The Pharma Innovation Journal 2022; SP-11(4): 230-235

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(4): 230-235

www.thepharmajournal.com
Received: 01-02-2022
Accepted: 04-03-2022

Pooja Rawat
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Sanju
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Corresponding Author
Pooja Rawat
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Forecasting food grains yield in Haryana: A time series approach

Pooja Rawat and Sanju

Abstract
The present study focuses on the forecasting of food grains yield in Haryana using autoregressive integrated moving average (ARIMA) technique. The annual data of Haryana on food grains yield were divided into the training data set from 1966-67 to 2015-16 and the testing data set from 2016-17 to 2019-20. Box Jenkins approach was used for model identification, the best possible ARIMA model is fitted among various competing models. Goodness of fit of each of these models is tested and the best model is used to forecast the yield of food grains. It is inferred that ARIMA (0,1,1) was found to be optimal and that the forecast values for the years 2020-21 to 2023-24 were estimated on the basis of this model which were 4.035, 4.094, 4.154 and 4.213 million tonnes respectively. The performances of these models validated with the R^2, RMSE, MAPE, MAPE and BIC. The forecasted values and percentage relative deviation comes within acceptable limits. Forecasted values are very useful for the policymakers and government agencies for proper policy decision regarding food security.

Keywords: ARIMA, forecasting, stationary, ljung-box, ACF, PACF

Introduction

Food grains production is the major activity in agriculture, covering about 80 percent of the cropped area in India and providing the main source of food. India is the second largest producer of food grains in the world. India is also gifted with the diverse agro climatic conditions suitable for production of many important food grain crops such as wheat, rice and bajra. Government policy in India has always given substantial importance to food grains production. Such support, particularly since the beginning of the green revolution in the mid-1960s, has contributed to noteworthy growth in this sector despite many constraints. Now-a-days India has become self-sufficient in food grains, as its production has reached to a record level of 308.65 million tonnes in 2020-21. Haryana is the second largest contributor to India's central pool of food grains. Haryana is among the top ten producers of food grains and stands at sixth place with a total production of 16.38 million tonnes from an area of 4.47 million hectares. Of total food grains produced in the state, contribution of wheat and rice was 11.3 and 4.15 million tonnes respectively. A number of studies have been undertaken to understand the growth pattern of food grains in Haryana using the statistical tool and techniques ranging from graphs, charts, and diagrams to regression analysis and time series analysis.

Forecasting helps the decision-makers to make their future decisions more correctly, whether from the economic or non-economic fields. Yield forecasting plays an important role in farming planning and management, domestic food supply, international food trade, ecosystem sustainability, and so on[5]. Forecasting production behaviors of the major crops plays an important role for policy makers to ensure food and nutrition security. The policy makers should have ideas about potential production scenarios of the major crops [10].

Verma *et al*. (2011) [11] worked on wheat, sugarcane, cotton and mustard crops for operational yield forecasting purpose under Crop Acreage Production Estimation (CAPE) project. Mishra *et al*. (2015) [9] studied the stability in production behavior and forecasted food grain production in India using ARIMA model. Chaudhuri *et al*. (2020) [2] used ARIMA model for forecasting of rice production in ten the highest rice harvesting states of the country. Kumar *et al*. (2021)[6] used univariate forecasting models such as random walk, random walk with drift, moving average, simple exponential smoothing and Autoregressive Integrated Moving Average (ARIMA) models for forecasting vegetable production in Haryana state. They concluded that ARIMA (2, 1, 1) was found to be optimal. Kumar *et al*. (2021) [7] fitted the ARIMA model using wheat yield data of Hisar for the period 1966-67 to 2018-19. The study revealed that ARIMA (0,1,1) model comes out to be the best model for forecasting wheat yield in Hisar.

Data description and methodology

The Haryana state comprised of 22 districts is situated between 74° 25' to 77° 38' E longitude and 27° 40' to 30° 55' N latitude. The total geographical area of Haryana is 44212 sq. km. In the present study secondary information of Haryana food grains yield gathered from the state Department of Agriculture and Farmers Welfare for the period 1966-67 to 2019-20. The validity of fitted model has been checked for testing data SETI 2016-17 to 2019-20, which is not included in the development of the models.

