Hydrol. Earth Syst. Sci. Discuss., 10, C5587–C5593, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C5587/2013/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Multi-step-ahead predictor design for effective long-term forecast of hydrological signals using a novel wavelet-NN hybrid model" *by* J.-S. Yang et al.

J.-S. Yang et al.

spyu@issas.ac.cn

Received and published: 12 October 2013

Dear referee,

Thank you for reviewing our manuscript. We appreciate all your invaluable suggestions and comments. We have modified the manuscript accordingly, and detailed corrections and explanations are listed below point by point. We believe that by presenting and publishing the interim results other scientist can apply the findings and discussions can be started to further develop the new methods.

Comment 1: Add more recent papers, especially Neuro-fuzzy and WNN applications

in rainfall-runoff and hydrosystems.

√ Following this suggestion, we have added more recent references about Neuro-Fuzzy and WNN applications in meteorological and hydrological signals forecast.

Comment 2: Methodology section would be improved by some further explanation of the wavelet technique as used in the study (e.g. the discrete form and which form of mother wavelet has been used – Daubechies or Morlet - and define it) and further explanation of the activation functions used in the ANN and membership functions used in neurofuzzy model. How these parameters are optimized. What are their threshold values. It is not very clear to readers who are unfamiliar with the techniques.

√ Following the referee's suggestion, we have given more explanation of the methods as used in the study. In the Methodologies section 2.1, the continuous wavelet form (Morlet wavelet) with its function and parameter definitions, the Morlet wavelet power spectrum testing process, the discrete wavelet form (Daubechies wavelet) with its definition were presented. The detailed decomposition process using the ten-level Daubechies wavelet (db10) was presented in Paragraph 1 Section 3.3.

In the Methodologies section 2.2, the membership functions (Gauss and Bell functions) and threshold rule numbers (2-5) in the ANFIS, the activation functions (Tangent Sigmoid and linear functions) in the 3-layer BP-ANN, and the principles of determining the optimal ANFIS and BP-ANN parameters (i.e. trying different transfer functions and rule numbers in ANFIS and trying different hidden layer numbers from 1 to 10 in BP-ANN) were presented.

Comment 3: Insufficient details on model selection. In an ideal world, each model would be able to fit any nonlinear function to an arbitrary accuracy. However, getting to the best model you can in the quickest amount of time with the least number of parameters seems to be very useful goal. Can the authors say anything about weather a certain type of model was faster to develop than others. In the manuscript, there is no information about calibration procedure adopted in modeling. How parameters are optimized, how different models are behaving in estimating runoff process. What are optimal architectures for different models they developed. More description of model selection and gereralisation is required. Which models are good in estimating peak flows. All these models are capable of fitting a complicated nonlinear function from asuitable model architecture. Instead of comparing model behavior based on statistics listed in Table-2, I would prefer to see more discussion on what additional insight or advantage be gained from wavelet may reveal the model dynamic or physical process of the system.

$\surd$ In the old manuscript, there were insufficient details on each type of NF model selection, that selected the commonly used Bell function and the default rule number value 5 in the layer one of ANFIS. Following the referee's suggestion in Comment 2 and 3, in the revised manuscript we determined the optimal four types of NF models (CDW-NF, CW-NF, DW-NF, NF) by trying different transfer functions and different rule numbers. Correspondingly, the Fig. 6, Table 2 and Fig. 7 were improved, and the advantage of CDW-NF model was more prominent.

As for the developing speed of different models, in the model training period the CDW-type models are the slowest to develop, because firstly the CWT method has to be used to reveal the optimal delay time nodes for determining the model input numbers, then the DWT method is used to optimize the model input by filtering some noisy details. Correspondingly, the pure models are the fastest to develop. However, in the model testing period, the established CDW- type and CW- type models (using 15-day and 28-day previous data to forecast St) are 15 times faster than the other models (using 1-day and 2-day previous data to forecast St) to forecast the future 1096 days river stages, because the CDW- type and CW- type models forecast 15 days new data in each forecast step, while the other models forecast one day new data in each step. What's more, the CDW-NF model's forecast performance in the testing period is very much better than the other models (Table 2 and Fig.7). Considering the huge difficulty to increase the long-range MSA prediction accuracy (see the very poor forecast

performances of most of models in Table 2 and Fig.7), it is worth to take a long time developing the CDW-NF model for effectively increasing the "true" long-term forecast accuracy. [Note: The long-range MSA forecast in our study is based on the "true" long-term forecast mechanism, explained in Referee #1's comment explanation, i.e. using the previous days' predicted data as model inputs in each forecast step in a "true" long-term forecast, while using the previous days' original/observed data as model inputs in each forecast step in a "seeming" long-term forecast]

In our study, the modeling and forecasting procedure containing DWT, NF(ANFIS), and BP-ANN are carried out in MATLAB. Thus, the detailed calibration procedure adopted in modeling and forecasting are achieved by using different MATLAB tool boxes. For DWT procedure, the 'wavedec' and 'wrcoef' functions are used. For NF(ANFIS) procedure, the 'genfis1', 'anfis' and 'evalfis' functions are used. For BP-ANN procedure, the 'newff', 'init', 'train' and 'sim' functions are used. Considering these tool boxes and functions have been maturely developed and commonly used by lots of researchers, we just have to set different function parameters according to our case study. Taking the CDW-NF and CDW-ANN hybrid models as examples, the detailed model training and testing procedure and corresponding parameter sets are presented below as MATLAB codes in the appendix. Following the referee's suggestion, in the revised manuscript we have given more detailed explanation of the key points in selecting optimal CDW-NF and CDW-ANN models, that trying two 'genfis1' parameters, i.e. different rule numbers and transfer functions, and one 'newff' parameter, i.e. different hidden layer nodes.

