www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; SP-10(7): 628-632

www.thepharmajournal.com
Received: 07-05-2021
Accepted: 10-06-2021

Ajay Kumar
Department of Mathematics, Guru Jambheshwar University of Science & Technology, Hisar, Haryana, India

Vinay Kumar
Department of Mathematics and Statistics, CCSHAU, Hisar, Haryana, India

Rahul
Department of Agricultural Meteorology, CCSHAU, Hisar, Haryana, India

Ajay Kumar Gautam
Department of Agricultural Statistics, ANDUAT, Kumarganj, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Rahul
Department of Agricultural Meteorology, CCSHAU, Hisar, Haryana, India

A time series approach to forecast wheat yield in Hisar district of Haryana

Ajay Kumar, Vinay Kumar, Rahul and Ajay Kumar Gautam

Abstract
The present study has been performed to forecast wheat yield in Hisar district of Haryana using a time series approach (Autoregressive Integrated Moving Average). The ARIMA model has been fitted using wheat yield data of Hisar for the period 1966-67 to 2018-19. The validity of fitted model has been checked for subsequent years i.e. 2016-17 to 2018-19, not included in the development of the models. In this study, it is found that ARIMA (0,1,1) model be the best model to forecast wheat yield in Hisar.

Keywords: ARIMA, Wheat, Forecast, ACF, PACF

Introduction

Now-a-days agriculture has become highly input and cost intensive. Under the changed scenario today, forecasting of various aspects relating to agriculture are becoming more essential. Crop forecasting is a formidable challenge. There is a worldwide research effort to develop methodologies for pre-harvest forecasting of crop yield. Reliable and timely forecasts provide useful input for proper, foresighted and informed planning, more so, in agriculture which is full of uncertainties. A popular and widely used statistical method for time series forecasting is the ARIMA model. ARIMA is an acronym that stands for Autoregressive Integrated Moving Average. It explicitly caters to a suite of standard structures in time series data and as such provides a simple yet powerful method for making skilful time series forecasts. Yule (1927) [11] was the first who launched the notion of stochasticity in time series by postulating that every time series can be regarded as the realization of a stochastic process. He and his co-workers first formulated the concept of ARIMA models. Panse (1952, 59, 64) [7] in a series of papers studied the trends in yields of rice and wheat with a view to compare the yield rates during the plan periods with that of the pre-plan periods. Box and Jenkins in early 1970's, pioneered in evolving methodologies for modelling the univariate case often referred to as Univariate Box-Jenkins (UBJ) ARIMA modelling. Box and Pierce (1970) [1] observed that if an appropriate model is chosen, there would be zero autocorrelation. Verma *et al*. (2011) [10] worked on wheat, sugarcane, cotton and mustard crops for operational yield forecasting purpose under Crop Acreage Production Estimation (CAPE) project. Padhan (2012) [6] forecasted annual productivity of thirty-four different agricultural products using ARIMA model. Biswas and Bhattacharyya (2013) carried out time-series modeling for forecasting area and production of rice in West Bengal using univariate series since independence. Kumarasinghe and Peiris (2018) developed ARIMA models for the total annual tea production in Sri Lanka based on time series data from 1964 to 2015. Kumar *et al*. (2019) [4] developed a model to forecast the yield of wheat in Haryana by using annual time series data.
Wheat is the main cereal crop in India. India is the second largest producer among wheat growing countries of the World. Haryana occupies third place for wheat production among the various states in India. Wheat crop has wide adaptability. It can be grown not only in the tropical and sub-tropical zones, but also in the temperate zone and the cold tracts of the far north, beyond even the 60-degree north altitude. Soils with a clay loam or loam texture, good structure and moderate water holding capacity are ideal for wheat cultivation.

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 the state is 44212 sq. km. The present study dealt with modeling the time-series yield of wheat crop in Hisar districts of Haryana.

