Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2020; Sp 9(5): 445-451
Received: 23-09-2020
Accepted: 19-10-2020

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

Ravita
Department of Obstetrics and Gynaecology, PGIMER, Chandigarh, Haryana, India

PK Muhammed Jaslam
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

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

Veer Sain
Department of Agricultural Economics, CCS Haryana Agricultural University, Hisar, Haryana, India

Corresponding Author:
Ajay Kumar
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, Haryana, India

Time-series modeling and forecasting of mustard yield in Haryana

Ajay Kumar, Ravita, PK Muhammed Jaslam, Deepankar and Veer Sain

Abstract
The importance of agriculture for Indian society and its role in economy, employment, food security, self-reliance and general well-being can hardly be over emphasized. Agriculture nowadays has become highly input and cost-intensive. Under the changed scenario today, forecasting of various aspects relating to agriculture are becoming more essential. In this study, Box-Jenkins' autoregressive integrated moving average (ARIMA) technique was applied to forecast mustard yield in Gurugram and Mahendragarh districts of Haryana. ARIMA (0, 1, 1) and (1, 1, 0) model has been found suitable for Gurugram and Mahendragarh districts respectively.

Keywords: ARIMA, forecast, modeling and mustard yield

Introduction

Time series are an integral part of our daily life. Time series data refers to observations on a variable that occurs in a time sequence. A basic assumption in any TS analysis is that some aspects of the past pattern will continue to remain in future. A lot of computation and data processing problems can be solved by time series analysis. The most widely used technique for modeling and forecasting the TS data is Box-Jenkins' Autoregressive integrated moving average (ARIMA) methodology. India is one of the largest rapeseed-mustard growing countries in the world, occupying first position in area and third position in production after the EU27 and China, and contributing around 12% of the world's total production. India's contributions to the world acreage and production are 28.3 and 19.8 percent, respectively (Source: www.mapsofindia.com/indiaagriculture). Rapeseed is a major oilseed crop in India, grown on nearly 13% of the cropped land. It is basically a winter crop and is grown in the *rabi* season from September-October to February-March in Haryana.

Yule (1927) [13] discovered the notion of stochasticity in time-series by postulating that every time series can be regarded as the realization of a stochastic process. The first concept of ARIMA models were formulated by him and his co-workers. Panse (1952, 59, 64) [9-11] in a series of papers studied the trends in yield(s) of rice and wheat with a view to compare the yield rates during the plan period(s) with that of the pre-plan period(s). Verma and Grover (2006) [12] applied ARIMA modelling on wheat yield in Haryana. ARIMA models were fitted for wheat yield forecasts in all districts of the state and further a comparison was made with remote sensing-based wheat yield forecasts and real-time yield as well. Kumar *et al.* (2017) [5] discussed modeling and forecasting of soybean yield in India using ARIMA analysis. Kumar *et al.* (2019) [4] developed a model to forecast the yield of wheat in Haryana by using annual time series data from 1980-81 to 2009-10. They applied various methods as a random walk, random walk with drift, linear trend, moving average, simple exponential smoothing, and ARIMA models and compared each other to find out the best model to forecast the yield. Mallick and Mishra (2019) [7] developed univariate ARIMA models to forecast interest rates of different maturities and stress points. They found that ARIMA (2, 1, 1) forecasting model of interest rates produced a better forecast, both in the case of in-sample and out-of-sample performances. Kumar and Verma (2020) [3] conducted a study to find out mustard yield forecast models for Bhiwani and Hisar districts of Haryana using autoregressive integrated moving average (ARIMA) technique. They found that ARIMA (0, 1, 1) and ARIMA (1, 1, 0) model is suitable for Bhiwani and Hisar districts respectively.

Materials & methods

Data Description

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 the state is 44212 sq. km. The present study dealt with modeling the time-series yield of mustard crop in Mahendragarh and

Gurugram districts of Haryana. The state Department of Agriculture and Farmers Welfare mustard yield data compiled for the period 1980-81 to 2018-19 of Mahendragarh and Gurugram districts were utilized for the training set. The yield data of post-sample period, *i.e.*, 2016-17 to 2018-19 have been used for validity testing of the developed mustard yield forecast models.

