Research Article

Indian Journal of Extension Education
Vol. 60, No. 1 (January–March), 2024, (95-99)

ISSN 0537-1996 (Print)
ISSN 2454-552X (Online)

ARIMA and ARIMAX Analysis on the Effect of Variability of Rainfall, Temperature on Wheat Yield in Haryana

Megha Goyal[1], Subodh Agarwal[1], Suman Ghalawat[1]* and Joginder Singh Malik[2]

[1]Assistant Professor, Department of Business Management, [2] Professor, Agricultural Extension Education, CCS Haryana Agricultural University, Hisar-125004, Haryana, India
*Corresponding author email id: suman.ghalawat@hau.ac.in

ARTICLE INFO

Keywords: Autocorrelation function (ACF), Partial autocorrelation function (PACF), Stationarity, Invertibility, Minimum temperature and rainfall, ARIMA and ARIMAX models

https://doi.org/10.48165/IJEE.2024.60118

Conflict of Interest: None

Research ethics statement(s): Informed consent of the participants

ABSTRACT

The national and state governments require crop production forecasts to make a variety of policy decisions on import-export, storage, distribution, price, and other factors. This article presents a pre-harvest forecasting method specially developed for crops grown in the western region of Haryana (India). The western region includes Hisar, Sirsa, and Bhiwani districts. For crop forecasting in the Hisar, Sirsa, and Bhiwani regions ARIMA and ARIMAX models have been framed. For the development of the ARIMAX model, climate data during the growing season of the crop were used as input along with the crop yield. The percentage difference between the root mean square error and the wheat yield estimations determined by the real-time yield(s) indicates how well-competing models performed in terms of forecasting. The ARIMAX model performs well at all time points with a lower measurement error compared to the ARIMA model.

INTRODUCTION

The present study was undertaken with the following objectives (i) Development of univariate ARIMA wheat yield prediction models and (ii) Fitting ARIMAX (weather parameters as regressors) models and testing the validity of the developed models. India's primary cereal crop is wheat. In the nation, there are around 29.8 million hectares of cropland. From 75.81 million tonnes in 2006–07 to 94.88 million tonnes in 2011–12 to 96.6 million tonnes in 2016–17, the nation's wheat production has expanded dramatically. India's wheat production decreased from 109.59 million tonnes in 2021 to 106.84 million tonnes in 2022. Due to the larger acreage and favorable weather (crop statics wheat), production could increase to 112 million tonnes this year. In their study of the impact of several climatic factors on wheat output, Kumar et al., (2001) discovered a negative correlation between maximum temperature and yield of late-planted wheat in the Tarai region. Goyal & Verma (2015) have used regression and principal component analysis to develop the wheat yield models on agro-climatic zone basis in Haryana State using spectral and weather data. Dharmaraja et al., (2020) predicted Bajra yield of Alwar district of Rajasthan using linear regression and time-series models. Goyal et al., (2021) used univariate time series autoregressive integrated moving average (ARIMA) model to analyze the trend and forecast of Agricultural Export in India.

Vikash et al., (2021) used ARIMA model to forecast the diverse range of vegetables in Haryana and Parveen et al., (2022) investigated the behaviour of the area, production and productivity of tomato crop in the Haryana and India by using different forecast models including ARIMA. Parkash et al., (2022) used SARIMA and other models to forecast sweet potato price. Pawan et al., (2018) forecast maize production for the year 2018 to 2022 based on the estimation of suitable ARIMA model. For univariate time series (crop yield) estimation, autoregressive integral moving average (ARIMA) model have been used in the past. However, this model cannot include exogenous variables according to ARIMA is preferred for more accurate estimation of crop yield. Sanjeev & Verma (2016) developed ARIMA and ARIMAX models for sugarcane

Received 13-12-2023; Accepted 29-12-2023
The copyright: The Indian Society of Extension Education (https://www.iseeiari.org/) vide registration number L-129744/2023

prediction in Karnal, Ambala and Kurukshetra districts of Haryana. Ahmer et al., (2023) compared the various models such as ARIMA, Sutte ARIMA, Holt-Winters and NNAR models for effective prediction of food grains production in India.

