International Journal of Innovation in Science and Mathematics
Volume 4, Issue 2, ISSN (Online): 2347–9051

A Comparative Analysis of Forecasting Reservoir Inflow using ARMA Model & Holt Winters Exponential Smoothening Technique

Naresh Vijayakumar
Post Graduate Student, Centre for Water Resources,
Anna University Email ID : nareshvijayakumar92@gmail.com

Soorya Vennila
Assistant Professor, Centre for Water Resources,
Anna University Email ID : soorya@annauniv.edu

***Abstract* -- Statistical analysis in hydrology includes estimating the future frequency or probability of hydrologic events based on the information contained in the past hydrologic records. The present study aims at applying time series analysis using Auto Regression Moving Average (ARMA) model and Holt Winters exponential smoothing technique for reservoir inflow forecast enabling better management of reservoir inflow. The Krishnagiri Reservoir in the state of Tamilnadu, India is selected as the focus area for this study. Using Box Jenkins approach, the model identification for ARMA model was done by Auto Correlation function plot (ACF) and Partial Auto Correlation Function Plot (PACF); then the parameters for the model are estimated by Maximum likelihood estimation method followed by diagnostic checking of estimated value of the model. ARMA (2, 4) shows better statistical parameters of generated annual reservoir inflow. One of the deficiencies of ARMA model is that it is not able to reproduce in a satisfactory manner the trend, seasonality, and periodicity component of a time series. In the present study, that deficiency was overcome by using Holt Winters Exponential Smoothening technique. The statistical parameters of generated reservoir inflow using exponential smoothening replicates the historical observed reservoir inflow with R^2= 0.9659 and the percentage of mean error to 19.3%. Although ARMA model is used as the best fit model for any time series analysis in hydrology, exponential smoothening technique gives a better matching between observed and computed values.**

***Keywords* – Auto Regression Moving Average, Exponential Smoothening, Krishnagiri Reservoir, Reservoir Inflow.**

I. Introduction

The dependence of India's agriculture on the monsoon and its consequent vulnerability has been recognized from the earliest time (Irrigation Commission Report, 1972). Water from rivers, and streams stored in reservoirs and tanks were brought to the agricultural fields to supplement monsoon rainfall through gravity irrigation by the early rulers & British Government in India. After Independence, accelerated efforts were made in the five year plans to increase the irrigated area through construction of massive reservoirs to meet the rising demands of food and fiber of the country (Raghava Rani, 2011 et al).

The application of statistical hydrology in earlier days was restricted to surface water problems, especially for analyzing the hydrologic extremes such as floods and droughts. However, during past three decades, the statistical domain of hydrology with the advent of fast computing technology has broadened to encompass the problems of both surface water and groundwater systems. With such a broad domain, statistics has emerged as a powerful tool for analyzing hydrologic time series. The main aim of time series analysis is to detect and describe quantitatively each of the generating processes underlying a given sequence of observations (Shahin et al., 1993). In hydrology, time series analysis is used for building mathematical models to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records, and to fill in missing data and extend records (Salas, 1993).

A number of successful applications of time series analysis to reservoir inflow forecasting have been reported in the literature review, Lilly (2013) applied time series analysis to generate a monthly synthetic stream flow generation for Krishna River at Srisailam dam site, India. S.K Jain (1999) applied time series analysis using ARMA to model reservoir inflow with a 32 year data. The limitations of the time series analysis using ARMA model was overcome by adding a momentum of noise term, randomized weight space to back propagation algorithm of an Artificial Neural Network to show a better match of observed inflow data with ARMA.

II. Study Objectives

Normally, the Auto Regressive Moving Average model has been used for modeling water resources time series because such models are accepted as standard representations of Stochastic Time series (Maier & Dandy 1997). The objective of the present work is to study the application of time series analysis using ARMA model to Krishnagiri reservoir inflow forecasting, and to overcome the limitations of ARMA model with Exponential Smoothening technique. The objective was attained through the following steps:

1. By determining the Hydrological Characteristics of Reservoir Inflow of Krishnagiri Reservoir Project (KRP) with observed annual data (1958-2010),
2. To Forecast the Reservoir inflow by Auto Regression Moving Average (ARMA) model and verifying the Computed data with the observed data over the years.
3. To use the exponential smoothing technique to overcome the deficiency in the ARMA model.

