Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), Vol. 75, No. 2, I_841-I_846, 2019.

INTEGRATING SWITCHED PREDICTION METHOD AND MULTIPLE ARTIFICIAL NEURAL NETWORKS FOR PREDICTING INFLOW AND SEDIMENT INTO RESERVOIRS

Sameh A. KANTOUSH[1], Peng An CHEN[2], Jihn Sung LAI[3], Gwo Fong LIN[4] and Tetsuya SUMI[5]

[1]Member of JSCE, Associate Professor, Water Resources Research Center, Disaster Prevention Research Institute, Kyoto University (Goka-sho, Uji-shi, 611-0011, Japan)
E-mail: kantoush.samehahmed.2n@kyoto-u.ac.jp
[2]Member of JSCE, Doctoral Student, Water Resources Research Center, Disaster Prevention Research Institute, Kyoto University (Goka-sho, Uji-shi, 611-0011, Japan)
E-mail: a0932195109@gmail.com
[3]Research Fellow, Hydrotech Research Institute, National Taiwan University (No. 158 Jhoushan Rd., Taipei,10617, Taiwan)
E-mail: jslai525@gmail.com
[4]Distinguished Professor, Department of Civil Engineering, National Taiwan University (No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617, Taiwan)
E-mail: gflin@ntu.edu.tw
[5]Member of JSCE, Professor, Water Resources Research Center, Disaster Prevention Research Institute, Kyoto University (Goka-sho, Uji-shi, 611-0011, Japan)
E-mail: sumi.tetsuya.2s@kyoto-u.ac.jp

Accurate predictions of the reservoir inflow and sediment concentration are necessary for real-time reservoir operation. This study used multiple artificial neural networks (ANNs), namely the back propagation networks (BPN) and four types of kernel function of support vector machines (SVM), to predict inflow and sediment concentration. These ANNs were calibrated and validated based on observed data of the Shihmen reservoir for typhoon events from 2008 to 2015. To avoid the risk of selecting multiple ANNs, the switched prediction method (SPM) is proposed to select the optimal predicting module time by time. This paper compares the predictions from SPM with optimal individual predictions and the ensemble means (EM) with respect to the root mean square error. The improvements in SPM compared with optimal individual ANN and EM are 3.8% and 10%, respectively. In conclusion, the uncertainty of the predictions could be effectively reduced by applying the switched prediction method.

Key Words : *reservoir inflow predicting, multiple artificial neural networks; switched prediction method*

1. INTRODUCTION

During typhoons and extreme flood events, the heavy rainfall always brings a large amount of flow rate (Q) and sediment concentration (S) into reservoirs. It causes reservoir sedimentation and reduces capacity[1)]. Chamoun et al. (2018)[2)] reveal that the Q and S significantly affect the behavior of turbidity current evolution. As the Q reduces, the turbidity current retained and silted. Hence, the reservoir Q and S are crucial for simulating the process of turbidity current and determining the optimal venting operation.

Due to the long computation time and the lack of observations, physically based mathematical models are not available for real-time predicting. Based on the aforementioned shortcomings, Artificial Neural Networks (ANNs) with superior ability in nonlinear processes have been applied in hydrology fields[3,4)]. ANNs, such as back propagation networks (BPN) and four types of kernel function of support vector machines (SVM) have been applied in hydrology and sediment transportation[5-8)]. To avoid the risk of the selection from multiple ANNs, a switched prediction method, which significantly improves the model generalization, is proposed[9)]. Chen et al. (2019)[10)] reveal that the switched prediction method could select

