Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2582-2845.7916

ISSN: 2582 – 2845

Ind. J. Pure App. Biosci. (2020) *8*(1), 16-25

Research Article

Indian Journal of Pure & Applied Biosciences

Peer-Reviewed, Refereed, Open Access Journal

A Time Series Analysis of Wastewater Inflow of Sewage Treatment Plant in Hubballi, India

Mallu B. Deshetti* and G. N. Kulkarni

Department of Agricultural Economics, UAS, Dharwad-580005

*Corresponding Author E-mail: mallu.deshetti@gmail.com

Received: 10.12.2019 | Revised: 17.01.2020 | Accepted: 24.01.2020

ABSTRACT

A time series stochastic Models have been used to analyze the inflow rate of wastewater to the sewage treatment plant (STP) of Gabbur village at Hubballi city. Based on the daily inflow data of 214 days (April 2019 to October 2019), many possible combinations of the orders 'p' and 'q' were made with the differencing one (d=1). On the basis of diagnostic check, ARIMA (0, 1, 3) was selected which has a combination of significant R – square value of 0. 380 and a least Normalized Bayesian information Criterion (BIC) value of 1.715. Linear regression model applied to the observed inflow and the predicted values of inflow obtained by the ARIMA model showed positive linear correlation. Forecasted inflow rate was high for 284 days, which infers that the future designs for STP may need modification to accommodate the high inflow and since the series has no seasonal trend, an average inflow may also occur for some days.

Keywords: *Wastewater, Sewage treatment plant, ARIMA*

INTRODUCTION

The safe treatment of sewage constitutes a huge responsibility; therefore Government has set up Sewage Treatment Plants (STP) based on the generation of waste, population and the respective area. The sewage treatment progresses slowly and it can be done efficiently, if it is planned according to the inflow changes of raw sewage. Hence the most important task in wastewater treatment is to monitor the variations in the quantity of inflow into STP. Variables like varied climatic conditions, population rise etc will affect the inflow rate of wastewater. Therefore forecasting of sewage inflow is necessary to determine the average and peak flow rates, which help in planning the size of collection and treatment facilities of STP for future conditions. Forecasting wastewater inflow is based on the current observed values of inflow recorded at regular intervals of time. Time series analysis of wastewater inflow into treatment plants is done in recent years (Ayesha Sulthana et al., 2013). Studies have also been carried out to forecast the inflow rate of reservoir using time-series model (Mays & Tung 1992). Box-Jenkins seasonal multiplicative models were fit to monthly inflow of Bekhme reservoir (Ali, 2009).

Cite this article: Deshetti, M.B. & Kulkarni, G.N. (2020). A Time Series Analysis of Wastewater Inflow of Sewage Treatment Plant in Hubballi, India, *Ind. J. Pure App. Biosci.* *8*(1), 16-25. doi: http://dx.doi.org/10.18782/2582-2845.7916

In a catchment, forecasting analysis was performed using the Box-Jenkins Autoregressive model for predicting inflows (Sales et al., 1994, Tao et al., 1994). A best fitted Autoregressive Integrated Moving Average (ARIMA) model is one which can give accurate prediction values to achieve success in controlling and planning of wastewater treatment in future. The rainfall forecasting was successively done by ARIMA modelling approach (Momani, & Naill, 2009). It consists of an integrated component (d), which performs differencing of the time series to make it stationary (Hosking 1981, Pankratz, 1983). Another two components are autoregressive (p) and moving average (q), AR component correlates the relation between the current value and the past value of time series. The moving average captures the duration of random shock in the series (Box et al., 1994). In the present study, best fitting ARIMA model for the time series inflow data of sewage treatment plant is determined.

Background of Box and Jenkins Model

Box-Jenkins model generate forecast values based on the statistical parameters of observed time–series data, these models have gained a remarkable attention in the field of operation research, management science and statistics. They are also known as Autoregressive Integrated Moving Average (ARIMA) models (Box and Jenkins 1976).

Autoregressive Models

The observed values Zt of time series are considered to be the outputs of an unobservable process (black box process); the input values at of this process are called independent random shocks. In this model, the observed value may depend upon previous outputs and inputs.

$$Zt = ø1zt\text{-}1 + ø2zt\text{-}2 + \ldots\ldots + øp\ Zt\text{-}p + at \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

In the above Autoregressive model of order “p”, the value of current output Zt (Observed value) depends upon the prior outputs “p” and the current inputs “at” (independent random shock). It is denoted by the notation AR (p).

