The Pharma Innovation Journal 2020; SP-9(4): 136-140

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2020; SP-9(4): 136-140

www.thepharmajournal.com
Received: 03-02-2020
Accepted: 05-03-2020

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

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

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

Forecasting mustard yield in Haryana with ARIMA model

Ajay Kumar and Urmil Verma

Abstract
The present study has been conducted to find out mustard yield forecast models for Bhiwani and Hisar districts of Haryana using autoregressive integrated moving average (ARIMA) technique. The main focus of the study is to forecast five years ahead mustard yield in the said districts. The selection criteria like Akaike information criteria and Bayesian information criteria acted as a guiding hint to decide the final models. The forecast figures and real-time yield(s) for both the districts are compared on the basis of mean absolute percentage error to check the validity performance of the developed models.

Keywords: Forecast, ARIMA (autoregressive integrated moving average), (MAPE) Mean absolute percentage error, (RMSE) Root mean square error, BIC (Bayesian information criteria), Mustard Yield.

Introduction

Forecasting helps the decision-makers to make their future decisions more correctly, whether from the economic or non-economic/ agricultural fields. In fact, national governments need forecasting to make different policy decisions on storage, pricing, marketing, import-export, etc. Continuous changes are usually characterized by the productivity of agricultural crops due to many factors such as variations in precipitation and economic, technological and agricultural conditions. Studying the essence and course of changes in that productivity is useful in evaluating efforts to increase agricultural production. Moreover, the productivity forecast of various agricultural crops allows accurate prediction of productivity levels to be made in the years to come. Thus, forecasting is one of the main tools in the field of agricultural production to make effective growth policies and successful economic plans.

Time series (TS) 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. Time series are an integral part of our daily life. 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.

Panse (1952, 59, 64) 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. Boken (2000) developed time series models for spring wheat yield forecasting for the Canadian Prairies. Suresh and Priya (2011) [12] used ARIMA models to predict the area, production and yield of sugarcane in Tamilnadu and found that ARIMA (1, 1, 1) for area & yield and ARIMA (2, 1, 2) was best for production. In 2016, Hossain and Abdulla found that ARIMA (0, 2, 1) model was best for predicting the potato production in Bangladesh using autoregressive integrated moving average (ARIMA) method. Kumar *et al.* (2017) [4] forecasted the productivity of sugarcane in Bihar by fitting ARIMA model and selected ARIMA (0, 1, 1) model as the best model. Nath *et al.* (2019) [7] used Box-Jenkins' time series modeling approach to forecast wheat production in India and found ARIMA (1, 1, 0) model as the best among all of them. The forecasted values indicated an average growth rate of approximately 4% per year in wheat production.

India has a very well-developed network for recording and aggregating crop statistics at various village levels. After groundnut, rapeseed-mustard is the largest edible oilseed which shares 27.8% in the India's oilseed economy. Indian mustard (*Brassica juncea* (L.) Czern & Coss.) is mainly cultivated in Gujarat, Uttar Pradesh, Madhya Pradesh, Rajasthan and Haryana. It is well grown in areas of rainfall of 25 to 40 cm.

It is primarily a winter crop and is grown in Haryana during the *rabi* season from September-October to February-March.

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 research dealt with time series modeling for mustard yield forecasts in Bhiwani and Hisar districts of Haryana. The state department of agriculture (DOA) mustard yield data from 1980-81 to 2015-16 were compiled from Statistical Abstracts of Haryana for the targeted purpose (Source: esaharyana.gov.in/State Statistical Abstract/). The mustard yield data from 1980-81 to 2010-11 were used to develop the model and next five years i.e. 2011-12 to 2015-16 have been used to check the validity of the developed ARIMA models.

Autoregressive Integrated Moving Average (ARIMA) procedure

Autoregressive integrated moving average method popularized by George Box and Gwilym Jenkins (1970) [2] is applied to time series analysis and forecasting. Autoregressive integrated moving average forecasts are based only on the previous value of the predicted variable. This approach covers both continuous and discrete data. The ARIMA model needs a minimum sample size of approximately 30-40 observations and only applies to stationary time series data. The mean, variance and autocorrelation function of a stationary time series is generally constant over time. Mostly, non-stationary time series arising in practice can be converted into stationary time series through differencing. Differencing is applied when the mean of a time series is fluctuating over time and log transformation is applied if the variance of a time series is fluctuating through time. The Autoregressive integrated moving average modelling have three stages 1) Identification stage 2) Estimation stage and 3) Diagnostic checking

At the identification stage, two graphical tools (autocorrelation function and partial autocorrelation function) are used to calculate the statistical relationships within a data set. These functions facilitate the selection of one or more relatively suitable ARIMA models. At the second step, one can correctly measure the coefficients of the model selected at the identification stage. This step also provides warning signals when some mathematical inequality requirements pertaining to stationarity/invertibility are not satisfied by the estimated coefficients. Finally, at third step, residuals are tested to see whether the random shocks are independent or not and finally to decide if an estimated model is statistically significant or not.

