J. Agr. Sci. Tech. (2022) Vol. 24(2): 261-273

Forecasting Wheat Production in Iran Using Time Series Technique and Artificial Neural Network

Z. Latifi[1*] and H. Shabanali Fami[2]

ABSTRACT

With the increase of the world population, the worries and concerns for food supply increase too. Wheat, as one of the most important agricultural products, which is widely consumed all over the world, has a very important role in people's nutrition, particularly among Iranians, the diet of whom is highly dependent on bread. Product forecasting is critical for any country so that decisions about storage, import or export, etc. can be planned. In this paper, several univariate time series models and the Artificial Neural Network (ANN) model are used to forecast wheat production in Iran. Annual wheat production, total annual precipitation, total applied fertilizer, population, and wheat cultivated area data were used in the period between 1961-1962 to 2018-2019. With the minimum values of 1.45894, 1.00329, 1.0448, and 1.09742 obtained for RMSE, AIC, HQC, and SIBC criteria, respectively, Autoregressive Integrated Moving Average (ARIMA) (1,1,1) was selected as the best univariate model. In testing the ANN models, total annual precipitation, total applied fertilizer, population, and wheat cultivated, area as input variables, and wheat production, as output variable, were used. Among several NN models, the Multilayer Perceptron Neural Network (MLP-NN) model with five hidden layers had the lowest MSE= 0.153 and was chosen in this study. Comparison between the ANN model and the ARIMA (1,1,1) model showed that RMSE= 0.391, MSE= 0.153, and MAPE= 0.4231 in the ANN model were much lower than that of the ARIMA (1,1,1) model. The results showed the power of ANN models to predict wheat production using efficient parameters, as compared to the ARIMA model.

Keywords: Agriculture Production, Autoregressive Integrated Moving Average Model.

INTRODUCTION

Agricultural products play a major role in providing the food needed for the world population; this is especially evident in developing countries including Iran (Salami, and Mohtashami, 2014). In Iran, as in other developing countries, agriculture is considered as one of the most important economic sectors that account for a significant and high percentage of production and employment (Latifi and Shabanali Fami, 2020; Farajzadeh and Shahvali, 2009). Wheat, due to its important role in the political and economic arena of different countries, is regarded as a strategic crop all over the world; particularly in the developing countries. The economic importance of wheat, in terms of production and nutrition, is higher than the other agricultural products in the world. Wheat is the most important agricultural product of Iran in terms of production and area under cultivation, and the Increasing wheat production is receiving more attention these days and is of great importance from the economic point of view and the supply of the main food (Shahriar and Ghashghaei,

[1] Department of Mechanical Engineering, Faculty of Engineering, Shohadaye Hoveizeh Campus of Technology, Shahid Chamran University of Ahvaz, Susangerd, Islamic Republic of Iran.
[*] Corresponding author; e-mail: z.latifi@scu.ac.ir
[2] Department of Agricultural Development and Management, Faculty of Agricultural Economics and Development, College of Agriculture and Natural Resources, University of Tehran, Karaj, Islamic Republic of Iran.

2018). Cereals, including wheat, can meet the humans' daily needs, including carbohydrates, proteins, fats, minerals, and some vitamins; provided that bran is not completely absorbed (Shewry and Hey, 2015). As economic and agricultural experts have acknowledged, wheat production in Iran does not have a favorable increasing trend because of the natural geographical problems and lack of financial and installation resources in the field of irrigation and drainage networks. On the other hand, traditional and rainfed cultivation is prevalent. Also, there is a lack of mechanized operations, lack of quality seeds and suitable chemical fertilizers, soil and water problems, and lack of access to technology and modern science, poor marketing, etc. (Ministry of Agriculture Jihad, 2020). Wheat production time series trend plot from 1961-1962 to 2018-2019 is shown in Figure 1. According to the Food and Agriculture Organization of the United Nations (FAO, 2020), between 1961-1962 and 2018-2019, wheat production had an increasing trend in the world and more than tripled.

Forecasting of agricultural products is essential for farmers, agricultural industries, and, specially, governments; through which, based on estimates of domestic agricultural products, they can make the necessary planning. According to production forecasts, planners must decide on the target level of production so that people's demands are met in the future. Low production can lead to more production gaps for a particular commodity in the country and thus it might lead to serious food insecurity, especially in emergency conditions such as floods, earthquakes, etc. Many studies have shown that providing planners and decision-makers with forecasts can lead to better production planning decisions (Zinyengerea, *et al*., 2011, Goodwin *et al*., 2010). Strategies for production and pricing systems and interregional food movements might need rearrangements.

In this area, statistics plays an important role in obtaining valid results. Several statistical and economic models have been developed to predict various topics, including agricultural products (Hanke and Wichern, 2008). Many studies, using different time series models, have predicted data. To mention some studies; some time series models have used and fitted to forecast gold prices by Deepika *et al*. (2012), accident cases by Balogun *et al*. (2015), agricultural production by Paul (2015), and Paul and Sinha (2016). Amin *et al*. (2014) developed various time series models to forecast wheat production of Pakistan. The best model, ARIMA (1,2,2) was selected. This model was used to forecast the data. Safa *et al*. (2015) used the

