Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 5, Issue, 02, pp.248-253, February, 2013

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

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

***Ayesha Sulthana, Balasubramanian S., and Latha K. C.**

Department of Water and Health, JSS University, Mysore, Karnataka

ARTICLE INFO

Article History:
Received 20th November, 2012
Received in revised form
14th December, 2012
Accepted 21st January, 2013
Published online 14th February, 2013

Key words:

Sewage Inflow,
Times series,
ARIMA,
Linear Regression.

ABSTRACT

Stochastic Models have been used to analyze the inflow rate of wastewater to the sewage treatment plant (STP) of Southern Mysore. Based on the daily inflow data of 217 days (November 2011 to June 2012), many possible combinations of the orders 'p' and 'q' were made with the differencing one (d=1). On the basis of diagnostic check, ARIMA (1, 1, 2) was selected which has a combination of significant R – square value of 0. 899 and a least Normalized Bayesian information Criterion (BIC) value of 1.681. 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 300 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.

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 influent into STP. Variables like varied climatic conditions, population rise, vacations and tourist inflow 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 (Chuchro, 2010). Studies have also been carried out to forecast the inflow rate of reservoir using time-series model (Mays and Tung 1992). Box-Jenkins seasonal multiplicative models were fit to monthly inflow of Bekhme reservoir (Ali, 2009). 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 modeling approach (Momani, and 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).

***Corresponding author:** ayeshasulthanaa@gmail.com

In the present study, best fitting ARIMA model for the time series inflow data of sewage treatment plant is determined.

Background and Overview of Box and Jenkins Model

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

$$Z_t = ø_1 z_{t-1} + ø_2 z_{t-2} + \ldots\ldots + ø_p Z_{t-p} + a_t \quad \ldots\ldots (1)$$

In the above Autoregressive model of order "p", the value of current output Z_t (Observed value) depends upon the prior outputs "p" and the current inputs "a_t" (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 Z_t (Observed Value) depends on the current input and prior inputs "q". It is denoted by the notation MA (q).

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots\ldots - \theta_q a_{t-q} \quad \ldots\ldots (2)$$

Mixed Autoregressive and Moving Average (ARMA) Models

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

$Z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots\ldots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots. - \theta_q a_{t-q}$ ……..(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.

$Y_t = Z_t - Z_{t-1}$ ……………….. (4)

After applying the ARMA model to the differenced time series, the differencing transformation is reversed to reclaim the original values obtained by the modeled values and "integration" ("d" times) is done. A process in which the d^{th} 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

The urban development authorities in India have constructed many Sewage Treatment Plants to treat the generated sewage. Accordingly Mysore Urban Development Authority (MUDA) has constructed three Sewage Treatment Plants at Rayankere, Vidyaranyapuram and Kesare, based on the Topography of the city. However, the efficiency of sewage treatment is affected by the frequency and increase quantity of the waste water inflow. Therefore predicting the inflow changes is necessary to have the anticipatory control over the wastewater treatment systems to manage the waste generated by the population growth. Many researchers have applied different formulas, physical laws and other empirical models to forecast the sewage inflow. To forecast the sewage inflow of Vidyaranyapuram STP of Mysore; the successful ARIMA model is developed.

Study Area

Mysore lies in the Southern Plateau of Karnataka. It covers an area of around 128.42 Sq. Km located between latitude 11°45' to 12°40' N and longitude 75°57' to 77°15' E. Mysore city was one among the earliest cities in India to have underground drainage systems. The Mysore City Corporation has three sewage treatment plants for the four divided drainage districts, named A, B, C and D. Sewage treatment plant at Rayankere for the drainage district A and D, Vidyaranyapuram sewage treatment plant for the drainage district B and the Kesare sewage treatment plant for the drainage district C.
The Sewage treatment plant (STP) of Vidyaranyapuram in Mysore was selected as a study area (Figure. 1). This STP covers an area of around 27.21 Sq. km between latitude of 12.284361 and longitude of 76.652760, with a sewer length of 7000 meters. The STP was constructed in 2002 and more than 50% of the sewage generated in Mysore is received by Vidyaranyapuram STP. The capacity of STP is 67.65 MLD comprising facultative aerated lagoons with sedimentation basins.