Methodology

Box and Jenkins (1970) [1] proposed a family of algebraic models from which, the one that seems appropriate for forecasting a given time series is selected. Univariate Box-Jenkins ARIMA models are based only on past patterns of the series being forecast and especially suited to short-term forecasting. The method applies to both discrete as well as to continuous data. However, the data should be available at equally spaced discrete time intervals. Also, building of an ARIMA model requires a minimum sample size of about 35-40 observations and applies only to stationary time series data. A stationary time series has mean, variance and autocorrelation function essentially constant over time. However, the most non-stationary series arising in practice can be transformed into stationary series through some simple operations.

ARIMA methodology is carried out in three stages, *viz.*, Identification, estimation and diagnostic checking. At the identification stage, two graphical devices estimated autocorrelation function and estimated partial autocorrelation function are used to measure the statistical relationships within a data series and are helpful in giving a feel for the pattern in the available data. At the estimation stage, one gets precise estimates of the coefficients of the model chosen at the identification stage. This stage also provides some warning signals if the estimated coefficients do not satisfy certain mathematical inequality conditions. At the diagnostic checking stage, testing is done to see if the estimated model is statistically adequate i.e. whether the error terms are white noise which means error terms are uncorrelated, with zero mean and constant variance. For this purpose, Ljung-Box test is applied to the original series or to the residuals after fitting a model. A good account on Ljung-Box test can be found in Box *et al.* (1994) [2]. The null hypothesis is that the series is white noise, and the alternative hypothesis is that one or more autocorrelations up to certain lags are not zero. The test statistics is given by:

$$Q^* = \text{n}(\text{n} + 2)\sum_{k=1}^{m}\frac{{r_k}^2}{n-k}$$

where n is the number of observations used to estimate the model and m is the maximum number of lag. The statistics Q^* approximately follows a chi-squared distribution with (n-k) degrees of freedom, where k is the number of parameters estimated in the ARIMA model and r_k is the autocorrelation function of residual at lag k. If it is not satisfactory, return to the identification stage again to tentatively select another model.

Results

Time-trend analysis often reflects an underlying pattern/behaviour in a time series which would otherwise be partly or nearly completely hidden by noise. The following time versus yield graphs (Figure 1) are showing overall increasing trend for mustard crop in Gurugram, Mahendergarh and Sirsa districts. The linear time-trend based model(s), i.e., T_r= a + bt, where T_r = Yield (q/ha), a = Intercept, b= Slope and t = Year, have been fitted and predictions T_r, based on this model, yielded a predictor variable *i.e.* 'trend yield'.

