

## NEURAL NETWORK AUTOREGRESSION AND CLASSICAL TIME SERIES APPROACHES FOR RICE YIELD FORECASTING

N. Bhardwaj<sup>1</sup>, P. K. M. Jaslam<sup>1\*</sup>, J. K. Bhatia<sup>2</sup>, B. Parashar<sup>3</sup> and Salinder<sup>4</sup>

<sup>1</sup>Department of Mathematics and Statistics, CCS HAU, Hisar (Haryana), India-125004

<sup>2</sup>DHRM, CCS Haryana Agricultural University, Hisar (Haryana), India -125004

<sup>3</sup>Department of Mathematics, JSS Academy of Technical Education, Noida (UP), India-201309

<sup>4</sup>Department of Agriculture & Farmers Welfare, Haryana Government, Panchkula (Haryana), India -134117

\*Corresponding author's email: pkjaslamagrico@gmail.com

### ABSTRACT

This study deals with the application of classical time series models such as double moving average method, exponential smoothing (Brown's double exponential smoothing method and Holt's double exponential smoothing method) along with neural network autoregression model for rice yield prediction in Karnal district of Haryana (India). The district annual time series data on rice productivity were divided into the training data set from 1980-81 to 2013-14 and the test data set from 2014-15 to 2019-20. Among the models fitted, the neural network model had a significantly lower absolute error and the root mean square error. During the test period, a markedly low mean absolute percent deviation was observed for the moving average (7x7) model. However, Future forecast values by this model surpass a fair threshold for rice yield in the country. Considering both the error analysis and model validation results it is found that neural network forecasting models are best fit for forecasting of rice yield in Karnal district followed by Holts double exponential method.

**Keywords:** Artificial neural network, Exponential smoothing, Moving average, Crop yield forecasting

<https://doi.org/10.36899/JAPS.2021.4.0310>

Published online December 15, 2020

### INTRODUCTION

As the highest produced cereal in India, rice is the largest staple food for the largest proportion of the population in the country. Rice crop is produced at an area of 43.79 million hectares with an annual production of 112.91 million tonnes with an average productivity of 2.58 tonnes per ha (2017–18). Normal estimates (2012-13 to 2016-17 average) of area, production and yield of rice crops reveals that approximately 90% of rice crops are grown during the southwest monsoon season (kharif) contributing 87% of annual rice production despite the likelihood of increased dry-season harvests with increased irrigation ([www.agricoop.nic.in](http://www.agricoop.nic.in)). It plays a fundamental role in regional food security and would remain so due to its greater variability to grow under diverse ecosystems. In Haryana State, an area of 14.22 lakh hectares under rice production, contributing 48.80 lakh tones of production with an average productivity of 3.48 tonnes per ha (2017-18) (Haryana Statistical Abstract). The state is blessed with remarkably higher yields and occupies the third place in rice productivity among the rice-growing states of the country. Figure 1 depicts the area and production of rice crop in Karnal district of Haryana. Although there is very slight variation in the rice crop area, there are lot of fluctuations in production data. This information underlines the

importance of studying the trend and predicting the future value of the district's rice crop yield.

On many accounts, proper trend analysis and yield forecast of such an important crop in an agrarian country are of great significance. Trend analysis and forecasting the crop yield are imperative for state agricultural planning and negotiating forward contracts. Critical production and productivity analysis are a prerequisite for proper ecological knowledge base and appropriate research / development efforts to harvest the maximum potential possible. Various statistical techniques were studied to forecast the crop production or yield in agriculture. To mention a few works; Verma *et al.* (2015), Celik *et al.* (2017), Ulla *et al.* (2018), Kumar *et al.* (2019) and Mehmood *et al.* (2019) are working on various forecasting aspects in agriculture. Boken (2000) made an attempt to obtain long-term forecast of average spring wheat yield for Saskatchewan using six forecasting techniques (linear trend, quadratic trend, simple exponential smoothing, double exponential smoothing, simple moving averaging, and double moving averaging). Zahra *et al.* (2015) evaluated the situation of rice regarding its area and yield in Punjab and reported that there is an increasing trend in the area and yield of rice in Punjab. They concluded that the increasing trend would provide a great opportunity for rice export from the country. Karadas *et al.* (2017) attempted to forecast the annual production of some oil seed crops in Turkey

using three exponential smoothing methods *viz.* Holt, Brown, Damped. The study found that Holt's two-parameter exponential smoothing method fits best among exponential smoothing methods. Adisa *et al.* (2019) used artificial neural network (ANN) models for predicting maize in the major maize producing provinces of South

