International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319- 3972; ISSN(E): 2319-3980
Vol. 4, Issue 6, Oct -Nov 2015, 53-62

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

TIME-SERIES MODELLING FOR SUGARCANE YIELD ESTIMATION IN HARYANA

SANJEEV[1], U. VERMA[2] & M. TONK[3]

[1]PG Scholar, CCS Haryana Agricultural University, Hisar, India

[2,3]Department of Mathematics, Statistics and Physics, CCS Haryana Agricultural University, Hisar, India

ABSTRACT

An efficient crop forecasting infrastructure is pre-requisite for information system about food supply, especially export–import policies, procurement and price-fixation. The ARIMA models have been fitted using the time-series sugarcane yield data for the period 1966-67 to 2009-10 of Karnal and Ambala districts and 1972-73 to 2009-10 of Kurukshetra district. Models have been validated using the data on subsequent years i.e. 2010 to 2014, not included in the development of the models.. After experimenting with different lags of the moving average and autoregressive processes; ARIMA (0,1,1) for Karnal and Ambala, ARIMA (1,1,0) for Kurukshetra districts have been fitted for crop yield forecasting. A perusal of the results indicates that the percent deviations of the forecast yield(s) from the observed yield(s) are within acceptable limits and favours the use of ARIMA models to get short-term forecast estimates.

KEYWORDS: Autocorrelation, Partial Autocorrelation, Differencing, Stationarity, Invertibility

INTRODUCTION

Various statistical approaches like regression, time-series and stochastic models are in vogue for arriving at crop forecasts. Every approach has its own advantages and limitations. Regression analysis is the most frequently used statistical technique for investigating and modelling the relationship between variables. The widespread availability of computers and good softwares have contributed greatly to the expanding use of regression. Some applications of regression involve regressor and response variables that have a natural sequential order over time and then the need of time series modelling arises for the analysis of such dependence.

Time series models have advantages in certain situations. They can be used more easily for forecasting purposes because the historical sequences of observations upon study variables are readily available at equally spaced intervals over discrete point of time. These successive observations are statistically dependent and time series modelling is concerned with techniques for the analysis of such dependence. The application of the Box-Jenkins (1) univariate 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. may be of immense importance.

The theory and practice of time-series analysis and forecasting have developed rapidly over the last several years. An approach to the modelling of stationary and non-stationary time series is discussed by Box and Jenkins, building on the earlier work of several authors beginning with Yule (2) and Wold (3). The availability of powerful computers and a variety of readily available softwares resulted in an impetus in the development of forecast models using time-series procedures.

India is one of the largest sugarcane producers in the world, producing around 255.36 million tonnes of cane per annum (2012-2013). The area, production and productivity averaged over 2008-09 to 2013-14 were 4.71 million ha, 325.79 million tonnes and 69118 kg/ha respectively (**Source:** Agriharyana.nic.in). Sugarcane ranks third in the list of the most cultivated crops in India after paddy and wheat. Production of sugar is the second largest agro-processing industry in the country after cotton and textiles. Sugar industry has been a focal point for socio-economic development in rural areas by mobilizing rural resources, generating employment and higher income, transport and communication facilities. Around 60-65 per cent of total cane area in the country is in the sub-tropics, and this covers U.P, Bihar, Haryana and Punjab.

The sugarcane producing area of Haryana lies along the border of Uttar Pradesh. Its share to area and production has been 3.65 and 6.35 percent of the total area and production of the country (2012-13). Ambala, Karnal, Rohtak, Jind, Sonipat, Gurgaon, Kurukshetra and Hisar districts are major sugarcane producers in Haryana. Despite Marginal decrease in area of sugarcane during 2011-12 and 2012-13 (4.63 million hectares to 4.39 million hectares), the production has decreased from 17.96 million tonnes to 16.22 million tonnes. Keeping in view the importance of the subject matter, an attempt has been made to develop ARIMA models for sugarcane yield prediction in Karnal, Ambala and kurukshetra districts of the state.