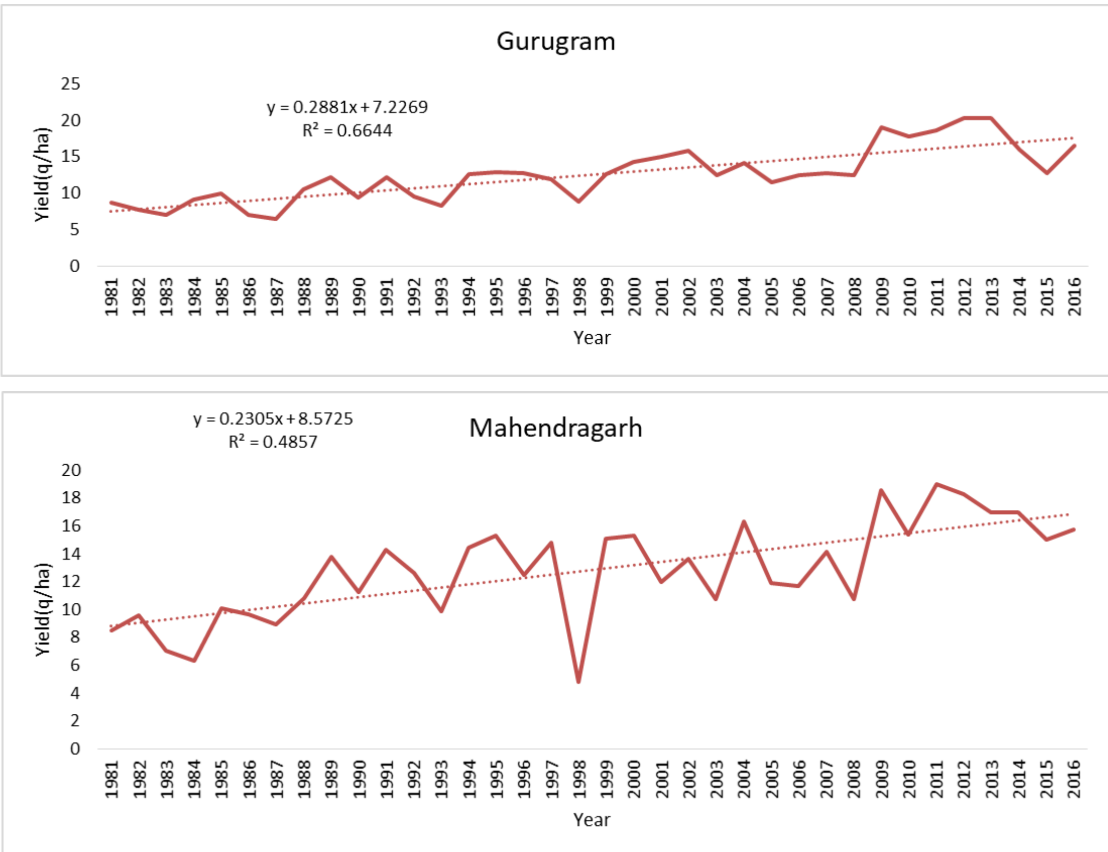


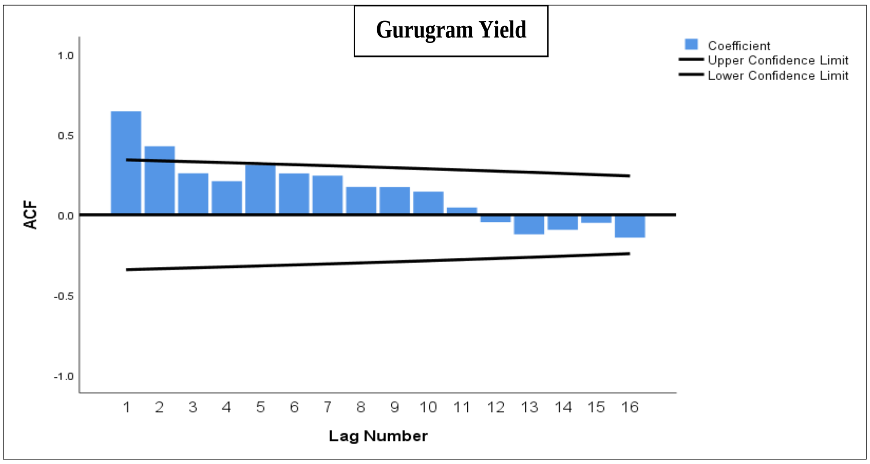
Fig 1: Time versus Yield graph(s) of mustard crop

Identification of order for AR and MA polynomials

At the identification steps, an appropriate order of AR and MA polynomials i.e. the values of p and q were determined with the help of acfs and pacfs of the stationary time series. The graphical presentation of mustard yield (q/ha) for Gurugram and Mahendragarh districts of Haryana in Figure 1 clearly shows that the data series are non-stationary. Nearly, all of the acfs upto $n/4^{th}$ lags significantly differ from zero reflecting the same non-stationarity condition (Table 1). The plotting of acfs in Figure 2 also indicates that the acfs decline gradually implying non-stationarity for all the districts. Thus, the series considered here were transformed into stationary series by differencing of order one of the original ones (Figure 4). Further, pacfs in Figure 3 show a significant spike at lag 1, just suggesting that the series may have an autoregressive component of order one. The same can be observed from the parameter values and corresponding t-test as well.

Table 1: Autocorrelations of mustard yield for all the districts

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig.
Gurugram					
1	0.74	0.16	21.60	1	<0.01
2	0.59	0.16	35.58	2	<0.01
3	0.51	0.16	46.32	3	<0.01
4	0.38	0.15	52.55	4	<0.01
5	0.34	0.15	57.71	5	<0.01
6	0.25	0.15	60.49	6	<0.01
7	0.22	0.15	62.74	7	<0.01
8	0.15	0.14	63.77	8	<0.01
9	0.16	0.14	65.01	9	<0.01
Mahendragarh					
1	0.40	0.16	6.27	1	0.01
2	0.48	0.16	15.70	2	<0.01
3	0.39	0.16	21.93	3	<0.01
4	0.18	0.15	23.29	4	<0.01
5	0.37	0.15	29.16	5	<0.01
6	-0.01	0.15	29.17	6	<0.01
7	0.06	0.15	29.36	7	<0.01
8	0.00	0.14	29.36	8	<0.01
9	0.01	0.14	29.36	9	<0.01