An estimated autocorrelation function r_k describe the correlation between ordered pairs $(\overline{Y}_t, \overline{Y}_{t+k})$ separated by various time spans (k=1, 2, 3,.) may be expressed as:

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2}$$

An estimated partial autocorrelation function ϕ_{kk} describe the correlation between ordered pairs $(\overline{Y}_t, \overline{Y}_{t+k})$ separated by various time spans (k = 1, 2, 3…) with the effect of intervening observations $(\overline{Y}_{t+1}, \overline{Y}_{t+2}, \ldots\ldots, \overline{Y}_{t+k-1})$ accounted for.

An ARIMA model can be written 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 parameters)

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

d - Order of differencing

B - Backshift operator

ε_t - White noise or error term

Results and discussion

First of all, the ARIMA model building leads to identify the stationary behavior of the data series. The graphical representation (Figure 1) showing the long-term behavior of the series indicates that the data series are non-stationary for both the districts. Almost, all autocorrelations upto $(n/4)^{th}$ lags significantly different from zero confirm non-stationarity (Figure 2). Differencing of order one i.e. d=1 was enough for getting an approximate stationary series in the said districts (Figure 4).

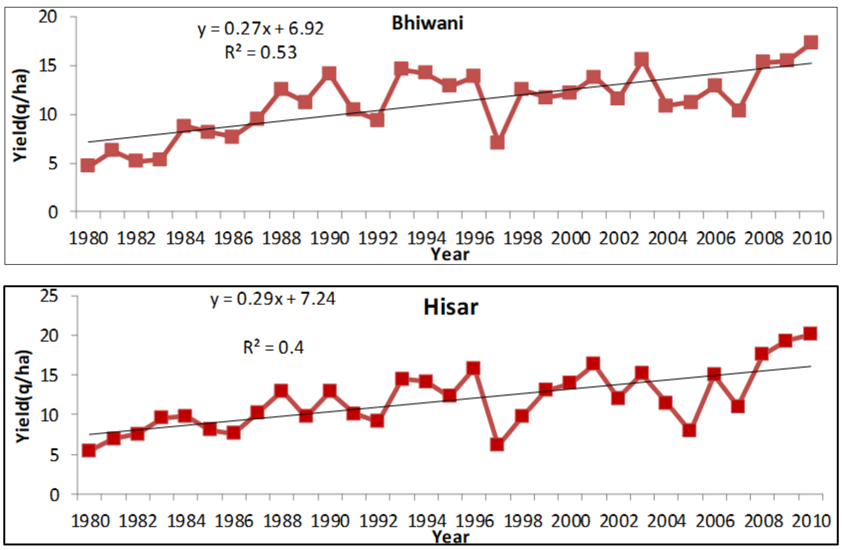

Fig 1: Time versus Yield graph(s) of mustard yield for Bhiwani and Hisar districts

