International Journal of Agriculture Sciences
ISSN: 0975-3710&E-ISSN: 0975-9107, Volume 8, Issue 45, 2016, pp.-1916-1922.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217

Research Article

STOCHASTIC BASED ARIMA MODEL FOR FORECASTING SEDIMENT YIELD FOR KAL RIVER IN KONKAN REGION OF INDIA

GHARDE K.D.[1*], KOTHARI MAHESH[2], MITTAL HEMANT[3] AND MAHALE D.M.[4]

[1&4]Department of Soil and Water Conservation Engineering, CAET, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli, 415712, India
[2&3]Department of Soil and Water Engineering, CTAE, Maharana Pratap University of Agriculture and Technology, Udaipur, Rajasthan,313001, India
*Corresponding Author: Email-ghardek@rediffmail.com

Received: April 24, 2016; Revised: June 20, 2016; Accepted: June 21, 2016; Published: October 09, 2016

Abstract- Forecasting of sediment yield and runoff 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 estimating and forecasting sediment yield. In present study, sediment yield on daily basis of Kal river is tributary of Savitri basin were forecasted using seasonal ARIMA model developed in SPSS software following initial analysis of time series, identification model, apply diagnostic checks, adopting a model and forecast the runoff by adopted model with checking the statistical indices such R, RMSE, CE, EV, MAD, MAPE. The sediment yield data for duration of 2003 to 2009 (7 years) measured at Birwadi station were used to calibrated, validate and forecast the sediment yield 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$(101,000)^{31}$ as best model for forecasting the sediment yield with R value more than 0.95 during calibration and forecasting period. The other statistical indices RMSE, CE, EV, MAD, MAPE were also found in appropriate range to forecast the sediment yield with goodness of fit. Hence, developed seasonal ARIMA model could be adopt for forecast the sediment yield on daily basis for sub tropical coastal region of Maharashtra by adopting the estimated parameters.

Keywords- Runoff, sediments¸ Auto correlation Function, Partial Correlation Function, Akai information criteria

Citation: Gharde K. D., et al., (2016) Stochastic based ARIMA Model for Forecasting Sediment Yield for Kal River in Konkan Region of India. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 8, Issue 45, pp.-1916-1922.

Introduction

The development and assumption for sediment transport function in river system are varying degree with complex in nature. But some of the assumptions made for forecasting sediment yield in natural river system may not be true for ever condition. Empirical solutions based on observations may be useful only for a particular site where the data were collected. Many of the sophisticated theoretical solution require large number of parameters that are difficult to obtain [1]. While adopting stochastic based approached for applying in natural system, it is essential to understand the concept of physical process occur in the nature. But, there should be conceptualized understanding between adopted system and natural system which considered for conceptualization.

The relationship between rainfall and runoff with sediment yield is an important issue in surface hydrology. The amount of sediment yield from rainfall is necessary to predict for avoiding risk and assessment of flood [2]. The quantity of sediment yield deposited or transported is totally depends on the rainfall amount, rainfall intensity, duration, distribution beside that, other parameters such as soil type, vegetation cover, soil moisture, and land slope etc [3]. There were several hydrologic models are available for prediction of sediment yield such as Physical based model (Soil and Water Assessment Tool [4], Black box models (Artificial Neural Network model, [5,6]), Stochastic process models (Autoregressive Moving Average, (ARMA), Auto regressive integrated moving average model etc [7], Physical based parameter distributive models (SHE [8], Kinematic waves models [9], etc. But each model have its own limitation and advantages in applying at particular topographic and geographical locations such as data available, processing pattern of data and output variability, etc.

The ARIMA widely adopted for forecasting the change in trends in climatic parameters, rainfall and runoff for small to large catchments and proven to be very good fitted hydrologic stochastic process model. The ARIMA model is statistical based stochastic process model analyzed the time series data for forecasting of runoff and sediment yield. 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 and sediment yield forecasting by many researchers [10, 11, 12 &13] mostly because of its simplicity and ease of use. Time series analysis allows identification of hidden deterministic behavior and thus understanding of cause and effect relationship in problems [14]. [15, and 16] applied the ARIMA model for forecasting the runoff on daily basis. This model provides an increased understanding of the behavior of the system. Stochastic based mathematical relationships are very useful in hydrology for building mathematical relationship to generate synthetic hydrological records, to forecast hydrological events, and to detect changes [17]. The steps involved in stochastic modeling are; identification and removal of significant trends present, identification of periodicity, separation of dependent stochastic components, and developed residuals for frequency distribution [18]. The unavailability of large time series data on runoff, rainfall and sediment yield hampered the application of various stochastic process model in forecasting runoff to sediment yields. 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 Kal river using daily time series data available for duration of 7 years (2003 to 2009). By considering this fact, the present study was undertaken to develop the seasonal ARIMA model for Kal river gauged at Birwadi hydrological station used to forecast sediment yield and checks it significance with statistical parameters in forecasting stream flow and sediment yield for Kal river tributary of Savitri Basin.