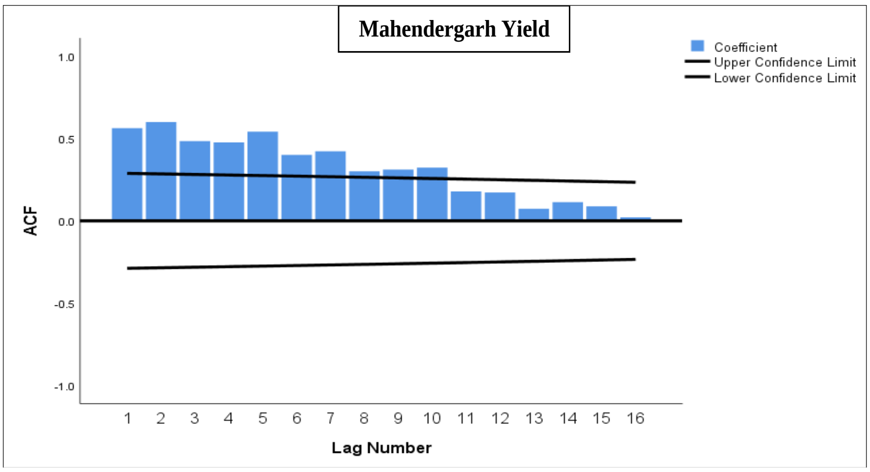
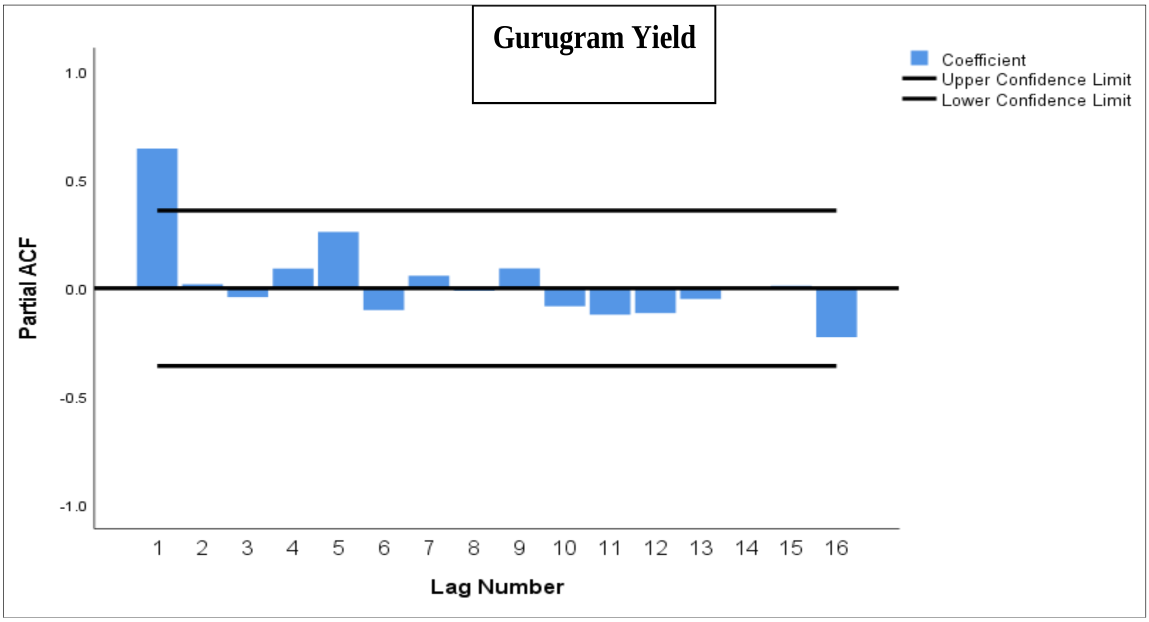


