**Hydrology and
Earth System
Sciences
Discussions**

HESSD

3, S184–S185, 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

First of all we would like to thank Raul Zurita for his constructive remarks that we can use to improve this paper.

In the first interactive comment there is a concern that the input data generated by the GA has no physical meaning. Apparently we did not fully explain what the role of the genetic algorithm was.

The initial data set was constructed such that all discharge levels and discharge variations occurred within the limits of the model's hydraulic constrictions. The design of the hydraulic model is based on (physical) dimensions from the 1D free surface flow drainage system in the Netherlands (Baambrugge). First of all the GA did not change or replace individual values of this training data to minimize the training error. This would indeed lead to unrealistic values.

The GA does select five random time periods from the training data and puts the cor-

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

EGU

responding input and output values in a subset. The new training set is based on simulation data from a computer model based on a real water system. Both the start and end date of each subset is randomly chosen, which results in a varying length of the subsets. In this paper the GA optimises the subsets by changing the start and end dates. These new five subsets together form a new training data set for another run with the ANN. Only small artificial effects occur at the boundaries of these subsets.

The second remark is about overfitting, which is a general problem of any ANN. Overfitting means that the training algorithm adjusts the weights of the ANN to fit every single data value in the data set and at the same time decreases it's capability to generalise new data sets. To prevent overfitting, cross-validation was used. The generalisation capacity was additionally tested with three different test sets which were constructed independently from the original training data set.

The last remark is about the configuration of the ANN. It is a general problem that ANN has many design parameters that must be chosen and optimised. It is true that a lot of energy has been put in finding the optimal configuration for this particular problem. In the designing process only a few rules of thumb were available and most of the optimisation was based on experience and trail-and-error. In the final publication we shall elaborate on this design problem, which is a general problem for any ANN.

Interactive comment on Hydrology and Earth System Sciences Discussions, 3, 285, 2006.

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

EGU