Research Journal of Applied Sciences 11 (9): 851-857, 2016
ISSN: 1815-932X

Predictive Modelling for Reservoir Water Level

Siti Rafidah M-Dawam and Ku Ruhana Ku-Mahamud
Faculty of Computer Sciences and Mathematics, UniversitiTeknologi MARA, Kedah, Malaysia
School of Computing, College of Arts and Sciences, Universiti Utara Malaysia,
Sintok, Kedah, Malaysia

Abstract: Neural Network (NN) has been the most popular technique used in predicting Reservoir Water Level (RWL). However, NN is a black-box modelling technique where the model can be established without knowledge of the mathematical relationship between the inputs and the corresponding outputs. Most researches on reservoir water release applied the NN techniques using discretized data. To discover the current Reservoir Water Level at time t (RWL_t) in relation to the previous rainfall event, this paper proposed a predictive modelling for RWL using regression and the temporal pattern of both RWL and rainfall. The sliding window technique has been used to segment the temporal data into various slices. The finding shows that the best input scenario for the current RWL is one day delay for RWL and two days delay for rainfall; comparing this to the actual data, the model has an error of 0.1628%. The model can be used to guide the reservoir operator predicting the present and immediate decisions on reservoir water release, especially in the absence of the supervisor or during emergency situations.

Key words: Predictive model, reservoir modelling, reservoir water release, sliding window, temporal data mining

INTRODUCTION

The growing relevance of Reservoir Water Release (RWR) is a consequence of its social repercussions and significant flood consequences. Reservoirs can bring two kinds of threats; during heavy rainfall that could cause severe flood and during drought that could jeopardise the water supply for domestic, industrial and agricultural purposes (Chiang and Tsai, 2012). Making decisions on RWR is critical and very much influenced by the changes of RWL. Studies on RWL decisions have demonstrated the ability in predicting RWR (Ashaary *et al*., 2015; Mokhtar *et al*., 2014; Rani and Parekh, 2014). Modelling the previous reservoir operator's decisions on water release can be used as guidance for the present decision. Another related activity to reservoir is reservoir operation which has been defined as "establishing and implementing decision rules that guide the amount of water to be released from the reservoir at any given point of time" (Stam *et al*., 1998) which is also known as reservoir operating policy or control strategy (Moeini *et al*., 2011). In this study, reservoir operations refer to the decisions of gate opening or closing for RWR. Reservoir operations comprise many components such as inflow, outflow, water storage level and gate operations. Inflows to the reservoir can be in terms of rainfall (Afshar and Salehi, 2011) and streamflow.

Reservoir outflows can be in two forms: uncontrolled spillways and gated spillways. When RWL exceeds the fully supply level, it discharges through the uncontrolled spillways. The gated spillways are controlled by certain reservoir operating rules which are stagnant and disregard the dynamic nature of the hydrology systems (Ishak *et al*., 2011a). Therefore, a non-structural approach such as classifying and/or predicting the reservoir inflow is crucial to facilitate RWR or gate opening decision. In most reservoir operations, the decision to sustain or release a certain amount of water has been laid down in their Standard Operating Procedure (SOP) where certain parameters need to be considered. However, the effectiveness of the SOP and operating policy are affected by several factors such as sedimentation, water usage, climate changes and urbanisation. Eventually, the decisions or judgements made are subjective, uncertain and vague, based on the operators' intuition and previous learning experiences. Therefore, the reservoir operating policy needs to be periodically re-evaluated and updated

Corresponding Author: Siti Rafidah M-Dawam, Faculty of Computer Sciences and Mathematics, UniversitiTeknologi MARA Kedah, Malaysia

to improve the reservoir operation. There is an urgent need to model the reservoir operators' previous experience facing such difficult situations. These experiences give significant information on the RWR decisions as the operators may involve in turnover and mobility processes.

