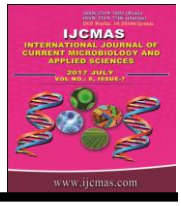
Int.J.Curr.Microbiol.App.Sci (2017) *6*(7): 2603-2610

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 6 Number 7 (2017) pp. 2603-2610**
Journal homepage: http://www.ijcmas.com



Original Research Article https://doi.org/10.20546/ijcmas.2017.607.307

Time Series Analysis of Monthly Rainfall for Gangetic West Bengal Using Box Jenkins SARIMA Modeling

G. Sathish[1*], Lakshmi Narasinhaiah[1], P. Mahesh Babu[2], Samrat Laha[3] and N. Bharath Kumar[4]

[1]Department of Agricultural Statistics, Faculty of Agriculture, India
[2]Department of Genetics and Plant Breeding, ANGRAU, Tirupati, Andhra Pradesh, India
[3]Department of Genetics and Plant Breeding, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur -741252, Nadia, West Bengal, India
[4]Scientist-B, Central Sericultural Reseach and Training Institute, Pampore, Jammu & Khasmir, India
**Corresponding author*

Keywords

SARIMA modeling, forecasting, Gangetic West Bengal, Rainfall.

Article Info

Accepted:
23 June 2017
Available Online:
10 July 2017

A B S T R A C T

West Bengal is a State where agriculture is mainly dependent on monsoon rainfall. But, erratic rainfall patterns cause a major negative impact on annual food-grain production. Time series forecasting has evolved as a major tool in different applications in hydrology and environmental management fields. The prediction of rainfall on time scales although scientifically challenging is nonetheless very important for decisive planning of agricultural strategies. In the present study, Box-Jenkins Seasonal ARIMA modeling was deployed in forecasting of monthly rainfall in Gangetic West Bengal up to 2020 based on data from 1960-2010 (a period of 50 years). The evaluation of validity of the performance of the selected model was carried out on the basis of the good-ness of fit (Chi-square), R^2 (coefficient of determination), RMSE (root mean square error), MAPE (mean absolute percentage error) and MAE (mean absolute error). The ARIMA model $(1, 1, 2)\ (0, 1, 1)^{12}$ fitted here was found to be most suitable for forecasting total monthly rainfall over the Gangetic West Bengal. This model is considered appropriate to forecast the monthly rainfall for the next ten years in the Gangetic West Bengal region to assist policy makers to establish priorities for water demand, storage, distribution and disaster management.

Introduction

The Lower Gangetic Plains has a gross cropped area of 6.96 m ha, out of which only 1.19 m ha is irrigated mainly by wells/tube wells, the rest being rain-fed. West Bengal, the leading producer of paddy and second largest producer of potato (30% of total potato production of the country), falls under this zone. Therefore, in purview of the importance and scarcity of available water resources, it has become imperative to study the behavior of Gangetic West Bengal rainfall, its fluctuations and its consistencies to forecast a fitting model for prediction of rainfall. Any modelling effort ought to be based on understanding of variability of past data (Mooley and Parthasarathy, 1984); Gregory (1989); Thapliyal (1990); Iyenger and Basak (1994); Iyenger and Raghukant (2003). A

general case of All India rainfall is documented in the works of (Gadgil *et al.*, 2002) and empirical modelling and forecasting has been presented by Sahai *et al.*, (2000); Yadav *et al.*, (2015); Pijush Basak (2016).

For many years, several efforts have been made to quantify the variability and forecast of monsoonal phenomenon at various temporal and spatial scales (Hartmann and Michelson, 1989; Ajay Mohan and Goswami, 2000).

The occurrence of rainfall over the Gangetic West Bengal state is an important phenomenon and its impact on the agriculture, economy and society is of profound significance.

Prediction of precise models, would help in implementation of advance policy formulation for efficient water management.

Materials and Methods

For the present study, the historical data for a period of fifty years (1960-2010) on mean monthly rainfall data was collected from the website www.indianwaterportal.com for 10 districts of Gangetic West Bengal region.