Materials and Methods

Study area

The present study was conducted for the Kal river tributary of Savitri basin comes under the Western part of Sahayandri Ghat of Konkan region and located in Maharashtra 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 runoff, rainfall and sediment yield data used measured at Birwadi were collected from Superintending Engineer, Unit of Hydrology Project, Nashik for duration of 7 years (2003 to 2009) of sediment yield, runoff and rainfall data. The statistical parameters of inputs data for models such as Mean, Maximum, Minimum, Standard deviation (SDV), ACF and PACF coefficient estimated using standard procedure [Table-1].

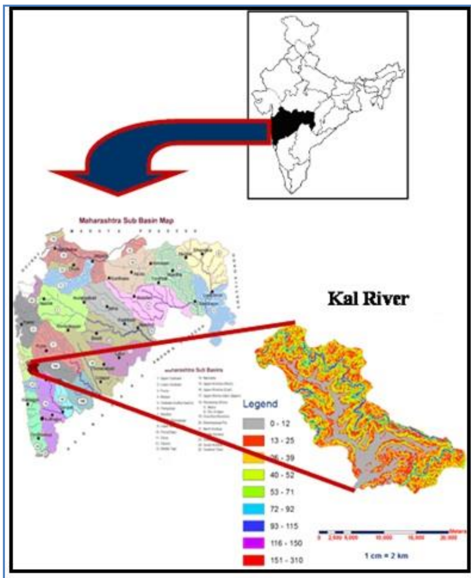

Fig-1. Location of Kal river

Table-1 *Estimated Statistical parameters of hydrologic data measured at Birwadi during 1991 to 2011*

Sr No	Data	Max.	Min	Mean	SD	ACF	PACF	SE
1	Rainfall, mm (1991 to 2011)	370	0.1	33.9	45.81	0.010	0.0036	0.020
2	Runoff, Cumecs (1991 to 2011)	2394.5	0.1	184.92	233.04	0.0036	0.0079	0.020
3	Sediment, t/day (2003 to 2007)	62123.3	0.8	475.28	2790.37	0.0193	0.0044	0.034

ARIMA model theory

The seasonal 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 [Eq-1].

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

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 [Eq-2]:

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

Where:
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 [Eq-3]:

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

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 [Eq-4].

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

Where,
B=backward operator,
d=the non-seasonal order of differences.
The seasonal difference of order D with period is written as S [Eq-5]

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

In general, the differencing operation may be done several times but in practice only one or two differencing operation are used [19]. Box and Jenkins (1974) [20], 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 [Eq-6]:

$$(1-\Phi_1 B^S - \Phi_2 B^{2S} - ..... - \Phi_p B^{pS})(1-B^S)x_t = (1-\Theta_1 B^S - \Theta_2 B^{2S} - ..... - \Theta_q B^{qS})\alpha_t \qquad ...[6]$$

The residuals α_t are in turn is represented by an ARIMA (p,d,q) model by [Eq-7],

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

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].

$$\Phi(B^S)\Phi(B)(1-B^S)^D(1-B)^d x_t = \Theta(B^S)\Theta(B)\varepsilon_t \qquad ...[8]$$

ARIMA Model Methodology

The ARIMA model was developed in SPSS software following the procedure suggested by Box and Jenkin (1976) [21] 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) Model parameter estimation. 4) Checks model appropriateness for forecasting and 5) Model adoption for forecasting.

Initial analysis of the data

The daily stream flow time series data used of hydrologic station Birwadi, Tal Mahad for duration of 7 years (2003 to 2009) for Kal river. The collected data were check 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) [22] reported a family of power transformations function which are monotonic in nature with observed data sets..

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

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 Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of ARIMA models are used in time series analysis and forecasting. These were 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 for 5 years (2003 to 2007) and calibration state for 2 years (2008 to 2009). This procedure helps in identify the form model and to estimates models parameters [21]. The maximum likelihood method was adopted for estimated parameters that maximize by the probability of observations. The adequacies of models were checks by assuming the residual is with white noise process and stationarity. The residual series analysis helps in diagnose the independent parameters using histogram, sample correlation and diagnostic checks (Ljung and Box, 1978) [23]. Ljung-Box, Q-test, is used to check the assumptions of model residuals and could be written as [Eq-10]:

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

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 runoff which performance better compared to other selected model in terms of the statistical performance indices.

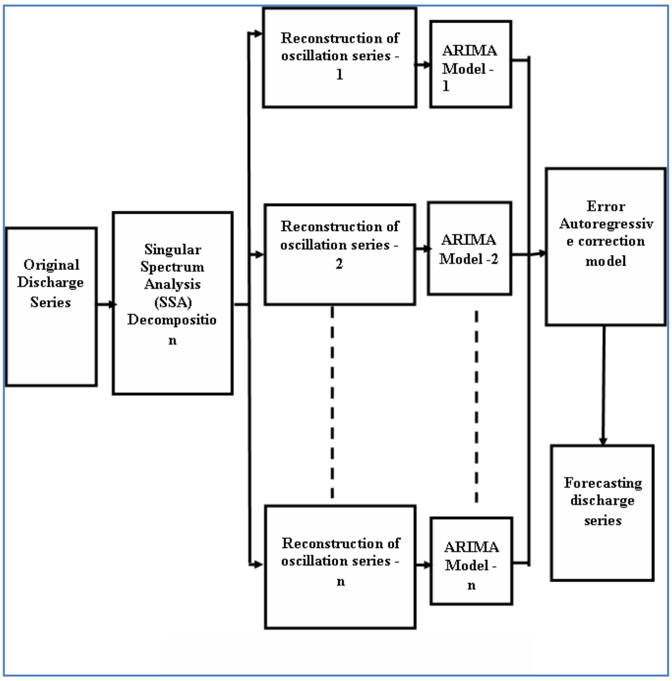

Fig-2 Flow diagram of ARIMA model execution

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 models were checked for validation of data set. The validation data sets were chosen for 2007 to 2009 time series to estimate runoff 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 runoff of Savitri basin.