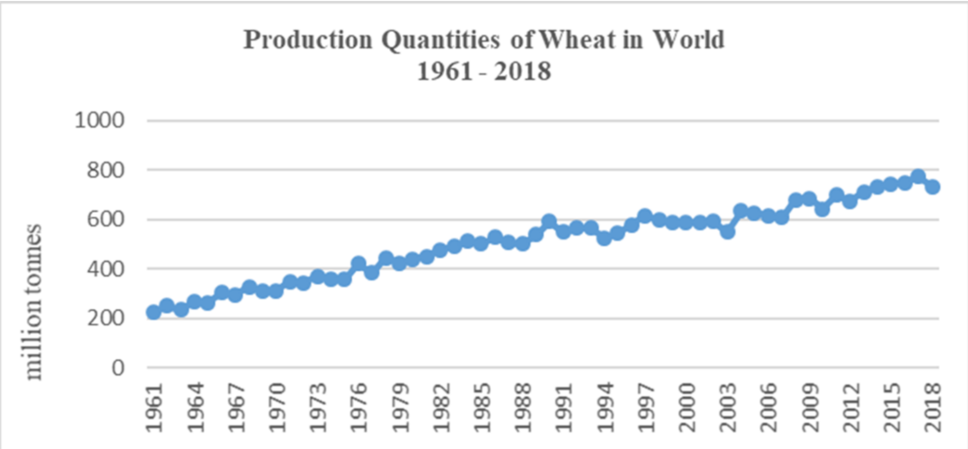

Figure 1. Wheat production time series trend plot (FAO).

Artificial Neural Network (ANN) to model wheat production in Canterbury, New Zealand. Some factors were selected as influential inputs into the model. The final ANN model can predict wheat production based on farm conditions, machinery conditions, and farm inputs. By using various time series models, Asif Masood *et al*. (2018) tried to forecast wheat production in Pakistan. Considering the values close to the production forecast with previous years, the ARIMA model was found to be appropriate. Niazian *et al*. (2018) applied ANN along with the MLR model to predict the seed yield of ajowan through seed yield components. The results showed that the performance of ANN was better than MLR. Nath *et al*. (2019) used Box-Jenkins' ARIMA model to forecast wheat production in India. By fitting ARIMA (1,1,0) model to the data, The results showed an increase in wheat production in the future. Niedbala and Kozlowski (2019) used three independent models for the prediction of yields of winter-cultivated wheat by Artificial Neural Networks with MLP topology, based on meteorological data (air temperature and precipitation) in Poland. The results showed that, among different factors, mean of air temperature had the greatest impact on winter wheat yield. Hashemi Nejad *et al*. (2020) tried to explore factors affecting wheat production risk in the bread supply chain in Iran. Using regression analysis, results revealed that wheat production risk was affected by population, wheat imports, rainfall, wheat guaranteed prices, harvested area, and wheat axial plan variables that population, import, rainfall, and the harvested area had a positive effect and guaranteed price and wheat axle plan harmed wheat production risk. Patryk *et al*. (2021) paid attention to environmental variables, such as climatic data, air temperature, and total precipitation, and soil parameters. Their study emphasized that the increasingly common use of remote sensing and photogrammetric tools enables the development of precision agriculture.

By analyzing the available data, the present study aimed to find the best model for predicting wheat production in Iran so that the results would be useful for farmers, researchers, and the government.

MATERIALS AND METHODS

In this study, the data of annual wheat production, total annual precipitation, total applied fertilizer (including nitrogen, potash, and phosphate fertilizers), population, and wheat cultivated area of Iran during 1961-1962 to 2018-2019 were used (FAO, 2020; CCKP, 2020). In the first step, several univariate time series models such as simple random walk, random walk with drift, linear trend, quadratic trend, simple moving average, simple exponential smoothing, double exponential smoothing, exponential trend, s-curve trend, and Autoregressive Integrated Moving Average (ARIMA) models were fitted to the data to select the best model for forecasting wheat production data. In univariate models, the total wheat production data was divided into training and testing data; training data included 1961-1962 to 2014-2015, and testing data included 2015-2016 to 2018-2019. In the second step, the ANN model was fitted to the data and, finally, the results of univariate models and ANN model were compared.

ARIMA Model

In the 1970s, Box-Jenkins developed the Autoregressive Integrated Moving Average (ARIMA) method, which was used by statisticians and economists to extract a model that would produce and predict time series. This method includes four stages of identification, estimation, diagnosis, and prediction. ARIMA method models static time series based on its past values and error sentences; therefore, it is a parametric method and no independent variables are used. ARIMA (p,q) model is shown as follows:

$$\begin{cases} Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} \\ +U_t + \theta_1 U_{t-1} + \theta_2 U_{t-2} + \cdots + \theta_q U_{t-q} \quad (1) \\ \qquad U_t \sim iid(0, \sigma_U^2) \end{cases}$$

Where, Z_t is the original series, for every t, we assume that is independent of $Z_{t-1} + Z_{t-2} + \cdots + Z_{t-p}$. The ultimate goal of the proposed Box-Jenkins model is prediction; therefore, the time series used must be static, because the instability of the time series cause the prediction of the future values of the series to be affected by a random or definite trend in them and thus affecting the results (Gujarati, 2004). Therefore, if we use model ARIMA (p,q) for a non-static time series accumulated of order d, model ARIMA (p,d,q) would be obtained. In the identification stage, p and q values are determined using Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values.