The time series data initially subjected for testing outliers using Grubb's test (Frank E. Grubbs, 1950) using R Studio package.

Grubb's test statistic:

$$G = \frac{\max \left| Y_t - \overline{Y} \right|}{s};$$

Where, $\overline{Y}$ is sample mean and s is standard deviation

To identify whether the time series is stationary or non-stationary, Augmented Dickey Fuller (ADF) test is carried out (David Dickey and Wayne Fuller, 1979).

$$ADF \text{ test equation}: \Delta z_t = \alpha + \beta t + \gamma z_{t-1} + \delta_1 \Delta z_{t-1} + \ldots\ldots + \delta_{p-1} \Delta z_{t-p+1} + a_t$$

Where α is constant, β is coefficient of time and p is log order of autoregressive process.

Subsequently the series has been tested for the presence of any auto-correlations using Ljung – Box Q test (G. M. Ljung; G. E. P. Box 1978), for which the test statistic is as follows

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-p}$$

Where, n is sample size, $\hat{\rho}_k^2$ is sample autocorrelation at p^{th} lag and h is the number of lags being test.

Most real time series show a trend. An average increase or decrease over time which means that they are Φ non-stationary i.e., they are integrated. Series also show cyclic behavior. Trends and cycles can be removed from a series through differencing. By differencing several times and/or at different lags, most series can be converted to a stationary series and then ARMA model for w_t is applied. Thus, the combined model for the original univariate time series, which involves auto-regression, moving average, and integration, is termed as *ARIMA(p, d, q)* model (model of orders *p, d,* and *q*) with *p* AR terms, *d* differences, and *q* MA terms.

The ARIMA model is often a parsimonious description of the behavior of a series. Given a set of time series data, one can calculate the mean, variance, autocorrelation function

(ACF), and partial autocorrelation function (PACF) of the time series.

The ARMA model

In general, we can combine the seasonal and non-seasonal operators into a multiplicative seasonal autoregressive moving average (MA) model, denoted by ARMA (p, q) × (P, Q)^s, and write as the overall model.

$$\Phi_P\left(B^s\right)\phi\left(B\right)x_t = \Theta_Q\left(B^s\right)\theta\left(B\right)w_t$$

……. Equation (1)

SARIMA models

The multiplicative seasonal autoregressive integrated MA model, or SARIMA model, of Box and Jenkins (1970) is given by:

$$\Phi_P\left(B^s\right)\phi\left(B\right)\nabla_s^D\nabla^d x_t = \alpha + \Theta_Q\left(B^s\right)\theta\left(B\right)w_t$$

……. Equation (2)

Where w_t is the usual Gaussian white noise process. The general model is denoted as SARIMA (p,d,q) × (P,D,Q)^s. The ordinary autoregressive and MA components are represented by polynomials ϕ (B) and θ(B) of orders p and q, respectively (Equation 1), and the seasonal autoregressive and MA components by Φ_P (B^s) and Θ_Q(B^s) (Equation 2), of orders P and Q and ordinary and seasonal difference components by ∇^d =(1-B)^d and ∇_s^D =(1 – B^s)^D.

Based on the nature of the above appropriate ARIMA models are worked out, but the final decision is made once the model is identified and diagnosed. In this step one can see whether the chosen model fits the data reasonably well or not. One simple test of the chosen model is to see if the residuals estimated from this model shows white noise; if so, one can accept the particular fit, only after iterative processing through Box Jenkins Methodology; if not, one can start the process afresh. Models are compared according to the minimum values of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) and maximum value of coefficient of determination (R^2).

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n}}$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|}{n} \times 100$$

$$BIC = -2 * \ln(L) + k * \ln(n)$$

Where Y_i, $\bar{Y}$ and $\hat{Y}_i$ are the values of the ith observation, mean and estimated values of the ith observation of the variable Y and k is the number of parameters in the statistical model, L is the maximized value of the likelihood function for the estimated model.