Africa. Celik (2019) concluded in his study to predict future production of mandarin in Turkey that ANN and ARIMA models were successful in forming the Mandarin production model and predicting production. Though, the ANN method provided better results to predict production.

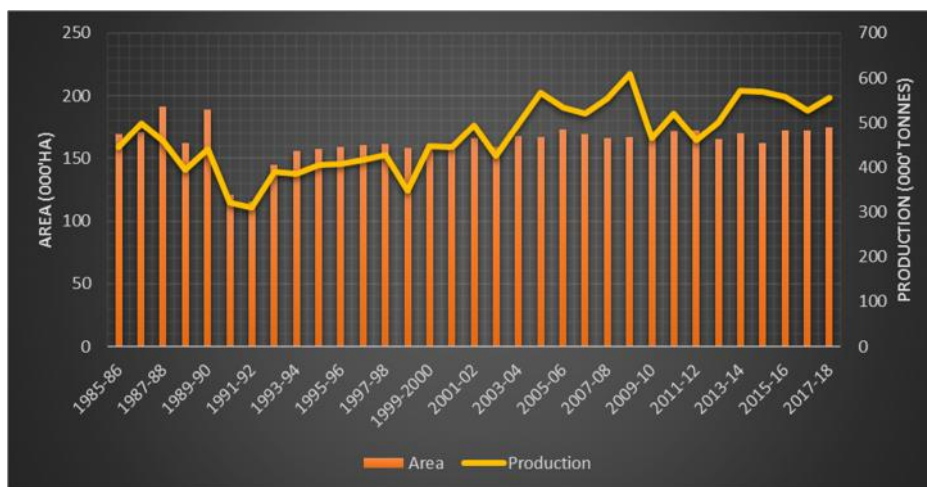


Figure 1. Area and production of rice crop in Karnal district

In the forecast analysis, there are two main types of quantitative methods that can be used: the method of causal or regression, and the method of time series. The first approach suggests that the forecasted variable or dependent variable is causally related to one or more independent variables. In the second method, the forecast process is based on previous observations, which means that it is related to the past in respect to variable and error values (Khayati, 2015). From the beginning, Artificial Neural Networks, commonly known as neural networks, were driven by the fact that the brain computing differs totally from conventional analytical techniques. ANN techniques have turned out to be a very vital tool for a wide variety of applications across many disciplines, including crop production prediction. Using of these sophisticated techniques may be gained importance for better forecasting crop statistics. In this context, this study deals with analyzing the trend and forecasting rice yield in Karnal district of Haryana state by means of neural network auto-regression and few classical time series modeling approaches.

## MATERIALS AND METHODS

**Source of Data:** Time series data for rice yield of Karnal district of Haryana, India, for 40 years (1980-1981 to 2019-2020) were collected from the Directorate of Agricultural Statistics and Crop Insurance, Govt. of Haryana. The first 34-year data were used for model construction (Training data set) and the remaining six-

years data 2013-14 and 2019-20 yield data were used to validate the model (Testing data set).

**Statistical approach:** Moving average (double moving average), exponential smoothing (Browns double exponential smoothing method and Holts double exponential smoothing method) and neural network modelling are used for rice yield forecasting. The comparative performance of different models was accessed on the basis of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and models validated on the test data on the basis of Mean Absolute Percent Deviation (RD %). Time series analysis performed in the Zaitun Time Series v 0.1.4 software according to the procedure given in the Zaitun Time Series Manual. In addition, MS-Excel and SPSS were also used for analysis. Zaitun Time Series is free software developed by the "Time Series" team as the final project of their four-year diploma degrees in Sekolah Tinggi Ilmu Statistik Jakarta, Indonesia.

