Advances in Research

21(5): 6-13, 2020; Article no.AIR.56002
ISSN: 2348-0394, NLM ID: 101666096

Development of State Space Models with Weather as Exogenous Input for Sugarcane Yield Prediction in Haryana

Ekta Hooda[1*] and Urmil Verma[1]

[1]Department of Mathematics and Statistics, Chaudhary Charan Singh Haryana Agricultural University, Hisar - 125 004, India.

Authors' contributions

This work was carried out in collaboration between both authors. Author EH designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors EH and UV managed the analyses of the study. Author UV managed the literature searches. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AIR/2020/v21i530202
Editor(s):
(1) Dr. Paola Deligios, University of Sassari, Italy.
Reviewers:
(1) Teguh Sri Ngadono, University of Mercu Buana, Jakarta, Indonesia.
(2) E. H. Etuk, Rivers State University, Nigeria.
Complete Peer review History: http://www.sdiarticle4.com/review-history/56002

Original Research Article

Received 10 February 2020
Accepted 15 April 2020
Published 27 April 2020

ABSTRACT

Parameter constancy is a fundamental issue for empirical models to be useful for forecasting, analyzing or testing any theory. Unlike classical regression analysis, the state space models (SSM) are time varying parameters models as they allow for known changes in the structure of the system over time and provide a flexible class of dynamic and structural time series models. The work deals with the development of state space models with weather as exogenous input for sugarcane yield prediction in Ambala and Karnal districts of Haryana. The state space models with weather as exogenous input using different types of growth trends viz., polynomial splines; PS(1), PS(2) and PS(3) have been developed however PS(2) with weather input was selected as the best suited model for this empirical study. Timely and effective pre-harvest forecast of crop yield helps in advance planning, formulation and implementation of policies related to the crop procurement, price structure, distribution and import-export decisions etc. These forecasts are also useful to farmers to decide in advance their future prospects and course of action. The sugarcane yield forecasts based on state space models with weather input showed good agreement with state Department of Agriculture and Farmers' Welfare yield(s) by showing nearly 4 percent average absolute relative deviations in the two districts.

**Corresponding author: E-mail: ektahooda@gmail.com;*

Keywords: Time series; yield forecasting; state space models; SSM; sugarcane.

1. INTRODUCTION

Time Series (TS) modelling is a dynamic research area which has attracted attention of researchers over last few decades. The main aim of time series modelling is to study the past behaviour of a time series to develop an appropriate model which describes the inherent structure of the series. The model is then used to generate future values for the series, i.e. to make forecasts. There are different methods and techniques to analyze and forecast the TS data. One of the most frequently used methodology is based on autoregressive integrated moving average (ARIMA) model given by Box and Jenkins [1]. Pindyck and Rubinfeld [2] and Makridakis et al. [3] emphasized the use of ARIMA models for forecasting econometric and financial time series.

2. LITERATURE REVIEW AND METHODOLOGY

In ARIMA methodology, the model parameters are regarded as constant. However, in practice this assumption is frequently violated and one find many situations where, the model parameters are time varying. One such class is varying coefficient models, where the response variable is allowed to depend linearly on some regressors, with coefficients as smooth functions of some other predictor variables, called the effect modifier. Varying coefficient models, where the effect modifier variable is calendar time, leads to time-varying coefficient models. In case of time-dependent parameters, state space (SS) modelling using Kalman filtering technique may be used successfully. State space models are time varying parameters models as they allow for known changes in the structure of the system over time. However, the state space models have received relatively little attention from practitioners even though they permit the inclusion of explanatory variables and testing for lead, lag and feedback relationships among them. A state space model consists of a measurement or an observation equation and a state or transition equation where the state equation formulates the dynamics of the state variables while the measurement equation relates the observed variables to the unobserved state vector. This area of mathematical statistics is relevant to many areas of econometric research, as we often encounter unobserved variables that may be included in a model. In addition, this framework is also relevant to financial research in context of many variants of stochastic volatility models.

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 [4], Kitagawa and Gersh [5] and Durbin and Koopman [6]. A good account on state space modelling is also given in the books by Aoki [7] and Commandeur and Koopman [8]. Hooda and Thakur [9] carried out probability of distributions of drought, normal and abnormal events (months and years) for Solan district in relation to crop planning in Himachal Pradesh and observed that the CV decreases as the months tend to become wet and increases for months having sporadic rainfall.