Data

To perform Time Series Analysis and forecasting of inflow of waste water, the recorded 217 days of daily inflow data which was read by the flow meter on hourly interval basis (From November 2011 to June 2012) 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)

$$Y_t = c + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + q_1 u_{t-1} + q_{2ut-2} + \ldots + q_q u_{t-q} + u_t \quad \ldots\ldots\ldots\ldots(5)$$

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 Y_t is non-stationary, first-difference of Y_t is taken so that ΔY_t becomes stationary.

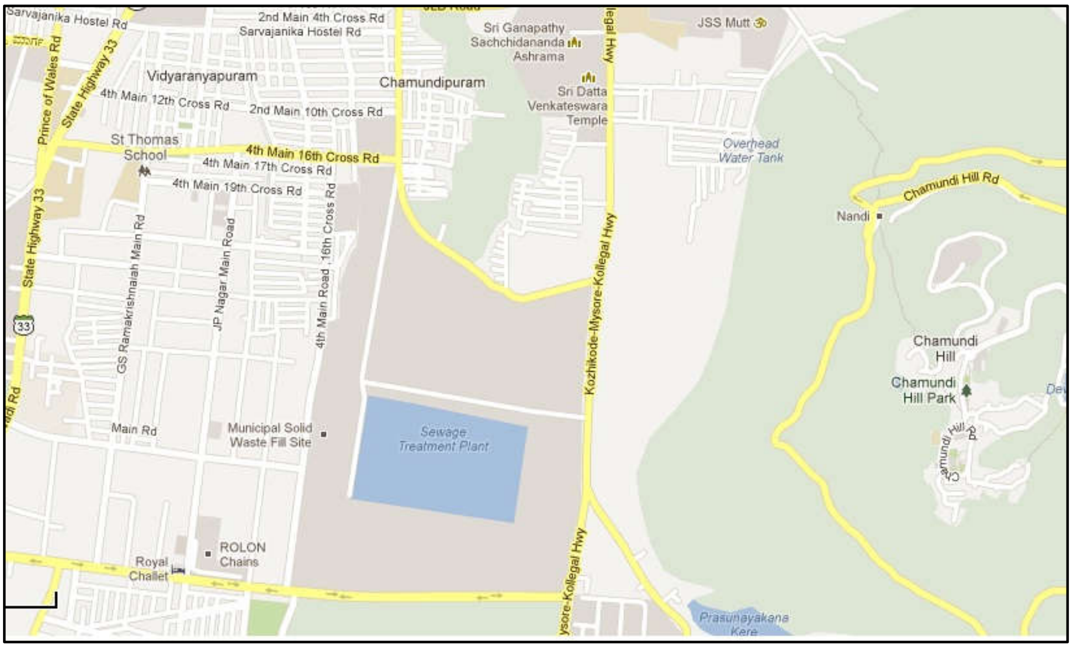

Figure 1. Map of Study area (From Google Maps)

$\Delta Y_t = Y_t - Y_{t-1}$
(d = 1 implies one time differencing)
The equation of ARIMA model of order (p, 1, q)

$$\Delta Y_t = c + a_1 \Delta Y_{t-1} + a_2 \Delta Y_{t-2} + \ldots + a_p \Delta Y_{t-p} + q_1 u_{t-1} + q_2 u_{t-2} + \ldots + q_q u_{t-q} + u_t \ldots\ldots\ldots\ldots(6)$$

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.

Pearson's correlation coefficient (R)

$$\frac{\sum_{t=1}^{N}[Q\text{obs}\ (t) - \bar{Q}\text{obs}]\ [Q\text{est}\ (t) - \bar{Q}\text{est}]}{\sqrt{\sum_{t=1}^{N}[Q\text{obs}\ (t) - \bar{Q}\text{obs}]^2[Q\text{est}\ (t) - \bar{Q}\text{est}]^2}}$$

Root mean square error (RMSE) RMSE

$$\sqrt{\frac{\sum_{i=1}^{n}(Q\text{obs}\ (t)\ - Q\text{est}\ (t))^2}{n}}$$

Where $Q_{obs}(t)$ is the observed inflow at time 't' and $Q_{est}(t)$ is the estimated inflow at time 't'. N is the total number of daily inflow data points. $\bar{Q}_{obs}$ is the mean observed inflow and $\bar{Q}_{est}$ is the mean estimated inflow.

Bayesian information criterion (BIC)
BIC = m in (RMSE) + n in (m)
Where "m" is the number of input-output patterns in RMSE and "n" is the number of parameters to be estimated in "m".