The state Department of Agriculture and Farmers Welfare wheat yield data compiled for the period 1966-67 to 2018-19 of Hisar district is utilized for the purpose. The validity of fitted model has been checked for subsequent years i.e. 2016-17 to 2018-19, not included in the development of the models.

Box-Jenkins autoregressive integrated moving average methodology

Box-Jenkins ARIMA is the most general class of models for forecasting time series data. ARIMA forecasts are based only on past values of the variable being forecast. They are not based on any other data series and especially suited to short-term forecasting. The method applies to both discrete as well as to continuous data. However, the data should be 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. Autocorrelation function (acf) refers to the way the observations in a time series are related to each other and is measured by the simple correlation between current observation Y_t and Y_{t-p}, the observation from *p* period before the current one.

An ARIMA model is expressed as:

$$\phi(B)\,(1-B)^d\, Y_t = \theta(B)\,\varepsilon_t$$

Where,

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ (Autoregressive parameter)

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ (Moving average parameter)

ε_t=white noise or error term

d=differencing term

B=Backshift operator i.e. $B^a Y_t = Y_{t-a}$

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. An estimated partial autocorrelation function shows the correlation between ordered pairs separated by various time spans with the effect of intervening observations. These functions act as guide for choosing one or more ARIMA models that seem to be appropriate.

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. Estimation of parameters for ARIMA model is generally done through iterative least squares method. 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) [3].

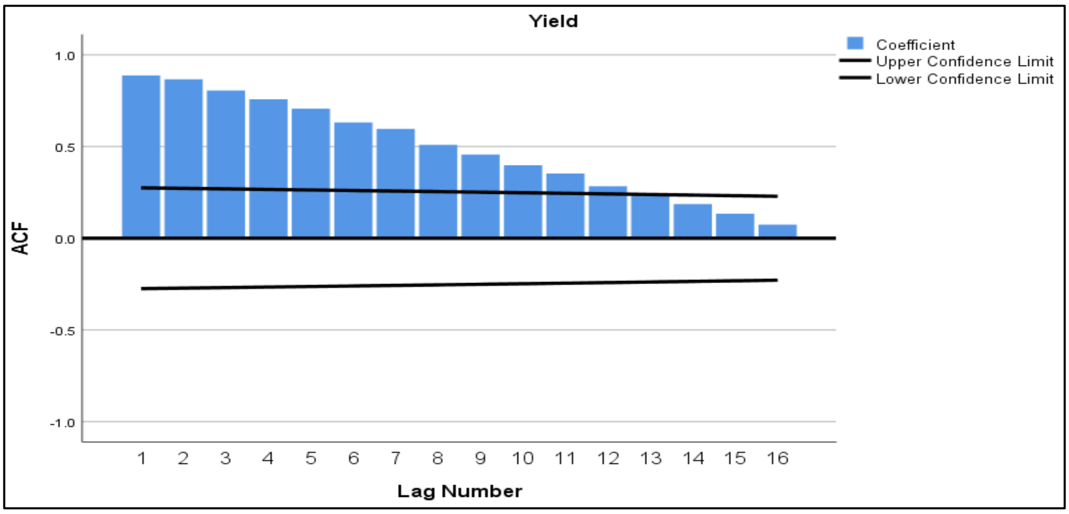
Result and Discussion

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 autocorrelation functions and partial autocorrelation functions of the stationary series. The wheat yield(s) 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 Wheat yield

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig
1	0.89	0.14	41.79	1	<0.01
2	0.87	0.14	82.48	2	<0.01
3	0.81	0.13	118.40	3	<0.01
4	0.76	0.13	150.82	4	<0.01
5	0.71	0.13	179.68	5	<0.01
6	0.63	0.13	203.21	6	<0.01
7	0.60	0.13	224.69	7	<0.01
8	0.51	0.13	240.75	8	<0.01
9	0.46	0.13	253.95	9	<0.01
10	0.40	0.12	264.24	10	<0.01