In RWR decisions, the data is in the form of temporal sequences where time is critical information related to each data whether in the form of month, day or hours (Mahamud *et al*., 2009). The changes in the patterns of the data can influence certain decision-making. Decision rules captured from the patterns provide invaluable information which can assist in making future decisions. The Temporal Data Mining (TDM) technique is required to uncover the values of the attributes involved from temporal sequences representing temporal information related to certain decisions by the algorithm formulation. The significant time delay between the cause of event and the actual event needs to be captured accurately.

This study focused on modelling RWL using regression based on the temporal data of RWL and rainfall in particularly on the reservoir's gate opening decisions. These temporal data were sliced using the sliding window technique to portray the delay between the rainfall and the increase of RWL. Then, the study applied the regression technique to discover a previous RWL and two-days of previous rainfall are the best predictor to the reservoir's gate opening decisions.

Literature review: Reservoir operation decisions are challenging and complex, especially during flood and drought events due to unpredictable inflow such as rainfall (Sattari *et al*., 2012). Thus, a few researches have focused on non-structural approaches predicting reservoir inflows. However, during flood or drought, the decision on RWR is not only based on the availability of water inflows but also on the previous release, demands, time, etc. Besides daily rainfall, several researches also considered water level as an input in the multipurpose reservoir forecasting model (Moeini *et al*., 2011). Rainfall (hydrological data) and water level are found to be correlated in the flood prediction model (Mahamud *et al*., 2009).

Many literature conducted on the RWR operation have utilized rainfall data and RWL as inputs (Nwobi and Igboanugo, 2013; Rani and Parekh, 2014) and have applied different methods and techniques of Artificial Intelligence and machine learning (Afshar and Salehi, 2011; Afshar, 2012; Alemu *et al*., 2010; Wei, 2012). Only a small number of researches conducted on RWR decisions highlighted on the time delay between the rainfall and the increase of RWL.

Ishak *et al*. (2011b) proposed the Reservoir Intelligent Decision Support System in the crisis condition. The system has three models: situation assessment, forecasting and decision models. In this model, the sliding window method is used to mine reservoir temporal patterns from the operational data as well as the previous experience of the reservoir operator. Results showed eight days' time lag relating to upstream rainfall and RWL. This study utilized NN to forecast and to classify RWL and concluded that the finest ANN model was 24-15-3. However, after two years, the model recommended five days' time lag with an 8-23-2 ANN model with a 0.007085% error. Nonetheless, the study utilized the discrete data of current RWL, tomorrow RWL and the changes of RWL as inputs in predicting the closing or opening of reservoir gates (Ishak *et al*., 2012).

Afiq forecasted the daily RWL of the Klang reservoir, Malaysia using Type 2 SVM regression. The input variables are rainfall and RWL which were used to determine the best time lag. Two days of rainfall and RWL were selected as the best time lag model with 1.64% error.

Nwobi-Okoye and Igboanugo developed Artificial Neural Network (ANN) models for predicting the water levels at the Kainji Dam which supplies water to Nigeria's largest hydropower generation station using a 10 year record of the daily water levels at the dam from 2001-2010. The ARIMA model with a relative error of 0.039% had the best prediction.

Rani and Parekh (2014) concluded that ANN using feedforward back propagation is an appropriate predictor for real-time water level forecasting of the Sukhi Reservoir, India. The inputs are the daily data of inflow, RWL and RWR and the best time lag is 10 day with a 0.82% error.

Mokhtar *et al*. (2014) applied NN to predict RWL and concluded a 5-25-1 NN model as the best architecture. The study found out that five days' observations of RWL are significant for the RWR decision with a 0.038756% error.

Ashaary *et al*. (2015) proposed a 4-17-1 NN architecture in forecasting the change of RWL stage. The changes and stage of RWL were used as the input patterns instead of the real value of RWL. The research showed that two days of delay have affected the changes in the stage of RWL.