METHODOLOGY

The 22 districts make up the Haryana State are located between 270 40' and 300 55' N latitude and 740 25' to 770 38' E longitude. The goal of the current study was to model time-series data on the wheat crop yield in the western zone of Haryana by comparing the districts of Hisar, Sirsa, and Bhiwani. The Statistical Abstracts of Haryana/Punjab were used to produce the State Department of Agriculture's (DOA) wheat yield data for the years 1978–1979 to 2019–20. The fortnightly weather variables used as input series (1978-79 to 2019-20), were taken for the study from Indian Meteorology Department.Weather data starting from the 1st fortnight of November to 1 month before harvest were utilized for the model building (crop growth period: 1st November to 15th April). The emphasis was placed on forecasting future values using historical time-series measurements, together with fortnightly meteorological parameters across the crop growth period as input series, keeping in mind the stated objectives. The training set consisted of time-series yield data for the wheat crop and meteorological parameters from 1978–1979 to 2016–17; the remaining data, which included the years 2017–18, 2018–19, and 2019–20, were utilized to assess the post–sample validity of the created ARIMA and ARIMAX models.

Unit Root Test is required to convert time series data into stationary form because they may not always be in that state. To accomplish this, one straightforward method is to difference the time series data. One way to do this is by using the Augmented Dickey-Fuller (ADF) t-statistic. The ADF test creates a parametric correction for higher-order correlation as follows by assuming that the y series follows an autoregressive of order p process and including p delayed difference terms of the dependent variable y to the right-hand side of the test regression:

$$\Delta y_t = \alpha y_{t-1} + x_t \delta + \beta_1 y_{t-1} + \beta_2 y_{t-2}) + \ldots\ldots + \beta_p y_{t-p} + v_t$$

In this case, the exogenous regressors x_t are optional and could be constant.

ARIMA model's standard functional form is Autoregressive Integrated Moving Average model i.e. ARIMA (p,d,q) as follows:

$$\phi_p(B)\Delta^d y_t = c + \theta_q(B)a_t \quad \ldots (1)$$

where, y = Variable under forecasting, B = Lag operator, a = Error term ($Y-\hat{y}$, where $\hat{y}$ is the estimated value of Y), t = time subscript, $\phi_p(B)$ = non-seasonal AR, Δ^d= non-seasonal difference, $\theta_q(B)$= non-seasonal MA, ϕ's and θ'$_s$ were the parameters to be estimated

The acronym ARIMAX stands for Autoregressive Integrated Moving Average with Exogenous Variables. From pure ARIMA modelling, it is a logical development to include independent variables that boost the value of explanation. Conceptually, regression and ARIMA modelling are merged. When the AR and MA terms in a pure ARIMA model are insufficient to provide a model with an acceptable level of overall explanatory power, it is only legitimate to look for additional driving events whose impact over time is not sufficiently embedded in the historical values of the dependent time series. Because the ARIMAX model also considers extra time series as input variables in addition to past values of the response series and previous errors, the model is frequently referred to as an ARIMAX model. An ARMAX form of the model is presented as:

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

where, x_t is a covariate at time t and β is its coefficient. β can only be interpreted conditional on the previous values of the response variable.

$$\phi(B) = 1-\phi_1(B)- \ldots\ldots\phi_p(B)^p$$

and $$\theta(B) = 1- \theta_1(B)-\ldots\ldots -\theta_q(B)^q$$

For ARIMA errors in case of non-stationary data, $\phi(B)$ is simply replaced $\nabla^d\phi(B)$ with where denotes the differencing operator. All the analysis has to be done through R software.

