Investigating the Spatial Relationship between the Upstream Gauging Stations and the Reservoir

Suriyati Abdul Mokhtar and Wan Hussain Wan Ishak
School of Computing, Universiti Utara Malaysia, Kedah, Malaysia
Email: suriyati87@gmail.com, hussain@uum.edu.my

Norita Md Norwawi
Faculty of Science and Technology, Universiti Sains Islam Malaysia, Ng. Sembilan, Malaysia
Email: norita@usim.edu.my

Abstract—**Reservoir operation involves many critical decisions due to unpredicted circumstances, such as drought or flood. The reservoir water release decision is one of the challenging tasks for the reservoir operator since the decision to deal with many complicated decision variables and multipurpose operation. Typically, the water that flow into the reservoir comes from the upstream river network. The gauging stations are located at different locations and the distance between the gauging stations and the reservoir are varied. This information indicates that there are spatial relationships between gauging stations and reservoir. This paper aims to investigate the spatial relationship between the upstream gauging stations and the reservoir by mapping temporal rainfall information on each gauging station with the reservoir water level. The mapping was established using Backpropagation Neural Network. The findings show that different location and distance of each gauging station with the Timah Tasoh reservoir does affect the water travel time from gauging stations to the reservoir.**

Index Terms—**backpropagation neural network, spatial-temporal data, reservoir water level forecasting**

I. Introduction

The reservoir is one of the structural flood mitigation approaches that aim to store water during heavy rainfall in order to maintain safe discharges at the downstream areas [1]. The reservoir can be defined as a natural or artificial lake, or pond where water is collected and stored that are used for multipurpose operation such as water supply, flood control, hydropower generation and other purposes like agricultural and recreation [2]. The reservoir water release decision is one of the challenging tasks for the reservoir operator in order to determine the quantities of water to be stored and to be released from a reservoir [3].

Typically, in both flood and drought situations, decisions regarding water release is made in accordance with the available water, inflows, demands, time, previous release and etc [4]. However, different reservoir has different objectives and purposes, thus required different operation rule [5]. Traditionally, the reservoir was operated based on the standard operating procedure (SOP) which provides reservoir engineer the guideline on how the reservoir should be operated. However, the reservoir operation is also influenced by several factors such as sediment, water usage, climate change, and urbanization at reservoir upstream. These factors reduce the efficacy of SOP and the operation rule. Eventually, reservoir operator has to apply heuristic procedures by embracing rule curves and subjective judgements [6]. Through these practices the reservoir operational database has become an experience rich repository where dependent decisions were recorded based on past human experience.

Typically, the water that flow into the reservoir comes from the upstream river network. The gauging station is one of the tools that record the river water level and rainfall. The gauging stations are located at different locations and the distance between the gauging stations and the reservoir are varied. This information indicates that there are spatial relationships between gauging stations and reservoir. However, as the locations of gauging stations are static, taking the location data into consideration is useless in a forecasting model. Nevertheless, the distance between the gauging stations and reservoir causes delay where rainfall recorded at gauging stations may take some time to reach the reservoir.

The focus of this paper is to investigate the spatial relationship between the upstream gauging stations and the reservoir by mapping temporal rainfall information on each gauging station with the reservoir water level. The mapping was established using Backpropagation Neural Network.

II. Literature Review

Data mining is a multidisciplinary field involves various research areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, high-performance computing, and data visualization [7]. Data mining is a part of the overall process of Knowledge Discovery in Database (KDD).

Manuscript received May 5, 2015; revised July 8, 2015.

doi: 10.18178/joams.4.6.503-506

Temporal data mining is one of the popular topics in data mining research. In general, temporal data mining is an important extension of the data mining techniques and it can be defined as a process of extracting implicit, potentially useful information from temporal database [5]. Wan Ishak *et al.* [2], [5] and Mokhtar *et al.* [8] for example apply temporal data mining, specifically sliding window technique to segment temporal pattern from reservoir operational database. Mohan and Revesz [9] apply temporal data mining to manage uncertain water reservoir data.