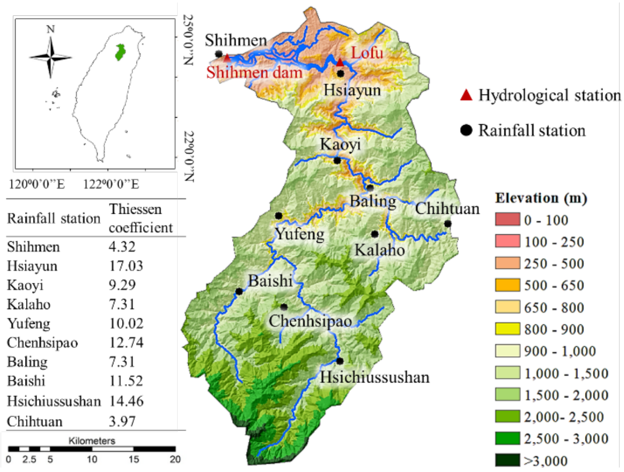

Fig.1 Locations and Thiessen coefficient of rainfall stations in the Shihmen reservoir catchment.

high-performance ANNs and filter out the noise for predicting reservoir inflow sediment concentration.

This paper proposes a model based on integrating switched prediction method (SPM) and multiple ANNs for predicting Q and S into the reservoir. These ANNs were calibrated and validated based on observed data in Shihmen reservoir for the typhoon events from 2008 to 2015. The root mean square error (RMSE) is used to evaluate the predictions from each ANN. To investigate the performance of each ANN, Typhoon Soulik, with high Q and S, is selected to indicate the comparison between the predictions with SPM to optimal individual predictions and the EM.

2. STUDY AREA & DATA COLLECTION

The Shihmen reservoir is located in the middle reach of the Dahan River in northern Taiwan with a catchment area of 763.4 km^2. The original design storage capacity is 309 million m^3 at the normal water level of EL. 245 m. Due to extensive reservoir sedimentation, almost 34% of storage capacity has been reduced. The locations and Thiessen coefficient of the ten rainfall stations are shown in **Fig.1**. The inflow and sediment concentration were measured in the Lofu hydrological station, which is located at the upstream boundary of Shihmen reservoir. **Table 1** presents the summary of hydrological data, including the date of the occurrence, duration, peak and average Q and S, in typhoon events from 2008 to 2015.

3. METHODOLOGY DEVELOPMENT

To accurately predict the reservoir Q and S, the proposed model by integrating the multiple ANNs and SPM is developed. The BPN, with 2 layers and 2 nodes in each layer, and four types of kernel function of SVM are conducted to predict Q and S for 1 to 3 hours lead times.

Table 1 Hydrological data in typhoon events from 2008 to 2015.

Event	Date	Inflow (m^3/s)		Sediment Concentration (ppm)	
		Peak	Average	Peak	Average
Fung-Wong	2008/7/26	2040	642	28235	8350
Sinlaku	2008/9/11	3447	826	37254	8970
Morakot	2009/8/5	1838	820	23864	8601
Saola	2012/7/30	5589	983	39100	9255
Soulik	2013/7/11	5458	955	86833	10807
Trami	2013/8/20	2412	942	72594	10915
Soudelor	2015/8/6	5634	946	18287	10322
Dujuan	2015/9/29	3802	970	17515	10288

Four kernel functions, including linear (LN), polynomial (PL), radial basis function (RBF) and sigmoid (SIG), are expressed in equations 1 to 4 as follows:
LN:

$$K(x_i, x_j) = x_i^T \cdot x_j \tag{1}$$

PL:

$$K(x_i, x_j) = (\gamma \cdot x_i^T \cdot x_j + r)^d, \gamma > 0 \tag{2}$$

RBF:

$$K(x_i, x_j) = (-\gamma ||x_i - x_j||), \gamma > 0 \tag{3}$$

SIG:

$$K(x_i, x_j) = \tanh(\gamma \cdot x_i^T \cdot x_j + r) \tag{4}$$

where γ is the gamma term, r is the bias term, and d is the polynomial degree term. With the above kernel function, the four types of SVM are named SVM-LN, SVM-PL, SVM-RBF and SVM-SIG.