Box-Jenkins modeling procedure

Autoregressive integrated moving average method popularized by George Box and Gwilym Jenkins (1970) [1] is applied to time series analysis and forecasting. ARIMA forecasts are based only on the preceding value of the predicted variable. If the mean, variance and autocorrelation function is constant over time then a time series is said to be stationary and a non-stationary time series can be converted into stationary time series through differencing. The Box-Jenkins ARIMA modeling for forecasting the future value of a time series follows 3 steps iterative procedure i.e. model identification, Estimation of parameters and checking the residual diagnostics [3, 4].

The most common and widely used stochastic model for time series forecasting is the ARIMA model. For a given, non-seasonal, non-stationary time series Xt having trend, the ARIMA model can be written as:

$$\phi (B) (1-B)dXt = \theta (B) \varepsilon t$$

Where,

$\phi(B) = 1 - \phi1B - \phi2\ B2 - \ldots - \phi p\ Bp$ (Autoregressive parameters)

$\theta (B) = 1 - \theta\ 1\ B - \theta2\ B2 - \ldots - \theta\ q\ Bq$ (Moving average parameters)

d - Order of differencing

B - Backshift operator

εt - White noise or error term

Auto-correlation function (ACF) and partial autocorrelation function (PACF) of a time series are plotted to understand the correlations between observations at different time lags. By looking at these two plots together can help us form an idea of what models to fit.

In order to select the best model among the various combination of models, measures like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Bayesian Information Criteria (BIC), estimated standard deviation Ljung Box test is also used to check goodness of fit. It can be applied to original series or to the residual after fitting a model.