Evaluate the Performance of Models

A number of statistical criteria have been suggested by researchers [24 and 25] to evaluate the performance of rainfall runoff models. The accuracy of a sediment yield model developed was evaluated by adopting multiple statistical indices such as Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Absolute Deviations (MAD), Coefficient of Efficiency (CE), and Volumetric Error (EV).

Results and Discussions

The ARIMA model developed for forecasting of sediment yield for Kal River was developed using SPSS and XLSTAT 2014 software adopting procedure discuss above. The sediment data of Birwadi station was collected from Hydrologic Project division, Data Storage Center, Nashik, Maharashtra.

Initial analysis

In present study, sediment yield and runoff data of 7 years (2003 to 2009) measured at Birwadi were used for developing the ARIMA models for Kal River. Out of available data series 5 years (2003 to 2007) data were used for diagnostic of the ARIMA model and of 2 years (2007 to 2009) data were used for calibration and forecasting of sediment yield. The statistical parameters of inputs data series performed such as the maximum, minimum, mean, standard error of data sets were estimated and presented in [Table-1]. The maximum, minimum, mean, standar deviation, ACF coefficient, PACF coefficient, and standard error of sediment yield for Kal river were 62123.35 tones/day, 0.8 tones/day, 475.28 tones/day, 2790.37, 0.0193, 0.0044, and 0.0349, respectively. The stationarity of data series was examined by applying Mann Kendall test and KPSS test analysis.

The seasonality was tested by applying the Dickey Fuller test. The test performed and results were presented in the [Table-2]. The Mann Kendall statistics (u_c) values of daily sediment yield data from Kal river were between z-table critical values (±1.96) at 5 per cent significance level. The same results also observed for Dickey Fuller and KPSS test. This suggest that, there is no linear trends, stationary, non seasonal in daily sediment yield for Kal river measured at Birwadi.

Model Identification

The stationarity of data series were checks through examining the time series plot. Stationary means that data fluctuates around a constant mean. The plot of runoff and sediment yield at Birwadi station of Kal river is given in [Fig-3]. The data shows stationarity and no need to apply the differencing to the data series with non seasonality. It is observed that, the fluctuating data series observed for sediment yield and seasonal affect observed in the series. Therefore, seasonal effect need to be taken in account.

Table-2 *Trends analysis of rainfall and runoff data of Kal river measured at Birwadi hydrologic station*

Sr. No.	Hydrological data	Mann Kendall Test*	Dickey Fuller test#	KPSS test **		Interference		
		τ value	τ value	τ value	P values		τ value	P values
1	Rainfall, mm	-0.002	-11.95	0.064	0.402	SNT	0.064	0.402
2	Runoff, cumecs	0.007	-9.97	0.078	0.277	SNT	0.079	0.279
3	Sediment yield, tones/day	0.038	-5.35	0.230	0.016	SNT	0.23	0.016

* **Mann Kendall Test (MKT) interference**: H_0= there is no trend in series, H_a= there is trend in the series, As the computed alpha (α=0.05) is greater than P value, hence there is no trend in the series. Or if the computed alpha (α=0.05) value is less than P value, hence there trend in the series.

\# **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) trend test interference**: H_0 : the series is stationary, ha: the series is not stationary, As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H_0.

****Dickey Fuller test** ADF (stationary) interference: H_0: there is a unit root for the series, H_a: There is no unit root for the series. The series is stationary, As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H_0, and accept the alternative hypothesis H_a.

$^{\$}$**SNT** = stationery series with no trend, NSDT = Non stationary decreasing in trend series.