The observed rainfall (R), Q and S are regarded as input candidates, then, the optimal input is determined by the correlation analysis. With the sorted factors, ranking by Pearson product-moment correlation coefficient, the number of input could be determined. The RMSE and correlation coefficient (CC) values were used as the performance criteria for optimizing the multiple ANNs. The results with minimum RMSE values or maximum CC values are regarded as the optimal input of the predicting module.

The cross validation[11)] is used to objectively evaluate the performance of each ANN. Each single typhoon event is selected as the testing set in turn, and other events are regarded as the training sets. Performance conclusions are then drawn on the basis of the averaged RMSE values by all testing events.

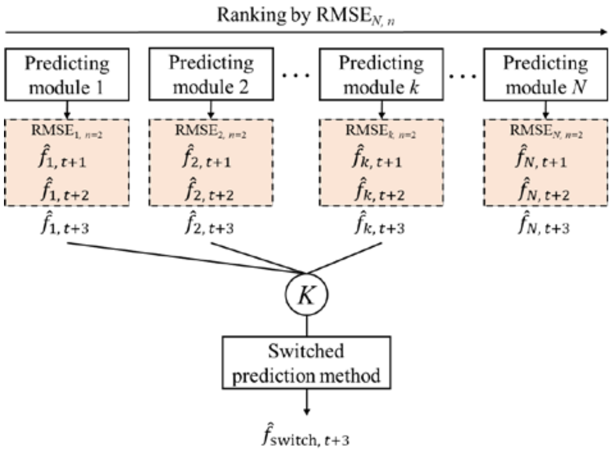

Fig.2 Structure of the switched prediction method.

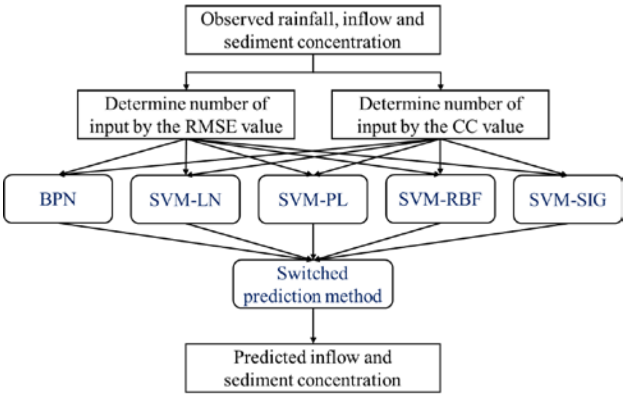

Fig.3 Flowchart of the developed combine models.

It is a difficult task to objectively select the optimal predicting module. Hence, this paper proposed the SPM for selecting the high performance predicting module time by time. The structure of the SPM is depicted in **Fig.2**. The model assumes that there are N predicting modules, $\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_N$. At time t+3, the k candidate ANNs, with the small RMSE in the past n hour, are averaged to be the switched predictor, $\hat{f}_{\text{Switch},t+3}$. The grid-search method[11)] is conducted to determine the optimal parameters. Based on the aforementioned process, the flowchart of the proposed model, integrating multiple ANNs and SPM, is shown in **Fig.3**. To considerable show the results of the improvement due to the application of SPM, the improvement percentage (IP) is calculated and expressed as follows:

$$\mathrm{IP} = \frac{R_{proposed} - R_{previous}}{R_{proposed}} \times 100\% \tag{5}$$

where $R_{proposed}$ and $R_{previous}$ is the RMSE of proposed and previous, respectively.

4. RESULTS AND DISCUSSION

(1) Comparison of the multiple ANNs

To identify the optimal input for each ANN, the Pearson product-moment correlation coefficient was examined and sorted (listed in **Table 2**). The determination on the number of input from the various input candidates over the RMSE and CC values.

The optimal number of input for Q and S are listed in **Table3**. The predicting module with different inputs, determined by RMSE and CC, are renamed (**Table 3**). To highlight the results of each ANN in high Q and S, the results from multiple ANNs in Typhoon Soulik is shown in **Fig.4**. It shows that the predictions from the SVM-PL are inaccurate at both peak and low Q. Meanwhile, the SVM-PL (1), SVM-PL (2) and BPN are seriously under predicted at the peak S. Due to the poor performance of these modules, it affects the performance of the EM.

