Full Length Research Article

© 2024, Centre for Advanced Research in Agricultural Sciences
Research Journal of Agricultural Sciences
Volume 15; Issue 03 (May–Jun 2024); pp 829–836

ISSN: 0976-1675 (P)
ISSN: 2249-4538 (E)

Simulation and Modeling for Mustard Yield (*Brassica juncea* L.) Forecasting in Haryana using Weather Variables

Ajay Kumar[1] and Raj Kumar*[1]

[1-2] College of Horticulture, Maharana Pratap Horticultural University, Karnal - 132 001, Haryana, India

Received: 02 Mar 2024; Revised accepted: 26 May 2024

Abstract

Simulation and Modeling is a discipline for developing a level of understanding of the interaction of the parts of a system, and of the system as a whole. The present study compares the efficacy of time series Intervention models and simulation in quantifying the pre-harvest mustard yield in Hisar, Bhiwani, Sirsa, and Fatehabad districts of Haryana. The fortnightly weather data on rainfall, minimum temperature and maximum temperature over the crop growth period (September-October to February-March) have been utilized from 1980-81 to 2010-11 for the models' building. The weather-yield data from 2011-12 to 2015-16 have been used to check the post-sample validity of the fitted models for mustard yield forecasts in comparison to those obtained from State Department of Agriculture crop yield(s) estimates. The statistical modeling approaches; Regression with ARIMA Errors (RegARIMA) and ARIMA-Intervention were applied for the purpose. Five-steps ahead forecast figures favour the use of RegARIMA models to obtain pre-harvest mustard yield forecasts in the districts under study. The forecasts generated by RegARIMA are remarkably close to the forecasts obtained through the simulation process.

***Key words*: ARIMA-Intervention, Forecast, Regression with ARIMA errors, Simulation

India is one of the largest rapeseed-mustard growing countries in the world, occupying first position in area and third position in production after the EU27 and China, and contributing around 12% of the world's total production. India's contributions to the world acreage and production are 28.3 and 19.8 percent, respectively (Source: www.mapsofindia.com/indiaagriculture). Rapeseed is a major oilseed crop in India, grown on nearly 13% of the cropped land. The crop grows well in areas receiving 25 to 40 cm of rainfall and this is provided by the monsoon rains during the sowing season of the crop in India. The major rapeseed-mustard growing states of India comprise Haryana, M.P., Rajasthan and U.P. and this collectively represent 81 per cent of the national acreage and contribute 82.9 per cent to the total rapeseed-mustard production. It is basically a winter crop and is grown in the *rabi* season from September-October to February-March in Haryana.

Regression analysis is the most frequently used statistical technique for investigating and modeling the relationship between variables. Building a regression model is an iterative process. Usually, several analyses are required as improvement in the model structure and flaws in the data are discovered [1]. The use and interpretation of multiple linear regression models often depends on the estimates of individual regression coefficients. However, in some situations, the problem of multicollinearity exists when there are near linear dependencies between/among the regressors. Some applications of regression involve regressor and response variables that have a natural sequential order over time and then the need of time series (TS) modeling arises for the analysis of such dependence.

Time series data refers to observations on a variable that occurs in a time sequence. A basic assumption in any TS analysis is that some aspects of the past pattern will continue to remain in future. The most widely used technique for modeling and forecasting the TS data is Box-Jenkins' Autoregressive integrated moving average (ARIMA) methodology. However, when the patterns of the time-series under study are affected by some external events such as incorporation of new environmental regulations, special promotion campaigns, introduction of new variety, severe disease of plant etc. then the forecasting performance of ARIMA model may be affected. Under such situations, it can be improved by employing some appropriate techniques such as Transfer Function/ARIMA-Intervention analysis. ARIMA-Intervention analysis may be used to account for the effect of the intervention event(s) on the series but wherein the input series (apart from the main variable) will be in the form of a simple indicator variable to indicate the presence or absence of the event.