An effective and timely method for predicting RWL can help in water-use formulation and scheduling for domestic, municipal and agricultural uses as well as in

Table 1: Sliced Reservoir Water Level (RWL)

Date	RWL_t	RWL_{t-1}	RWL_{t-2}	RWL_{t-3}	RWL_{t-4}	RWL_{t-5}	RWL_{t-6}	RWL_{t-7}	RWL_{t-8}
12-Feb-97	29.275	29.255	29.220	29.165	29.130	29.105	28.955	28.950	28.930
13-Feb-97	29.335	29.275	29.255	29.220	29.165	29.130	29.105	28.955	28.950
14-Feb-97	29.335	29.335	29.275	29.255	29.220	29.165	29.130	29.105	28.955
15-Feb-97	29.280	29.335	29.335	29.275	29.255	29.220	29.165	29.130	29.105

Table 2: Sliced Averaged Rainfall Data (RF)

Date	RF_t	RF_{t-1}	RF_{t-2}	RF_{t-3}	RF_{t-4}	RF_{t-5}	RF_{t-6}	RF_{t-7}	RF_{t-8}
12-Feb-97	20.25	7.330	5.380	13.00	0.000	46.25	24.50	10.00	16.17
13-Feb-97	13.88	20.25	7.330	5.380	13.00	0.000	46.25	24.50	10.00
14-Feb-97	8.25	13.88	20.25	7.330	5.380	13.00	0.000	46.25	24.50
15-Feb-97	1.00	8.250	13.88	20.25	7.330	5.380	13.00	0.000	46.25

disaster monitoring, response and control in areas prone to floods. The number of feature groups and the number of elements in each feature group used as inputs greatly influence the forecasting of RWL accurately (Rani and Parekh, 2014).

Study area: In this study, the TimahTasoh reservoir, situated in the northern part of Malaysia was used as a case study. The TimahTasoh reservoir is one of the largest multipurpose reservoirs in northern Peninsular Malaysia. It serves as a reservoir for flood mitigation as well as water supply and recreation. Water from TimahTasoh is used for domestic, industrial and irrigation purposes. The reservoir operation data from 1997 until 2006 is collected from the Department of Irrigation and Drainage (DID) which is in charge of monitoring and managing the reservoir. The data consists of operational and hydrological data. The operational data has the daily RWLs measured in metre (m) unit and the reservoir gate opening decisions in terms of quantity, size and duration. The hydrological data has the daily rainfall readings measured in milimetre (mm), recorded from five gauging stations.

Data preparation: Data preprocessing is often required before using any data mining process to improve the results' performance. In order to handle the missing value, the interpolation technique is used by computing:

$$f(x) = f(x_0) + (x - x_0)\frac{f(b) - f(a)}{b - a}$$

Where:

f(x) = The missing value
$f(x_0)$ = The value before the missing value
x = The point of missing value
x_0 = The point of value before the missing value
f(a) = The constant value before the missing value
f(b) = The constant value after the missing value
a = The constant point before the missing value
b = The constant point after the missing value

The rainfall data is recorded daily through five gauging stations. The data is averaged by the number of stations that have rain based on:

$$\text{AvgRF} = \frac{\text{total\_rain}}{\text{num\_of\_stations\_with\_rain}}$$

Next, the change-point detection technique is applied where there is a transition from gate closed to gate opened. An initial total of 498 of gate opening decisions were detected from 10 years of operation (1997-2006). Once this change-point is detected, a window slice will be formed which includes preceding t days according to the window size to ensure the time delay captured as shown in Table 1 and 2.

Once the rainfall and RWL data is ready, the TDM technique is used to extract the temporal patterns of the reservoir gate opening events. These temporal data usually represent sequences of events which are usually the impacts of certain causes. The temporal information of RWL and rainfall is preserved by using a sliding window technique (Mokhtar *et al*., 2014). Figure 1 and 2 show the delaying effect of the rainfall on RWL.

Next in this study, the data normalization process is applied to the RWL data and RF data which were used as the input data. The gate opening at RWL_t is used as the target output. Data normalization is one of the preprocessing procedures in data mining where the attribute data is scaled so as to fall within a small specified range. In real application because of the differences in the range of attributes' values, one attribute might overpower the other one. Normalization prevents the outweighing attributes with a large range. The goal is to equalize the size or magnitude and the variability of these attributes.