Moving Average Models

In the Moving Average model of order “q”, the current output Zt (Observed Value) depends on the current input and prior inputs “q”. It is denoted by the notation MA (q).

$$Zt = at - θ1at\text{-}1 \text{-} θ2at\text{-}2 \text{-}\ldots\ldots \text{-} θq\ at\text{-}q \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

Mixed Autoregressive and Moving Average (ARMA) Models

Autoregressive Moving Average Model (ARMA) of order (p, q) involves elements of both AR and MA processes.

$$Zt = ø1zt\text{-}1 + ø2zt\text{-}2 + \ldots\ldots + øpZt\text{-}p + at - θ1at\text{-}1 \text{-} θ2at\text{-}2 \text{-}\ldots \text{-} θq\ at\text{-}q \ldots\ldots (3)$$

Autoregressive Integrated Moving Average Models (ARIMA)

The Box-Jenkins models require a stationary time series data; therefore a non-stationary data is always transformed to induce mean stationarity. A difference of order one leads to the subtraction of each observed value with the neighboring value, which gives the new time series. Hence “d” is referred as the order of differencing to achieve stationarity.

$$Yt = Zt - Zt\text{-}1 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

After applying the ARMA model to the differenced time series, the differencing transformation is reversed to reclaim the original values obtained by the modelled values and “integration” (“d” times) is done. A process in which the dth order differencing is involved is called an integrated process of order d, it is denoted by the notion I (d). A combination of AR, MA and I models is called an ARIMA (p, d, q) model of order (p, d, q).

Background of the present study

Hubballi-Dharwad is presently maintained by HDMC. HDMC has constructed presently 2 sewage treatment plants at Gabbur village in Hubballi and Madihal village of Dharwad, based n the topography of the twin cities. The underground sewerage system connected to the STP covers about 36 percent of the total road length of Hubballi Township. The efficiency of the sewage treatment is affected by the frequency and increase in wastewater inflow. Therefore predicting the inflow changes is necessary to have the anticipatory control over the sewage treatment systems to manage the waste generated by the population growth. Many researchers have applied different formulas, other empirical models and physical laws to forecast the sewage inflow. To forecast the sewage inflow of Gabbur STP of Hubballi city; the successful ARIMA model is developed

Data

To perform Time Series Analysis and forecasting of inflow of waste water, the recorded 214 days of daily inflow data which was read by the flow meter on hourly interval basis (From April 2019 to October 2019) was collected from the STP. Average of daily inflow was calculated and the obtained time series was used for further analysis.

Model Development

ARIMA model used in this study consists of the following steps: Identification, Estimation, Diagnostic checking and Forecasting. The model was estimated using the software SPSS Statistics 20. The equation of ARIMA model of order (p, d, q)

$$Yt = c + a1\ Yt\text{-}1 + a2\ Yt\text{-}2 + \ldots + apYt\text{-}p + q1ut\text{-}1 + q2ut\text{-}2 + \ldots + qqut\text{-}q + ut \quad \ldots\ldots\ldots(5)$$

Where Yt is Inf low of STP, 'U*t*' are independently and normally distributed with zero mean and constant variance σ 2 for t = 1,2,..., n, '*a*'and'*q*' are the coefficients to be estimated. If *Yt* is non-stationary, first-difference of *Yt* is taken so that *ΔYt* becomes stationary.

$$\Delta Yt = Yt - Yt\text{-}1$$

(d = 1 implies one time differencing) The equation of ARIMA model of order (p, 1, q)

$$\Delta Yt = c + a1\ \Delta Yt\text{-}1 + a2\ \Delta Yt\text{-}2 + \ldots + ap\Delta Yt\text{-}p + q1ut\text{-}1 + q2ut\text{-}2 + \ldots + qqut\text{-}q + ut \ldots\ldots\ldots\ldots \quad (6)$$

Model Performance Criteria

Many performance criterions like R-Square, Stationary R-Square Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC) were used to select the best fitting ARIMA model.

RESULTS

Model Identification and Estimation

A graph was plotted for 284 days daily inflow data of STP to check the stationary. The data was found to be non-stationary, shown in Figure 1. Therefore, the first order differencing of the series was carried out. The obtained differenced data was examined for stationary in mean by computing the autocorrelation and partial autocorrelation coefficients (ACF and PACF) for various orders of Yt. After examining Table 1, Figure 2 and 3, it was concluded that Yt was stationary in mean. Various orders of 'p' and 'q' were tried with the difference of one (d=1) to select the best fitting ARIMA model. Among the various ARIMA models, the best fitting model was chosen based on the high Stationary R-Square value, Good R- Square value and low values of RMSE, MAPE and Normalized BIC.The best suitable model for inflow rate of STP was found to be ARIMA (0, 1, 3), with the low normalized BIC value and good R Square. The parameters of selected model are given in Table 2.

