J. Indian Soc. Coastal agric. Res. 34(1): 110-119 (2016)

Developed Seasonal ARIMA Model to Forecast Streamflow for Savitri Basin in Konkan Region of Maharshtra on Daily Basis

K. D. GHARDE*, M. KOTHARI and D. M. MAHALE
College of Agricultural Engineering and Technology
Dr. B.S.K.K.V., Dapoli - 415 712, Ratnagiri, Maharashtra

Received: 12.12.2015 Accepted: 05.06.2016

Forecasting of streamflow is very useful for designing of soil and water conservation structures and planning of watershed activities, design of hydraulic structures, etc. Stochastic based model are found to be very reliable in estimation and forecasting streamflow on short and long term basis. In present study, streamflow on daily basis of Savitri basin were forecasted using seasonal ARIMA model developed in SPSS software following initial analysis of time series, identification model, apply diagnostic checks, adoption of model and forecast the stream flow by adopted model with checking the statistical indices such as correlation coefficient (R), root mean square errpr (RMSE), coefficient of efficiency (CE), volumetric error (EV), mean absolute deviation (MAD) and mean absolute percentage error (MAPE). The streamflow data for duration of 1992 to 2011 (20 years) measured at Mahad were used to calibrate, validate and forecast the streamflow by the ARIMA model. It is observed that, the data used in ARIMA model is stationary, no trends and no seasonality found in data sets on daily basis. The two models were identified as ARIMA $(111,111)^{31}$ and ARIMA $(202,000)^{31}$ as best model for forecasting the streamflow with R value more than 0.9 during calibration and forecasting period. The other statistical indices RMSE, CE, EV, MAD, MAPE were also found in appropriate range to forecaste the runoff with goodness of fit. Hence, developed ARIMA model could be adopted to forecast the runoff for sub tropical coastal region of Maharashtra on short and long term.

(***Key words:*** *Auto correlation, Calibration, Forecast, Partial Autocorrelation, Rainfall, Streamflow*)

Hydrologic models are simple especially widely used in understanding and quantifying the land use changes for rainfall runoff relationship and to provide the information for decision making in resources' management (Muthikrishana *et al*., 2006). The rainfall runoff modelling is very important in hydrology for estimation of water resources in river basin or catchment for meeting flood, potential assessment, its management and forecasting etc. There were several hydrologic models are available for prediction of runoff and sediment yield such as Physical based model (Soil and Water Assessment Tool, Arnolds *et al*., 1998), Black box models [(Artificial Neural Network model, ASCE 2000 (a & b)], Stochastic process models (Autoregressive Moving Average, (ARMA), Auto regressive integrated moving average model (ARIMA) etc, Box and Jenkins, 1970), Physical based parameter distributive models (SHE, Abbott *et al.* 1986), Kinematic waves models (Millere and Cunge, 1975), etc. But each model have its own limitation and advantages in applying at particular topographic and geographical locations such as available of data, processing pattern of data and output variability etc.

The seasonal ARIMA (SARIMA) model is statistical based stochastic process model analysed the time series data for forecasting of streamflow on long and short term. The time series are handling with mathematical models for predicting the new records and identified the changes in trends of hydrological records. The ARIMA have been used widely in river flow forecasting by many researchers (Davidson, *et al*., 2002; Gorman and Toman, 1966, Lall and Bosworth, 1993, Galeati, 1990) mostly because of its simplicity and ease of use. Time series analysis allows identification of hidden deterministic behaviour and thus understanding of cause and effect relationship in problems. Mahdi *et al.* (2013) and Nigam *et al.* (2013) applied the ARIMA model for forecasting the streamflow on daily basis. The univariate model provides an increased understanding of the behaviour of the system. The ARIMA widely adopted for forecasting the change in trends in climatic parameters, rainfall and streamflow for small to large catchments and proven to be very good fitted hydrologic stochastic process model. By considering this fact the present study was undertaken to develop the seasonal ARIMA model for sub tropical coastal Konkan region of Maharashtra and checks it significance with statistical parameters in forecasting stream flow for Savitri Basin.

**Corresponding Author :* E-mail: aryanavipsha2010@gmail.com

MATERIALS AND METHODS

Study area

The present study was conducted for the Savitri river basin which comes under the Western part of Sahayandri Ghat of Konkan region and located in Maharshtra State of India (Fig.1). The latitude and longitude of the study area is $18^0 20'$N to $17^0 51'$N and $73^0 22'$ E to $73^0 41'$E respectively and elevation ranges from 6.50 m to 1366.23 m above mean sea level**.** The streamflow and rainfall measured at Birwadi, Kangule, Bhave, Kotheri, etc were collected from superintending Engineer, Unit of Hydrology Project, Nashik. The mean aerial average daily rainfall and streamflow for duration of 20 years (1992 to 2011) of all stations were considered. The statistical parameters of inputs data for models such as Mean, Maximum, Minimum, Standard deviation (SDV), auto correlation function (ACF) and partial auto correlation function (PACF) coefficient estimated using standard procedure (Table 1).

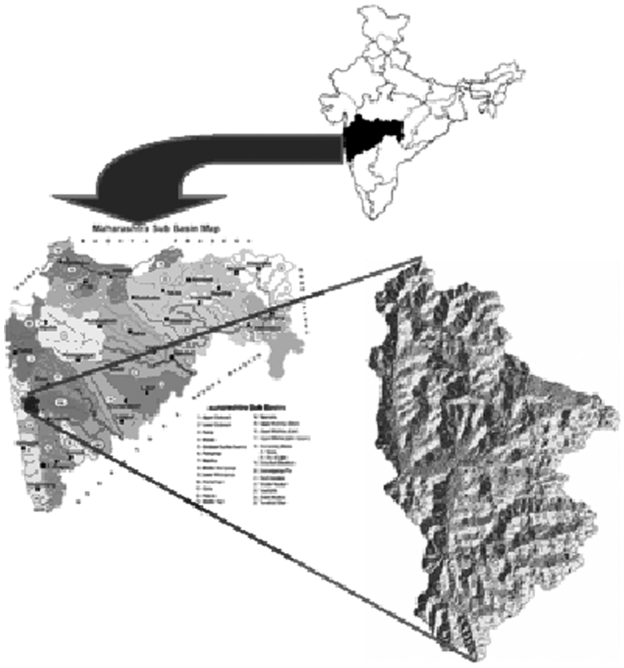

Fig. 1. Location of Savitri Basin

ARIMA model theory

The multiplicative ARIMA model, as a short term, stands for Autoregressive Integrated Moving Average. The acronyms AR (p) is known for an Autoregressive model of order (p), and represented by:

$$x_t = \sum_{j=1}^{p} \Phi_j x_{t-1} + \varepsilon_t \qquad ...1$$

Where,

x_t = observation at time=t,

Φ_j = j^{th} autoregressive parameter.