Pierce [2] discussed simultaneous least squares estimation of the regression and the time series parameters to treat the problem of correlated errors in regression and had shown that asymptotically the estimates obtained in this manner possess normal distributions, whether or not the errors themselves are normally distributed. Tsay [3] described time series regression models, in which regression equation errors accompany non-stationary or stationary moving average models. Bianco *et al.* [4] estimated the regression model's

***Correspondence to:** Raj Kumar, E-mail: rajk.apeco@mhu.ac.in; Tel: +91 9034688638

Citation: Kumar A, Kumar R. 2024. Simulation and modeling for mustard Yield (*Brassica juncea* L.) forecasting in Haryana using weather variables. *Res. Jr. Agril. Sci*. 15(3): 829-836.

parameters with ARIMA errors by minimizing a conveniently robustified likelihood function. Fisher and Planas [5] discussed about the large-scale fitting of regression model with ARIMA errors using economic time series data. Patowary *et al.* [6] studied regression with ARIMA errors to yearly production of wheat in India for the period of 1960-2016 and observed the fitted model as more accurate than autoregressive integrated moving average (ARIMA) model.

Aryani *et al.* [7] used two methods, ARIMAX and regression with ARIMA error, to forecast the influence of predictor variables on the profitability of Islamic banks. Box and Tiao [8] introduced intervention modeling to study and quantify the impact of air pollution controls on smog-producing oxidant levels in the Los Angeles area and economic control on the U.S. consumer price index. Wiorkowski and Heckard [9] applied intervention analysis to find the magnitude of the effect of fuel shortage and the max speed limit on all roads for the state of Texas. Girard [10] used ARIMA model with intervention in order to analyze the epidemiological situation of whooping-cough in England and Wales for the period of 1940-1990. Larson *et al.* [11] studied the effects of cotton defoliation and harvest timing on yield and efficiency, and consistently demonstrated the harmful effects of early crop termination. Prestemon [12] simulated the statistical power of univariate and bivariate methods of shock detection using time series intervention models and derived that bivariate methods are several times more statistically robust than univariate methods. Oyatoye and Fabson [13] compared the efficacy of simulation and time series models in quantifying the bullwhip effects in supply chain management. Chaudhuri and Dutta [14] applied different ARIMA models to identify the trends in the concentrations of few atmospheric pollutants and meteorological parameters over an urban station Kolkata, India and results revealed that the ARIMA (0, 2, 2) is the best statistical model for forecasting the daily concentration of pollutants as well as the meteorological parameters over Kolkata. Ray *et al.* [15] dealt with time series intervention modeling of cotton yield for Gujarat, Maharashtra and India as a whole. Ray *et al.* [16] applied time series intervention-based trend impact analysis for wheat yield scenario in India and developed a new trend impact analysis (TIA) approach.

MATERIALS AND METHODS

The Haryana state comprises 22 districts and situated between 74° 25' to 77° 38' E longitude and 27° 40' to 30° 55' N latitude. The total geographical area of the state is 44212 sq. km. The present study dealt with modeling the time-series yield of mustard crop in Hisar, Bhiwani, Sirsa and Fatehabad districts of Haryana. The state Department of Agriculture and Farmers Welfare mustard yield data compiled for the period 1980-81 to 2015-16 of Hisar, Bhiwani, Sirsa and 1997-98 to 2015-16 of Fatehabad districts were utilized for the purpose. The mustard yield data from 1980-81 to 2010-11 along with weather data (collected from IMD, Delhi and different meteorological stations in Haryana) of the same period were used for the training set. The weather-yield data of post-sample period, i.e., 2011-12 to 2015-16 have been used for validity testing of the developed mustard yield forecast models.

ARIMA-intervention modeling

The special kind of ARIMA model with input series is called an ARIMA-Intervention model or interrupted time series model. In an intervention model, the input series may be an indicator variable that contains discrete values that flag the occurrence of an event affecting the response series.

The ARIMA-Intervention model may also be written in a more parsimonious form in case the input series X_t is replaced by leading indicator I_t.

$$Y_t = a + \frac{\omega(B)}{\delta(B)} B^b I_t + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

Where;

Y_t = Dependent variable,

I_t = Indicator variables coded according to the type of intervention,

$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \ldots - \omega_s B^s$ (Impact parameter)

$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r$ (Slope parameter)

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ (AR parameter)

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ (MA parameter)

ε_t = white noise or error term

and r, s, p, q and b are constants. B is backshift operator and b is the delay parameter.

