



Forecasting Area, Production and Productivity of Vegetable Crops in Nepal using the Box-Jenkins ARIMA Model

Rabin Thapa^{1,a,*}, Shivahari Devkota^{1,b}, Sandip Subedi^{2,c}, Babak Jamshidi^{3,d}

¹Department of Agricultural Economics and Agribusiness Management, Faculty of Agriculture, Agriculture and Forestry University, Chitwan, Nepal

²National Center for Potato, Vegetable and Spice Crops Development, Ministry of Agriculture and Livestock Development, Nepal

³Social Development and Health Promotion Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

*Corresponding author

ARTICLE INFO

ABSTRACT

Research Article

Received : 10/08/2021
Accepted : 29/10/2021

Keywords:

Autoregressive Model
Box-Ljung test
Jarque-Bera test
Policy
Time-series analysis

Forecasting of vegetable area, production, and productivity of Nepal was made from the historical data of 1977/78 to 2019/20 by using the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) models. The best fitted ARIMA models were chosen based on the minimum value of the selection criterion, Akaike information criteria (AIC) and Bayesian information criteria (BIC). ARIMA (0, 2, 1) model was found suitable for all areas and production, whereas ARIMA (0, 2, 0) model was best fitted for forecasting vegetable productivity. The model was cross-validated by comparing the point prediction with the actual test series data from 2015/16 to 2019/20. The performances of models were determined from the mean absolute percent error (MAPE) value. The MAPE was found to be 2.70%, 2.40%, and 3.80%, respectively for the prediction of area, production, and productivity. The forecast was made for the immediate five years (2020/21 to 2024/25), and it showed an increasing value for area and production while the forecasts of productivity had lower values. The vegetable production policy in Nepal should be planned following accurate forecasts to increase production in the upcoming years. Awareness among the vegetable growers should be raised in the following years with appropriate extension programs.

^a rabinth007@gmail.com

^b <https://orcid.org/0000-0003-3625-2587>

^c devkota.shivahari@gmail.com

^d <https://orcid.org/0000-0001-7587-0623>

^c agsandipsbedi@gmail.com

^d <https://orcid.org/0000-0003-0228-2446>

^d babak.j6668@gmail.com

^d <https://orcid.org/0000-0002-5995-367X>



This work is licensed under Creative Commons Attribution 4.0 International License

Introduction

Agriculture is an important sector of Nepal as it contributes 27.08% to the national GDP, and almost 60.4% of the total population in the country is involved in agriculture (AITC, 2021). Nepal is a developing country with a significant contribution to agriculture (GC and Ghimire, 2018). In Nepal, agriculture is practiced as one of the main economic activities, and especially farmers are involved in vegetable farming, making it one of the major agricultural practices (Rai et al., 2019). Vegetable production bears a significant contribution to the economy, thus determining the economic status of the farmers (Ghimire et al., 2018). Furthermore, along with a huge nutritious value, vegetables provide almost 5 to 10 times benefit making them a high-value crop (Gurung et al., 2016). In the year 2019/20, the total cultivated area of the country was about 3.09 million hectares, among which the vegetable crops occupied 9.09 % of total cultivated agricultural land (AITC, 2021), and the annual vegetable production was found to be 3,962,383 metric tons with

average productivity of 14.09 Mt/ha. The average productivity of vegetables from 2008/09 to 2018/19 was 13.42 Mt/ha (MoALD, 2019).

The top exported agricultural commodities of Nepal are lentil (29.6%), cardamom (7%), wheat (6.7%) and tea (4%) followed by vegetables, making them the fifth most important commodity in terms of export (CBS, 2013). In 2019/20, Nepal imported edible vegetables, roots and tuber crops worth NPR 33,057,091 and exported them worth NPR 1,073,127 with a negative trade balance of NPR - 31,983,964 (MoF, 2019). This shows the considerable potential of import substitution in vegetables. The benefit from vegetable farming has shifted the food demand from cereals towards vegetables and fruits (Ghimire et al., 2018). Similarly, the increasing demand for vegetables could be attributed to the increasing population (Alhassan and Kilishi, 2019) as well as due to the global COVID-19 pandemics.

Considering the production potential and increasing demand for vegetables, commercial vegetable farming has been promoted by various programs. Agriculture Perspective Plan (APP, 1995-2014) targeted the growth rate of GDP (Horticultural sector) to 5.5 percent per annum by 2014/2015 and that of vegetables to 5.42 percent per annum (NPC, 1995). Under the outcome Profitable Commercialization, ADS (Agriculture Development Strategy) has focused on developing the value chain of vegetables through production, marketing and policy/institutional support as the main activity (ADS, 2015). The strategy has given precedence to few value chains over the first five years, and the vegetable sector falls among the top five sectors whose value chain has been prioritized.