Diagnostic Checking

The model was verified by examining the residuals ACF and PACF of various orders, which indicated the "good fit" of the model (Figure.4). Autocorrelations up to 24 lags were computed and their significance was tested by Box-Ljung statistic (Table 3).

Autocorrelations up to 24 lags were computed and their significance was tested by Box-Ljung statistic (Table 3). At none of the reasonable level any of the autocorrelation was not significantly different from zero. Therefore, this concludes that the selected ARIMA (0, 1, 3) model is the best fitted model for inflow rate of Gabbur STP.

The linear regression model was applied to the observed inflow and predicted inflow values of ARIMA model, there was no much variation in the mean of observed and predicted data (Table 4). The correlation coefficient of predicted inflow was 0.793, which suggests a positive linear correlation. The coefficient of determination obtained was

0.629 (Table 5), therefore about 63% of the variation in the predicted inflow is explained by the observed inflow.

The Normal P-P Plot of Regression Standardized Residual showed a random scatter of the points with a constant variance without any outliers. Since the points are closer to the diagonal line (Figure 5), it is understood that the residuals are approximately normally distributed.

Forecasting

The best fitted ARIMA (0,1,3) was used to forecast the inflow rate till 284 days, the inflow values obtained showed increase in inflow, which predicts that the excess inflow in STP may interrupt with the collection and treatment facilities of the plant. The forecasted values are tabulated in Table 6; the observed and predicted values with the confidential limits are shown in the Figure.6.

Table 1: ACF and PACF of Daily Inflow Data

Autocorrelation			Box-Ljung Statistic			Partial Autocorrelation		
Lag	**Value**	**Std. Error**	**Value**	**df**	**Sig**	**Lag**	**Value**	**Std. Error**
1	-0.452	0.068	44.121	1	0.000	1	-0.452	0.69
2	0.052	0.068	44.707	2	0.000	2	-0.191	0.69
3	-0.181	0.068	51.842	3	0.000	3	-0.312	0.69
4	0.125	0.068	55.253	4	0.000	4	-0.153	0.69
5	-0.059	0.067	56.008	5	0.000	5	-0.150	0.69
6	0.069	0.067	57.072	6	0.000	6	-0.075	0.69
7	-0.032	0.067	57.305	7	0.000	7	-0.041	0.69
8	-0.83	0.067	58.854	8	0.000	8	-0.176	0.69
9	0.118	0.067	61.982	9	0.000	9	-0.009	0.69
10	-0.58	0.067	62.739	10	0.000	10	-0.049	0.69
11	0.093	0.066	64.682	11	0.000	11	0.053	0.69
12	-0.088	0.066	66.437	12	0.000	12	0.028	0.69
13	-0.011	0.066	66.463	13	0.000	13	-0.053	0.69
14	-0.029	0.066	66.661	14	0.000	14	-0.060	0.69
15	0.075	0.066	67.966	15	0.000	15	-0.019	0.69
16	-0.063	0.066	68.895	16	0.000	16	-0.088	0.69

Table 2: ARIMA model summary

	Model Fit statistics						**Ljung-Box Q(18)**			**Number of outliers**
Best fit Modal	**Stationary R- squared**	**R-squared**	**RMSE**	**MAPE**	**MaxAE**	**Normalized BIC**	**Statistics**	**DF**	**Sig.**	
	0.353	0.380	2.242	33.609	8.103	1.715	14.722	15	0.472	0

