SELECTION OF THE BEST MODEL FOR THE RAINFALL DATA IN INDIA

Venkatramana Reddy S[1], Karthick Kumar Reddy G[2], Thasleema S.C[3], Ramkumar T.K[4], Sarojamma B[5]

***Abstract:** Forecasting Rainfall is beneficial to farmers, cultivators and Scientists. Year by year rainfall is change and from decade to decade the rainfall level is decreasing. In this paper, we are fitted two ARMA models for the rainfall data at different auto regressions with same moving averages. For the fitted ARMA models, the fit data best or not is tested using χ^2- test for goodness of fit. Between ARMA models which model is the best is identified by using some statistical measure using error.*

Keywords: Rainfall, Forecasting, ARMA model, Chi square test, MSE, MAPE, ME.

1. INTRODUCTION

Time series analysis and forecasting techniques plays an important role in meteorological Sciences, such as Climate conditions like Humidity, Rainfall, temperature, droughts, floods, environmental management fields etc. Among all, rainfall takes an important place because for year to year rainfall is not reached the sufficient levels. Decade to decade rainfall is going back to sufficient levels. Therefore, forecast on rainfall plays an important place for meteorological phenomena.

Md. Mahsin et al. [1] studied on Rainfall in Dhaka Division of Bangladesh using time series. In their study, they used Box-Jenkins methodology to build seasonal ARIMA model for monthly rainfall data taken for Dhaka station for the period from 1981 – 2001 with a total of 354 readings. RMSE values were calculated for Box-Jenkins model and selected the ARIMA (0, 0, 1) $(0, 1, 1)^{12}$ model. The selected model RMSE value is less comparatively Box- Jenkins model. Therefore, they concluded that the model forecasts rainfall sufficiently accurate.

Time series Analysis model for annual rainfall data in lower Kaduna Catchment Kaduna, Nigeria is given by Attah et al. [2]. In their paper they concluded that ARMA (1, 1) model is adequate for the forecasting of feature annual rainfall data.

Venkatramana Reddy et al. [3] fitted ARMA (3, 2) model for an average rainfall data in India during the period 1986-2006. Auto covariance and Auto correlations were calculated for data to measure linear relationship using single time series. Diagnostic check was attempted using portmanteau test to check the fitted ARMA best or not.

Pierini et al. [4], in their paper "Discriminating climatological regimes in rainfall time series by using the Fisher-Shannon method" investigated the complex dynamics of 8 long monthly rainfall in central Argentina, recorded from 1860 to

2006 using Fisher-Shannon (F.S.) information plane, defined by the Fisher information measure (FIM) and the Shannon entropy power (N_x).

2. METHODOLOGY

There are many types of ARIMA (p, d, q) (Auto Regressive Integrated Moving Average) models, where developed by changing parameters p, d, q.
AR: p = order of the autoregressive part.
I: d = degree of first differencing involved.
MA: q = order of the moving average part.

If you ignore differencing, then 'I' vanishes from ARIMA model, and now the model becomes ARMA model form. ARMA (p, q) (Auto Regressive Moving Average) model is generated by p^{th} order of auto regressive and q^{th} order of moving average.

AR: Auto regression tells about the relationship between the current time series values and past time series values.

MA: Moving Average tells about the relationship between the current time series values with current and past error terms.

The basic elements of AR and MA models can be combined to produce a great variety of models. An ARMA (p, q) model with higher-order terms is written as

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots\dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots\dots - \theta_q e_{t-q}$$

The above model is also written with the help of back shift operator

$$(1 - \phi_1 B - \phi_2 B^2 - \dots\dots - \phi_p B^p)\, Y_t = C + (1 - \theta_1 B - \dots\dots - \theta_q B^q)\, e_t$$

In this paper, we are fitted two types of models for rainfall data from 1990 to 2006. They are ARMA (2, 3) and ARMA (1, 3).

ARMA (2, 3): This model is a mixture of 2^{nd} order Auto regression and 3^{rd} order of the moving average part. The equation becomes
$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3}$
(or)
$(1 - \phi_1 B - \phi_2 B^2) Y_t = C + (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)\, e_t$

Where B is Back shift operator and
$B\, Y_t = Y_{t-1}$, $B^2 Y_t = Y_{t-2}$, $Be_t = e_{t-1}$, $B^2 e_t = e_{t-2}$, $B^3 e_t = e_{t-3}$

ARMA (1, 3): It is a mixture of 1^{st} order auto regression with 3^{rd} order moving average
$Y_t = C + \phi_1 Y_{t-1} e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3\, e_{t-3}$
(or)
$(1 - \phi_1 B)\, Y_t = C + (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)\, e_t$

Where $B\,Y_t = Y_{t-1}, Be_t = e_{t-1}, B^2e_t = e_{t-2}, B^3e_t = e_{t-3}$

Goodness of fit: The fitted model to data is either good or not, is tested by using χ^2-goodness of fit to a model. χ^2 - test for goodness of fit is

$$\chi^2 = \sum_{t=0}^{n} \frac{(Y_t - \hat{Y}_t)^2}{\hat{Y}_t} \sim \chi_n^2\ \alpha\ \%\ \text{l.o.s.}$$

Where Y_t = original time series value at time 't'

$\hat{Y}_t$ = estimated time series value at time 't'.

