

## Development of Fuzzy Based Autoregressive Moving Average Exogenous Input Model for Water Flow in Nigerian Kanji Hydro-Power Dam

<sup>1</sup>Araromi D. O., <sup>1\*</sup>Olu-Arotiowa O. A., <sup>2</sup>Adeoye O. K., <sup>1</sup>Sulayman A. A., <sup>1</sup>Ajala O.O. and <sup>3</sup>Adabanija M. A.

<sup>1</sup> Department of Chemical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

<sup>2</sup> Department of Chemical Engineering, Federal University of Technology, Akure, Nigeria

<sup>3</sup> Department of Earth Sciences, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

\* Correspondence Author: oaolu-arotiowa@lautech.edu.ng

---

Submitted on: 12/07/2021

Accepted on: 22/09/2021

---

### Abstract

*Design of robust control system for any system requires model-driven approach. Therefore, it becomes imperative to develop a dynamic model suitable for controller design on safety operation of hydropower dam for power production in Kanji dam in Nigeria. Model for reservoir flow was developed in MATLAB environment using Fuzzy Based Autoregressive Moving Average Exogenous Input (FARMAX) model structure in this study. The data used for model development covered a period of ten years (2003-2013). It consists of water inflow (WI), water outflow (WO) and spillage (S). WI and S are input variables while WO was the output variable. The model obtained using the unsmoothed data with an outlier gave -14.115%, -0.302 and 610.317 for fit,  $R^2$  and RMSE, respectively. Unsmoothed data with no outlier gave -13.802%, -0.295 and 608.643 corresponding to fit,  $R^2$  and RMSE, respectively. The model obtained using the smoothed data in the presence of an outlier gave 80.533%, 0.962 and 104.113 for fit,  $R^2$  and RMSE, respectively. Smoothed data in the absence of outlier gave 81.533%, 0.962 and 99.637 for fit,  $R^2$  and RMSE, respectively. FARMAX has the best fit value of 87.8774% when number of rules was equal to 3 with optimal model order of 3 1 4 3. The model can serve as a decision support system in evaluating the optimal reservoir operation policies in real time.*

**Keywords:** Electricity generation; FARMAX model; Reservoir flow; Spillage; Hydropower; Kanji dam.

### Introduction

The use of hydropower as a renewable energy resource is becoming increasingly popular because of its environmental adaptivity, ease of operation and cost-effectiveness for electricity generation. Its operation is free of greenhouse emission that causes climate change and fly-ash associated with thermal power plants (Salami, *et al.*, 2012; Abdulkadir, *et al.*, 2013; Ahmed, *et al.*, 2014; Priyabrata, *et al.*, 2012). It is important to have an effective management system for hydropower reservoirs for improved electric power supply and this can only be achieved when there is sufficient understanding of the existing interactions between reservoir variables and energy generation. Improved electric power supply to consumers can bring about an increase in the domestic product of the country and improved the standard of living (Nwobi-Okoye and Igboanugo, 2013). Abdulkadir, Salami, Anwar, and Kareem, (2013) pointed out that Nigeria is blessed with seventy micro dams, one hundred and twenty-six mini dams and eighty-six small sites along with numerous river systems that can be used to develop hydropower plants.

Kanji hydropower station was commissioned in 1968 with the installed capacity of 320MW and by 1978; the station had 8 plants with a capacity of 760 MW (Jimoh, 2008). In Kanji hydropower station, a visible

problem is the variability in water level and consequent variation in electric power outputs from the generators. According to Tangirala, (2015) the major hydropower reservoir variables affecting the energy generation were reservoir inflow, storage, reservoir elevation, turbine release, net generating head, tailrace level and losses due to evaporation. The relevance of these variables varies from month to month and during low flows, the variables are more significant in power generation (Abdulkadir, *et al.*, 2013). The reservoir inflow parameters which comprise of peak inflow, average inflow and minimum inflow, largely determine the amount of water that enters into the reservoir. The inflows in turn control the level of water in the reservoir, the pressure head and the reservoir storage balance. Also, the outflow parameters comprising of peak outflow, minimum outflow and average outflow are crucial to reservoir balance, the reservoir useful life and reservoir discharge (Ifabiyi, 2011). Kanji hydropower station is characterized by seasonal and inter-annual variations of inflows leading to two distinct peaks called black (December or January) and white (August or September) floods (Nwobi-Okoye and Igboanugo, 2013). Besides, reservoir sedimentation of solid particles causes a reduction in capacity, false water level and aids the occurrence of flooding (Salami, *et al.*, 2012).

In order to provide power at rated frequency and voltage, various control systems are required by the hydropower generation plant. Voltage is maintained within the prescribed limit by the control of excitation of the generator while that of frequency is maintained by removing mismatch between generation and load demand due to the river flow and head through a turbine. Power can be generated at the desired rate by controlling flow through the turbine while dams are maintained safely through controlling spillway gates (Priyabrata, *et al.*, 2012). The design of a controller for any system requires a model-driven approach which must be simple but adequate for the system of interest Zeigler *et al.*, (2017). Therefore, it becomes crucial to develop a dynamic model that establishes an input-output relationship for the design of a robust controller for the safe operation of hydropower dam for power production in Kanji dam in Nigeria.

Different soft computing techniques have been used in recent times to model hydropower reservoir variables affecting energy generation such techniques include Mixed Integer Programming, Dynamic Programming, Stochastic Modeling, Nonlinear Approach and Successive Linear Programming. The nonlinear method for solving hydrogeneration scheduling showed good results and is high in precision and execution time for large scale systems. In this work, a dynamic model for reservoir flow prediction for the Kanji hydroelectric station is presented. The model was developed by using Fuzzy Based Autoregressive Moving Average Exogenous Input (FARMAX) model structure. The data used consist of water inflow (WI), water outflow (WO) and spillage (S). WI and S were used as input variables while WO as the output variable. The model can be used to design an adaptive control scheme to regulate water outflow from the reservoir to the power plant in the presence of disturbance (inflow stream) by using spillage as a manipulated variable for safe operation and regular supply of electricity.