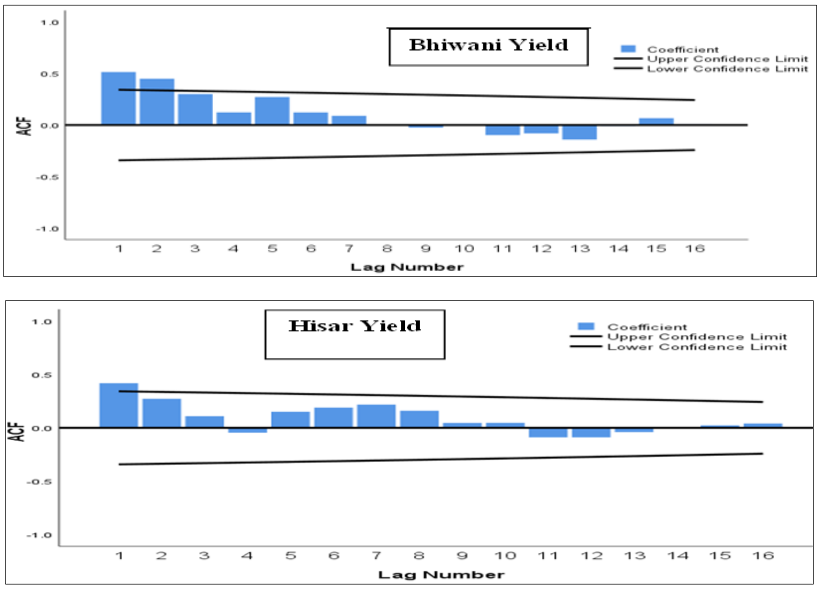

Fig 2: Autocorrelations of mustard yield for Bhiwani and Hisar districts

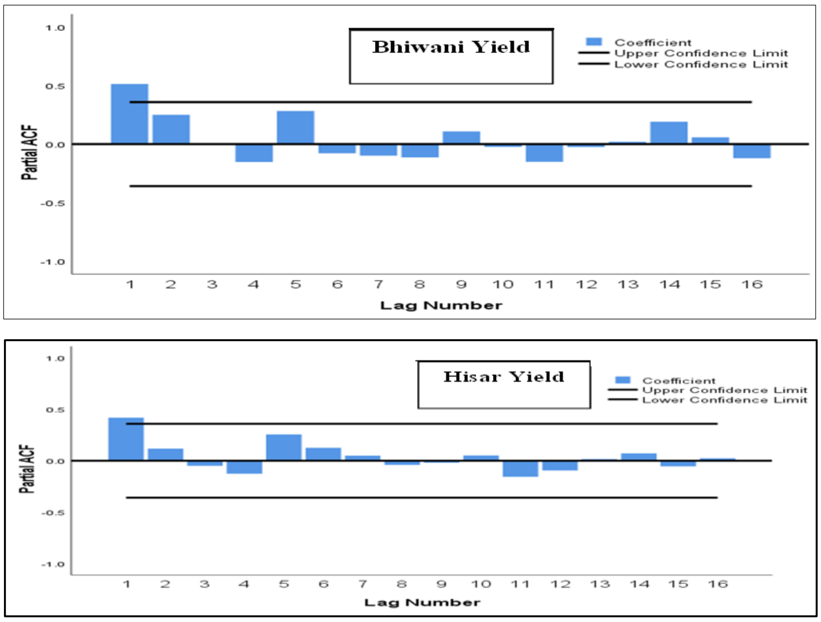

Fig 3: Partial autocorrelations of mustard yield for Bhiwani and Hisar districts

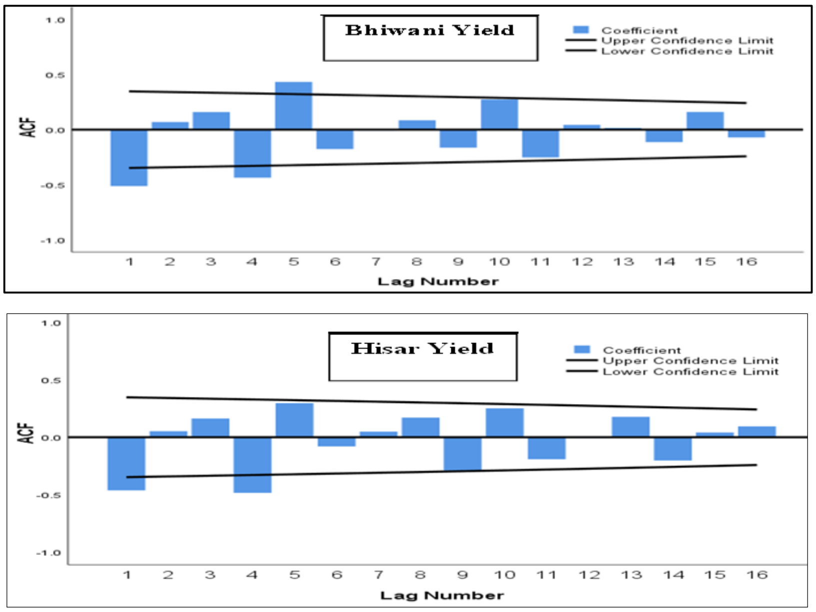

Fig 4: Autocorrelations of mustard yield after 1st differencing for Bhiwani and Hisar districts

After trying with different lags of AR and MA orders; the models ARIMA (1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1) were considered in the identification stage. ARIMA estimation was carried out using a non-linear least squares (NLS) approach due to Marquardt (1963) [6]. Schwarz's Bayesian information criterion (SBIC, 1978), Mean absolute percentage error, Root mean square error and residual variance gave the way to select the appropriate models. Parameter estimates of the finally selected ARIMA models are given in Table 1. Approximate 't' values were calculated for residual autocorrelation coefficients using Bartlett's approximation for the standard error of the estimated autocorrelations. The residual acfs along with the associated 't' tests and Chi-squared test suggested by Ljung and Box (1978) were used for the checking of random shocks to be white noise. All Chi-Squared statistics in this concern are shown in Table 3.