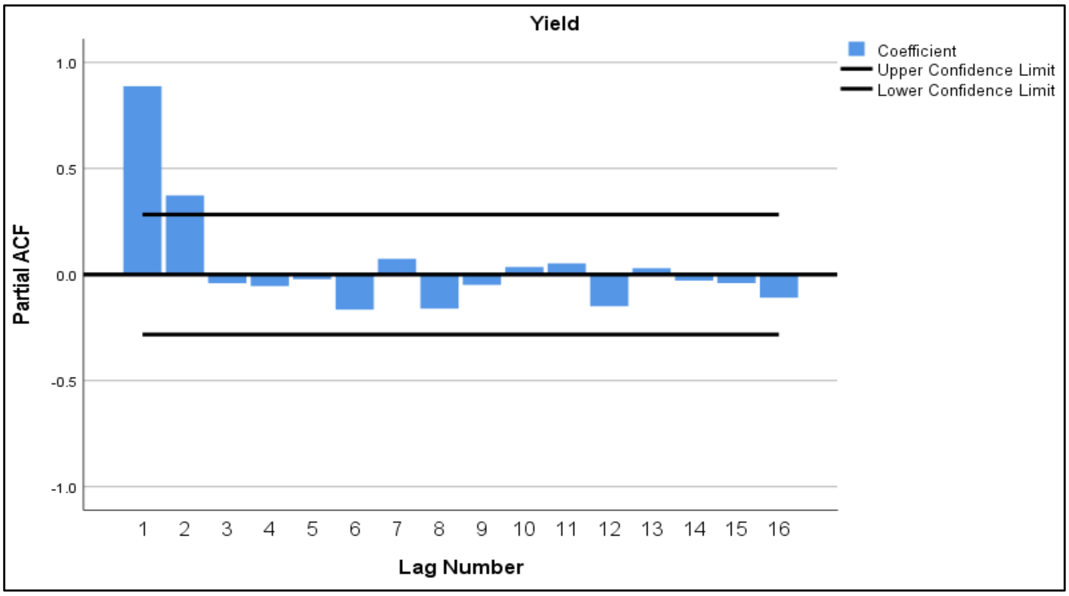
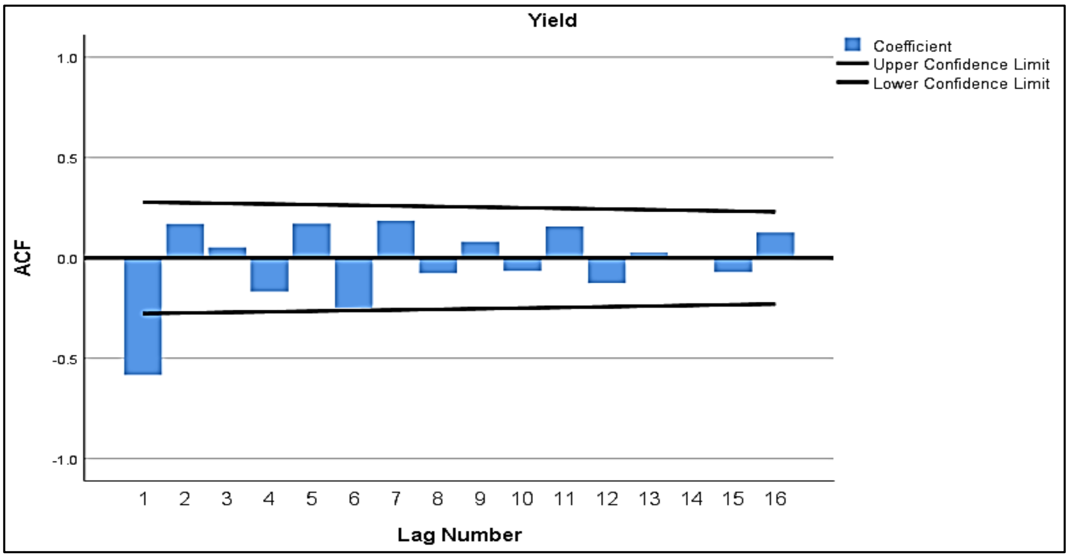


