International Journal of Agriculture Sciences
ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 9, 2022, pp.-11661-11664.
Available online at https://bioinfopublication.org/pages/jouarchive.php?id=BPJ0000217

Research Article

MODELLING AND PREDICTING OF AREA, PRODUCTION AND PRODUCTIVITY OF ONION CROP IN ANDHRA PRADESH

R. SHAFIYA[1], G. MOHAN NAIDU[1]*, G.M. RISHITHA[2] AND S. GOVINDA RAO[3]
[1]Department of Statistics and Computer Applications, S. V. Agricultural College, Tirupati, 517502, Acharya N. G. Ranga Agricultural University, Lam, 522034, India
[2]Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632 014, Tamil Nadu, India
[3]Department of Statistics and Computer Applications, Agricultural College, Naira, 532185, Acharya N. G. Ranga Agricultural University, Lam, 522034, India
*Corresponding Author: Email - naidu_svag2001@yahoo.com, govinda.seepana@gmail.com

Received: September 03, 2022; Revised: September 26, 2022; Accepted: September 27, 2022; Published: September 30, 2022

Abstract: Onion (*Allium cepa* L) is one of the important condiments widely used year-round in all households. The present study attempted to analyze the trends and forecasts in area, production and productivity of onion (*Allium cepa* L) crop grown during the period 1978-79 to 2017-18 in Andhra Pradesh. Trends and forecasts in area, production and productivity were studied using different linear, non-linear and time series models. Based on model selection criteria such RMSE and MAE, cubic, ARIMA (2, 1, 1), and ARIMA (2, 0, 1) models were chosen to fit the trends and forecasts in area, production, and productivity. The forecasted values of area and production of onion had shown an increasing trend, while productivity had a slightly fluctuating trend.

Keywords: *Area, Productivity, Time series models, Ljung-Box Q test, RMSE and MAE*

Citation: R. Shafiya, *et al*., (2022) Modelling and Predicting of Area, Production and Productivity of Onion Crop in Andhra Pradesh. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 9, pp.- 11661-11664.

Academic Editor / Reviewer: Amit Anil Shahane

Introduction

Horticulture is the agricultural science of growing fruits, vegetables, flowers, and other plants in a way that maximizes agricultural productivity while conserving resources and protecting the environment. India ranks second in fruits and vegetable production in the world, after China. It is also said that horticulture can be viewed as 'the mirror to society'. Horticultural crops provide year-round source of income for small farmers and agricultural labour. The cultivation and processing of these crops provide sustainable job possibilities for the rural and peri-urban residents. Furthermore, marketing creates employment opportunities for the urban poor ensuring a better standard of living. Onion (*Allium cepa* L) is one of the important condiments widely used year-round in all households. Onion is one of the most popular vegetables that form a daily diet. Its demand is worldwide. Onion is exported from India to 38 countries in varying quantities. Thus, from a food, nutritional, medicinal, and economic security point of view, study of production behaviour and its future prediction is of utmost importance. Indian onions are famous for their pungency and are available round the year. According to the Agricultural and Processed Food Products Export Development Authority (www.apeda.gov.in, 2021), India is the world's second largest onion growing country. The Major onion producing states area Maharashtra, Karnataka, Madhya Pradesh, Gujarat, Bihar, Andhra Pradesh, Rajasthan, Haryana, and Telangana. There is a lot of demand of Indian onion in the world. It is a very important vegetable crop in Andhra Pradesh. Numerous researchers calculated growth rates with the aid of parametric models. Naidu (2015) [1] applied various non-linear models to project the area, production, and productivity of sugarcane in Andhra Pradesh. Rao and Naidu (2021) [2] studied various non-linear models to forecast the area, production, and productivity of wheat crop in India. Singh and Supriya (2017) [3] evaluated the growth trends in area, production, and productivity of wheat crop in India and Uttar Pradesh for the period of 1971 to 2010 by making use of different growth models such as linear, logarithmic, sigmoid, compound, power, cubic, quadratic, inverse, growth, and exponential models.