Although various strategies and plans are made for the betterment of the agriculture sector, a sudden decrease in production lowers the farmers' income, decreases the marketable surplus and ultimately increases in price could be foreseen. Likewise, a boost in agricultural production can lead to a sharp fall in prices and affects the farmers' incomes (Tripathi et al., 2014). An accurate forecast could suggest appropriate surplus and deficit management, stabilize the price and ensure profits for the farmers (Kumar, 2020). Furthermore, forecasting could be of immense importance in reinforcing the policy decisions, ensuring food security, managing import/export and implementing price policy (Badmus and Ariyo, 2011; Sharma et al., 2018). Besides, land use allocation, food safety, choosing high yielding varieties, conducting training for improved cultural practices, adequate supply of inputs, adoption of latest technologies, and security and environmental issues could also be addressed by the appropriate forecast (Mahapatra and Dash, 2020; Tripathi et al., 2014). Thus, it is imperative to understand the trend of vegetable production, productivity and area over time, and forecasting these parameters could be of great essence. However, preliminary study on analysis of the recent trends and poor forecasting of vegetable production in the country has limited the effectual planning and program development (Ghimire et al., 2018).

Various literature mentions different statistical and econometric forecasting models that could be suitable for making a forecast on different issues, like agricultural production (Ali et al., 2015). Several techniques like simulation modelling and remote sensing are essentially being used to forecast the productivity and production area (Tripathi et al., 2014). But sometimes, forecasting is felt necessary before crop harvest or in some cases even before the plantation of crops. When using forecasting models for univariate time-series data, the Autoregressive Integrated Moving Average (ARIMA) forecasting model is mostly recommended and widely used (Badmus and Ariyo, 2011). Its popularity could be attributed to its statistical properties and the well-known Box and Jenkins methodology (Mahapatra and Dash, 2020). Allen (1994) reported that the ARIMA model performs markedly better for univariate time-series analysis than other methods.

Applying the ARIMA model, Iqbal et al. (2005) forecasted the area and production of wheat in Pakistan up to the year 2022; Debnath et al. (2013) predicted the area, production and yield of cotton in India from 2011 to 2020; Kumar and Anand (2014) forecasted the sugarcane

production in India from 2013 to 2017; Ali et al. (2015), using the data from 1948 to 2012, forecasted productions and yields of both crops from 2013 to 2030; Ilic et al. (2016) forecasted corn production in Serbia from 2015 to 2017; Hossain and Abdulla (2017) forecasted the potato production in Bangladesh from 2014 to 2023; Celik et al. (2017) forecasted the production of groundnut in Turkey from 2016 to 2030; Sharma et al. (2018) forecasted maize production in India from 2017 to 2022; Nath et al. (2019) forecasted wheat production in India from 2018 to 2027; Mahapatra and Dash (2020) forecasted the productivity of black gram in India for the year 2016-17 to 2018-19. These studies show the applicability of the ARIMA model for forecasting the production, productivity and area of agricultural products. In this context, we can be certain about using the ARIMA model for forecasting purposes. The model captures the effect of its past data, and considers the current and past information of error term and hence cogitates all sorts of information delimited with the univariate time-series during forecast (Padhan, 2012). Thus, to analyze and study the trend of vegetable production, Man-Kendall's test and Sen's slope estimation was used and to forecast the area, production and productivity of vegetables from 2021 to 2025, ARIMA Model of order (0,2,1) for area and production and (0,2,0) for productivity was used, along with proper cross-validation.

Materials and Methods

Data Collection

Our study was aimed at carrying out an in-depth analysis of the trend and forecast the future value points of the vegetable crops area, production and productivity in Nepal. For this purpose, the necessary time-series data were obtained from secondary sources; the 'Statistical Yearbook' published by the Ministry of Agriculture and Livestock Development (MoALD), Nepal, Krishi Diary published by AITC (2021) and other reports published by the Department of Agriculture, Nepal. The time-series data included the area, production and productivity of vegetable crops from the year 1977/78 to the year 2019/20, making a total of 43-year data points, which were further subjected to data analysis. In total, 85% of the total time-series were used as a training dataset for model estimation, and the remaining 15% of the data were used as a test dataset for cross-validation of the proposed model. Thus, the dataset from the year 1977/78 to the year 2014/15 was used for the prediction of the ARIMA model and the dataset from year 2015/16 to the year 2019/20 was used for cross-validation of the model. The best fit model for the time-series data was analyzed using Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model using the R Studio software. The selected best fit model was used to forecast the area, production and productivity of vegetable crops in Nepal from 2020/21 to 2024/2025.