RESULTS

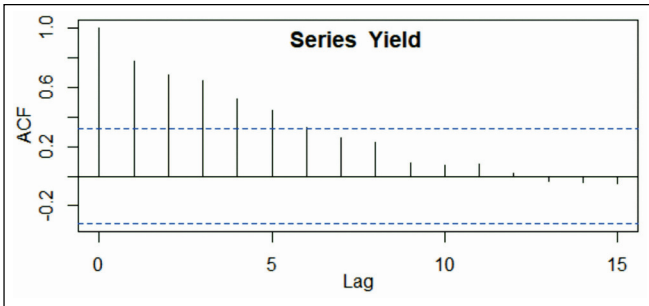
The ARIMA models were fitted using the time series wheat yield data for the period 1978-79 to 2016-17 of Hisar, Sirsa and Bhiwani districts. Nonetheless, from 1978–1979 to 2016–2017, the fortnightly meteorological data throughout the crop growth period were used to create the ARIMAX models. The fortnightly weather data on maximum temperature, minimum temperature and rainfall for the above mentioned period were used as exogenous input with ARIMA models.

For all three districts, it was discovered that the wheat yield (s) statistics were non-stationary (Table 1). Which model best describes the time series value can be determined by the behavior of the ACF and PACF. Nearly all autocorrelations up to lag 10 differ significantly from zero. The ACFs fall gradually, implying non-stationarity, according to the charting of the ACFs and PACFs, which is also shown below. The series may have an autoregressive component of order one, as the pacfs indicated by the occurrence of one major spike at lag one for all districts. The first differencing of the original data series converted the non-stationary data series of all the districts into stationary series. For obtaining an approximation, order one (i.e., d=1) differencing was sufficient.

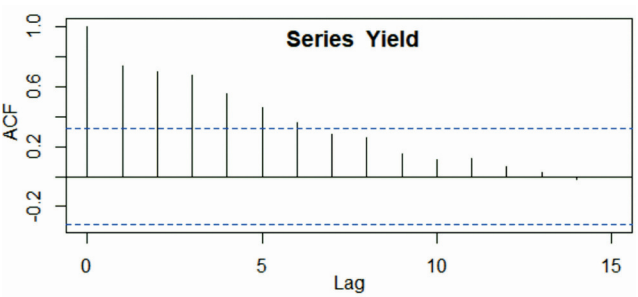
In the identification step, the models ARIMA (0,1,1), ARIMA (0,1,1), and ARIMA (0,1,2) were each tentatively considered for the districts of Hisar, Sirsa, and Bhiwani, respectively (Tables 2 and 3). ARIMA estimation was then completed using a non-linear least squares (NLS) method. In order to determine whether the residuals from the fitted models were white noise, a diagnostic check was lastly carried out. Any systematic pattern in the residuals was ruled out by all Chi-Squared statistic (s) in this study determined using the Ljung-Box (1978) formula (Table 4). As a result, after experimenting with various delays for the moving average and autoregressive processes, it was determined that ARIMA

Table 1. Unit Root Test using ADF

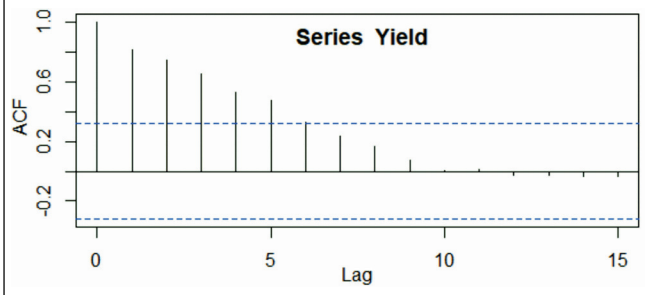
Districts	Variable	ADF	P-value	Remarks
Hisar	Yield	-2.5087	0.3745	Non-Stationary
Sirsa	Yield	-2.0112	0.5689	Non-Stationary
Bhiwani	Yield	-2.0287	0.5621	Non-Stationary