As the referee comments, all the 12 types of models we established are capable of fitting a complicated nonlinear function from a suitable model architecture in the model training period (see the overall high R2 of each model in the training period in Table2). However, because of the excessive error accumulation step by step in the testing period in a "true" long-time forecast, most of models behave very poor forecast performances in the testing period although they are very capable of fitting the complicated nonlinear hydro-system in the training period. Following the referee's suggestion, we take a deep analysis at the model performances on estimating peak values. We find that in the testing period after the 1096-day long-term forecast most of the models show very poor forecast performances with very low R2 (<0.05), except for the CDW-NF model with a relatively prominent high R2 (0.2∼0.3). Except for the CDW-NN (CDW-NF and CDW-ANN) models, the predicted 1096-day data curves from the other 10 types of models show almost straight lines when the forecasting step increases. Unfortunately, unlike the CDW-NF, the CDW-ANN predictions show poorer, even negative, R values. Because of the overall poor performance of the long-term forecast in R values, most of the models' predictions can not well estimate the peak values. Even in the CDW-NF predictions with the best R values, the modeling capability in estimating signal peak values still needs further research to increase a lot (see Fig. 6). Thus, we have pointed this situation in Section 4.2 in the revised manuscript. In addition, we have discussed the different forecast performance dynamics in the short-term, mid-term and long-term forecast procedures.

In view of the above analysis, obtaining an accurate "true" long-term forecast is proven to be almost an impossible work, but the novel CDW-NF hybrid model we established seems to take an important step forward for significantly increasing the "true" long-term forecast performances in R, RMSE and MAE, which is the most important purpose of this study. However, the other important steps forward for increasing the "true" long-term forecast performances both in the correlations (R, RMSE and MAE) and in the peak value estimation and physical process revealing (as the referee's suggestion) may need deeper and more advanced development of model algorithms, which will always be a challenging work in future researches. In addition, to accurately estimate the estuarine daily river stage peak value in our case study, we think it is not enough to use single time series models, and all the impact factors of tidal fluctuation, upstream flood dynamics, water impounding and releasing from the upstream hydroelectric projects, and precipitation dynamics along the river need to be consider in future modeling.

Appendix: % MATLAB codes for CDW-NF model training and testing

%CDW-NF model training period:

load('CDW-NFdata.txt'); x=CDW-NFdata (:,1:2); y=CDW-NFdata (:,3); epoch_n = 200; in_fis=genfis1([x y],5,'gbellmf')%'5'and'gbellmf'are rule number and transfer function; out_fis=anfis([x y],in_fis,epoch_n); A=evalfis(x,out_fis); [R,P]=corrcoef(A,y); E1=y-A; MSE=mse(E1); MAE=mae(E1)

%CDW-NF model testing period:

load('p11.txt') ; load('t1.txt'); for i=3653:15:4748; j=i-28; k=j+14; l=i-15; m=l+14; M=[p11(j:k,:) p11(l:m,:)]; B=evalfis(M,out_fis); t1=[t1;B]; [C1,L1]=wavedec(t1,10,'db10'); D3=wrcoef('d',C1,L1,'db10',3) ; D4=wrcoef('d',C1,L1,'db10',4); D8=wrcoef('d',C1,L1,'db10',8); p11=D3+D4+D8; end; b=t1(3653:4748,:); load('o.txt');[R,P]=corrcoef(b,o); E2=o-b; MSE=mse(E); MAE=mae(E)

% MATLAB codes for CDW-ANN model training and testing

%CDW-ANN model training period:

load('CDW-ANNdata.txt'); x=CDW-ANNdata (1:2, :); y=CDW-ANNdata (3, :); net_1=newff(minmax(x),[3,1],{'tansig','purelin'},'trainlm');%'3'is hidden layer nodes; net_1.trainParam.show=50; net_1.trainParam.lr=0.01; net_1.trainParam.mc=0.9; net_1.trainParam.epochs=5000; net_1.trainParam.goal=0.01; net_1.trainParam.mu_max=1e30; net_1.trainParam.min_grad=1e-020; net_1=init(net_1); [net_1,tr]=train(net_1,P,T) ; A = sim(net_1,P); E=y-A; [R,P]=corrcoef(A,y); MSE=mse(E); MAE=mae(E); inputWeights=net_1.LW{1,1}; inputbias=net_1.b{1}; layerWeights=net_1.LW{2,1}; layerbias=net_1.b{2}

%CDW-ANN model testing period:

load('p11.txt') ; load('t1.txt'); p11=p11'; for i=3653:15:4748; j=i-28; k=j+14; l=i-15; m=l+14; M=[p11(:,j:k); p11(:,l:m)]; B=sim(net_1,M); b1=B'; t1=[t1;b1]; [C1,L1]=wavedec(t1,10,'db10'); D3=wrcoef('d',C1,L1,'db10',3) ;

```
D4=wrcoef('d',C1,L1,'db10',4); D8=wrcoef('d',C1,L1,'db10',8) ; td=D3+D4+D8; p11=td';
end; b=t1(3653:4748,:); load('o.txt');[R,P]=corrcoef(b,o); E2=o-b; MSE=mse(E);
MAE=mae(E)
```

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 9239, 2013.