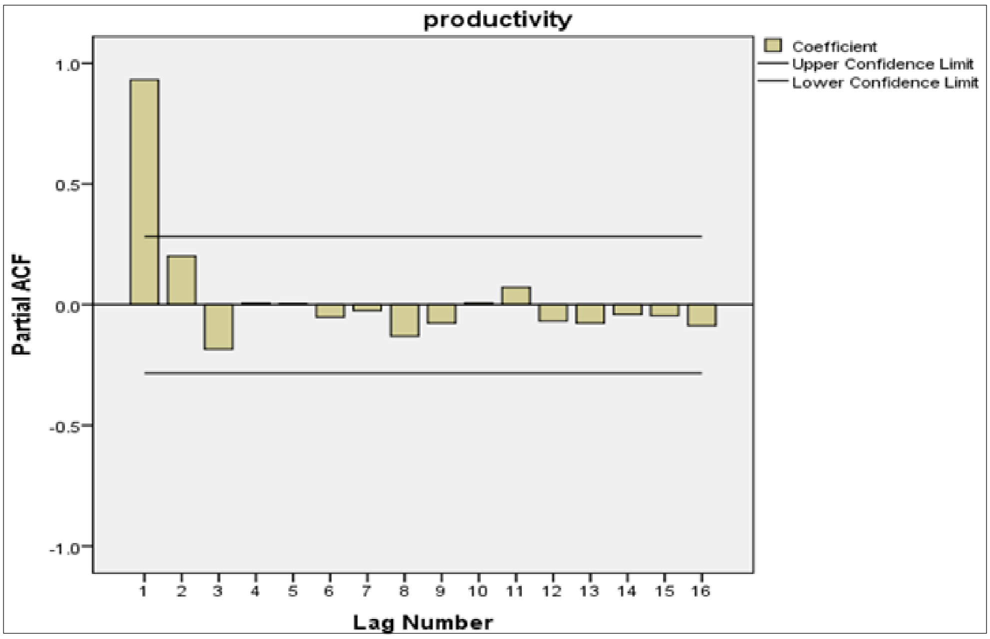
Result and Discussion

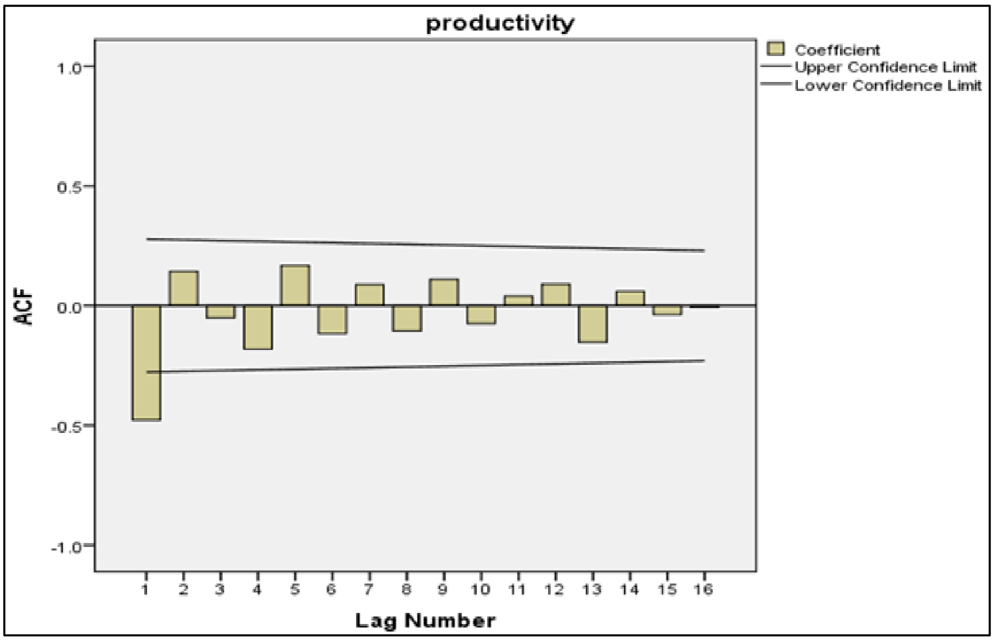
Identification of order for autoregressive and moving average polynomial

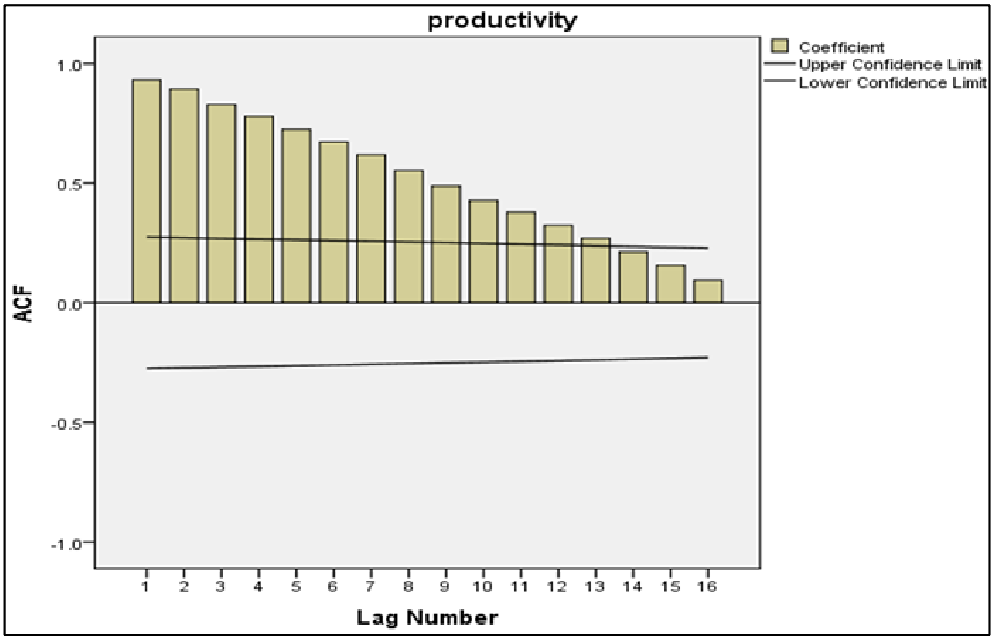
Identification involves the determination of the appropriate order of AR and MA polynomials i.e. values for p and q. The orders were determined from the ACF and PACF of the stationary time series. The food grains yield data was found to be non-stationary from the Table 1 and Figure 1. Almost all the autocorrelations upto (n/4) Th lags significantly different from zero confirmed non-stationarity. The plotting of the ACFs (Figure 1) also indicates that the ACFs decline gradually implying non-stationarity. Thus, the crop yield series considered here were found to be non-stationary. The non-stationary data series was transformed into stationary series by the first differencing of the original data series. Differencing of order one i.e. d=1 was enough for getting an approximate stationary series.