For various types of agricultural crops, researchers conducted numerous experiments to fit ARIMA models in the agriculture sector around the world. The relevant work for forecasting by using Box-Jenkins (1976) [4] ARIMA model, from which we get the idea about forecasting techniques for different types of agricultural productions such as Paswan *et al.* (2022) [5] estimated the best fitted models for forecasting sugarcane crop using Box-Jenkins ARIMA model. The results show that the ARIMA (1, 1, 0) model is the most suitable for forecasting. Rajan and Palanivel (2018) [6] aimed at presenting models for forecasting time series data regarding area, production, and productivity of cotton (*Gossypium hirsutum*) in Tamil Nadu by using Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) models. Thapa *et al.* (2022) [7] forecasted area, production, and productivity of vegetable crops in Nepal using the Box-Jenkins ARIMA model and concluded that ARIMA (0, 2, 1) model was found suitable for area and production whereas ARIMA (0, 2, 0) model for forecasting vegetable productivity based on Akaike information Criterion (AIC) and Schwarz Bayesian Criterion (SBC).

Materials and Methods

The secondary data collected for a period of 40 years from 1978-79 to 2017-18 of onion have been used to study the linear, non-linear and time series models. The time series data have been taken from the Economic and Political Weekly Research Foundation (EPWRF) and indiastat.com.

The current study tried to fit linear, non-linear and time series models such as linear, quadratic, cubic, compound, power, exponential, logistic and time series models to study trend and to predict area, production, and productivity. Model diagnostic checking was done through minimum of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) with highest $\bar{R}^2$. The models along with their equations are given below:

Linear and Non-Linear Growth Models

Linear model: $y_t=a+bt$

Quadratic model: $y_t=a+bt+ct^2$
Cubic model: $y_t=a+bt+ct^2+dt^3$
Compound model: $y_t=ab^t$
Power model: $y_t=at^b$
Exponential model: $y_t=ae^{bt}$
Logistic model: $yt=a/(1+be^{-ct})$

Time Series Models

A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally spaced intervals with a consistency in the activity and the method of measurement.

Auto Regressive Process

Yule (1926) [8] carried out the original work on auto regressive process. Auto regressive processes as their name suggests regressions on themselves.
Specially, a p^{th} order autoregressive y_t process satisfied the equation

$$y_t=\mu+\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+.....+\varphi_p y_{t-p}+\varepsilon_t \quad (1)$$

The current value of the series y_t is a linear combination of the 'p' most recent past values of itself plus an "innovation" term ε_t that incorporates everything new in the series at time 't' that is not explained by the past values. Thus, for every t, we assume ε_t that is independent of
$y_{t-1}, y_{t-2},.....,y_{t-p}$

Moving Average Process

Moving average modes were first considered by Wold (1938) [9]. The moving average series can be written as

$$y_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-.....-\theta_q\varepsilon_{t-q} \quad (2)$$

We call such series as moving average series with order q and it can be abbreviated as MA(q), where y_t is the original series and ε_t is the series of errors.

Auto Regressive Integrated Moving Average (ARIMA) model

The Box and Jenkins (1976) procedure is the milestone of the modern approach to time series analysis. Given an observed time series, the aim of the Box and Jenkins procedure is to build an ARIMA model. Passing by opportune preliminary transformation of the data, the procedure focuses on stationary processes.
In this study, it is tried to fit the Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) model. This model is the generalized model of the non-stationary ARMA model denoted by ARMA(p,q) can be written as

$$y_t=\varphi^1 y_{t-1}+\varphi_2 yt\text{-}2+.....+\varphi_p y_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-.....-\theta_q\varepsilon_{t-q} \quad (3)$$

Where y_t is the original series, for every t, we assume that ε_t is independent of
$y_{t-1}, y_{t-2},.....,y_{t-p}$
Fortunately, for practical purposes, we can usually take the values, d = 1 or at most 2.
The equation of the ARIMA model is given by

$$y_t=\mu+\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+.....+\varphi_p y_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-.....-\theta_q\varepsilon_{t-q} \quad (4)$$

Box and Jenkins procedure's steps

Preliminary analysis

Create conditions such that the data at hand can be considered as the realization of a stationary process

Identification: Specify the orders p, d, q of the ARIMA model so that it will be clear regarding the number of parameters to estimate. Recognizing the behavior of empirical autocorrelation function plays an extremely important role.

Estimate: Efficient, consistent and sufficient estimate of the parameters of the ARIMA model

Diagnostics: Check whether the model is a good one using tests on the parameters and residuals of the model. Note that when the model is rejected, still this is a very useful step to obtain information to improve the model.