Trend Analysis

The non-parametric Mann-Kendall (MK) test is most commonly utilized for a monotonic time-series trend. The Mann-Kendall test, and Sen's slope estimator, were used to analyze the trend in our time-series data. The Mann-Kendall test was used to analyze the increasing or

decreasing trend in time-series. In contrast, Sen's slope estimator was used to estimate, on average, how much area, production and productivity had changed in each year. The descriptive properties of time-series were also analyzed. The Mann-Kendall test statistic is given as:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$

Where, n is no. of data points, x_j and x_i are the data value in time-series i and j ($j>i$) respectively and $\text{sgn}(x_j - x_i)$ is the function as:

$$\text{sgn}(x_j - x_i) = \begin{cases} +1, & \text{if } x_j - x_i > 0 \\ 0, & \text{if } x_j - x_i = 0 \\ -1, & \text{if } x_j - x_i < 0 \end{cases}$$

Similarly, Sen's slope was estimated as:

$$Q = \frac{x_j - x_i}{j - i}, \text{ where } i < j$$

Analytical Model

In our study, we used area, production and productivity of vegetable crops in Nepal as our target variables. An autoregressive integrated moving average model, also known as the ARIMA model, is used to understand better or forecast future values of time-series. The ARIMA model given by Box-Jenkins (1970) is a form of regression analysis that gauges the strength of dependent variable relative to its independent variable and has three different processes; autoregressive (AR) of order p, differencing of degree d to make the time-series stationary and moving average (MA) of order q, and written as ARIMA (p, d, q). The Box-Jenkins methodology of the ARIMA model uses the non-stationary time-series to forecast future values. The given non-stationary time-series are converted to stationary by differentiating process in the ARIMA model which is usually known as 'd' degree of differentiating. Generally, a stationary stochastic process can be used for forecasting future point values in time-series (Verbeek, 2004). A stochastic process to be stationary should satisfy the following criteria:

$$E(Y_t) = \mu; E[(Y_t - \mu)^2] = \sigma^2; Cov(Y_t, Y_{t+k}) = \gamma(k)$$

The above equation implies that the mean, variances and autocovariances are unaffected by a change in time origin.

The ARIMA model of order (p, d, q) can be expressed in econometric model form as shown below.

$$Y_t = \mu + \sum_{i=1}^p \theta_i Y_{t-i} + \sum_{j=1}^q \alpha_j \varepsilon_{t-j} + \varepsilon_t$$

Where, Y_t is given variable at time t, μ is the mean of series, the $\theta_1, \dots, \theta_p$ are the parameters of the AR model, the $\alpha_1, \alpha_2, \dots, \alpha_q$ are the parameters of the MA model and the $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ are white noise residuals. Similarly, the mathematical form of the ARIMA model of order (p, d, q) can be expressed as:

$$\Phi_p(L)(1-L)^d Y_t = c + \theta_q(L)\varepsilon_t$$

Where, $\Phi_p(L)$ and $\theta_q(L)$ are polynomials in L of degrees p and q respectively $c = \text{constant}$, L = a Lag operator (some uses as Backshift operator) (Verbeek, 2004), d = order of difference operator, p = order of non-seasonal AR operator; and q = order of non-seasonal MA operator.

The stationarity conditions and invertibility of the time-series were met only if all the roots of the characteristic equations $\Phi_p(L) = 0, \theta_q(L) = 0$ lied outside the unit circle (Box et al., 2008).

Model Prediction and Validation

Box-Jenkins's methodology was used for the model prediction and model validation of the production, area and productivity of vegetable crops in Nepal. The methodology comprised of the following process:

Model Identification

The time-series data stationarity check was done using the Augmented Dickey-Fuller (ADF) statistical test. When the obtained value of the ADF test statistic was greater than the critical value for the error level of 5%, we rejected the null hypothesis i.e., there is in the time-series data. The differencing of 'd' order was done until the time-series became stationary, as confirmed by the ADF test. After that, we plotted the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to observe where the spikes showed significance. Finally, using the auto-ARIMA function of the R Studio forecasting package, the appropriate order (p, d, q) of the ARIMA model was determined for each of the time-series data of the area, production and productivity.

Model Estimation

After selecting the appropriate model, the estimation of model parameters was done using the Maximum Likelihood (ML) method. ML is one of the best and well-known methods for estimating the parameters of an ARIMA model (Box et al., 2008). A computer-aided statistical packages was used to facilitate the process of estimation.

Diagnostic checking

The best model of the ARIMA (p, d, q) was obtained using several diagnostic tests. The various tests are given as:

- The best fit for the ARIMA model was estimated according to two performance criteria; Akaike information criterion (AIC) and Bayesian information criterion (BIC). The model with the lowest value of the above criterion was chosen as the best model. Among AIC and BIC; the AIC is the best and important criteria to select the best model (Verbeek, 2004).
- Along with AIC and BIC criteria, the goodness of fit was estimated by plotting the autocorrelation function (ACF) of residuals of the fitted model. For the best-fitted model, the residuals should show white noise (i.e., zero mean, constant variance, and zero correlation). If not so, that indicated evidence of autocorrelation of residuals (ε_t). The autocorrelation of residuals was checked by the Box-Ljung Q test statistical tool, which tests for the autocorrelation of residuals. The Box-Ljung test statistic is given as:

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k}$$

Where n is the number of samples, r_k is the estimated autocorrelation of series at lag $k=1, 2, \dots, m$ and 'm' is the number of lags considered. The insignificant p-value at a given level of significance validated the absence of correlation of residuals. Hence, in this way, it can be considered as our best-fitted model (Gujrati and Porter, 2009).

