Inflow Forecasting Analysis for Nagarjunasagar Dam Andhra Pradesh using Time series Statistical models

B. Pradeepakumari, K. Srinivasu

Abstract: ***Krishna river based dams are an important part of socio economic growth of Andhra Pradhesh. Nagarjuna sagar dam is one of them. Historically it has been providing unprecedented value to Andhra Pradesh. Analysis of this dam's history like patterns of water inflow, outflow, capacity of dam at various times is performed here. This detailed analysis provides deeper insight for last fifteen year's of dam data. Seasonal trends in data show how the water is used, and with what pattern dam inflow occurs. Also, auto correlation of various dam variables is calculated to improve forecasting models. Additionally, here data is cleaned, outliers are handled, and missing values are filled. This pretreatment of data leads to better analytical perspectives from data. Such analytical study is helpful in prediction of future trends related this reservoir.***

Keywords: ***Nagarjuna Sagar Dam, Water inflow and outflow, Dam Capacity, Data Analysis.***

I. INTRODUCTION

Krishna river based dams are important part of socio economic growth of Andhra Pradhesh. Nagarjuna sagar dam is one of them. Historically it has been providing unprecedented value to Andhra Pradesh.

In recent times data analysis is helping humans to have various novel perspectives to data. Innovative visualizations, trend extraction, and data cleaning techniques allow better understanding of data. It also helps decision making for data forecasting models. In case of dam data, accurate forecasts help in improved flood routing and management.[1,2]

A dam's data about inflow, outflow, capacity and level varies with time. Each day data is collected by officials. Such kind of data is known as time series [3]. In time series value of a variable is dependent on time. Also, current value of a variable can be determined by its historical value. So, time series variable is correlated with itself, also known as auto correlation. Analysis of auto correlations and partial auto correlation lead to better future predictions for a variable. Here auto correlation analysis and partial auto correlation analysis for various data variables is carried out.

This paper is organized in following manner. Next section describes data collection and data cleaning performed on the data. Then analysis of various variables from data is performed.

Revised Manuscript Received on May 06, 2019

B. Pradeepakumari, Department of Civil Engineering, Acharya Nagarjuna University, Guntur, India.

K. Srinivasu, Department of Civil Engineering, RVR&JC Engineering College, Guntur, India

Later data is analyzed for time series patterns using autocorrelation factor and partial auto correlation factor. Further moving average, and differencing techniques are applied on data for pointers to forecasting.

II. DATA COLLECTION AND CLEANING

Last fifteen years of data is collected for Nagarjun Sagar Dam. Various Dam data is publically made available by different state governments and Government of India. Data collected is having single entry for each day since 01 January 2003. Data is collected up to 11 August 2018 for this work. Details of data are shown in table 1. Available data on website is provided in form of current date to previous dates for each month. (source of data: www.cadarsms.cgg.gov.in [9]). So, data for each month for all selected years is collected one by one. Then all data is sequentially combined in a single file. Data contains four variable values namely water inflow, water outflow, current dam capacity and current water level of the reservoir [4].

Sr. No.	Parameter	Value
1	Start Date	01-January-2003
2	Last Date	11-August-2018
3	Total Rows	6527
4	Total Variables (Columns)	4
5	Frequency of sampling	24 Hrs (in most cases)

Table.1 Description of Collected Data

Detailed satellite image of Nagarjuna Sagar Dam is shown in figure 1.

Fig. 1. **Nagarjuna Sagar Dam Satellite Image from Google Maps**

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Various data cleaning steps are performed here. First, date and time from website data is converted to corrected format. There are repeated entries for date when frequency of sampling is more than 24hrs. In those cases all records of same day are selected and they are replaced by a single record for that day using averaging. This allows having unique dates in data and makes time series analysis of data easier [5].

There are lots of missing data in collected samples. Missing values are replaced with previous observation value. This strategy allows minimal disturbance in pattern of data due to missing values. Additionally, extreme or outlier values in data are replaced with mean value.

