Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 8 [3] February 2019 : 63-69

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.876
Universal Impact Factor 0.9804
NAAS Rating 4.95

ORIGINAL ARTICLE **OPEN ACCESS**

Wheat Yield Forecasting in Haryana: A Time Series Approach

Ajay Kumar[1], Deepankar[1], P.K Muhammed Jaslam[1*], Anil Kumar[2]
[1]Department of Mathematics and Statistics, College of Basic Sciences and Humanities,
CCS Haryana Agricultural University, Hisar-125004
[2]Department of Statistics, Central University of Haryana, Mahendergarh-123031
*Email: jaslam.stat@hau.ernet.in

ABSTRACT

India is an agrarian country in which Agriculture is the single most important contributor to the Indian economy because of its significance in food security, trade and industry. Wheat is the most important food grains of human in India. Wheat crop is India's prime most staple harvest, placed after Rice. It is mostly consumed in the north and north-west parts of the country. At the time of independence in 1947, production and productivity of wheat were quite low at 6.46 million tones and 663 kg/hectare respectively. At present, India is the second largest producer of wheat in world after China. The purpose of this study was to fit a model that forecast the yield of wheat in Haryana by using annual time series data from 1980-81 to 2009-10. Random walk, random walk with drift, linear trend, moving average, simple exponential smoothing and ARIMA models were employed and compared for finding out a best model to forecast the yield of wheat in Haryana. A software Statgraphics is used to forecast the time series data. The best fitted model was selected based on performance in goodness of fit criterion; Akaike Information Criterion (AIC). This study found that on the basis of AIC, ARIMA(0,2,2) be the best model to forecast wheat yield in Haryana. The forecast value of yield was obtained as 4620.91 kg/hectare in 2017, 4669.28 kg/hectare in 2018, 4717.64 kg/hectare in 2019, 4766 kg/hectare in 2020, 4814.37 kg/hectare in 2021 and 4862 kg/hectare in 2022.
Keywords: *ARIMA; Time Series Models; Forecast; Wheat yield; Exponential smoothing.*

Received 12.11.2018 Revised 24.12.2018 Accepted 09.01.2019

INTRODUCTION

Agriculture is the backbone of India's economy and it contributes to the economic and social well being of the nation through its influence on the gross domestic product (GDP), employment and foreign exchange earnings. Agriculture plays an important role in the betterment of the large proportion of the rural population in particular and overall economy in general. Agricultural development is desired in almost every part of the world today. The race between increasing population and food supply is a real grim.

Forecast is an approach that can help decision-makers, whether from the economic or non-economic fields in making their future decisions with greater accuracy. In fact, forecast is needed by national governments for the establishment of various policy decisions related to storage, distribution, pricing, marketing, import-export, etc. The productivity of agricultural crops is generally characterized by continuous changes due to many factors such as the precipitation fluctuations and the economic, technical and agricultural conditions. The study of the nature and direction of changes in that productivity is helpful in the evaluation of the efforts made to increase agricultural production. Also, the forecast of productivity of various agricultural crops allows making accurate predictions about productivity levels during the coming years. Therefore, forecast represents one of main tools of making efficient development policies and successful economic plans in the field of agricultural production.

Next to rice, Wheat is the most important food-grain of India and is the staple food of millions of Indians, particularly in the northern and north-western parts of the country. Wheat is the main cereal crop and a mainly rabi (winter) season crop in India. The major increase in the productivity of wheat has been observed in the state Haryana, Punjab and UP. Higher area coverage is reported from Madhya Pradesh in recent years. The time of sowing and harvesting differs in different regions due to climatic variations. The sowing of wheat crop normally begins in the September-October in Karnataka, Maharashtra, Andhra

Pradesh, Madhya Pradesh and West Bengal; October-November in Bihar, Uttar Pradesh, Punjab, Haryana and Rajasthan and Nov-Dec in Himachal Pradesh and Jammu & Kashmir. The harvesting is done in Jan-Feb in Karnataka, Andhra Pradesh, M.P., and in West Bengal; March-April in Punjab, Haryana, U.P. and Rajasthan and in April-May in Himachal Pradesh and J&K. The growing period is variable from one agro climatic zone to other that effects the vegetative and reproductive period leading to differences in potential yield. The important factors affecting the productivity are seeding time and methodology, crop establishment and climatic conditions during the growing season.