ε_t = independent random variable represent the error term at time t,

t-1 = time series at the time (t-1),

p = order of autoregressive process.

The acronyms MA (q) is known for a moving average model of the order q and is represented by:

$$x_t = \varepsilon_t - \sum_{j=0}^{q} \Theta_j x_{t-1} \qquad ...2$$

Θ_j = j^{th} moving average parameter,

q = order of moving average process.

The combination between AR (p) and MA (q) models is called the ARMA (p, q) model, and is represented by:

$$x_t = \sum_{j=1}^{p} \Phi_j x_{t-1} - \sum_{j=0}^{q} \Theta_j x_{t-1} \qquad ...3$$

To achieve stationary case in the time series, it may be differenced ARMA model for d times to obtain ARIMA (p, d, q), similarly an ARMA model may be seasonal differenced for D times to obtain seasonal ARIMA (P, D, Q)S for S seasonal period. So when they are coupled together that will give ARIMA (p, d, q) x (P, D, Q)S. The regular difference is written as

$$(1 - B)^d x_t \qquad ...4$$

Where,

B = backward operator,

d = the non-seasonal order of differences.

The seasonal difference of order D with period is written as S

$$(1 - B^S)^D x_t \qquad ...5$$

In general, the differencing operation may be done several times but in practice only one or two differencing operation are used (Salas *et al.,* 1980). Box and Jenkins (1974), generalized the above model and obtained the multiplicative Autoregressive Integrated Moving Average where the general form is {seasonal ARIMA (p, d, q) x (P, D, Q)s } which is written as:

$$(1 - \Phi_1 B^S - \Phi B^{2S} - .... - \Phi_p B^{pS})(1 - B^S) x_t \qquad ...6$$
$$= (1 - \Theta_1 B^S - \Theta_2 B^{2S} - .... - \Theta_q B^{qS}) \alpha_t$$

Table 1. *Statistical parameters of input datasets for Svaitri basin*

Sr No	Data	No of sample	Max	Min	Mean	SD	ACF	PACF	SE
1	Rainfall, mm	2483	366.9	0.1	28.14	36.56	0.0093	0.0145	0.020
2	Streamflow,cumecs	2383	4932.3	0.01	395.1	532.31	0.0146	0.0045	0.019

The residuals α_t are in turn is represented by an ARIMA (p, d, q) model,

$$(1 - \Phi_1 B - \Phi_2 B^2 - .... - \Phi_p B^p)(1 - B)^d \alpha_t = (1 - \Theta_1 B - \Theta_2 B^S - .... - \Theta_q B^q)\, \varepsilon_t \quad ...7$$

The general multiplicative ARIMA (p, d, q) x (P,D,Q)S model is obtained by solving Eq.6 for α_t and replacing in Eq.7 as get Eq 8

$$\Phi(B^S)\, \Phi(B)\, (1 - B^S)^D\, (1 - B)^d\, x_t = \Theta(B^S)\, \Theta(B)\, \varepsilon_t \quad ...8$$

ARIMA Model Methodology

The ARIMA model was developed in SPSS software following the procedure suggested by Box and Jenkin (1976) and representative steps for development of model is given in flow chart (Fig. 2). The methodology consists of five steps: 1) Initial analysis of the input data, 2) Model identification. 3) Estimation of model parameters. 4) Diagnostic checking for the identified model appropriateness for modeling and 5) Adoption of the model (*i.e.* forecasting).

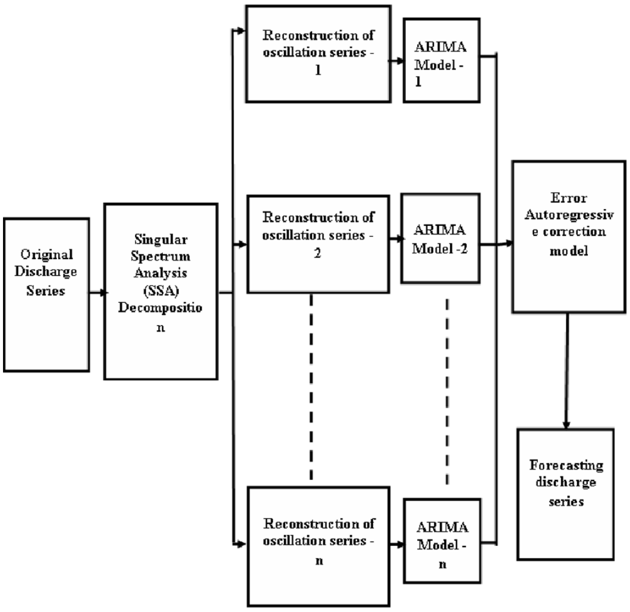

Fig. 2. *Flow diagram of ARIMA model execution*

Initial analysis of the data

The daily stream flow time series data used of hydrologic station Mahad, Tal Mahad for duration of 20 years (1992 to 2011) for Savitri basins. The collected data were checked for seasonal variation by plotting ACF and PACF function and model were identified. The Box- Cox transformation is used for making data to be static. The Turkey (1957) was introduced a family of power transformations such that the transformed values are a monotonic function of the observations over some admissible range and indexed.

$$(y_i^{(\lambda)} = \begin{cases} y_i^{\lambda}; \lambda \neq 0 \\ \log y_i; \quad = 0 \end{cases} \quad ...9$$

Where, y_i = data which shall be normalized, $y_i^{(\lambda)}$ = transformed amount and λ = real value that should be set so that the distribution of the data as much as possible closer to the normal distribution.