Data Description

The Haryana state comprising of 21 districts is situated between 74^o 25'E to 77^o 38' E longitude and 27^o 40' N to 30^o 55' N latitude. The total geographical area of the state is 44212 sq. km. The present study dealt with modeling the time-series data related to the yield of sugarcane crop in Karnal, Ambala and kurukshetra districts of Haryana. The sugarcane yield data for the period 1966-67 to 2013-14 of Karnal and Ambala districts and 1972-73 to 2013-14 for Kurukshetra district were compiled from the Statistical Abstracts of Haryana/Punjab. The emphasis has been given in predicting a future value on the basis of previous time-series observations. The time-series yield data from 1966-67 (or 1972-73) to 2009-10 of sugarcane crop have been used for the training set and the remaining data i.e. 2010, 2011, 2012, 2013 and 2014 have been used for the post-sample validity checking of the developed ARIMA models.

Analysis Using the Box-Jenkins Method

Univariate Box-Jenkins(UBJ) ARIMA forecasts are based only on past values of the variable being forecast. The method applies to both discrete as well as to continuous data. However, the data should be available at equally spaced discrete time intervals. Before attempting to choose an appropriate ARIMA model for forecasting, it is necessary to make the data series stationary. One of the simplest transformations called 'differencing' is used when the mean of a series is changing over time and log transformation is used if the variance of a series is changing through time.

The estimated autocorrelation function and partial autocorrelation function are very important tools at the identification stage. An estimated autocorrelation function r_k shows the correlation between ordered pairs ($\overline{Y}_t, \overline{Y}_{t+k}$) separated by various time spans ($k=1,2,3,\ldots$). An estimated partial autocorrelation function $\hat{\phi}_{kk}$ shows the correlation between ordered pairs ($\overline{Y}_t, \overline{Y}_{t+k}$) separated by various time spans (k = 1, 2, 3…….) with the effect of intervening observations ($\overline{Y}_{t+1}, \overline{Y}_{t+2}, \ldots \overline{Y}_{t+k-1}$) accounted for.

The general functional form of ARIMA model used is :

Autoregressive Integrated Moving Average model i.e. *ARIMA (p,d,q)*

$\phi_p(B)\, \Delta^d\, Y_t = c' + \theta_q(B)\, a_t$, where c' = 0 if Y_t is adjusted for its mean 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

$(1-B)^d$ = non-seasonal difference

$\theta_q(B)$ = non-seasonal MA

ϕ's and θ's are the parameters need to be estimated

Further, at the estimation stage, it is attempted to obtain precise estimates of a small number of parameters of the model. Linear least-squares may be used to estimate only pure AR models. All other models require a non-linear least squares (NLS) method. Thirdly, the diagnostic tests are performed to see the random shocks to be independent or not.

RESULTS AND DISCUSSIONS

The UBJ methodology has been applied for sugarcane yield prediction in Haryana. UBJ- identification involves the determination of the appropriate orders of AR and MA polynomials i.e. the values of p and q. The orders were determined from the autocorrelation functions and partial autocorrelation functions of the stationary series. The graphical representation of sugarcane yield (q/ha) of Karnal, Ambala, and Kurukshetra districts in Figures 1 to 3 clearly indicates that the data series are non-stationary. Almost all the autocorrelations upto 10/12 lags significantly different from zero in Tables 1 to 3 confirm non-stationarity. Thus the series considered here were found to be non-stationary. Differencing of order one was enough for getting an appropriate stationary series (Figures 4 to 6).

