International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:5, No:4, 2011

Drainage Prediction for Dam using Fuzzy Support Vector Regression

S. Wiriyarattanakun, A. Ruengsiriwatanakun and S. Noimanee

***Abstract*—**The drainage Estimating is an important factor in dam management. In this paper, we use fuzzy support vector regression (FSVR) to predict the drainage of the Sirikrit Dam at Uttaradit province, Thailand. The results show that the FSVR is a suitable method in drainage estimating.

***Keywords*—**Drainage Estimation, Prediction.

I. INTRODUCTION

THE drainage estimating is the most important factor in planning, designing and managing including controlling reservoir and dam for Irrigation Engineers. In addition, this information may be used in flood or drought warning. Drainage estimating (m^3 / s) is a difficult task because drainage is varied based on consist of Inflow (m^3 / s) is nature water into Dam. Level (m) is level water in dam which reference form mean sea level. Storage (m^3) is all water amounts in dam. Water for electric power generate (m^3 / s) is water quantity pass to dynamo for electric power generate. And evaporate loss (m^3). This occurs to be uncertain. So it trouble to drainage estimating of dam.

There are several estimating systems for water management that use back propagation (BP) neural networks. For example, in [1], they predict Runoff coefficient at Eastern Botswana. In [2], they develop a water level model for flood warning system. In [3], they predict Runoff in water management. In addition, this several water management of dam management used neural network.

However, Back-Propagation (BP) neural network, suffers from difficulty in selecting a large number of controlling parameters which include relevant input variables, hidden layer size, learning rate, momentum term.

From the trouble of Back-Propagation (BP) neural network, Vapnik et al. [4] developed support vector machines (SVMs). The introduction of Vapnik's -insensitive loss function, SVMs have been extended to solve non-linear regression estimation problems. They have been shown to exhibit excellent performance in time series forecasting [5]. In [6], the comparison results between support vector machines regression (FSVMR) and BP neural network show that FSVMR performs better than BP. In several regression applications, the FSVMR is a popular tool. For example, in [7, 8], they found that FSVMR is the best tool in the finance forecasting and Mackey-glass data. In [9], the FSVMR is used to develop an ocean model estimate the tidal force and its direction effecting by the wind in the Gulf of Thailand. In Runoff prediction, there are a few researchers using FSVMR as a tool. For example, in [10], one-lead-day rainfall forecasting and Runoff forecasting are performed using FSVR, in which the input data are pre-processed by Singular Spectrum Analysis, resulting in a high-dimensional input space. The relationship model between rainfall and river discharge is made by FSVR. Bray and Han applied FSVR to forecast Runoff, focusing on the identification of an appropriate model structure and relevant parameters [11]. In [12], sequential elimination approach is used to identify the optimal training data set and then FSVMR is performed to forecast the water level.

This paper focuses on the application of FSVR in drainage estimating of Sirikrit dam at Uttaradit province, Thailand. This paper is divided into 5 sections. Section 2 provides a brief introduction to general principles of FSVR and its application in forecasting. The procedures of employing FSVR for drainage estimating are presented in section 3 in details by raising an experiment. Section 4 experimental results followed by the conclusions drawn from this study in the last section.

II. THEORIES

Here a brief description of FSVR is given. For a more detailed description the reader is referred to Vapnik [4,13], Scholkopf and Smola [14] and Cristianini and Shawe-Taylor [15].

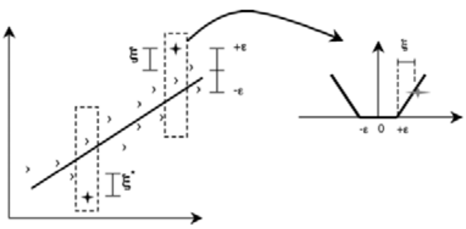

Fig. 1 In FSVR, a tube with radius ε is fitted to the data. The trade-off between model complexity (flatness) and points lying outside the tube (slack variables ξ) is determined by minimizing Eq. (4). The points outside the ε zone are support vectors (black stars). On the right, the ε-insensitive loss function is shown in which the slope is determined by C (the star represents a support vector).

S. Wiriyarattanakun is with the Computer science Department, Uttaradit Rajabhat University, Uttaradit, Thailand 53000 (corresponding author to provide phone: +66 55 411096 ext 1303; fax: +66 55 411096 ext 1312; e-mail: sopon_oil@ uru.ac.th).