Table 1: Autocorrelation for Food grains yield

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	DF	Sig.
1	.931	.137	46.014	1	.000
2	.894	.136	89.304	2	.000
3	.829	.134	127.312	3	.000
4	.779	.133	161.597	4	.000
5	.725	.132	191.948	5	.000
6	.671	.130	218.580	6	.000
7	.619	.129	241.741	7	.000
8	.553	.127	260.704	8	.000
9	.489	.126	275.881	9	.000
10	.427	.124	287.748	10	.000
11	.379	.122	297.317	11	.000
12	.323	.121	304.472	12	.000
13	.270	.119	309.580	13	.000
14	.213	.118	312.863	14	.000
15	.156	.116	314.673	15	.000
16	.094	.114	315.353	16	.000


productivity
Coefficient
Upper Confidence Limit
Lower Confidence Limit
Partial ACF
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Lag Number


productivity
Coefficient
Upper Confidence Limit
Lower Confidence Limit
ACF
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Lag Number


productivity
Coefficient
Upper Confidence Limit
Lower Confidence Limit
ACF
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Lag Number

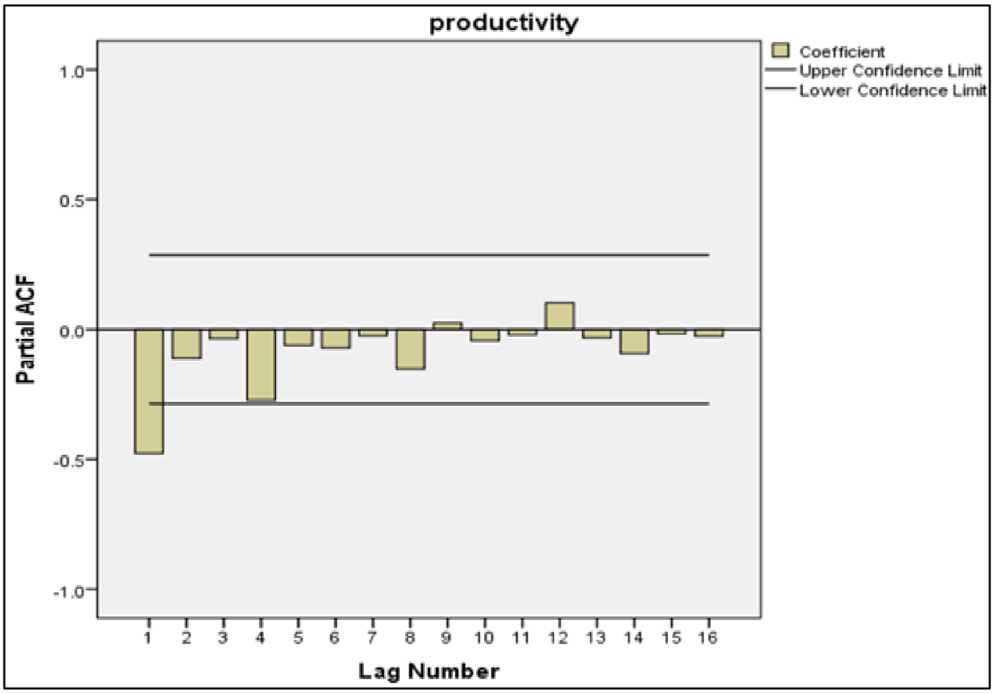


Fig 2: Autocorrelation and Partial autocorrelation for Foodgrains yield after first differencing

Parameter Estimates

After experimenting with different lags of the moving average and autoregressive processes; ARIMA(0,1,1), ARIMA(1,1,1), ARIMA(1,1,0), ARIMA(2,1,0) and ARIMA(2,1,1) were fitted for estimating the food grains yield. The models ARIMA (1, 1, 0) and ARIMA (0, 1, 1) were considered in the identification stage and ARIMA estimation was carried out using least square approach. Parameter estimates of the selected models shown in Table 2 are less than one (needed for convergence) and also satisfy the stationarity and inevitability conditions under ARIMA structure.

Table 2: Tentative ARIMA models for Food grains yield

Model		Estimate	Std. Error	Approx. Prob.
ARIMA (1,1,0)	AR(1)	-0.491	0.128	<0.01
ARIMA (0,1,1)	MA(1)	0.618	0.124	<0.01
ARIMA (1,1,1)	AR(1)	0.130	0.232	0.579
	MA(1)	0.724	0.174	<0.01
ARIMA (2,1,0)	AR(1)	-0.529	0.147	<0.01
	AR(2)	-0.088	0.161	0.586
ARIMA (2,1,1)	AR(1)	0.326	0.182	0.79
	AR(2)	0.277	0.177	0.124
	MA(1)	0.995	1.274	0.439

By comparing approx. probability in table 2 ARIMA (1, 1, 0) and ARIMA (0, 1, 1) has been selected since these are comes out to be highly significant.