Linear Regression model was fit to the predicted inflow of the selected ARIMA model and observed inflow to investigate the variation between them; this was taken as another criterion to assess the predicted values generated by the selected model.

RESULTS

Model Identification and Estimation

A graph was plotted for 217 days daily inflow data of STP to check the stationarity. The data was found to be non-stationary, shown in Figure 2. Therefore, the first order differencing of the series was carried out. The obtained differenced data was examined for stationarity in mean by computing the autocorrelation and partial autocorrelation coefficients (ACF and PACF) for various orders of Y_t. After examining Table 1, Figure 3 and 4, it was concluded that Y_t 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 (Table 2).The best suitable model for inflow rate of STP was found to be ARIMA (1, 1, 2), with the low normalized BIC value and good R-Square. The parameters of selected model are given in Table 3 and Table 4.

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	-0.348	0.068	26.557	1	0	1	-0.348	0.068
2	-0.144	0.067	31.149	2	0	2	-0.302	0.068
3	0.004	0.067	31.153	3	0	3	-0.208	0.068
4	-0.05	0.067	31.708	4	0	4	-0.232	0.068
5	0.176	0.067	38.651	5	0	5	0.03	0.068
6	-0.171	0.067	45.175	6	0	6	-0.16	0.068
7	0.087	0.067	46.897	7	0	7	0.004	0.068
8	0.002	0.066	46.898	8	0	8	-0.013	0.068
9	-0.095	0.066	48.955	9	0	9	-0.094	0.068
10	0.145	0.066	53.73	10	0	10	0.056	0.068
11	-0.107	0.066	56.339	11	0	11	-0.034	0.068
12	0.002	0.066	56.34	12	0	12	-0.06	0.068
13	0.068	0.066	57.428	13	0	13	0.041	0.068
14	-0.079	0.065	58.876	14	0	14	-0.041	0.068
15	0.064	0.065	59.832	15	0	15	-0.008	0.068
16	-0.02	0.065	59.929	16	0	16	0.036	0.068

Table 2. Suggested Models

Model (p, d, q)	Stationary R-Squared	R-Squared	RMSE	MAPE	Normalized BIC
(0,1,1)	0.251	0.896	2.227	3.083	1.651
(0,1,2)	0.265	0.898	2.211	3.078	1.662
(1,1,0)	0.120	0.878	2.419	3.301	1.817
(1,1,1)	0.262	0.898	2.215	3.078	1.665
(1,1,2)	0.273	0.899	2.205	3.074	1.681
(2,1,0)	0.202	0.888	2.316	3.222	1.755
(2,1,1)	0.270	0.899	2.211	3.075	1.687
(2,1,2)	0.274	0.899	2.209	3.070	1.710

Table 3. ARIMA model summary

	Model Fit statistics						Ljung-Box Q(18)			Number of Outliers
	Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	Normalized BIC	Statistics	DF	Sig.	
Best Fit Model	0.273	0.899	2.205	3.074	15.978	1.681	12.074	15	0.673	0

Table 4. Model Statistics

		Estimate	SE	t	Sig.
Coefficients	Constant	0.001	0.001	1.648	0.101
	AR Lag 1	-0.651	0.199	-3.27	0.001
	Difference	1			
	MA Lag 1	-0.074	0.179	-0.415	0.679
	Lag 2	0.563	0.108	5.209	0