Results and Discussion

The non-significance of the Grubb's test indicates that there are outliers present in the data. The observed outliers are replaced with median value (Lukasz Komsta, 2006). The next step indicates that the data sets were non-stationary in nature (in terms of both mean and variance) and they reflected seasonal cycles. This was confirmed when the ACF and PACF plots of the original data were

prior to any transformation and differencing, they were obtained. In order to fit an ARIMA model, a stationary series (in terms of both in mean and variance) is needed. To establish the stationarity of the variance of the time series, first difference (d=1) of the original data was done in order to establish stationarity in the series with no non-seasonal impact. Stationarity of the mean could be attained by differencing the series. However, for SARIMA, first difference (D=1) of the original data was done in order to establish stationarity in the series with no seasonal impact.

The ACF and PACF plots for the differenced series were obtained again to check the stationary (Figure 2). The figure confirms that the ACF and PACF plots for the differenced and de-seasonalized rainfall data were nearly stable and the SARIMA model (p, 1, q) (P, 1, Q)12 could be identified for further analysis (Hillmer and Tiao, 1982).

In the next step, model parameters p, q, P and Q were identified. The ACF and PACF plots of the SARIMA model (p,0,q)(P, 1, Q)12 with first order seasonal differencing (Figure 2) suggested that at the initial stage the tentative model should be (1,1,1)(1,1,1)12 because there was one auto-regressive and one MA parameter in the plots. We found similar condition for data set of total Gangetic West Bengal. In SARIMA modeling it is necessary to minimize the residual sum of squares (RSS) between the actual and predicted values to represent the data most appropriately. The criteria for choosing the best SARIMA model should have the least number of parameters to acquire the minimum AIC along with the minimum RSS. Therefore, in the stage of identifying the number of autoregressive and MA parameters, an SARIMA model (p, 1, q) (P, 1, Q)12 with the least number of parameters was attempted. We evaluated different SARIMA models to obtain the best model among them. The performance evaluations and validity of selected models were carried out on the basis of the diagnostic values was considered as best model for Gangetic West Bengal.

As mentioned earlier, monthly rainfall data of total Gangetic West Bengal 50 years (1960-2010) data were used for model calibration and for the years 2010-2020 were used for forecasting.

The AIC values for best fitted model were estimated using Equation (4); the SARIMA model (1, 1, 2) $(0,1,1)^{12}$ provided the best results (Table 1) for the study area under study.