χ^2 calculated value is compared with χ^2 table value at (n-1) degrees of freedom at α % level of significance. If χ^2 calculated value greater than χ^2 table value at α % l. o. s. then we say that the fitted model is not good for data.

Statistical Measures using Error: If both models are fit good to data then which model is better comparing the two is identified by some statistical measures using error. Error means difference between actual observation and predicted observation. Error is denoted with 'e_t'.

$e_t = Y_t - \hat{Y}_t$

Where Y_t = actual observation of time series value at time 't'.

$\hat{Y}_t$ = forecasted observation of time series value at time 't'.

If there are observations and forecasts for 'n' time periods, then there will be 'n' error terms, and the following standard statistical measures can be obtained.

Mean error (ME): The mean error is the average of errors obtained from 'n' time periods. ME may be small since positive and negative errors tend to offset one another. ME tells about forecasting bias on under or over forecasting.

$$\text{Mean error (ME)} = \frac{1}{n}\sum_{t=1}^{n} e_t$$

Mean absolute error (MAE): MAE gives much indication as to the size of the typical errors. MAE is defined by first making each error positive by taking its absolute value, and then averaging the results.

$$\text{Mean absolute error (MAE)} = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Mean square error (MSE): MSE is widely used by all statisticians and also non-statisticians. It also gives much indication as to the size of the typical errors. The definition of MSE is the error made positive by squaring each one, and the squared errors are averaged. The MSE has the advantage of being easier to handle mathematically.

$$\text{Mean square error (MSE)} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

Mean percentage error (MPE): Before going to MPE, we need to define a relation or percentage error as

$$PE_t = \frac{Y_t - \hat{Y}_t}{Y_t} \times 100$$

It is error divided with original time series value and multiplied with '100'. Average of PE_t (percentage error) is mean percentage error.

$$\text{Mean percentage error (MPE)} = \frac{1}{n}\sum_{t=1}^{n} PE_t$$

Mean absolute percentage error (MAPE): Percentage errors are generally positive and also negative. But Absolute percentage errors are calculated by taking modulus to the percentage errors. On calculating average absolute percentage errors, we obtain mean absolute percentage error.

$$\text{Mean absolute percentage error (MAPE)} = \frac{1}{n}\sum_{t=1}^{n} |PE_t|$$

Where $e_t = Y_t - \hat{Y}_t$

$$PE_t = \frac{Y_t - \hat{Y}_t}{Y_t} \times 100$$

Where Y_t = original time series value at time 't'.

$\hat{Y}_t$ = estimated time series value at time 't'.

Empirical investigations: There are so many models generated using general ARMA (p, q) by changing order of 'p' and moving average order of 'q'. In this paper, between ARMA (2, 3) and ARMA (1, 3), which model is the best, estimated by using some statistical techniques. The fitted auto regressive moving average model ARMA (2, 3) for data is

$Y_{t1} = 52.5984 + 0.816\ Y_{t-1} - 0.36\ Y_{t-2} - 0.893\ e_{t-1} + 0.5363\ e_{t-2} - 0.216\ e_{t-3}$.

The another ARMA model, ARMA (1, 3) model fitted to data is

$Y_{t2} = 67.0945 + 0.6\ Y_{t-1} - 0.893\ e_{t-1} + 0.5363\ e_{t-2} - 0.216\ e_{t-3}$.

Goodness of fit: χ^2 goodness of fit is used to test both models, which fit is good for the model or not.

$\chi_1^2 = 0.9792$

The calculated χ^2 value for ARMA (2, 3) is 0.9792, which is less than χ^2 - table value 26.29 (i.e. χ^2; 16 degrees of freedom at 5% level of significance).

χ_1^2 Calculated (0.9792) < χ^2 Table value (26.29). Therefore, we conclude that ARMA (2, 3) is good fit for data.

$\chi_2^2 = 0.9643$

The above value is χ^2 calculated value for model ARMA (1, 3). χ_2^2 is less than χ^2 table value 26.29 (i.e. χ^2 ;16 degrees of freedom at 5% level of significance).

χ_2^2 Calculated (0.9643) < χ^2 Table value (26.29).

Therefore, we conclude that ARMA (1, 3) model is also good fit for data.

Both ARMA models ARMA (2, 3) and ARMA (1, 3) are good fit for data. Between ARMA (2, 3) and ARMA (1, 3), which model is the best is estimated by some statistical measures using error given in Table-I.

Table-I

Type of error	ARMA (2,3)	ARMA (1,3)
Mean error(ME)	-0.2231	-0.1.29
Mean absolute error (MAE)	5.7261	6.6143
Mean square error (MSE)	53.3035	58.0351
Mean percentage error (MPE)	0.0146	0.0233
Mean absolute percentage error (MAPE)	5.9480	6.8801

From the above statistical tools using error, we clearly observe that ARMA (2, 3) model is the best model than ARMA (1, 3) model. The descriptive statistics of the data is listed in Table-II.

