Hydrol. Earth Syst. Sci. Discuss., 6, C2536–C2537, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C2536/2009/

Hydrology and Earth System Sciences Discussions

Interactive comment on "An experiment on the evolution of an ensemble of neural networks for streamflow forecasting" *by* M.-A. Boucher et al.

Anonymous Referee #1

Received and published: 4 November 2009

This paper has employed an ensemble of fifty 3-layer neural networks for streamflow forecasting. Each of fifty neural networks has the same structure, i.e. four neurons in the input layer, five neurons in the hidden layer, and one neuron in the output layer. The bootstrap technique is used to obtain fifty different training datasets from original data sets. With each of fifty different training datasets, one set of the neural network weights is obtained and different from each other, thus leading to different streamflow forecasts to construct the forecast ensemble. Considering that the optimization of the neural network weights is an iterative process, thus this paper is aimed to assess how the performance of the forecast ensemble evolves with iteration ("epoch"). The criteria used to assess the performance of forecast ensemble are rather comprehensive, including CRPS (continuous ranked probability score), MAE (mean absolute error) of

the average value of forecast ensemble, the logarithmic score, the rank histogram, and the reliability diagram.

Some specific comments are made as follows. (1) About the physical validity of the fifty bootstrapped training dataset used for training neural networks. This paper has done lots of work to treat the original observation datasets. One is to separate the original datasets into two sub-sets, one is for training and the other for testing. The second treatment of data is use the bootstrap technique to obtain fifty training samples from the training dataset obtained by self organizing map. A question thus arises. As we know, the observed daily rainfall-runoff process is mainly a deterministic time series with its own intrinsic correlation structure. However, it seems that the samples generated through these twice treatments may no longer be the physically realistic/valid time series from the hydrological viewpoint. Should the physical validity be kept for the any daily rainfall-runoff input data in training neural networks? Probably more details about the properties of the generated data samples should be provided in the paper. (2) The paper indeed has adopted many criteria to assess how the performance of forecast ensemble evolves with the training process. However, the purposes and functions of these criteria are not clearly distinguished. What is the difference between the reliability component of CRPS and the reliability diagram? (3) On page 12, the paper stated that "for all basins and every epoch, the CRPS values are lower than the MAE values. It indicates that the ensemble of neural networks performs better when taken as a whole than when aggregated in a single averaged predictor." The statement is unfair for using the average value of forecast ensemble as the point forecast, since CRPS and MAE are just different indicators for assessing the same forecast ensemble; they are different just because their definitions are different. (4) On page 17, in the reference of "Maier, H.-R. and Dandy, G.-C.", "edelling issues" probably should be "modeling issues". (5) In Table 1, the data for the Sanjuan Basin should be checked, as the mean daily precipitation 3.47mm is much less than the mean daily streamflow 7.10mm. (6) What is the testing dataset used for?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 6265, 2009.