Model Adequacy

The aforementioned process is the core of the Box-Jenkins methodology. Once the best-fitted model of ARIMA was selected, we checked for the adequacy of the model. It was checked on the test dataset from 2015/16 to 2019/2020 to estimate the model's goodness of fit. The selected ARIMA model forecasted the predicted value for the same time frame. Mean absolute percentage error (MAPE) was used to estimate model adequacy.

The formula to calculate the MAPE is as follows:

$$\text{MAPE} = \left(\frac{1}{n} \right) \times \sum (|\text{actual-forecast}|/|\text{actual}|) \times 100$$

Where, n = Sample size, actual = the actual data value and forecast = the forecasted data value.

Model Forecasting

The model was used to forecast the production, area and productivity of vegetable crops in Nepal for the next five years i.e., from 2020/2021 to 2024/2025 production period.

The forecasting was done by using the software R Studio Version 4.1 and STATA 16.

Results and discussion

Trend Analysis

Descriptive statistics for the time-series data of the vegetable area, production, and productivity of Nepal are given in Table 1. The time-series data is plotted in Figure 1. The time-series data for the vegetable area, production, and productivity are non-normal, which can be assessed from the skewness and kurtosis values. Hence, a non-parametric Mann-Kendall test for trend analysis was performed to test the significance of the trend. As evident from the values of Mann-Kendall's Z statistics and Sen's slope estimate (Q), the time-series data for all the parameters selected for analysis showed a significant and positive trend.

The three-time series data of the area, production and productivity have approximately equal Mann-Kendall Z scores, therefore approximately equal P-values indicating that the trends in the time-series are highly significant.

While the Sen's slope statistics from the table, estimating how much the intended variable averagely changes with one unit increase in time, shows that the rate of increase in area, production and productivity of vegetables is 3875 ha/year, 70953 mt per year and 0.266 mt/ha per year which indicates an unexploited growth potential of vegetable crops in Nepal. To exploit this potential, the Government of Nepal has started a 10-year long Project, PMAMP (Prime Minister Agricultural Modernization Project), by building superzones, zones, blocks and pockets for different vegetable crops.

Building ARIMA Models

The ARIMA model fits the time-series for which either some n^{th} differences are stationary, or some Box-Cox transformation makes them stationary. Here, the stationarity of the time-series data was checked through the first and second differences of the original data. Regarding our variables, the second differencing led to stationary time-series, and there is no need to transform data using the Box-Cox formula (Figure 1). The first difference of the curves lacks the stability of variance. The first differential plots illustrate those fluctuations (variance) are proportional to the local averages, while the second differential seems to meet the requirements graphically.

To confirm the visual results, we performed the Augmented Dickey Fuller (ADF) test. The test statistic showed that the original time-series variables were non-stationary since the test statistic was non-significant. The stationarity of the three variables' time-series was confirmed at second degree differencing by ADF test which showed highly significant test statistic for respective variables (Table 2). Once the series was stabilized, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots helped us find the degree of the desired model (Figure 2) i.e., the p and q order of the model. The slow decaying of the ACF values demonstrates that the original time-series are not stationary, and some modifications are required. The ACF and PACF of the differenced time-series of area and production, respectively, due to fast and gradual declining, show that the MA model fits their time-series. For the productivity rate, ARMA seems to be the best-fitted model.

After accomplishing appropriate models, model parameters (coefficients) of the variables were obtained through the maximum likelihood method (MLE), and later the optimal model was selected based on performance metrics -- Akaike information criteria (AIC) and Bayesian information criterion (BIC). The use of 'auto.arima' in RStudio further confirmed the best-fitted models.

From Table 2, the following models for the area, production and productivity of vegetables can be devised, respectively.

Table 1. Descriptive statistics and Mann-Kendall trend analysis test for the time-series data of vegetable cultivation in Nepal from 1977/78 to 2014/15

Parameter	Mean	Median	SD	Skewness	Kurtosis	MKT	Sen's slope (Q)
Area (ha)	162317.2	142446	47521.64	0.737	2.6482	8.33 (0.0000)	3875.92
Production (mt)	1595847	1334933	945746.4	0.755	2.3510	8.75 (0.0000)	70953
Productivity (mt/ha)	9.07792	9.229715	2.92	0.0359	1.6068	8.373 (0.0000)	0.266

SD: Standard deviation, MKT: Mann-Kendall trend Z (P-value)