Wheat cultivation has been suffering from various problems, such as traditional methods of farming, low yields, shortage of key inputs and shortage of irrigation water. On the other hand, farmers do not know future prospect of wheat production and prices while deciding to cultivate this and other crops. Therefore, the objective of this paper is to determine future prospects of wheat in the Haryana using past trends.

Various ARIMA model techniques has been used extensively in the literature for forecasting purpose. Efforts have been made to forecast production and productivity of sugarcane employing ARIMA models [14], forecasting agricultural production at state level [4], forecasting production of oilseeds [3], forecasting and modeling of wheat yield in Pakistan [12] and forecasting productivity in India [9]. Models developed by Mehta *et al*. [8] and Agarwal *et al*. [1] were also used to forecast yields of various crops, and studies done by Bazgeer et al. [2]. Verma et al. [13] have developed and used different indicators in the context of crop yield prediction. Saeed *et al*. [10] forecasted wheat production in Pakistan using ARIMA models, Mandal [7] forecasted Sugarcane production in India, Sahu [11] forecasting irrigated crops like Potato, Mustard and Wheat using ARIMA models.

Karim *et al.* [5] applied regression modeling to forecast wheat production of Bangladesh districts. They used seven model selection criteria's and found that different models were identified for different districts for wheat production forecasts. Iqbal *et al.* [6] used ARIMA model for forecasting wheat area and production in Pakistan. They used ARIMA (1,1,1) model for wheat area forecasting and ARIMA (2,1,2) model for wheat production forecasting. They have found that for 2000-2001 forecasts of wheat area was about 8451.5 thousand hectares.

MATERIAL AND METHODS

Data and Methods

The Haryana state comprising of 22 districts is situated between 74o 25' to 77o 38' E longitude and 27o 40' to 30o 55' N latitude. The total geographical area of the state is 44212 sq. km. The study was carried out on the basis of wheat productivity of Haryana during the years 1980-81 to 2009-10 collected from secondary sources (Ministry of Agriculture, Govt. of India). ARIMA model is the most general form of stochastic models for analyzing time series data. The ARIMA models include autoregressive (AR) terms, moving average (MA) terms, and differencing (or integrated) operations. The model is called AR model if it contains only the autoregressive terms. Model is known as MA model if it involves only the moving average terms. It is known as ARMA models when both autoregressive and moving average terms are involved. Finally when a non-stationary series is made stationary by differencing method, it is known as ARIMA model. The general form of ARIMA is denoted by ARIMA (p,d,q), where 'p' represents the order of autoregressive process, 'q' represents the order of moving average process, while 'd' shows the order of differencing the series to make it stationary.

Time-series analysis is divided into four main forecasting models, namely the deterministic models, the smoothing models, the analytical models and the stochastic models. In this study, we use smoothing model and one type of stochastic models which is the Autoregressive Integrated Moving Average (ARIMA) model. In ARIMA modeling, the order of AR(p) is identified by partial autocorrelation function (PACF) while the order of MA(q) is identified by autocorrelation function. The order of ARIMA (p, d, q) is also identified by model selection criteria's i.e. Schwarz Bayesian information criteria (SBIC) and Akaike's Information Criteria (AIC).

One of the important issues in time series forecasting is to specify model. Time series model is specified on the basis of some information criteria's which includes AIC, BIC likelihood etc. Akaike's (1973) introduced AIC criteria for model specification. AIC is mathematically defined as;

$$AIC = -2\log(\text{maximum likelihood}) + 2k$$

where $k = p + q + 1$ (if model includes intercept) otherwise $k = p + q$. Model specified well if its AIC value is minimum as other fitted models. Other model specification criterion is SBIC and is computed as;

$$SBIC = -2\log(\text{maximum likelihood}) + 2k\log(n)$$

Model which has minimum SBIC value specified well as other fitted models.

It is also important to evaluate and test the results obtained from the forecasting in order to identify the degree of efficiency of the models, and in order to choose the best model among them. There are different indicators that can be used in the evaluation and judgment of the degree of efficiency and accuracy of the model. The main indicators are the Mean Absolute Error (MAE), the Mean Absolute Deviation (MAD), the Root Mean Squared Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the correlation coefficient between actual and forecasted values.