Usage of the model: If the model passes the diagnostic steps, then it can be used to interpret a phenomenon forecast.

Ljung-Box test

Ljung-Box test can be used to check autocorrelation among the residuals. If a model fit well, the residuals should not be correlated and the correlation should be small. In this case the null hypothesis is $H_o:\rho_1(e)=\rho_2(e)=......=\rho_k(e)=0$ which is tested with the Box-Ljung statistic

$$Q=n(n+1)\sum_{i=1}^{k}\frac{\rho^2{}_k(e)}{n-k}$$

Where, n is the number of observations used to estimate the model. This statistic approximately follows the chi-square distribution with (k-q)df, where q is the number of parameters should be estimated in the model. If Q is large (significantly large from zero), it is said that the residuals autocorrelation is as a set are significantly different from zero and random shocks of estimated model are probably auto-correlated. So, one should then consider reformulating the model.

The most useful forecast evaluation criteria are RMSE, MAE and $\overline{R}^2$.Models are compared according to the minimum values of RMSE and MAE with highest $\overline{R}^2$.

$$RMSE=\sqrt{\frac{\sum_{t=1}^{n}\left(y_t-\hat{y_t}\right)^2}{n}} \text{ and } \qquad MAE=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t|$$

Where y_t is the actual values and $\hat{Y}_t$ is the predicted values
Adjusted $R^2=\overline{R}^2=1-\left[(1-R^2)\frac{(n-1)}{n-p-1}\right]$
Where n is the number of observations and p is the number of parameters.

Results and Discussion

To understand the behaviour of the area, production, and productivity of onion crop in Andhra Pradesh, time series data was analyzed by different types of nonlinear models for the purpose of modeling and predicting *viz*., linear, quadratic, cubic, compound, power, exponential and logistic. ARIMA modeling techniques were also used and analyzed using the available information from 1978-79 to 2017-18. It was observed from the [Table-1], that the area of the onion crop revealed that among the linear, non-linear and the time series models, the maximum adjusted R^2 of 89.2 per cent was observed in case of cubic model with comparatively lower values of RMSE (6.959) and MAE (3.557). The cubic model is a sophisticated model for accurate forecasting.

Table-1 Characteristics of fitted linear and time series models for area under onion crop

Model	Parameters				Criteria		
	a	b	c	d	$\overline{R}^2$	RMSE	MAE
Linear	6.507	1.198			0.731	8.236	5.513
Quadratic	17.081	-0.312	0.037		0.802	6.971	3.627
Cubic	15.862	0.024	0.017	0	0.892	6.959	3.557
Compound	12.882	1.038			0.890	7.350	3.775
Power	8.389	0.434			0.660	10.699	6.568
Exponential	12.882	0.037			0.890	7.350	3.775
Logistic	0.078	0.963			0.890	7.350	3.775
Time Series Models							
ARIMA (1, 1, 1)					0.791	7.677	4.096
ARIMA (1, 1, 2)					0.816	7.289	3.571
ARIMA (2, 0, 2)					0.783	8.031	4.135
ARIMA (2, 1, 1)					0.807	7.462	3.831
ARIMA (2, 1, 2)					0.823	7.251	3.840

Hence, the fitted cubic model equation can be written as
$\hat{Y}_t=15.862+0.024t+0.017t^2+0.0003t^3$
The parameters of the cubic model were presented in [Table-2]. Actual and predicted values of onion area are depicted in [Fig-1].

Table-2 *Parameters of the best fitted model*

Model	Term	Coefficient	Sig.
Cubic	a	15.862**	0.004
	b	0.024	0.982
	c	0.017	0.784
	d	0.0003	0.735