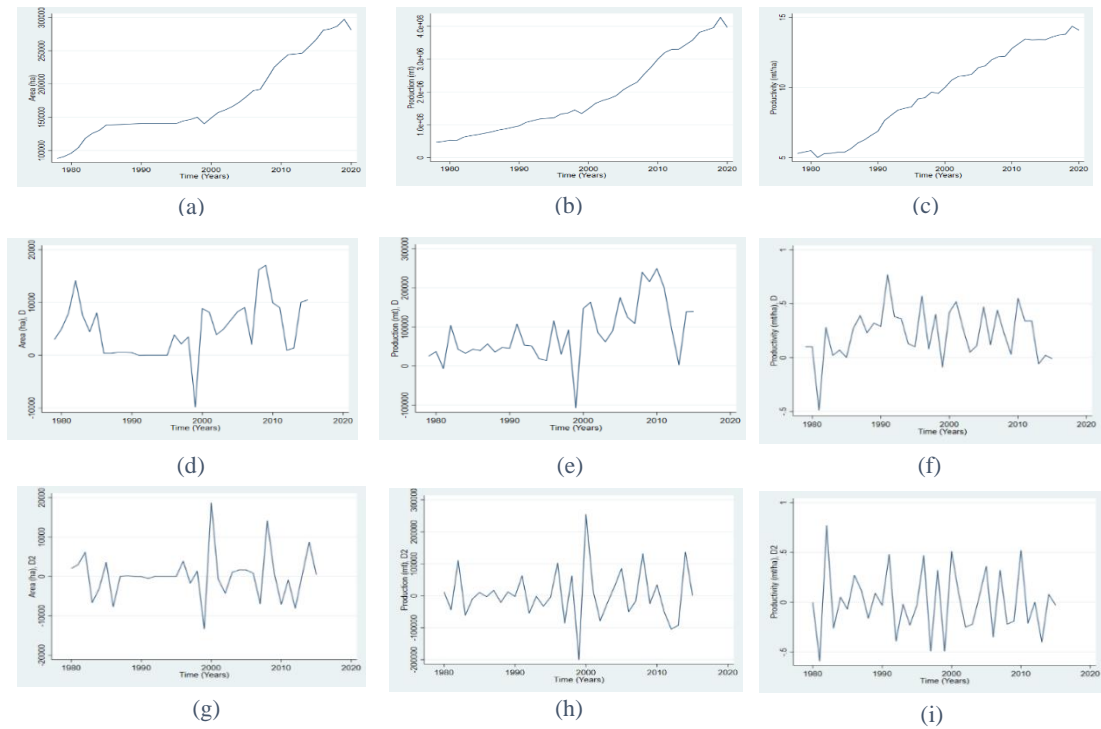


Figure 1. (a), (b) and (c): Trend of vegetable cultivation area, production and productivity of Nepal (1977/78-2019/20). (d), (e) and (f): First order differencing of train data (1977/78 to 2014/15). (g), (h) and (i): Second order differential of train data (1977/78 to 2014/15)