Fig 1: Autocorrelation and Partial autocorrelation for Wheat yield



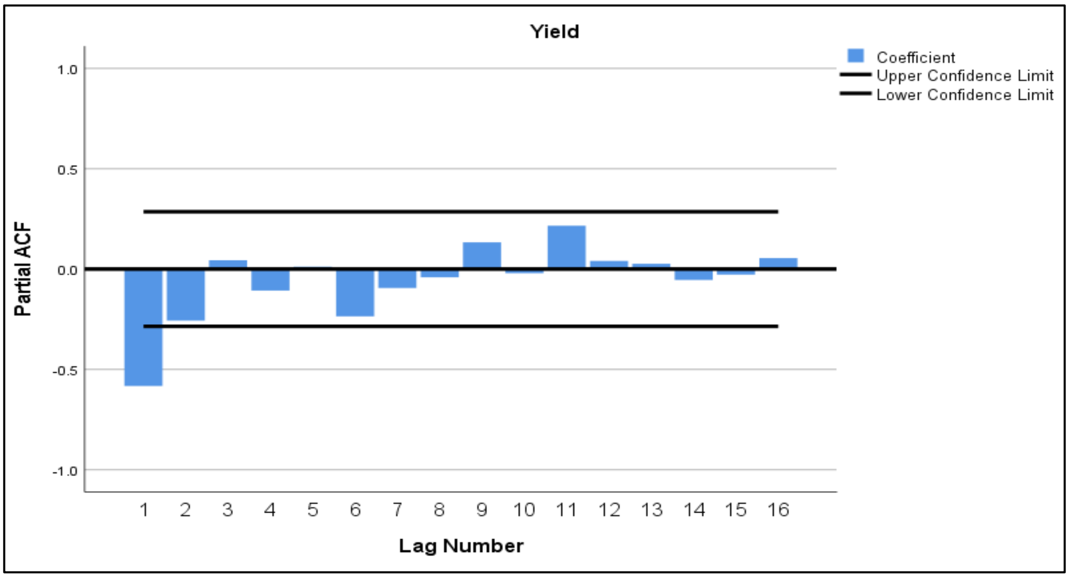


Fig 2: Autocorrelation and Partial autocorrelation for Wheat yield after first differencing

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(0,1,2) ARIMA(2,1,0) ARIMA(2,1,1) were fitted for estimating the wheat yield. The models ARIMA (2,1,1), ARIMA (1,1,0) and ARIMA (0,1,1) were considered in the identification stage and ARIMA estimation was carried out using a non-linear least squares (NLS) approach. The relatively popular method for estimation of parameters under ARIMA structure is due to Marquardt who has designed a powerful algorithm for estimation through iterative improvement. Parameter estimates of the selected models shown in Table 2 are less than one (needed for convergence) and also satisfy the stationarity and invertibility conditions under ARIMA structure.

Table 2: Tentative ARIMA models for Wheat yield

Model		Estimate	Std. Error	Approx Prob.
ARIMA(1,1,0)	AR(1)	-0.59	0.12	<0.01
ARIMA(0,1,1)	MA(1)	0.63	0.12	<0.01
ARIMA(1,1,1)	AR(1)	-0.37	0.22	0.10
	MA(1)	0.36	0.20	0.11
ARIMA(0,1,2)	MA(1)	0.69	0.14	<0.01
	MA(2)	-0.25	0.15	0.12
ARIMA(2,1,0)	AR(1)	-0.74	0.14	<0.01
	AR(2)	-0.25	0.14	0.09
ARIMA(2,1,1)	AR(1)	-1.44	0.19	<0.01
	AR(2)	-.62	0.12	<0.01
	MA(1)	-.82	0.21	<0.01

Table 3: Selection Criteria vales for choosing ARIMA models

Models	RMSE	R-squared	MAE	MAPE	BIC
ARIMA(1,1,0)	314.21	0.91	237.65	7.90	11.66
ARIMA(2,1,1)	308.80	0.92	247.25	8.31	11.70
ARIMA(0,1,1)	306.65	0.92	237.34	7.91	11.77

After experimenting with different lags of the moving average and the autoregressive processes, ARIMA (0,1,1) was found to be the best fit for wheat yield estimation.