Ravichandran and Prajneshu [10] applied Box-Jenkins' ARIMA and state space modelling approaches using Kalman filtering technique for analysing all-India marine products export data. Verma and Grover [11] worked on ARIMA methodology for modelling wheat yield in Haryana and comparison of ARIMA based forecasts was also made with remote sensing based yield forecasts and the real time crop yield data. Hooda [12] conducted a probability analysis of monthly rainfall for agricultural planning at Hisar district of Haryana using monthly rainfall data of 46 years and catergorized rainfall events as normal, abnormal and drought.

Mwanga et al. [13] proposed seasonal ARIMA models to forecast quarterly yields of sugarcane in Kenya based on yields data from 1973-2015. They found SARIMA $(2,1,2)\ (2,0,3)_4$ to be the best model for the quarterly sugarcane yield. 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 [14] have worked on unobserved component models to study sugarcane yield trend in Haryana.

Timely and effective pre-harvest forecast of crop yield helps in advance planning, formulation and implementation of policies related to the crop procurement, price structure, distribution and import-export decisions etc. These forecasts are also useful to farmers to decide in advance their

future prospects and course of action. Keeping in view the importance of the subject matter, the work has been carried out for sugarcane crop in Ambala and Karnal districts of Haryana. Though, the SS models with exogenous variables have not been used so far in Indian agricultural setup. The present study investigates the use of weather variables as exogenous input under state space framework for improvement of forecast accuracy achieved by simple SS models. Kalman Filtering and Smoothing have been used for parameter estimation as it provides the optimal estimates of the states. Inclusion of exogenous variables with state space formulations in agriculture may open a new era for agricultural commodity forecasting.

2.1 Data Description and State Space Modelling

Haryana is one amongst the northern states in India and is adjacent to national capital New Delhi. It is surrounded by Himachal Pradesh in the North, Rajasthan in the South, Uttar Pradesh in the East and Punjab in the West. Despite recent industrial development, Haryana is primarily an agricultural state, with nearly 70% of its residents directly or indirectly engaged in agriculture. Haryana comprises 22 districts with a total geographical area of 44,212 km^2. It is self-sufficient in food production and is the second largest contributor to India's central pool of food grains. Most of the sugarcane growing districts of Haryana are situated along the border of Uttar Pradesh.

The time series data on sugarcane yield from 1980-81 to 2016-17 of Ambala and Karnal districts have been compiled from Statistical Abstracts of Haryana however the seven years data *i.e.,* 2010-11 to 2016-17 have been used to check the validity of the developed models for district-level sugarcane yield prediction in comparison to real-time yield obtained from state Department of Agriculture and Farmers' Welfare. The daily weather data on maximum temperature, minimum temperature and rainfall for 37 years (i.e., 1980-81 to 2016-17) were obtained from India Meteorological Department (IMD), Delhi and different meteorological stations in Haryana. Temperature and rainfall are the important weather parameters influencing crop growth through different physiological processes. The fortnightly weather data were computed for minimum temperature, maximum temperature and rainfall, for inclusion in State Space modelling.

2.2 State Space Models with Exogenous Input

State Space models deal with dynamic time series which involve unobserved variables that describe the evolution in the state of the underlying system. SS models are time varying parameters models as they allow for known changes in the structure of the system over time [6]. These models are widely used in variety of fields such as econometrics, engineering and agriculture etc. The general State Space model with exogenous input includes non-stationary SSMs with time-varying system matrices and the state equations with a diffuse initial condition. It can be formulated as:

$$\boldsymbol{Y_t} = \boldsymbol{Z_t\alpha_t} + \boldsymbol{X_t\beta} + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon_t} \sim N(0, \sigma^2)$$

(Observation Equation)

$$\boldsymbol{\alpha_t} = \boldsymbol{T\alpha_{t-1}} + \boldsymbol{W_t\gamma} + \boldsymbol{\eta_t}, \boldsymbol{\eta_t} \sim N(0, \boldsymbol{Q})$$

(State Transition Equation)

$$\boldsymbol{\alpha_1} = \boldsymbol{\alpha}\,,$$

(Unknown Initial Condition)