***Significant at 1% level*

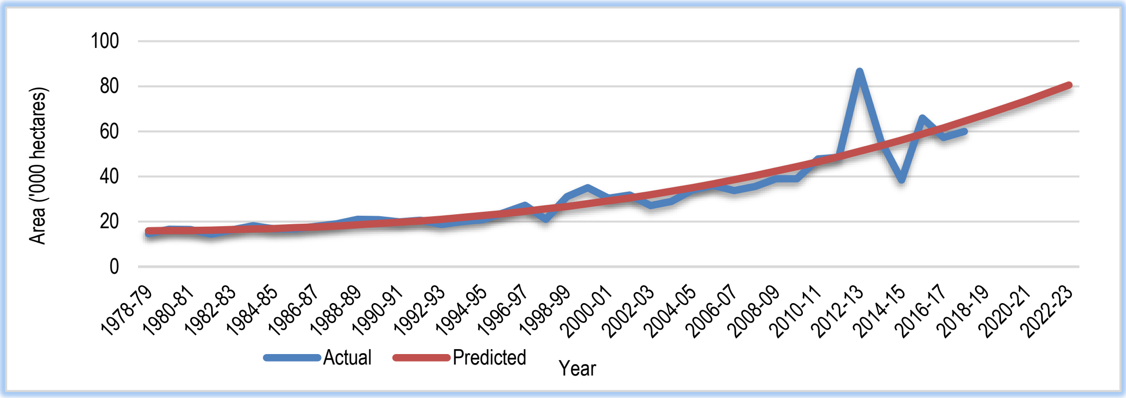

Fig-1 Actual and predicted values of onion area in Andhra Pradesh

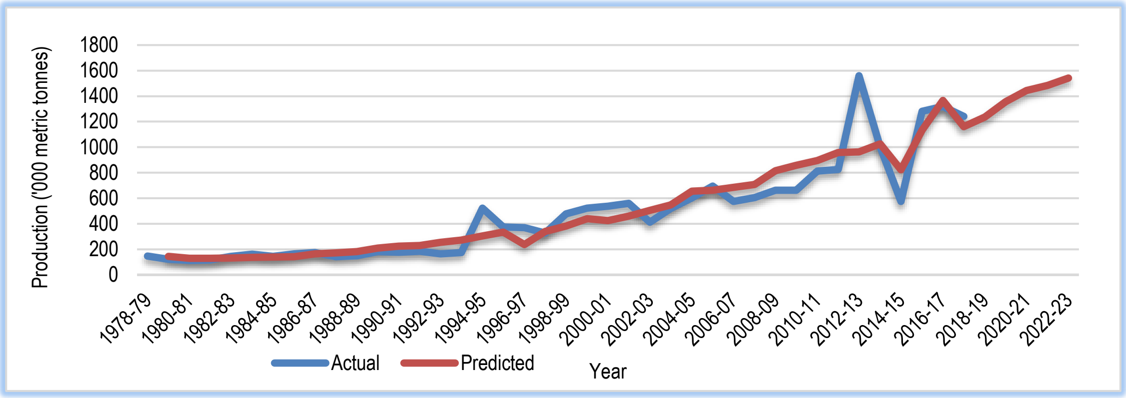

Fig-2 Actual and predicted values of onion production in Andhra Pradesh

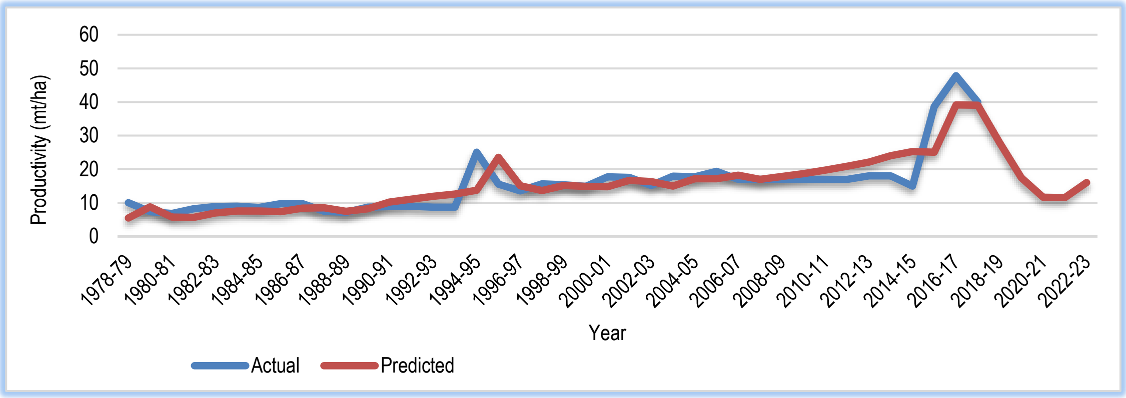

Fig-3 Actual and predicted values of onion productivity in Andhra Pradesh

Table-3 Characteristics of fitted linear and time series models for production under onion crop