Regression with ARIMA errors

The general form of multiple regression is

$$Y_t = a + b_1 X_{1,t} + b_2 X_{2,t} + \ldots + b_k X_{k,t} + e_t \text{ ....} (1)$$

Where;

Y_t is modeled as a function of the k explanatory variables $X_{1,t}$,…, $X_{k,t}$. The key assumption is that the error term e_t is an uncorrelated series, that is, it is white noise. The general assumptions for the errors are as follows:

1. The errors have mean zero
2. The errors are uncorrelated with each other
3. The errors are uncorrelated with each predictor

If e_t contains autocorrelations, the ARIMA models may be combined to regression to handle the autocorrelations with the regression model to describe the explanatory relationship. The resulting model is a Regression with ARIMA Errors (RegARIMA) methodology propounded by Pankratz [17].

In equation (1), the autocorrelated errors e_t may be treated as:

$$(1-B)^d \phi(B)\, e_t = \theta(B)\, \varepsilon_t \text{ ....} (2)$$

Using equations (1) and (2), RegARIMA is finally expressed as:

$$Y_t = a + b_1 X_{1,t} + b_2 X_{2,t} + \ldots + b_k X_{k,t} + (\theta(B)\, \varepsilon_t / (1-B)^d \phi(B))$$

Simulation

Simulation is a method of solving decision making problems by designing, constructing and manipulating a model of real system. Simulation duplicates the essence of a system or activity without actually obtaining reality. Most simulations are random number driven. For each application of random numbers in a simulation, a distribution must be chosen. The distribution determines the likelihood of different values occurring. A distribution is uniquely specified by the name of its family (such as uniform, exponential, or normal etc.) and its parameter values (such as the mean and standard deviation). The key to modeling any random event is based on generating random numbers uniformly distributed between 0 and 1, denoted as U(0,1).

RESULTS AND DISCUSSION

ARIMA-intervention modeling

The ARIMA models with input series (ARIMA-Intervention model), is a generalization of distributed lag models and useful for capturing the contributions from lagged values of the predictor series. Several ARIMA models were tried with alternative combinations of weather variables to fit ARIMA-Intervention models. ARIMA (1,1,0) and ARIMA (0,1,1) for Bhiwani, Fatehabad, Hisar and Sirsa districts with fortnightly weather variables i.e., Tmx_3, Tmn_{10}, Arf_{12}, Tmx_9 and Arf_9 were fitted as ARIMA-Intervention models. Marquardt algorithm was used to minimize the sum of squared residuals and AIC & BIC guided to select the final ARIMA-Intervention models. The results achieved are presented in (Table 1-3).

Table 1 Parameter estimates of ARIMA-Intervention models for all the districts

District / Model Bhiwani	Parameter	Estimate	Standard error	t-Value	Pr> |t|
ARIMA (0,1,1) with Tmx_3	Constant	-10.20	6.21	-1.64	0.11
	MA	0.53	0.18	-2.93	0.01
	ω_0	0.66	0.37	1.78	0.09
	Tmx_3	-0.90	0.14	-6.32	<0.01
ARIMA (1,1,0) with Arf_9	Constant	0.07	0.34	0.21	0.83
	AR	-0.61	0.18	-3.34	<0.01
	ω_0	0.07	0.05	1.51	0.12
	Af_9	-0.83	0.26	-3.23	<0.01
		Fatehabad			
ARIMA (0,1,1) with Tmx_3	Constant	-20.72	9.36	-2.22	0.04
	MA	0.72	0.23	3.13	<0.01
	ω_0	1.21	0.50	2.41	0.02
	Tmx_3	-0.74	0.39	-1.89	0.07
ARIMA (1,1,0) with Arf_{12}	Constant	1.32	0.50	2.61	0.02
	AR	-0.40	0.21	-1.90	0.07
	ω_0	-0.18	0.06	-3.04	0.01
	Arf_{12}	-0.56	0.26	-2.13	0.04
		Hisar			
ARIMA (0,1,1) with Tmn_{10}	Constant	3.46	1.51	2.29	0.03
	MA	1.00	0.12	8.01	<0.01
	ω_0	-0.72	0.33	-2.20	0.04
	Tmn_{10}	-0.81	0.24	-3.45	<0.01
ARIMA (1,1,0) with Tmn_{10}	Constant	4.25	1.66	2.57	0.02
	AR	-0.47	0.19	-2.51	0.02
	ω_0	-0.86	0.35	-2.48	0.02
	Tmn_{10}	-0.81	0.19	-4.34	<0.01
		Sirsa			
ARIMA (1,1,0) with Tmn_{10} and Tmx_9	Constant	-8.58	4.59	-1.87	0.07
	AR	-0.67	0.17	-3.98	<0.01
	ω_0	-0.73	0.25	-2.87	0.01
	Tmn_{10}	-0.76	0.24	-3.22	<0.01
	Tmx_9	0.52	0.20	2.63	0.01
ARIMA (1,1,0) with Tmn_{10}	Constant	3.08	1.26	2.46	0.02
	AR	-0.70	0.16	-4.35	<0.01
	ω_0	-0.59	0.28	-2.12	0.04
	Tmn_{10}	-0.81	0.29	-2.77	0.01