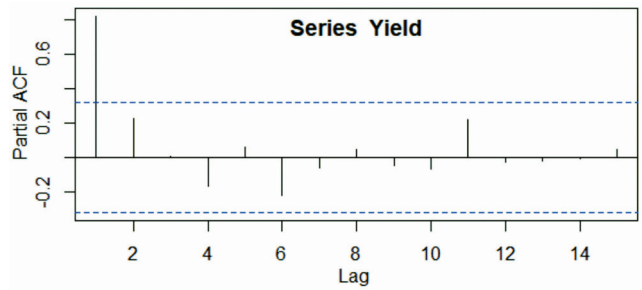
Auto correlations: Hisar



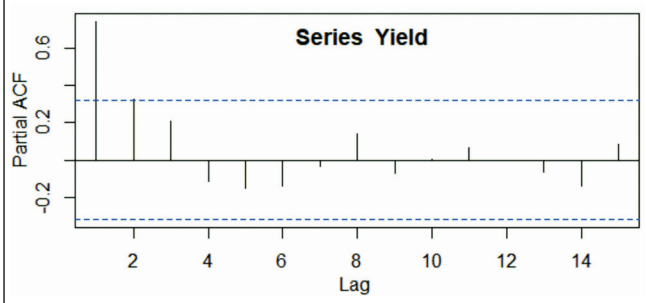
Auto correlations: Sirsa



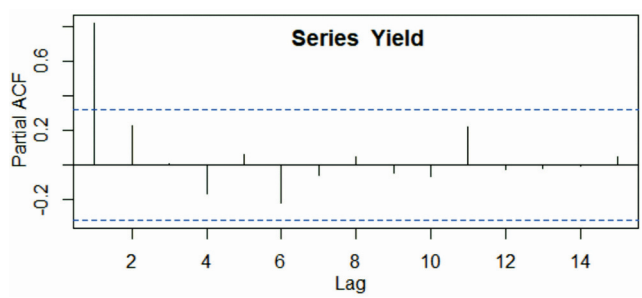
Auto correlations: Bhiwani



Partial auto correlations: Hisar



Partial auto correlations: Sirsa



Partial auto correlations: Bhiwani

(0,1,1) for the districts of Hisar and Sirsa and ARIMA (0,1,2) for the district of Bhiwani provided the best fit for estimating wheat yield. The following models were utilized to get the projections for wheat yield for the post-sample periods 2017–18, 2018–19, and 2019–20 were obtained using the models below.

ARIMAX Modeling

The best weather contributors, the average maximum temperature of the third, or tmx3, the average minimum temperature of the third and ninth, or tmn3, and the accumulated rainfall of the ninth fortnight, or arf9, over the crop growth period, were used as input series with ARIMA modeling in an attempt to improve the predictive performance. This was done while taking into account the non-stationary behavior of the series under consideration. The 27 meteorological variables—average maximum temperature, average minimum temperature, and total rainfall-were regressively analyzed to choose the useful variables listed above. To find the best weather predictors for wheat production, these meteorological variables were computed for nine fortnights throughout the crop growth period, which ran from November 1 to March 15.

Table 2. The wheat yield (q/ha) parameter estimations from the fitted ARIMA models

Models		Estimate	Standard error	t-ratio	Approx. Prob.
Hisar	Constant	58.83	20.91	2.81	0.00
ARIMA (0,1,1)	MA (1)	-0.63	0.12	-5.24	0.00
Bhiwani	Constant	54.76	23.65	2.31	0.02
ARIMA (0,1,1)	MA (1)	-0.50	0.17	-3.00	0.00
Sirsa	Constant	67.54	16.22	4.16	0.00
ARIMA (0,1,2)	MA (1)	-0.91	0.34	-3.79	0.00
	MA (2)	0.18	0.50	1.35	0.17

Table 3. The fitted ARIMAX models parameter estimations for the three districts wheat yield

	Models			Estimate	S.E.	t-ratio	Sig.
ARIMA (0,1,1)	Hisar yield	Constant		61.54	16.78	3.67	0.00
	with arf_9	MA	Lag 1	-0.67	0.11	-6.18	0.00
	arf_9	Numerator	Lag 0	-9.91	3.60	-2.76	0.00
ARIMA (0,1,1) with tmn_3	Sirsa yield	Constant		57.65	22.79	2.53	0.01
		MA	Lag 1	-0.50	0.17	-3.00	0.00
	Tmn_3	Numerator	Lag 0	-29.22	19.62	-1.48	0.31
ARIMA (0,1,2) with tmx_3, tmn_9 and arf_9	Bhiwani yield	Constant		70.71	19.33	3.65	0.00
		MA	Lag 1	-1.31	0.35	-3.79	0.00
		MA	Lag 2	0.68	0.50	1.35	0.17
	Tmx_3	Numerator	Lag 0	-39.45	19.69	-2.00	0.04
	Tmn_9	Numerator	Lag 0	-11.60	3.04	-3.81	0.00
	arf_9	Numerator	Lag 0	-33.74	24.85	-1.35	0.17