Some models used in the study given in Table 1;

Sr. No	Model name	Model equation
1.	Random walk with drift	$y_t = y_{t-1} + \theta_0 + e_t$
2.	Linear trend	$y_t = a + bt + e_t$
3.	Autoregressive	$y_t = C + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + e_t$
4.	Moving Average	$y_t = C - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$
5.	Simple exponential smoothing	$\hat{y_t} = \alpha y_t + (1-\alpha) y_{t-1}$

Results and Discussion

In this research paper, software Statgraphics is used to forecast wheat production in Haryana. In the Table 2, different models and results are presented with model selection and validity criteria's. Model with the lowest value of the Akaike Information Criterion (AIC) is model G i.e. ARIMA(0,2,2) which has been used to generate the forecasts.

Model Comparison

Number of observations = 30

Models:

(A) Random walk

(B) Random walk with drift = 63.8966

(C) Constant mean = 3530.1

(D) Linear trend = 2506.43 + 66.0434 t

(E) Simple moving average of 2 terms

(F) Simple exponential smoothing with alpha = 0.7532

(G) ARIMA(0,2,2)

(H) ARIMA(2,0,0)

(I) ARIMA(2,2,1)

(J) ARIMA(1,2,1)

Table 2: Best model selection and validity model testing of Wheat production forecasting

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*	*AIC*	*HQC*	*SBIC*
(A)	215.913	156.31	4.45111	63.8966	1.7878	10.7498	10.7498	10.7498
(B)	209.892	164.699	4.70374	-9.40857	-0.05849	10.7599	10.7748	10.8066
(C)	621.919	512.807	16.2058	-1.81899	-3.58824	12.9323	12.9472	12.979
(D)	224.699	195.98	5.85002	3.63798	-0.54258	10.9629	10.9927	11.0563
(E)	217.304	156.0	4.40813	102.536	2.92012	10.8293	10.8442	10.876
(F)	207.282	142.699	4.04788	82.8645	2.34223	10.7348	10.7498	10.7815
(G)	193.111	151.27	4.23863	-21.3289	-0.589514	10.6599	10.6897	10.7533
(H)	194.786	140.177	4.02101	21.7084	0.587142	10.6771	10.707	10.7705
(I)	193.589	148.706	4.1078	-14.8578	-0.405213	10.7315	10.7763	10.8716
(J)	201.18	152.514	4.28318	0.673851	0.0309229	10.7417	10.7716	10.8351

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(A)	215.913	NS	NS	NS	NS	NS
(B)	209.892	NS	NS	NS	NS	NS
(C)	621.919	NS	***	***	***	**
(D)	224.699	NS	**	***	NS	NS
(E)	217.304	NS	NS	NS	NS	NS
(F)	207.282	NS	NS	NS	NS	NS
(G)	193.111	NS	NS	NS	NS	NS
(H)	194.786	NS	NS	NS	NS	NS
(I)	193.589	NS	NS	NS	NS	NS
(J)	201.18	NS	NS	NS	NS	NS

Key:
RMSE = Root Mean Squared Error
RUNS = Test for excessive runs up and down
RUNM = Test for excessive runs above and below median
AUTO = Box-Pierce test for excessive autocorrelation
MEAN = Test for difference in mean 1st half to 2nd half
VAR = Test for difference in variance 1st half to 2nd half
NS = not significant (p >= 0.05)
* = marginally significant (0.01 < p <= 0.05)
** = significant (0.001 < p <= 0.01)
*** = highly significant (p <= 0.001)

This table compares the results of fitting different models to the data. Model G i.e. ARIMA(0,2,2) with the lowest value of the Akaike Information Criterion (AIC) has been used to generate the forecasts. The table also summarizes the results of five tests run on the residuals to determine whether each model is adequate for the data. An NS means that the model passes the test. One * means that it fails at the 95% confidence level. Two *'s means that it fails at the 99% confidence level. Three *'s means that it fails at the 99.9% confidence level. Note that the currently selected model, model G, passes 5 tests. Since no tests are statistically significant at the 95% or higher confidence level, the current model is probably adequate for the data.