Fig.1 Observed, fitted and forecasted values for Gangetic West Bengal

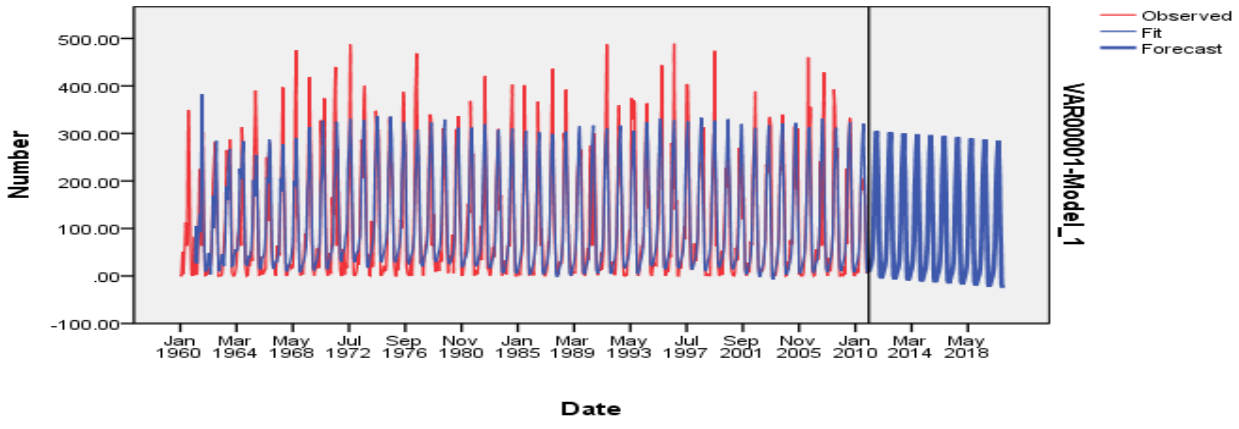


Fig.2 ACF and PACF residuals plot for Gangetic West Bengal

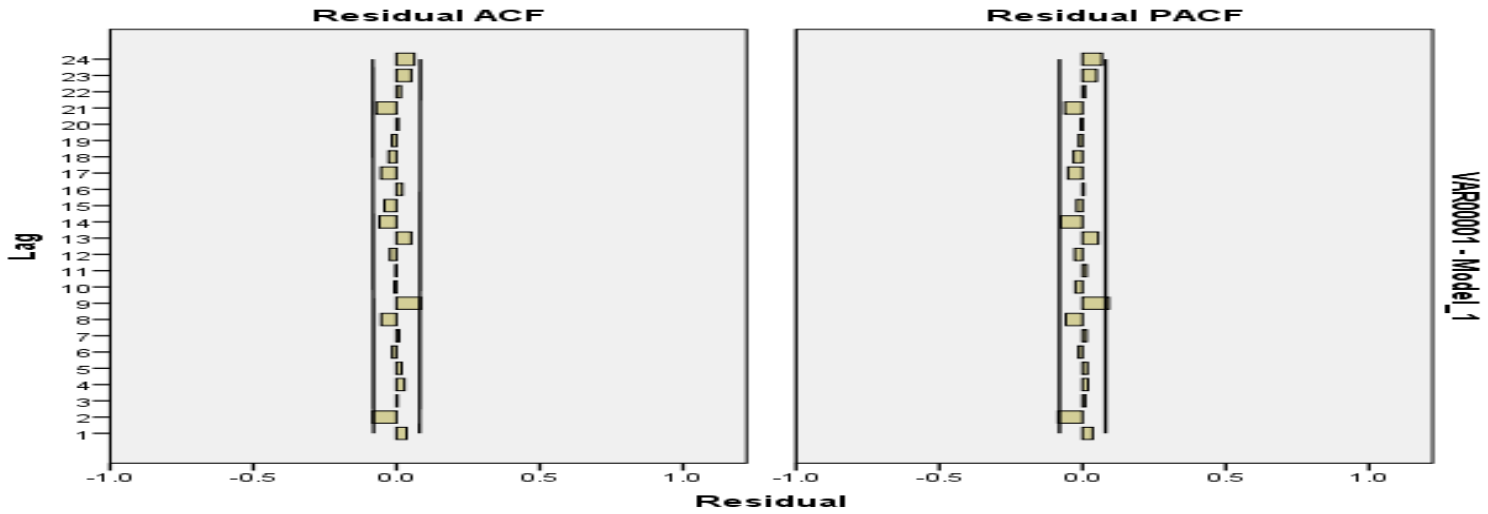


Fig.3 SEASONAL ARIMA (1, 1, 2)(0,1,1)^{12} model validation for two years (2008-2010)