Fig 2: Autocorrelations of mustard yield for all the districts



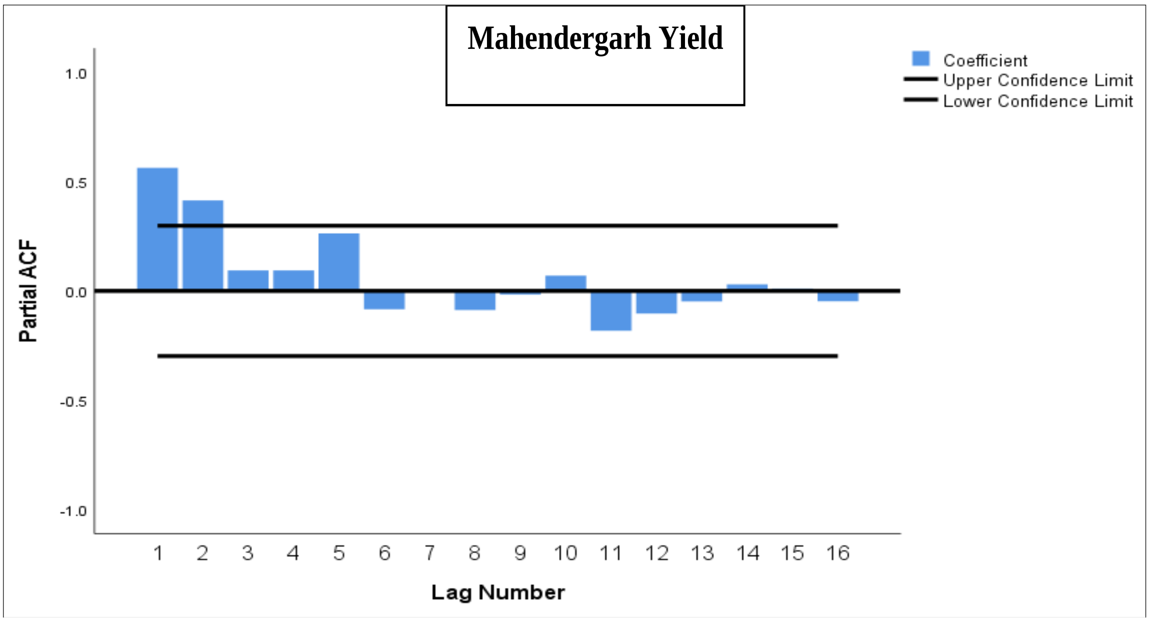
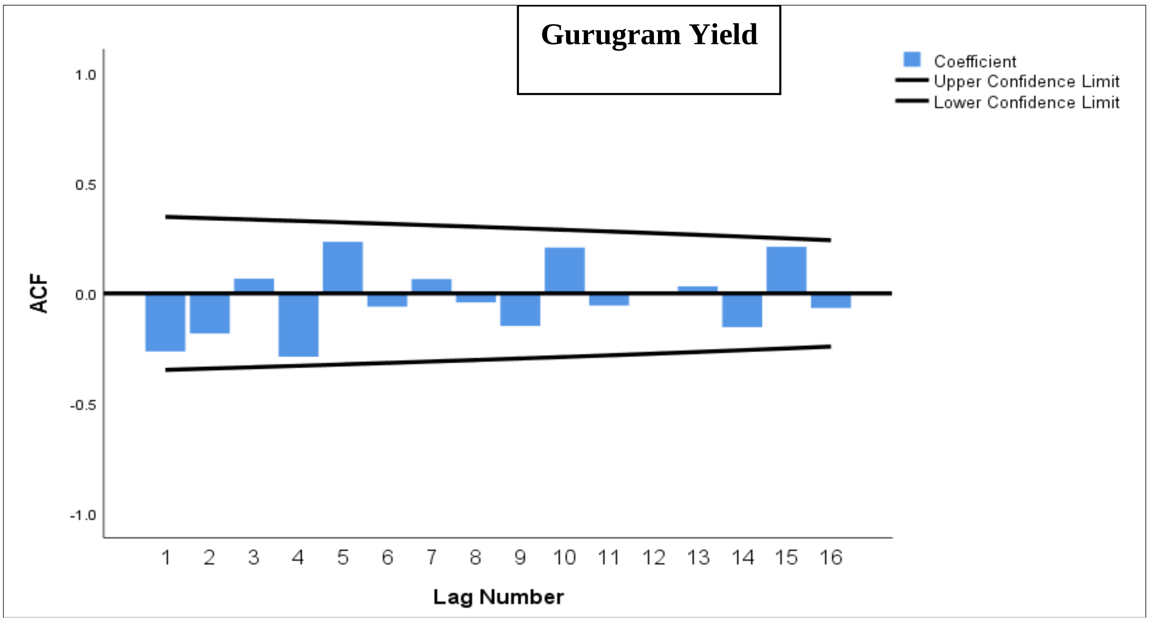