Model	Parameter				Criteria		
	a	b	c	d	$\bar{R}^2$	RMSE	MAE
Linear	-95.269	28.442			0.774	174.636	126.925
Quadratic	139.334	-5.073	0.817		0.840	144.997	84.194
Cubic	108.640	3.392	0.308	0.008	0.836	144.653	84.586
Compound	96.382	1.067			0.906	145.742	82.775
Power	43.765	0.768			0.705	231.459	140.495
Exponential	96.382	0.065			0.906	145.742	82.775
Logistic	0.01	0.937			0.906	145.742	82.775
Time Series Models							
ARIMA (1,1,1)					0.835	158.008	87.066
ARIMA (1,1,2)					0.852	151.175	82.973
ARIMA (2,0, 2)					0.812	172.448	118.017
ARIMA (2, 1, 1)					0.910	143.641	82.424
ARIMA (2,1,2)					0.866	146.684	86.575

The data shown in [Table-3] for the production under the onion crop revealed that among various models fitted to the production, the maximum adjusted R^2 of 91 per cent was observed in case of ARIMA (2, 1, 1) model with comparatively lower values of RMSE (143.641) and MAE (82.424). The parameter estimates of the fitted ARIMA (2, 1, 2) model are given in [Table-4]. Based on $\bar{R}^2$, RMSE and MAE, the model was selected and used to generate the forecast values.

The fitted ARIMA (2,1,1) model is

$$Y_t = -5.210 - 0.035y_{t-1} - 0.404y_{t-2} - 0.994\varepsilon_{t-1} + \varepsilon_t$$

Among all the estimated values of the parameters, second auto regressive term has statistically significant effects on onion production at 1 per cent level of significance and it can be presented in [Table-4]. Actual and predicted values are depicted in the [Fig-2].

Table-4 *Summary statistics of the ARIMA (2,1,1) model for onion production*

Term	Coefficient	SE	t-value	p-value
Constant	-5.210	6.465	-0.806	0.426
AR(1)	-0.035	0.166	-0.207	0.837
AR(2)	-0.404**	0.163	-2.483	0.010
MA(1)	0.994	1.355	0.733	0.468

***Significant at 1% level*

The productivity data under the onion crop presented in [Table-5] revealed that among the linear, non-linear and the time series models, the best fitted ARIMA model to forecast the onion productivity in Andhra Pradesh was ARIMA (2, 0, 1) model with highest

Table-5 Characteristics of fitted non-linear and time series models for productivity under onion crop

Model	Parameter				Model Criteria		
	a	b	c	d	$\overline{R}^2$	RMSE	MAE
Linear	3.946	0.566			0.537	5.933	3.995
Quadratic	9.677	-0.252	0.0200		0.600	5.435	3.606
Cubic	3.565	1.434	-0.082	0.002	0.644	5.059	3.716
Compound	6.801	1.035			0.709	5.655	3.467
Power	4.617	0.396			0.517	6.874	3.621
Exponential	6.801	0.034			0.709	5.655	3.467
Logistic	0.147	0.966			0.709	5.655	3.467
Time Series Models							
ARIMA (1, 1, 1)					0.706	5.022	2.813
ARIMA (1, 1, 2)					0.721	4.966	2.737
ARIMA (2, 0, 1)					0.754	4.623	2.243
ARIMA (2, 0, 2)					0.744	4.781	3.009
ARIMA (2, 1, 2)					0.720	5.052	2.753

$\overline{R}^2$ of 75.4 per cent, minimum values of RMSE (4.623) and MAE (2.243). The parameter estimates of the fitted ARIMA (2, 0, 1) are given in [Table-6]. Actual and predicted values are shown in the [Fig-3].

Therefore, the fitted ARIMA (2, 0, 1) model is

$\hat{Y}_t = 5.025+1.470y_{t-1} -0.784y_{t-2} -0.996\varepsilon_{t-1} + \varepsilon_t$

From the [Table-6] it is obvious that among all the estimated values of the parameters, auto regressive terms have statistically significant effects on papaya productivity at 1 per cent level of significance.