A. Ruensiriwattanakun is with the Computer science Department, Uttaradit Rajabhat University, Uttaradit, Thailand 53000

S. Noimanee is with the Computer engineering Department, Chiang mai University, Chiang mai, Thailand 50000

Like most linear regression models e.g. PLS, the FSVR algorithm developed by Vapnik [4,13] relies on estimating a linear regression function:

$$f(\boldsymbol{x}) = w^T x + b \ (w,\ \mathrm{x} \in R^d \, (d - \text{dimension input space})) \quad (1)$$

Where *w* and *b* are the slope and offset of the regression line. In case of FSVR, the regression function is calculated by minimizing:

$$\frac{1}{2} w^T w + \frac{1}{n} \sum_{i=1}^{n} c(f(\boldsymbol{x}_i), y_i) \quad (2)$$

Where $(1/2)\|w\|^2$ is the term characterizing the model complexity (smoothness of $f(\boldsymbol{x}_i)$) and $c(f(\boldsymbol{x}_i), y_i)$ the loss function determining how the distance between $f(\boldsymbol{x}_i)$ and the target values y_i should be penalized. In this so-called primal formulation, several different loss functions [14] are available, but in this paper we adopted the commonly used *ε*-insensitive loss function which was introduced by Vapnik [4]. This *ε*-insensitive loss function is defined by:

$$c(f(\boldsymbol{x}_i), y_i) = \begin{cases} 0, & |y - f(\boldsymbol{x})| \leq \varepsilon \\ if\,|y - f(\boldsymbol{x})| - \varepsilon, & \text{otherwise} \end{cases} \quad (3)$$

In fact, this particular constraint defines a tube with radius *ε* around the hypothetical regression function (see Fig. 1) in such way that if a data point is positioned in this tube the loss function equals 0, while if a data point lies outside the tube, the loss is proportional to the magnitude of the Euclidean difference between the data point and the radius *ε* of the tube. In this particular case, the minimization of Eq. (2) is equivalent to solving the following constrained optimization problem:

$$\min_{w,\xi_i,\xi_i^*} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{\ell} s_i(\xi_i + \xi_i^*) \quad (4)$$

$$\textit{Subject to} \begin{cases} y_i - w^T x_i - b \leq \varepsilon + \xi_i \\ w^T(x_i) + b - y_i \leq +\xi_i^* \\ \varepsilon, \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

Where the constant $C > 0$ determines the trade-off between the model complexity of $f(\boldsymbol{x})$ and the amount up to which deviations larger than *ε* are tolerated. The slack variables ξ_i, ξ_i^* are introduced for the situation that the target value exceeds, this with respect to the origin of the original data space, more than *ε* above (ξ_i) and more than *ε* below the target (ξ_i^*), see Fig. 1. The points lying outside the *ε* tube are named support vectors (SVs), because these establish ('support') the fundaments of the estimated regression function. This implies that all other data points are in fact not important for inclusion into the model and can be removed after the FSVR model has been constructed. Hence, usually (much) less training objects do constitute the regression model; therefore, such a solution is referred to as 'sparse'.

The constrained optimization problem given by Eqs. (4) and (5) can be reformulated into dual problem formalism (Eq. (6)) by using Lagrange multipliers. In this paper we adopted the strategy outlined by Vapnik [4], which leads to the solution:

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) K(\boldsymbol{x}_i, \boldsymbol{x}) + b \quad (6)$$

where α_i and α_i^* (with $0 \leq \alpha_i, \alpha_i^* \geq C$) are the Lagrange multipliers and $K(\boldsymbol{x}_i, \boldsymbol{x})$ represent the so called kernel function [4]. Intuitively, the primal formulation is suitable to solve problems where many objects (samples) are available, this with respect to the number of variables at hand. The dual Lagrangian formalism, on the other hand, eliminates the curse of dimensionality, and hence, is even suitable to find solutions for ill-posed problems. In the context of Eq. (6), data points with nonzero α_i and α_i^* value are SVs. It has been shown that a suitable kernel function makes it possible to map a non-linear input space to a high-dimensional feature space where linear regression can be performed [4]. Several kernel functions have been proposed in literature, but the particular choice of a kernel to map the non-linear input space into a linear feature space depends highly on the nature of the data representing the problem at hand. In this paper the focus is put on two widely used kernel functions, namely, radial basis function (RBF) which are defined in Eqs. (7):

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right) \quad (7)$$

In case of the RBF kernel the parameter *σ* represents the kernel. The kernel parameter earlier mentioned parameters *C* and *ε* need to be selected properly by the user, because the generalization performance of the FSVR model heavily depends on the right setting of these three parameters. Hence, a reliable and robust parameter selection optimization strategy is a pre-requisite to obtain a well-performing and robust FSVR regression model.