Fig 3: Partial autocorrelations of mustard yield for all the districts



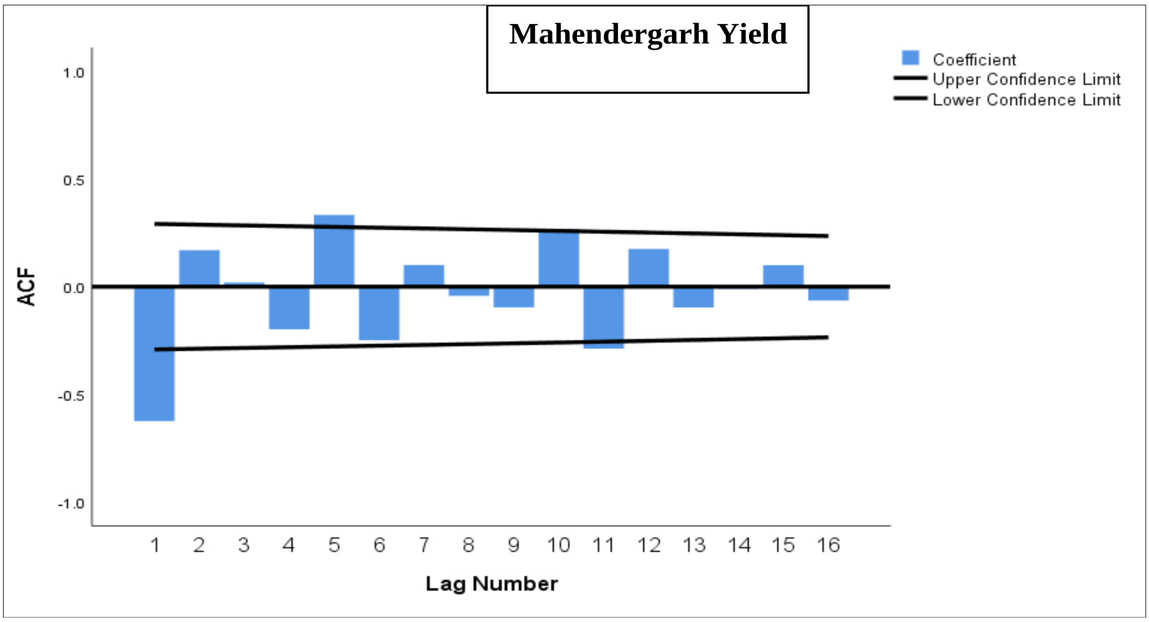


Fig 4: Autocorrelations of mustard yield after 1st differencing for all the districts

Parameter Estimation

After trying with different lags of AR and MA orders; the models ARIMA (1, 1, 0) and ARIMA (0, 1, 1), were considered at the identification stage. ARIMA estimation was carried out using non-linear least squares (NLS) approach. The relatively popular method due to Marquardt (1963) [8] was used for the purpose. Parameter estimates of the fitted ARIMA models are given in Table 2 subsequently followed by the related results shown in Tables 3 and 4.

Table 2: Parameter estimates of ARIMA models for mustard yield in all the districts

District (s)	Models		Parameter Estimate	Standard Error	Approx. Prob.
Gurugram	ARIMA (1,1,0)	AR(1)	-0.26	0.18	0.16
	ARIMA (0,1,1)	MA(1)	0.59	0.16	<0.01
Mahendragarh	ARIMA (1,1,0)	AR(1)	-0.58	0.14	<0.01
	ARIMA (0,1,1)	MA(1)	1.00	98.23	0.99

Table 3: Selection criteria values for choosing ARIMA models

District(s)	Models	RMSE	MAPE	BIC
Gurugram	ARIMA (0,1,1)	2.20	16.67	1.80
Mahendragarh	ARIMA (1,1,0)	3.14	23.48	2.50

ARIMA (0,1,1) for Gurugram and ARIMA (1,1,0) for Mahendragarh districts were fitted for district-level mustard yield(s) estimation. These models were used to obtain mustard yield forecasts for the post-sample period 2016-17 to 2018-19 as has been given in Table 6.

Table 4: Results on Stationarity and Invertibility conditions for AR and MA coefficients of fitted ARIMA models

District(s)	Model	Stationarity	Invertibility
Gurugram	ARIMA (0,1,1)	*	0.59.
Mahendragarh	ARIMA (1,1,0)	-0.58	**

*Stationarity condition is not applicable since the model is MA model

**Invertibility condition is not applicable since the model is AR model

Parameter estimates of the fitted models satisfied the stationarity and invertibility conditions since absolute value of AR and MA coefficients is less than one for all the districts.