Table 2 Selection criteria values for ARIMA-Intervention models

District(s)	Models	RMSE	AIC	BIC
Bhiwani	ARIMA (1,1,0) with Tmx_3	1.50	133.95	139.27
	ARIMA (1,1,0) with Arf_9	3.93	135.00	140.32
Fatehabad	ARIMA (0,1,1) with Tmx_3	2.18	125.79	130.82
	ARIMA (1,1,0) with Arf_{12}	6.68	128.31	133.34
Hisar	ARIMA (0,1,1) with Tmn_{10}	1.58	151.64	157.25
	ARIMA (1,1,0) with Tmn_{10}	4.31	154.99	160.60
Sirsa	ARIMA (1,1,0) with Tmn_{10} and Tmx_9	1.32	134.54	141.54
	ARIMA (1,1,0) with Tmn_{10}	3.56	139.88	145.48

Table 3 Percent relative deviations of post-sample mustard yield forecasts from real-time yield(s) based on ARIMA-Intervention models

District / Model	Forecast year	Observed yield (q/ha)	Fitted yield (q/ha)	Percent relative deviation
Bhiwani	2011-12	12.00	13.36	-11.36
	2012-13	16.40	15.28	6.84
ARIMA (0,1,1) with Arf_6	2013-14	15.16	14.89	1.80
	2014-15	13.98	15.87	-13.54
	2015-16	14.61	15.61	-6.85
	Av. Abs. percent dev.			8.08
Fatehabad	2011-12	18.66	18.92	-1.41
	2012-13	15.99	17.66	-10.42
ARIMA (0,1,1) with Tmx_3	2013-14	18.53	14.11	23.83
	2014-15	15.37	16.07	-4.57
	2015-16	13.55	12.63	6.80
	Av. Abs. percent dev.			9.41
Hisar	2011-12	17.07	17.02	0.31
	2012-13	16.78	14.86	11.46
ARIMA (0,1,1) with Tmn_{10}	2013-14	16.26	17.41	-7.05
	2014-15	14.17	12.59	11.15
	2015-16	18.16	15.93	12.26
	Av. Abs. percent dev.			8.45
Sirsa	2011-12	16.78	19.38	-15.49
	2012-13	16.47	16.81	-2.09
ARIMA (1,1,0) with Tmn_{10} and Tmx_9	2013-14	17.37	17.77	-2.28
	2014-15	15.00	14.42	3.87
	2015-16	17.09	18.28	-6.94
	Av. Abs. percent dev.			6.13

Regression with ARIMA errors modeling

The identification stage of regression with ARIMA errors required the checking of data stationarity followed by determination of tentative models with the help of acfs and pacfs plots. Several RegARIMA models along with fortnightly weather variables viz., Tmx_4, Tmx_3, Tmx_8, Arf_7, Arf_6 and Tmx_6 selected on the basis of stepwise regression method, were tried. Using the model selection criteria like AIC, BIC and root mean square error etc., RegARIMA with Tmx_4 and ARIMA (1,1,0) for Bhiwani and RegARIMA with Tmx_3, Arf_7 and ARIMA (0,1,1) for Fatehabad, Hisar and Sirsa districts were selected for pre-harvest mustard yield(s) estimation. The results obtained in this regard are shown in (Table 4-6).

Table 4 Parameter estimates of Regression with ARIMA Errors models for all the districts

