Hydrol. Earth Syst. Sci. Discuss., 2, S1241–S1244, 2005
www.copernicus.org/EGU/hess/hessd/2/S1241/
European Geosciences Union

Hydrology and Earth System Sciences Discussions

Interactive comment on "How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?" *by* Y. Tang et al.

Anonymous Referee #4

Received and published: 10 January 2006

General comments:

The paper provides a very comprehensive evaluation of 3 different multi-objective optimization algorithms for parameter estimation in hydrological modeling. The overall quality of the paper is good, technically sound, and, in general, well written. The paper provides a valuable contribution to the research on multi-objective calibration of hydrological models that has gained increasing interest in recent years.

Specific comments:

1. All three analysed algorithms are termed "evolutionary algorithms". I'm not aware of the precise definition of evolutionary algorithms (if any at all), but while the Epsilon-

NSGAII and SPEA2 algorithms both are based on traditional evolutionary operators, the MOSCEM algorithm uses very different operators.

2. In the description of the algorithms in Section 2 the algorithmic parameters are described. For the Epsilon-NSGAII algorithm one of the parameters is the maximum run time (or maximum number of model evaluations). However, this parameter is not included for the other 2 algorithms, although it is used as stopping criterion in the analysis.

3. In case study 3 a model with 36 parameters is calibrated. Since calibration is performed on objective functions based only on the total runoff, it is to be expected that a large number of these parameters would be quite insensitive to the objective functions and probably also exhibit significant correlations. How robust are the different algorithms to parameter insensitivity and correlations? And would this affect the conclusions of the performance of the algorithms for this case study? In practice, one would perform a preliminary sensitivity analysis to reduce the number of parameters for the calibration.

4. The description of performance metrics in Section 4.2 is not so clear. The 2 unary measures are based on, respectively, a distance measure and a volume measure. However, both measures are sensitive to the units and scales of the objective functions, and hence I would expect that some kind of normalization is necessary when evaluating the measures. A figure that shows how the measures are defined could be included.

5. The discussion of computational time required for the different algorithms and test cases in Section 5 is a bit unclear. All three methods uses the same number of model evaluations, so any differences in computational time is due to the differences in the time spent for algorithmic processing. It is to be expected that differences in algorithmic processing has a larger effect on the differences in total computational time for very cheap model evaluations (such as the test functions in case study 1), whereas the overhead from algorithmic processing is more or less negligible for expensive mod-

HESSD

2, S1241–S1244, 2005

els (such as the model used in case study 3). This can be seen from Tables 4 and 6, where computational differences are more pronounced for case study 1 than case study 2. Computational time is unfortunately not shown for case study 3 but differences would probably be less pronounced than for case study 2. Computational time is now discussed in Section 5.1 and 5.3. I suggest restructuring and reformulating this discussion according to the above.

6. Figure 7 and Figure 11 show results for the best run for each algorithm. However, it is not clear how “best” is defined in this case.

7. In the last paragraph of Section 5, it is stated that the results for Epsilon-NSGAII are conservative because a small initial population size is used as compared to the other algorithms. What are the arguments for choosing this initial population size? Elaboration of this aspect could be included in the paper.

8. Considering the very comprehensive analysis that has been conducted, I find the conclusions a bit vague. Besides effectiveness and efficiency, robustness is a very important property of a search algorithm when applied to hydrological model calibration. This aspect is nicely discussed in the paper, but not highlighted in the conclusions (or in the abstract). The fact that SPEA2 would require extensive trial-and-error analysis to determine appropriate algorithmic parameters is a severe limitation of its practical use. So rather than stressing that “overall, SPEA2 is an excellent benchmark algorithm” (p. 2496, l. 12) I would prefer a conclusion related to robustness and applicability in hydrological modeling practice.

Technical corrections:

1. p. 2470, line 23: I think that “as well as or better” should be “as well as or worse”.

2. p. 2474, line 6: Remove “also”.

3. p. 2474, line 10: Include “the” between “from” and “current”.

4. p. 2478, lines 11-16: In each sentence “this study” is used. Reformulate.

HESSD

2, S1241–S1244, 2005

5. p. 2483, line 21: Remove “the” before “Jasper”.
6. p. 2488, line 19: Include “from” between “different” and “those”.
7. p. 2493, line 26: Use “number of complexes” instead of “complex number”.
8. p. 2494, line 6: Use “number of complexes” instead of “complexes”.
9. Table 3: Use “12/50” instead of “12”.

Interactive comment on Hydrology and Earth System Sciences Discussions, 2, 2465, 2005.