

## ARIMA and State-Space models for sugarcane (*Saccharum officinarum*) yield forecasting in Northern agro-climatic zone of Haryana

### Ekta Hooda\*

Department of Mathematics & Statistics, CCS Haryana Agricultural University, Hisar-125004 (Haryana), India

### Urmil Verma

Department of Mathematics & Statistics, CCS Haryana Agricultural University, Hisar-125004 (Haryana), India

### B. K. Hooda

Department of Mathematics & Statistics, CCS Haryana Agricultural University, Hisar-125004 (Haryana), India

\*Corresponding author. E-mail: [ektahooda@gmail.com](mailto:ektahooda@gmail.com)

### Abstract

Advance estimates of significant cereal and commercial crops are given by the Directorate of Economics and Statistics and the central Ministry of Agriculture, Cooperation & Farmers' Welfare. However, the final estimates are released a few months after the actual harvest of the crops. In this study, ARIMA and State-Space models have been developed for sugarcane yield forecasting in Ambala and Karnal districts of Haryana. The above-mentioned models have been developed using yield data of sugarcane crop for the time period 1966-67 to 2009-10 of Ambala and Karnal districts. The validity of fitted models has been tested over the years 2010-11 to 2016-17. The forecasting performance of the developed models has been studied using percent deviations of sugarcane yield forecasts in relation to the actual yield, and root means squared errors. It has been observed that state-space models outperform the popular ARIMA models for forecasting of sugarcane yield in Northern Agro-climatic Zone of Haryana.

**Keywords:** ACF, ARIMA, PACF, State-Space models, Stationarity, Yield forecasting

### Article Info

<https://doi.org/10.31018/jans.v12i1.2229>

Received: February 16, 2020

Revised: March 8, 2020

Accepted: March 10, 2020

### How to Cite

Hooda, E. *et al.* (2020). ARIMA and State-Space models for sugarcane (*Saccharum officinarum*) yield forecasting in Northern agro-climatic zone of Haryana. *Journal of Applied and Natural Science*, 12(1): 53 - 58  
<https://doi.org/10.31018/jans.v12i1.2229>

## INTRODUCTION

India has a well-established system for collecting agricultural statistics, and forecasting of crop production is one of the most important aspects of agricultural statistics system. Advance estimates of major cereal and commercial crops are issued by the central government of India through the Ministry of Agriculture, Cooperation, and Farmers' Welfare. However, the final estimates are provided a few months after the actual harvest of the sugarcane crop. Accordingly, one of the impediments of the ordinary techniques is practicality and nature of the estimates. Consequently, there is always a considerable scope of progress in the regular system of estimation.

Forecasts can be obtained using various statistical approaches like regression, time-series, and stochastic models. 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. Time series modelling arises for the analysis of dependence when regressor and response variables have a natural sequential order over time. Time series models can be effectively utilized for predication purposes as the historical sequences of observations are promptly accessible at equally spaced intervals over discrete points of time. These successive observations are statistically dependent, and TS modelling is concerned with procedures for the analysis of such dependence. Autoregressive integrated moving average (ARIMA) models given by Box and Jenkins (1976) are of immense importance for forecasting a variety of variables in the field of agriculture.

The state-space models are frequently used to take into account the time dependency of the underlying parameters, which may further enhance the predictive accuracy of the most popularly used ARIMA models with parameter constancy. Expositions of the state space approach to multivariate forecasting can be found in Akaike (1976), Kitaga-

wa and Gersh (1984), and Durbin and Koopman (2002). A good account of state-space modelling is also given in the books by Aoki (1987) and Commandeur and Koopman (2007). Ravichandran and Prajneshu(2001) used Box-Jenkins ARIMA, and state-space approaches for modelling all-India marine products export data. Piepho and Ogutu (2007) studied the simple state-space models in a mixed model framework.