III. Time Series Modeling

Time series modeling is the analysis of a temporally distributed sequence of data or the synthesis of a model for prediction in which time is an independent variable. Most

International Journal of Innovation in Science and Mathematics
Volume 4, Issue 2, ISSN (Online): 2347–9051

statistical analyses of hydrologic time series at the usual time scale encountered in water resources studies are based on the following fundamental assumptions: the series is homogenous, stationary, and free from trends and shifts, non-periodic with no persistence (Adeloye and Montaseri, 2002). Homogeneity implies that the data in the series belong to one population, and hence have a time invariant mean. Stationarity, on the other hand, implies that the statistical parameters of the series computed from different samples do not change except due to sampling variations. A time series is said to be strictly stationary if its statistical properties do not vary with changes of time origin. A less strict type of stationarity (called weak stationarity or second-order stationarity) is that in which the first- and second-order moments depend only on time differences (Chen & Rao, 2002). In nature, strict stationary time series does not exist, and weak stationary time series is practically considered as stationary time series. A time series is said to have trends, if there is a significant correlation (positive or negative) between the observations and time. In an annual time series, periodicity effect is not discernible and hence half-annual or monthly time series normally encountered in hydrology can be used for analyzing the periodicity. A Systematic Approach to hydrologic time series modeling may be composed of six main phases (Salas & smith, 1980) Identification of model composition, Selection of Model Type, Identification of model form, Estimation of the parameters of model, testing goodness of fit of the model, and Evaluation of uncertainties.

IV. Study Area

Krishnagiri Reservoir Project (KRP) is located near the village Periamuthur in Krishnagiri taluk of Krishnagiri District in northern Tamil Nadu (Fig 1). The dam is constructed of a combination of masonry and earthen portions across Pennaiyar River. This project is planned to irrigate 3647 Ha of command area benefitting 16 villages lying in the Krishnagiri District. Data of monthly inflow into the reservoir over a period of 54 years from 1958–2012, obtained from the Water Resources Department, Government of Tamilnadu, was used for the study to determine the characteristics of hydrology at Krishnagiri Reservoir project and to forecast the reservoir inflow for next 5 years thereby anticipating the dependency of the reservoir inflow for the command area for irrigation and other purposes.

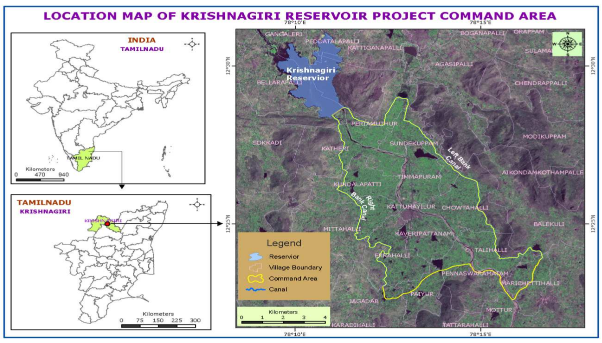

Fig. 1. Study Area Location

V. Materials and Method

A time series is a set of observations generated sequentially in time. The main components of a hydrological time series are: trends and other deterministic changes, cycles or periodic changes, and components representing the stochastic or random variations. Auto Regression moving average (ARMA) models are linear stochastic models. They are analogous to conceptual models of parametric hydrology, based on linear reservoirs. ARMA models cannot be estimated exactly as they are constituted by several random effects. Linear stochastic models are selected to forecast data of one or more time periods ahead, and to generate a synthetic data sequence of the time series (Ghanshyam Das, 2009).

Autoregressive moving average ARMA (p,q) model

A reasonable extension to AR (p) and MA (q) is a mixed model of the form

$$S_t = \beta_{p,1} S_{t-1} + \ldots + \beta_{p,p} S_{t-p} + \varepsilon_t + R_t - \alpha_{q,1} R_{t-1} - \ldots - \alpha_{q,q} R_{t-q}$$
$$= \sum_{k=1}^{p} \beta_k S_{t-k} + R_t - \sum_{k=1}^{q} \alpha_k R_{t-k}$$

Where S_t = Stationary Stochastic Component of the original series

β_k = autoregression model parameter,

k = 1, 2, 3…*p*

α_k= moving average model parameter, k= 1,2,3..*q*

p= order of auto regression process

q= order of moving average process

R_t= Independent random effect at time t of residuals.