Diagnostic checking

The model verification concerns with checking the residuals to see if they contained any systematic pattern which can be removed to improve the chosen ARIMA models. Approximate t-values were calculated for residual acfs using Bartlett's approximation for the standard error of the estimated autocorrelations. All Chi-Squared statistic(s) in this concern were calculated using the Ljung-Box (1978) [6] formula as has been shown in Table 5. The graphical Figure 5 shows that none of the residual acfs in any of the districts were significantly different from zero at a reasonable level. This ruled out any systematic pattern in the residuals.

The fitted ARIMA (0,1,1) model for Gurugram districts may be elaborated as:

$(1 – B)\, Y_t = (1 - \theta_1 B) a_t$

$Y_t - Y_{t-1} = a_t - \theta_1 B a_t$

$Y_t = Y_{t-1} - \theta_1\, a_{t-1} + a_t$

For Mahendragarh district; fitted ARIMA (1,1,0) model is expressed as:

$(1 - \phi_1 B)\,(1 – B)\, Y_t = a_t$

$(1 - \phi_1 B)\,(Y_t - BY_t) = a_t$

$Y_t - \phi_1 Y_{t-1} - Y_{t-1} + \phi_1 Y_{t-2} = a_t$

$Y_t = (1 + \phi_1)\, Y_{t-1} - \phi_1\, Y_{t-2} + a_t$

Table 5: Diagnostic checking of residual autocorrelations based on fitted ARIMA models

District(s)	Model	Ljung-box Q Statistic		
		Statistic	df	Sig
Gurugram	ARIMA (0,1,1)	9.46	17	0.93
Mahendragarh	ARIMA (1,1,0)	23.74	17	0.13

