International Journal of Agricultural Sciences
Volume **10** | Issue 1| January, 2014 | 158-161

***RESEARCH PAPER***

# Forecasting wheat production using ARIMA model in Punjab

BARUN BISWAS[1], L.K. DHALIWAL*, SOM PAL SINGH AND S.K. SANDHU
School of Climate Change and Agricultural Meteorology, Punjab Agricultural University,
LUDHIANA (PUNJAB) INDIA (Email : dhaliwal1969@yahoo.com)

**Abstract :** In the present study, area, production and productivity data of wheat for Punjab was analyzed by time-series method. Yearly wheat yield data for the period of 1950-51 to 2009-10 was used as input to forecast the yield upto the year 2020-21.The Box-Jenkins ARIMA method was put into use to forecast the yield. The validity of the model were tested by standard statistical techniques. The past 60 years data revealed that wheat yield was increased from 0.8 t $ha^{-1}$ in 1950-51 to 4.3 t $ha^{-1}$ in 2009-10. The model projected 15.3 per cent increase in wheat production in the years to come by 2020-21 in Punjab. Based on ARIMA output, wheat production of Punjab is likely to increase from 15844.7 thousand tons in 2010-11 to 18271.7 thousand tons in 2020-21.

**Key Words :** Wheat production, Forecasting, ARIMA model

**View Point Article :** Biswas, Barun, Dhaliwal, L.K., Singh, Som Pal and Sandhu, S.K. (2014). Forecasting wheat production using ARIMA model in Punjab. *Internat. J. agric. Sci.*, **10** (1): 158-161.

**Article History : Received :** 11.04.2013**; Revised :** 04.10.2013**; Accepted :** 01.11.2013

## INTRODUCTION

Policy making is the process of setting goals, developing strategies and outlining tasks and schedules to accomplish the goals. Crop yield prediction is important for advanced planning, formulation and implementation of policies related with food procurement, distribution and import-export decision. Long-term projections of agricultural produce and operational area help the policy makers regarding resource allocation, price fixation, procurement, marketing and storage. Climate change during last decades aggressively limited our resources for future agriculture. Therefore, proper utilization of available resources is utmost challenging task for the policy makers to pave a smooth path towards sustainable agriculture.

In statistics and econometrics, an autoregressive integrated moving average (ARIMA) model is a univariate time series model which is used for statistical forecasting of crop production. However, because of its power and flexibility, it is a complex technique which requires a great deal of experience. ARIMA produces satisfactory result depending on the researcher's expertise (Bails and Peppers, 1993).

Punjab, known as 'wheat bowl of India', contributes more than 80 per cent of its produce to the central pool. Despite of present anomalies in climate, there is an ample scope for increasing wheat productivity. To meet this challenge there is a need to exploit the potential of rain-fed areas of the state stretching towards lower Shivaliks where prolonged dry spell renders growth of crop. The major objectives of this study are :

– To assess the trends in acreage, yield per hectare and production in Punjab during last 60 years and

– To forecast the production potential of wheat up to 2020-21.

## MATERIAL AND METHODS

The ARIMA methodology developed by Box and Jenkins (1976) has gained enormous popularity in many areas and research practices confirmed its power and flexibility (Hoff, 1983; Pankratz, 1983; Vandaele, 1983). In general, an ARIMA model is characterized by the notation ARIMA ($p$, $d$, $q$) where, $p$, $d$ and $q$ denote orders of auto-regression,

*** Author for correspondence**
[1]Department of Agricultural Meteorology, C.C.S. Haryana Agricultural University, HISAR (HARYANA) INDIA

integration (differentiation) and moving average, respectively. ARIMA technique comprises of linear time series function of past actual values and random shocks. For instance, given a time series process $\{Y_t\}$, a first order auto-regressive process is denoted by ARIMA (1,0,0) or simply AR(1) and is given by:

$$\hat{Y}_t = \pi + \Phi Y_{t-1} + \varepsilon_t$$

and a first order moving average process is denoted by ARIMA (0,0,1) or simply MA(1) and is given by:

$$\hat{Y}_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Alternatively, the ultimately derived model may be a mixture of these processes and of higher orders as well. Thus a stationary ARMA (*p, d, q*) process is defined by the equation:

$$Y_t = « + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \varepsilon_2 \theta_{t-2} + ... - \varepsilon_q \theta_{t-q} + \varepsilon_t$$

where $\varepsilon_t$'s are independently and normally distributed with zero mean and constant variance $\sigma^2$ for t = 1,2,...n. Note here that the values of p and q, in practice lie between 0 and 3.

