The Andhra Agric. J 60(2): 405-408, 2013

Forecasting of Guntur District Rainfall by using ARIMA models

D Ramesh, V Srinivasa Rao, D V S Rao and G Raghunadha Reddy
Department of Statistics & Mathematics, Agricultural College, Bapatla 522 101, Guntur Dist, A P

ABSTRACT

This paper attempted to the Auto Regressive Integrated Moving Average (ARIMA) models for the rainfall data of Guntur district of Andhra Pradesh . The data was collected from the Office of Chief Planning Officer, Guntur for monthly rainfall data covering the years 1980-2011. ARIMA (0,1,1) was identified as best model based on criteria like Akaike information criterion (AIC), Schwarz Bayesian information criterion (SBC), Mean Absolute Percent Error (MAPE) and parameter estimation was done to the best model ARIMA (0,1,1). And also an attempt was made to forecast the rainfall up to 2015.

Key words : ARIMA=Auto Regressive Integrated Moving Average, ACF=Auto Correlation Function, PACF=Partial Auto Correlation Function. AIC=Akaike information criterion, SBC=Schwarz Bayesian information criterion, MAPE=Mean Absolute Percent Error.

In India, about 70% of the population is being dependent on agriculture and allied sectors. Indian agriculture is predominantly a rain fed agriculture, hence the study of rainfall pattern and estimation of parameter for suitable stochastic model is very much useful to predict the rainfall.

Lot of work was done on forecasting by using ARIMA models on various crops and various regions. Forecast of paddy in Tamilnadu and food grains in India (Balasubramanian and Dhanavanthan, 2002). Potato price forecasting using seasonal ARIMA approach (Chandran *et al*. 2007). Forecasting of exports of industrial goods from Punjab-an application of univariate ARIMA approach (Kumar and Gupta, 2010). Forecasting the area, production and productivity of Sugarcane in Tamilnadu using ARIMA models (Suresh and Priya, 2011).

This study was confined to Guntur district of Andhra Pradesh since the district enjoys the benefit of both South-West monsoon (55.33%) and North-East monsoon (19.53%). The major crops grown in the district are Paddy, Blackgram, Chillies, Cotton and Tobacco. Most of the crop yields are much depend on the rainfall (Hand Book of Statistics-2010).

Development of suitable model and its parameter estimation for forecasting of the rainfall plays an important role particularly in the regions where most of the cropped area is rainfed and for management practices like transplantation and harvest are very important for the farmers.

MATERIAL AND METHODS

The Rainfall data was collected from the Office of Chief Planning Officer, Guntur for monthly rainfall data covering the years 1980-2010. By using SAS 9.3 software fitted the models i.e. ARIMA (1,1,0), ARIMA (2,1,0), ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (1,1,2) and ARIMA (2,1,2). The best model was identified based on the criterions i.e., lowest AIC, SBC and MAPE. After selection of best model, estimation of parameters was done for prediction of rainfall.

Description of the Model:

Auto Regressive Integrated Moving Average (ARIMA) model was introduced by Box and Jenkins. In general, an ARIMA model is characterized by the notation ARIMA (p, d, q) where p, d, q denote orders of auto-regression, integration (differencing) and moving average respectively. In ARIMA, time series is a liner function of past actual values and random shocks. A stationary ARIMA (p, d, q) process is defined by the equation

$$Yt = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \ldots - \omega_q \varepsilon_{t-q} + \varepsilon_t$$

Where Yt is the response (dependent) variable at time t. Y_{t-1}, Y_{t-2}, Y_{t-p} are the response (dependent) variable at lags t-1, t-2….t-q. An assumption about error term is same as standard regression model. $\omega_1, \omega_2, \ldots \omega_q$ are the coefficients to be estimated.

Model Identification:

Stationary:

A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the disturbance or lag between the two time periods and not on the actual time at which the covariance is computed (Damodar Gujarati, 1995).

There are several ways to ascertain this. The most common method is to check stationary through examining the graph or time plot of the data. If the data is Non-stationary, it will corrected through appropriate differencing of the data. The newly constructed differenced variable can now be examined for stationarity. The graph of differenced variable was stationary.

The next step in the identification process is to find the initial values for the orders of non-seasonal parameters, p and q. They are obtained by looking for significant autocorrelation and partial autocorrelation coefficients from ACF and PACF charts.

Estimation:

At the identification stage, one or more models are chosen that seem to provide statistically adequate representations of the available data. Then precise estimates of parameters of the model are obtained by least-squares. Standard computer package SAS 9.3 etc... is available for finding the estimates of relevant parameters using iterative procedures.

Diagnostic checking:

- **Akaike information criterion (AIC) :** The Akaike information criterion is a measure of the relative goodness of fit of a statistical model. AIC values provide a means for model selection. AIC can tell nothing about how well a model fits the data in an absolute sense. In general case, the AIC is

  $$AIC = 2k - 2 \ln(L)$$

where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

- **Schwarz Bayesian information criterion (SBC):** In statistics, the Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models. It is based, on the likelihood function, and it is closely related to Akaike information criterion (AIC).

The formula for the BIC is

$$BIC = -2 \log(L) + m \log n$$

L is the likelihood of the data with a certain model

n is the number of observations

m is the number of parameters in the model

- **Mean Absolute Percentage Error (MAPE) :**

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

$$M = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

where At is the actual value and Ft is the forecast value.

RESULTS AND DISCUSSION

It was observed from the rainfall data of Guntur district during the period 1980-2011, highest average rainfall was observed in the month of August (162.65mm), lowest average rainfall was in the month of January (7.46mm), highest average rainfall was observed in the year of 2010 (1460.34mm), lowest average rainfall was in the year of 2000 (569.64mm), highest variability in the rainfall was observed in the year of 2000 and the lowest variability was observed in the year of 2002. By using ARIMA models, it was observed that the time series data on Rainfall was not stationary during the period. It was made stationary by taking the first order differencing technique. The ACF and PACF graphs were used to fit the models. The appropriate model was chosen based on AIC, SBC and MAPE criterion.