This study applied the z-score normalization where the values for the attributes of RWL (X) and RF (X) are normalized based on the mean and standard deviation of (X). A value of X is normalized to X_{new} by computing:

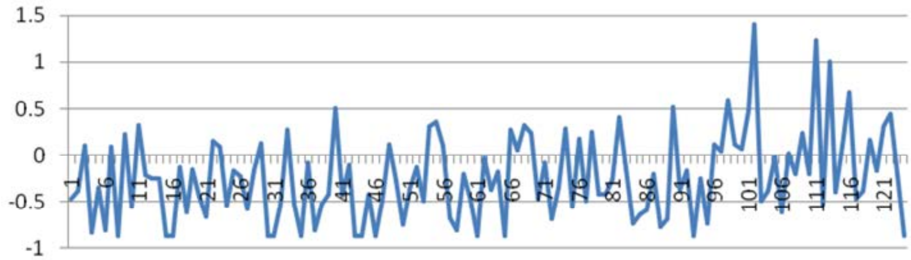

Fig. 1: Rainfall measurement for 10 year

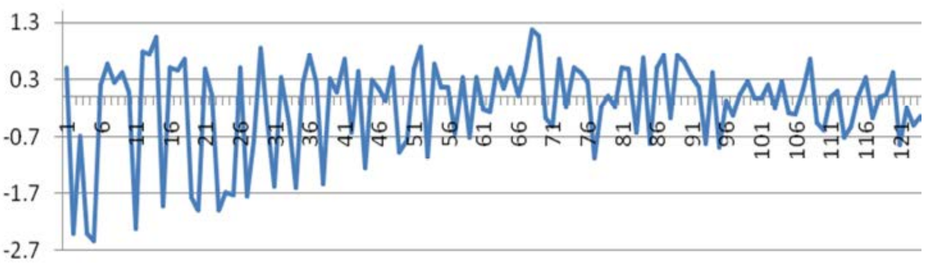

Fig. 2: Water level at TimahTasoh Reservoir for 10 year

Table 3: Sliced Reservoir Water Level (Normalized)

Date	RF_t	RF_{t-1}	RF_{t-2}	RF_{t-3}	RF_{t-4}	RF_{t-5}	RF_{t-6}	RF_{t-7}	RF_{t-8}
12-Feb-97	29.275	0.592	0.475	0.323	0.262	0.253	-0.165	-0.1	-0.106
13-Feb-97	29.335	0.662	0.597	0.511	0.381	0.337	0.336	-0.083	-0.038
14-Feb-97	29.335	0.871	0.667	0.631	0.568	0.454	0.42	0.423	-0.021
15-Feb-97	29.28	0.871	0.875	0.699	0.687	0.638	0.537	0.507	0.486

Table 4: Sliced Averaged Rainfall Data (Normalized)

Date	RF_t	RF_{t-1}	RF_{t-2}	RF_{t-3}	RF_{t-4}	RF_{t-5}	RF_{t-6}	RF_{t-7}	RF_{t-8}
12-Feb-97	20.25	-0.461	-0.617	-0.194	-1.041	1.938	0.559	-0.355	0.150
13-Feb-97	13.88	0.301	-0.502	-0.644	-0.194	-1.037	1.981	0.601	-0.305
14-Feb-97	8.25	-0.075	0.256	-0.528	-0.691	-0.200	-1.043	2.036	0.764
15-Feb-97	1.00	-0.407	-0.118	0.234	-0.563	-0.691	-0.193	-1.015	2.369

$$X_{new} = \frac{X - \bar{X}}{S}$$

Where:

X^{-} = The mean of attribute

S = The standard deviation of the RWL attribute and averaged rainfall.

The normalized values for RWLand RF are shown in Table 3 and 4. This method of normalization is useful because the actual minimum and maximum values of the attributes are unknown.

MATERIALS AND METHODS

Multiple regression is used to explore the relationship between one continuous dependent variable and a number of independent variables or predictors (usually continuous). This study applied multiple regression in order to identify which slices of RWL and RF can best be the input predictors to predict RWL_t.