Artificial Neural Network

Artificial Neural Network (ANN) is used to forecast and model non-linear time series data. This model involves input, hidden, and output nodes. The best network model can be achieved from the appropriate integration of the number of input, hidden, and output nodes that are influenced by the weighted connectivity of each node. Multi-Layer Perceptron (MLP), a back propagation algorithm, is the most common learning algorithm. In MLP models, the output is the function of the linear combination of hidden unit activations; each one is a non-linear function of the weighted sum of inputs (Azadeh *et al*., 2006).

$$\hat{Z} = F(\alpha_0 + \sum_{j=1}^{m} H(\beta_j + \sum_{i=1}^{n} x_i w_{ij})\, \alpha_j) \quad (2)$$

Where, $\hat{Z}$ is the network output; α_0, and β_j are the output and hidden units biases, respectively. w_{ij} is the weight from input layer i to hidden unit j, and α_j is weight from hidden unit j to output. x_i is the input vector for unit i. F, and H are the output and hidden unit activation functions (Cheng and Titterington, 1994).

Diagnostic Measures

In this paper, Statgraphics, and Matlab software are used for analyzing wheat production in Iran. If the residuals are obtained randomly, then the models fitted to the data will be acceptable. After fitting different suitable models, the ACF and PACF of these models' residuals are estimated. Three tests were used to test for residual randomness based on ACF and PACF, including the followings:

1. Runs above and below median and counts the number of times the series goes above or below its median. This number is compared to the expected value for a random time series. Small P-values (less than 0.05 if operating at the 5% significance level) indicate that the residuals are not purely random.
2. Runs up and down and counts the number of times the series goes up or down. This number is compared to the expected value for a random time series. Small p-values indicate that the residuals are not purely random.
3. Ljung-Box Test, which constructs a test statistic based on the first k residual autocorrelations. As with the other two tests, small P-values indicate that the residuals are not purely random (Forecasting Statgraphics 18, 2017).

Since the P-values for all three tests are well above 0.05, there remains no reason for doubting that the residuals are white noise (Box *et al*., 2008).

To measure the accuracy of the fitted model the following methods are used:

1. RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} e_{n+i}^2}{m}}$$

Where, e_{n+i} is the residual term of $(n+i)^{the}$ observation, and m is the number of observations.

2. MAE (Mean Absolute Error): $MAE = \frac{\sum_{i=1}^{m} |e_{n+i}|}{m}$

3. MAPE (Mean Absolute Percentage Error): $MAPE = 100\% \frac{\sum_{i=1}^{m} \left|\frac{e_{n+i}}{Z_{t+i}}\right|}{m}$

4. ME (Mean Error): $ME = \frac{\sum_{i=1}^{m} e_{n+i}}{m}$

5. MPE (Mean Percentage Error): $MPE = 100\% \frac{\sum_{i=1}^{m} \frac{e_{n+i}}{Z_{t+i}}}{m}$

Better models have smaller RMSE, MAE, and MAPE values that measure the variance of the forecasting errors. ME and MPE are measures of bias and should be close to 0 (Forecasting Statgraphics 18, 2017); therefore, the minimal values of these measures suggest a better model with minimum forecasting error (Karim and Akhter, 2010).

6. AIC (Akaike Information Criteria): $AIC = -2\ln(\hat{L}) + 2k$

Suppose $\hat{L}$ as the maximum value of the likelihood function for the model and k as the number of estimated parameters in the model (Burnham and Anderson, 2002; Akaike, 1974). The model is defined well in case that its AIC value, compared to other fitted models, is minimal (Tsay, 2005).

7. HQC (Hannan-Quinn Criteria): $\mathrm{HQC} = -2L_{max} + 2k\ln(ln(m))$

8. SBIC (Schwarz Bayesian Information Criteria): $\mathrm{SBIC} = -2\ln(\hat{L}) + 2k\ ln(m)$

Where, L_{max} is the log-likelihood. The model with the minimum SBIC value is specified well as other fitted models (Tsay, 2005).

RESULTS AND DISCUSSION

Descriptive statistics of wheat production data are shown in Table 1. There are descriptive statistics such as average, standard deviation, coefficient of variation, minimum, maximum, range, standard skewness, and standard kurtosis. Standard deviation is one of the scattering indices that averagely shows how far the data is from the mean. If the standard deviation of a set of data is close to zero, it indicates that the data are close to the mean and thus have little scatter; whereas, a large standard deviation indicates a significant dispersion of data. The standard deviation value was 3707530 tons that indicates a significant dispersion of the data. The coefficient of variation expresses the scattering rate per unit of average. Of particular interest are the standardized skewness and standardized kurtosis, which can be used to determine whether the data come from a normal distribution or not. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate any statistical test regarding the standard deviation (Forecasting Statgraphics 18, 2017). In this case, the standardized skewness value would be within the range expected for the data from a normal distribution. The standardized kurtosis value would be within the range expected for the data from a normal distribution (Statgraphics output).

Estimation of the Univariate Time Series Model

The results of fitting different univariate models to the data of wheat production are compared in Table 2. The model with minimal AIC, HQC, and SBIC values is model M [ARIMA (1,1,1)], which has been selected to generate the forecast data.

The output of the tests that are run on the

Table. Summary statistics for annul wheat production data of Iran (1961-1962 to 2018-2019).

Average	Standard deviation	Coefficient of variation	Minimum	Maximum	Range	Standard skewness	Standard kurtosis
8165160	3707530	45.4067%	2468140	15886600	13418500	1.17502	-1.44854

Table 2. Selecting the best model based on criteria.