Table 4. Diagnostic checking of residual autocorrelations of wheat yield

District(s)	Model	Ljung-box Q Statistic	d.f.	Sig.
Hisar	ARIMA (0,1,1)	5.17	7	0.64
	ARIMAX (0,1,1)	5.99	7	0.54
Sirsa	ARIMA (0,1,1)	6.02	7	0.53
	ARIMAX (0,1,1)	4.95	7	0.66
Bhiwani	ARIMA (0,1,2)	4.26	6	0.64
	ARIMAX (0,1,2)	7.32	6	0.29

To estimate wheat yield, the ARIMAX models (Table 3) were developed using the ARIMA models with alternative combinations of explanatory variables, namely ARIMA (0,1,1) for Hisar & Sirsa and ARIMA (0,1,2) for Bhiwani districts, along with fortnightly weather variables (tmx3, tmn3, tmn9, and arf9 over the crop growth period) as input series.

The Marquardt approach (1963) was used to minimize the sum of squared residuals. Schwarz's Bayesian Criterion, SBC (1978), residual variance, Akaike's Information Criterion, AIC (1969), and log likelihood were used to develop the criteria for estimating the AR and MA coefficients in the model. The residual acfs was used in conjunction with the 't' tests and Chi-squared test advised by Ljung and Box to determine if random shocks were white noise (Table 4).

DISCUSSION

Forecasts for the wheat yield in the years 2017–18, 2018–19, and 2019–20 were generated using the ARIMA and ARIMAX models. Individually, either of the fitted models could offer the

Table 5. Model fit statistics of ARIMA and ARIMAX models

District(s)	Model	Model fit statistics			
		RMSE	MAPE	AIC	BIC
Hisar	ARIMA (0,1,1)	322.61	6.69	539.95	544.78
	ARIMAX (0,1,1)	292.57	6.30	534.83	541.27
Sirsa	ARIMA (0,1,1)	276.51	6.60	528.32	533.16
	ARIMAX (0,1,1)	268.49	6.46	528.16	532.60
Bhiwani	ARIMA (0,1,2)	339.62	6.81	546.23	552.67
	ARIMAX (0,1,2)	280.16	5.78	539.21	550.48

Table 6. Estimated wheat yield(s) for each district, based on ARIMA and ARIMAX models and corresponding percentage deviations (RD%) = 100 (observed yield - est. yield)/ observed yield)

Models	Year	Observed yield (q/ha)	ARIMA		ARIMAX	
			Estimated yield (q/ha)	Percent relative deviation	Estimated yield (q/ha)	Percent relative deviation
Hisar ARIMA (0,1,1) & ARIMA (0,1,1) with arf_9	2017-18	49.14	47.22	3.90	48.89	0.51
	2018-19	49.55	47.81	3.51	48.42	2.27
	2019-20	45.60	48.40	-6.14	47.56	-4.30
Sirsa ARIMA (0,1,1) & ARIMA (0,1,1) with tmn_3	2017-18	52.34	55.65	-6.33	50.87	2.80
	2018-19	50.62	52.07	-2.86	51.56	-1.85
	2019-20	48.63	52.22	-7.37	52.23	-7.41
ARIMA (0,1,2) & ARIMA (0,1,2) with tmx_3, tmn_9 & arf_9	2017-18	43.26	42.49	1.79	43.17	0.20
	2018-19	44.25	43.03	2.75	44.70	-1.02
	2019-20	41.86	43.58	-4.11	45.04	-7.60