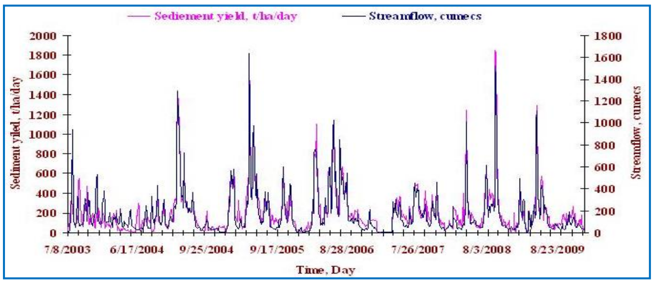

Fig-3 Time series plot of sediment yield and runoff of Kal river

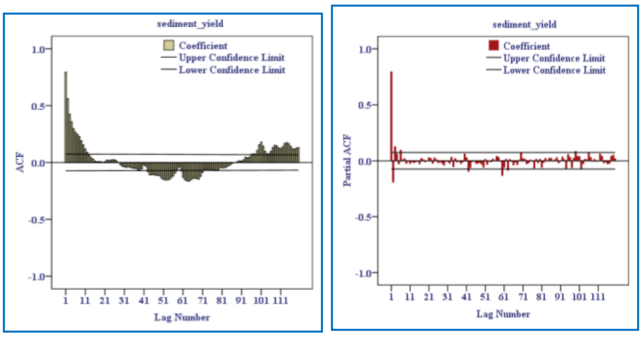

Fig-4(a) ACF for sediment yield Fig-4(b) PACF for Sediment yield

Fig-4(a,b) ACF and PACF for sediment yield of Kal river before transformation

The estimated autocorrelation (ACF) and PACF before differencing are presented in [Fig-4(a&b)] whereas the ACF and PACF after transformation are presented in [Fig-5(a&b)]. The ACF and PACF coefficient estimated are presented in [Table-2]. From the above [Fig-4] it is observed that, the ACFs for daily sediment yield time series follows an attenuating sine wave pattern that reflects the random periodicity of the data and possibly indicates the need of non seasonal and or seasonal AR terms in the model. The data sequences have the cyclic seasonal component and it needed seasonal differencing. By considering one as the difference operator, the seasonal wave pattern in the ACFs was nullified [Fig-4] All the ACF graphs of daily data series were significantly difference from zero. This indicates that, data have linear dependence and the ACFs did not cut off rather damp out. This may suggested the presence of autoregressive term (AR) terms. The PACFs graphs shows the significant value at some lag but it tail off. This means moving average model exist. It is also observed that, ACF is significant at multiple of 31 terms in seasonal values and indicate AR term required in ARIMA model development but at attenuate value. The value of PACF at multiple of 31 shows the presences of MA term but it dump out at some places [Fig-5].

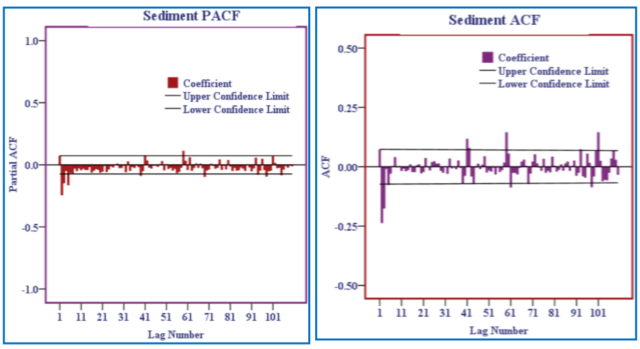

Fig-5(a) ACF for sediment yield Fig-5(b) PACF for sediment yield

Fig-5(a&b) ACF and PACF for sediment yield of Kal river after transformation.

Table-3 *ARIMA model Selection Criteria by AIC, BIC and FPE for sediment yield estimation for Kal river*

Sr. No.	ARIMA Model	AIC	BIC	FPE
1	ARIMA(101,000)^{31}	7049.11	7062.92	814.92
2	ARIMA(111,111)^{31}	6976.97	6999.97	759.34

The alternatively models identified on the basis of ACFs and PACFs are presented in [Table-3]. The best models were also identified by adopting AIC, BIC and FPE criteria applying the principle of parsimony. The model that has minimum AIC, BIC and FPE values was assumed to be parsimonious. From the [Table-3] it is observed that among the identified models ARIMA (111,111)^{31} model shows lowest AIC, BIC, FPE values 6976.97, 6999.97 and 759.34, respectively. The other model considered was ARIMA (101,000)^{31} for identified the models and estimate the models parameters.

Estimation of models parameters

The tentative parameters can be tested for each identified model using t value and p values. The identified models parameters estimated for model ARIMA (111,111)[31] and ARIMA (101,000)[31] are presented in [Table-4] and [Table-5], respectively. It is observed from [Table-4] that Hessian standard error (0.007) and Asymptotic standard error are (0.0001) lowest in seasonal moving average. The autoregressive (φ_1), seasonal auto regression (Φ_2), moving average and seasonal moving average values for model ARIMA (111,111)[31] are 0.419, 0.019, 0.339 and -1.00 respectively. The T value for ARIMA (111,111)[31] observed lowest in case of SAR (1) i. e 0.29 and height for MA(1) i.e 7.55. This indicates that standard error of MA1 coefficient is large relative to the value of the coefficient itself, so the t-value of 0.29 is too small to declare statistical significance. This tends to conclude that coefficient is not different from zero.

From [Table-5] it is observed that Autoregressive (φ_1), and Moving average (θ_1) values for ARIMA (101,000)[31] were 0.958, 0.608, respectively. The Hessian standard error was highest for autoregressive (MA1) i.e., 0.023 and lowest for AR1 i. e. 0.01. Asymptotic standard error was observed highest for ARIMA(101,000)[31], MA1(0.029) and lowest for AR(0.011). it is also observed that T value is less for MA (23.38). It is revealed from both the [Table-3] and [Table-4] that $|T_{cal}| < T_{table}$ and alpha value is less than level. Hence, hypothesis cannot be rejected and it can be concluded that coefficient are different from the zero. Therefore, these parameters were passive for the above-mentioned models and these can be used for forecasting sediment yield of Kal river.