Spatial data mining is the process of discovering interesting and useful patterns from large spatial datasets [10]. The process of extracting a pattern from spatial data sets is more difficult compared to traditional numeric and categorical data. Manikandan dan Srinivasan [11] presented spatial data mining and Prim's Algorithm for mining spatially co-located moving objects. They use R-TREE for mining the spatial co-location patterns in order to reduce computation time. The results of this study show the proposed techniques performed better in terms of time and memory space.

Spatial temporal data mining or also known as spatio-temporal data mining is an integration between spatial and temporal data mining that focuses on finding spatial and temporal relationships from spatio-temporal dataset [10]. Two important attributes of spatial temporal data mining are location and time where location refers to spatial relationships and time refers to temporal relationship of the data.

Rashid and Hossain [12] discuss the challenging issues of spatial-temporal data mining. The ability for analyzing the huge amount of data is still inadequate and there is a need for adapting data mining tools. Currently, spatial-temporal data mining has been widely discussed in reservoir operation and management research [9], [13].

III. Methodology

In this study, the Timah Tasoh reservoir was used as a case study. The Timah Tasoh reservoir is one of the largest multipurpose reservoirs in Northern Peninsular Malaysia. Timah Tasoh located on Sungai Korok in the state of Perlis, about 2.5km below the confluence of Sungai Timah and Sungai Tasoh. The Timah Tasoh reservoir serves as flood mitigation in conjunction to other purposes: water supply, irrigation, and recreation.

In this study daily reservoir water level (WL) and rainfall (RF) data from five upstream gauging stations from 1999-2012 are used. The gauging stations are Padang Besar(PB), Tasoh (TH), Lubuk Sireh (LS), Kaki Bukit (KB) and Wang Kelian (WK) (Fig. 1). The data is preprocessed and normalized into the range -1 and 1. Standard backpropagation neural network with bias, learning rate and momentum is used to model the relationship between the RF from upstream gauging stations and the reservoir WL. The modeling is based on the temporal pattern of RF. The temporal information of the reservoir water level data is preserved by using a sliding window technique [14]. This process is called segmentation process.

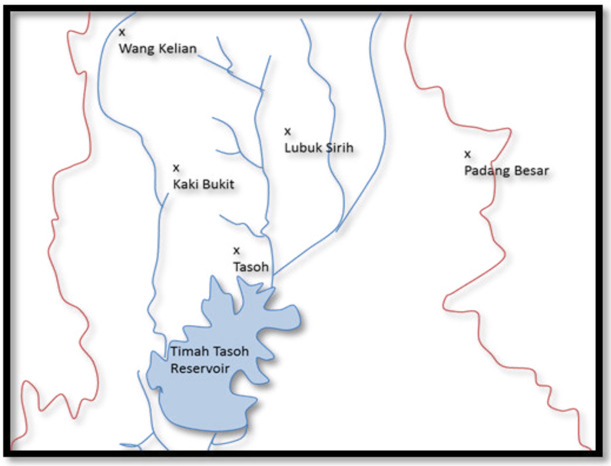

Figure 1. Timah Tasoh reservoir and five gauging stations.

In this study five data sets have been formed for each gauging station. Each data set represents the different window size. Each window size represents time duration of the delays. For example, window size (w) 3 represents three days of delays. Each data set consists of N number of input columns and 2 output columns. The N is equal to the window size that represent the RF at different time t, t-1, t-2, t-w, where t represent the time and w is the window size. The output is the reservoir WL at t is either normal (0), alert (1), warning (2) and danger (3). The output is encoded based on Binary-Coded-Decimal (BCD) scheme [15]. Table I shows a BCD representation of each output class.

TABLE I. Output Coding Using BCD

Output Class	BCD Representation
0	-1,-1
1	-1,1
2	1,-1
3	1,1

Table II summarizes the number of instances extracted for each data set. Redundant and conflicting instances are then removed.

TABLE II. Dataset and the Number of Instances by Gauging Stations

Data Set	Window Size	Padang Besar	Tasoh	Lubuk Sireh	Kaki Bukit	Wang Kelian
1	2	25	25	25	25	25
2	3	110	106	107	100	114
3	4	313	296	336	274	386
4	5	665	620	755	587	910
5	6	1125	779	1355	1034	1576

Each data set is then divided randomly into three data sets: Training set (80%), validation set (10%) and testing set (10%). The training set is used in the training phase of NN, while validation set is used to validate the NN performance during the training. Testing set is used to test the performance of NN after the training has completed.