Estimation of model parameter and to identify the basic model

The most important analytical tools used with time series analysis and forecasting are the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of models identified in SPSS software to develop the SARIMA model before and after transformation of data series. The data series were divided in two segments as forecasting stage 13 years (1992 to 2004) and calibration state 7 years (2005 to 2011). It not only helps to guess the form of the model, but also to obtain approximate estimates of the Parameters (Box and Jenkins, 1976). Once identifying a tentative model is completed, the next step is to estimate the parameters in the model using maximum likelihood estimation. The parameters estimated that maximize by the probability of observations is main goal of maximum likelihood. The next, is checking on the adequacy of the model for the series. The assumption is that, the residual is a white noise process and that the process is stationary. The independent model diagnostic checking is accomplished in this work, through careful analysis of the residual series, the histogram of the residual, sample correlation and a diagnosis test (Ljung and Box, 1978). Ljung-Box, Q-test, is used to check the assumptions of model residuals and could be written as:

$$Q = n(n+2) \sum_{k=1}^{h} \left(\frac{r_k^2}{(n-k)}\right) \quad ...10$$

Where:

h= the maximum lag being considered,

n=the number of observations in the series and

r_k = the autocorrelation at lag k.

The statistic Q has a chi-square distribution with degrees of freedom (h-m) where m is the number of parameters in the model which has been fitted to the data, the chi square value has been compared with the tabulated values; in order to evaluate the valid model otherwise the model will be rejected.

Identification of best model

For successful models, it should be noted that a model with the less number of variables gives the best forecasting results, *i.e.* for a time series having more than one successful ARIMA models. The model with fewer variables (number of AR and / or MA) and this is achieved by using Akaike's Information Criterion (AIC), Bayesian information criterion (BIC) and Final Prediction Error criteria (FPE) in order to select the best ARIMA model among successful models. The smallest value of AIC, BIC and FPE should be chosen as best models.

Diagnostic checks of selected model

The final model selected based on statistical parameters such Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were considered to examine the performance of the model during forecasting period. The model was selected for further forecasting the streamflow which performance better compared to other selected model in terms of the statistical performance indices..

Adoption of model

The best ARIMA model identified on the basis of AIC, BIC, FPE, diagnostic checks and model parameter, best suited model were selected. The best suited model were checked for validation of data set. The validation data sets were chosen for 2005 to 2011 (883 days) time series to estimate srteamflow from Savitri basin. After checking the performance of model for validation period, the bests suited of the model in validation period will be adopted for forecasting streamflow of Savitri basin.

Evaluate the Performance of Models

A number of statistical criteria have been suggested by researchers (Abraham and Ledolter, 1983, Nash and Sutcliffe, 1970) to evaluate the performance of rainfall runoff models. To assess the accuracy of a rainfall-runoff model, more than one criterion should be used. The model performances were evaluated using Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Absolute Deviations (MAD), Coefficient of Efficiency (CE), and Volumetric Error (EV).

RESULTS AND DISCUSSIONS

The seasonal ARIMA (p, d, q) (P, D, Q) model which is called seasonal multiplicative autoregressive integrated moving average model or Box Jenkins seasonal models were developed for Savitri basin using daily time series data adopting five stage procedure. The ARIMA model for Savitri basin was developed using the statistical SPSS and XLSTAT-2014 software.

Initial analysis of input time series

In present study rainfall and runoff data of 20 years (1992 to 2011) were used for developing the ARIMA models for Savitri basin. The 13 year (1992 to 2004) runoff data were used for diagnose of ARIMA model and 7 year (2005 to 2011) runoff data were used for calibration of ARIMA models. The maximum, minimum, average, standard error of rainfall and runoff for Savitri basin were found as 366.9 mm, 0.1 mm, 28.14 mm and 0.0201 and 4932.34 cumecs, 0.01 cumecs, 395.12 cumecs and 0.0194, respectively (Table 1). The trends and stationary of data series was also examined by applying Mann Kendall test and KPSS test. The seasonality was tested by Dickey Fuller test. The Mann Kendall statistics (u_c) values of daily runoff data for Savitri basin were between z-table critical values (±1.96) at 5 per cent significance level. The similar results were observed also for Dickey fuller and KPSS test. This suggests that there is no linear trend and data are stationary. The seasonal variation was found in daily rainfall and streamflow used for developed ARIMA model of Savitri basin.

Model Identifications

The most common method to check stationarity is through examining the time series plot of the data. The plot of time series rainfall and streamflow at Mahad station of Savitri basin is given in Fig 3. The data shows stationary and no need to apply the differencing of the data sereis as no much seasonal vraition exist. It is also observed that the fluctuation in data series was observed for rainfall and runoff. The seasonal effect was observed in the sereis. Therefore, seasonal effect need to be taken care by differencing.

The identification of models also involve the plotting of sample autocorrelation finction (ACF) and sample partial autocorrelation function (PACF) to determine whether the series is stationary or not so as to make a decision about best fit and appropriate model for the data. The ACF and PACF are random variables and will not give the same picture as the theoretical functions.

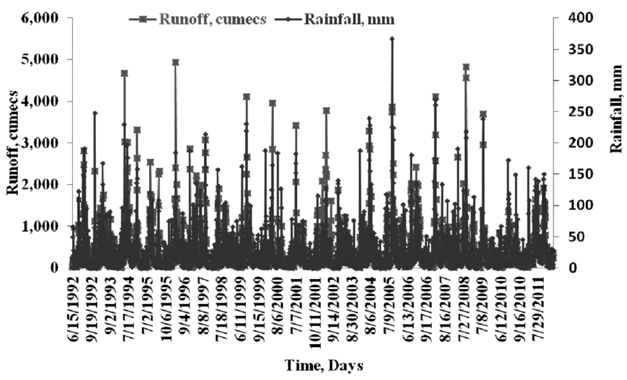

Fig. 3. Time series plot of runoff and rainfall for Savitri basin

This makes the model identification more difficult and can involves much trial and error (Nazuha *et al*., 2010).