The ARMA models are suitable for stationary hydrologic series. If the time series is non stationary, the trend and periodic or seasonal fluctuation can be removed by taking the differences and the ARMA model can be applied to the resultant series. This type of model is termed as the ARIMA model. The main steps of ARMA model application are as follows:

• *Model Identification*.

The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled. Stationarity can be assessed from an autocorrelation plot. Autocorrelation plots are a commonly-used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero. Autocorrelation plots are also used in the model identification stage for fitting MA (q) models. Partial autocorrelation plots are also a commonly used tool for model identification in Box-Jenkins models. The partial autocorrelation at lag *k* is the autocorrelation between X_t and X_{t-k} that is not accounted for by lags 1 through $k-1$. The partial autocorrelation of an AR (*p*) process is zero at lag *p*+1 and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order. We look for the point on the plot where the partial autocorrelations essentially become zero. Placing a 95 % confidence level for statistical significance is helpful for this purpose. To

IJISM Source of Knowledge
International Journal of Innovation in Science and Mathematics
Volume 4, Issue 2, ISSN (Online): 2347–9051

determine the value of p and q we use the graphical properties of the autocorrelation function and the partial autocorrelation function.

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

• Parameter estimation, which involves efficient use of the data to make inferences about parameters conditional on the adequacy of the entertained model. Estimating the parameters for the Box-Jenkins models is a quite complicated non-linear estimation problem. The main approaches to fitting Box-Jenkins models are linear least squares and maximum likelihood estimation. Maximum likelihood estimation is generally the preferred technique.

Estimation of Parameters of an MA (q) series

The theoretical autocorrelation function in terms the parameters of an MA (q) process is given by

$$\rho_h = \begin{cases} \dfrac{\alpha_h + \alpha_1\alpha_{h+1} + \cdots + \alpha_{q-h}\alpha_q}{1+\alpha_1^2+\alpha_2^2+\cdots+\alpha_q^2} & 1 \le h \le q \\ 0 & h > q \end{cases}$$

To estimate a_1, $a_2 \ldots$ a_q we solve the system of equations:

$$r_h = \frac{\hat{\alpha}_h + \hat{\alpha}_1\hat{\alpha}_{h+1} + \cdots + \hat{\alpha}_{q-h}\hat{\alpha}_q}{1+\hat{\alpha}_1^2+\hat{\alpha}_2^2+\cdots+\hat{\alpha}_q^2} \quad 1 \le h \le q$$

This set of equations is non-linear and generally very difficult to solve; for $q = 1$ the equation becomes:

$$r_1 = \frac{\hat{\alpha}_1}{1+\hat{\alpha}_1^2}$$

$$\left(1+\hat{\alpha}_1^2\right) r_1 - \hat{\alpha}_1 = 0$$

or

$$r_1\hat{\alpha}_1^2 - \hat{\alpha}_1 + r_1 = 0$$

Thus, this equation has the two solutions,

$$\hat{\alpha}_1 = \frac{1}{2r_1} \pm \sqrt{\frac{1}{4r_1^2} - 1}$$

One solution will result in the MA (1) time series being invertible

For $q = 2$ the equations become:

$$r_1 = \frac{\hat{\alpha}_1 + \hat{\alpha}_1\hat{\alpha}_2}{1+\hat{\alpha}_1^2+\hat{\alpha}_2^2} \quad r_2 = \frac{\hat{\alpha}_2}{1+\hat{\alpha}_1^2+\hat{\alpha}_2^2}$$

Estimation of parameters of an ARMA (p,q) series we use a similar technique as MA (q). That is, obtain an expression for ρ_h in terms $\beta_1, \beta_2, \ldots, \beta_p$; $\alpha_1, \alpha_1, \ldots, \alpha_q$ of and set up $q + p$ equations for the estimates of $\beta_1, \beta_2, \ldots, \beta_p$; $\alpha_1, \alpha_2, \ldots, \alpha_q$ by replacing ρ_h by r_h.