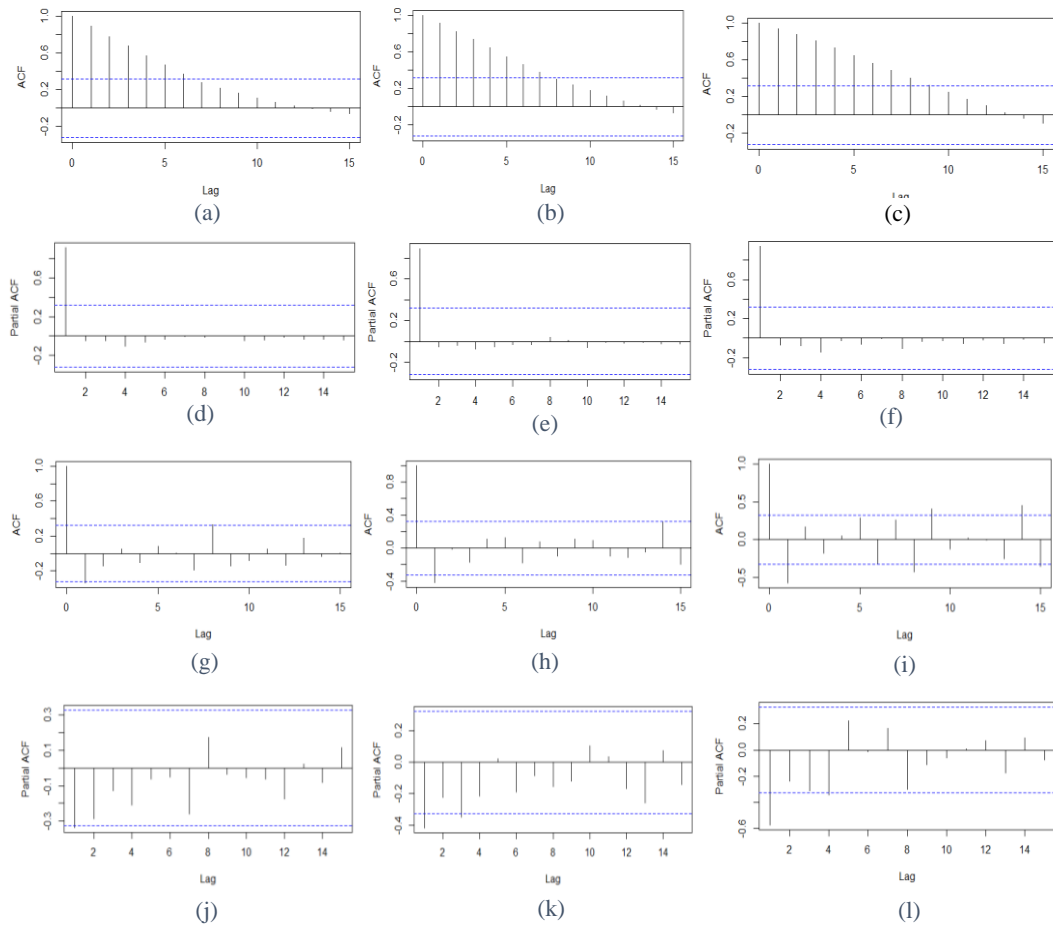


Figure 2. (a), (b) and (c): Autocorrelation Function (ACF) plot of the original time-series (d), (e) and (f): Partial Auto-correlation Function (PACF) plot of the original time-series data. (g), (h) and (i): ACF of the second differentiated data. (j), (k) and (l): PACF of the second differentiated data (From left to right: area, production and productivity)

Table 2. Fitted ARIMA models, and their selection criteria values (AIC and BIC) and estimates of parameters for time-series data on vegetables area, production, and productivity for Nepal (1977-78 to 2014-15)

Parameter	AR	AIC	BIC	Estimate	SE	Z	ADF
Area	(0,2,1)	720.22	723.39	-0.6671 (MA1)	0.144	2.7461 (<0.01)	0.001 (2 nd order)
Production	(0,2,1)	906.02	909.18	-0.7362 (MA1)	0.124	-5.9367 (<0.001)	0.001 (2 nd order)
Productivity	(0,2,0)	0.45	3.68	-	0.038	-	0.001 (2 nd order)

AR: ARIMA model, Z: SE: Standard error, Z-value (Pr(>|z|)), ADF: ADF test (Order of differential)

Table 3. Test for autocorrelation in the ACF and PACF of the residuals and test for heteroscedasticity of the model residuals for the three-time series models

Parameter	AR	Box-Ljung test χ^2 (p-value)	Jarque-Bera test χ^2 (p-value)	Significance (at 5%)
Area	(0,2,1)	14.96 (0.7785)	3.7816 (0.151)	Both non-significant
Production	(0,2,1)	18.936 (0.526)	5.2487 (0.07249)	Both non-significant
Productivity	(0,2,0)	28.236 (0.1039)	2.498 (0.2868)	Both non-significant

AR: ARIMA model

Table 4. Validation (2015-16 to 2019-2020) of ARIMA models for vegetable area, production, and productivity of Nepal

Parameter	Year	Actual value	Point prediction	Relative error	MAPE in prediction
Area (Million hectares)	2015-16	0.2808067	0.275015	-2.10%	2.70%
	2016-17	0.282809	0.283094	0.40%	
	2017-18	0.286864	0.291173	1.50%	
	2018-19	0.297195	0.299251	0.70%	
	2019-20	0.281132	0.307330	9.10%	
Production (Million tons)	2015-16	3.819809	3.709971	-0.20%	2.40%
	2016-17	3.884797	3.839857	-0.90%	
	2017-18	3.958230	3.969744	0.20%	
	2018-19	4.271270	4.099630	-4.20%	
	2019-20	3.962383	4.229517	6.50%	
Productivity (mt/ha)	2015-16	13.603	13.4063	-1.50%	3.80%
	2016-17	13.7364	13.4009	-2.50%	
	2017-18	13.7983	13.3954	-3.00%	
	2018-19	14.372	13.39	-7.00%	
	2019-20	14.0944	13.3846	-5.00%	

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - 0.6671 \varepsilon_{t-1},$$

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t - 0.7362 \varepsilon_{t-1},$$

and

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t$$

The rightmost column in Table 2 establishes that the obtained coefficient of the terms ε_{t-1} are 95% significant. The table also shows the performance of these models during crop years 2015/16 -2019/20 and the forecasts that these models lead to, for the next 5 years.