III. Procedure of FSVMR for Forecasting

A. *Process*

The input data set is normalized so that they are in the same range, i.e., [-1,1]. We also map the desired output data set to [-1, 1]. However, after the algorithm computes the actual output, we have to map the actual output value back to its normal range.

B. *Data collection*

The Sirikrit Dam shown in figure 2 Located at Phasom Village, Tambon Tha Pla, 60 kilometers away from the province town, constructed to contain the Nan River in the area of Tambon Phaluad. It is an earth dam, the ridge of which is clay; its is 113.60 meters tall and 810 meters long; its ridge is 12 meters wide, while its base is about 630 meters wide. The area above the dam has a length along the Nan River of about 129 kilometers; the widest portion is about 20 kilometers; the water-surface area is about 178,000 rai; the maximum capacity is about 10,500 million cubic meters, capable of generating.

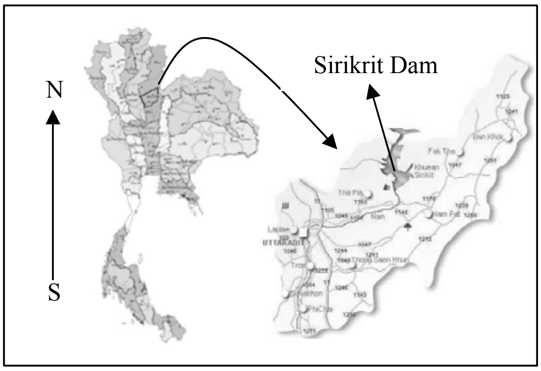

Fig. 2 The Sirikrit Dam is located in Uttaradit province at the north of Thailand.

This paper used the data set consist of Inflow, Level, Storage, water for electric power generate and evaporate loss. That total up is five patters. Which it is features. For used drainage estimating by fuzzy support vector regression. We select data set five patters form January 1996 till December 2006. We implemented 10% cross validation on the data set. The data set in January 2007 till December 2008 shown in figure 8 is utilized as a blind test data set.

C. *Performance criteria*

The prediction performance is evaluated using the mean absolute error (MAE) and mean square error (MSE) i.e.,

$$MAE = \left(\frac{\sum |D - y|}{n} \right) \qquad (8)$$

Where n represents the total number of data points in the data set. y represents the predicted value. D is denoted as the desired value which is the drainage of Sirikrit Dam.

IV. Experimental and the Results

D. *FSVR for drainage estimating*

Table I shows the setup of Runoff data set used for the 10% cross validation. In each cross validation set, there are approximately 4,018 samples and 402 samples in the training and testing data set, respectively. We used the real value of the drainage at Sirikrit Dam as a desired output. For the input of each sample point is comprised of inflow, level water, storage water for electric power generate, and evaporate loss. We used the input data from January 1996 till December 2006.

The model with C = 10,000, ε =0.0001 and σ = 1.2 (kernel parameter), gives minimum average MAE of 10% cross validation on the testing data set

We compare the results of the FSVR with that of the Back-propagation and polynomial regression that give the best average MAE on the testing data set in the 10% cross validation. The average MAE of training and testing data set from FSVR model Back-propagation model and Polynomial regression model are shown in Table II. The MAE of testing data set 7 in 10% cross validation is minimum.

TABLE I
DATA SETUP TO FSVR FOR DRAINAGE ESTIMATING

Input Feature					Desired Output
Level	Storage	Inflow	Evaporate loss	Water for electric generate	Drainage
1/1/1996	1/1/1996	1/1/1996	1/1/1996	1/1/1996	1/1/1996
2/1/1996	2/1/1996	2/1/1996	2/1/1996	2/1/1996	2/1/1996
3/1/1996	3/1/1996	3/1/1996	3/1/1996	3/1/1996	3/1/1996
⋮	⋮	⋮	⋮	⋮	⋮
31/12/2006	31/12/2006	31/12/2006	31/12/2006	31/12/2006	31/12/2006

The result form table 2. We select dataset 7 of FSVR while dataset 3 of Back-propagation. And dataset 2 of Polynomial regression. That is minimum MAE for test to blind test dataset.