In the ARMA (1, 1) process the expression for ρ_1 and ρ_2 in terms of β_1 and α_1 are:

$$\rho_1 = \frac{\left(1+\alpha_1\beta_1\right)\left(\alpha_1+\beta_1\right)}{1+\alpha_1^2+2\alpha_1\beta_1}$$

$$\rho_2 = \rho_1\beta_1$$

Further,

$$\sigma^2 = Var\left(u_t\right) = \frac{1-\beta_1^2}{1+\alpha_1^2+2\alpha_1\beta_1}\sigma_x\left(0\right)$$

Thus the expression for the estimates of β_1, α_1, & σ^2 are:

$$r_1 = \frac{\left(1+\hat{\alpha}_1\hat{\beta}_1\right)\left(\hat{\alpha}_1+\hat{\beta}_1\right)}{1+\hat{\alpha}_1^2+2\hat{\alpha}_1\hat{\beta}_1}$$

$$r_2 = r_1\hat{\beta}_1$$

And

$$\hat{\sigma}^2 = \frac{1-\hat{\beta}_1^2}{1+\hat{\alpha}_1^2+2\hat{\alpha}_1\hat{\beta}_1}C_x\left(0\right)$$

Hence

$$\hat{\beta}_1 = \frac{r_2}{r_1} \quad \text{and}$$

$$r_1\left(1+\hat{\alpha}_1^2+2\hat{\alpha}_1\hat{\beta}_1\right) = \left(1+\hat{\alpha}_1\hat{\beta}_1\right)\left(\hat{\alpha}_1+\hat{\beta}_1\right)$$

or

$$r_1\left(1+\hat{\alpha}_1^2+2\hat{\alpha}_1\frac{r_2}{r_1}\right) = \left(1+\hat{\alpha}_1\frac{r_2}{r_1}\right)\left(\hat{\alpha}_1+\frac{r_2}{r_1}\right)$$

$$\left(r_1-\frac{r_2}{r_1}\right)\hat{\alpha}_1^2 + \left(2r_2-1-\frac{r_2^2}{r_1^2}\right)\hat{\alpha}_1 + \left(r_1-\frac{r_2}{r_1}\right) = 0$$

This is a quadratic equation which can be solved

• Diagnostic check involves checking the fitted model in its relation to the data with the intent to reveal model inadequacies and achieve model improvement. Model diagnostics for Box-Jenkins models is similar to model validation that is, the error term ε_t is assumed to follow the assumption for a stationary univariate process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance and the residuals need to be uncorrelated. If the Box-Jenkins model is a good model for the data, the residuals should satisfy these assumptions. If these assumptions are not satisfied, we need to fit a more appropriate model. That is, we go back to the model identification step and try to develop a better model. Hopefully the analysis of the residuals can provide some clues as to a more appropriate model.

VI. Results & Discussions

Yearly synthetic flow generation model of Reservoir inflow at Krishnagiri reservoir project was developed using 50 years of historical data. Yearly inflow data needs transformation to have normality. Transformation technique used for yearly reservoir inflow for all the years is logarithmic transformation technique. The time series was tested for homogeneity and stationarity using Student t test. The Skewness test of normality being 0.086 and Fillibien test of normality is 0.9922 shows that transformed data exhibits stationarity. Once the data has been checked for normality and stationarity, the next step is to identify the model for fitting the transformed data to forecast; for this we use Autocorrelation plot and Partial Auto Correlation plot to determine the model for ARMA. From the following figures (Fig 2 & 3).