## **Methodology**

### **The study area**

Kanji hydropower reservoir in Nigeria is located in the lower Niger basin majorly for hydroelectric power generation. The Niger River, in which the Kanji dam impounds, is the third-longest river in Africa after the Nile and Congo/Zaire Rivers with a total length of about 4200 km and an active catchment area of just

about 1.5 million km<sup>2</sup>. It stretches across Benin, Burkina Faso, Cameroon, Chad, Guinea, Ivory Coast, Mali, Niger, and Nigeria. Kanji Dam was constructed in 1969 across the River Niger on Kanji Island to impound water for electricity generation. Kanji Hydro Electric Power PLC, with an installed capacity of 760 MW and provisions for expansion to 1156 MW, is Nigeria's largest hydropower station (Abdulkadir, *et al.*, 2013).

The location is between the latitudes 9°50' N and 10°35'N and longitudes 4°26'E and 4°40'E. The height of the dam from its toe to the crest is 65.5 m, its length is 135 km and the width at its widest point is 30 km. The maximum tail water elevation is 104m (341ft), the minimum headwater elevation at which the plant can operate is 132m, the reservoir surface area is about 1270 km<sup>2</sup> and the total storage capacity is 15 billion cubic meters (Adebayo and Bolaji, 2014; Nwobi-Okoye and Igboanugo, 2013). In compliance with the international law on dams across international rivers, Kanji dam has two navigational locks (the upper and lower locks) opened for the passage of barges or boats from the upstream to the downstream of the dam (Nwobi-Okoye, *et al.*, 2013)) in predicting water levels at Kainji dam using Artificial Neural Networks)

Due to dam construction, Kanji Lake was formed to acts as a reservoir for the dam. The lake is 136 km long and characterized by two flooding seasons, the white flooding (around July) and black flooding (arrives in December). The white flooding is due to the inflow of flood into the reservoir from rains within the catchment areas of the river within Nigeria. The Black flooding is a result of the inflow of flood into the reservoir from rains in the catchment areas of the river outside Nigeria such as Guinea, Mali, Niger, etc. (Nwobi-Okoye, *et al.*, 2013),2013)

Ten (10) -year daily water flows data were obtained from Kanji Hydroelectric Power Company PLC, Kanji, New Bussa, and Niger State, Nigeria. The data were used to model the daily water outflow from given inflow and spillage using Takagi-Sugeno Fuzzy System. The "C-means clustering" and recursive least squares were used to train the premises and consequents parts of the Takagi-Sugeno fuzzy system respectively.

### **Data pretreatment**

The data collected, which covered a period of 10 years (2003-2013), include reservoir inflow (m<sup>3</sup>/s), reservoir outflow (m<sup>3</sup>/s) and spillage (m<sup>3</sup>/s). The data were subjected to quality checks and pre-treatment steps before presenting it to the model estimation algorithms. Pre-treatment involves pre-filtering, removal of non-zero means, trends/drifts, outliers and filling of missing data and transformations of the data because modeling requirements demand consistent or efficient estimates (e.g., stationary).

The missing data were filled by using the approach of linear interpolation and matching patterns. Linear interpolation was used between the last good measurement and the next first good measurement while in the matching pattern approach; a similar pattern in the dataset was copied to the missing time frame. The data values that do not appear to be consistent with the rest of the data were removed by calculating the mean and the standard deviation of each data column in the matrix. In this case, an outlier was defined as a value that is more than three times the standard deviations. Smoothing of data was achieved by using locally weighted scatter plot smooth (LOWESS). In this method, each smoothed value was determined by

neighboring data points defined within the span. The span was determined as 10% of the total number of data points in the dataset. A regression weight function was defined for the data points contained within the span using a linear polynomial. The weights are given by the tricube function shown below.

$$w_1 = \left(1 - \left|\frac{x-x_i}{d(x)}\right|^3\right)^3 \quad (1)$$

$x$  is the predictor value associated with the response value smoothed,  $x_i$  are the nearest neighbors of  $x$  as defined by the span, and  $d(x)$  is the distance along the abscissa from  $x$  to the most distant predictor value within the span. The characteristics of the weights are that the data point smoothed has the largest weight and the most influence on the fit while the data points outside the span have zero weight and no influence on the fit.

### Model development

In this research, Autoregressive Moving Average Exogenous Input (ARMAX) structure was used. It has the following formulations:

$$y[k + 1] = \frac{B_1(q^{-1})}{A(q^{-1})} u_1[k] + \frac{B_2(q^{-1})}{A(q^{-1})} u_2[k] + \frac{C(q^{-1})}{A(q^{-1})} e[k] \quad (2)$$

where  $u_1$  and  $u_2$  are input sequences that represent reservoir inflow ( $m^3$ ) and spillage ( $m^3$ ), respectively.  $y$  is the output sequence representing reservoir outflow ( $m^3$ ),  $e(k)$  is white noise sequence with zero mean value and variance  $\sigma^2$ , i.e.,  $e(k) \sim N(0, \sigma^2)$ . The polynomials A, B, and C are defined in terms of the backward shift operator (Tangirala, 2015) as follows:-

$$\left. \begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_{na} q^{-na} \\ B_1(q^{-1}) &= b_{10} + b_{11} q^{-1} + \dots + b_{1nb_1} q^{-nb_1} \\ B_2(q^{-1}) &= b_{20} + b_{21} q^{-1} + \dots + b_{2nb_2} q^{-nb_2} \\ C(q^{-1}) &= c_0 + c_1 q^{-1} + \dots + c_{nc} q^{-nc} \end{aligned} \right\} \quad (3)$$