Table 5: Three significant independent variables on RWLt

Variables	B	Sig.
RWL_{t-1}	0.909*	0
RF_{t-1}	0.003*	0
RF_{t-2}	0.002*	0

The result showed that there is a statistically significant difference in three independent variables on R at $p<0.05$ which are RWL_{t-1}, Rf_{t-1}, Rf_{t-2} as shown in Table 5. However, the regression analysis requires certain assumptions to be fulfilled. First is to determine the adequate sample size. A calculation of the sample size proposed by Tabachnick and Fidell (2007) is used by taking into account the number of independent variables that will be used in this study, based on:

$$N>50+8\text{ m}$$

where, m is the number of independent variables. In this study, there are three independent variables

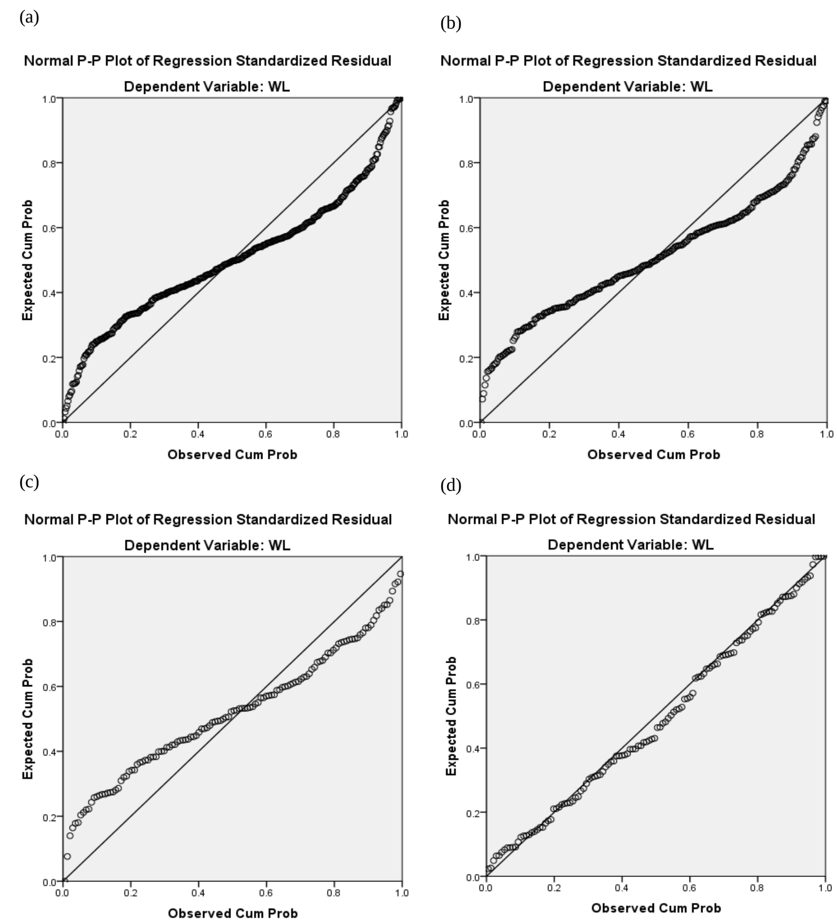

Fig. 3: The predicted water levels as compared to the real water levels

and the sample size used is 498 which exceeds the minimum requirement for the sample size which is 74.

Next is to test on the multicollinearity problems among the independent variables by using a Spearman's correlation coefficient. According to Pallant (2007), multicollinearity exists when the independent variables are highly correlated (r = 0.9 and above). In this study, the correlation coefficients of three independent variables are below 0.9. There was a positive correlation between the three independent variables, RWL_{t-1}, RF_{t-1}, RF_{t-2}, r = 0.894, 0.119 and 0.261, respectively and n = 498, $p<0.005$ with the higher the previous water level and RF associated, the more likely RWL_t increases.