Table 3: ARIMA Model Summary

Parameter	*Estimate*	*Standard Error*	*t*	*P-value*
MA(1)	1.47019	0.157156	9.35498	0.000000
MA(2)	-0.516541	0.147366	-3.50515	0.001674

On the basis of above ARIMA model, the estimated wheat forecast model is:

$$y_t = 2y_{t-1} - y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

which is model ARIMA(0,2,2) and assumes that the best forecast for future data is given by a parametric model relating the most recent data value to previous data values and previous noise. The output summarizes the statistical significance of the terms in the forecasting model. Terms with P-values less than 0.05 are statistically significantly different from zero at the 95.0% confidence level. The P-value for the MA(2) term is less than 0.05, so it is significantly different from 0. The estimated standard deviation of the input white noise equals 193.674.

The table also summarizes the performance of the currently selected model in fitting the historical data. It displays:

- root mean squared error (RMSE)
- mean absolute error (MAE)
- mean absolute percentage error (MAPE)
- mean error (ME)
- mean percentage error (MPE)

Each of the statistic is based on one-ahead forecast error, which is the differences between the data value at time t and the forecast of that value made at time t-1. The first three statistic measure the magnitude of the errors. A better model will give a smaller value. The last two statistic measure bias. A better model will give a value close to 0.

Table 4: Estimated Autocorrelations for residuals

Model: ARIMA(0,2,2)

			Lower 95.0%	*Upper 95.0%*
Lag	*Autocorrelation*	*Standard Error*	*Prob. Limit*	*Prob. Limit*
1	-0.0616097	0.188982	-0.370399	0.370399
2	-0.011386	0.189698	-0.371802	0.371802
3	0.153216	0.189723	-0.37185	0.37185
4	-0.191508	0.194091	-0.380413	0.380413
5	-0.115522	0.200726	-0.393417	0.393417
6	0.0428371	0.203087	-0.398044	0.398044
7	0.0276818	0.20341	-0.398676	0.398676
8	0.193337	0.203544	-0.39894	0.39894
9	-0.0924659	0.21	-0.411594	0.411594

This table shows the estimated autocorrelations between the residuals at various lags. The lag k autocorrelation coefficient measures the correlation between the residuals at time t and time t-k. Also

shown are 95.0% probability limits around 0. If the probability limits at a particular lag do not contain the estimated coefficient, there is a statistically significant correlation at that lag at the 95.0% confidence level.
In this case, none of the autocorrelations coefficients are statistically significant, implying that the time series may well be completely random (white noise).