Table 2 Pearson product-moment correlation coefficient for the inflow and sediment concentration predicting inputs.

Rank	Inflow predicting		Sediment concentration predicting	
	Factor	CC	Factor	CC
1	Q_t	0.97	S_t	0.85
2	Q_{t-1}	0.90	Q_t	0.76
3	R_{t-4}	0.86	S_{t-1}	0.73
4	R_{t-3}	0.85	R_{t-3}	0.73
5	R_{t-5}	0.85	R_{t-4}	0.73
6	R_{t-2}	0.81	R_{t-5}	0.72
7	R_{t-6}	0.81	R_{t-2}	0.70
8	Q_{t-2}	0.81	Q_{t-1}	0.70
9	R_{t-7}	0.77	R_{t-6}	0.69
10	R_{t-1}	0.74	R_{t-8}	0.68

Table 3 The number of factors as input to the multiple ANNs for inflow and sediment concentration prediction.

Predicting module	Determined performance criteria	Inflow predicting		Sediment concentration predicting	
		Number of input	Named	Number of input	Named
BPN	RMSE	7	BPN (1)	1	BPN
	CC	4	BPN (2)	1	
SVM-LN	RMSE	7	SVM-LN (1)	2	SVM-LN
	CC	8	SVM-LN (2)	2	
SVM-PL	RMSE	8	SVM-PL	1	SVM-PL (1)
	CC	8		2	SVM-PL (2)
SVM-RBF	RMSE	7	SVM-RBF (1)	1	SVM-RBF (1)
	CC	8	SVM-RBF (2)	2	SVM-RBF (2)
SVM-SIG	RMSE	8	SVM-SIG	2	SVM-SIG
	CC	8		2	

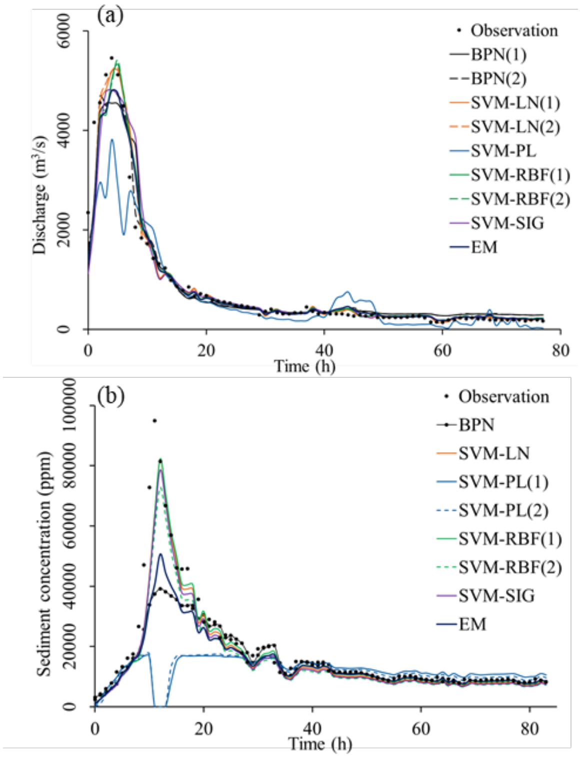

Fig.4 Comparison of the observed data with predictions resulting from multiple ANNs and ensemble means for (a) Inflow and (b) sediment concentration at the 1 hour lead time.

Table 4 The RMSE values obtained from multiple ANNs.