Table-6 *Summary Statistics of the ARIMA (2, 0, 1) model for onion productivity*

Term	Coefficient	SE	t-value	Sig.
Constant	5.025**	0.889	5.652	0.000
AR(1)	1.470**	0.176	8.340	0.000
AR(2)	-0.784**	0.168	-4.677	0.000
MA(1)	0.996	3.095	0.322	0.749

*** Significant at 1% level*

The ARIMA models verification (or) diagnosed by the Ljung-Box Q statistic. The p-value associated with Q statistics should be large (ρ-value>α). The results of estimation are reported in [Table-7] indicating that the autocorrelation function of the residuals was not significant. Hence, we conclude that the model was adequate.

Table-8 *Predicted values of area, production and productivity of onion crop*

Year	Area('000 hectares)	Production ('000 metric tonnes)	Productivity (mt/ha)
2018-19	67.437	1236.76	27.94
2019-20	70.540	1355.33	17.48
2020-21	73.758	1443.25	11.70
2021-22	77.095	1484.34	11.55
2022-23	80.552	1541.74	16.00

Conclusion

The study revealed the best fitted models for area, production and productivity of onion crop up to 2022-23. When compared with other models, Cubic, ARIMA (2, 1, 1) & ARIMA (2, 0, 1) models were taken as best fitted models to predict the area, production, and productivity of onion for the next 5 years (2018-19 to 2022-23) of Andhra Pradesh. The forecasted pattern of area and production of onion had shown an increasing trend, while productivity had a decreasing trend and increased in the recent year. The forecasted results for the year 2022-23 were approximately 80.55 thousand hectares of onion area, production of 1541.74 thousand metric tonnes and 16.00 mt/ha of productivity.

Application of research: To boost the yields of these crops, policies, Sustainable agriculture, new technologies and scientific cultivation methods need to be introduced.

Research Category: Agricultural Statistics and Computer Applications

Abbreviations: RMSE-Root Mean Square Error
ARIMA- autoregressive integrated moving average

Acknowledgement / Funding: Authors are thankful to Department of Statistics and Computer Applications, S. V. Agricultural College, Tirupati, 517502, Acharya N. G. Ranga Agricultural University, Lam, 522034, Andhra Pradesh, India. Authors are thankful to Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632 014, Tamil Nadu, India and Department of Statistics and Computer Applications, Agricultural College, Naira, 532185, India

****Research Guide or Chairperson of research: Dr S. Govinda Rao**
University: Acharya N. G. Ranga Agricultural University, Lam, 522034, Andhra Pradesh, India
Research project name or number: Research station study

Author Contributions: All authors equally contributed

Author statement: All authors read, reviewed, agreed and approved the final manuscript. Note-All authors agreed that- Written informed consent was obtained from all participants prior to publish / enrolment

Study area / Sample Collection: S. V. Agricultural College, Tirupati, 517502

Cultivar / Variety / Breed name: Onion (*Allium cepa* L)

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Ethical Committee Approval Number: Nil

References

[1] Naidu G.M. (2015) *International Journal of Agricultural and Statistical Sciences*, 11(2), 215-218.
[2] Rao S.G. and Naidu G.M. (2021) *International Journal of Agricultural and Statistical Sciences*, 17(2), 539-543.
[3] Singh M. and Supriya K. (2017) *International Journal of Current Microbiology and Applied Sciences*, 6 (11), 2587-2595.
[4] Box G.E.P. & Jenkins G.M. (1976) *Time Series Analysis, Forecasting and Control. San Francisco, Holden-Day, California, USA.*
[5] Paswan S., Paul A., Paul A. and Noel A.S. (2022) *The Pharma Innovation Journal*, 11(4), 1947-1956.
[6] Rajan M.S. and Palanivel M. (2018) *International Journal of Education and ScienceResearch Review*, 4(3), 1-9.
[7] Thapa R., Devkota S., Subedi S. and Jamshidi B. (2022) *Turkish Journal of Agriculture-Food Science and Technology*, 10(2), 174-181.
[8] Yule G.U. (1926) *Journal of the Royal Statistical Society*, 89(1), 1-63.
[9] Wold H.O.A. (1938) *A study of the analysis of Time Series.* (2nd Edition, 1954). Uppsala: Almqvist and Wiksells)