Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol No. 24(2) : 200-202 (June 2022)

https://journal.agrimetassociation.org/index.php/jam

Short Communication

ARIMA and ARIMAX models for sugarcane yield forecasting in Northern Agro-climatic zone of Haryana

PUSHPA*, CHETNA, ADITI and URMIL VERMA

Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, India
**Corresponding author's email*: pushpaghiyal@gmail.com

Various statistical approaches like regression, time-series and stochastic models are in vogue for arriving at crop forecasts. Though, the performance of Lasso regression was better than stepwise regression to some extent (Kumar *et al*., 2019). Every approach has its own advantages and limitations. The application of the autoregressive integrated moving average (ARIMA) models in the field of agriculture for forecasting a variety of study variables of interest for different crops/regions etc. When an ARIMA model includes other time series as input variables, the model is referred to as an ARIMAX model. In the present study was undertaken with the following objectives i) Development of ARIMA models for sugarcane yield forecasting and ii) Fitting ARIMAX models and testing the post-sample validity of the developed models. Ghosh *et al.,* (2014) developed wheat crop yield forecast model using the statistical model. Paul *et al.* (2009) have conducted the time-series analysis for modeling and forecasting of spices export data in India. Anggraeni *et al.* (2015) compared the performance of ARIMA as univariate time series method and ARIMAX as multivariate method for forecasting the demand of Moslem kids clothes.

The present study dealt with modeling the time-series data related to sugarcane yield in Karnal, Kurukshetra and Ambala districts of Haryana. The sugarcane yield data from 1966-67 (Karnal and Ambala) and 1972-73 (Kurukshetra) to 2013-14 and weather data from 1978-79 to 2013-14 were used for the training set and the remaining data *i.e.* 2014-15 to 2018-19 were used for the post-sample validity checking of the developed ARIMA and ARIMAX models. The sugarcane yield data of Karnal, Ambala and Kurukshetra district were compiled from the Statistical Abstracts of Haryana. The weather data were collected from Meteorological Centre, Chandigarh of IMD.

Box-Jenkins (1976) ARIMA forecasts are based on past values of the variable being forecast. An ARIMA model requires a minimum sample size of about 35-40 time-series observations and applies only to stationary time series data. A stationary time series has mean, variance and auto-correlation function essentially constant over time. The general functional form of ARIMA (p,d,q) model used is:

$$\hat{Y}_t = u + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} - \theta_1 e_{t-1} - ...... - \theta_q e_{t-q}$$

Where, p is the order of autoregressive terms, d is the number of non-seasonal differences needed for stationarity, and q is the order of moving average.

Y_t= Variable under forecasting, e = Error term (Y_t - $\hat{Y}_t$, where $\hat{Y}_t$ is the estimated value of Y_t)

t = the time subscript, $\phi_p(B)$ = Non-seasonal AR and $\theta_q(B)$ = Non-seasonal MA.

ARIMA (1,1,1) = First order autoregressive and moving average model with first order of non-seasonal differencing is written as

$$Y_t = u + Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) - \theta_1 e_{t-1}$$

ARIMA (1,1,0) = First order autoregressive model with first order of non-seasonal differencing is written as

$$Y_t = u + Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2})$$

ARIMA (0,1,1) = First order moving average model with first order of non-seasonal differencing is written as $Y_t = u + Y_{t-1} - \theta_1 e_{t-1}$

ARIMAX is an acronym for auto-regressive integrated moving average with exogenous variables. An ARMAX form of the model is presented as:

Article info - DOI: *https://doi.org/10.54386/jam.v24i2.1086*
Received: 02 February 2022; Accepted: 10 April 2022; Published online: 28 May 2022

$$\phi(B)Y_t = \beta x_t + \theta(B)a_t \quad \text{or} \quad Y_t = \frac{\beta}{\phi(B)}x_t + \frac{\theta(B)}{\phi(B)}a_t$$

Where, x_t is a independent variable and β is coefficient of x_t. B can only be interpreted conditional on the previous values of the response variable. The fortnightly weather data on maximum temperature, minimum temperature and rainfall were used as exogenous input variables to developed ARIMAX models. For example first fortnightly of January for weather data on maximum temperature, minimum temperature and rainfall denoted as tmx_1, tmn_1 and arf_1 *i.e.* tmx_7 is represent for 1st fortnight of April for maximum temperature, tmn_1 is represent for 1st fortnight of January and tmn_{10} is represent for 2nd fortnight of May for minimum temperature, arf_8 is represent for 2nd fortnight of April and arf_{11} is represent for 1st fortnight of June for accumulated rainfall.

