td>2002</td>
<td>Jun</td>
<td>180.100</td>
<td>176.648</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>146.200</td>
<td>290.073</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>259.800</td>
<td>252.800</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>151.200</td>
<td>158.999</td>
</tr>
<tr>
<td>2003</td>
<td>Jun</td>
<td>179.900</td>
<td>170.149</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>317.800</td>
<td>286.14</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>252.300</td>
<td>251.281</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>169.500</td>
<td>159.905</td>
</tr>
<tr>
<td>2004</td>
<td>Jun</td>
<td>158.700</td>
<td>163.889</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>242.100</td>
<td>279.159</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>248.700</td>
<td>246.166</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>124.600</td>
<td>157.76</td>
</tr>
<tr>
<td>2005</td>
<td>Jun</td>
<td>143.200</td>
<td>157.86</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>334.100</td>
<td>270.755</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>190.100</td>
<td>239.291</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>206.900</td>
<td>153.976</td>
</tr>
</tbody>
</table>

Figure 1: Original and Predicted Rainfall
Forecasts using the proposed method are presented in figure 1. It can be observed that the forecasted rainfall is very close to the original rainfall (Fig 1). Therefore the proposed method can be used to forecast the monsoon rainfall in India.

CONCLUSION

From the above study it is observed that, the proposed method can be treated as an alternative method for forecasting monsoon rainfall in India. The proposed method performs well at forecasting of structural changes and seasonal variations in the historical rainfall. MAE, RMSE and MAPE values are useful in comparing this method with any other method. However, this technique does not guarantee perfect forecasts. Nevertheless, it can be successfully used for forecasting long time series data and it should be updated from time to time with incorporation of current data.

REFERENCES


http://www.tropmet.res.in/~kolli/mol/Forecasting/frameindex.html

[MS received 13 March 2012; MS accepted 27 May 2012]