III. Analysis of Water Inflow

Inflow water in a dam is a significant parameter to predict decision making for flood prediction and dam gate operations. It depends on how much rainfall occurred in river basin nearby to a dam area. Total inflow ultimately affects a overall dam decision system. There are seasonal changes in inflow values. This can be clearly seen in figure 2For inflow values of Nagarjun sagar dam for last fifteen years, it can be concluded that each year inflow, outflow and capacity is grossly seasonal. But there are some uneven patterns are observed in data over the years.

IV. Analysis of Dam Capacity

Capacity of a dam provides estimate value of current water storage inside a dam. Capacity of dam also follows a seasonal change. In rainy season capacity increase and then it slowly keeps on depleting. This is due to continuous use of water. This repeated pattern can be clearly seen in figure 3.

For capacity values of Nagarjun sagar dam for last fifteen years, it can be observed that in 2003, 2004, 2015 and 2016, dam water storage was minimal throughout the year. On contrary, second half of 2010 and first half of 2011 experienced significantly large values of consistent storage. Additionally, year 2011 experienced consistent fluctuations in water storage [6,7].

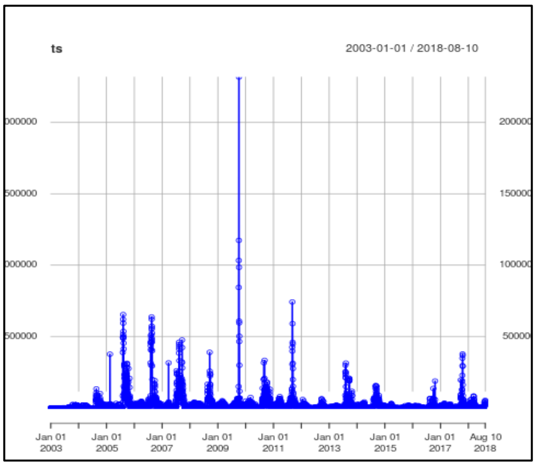

Fig-2: Water Inflow Time Series Plot

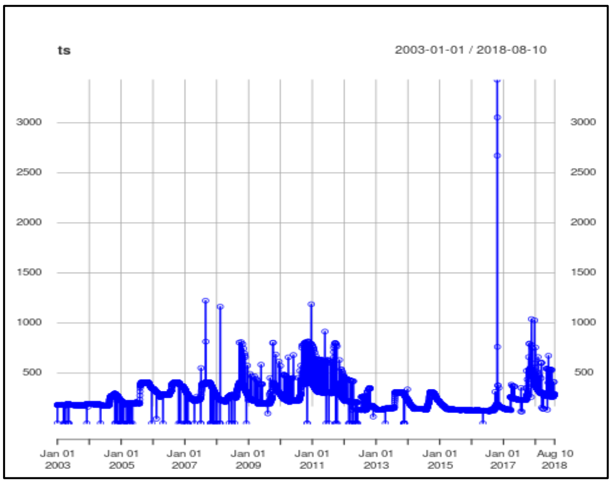

Fig-3: Time series plot for capacity of Dam

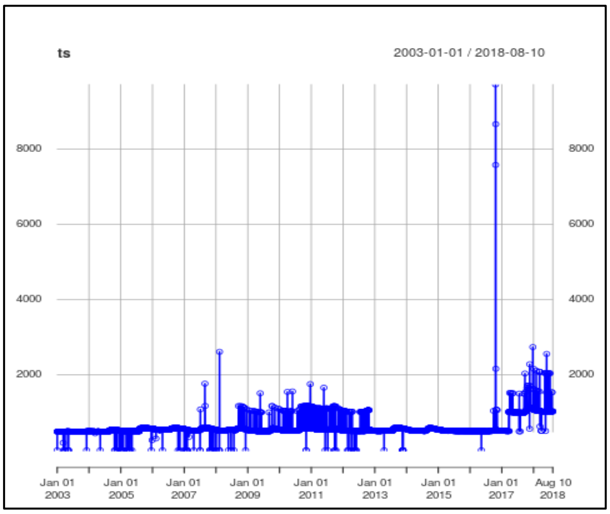

Fig-4: Water Level Time Series Plot

V. Analysis of Water Level

Water level in dam determines total capacity or contents of water in a dam. It is dependent of consistent inflow and current storage or capacity in dam. From figure 4 it can be easily seen that capacity of dam and water level are closely related in trend. Both of them follow similar seasonal pattern.