Table-II

	Max	Min	Mean	SD	Skewness	Kurtosis	CV
Jan	31.3	5	12.85238	6.941441	1.125387	1.219187	54.00899
Feb	36.5	3.4	17.11905	9.931123	0.44579	-0.39212	58.01212
Mar	32.7	4.3	18.12381	8.98889	0.194234	-1.12036	49.59713
Apr	46.6	15.2	29.74286	8.816683	0.062843	-0.05288	29.64303
May	108.7	30.3	62.25714	24.40177	0.377103	-1.2722	39.19513
Jun	230.4	133.1	175.8524	24.58783	0.248636	0.280139	13.98208
Jul	361.8	133.9	282.6095	48.21319	-1.71833	5.705667	17.06
Aug	301.5	198	256.2429	36.1241	-0.63095	-1.17375	14.0976
Sep	237.3	115	172.1333	36.75615	0.039332	-1.1715	21.3533
Oct	131.3	43	84.58095	24.26023	0.112137	-0.41792	28.68286
Nov	64.5	15.4	31.65238	13.58257	0.887545	0.580874	42.91168
Dec	60.8	2.7	14.29048	13.61176	2.922121	10.12728	95.25055

3. SUMMARY AND CONCLUSIONS

In this paper the fitted two ARMA models are ARMA (2, 3) and ARMA (1, 3). The fitted ARMA (2, 3) model is a mixture of 2^{nd} order Auto regression with 3^{rd} order Moving average.

$Y_{t1} = 52.5984 + 0.816\, Y_{t-1} - 0.36\, Y_{t-2} - 0.893\, e_{t-1} + 0.5363\, e_{t-2} - 0.216\, e_{t-3}$.

ARMA (1, 3) model is a mixture of 1^{st} order Auto regression with 3^{rd} order Moving average.

$Y_{t2} = 67.0945 + 0.6\, Y_{t-1} - 0.893\, e_{t-1} + 0.5363\, e_{t-2} - 0.216\, e_{t-3}$.

χ^2 - test is used for testing the goodness of fit for the above two fitted models. χ_1^2 Value for ARMA (2, 3) is 0.9792, which is less than χ^2 table value. χ_2^2 Value for ARMA (1, 3) is 0.9643, which is less than χ^2 table value. Therefore, we conclude that both models i.e. ARMA (2, 3) and ARMA (1, 3) are fitted good for data. Between ARMA (2, 3) and ARMA (1, 3), the better model is estimated by some statistical measures using errors. The statistical measures using errors are ME, MAE, MSE, MPE and MAPE. ARMA (2, 3) model is the best model than ARMA (1, 3) model for rainfall data during the years 1990 to 2006.

4. REFERENCES

1. Md. Mahsin, Yesmin Akhter and Monira Begum, "Modeling rainfall in Dhaka Division of Bangladesh using Time series Analysis", Journal of Mathematical Modeling and Application, Vol. 1, No. 5, 2012, pp. 67-73.
2. D.A.Attah and G.M.Bankole, "Time series Analysis model for annual rainfall data in lower Kaduna Catchment Kaduna, Nigeria", International journal of research in Chemistry and Environment, Vol. 2, Issue 1, 2012, pp. 82-87.
3. S.Venkatramana Reddy, G.Karthick Kumar Reddy, B.Sarojamma and T.K.Ramakumar, "ARMA model fitted to rainfall data in India", Proc. of the International Conference on Mathematical Modeling and Applied Soft computing 2012, Vol. 2, 2012, pp. 229-236.
4. J.O.Pierini, B.Scian, M.Lovallo and L.Telesca, "Discriminating Climatological regimes in rainfall time series by using the Fisher-Shannon method", International Journal of the Physical Sciences, Vol. 6, 2011, pp. 7799-7804.
5. S.Makridakis, S.C.Wheelwright, and R.J.Hyndman, Forecasting: Methods and Applications, 3/e, Johnwiley & Sons (Asia) pte. Ltd., 2005.
6. Indian Institute of Tropical Meteorology data site address: http://www.tropmet.res.in/static_page.php?page_id=52.

****

----

[1]*Author 1, S. Venkatramana Reddy, Department of Physics,*
Sri Venkateswara University, Tirupati-517502, India. drsvreddy123@gmail.com
[2]*Author 2, G. Karthick Kumar Reddy, Department of Physics,*
Sri Venkateswara University, Tirupati-517502, India. guddetikarthick@yahoo.com
[3]*Author 3, S.C. Thasleema, Department of Statistics,*
Sri Venkateswara University, Tirupati-517502, India.
scthasleema@gmail.com
[4]*Author 4, T.K. Ramakumar, National Atmospheric Research Laboratory, Dept. of space, ISRO, Govt. of India, Gadanki-517112, India. Email:tkram@narl.gov.in*
[5]*Author 5, B. Sarojamma, Department of Statistics, Sri Venkateswara University, Tirupati-517502, India.*
**Author for correspondence Email: saroja14397@rediffmail.com*