The SSM is a dynamic version of the standard regression model in which the overall regression vector is divided into two parts; a time-invariant part **β** and a time-varyin g part $\boldsymbol{\alpha_t}$. The observation equation shows that the response value $\mathbf{Y_t}$ at time t is decomposed into three parts, where $\mathbf{X_t}\boldsymbol{\beta}$ and $\mathbf{Z_t}\boldsymbol{\alpha_t}$ are the contributions from the regression variables that are associated with the time-invariant and time-varying regression coefficients, respectively and $\boldsymbol{\varepsilon_t}$ is a value from a sequence of independent, zero mean, Gaussian noise variables. The time-varying part $\boldsymbol{\alpha_t}$ is called the state, which evolves in time as a first-order vector autoregression. The state transition equation, which describes the time evolution of $\boldsymbol{\alpha_t}$ shows that the new instance of $\boldsymbol{\alpha_t}$ is obtained by multiplying its previous instance $\boldsymbol{\alpha_{t-1}}$, by a square matrix **T**(called the state transition matrix) and by adding two more terms, i.e. a regression term $\boldsymbol{W_t\gamma}$ where $\boldsymbol{W_t}$ denotes the design matrix and $\boldsymbol{\gamma}$ is the state regression vector and a random disturbance vector $\boldsymbol{\eta_t}$. The state disturbance vectors $\boldsymbol{\eta_t}$ are assumed to be independent, zero mean, Gaussian random vectors with covariance **Q**. The state transition equation is initialized at t=1 with an unknown vector$\boldsymbol{\alpha}$. This type of initial condition is called a diffuse initial condition.

2.3 Polynomial Spline Trend

State Space procedure offers some models that govern the predefined trend components for different types of data. The most widely used trend types are Random Walk trend, Local Linear trend, ARIMA trend, Polynomial Spline trend and Decay & Growth trends. Amongst these, Polynomial Spline is the most commonly used trend. The polynomial spline trend is a general-purpose tool for extracting a smooth trend from the noisy data. A spline is a piece wise function that interpolates a set of knots or a function that goes through a set of points. Interpolation is a type of estimation and is often required to estimate the value of that function for an intermediate value. It is a form of interpolation where the interpolant is a special type of piecewise polynomial called 'spline'. Spline interpolation uses low-degree polynomials in each of the intervals and chooses the polynomial pieces such that they fit smoothly together.

The simplest spline has degree 0 and is called a step function. The next one is called a linear spline having degree 1. It represents a set of line segments between the two adjacent data points. The resulting curve is clearly not smooth, with sharp corners at the data points. The order of the spline rests on the order of the interpolating polynomial. The order-1 spline corresponds to Brownian motion (continuous-time random walk) and the order-2 spline corresponds to integrated Brownian motion (continuous-time random walk) and the order-3 spline provides a locally quadratic trend. The system matrices for the orders up to 3 are described as follows:

Order-1 spline: **Z** = (1), **T** = (1) and **Q** = $\sigma^2(h)$

Order-2 spline: **Z** = (1 0), **T** = (1 h, 0 1) and **Q** = $\sigma^2(\frac{h^3}{3}, \frac{h^2}{2}, \frac{h^2}{2}, h)$

Order-3 spline: **Z** = (1 0 0), **T** = (1 h, $\frac{h^2}{2}$, 0 1 h, 0 0 1),

and **Q**(i, j) = $\sigma^2 \frac{h^{6-i-j+1}}{(6-i-j+1)(3-i)!(3-j)!}$ 1≤i, j≤3

where, *h* denotes the difference between the successive time points.

The system matrices for higher orders are similarly defined by De Jong and Mazzi [15].

The PROC SSM procedure in SAS 9.4 has been used for developing the state space models with weather input and making post-sample prediction.

The Kalman Filter and Smoother (KFS) algorithm is the main computational tool for using PROC SSM for data analysis. For SSMs with a diffuse initial condition or when the regression variables are present, Diffuse Kalman Filter and Smoother (DKFS) is needed.

2.4 Goodness of Fit of the Developed Models

Reliability in numerical models are quantified and built by verification and validation. In the verification process, the developer's conceptual description of the model is properly judged in the implementation phase. The process of determining the degree, to which a model is an accurate representation of the real world, is known as validation. In other words, the process to collect evidence of a model's correctness or accuracy for a specific scenario is the validation testing. An evidence of sufficiently accurate model is provided by verification and validation. The forecasting performance of the introduced model is compared in relation to the state Department of Agriculture and Farmers' Welfare yield estimates using the following measure:

2.5 Relative Deviation (RD%)

It measures the deviation (in percentage) of yield forecasts from the actual yield data. The formula for calculating the percent relative deviation is:

$$RD\ (\%) = \left(\frac{Y_i - \hat{Y}_i}{Y_i}\right) x\ 100$$

Where Y_i and $\hat{Y}_i$ are the observed and predicted values and n is the number of years for which prediction has been done.