Model[a]	RMSE	MAE	MAPE	ME	MPE	AIC	HQC	SBIC
(A)	1.76222	1.02237	12.786	0.204051	0.960703	1.13315	1.13315	1.13315
(B)	1.76593	0.983245	12.5972	-1.09075E-16	-2.13781	1.17184	1.18567	1.20736
(C)	3.70753	3.14473	51.482	-1.01069E-15	-26.7573	2.65522	2.66905	2.69074
(D)	1.74945	1.19668	13.7903	-8.26925E-16	-3.34993	1.18757	1.21525	1.25862
(E)	1.74397	1.23766	14.6387	-5.8191E-16	-2.50275	1.21578	1.25729	1.32235
(F)	2.01326	1.39553	16.2325	0.112803	-2.13274	1.46848	1.49615	1.53953
(G)	3.15775	2.53492	34.1329	0.642277	-7.51456	2.36868	2.39636	2.43973
(H)	1.76048	1.13828	13.6493	0.30912	2.08502	1.16566	1.1795	1.20118
(I)	1.67147	1.0487	12.7207	0.299341	2.05011	1.06189	1.07573	1.10987
(J)	1.78783	1.07574	13.4952	0.115115	0.138071	1.19648	1.21032	1.23201
(K)	1.67398	0.981683	13.0894	-0.137132	-4.64088	1.09937	1.12704	1.17042
(L)	1.88294	1.13831	14.1924	0.0833133	-0.777149	1.30015	1.31399	1.33567
(M)	**1.45894**	**1.00611**	**12.6793**	**-0.114894**	**-4.48429**	**1.00329**	**1.0448**	**1.09742**
(N)	1.56819	0.956025	11.9992	-0.0795202	-3.05529	1.03128	1.14198	1.31548
(O)	1.54853	0.971844	12.3016	-0.112574	-4.16063	1.04703	1.11622	1.22465
(P)	1.61492	1.10924	13.4262	-0.129042	-4.85462	1.06202	1.10353	1.16859
(Q)	1.53634	0.941404	12.0939	-0.149372	-4.89793	1.0657	1.14873	1.27885

[a] (A) Random walk; (B) Random walk with drift= 0.204051; (C) Constant mean= 8.16516; (D) Linear trend= 2.44046+0.194058t; (E) Quadratic trend= 1.81334+0.25677t+-0.00106292t^2; (F) Exponential trend= Exp (1.19985+0.0267136t); (G) S-curve trend= Exp (2.15105+-2.03666/t); (H) Simple moving average of 2 terms; (I) Simple exponential smoothing with alpha= 0.6588; (J) Brown's linear exponential smoothing with alpha= 0.3307; (K) Holt's linear exponential smoothing with alpha= 0.6236 and beta= 0.0139; (L) Brown's quadratic exponential smoothing with alpha= 0.2021; (M) ARIMA (1,1,1) with constant; (N) ARIMA (3,1,4) with constant; (O) ARIMA (0,1,4) with constant; (P) ARIMA (0,1,2) with constant, (Q) ARIMA (2,1,3) with constant.

residuals to determine the suitability of each model for the data are shown in Table 3. An OK means that the model accepted the hypothesis; that is the residuals are purely random. Symbols "*", "**", and "***" mean that the model rejected the hypothesis at the 95, 99, and 99.9% confidence level, respectively, which means that the residuals are not purely random. In Table 3, the model M i.e. ARIMA (1,1,1), passed 4 tests (Statgraphics output).

Summary of ARIMA Model Forecast

Table 4 is the summary of the best univariate model i.e. ARIMA (1,1,1). This table shows the parameters of the model and their significance. At the 95% confidence level, if the p-value is less than 0.05, statistically the parameter is significantly different from zero. P-values for AR (1), MA (1), and the constant parameters are less than 0.05; so, it is significantly different from zero. The estimated standard error of the input white noise equals to 1.5683 (Statgraphics output).

Estimated Autocorrelations and Partial Autocorrelations for Residuals of ARIMA

The plots of ACF, PACF, and residual's normal probability are given in Figure 2. Besides, three tests used to test for residual randomness based on ACF and PACF (Table 3) showed that the residuals are white noise. The estimated ACF (PACF) between the residuals at different lags, and also 95% probability limits around zero, are shown in Figure 2. The lag j ACF (PACF) coefficient measures the correlation between the residuals at the time i and i-j. The probability limits show that, at the 95% confidence level, if the probability limits at lag j do not contain the j estimated ACF

Table 1. Tests to adequate the best model for data.[a]

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	1.76222	OK	OK	OK	OK	***
(B)	1.76593	OK	OK	OK	OK	***
(C)	3.70753	**	***	***	***	**
(D)	1.74945	*	***	***	OK	***
(E)	1.74397	*	***	***	OK	***
(F)	2.01326	OK	***	***	OK	***
(G)	3.15775	*	***	***	***	***
(H)	1.76048	*	OK	OK	OK	***
(I)	1.67147	OK	OK	OK	OK	***
(J)	1.78783	*	OK	OK	OK	***
(K)	1.67398	OK	OK	OK	OK	***
(L)	1.88294	**	OK	OK	OK	***
(M)	**1.45894**	**OK**	**OK**	**OK**	**OK**	*******
(N)	1.56819	OK	OK	OK	OK	***
(O)	1.54853	OK	OK	OK	OK	***
(P)	1.61492	OK	OK	OK	OK	***
(Q)	1.53634	OK	OK	OK	OK	***

[a] RMSE= Root Mean Squared Error; RUNS= Test for excessive runs up and down; RUNM= Test for excessive runs above and below median; AUTO= Ljung-Box 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; OK= Not significant (P>= 0.05). * Marginally significant (0.01< P<= 0.05); ** Significant (0.001< P<= 0.01), *** Highly significant (P<= 0.001).

Table 2. ARIMA Model Summary.