**Double Moving Average:** There are several methods which can be used to smooth time series data by moving averages. They are the single moving average and the double moving average methods. Both of them use several past data points to forecast the future. In this study we have used double moving average method. The second moving average denoted by MA (T x T) is determined from average of the first moving average.

The method to compute double moving average is:

$$1. \text{ Compute single moving average } S'_n = \frac{y_1 + y_2 + \dots + y_{n-T+1}}{T}$$

2. Calculate adjustment, which is the difference between single MA & double MA ( $S'_n - S''_n$ ) where

$$S''_n = \frac{S'_n + \dots + S'_{n-T+1}}{T}$$

3. Adjust trend from  $n$  to  $n+m$ , if you want to forecast  $m$  period ahead.

**Exponential Smoothing methods:** It's a specific kind of MA technique that is applied to time series data, used to make a smooth data for projection, or to make forecast. This method weights preceding observations by diminishing weights exponentially to the future forecast values.

**Double Exponential Smoothing (Browns):** This method is often referred to as Brown's linear one-parameter method. The following equations are used in double exponential smoothing with Browns method:

$$\text{Single smoothing statistic equation: } S'_n = \alpha Y_n + (1 - \alpha) S'_{n-1}$$

$$\text{Double smoothing statistic equation: } S''_n = \alpha Y_n + (1 - \alpha) S''_{n-1}$$

The equation to estimate forecasting  $m$  forward period with double exponential smoothing with Brown method is  $\hat{Y}_{n+m} = \beta_{0,n} + \beta_{1,n}m$

$$\text{Where, } \beta_{0,n} = 2S'_n - S''_n \text{ and } \beta_{1,n} = \frac{\alpha}{1-\alpha} (S'_n - S''_n)$$

**Double Exponential Smoothing (Holts):** This method is similar to Browns method, but Holts Method uses different parameters than the one used in original series to smooth the trend value. Exponential smoothing prediction can be achieved by using two smoothing constants (with values between 0 and 1) and the following three equations:

$$S_n = \alpha Y_n + (1 - \alpha)(S_{n-1} + T_{n-1})$$

$$T_n = \gamma(S_n - S_{n-1}) + (1 - \gamma)T_{n-1}$$

$$\hat{Y}_{n+m} = S_n + T_n m$$

**Neural Network Time Series Forecasts:** Artificial Neural Networks, or often called Neural Networks, is a computational technique that has made substantial progress recently. Neural networks in a number of scientific fields have shown their ability to handle various problems. Neural networks have a powerful ability known as universal approx. A feed-forward neural network is equipped with lagged input values of  $y$  and a single hidden layer with nodes of height. The inputs are for lags 1 to  $p$ , and  $m$  to  $mP$  is lagged where  $m = \text{frequency}(y)$ . When shipping  $x_{reg}$  the columns will also be used as inputs. If the values in  $y$  or  $x_{reg}$  are absent, the corresponding rows (and any others that depend upon them as lags) are omitted from the fit. A total of repeat networks is fitted, each with random starting weights. These are then averaged when computing forecasts (Hyndman *et al.*, 2019).

To measure the adequacy of the fitted models, Mean Absolute Error (MAE) and Root Mean Squared

Error (RMSE) are utilized and it can be computed as follows:

$$M = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad R = \sqrt{\frac{1}{n} \sum e_i^2}$$

where,

$Y_i$  = Observed value of  $y$

$\hat{Y}_i$  = Estimated values from the estimated equation that correspond to each  $y$  value

$e$  = The error term ( $Y - \hat{Y}_i$ )

$n$  = Sample size.