In this study, five NN models were developed for each gauging station. Each neural network model is trained with one data set. Each model is trained with different combination of hidden unit, learning rate and momentum [16]. The training is controlled by three conditions (1) maximum epoch (2) minimum error, and (3) early stopping condition.

IV. Findings and Discussion

Table III summarizes the best results of training, validation and testing for the data sets that represent each gauging station. The result reveals that the best window size for PB, TH, LS, KB and WK is 6, 3, 4, 3, and 6 respectively. This indicates that the rainfall recorded at each gauging station took 6, 3, 4, 3, and 6 days respectively to give effect to the Timah Tasoh reservoir. This is due to the different location and distance of each gauging station with the Timah Tasoh reservoir. Values for the NN parameters of each gauging stations that were achieved after the training phase are shown in Table IV.

TABLE III. Summary of the Findings

Station	*w*	Training			Validation			Testing		
		MSE	%		MSE	%		MSE	%	
PB	6	0.3582	97.44	81.76	0.3099	98.23	83.19	0.3941	96.46	80.53
TH	3	0.1906	95.24	95.24	0.7558	81.82	81.82	0.8431	72.73	81.82
LS	4	0.3651	98.52	78.76	0.3563	96.32	83.09	0.3581	97.79	80.15
KB	3	0.3752	91.25	90	0.3198	100	80	0.3994	90	90
WK	6	0.3550	98.33	80.08	0.4030	95.57	80.38	0.3525	97.47	80.38

TABLE IV. Summary of NN Specification

Station	#Input	#Output	Hidden units	Learning Rate	Momentum
PB	6	2	7	0.2	0.2
TH	3	2	17	0.4	0.2
LS	4	2	19	0.1	0.4
KB	3	2	11	0.6	0.1
WK	6	2	15	0.2	0.1

Based on the findings, the spatial-temporal representation of the rainfall (RF) at each gauging station and the reservoir water level (WL) can be expressed as follows:

$$WL(t+1) = f(RF_{PB}(t), RF_{PB}(t+1), RF_{PB}(t+2), RF_{PB}(t+3), RF_{PB}(t+4), RF_{PB}(t+5), RF_{TH}(t), RF_{TH}(t+1), RF_{TH}(t+2), RF_{LS}(t), RF_{LS}(t+1), RF_{LS}(t+2), RF_{LS}(t+3), RF_{KB}(t), RF_{KB}(t+1), RF_{KB}(t+2), RF_{WK}(t), RF_{WK}(t+1), RF_{WK}(t+2), RF_{WK}(t+3), RF_{WK}(t+4), RF_{WK}(t+5), WL(t))$$

where WL (t+1) is the reservoir water level at time t+1 and RF_{PB}, RF_{TH}, RF_{LS}, RF_{KB}, and RF_{WK} represent the rainfall at gauging stations of the Padang Besar, Tasoh, Lubuk Sireh, Kaki Bukit, and Wang Kelian respectively. Besides WL (t) represent the reservoir water level at time t. Overall, there are 22 input variables in this model.

V. Conclusion

In this study the spatial-temporal model has been established for the rainfall recorded at the upstream gauging stations and the reservoir water level. The findings show that the different location of gauging stations has an effect in term of delay of inflow into the reservoir, thus affect the reservoir water level.

The spatial-temporal model established in this study can be used in water release decision making. Reservoir operator can use the model to forecast the future water level and decide early water release so that reservoir can have enough space for incoming inflow. In addition, the water release can be controlled within the safe carrying capacity of downstream river. Thus the flood risk downstream due to extreme water release from the reservoir can be reduced.

Acknowledgment

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

References

[1] K. Smith and R Ward, *Floods: Physical Processes and Human Impacts*, England: John Wiley, 1998, ch. 7, pp. 205-237.

[2] W. H. Wan Ishak, K. R. Ku Mahamud, and M. N. Norwawi, "Intelligent decision support model based on neural network to support reservoir water release decision," in *Proc. ICSECS 2011,* Part I, Communications in Computer and Information Science (CCIS), 2011, pp. 365-379.