The next step is to identify the value of p and q which are the AR (p) and MA (q) components for both seasonal and non seasonal series. The identification provides the clues about choice of the order of (p, d, q) (P, D, Q). However, in practice the degree of differencing d and D are assumed 1 while autocorrelation and partial autocorrelation function are plotted to guess the order of p, q, and P, Q. For this purpose, the ACF and PACF coefficients are computed. The estimated autocorrelation (ACF) and PACF before differencing are presented in Fig. 4.

The plot of ACF and PACF (Fig. 4) for daily sequence data were drawn to gather information about the seasonal and non-seasonal AR and MA operators concerning the daily time series. The ACF graph shows an attenuating sine wave pattern that reflects the random periodicity of the data and possibly indicates need for non-seasonal and /or seasonal AR term in the model. For this data sequences with the cyclic seasonal component, seasonal differencing need to be done. By taking the seasonal differencing operator as 1, 2, the seasonal wave pattern in the ACF was removed (Fig. 5). All the ACF for daily runoff data from Mahad station of Savitri basin were significantly different from zero. There exists a significant linear dependence between daily observations. However, the ACF did not cutoff rather damped out. This may suggest the presence of AR term.

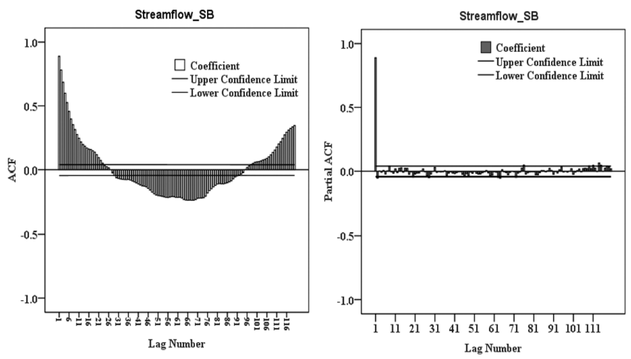

Fig. 4. *ACF and PACF for runoff of Savitri basin before differencing*

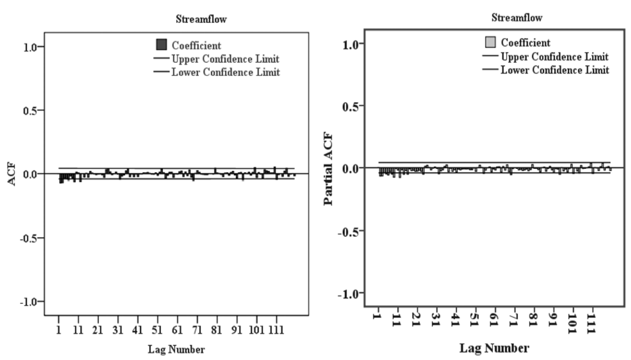

Fig. 5. *ACF and PACF for runoff of Savitri basin after differencing*

The PACF (Fig. 5) possess the significant value at some lag but rather tail off. This may imply the presence of MA term. The ACF have significant lag but multiple of 31. This may stress that, seasonal AR term are required but these value alternate. There are peaks in the graphs of the PACFs at lags 2 but that are multiple of 31, which may suggest seasonal MA term. But these peaks may damp out. For seasonal ARIMA, the general notations is ARIMA (p, d, q) $(P.D, Q)^S$. Based on the pattern and above discussion, the respective values of P, D, Q were determined for ARIMA (1,1,1) $(1,1,1)^{31}$. The alternative models were also selected by inspecting and considered for the principle of parsimony. The diagnostic checks were applied in order to determine whether the residual of alternative models were independent, homoscedesic, and normally distributed. The other models identified were ARIMA(1,0,2)$(0,0,0)^{31}$, ARIMA(2,0,2) $(0,0,0)^{31}$. ARIMA(2,2,2) $(0,0,0)^{31}$, ARIMA(0,1,1) $(0,1,1)^{31}$, ARIMA(1,0,1) $(0,1,1)^{31}$, and ARIMA(1,1,0) $(1,0,1)^{31}$.

Selection of models from information criteria

The identification and selection of non Gaussian parsimonious ARIMA models are based on Akaike Information Criteria (AIC), (Akaike, 1974), the Bayesian information criteria, (BIC) (Schwarz, 1978) and Final Prediction Error (FIC), (Akaike, 1970). The most parsimonious models are those which gives minimum AIC, BIC and FIC values. The alternative ARIMA models were estimated by considering the AIC, BIC and FPE criteria. The models that have minimum AIC, BIC and FPE values are assumed to be parsimonious. The estimated value of AIC, BIC and FPE are presented in Table 2. From the table it is observed that ARIMA $(202)(000)^{31}$ have lowest AIC, BIC and FPE value, but error of prediction was found in case of ARIMA $(111)(111)^{31}$. Therefore, the model ARIMA $(111)(111)^{31}$ was selected and identified as best model to predict the runoff for Savitri basin.

Estimation of model parameters

Each ARIMAs tentative model parameters can be tested using t-value and p values. Dividing the coefficient by its standard error calculates a t-value. The standard error (SE) of coefficient is the standard deviation of the estimate of a regression coefficient. Its value is always positive and smaller value indicates a more precise estimate. The standard error of a coefficient helps in determining whether the value of the coefficient is significantly different than zero. The estimated model parameter presented in Table 3 and Table 4 were analyzed at 5 per cent significant level by using t-test to select the best fit model to the data. The values of parameters concerning with the test models were standard errors, t-ratios and probabilities (at 95 per cent confidence level) for model ARIMA $(111,111)^{31}$ and ARIMA $(202,000)^{31}$ are given in Table 3 and Table 4, respectively. The standard errors calculated for the both models by Hessian and Asymptotic methods were found to be small as compared to model parameters values. Therefore, all of the parameters are significant and these parameters should be included in the model. The model ARIMA $(111,111)^{31}$ shows parameters are in confidence limits as compared to ARIMA $(202,000)^{31}$. Hence, model ARIMA $(111,111)^{31}$ was adopted for the diagnostic analysis and forecasting of the runoff.