Reliability in numerical models are quantified by validation. The process of determining the degree to which a model is an accurate representation of the real world, is known as validation. The forecasting performance of the developed model is compared in relation to the state Department of Agriculture yield estimates using Percent Relative Deviation (RD%). This measure the deviation (in percentage) of forecast yield from the observed yield and is measured as:

$$\text{Percent Relative Deviation} = \{(\text{observed yield} - \text{forecasted yield}) / \text{observed yield}\} \times 100$$

## RESULTS AND DISCUSSION

In this study, various time series analyses were used to forecast rice yield in Karnal district from 2020-21 to 2024-25. Figure 2 shows the time series of rice yield for the period 1980 to 2019, and it can be seen from the graph that a stochastic trend has been observed in the series. To infer the trend more precisely, time series autocorrelation (ACF) and partial autocorrelation (PACF) functions were examined. In order to obtain the stationary status of the series, the first differences of the series are considered. ACF and PACF graphs of the first-difference series created to include a stationary state are shown in Figure 3.a) and Figure 3.b) respectively, and the graphs show evidence of a stationary state.

The actual and predicted value of the yield is given in Figure 4. The double moving average, MA (7 x 7) has lower values of MAE (302.62) and RMSE (394.97) compared to all other moving average length (order). Analysis of exponential smoothing techniques that weights preceding observations by lessening weights exponentially to estimate future values also carried out with optimal values of the smoothening parameters obtained through grid search method. The MAE and RMSE values of Browns double exponential smoothing method with  $(\alpha) = 0.1$  & Holts double exponential smoothing method with  $\alpha$  parameter (for data) = 0.1 and  $\gamma$  parameter (for trend) = 0.1 have been shown in figure 5. The ANN method with 1 input layer and 1 neural hidden layer with hyperbolic tangent activation function has significantly lower MAE (218.65) and RMSE (274.94) values compared to the rest of the models.

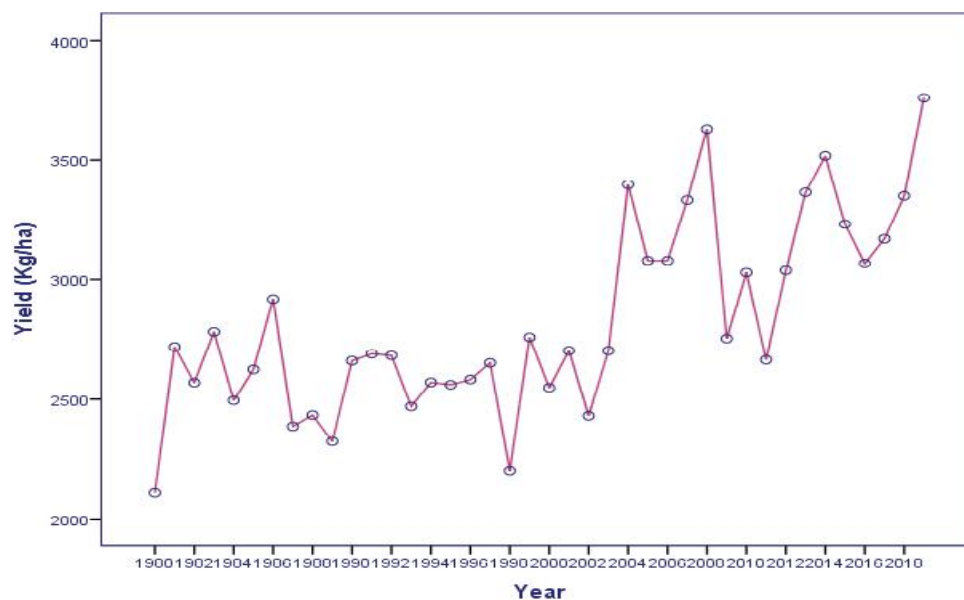


Figure 2. Rice Yield, 1980-2019 (Kg/ha)