appropriate relationship(s) needed to accurately estimate the wheat production in the district under consideration. On the basis of mean absolute percentage error (MAPE), root mean square error (RMSE), AIC, BIC, and other metrics, the predictive ability of the competing models were evaluated (Tables 5 and 6). For estimating wheat yields, the accuracy level attained by ARIMA model(s) with weather as input series was deemed sufficient, i.e., ARIMA models with weather variable(s) as input series could more effectively explain the crop yield data. In order to obtain short-term forecasts of wheat yield in the Haryana districts of Hisar, Sirsa, and Bhiwani, three-steps ahead (out-of-model development period, i.e. 2017–18, 2018–19, and 2019–20) predicted values favor the use of ARIMAX models. In terms of percent variation, yield estimations generated on ARIMA and ARIMAX were compared to DOA yields (Table 6). Similar to this, Sanjeev & Verma (2016) created ARIMA and ARIMAX models for sugarcane forecasting and discovered that ARIMAX models outperformed ARIMA models.

CONCLUSION

Summarizing the aforementioned findings, it was discovered that, unlike regression, ARIMA models occasionally couldn't produce strong results. The focus was therefore on determining if an ARIMA model with additional time series as input variable(s) improves predictions of pre-harvest crop output. In this empirical study, it was discovered that ARIMAX, an ARIMA model that uses weather variables as input series, consistently outperformed ARIMA models in capturing the percent relative deviations relating to pre-harvest wheat yield forecasts in the Hisar, Sirsa, and Bhiwani districts of Haryana. The ARIMAX models outperformed the ARIMA models, exhibiting lower error metrics throughout all time intervals. While DOA yield estimates are obtained considerably later, after the crop has actually been harvested, the developed models can also accurately predict wheat yield far in advance of the crop's actual harvest.

REFERENCES

Ahmer, S. A., Singh, P. K., Ruliana, R., Pandey, A. K., & Gupta, S. (2023). comparison of ARIMA, Sutte ARIMA, Holt-Winters and NNAR models to predict food grain in India. *Forecasting 5*(1), 138-152.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics, 21*, 243-47.

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

Dharmaraja, S., Jain, V., Anjoy, P., & Chandra, H. (2020). Empirical analysis for crop yield forecasting in India, *Agricultural Research, 9*(1), 132-138.

Goyal, M., & Verma, U. (2015). Development of weather-spectral models for pre-harvest wheat yield prediction on agro-climatic zone basis in Haryana (India). *International Journal of Agricultural and Statistical Sciences, 11*(1), 73-79.

Goyal, M., Goyal, S. K., Agarwal, S., & Kumar, N. (2021). Forecasting of Indian agricultural export using ARIMA model. *Journal of Community Mobilization and Sustainable Development, 16*(3), 655-658.

Kumar, N. P., Muhammed, J. P. K., & Aniket, C. (2022). Modelling and forecasting of area, production and productivity of tomatoes in Haryana and India. *Indian Journal of Extension Education, 58*(2), 205-208.

Kumar, S., Mishra, H. S., Sharma, A. K., & Kumar, S. (2001). Effect of weather variables on the yield of early, timely and late sown wheat in the Tarai region. *Journal Agricultural Physics, 1*(1), 58-62.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297-303.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters, *Journal of Society for Industrial and Applied Mathematics, 11*(2), 431-441.

Sanjeev and Verma, U. (2016). ARIMA versus ARIMAX modelling for sugarcane yield prediction. *International Agricultural Statistics of Science in Haryana, 12*(2), 327-334.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics, 6*(2), 461-64.

Sharma, P. K., Dwivedi, S., Ali, L., & Arora, R. K. (2018). Forecasting maize production in India using ARIMA Model. *Agro Economist - An International Journal, 5*(1), 1-6.

Sreekumar, J., & Sivakumar (2022). Forecasting of Sweet Potato (Ipomoea batatas L.) Prices in India. *Indian Journal of Extension Education, 58*(2), 15-20.

Vikash, Meena, S. S., & Verma, R. K. (2022). Forecasting of vegetable production in Haryana by ordinary least square method and ARIMA Model. *Indian Journal of Extension Education, 58*(4), 71-75.