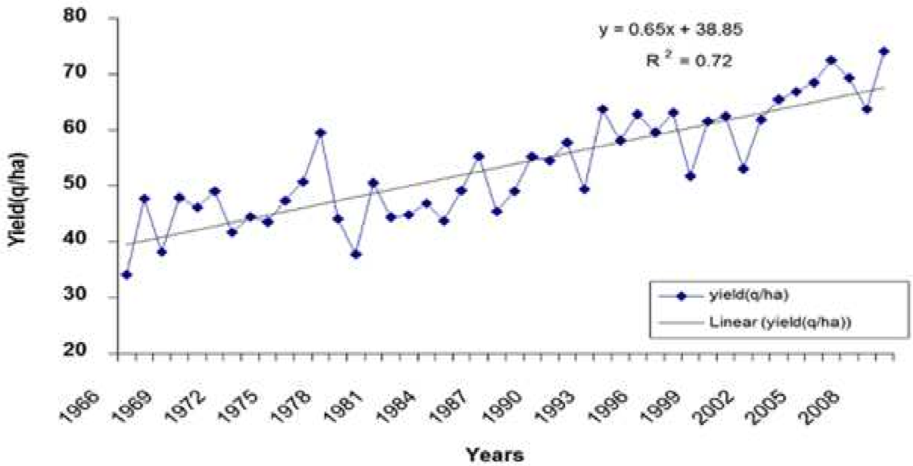


Figure 1: Annual Sugarcane Yield (q/ha) of Karnal District

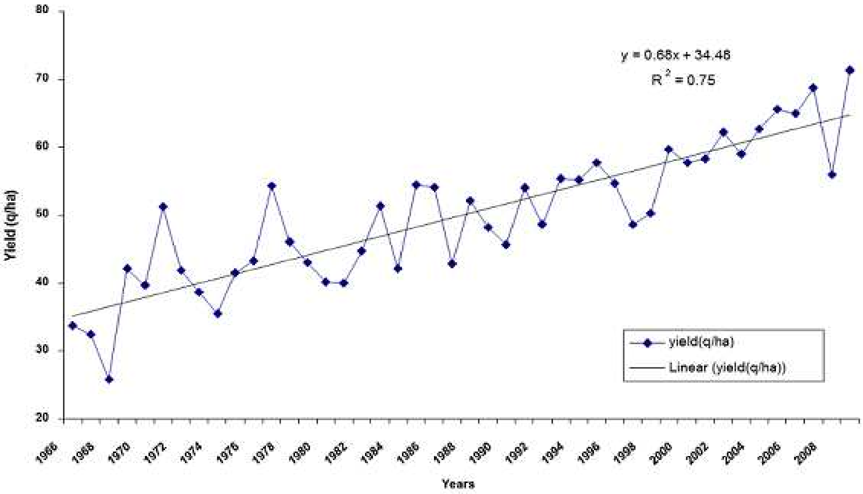


Figure 2: Annual Sugarcane Yield (q/ha) of Ambala District

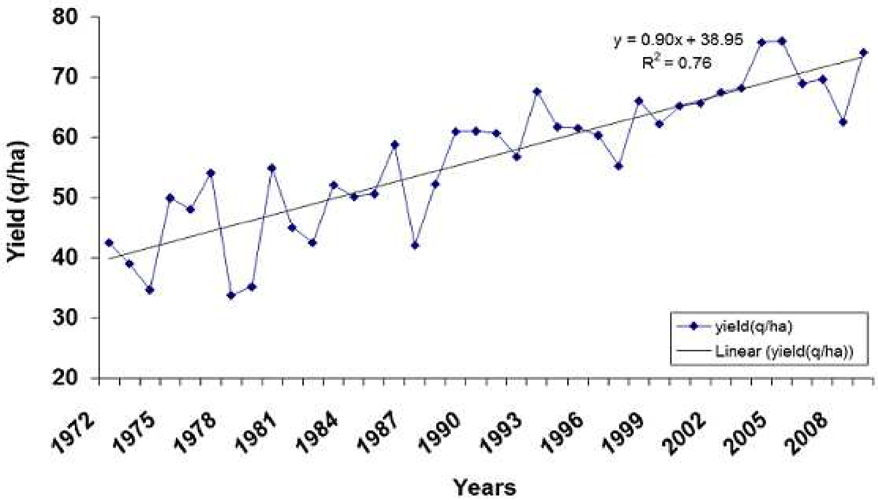


Figure 3: Annual Sugarcane Yield (q/ha) of Kurukshetra District

Table 1: Autocorrelations: Karnal Sugarcane Yield

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig.
1	0.77	0.14	31.26	1	.000
2	0.72	0.21	59.33	2	.000
3	0.65	0.25	82.37	3	.000
4	0.59	0.28	102.06	4	.000
5	0.51	0.31	117.07	5	.000
6	0.50	0.33	132.10	6	.000
7	0.45	0.34	144.17	7	.000
8	0.38	0.35	153.28	8	.000
9	0.34	0.36	160.70	9	.000
10	0.28	0.37	165.94	10	.000
11	0.23	0.37	169.45	11	.000
12	0.19	0.38	172.08	12	.000