India, with an annual production of 350 million tonnes, is the second-largest sugarcane producer in the world after Brazil. Sugarcane ranks third in the list of most cultivated crops in India after paddy and wheat. Mwangi *et al.* (2017) proposed seasonal ARIMA models to forecast quarterly yields of sugarcane in Kenya based on yields data from 1973-2015. Assuming the level and trend components to be locally linear as well as when level and trend components remain constant without any persistent upward or downward drift, Hooda and Verma (2019) developed unobserved component models to study trend in sugarcane yield of five districts (Ambala, Karnal, Panipat, Yamunanagar and Kurukshetra) in Haryana. In this study, ARIMA and state-space models have been developed for sugarcane yield prediction in Karnal and Ambala districts of the state of Haryana. The models have been developed using yield data of sugarcane crop for the period 1966-67 to 2009-10 of Karnal and Ambala districts. The validity of fitted models have been tested for subsequent years, i.e., 2010-11 to 2016-17, not included in the development of the models.

**MATERIALS AND METHODS**

Haryana is one of the northernmost states in India and is adjacent to the national capital New Delhi. It is surrounded by Himachal Pradesh (HP) in the north, Rajasthan (RJ) in the south, Uttar Pradesh (UP) in the east, and Punjab (PB) in the west. As per haryanahighway.com, in spite of recent industrial development, Haryana remains an agricultural state primarily, with near about 70% of its residents directly or indirectly involved in the agriculture sector. Haryana is self-sufficient when it comes to food production and is the second largest contributor to India's central pool of food grains. It comprises of 22 districts with a total geographical area of 44,212 kms<sup>2</sup>.

The time-series data on sugarcane yield from 1966-67 to 2016-17 of Ambala and Karnal districts compiled from Statistical Abstracts of Haryana were used for the present study. The data for the last seven years, i.e., 2010-11 to 2016-17, have been used to check the validity of the developed ARIMA and state-space models for district-level sugarcane yield prediction in comparison to the actual yield obtained from state Department of Agriculture and Farmers' Welfare.

**ARIMA Model:** A stationary time series has

mean, variance, and Auto-Correlation Function essentially constant over time. Though prerequisite, the stationarity prerequisite for the applicability of the Box-Jenkins approach has all the earmarks of being quite restrictive. Notwithstanding, most non-stationary time series arising in practice can be transformed into stationary series through some basic operations. The Box-Jenkins methodology for developing an ARIMA model consists of the three stages, viz., identification, estimation, and the diagnostic checking stage.

The estimated ACF and Partial Auto-Correlation Function (PACF) are used at the model identification stage. These functions act as a guide for choosing one or more ARIMA models that seem to be appropriate for the given time series. At the second stage, we estimate the parameters of the model chosen at the identification stage. This stage additionally gives some warning signals in case the estimated coefficients don't fulfill certain mathematical inequality conditions. At the third stage, the residuals are used to test the independence of random shocks and to check the adequacy of the estimated model.

The general ARIMA(p,d,q)model for the time series  $Y_1, Y_2, Y_3, \dots$  may be expressed as  $\Phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t$  .....Eq. (1)

where,  $Y_t$  is the yield in the t-th year and error component  $a_t \sim WN(0, \sigma^2)$ ; B is the backward shift operator defined by  $BY_t = Y_t - Y_{t-1}$  and d is the order of differencing;  $\Phi_p(B)$  and  $\theta_q(B)$  are polynomials of order p, and q respectively defined as  $\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 \dots \dots \dots \Phi_p B^p$  and  $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \dots \dots \dots \theta_q B^q$

Here,  $\Phi_1, \Phi_2, \dots, \Phi_p$  are autoregressive coefficients, and  $\theta_1, \theta_2, \dots, \theta_q$  are the moving average coefficients. The first step in developing an ARIMA model is to examine whether the time series is stationary or non-stationary. The time series plots, ACF, and DickeyFuller test can be used to test the series for stationarity. A mean non-stationary series can be transformed into a stationary series by proper order of differencing, while natural logarithm may be used for transforming a variance non-stationary series. The appropriate values of p and q are determined by examining the ACF and PACF plots of the resulting stationary series.