In the present study, the annual production data of wheat for the years 1950-51 to 2009-10 were used (Anonymous 2010a) to fit an appropriate ARIMA model. The Box-Jenkins methodology for analyzing and modeling a time series involves following steps of model identification, parameter estimation and model validation. Identification of suitable ARIMA model with lowest AIC (*Akaike Information Criterion*) and parameter estimation for forecasting purpose is a tedious job. It is not feasible to simply fit every potential model and choose the one with the lowest AIC. To overcome the above barrier, the function '*auto.arima()* (Hyndman and Khandakar, 2008) available in forecast (Hyndman, 2010 ) library of statistical language tool R (ver. 2.12.2) (Annomyus, 2010b) was used. This function automatically checks the possible models and selects one with lowest AIC value by using appropriate algorithms. To use the above function, first raw data was converted to time series object using *tseries* (Trapletti and Hornik, 2010) library in R.

For the projections to be more precise and realistic we need to specify certain preconditions considered for time series analysis. This includes absence of abnormal climatic conditions and unusual developments such as wide spread flood, viral attacks, effects of pest and diseases.

## RESULTS AND DISCUSSION

The production and yield of wheat in Punjab have observed a several fold increase during past 60 years (Fig. 1). A four fold increase in wheat productivity from 0.8 t ha$^{-1}$ in 1950- 51 to 4.3 t ha$^{-1}$ in 2009-10 represented an increase of nearly 6.6 per cent rate per year, while production increased at a yearly rate of 17.5 per cent. However, cultivated area merely doubled during this period. Highest production as well as yield of wheat was recorded during 1999-2000. Increase in production of wheat in this region is mainly attributed to adoption of HYVs, high dose of fertilizers, improved plant protection measures, efficient irrigation facility, etc. (Singh and Kohli, 2005). As close observation of the production and yield of last decade showed that overall negative trend was registered over these years. Possible cause of such disparity in production might be abnormal climate change, excessive chemical use in recent years etc.

Estimation of ARIMA model involves transformation of the forecasting variables into a stationary series. The most common method is to check stationary series through the graph or time plot of data (Mandal, 2005). Fig. 1 revealed that the data is non-stationary. Non-stationary in mean is corrected through appropriate differencing of the data. ARIMA models were selected using R statistical language. The function *auto.arima()* in forecast library of the language worked out that difference of order 1 was sufficient to achieve stationarity in mean for all the cases. Model diagnostics for all three variables *viz*., yield, production and area is presented in Table 1. A lower value of RMSE (root mean square error) and MAPE (mean absolute percentage error) indicated that identified models were accurate enough to forecast above variables. The projected results are hereunder.

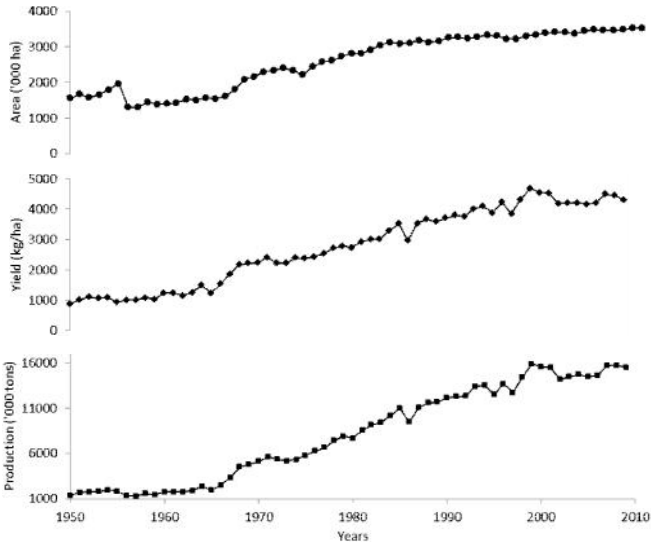


**Fig. 1 : Past trend in production, average yield and area of wheat in Punjab**

### Yield projections:

Point forecast for yield (kg ha$^{-1}$) of wheat crop register an increase of 13.4 per cent by the end of following decade. From 2010-11 to 2020-21 average yield per hectare would increase from 4458.2 to 5053.7 kg ha$^{-1}$ (Table 2). Considerable increase in wheat yield can be achieved in spite of climate change by the policy interventions to encourage cultivation. At 5 per cent probability, the increase in minimum and maximum yield is due to better technology

availability, suitable government policies and irrigation facilities.