District / Model bhiwani	Parameter	Estimate	Standard error	t-Value	Pr> |t|
RegARIMA with Tmx_4 and ARIMA (1,1,0)	Constant	-14.64	9.23	-1.59	0.13
	AR	-0.65	0.17	-3.90	<0.01
	Tmx_4	0.56	0.34	1.62	0.08
RegARIMAwith Tmx_8 and ARIMA (0,1,1)	Constant	-7.83	3.73	-2.10	0.05
	MA	1.00	0.20	5.08	<0.01
	Tmx_8	0.41	0.19	2.13	0.04
	Fatehabad				
RegARIMA with Tmx_3 and ARIMA (0,1,1)	Constant	-30.86	11.48	-2.69	0.01
	MA	1.00	0.24	4.25	<0.01
	Tmx_3	1.03	0.38	2.71	0.01
RegARIMA with Tmx_6 and ARIMA (1,1,0)	Constant	-6.26	4.16	-1.50	0.15
	AR	1.00	0.27	3.64	<0.01
	Tmx_6	0.31	0.19	1.59	0.13
	Hisar				
RegARIMAwith Arf_7 and ARIMA (0,1,1)	Constant	0.86	0.15	5.89	<0.01
	MA	1.00	0.17	5.86	<0.01
	Arf_7	-0.33	0.11	-2.90	0.01
RegARIMAwith Arf_6 and ARIMA (0,1,1)	Constant	0.20	0.08	2.38	0.02
	MA	1.00	0.15	6.71	<0.01
	Arf_6	0.23	0.10	2.27	0.03
	Sirsa				
RegARIMA with Arf_7 and ARIMA (0,1,1)	Constant	0.85	0.13	6.32	<0.01
	MA	1.00	0.19	5.32	<0.01
	Arf_7	-0.31	0.07	-4.24	<0.01
RegARIMA with Arf_6 and ARIMA (0,1,1)	Constant	0.27	0.07	4.04	<0.01
	MA	1.00	0.09	10.71	<0.01
	Arf_6	0.21	0.08	2.53	0.02

Table 5 Selection criteria values for Regression with ARIMA Errors models

District(s)	Models	RMSE	AIC	BIC
Bhiwani	RegARIMA with Tmx_4 and ARIMA (1,1,0)	1.43	133.30	137.29
	RegARIMA with Tmx_8 and ARIMA (0,1,1)	1.55	126.79	130.79
Fatehabad	RegARIMA with Tmx_3 and ARIMA (0,1,1)	1.78	117.05	120.71
	RegARIMA with Tmx_6 and ARIMA (1,1,0)	1.82	118.87	122.53
Hisar	RegARIMA with Arf_7 and ARIMA (0,1,1)	1.17	151.28	155.49
	RegARIMA with Arf_6 and ARIMA (0,1,1)	1.36	145.54	149.74
Sirsa	RegARIMA with Arf_7 and ARIMA (0,1,1)	0.45	129.45	133.66
	RegARIMA with Arf_6 and ARIMA (0,1,1)	1.28	139.55	143.76

Table 6 Post-sample mustard yield forecasts based on Regression with ARIMA Errors models for all the districts

District / Model	Forecast year	Observed yield (q/ha)	Fitted yield (q/ha)	Percent relative deviation
Bhiwani	2011-12	12.00	13.13	-9.38
RegARIMA with Tmx_4 and ARIMA	2012-13	16.40	13.89	15.33
(1,1,0)	2013-14	15.16	13.69	9.67
	2014-15	13.98	14.62	-4.57
	2015-16	14.61	14.82	-1.41
	Av. Abs. percent dev.			8.07
Fatehabad	2011-12	18.66	18.30	1.92
RegARIMA with Tmx_3 and ARIMA	2012-13	15.99	17.73	-10.91
(0,1,1)	2013-14	18.53	15.04	18.81
	2014-15	15.37	14.62	4.87
	2015-16	13.55	13.41	1.02
	Av. Abs. percent dev.			7.51
Hisar	2011-12	17.07	17.34	-1.60
RegARIMA with Arf_7 and ARIMA	2012-13	16.78	16.76	0.12
(0,1,1)	2013-14	16.26	17.62	-8.36
	2014-15	14.17	15.53	-9.60
	2015-16	18.16	16.39	9.75
	Av. Abs. percent dev.			5.89
Sirsa	2011-12	16.78	16.90	-0.73
RegARIMA with Arf_7 and ARIMA	2012-13	16.47	16.39	0.51
(0,1,1)	2013-14	17.37	17.24	0.76
	2014-15	15.00	15.29	-1.93
	2015-16	17.09	16.14	5.55
	Av. Abs. percent dev.			1.90

Comparison of the fitted models

Mustard yield forecasts for the post-sample years 2011-12, 2012-13, 2013-14, 2014-15 and 2015-16 have been obtained on the basis of Regression with ARIMA Errors and ARIMA-Intervention analyses. The performance(s) of the compending models were examined in terms of average absolute percent deviations and RMSEs of mustard yield forecasts in relation to real-time yield(s). Comparative view in terms of per cent relative deviations and root mean square errors has been presented in (Table 7).