Table 3: Selection Criteria vales for choosing ARIMA models

Models	R-squared	RMSE	MAPE	MAE	BIC
ARIMA (1,1,0)	0.969	166.542	6.833	128.970	10.389
ARIMA (0,1,1)	0.970	164.228	6.928	130.825	10.361

On the basis of R^2, RMSE, MAPE, MAE and BIC given in the above Table 3 ARIMA (0, 1, 1) was found to be the best fit for food grains yield estimation.

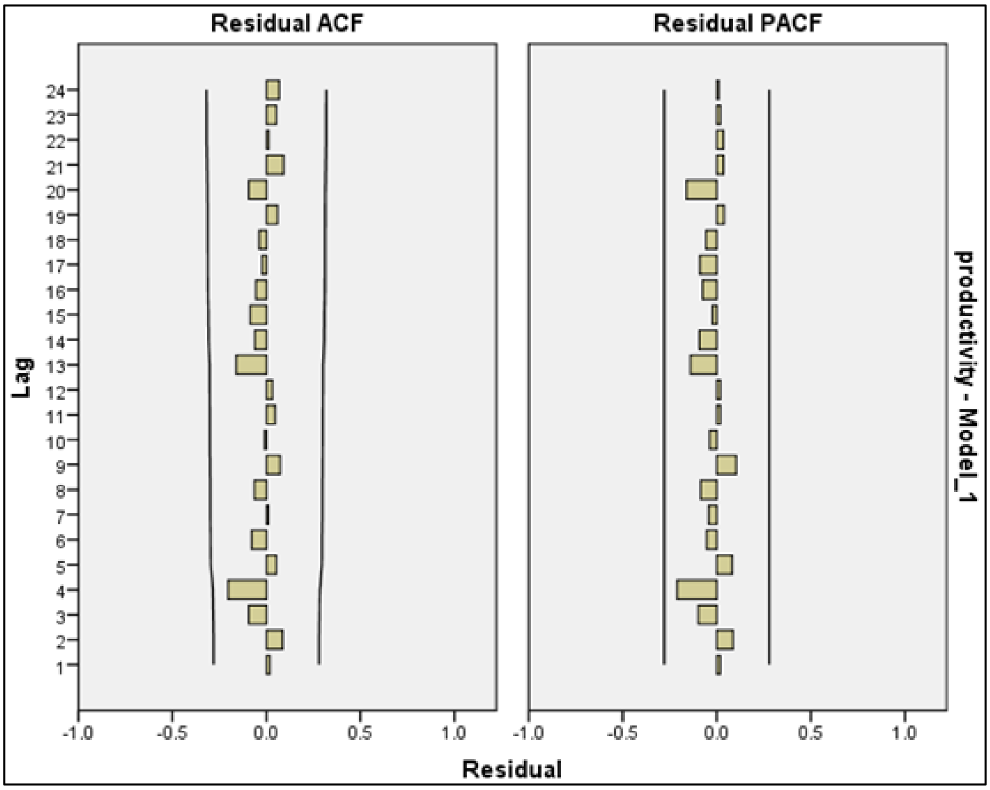


Fig 3: Residual ACF and PACF plot based on fitted ARIMA model

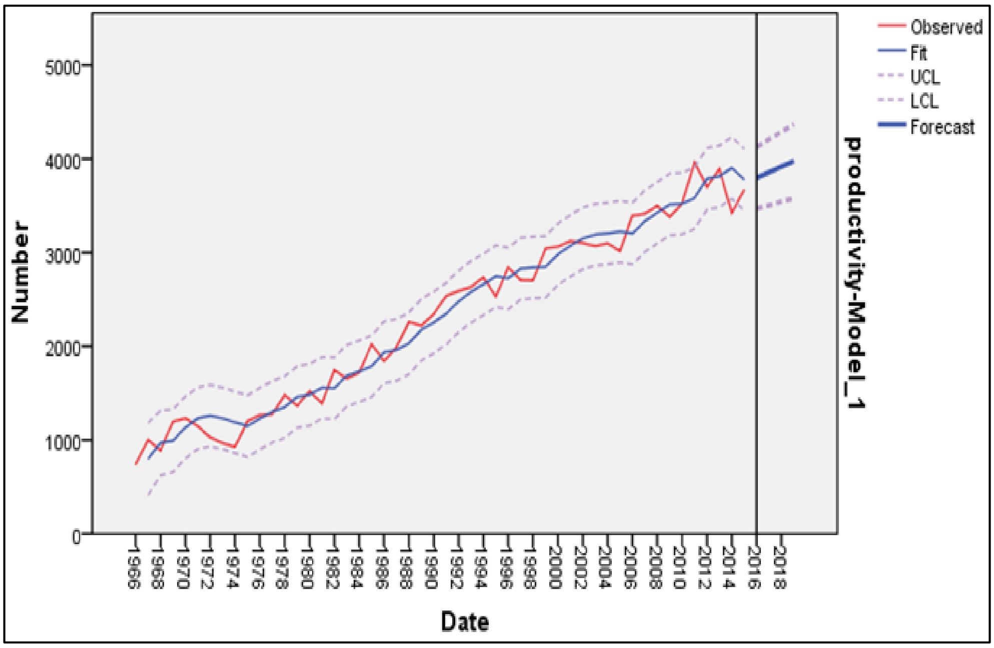


Fig 4: Observed and Estimated yield graph based on fitted ARIMA model

Table 4: Estimated Food grains yield based on ARIMA (0, 1, 1) model and their associated percentage deviation

Year	Observed Yield	Estimated Yield	Percent relative deviation
2016-17	3736	3797	1.632
2017-18	3991	3857	3.357
2018-19	3979	3916	1.583
2019-20	4036	3976	1.486

Table 5: Forecasted values for the period 2020-21 to 2023-24

Year	2020-21	2021-22	2022-23	2023-24
Forecast	4035	4094	4154	4213

Conclusion

In this study, increasing trend pattern was observed for food grains yield in Haryana. For modelling and forecasting of food grains yield ARIMA (0, 1, 1) was found best among various tried ARIMA model. Comparative plot for observed and estimated values obtained by selected model shows good fit of the selected model. This model provides a forecasted yield estimate of 4.035, 4.094, 4.154, 4.213 million tonnes for the year 2020-21 to 2023-24. Respectively. These forecasted estimates will be helpful to the government, agro-based industries, traders and agriculturists alike.

References

1. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden Day, San Franscisco. Economics. 1970;14:11-38.