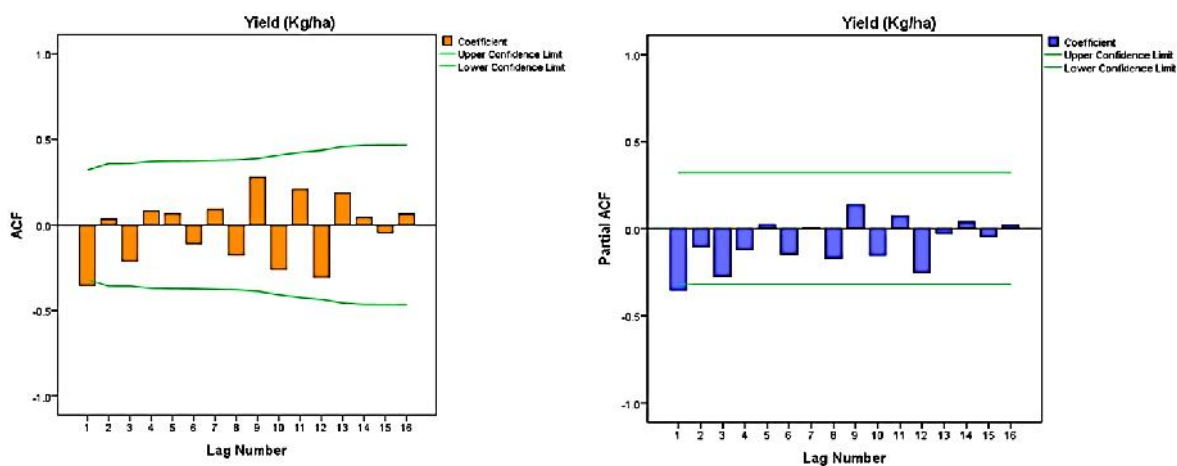


Figure 3. a) ACF plot of first differences b) PACF plot of first differences

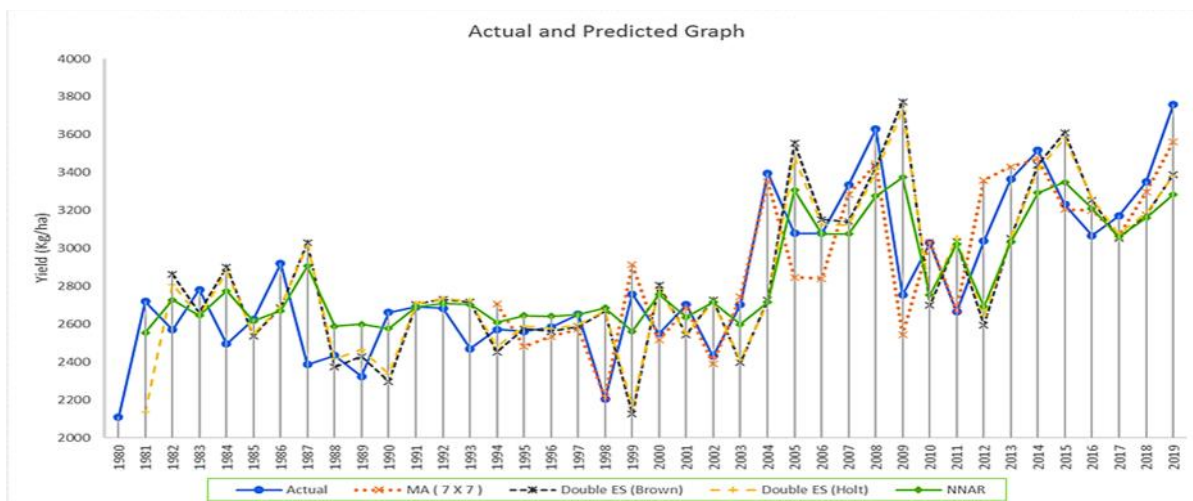


Figure 4. The actual and predicted value of the rice yield (kg/ha)



