Hydrol. Earth Syst. Sci. Discuss., 10, 13889–13895, 2013
www.hydrol-earth-syst-sci-discuss.net/10/13889/2013/
doi:10.5194/hessd-10-13889-2013

Hydrology and Earth System Sciences Discussions
Open Access

Comment on “A hybrid model of self organizing maps and least square support vector machine for river flow forecasting” by Ismail et al. (2012)

F. Fahimi and A. H. El-Shafie

Department of Civil and Structural Engineering, Universiti Kebangsaan Malaysia, UKM, Bangi, Selangor, Malaysia

Received: 9 June 2013 – Accepted: 25 October 2013 – Published: 15 November 2013

Correspondence to: F. Fahimi (fahimi@vlsi.eng.ukm.my)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Without a doubt, river flow forecasting is one of the most important issues in water engineering field. There are lots of forecasting techniques, which have successfully been utilized by previously conducted studies in water resource management and water engineering. The study of Ismail et al. (2012) which has been published in Journal of Hydrology and Earth System Sciences in 2012 was a valuable research that investigated the combination of two effective methods (self-organizing map and least squares support vector machine) for river flow forecasting. The goal was to make a comparison between the performances of SOM-LSSVM, autoregressive integrated moving average (ARIMA), artificial neural network (ANN) and least squares support vector machine (LSSVM) models for river flow prediction. This comment attempts to focus on some parts of the original paper that need more discussion. The emphasis here is to provide more information about the accuracy of the observed river flow data and the optimum map size for SOM mode as well.

1 Introduction

Recently, predicting river flow has become one of the indispensable parts of water resource management and water engineering. During recent decades, a considerable number of studies have been carried out for forecasting the river flow in different river basins and various methods including self-organizing map were applied to achieve more accurate and reliable results. In general, as an unsupervised learning method, the self organizing map is a kind of artificial neural network (ANN) model for clustering and classification of input data, prediction, and also data mining (Kohonen, 1998; Alhoniemi et al., 1999; Vesanto and Alhoniemi, 2000). In 2012, Ismail et al. carried out an inclusive study to improve the forecasting of river flow by using four different methods, i.e. SOM-LSSVM, ARIMA, ANN, and LSSVM models. However, the main contribution of this study was to improve the efficiency of the river flow prediction by employing self-organizing map (SOM) model for clustering input data and coupling this method with LSSVM model.

They examined 516 monthly recorded flow data of Bernam River in Malaysia from the beginning of 1966 to the end of 2008. The study area of the research was about 1090 km^2. They mentioned that Mann–Kendall test were used as the input data in order to find other river flow trends. For the analysis procedure, the data was divided into training (456 monthly recorded data, equivalent to 88 % of total data) and testing (60 monthly recorded data, equivalent to 12 % of total data) sets. It should be noted that the data was normalized in the range of [0.1, 0.9]. They also tested 8 different input data types. Mean absolute error (MAE), rout mean square error (RMSE) and correlation coefficient (R) were used to evaluate the performance of models.

During the review of this study, authors found that some parts of this research might be more useful if there would be more discussion on them. Consequently, following comments are proposed to be discussed more to enhance the contributions of the original study in presenting a new and accurate hybrid model for forecasting river flow.

2 Comments

To summarize, the following issues related to the authors' study should be properly addressed to clarify the points:

1. As mentioned before, in this study, authors tried to examine the performance of four different methods including the autoregressive integrated moving average (ARIMA), artificial neural network (ANN), least squares support vector machine (LSSVM) and SOM-LSSVM models for forecasting Bernam river flow. They indicated that 516 monthly Bernam river flow data were chosen as the real data for developing and testing the models. In order to make a better evaluation of each model performance, the predicted and observed river flow in testing period were illustrated in Fig. 8 of mentioned paper (Fig. 1). Normally, the predicted river flow

obtained from each of the four models has to be compared with a single unique observed dataset. However, as can be seen in this figure the observed river flow used for comparison with the SOM-LSSVM model is different from the other observed river flow data for the other three methods.

2. In the original research, authors utilized only four map sizes including 2×2, 3×3, 4×4 and 5×5 for SOM models then stopped increasing the size at 5×5 and showed the results of the proposed hybrid model for training and testing sets in Table 4 of the mentioned paper (Table 1). They, therefore, indicated that 5×5 is the optimal map size and would present the best results. Whereas, according to general trend of the results, during testing period, more accurate results could be expected for larger map sizes and it seems that even more optimum results might be obtained by using larger map sizes than 5×5. Moreover, based on the utilizing self organizing map model in Water Resources and Hydrology field, most of the previous studies proposed larger map sizes for similar researches. For example, Lin and Chen (2006) used the map size of 12×12 or in another study Abrahart and See (2000) used a map size of 8×8 to cluster the modelling domain into distinct types. As a different example, Chon et al. (1996) used a self organizing map with a 9×9 neurons map size.

3. In Sect. 3.2 (Artificial Neural Network), general concept of artificial neural network were presented. Three-layer MLP were expressed through Eq. (2) and also Fig. 1. But, it seems that some mentioned points should be corrected. In the proposed definition for Eq. (2), w_i is defined two times, while it seems that the second one which refers to the connection weights between hidden and output layer nodes must be w_j. Moreover, after that, the most common type of $f(\cdot)$ and $g(\cdot)$ are introduced in the text as the linear function and the Sigmoid function respectively, while according to Eq. (2) and Fig. 1 it seems that the Sigmoid function is the most common type of $f(\cdot)$ and the linear function is the most common type of $g(\cdot)$

HESSD
10, 13889–13895, 2013

3 Summary

Some parts of the original study are reviewed and the unclear points are discussed. In spite of the discussed issues and points, the mentioned study successfully proposed a new approach in river flow prediction by coupling different methods and also made an improvement in forecasting hydrological variables. Undoubtedly, any response from the authors would improve their work and facilitate a better understanding for the readers.