Table 2: Autocorrelations: Ambala Sugarcane Yield

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig.
1	0.79	0.14	32.56	1	.000
2	0.69	0.21	58.32	2	.000
3	0.55	0.25	75.03	3	.000
4	0.48	0.28	88.23	4	.000
5	0.49	0.29	102.11	5	.000
6	0.51	0.31	117.78	6	.000
7	0.46	0.33	130.75	7	.000
8	0.42	0.34	141.86	8	.000
9	0.30	0.35	147.53	9	.000
10	0.26	0.36	152.07	10	.000
11	0.23	0.36	155.76	11	.000
12	0.20	0.37	158.69	12	.000

Table 3: Autocorrelations: Kurukshetra Sugarcane Yield

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	df	Sig.
1	0.76	0.15	27.13	1	.000
2	0.66	0.22	48.21	2	.000
3	0.60	0.26	66.15	3	.000
4	0.55	0.29	81.53	4	.000
5	0.56	0.32	97.84	5	.000
6	0.50	0.34	111.08	6	.000
7	0.43	0.36	121.30	7	.000
8	0.38	0.37	129.66	8	.000
9	0.39	0.38	138.58	9	.000
10	0.27	0.39	143.03	10	.000

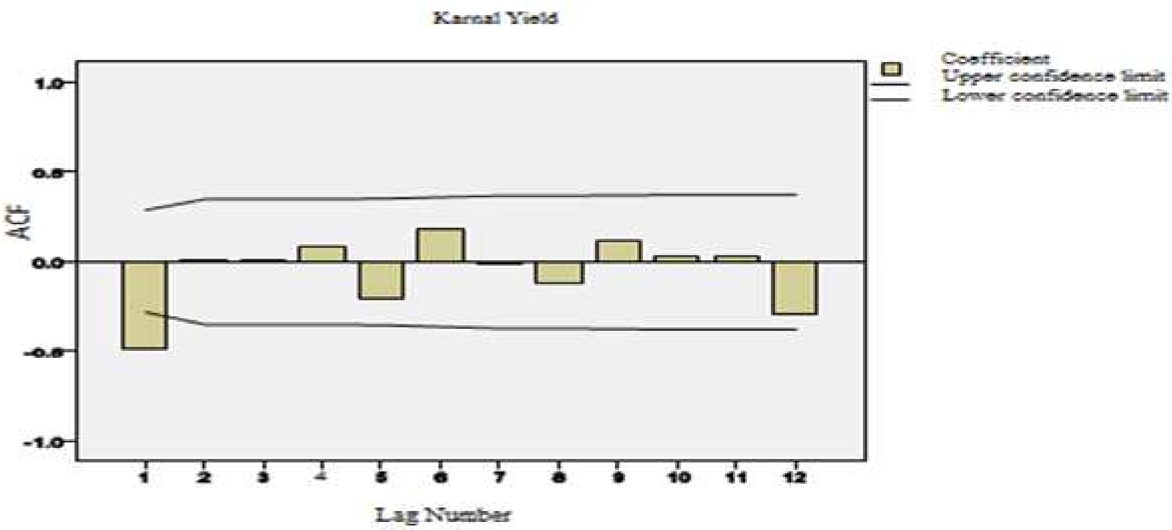


Figure 4: Autocorrelations: Karnal Sugarcane Yield Transformation: Difference(1)