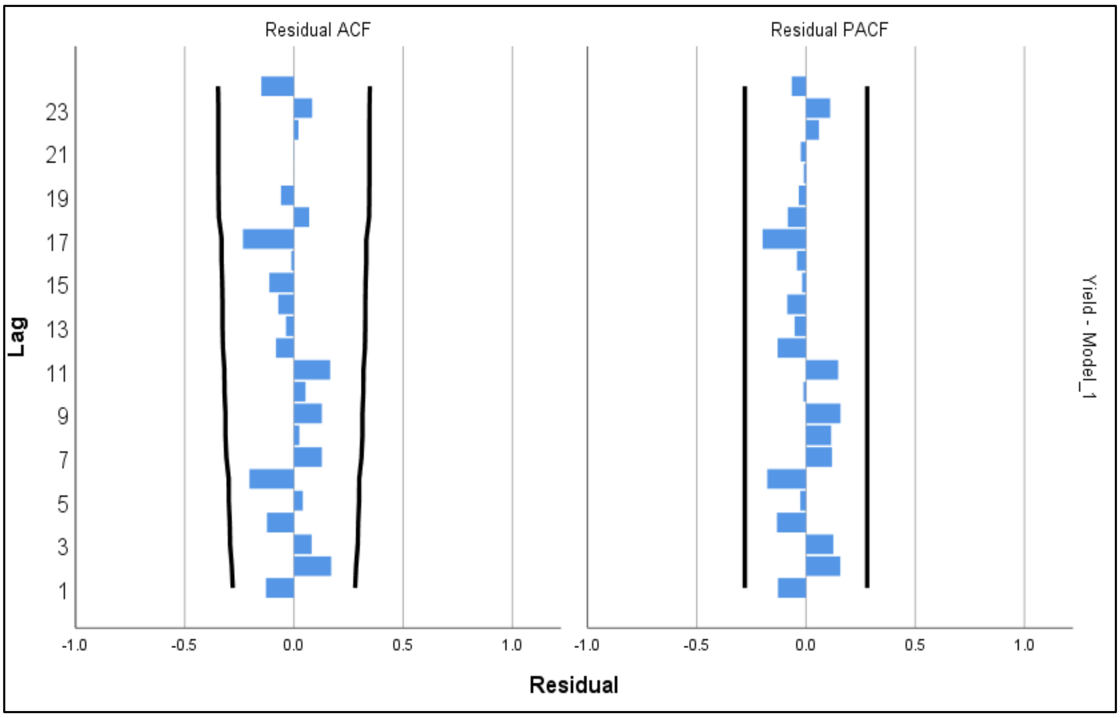


Fig 3: Residual acf and pacf plot based on fitted ARIMA model

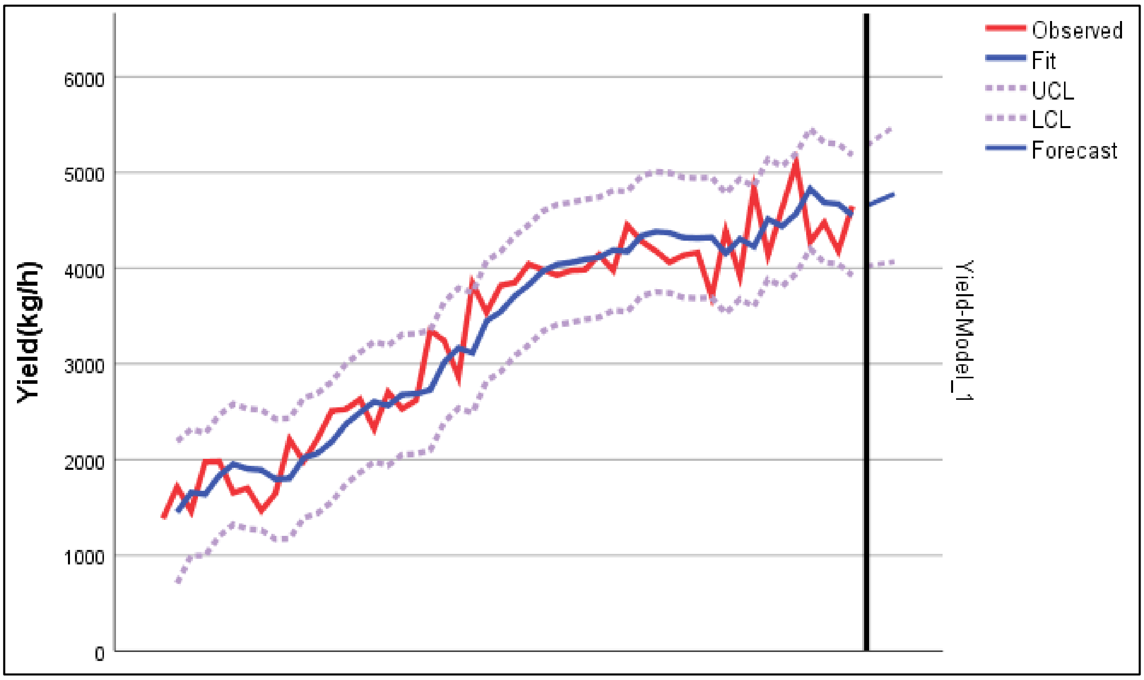


Fig 4: Observed and Estimated yield graph based on fitted ARIMA model

Table 4: Estimated Wheat yield based on ARIMA (0,1,1) model and their associated percentage deviation (RD%) = 100×(Obs. Yield - Est. Yield)/Obs. Yield).

Forecast Year	Observed Yield (kg/ha)	Estimated Yield (kg/ha)	Percent relative deviation
2016-17	4758	4651	2.25
2017-18	4914	4714	4.07
2018-19	4955	4777	3.59

Conclusion

In this study, wheat yield for Hisar district is forecasted using ARIMA techniques. From the above results and summary, ARIMA (0, 1, 1) provides satisfactory results for subsequent years. So, it is concluding that ARIMA (0, 1, 1) model be the best model to forecast wheat yield in Hisar.

References

1. Box GEP, Pierce AD. Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statistical association 1970;65(332):l1509-1526.
2. Box GEP, Jenkins GM. Time series analysis: Forecasting and Control, Holden Day, San Francisco 1976.
3. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: Forecasting and control, Holden Day, San Francisco 1994.
4. Kumar A, Deepankar PKMJ, Kumar A. Wheat yield forecasting in Haryana: A time series approach, Bulletin of Environment Pharmacology and Life Sciences 2019;8(3):63-69.
5. Kumarasinghe HPASS, Peiris BL. Forecasting annual tea production in Sri Lanka, Tropical Agricultural Research 2018;29(2):184-193.
6. Padhan PC. Application of ARIMA Model for forecasting agricultural productivity in India, Journal of Agriculture & Social Sciences 2012;8:50-56.
7. Panse VG. Trends in areas and yields of principal crops in India, Agriculture Situation in India 1952;7:144-48.
8. Panse VG. Recent trends in the yield of rice and wheat in India, Indian Journal of Agricultural Economics 1959;14:11-38.
9. Panse VG. Yield trends of rice and wheat in first two five year plans in India, Journal of Indian Society of Agricultural Statistics 1964;16(1):1-50.
10. 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.
11. 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, 226-298.