The ARIMA(p,d,q) model contains p autoregressive and q moving average parameters, which can be estimated using least-squares or maximum likelihood methods. The diagnostic tests are performed to check if the residuals are independent or not. The residual ACFs for an appropriately built ARIMA model ought to have autocorrelation coefficients that are all statistically zero. As per Pankratz (1983), all residual ACFs ought to be zero; however, every one of them need not be essentially zero because of sampling error. For this purpose, the serial autocorrelations of residu-

als up to a specified lag  $k$  are generally tested using Ljung-Box Statistic. Ljung and Box (1978) suggested the following test statistic based on all the residual autocorrelations

$$Q = n(n+2) \sum_{j=1}^k (n-j)^{-1} r_j^2 \dots\dots\dots \text{Eq. (2)}$$

where,  $n$  is the total number of observations used to estimate the model,  $r_j$  autocorrelation at lag  $j$ , and  $k$  is the number of lags being used. The statistic  $Q$  approximately follows a Chi-squared distribution with  $(k-m)$  degrees of freedom, where  $k$  is the number of residual autocorrelation and  $m$  is the number of parameters estimated in the ARIMA model.

The accuracy of post-sample forecasts has been assessed using percent Relative Deviation (RD %) and Root Mean Square Error (RMSE).

**State-Space model** : The SS model comprises of an observation or measurement equation and a state or transition equation where the state equation details the dynamics of the state variables while the measurement equation relates the observed variables to the unobserved state vector. The SS model, formulated and described by Akaike (1976) have been used in the present study.

Let  $y_t$ :  $r \times 1$  be a vector of observed variables (after differencing if needed and subtracting the sample mean) and  $z_t$ :  $s \times 1$  ( $s \geq r$ ) be the state vector, where the first  $r$  components of  $z_t$  consist of  $y_t$  and the last  $s-r$  components are conditional predictions of future  $y_t$ . The state-space model formulated in terms of the state transition equation is

$$z_{t+1} = Fz_t + Ge_{t+1} \dots\dots \text{Eq.(3)}$$

where, the matrix  $F$ :  $s \times s$  is the transition matrix and the matrix  $G$ :  $s \times r$  is known as the input matrix.

For the purpose of model identification, the first  $r$  rows and  $r$  columns of  $G$  are set to an identity matrix of order  $r \times r$ . The vector  $e_t$  is a sequence of independently and normally distributed random vectors of dimension  $r$  with mean 0 and covariance matrix  $\Sigma_{ee}$ . Notwithstanding the state transition equation, SS models also consist of an observation equation that gives the observed values  $y_t$  as a function of the state vector  $z_t$ . The measurement equation used by the SAS STATE SPACE procedure in this study is

$$y_t = Hz_t \dots\dots\dots \text{Eq. (4)}$$

where,  $H = [I, 0]$  and  $I_r$  is an  $r \times r$  identity matrix and 0 is an  $r \times (s-r)$  null matrix.

The methodology used by the SAS STATE SPACE procedure also assumes the input series to be stationary. Therefore, the first step is to examine the data and test the requirement of differencing. The SAS STATE SPACE procedure employs a canonical correlation analysis for the identification of the state space model. The identification of the canonical SS model is practiced in two steps. The initial step involves the determination of the measure of past data to be utilized in the canonical correlation analysis. This is accomplished by fitting successively higher-order vector autoregressive (VAR) models and figuring Akaike information criterion (AIC) for each fitted model. The optimum lag ( $p$ ) into the past is chosen as the order of VAR model for which AIC is least.

The subsequent step includes the selection of state vector via means of canonical correlation analysis between the set of present and past values and the set of present and future values. The canonical correlation coefficients are computed for the sample covariance matrices of the set of successively increasing number of present and future values and the fixed set of present and past values. In the event that the smallest canonical correlation coefficient of the sample covariance matrix, that corresponds to the component being evaluated for inclusion in the state vector is non-zero. At that point, that specific component is included in the state vector. Once the state vector is resolved, the state space model is fitted to the data. The parameters in  $F$ ,  $G$  and  $\Sigma_{ee}$  are estimated utilizing maximum likelihood (ML) procedure.

The state-space forecasts are obtained through the Kalman filter (Harvey, 1989), which updates the knowledge of the system each time a new observation is brought and minimizes the error terms. The  $m$ -step ahead forecast of  $z_{t+m}$  i.e.  $z_{t+m/t}$  denotes the conditional expectation of  $z_{t+m/t}$  given the information available at time  $t$  i.e.  $y_{t+m/t} = Hz_{t+m/t}$ , where the matrix  $H = [I, 0]$ .