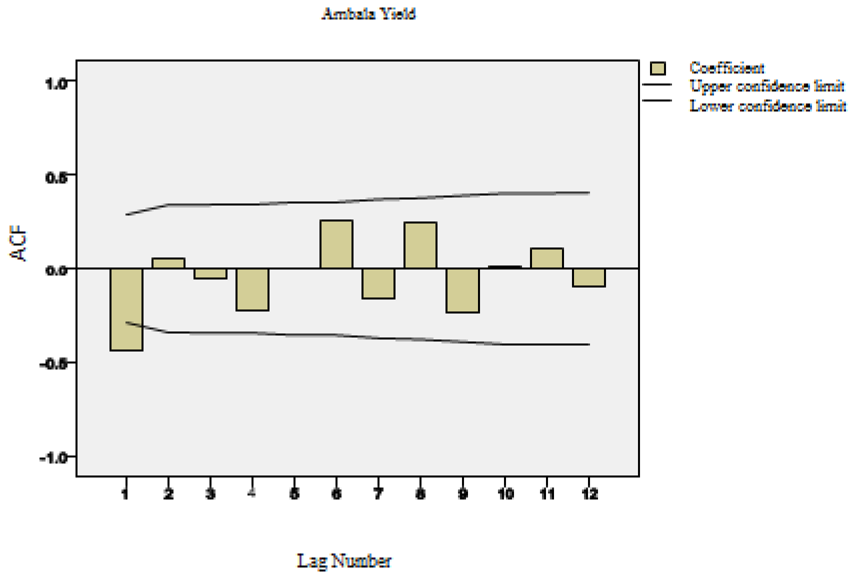


Figure 5: Autocorrelations: Ambala Sugarcane Yield Transformation: Difference (1)

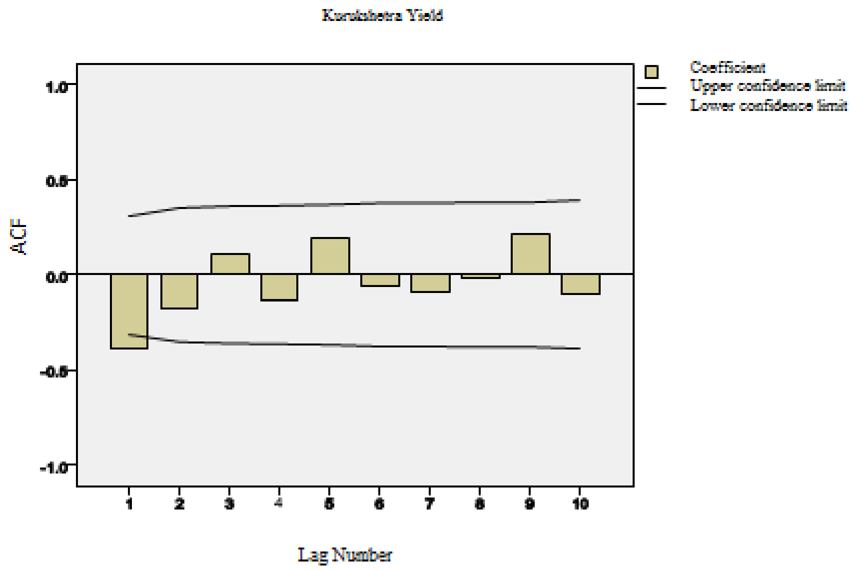


Figure 6: Autocorrelations: Kurukshetra Sugarcane Yield Transformation: Difference (1)