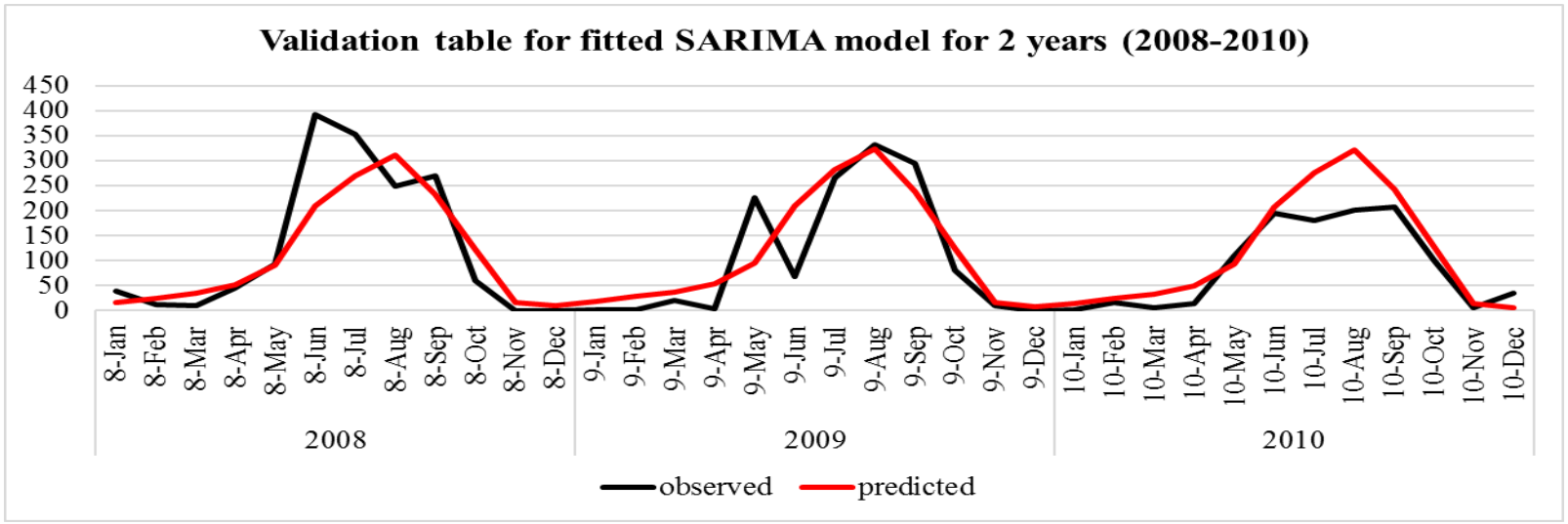


Table.1 SEASONAL ARIMA (1, 1, 2)(0,1,1)^{12} model parameters

ARIMA Model Parameters					
		Estimate	**SE**	**T**	**Sig.**
Constant		-.007	.005	-1.417	.157
AR	Lag 1	.771	.087	8.874	.000
Difference		1			
MA	Lag 1	1.813	.070	25.920	.000
	Lag 2	-.818	.068	-12.124	.000
Seasonal Difference		1			
MA, Seasonal	Lag 1	.996	.227	4.392	.000

Table.2 SEASONAL ARIMA (1, 1, 2)(0,1,1)^{12} model Diagnostics

Model Statistics												
Model	**Number of Predictors**	**Model Fit statistics**							**Ljung-Box Q(18)**			**Number of Outliers**
		Stationary R2	**R^2**	**RMSE**	**MAPE**	**MAE**	**MaxAPE**	**Normalized BIC**	**Statistics**	**DF**	**Sig.**	
VAR00001-Model_1	0	.741	.742	63.449	1025.166	42.888	62120.624	8.354	19.744	14	.138	0

As discussed earlier, the SARIMA (1, 1, 2)(0,1,1)^{12} model could be written in the following form (Equation Write according to fitted model)

$$(1)(1-0.77B)\nabla^{12}\nabla\hat{x}_t = (1-0.996B^{12})(1-1.81B-0.818)\hat{w}_t$$

……...Equation (4)

The goodness of fit of the SARIMA model (1, 1, 2)(0,1,1)^{12} was tested using the Ljung-Box statistic as shown in Equation (4). The goodness of fit values for the autocorrelations of residuals from the (1, 1, 2)(0,1,1)^{12} model up to lag 24 was ≥0.05 for the study period. These results substantiate the acceptance of the null hypothesis of model adequacy at the 5% significance level and the set of autocorrelations of residuals was considered white noise.

The SARIMA model (1, 1, 2) (0,1,1)^{12} was also tested for its validity to forecast was made up to 2020. The fitted was model checked for validation for two years (2008-2010), validation results obtained using the model are shown in (Figure 3). The observed mean rainfall was found to be closely aligned to the forecasted values of the mean rainfall for the Gangetic West Bengal shown in (Figure 1). From the results presented in this study, it is apparent that the chosen model should be sufficiently accurate to forecast rainfall in this region.