**RESULTS AND DISCUSSION**

The Box Jenkins' ARIMA methodology and the

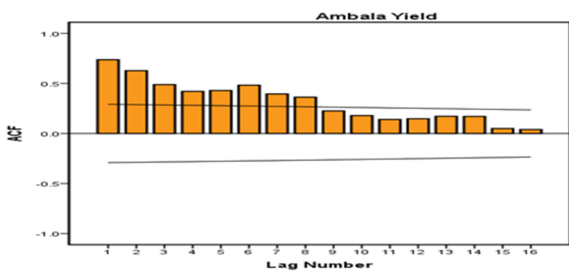


Fig. 1(a). ACF plot for sugarcane yield of Ambala district.

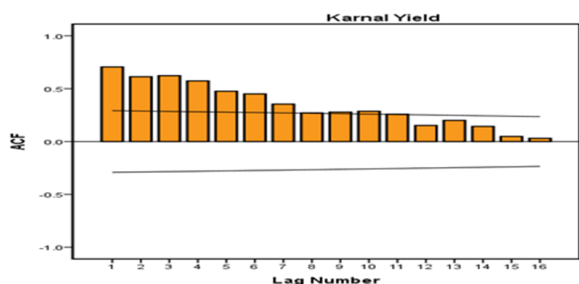


Fig. 1(b). ACF plot for sugarcane yield of Karnal district.

**Table 1.** Tentative ARIMA Models for sugarcane yield in Ambala and Karnal districts.

District (s)	Model	Parameter	Parameter Estimates	Standard error	t-value	p-value
Ambala	ARIMA(1,1,1)	$f_1$	0.168	0.195	0.857	0.396
		$\Theta_1$	0.824	0.122	6.760	< 0.001
	ARIMA(0,1,1)	$\Theta_1$	0.739	0.104	7.075	< 0.001
Karnal	ARIMA(1,1,0)	$f_1$	-0.411	0.132	-3.103	0.003
	ARIMA(1,1,1)	$f_1$	-0.022	0.179	-0.123	0.903
		$\Theta_1$	0.843	0.115	7.307	< 0.001
	ARIMA(0,1,1)	$\Theta_1$	0.852	0.095	8.952	< 0.001
	ARIMA(1,1,0)	$f_1$	-0.481	0.127	-3.779	< 0.001

**Table 2.** Selection criteria values for Fitted ARIMA models for Ambala and Karnal districts.

District (s)	Model	RMSE	MAPE	BIC
Ambala	ARIMA (1,1,1)	5.137	8.856	3.511
	ARIMA(0,1,1)	5.118	9.008	3.424
	ARIMA(1,1,0)	5.557	10.080	3.589
Karnal	ARIMA (1,1,1)	5.723	9.537	3.727
	ARIMA(0,1,1)	5.663	9.552	3.627
	ARIMA(1,1,0)	6.471	10.360	3.893

**Table 3.** Residual autocorrelations checking based on ARIMA Models for Ambala and Karnal districts.

District (s)	Model	Ljung-box Q statistic		
		Statistic	DF	p-value
Ambala	ARIMA (0,1,1)	24.108	17	0.117
Karnal	ARIMA (0,1,1)	14.234	17	0.650

**Table 4.** Akaike information criterion for Autoregressive models.

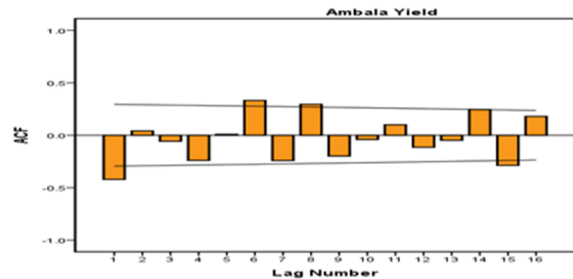
Lag/District(s)	Ambala	Karnal
0	158.33	165.97
1	151.95	157.79
2	152.75	153.60
3	153.98	153.04
4	148.41	154.84
5	140.37	155.29
6	142.23	157.24
7	142.49	158.92
8	144.43	160.54
9	146.37	162.51
10	148.21	164.37