After data collection, the RF and RWL sets are cleaned up from outliers. According to Tabachnick and Fidell (2007), outliers are those withstandardized residual values above 3.3 (or <-3.3). Outliers are checked by inspecting the Mahalanobis distances that are produced by the multiple regression analysis. A critical Chi-square value is determined using the number of independent variables as the degree of freedom. In this study, the critical value is 16.27 for three independent variables (Tabachnick and Fidell, 2007). Thus, the number of sample is reduced from 498 cases to 124 cases after applying the Mahalanobis distances for removing the outliers for three times. Figure 3 shows the normality assessment and transformation of data. (A)-(C) The shape of curves formed by the points is not linear, nor does it remotely match the straight diagonal which both indicate that the data is not normally distributed. (D) A P-P Plot of normally distributed data where a majority of the data points lie along the diagonal line.

RESULTS AND DISCUSSION

From the conducted experiments, the RWL_t can be calculated as:

$$RWL_t = RWL_{t-1}+RF_{t-1}+ RF_{t-2}$$

A regression model for RWL can be calculated as:

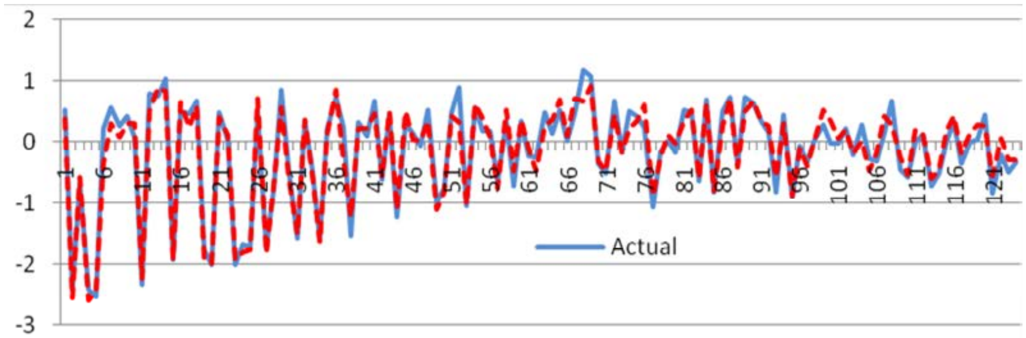

Fig. 4: The predicted water levels as compared to the real water levels

$$RWL = (2.571)+(0.909)RWL_{t-1}+(0.003)RF_{t-1}+(0.002)RF_{t-2}$$

The RWL_t for TimahTasoh can be explained by one day of previous RWL (RWL_{t-1}) and two days of previous rainfall (Rf_{t-1} and RF_{t-2}) with a 0.1628% error. The actual and predicted observation on RWL can be seen in Fig. 4.

Several RWL patterns/models have been tested. The results in terms of error rate are shown in Table 4. Model A which was proposed by Mokhtar *et al*. (2014), utilized five windows of RWL, resulted in 0.093543 error. Model B by Afiqanticipated two days of RF and RWL, resulted in 0.082624 error. Model C by Ishak *et al*. (2011b) suggested eight days of RF, resulted in 0.253844 error and Model D by Norwawi *et al*. (2005) recommended two days of RF, resulted in 0.265880 error.

Based on the result, the predictive model can be considered good where the error is <10% with less number of IVs. The predicted model is considered reliable and can be used by reservoir operators because the predicted RWL lines are very similar to the real RWL in most events.

CONCLUSION

This study has revealed the applicability of the TDM technique for extracting decision patterns from the historical data. A predictive mathematical model can be derived from the extracted patterns where the future decisions can be made by the reservoir operator besides relying on weather events. The sliding window technique has been successfully applied on RWL which can lead to RWR. This information is vital for the reservoir management to plan the early water release.

ACKNOWLEDGEMENTS

The researchers wish to thank the Ministry of Higher Education Malaysia for funding this study under the Long Term Research Grant Scheme (LRGS/b-u/2012/UUM/TeknologiKomunikasidanInformasi) and the Department of Irrigation and Drainage Malaysia for supplying hydrology and reservoir operational data.