Table 3: Residual of ACF and PACF of daily inflow of STP

Lag	ACF		PACF	
	Mean	SE	Mean	SE
Lag 1	-0.015	0.069	-0.015	0.069
Lag 2	-0.009	0.069	-0.009	0.069
Lag 3	0.075	0.069	-0.075	0.069
Lag 4	0.096	0.069	0.094	0.069
Lag 5	0.011	0.070	0.012	0.069
Lag 6	0.071	0.070	0.069	0.069
Lag 7	-0.023	0.070	-0.008	0.069
Lag 8	-0.048	0.070	-0.055	0.069
Lag 9	0.098	0.070	0.107	0.069
Lag 10	-0.006	0.071	-0.022	0.069
Lag 11	0.060	0.071	0.059	0.069
Lag 12	-0.086	0.071	-0.069	0.069
Lag 13	-0.073	0.071	-0.094	0.069
Lag 14	-0.038	0.072	-0.024	0.069
Lag 15	0.034	0.072	-0.008	0.069
Lag 16	-0.042	0.072	-0.037	0.069
Lag 17	0.046	0.072	0.061	0.069
Lag 18	-0.098	0.072	-0.095	0.069
Lag 19	-0.074	0.073	-0.066	0.069
Lag 20	0.015	0.073	0.010	0.069
Lag 21	0.065	0.073	0.050	0.069
Lag 22	0.037	0.073	0.069	0.069
Lag 23	0.034	-0.115	0.061	0.069
Lag 24	0.074	0.074	-0.106	0.069

Table 4: Descriptive Statistics

	Mean	Std. Deviation	N
Predicted	7.46	1.953	213
Observed	7.45	2.82	213

Table 5: Linear Regression Model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	0.803	0.645	0.643	1.209	0.645	383.335	1	211	0.000	0.376

Table 6: Forecasted Inflow rate of STP

Period	Observed Inflow	Predicted Inflow	Lower Confidential Limit	Upper Confidential Limit
DAY 1-20	12.37-1.87	8.31-0.00	3.85- -0.30	12.77-0.00
DAY 21-40	7.26-2.74	5.93-4.41	1.52-0.00	10.34-8.82
DAY 41-60	6.54-1.22	5.41-3.86	1- -0.55	9.82-8.26
DAY 61-80	8.83-2.83	7.14-4.46	2.73-0.05	11.55-8.87
DAY 81-100	12.45-2.73	9.25-5.84	4.60-1.43	13.41-10.25
DAY 101-120	10.54-5.13	8.86-6.51	4.45-2.11	13.27-10.92
DAY 121-140	15.58-0.64	10.65-5.05	6.25-0.64	15.06-9.46
DAY 141-160	12.79-5.33	10.95-9.17	6.54-4.76	15.36-13.58
DAY 161-180	11.38-7.53	10.52-9.25	6.11-4.60	14.93-13.41
DAY 181-200	12.13-1.19	10.28-6.51	5.87-2.10	14.69-10.92
DAY 201-214	16.63-1.19	11.07-6.51	6.66-2.10	15.48-10.92
DAY 215-234	--------	8.67-8.21	3.89-3.40	13.94-12.71
DAY 235-254	--------	8.92-8.68	3.38-3.06	14.77-13.99
DAY 255-274	--------	9.16-8.93	3.05-2.78	15.54-14.81
DAY 275-284	--------	9.28-9.17	2.76-2.65	15.92-15.58