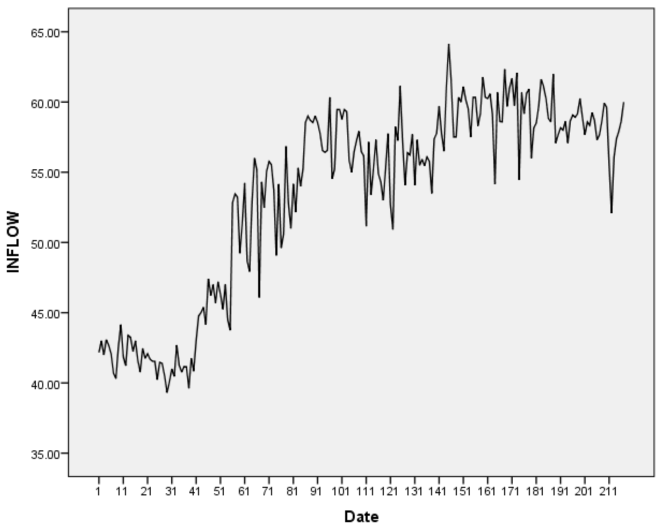

Figure 2. Time plot of Daily Inflow in STP

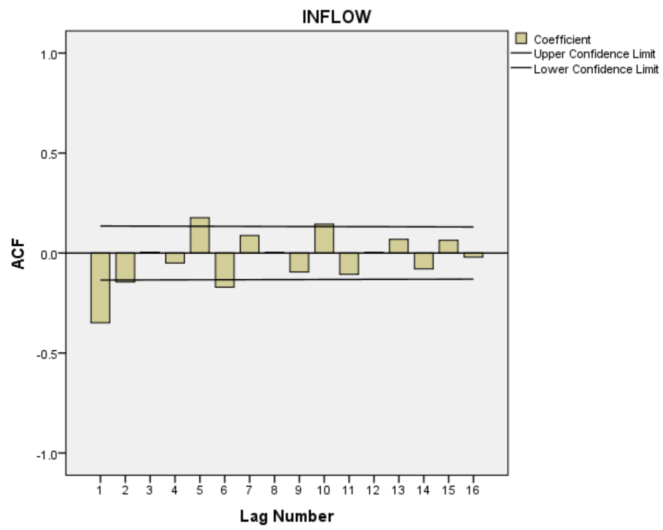

Figure 3. ACF of Differenced Daily Inflow Data

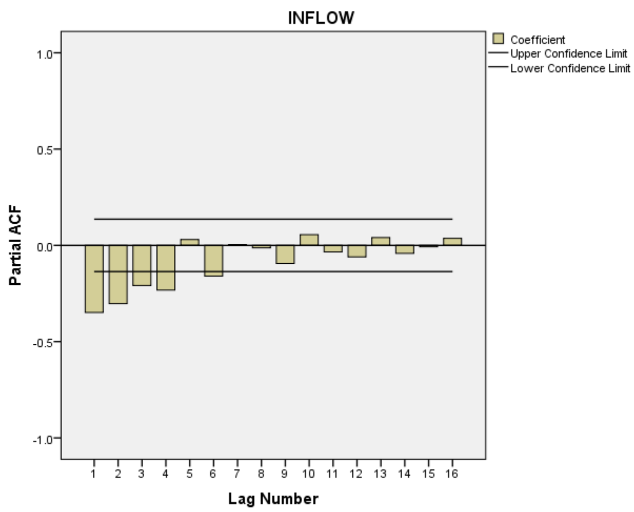

Figure 4. PACF of Differenced Daily Inflow Data

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.5). Autocorrelations up to 24 lags were computed and their significance was tested by Box-Ljung statistic (Table 5).

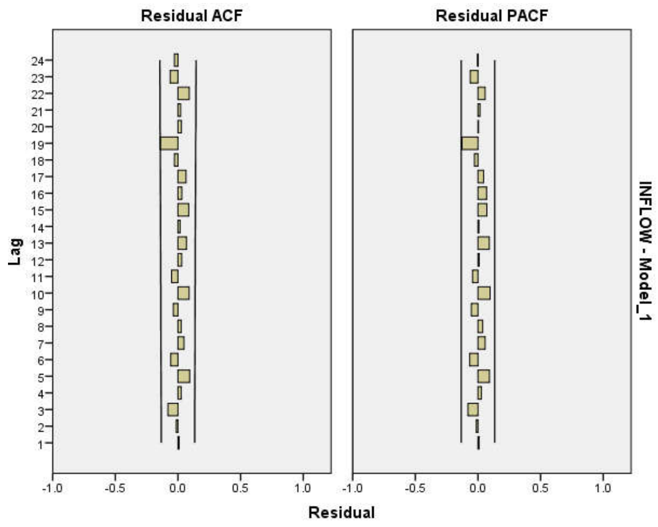

Figure 5. Residuals of ACF and PACF

Autocorrelations up to 24 lags were computed and their significance was tested by Box-Ljung statistic (Table 5). At none of the reasonable level any of the autocorrelation was not significantly different from zero. Therefore, this concludes that the selected ARIMA (1, 1, 2) model is the best fitted model for inflow rate of Vidyaranyapuram STP.