**Table 5.** Yule-Walker estimates of selected Autoregressive models.

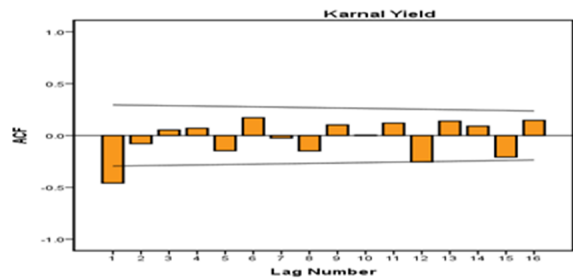
District (s)/Lag	1	2	3	4	5
Ambala	-0.75	-0.48	-0.49	-0.66	-0.46
Karnal	-0.71	-0.52	-0.24	-	-

state space methodology assume that the time series being modelled is stationary. The sugarcane yield data of both the districts were checked for stationarity using ACFs. The ACFs of sugarcane yield series decayed slowly, indicating the presence of non-stationarity for both the districts (Fig.1(a) and 1(b)). As indicated by the ACFs of the differenced series, the first order differencing was found sufficient for getting stationary series in both the districts (Fig. 2(a) and 2(b)).

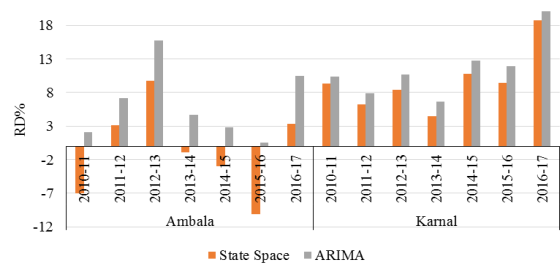
The orders of autoregressive (AR) and moving average (MA) components were determined through ACFs and PACFs of the stationary time



**Fig. 2(a).** ACF plots after differencing of order 1 for Ambala district.



**Fig. 2(b).** ACF plots after differencing of order 1 for Karnal district.



**Fig. 3.** Post sample (RD%) of forecasts based on ARIMA & SS models.

series. The sum of squared residuals was minimized utilizing the Marquardt algorithm (1963). The residual ACFs, along with the Chi-square test given by Ljung and Box (1978) were used to work out the random shocks as white noise. Consider-



**Table 6.** Parameter estimates of Stat- Space model.

District (s)	Parameter	Estimate	Std. Error	t-value
Ambala	F(5,1)	-0.52	0.16	-3.24
	F(5,2)	-0.81	0.17	-4.84
	F(5,3)	-0.13	0.23	-0.56
	F(5,4)	-0.34	0.23	-1.50
	F(5,5)	-0.82	0.24	-3.49
	G(2,1)	-0.74	0.15	-4.98
	G(3,1)	-0.01	0.18	-0.03
	G(4,1)	-0.18	0.17	-1.07
Karnal	G(5,1)	-0.07	0.15	-0.46
	F(2,1)	-0.16	0.21	-0.76
	F(2,2)	-0.16	0.26	-0.60
	G(2,1)	-0.75	0.15	-4.93

**Table 7.** Post-sample sugarcane yield forecasts based on ARIMA(0,1, 1) and State-Space models.

District (s)	Years	Actual Yield (q/ha)	ARIMA		State Space	
			Forecasted Yield (q/ha)	Relative Deviation (%)	Forecasted Yield (q/ha)	Relative Deviation (%)
Ambala	2010-11	67.22	65.78	2.14	71.92	-7.00
	2011-12	71.58	66.47	7.14	69.37	3.08
	2012-13	79.68	67.17	15.70	71.91	9.75
	2013-14	71.23	67.86	4.73	71.86	-0.88
	2014-15	70.55	68.56	2.82	72.67	-3.01
	2015-16	69.60	69.25	0.50	76.65	-10.13
Karnal	2016-17	78.13	69.95	10.47	75.53	3.33
	2010-11	79.77	71.52	10.34	72.34	9.32
	2011-12	78.38	72.20	7.88	73.49	6.23
	2012-13	81.60	72.87	10.70	74.78	8.36
	2013-14	78.81	73.55	6.67	75.26	4.51
	2014-15	85.04	74.22	12.72	75.85	10.81
	2015-16	84.54	74.90	11.92	76.54	9.46
	2016-17	95.00	75.57	20.45	77.21	18.73