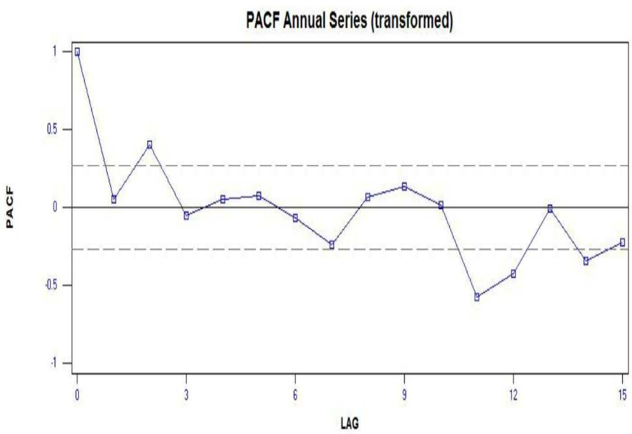

Fig. 1 . Partial Auto Correlation Plot

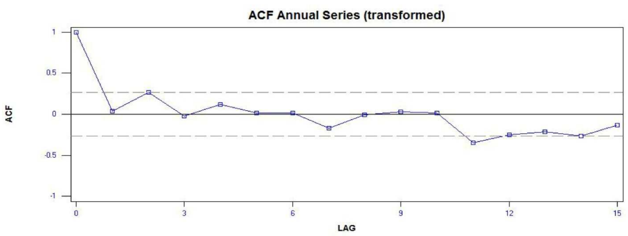

Fig. 2. Auto Correlation Plot

We can relate from the ACF and PACF plots the transformed data tails off after lag 4 in ACF plot and 2 in PACF plot so we choose a model of order ARMA (2, 4). Once the model has been identified we go for parameter estimation for the ARMA (2, 4) model using method of moments (maximum Likelihood estimation). With the estimated parameter for each model, the transformed data is fitted to each model and results were obtained for 58 years in order to verify the accuracy of the model; the residual diagnosis is done to check for following criterion. The residuals need to be uncorrelated, for this an ACF plot for the residual of the model is done and shown in Fig (4).

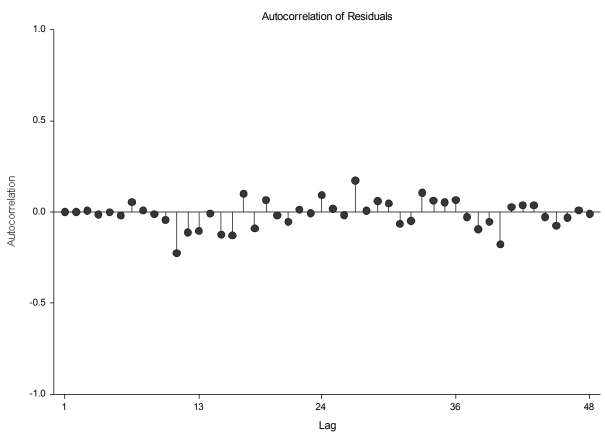

Fig. 3. ARMA (2, 4)

The plot shows that residual estimated by the model exhibit an uncorrelation between them. Secondly, the residuals need to have normality, so a normal distribution plot for the model is shown below. Fig (5)

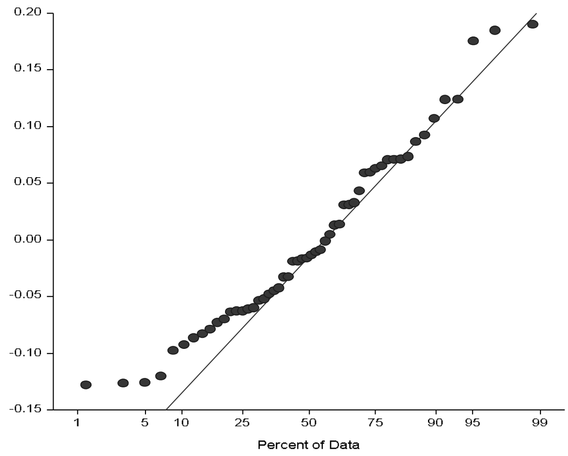

Fig. 5. ARMA (2, 4)

From the study, ARMA (2, 4) generated 58 years reservoir inflow was compared with observed in flow in Figure 6. The fit is not satisfactory.

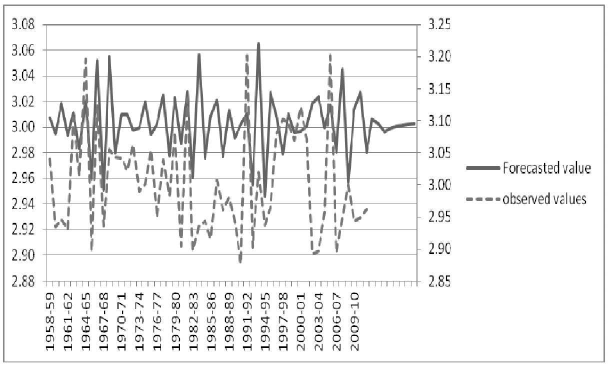

Fig. 6. ARMA (2, 4)

Although the conventional model of time series analysis using ARMA model for forecasting the data is used widely, it has some limitations in modeling the data. Some deficiencies of ARMA model are

1. The model doesn’t account fully for trend, periodicity and seasonality in predicting the observed data and it also eliminates the trend component before modeling the autocorrelation.
2. The models are fundamentally “Backward looking” as such they are generally poor at predicting turning points, unless the turning points represents a return to a long run equilibrium.