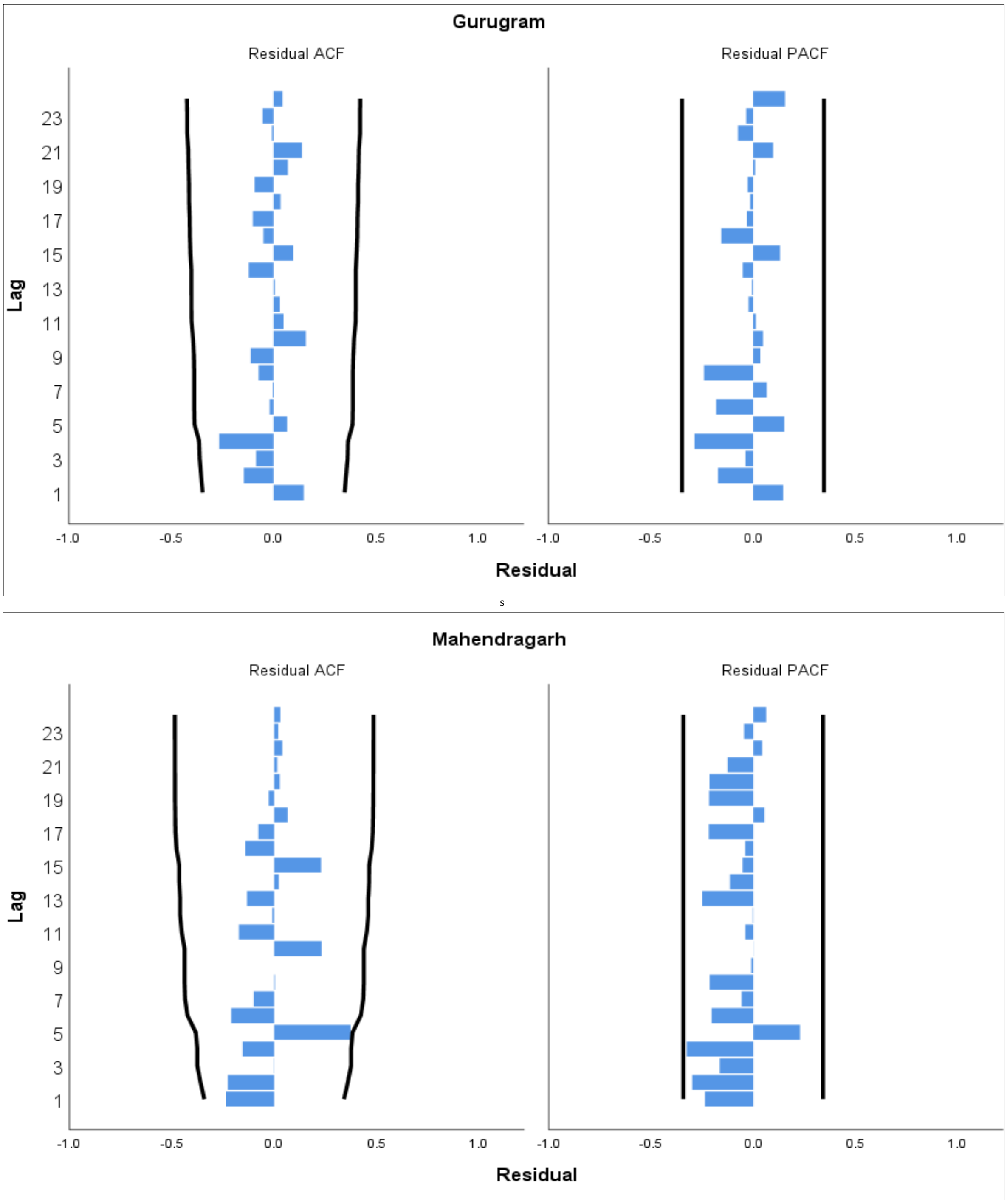


Fig 5: Residual acfs and pacfs plots based on fitted ARIMA models

Table 6: District-specific estimated mustard yield(s) based on ARIMA models and their associated percent relative deviations (RD%) = 100×(Obs. Yield-Fitted Yield)/Obs. Yield)

District/Model	Forecast Year	Observed Yield (q/ha)	Fitted Yield (q/ha)	Percent Relative Deviation
Gurugram ARIMA (0,1,1)	2016-17	20.03	-6.24	-6.24
	2017-18	23.25	6.84	6.84
	2018-19	22.36	1.48	1.48
Av. Abs. percent dev.				4.85
Mahendragarh ARIMA (1,1,0)	2016-17	19.58	17.85	8.84
	2017-18	18.83	18.07	4.04
	2018-19	20.54	18.34	10.71
Av. Abs. percent dev.				7.86

Discussion

The forecast performance(s) of alternative models were observed in terms of per cent deviations of mustard yield forecasts about the real-time yield(s). It has been observed that ARIMA technique is appropriate to forecast the mustard yield. The ARIMA model can be improved the forecast accuracy when it incorporates by several variables as weather variables, diseases effects etc. Finally, the fitted models are accomplished of providing satisfactory estimates of mustard yield well in advance of the crop harvest.

References

1. Box GEP, Jenkins GM. Time series analysis: Forecasting and control, 1st ed., Holden Day, San Francisco, 1970.
2. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: Forecasting and control, Holden Day, San Francisco, 1994.
3. Kumar A, Verma U. Forecasting mustard yield in Haryana with ARIMA model, The Pharma Innovation 2020;9(4):136-140.
4. Kumar A, Deepankar PK, MJ, Kumar A. Wheat yield forecasting in Haryana: A time series approach, Bulletin of Environment Pharmacology and Life Sciences 2019;8(3):63-69.
5. Kumar M, Raman RK, Kumar S. Forecasting of soybean yield in India through ARIMA model, International Journal of Pure and Applied Bioscience 2017;5(5):1538-1546.
6. Ljung GM, Box GEP. On a measure of lack of fit in time series models, Biometrika 1978;65:297-303.
7. Mallick AK, Mishra AK. Interest rates forecasting and stress testing in India: An ARIMA approach, Palgrave Communications 2019;5(32):1078-1088.
8. Marquardt DW. An algorithm for least-squares estimation of non-linear parameters. Journal of the Society for Industrial and Applied Mathematics. 1963;11(2):431-441.
9. Panse VG. Trends in areas and yields of principal crops in India. Agriculture Situation in India 1952;7:144-48.
10. Panse VG. Recent trends in the yield of rice and wheat in India. Indian Journal of Agricultural Economics 1959;14:11-38.
11. Panse VG. Yield trends of rice and wheat in first two five years plans in India. Journal of Indian Society of Agricultural Statistics 1964;16(1):1-50.
12. Verma U, Grover D. ARIMA wheat yield modelling in Haryana. Research Bulletin, Department of Mathematics and Statistics, CCS HAU, Hisar, Haryana, 2006, 1-51.
13. Yule GU. On the method of investigating periodicities in disturbed series, with special reference to wolfers sunspot numbers, Philosophical Transactions of the Royal Society London, series A 1927;226:267 298.