

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

### **Anonymous Referee #1**

Received and published: 13 April 2006

Overall this is a really interesting idea and worthy of publication. I've put minor revisions because I think it needs more explanation in places.

Some specific comments:

1. I have lots of small editorial comments (typos, grammar, etc.). I will spare you these unless you specifically want them (I'm assuming the publishing process will pick these up?). I can, however, list them in a separate posting as they are not so interesting so let me know.
2. Page 288 - What does this sentence mean: “However, locations <do> not always match with the problem area”. I just want more information/an example.
3. What is the size of the five subsets? How do you know what size to select? What was the reason for the choice of five subsets? At the end of page 288, it sort of implies

Full Screen / Esc

Print Version

Interactive Discussion

Discussion Paper

that the five subsets equal the original training data but only the order is different. It's then clear on page 290 that this is not the case but maybe you could rewrite this so it's a bit clearer. Also is there the possibility for a given observation to be used more than once in your training data set?

4. How many previous values of discharge are used and how sensitive is the ANN to this choice?

5. What does the sentence on page 290 (section 4) mean: "Three differently shaped test sets were constructed." What does "differently shaped" mean? Just needs more explanation.

6. I can understand using this in situations where you have too much data and you want to pull out a new optimum subset but you imply on page 289 that this is a situation where you presume there is not enough data. Isn't this method decreasing the size of your training set and therefore giving you even less data? Perhaps I am just not understanding this completely.

7. Is it possible to provide a frequency analysis of the original dataset and the optimised one in order to understand how different the optimised dataset is? This would be clearer for me than figures 4 and 5, although perhaps these just need more explanation. I understood the bit about the sharp edges though so perhaps this is just something to think about.

---

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

[Full Screen / Esc](#)[Print Version](#)[Interactive Discussion](#)[Discussion Paper](#)