Table 1: Parameter estimates of tentative ARIMA models

Districts	Models		Parameter Estimate	Standard error	Approx. prob.
Bhiwani	ARIMA (0,1,1)	MA	0.69	0.15	<0.01
	ARIMA (1,1,0)	AR	-0.50	0.17	0.01
	ARIMA (1,1,1)	AR	-0.09	0.29	0.76
		MA	0.64	0.23	0.01
Hisar	ARIMA (0,1,1)	MA	1.00	18.08	0.96
	ARIMA (1,1,0)	AR	-0.52	0.18	0.01
	ARIMA (1,1,1)	AR	0.01	0.26	0.97
		MA	0.99	2.21	0.66

Table 2: Results on Stationary and Inevitability conditions for AR and MA coefficients of fitted ARIMA models

Districts	Model	Stationary	Inevitability
Bhiwani	ARIMA (0,1,1)	*	0.69
Hisar	ARIMA (1,1,0)	-0.52	**

- 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 selected models satisfied the stationary and inevitability conditions since absolute value of AR and MA coefficients is less than one.

Table 3: Diagnostic checking of residual autocorrelations of mustard yield for Bhiwani and Hisar districts

Districts	Models	Model Fit Statistic			Ljung-Box Q Statistic		
		RMSE	MAPE	Normalized BIC	Statistic	df	Sig.
Bhiwani	ARIMA(0,1,1)	1.46	9.20	1.97	18.06	17	0.38
Hisar	ARIMA(1,1,0)	1.61	7.42	2.62	20.69	17	0.24

Ljung-Box Q Statistic(s) given in Table 3 favors the acceptance of random shocks to be as white noise pertaining to the fitted models. Finally, a comparison between ARIMA model based mustard yield estimates with real-time yield(s) was observed in terms of percent relative deviation (RD%). The results presented in Table 4 indicate that the deviations of predicted yield(s) from the actual yield(s) are within acceptable limits and thus favouring the use of ARIMA models to get short-term forecast estimates in the districts under consideration.

Table 4: Post-sample mustard yield forecasts based on fitted models for Bhiwani and Hisar districts

Districts/Model	Year	Observed Yield (q/ha)	Estimated Yield (q/ha)	Percent relative deviation
Bhiwani ARIMA (0,1,1)	2011	12.00	14.29	-19.05
	2012	16.40	14.57	11.18
	2013	15.16	14.85	2.06
	2014	13.98	15.13	-8.21
	2015	14.61	15.41	-5.47
Av. Abs. percent dev.				9.19
Hisar ARIMA (1,1,0)	2011	17.07	16.16	5.29
	2012	16.78	16.97	-1.15
	2013	16.26	17.07	-5.00
	2014	14.17	17.54	-23.80*
	2015	18.16	17.82	1.88
Av. Abs. percent dev.				7.42

RD (%) = {(Observed yield -Estimated yield)/ Observed yield} × 100

***** As per IMD 2015, forty percent of the state's net sown area (2.24 million ha) was affected by unseasonal rainfall and hailstorm in March (India, Ministry of Agriculture 2015c).

References

1. Boken VK. Forecasting spring wheat yield using time series analysis: A case study for the Canadian Prairies, Agronomy Journal. 2000; 92:1047-1053.
2. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden Day, San Franscisco. Economics. 1970; 14:11-38.
3. Hossain Md M, Abdulla F. Forecasting Potato Production in Bangladesh by ARIMA Model, Journal of Advanced Statistics, 2016, 1(4).
4. Kumar M, Raman RK, Kumar S. Sugarcane Productivity in Bihar- A Forecast through ARIMA Model, Int. J. Pure

App. Biosci. 2017; 5(6):1042-1051.
5. Ljung GM, Box GEP. On a measure of lack of fit in time series models, Biometrical. 1978; 65:297-303.
6. Marquardt DW. An algorithm for least-squares estimation of non-linear parameters, J Society for Industrial and Applied Mathematics. 1963; 2:431-41.
7. Nath BS, Dhakre DS, Bhattacharya D. Forecasting wheat production in India: An ARIMA modelling approach, Journal of Pharmacognosy and Phytochemistr. 2019; 8(1):2158-2165.
8. Panse VG. Trends in areas and yields of principal crops in India, Agriculture Situation in India. 1952; 7:144- 48.
9. Panse VG. Recent trends in the yield of rice and wheat in India, Indian Journal of Agricultural Economics. 1959; 14:11-38.
10. 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.
11. Schwarz G. Estimating the dimension of a model, The annals of Statistics. 1978; 62:461-64.
12. Suresh KK, Priya SRK. Forecasting Sugarcane Yield of Tamilnadu Using ARIMA Models, Sugar Tech. 2011; 13(1):23-26.