Table-4 *Estimated models parameters of ARIMA(111,111)[31] for Sediment yield modeling of Kal river*

Parameter	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)	Asympt. standard error	Lower bound (95%)	Upper bound (95%)	T ratio
AR(1)	0.42	0.10	0.22	0.62	0.04	0.35	0.49	6.07
SAR(1)	0.02	0.01	-0.16	0.19	0.04	-0.06	0.09	0.29
MA(1)	0.37	0.01	0.18	0.56	0.03	0.30	0.43	7.55
SMA(1)	-1.00	0.01	-1.01	-0.98	0.00	-1.00	-1.00	-285.7

Table-5 *Estimated models parameters of ARIMA (101,000)[31] for Sediment yield modeling of Kal river*

Parameter	Value	Hessian standard error	Lower bound (95%)	Upper bound (95%)	Asympt. standard error	Lower bound (95%)	Upper bound (95%)	T ratio
AR(1)	0.958	0.010	0.938	0.979	0.011	0.937	0.979	91.24
MA(1)	0.608	0.023	0.563	0.652	0.029	0.550	0.666	23.38

Diagnostic checks of adopted models

The diagnostic checks was performed by using Ljung Box, Box Pierce and McLeod-Li test were used for testing the white noise residual. The hypothesis null is that residual should be white noise [28, 29]. It means the residual should be independent, homoscedastic (having constant variance), and normaly distributed. It can reject hypothesis null if p-value in Chi Square statistics is greater than alpha at 5 per cent level of confidence.

The normality test and white nose test results for ARIMA (111,111)[31] and ARIMA (101,000)[31] models are presented in [Table-6] and [Table-7], respectively. It is observed from above tables that, alpha p-value for both ARIMA models were found less than alpha level. So, the null hypothesis cannot be rejected and it can be concluded that, the residual is significant as white noise for both models.

The selected ARIMA models were evaluated for their performance during the diagnostic stage. The two models i.e ARIMA(111,111)[31] and ARIMA(101,000)[31] are primarily identified. The diagnose was made on the basis of time series plot of ACFs, PACFs, normality and white noise test using Ljung-Box, Box-Pierce and McLeod-Li approaches. It is observed from above test that the diagnostic models are performing well in forecasting the runoff. The statistical parameter adopted to diagnose the statistical performance by root mean square error and mean absolute percent error is presented in [Table-8]. It is revealed from the [Table-8] that ARIMA(101,000)[31] model have low RMSE and MAD i.e 8.26 t/ha/year and 0.267 per cent, respectively compared to other model. Hence, the ARIMA (101,000)[31] was model found more appropriate in forecasting the sediment yield for Kal river.

Table-6 *Normality test and white noise tests of ARIMA (111,111)[31] sediment yield modeling for Kal river*

Statistic	DF	α,Value	p-value
Jarque-Bera	2	40.53	< 0.0001
Box-Pierce	6	3538.76	< 0.0001
Ljung-Box	6	3562.67	< 0.0001
McLeod-Li	6	4354.59	< 0.0001
Box-Pierce	12	3538.76	< 0.0001
Ljung-Box	12	3562.67	< 0.0001
McLeod-Li	12	8858.52	< 0.0001

Table-7 *Normality test and white noise tests of ARIMA (101,000)[31] sediment yield modeling for Kal river*

Statistic	DF	α, Value	p-value
Jarque-Bera	2	25.46	< 0.0001
Box-Pierce	6	3126.38	< 0.0001
Ljung-Box	6	3156.61	< 0.0001
McLeod-Li	6	3231.49	< 0.0001
Box-Pierce	12	6059.76	< 0.0001
Ljung-Box	12	6150.51	< 0.0001
McLeod-Li	12	6287.41	< 0.0001

Performance of ARIMA models

In the present study the runoff data of Kal river were used for sediment yield modelling using ARIMA model from June 2003 to 2009. The data series from 2003 to 2007 (5 years) were used for diagnose analysis of the model whereas data of time period 2007 to 2009 (2 years) were used for forecasting the sediment yield using ARIMA model through SPSS 16.1 and XLSTAT-2014 software. The sediment yield generated in adopted ARIMA models were checked for their statistical performance during diagnostic period and forecasting period. The performances of the models were evaluated using regression coefficient (R). root mean square error (RMSE), coefficient efficiency (CE), volumetric efficiency (EV), mean average deviation (MAD) and mean average percentage error (MAPE) and are presented in [Table-8] for diagnostic and forecasting period. The R values of ARIMA (111)(111)[31], ARIMA (101,000)[31] adopted ARIMA models during diagnostic period were 0.986, and 0.988, respectively whereas RMSE was 8.69 tones/day and 8.27 tones/day, respectively. The CE, EV, MAD and MAPE for ARIMA (111,111)[31] were 98.95 per cent, 0.19, 53.48 and 86.32 per cent respectively and for ARIMA(101,000)[31] they were 99.05 per cent, 0.97, 0.267, and 154.85 per cent respectively. Hence, it concluded that during diagnostic period the ARIMA model performance was found very good to generate the runoff for all selected models. But ARIMA (111,111)[31] was found to perform better in terms of MAPE, RMSE, EV, MAD and CE as compared to ARIMA (101,000)[31] model.