Where:

$a_1 \dots a_{na}$  are the parameters associated with number ( $na$ ) of past output,  $y$  used in forming regressor variables

$b_{11} \dots b_{1nb_1}$  are the parameters associated with number ( $nb_1$ ) of past input,  $u_1$  used in forming regressor variables

$b_{21} \dots b_{2nb_2}$  are the parameters associated with number ( $nb_2$ ) of past input,  $u_2$  used in forming regressor variables

The vector [ $na \quad nb_1 \quad nb_2 \quad nc$ ] defines the order of the model

The nonlinear variance of Equation (2) was defined according to:

$$y(k + 1) = F(y(k) \dots y(k - na + 1), u_1(k) \dots u_1(k - nb_1 + 1), u_2(k) \dots u_2(k - nb_2 + 1), e(k) \dots e(k - nc + 1)) \quad (4)$$

The function  $F(\cdot)$  was accounted for by using Takagi-Sugeno (TS) fuzzy system approximated as Multi Input Single Output (MISO) TS fuzzy system according to:

$$R^i: \text{ IF } Z_i \text{ is } F_{i_1}^i \text{ AND } \dots \text{ AND } Z_v \text{ is } F_{i_v}^i \\
 \text{ THEN } y_i(k+1) = a_{i_1} y(k) + \dots + a_{i_{na}} y(k-na+1) + b_{i_{11}} u_1(k) \dots + b_{i_{1nb_1}} u_1(k-nb_1+1) + b_{i_{21}} u_2(k) + \dots + b_{i_{2nb_2}} u_2(k-nb_2+1) + c_{i_1} e(k) \dots + c_{i_{nc}} e(k-nc+1) \quad (5)$$

Where:

$R^i$  is the  $i^{th}$  fuzzy inference rule,

$a_{i_1}, \dots, a_{i_{na}}, b_{i_{11}}, \dots, b_{i_{1nb_1}}, b_{i_{21}}, \dots, b_{i_{2nb_2}}, c_{i_1}, \dots, c_{i_{nc}}$  are the consequent parameters associated with the model, and  $\mathbf{z}(k)$  is regressor vector which is given as:

$$\mathbf{z}(k) = [y(k-1), \dots, y(k-na+1), u_1(k-1), \dots, u_1(k-nb_1+1), u_2(k-1), \dots, u_2(k-nb_2+1), e(k), \dots, e(k-nc+1)].$$

$F_j^i$  ( $j = 1, 2, \dots, n$ ) is the fuzzy set which was represented by a membership function,  $\mu_j^i$  and  $n$  defines the overall order of the model and equals to the sum of the elements in vector  $[na \ nb_1 \ nb_2 \ nc]$ . The membership functions used in the characterization of each fuzzy set were Gaussian membership functions. Gaussian membership function was adopted to reduce the effect of noise in the signal. It has low sensitivity to small changes in input parameters than the singleton membership function (Nafisi, Eghbal, Motlagh and Yavari, 2011). The two parameters that characterized Gaussian membership functions as shown in Equation 6 are the center,  $c$  and the width,  $\sigma$

$$\mu_j^i = \exp \left[ -\frac{1}{2} \left( \frac{u_j - c_j^i}{\sigma_j^i} \right)^2 \right] \quad (6)$$

The parameters of the premise part of the fuzzy rules and the number of the rules were determined by fuzzy C- mean clustering whiles the parameters of the consequent part were determined by the recursive least square (RLS) technique. A crisp output was chosen using the center of each of the output membership functions.

The Defuzzification strategy employed to provide a means to choose a single output is the center of gravity (COG) given in Equation (7). COG was chosen as a defuzzifier to allow smooth and continuous changes in the output parameters ( Karray and De Silva, 2004).

$$y = \frac{\sum_{i=1}^R \prod_{j=1}^n \exp \left( -\frac{1}{2} \left( \frac{z_j(k) - c_j^i}{\sigma_j^i} \right)^2 \right) g_i(z(k))}{\sum_{i=1}^R \prod_{j=1}^n \exp \left( -\frac{1}{2} \left( \frac{z_j(k) - c_j^i}{\sigma_j^i} \right)^2 \right)} \quad (7)$$

Where:

$$g_i(z(k)) = a_{i_1} y(k) + \dots + a_{i_{na}} y(k-na+1) + b_{i_{11}} u_1(k) + \dots + b_{i_{1nb_1}} u_1(k-nb_1+1) + b_{i_{21}} u_2(k) + \dots + b_{i_{2nb_2}} u_2(k-nb_2+1) + c_{i_1} e(k) \dots + c_{i_{nc}} e(k-nc+1) \quad (8)$$

The normalized membership function  $\xi$  is determined according to

$$\xi_i(z(k)) = \frac{\prod_{j=1}^n \exp\left(-\frac{1}{2}\left(\frac{z_j(k)-c_j^i}{\sigma_j^i}\right)^2\right)}{\sum_{i=1}^R \prod_{j=1}^n \exp\left(-\frac{1}{2}\left(\frac{z_j(k)-c_j^i}{\sigma_j^i}\right)^2\right)} \quad (9)$$

Equations (8) and (9) were combined to obtain

$$\begin{aligned} y = & \sum_{i=1}^R a_{i1} \xi_i(z(k))y(k) + \dots + \sum_{i=1}^R a_{ina} \xi_i(z(k))y(k - na + 1) + \\ & \sum_{i=1}^R b_{i11} \xi_i(z(k))u_1(k) + \dots + \sum_{i=1}^R b_{i1nb1} \xi_i(z(k))u_1(k - nb_1 + 1) + \\ & \sum_{i=1}^R b_{i21} \xi_i(z(k))u_2(k) + \dots + \sum_{i=1}^R b_{i2nb2} \xi_i(z(k))u_2(k - nb_2 + 1) + \sum_{i=1}^R c_{i1} \xi_i(z(k))e(k) + \\ & \dots + \sum_{i=1}^R c_{inc} \xi_i(z(k))e(k - n_c + 1) \end{aligned} \quad (10)$$

Equation (10) in compact form gives:

$$\hat{y}(k) = \boldsymbol{\theta}^T(k)\boldsymbol{\xi}(z(k)) \quad (11)$$

in which  $\boldsymbol{\xi}(z)$  is a vector containing convolution regressor variables and normalized membership function as given in Equation (9) and  $\boldsymbol{\theta}$  is a vector of unknown parameters.