Predicting module	Lead time (h) 1	2	3
Inflow (m^3/s)			
BPN(1)	196.18	287.45	358.51
BPN(2)	177.48	295.00	383.22
SVM-LN(1)	169.91	291.95	369.22
SVM-LN(2)	172.67	297.52	372.53
SVM-PL	600.44	586.00	568.51
SVM-RBF(1)	163.20	268.72	331.98
SVM-RBF(2)	166.50	277.43	339.77
SVM-SIG	210.00	309.61	381.62
EM	176.94	278.51	341.70
SPM	157.87	255.69	317.70
Sediment concentration (ppm)			
BPN	5459.13	6639.07	7395.70
SVM-LN	4195.30	5738.19	6885.43
SVM-PL(1)	8119.89	8383.93	8736.32
SVM-PL(2)	8025.94	8272.45	8681.35
SVM-RBF(1)	4252.90	5867.98	7143.31
SVM-RBF(2)	4329.96	5782.62	6924.04
SVM-SIG	4219.93	5743.81	6892.95
EM	4936.54	6119.43	7093.58
SPM	4148.59	5559.30	6480.84

As shown in Table 4, the SVM-RBF (1) and SVM-LN are the individual optimal predicting module for Q and S predicting, respectively.

(2) Improvement due to the application of SPM

To select objectively the optimal predict module time by time, the SPM is proposed. Firstly, identification of optimal parameters, *n* and *k* are examined based on various parameters combinations. With the grid-search method, we varied *n* and *k* between one and five. The optimal parameters for accurate prediction of Q and S are presented in **Table 5**.

Once the optimal parameters identified, the results of SPM are compared with those of EM and individual optimal predicting module as well. **Fig.5** clearly shows that the performance of SPM and individual predicting module are better than EM at the peak inflows. Moreover, the performance of SPM is better than EM and individual predicting module at the turning place, especially in long lead time predicting.

Fig.6 presents the comparison of the measured and predicted S based on SPM, EM and the individual optimal predicting module for 1 to 3 hours lead times. Similar results to Q prediction, the EM yield the poor performance at the peak S. Due to the predictions from SVM-LN in high S is better than other ANNs.

Compare to the individual predicting module; the SPM produces reliable predictions, accurately matching the measured S, especially at low S. **Table 4** clearly indicates RMSE values from SPM are the lowest (i.e., best); it means that the application of SPM could effectively increase accuracy. Due to the predictions are obtained by the previously observed data, it is difficult to predict the turning place. **Table 6** summarizes the IP for 1 to 3 hours lead times in Q and S. The average of the IP of SPM instead of individual predicting module and EM is 3.75% and 9.96%, respectively. The improvement shows that the SPM could select high performance predicting module.

Table 5 Parameters used in the SPM.

Lead time (h)	Inflow (m^3/s) *n*	*k*	Sediment concentration (ppm) *n*	*k*
1	1	3	2	3
2	1	2	1	3
3	1	5	1	2

Table 6 The improving percentage for the SPM instead of the individual optimal predicting module and EM.

Lead time (h)	Improvement percentage (%) SPM instead of individual optimal predicting module	SPM instead of ensemble means
Inflow (m^3/s)		
1	3.26	10.78
2	4.85	8.19
3	4.30	7.02
Sediment concentration (ppm)		
1	1.11	15.96
2	3.12	9.15
3	5.88	8.64
Average	**3.75**	**9.96**