3. RESULTS AND DISCUSSION

The state space models with weather as exogenous input have been developed using the sugarcane yield data for the period 1980-81 to 2009-10. Fortnightly weather data on maximum temperature, minimum temperature and rainfall over the crop growth period have been utilized from 1980-81 to 2009-10 for building the state space models with weather input. The weather-yield data from 2010-11 to 2016-17 have been

Table 1. Parameter estimates of polynomial splines for Ambala district

Weather variables	PS(1)			PS(2)			PS(3)		
	Estimate	Std. error	t-value	Estimate	Std. error	t-value	Estimate	Std. error	t-value
TMX_9	0.304	0.408	0.75	0.322	0.375	0.86	0.371	0.385	0.96
TMX_{11}	0.188	0.393	0.48	0.302	0.363	0.83	0.243	0.379	0.64
TMX_{18}	-1.224	0.842	-1.45	-0.135	0.740	-1.83	-1.212	0.772	-1.57
TMN_{15}	0.539	1.293	0.42	0.924	1.155	0.80	0.667	1.222	0.55
TMN_6	0.555	0.712	0.78	0.770	0.589	1.31	0.530	0.690	0.77
TMN_{19}	0.639	0.451	1.42	0.523	0.427	1.22	0.524	0.431	1.21
ARF_4	-0.042	0.045	-0.93	-0.055	0.038	-1.46	-0.041	0.047	-0.86
ARF_{16}	0.006	0.011	0.53	0.007	0.009	0.72	0.008	0.011	0.80

Table 2. Parameter estimates of polynomial splines for Karnal district

Weather variables	PS(1)			PS(2)			PS(3)		
	Estimate	Std. error	t-value	Estimate	Std. error	t-value	Estimate	Std. error	t-value
TMX_{10}	0.302	0.419	0.72	0.329	0.389	0.85	0.415	0.384	1.08
TMX_{14}	-0.354	0.697	-0.51	-0.347	0.623	-0.56	-0.425	0.614	-0.69
TMX_{17}	0.285	1.151	0.25	0.269	1.053	0.26	0.315	1.033	0.30
TMN_8	-0.920	0.871	-1.06	-0.959	0.779	-1.23	-0.986	0.785	-1.26
TMN_{18}	0.210	0.861	0.24	0.133	0.792	0.17	0.121	0.769	0.16
ARF_7	-0.124	0.087	-1.43	-0.118	0.077	-1.53	-0.131	0.076	-1.17
ARF_8	0.126	0.089	1.42	0.114	0.076	1.50	0.135	0.075	1.78
ARF_{17}	0.023	0.023	0.99	0.023	0.021	1.15	0.024	0.020	1.21

used to check the post-sample validity of the fitted models for district-level sugarcane yield prediction.

3.1 Selection of Weather Variables for Ambala and Karnal Districts

The total growth period (2^{nd} fortnight of February to first fortnight of October) spread over 16 fortnights for three weather parameters i.e. average maximum temperature, average minimum temperature and accumulated rainfall turned up with 48 weather variables. Out of the 48 weather variables, eight variables were selected using stepwise regression for both the districts Draper and Smith, [16]. For Ambala district, three variables each from average maximum temperature (TMX_9, TMX_{11}, TMX_{18}) and average minimum temperature (TMN_6, TMN_{15}, TMN_{19}) and two from accumulated rainfall (ARF_4 and ARF_{16}) were found to have relatively better contribution towards yield. While for Karnal, three were from average maximum temperature (TMX_{10}, TMX_{14}, TMX_{17}), two from average minimum temperature (TMN_8 and TMN_{18}) and three from accumulated rainfall (ARF_7, ARF_8 and ARF_{17}). Here, TMX_i is the i^{th} day maximum temperature, TMN_j being the j^{th} day minimum temperature and ARF_k is the k^{th} day rainfall (i, j, k = 1, 2, 3,….,16 fortnights over crop growth period).

3.2 SSMs for Sugarcane Yield of Ambala and Karnal Districts

Growth trend models of polynomial spline; PS(1), PS(2) and PS(3) of orders 1, 2 and 3 respectively, along with selected weather variables were tried to get the best suited SSMs for both the districts. Parameter estimates of polynomial splines and the MLEs of alternative SSMs are shown in Tables 1-3 followed by the values of AIC, BIC and log likelihood in Table 4. for assessing the goodness of fit of the developed models.