### Parameter estimation

Recursive least squares (RLS) algorithm was employed to determine unknown parameters of the model (parameters of the consequent part). This is because it gives faster convergence and smaller errors (Paleologu, Benesty, and Silviu, 2008). It also offers a variation profile for the adaptation gain and saves the computation time by using the results obtained at a time  $(k - 1)$  to get the estimates at time  $k$ . Recursive version of the least square method where the vector of parameters was updated in each step is given according to general formula (Landau *et al.*, 2011)

$$\begin{aligned} \begin{bmatrix} \text{New estimated} \\ \text{parameter(vector)} \end{bmatrix} &= \begin{bmatrix} \text{Previous estimated} \\ \text{parameter(vector)} \end{bmatrix} + \begin{bmatrix} \text{Adaptive gain} \\ \text{(matrix)} \end{bmatrix} \times \begin{bmatrix} \text{Measurement} \\ \text{function(vector)} \end{bmatrix} \times \\ \begin{bmatrix} \text{Prediction error} \\ \text{function(scalar)} \end{bmatrix} & \end{aligned} \quad (12)$$

The unknown parameters of the model were determined by minimizing the sum of the square of the difference between the actual and predicted output values with possible weighting that measures the degree of precision presented in Equation (13). Z

$$J_N = z \sum_{i=1}^N [y(k) - \theta^T(k)\xi(z(k))]^2 \quad (13)$$

Where:

$y(k)$  is the process output in  $k$ -th step (outflow stream),  $\theta^T(k)\xi(z(k))$  is the predicted process output. Therefore, Equation (12) is translated to Equation (14) as follow:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\xi(k)\varepsilon(k) \quad (14)$$

In Equation (14),

$$\varepsilon(k) = [y(k) - \hat{\theta}(k-1)\xi(k)];$$

$P(k)$  is the covariance matrix or adaptation gain matrix defined according to

$$P(k) = \frac{1}{\lambda} ((I - P(k-1)\xi(k)\xi(k)^T) (\lambda I + \xi(k)\xi(k)^T) P(k-1)) \quad (15)$$

$I$  is the identity matrix and  $\lambda$  is the forgetting factor.

### Model order selection

The selection of orders of the model is essential whenever a data-driven model is being developed. Model order was selected by starting with a low order of vector [1, 1, 1, 1]. The equal increments were made on each element of the vector until the best fit was obtained. To avoid overfitting and large modeling error, the appropriate model order was determined by the factorial method.

### Model assessment and validation

Goodness-of-predictions were measured using the normalized root mean square (NRMS) expressed in a percentage called fit, root mean square error (RMSE) and the coefficient of determination denoted by  $R^2$  defined according to

$$\text{fit} = 100 \left( 1 - \frac{\|\hat{y}-y\|_2}{\|y-\bar{y}\|_2} \right) \quad (16)$$

$$\text{RMSE} = \frac{\|y-\hat{y}\|_2}{\sqrt{N}} \quad (17)$$

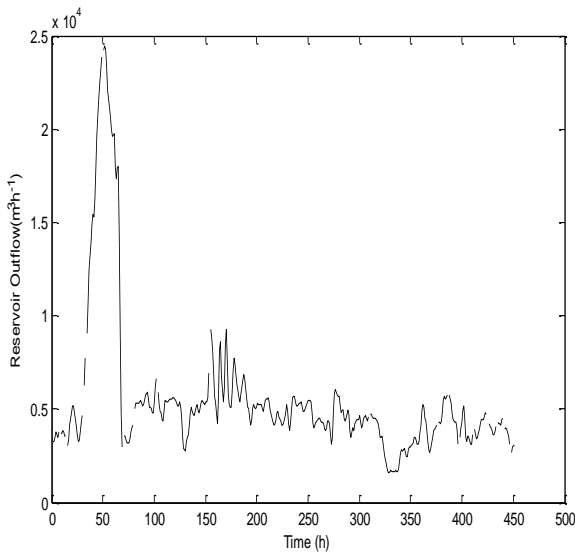
$$R^2 = 1 - \frac{\|\hat{y}-y\|_2^2}{\|y-\bar{y}\|_2^2} \quad (18)$$

### Results and Discussion

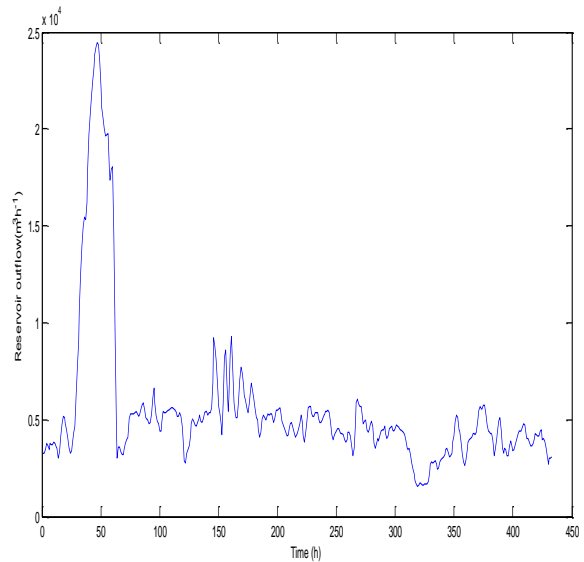
The proposed inflow and spillage prediction, Fuzzy-ARMAX (FARMAX) model was developed according to the procedure shown in sub-section 2.2-2.6. All the development made in this study was implemented with the aid of the Fuzzy logic toolbox in MATLAB.