VII. Exponential Smoothening Technique

When computed reservoir inflow using ARMA model do not match with the observed inflow, an attempt is made in this paper to reduce the above deficiency with exponential smoothening technique to have a better match. Exponential smoothing is a procedure for continually revising a forecast in the light of more recent experience.

International Journal of Innovation in Science and Mathematics
Volume 4, Issue 2, ISSN (Online): 2347–9051

This is a very popular method to produce a smoothed Time Series. Whereas in Auto Regression Moving Average the past observations are weighted equally, Exponential Smoothing assigns exponentially decreasing weights as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. The exponential smoothening technique has three methods- single, double, triple exponential smoothening. In single exponential smoothening the data fluctuations in the series are smoothened whereas the double smoothening method is used when the data exhibits the trend and periodicity. The periodicity is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. In triple smoothening periodicity, trend and seasonality are smoothened where a third parameter is introduced to smoothen the seasonality; this method is also known as Holt Winters method after the name of inventor. In the present study we use Holt Winters method to smoothen the transformed data for forecasting; there are two models under this method; one is Multiplicative seasonal model and the other is Additive seasonal model. We use these two models and compare the results and evaluate it based on RMSE from the model.

Multiplicative Seasonal Model:

Additive Seasonal Model:

$$y_t = (b_1 + b_2 t) S_t + \varepsilon_t$$

$$y_t = b_1 + b_2 t + S_t + \varepsilon_t$$

Where

b_1 is smoothened observation component

b_2 is a linear trend component

S_t is a multiplicative and Additive seasonal factor

εt is the random error component

The next procedure in determining the parameters for each component is done by the following steps,

[1] Smoothened observation component:

$$b_1 = \alpha\, y_t / (S_t - L) + (1-\alpha)(b_{t-1} + {}_2b_{t-1})$$

[2] Trend Component:

$$b_2 = \gamma\, (b_t - b_{t-1}) + (1-\gamma)\, {}_2b_{t-1}$$

[3] Seasonal factor:

$$S_t = \beta\, (y_t / b_t) + (1-\beta)\, S_{t-L}$$

Where α, β, and γ are constants that must be estimated in such a way that the MSE of the error is minimized and L is the length of time series. In the present study more weight age was given to recent observation; therefore the constant value was 0.999 for α, β, and γ. using the above procedure for model parameter estimation, the transformed data is used in the two models: Multiplicative and Additive seasonal model. The following table shows the summary parameters of each model.

Table 1 Summary of Holt Winters Exponential Smoothening Method

Summary	Multiplicative seasonal model	Additive seasonal model
Mean	3.008	3.002852
R-Squared	0.965593	0.965939
Mean (Error)	0.00585	0.005832289
Mean (Percent Error)	0.1959114	0.1953426

From the above table we get to know that Additive seasonal model has a least error when compared with Multiplicative seasonal model. By using both the multiplicative and the additive seasonal model, the Computed data is compared with the historical data. (Figs 7 & 8)