Parameter	Estimate	Standard Error	t	P-value
AR(1)	0.507459	0.122073	4.15702	0.000116
MA(1)	0.986383	0.0185478	53.1805	0.000000
Constant	0.0971988	0.024779	7.96408	0.000000

(PACF) coefficient, there is a significant correlation at lag j. None of the ACF (PACF) coefficients in Figure 2 are significant; indicating that the wheat production data are completely random (white noise) (Statgraphics output).

Test Data and Forecasting of the ARIMA Model

The test data set is wheat production data from 2015-16 to 2018-19. To test the fitted model (ARIMA (1,1,1)), the predicted values and the actual values are shown in Table 5. During this period, the actual data, and data forecasted by ARIMA (1,1,1), along with the residuals, are shown in Table 5.

The forecasted data of wheat production based on the ARIMA (1,1,1) model are shown in Table 6 for the next twelve years (2019-2020 to 2030-2031). For these periods, it shows 95% prediction intervals for the forecasted data. Assuming the ARIMA (1,1,1) model to be appropriate for the wheat production data, these prediction intervals indicate that with 95% confidence, the real data at a selected future time are within this distance. The actual and forecasted data, with 95% limits plot, is shown in Figure 3 (Statgraphics output).

Estimation of the ANN Model

Many factors affect production of wheat. Due to the limitations to get access to data for some factors in the selected period

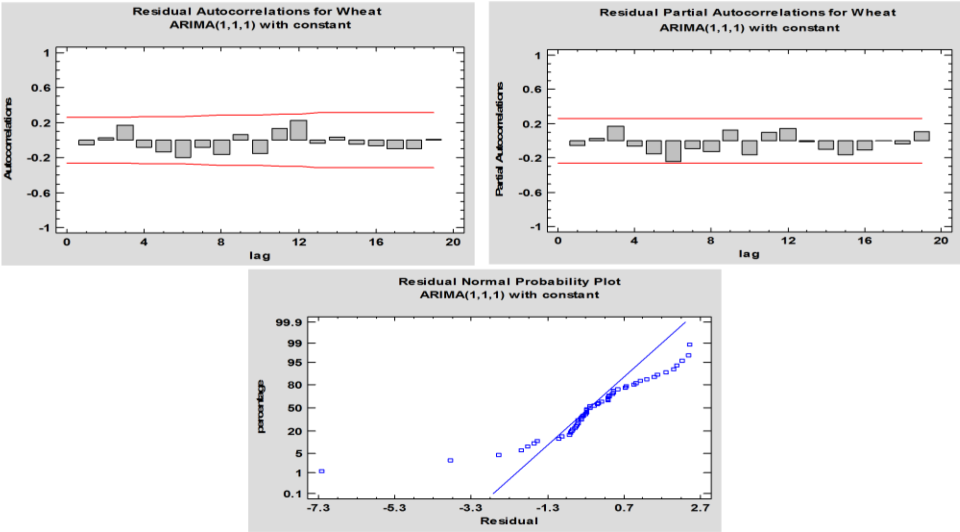

Figure 2. Residual Autocorrelations, Partial Autocorrelations, and normal probability plots for wheat production in Iran.

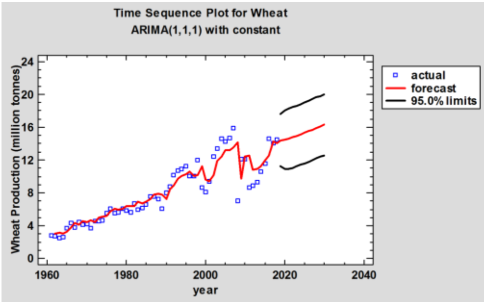

Figure 3. The actual and forecasted data plot of wheat production based on ARIMA (1,1,1) model.

Table 5. The predicted values and the actual values of wheat production (million tons) in 2015-2016 to 2018-2019.

Period	Actual Data	Forecasted Data	Residual
2015-16	11.5223	12.017	-0.494688
2016-17	14.592	12.5863	2.00571
2017-18	14.0	14.2685	-0.268548
2018-19	14.5	14.0617	0.438326

Table 6. The predicted values of wheat production (million tons) based on ARIMA (1,1,1) model for 2019-20 to 2030-31.

Period	Forecast	Lower 95% Limit	Upper 95% Limit
2019-20	14.4186	11.2743	17.5628
2020-21	14.4744	10.9289	18.02
2021-22	14.6	10.9483	18.2517
2022-23	14.7609	11.077	18.4449
2023-24	14.9398	11.2444	18.6351
2024-25	15.1277	11.4275	18.8279
2025-26	15.3203	11.6175	19.0231
2026-27	15.5152	11.8107	19.2198
2027-28	15.7114	12.0054	19.4173
2028-29	15.9081	12.201	19.6152
2029-30	16.1051	12.3969	19.8133
2030-31	16.3023	12.593	20.0115