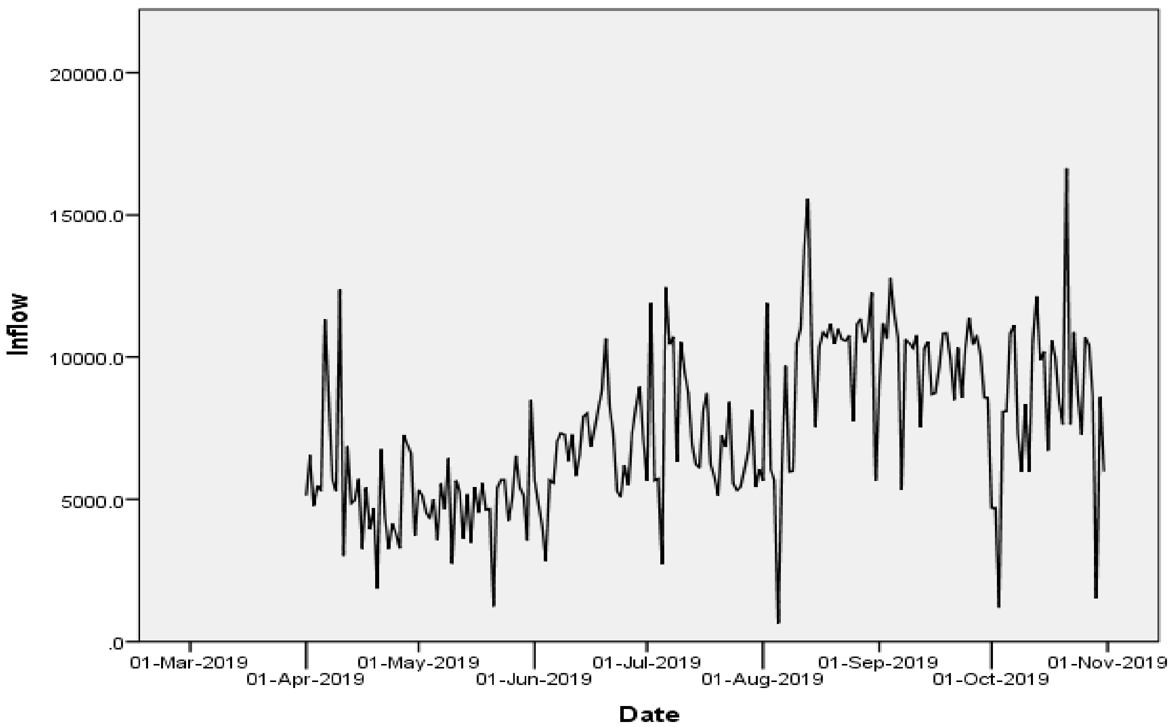

Fig. 1: Time plot of daily inflow in STP

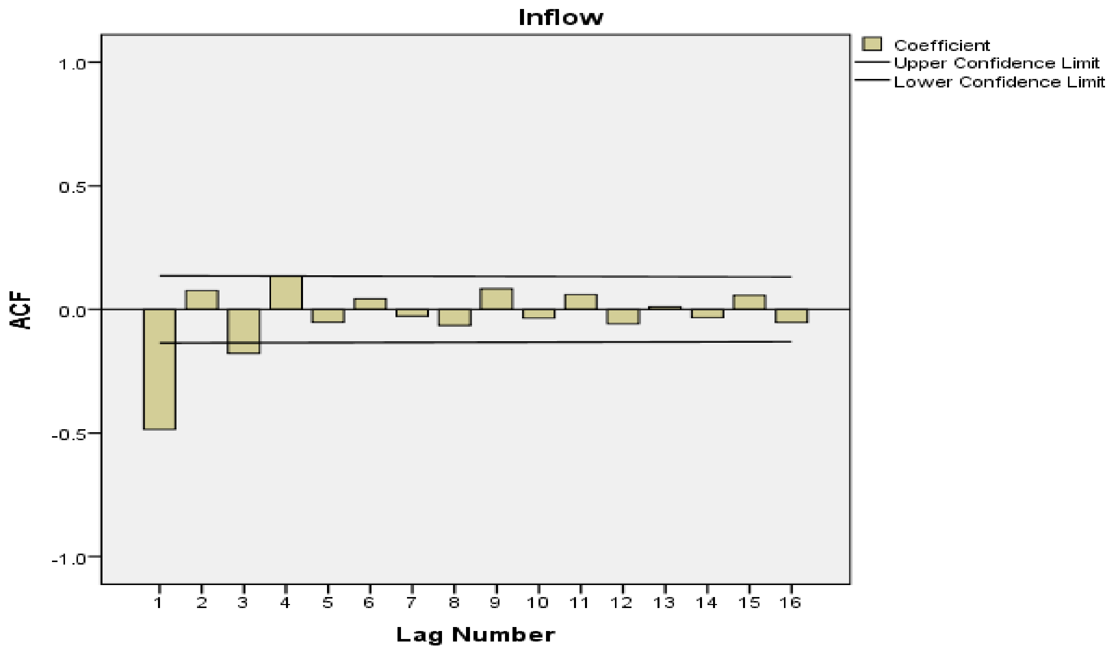

Fig. 2: ACF of differentiated daily inflow data

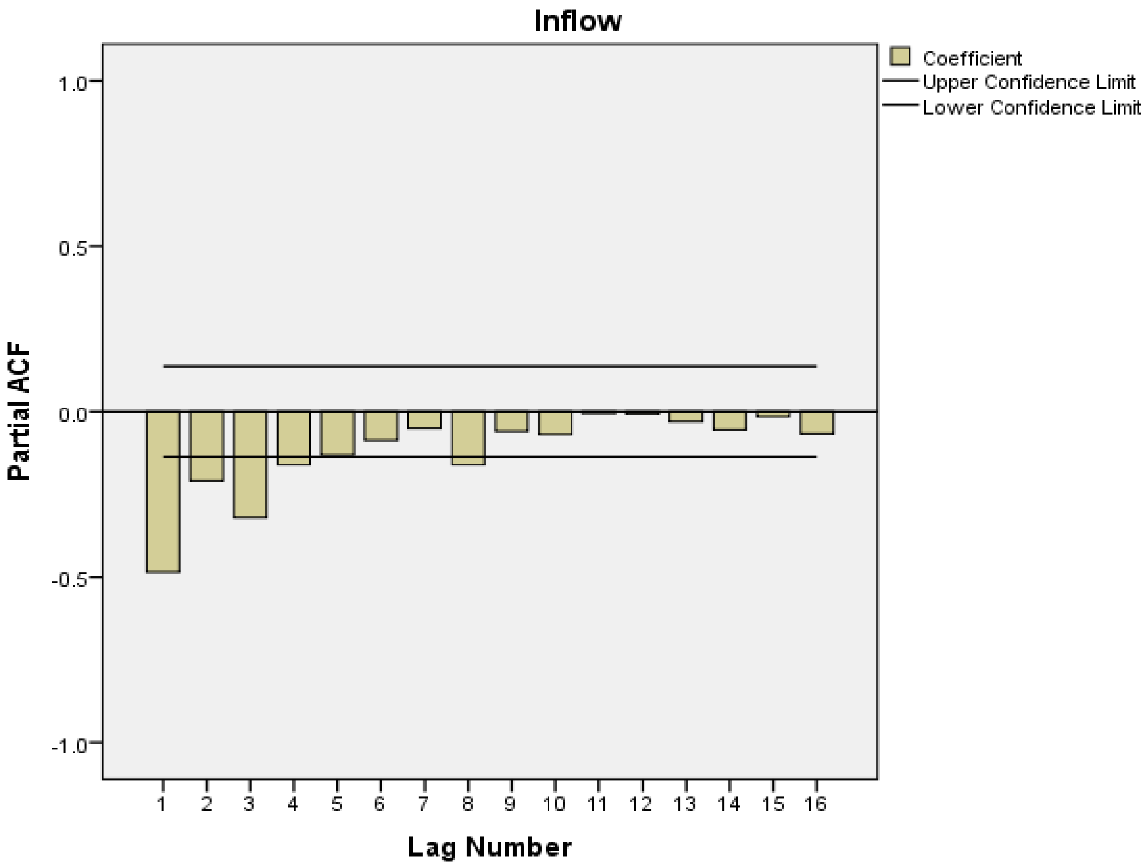

Fig. 3: PACF of differentiated daily inflow data

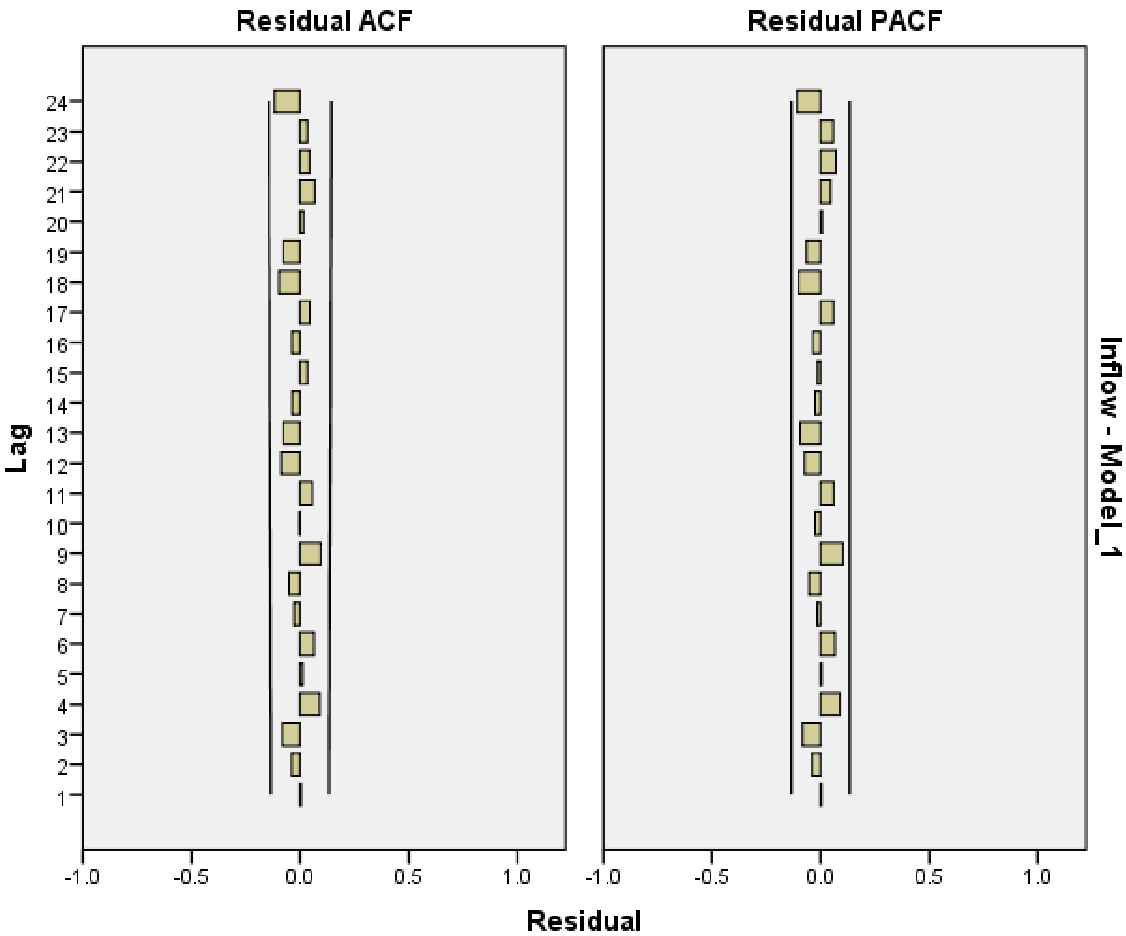

Fig. 4: Residuals ACF and PACF

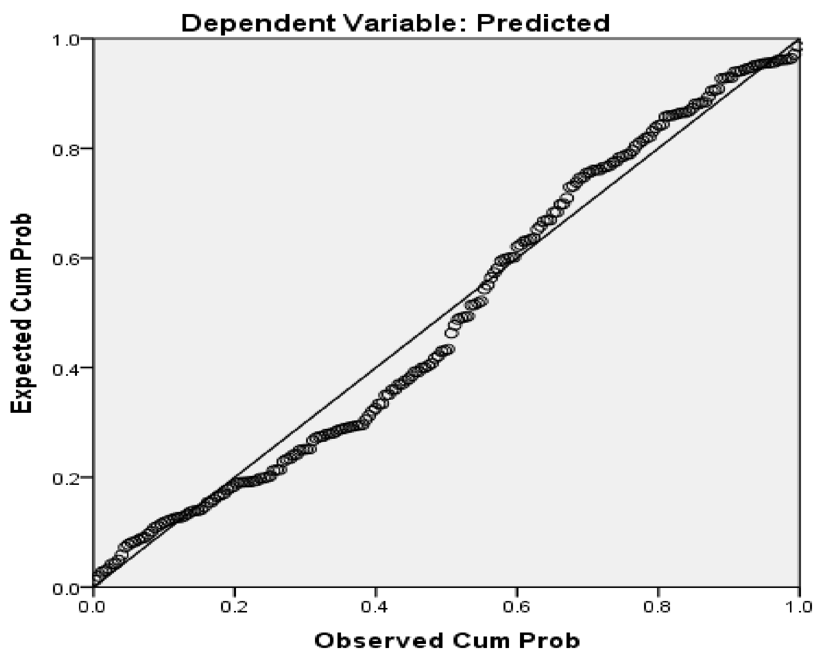

Fig. 5: Normal P-P Plot of Regression standardized Residual

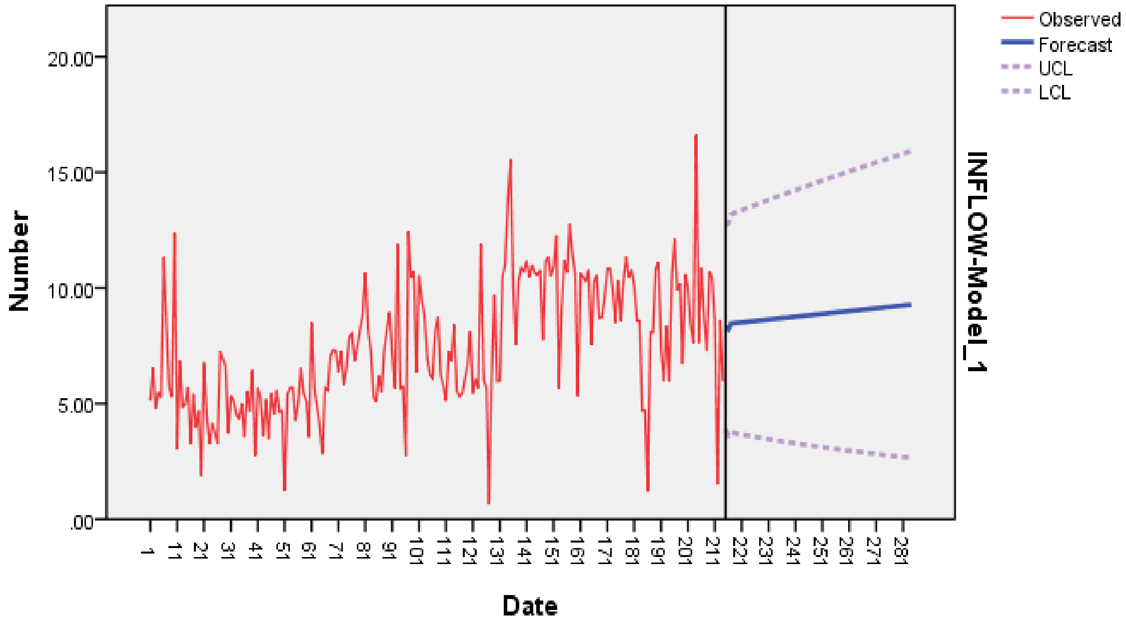

Fig. 6: Forecasted Model

CONCLUSION

This study describes the importance of time series ARIMA modelling for the planning of sewage inflow into the treatment systems. The sewage inflow data of Gabbur STP was found to be nonstationary, which was transformed to first order differencing to make it stationary. Eight ARIMA models of various orders of 'p' and 'q' were applied on to the transformed data to select the best fitted model. Based on the diagnostics like