2. Chaudhuri T, Verma P, Khetan M. Modeling and Forecasting of Rice Production in Some Major States of India using ARIMA, International Journal for Research in Applied Science & Engineering Technology. 2020;8(12):749-757.
3. GEP Box, GM Jenkins GC Reinsel. Time Series Analysis: Forecasting and Control, third ed., Prentice Hall, Englewood Clifs, 1967.
4. GEP Box, GM Jenkins, GC Reinsel. Time Series Analysis: Forecasting and Control, third ed., Holden-Day, San Francisco, 2009.
5. Guo WW, Xue H. Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models. Mathematical Problems in Engineering, 2014. http://dx.doi.org/10.1155/2014/857865.
6. Kumar M, Jaslam PK, Kumar S, Dhllion A. Forecast and error analysis of vegetable production in Haryana by various modeling techniques, Journal of Applied and Natural Science. 2021;13(3):907-912.
7. Kumar A, Kumar V, Rahul, Gautam AK. A time series approach to forecast wheat yield in Hisar district of Haryana, The Pharma Innovation Journal. 2021;10(7):628-632.
8. Marquardt DW. An algorithm for least-squares estimation of non-linear parameters, J Society for Industrial and Applied Mathematics. 1963;2:431-41.
9. Mishra P, Sahu PK, Padmanaban K, Vishwajith KP, Dhekale BS. Study of instability and forecasting of foodgrain production in India, International Journal of Agriculture Sciences. 2015;7(3):474-481.
10. Mishra P, Al Khatib MG, Sardar I, *et al*., Modeling and forecasting of sugarcane production in India, Sugar Tech 2021, 1-8, https://doi.org/10.1007/s12355-021- 01004-3.
11. Verma U, Dabas DS, Grewal MS, Singh JP, Hooda RS, Yadav M, *et al*. Crop yield forecasting in Haryana: 1986 to 2010, Summary Report, Department of Soil Science, CCS HAU, Hisar (Haryana). 2011, 1-148.