[3] R. A. Wurbs, "Reservoir-system simulation and optimization models," *Journal of Water Resources Planning and Management*, vol. 119, no. 4, pp. 455-472, 1993.

[4] S. K. Jain and V. P. Singh, "Reservoir operation," in *Water Resources Systems Planning & Management*, S. K. Jain and V. P. Singh, ed. 2003, pp. 615-679.

[5] W. H. Wan Ishak, K. R. Ku Mahamud, and N. M. Norwawi, "Modelling reservoir water release decision using temporal data mining and neural network," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 8, pp. 422-428, 2012.

[6] L. L. Ngo, "Optimising reservoir operation: A case study of the Hoa Binh Reservoir, Vietnam," PhD Thesis, Institute of Environment & Resources Technical University of Denmark, 2006.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 2nd, Morgan Kaufmann, 2006, pp. xxi.

[8] S. A. Mokhtar, W. H. Wan Ishak, and N. M. Norwawi, "Modelling of reservoir water release decision using neural network and temporal pattern of reservoir water level," in *Proc. the Fifth International Conference on Intelligent Systems, Modelling and Simulation*, 2014, pp. 127-130.
[9] A. Mohan and P. Revesz, "Temporal data mining of uncertain water reservoir data," in *Proc. the Third ACM SIGSPATIAL International Workshop on Querying and Mining Uncertain Spatio-Temporal Data*, 2012, pp. 10-17.
[10] D. Kalyani and S. K. Chaturvedi, "A Survey on spatio-temporal data mining," *International Journal of Computer Science and Network,* vol. 1, no. 4, 2012.
[11] G. Manikandan and S. Srinivasan, "An efficient algorithm for mining spatially co-located moving objects," *American Journal of Applied Sciences*, vol. 10, no. 3, pp. 195-208, 2013.
[12] A. N. M. B. Rashid and M. A. Hossain, "Challenging issues of spatio-temporal data mining," *Computer Engineering and Intelligent Systems*, vol. 3, no. 4, pp. 55-63, 2012.
[13] F. J. Chang and Y. T. Chang, "Adaptive neuro-fuzzy inference system for prediction of water level in reservoir," *Advances in Water Resources*, vol. 29, no. 1, pp. 1-10, 2006.
[14] W. H. Wan Ishak, K. R. Ku Mahamud, and N. M. Norwawi, "Mining temporal reservoir data using sliding window techniques," *CiiT International Journal of Data Mining Knowledge Engineering*, vol. 3, no. 8, pp. 473-478, 2011.
[15] C. C. Chong and J. C. Jia, "Assessments of neural network output codings for classification of multispectral images using hamming distance measure," in *Proc. the 12th IAPR International. Conference on Pattern Recognition*, 1994, pp. 526-528.
[16] F. Mat-Yamin, W. H. W. Ishak, and S. N. Othman, "Procedure in developing neural network model," presented at the International Conference on Technology Management (ICTM), Palm Garden Hotel, IOI Resort, Putrajaya, Malaysia, December 4-5, 2006.

Suriyati Abdul Mokhtar is currently completing her MSc research study at Universiti Utara Malaysia. Prior to that she obtain her bachelor degree in Information Technology from Universiti Utara Malaysia.

Wan Hussain Wan Ishak received the Bachelor in Information Technology and Master of Science in Information Technology from Universiti Utara Malaysia in 2000 and 2003 respectively. Currently, he is a Senior Lecturer at Universiti Utara Malaysia.

Norita Md Norwawi is a Professor at Universiti Sains Islam Malaysia. She obtained her Bachelor in Computer Science in 1987 from the University of New South Wales, Australia. She received her Masters degree in Computer Science from National University of Malaysia in 1994. In 2004, she obtained her PhD specializing in Temporal Data Mining and Multiagent System from Universiti Utara Malaysia. As an academician, her research interests include artificial intelligence, multi-agent system, temporal data mining, text mining, knowledge mining, information security and digital Islamic application and content. Her works have been published in international conferences, journals and won awards on research and innovation competition in national and international level.