The polynomial spline PS(2) model with weather input bearing the lowest AIC and BIC values, was found to be the best suited model for both Ambala and Karnal districts. The post-sample sugarcane yield prediction in this regard are shown in Table 5. The average absolute percent deviations from real-time yield data based on the selected models were found to be 4.78 and 4.74 for Ambala and Karnal districts.

Residual histogram and Quantile-Quantile plots for Ambala and Karnal districts were prepared for examining normality assumptions of the residuals (not shown in the text). Histogram shows approximate behaviour with slight deviation from normality but within acceptable range. The Q-Q plot infers the same as well. Standardized residual plot appears fine as all the residuals have magnitude within ±3σ limits and are scattered around zero. On the whole, these plots do not exhibit serious violations of the model assumptions.

The sugarcane yield prediction of the post-sample years 2009-10 to 2016-17 have been obtained on the basis of fitted SSMs with weather input. The predictive performance(s) of the models were observed in terms of percent relative deviations and RMSEs of sugarcane yield forecasts in relation to observed yield(s). The level of accuracy achieved by the SSMs with weather as exogenous input was considered adequate by capturing lower values of percent relative deviations from real-time yield. Comparative view on the district-specific yield forecasts along with percent relative deviations is shown in Table 5 followed by Fig. 1 reflecting the same view.

Table 3. Maximum likelihood estimates of unknown parameters of different SSMs for Ambala and Karnal districts

Component	Type	Parameter	Ambala		Karnal	
			Estimate	Std. error	Estimate	Std. error
Growth	PS(1) trend	Level Variance	5.658	3.899	7.020	4.343
White noise	Irregular	Variance	11.144	5.194	16.267	6.654
Growth	PS(2) trend	Level Variance	1.053E-8	.	0.037	0.124
White noise	Irregular	Variance	13.595	4.299	17.662	6.522
Growth	PS(3) trend	Level Variance	0.0004	0.003	0.005	0.008
White noise	Irregular	Variance	13.881	4.665	16.407	5.499

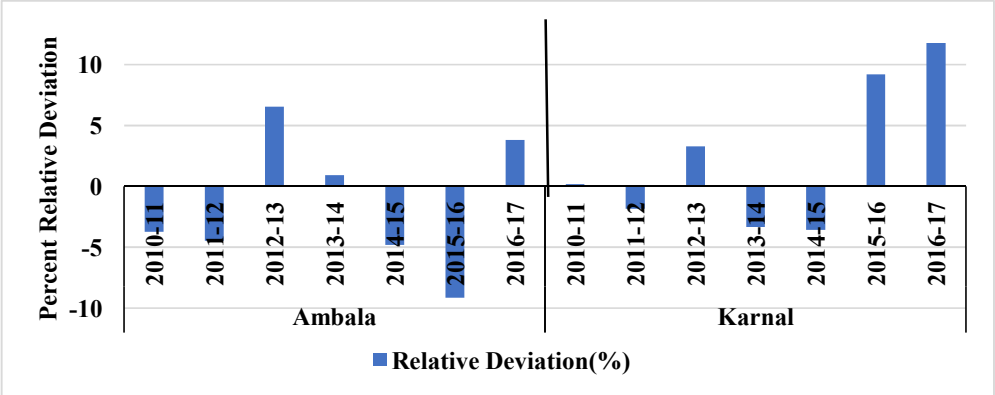

Fig. 1. Percent deviations of post-sample sugarcane yield(s) prediction from real-time yield(s) based on SS models with exogenous input

Table 4. Selection fit criteria of SSMs with weather input for sugarcane yield of Ambala and Karnal districts

Model fit statistic(s)		Ambala			Karnal		
		PS(1) with weather input	PS(2) with weather input	PS(3) with weather input	PS(1) with weather input	PS(2) with weather input	PS(3) with weather input
Diffuse	Log likelihood	-86.39	-82.60	-85.14	-91.97	-88.25	-89.66
	AIC	176.78	169.21	174.28	187.93	180.50	183.33
	BIC	178.87	171.20	176.17	190.02	182.49	185.22
Profile	Log likelihood	-84.24	-80.17	-84.14	-89.12	-85.28	-87.74
	AIC	190.47	184.33	194.86	200.23	194.56	201.47
	BIC	205.89	201.15	213.07	215.64	211.38	219.69