**Table 8.** District-specific Av. Abs. percent RDs and RMSEs of post-sample sugarcane yield forecasts.

District (s)	Av. Abs. Percent RD		RMSE	
	ARIMA	SS	ARIMA	SS
Ambala	6.21	5.31	6.18	4.61
Karnal	11.52	9.63	10.74	9.28

ing various blends of moving average and autoregressive orders (Tables 1 and 2), ARIMA (0,1,1) model was found satisfactory for both the districts. The parameter estimates of fitted ARIMA models are presented in Table 1. Based on the estimates in Table-1, the model equations for Ambala and Karnal can be written as:  
 Ambala :  $(1-B)Y_t = (1-0.739B) a_t$  or  $Y_t = Y_{t-1} - 0.739 a_{t-1} + a_t$  .....Eq. (5)  
 Karnal :  $(1-B)Y_t = (1-0.825B) a_t$  or  $Y_t = Y_{t-1} - 0.825 a_{t-1} + a_t$  .....Eq. (6)  
 Where, B is the Backshift operator. The significance level of the moving average parameter  $\Theta_1$  has been found to be satisfactory (p-value is < 0.001) for both Ambala and Karnal districts. The ARIMA (0,1,1) model has only one moving average component and that is also less than one. Therefore, the developed models also satisfy the invertibility condition required for an ARIMA model. The Ljung and Box statistic based on 17 degrees of freedom is equal to 24.108 for Ambala and 14.034 for Karnal (Table-3). Both these val-

ues are non-significant, indicating the residuals to be white noise.

**State-Space modelling:** The 49 years sugarcane yield time series data for Ambala and Karnal districts were used for building the state-space models. The vector of observed variables  $y_t$  for Ambala and Karnal districts, after differencing and subtracting the mean from can be expressed as:  
 Ambala :  $y_t = (1-B)Y_t - 0.874$   
 Karnal :  $y_t = (1-B)Y_t - 0.659$   
 where B represents the backshift operator. Before the identification of SS model, it is important to decide the measure of past information to be utilized in the canonical correlation analysis. This is accomplished by fitting higher-order vector autoregressive models successively and processing AIC for each fitted model. Based on AIC values (Table-4), the autoregressive orders of five and three appeared to be appropriate for Ambala and Karnal districts, respectively. The least AIC values for respective autoregressive models provided the number of autocovariance matrices to be analyzed under the canonical correlation phase. Also, the Yule-Walker estimates obtained for the selected AR models are given in Table 5. Canonical correlation analysis selected the state vector after the autoregressive order selection process had determined the number of lags to be used under analysis. The preliminary estimate of the parameters of state-space models was formed

using the information from the canonical correlation and preliminary autoregression analyses (Table 6).

The state space models in fitted form for the two districts under study may be presented as follows:

**Ambala**

$$\begin{bmatrix} \mathcal{Y}_{t+1} \\ \mathcal{Y}_{t+2|t+1} \\ \mathcal{Y}_{t+3|t+2} \\ \mathcal{Y}_{t+4|t+3} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -0.52 & -0.81 & -0.13 & -0.34 & -0.82 \end{bmatrix} \begin{bmatrix} \mathcal{Y}_t \\ \mathcal{Y}_{t-1} \\ \mathcal{Y}_{t-2} \\ \mathcal{Y}_{t-3} \\ \mathcal{Y}_{t-4} \end{bmatrix} + \begin{bmatrix} 1 \\ -0.74 \\ -0.01 \\ -0.18 \\ -0.07 \end{bmatrix} [9.31]$$

**Karnal**

$$\begin{bmatrix} \mathcal{Y}_{t+1} \\ \mathcal{Y}_{t+2|t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.16 & -0.16 \end{bmatrix} \begin{bmatrix} \mathcal{Y}_t \\ \mathcal{Y}_{t+1|t} \end{bmatrix} + \begin{bmatrix} 1 \\ -0.75 \end{bmatrix} [30.55]$$