$$Y_t = 0.001 - 0.651Y_{t-1} - 0.074u_{t-1} + 0.563\ u_{t-2} + u_t$$

Table 5. Residual of ACF and PACF of Daily Inflow of STP

	ACF		PACF	
Lag	Mean	SE	Mean	SE
Lag 1	0.009	0.068	0009	0.068
Lag 2	-0.013	0.068	-0.013	0.068
Lag 3	-0.080	0.068	-0.080	0.068
Lag 4	0.026	0.069	0.027	0.068
Lag 5	0.094	0.069	0.092	0.068
Lag 6	-0.057	0.069	-0.065	0.068
Lag 7	0.048	0.070	0.057	0.068
Lag 8	0.025	0.070	0.038	0.068
Lag 9	-0.037	0.070	-0.053	0.068
Lag 10	0.089	0.070	0.097	0.068
Lag 11	-0.051	0.070	-0.043	0.068
Lag 12	0.030	0.070	0.009	0.068
Lag 13	0.067	0.071	0.090	0.068
Lag 14	0.017	0.071	0.007	0.068
Lag 15	0.086	0.071	0.071	0.068
Lag 16	0.030	0.071	0.068	0.068
Lag 17	0.065	0.071	0.045	0.068
Lag 18	-0.028	0.072	-0.028	0.068
Lag 19	-0.143	0.072	-0.128	0.068
Lag 20	0.028	0.073	0.004	0.068
Lag 21	0.021	0.073	0.016	0.068
Lag 22	0.090	0.073	0.057	0.068
Lag 23	-0.062	0.074	-0.062	0.068
Lag 24	-0.028	0.074	-0.002	0.068

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 6). The correlation coefficient of predicted inflow was 0.949, which suggests a strong positive linear correlation. The coefficient of determination obtained was 0.901 (Table 7), therefore about 90.1% of the variation in the predicted inflow is explained by the observed inflow.

Table 6. Descriptive Statistics

	Mean	Std. Deviation	N
PREDICTED	53.47	6.811	215
OBSERVED	53.433	6.89927	215

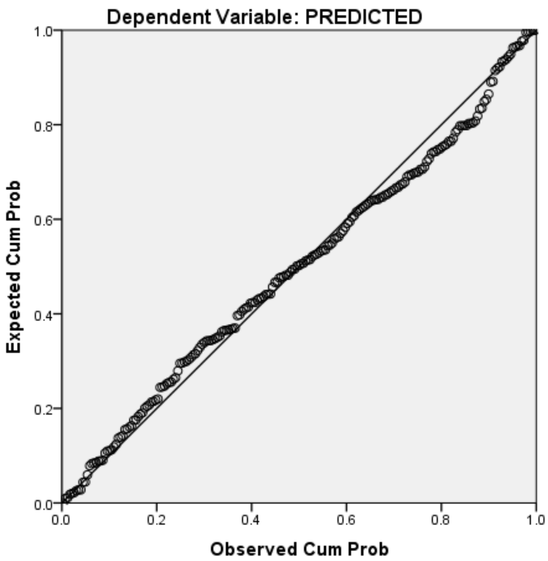

Figure 6. Normal P-P Plot of Regression Standardized Residual

Predicted = 3.401 + 0.937 (Observed) The above regression equation appears to be the useful one for making predictions, since the R-Square value is close to one.

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 close to the diagonal line (Figure 6), it is understood that the residuals are approximately normally distributed.

Forecasting

The best fitted ARIMA (1,1,2) was used to forecast the inflow rate till 300days, 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 8; the observed and predicted values with the confidential limits are shown in the Figure.7.

DISCUSSION

This study describes the importance of time series ARIMA modeling for the planning of sewage inflow into the treatment systems. The sewage inflow data of Vidyaranyapuram STP was found to be non-stationary, 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 (1, 1, 2) was found to be the best fitted model. The linear regression model applied to the observed and predicted values showed the approximately

Table 7. 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	.949[a]	0.901	0.9	2.151	0.901	1933.707	1	213	0	1.877