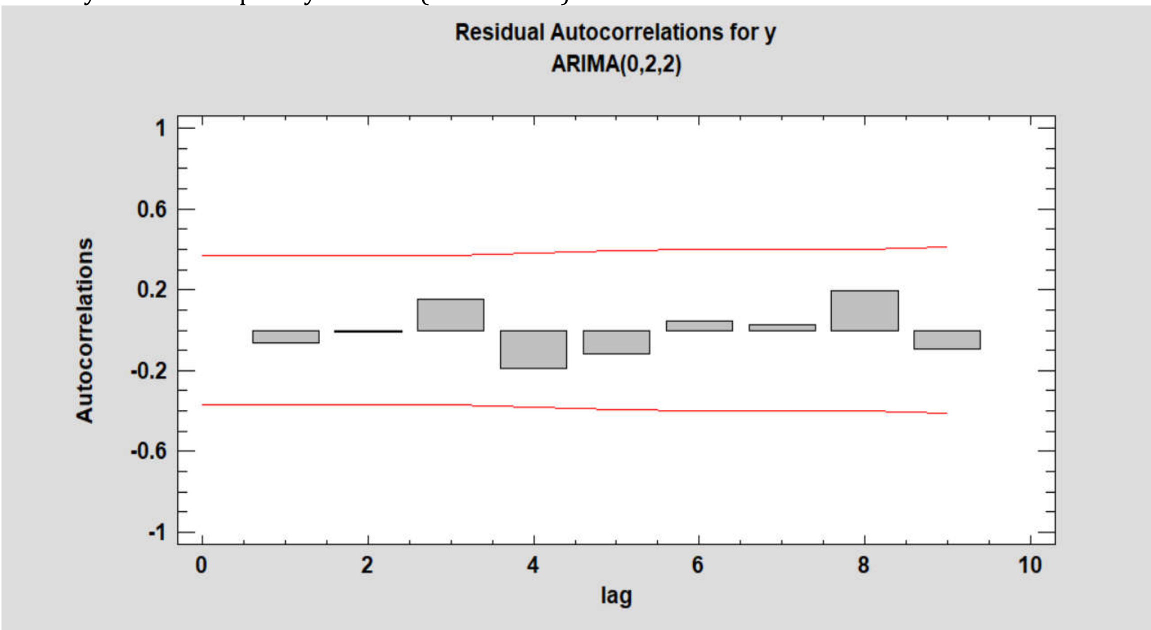

Figure 1: Residuals Autocorrelation Plot of Wheat Yield Model for 1981-2010

Table 5: Estimated Partial Autocorrelations for residuals

Model: ARIMA(0,2,2)

	Partial		*Lower 95.0%*	*Upper 95.0%*
Lag	*Autocorrelation*	*Standard Error*	*Prob. Limit*	*Prob. Limit*
1	-0.061609	0.188982	-0.370399	0.370399
2	-0.015239	0.188982	-0.370399	0.370399
3	0.152178	0.188982	-0.370399	0.370399
4	-0.177786	0.188982	-0.370399	0.370399
5	-0.138178	0.188982	-0.370399	0.370399
6	0.007031	0.188982	-0.370399	0.370399
7	0.092290	0.188982	-0.370399	0.370399
8	0.219994	0.188982	-0.370399	0.370399
9	-0.141263	0.188982	-0.370399	0.370399

This table shows the estimated partial autocorrelations between the residuals at various lags. The lag k partial autocorrelation coefficient measures the correlation between the residuals at time t and time t + k having accounted for the correlations at all lower lags. It can be used to judge the order of autoregressive model needed to fit the data. Also shown are 95.0% probability limits around 0. If the probability limits at a particular lag do not contain the estimated coefficient, there is a statistically significant correlation at that lag at the 95.0% confidence level. In this case, none of the partial autocorrelations coefficients is statistically significant at the 95.0% confidence level.