In the same sequence performance of selected ARIMA models were evaluated during forecasting period and presented in [Table-8]. From the above table it is observed that adopted models performing comparatively satisfactorily. The R values for ARIMA (111,111)[31] and ARIMA (101,000)[31] was 0.976 and 0.978, respectively. The RMSE, CE, EV, MAD and MAPE for ARIMA (111,111)[31] model were 96.40 tones/day, 95.24 per cent, 0.186, 0.399 and 54.54 per cent, respectively and for ARIMA (101,000)[31] they were 35.94 tones/day, 95.36 per cent, 0.093, 0.199, and 123.976 per cent, respectively. It reveals that ARIMA (111,111)[31] is comparatively found performances better to the ARIMA (101,000)[31]. The similar results regarding the statistical performance were reported by Kachroo (1986) [30] for forecasting sediment yield using ARIMA model.

Validation of ARIMA models

The sediment yield forecasted during diagnostic period and forecasting period were compared and validated with observed sediment yield by plotting the scatter plot and time series plots. The scatter plot of estimated and observed sediment yield during diagnostic period and forecasting period for ARIMA (101,000)[31] and ARIMA (111,111)[31] are presented in [Fig-6] to [Fig-9]. It is interpreted that the estimated and observed sediment yield closely matches and R^2 values are more than 0.95 during diagnostic period and forecasting period. It is also observed that model perform well for low sediment yield load as compared to large sediment yield load in all the cases. The R^2 for ARIMA (111,111)[31] during diagnostic and forecasting period was observed as 0.977 and 0.957, respectively, whereas for ARIMA (101,000)[31] it is 0.973 and 0.952, respectively.

The residual error plot for ARIMA (101,000)[31] is presented in [Fig-10] and [Fig-11] for diagnostic and forecasting period. It is observed that ARIMA (101,000)[31] shows minimum error during diagnostic and forecasting period. The results of ARIMA models with original observed runoff series were plotted and presented in [Fig-12] to [Fig-15] for ARIMA (101,000)[31] and ARIMA(111,111)[31] for a diagnostic and forecasting period respectively. It is observed form the [Fig-12] and [Fig-15] that ARIMA (101,000)[31] performed better in prediction and forecasting the runoff over the ARIMA (111,111)[31]. Hence ARIMA (101,000)[31] was selected for further forecasting the stream flow of Kal river.

Table-8 *Sensitivity analysis of selected ARIMA models for Kal river to estimated the sediment during diagnostic period*

Sr. No	ARIMA Model	R	RMSE	CE	EV	MAD	MAPE
A)	During Diagnostic period (2003-2007)						
1	ARIMA (101,000)[31]	0.98	8.27	99.05	0.97	0.267	154.85
2	ARIMA (111,111)[31]	0.98	8.69	98.95	0.19	53.48	86.32
B)	During forecasting (2008-2009)						
1	ARIMA (101,000)[31]	0.98	35.94	95.36	0.09	0.19	122.97
2	ARIMA (111,111)[31]	0.98	96.40	95.24	0.18	0.39	54.54

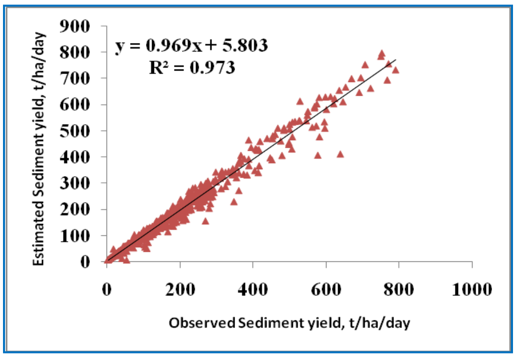

Fig-6 Scatter plot of observed and estimated sediment yield during diagnostic period of ARIMA (101,000)[31] for Kal river

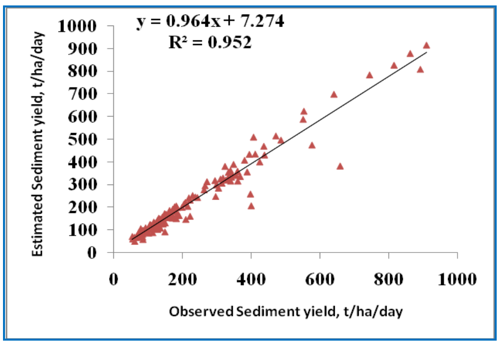

Fig-7 Scatter plot of observed and estimated sediment yield during forecasting period of ARIMA (101,000)[31] for Kal river