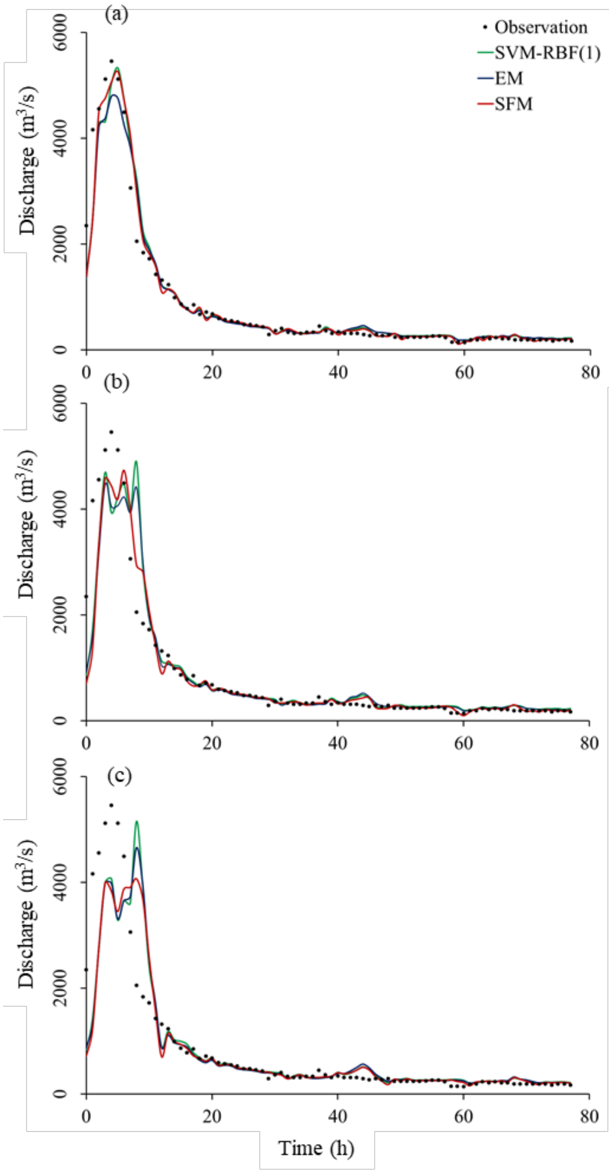

Fig.5 Comparison of the observed Q with Q predictions resulting from the individual predictions (SVM-RBF (1)) and ensemble means at the (a) 1 hour, (b) 2 hour and (c) 3 hour lead time.

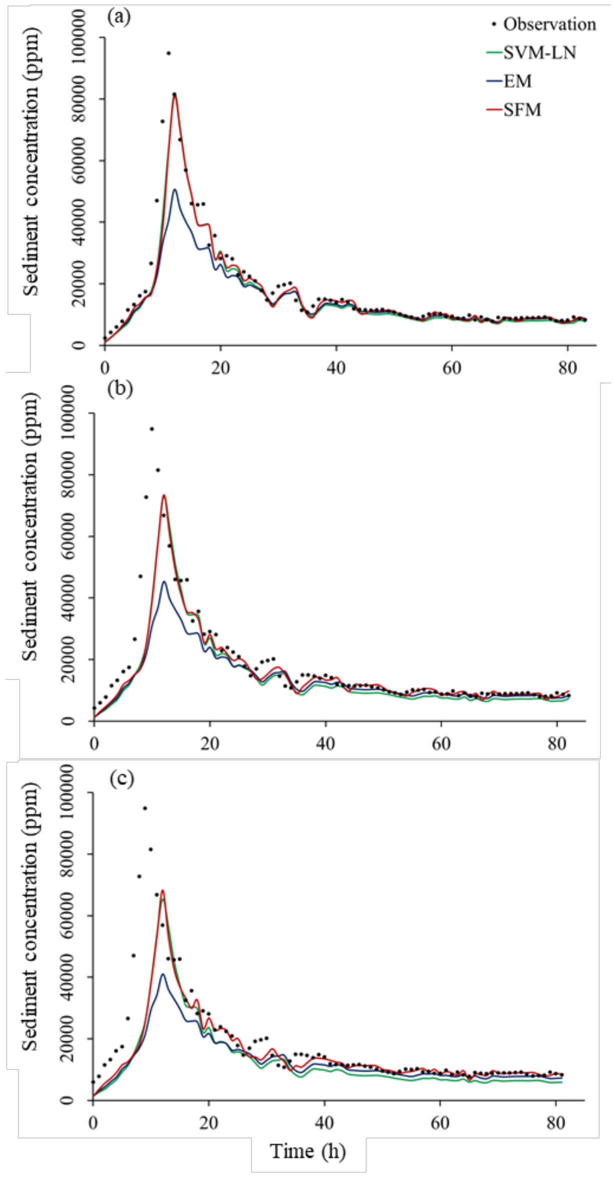

Fig.6 Comparison of the observed S with S predictions resulting from the individual predictions (SVM-LN) and ensemble means at the (a) 1 hour, (b) 2 hour and (c) 3 hour lead time.