Table 7 Comparative view in terms of average absolute percent deviations and root mean square error(s) of mustard yield forecasts with real time yield(s) for all the districts

	Average Absolute Percent Deviations		Root Mean Square Error(s)	
District(s)	Regression with ARIMA errors	ARIMA-Intervention model	Regression with ARIMA errors	ARIMA-Intervention model
Bhiwani	8.07	8.08	1.43	1.48
Fatehabad	7.51	9.41	1.78	2.18
Hisar	5.89	8.45	1.18	1.58
Sirsa	1.90	6.13	0.45	1.32

The overall results indicate the preference of using Regression with ARIMA Errors models over the ARIMA-Intervention models for obtaining mustard yield forecasts in the districts under study. RegARIMA models performed well in most of the time-regimes and consistently showed the superiority over compending models in capturing lower percent relative deviations pertaining to mustard yield forecasts in Haryana. This suggests that RegARIMA models offer more reliable and precise predictions, making them the preferred choice for agricultural yield forecasting in this context.

Regression diagnostics of the regression with ARIMA errors model

Any graph suitable for displaying the distribution of a set of data may be used for judging the normality of distribution of a group of residuals. Regression diagnostics of RegARIMA models for all the districts have been given in (Fig 1).

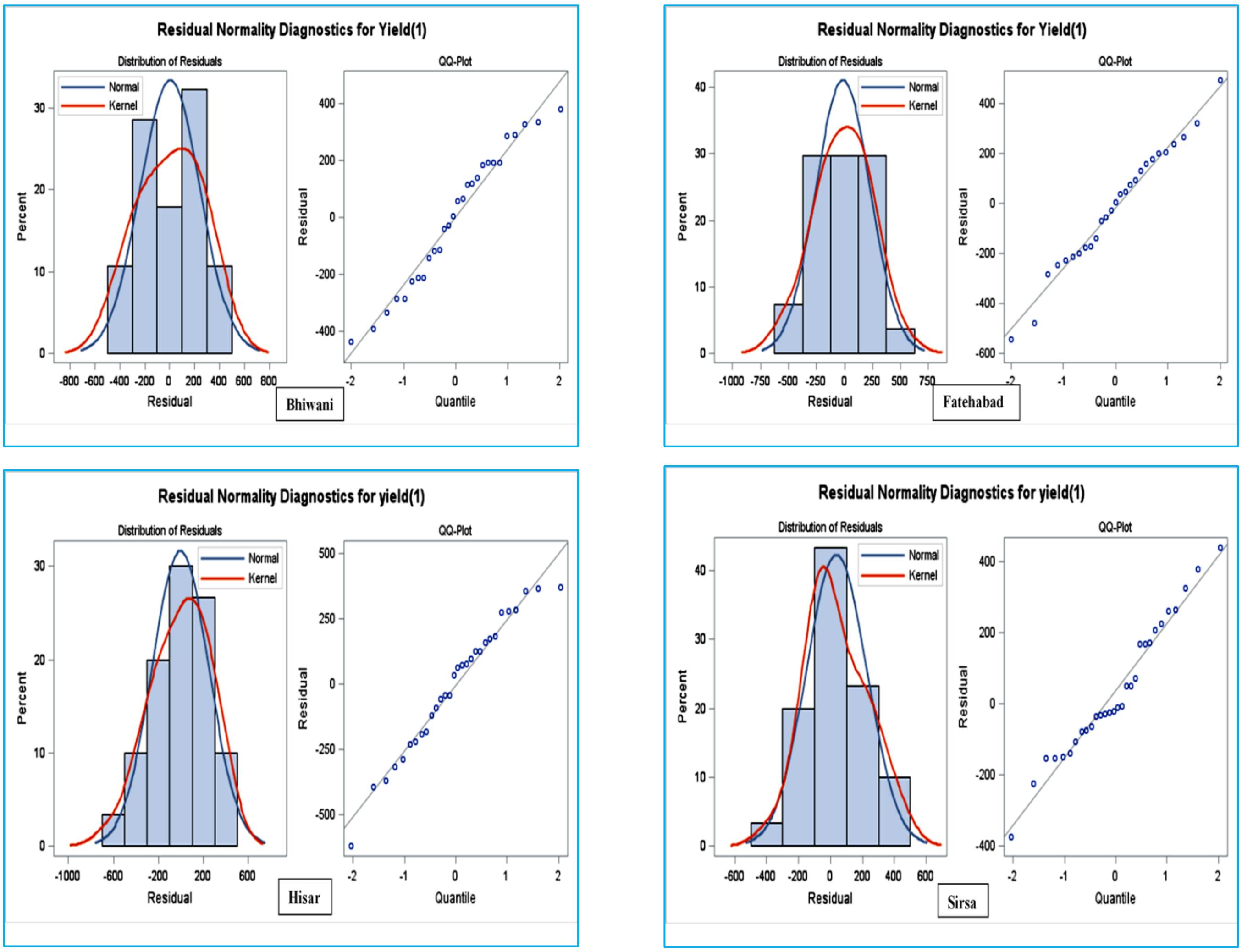

Fig 1 Regression diagnostics of the fitted RegARIMA models for all the districts

Simulation

Student's t copula procedure in SAS has been used to simulate the results achieved from Regression with ARIMA Errors. The copula function combined the marginal distributions of variables into a specific multivariate distribution. This approach enabled the modeling of dependencies between the variables, providing a more comprehensive and accurate simulation of the mustard yield forecasts in Haryana. By capturing the joint behavior of the variables, the robustness and reliability of the yield predictions generated by the RegARIMA models. The results pertaining to simulation for the post-sample period described in (Table 8-10).

Table 8 Post-sample mustard yield(s) along with simulated yield(s) and percent relative deviations based on regression with ARIMA Errors models for all the districts