studied, some assumed important factors were examined viz., total annual precipitation (mm), total applied fertilizer (covering nitrogenous, potash, and phosphate fertilizers) (tons), population, and wheat cultivated area (ha) since 1961-1962 to 2018-2019 in Iran. These factors were found to have significant correlations with the wheat production (FAO, 2020; CCKP, 2020). To run the ANN model, total annual precipitation, total applied fertilizer, population, and wheat cultivated area since 1961-1962 to 2018-2019, as input variables, and wheat production during the same period, as output variables, were applied. ANN model can be successfully tuned to explain the influence of direct and indirect effects on wheat production. The sample size used in this research was 58. Seventy percent of the data (40 samples) that were randomly used for running the model, while for each validation and testing, 15% of the remaining data (9 samples) were selected and used. After many trials by Matlab software, the MLP Neural Network model with five hidden layers was chosen. This model had the minimum value of MSE and, consequently, was selected as the optimum model. The MSE measures of the selected ANN-MLP model was estimated to be 0.298, 0.153, and 0.487 on training, validation, and testing data that were the lowest MSE values among several NN models considered in this research. Furthermore, it is lower than the ARIMA (1,1,1) model. As shown in Figure 4, wheat production estimated by the ANN-MLP for 99 and 99% of the actual variabilities in training and validation data, respectively. The correlation between observed and predicted wheat production is very high, with R^2= 0.98 and r= 0.99 (training). The observed and predicted values of wheat production by the ANN-MLP model are indicated in Figure 5. It is inferred that the predicted value is very close to the actual value. The values of RMSE, MSE, and MAPE of the ANN-MLP model were much lower when compared to the ARIMA (1,1,1) model, as shown in Table 7.

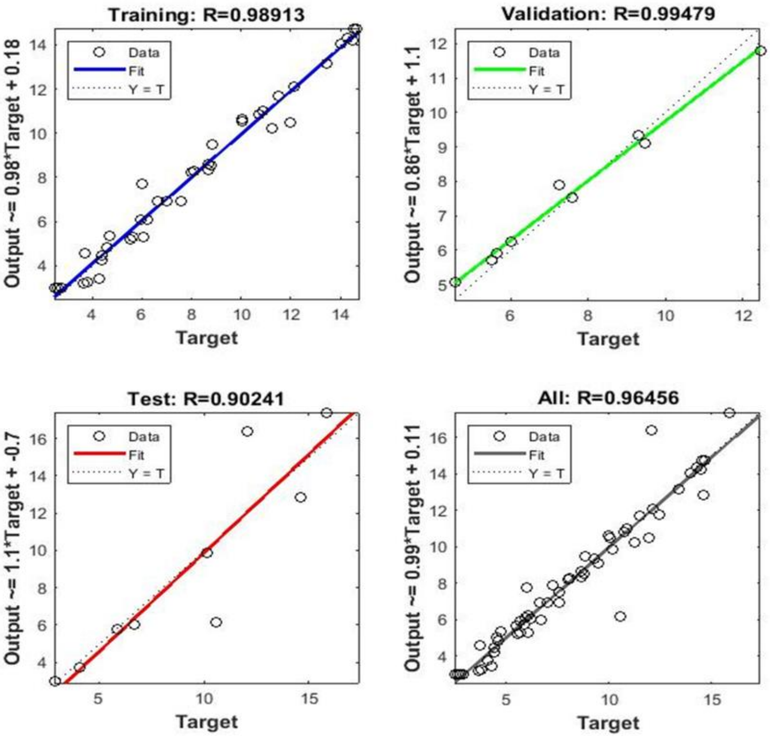

Figure 4. Relationships between observed and predicted wheat production in training, validation, test, and all data using the Artificial Neural Networks model.

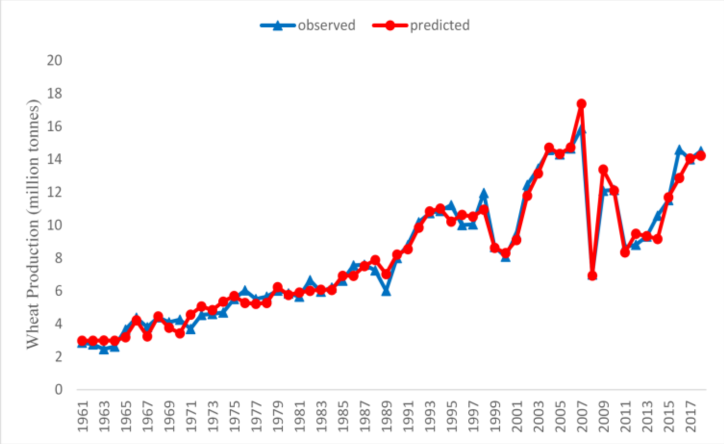

Figure 1. Observed and predicted values of wheat production based on the Artificial Neural Networks model.

Table 3. MSE, RMSE, and MAPE of the ARIMA and ANN models.

	MSE	RMSE	MAPE
ARIMA (1,1,1)	2.12851	1.45894	12.6793
ANN-MLP	0.1525	0.3905	0.4231

DISCUSSION

The production of wheat, as a vital natural resource, Must be planned properly In Iran; wheat is a staple food and its production is very important for the country. In this study, first, different univariate models of the time series were fitted to the wheat production data, and the best model was selected. Based on the minimum AIC, HQC, and SBIC values, the best model was ARIMA (1,1,1). Using this model, wheat production is forecasted for the next twelve years, which predictions showed that in the absence of any shocks and changes in major agricultural prices, agricultural policies, and consumer food patterns, wheat production continues to increase from 2019-20 to 2030-31. Also, to get better results in this study, we have used ANN as a prediction tool and chose four significant factors (total annual precipitation, total applied fertilizer, population, and wheat cultivated area) in wheat production. The MLP Neural Network model with five hidden layers with the lowest MSE was selected. Comparison between the ANN model and the ARIMA (1,1,1) model showed that RMSE=0.391, MSE=0.153, and MAPE=0.4231 of the ANN model were much lower than that of the ARIMA(1,1,1) model (Table 7). As shown in Table 7, using some variables such as precipitation, applied fertilizer, population, and wheat cultivated area would improve the ability of decision-makers to look at the problem from various perspectives and develop solutions. So, these findings are in line with other studies carried out by Safa et al. (2015), Niedbala and Kozlowski (2019), and Patryk et al. (2021). It is suggested that with accurate data, the other parameters such as climate data, soil parameters, and so on, should be studied to obtain more accurate results for wheat production. Also, comparing the results of this research with the study of Hashemi Nejad et al. (2020), the government and the Ministry of Agriculture, taking into account the increase in population in the future, must make the right planning and policies to ensure food security of the people. Besides, by using the exact data, the governments can handle the storage, transport, and distribution of wheat. In general, accurate predicting plays an important role in reducing food instability and price determination.