Diagnostic analysis of ARIMA models

The next step of series modeling approach is diagnostic checking. It is aimed at examining the accuracy of the chosen tentative model in ensuring that the modeling assumptions are satisfied. Several procedures can be applied to check the accuracy of the model as to whether the model satisfies the stability or stationary conditions, as required in stochastic modeling (Ayob and Amat, 2004).

For this Ljung Box is used for testing of white noise residual. Hypothesis null is that, residual should be white noise. In other word, the residual series should be independent, homoscedastic (having constant variance), and normally distributed. We can reject the null hypothesis if p value in Chi-square statistics is greater than alpha at 95 per cent confidence level.

The chi-square test by Box-Pierce (1970), Ljung and Box (1970), and McLeod and Li (1984) is termed as portmanteau test and was performed with different lag and degree of freedom for both tentative selected ARIMA models. The Box-Pierce test indicates that if residuals are white noise, the Q-statistic follows χ^2 distribution with $(h-m)$ degrees of freedom. If a model is fitted, then m is the number of parameters. However, no model is fitted here, so our m=0. If each r_k value is close to zero, then Q will be very small; otherwise, if some r_k values are large – either negatively or positively – then Q will be relatively large. It compared Q to the χ^2 distribution, just like any other significance test. In this study, the selected ARIMA models have p-value less than alpha value (Table 5 and Table 6). So the null hypothesis cannot be rejected and it is concluded that white noise for both models was observed.

The models performance was also checked by performing the RMSE and MAPE during diagnostic

Table 2. *Identification of ARIMA(p d q)(P D Q) model on the basis of AIC, BIC and FPE for Savitri basin for runoff modelling*

Sr. No.	ARIMA (p d q) (PDZQ) Model	AIC	BIC	FPE
1	ARIMA$(1\ 1\ 1,\ 1\ 1\ 0)^{31}$	29219.79	29301.39	9280.60
2	ARIM $(1\ 1\ 1,\ 1\ 1\ 1)^{31}$	29224.06	29253.24	7619.81
2	ARIMA$(1\ 1\ 0,\ 1\ 0\ 1)^{31}$	29669.78	29681.41	9084.16
4	ARIMA$(10\ 1,\ 011)^{31}$	29387.84	29411.11	8118.65
5	ARIMA$(1\ 0\ 2,\ 0\ 0\ 0)^{31}$	29071.67	29084.95	7087.76
6	ARIMA$(2\ 0\ 2,\ 0\ 0\ 0)^{31}$	28616.30	28645.39	5869.77
7	ARIMA$(2\ 2\ 2,\ 0\ 0\ 0)^{31}$	29227.51	29256.59	7612.86

Table 3. *Final estimated model parameters for ARIMA $(111,111)^{31}$*

Parameter	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)	Asympt. standard error	Lower bound (95%)	Upper bound (95%)	T
AR (1)	0.461	0.029	0.405	0.517	0.020	0.421	0.500	15.89
SAR(1)	0.081	0.029	0.025	0.138	0.022	0.038	0.124	2.79
MA(1)	0.421	0.028	0.367	0.475	0.018	0.385	0.457	15.03
SMA(1)	-1.000	0.003	-1.006	-0.994	0.000	-1.000	-1.000	-333.3

P<0.0001 for all

Table 4. *Final model estimated parameter for ARIMA(202,000)*[31]

Parameter	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)	Asympt. standard error	Lower bound (95%)	Upper bound (95%)	T
AR(1)	0.443	0.015	0.413	0.472	0.018	0.407	0.478	29.53
AR(2)	0.424	0.012	0.400	0.447	0.018	0.388	0.459	35.33
MA(1)	1.620	0.002	1.616	1.623	0.000	1.620	1.620	810.0
MA(2)	1.000	0.002	0.996	1.004	0.000	1.000	1.000	500.0

(P<0.0001 for all)

Table 5. *Modified Box Pierce (ljung Box) Chi Square Statistics for ARIMA(111,111)*[31]

Lag	6	12	24	30
Chi-Square	4539.179	5580.225	5580.947	255580.947
DF	6	12	24	30
p-Value	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 6. *Modified Box Pierce (ljung Box) Chi Square Statistics for ARIMA (202,000)*[31]

Statistic	DF	Value	p-value	Lag
Jarque-Bera	2	6201.788	< 0.0001	6
Box-Pierce	6	6401.010	< 0.0001	6
Ljung-Box	6	6412.234	< 0.0001	6
Box-Pierce	12	6401.010	< 0.0001	12
Ljung-Box	12	6412.234	< 0.0001	12

period for checking its consistency of selected of models. The RMSE and MAPE for both ARIMA (111,111)[31] and ARIMA (202,000)[31] were observed 87.18 cumecs and 14.93 per cent and 76.55 cumecs and 21.61 per cent respectively (Table 7). The lower value of RMSE and MAPE indicates the goodness of fit of the models. Therefore, it is interfered that ARIMA (111,111)[31] best fit compared to ARIMA (202, 000)[31]. Hence, ARIMA (111,111)[31] model could be used for forecasting the runoff of Savitri basin.