### Data pretreatment

Figure 1 showed the plot of the segment of the raw data used for the modeling, the gap on the profile indicates that some data were missing. Figure 2 shows the same segment of raw data when the data were filled up. Figure 3 shows the profile of smoothed data and the original data. The smoothed data represent the original data without process noise and as seen in the figure, the two curves follow the same pattern indicating that smoothed data is closer to the given data. The impact of smoothing and outlier removal on the fitness of data is presented in Table 1. Root mean square error (RMSE), correlation coefficient ( $R^2$ ) and fit were used as performance measures. The worst fit was obtained when the data was not smoothed in the presence of an outlier as indicated by negative values of performance measures. No significant improvement was observed when the outlier was removed without data smoothing as indicated by negative values of performance measures. Tremendous improvement was achieved when the data was smoothed with no removal of outliers. The result gave an RMSE value of 104.1127,  $R^2$  value of 0.9621 and a fit value of 80.5333%. There was further improvement when outliers were removed and data were smoothed. The result gave an RMSE value of 99.6368,  $R^2$  value of 0.9621 and a fit value of 81.5333%. The result showed that the best fit was obtained when data were smoothed with no outlier.

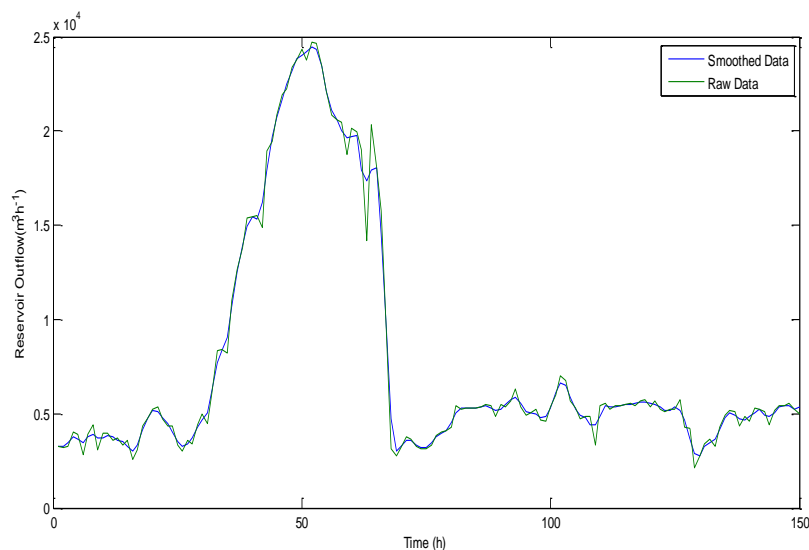


**Figure 1:** Raw data showing some missing data



**Figure 2:** Profile of completely filled data





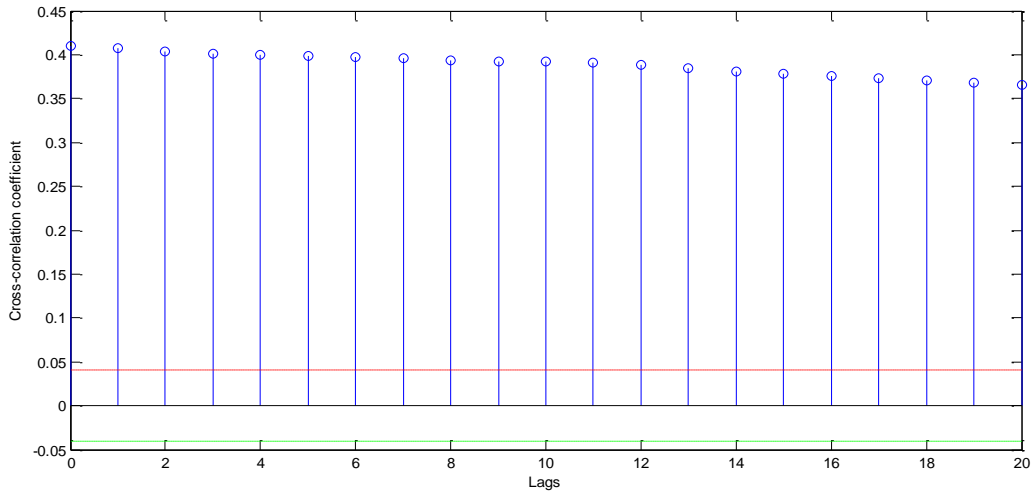
**Figure 3:** Profile of smooth data and original data