REFERENCES

Afshar, A. and A. Salehi, 2011. Gated spillways operation rules considering water surface elevation and flood peak: application to Karkheh Dam. World Environ. Water Resour. Congress, 2011: 3007-3015.

Afshar, M.H., 2012. Large scale reservoir operation by constrained particle swarm optimization algorithms. J. Hydro Environ. Res., 6: 75-87.

Alemu, E.T., R.N. Palmer, A. Polebitski and B. Meaker, 2010. Decision support system for optimizing reservoir operations using ensemble streamflow predictions. J. Water Resour. Plann. Manage., 137: 72-82.

Ashaary, N.A., W.H.W. Ishak and K.K.R. Mahamud, 2015. Neural network application in the change of reservoir water level stage forecasting. Indian J. Sci. Technol., 8: 1-6.

Chiang, J.L. and Y.S. Tsai, 2012. Reservoir drought prediction using support vector machines. Appl. Mech. Mater., 145: 455-459.

Ishak, W.H.W., K.K.R. Mahamud and N.M. Morwawi, 2011a. Conceptual model of intelligent decision support system based on naturalistic decision theory for reservoir operation during emergency situation. Intl. J. Civil Environ. Eng., 11: 6-11.

Ishak, W.H.W., K.K.R. Mahamud and N.M. Norwawi, 2011b. Intelligent Decision Support Model Based on Neural Network to Support Reservoir Water Release Decision. In: Software Engineering and Computer Systems, Jasni, M.Z., M.B.W.M. Wan and E.E. Qawasmeh (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-22169-9, pp: 365-379.

Ishak, W.W., K.R.K. Mahamud and N.M. Norwawi, 2012. Modelling reservoir water release decision using temporal data mining and neural network. Intl. J. Emerg. Technol. Adv. Eng., 2: 422-428.

Mahamud, K.K.R., N. Zakaria, N. Katuk and M. Shbier, 2009. Flood pattern detection using sliding window technique. Proceedings of the 2009 Third Asia International Conference on Modelling and Simulation, May 25-29, 2009, IEEE, Changlun, Malaysia, ISBN:978-1-4244-4154-9, pp: 45-50.

Moeini, R., A. Afshar and M.H. Afshar, 2011. Fuzzy rule-based model for hydropower reservoirs operation. Intl. J. Electr. Power Energy Syst., 33: 171-178.

Mokhtar, S.A., W.H.W. Ishak and N.M. Norwawi, 2014. Modelling of reservoir water release decision using neural network and temporal pattern of reservoir water level. Proceedings of the 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, January 27-29, 2014, IEEE, Sintok, Malaysia, ISBN:978-1-4799-3858-2, pp: 127-130.

Norwawi, M.N., K.K.R. Mahamud and D. Safaai, 2005. Recognition decision-making model using temporal data mining technique. J. ICT., 4: 37-56.

Nwobi, O.C.C. and A.C. Igboanugo, 2013. Predicting water levels at Kainji Dam using artificial neural networks. Nigeria J. Technol., 32: 129-136.

Pallant, J., 2007. SPSS Survival Manual: A Step by Step Guide to Data Analysis using SPSS for Windows. Open University Press, New York, USA.,.

Rani, S. and F. Parekh, 2014. Predicting reservoir water level using artificial neural network. Intl. J. Innov. Res. Sci. Eng. Technol., 3: 14489-14496.

Sattari, M.T., K. Yurekli and M. Pal, 2012. Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. Appl. Math. Modell., 36: 2649-2657.

Stam, A., K.A. Salewicz and J.E. Aronson, 1998. An interactive reservoir management system for Lake Kariba. Eur. J. Oper. Res., 107: 119-136.

Tabachnick, B.G. and L.S. Fidell, 2007. Using Multivariate Statistics. 5th Edn., Pearson Education, Boston, MA., ISBN-13: 9780205465255, Pages: 980.

Wei, C.C., 2012. Discretized and continuous target fields for the reservoir release rules during floods. Water Resour. Manage., 26: 3457-3477.