Various Statistical diagnostic measures such as R^2 (0.742), root mean square error (63.44), mean absolute percentage error (1025.16), MAE (42.88), minimum of BIC value (8.35) are given in (Table 2).

Time series analysis is an important tool in modeling and forecasting rainfall data. In this study we used the SARIMA model to simulate and forecast mean rainfall for total Gangetic West Bengal. The SARIMA model (1, 1, 2)(0,1,1)^{12} was developed considering step-wise analysis, non-seasonal and seasonal parameters, and various diagnostic checks. Interestingly, SARIMA model (1, 1, 2)(0,1,1)^{12} fitted. The forecasting results for the upcoming 10 years are considered to be precise and accurate. This will certainly assist policy makers and decision makers in planning for any kind of disaster or extreme condition in every district town of the Gangetic West Bengal by generating scenarios for the next few years.

References

Ajay Mohan RS and Goswami BN (2000). A common spatial mode for intra-seasonal and inter- variation and predictability of the Indian summer monsoon. Current Science, 79(8): 1106-1111.

Dickey, D. A.; Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root". Journal of the American Statistical Association. 74 (366a): 427–431. doi:10.1080/01621459.1979.10482531

Gadgil S, Srinivasan J, Nanjundiah RS, Krishna Kumar K, Munot AA and Rupa Kumar K (2002). On forecasting the Indian summer monsoon: the intriguing season of 2002. Current Science, 83(4): 394-403.

Gregory S (1989). Macro-regional definition and characteristics of Indian summer monsoon rainfall, 1871-1975. International Journal of Climatology, 9:465-483.

Grubbs, Frank E. (1950). "Sample criteria for testing outlying observations". Annals of Mathematical Statistics. 21 (1): 27–58. doi:10.1214/aoms/1177729885.

Hartmann, DL; Michelson, ML. (1989). Intraseasonal periodicities in Indian rainfall. Journal of the Atmospheric Science, 46: 2838-2862.

Hillmer, S. C.; Tiao, G. C. (1982) Journal of

the American Statistical Association Vol. 77, Iss. 377.

Iyenger, RN; Basak, P. (1994). Regionalization of Indian monsoon rainfall and long-term variability signals. International Journal Climatology, 14: 1095-1114.

Iyenger, RN; Raghukant, STG. (2003). Empirical modeling and forecasting of Indian monsoon rainfall. Current Science, 85(8): 1189-1201.

Ljung; G. M.; Box (1978), G. E. P. "On a Measure of a Lack of Fit in Time Series Models". Biometrika. 65 (2): 297–303. doi:10.1093/biomet/65.2.297

Lukasz Komsta, (2006) ported from R package "outliers". See R News, 6(2):10-13.

Mooley DA and Parthasarathy B (1984). Fluctuations in all- India summer monsoon rainfall during 1871-1978. Climatic Change, 6(3): 287-301.

Pijush Basak (2016). Monsoonal rainfall of Gangetic West Bengal:Empirical Modelling and Forecasting. International Journal of Earth and Atmospheric Science, Vol 3 Issue3:57-62.

Sahai, AK; Soman, MK; Satyen, V. (2000). All India summer monsoon rainfall prediction using an artificial neural network. Climate Dynamics, 16(4): 291-302.

Thapliyal V (1990). Long range prediction of summer monsoon rainfall over India: Evaluation and development of new models. Mausam, 41: 334-346.

Yadav BP, Naresh Kumar and Tomar CS (2015). Analysis of heavy precipitation events over western Himalayan region. International Journal of Earth and Atmospheric Science, 2(3):90-96.

How to cite this article:

Sathish, G., Lakshmi Narasinhaiah, P. Mahesh Babu, Samrat Laha and Bharath Kumar, N. 2017. Time Series Analysis of Monthly Rainfall for Gangetic West Bengal Using Box Jenkins SARIMA Modeling. *Int.J.Curr.Microbiol.App.Sci.* 6(7): 2603-2610.
doi: https://doi.org/10.20546/ijcmas.2017.607.307