It is necessary to check out the presumptions of the ARIMA model before using it for prediction: The residuals should show white noise (i.e., zero mean, constant variance, and zero correlation). The autocorrelation of residuals was checked using the Box-Ljung Q test. The insignificant p-value resulted from Box-Ljung test statistics at 0.05 level of significance (Table 3) validated the absence of correlation of residuals of respective models. Also, the Jarque-Bera (JB) test of normality (Table 3) of residuals showed a non-significant p-value at 0.05 level of significance, indicating the normality of residuals. The absence of the autocorrelation of residuals and the normality of the residuals showed the accomplished models are the best fit for the given time-series. Further prediction and forecasting of the variables can be done.

Forecast Using ARIMA Models

After determining the appropriate model and using appropriate, the selected models are rendered fit for further forecasting. Using the ARIMA model of order (0,2,1) for area and production and order (0,2,0) for productivity, the actual data from test series and predicted values from the model for area, production and productivity from the year 2015/16 to 2019/20 were compared (Table 4). The lower value of Mean Absolute Percentage Error (MAPE) 2.70%, 2.40% and 3.80% for prediction of the area, production and productivity, respectively, implies a very good model accuracy for further prediction. Thus, our cross-validation showed that the model is suitable for forecasting the values of the future.

The forecasted values of the area, production and productivity of vegetables from 2020/21 to 2024/25 are presented in Table 5, along with 95% (low and high) prediction intervals. From the table, it can be devised that both area and production showed increasing forecasted values for the vegetable crops with successive years. The productivity, unlike the other two parameters, showed lower forecasted values. The prediction for the year 2024/25 resulted in approximately 283416 hectares of vegetable area, 4398529 metric tons of vegetables and 12.706 mt/ha productivity.

Table 5. Forecasts (2020-21 to 2024-25) by the ARIMA models for vegetable area, production, and productivity of Nepal

Parameter	Year	Forecast	95% interval prediction (Low-High)
Area (Hectares)	2020-21	281588.9	269397.5 - 293780.2
	2021-22	282045.7	262033.1 - 302058.4
	2022-23	282502.6	254525.8 - 310479.4
	2023-24	282959.5	246596.2 - 319322.8
	2024-25	283416.4	238179.8 - 328652.9
Production (Metric tonnes)	2020-21	4049612	3848652 - 4250573
	2021-22	4136841	3832422 - 4441261
	2022-23	4224071	3826084 - 4622057
	2023-24	4311300	3822336 - 4800263
	2024-25	4398529	3818630 - 4978427
Productivity rate (mt/ha)	2020-21	13.8168	13.1531 - 14.4806
	2021-22	13.5393	12.0551 - 15.0235
	2022-23	13.2617	10.7782 - 15.7453
	2023-24	12.9842	9.3486 - 16.6197
	2024-25	12.7066	7.7841 - 17.6291

Conclusion

In the current study, the ARIMA of order (0,2,1) was the best-fitted model for the area and production forecasting, while (0,2,0) was the best one for productivity forecasting. The ARIMA models projected an increment in the area, production and decreased productivity of vegetables for the duration of 5 years (2020/21 to 2024/25). The prediction for 2024/25 resulted in approximately 283416 hectares of vegetable area, 4398529 metric tons of vegetables and 12.706 mt/ha productivity. Thus, in order to shape the production, the vegetable growers should be given proper information via training, extension activities, policy tools, etc. Furthermore, the vegetable production policy in Nepal should be planned according to the forecasts for an increased production in the upcoming years.

