Hydrology and Earth System Sciences Discussions

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

Anonymous Referee #2

Received and published: 28 April 2006

The paper tackles an interesting topic, but lacks few details for it to be more useful to researchers and practitioners. The authors proved that the use of GA improves the training set selection and eventually the performance of ANN. However, they do not fully justify the use of GAs with respect to other optimisation techniques.

Detailed comments:

Page 287: The description of the GA is very short and does not really explain what the GA does to inputs into ANN. "In this paper the GA is used to feed an ANN model which mimics a hydrological model". The question here is: What is the GA feeding an ANN model with? "Each child consists of a set of five starting points corresponding with the five sub-sets of data of the training data." Does it mean that the whole data record is divided in 5 sub-sets and the GA selects the starting point of each of them? A pictorial representation of the GA chromosome would be useful here to explain this. In general I

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

EGU

would expect a short introduction to GAs here (explain the basic principles and provide references for more in-depth understanding of GAs) and then a detailed discussion of the main GA issues (i.e., coding, operators, fitness function, penalty function, termination conditions, etc). The authors later explain in detail how the GA is used in this particular study, therefore, only a general introduction to GAs is needed here.

Later on the authors apply their proposed technique to a hydraulic not hydrologic model. Please be consistent.

The statement "The GA solves this problem in a reasonable time without restricting itself to local minima." is not entirely correct. The GA does not guarantee that a solution will not be trapped in a local optimum. The only valid claim here would be that GA has the global search capability (but there are not guarantees).

Page 288: Based on the sentence "The RMSE is a common measure for the ANN's performance" I assume that the objective (GA fitness) function is to minimise RMSE? Any formal optimisation paper should clearly state what the objective(s) and constraints are. Please specify.

Page 289: "Ěupstream discharge boundary" should probably read "Ěupper discharge limit".

The sentence "The discharge frequency is slowly increasing from days to hours" is not clear. Why is it increasing? How? Please expand.

What does "Three differently shaped test sets were constructed" means. What tests sets are the authors referring to?

Thirty generations with the size of population as low as 10 is quite uncommon in GA practice. Why are these parameters selected as such? Would any other technique be suitable for this task. Have the authors considered any other optimisation technique and compared it to the GA results. Some justification for the use of GA is needed here.

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

EGU

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