(3) Coupling ANN and 3D numerical modeling for turbidity current simulation

Turbidity currents venting through the dam is an effective strategy to reduce reservoir sedimentation. To understand the propagation of the turbidity current movement within the reservoir, the predicted inflow boundary conditions, Q and S, are essential to simulate future scenarios. Based on the proposed model in this paper, the high accuracy hourly S and Q could be regarded as the inflow boundary condition of the 3D numerical model based on TELEMAC 3D. Due to the Q and S significantly affect the behavior of turbidity current evolution, the accuracy predictions could increase the accuracy of the real-time simulation. The ANNs can predict the velocity and arrival time of turbidity current and outflow sediment concentration. However, the ANN cannot predict the whole process of the turbidity current and the ANN lack of the physical theory. In the future, the 3D numerical model is crucial to solve the aforementioned problems. The model setup and parameter determination will be calibrated based on the physical model.

Fig. 7 describes the future study based on coupling machine learning, physical processes from the field and 3D numerical model. Firstly, the multiple predicting modules will be developed and using switched prediction method to yield the accuracy predicted Q and S. Then, the 3D numerical model will be constructed in step 2 for determining the venting operation timing and optimal venting operation from different sediment bypass tunnel and different height of outlets at the dam.

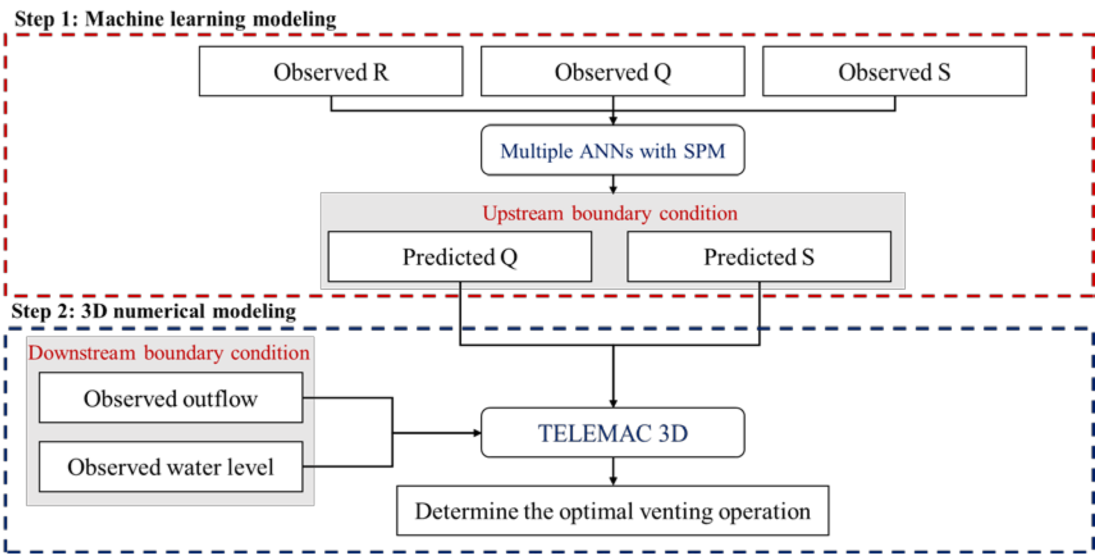

Fig.7 Flow chart of coupling ANN and 3D numerical modeling for turbidity current simulation

5. CONCLUSIONS

Since the construction of the reservoir, the estimation of inflow and associated suspended sediment concentration in the Shihmen reservoir is challenging during the typhoon events. The accurate prediction is necessary for turbidity current simulation and optimizing the useful life of the reservoir. This paper used multiple ANNs, BPN and four types of kernel function of SVM, to establish predictors. The different ANNs and different inputs could lead to different predictions. In general, EM is the objective way to obtain the prediction. However, each ANN has its weak points in a certain situation, which causes the poor performance of EM.