**Table 1:** Impact of smoothing and outlier on the fitness of data

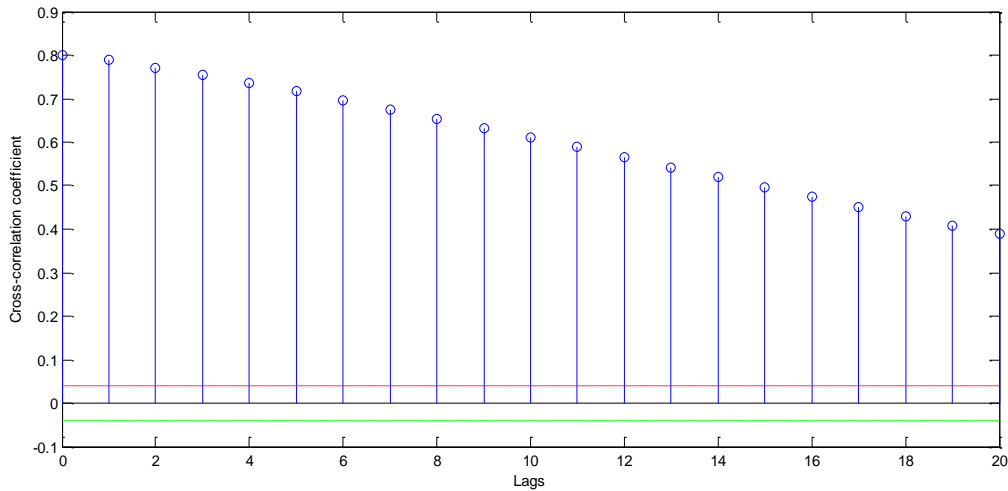
RMSE	R <sup>2</sup>	Fit	Remark
610.3174	-0.3022	-14.115	Outlier and no smoothing
608.6432	-0.2951	-13.802	No outlier and no smoothing
104.1127	0.9621	80.5333	Outliers with data smoothed
99.6368	0.9621	81.5333	No Outlier with data smoothed

**Identification of time lag**

Cross-correlation plots between output reservoir flow and the inputs (reservoir inflow and spillage) were shown in Figures 4 and 5 respectively. The peaks were obtained at  $k = 0$  on both plots meaning that there was no estimated time lag between the process input and the corresponding output. The horizontal dotted lines are confidence intervals with a probability of 0.99 (99%) corresponds to a 2.58 standard deviation. The vertical lines represent the magnitude of correlation at different time lags. The time lag of a system corresponds to the time where there is the longest vertical line (Brian, *et al.*, 2006). As can be observed in Figures 5 and 6, the longest vertical line is at  $k = 0$  for both inputs, implying that there was no estimated time lag between both inputs and the output.



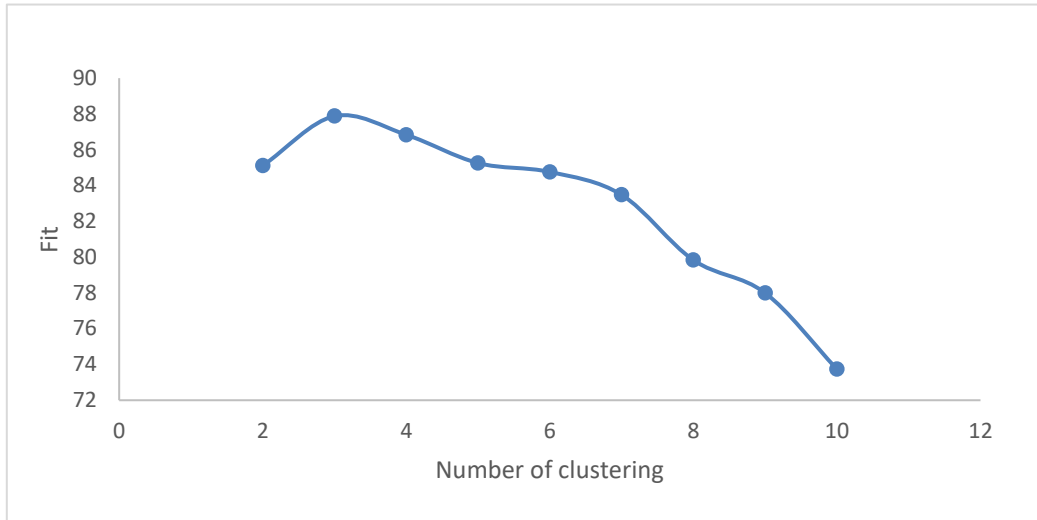
**Figure 4:** Cross-correlation for process output



**Figure 5:** Cross-correlation for reservoir inflow and spillage

### Effect of cluster centre or number of rule

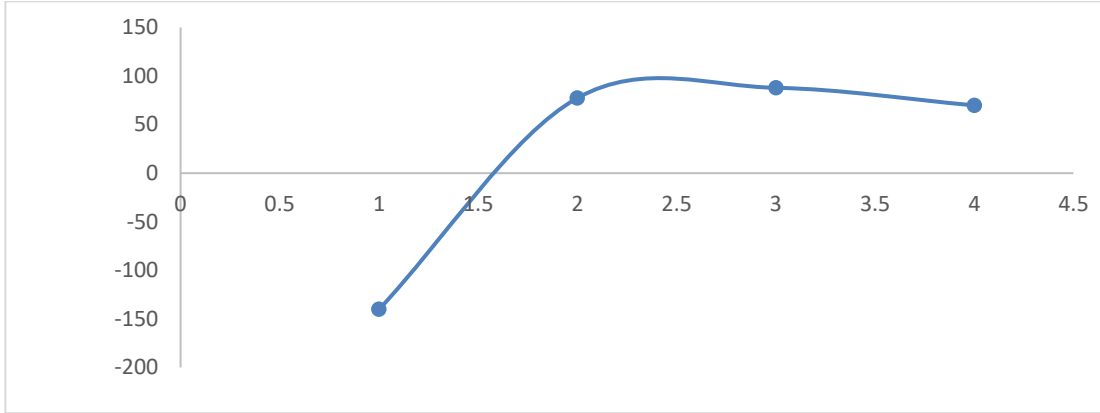
The relationship between the fit and numbers of clustering was shown in Figure 6. The clustering was varied between 2 and 10. One clustering was avoided because there was a loss in the rank of the working matrix when one clustering was used. The value of fit increased when the clustering number was increased from 2 to 3. There was a drop in the value of fit with further increase clustering number with some undulating moves. The best fit was obtained when the number of clustering was equal to 3. This value was used as the number of rules in the fuzzy system for modeling.