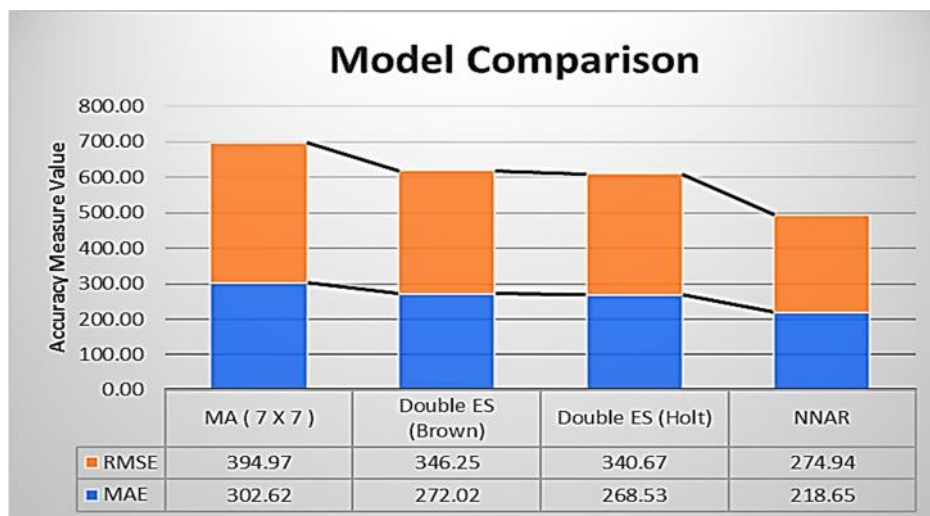


Figure 5. Error estimation plot

The result of the comparison between observed and model-based estimates for the test data set in terms of RD percent is shown in Table 1. Although the average absolute percent deviation for the MA (7x7) model is very low (2,759 percent), this model cannot be

recommended as future forecast values are beyond imaginary for rice yield. Considering both model adequacy and model validation, the ANN method followed by Holts double exponential method can more accurately forecast the future rice yield.

Table 1. Comparison between observed and model-based estimates.

Test Year	Observed Yield (Kg/ha)	MA ( 7 X 7 )		Double ES (Brown)		Double ES (Holt)		NNAR	
		Estimated Yield (Kg/ha)	R D (%)	Estimated Yield (Kg/ha)	R D (%)	Estimated Yield (Kg/ha)	R D (%)	Estimated Yield (Kg/ha)	R D (%)
2014-15	3517	3473	1.25	3441	2.15	3416	2.88	3293	6.38
2015-16	3232	3205	0.83	3611	-11.73	3581	-11.63	3349	8.10
2016-17	3068	3199	-4.28	3254	-6.05	3261	-6.29	3211	-4.65
2017-18	3171	3065	3.35	3052	3.75	3076	3.01	3063	3.41
2018-19	3350	3296	1.61	3177	5.18	3187	4.87	3161	5.63
2019-20	3760	3564	5.23	3389	9.86	3383	10.04	3285	12.63
<b>Avg. Abs. percent dev.</b>			<b>2.759</b>		<b>6.453</b>		<b>6.453</b>		<b>6.800</b>

At the next stage, the forecasting series were graphed together with observation values of the original series. Obtained graph is depicted in Figure 6. Rice crop productivity (yield) is forecasted for the year 2020-21 to 2024-25 based on various developed models and presented in Table 2.

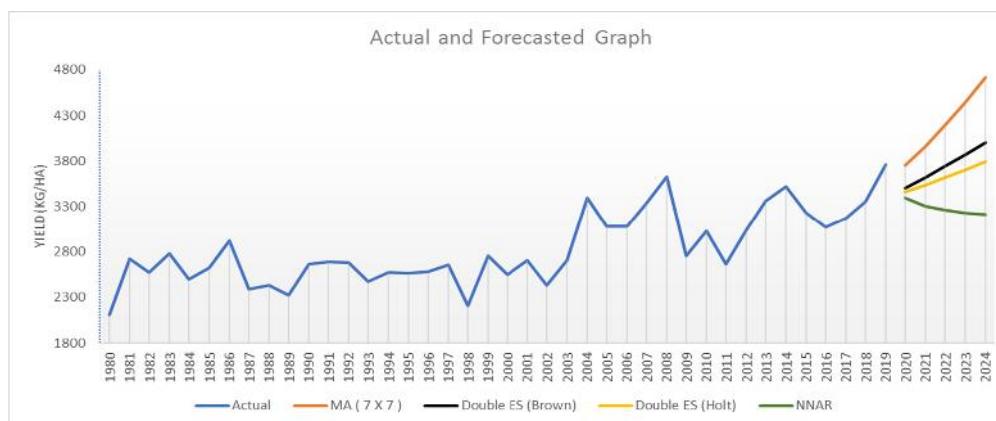


Figure 8. forecasting series with observation values of the original series

**Table 2. Forecasted values of rice yield for Karnal district.**

Forecast Year	MA (7 X 7)	Double ES (Brown)	Double ES (Holt)	NNAR
2020-21	3752	3503	3457	3391
2021-22	3961	3621	3536	3305
2022-23	4191	3744	3619	3260
2023-24	4443	3871	3706	3231
2024-25	4716	4002	3797	3210