District(s)	Forecast year	Observed yield (q/ha)	Simulated yield (q/ha)	Percent relative deviation
Bhiwani	2011-12	12.00	13.24	-10.37
	2012-13	16.40	13.71	16.41
	2013-14	15.16	13.82	8.87
	2014-15	13.98	13.99	-0.07
	2015-16	14.61	14.44	1.18
	Av. Abs. percent dev.			7.38
Fatehabad	2011-12	18.66	16.93	9.23
	2012-13	15.99	16.25	-1.63
	2013-14	18.53	16.40	11.47
	2014-15	15.37	16.04	-4.34
	2015-16	13.55	15.84	-16.91
	Av. Abs. percent dev.			8.72

Hisar	2011-12	17.07	17.04	0.18
	2012-13	16.78	15.19	9.49
	2013-14	16.26	15.35	5.59
	2014-15	14.17	15.51	-9.42
	2015-16	18.16	16.36	9.90
	Av. Abs. percent dev.			6.92
Sirsa	2011-12	16.78	15.43	8.02
	2012-13	16.47	16.42	0.33
	2013-14	17.37	16.07	7.46
	2014-15	15.00	15.47	-3.11
	2015-16	17.09	16.72	2.15
	Av. Abs. percent dev.			4.21

Table 9 Comparative view in terms of average absolute percent deviations of mustard yield forecasts based on fitted and simulated models

	Average absolute percent deviations		Root mean square error(s)	
District(s)	Regression with ARIMA errors models	Simulated regression with ARIMA errors models	Regression with ARIMA errors models	Simulated regression with ARIMA errors models
Bhiwani	8.07	7.38	1.43	1.46
Fatehabad	7.51	8.72	1.78	1.63
Hisar	5.89	6.92	1.18	1.29
Sirsa	1.90	4.21	0.45	0.88

Table 10 Post-sample mustard yield forecasts along with simulated yield(s) and percent relative deviations based on proposed regression with ARIMA Errors models for all the districts

District(s)	Forecast year	Fitted yield (q/ha)	Simulated yield (q/ha)	Percent relative deviation
Bhiwani	2011-12	13.13	13.24	-0.91
	2012-13	13.89	13.71	1.27
	2013-14	13.69	13.82	-0.89
	2014-15	14.62	13.99	4.30
	2015-16	14.82	14.44	2.55
	Av. Abs. percent dev.			1.98
Fatehabad	2011-12	18.30	16.94	7.45
	2012-13	17.73	16.25	8.37
	2013-14	15.04	16.40	-9.04
	2014-15	14.62	16.04	-9.68
	2015-16	13.41	15.84	-18.11
	Av. Abs. percent dev.			10.53
Hisar	2011-12	17.34	17.04	1.75
	2012-13	16.76	15.19	9.39
	2013-14	17.62	15.35	12.87
	2014-15	15.53	15.51	0.16
	2015-16	16.39	16.36	0.17
	Av. Abs. percent dev.			4.87
Sirsa	2011-12	16.90	15.43	8.69
	2012-13	16.39	16.42	-0.19
	2013-14	17.24	16.07	6.75
	2014-15	15.29	15.47	-1.16
	2015-16	16.14	16.72	-3.61
	Av. Abs. percent dev.			4.08