**Figure 6:** Effect of number of clustering on model performance

**Order of the model**

The numbers of past output, past input 1, past input 2 and past error represented by  $n_a, n_{b1}, n_{b2}$  and  $n_c$ , respectively form regressor variable constituted the ARMAX model order. The noise  $e(t)$  used throughout the identification process is Gaussian white with zero mean and variance  $\sigma^2 = 0.0001$ . When these numbers were varied equally from (1 1 1 1) to (4 4 4 4) and model performance was measured in terms of fit, the results obtained are presented in Figure 8. Model structure (1 1 1 1) gave very poor model fitting as indicated by a negative value of fit. There was an improvement in fitting with variation between (2 2 2 2) and (3 3 3 3) and between (3 3 3 3) and (4 4 4 4) there was a drop in the fit value indicating that increasing some value beyond 3 led to poor model performance. It is pertinent to note that increasing model order will increase the complexity of the model which may lead to a model with high demand for computational resources and over fitting. When final model order selection was selected with the help of factorial design implemented in expert design software, the results obtained are shown in Figure 7. The best fit (87.8774 %) was obtained for model order structure (3 1 4 3). This means that three (3) pasts outflow, one (1) past inflow, four (4) past spillage and three (3) past error values (residual) will be needed in predicting the current outflow. This observation may not be out of place considering the natural occurrence of the system under consideration. The model will be good for scheduling policy in the power generation to plan ahead of shortfall in water volume and for dam diagnosis and retrofitting to avoid flooding during the period of heavy rain. However, model order reduction may need to be carried out on the model before it is suitable for real-time applications like control, fault detection systems and online optimization to reduce computational resources.



**Figure 7:** Effect of model order on model performance

### Fuzzy Based Autoregressive Moving Average Exogenous Input (FARMAX) Model

The FARMAX model obtained consists of three rules ( $R=3$ ). The entire set of fuzzy rules for the FARMAX model with  $i = 1,2,3$  is of the form:

*Rule<sub>i</sub>*: IF  $y(k-1)$  is  $\mu_{y_1}^i$  and  $y(k-2)$  is  $\mu_{y_2}^i$  and  $y(k-3)$  is  $\mu_{y_3}^i$  and  $u_1$  is  $\mu_{u_{11}}^i$  and  $u_2(k-1)$  is  $\mu_{u_{21}}^i$  and  $u_2(k-2)$  is  $\mu_{u_{22}}^i$  and  $u_2(k-3)$  is  $\mu_{u_{23}}^i$  and  $u_2(k-4)$  is  $\mu_{u_{24}}^i$  and  $e(k-1)$  is  $\mu_{e_1}^i$  and  $e(k-2)$  is  $\mu_{e_2}^i$  and  $e(k-3)$  is  $\mu_{e_3}^i$  Then  $f = a_{oi} + a_{1i}y(k-1) + a_{2i}y(k-2) + a_{3i}y(k-3) + b_{1i}u_1(k-1) + b_{2i}u_2(k-1) + b_{22i}u_2(k-2) + b_{23i}u_2(k-3) + b_{24i}u_2(k-4) + c_1e(k-1) + c_2e(k-2) + c_3e(k-3)$

The membership function used in the premise part of the rule *Rule<sub>i</sub>* is given

$$\mu_i(x, k) = \exp\left(-\frac{1}{2}\left(\frac{x(k)-c_i}{\sigma_i}\right)^2\right) \quad (19)$$

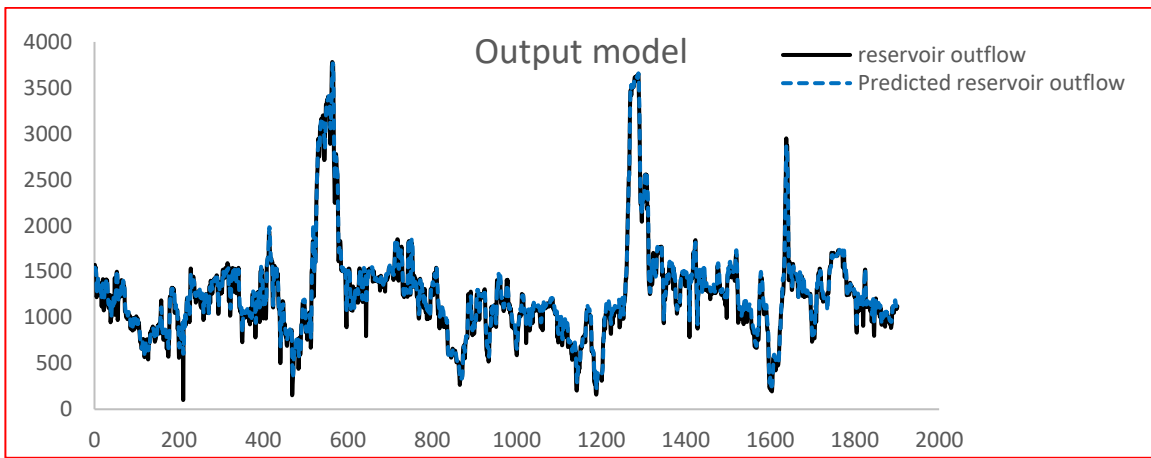
The values of parameters, sigma ( $\sigma_i$ ) and cluster center ( $c_i$ ) with  $i = 1,2,3$ , of the membership function given by Equation (19) are presented in Table 2. Variations are observed for all the parameters at the earlier time but with time progression the value of some parameters level out to remain constant while other parameters maintain their variation trends. Variations are more pronounced in rule 1 and rule 2 than in rule 3. The variations in parameters in the time evolution have established that the dam is characterized by complex and uncertain dynamics (Abdulkadir, *et al.*, 2013).