The prediction result from FSVR model, Back-propagation model and Polynomial regression model on the blind test data set is shown Table 3 and figure 3 and 4.

TABLE II
MAE VALUES OF COMPARATIVE METHODS (M^3/S)

Data set	FSVR		SVR		Back-propagation		Polynomial	
	Training data	Testing data	Training data	Testing data	Training data	Testing data	Training data	Testing data
1	14.25	27.41	23.28	35.60	27.85	41.62	51.41	91.15
2	19.81	22.64	21.84	27.07	45.15	32.08	**46.70**	**75.80**
3	15.46	30.41	33.05	36.65	**21.71**	**27.04**	53.70	84.58
4	18.89	23.74	25.00	25.48	23.69	39.98	63.94	109.47
5	17.21	30.58	26.81	32.79	33.19	32.82	57.39	94.71
6	30.14	45.1	43.23	47.77	27.63	48.17	49.61	78.90
7	**12.98**	**20.98**	**16.16**	**21.95**	25.88	37.04	45.03	85.23
8	20.65	32.54	37.24	34.16	25.68	31.74	52.42	113.84
9	20.78	23.58	21.72	27.95	23.65	29.85	69.47	86.08
10	17.56	30.87	19.79	33.77	27.40	41.68	51.09	95.37
Average MAE	**18.773**	**28.785**	**26.81**	**32.32**	**28.18**	**36.20**	**54.08**	**91.51**

TABLE III
MAE VALUES OF COMPARATIVE METHODS (M^3/S)

dataset	Blind test			
	FSVR	SVR	Back-propagation	Polynomial
2007	**33.81**	35.24	39.65	92.84
2008	**35.92**	38.51	39.06	105.82
Average MAE	**34.865**	36.98	39.355	99.33

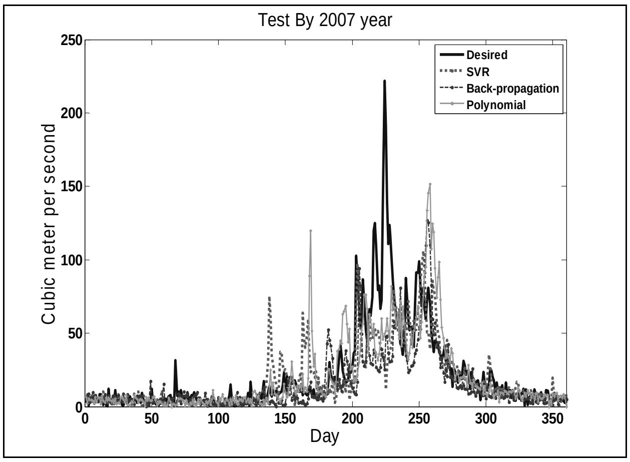

Fig. 3 drainage estimation result in the year 2007 (blind data set)

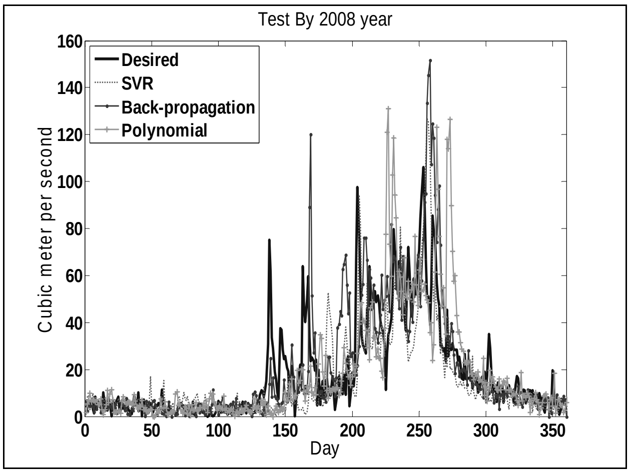

Fig. 4 drainage estimation result in the year 2008 (blind data set)

V. EXPERIMENTAL AND THE RESULTS

In this paper, we implement Fuzzy support vector regression (FSVR) to predict the drainage for Sirikrit Dam at Uttaradit province, Thailand. We found that the average MAE of the best FSVR model is 26.81 m^3/s and 32.32 m^3/s in the training and testing data set, respectively. While the average MAE of best Back-propagation model is 28.18 m^3/s and 36.20 in the training and testing data set, respectively, And the average MAE of best Polynomial regression model is 54.08 m^3/s and 91.51 in the training and testing data set, respectively. While

The MAE of the blind test data set in 2007year from the best FSVR, best Back-propagation model and best polynomial regression model are 33.81 m^3/s 39.65 m^3/s and 92.84 m^3/s, respectively. And 2008year are 35.92 m^3/s, 39.06 m^3/s and 105.82 m^3/s, respectively. This shows that the FSVR is more effective and efficient in drainage estimating than the Back-propagation and Polynomial regression.