CONCLUSIONS

The results of this study showed the power of ANN models to predict wheat production indicating efficient result than using the ARIMA model. Thus, this ANN-MLP model for the prediction of wheat production in the future can be a good solution to ensure the food security of people in the country. Due to the high importance of the harvested area in wheat production, there is a risk that farmers will not significantly devote their arable land to wheat production. It is suggested that, according to experts, a limit should be set not only for the wheat cultivated area but also for all factors affecting it. Through proper promotion among the villagers, steps should be taken to spread the proper and efficient use of the facilities. Also, it is suggested that the government increase its declared rate every year in proportion to production costs to be an incentive for more farmers to produce. This study and its results can be useful for the government, the Ministry of Agriculture, and all researchers who are eager to study more in this field, and also designing their management methods and economic precautions. It is suggested that, according to experts, a limit should be set not only for the wheat cultivated area but also for other factors affecting it such as total applied fertilizer, and population. Through proper promotion among the villagers, steps should be taken to spread the proper and efficient use of the facilities. Due to necessity of using annual data in this study and the need to investigate many observations, disability to access appropriate data of other effective variables on an annual at the large scale level were one of the limitations of the study. This inhibited the researchers to examine the relationship between production and these variables. On the other hand, because the study was carried out on a large scale and across the whole country, the local climatic variations in the country could not been taken into account. Therefore, regional climatic variations could be studied in the future.

REFERENCES

1. Akaike, H. 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.*, **19(6):** 716–723.
2. Amin, M., Amanullah, M. and Akbar, A. 2014. Time Series Modeling for Forecasting Wheat Production of Pakistan. *J. Anim. Plant Sci.*, **24(5):** 1444-1451.
3. Asif Masood, M., Irum R. and Saleem A. 2018. Forecasting Wheat Production Using Time Series Models in Pakistan. *Asian J. Agric. Rural Dev.*, **8(2):** 172-177.
4. Azadeh, A., Ghaderi, S. F., Anvari, M. and Saberi, M. 2006. Measuring Performance Electric Power Generations Using Artificial Neural Networks and Fuzzy Clustering. *Proceedings of the 32nd Annual Conference of the IEEE Industrial Electronics Society IECON'06*, Conservatoire National des Arts & Metiers, Paris, France.
5. Balogun, O. S., Awoeyo, O. O., Akinrefon, A. A. and Yami, A. M. 2015. On the Model Selection of Road Accident Data in Nigeria: A Time Series Approach. *Am. J. Res. Commun.*, **3(5):** 139-177.
6. Box, G. and Jenkins, G. 1970. *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco. http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015.pdf.
7. Box, G., Jenkins, G. and Reinsel, G. C. 2008. Multivariate Time Series Analysis. *In Time Series Analysis.* https://Doi.org/10.1002/9781118619193.ch14.
8. Burnham, K. P. and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer, New York. http://ecologia.ib.usp.br/bie5782/lib/exe/fetch.php?media=bie5782:pdfs:burnham_anderson2002.pdf.

