Hydrol. Earth Syst. Sci. Discuss., 3, S267–S269, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S267/2006/

Hydrology and Earth System Sciences Discussions

HESSD

3, S267–S269, 2006

Interactive Comment

Interactive comment on "Optimising training data for ANNs with Genetic Algorithms" *by* R. G. Kamp and H. H. G. Savenije

R. G. Kamp and H. H. G. Savenije

Received and published: 29 May 2006

First of all we would like to thank the referee for his valuable comments on Genetic Algorithms given to this paper. Below we provide answers to queries raised:

1. The hypothesis in this paper was that the initial training dataset should contain sufficient input/output data for the ANN because the total input space was covered (see point 8.). However it was not known which part of the dataset contributes most to adequate results, how many (subsets) are needed and in which sequence they must be placed. This is a complex optimization problem. GAs have global optimization capabilities and have advantages above other search techniques, including the ability to deal with qualitatively different types of domains, such as continuous variable domains and discrete variable domains.

2. The description of the role of GA in relation to ANN is, also according to earlier comments, not clearly described in this paper. In answers to previous questions we explained the role of GA in more detail and in the final paper we will pay more attention to this procedure. In essence the GA selects an optimal training dataset for the ANN. The new training dataset is constructed by copying subsets from the initial training dataset. The GA is used to optimize these subsets.

3. It is true that considering the context of this paper, it is better to include a short introduction to GAs, e.g.: Short introduction to GAs Genetic Algorithms are search algorithms that mimic the behavior of natural selection. GAs attempt to find the best solution to a problem by generating a collection (population) of potential solutions (individuals) to the problem. The best solution is the maximum of a function. Through mutation and recombination (crossover) operations, possibly better solutions are generated out of the current set of potential solutions. This process continues until an acceptably good solution is found. An optimal solution is however not guaranteed.

4. The flow model used in this paper should have been referred to as a hydraulic model.

5. In the final paper we will emphasize that GAs have global search capacity and an optimal solution is not guaranteed (see point 3, short introduction to GAs).

6. The objective function of the GA minimizes the RMSE of the simulated dataset of the ANN. The design parameters consist of begin and endpoints of the subsets which are copied from the initial training dataset. The parameter's domain therefore equals the time period of the initial training dataset. No inequality constraints and equality constraints, nor a penalty function was used. (Houck, C., Joines, J., Kay, M., 1995. A genetic algorithm for function optimization: a MATLAB implementation. NCSU-IE TR 95-09).

7. On page 289 is meant the upstream boundary of the flow model. In a 1D-flow model the upstream boundary is usually a discharge boundary and the downstream boundary a level- or a combined QH-boundary.

HESSD

3, S267–S269, 2006

Interactive Comment

8. The upstream discharge fluctuates between a minimal and maximal value to cover the input space. The discharge frequency (or tide frequency) slowly increases to encapsulate the whole input space of the input parameter to produce floods of different durations. It is a sin-function with a decreasing time period resulting in a quick oscillation at the end.

9. We have not been clear enough on the explanation of the test sets. Three test sets were used to test the performance of the ANN. The test sets are independent of the initial training set. With “differently shaped” was meant that the test sets were different (e.g. not subtracted) from the initial training set. This condition is important for a independent and objective testing procedure.

10. The size of the population of 10 is low. Calculations with higher numbers however did not give any better results, but did slow down the calculation.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 285, 2006.