ACKNOWLEDGMENT

The authors would like to thank Mr. Wisod patpin from Sirikrit Dam at Uttaradit province, Thailand., for the data set use in this work.

We also would like to thank Mr. Wenit Mrapik at Hydrology & water management center of Royal Irrigation Department in Phitsanulok, Thailand for the valuable information.

The authors are very grateful to the Uttaradit Rajabhat University for their financial support.

REFERENCES

[1] B.P. Parida a,*, D.B. Moalafhi b, P.K. Kenabatho "Forecasting Runoff coefficients using ANN for water resources management: The case of Notwane catchment in Eastern Botswana " Physics and Chemistry of the Earth 31 , 2006, pp.928–934.

[2] T. gtokelj, R Golob "Application of neural networks for hydro power plant water inflow forecasting " *2000 IEEE.* Neurel-2000*,* 5th Seminar on Neural Network Applications in Electrical Engineering.

[3] Y. B. Dibike and D. P. Solomatlne. River Flow Forecasting Using Artificial Neural Networks. Phys. Chem. Earth (B), Vol. 26, No. 1, 2001, pp. 1-7,

[4] Vapnik VN, GolowichSE, Smola AJ. Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems 1996, 9:281-7.

[5] S. Mukherjee, E. Osuna, F. Girosi, Nonlinear prediction of chaotic time series using support vector machines, in: NNSP'97: Neural Networks for Signal Processing VII: Proceedings of the IEEE Signal Processing Society Workshop, Amelia Island, FL, USA ,1997, pp.511–520.

[6] Francis E.H. Tay , Lijuan Cao "Application of support vector machines in financialtime series forecasting" Omega 29 , 2001, pp. 309–317.

[7] Yongsheng Ding , Xinping Song , Yueming Zen "Forecasting financial condition of Chinese listed companies based on support vector machine " Expert Systems with Applications , 2007, pp23-32,.

[8] U. Thissen, R. van Brakel, A.P. de Weijer, W.J. Melssen, L.M.C. Buydens "Using support vector machines for time series prediction " Chemometrics and Intelligent Laboratory Systems 69, 2003, pp.35– 49.

[9] Lt. Udomsak Boonprasert R.N. “Development of the Ocean Model for Search and Rescue Using Support Vector Machine” master's thesis, Dept. Electrical Engineering,Univ. Chiang mai,2003

[10] Sivapragasam, C., Liong, S.-Y., Pasha, M.F.K., Rainfall andRunoff forecasting with SSA–SVM approach. Journal of Hydroinformatics 3(3), 2001, pp.141–152,.

[11] Bray, M., Han, D.,. Identification of support vector machines for Runoff modeling. Journal of Hydroinformatics 6 (4), 2004, pp.265–280.

[12] Sivapragasam, C., Liong, S.-Y., Identifying optima training data set – a new approach. In: Liong, S.Y.,Phoon, K.K., Babovic, V. (Eds.), Proceedings of the Sixth International Conference on Hydroinformatics, Singapore, 2004.

[13] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, USA, 1998.

[14] B. Sch¨olkopf, A.J. Smola, Learning with Kernels, MIT Press, Cambridge,2002.

[15] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, UK, 2000.

S. Wiriyarattanakun was born at Kampangphet, Thailand on 30 June, 1983. The Degree of highest education was Master of Computer Engineering from Chiang Mai University, Chiang Mai, Thailand, and 2 year degree was earned.

He is lecturer of Department of Computer Science, Faculty of Science and Technology at Uttaradit Rajabhat University, Uttaradit, Thailand. The previous research experience were (1) Run-off forecasting using fuzzy fuzzy support vector regression, and (2) Image Light Intensity Normalization form scanner. For current research were approximate values of smoke from automobile exhaust.