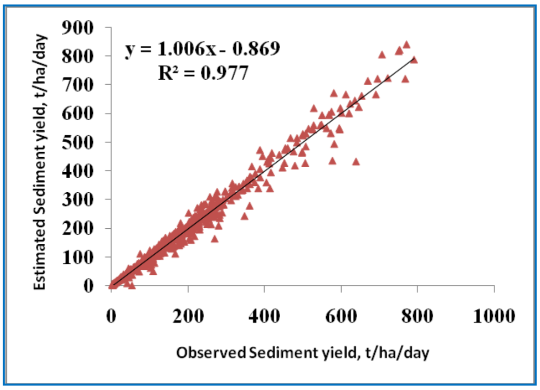

Fig.-8 Scatter plot of observed and estimated sediment yield during diagnostic period of ARIMA (111,111)[31] for Kal river

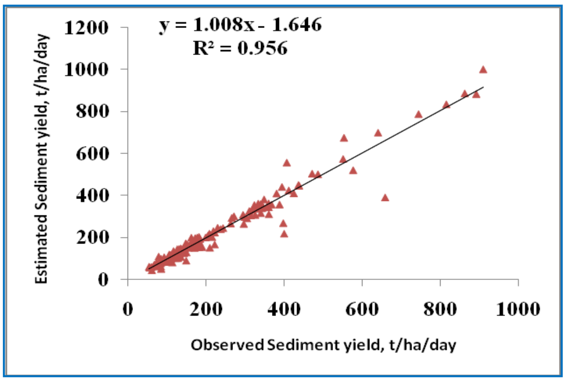

Fig-9 Scatter plot of observed and estimated sediment yield during forecasting period of ARIMA (111,111)[31] for Kal river

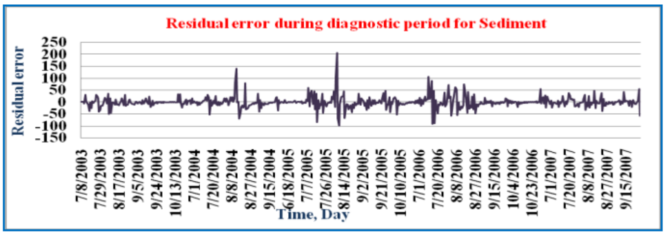

Fig-10 Residual error plot of ARIMA(101,000)[31] during diagnostic period of sediment yield for Kal river

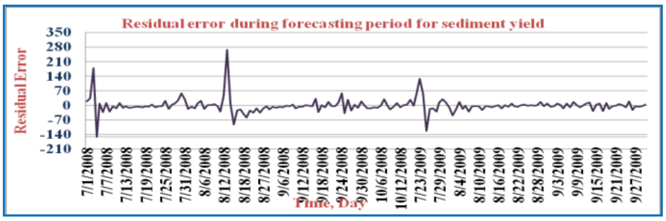

Fig-11 Residual error plot of ARIMA(101,000)[31] during forecasting period of sediment yield for Kal river

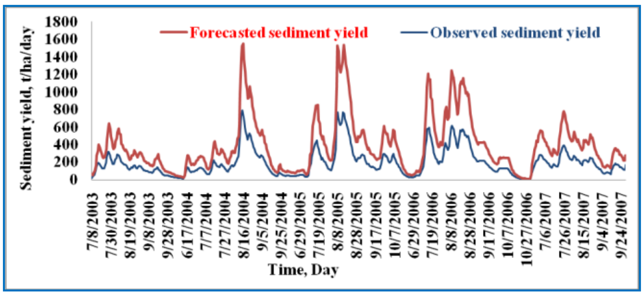

Fig-12 Comparative plot of observed and estimated sediment yield of ARIMA(101,000)[31] for Kal river during diagnostic period

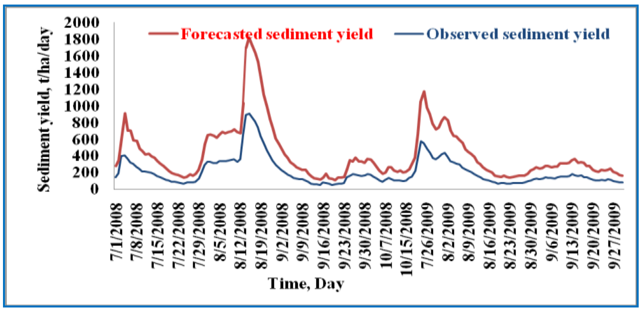

Fig-13 Comparative plot of observed and estimated sediment yield of ARIMA (101,000)31 for Kal river during forecasted period

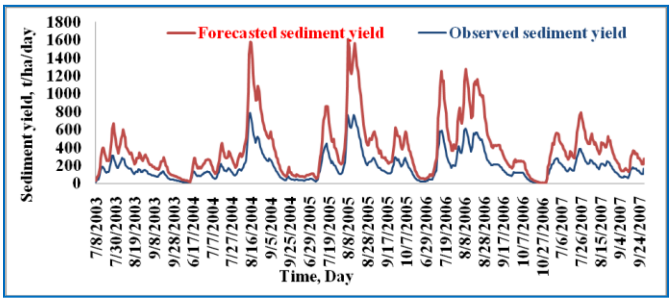

Fig-14 Comparative plot of observed and estimated sediment yield of ARIMA (111,111)31 for Kal river during diagnostic period