Table 7. *ARIMA Models performance during developmental stage for Savitri basin*

ARIMA Model	RMSE	MAPE
ARIMA(111,111)[31]	87.18 cumecs	14.93 per cent
ARIMA(202,000)[31]	76.55 cumecs	21.61 per cent

Statistical performance of ARIMA models

The statistical performance of selected ARIMA models were also performed for Coefficient of determination (R), Root mean square error (RMSE), Nash shift coefficient (CE), volumetric efficiency (VE), mean absolute deviation (MAD) and Mean average percentage error (MAPE) during the diagnostic and forecasting (Calibration) period. The results for both selected ARIMA models are presented in Table 8. It is observed that for ARIMA (111,111)[31] model R value is higher (0.94) as compared to ARIMA (202,000)[31] during diagnostic and forecasting period. The value of R approaching to 1 indicates the best fit model for prediction. The minimum RMSE value indicates the model performed well. The CE value approaching to 100 reflects model performance to be best. The EV approaching to one indicate the model prediction approaches to real estimates. The deviation between data series less the model performance is good. The MAPE of two data series is minimum the model value reaches to real. These are the indices which interpret the performance of models. The negative value of MAD indicates that model over predict. Hence, from Table 8 it is interfered that on an overall ARIMA (111,111)[31]performed better as compared to ARIMA (202,000)[31] for prediction of runoff of Savitri basin.

The scatter plot of ARIMA (111,111)[31] and ARIMA (202,000)[31] were ploted for diagnostic and forecasting period as shown in Fig. 6 to Fig. 9. It is observed that ARIMA (111,111)[31] performed better during diagnostic period and for forecasting the streamflow for Savitri basin over ARIMA (202,000)[31] with R^2 0.85 in diagnostic

***Table 8.** Statistical performance of ARIMA models during diagnostic (1992 to 2004) and forecasting (2005 – 20011) period for Savitri basin*

Sr. No.	Model	Statistical parameters					
		R	RMSE	CE	EV	MAD	MAPE
A	During diagnostic period (1992 to 2004)						
1	ARIMA(111,111)[31]	0.943	50.69	88.95	0.085	0.0279	29.516
2	ARIMA(202,000)[31]	0.926	58.89	85.091	0.916	-0.317	199.42
B	Forecasting period (2005 to 2011)						
1	ARIMA(111,111)[31]	0.938	83.158	83.158	0.0519	-0.0365	146.51
2	ARIMA(202,000)[31]	0.896	69.097	80.00	0.249	-0.175	266.72