VI. Time Series Modelling of Data

A. Auto Correlation Factor (ACF) Graphs

Auto correlation factor graphs are used to find correlation of time series with its historical values. Factors represented in this graph are keys to detect auto regression factor value to be applied for forecasting of a time series [8]. Magnitude of each factor determines the magnitude of correlation. If the magnitude of a factor is greater than red dashed line (threshold) then, those factors are considered as significantly correlated.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Equation (1) depicts equation for finding auto correlation value (ρ) of m^{th} order for Y series with n samples. This equation is non-normalized as time series may have standard deviation zero [5,8].

$$\rho(m) = \frac{1}{(n-m)} \sum_{t=m+1}^{n} (Y_t - \mu_y)(Y_{t-m} - \mu_y) \ldots(1)$$

where μ_y is mean of series Y.

In figure 5, ACF for inflow is calculated. It can be clearly seen that third factor is having significant positive value of correlation. Thus in forecasting of this time series, last three values in time will have significant role in predicting current value. Another part is partial auto correlation factor (PACF) calculation [8]. PACF (φ) of order 'm' for series Y is depicted by using equation (2). 'corr' function finds provides correlation value for parameters. Figure 6 shows PACF for inflow values.

$$\varphi(mm) = corr(Y_t - P(Y_t \mid Y_{t+1}, \ldots, Y_{t+m-1}),$$

$$Y_{t+m} - P(Y_{t+m} \mid Y_{t+1}, \ldots, Y_{t+m-1})) \ldots(2)$$

Where, P is best linear projection between two entities optimized using root mean squared error

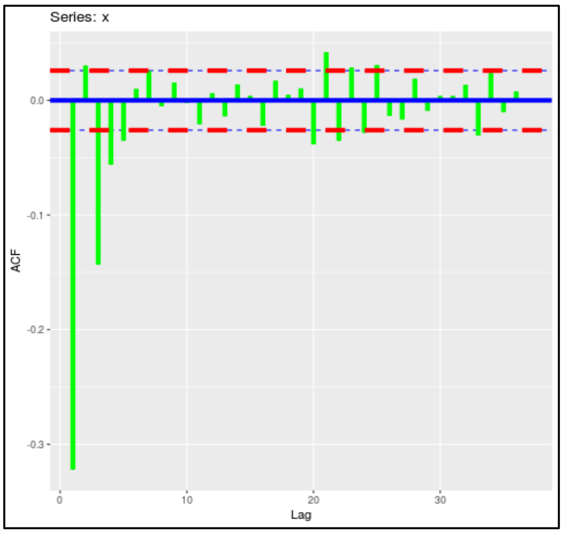

Fig-5: Time Series ACF Plot

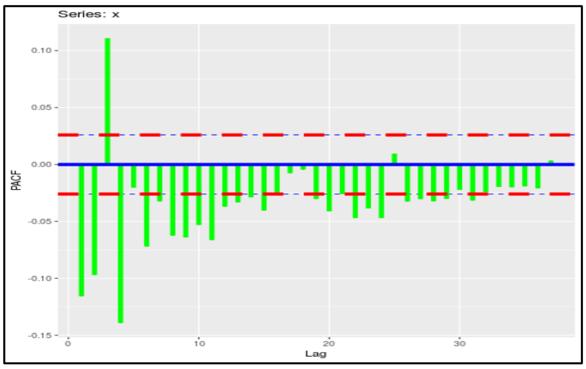

Fig-6: Time Series PACF Plot

B. Moving Average

In this technique time series data is analyzed using a window. A fixed size window of data is chosen. Then samples selected in this window are averaged to find moving average [5,8]. This window moves one place in data at a time to cover all available data. As this window is moving this averaging method is known as moving average. In time series data it plays an important role. Simple mean or average value will take away the effect of time from data. But moving average is based on fixed window so it maintains timely patterns of provided data. Equation (3) described moving average calculation for series Y in time frame of size w.