**Conclusion:** Crop forecasting is a formidable challenge, and it is important in every region in order to planning and policy making future for food providing distribution, pricing and also its import and export. For the rice yield forecasting, various time series models and artificial neural network forecasting methods were used. Considering both the error analysis and model validation results it is found that neural network forecasting models are best fit for forecasting of rice yield in Karnal district followed by Holts double exponential method. This ANN model provided a forecasted yield estimate of 3391, 3305, 3260, 3231, 3210 Kg/ha for the forecast year 2020-21 to 2024-25, respectively. The government, agro-based industries, traders and farmers alike will be beneficial with these forecasted estimates.

**Acknowledgements:** We would like to thank the editor and anonymous reviewers who have improved the manuscript by making valuable comments.

## REFERENCES

- Adisa, O.M., J.O. Botai, A.M. Adeola, A. Hassen, C.M. Botai, D. Darkey and E. Tesfamariam (2019). Application of artificial neural network for predicting maize production in South Africa. *Sustainability*, 11, 1145: doi:10.3390/su11041145.
- Boken, V.K. (2000). Forecasting spring wheat yield using time series analysis: a case study for the Canadian Prairies, *Agron. J.*, 92(6):1047-1053.
- Celik, S., K. Karadas and E. Eydurán (2017). Forecasting the production of groundnut in Turkey using ARIMA model. *J. Anim. Plant Sci.* 27(3): 920-928.
- Celik, S (2019). Prediction of mandarin production in turkey through artificial neural networks and time-series analysis. *Int. J. Trend Res. Dev.*, 6(5): 85-90.
- Department of Agriculture, Cooperation & Farmers Welfare, Ministry of Agriculture & Farmers Welfare, Government of India (<http://www.agricoop.nic.in>).
- Karadas, K., S. Celik, E. Eydurán and S. Hopoglu (2017). Forecasting production of some oil seed crops in Turkey using exponential smoothing methods. *J. Anim. Plant Sci.*, 27(5): 1719-1729.
- Khayati, A (2015). Forecasting major vegetable crops productions in Tunisia. *Int. J. Res. Bus. Stud. and Manag.* 2(6):15-19.
- Kumar, A., Deepankar, P. K. M. Jaslam and A. Kumar (2019). Wheat yield forecasting in Haryana: A time series approach. *Bull. Envir. Pharmaco. Life Sci.*, 8(3): 63-69.
- Mehmood, Q., M.H. Sial, M. Riaz and N. Shaheen (2019). Forecasting the production of sugarcane in Pakistan for the year 2018-2030, using Box-Jenkin's methodology. *J. Anim. Plant Sci.* 29(5): 1396-1401.
- State Statistical Abstract of Haryana. (2017), Department of Economic and Statistical Analysis, Haryana. (<http://esaharyana.gov.in/en-us/State-Statistical-Abstract-of-Haryana>)
- Ullah, A., D. Khan and S. Zheng (2018). Forecasting of peach area and production wise econometric analysis. *J. Anim. Plant Sci.* 28(4): 1121-1127
- Verma, U and A. Goyal (2015). Linear mixed modeling for mustard yield prediction in Haryana State (India). *J. of Math. Stat. Sci.*: 96-105.
- Zahra, N., N. Akmal, S. Farooq, N. Habib, S. Siddiqui & I. Raza (2015). Trend Analysis of Rice Area and Yield in Punjab. *Pakistan J. Agric. Res.*, 28(4):439-444.

Copyright of JAPS: Journal of Animal & Plant Sciences is the property of Knowledge Bylanes and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.