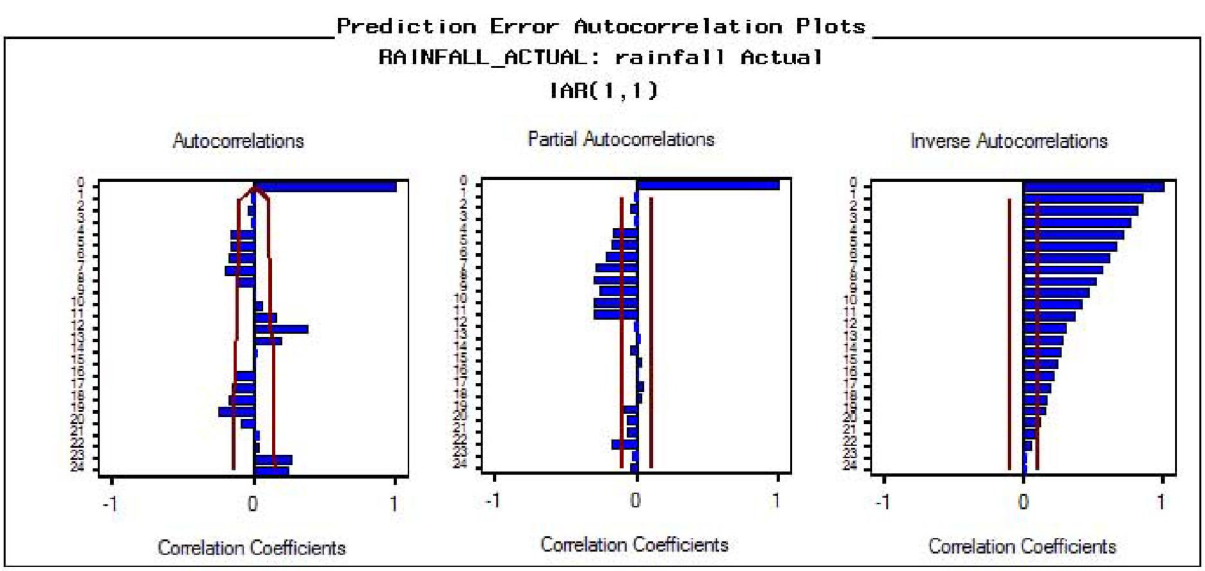
From, the Table-1 it was observed that ARIMA (0,1,1) model was selected as best model because of its lowest AIC, SBC and MAPE values. So, parameters was estimated to ARIMA (0,1,1) model (shown in Table-2), which plays an important role in prediction of rainfall. The ACF and PACF graphs were shown in Fig.2.

Table-1

	MAPE	AIC	SBC
ARIMA (1,1,0)	1687.3	3257.7	3265.5
ARIMA (2,1,0)	1738.3	3259.0	3270.8
ARIMA (0,1,1)	**1587.3**	**3256.7**	**3264.6**
ARIMA (0,1,2)	1716.7	3258.6	3270.4
ARIMA (1,1,2)	1596.8	3260.7	3276.3
ARIMA (2,1,2)	1849.4	3262.6	3282.2

Table-2 Parameter Estimates

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.33875	2.9129	0.1163	0.9075
Moving Average, Lag 1	0.30247	0.0496	6.0933	<.0001
Model Variance (sigma squared)	6457	.	.	.



The graphs of sample ACF and PACFs reveled that the spikes cutoff after lag1 both in ACF and PACF plots. Hence the value of p = 0 and q = 1 . The ARIMA (0,1,1) reveals that the rainfall is not showing any trend and it is difficult to forecast the future rainfall with high accuracy. It is highly random. However, using ARIMA (0,1,1) the rainfall was predicted upto 2015 as shown in the table (3).

Table 3. Forecasting values for Average annual rainfall data by ARIMA (0,1,1) model.

Year	Actual Rainfall (mm)	Forecasted Rainfall (mm)
2001	73.8089	74.0011
2002	47.4699	73.0863
2003	76.0751	80.0353
2004	62.9403	61.5462
2005	81.9341	64.2732
2006	70.773	70.7926
2007	81.0495	74.7219
2008	82.3487	77.7808
2009	47.9603	77.8742
2010	121.6951	83.9028
2011	54.5083	65.3677
2012	.	88.8707
2013	.	88.1214
2014	.	74.1597
2015	.	90.4178

Table-2 reveals that for any given period of time the average assured rainfall was 0.33mm with a standard error of 2.92. This model reveals that the rainfall depends only on the previous year rainfall only.

LITERATURE CITED

Balasubramanian P and Dhanavanthan P 2002 Seasonal modeling and forecasting of crop production. *Statistics and Applications.* 4 (2): 107–118.

Chandran K P and Pandey N K 2007 Potato price forecasting using seasonal ARIMA approach. *Potato Journal*. 34(1-2): 137-138.

Damodar G 1995 *Basic Econometrics*, Mc Graw Hill, Inc. New York.

Hand Book of Statistics Guntur District 2010.

Kumar G and Gupta S 2010 Forecasting exports of industrial goods from Punjab-An application of univariate ARIMA model. *Annals of university of Petrosani*, Economics. 10(4): 169-180.

Suresh K K and Priya K 2011 Forecasting Sugarcane yields of Tamilnadu using ARIMA models. *Sugar tech*. 13(1): 23-26.

(Received on 24.05.2012 and revised on 29.10.2012)