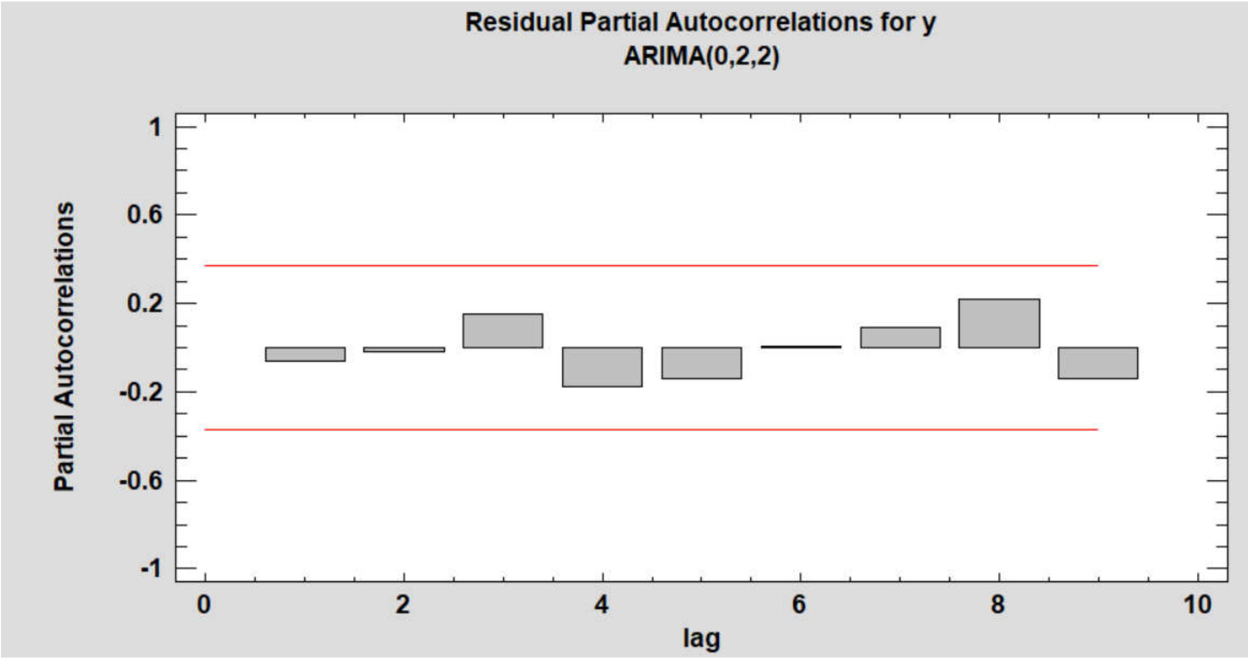

Figure 2: Residuals Partial Autocorrelation Plot of Wheat Yield Model for 1981-2010

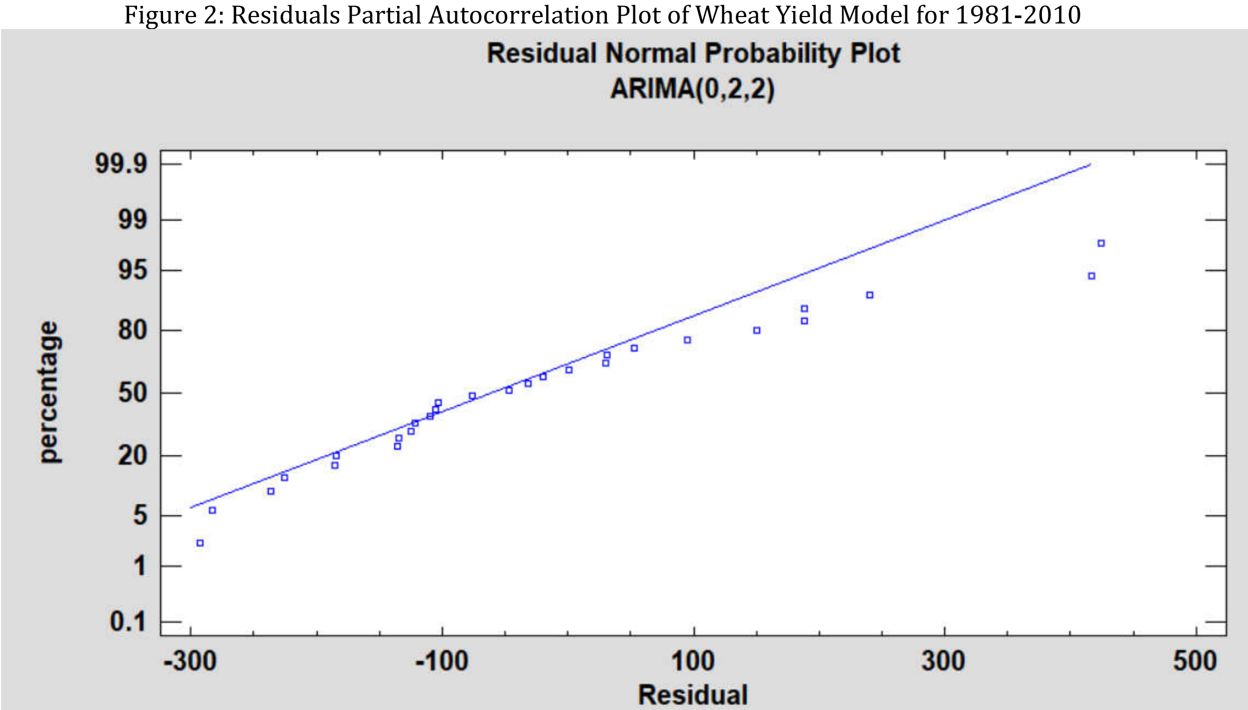

Figure 3: Residual Normal Probability Plot of Wheat Yield Model for 1981-2010

Table 6: Six ahead step forecast and residual for wheat yield data

Period	Observed data (kg/hectare)	*Forecast* (kg/hectare)	Residual
2011	4624	4330.73	294.00
2012	5030	4379.10	650.90
2013	4452	4427.46	024.54
2014	4722	4475.82	246.18
2015	3995	4524.19	-529.19
2016	4407	4572.55	-165.55