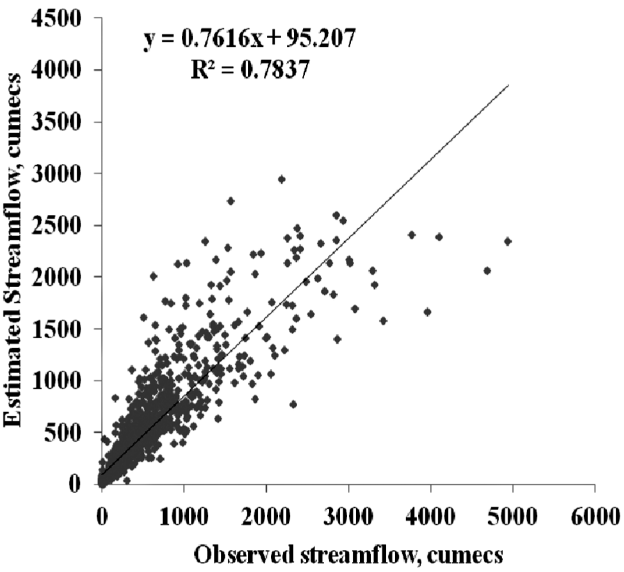

__Fig. 6.__ Scatter plot of estimated and observed runoff for dignostic period of ARIMA (111,111)[31] model

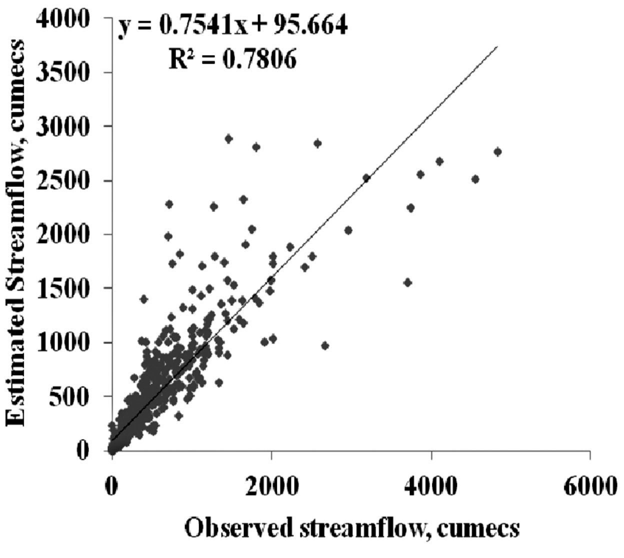

__Fig. 7.__ Scatter plot of estimated and observed runoff for forecasting period of ARIMA (111,111)[31] model

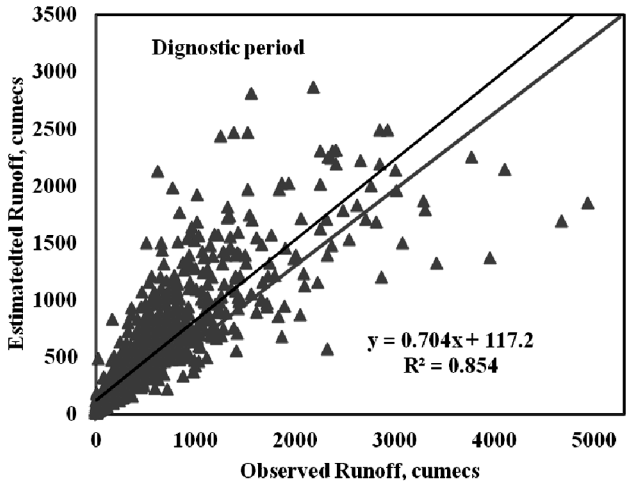

__Fig. 8.__ Scatter plot of estimated and observed runoff for dignostic period of ARIMA (202,000)[31] model

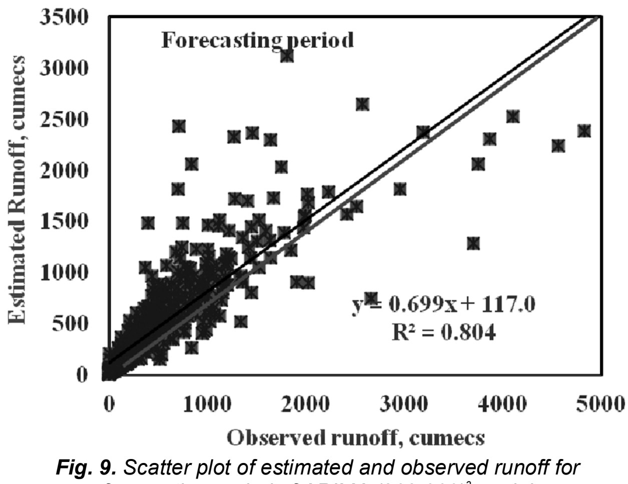

__Fig. 9.__ Scatter plot of estimated and observed runoff for forecasting period of ARIMA (202,000)[3] model

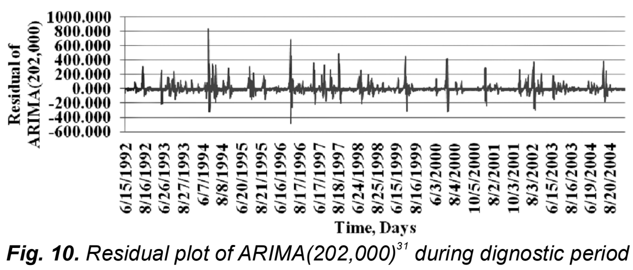

__Fig. 10.__ Residual plot of ARIMA(202,000)[31] during dignostic period

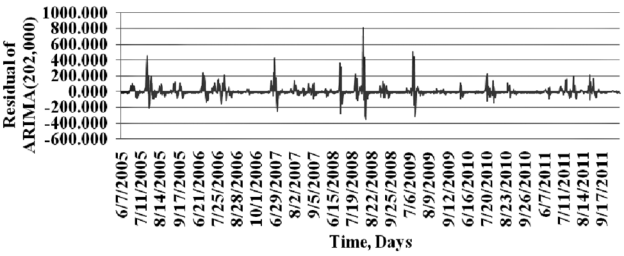

Fig. 11. *Residual plot of ARIMA(202,000)[31] during forecasting period*

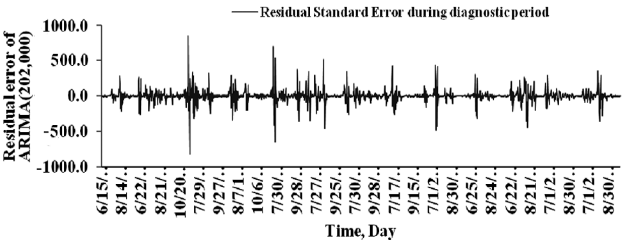

Fig. 12. *Residual plot of ARIMA(111,111)[31] during diagnostic period*

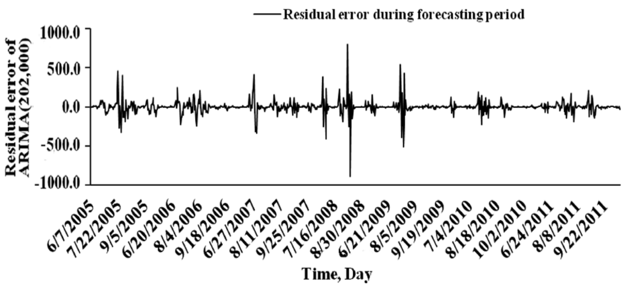

Fig. 13. *Residual plot of ARIMA(111,111)[31] during forecasting period*

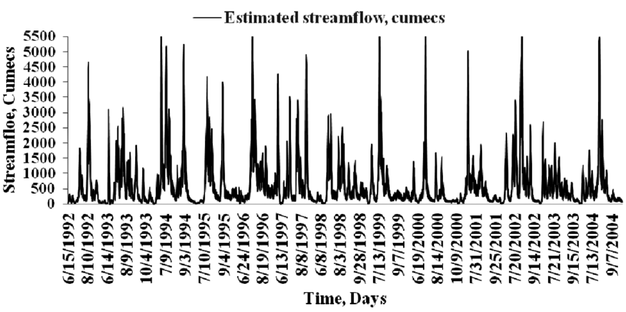

Fig. 14. *Comparative plot of observed and estimated runoff by ARIMA(111,111)[31] of Savitri basin during dignostic period*

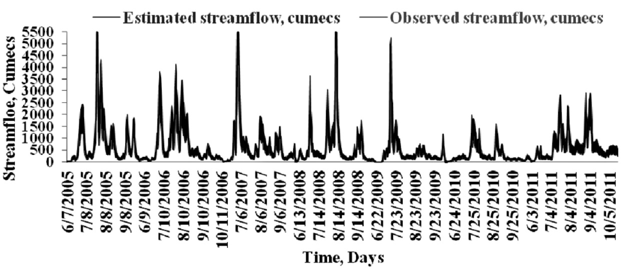

Fig. 15. *Comparative plot of observed and estimated runoff by ARIMA(111,111)[31] of Savitri basin during forecasting period*

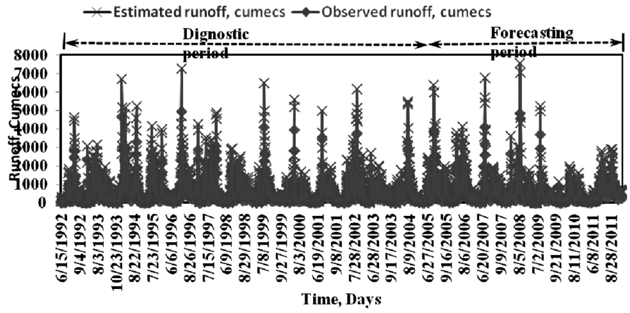

Fig. 16. *Comparative plot of observed and estimated runoff by ARIMA(202,000)[31] model for Savitri basin during dignostic and forecasting period*

period and 0.80 for forecsting period for the runoff. The residual error plot of ARIMA (111,111)[31] and ARIMA (202,000)[31] are presented in Fig. 10 to Fig. 13 for diagnostic and forecasting period, resepctively. It is also observed that ARIMA (111,111)[31] shows minimum error compaired to ARIMA (202,000)[31] during diganostic and forecasting periods also. The comparing the results of ARIMA models with original observed runoff series is presented in Fig.14 to 16 for ARIMA (111,111)[31] and ARIMA (202,000)[31]. It is observed form the Fig.14 and Fig. 16 that ARIMA (111,111)[31] performed better in prediction and forecasting the runoff over the ARIMA (202,000)[31]. Hence ARIMA (111,111)[31] was selcted for further forecasting of the stream flow of Savitri basin.