CONCLUSION

From the above results, it is inferred that the mustard yield forecasts based on simulated RegARIMA models are quite consistent to the yield forecasts obtained from fitted RegARIMA models pertaining to real-time yield(s) for all the districts under consideration. The forecasts obtained by RegARIMA are remarkably close to the forecasts obtained through the simulation process. Student's t-Copula has proved quite effective to model a diverse range of variations for the input parameters. By capturing the dependencies and joint behavior of the variables, thereby ensuring that the simulated forecasts align closely with the real-time forecasts. This demonstrates the utility of combining RegARIMA models with the Student's t-copula in achieving precise and dependable mustard yield predictions.

Acknowledgement and data availability

The authors would like to acknowledge the "Department of Economic and Statistical Analysis" and "Department of Agrometeorology, CCS, Haryana Agricultural University, Hisar" for providing online mustard yield data and weather data (maximum and minimum temperature, rainfall) respectively.

LITERATURE CITED

1. Draper NR, Smith H. 2003. *Applied Regression Analysis*. 3rd Edition, John Wiley & Sons, New York.
2. Pierce DA. 1971. Least squares estimation in the regression model with autoregressive moving average errors. *Biometrika* 58(2): 299-312.
3. Tsay RS. 1984. Regression models with Time Series errors. *Journal of the American Statistical Association* 79(385): 118-124.
4. Bianco AM, Ben MG, Martinez EJ, Yohai VJ. 1996. Robust procedures for regression models with ARIMA Errors. *A. Prat (Ed.).* COMPSTAT. pp 27-38.
5. Fischer B, Planas C. 2000. Large scale fitting of regression models with ARIMA errors. *Journal of Official Statistics* 16(2): 173-184.
6. Patowary AN, Bhuyan PC, Dutta MP, Hazarika J, Hazarika PJ. 2017. Development of a time series model to forecast wheat production in India. *Environment and Ecology* 35(4D): 3313-3318.
7. Aryani S, Aidi MN, Syafitri UD. 2018. Analysis of the profitability of Islamic banking using ARIMAX model and Regression with ARIMA errors model. *International Journal of Scientific Research in Science, Engineering and Technology* 4(8): 49-53.
8. Box GEP, Tiao GC. 1975. Intervention with applications analysis to economic and environmental problems. *Journal of the American Statistical Association* 70: 70-79.
9. Wiorkowski J, Heckard RF. 1977. The use of time series analysis and intervention analysis to assess the effects of external factors on traffic indices: A case study of the effects of the speed limit reduction and energy crisis in the state of Texas. *Accident Analysis and Prevention* 9(4): 229-247.
10. Girard DZ. 2000. Intervention time series analysis of pertussis vaccination in England and Wales. *Health Policy* 54: 13-25.
11. Larson JA, Gwathmey CO, Hayes RM. 2002. Cotton defoliation and harvest timing effects on yields, quality and net revenues. *Journal of Cotton Science* 6: 13-27.
12. Prestemon PJ. 2009. Statistical power of intervention analysis: Simulation and empirical application to treated lumber prices. *Forest Science* 55(1): 48-63.
13. Oyatoye OE, Fabson OVT. 2011. A Comparative study of simulation and time series model in quantifying bullwhip effect in supply chain. *Serbian Journal of Management* 6(2): 145-154.
14. Chaudhuri S, Dutta D. 2014. Mann-Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. *Environment Monitoring Assessment* 186: 4719-4742.
15. Ray M, Ramasubramanian V, Kumar A, Rai A. 2014. Application of time series intervention modelling for modelling and forecasting cotton yield. *Statistics and Applications* 12: 61-70.
16. Ray M, Rai A, Singha KN, Ramasubramanian V, Kumar A. 2017. Technology forecasting using time series intervention-based trend impact analysis for wheat yield scenario in India. *Technological Forecasting and Social Change* 118(C): 128-133.
17. Pankratz A. 1991. *Forecasting with Dynamic Regression Models*. John Wiley and Sons, New York.