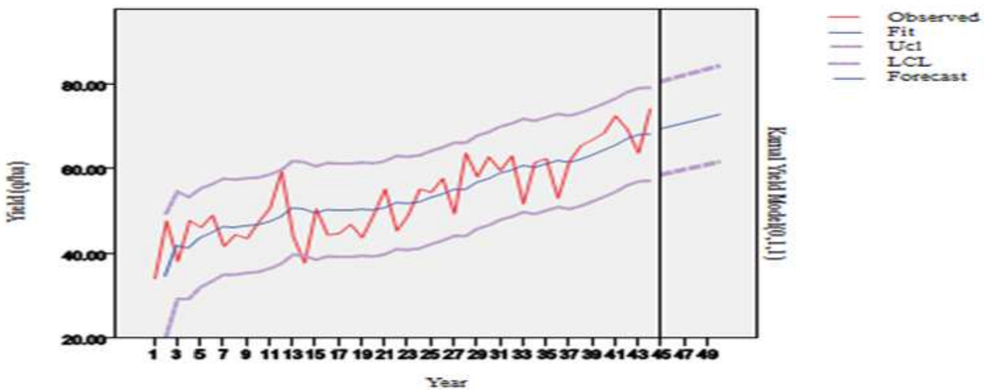
The models ARIMA (1,1,1), ARIMA (1,1,0) and ARIMA (0,1,1) were considered in the identification stage and parameter estimation was carried out using a non-linear least squares (NLS) approach. Marquardt algorithm (4) was used to minimize the sum of squared residuals. Log Likelihood, Akaike's Information Criterion, AIC (5), Schwarz's Bayesian Criterion, SBC (6) and residual variance decided the criteria to estimate AR and MA coefficients in the model. Approximate 't' values were calculated for residual autocorrelation coefficients using Bartlett's approximation for the standard error of the estimated autocorrelations. The residual acf along with the associated 't' tests and Chi-squared test suggested by Ljung and Box (7) were used for the checking of random shocks to be white noise. After experimenting with different lags of the moving average and autoregressive processes; ARIMA (0,1,1) for Karnal and Ambala and ARIMA (1,1,0) for Kurukshetra district/s were fitted for pre-harvest crop yield forecasting. Parameter estimates of the fitted ARIMA models are given in Table 4. All Chi-Squared statistic in this concern were calculated using the Ljung-Box formula as has been shown in Table 5. The observed, estimated and forecast yield(s) of Karnal, Ambala and Kurukshetra districts alongwith lower and upper confidence limits are depicted in Figure 7.

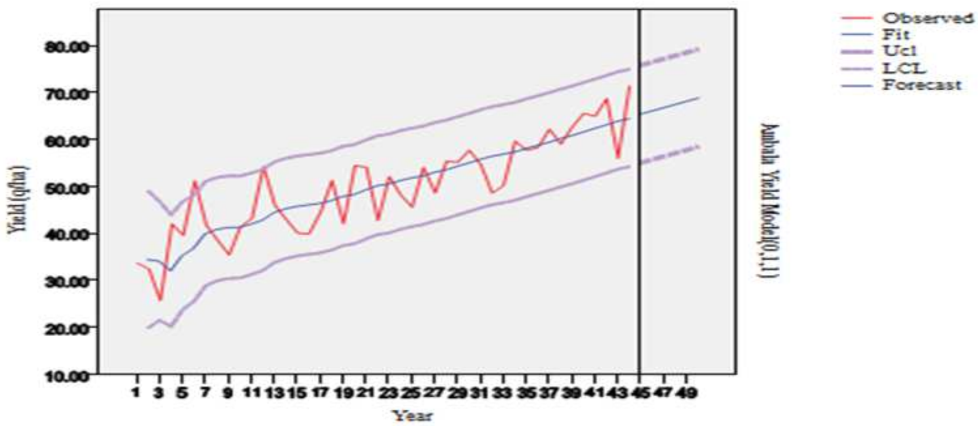
Table 4: Parameter Estimates of ARIMA Models for Sugarcane Yield (q/ha) of Karnal, Ambala and Kurukshetra Districts

District(s)	Model	Estimate	Standard error	t-ratio	Approx. prob
Karnal	**ARIMA(0,1,1);** MA(1)	0.74	0.11	6.68	0.00
Ambala	**ARIMA(0,1,1);** MA(1)	0.86	0.11	7.32	0.00
Kurukshetra	**ARIMA(1,1,0);** AR(1)	-0.41	0.14	-2.76	0.01

Table 5: Diagnostic Checking of Residual Autocorrelations of ARIMA Models Based on Sugarcane Yield of All The Districts

District(s)	Model	Model Fit Statistic				Ljung-Box Q Statistic		
		R-Squared	RMSE	MAPE	SBC	Statistic	df	Sig.
Karnal	ARIMA(0,1,1)	0.65	5.63	8.86	3.63	11.57	17	0.82
Ambala	ARIMA(0,1,1)	0.71	5.41	8.99	3.58	21.30	17	0.21
Kurukshetra	ARIMA(1,1,0)	0.70	6.27	9.38	4.07	14.39	16	0.56





