Hydrol. Earth Syst. Sci. Discuss., 7, C4620–C4623, 2011
www.hydrol-earth-syst-sci-discuss.net/7/C4620/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "A hybrid least squares support vector machines and GMDH approach for river flow forecasting" *by* R. Samsudin et al.

R. Samsudin et al.

ruhaidah@utm.my

Received and published: 10 January 2011

Response to Reviewer We are grateful for the helpful comments and suggestions made by the reviews. The revised version has been adapted to their comments and suggestions.

Comment from Reviewer 2 The proposed hybrid technique does not take advantage of the two techniques. Furthermore, the fashion in which the hybrid has been constructed seem to be based on over-fitting. Figure 4 seem to indicate that GMDH firstly carries out polynomial fitting. After the fitting has been carried out, SVM is applied on GMDH fitted outputs to fit one again (second time!) to produce an output. This is an open invitation to overfitting and I do not see anywhere in the text how is this issues addressed,

if at all.

Response to Reviewer 2 Thank you for your comment about overfitting of the hybrid model. In our study, for avoid overfitting of the hybrid model , two sets of input data are used during the training process: (1) the primary training data, and (2) the control data which is used to stop the building process when overfitting occurs.

Here is an outline of the training method to avoid overfitting in propose model:

1. Construct each pair of input and output variables and divide the training and checking data sets. Preprocess the data to normalize the data sets. 2. Use GMDH to determine new input variables for every layer. To find new input variables for the next layer, the output variable which gives the smallest of root mean square error (RMSE) from GMDH model is combined with the input variables with = +1. The resulting variables are screened according RMSE criteria and only those having good predicting power are kept. 3. The new input of the neurons in the hidden layers are used as input for the LSSVM model.

4. Each new inputs is estimated by deciding which variable best evaluates the dependent variable . The criterion based on mean square error (MSE) used to evaluate the new variable for checking data (control the data) is defined as follow k = 1,2, . . . (the number of layer)

5. The value of MSE is kept, and jump back to step 2 to construct the next layer. At each layer, the value of MSE (k) is compared with its previous value and the process is continued as long as MSE(k) decreases or until a given model complexity is reached. An increasing MSE(k) is an indication of the model becoming overly complex, thus over-fitting the estimation data and performing poorly on the new selection data. Note, when over-fitting begins, the error as measured with the control data will being to increase, thus stopping the training.

For example, Fig. 1 shows the error (MSE) relative to the layer number for Selangor

HESSD

7, C4620–C4623, 2011

river (for input structure M4). MSE(k) which is picked as the best of the GLSSVM in each layer, as shown in Figure 1. The Figure showing that GLSSVM converges to a minimum value at k = 3. Therefore the output of the layer 3 gives the optimal performances. At the next k = 4 layer the value of MSE is increased, therefore in accordance with the performance criterion the GLSSVM has been over-fitted. Because a minimum of MSE was reached at the previous layer, we stop the training algorithm and conclude that a desirable GLSSVM is selected at the final approximate model.

The other comment from review 2 On a different note, the paper is entitled " A Hybrid Least Squares Support Vector Machines and GMDH Approach". However, large body of the paper covers ARIMA and ANN. The authors should at least consider changing the title to accurately reflect scope of the paper.

Response to Reviewer 2 In our study, we try to develop the new hybrid model. We feel the best title is for my study is “A Hybrid Least Squares Support Vector Machines and GMDH Approach for River Flow Forecasting “. The other models like ARIMA, ANN, GMDH and LSSVM models just for comparison to show that the new hybrid models more better than the other models.

We would like to thank the Reviews for all their useful suggestions and comments.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 3691, 2010.

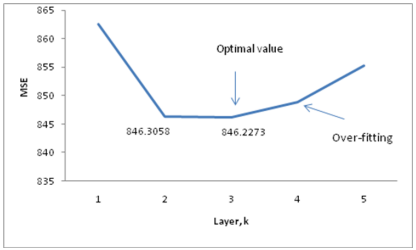

Figure 1: The value of criterion MSE calculated for the GLSSVM model

Fig. 1.

HESSD

7, C4620–C4623, 2011