Table 5. Post-sample sugarcane yield prediction based on PS(2) model for Ambala and Karnal districts

Years	Ambala			Karnal		
	Actual yield (q/ha)	Predicted yield (q/ha)	Relative deviation (%)	Actual yield (q/ha)	Predicted yield (q/ha)	Relative deviation (%)
2010-11	67.22	69.73	-3.73	79.77	79.62	0.18
2011-12	71.58	74.79	-4.48	78.38	79.82	-1.84
2012-13	79.68	74.46	6.54	81.60	78.92	3.28
2013-14	71.23	70.58	0.90	78.81	81.44	-3.34
2014-15	70.55	73.94	-4.81	85.04	88.09	-3.59
2015-16	69.60	75.98	-9.17	84.54	76.76	9.21
2016-17	78.13	75.14	3.82	95.00	83.81	11.78
Av. Abs. Percent RD			4.78			4.74

4. CONCLUSION

A perusal of the results indicates that SSMs with weather input performed well with low error metrics in most of the time regimes. The sugarcane yield forecasts based on SSMs with weather input showed good agreement with state Department of Agriculture and Farmers' Welfare yield(s) by showing nearly 4 percent average absolute relative deviations in both the districts. Moreover, the developed models are capable of providing the reliable estimates of sugarcane yield well in advance of the crop harvest while the state department yield estimates are obtained quite late after the actual harvest of the crop.

Thus, the state space models may be effectively used pertaining to Indian agriculture data, as it takes into account the time dependency of the underlying parameters which may further enhance the predictive accuracy of the time-series models with parameter constancy. Though, it is emphasized that the selection of weather variables, VAR order, PS order etc. are quite sensitive to the model results. A proper care must be taken in identifying these features, otherwise the results may mislead the decision makers.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Box GEP. Jenkins GM. Time series analysis: Forecasting and control, holden day, San Franscisco; 1976.
2. Pindyck R, Rubinfeld D, Econometric models and economic forecasts (2nd ed.), New York: McGraw-Hill; 1981.
3. Makridakis S, Andersen A, Carbone R, Fildes R, Hibon M, Lewandowski R, Newton J, Parzen E, Winkler R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition, Journal of Forecasting. 1982; 1(2):111-153.
4. Akaike H. Canonical correlations analysis of time series and the use of an information criterion in advances and case studies in System Identification, Academic Press, New York; 1976.
5. Kitagawa G, Gersch W, A smoothness priors-state space modeling of time series with trend and seasonality. Journal of American Statistical Association. 1984;79: 378-389.
6. Durbin J, Koopman SJ, Time series analysis by state space methods, Oxford University Press, Oxford, USA; 2002.
7. Aoki M, State space modeling of time series, springer, Berlin; 1987.
8. Commandeur JJF, Koopman S. An introduction to state space time series analysis, Oxford University Press, Oxford, USA; 2007.
9. Hooda BK, Thakur BC. Probability analysis of rainfall at Nauni, Himachal Pradesh, Indian Journal of Soil Conservation. 1998;26(2):153-155.
10. Ravichandran S. Prajneshu state space modelling versus arima time series modeling. Journal of the Indian Society of Agricultural Statistics. 2001;54(1):43-51.
11. Verma U, Grover D. ARIMA wheat yield modelling in Haryana, Research Bulletin, Department of Mathematics and Statistics, CCS HAU, Hisar, Haryana. 2006;1:1-51.
12. Hooda BK. Probability analysis of monthly rainfall for agriculture planning at Hisar, Indian Journal of Soil Conservation. 2006; 34(1):12-14.
13. Mwanga D, Ongala J. Orwa G. Modeling sugarcane yields in the kenya sugar industry: A SARIMA Model Forecasting Approach, International Journal of Statistics and Applications. 2017;7(6):280-288.
    [ISSN: 2168-5193]
    [e-ISSN: 2168-5215]
14. Hooda E, Verma U, Unobserved components model for forecasting sugarcane yield in Haryana. Journal of Applied and Natural Science. 2013;11(3): 661–665.
15. De Jong P, Mazzi S. Modelling and smoothing unequally spaced sequence data, Statistical Inference for Stochastic Processes. 2001;4:53-71.
16. Draper NR, Smith H. Applied regression analysis, 3rd edition, John Wiley & Sons. 1981;736.
    [ISBN:0-471-17082]

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/56002