Hydrol. Earth Syst. Sci. Discuss., 6, C2745–C2748, 2009
www.hydrol-earth-syst-sci-discuss.net/6/C2745/2009/

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.

M.-A. Boucher et al.

marie-a.boucher.1@ulaval.ca

Received and published: 26 November 2009

We are very grateful to anonymous referee number one for his rapid answer and very useful comments. Our responses follow and the manuscript will later be improved accordingly.

1)About the physical validity of the bootstrapped series

All watersheds used in this study have a response time of about three days. Because multilayer perceptrons don't account for the temporal correlation between the inputs and the output, we had to recreate it. To do so, precipitation for the three previous days (according to the response time) was included in the inputs, as well as the streamflow

for the previous day. Such input selection allows the network to use prior information to forecast one-day-ahead streamflow, therefore accounting for the temporal correlation in daily precipitation and streamflow. This procedure was applied to the bootstrapped series. To do this, we first use the entire database, with all entries in chronological order, to produce a n by 5 matrix, n being the number of streamflow observations in the whole database. The fifth column is the observed streamflow for day t. The first three columns are the observed precipitation values for the three previous days (days t-1, t-2 and t-3, all in chronological order). The fourth column is the observed streamflow at time t-1. Then, we perform the Kohonen mapping, separation of the database and bootstrap using the row indices. This ensures that the streamflow observed at time t is accompanied by the precipitation and streamflow observations for the previous days in chronological order. Because the response time is very short (three days), there is no need to provide the network with observations further in the past.

Considering the above explanation, we ensure that the temporal correlation in the series was preserved and that the forecasts really are meaningful physically speaking.

2)About the performance criteria

The reliability component of the CRPS and the reliability diagram are closely related since they investigate the same characteristic of the forecasts. However, they can serve different purposes. For instance, the reliability diagram allows one to appreciate the reliability of various confidence intervals separately, while the reliability component of the CRPS is not linked to any particular level of confidence and provides a general numerical assessment for the reliability of the system. On the other hand, it is interesting to split the CRPS in two components not only for the reliability component but also for the resolution component, another characteristic of the forecasting system that is not clearly described by the other performance assessment tools used in this study, aside for the MAE for point forecasts. In fact, the rank histogram accounts for this characteristic to a certain extent, but as pointed out in Hamill and Colucci (1997), it can be misleading if used on its own. Finally, we believe in the pertinence of using many performance assessment tools that can confirm each other diagnostic.

3)About the comparison between the CRPS and the MAE

As it is mentioned in the manuscript, the CRPS reduces to the MAE for a point forecast. This result has been formally demonstrated by Gneiting and Raftery (2007), based on previous mathematical proofs by Barighaus and Franz (2004) and by Székely and Rizzo (2005).

Therefore, the fact that the MAE and the CRPS are completely equivalent for a point or a deterministic forecast is irrefutable. As explained by Gneiting and Raftery (2007), and mentioned in our manuscript, this property is very useful, since it allows us to compare the performance of a point forecast (the ensemble mean) with the performance of the probabilistic forecasts.

4)Spelling mistake

We thank the referee for pointing the mistake, which will of course be corrected.

5)Anomaly in Sanjuan's mean precipitation and streamflow

The reviewer correctly noted noted that the annual mean of the daily streamflows for Sanjuan watershed is higher than the annual mean of the daily precipitation. This oddity depends on the basin being located in the Canadian Coastal Mountains where only few rain gauges are available. However, the good performance of the neural networks confirms that the rainfall temporal sequence is correctly depicted by the available rain gauge networks.

We would also like to stress that a global underestimation of the rainfall intensities has no impact on the model performance since all data are standardized before being fed to the neural networks. This procedure ensures that all input data have the same range of values.

6)Utility of the testing data set

The purpose of the testing dataset is to validate the networks parameters (weights and biases). They have been optimized during the training using a portion of the data (the training dataset). Then, we want to test those parameters on a second set of data that was not used during the optimization.

References:

Hamill, T.M. and Colucci, S.J. 1997: Verification of Eta-RSM Short-Range Ensemble Forecasts, Monthly Weather Review, 125, 1312-1327.

Gneiting, T. and Raftery, A.E. 2007: Strictly Proper Scoring Rules, Prediction and Estimation, Journal of the American Statistical Association, 102 (477), 359-378.

Baringhaus, L. and Franz, C. 2004: On a new multivariate two-sample test, Journal of Multivariate analysis, 88, 190-206.

Székely, G.J. and Rizzo, M.L. 2005: A New Test for Multivariate Normality, Journal of Multivariate Analysis, 93, 58-80.

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