Table 8. Forecast of Inflow Rate of STP

Period	Observed Inflow	Predicted Inflow	Lower Confidential Limit	Upper Confidential Limit
DAY 1 - 20	42.17 - 41.75	42.28 - 42.39	38.37 - 39.05	46.48 - 45.93
DAY 21 - 41	42.08 - 43.08	42.07 - 40.87	38.75 - 37.65	45.59 - 44.29
DAY 42 - 62	44.77 - 48.69	42.31 - 51.66	38.98 - 47.59	45.85 - 55.98
DAY 63 - 83	47.93 - 55.31	50.53 - 52.68	46.54 - 48.53	54.75 - 57.09
DAY 84 - 104	54.00 - 55.85	54.34 - 58.99	50.06 - 54.34	58.89 - 63.93
DAY 105 - 125	55.00 - 61.14	57.88 - 55.66	53.32 - 51.28	62.73 - 60.32
DAY 126 - 146	57.16 - 61.44	58.36 - 60.65	53.76 - 55.88	63.25 - 65.73
DAY 147 - 167	57.50 - 58.58	60.50 - 58.37	55.74 - 53.77	65.57 - 63.26
DAY 168 - 188	62.33 – 62.00	59.67 - 59.74	54.97 - 55.03	64.66 - 64.74
DAY 189 - 209	57.08 - 59.93	60.84 - 58.69	56.05 - 54.07	65.93 - 63.61
DAY 210 - 216	59.62 - 58.67	59.23 - 57.52	54.57 - 52.99	64.19 - 62.34
DAY 217 - 237	---------------	57.75 - 59.56	53.20 - 51.63	62.59 - 68.36
DAY 238 - 258	---------------	59.65 - 61.51	51.59 - 51.08	68.61 - 73.45
DAY 259 - 279	---------------	61.60 - 63.52	51.06 - 50.92	73.68 - 78.30
DAY 280 - 300	---------------	63.62 - 65.60	50.92 – 51.00	78.53 - 83.10

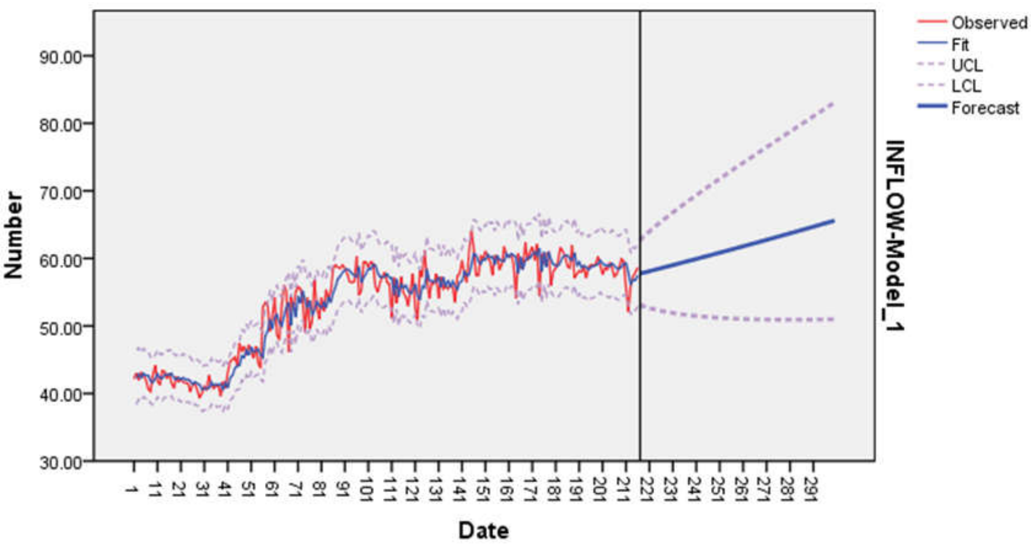

Figure 7. Forecasted Model

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.

Conclusion

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 217 days of Vidyaranyapuram STP. The best fitted ARIMA (1, 1, 2) model forecasted the increase in inflow up to 83 MLD on the 300th day; therefore this study can be considered for future design planning of Vidyararnyapuram STP of Mysore to treat the influent waste efficiently by saving time, energy and cost.

Acknowledgements

The authors wish to thank the Vidyaranyapuram Sewage Treatment Plant for supplying the flow meter data used in this study. The authors also would like to thank Mysore City Corporation for granting the permission to collect data from STP used in this study.

REFERENCES

Ali,S.T.,(2009):"Fitting Seasonal Stochastic Models to Inflow of Bekhme Reservoir", The Iraqi Journal for Mechanical and Material Engineering, Special Issue(A), College of Engineering, University of Babylon.

Box, G.E.P. and Jenkins, 1976. Time Series Analysis; Forecasting and Control, Holden-Day Inc. USA, pp: 574.

Box, G.E.P., G.M. Jenkins and G.C. Reinsel, 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 LW and Tung Y 1992 *Hydro-systems Engineering and Management*, McGraw-Hill Inc, USA.

Momani M and 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. and 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. and 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.

****