Table 1: Selection criteria values of ARIMA models considered for all the districts

District (s)	Model fit statistic (s)	ARIMA (0,1,1)	ARIMA (1,1,0)
Karnal	RMSE	5.41	6.17
	MAPE	7.42	8.09
	BIC	3.59	3.87
Ambala	RMSE	5.72	5.85
	MAPE	9.50	9.96
	BIC	3.64	3.78
Kurukshetra	RMSE	6.84	7.83
	MAPE	10.39	10.73
	BIC	3.94	4.15

Table 2: District-level estimated sugarcane yield (q/ha) based on ARIMA and ARIMAX models and their percent relative deviations.

District/ Model	Forecast Year	Observed yield ($q\ ha^{-1}$)	Estimated yield ($q\ ha^{-1}$)	Percent relative deviation	ARIMA with input series	Estimated yield ($q\ ha^{-1}$)	Percent relative deviation
Karnal ARIMA (0,1,1)	2015-16	85.04	76.94	6.00	ARIMA (0,1,1) with tmn_1 and arf_8	84.92	0.14
	2016-17	84.54	77.68	6.93		89.63	-6.02
	2017-18	95.00	78.14	1.96		90.55	4.68
	2018-19	93.13	79.14	15.02		86.69	6.92
	2019-20	83.99	79.88	4.89		83.71	0.33
Ambala ARIMA (0,1,1)	2015-16	70.55	69.84	1.01	ARIMA (0,1,1) with tmn_{10}	71.26	-1.01
	2016-17	69.60	70.54	-1.35		72.41	-4.04
	2017-18	78.13	70.26	10.07		72.95	6.63
	2018-19	81.21	70.37	13.35		73.73	9.21
	2019-20	72.14	70.33	2.51		73.29	-1.59
Kurukshetra ARIMA (0,1,1)	2015-16	81.64	73.58	9.87	ARIMA (0,1,1) with tmx_7 and arf_{11}	76.7	6.05
	2016-17	82.56	76.94	6.81		77.69	5.90
	2017-18	85.57	74.14	13.36		77.11	9.89
	2018-19	92.43	82.04	11.24		82.49	10.75
	2019-20	81.09	71.13	12.28		81.74	-0.80

Table 3: Residual autocorrelations checking based on ARIMAX models for all the districts

District (s)	Model	Ljung-box Q statistic		
		Statistic	d.f.	Sig
Karnal	ARIMA (0,1,1)	8.10	17	0.96
	ARIMA(0,1,1) with tmn_1 and arf_8	17.02	17	0.45
Ambala	ARIMA (0,1,1)	29.40	17	0.31
	ARIMA(0,1,1) with tmn_{10}	25.71	17	0.08
Kurukshetra	ARIMA (0,1,1)	13.06	17	0.73
	ARIMA(0,1,1) with tmx_7 and arf_{11}	9.49	17	0.93

The sugarcane yield (s) data were found to be non stationary for all the three districts. The non-stationary data series of all the districts were transformed into stationary series by the first differencing of the original data series. The models ARIMA (1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1) were tentatively considered in the identification stage and ARIMA estimation was carried out using a non-linear least squares (NLS) approach. Lastly, the diagnostic check was performed to see whether the residuals from the fitted models were white noise. All Chi-Squared statistic (s) in this concern calculated using the Ljung-Box (1978) formula ruled out any systematic pattern in the residuals. Thus, after experimenting with different lags of the moving average and the autoregressive processes, ARIMA (0,1,1) for Karnal, Ambala and Kurukshetra districts were found to be the best fit for sugarcane yield estimation. The above models were used to obtain the sugarcane yield forecasts for the post sample period(s) *i.e.* 2015-16, 2016-17, 2017-18, 2018-19, and 2019-20.