This table shows the forecasted values for time period 2011-2016. During the period where actual data is available, it also displays the predicted values from the fitted model and the residuals (data-forecast). forecast the yields for next years as

Table 7: Forecasted value of the Wheat yields for next years

Period	Forecast(kg/hectare)
2017	4620.91
2018	4669.28
2019	4717.64
2020	4766.00
2021	4814.37
2022	4862.00

CONCLUSION

The best model is selected on the basis of model selection criteria AIC. Main interest of developing time series model as other studies is that the model fitted is also satisfied residual assumptions i.e. normality, independence and no autocorrelation. On the basis of model selection criteria, we have found that the best model for wheat production forecasting of Haryana is ARIMA (0, 2, 2). On the basis of this model, we have found that wheat yield of Haryana would become 4620.91 kg/hectare in 2017, 4669.28 kg/hectare in 2018, 4717.64 kg/hectare in 2019, 4766 kg/hectare in 2020, 4814.37 kg/hectare in 2021 and 4862 kg/hectare in 2022.

REFERENCES

1. Agarwal, R.; Jain, R. C. and Mehta, S. C. (2001) Yield Forecast Based on Weather Variables and Agricultural Inputs on Agroclimatic Zone Basis. *Indian Journal of Agricultural Sciences,* 71(7), pp. 487-490.
2. Bazgeer, S., Gh. Kamali and A. Mortazavi (2007) Wheat yield prediction through agro meteorological indices for Hamedan, Iran. *Biaban,* **12**, 33-38.
3. Chandran, K. P. and Prajneshu (2005) Non-parametric regression with jump points methodology for describing country' soil seed yield data. *Journal of Indian Society of Agricultural Statistics,* 59(**2**): 126-130.
4. Indira, R., and A. Datta, (2003) Univariate forecasting of state-level agricultural production. *Economic and Political Weekly*, 38: 1800–1803.
5. Iqbal, N., K. Bakhsh, K. Maqbool, and A. S. Ahmad (2000) Use of the ARIMA model for forecasting wheat area and production in Pakistan. *Int. J. Agri*. Biol. 2: 352-354.
6. Karim, R., A. Awal and M. Akhter (2005) Forecasting of wheat production in Bangladesh. *J.Agri. Soc. Sci*. 1: 120–122.
7. Mandal B. N. (2005) Forecasting Sugarcane Production in India using ARIMA model. *http://interstat.statjournals.net.*
8. Mehta, S. C., Agarwal, Ranjana and Singh, V. P. (2000) Strategies for composite forecast. *Journal of Indian Society of Agricultural Statistics*, 53(**3**), pp. 262-272.
9. Padhan, P. C. (2012) Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture & Social Sciences*, 8: 50-56.
10. Saeed, N., A.Saeed, M. Zakria and T. M. Bajwa (2000) Forecasting of wheat production in Pakistan using ARIMA models. *International journal of Agricultural Biology*, 2(**4**): 352-353.
11. Sahu, P. K. (2010) Forecasting production of major food crops in four major SAARC countries. *International Journal of Statistical Sciences*, Volume 10, pp. 71-92.
12. Ullah, S. B. Din, and G. Haider (2010) Modeling and forecasting wheat yield of Pakistan. *International Journal of Agriculture and Applied Sciences*. 2(**1**): 15-19.
13. Verma, U., Dabas, D. S., Grewal, M. S., Singh, J. P., Hooda, R. S., Yadav, M., Kalubarme, M. H., Sharma, M. P. and Prawasi, R. (2011) Crop yield forecasting in Haryana: 1986 to 2010. *Summary Report*, pp.1-148. Department of Soil Science, CCS HAU, Hisar
14. Yaseen, M., M. Zakria, I. Shahzad, M. I. Khan and M. A. Javed (2005) Modeling and forecasting the sugarcane yield of Pakistan. *International Journal of Agricultural Biology*, 7(**2**): 180–183.

CITATION OF THIS ARTICLE

Ajay Kumar, Deepankar, P.K Muhammed Jaslam, Anil Kumar. Wheat Yield Forecasting in Haryana: A Time Series Approach. Bull. Env. Pharmacol. Life Sci., Vol 8 [3] February 2019: 63-69