CONCLUSIONS

The stochastic based seasonal ARIMA model was developed using SPSS software for forecasting the streamflow for Savitri Basin in Maharashtra. The performance of model during diagnostic and calibration stage were found to be satisfactory with R value more than 0.9, RMSE is less, CE is also approaching to 100 and EV is less. These indicate the model could be used for forecasting the streamflow on daily basis with good performance on short term basis.

REFERENCES

Abbott, M. B., Bathurst, Cunge, J. A. Connell, P. E., and Rasmussen, J. (1986). An Introduction to the European Hydrological System - System Hydrologique Europeen, "SHE", 2: Structure of a Physically-Based Distributed Modeling System. *Journal of Hydrology* **87**: 61-77.

Abraham, B. and Ledoltor, J. (1983). Statistical Methods for Forecasting, *John Wiley and Sons Inc.,* New York. pp 472.

Akaike, H. (1970). Statistical predictor identification. *Annals of Institutional Mathematics* **22**: 203-217.

Akaike, H. (1974). Information theory and an extension of the maximum likelihood principle. In: Second international symposium on *Information Theory*, B. N. Petrov & F. Csaki (eds.), Budapest, Hungary, Akademai Kiado. pp 267-281.

Arnold, J. G., Srinivasan, R. ,Muttiah, R. S., and Williams. J. R. (1998). Large Area Hydrologic Modeling and Assessment Part I: Model development. *Journal of the American Water Resources Association* **34**: 73-89.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). Artificial neural networks in hydrology. I: preliminary concepts. *Journal of Hydrology Engineer ASCE* **5**(2): 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000b). Artificial neural networks in hydrology. II: hydrologic applications. *Journal of Hydrology Engineer ASCE* **5**(2):124-137.

Ayob K. And Amat, S. D. (2004). Water use trends at university Tekologi Malasiya:Application of ARIMA model. *Journal of Technology* **41**(B): 47-56.

Box, G. E. P. and Jenkins, G. (1974). Time series analysis, forecasting and control. Holden Day, San Francisco.

Box, G. E. P. and Jenkins, G. M., (1970). Time series Analysis, Forecasting and Control. Holden-Day, San Francisco, CA.

Box, G. E. P. and Pierce. D. A. (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association* JSTOR 2284333 **65**: 1509–1526.

Box, G. E. P. and Jenkins. G. M. (1976). Time series analysis forecasting and control. Holden-Day, Oakland, Calf.

Davidson, J. W., Savic, D. A. and Walters, G. A. (2002). Symbolic and Numerical Regression: Experiments and Applications. *Journal of Informatic Science*.

Galeati, G. (1990). A Comparison of Parametric and Non-parametric Methods for Runoff Forecasting. *Hydrology Science Journal* **35**: 79–94.

Gorman, J. W. and Toman, R. J. (1966). Selection of Variables for Fitting Equations to Data, *Technometrics* **8**(1): 27-51.

Lall, U. and Bosworth, K. (1993). Multivariate Kernel Estimation of Functions of Space and Time Hydrologic Data. In: *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, (ed.) Hipel, K. Kluwer, Waterloo.

Ljung, G. M. and Box, G. E. P. (1978). On measure of lack of fit in time series models. *Biometrika* **65**(2): 279-303.

Mahdi, Moharrampour, Abdulhamid Mehrabi, Hooman Hajikandi, Saeed Sohrabi, and Javad Vakili. (2013). Comparison of Support Vector Machines (SVM) and Autoregressive integrated moving average (ARIMA) in daily flow forecasting. *Journal of River Engineering* **1**(1):1-8.

McLeod, A. I. and Li. W. K. (1984). Diagnostic checking ARIMA time series models using squared residual autoregressive models. *Journal of Time Series Analysis* **4**: 269-273.

Millere, W. A. and Cunge, J. A. (1975). Simplified equation of unsteady flow. In: *Unsteady flow in Open Channel*. Vol –I, K. Mahamood and V. Yejevich (eds.), Water Resources Publication, Fods, Collins, Colorado.

Muthukrishnan, S. J. Harbor, Kyoung Jae Lim, and Bernard A. Engel. (2006). Calibration of a Simple Rainfall-runoff Model for Long-term Hydrological Impact Evaluation. *URISA Journal* **18**(2): 35-42.

Nash, J. F. and Sutcliffe. J. V. (1970). River flow forecasting through conceptual models. *Journal Hydrology Science* **44**: 399-417.

Nazuha, M., Ruzaidah, S. and Zamzulani, M. (2010). Malaysia Crude Oil Production Estimation: an Application of ARIMA model. International Conference on *Science and Social Research* (CSSR 2010).

Nigam, R., Nigam, S. and Kapur, S. (2013). Time series modeling of tropical runoff. *International Journal of Pure and Applied Research in Engineering and Technology*. **1**(7): 13-19

Salas, J. D., Delleur, J. W., Yevjevich, V. and Lane, W. L., (1980). Applied modeling of hydrologic time series. *Water Resources*, Littleton. p 484.

Schwars, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**: 461–464.

Turkey, J. W. (1957). The comparative anatomy of transformations. *Annals of Mathematical Statistics* **28**: 602-632.