References

- ADS. 2015. Agriculture Development Strategy 2015 to 2035 Part 1. Agriculture Development Strategy 2015 to 2035. <http://nnfsp.gov.np/PortalContent.aspx?Doctype=Resources andID=325>
- AITC. 2021. Krishi Diary. Ministry of Agriculture and Livestock Development.
- Alhassan A, Kilishi AA. 2019. Weak economic institutions in Africa: a destiny or design? *International Journal of Social Economics*, 46(7): 904–919. <https://doi.org/10.1108/IJSE-12-2018-0651>
- Ali S, Badar N, Fatima H. 2015. Forecasting Production and Yield of Sugarcane and Cotton Crops of Pakistan for 2013-2030. *Sarhad Journal of Agriculture*, 31(1): 1–10.
- Allen PG. 1994. Economic forecasting in agriculture: Reply. *International Journal of Forecasting*, 10(4): 601–602. [https://doi.org/10.1016/0169-2070\(94\)90028-0](https://doi.org/10.1016/0169-2070(94)90028-0)
- Badmus MA, Ariyo OS. 2011. Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model. *Asian Journal of Agricultural Sciences*, 3(3): 171–176. <http://maxwellsci.com/print/ajags/v3-171-176.pdf>
- CBS. 2013. National Sample Census of Agriculture 2011/12. http://www.fao.org/fileadmin/templates/ess/ess_test_folder/World_Census_Agriculture/Country_info_2010/Reports/Reports_5/NPL_EN_REP_2011-12.pdf
- Celik S, Karadas K, Eyduran E. 2017. Forecasting the production of groundnut in Turkey using arima model. *Journal of Animal and Plant Sciences*, 27(3): 920–928.
- Debnath MK, Kartic B, Mishra P. 2013. Forecasting of Area, Production and Yield of Cotton in India and Karnataka Using ARIMA Model. *Journal of Space Science and Technology*, 2(1): 16–20. <https://doi.org/10.5958/2322-0430.2017.00236.0>
- GC A, Ghimire K. 2018. A SWOT analysis of Nepalese agricultural policy. *International Journal of Agriculture, Environment and Food Sciences*, 2: 119–123. <https://doi.org/10.31015/jaefs.18020>
- Ghimire D, Lamsal G, Paudel B, Khatri S, Bhusal B. 2018. Analysis of trend in area, production and yield of major vegetables of Nepal. *Trends in Horticulture*, 1(2): 1–11. <https://doi.org/10.24294/th.v1i2.914>
- Gujrati D, Porter DC. 2009. *Basic Econometrics* (Fifth ed.). Irvin: McGraw-Hill.
- Gurung B, Thapa RB, Gautam DM, Karki KB, Regmi PP. 2016. Commercial Vegetable Farming: An Approach for Poverty Reduction in Nepal. *Agronomy Journal of Nepal*, 4: 92–106. <https://doi.org/10.3126/ajn.v4i0.15518>
- Hossain M, Abdulla F. 2017. Forecasting Potato Production in Bangladesh by ARIMA Model. *Journal of Advanced Statistics*, 2(4): 125–136. <https://doi.org/10.22606/jas.2017.24001>
- Ilic I, Jovanovic S, Jankovic-Milic V. 2016. Forecasting corn production in Serbia using ARIMA model. *Ekonomika Poljoprivrede*, 63(4): 1141–1156. <https://doi.org/10.5937/ekopolj1604141i>
- Iqbal N, Bakhsh K, Maqbool A, Abid Shohab A. 2005. Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan. *Journal of Agriculture and Social Science*, 2(1988): 120–122.
- Kumar M, Anand M. 2014. An Application Of Time Series Arima Forecasting Model For Predicting Sugarcane Production In India. *Studies in Business and Economics*, 9(1): 81–94.
- Kumar RR. 2020. Forecasting of Potato Pvegetableness in India: An Application of Arima Model. *Economic Affairs*, 65(4): 473–479. <https://doi.org/10.46852/0424-2513.4.2020.1>
- Mahapatra SK, Dash A. 2020. ARIMA Model for Forecasting of Black Gram Productivity in Odisha. *Asiatic Society for Social Science Research (ASSSR)*, 2(1): 131. <https://doi.org/10.46700/asssr/2020/v2/i1/196124>
- MoALD. 2019. *Statistical Information in Nepalese Agriculture*. In Ministry of Agriculture and Livestock Development. Planning and Development Cooperation Co-ordination Division, Statistics and Analysis Section. <https://nepalindata.com/resource/statistical-information-nepalese-agriculture-207374-201617/>

- MoF. 2019. Nepal Foreign Trade Statistics, Fiscal Year 2019/20 (2076/77) (Vol. 20).
- Nath B, Dhakre D, Bhattacharya D. 2019. Forecasting wheat production in India: An ARIMA modelling approach. *Journal of Pharmacognosy and Phytochemistry*, 8(1): 2158–2165. https://www.researchgate.net/publication/331471229_Forecasting_wheat_production_in_India_An_ARIMA_modelling_approach
- NPC. 1995. Nepal agriculture perspective plan (1995/96-2014/15) - Summary. In Agricultural Projects Service Centre (p. 53). http://lib.icimod.org/record/4168/files/APROSC_Nepalagricultureperspectiveplan630AGN.pdf
- Padhan PC. 2012. Application of ARIMA Model for Forecasting Agricultural Productivity in India. *Journal of Agriculture and Social Sciences*, 50–56.
- Rai MK, Paudel B, Zhang Y, Khanal NR, Nepal P, Koirala HL. 2019. Vegetable farming and farmers' livelihood: Insights from Kathmandu Valley, Nepal. *Sustainability (Switzerland)*, 11(3): 1–17. <https://doi.org/10.3390/su11030889>
- Sharma PK, Dwivedi S, Ali L, Arora RK. 2018. Forecasting maize production in India using ARIMA model. *Agro-Economist*, 5(1): 1–6. <https://doi.org/10.30954/2394-8159.01.2018.1>
- Tripathi R, Nayak AK, Raja R, Shahid M, Kumar A, Mohanty S, Panda BB, Lal B, Gautam P. 2014. Forecasting Vegetables Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models. *Advances in Agriculture*, 1–9. <https://doi.org/10.1155/2014/621313>
- Verbeek M. 2004. *A Guide to Modern Econometrics* (Second ed.). West Sussex, England: John Wiley and Sons, Ltd.