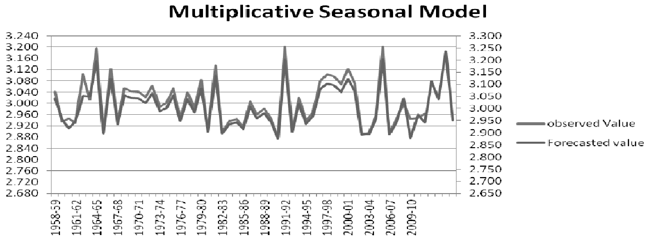

Fig. 7. Multiplicative seasonal model time series graph

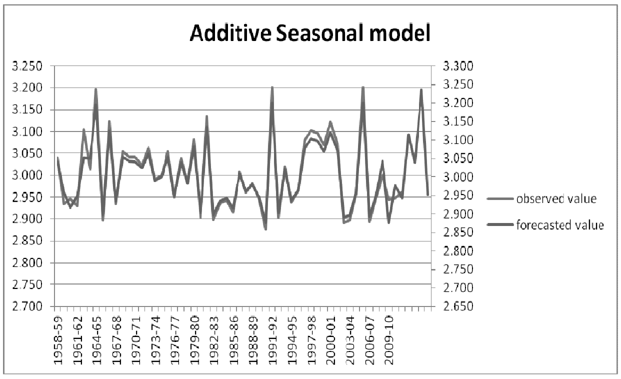

Fig. 8. Additive Seasonal model Time series Graph

From the figure 8, the Computed and observed value of transformed reservoir inflow of Krishnagiri Reservoir project are roughly coinciding indicating that transformed reservoir inflow is firmly smoothened by trend, seasonality, and periodicity by additive seasonal model for forecasting the reservoir inflow for next 4 years (from 2010 -2014).

VIII. Conclusion

Auto Regression Moving average model of order (2, 4) selected is the best fit for yearly reservoir inflow generation of Krishnagiri Reservoir project. The statistical parameters of generated yearly inflow are almost comparable with historical yearly inflow data but the ARMA model doesn’t account trend, seasonality, and periodicity in a satisfactory manner; so we use Exponential smoothing technique where it attempts to estimate the trend as a part of the modeling process whereas ARMA model attempts to eliminate the trend before modeling the auto correlation. In this particular study the Limitation of ARMA model is overcome with Exponential Smoothening Technique. With Additive Seasonal Model of Exponential Smoothening Technique we could obtain the best Computed results for the Reservoir inflow of Krishnagiri Reservoir Project.

IX. ACKNOWLEDGMENT

I express my deep sense of gratitude to **Dr. R. Sakthivaidvel** Emeritus Professor, Centre for Water Resources, Anna University for his valuable guidance and constant encouragement for writing this paper work. I would also like to impart my deepest gratitude for the new suggestions, and constant motivation throughout the project work.

REFERENCES

[1] Chen, H. and Rao, A. (2002) Testing Hydrologic Time Series for Stationarity J. Hydrol. Eng 10.1061/ (ASCE) 1084-0699(2002)7:2(129), 129-136

[2] Government of India (1972) Report of the Irrigation Commission, Ministry of Irrigation and Power, New Delhi, http://finance.bih.nic.in/Documents/Reports/Economic-Survey-2015-EN.pdf. Accessed on 14 September 2015

[3] Ghanshyam Das (2009) Hydrology and Soil Conservation engineering including watershed management, India

[4] Lilly Padmaja Joshi et al (2014) Generation of Monthly synthetic flow Sequences for Krishna River at Srisailam Dam Site, International Conference on Hydrology and Watershed Management, 2:28-34.

[5] Maier, H. and Dandy, G. (1997) Statistical analysis & Time Series Modeling in Hydrology, Water Resources Research 33: 0043-1397

[6] McMahon, T A and Adeloye Adebayo (2004) Water Resources Yield, Water Resources Publications, 2005: 234

[7] Rhagav Rani and Venkateswara Rao (2011) Modernization of an existing Irrigation project by Performance Evaluation Using performance indicators, International E journal of Mathematics and Engineering, 141: 1122-1134

[8] Salas, J. D. and Smith, R.A., (1980). Uncertainties in hydrologic time series analysis Paper presented at the ASCE Spring Meeting, Portland, Oregon, and Preprint 80-158.

[9] Salas, J, D., Delleur, J, W. Yevjevich, V. Lane,W . L. (1985) Applied Modeling of Hydrologic Time series, Water Resources Publications, Colorado, USA, Second Edition: 172-184.

[10] Shahin (1993) Statistical Analysis in Water Resources Engineering, A Rotterdam, Springer, 175: 117-124.

[11] S. k Jain and Das A (1999) Application of ANN Prediction for Reservoir inflow and Operation, Journal of water resources planning and management, 125: 263-271

AUTHOR'S PROFILE

Mr. Naresh Vijayakumar
Post Graduate Student in Integrated Water Resources Management, Centre for Water Resources, Anna University,
Email ID : nareshvijayakumar92@gmail.com,
DOB: 07th August, 1992.

Dr. Soorya Vennila
Assistant Professor, Centre for Water Resources, Anna University, soorya@annauniv.edu, She is Specialized in issues related in water resources management and has been teaching various subjects related to Water resources management ranging from sociological aspects to qualitative & quantitative research.