**Production projections:**

Using ARIMA (0, 1, 1) the 11 year advance production projection and their 95 per cent confidence interval are calculated and presented in Table 2. The projected results in wheat production indicated trends in production in the coming years. A considerable increase of about 15.3 per cent can be achieved by 2020-21 in Punjab. On the basis of model output an increase in the production from 15844.7 thousand tons in 2010-11 to 17058.2 thousand tons in 2015-16 and 18271.7 thousand tons in 2020-21 is expected (Table 2). The minimum production projection with 95 per cent confidence showed that wheat production may decrease in the coming years and will reach 15051.0 thousand tons during 2020-21 which is at par with the current production level. This may occur due to adverse effect of temperature rise, government attention to the crop and decrease in cultivable land. However, maximum production projections registered a significant increase in production levels which may reach upto 21492.5 thousand tons by 2020-21. Such hike in wheat production was also projected for wheat production in Pakistan (Saboor *et al*., 2003; Saeed, 2000).

**Area projections:**

ARIMA model projected that area under wheat cultivation will increase from 3571.7 thousand hectares in 2010-11 to 3908.5 thousand hectares in 2020-21 (Table 2). This nominal increase may be due to land reclamation and government policy towards wheat production. With 95 per cent confidence interval, the maximum area under wheat would increase from 3806.4 thousand hectares to 4687.0 thousands hectares in 2020-21. The increase in maximum area under wheat is due to increasing irrigation facilities and increasing population pressure of India. On the other hand, decrease in minimum area under wheat production may be due to adverse climatic conditions, unfavourable government policies and land degradation due to changing climatic patterns.

**Conclusion:**

ARIMA model offers a good technique for predicting the magnitude of any variable. Its strength lays the fact that the method is suitable for any time series with any pattern of change. Its limitations include the requirement of a long time series. Above analysis showed a clear direction towards an increased average yield per hectare in the following decade. Such breakthroughs can only be achieved using improved cultivars of wheat, increased water supply, proper farm planning, adequate plant protection measures and mitigation of adverse effect of climate change. Expansion of the area under wheat would directly increase its production in future. Total cultivated area can only be increased through suitable government policies like increase in minimum support price (MSP), crop weather insurance to cover

**Table 1 : Model diagnostics**

| Variable | Model Order | $\sigma^2$ | AIC | RMSE | MAPE |
|---|---|---|---|---|---|
| Yield | ARIMA (0,1,1) | 34862 | 790.69 | 185.15 | 6.06 |
| Production | ARIMA (0,1,1) | 372312 | 930.30 | 605.06 | 8.65 |
| Area | ARIMA (0,1,0) | 14344 | 736.13 | 118.76 | 3.51 |

$\sigma^{2:}$ Variance AIC: Akaike Information Criterion
RMSE: Root Mean Square Error MAPE: Mean Absolute percentage Error