$$MA_{yw} = \frac{1}{w} \sum_{i=0}^{w} Y_i \quad \text{for } w \in \{1,n\} \ldots(3)$$

C. Differencing

In time series analysis, differencing is method to remove the additive effect. Such differencing, leads to a stationary time series. Some very highly un-stationary time series may require differing more than once. Equation (4) describes differencing for a time series Y.

$$Y' = Y_t - Y_{t-1} \text{ for all } t \in \{1,n\} \ldots(4)$$

In this case inflow with single differencing was made stationary. It can be seen from figure 7

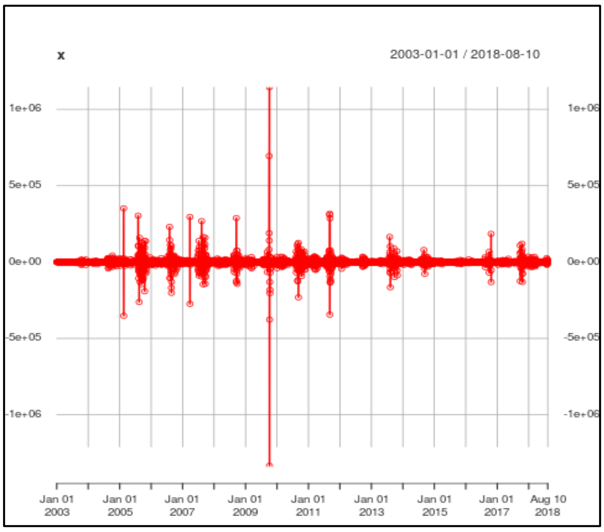

Fig-7: Time Series after Differencing

VII. CONCLUSION

Nagarjuana Sagar Dam has significant contribution towards social and economic development of Andhra Pradesh social as well as economic devlopment. So, this dam's data for last fifteen years is analysed here. This analysis provides ground prediction of future values for various dam parameters like inflow, outflow, capacity and level.

Here in this work detailed analysis reveals patterns in Auto Correlation factors. Additionally application of differencing to time series has proved to stabilize the series. These key observations help in application of forecasting models for the data.

In future various forecasting models will be applied on this data. These models can belong to different categories like civil engineering graph based models, statistical models (ARIMA), and machine learning models like Reservoir computing and CapsNet.

REFERENCES

1. Barakat, Moataz Kh, and Ahmed Z. Nooh. "Reservoir quality using the routine core analysis data of Abu Roash “C” in Badr El Din-15 oil field, Abu Gharadig basin, north Western Desert, Egypt." Journal of African Earth Sciences 129 (2017): 683-691.
2. Zhu, Jun, et al. "Dam-break flood routing simulation and scale effect analysis based on virtual geographic environment." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8.1 (2015): 105-113.
3. Ahmed, Tarek. Reservoir engineering handbook. Gulf Professional Publishing, 2018.
4. Martin, James L., and Steven C. McCutcheon. Hydrodynamics and transport for water quality modeling. CRC Press, 2018.
5. Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.
6. Gallegos, Tanya J., and Brian A. Varela. Trends in hydraulic fracturing distributions and treatment fluids, additives, proppants, and water volumes applied to wells drilled in the United States from 1947 through 2010: Data analysis and comparison to the literature. USGS, 2014.
7. Azad, Wan Hazdy, et al. "2 Dimensional Hydrodynamic Flood Routing Analysis on Flood Forecasting Modelling for Kelantan River Basin." MATEC Web of Conferences. Vol. 87. EDP Sciences, 2017.
8. Weigend, Andreas S. Time series prediction: forecasting the future and understanding the past. Routledge, 2018
9. Source of data: www.cadarsms.cgg.gov.in