9. Cheng, B. and Titterington, D. M. 1994. Neural Networks: A Review from a Statistical Perspective. *Stat. Sci.*, **9(1):** 2–30.
10. Deepika, M. G., Gautam, N. and Rajkumar, M. 2012. Forecasting Price and Analyzing Factors Influencing the Price of Gold Using ARIMA Model and Multiple Regression Analysis. *IJRMEC*, **2(11):** 548-563.
11. FAO. 2020. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data/QC
12. Farajzadeh, Z., and Shahvali, O. 2009. Forecasting Agricultural Crops Prices: Case Study of Cotton, Rice and Saffron. *Agricultural Economics and Development*, **17(67):** 43-71. (in Persian)
13. Forecasting Statgraphics 18. 2017. *Statgraphics Technologies*. Inc. https://cdn2.hubspot.net/hubfs/402067/statdisplay.pdf.
14. Goodwin, P., Onkal. D. and Thomas, M. 2010. Do Forecasts Expressed as Prediction Intervals Improve Production Planning Decisions? *Eur. J. Oper. Res.*, **205**: 195–201.
15. Gujarati, N. D. 2004. *Basic Econometrics.* 4th Edition, Tata McGraw-Hill, New Delhi, pp. 748-807.
16. Hanke, J. E. and Wichern, D. W. 2008. *Business Forecasting.* Pearson Education, New Delhi. https://www.pearson.com/us/higher-education/program/Hanke-Business-Forecasting-9th-Edition/PGM58861.html.
17. Hashemi Nejad, A., abdeshahi, A., Ghanian, M. and Khosravipour, B. 2020. Analyzing Factors Affecting Wheat Production Risk in Iran. *Journal of Agricultural Economics and Development*, **33(4):** 329-338. (in Persian).
18. Karim, R. and Akhter, M. 2010. Forecasting of Wheat Production in Bangladesh. *Bangladesh J. Agric. Res.*, **35(1):** 17-28.
19. Latifi, Z., and Shabanali Fami, H. 2020. Investigating the Technical Efficiency of Date Palm Cultivation Using the Data Envelopment Analysis. *IJAMAD*, **10(4):** 383-399.
20. Ministry of Agriculture Jihad. 2020. Agricultural Statistics-Crops. https://www.maj.ir/
21. Nath, B., Dhakre, D. S. and Bhattacharya, D. 2019. Forecasting Wheat Production in India: an ARIMA Modelling Approach. *J. Pharmacogn. Phytochem.*, **8(1):** 2158–2165.
22. Niazian, M., Sadat-Noori, S. A. and Adibpour, M. 2018. Modeling the Seed Yield of Ajowan (Trachyspermum ammi L.) Using Artificial Neural Network and Multiple Linear Regression Models. *Ind. Crops Prod.*, **117:** 224-234.
23. Niedbala, G. and Kozlowski, J. R. 2019. Application of Artificial Neural Networks for Multi-Criteria Yield Prediction of Winter Wheat. *J. Agr. Sci. Tech.*, **21(1):** 51-61.
24. Patryk, H., Piekutowska, M. and Niedbala, G. 2021. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land*, 10(6): 609.
25. Paul, R. K. 2015. ARIMAX-GARCH-WAVELET Model for Forecasting Volatile Data. *Model Assist. Stat. Appl.*, **10(3):** 243-252.
26. Paul, R. K. and Sinha, K. 2016. Forecasting Crop Yield: A Comparative Assessment of ARIMAX and NARX Model. *RASHI*, **1(1):** 77-85.
27. Safa, M., Samarasinghe, S. and Nejat, M. 2015. Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New Zealand. *J. Agr. Sci. Tech.*, **17(4):** 791-803.
28. Salami, H. and Mohtashami, T. 2014. The Projection Model of Iran's Crop Production in 2025. *Iranian Journal of Agricultural Economics and Development Research*, **45(4):** 585-599. (In Persian).
29. Shahriar, N. and Ghashghaei, Sh. (2018). Forecasting Global Wheat Prices and Influences on Currency Reserve in Iran. *J. Financ. Econ. (Financial Economics and Development)*, **11(41):** 225-241.
30. Shewry, P. R. and Hey, S. J. (2015). The Contribution of Wheat to Human Diet and Health. *Food Energy Secur.*, **4(3):** 178–202.
31. Tsay, R. 2005. Multivariate Time Series Analysis and Its Applications. Doi:10.1002/0471746193.ch8.
32. World Bank Group. 2020. *Climate Change Knowledge Portal (CCKP).* https://climateknowledgeportal.worldbank.org.
33. Zinyengerea, N., Mhizha, T., Mashonjowa, E., Chipindu, B., Geerts, S. and Raes, D.

2011. Using Seasonal Climate Forecasts to Improve Maize Production Decision Support in Zimbabwe. *Agric. For. Meteorol.*, **151:** 1792–1799.

پیش بینی تولید گندم در ایران با استفاده از تکنیک سری زمانی و شبکه عصبی مصنوعی

ز. لطیفی، و ح. شعبانعلی فمی

چکیده

با افزایش جمعیت جهان، نگرانی‌ها برای تأمین غذا نیز افزایش می‌یابد. گندم، به عنوان یکی از مهمترین محصولات کشاورزی که به طور گسترده در سراسر جهان مصرف می‌شود، نقش بسیار مهمی در تغذیه مردم دارد، به ویژه در رژیم غذایی ایرانیان که به نان بستگی زیادی دارد. پیش بینی محصول برای هر کشوری امری حیاتی است، لذا بر این اساس، تصمیم گیری در مورد ذخیره سازی، واردات یا صادرات و غیره می‌تواند برنامه ریزی شود. در این مقاله، مدل‌های مختلف سری زمانی تک متغیره و مدل شبکه عصبی مصنوعی برای پیش بینی تولید گندم در ایران استفاده شده است. داده‌های تولید سالیانه گندم، مجموع بارندگی سالیانه، مجموع کود مصرفی، جمعیت و زمین زیرکشت گندم از سال 62-1961 تا 19-2018 مورد استفاده قرار گرفته است. با مینیمم مقادیر 1/45894، 1/00329، 1/0448 و 1/09742 به ترتیب برای معیارهای RMSE، AIC، HQC و SBIC، مدل ARIMA (1,1,1)به عنوان بهترین مدل تک متغیره انتخاب گردید. در پیاده‌سازی مدل‌های شبکه عصبی مصنوعی (ANN)، مجموع بارندگی سالیانه، مجموع کود مصرفی، جمعیت و زمین زیرکشت گندم به عنوان متغیرهای ورودی و تولید گندم به عنوان متغیر خروجی مورد استفاده قرار گرفت. مدل شبکه عصبی پرسپترون چندلایه (MLP-NN) با پنج لایه پنهان که دارای کمترین مقدار MSE=0.153 در میان چندین مدل شبکه عصبی در این مطالعه بود انتخاب گردید. مقایسه میان مدل ANN و مدل ARIMA (1,1,1) نشان داد که در مدل ANN، مقادیر RMSE=0.391، MSE=0.153 و MAPE=0.4231 بسیار کمتر از مدل ARIMA (1,1,1) می‌باشد. نتایج نشان دهنده قدرت مدل‌های ANN در پیش بینی تولید گندم با استفاده از پارامترهای کارامد در مقایسه با مدل ARIMA می‌باشد.