### Model validation

The developed model was validated using a dataset different from the one used in developing the model to give proper calibration to the model. Out of 4000 datasets collected, 2500 dataset was used for training while the rest was used for validation of the model. Visual assessment from Figure 8 indicated that there is a close match between reservoir outflow and predicted outflow. This also indicated the good predictive power of the model. The RMSE (81.3432), fit (87.8774%) and correlation coefficient (0.9901) were obtained in the model validation task and all these indicate good predictive power of the model.

**Table 2:** Parameters of membership functions

	Width of the membership function ( $\sigma$ )			Cluster Center (C)		
	Rule 1	Rule 2	Rule 3	Rule 1	Rule 2	Rule 3
y(k-1)	3035.320	18587.604	5393.599	1387.979	4113.550	1348.439
y(k-2)	3031.980	18562.970	5392.696	1392.196	4107.217	1349.782
y(k-3)	3052.686	18386.769	5374.071	1429.791	4066.721	1396.333
u1(k-1)	135.259	140.763	139.419	6.029	1.048	3.739
u2(k-1)	207.977	2599.599	330.097	114.221	668.598	146.258
u2(k-2)	207.787	2590.302	329.782	114.307	666.453	146.099
u2(k-3)	209.207	2555.999	328.988	116.286	658.721	147.518
u2(k-4)	211.928	2506.110	327.875	120.975	648.653	150.888
e(k-1)	66.545	54.986	67.385	33.310	6.795	37.853
e(k-2)	66.530	55.051	67.581	33.154	6.795	37.559
e(k-3)	66.331	55.968	67.820	33.368	6.657	37.821



**Figure 8:** Reservoir flow model

**Conclusion**

A model for reservoir flow prediction for the Kanji hydroelectric station was developed using the ARMAX model structure. A fuzzy inference system paradigm in a recursive least square manner was adopted to accommodate uncertainty in the data used for the model development. Forgetting factor, the number of fuzzy rules were found to play a major effect in arriving at a good predictive model. The model structure that gave the best performance metrics (RMSE=81.3432, fit=87.8774% and correlation coefficient=0.9901) has a model order structure (3 1 4 3). The high order justified the complex nature of the dam. The model will be good for scheduling policy in the power generation to plan ahead of shortfall in water volume and

for dam diagnosis and retrofitting to avoid flooding during the period of heavy rain. The model will consequently serve as a decision support tool to enhance the quality and quantity of power being generated and in the process of planning and management of a water resources system. However, the model will not be useful in its present form for a real-time operation like water flow control and real-time fault detection systems due to its high model order. The future work will target model order reduction.

## References

- Abdulkadir, T. S., Salami, A. W., Anwar, A. R., and Kareem, A. G. (2013). Modelling of Hydropower Reservoir Variables For Energy Generation: Neural Network Approach. *Ethiopian Journal of Environmental Studies and Management*, 6(3), 310-316.
- Adebayo, W. S., and Bolaji, F. S. (2014). Teleconnections of Seasonal Reservoir Inflow with ENSO indices: Application to Kainji Hydropower Generation Reservoir in Nigeria. *Report and Opinion*, 6(11).
- Ahmed, M. I., and Abed, S. Y. (2014). Simulation Model FOR Stage –IV Koyna Hydropower Plant International. *Journal of Advances in Engineering and Technology*, 6(6), 2373-2381.
- Brian, R., and Ben, B. (2006). *Process Dynamics and Control: Modeling for control and prediction*. USA: John Wiley and Son Ltd.
- Ifabiyi, I. P. (2011). Contributions of Reservoir Elements to Monthly Electricity Generation in the Jebba Hydropower Reservoir, Nigeria. *Ozean Journal of Applied Sciences*, 4(3), 251–264.
- Jimoh, O. D. (2008). Optimized Operation of Kainji Reservoir. *AU J.T*, 12(1), 34-42.
- Karray O, and De Silva C. (2014) Soft Computing and Intelligent Systems Design: Theory, Tools and Applications 1st Edition
- Landau, I. D., Lozano, R., M'Saad, M., and Karimi, A. (2011). *Adaptive Control Algorithms, Analysis and Applications* (2nd ed.). New York: Springer-Verlag London Limited.
- Nafisi VR, Eghbal M, Motlagh MJ, and Yavari F. Fuzzy logic controller for haemodialysis on human body model. *J Med Signals Sens* 2011;1:36-48
- Nwobi-Okoye, C. C., and Igboanugo, A. C. (2013). Predicting water levels at Kainji dam using Artificial Neural Networks. *Nigerian Journal of Technology*, 32(1), 129-136
- Paleologu, C. Benesty, J and Ciochina, S ( 2008). A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification, *IEEE Signal Processing Letters*, .( 15), 597-599.
- Priyabrata, A., Pankaj, K. R., and Asis, M. (2012). Safe and Efficient Control of Hydro Power Plant by Fuzzy Logic. *International Journal of Engineering Science and Advanced Technology*, 2(5), 1270 – 1277.
- Salami, A. W., and Sule, B. (2012). Optimal Water Management Modelling for Hydropower System on River Niger in Nigeria. *International Journal of Engineering*, 185-192.
- Tangirala, A. K. (2015). *Principle of System Identification-Theory and Practice*. New York, USA: CRC Press is an imprint of Taylor and Francis Group, an Informa business.
- Zeigler, B.P. Bernard, Nutaro JJ and Seo C, 2017. Combining Devs and model checking: concepts and tools for integrating simulation and analysis; *International Journal Of Simulation and Process Modeling*. 1:2-15.