To reduce the risk of selecting the ANNs and input, the SPM is proposed. The results indicate that SPM is capable to yield accurate hourly S and Q as compared with the optimal individual predictions and the EM for 1 to 3 hours lead times. Based on the aforementioned results, using the SPM could effectively select high performance modules and filter out the noises. Hence, the uncertainty of the predictions could be effectively reduced by using the SPM and help the turbidity current simulation in reservoirs.

ACKNOWLEDGMENT: The authors thank the support of Taiwan Water Resources for providing data. The authors also acknowledge the Hydrotech Research Institute, National Taiwan University for providing technique support.

REFERENCES

1) Pandey, A., Chaube, U. C., Mishra, S. K. and Kumar, D. : Assessment of reservoir sedimentation using remote sensing and recommendations for desilting Patratu Reservoir, India, *Hydrol. Sci. J.*, Vol. 61, No. 4, pp. 711-718, 2016.
2) Chamoun, S., De Cesare G. and Schleiss, A. J. : Influence of Operational Timing on the Efficiency of Venting Turbidity Currents, *J. Hydraul. Eng.*, Vol. 144, No. 9, pp. 04018062, 2018.
3) Kokpinar, M. A., Altan-Sakarya, A. B., Kumcu, S. Y. and Gogus, M. : Assessment of sediment yield estimations for large watershed areas: a case study for the Seyhan, Demirköprü and Hirfanlı reservoirs in Turkey, *Hydrol. Sci. J.*, Vol. 60, No. 12, pp. 2189-2203, 2015.
4) Choubin, B., Khalighi-Sigaroodi, S., Malekian, A. and Kişi, Ö. : Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals, *Hydrol. Sci. J.*, Vol. 61, No. 6, pp. 1001-1009, 2016.
5) Yadav, A., Chatterjee, S. and Equeenuddin, S. M. : Suspended sediment yield estimation using genetic algorithm-based artificial intelligence models: case study of Mahanadi River, India, *Hydrol. Sci. J.*, Vol. 63, No. 8, pp. 1162-1182, 2018.
6) Kokpinar, M. A., Altan-Sakarya A. B., Kumcu, S. Y. and Gogus, M. : Assessment of sediment yield estimations for large watershed areas: a case study for the Seyhan, Demirköprü and Hirfanlı reservoirs in Turkey, *Hydrol. Sci. J.*, Vol. 60, No. 12, pp. 2189-2203, 2015.
7) Chang, M. J., Chang, H. K., Chen, Y. C., Lin, G. F., Chen, P. A., Lai, J. S. and Tan, Y. C. : A Support Vector Machine Forecasting Model for Typhoon Flood Inundation Mapping and Early Flood Warning Systems, *Water*, Vol. 10, No. 12, pp. 1734, 2018.
8) Sun W. and Trevor B. : Multiple model combination methods for annual maximum water level prediction during river ice breakup, *Hydrol. Process.*, Vol. 32, pp. 421-435, 2018.
9) Lian, C., Zeng, Z., Yao, W. and Tang, H. : Multiple neural networks switched prediction for landslide displacement, *Eng. Geol.*, Vol. 186, pp. 91-99, 2015.
10) Chen, P. A., Chang, M. J., Lee, F. Z., Lin, G. F. and Lai, J. S. : A sediment concentration forecasting model integrating multiple artificial neural networks and the switched prediction method. *3rd International Workshop on Sediment Bypass Tunnels, Taipei, Taiwan,* CD-ROM, 2019.
11) Geisser, S. : *Predictive Inference*, Chapman and Hall, 1993.
12) Brad, J. F. : An efficient point algorithm for a linear two-stage optimization problem, *Oper. Res.*, Vol. 31, pp. 670-684, 1983.

(Received May 31, 2019)
(Accepted August 7, 2019)