high R – square value and low Normalized Bayesian information Criterion, ARIMA (0, 1, 3) was found to be the best fitted model. The linear regression model applied to the observed and predicted values showed the approximately similar mean with positive linear correlation. By this linear regression analysis it was understood that there was no much variation between the observed and predicted data.

This study will help in analyzing the variations in the sewage inflow of STP. One must estimate the design flow for short term and long term wastewater treatment to monitor the sewage load. The estimation of inflow and model are based on the hand book values of STP, the aim of the work is to increase the forecasting efficiency and decrease the error by utilizing the inflow data of 214 days of Gabbur STP. The best fitted ARIMA (0, 1, 3) model forecasted the increase in inflow up to 40 MLD on the 284th day; therefore this study can be considered for future design planning of Gabbur STP of Hubballi to treat the influent waste efficiently by saving time, energy and cost.

Acknowledgements

The authors wish to thank the Gabbur Sewage Treatment Plant for supplying the flow meter data used in this study. The authors also would like to thank Hubli-Dharwad Municipal Corporation for granting the permission to collect data from STP used in this study.

REFERENCES

Ayesha, S., Balsubramanian, S., & Latha, K. C. (2013). A time series analysis of wastewater inflow of sewage treatment plant in Mysore, India. *Inter. J. Current Res.*, *5*(2), 248-253

Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (1994). *Time Series Analysis; Forecasting and Control*, Pearson Education, Delhi.

Chuchro, M. (2010). The structure of influent time series in wastewater treatment plants Environmental Engineering III, Edited by Artur Pawłowski, CRC Press 2010, Pages 167–172, Print ISBN: 978-0-415-54882-3.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika 68*(1), 165-176.

Mays, L.W., & Tung, Y. (1992). *Hydro-systems Engineering and Management*, McGraw-Hill Inc, USA.

Momani, M., & Naill, P. E. (2009). Time series model for rainfall data in Jordan: Case study for using time series analysis; *American J. Env. Sci. 5*(5), 599–604.

Pankratz, A. (1983). Forecasting with univariate Box- Jenkins models: concepts and cases, John Wiley, New York.

Sales, P. R., Pereira H. D. B., & Vieira, A. M. (1994). 'Linear Procedures for Time-series Analysis in Hydrology'. In Stochastic and Statistical Methods in Hydrology and Environmental Engineering, The Netherlands: Kluwer Academic Publishers, *3*, 105 - 117.

Tao, T., Corbu, I., Penn, R., Benzaquen, F., & Lai, L. (1994). 'Seeking User input in Inflow Forecasting'. In Time Series Analysis in Hydrology and Environmental Engineering, Netherlands: Kluwer Academic Publishers, *3,* 99 - 104.