ARIMA (0,1,1) along with fortnightly weather variables selected on the basis of stepwise regression method (viz., tmx_7, tmn_1, tmn_{10}, arf_8 and arf_{11} over the crop growth period) as input series

Table 4: Model fit statistics of ARIMA and ARIMAX models.

District (s)	Model	Model fit statistic		
		RMSE	MAPE	SBC
Karnal	ARIMA (0,1,1)	5.51	8.42	3.58
	ARIMA (0,1,1) with tmn_1 and arf_8	4.60	5.97	3.48
Ambala	ARIMA (0,1,1)	5.62	9.49	3.54
	ARIMA (0,1,1) with tmn_{10}	4.16	6.16	3.14
Kurukshetra	ARIMA (0,1,1)	6.86	10.4	3.94
	ARIMA (0,1,1) with tmx_7 and arf_{11}	5.21	6.25	**3.72**

Table 5: Comparative view in terms of RMSEs of sugarcane yield forecasts based on ARIMA and ARIMAX models.

Distict(s)	RMSE(s)	
	ARIMA model	ARIMAX model
Karnal	7.44	4.18
Ambala	6.07	4.30
Kurukshetra	9.32	6.62

were utilized in fitting ARIMAX models for Karnal, Ambala and Kurukshetra districts sugarcane yield estimation.

Summarizing the results, sugarcane yield forecasts for the years 2015-16, 2016-17, 2017-18, 2018-19 and 2019-20 were obtained on the basis of ARIMA and ARIMAX models as shown in tables 2. It is expressed that ARIMA with weather variables *i.e.* ARIMA (0,1,1) with tmn_1 and arf_8, ARIMA (0,1,1) with tmn_{10} and ARIMA (0,1,1) with tmx_7 and arf_{11} models consistently showed the superiority over ARIMA(0,1,1) models in capturing lower percent deviations, mean absolute percentage error (MAPE), RMSE and RD% *etc.* (Tables 4&5) for sugarcane yield prediction in Karnal, Ambala and Kurukshetra districts of Haryana. The ARIMAX models performed well with lower error metrics as compared to the ARIMA models in all time regimes. The developed models are capable of providing the reliable estimates of district-level sugarcane yield well in advance of the crop harvest and may be used successfully for different stakeholders.

ACKNOWLEDGEMENTS

The authors are highly thankful to CCS Haryana Agricultural University Hisar, Haryana, India for conducting above study.

Conflict of Interest Statement: The author (s) declares (s) that there is no conflict of interest.

Disclaimer: The contents, opinions and views expressed in the research article published in Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Anggraeni, W., Vinarti, R.A., Kurniawati, Y.D. (2015). Performance comparisons between ARIMA and ARIMAX method in moslem kids clothes demand forecasting: Case Study. *Procedia Com. Sci.*, 72(3): 630 – 637.

Box, G.E.P., G.M. Jenkins (1976). Time series analysis. Forecasting and Control, *Holden Day, San Francisco*.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K.K. and Rathore, L.S. (2014).Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal. *J. Agrometeorol.,* 16 (1): 1-8.

Kumar, S.; Attri, S.D. and Singh, K.K. (2019). Comparison of Lasso and stepwise regression technique for wheat yield prediction. *J. Agrometeorol*., 21 (2): 188-192.

Ljung, G.M., G.E.P. Box (1978). On a measure of lack of fit in time series models, *Biometeorol.*, 65: 297-303.

Paul, R.K., Prajneshu and Ghosh, H. (2009). GARCH nonlinear time series analysis for modeling and forecasting of India's volatile spices export data. *J. Indian Soc. Agril. Stat*. 63(2):123-131.