References

Abrahart, R. J. and See, L.: Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments, Hydrol. Process., 14, 2157–2172, 2000.

Alhoniemi, E., Hollmen, J., Simula, O., and Vesanto, J.: Process monitoring and modeling using the self-organizing map, Integr. Comput.-Aid. E., 6, 3–14, 1999.

Chon, T. S., Park, Y. S., Moon, K. H., and Cha, E. Y.: Patternizing communities by using an artificial neural network, Ecol. Model., 90, 69–78, 1996.

Ismail, S., Shabri, A., and Samsudin, R.: A hybrid model of self organizing maps and least square support vector machine for river flow forecasting, Hydrol. Earth Syst. Sci., 16, 4417–4433, doi:10.5194/hess-16-4417-2012, 2012.

Kohonen, T.: The self-organizing map, Neurocomputing, 21, 1–6, 1998.

Lin, G. and Chen, L.: Identification of homogenous regions for regional frequency analysis using the self-organizingmap, J. Hydrol., 324, 1–9, 2006.

Vesanto, J. and Alhoniemi, E.: Clustering of the self-organizing map, IEEE T. Neural Networ., 11, 586–600, 2000.

HESSD
10, 13889–13895, 2013

Table 1. The result for training and testing using a hybrid model of SOM-LSSVM for different map sizes (Ismail et al., 2012).

Map Sizes	Data	Training			Testing		
		MAE	RMSE	*R*	MAE	RMSE	*R*
2 × 2	M1	0.0680	0.0903	0.7964	0.0740	0.0872	0.7508
	M2	0.0655	0.0860	0.8205	0.0767	0.0963	0.6574
	M3	0.0758	0.1031	0.7259	0.0785	0.1020	0.6072
	M4	0.0686	0.0931	0.7925	0.0752	0.0975	0.6456
	M5	0.0770	0.1045	0.7250	0.0784	0.0988	0.6322
	M6	0.0869	0.1135	0.6495	0.0794	0.1022	0.5931
	M7	0.0758	0.1051	0.7126	0.0764	0.1011	0.6082
	M8	0.0212	0.0333	0.9782	0.0441	0.0620	0.8766
3 × 3	M1	0.0747	0.0997	0.7445	0.0640	0.0860	0.7376
	M2	0.0532	0.0681	0.8908	0.0683	0.0879	0.7254
	M3	0.0760	0.1029	0.7271	0.0736	0.0917	0.7099
	M4	0.0736	0.0995	0.7517	0.0733	0.0908	0.7019
	M5	0.0721	0.1008	0.7504	0.0734	0.0951	0.6650
	M6	0.0685	0.0914	0.8049	0.0794	0.1067	0.5421
	M7	0.0735	0.0974	0.7599	0.0703	0.0971	0.6474
	M8	0.0278	0.0378	0.9705	0.0431	0.0622	0.8734
4 × 4	M1	0.0537	0.0751	0.8640	0.0557	0.0691	0.8457
	M2	0.0561	0.0756	0.8642	0.0727	0.0884	0.7299
	M3	0.0649	0.0885	0.8105	0.0741	0.0937	0.6841
	M4	0.0696	0.0921	0.7947	0.0800	0.1010	0.6159
	M5	0.0647	0.0920	0.7990	0.0686	0.0916	0.7005
	M6	0.0645	0.0879	0.8253	0.0737	0.0992	0.6247
	M7	0.0701	0.0933	0.7830	0.0620	0.0894	0.7117
	M8	0.0348	0.0507	0.9485	0.0435	0.0647	0.8651
5 × 5	M1	0.0659	0.0950	0.7715	0.0637	0.0876	0.7274
	M2	0.0446	0.0612	0.9127	0.0668	0.0838	0.7560
	M3	0.0646	0.0920	0.7911	0.0690	0.0865	0.7361
	M4	0.0550	0.0762	0.8762	0.0680	0.0870	0.7306
	M5	0.0555	0.0782	0.8646	0.0727	0.0931	0.6855
	M6	0.0385	0.0605	0.9231	0.0661	0.0850	0.7462
	M7	0.0761	0.0982	0.7848	0.0731	0.0969	0.6490
	M8	0.0401	0.0567	0.9299	0.0370	0.0492	0.9222

HESSD
10, 13889–13895, 2013

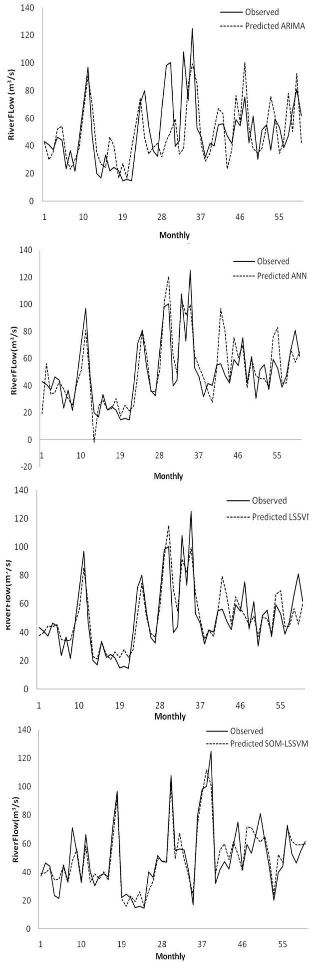

Fig. 1. Predicted and observed river flow during testing period by ARIMA, ANN, LSSVM and SOM-LSSVM for Bernam River (Ismail et al., 2012).

HESSD

10, 13889–13895, 2013

F. Fahimi and A. H. El-Shafie