The forecast yield(s) along with the percent relative deviations for both the districts are presented in Table-7. Average absolute percent deviations and RMSEs of sugarcane yield forecasts in relation to the observed yield(s) from Dept of Agriculture and Farmers' Welfare were observed for checking the forecasting performance(s) of the contending models. The sugarcane yield prediction of the post sample period, i.e. 2010-11 to 2016-17 were obtained on the basis of fitted ARIMA and SS models to check the validity of the developed models. However, the average absolute percent deviations and the root mean square errors (RMSEs) of sugarcane yield forecasts based on both the models are depicted in Table 8. The average percent relative deviations and RMSEs are smaller for the SS models as compared to the selected ARIMA models for the districts indicating superiority of SS model over the ARIMA model. Fig. 3 shows the comparative view of post-sample sugarcane yield percent relative deviations based on ARIMA and state space models.

Sugarcane yield forecasting using time series techniques i.e. ARIMA and state space models provide far better results than time-trend based/linear mixed models applied on sugarcane crop of Haryana pertaining to percent relative deviations of the forecasts (Suman and Verma, 2018). ARIMA and state space models have also been observed superior over traditional regression analysis for obtaining advance estimates of the crop.

**Conclusion**

ARIMA and state-space models both provided suitable relationships to predict sugarcane yield in the districts under consideration. The forecasting performance(s) of the contending models were observed in terms of Av. Abs. Percent Deviations and RMSEs. The level of accuracy achieved by state-space model (s) was considered adequate, i.e., state-space models could better explain the sugarcane yield. Seven-

steps ahead estimated values (2010-11 to 2016-17) of sugarcane yield favour the use of SS models for sugarcane yield forecasting in the study region. The SS models performed well with lower error metrics as compared to the ARIMA models in all-time regimes. The developed models are fit for giving reliable estimates of sugarcane crop yield well ahead of time of the harvest collection. On the other hand, the DOAFA yield estimates are obtained quite late after the actual harvest of the crop.

**REFERENCES**

1. Akaike, H. (1976). Canonical correlations analysis of time series and the use of an information criterion in advances and case studies in system identification (R. Mehra and D.G. Lainiotis (Eds.)). Academic Press, New York.
2. Aoki, M. (1987). State Space modeling of time series. Springer, Berlin.
3. Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: Forecasting and control. *Holden Day, San Francisco*.
4. Commandeur, J.J.F. and Koopman, S. (2007) An introduction to state space time series analysis, *Oxford University Press, Oxford, USA*.
5. Durbin, J. and Koopman, S. J. (2002) A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89, 603-616.
6. Harvey, A.C. (1989) Forecasting, *Structural Time Series and the Kalman Filter, Cambridge University Press, Cambridge, UK*.
7. Hooda E. and Verma U. (2019). Unobserved components model for forecasting sugarcane yield in Haryana. *Journal of Applied and Natural Science*. 11 (3): 661-665. doi.org/10.31018/jans.v11i3.2144.
8. Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality. *J. Amer. Statist. Assoc.* 79: 378-389.
9. Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 6: 297-303.
10. Marquardt, D.W. (1963). An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Ind. Appl. Math.* 2 : 431-441.
11. Mwanga, D., Ongala, J. and Orwa, G. (2017) Modeling sugarcane yields in the Kenya sugar industry: A SARIMA model forecasting approach, *International Journal of Statistics and Applications*, ISSN: 2168-5193 e-ISSN: 2168-5215, 7(6), 280-288.
12. Pankratz, A. (1983) Forecasting with univariate Box-Jenkins models: concepts and cases, *John Wiley & Sons, New York, USA*.
13. Piepho, H.P. and Ogutu, J.O. (2007). Simple state-space models in a mixed model framework. *Amer. Statist.* 61 : 224 -232.
14. Ravichandran, S. and Prajneshu (2001). State space modelling versus ARIMA time series modelling, *Journal of the Indian Society of Agricultural Statistics*, 54 (1), 43-51.
15. Suman and Verma U. (2018). Linear mixed models for sugarcane yield estimation in Haryana (India). *Int. J. Agricultt. Stat. Sci.*, 14(1), 43-47.