Hydrol. Earth Syst. Sci. Discuss., 6, C3108–C3111, 2010
www.hydrol-earth-syst-sci-discuss.net/6/C3108/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology – Part 2: Application" *by* A. Elshorbagy et al.

Anonymous Referee #1

Received and published: 13 January 2010

General comments

Model uncertainty and model structure optimization are among the most relevant topics in hydrology nowadays. It is felt that machine learning techniques could be a useful alternative in many cases where data are not sufficient to support a fully fledged process model, where a less time-consuming approach is needed, or where machine learning techniques could help to elucidate the inherent structure of large data sets and thus help to optimize model structure. Numerous techniques have been suggested and tested in various case studies. However, there is a lag of comprehensive stud-

ies summarizing results from various case studies and comparing more than two or three approaches. The two joint papers submitted by Elshorbagy et al. address that issue which is expected to contribute substantially to a better understanding and more efficient use of these recently developed techniques. The authors tested six different techniques (artificial neural networks, genetic programming, evolutionary polynomial regression, support vector machines, model trees and k-nearest neighbors), applied to five different hydrological data sets.

The authors are right that there is a need for comprehensive model comparisons instead of the often-published pairwise model comparison. In fact, the paper aims at becoming a benchmark paper with that regard (p. 7058, l. 17). That is a highly commendable, but also a highly ambitious goal. In fact, the two papers present a tremendous amount of work. However, still the study suffers from one of the generic problems of machine learning techniques: The single approaches can be parameterized in very different ways, having much of an influence on the model performance. For example, results from different artificial neural networks might differ substantially depending on the chosen initializations, the learning rate, the number of hidden nodes, the type of activity function, and the learning algorithm used. Although this issue is addressed occasionally in the paper (e.g., with respect to SVM), it is not considered in a systematic way. Consequently, readers would like not only to have a comparison between single realizations of different techniques, but to get some information if these differences are significant or not. It has been often argued that differences between different machine learning techniques might be small compared to the variability encountered within different realizations of models of the same type. Only when that is considered, the paper could really become a benchmark paper.

I do not feel happy with the study being split into two papers. On the one hand, the papers are very comprehensive and cannot be presented as two stand-alone papers (methods described in the 1st papers, results presented in the 2nd paper, references nearly exclusively given in the 1st paper). On the other hand, they are partly redundant

to each other (e.g., section 2 of 2nd paper and section 4 of the 1st paper). Thus, I would suggest the following:

1. Skip sections 3 and 5 in the 1st paper, and make it a pure review paper. It could be published nearly as it is (but see specific comments below) and would be a valuable source of information.

2. Restrict the model comparison to a single paper. That would require condensing the description of models, of the data set (refer more to relevant papers), and of the results. I recommend focusing more on generic features rather than to the performance of single models on single data sets (see comments below).

3. All the necessary details that cannot be presented in a HESS paper should be compiled in a technical report. See, e.g., the report by Maier and Dandy (1995) that complemented their paper in Water Resources Research 1996 on artificial neural networks.

Specific comments

1. The term “data driven modeling techniques” in the title is too generic. I suggest using the term “machine learning techniques” instead. At least, the latter term should be introduced in the text.

2. The paper comes along with only 12 references.

3. It is well known that different measures of performance in fact measure only single aspects of performance (e,g„ see Janssen & Heuberger (1995), Ecol. Mod. 83: 55-66). On the other hand, most of the different measures are not independent from each other. In the climate modeling community, the Taylor diagram has been developed to visualize comprehensive information about model performance (Taylor, K.E. (2001): Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106: 7183-7192). I suggest using that diagram.

4. What fraction of the total variance is explained by the respective models? That could

be given, e.g., by the Nash-Sutcliffe model efficiency.

Technical corrections

1. P. 7098, l. 22: Time needed for the GP depends on the used computer.

2. P. 7101, l. 5-6: The ANN does not force highly non-linear structures. To the contrary, it starts with a very smooth and nearly linear structure that becomes highly nonlinear only during later stages of the learning procedure.

3. P. 7103, l. 17 (and others): How do you measure the nonlinearity of the datasets? Could you give numbers for that?

4. P. 7110, l. 10: Why is the R statistic a key indicator in that study?

5. P. 7111, l. 26: Why do you scale to 0.5 standard deviation, not to 1.0 standard deviation, as is usually done?

6. P. 7130, table 15: What does the grey shading and underlining denote?

7. P. 7132, fig. 1: Please use barplots instead of lines.

8. P. 7122-7128, tab. 7-9, 11, 13: Confidence intervals would be more useful than the range. In addition, I suggest to summarize these results in a matrix of barplots with error bars.

9. In general: Fonts are much too small in most of the graphs.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 7095, 2009.