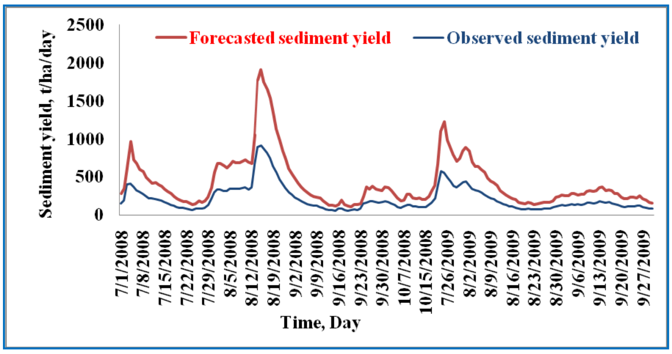

Fig-15 Comparative plot of observed and estimated sediment yield of ARIMA (111,111)31 for Kal river during forecasted period

Conclusions

The stochastic based seasonal ARIMA model was developed for forecasting sediment yield of Kal river a tributary of Savitri basin using data for period of 7 years (2003 to 2009) by adopting standard statistical procedure in SPSS software. The seasonal ARIMA(111,111)31 and ARIMA(101,000)31 were identified and found more appropriate in forecasting sediment yield on seasonal basis with R value were approaches to 0.95, RMSE were approaching to minimum value, CE were more than 95 per cent, CE were nearer to zero and minimum MAPE were observed for both model during diagnostic and forecasting period. The ARIMA model developed for forecasting the sediment yield were found to be more appropriate on the basis of statistical indices and observed time series data series. Hence, developed can be adopted for forecasting sediment yield for Konkan region of Maharashtra using the estimated parameters.

Conflict of Interest: None declared

References

[1] Yang C.T. (2002) 'Sediment Transport Modeling-Combination of Theoretical and Practical Approach, In: Modeling Erosion, Sediment Transport and Sediment Yield, Eds. Summer, W. and Walling, D.E., IHP-VI Technical Document in Hydrology, No. 60, UNESCO.

[2] Mahesh Kothari and Gharde K.D. (2015) *J. of Earth System Science.*, ISSN-0253-4126. 124(5), 933-943.

[3] Gharde K.D., Mahesh Kothari, Mittal H.K., Singh P.K., and Dahiphale P.A. (2015) Sediment Yield Modeling of Kal river In Maharashtra

[4] Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams. J. R., (1998) *J. of the Am. Water Resources Association*. 34, 73-89.

[5] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) *J. Hydrol. Eng. ASCE.* 5(2),115-123.

[6] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000b) *J. Hydrol. Eng. ASCE.* 5(2),124-137.

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

[8] Abbott M.B., Bathurst Cunge J.A., Connell P.E. and Rasmussen J. (1986b) *J. of Hydrology*. 87, 61-77.

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

[10] Davidson J.W., Savic D.A. and Walters G.A. (2002 *J. Inform. Sci*. 15(12), 95-117.

[11] Gorman J. W. and Toman R. J. (1966) *Technometrics*, 8(1), 27-51.

[12] 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.

[13] Galeati G. (1990) *Research Journal of Recent Sciences, ISSN 2277-2502.* 4(ISC-2014), 120-130.

[14] Schwars G. (1978) *Annals of Statistics,* 6, 461–464.

[15] 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. *J of River Engineering*. 1(1):1-8.

[16] Nigam Rashmai, Nigam Sudhir and Kapur Sangita (2013) *Int. J. of pure and Applied Research in Engineering and Technology*, 1(7),13-19

[17] Maidment David R. (1993) Handbook of Hydrology, McGraw-Hill, New York.

[18] Gangyan Z., Goel N.K. and Bhatt V.K. (2002) *Hydrological Sciences Journal,* 47, 93-105

[19] Salas J.D., Delleur J.W., Yevjevich V. and Lane W.L. (1980) *Water Resources*, Littleton. 484.

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

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

[22] Turkey J. W. (1957) *Annals of Mathematical Statistics,* 28,602-632.

[23] Ljung G.M. and Box. G. E. P. (1978) *Biometrika,* 65(2), 279-303.

[24] Abraham B. and Ledoltor J. (1983) Statistical Methods for forecasting. *John Wiley and Sons Inc.* New York pp:472.

[25] Nash J.F. and Sutcliffe. J.V. (1970) *J. hydrol Sci*., 44, 399-417

[26] Janacek G. and Swift L. (1983) Time series forecasting, simulation, Application. Ellis Horwood. New York. USA.

[27] 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).

[28] Ayob K. and Amat. S. D. (2004) *J. of Technology*., 41(B), 47-56.

[29] Mohammad M.E., Nadhir AL-ANSARI, Sven KNUTSSON (2012) Application of SWAT Model to Estimate the Runoff and Sediment Load from the Right Bank Valleys of Mosul Dam Reservoir. ICSE-6 Paris - August 27-31.

[30] Kachroo R.K. (1986) HOMS Workshop on River Flow Forecasting, Nanjing, China, Unpublished Internal Report, Dept. of Engg. Hydr., University College Galway, Ireland.