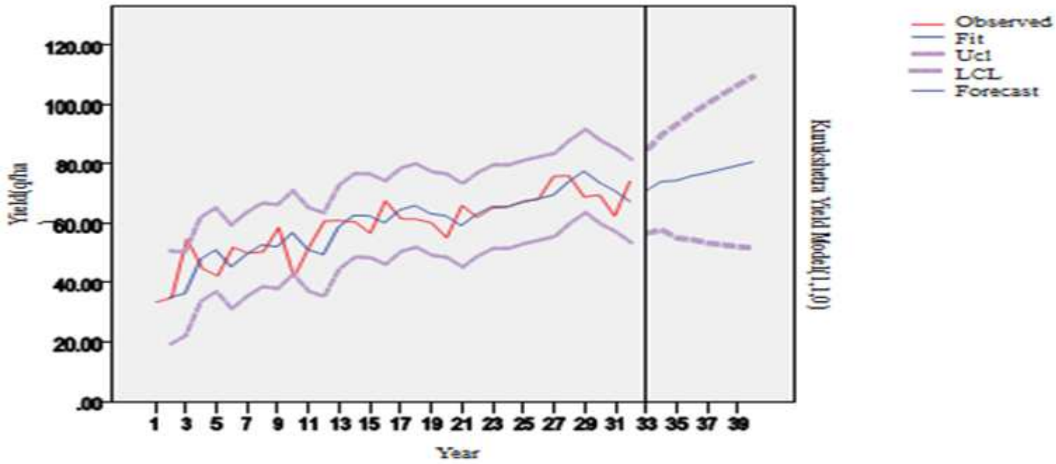


Figure 7: Observed, Estimated and Forecast Yield(s) of Karnal, Ambala and Kurukshetra Districts

Finally, a comparison between ARIMA based yield estimates with observed yield(s) was seen in terms of percent relative deviation (RD%). The results presented in Table 6 indicate that the deviation of predicted yield from the actual yield is very low, favouring the use of ARIMA models to get short-term forecast estimates.

Table 6: District-Specific Estimated Sugarcane Yield(s) (Est. Yield) Based on ARIMA Models and their Associated Percentage Deviations (RD%)=100×(Observed Yield - est. Yield)/ Observed Yield)

District/Model	Year	Observed Yield (q/ha)	Estimated Yield (q/ha)	Percent Relative Deviation
Karnal ARIMA(0,1,1)	2010	76.09	69.59	8.54
	2011	79.77	71.85	9.92
	2012	78.38	74.85	4.50
	2013	81.60	76.77	5.91
	2014	76.30	79.15	-3.73
Ambala ARIMA(0,1,1)	2010	67.22	65.40	2.70
	2011	71.58	66.25	7.44
	2012	79.67	67.40	15.40
	2013	77.70	69.15	11.00
	2014	67.60	73.14	-8.19
Kurukshetra ARIMA(1,1,0)	2010	69.93	70.63	-1.00
	2011	77.09	72.75	5.63
	2012	83.28	75.36	9.51
	2013	76.00	82.25	-8.22
	2014	72.90	80.24	-10.06

REFERENCES

1. Box GEP, Jenkins GM (1976) Time series analysis: Forecasting and control. Holden Day, San Franscisco : 575.
2. Yule GU (1927) On a method of investigating periodicities in disturbed series. Philosophical Transactions (A) 226 : 267-98.
3. Wold H (1938) A study in the Analysis of Stationary Time Series. Stockholm, Almquist and Wiksell.
4. Marquardt DW (1963) An algorithm for least-squares estimation of non-linear parameters. J Society for Industrial and Applied Mathematics 2 : 431-41.

5. Akaike H (1969) Fitting autoregressive models for prediction. Ann Inst Statist Math 21 : 243-47.
6. Schwarz G (1978) Estimating the dimension of a model. The annals of Statistics 62 : 461-64.
7. Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. Biometrika 65: 297-303.