**Table 2 : Projected yield, production and area of wheat in Punjab**

| Year | Projected yield (kg/ha) | | | Projected production ('000 tons) | | | Projected area ('000 hectares) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Point forecast | W.C.I.* | | Point Forecast | W.C.I.* | | Point Forecast | W.C.I.* | |
| | | Minimum | Maximum | | Minimum | Maximum | | Minimum | Maximum |
| 2010-11 | 4458.15 | 4092.20 | 4824.11 | 15844.69 | 14648.77 | 17040.61 | 3571.68 | 3336.94 | 3806.42 |
| 2011-12 | 4517.71 | 4088.86 | 4946.56 | 16087.39 | 14562.77 | 17612.01 | 3605.36 | 3273.38 | 3937.33 |
| 2012-13 | 4577.26 | 4093.63 | 5060.90 | 16330.09 | 14536.02 | 18124.17 | 3639.03 | 3232.45 | 4045.61 |
| 2013-14 | 4636.82 | 4103.99 | 5169.64 | 16572.80 | 14544.75 | 18600.84 | 3672.71 | 3203.23 | 4142.19 |
| 2014-15 | 4696.37 | 4118.54 | 5274.21 | 16815.50 | 14577.82 | 19053.18 | 3706.39 | 3181.49 | 4231.28 |
| 2015-16 | 4755.93 | 4136.34 | 5375.51 | 17058.20 | 14628.91 | 19487.49 | 3740.07 | 3165.08 | 4315.06 |
| 2016-17 | 4815.48 | 4156.79 | 5474.18 | 17300.90 | 14694.04 | 19907.76 | 3773.75 | 3152.68 | 4394.81 |
| 2017-18 | 4875.04 | 4179.43 | 5570.65 | 17543.61 | 14770.53 | 20316.68 | 3807.42 | 3143.48 | 4471.37 |
| 2018-19 | 4934.59 | 4203.93 | 5665.25 | 17786.31 | 14856.42 | 20716.19 | 3841.10 | 3136.88 | 4545.32 |
| 2019-20 | 4994.15 | 4230.04 | 5758.25 | 18029.01 | 14950.30 | 21107.72 | 3874.78 | 3132.47 | 4617.09 |
| 2020-21 | 5053.70 | 4257.55 | 5849.85 | 18271.71 | 15051.04 | 21492.38 | 3908.46 | 3129.92 | 4687.00 |

*With confidence interval (95 %)

weather related damages, cheap irrigation facilities and adequate supply of different farm inputs. So, the above forecasting technique could be a handy tool for policy makers of different public institutes to filter their approach towards wheat crop in Punjab. Further analysis can be done to forecast yield of different major crops of Punjab and other major growing regions of the country.

## REFERENCES

**Bails, D.G. and Peppers, L.C. (1993).** *Business fluctuations: Forecasting techniques and applications*, Englewood Cliffs, Prentice-Hall, New Jercy, pp 607.

**Box, G.E.P. and Jenkins, G.M. (1976).** *Time series analysis: Forecasting and control*, Holden-Day, San Francisco. pp. 575.

**Hoff, J.C. (1983).** *A practical guide to Box-Jenkins forecastin*g, Lifetime Learning Publications, London. p 316.

**Hyndman, R.J. and Khandakar, Y. (2008).** Automatic time series forecasting: The forecast package for R", *J. Statistical Software*, **27** (3) : 1-22.

**Pankratz, A. (1983).** Forecasting with univariate Box-Jenkins models: Concepts and cases, Wiley, New York. pp. 562.

**Saboor, A., Tanwir, F. and Inayat, S. (2003).** Time series analysis of production potential of wheat in Pakistan up to the year 2010-2011, *Internat. J. Agric. & Biol.*, **5** (3) : 329-331.

**Saeed, N., Saeed, A., Zakria, M and Bajwa, T. M. (2000).** Forecasting of wheat production in Pakistan using ARIMA models, *Internat. J. Agric. & Biol.*, **2** (4) : 352-353.

**Singh, N. and Kohli, D.S. (2005).** The green revolution in Punjab, India: The economics of technological change, *J. Punjab Studies*, **12** (2) : 285-306

**Vandaele, W. (1983).** Applied time series and Box-Jenkins models", Academic Press, New York. pp. 417.

### ■ WEBLIOGRAPHY

Annoymous (2010a). State wise area, production and yield of wheat, *Indiastat*, URL: *http:// www. Indiastat. com/ agriculture/2/ foodgrains/17180/wheat/17195/stats.aspx*.

Annoymous (2010b). R: A language and environment for statistical computing", R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, URL: *http://www.R-project.org/*.

**Hyndman, R.J. (2010).** Forecast: forecasting functions for time series, R package version 2.17, URL: *http://CRAN.R-project.org/ package=forecast*.

**Mandal, B.N. (2005).** *Forecasting sugarcane production in India with ARIMA model*, *Interstat*, URL: *interstat.statjournals.net/ YEAR/2005/articles/0510002.pdf*.

**Trapletti, A. and Hornik, K. (2010).** "tseries: Time Series Analysis and Computational Finance", R package version 0.10-25, URL: *http:/ /CRAN.R-project.org/package=tseries*.