Hydrol. Earth Syst. Sci. Discuss., 2, S1290–S1294, 2005
www.copernicus.org/EGU/hess/hessd/2/S1290/
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.

**Y. Tang et al.**

Received and published: 17 January 2006

Our discussion below first lists Referee #4's specific comments and then our response to each comment.

1. All three analyzed 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.

Authors' Response: The definition of evolutionary algorithms is quite broad as defined in the Handbook of Evolutionary Computation (Back et al. 2000). The Handbook defines evolutionary algorithms as population-based search algorithms that use solution variation and selection operators in optimization. MOSCEM employs population-based search, traditional selection operators, and its solution generation via complexes is analogous to a real valued recombination operator.

Back, T., D. Fogel, and Z. Michalewicz. 2000. Handbook of Evolutionary Computation, Bristol, UK.

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.

Authors' Response: The final manuscript will be edited to clarify our use of the run time parameter.

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.

Authors' Response: We fully agree that parameter correlations and sensitivities are an issue in hydrologic model calibration. These issues are related to the initial problem formulation for hydrologic model applications. With such a large number of parameters and a single output variable, it is to be expected that some of the parameters are insensitive to the objective function(s) selected. Different optimization runs will therefore result in different combinations of parameter sets that provide equal objective function value. It should be noted that the focus of this study is testing the relative performances of the three evolutionary multiobjective algorithms assuming an initial problem formulation. Each of the algorithms faced the same parameter correlations and sensitivities in their search for Pareto optimal points. The fact that different parameter combinations are obtained is not a problem in this context. It is, however, a considerable problem if the actual parameter values should not only be within a feasible range, but they should for example be used within a regionalization study. This is issue is discussed at length for example in Wagener and Wheater (Journal of Hydrology, in Press, available online), but beyond the scope of this paper.

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.

Authors' Response: A figure will be added to the final manuscript illustrating both unary measures.

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 models (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.



HESSD

2, S1290–S1294, 2005

Authors' Response: We fully agree that the differences in computational times presented in this paper highlighted the time spent for algorithmic processing. A key result in our paper (see lines 5-13, p. 2488) shows that MOSCEM required several days to solve the test function suite using its largest and most effective population size whereas the other algorithms required minutes to hours. This large difference in computational times highlights that MOSCEM's operators use a matrix inversion that yields very large algorithm run times for large population sizes and large numbers of complexes. MOSCEM does perform better with larger populations and increased numbers of complexes, but users should be aware that this will result in a substantial computational cost.

Timings were not shown for the Shale Hills case study because all of the algorithms were given 7-days to complete 5000 evaluations (see discussion in lines 10-18 on p. 2484). The algorithmic computational costs associated with SPEA2 and Epsilon-NSGAII were negligible relative to the model evaluation times for the Shale Hills case study. We note in our text that the algorithmic computational costs of using MOSCEM with a large population size and large numbers of complexes were not negligible even for the Shale Hills test case. We used the largest population size and number of complexes that would allow MOSCEM to complete its search within the 7 days allotted.

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.

Authors' Response: Our rationale for using initially small population sizes is discussed in detail in the manuscript on p. 2472, lines 1-20. The Epsilon-NSGAII uses a series of "connected runs" where small populations are exploited to pre-condition search with successively adapted population sizes. The algorithm automatically adapts its popula-



HESSD

2, S1290–S1294, 2005

Interactive Comment

tion size commensurate with problem difficulty and reduces the trial-and-error analysis often associated with determining the population size. See p. 2472, lines 1-20 for more detail.

The best runs were judged based on their final unary metrics values. We will clarify this in the final manuscript.

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.

Authors’ Response: We agree with the above comment and will add more details on SPEA2’s robustness and applicability in the final submitted manuscript’s abstract and conclusions. Additionally, Technical Corrections 1-9 will also be addressed in the final manuscript.

References Used in Response:

Wagener, T. and Wheater, H.S. In Press. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology, In Press, Corrected Proof, Available online 2 September 2005

---

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