

Hydrologic Model Parameter Optimisation

W. J. Cohen^a, Dr R. B. Ollington^b and Dr F.L.N. Ling^a

^a *Entura, Cambridge, Tasmania*

^b *University of Tasmania, Sandy Bay, Tasmania*

Email: Bill.Cohen@entura.com.au

Abstract: Deterministic rainfall-runoff models require parameter calibration with the aim of matching the modeled streamflow record to an observed record as closely as possible. Much attention has been focused on automated methods of model calibration using meta-heuristics, such as genetic algorithms. In this study, two meta-heuristic algorithms were compared: the shuffled complex evolution (SCE) algorithm and a real coded genetic algorithm (GA). The GA was modified with a fitness scaler based on a Boltzmann distribution, which was used to adjust the level of elitism of the selection operator. This was found to improve the performance of the GA. The algorithms were compared using a range of configurations on two test problems - the two parameter Rosenbrock function and the six parameter Hartman function. A relatively small number of objective function evaluations were allowed so as to test the ability of the functions to converge under adverse conditions (that is, limited computational runtime). These functions were found to be best solved with different strategies. The Hartman function was found to respond better to a more elitist strategy, and the Rosenbrock function was solved with a more egalitarian strategy with higher population diversity. On both functions, the best configuration of the SCE algorithm was found to perform favorably compared to the best configurations of the GA.

The best performing configurations of these algorithms were applied to the calibration of a rainfall runoff model of a catchment in Tasmania. The model was a two-tap Australian Water Balance Model (AWBM) consisting of eight parameters, with 10 years of data and a six month warm up period. 20 trials were performed on each algorithm, and in each trial the data were randomly split into train and test sets of nine and one year's data respectively. The more egalitarian strategies (which maintain higher population diversity, as taken from the algorithm configurations found to best suit the Rosenbrock function) were found to be favorable for both the SCE and the GA. When comparing results on the test data, this difference was marginal. The difference in performance was more apparent on the results of the SCE algorithm.

A recommendation of this paper is to develop a catalogue of the performance of several algorithms under different configurations on a range of test objective functions. This could be used to assist in the configuration of meta-heuristics on a range of applied problems, such as the calibration of additional rainfall runoff models.

Keywords: *Hydrology, modeling, calibration, optimisation*

1. INTRODUCTION

Calibration of a rainfall runoff model against an observed record is a computationally complex problem, given that there are often a large number of parameters used in the models and these parameters are typically in real space. A common approach to solving the calibration problem is to use meta-heuristics, such as genetic algorithms (Gill et al. 2006; Seibert 2000; Duan et al. 1993). These techniques have been found to be effective at calibrating models within a practical computational runtime.

Meta-heuristics are also known as global search algorithms due to their ability to locate the global optimum in the parameter search space. They are used to find the optimum (either the minimum or the maximum) of a given *objective function* in a relatively short time: in this context, the objective function is either the calibration metric of a rainfall runoff model (an applied problem), or a test function. Meta-heuristics differ from local search methods (such as a hill climbing algorithm) which typically use a 'greedy' approach to find the best solution. Though local solutions can be fast, they are prone to getting stuck in local optima without finding the best solution to the problem. The terms 'meta-heuristic' and 'algorithm' are used interchangeably in this paper.

Two common algorithms used in the calibration of rainfall runoff models are the shuffled complex evolution (SCE) (Duan et al. 1993) and genetic algorithms (GA) (Seibert 2000). Although they are good at searching the entire parameter space, GAs have been found to have problems with final convergence of the global optimum (Kuczera 1997; Q. Wang et al. 1995). A common technique used to address this is to augment the GA with local search (Seibert 2000), however this is not always known to work (Kuczera 1997; Q. Wang et al. 1995) as the objective function surface may be particularly noisy (ie have many local optima) in the region of the global optimum.

There have been various comparisons between SCE and different variations of GAs. Kuczera (1997) found SCE to be more robust than traditional GAs, though notes that other GA variants may perform better. Wang et al. (2010) found mixed results when comparing SCE with two variants of GAs when calibrating rainfall runoff models – results were dependent upon the objective function used.

Seibert (2000) gives Janikow and Michalewicz's (1991) findings that real coded genetic algorithms are faster, more accurate and give more consistent results than traditional binary encoded GAs, and presents such an algorithm. Seibert gives additional modifications: instead of traditional one or two point crossovers, each parameter is determined independently of the others; the cross-over operator takes (under certain probabilities) a parameter value from either parent (as with traditional GAs) or a value on the interval of the parent's values (randomly selected); and the algorithm has an additional single dimension local search cross-over/mutation (Brent's method). It is thought that Seibert's algorithm could perform well against the SCE, given the advantages of real coded genetic algorithms and its use of the local search mutation operator. The modified cross-over operator may also be able to hone in on the global optimum more effectively than a traditional GA. Fitness scaling has been found to improve the performance of GAs (Dandy et al. 1996; Maza & Tidor 1991), so Seibert's GA has been modified with a fitness scaler based on a Boltzmann distribution.

When optimising a problem such as the calibration of a rainfall runoff model, the configuration of a meta-heuristic is just as important as the choice of meta-heuristic itself. Different optimisation problems will be best solved with different algorithms under different configurations. It is important to achieve a balance between run time and convergence. If the meta-heuristic is configured to converge on a solution too quickly, it may be little better than a local search algorithm which fails to find the global optimum. If it converges too slowly, it is effectively inefficient, with wasted processing time. Finding this balance is not trivial: different problems will be able to converge at different rates.

Every objective function is likely to have a most suitable method for solving it, however it can be difficult to determine what this method is without considerable information about the objective function response surface (Duan et al. 1993). Test functions are often used in evaluating the performance of meta-heuristics (Bessaou & Siarry 2001; Duan et al. 1993), and give two advantages over applied problems: they are typically faster to execute and have known global optima. It is thought that if similarities between model calibration problems and a given test problem can be identified, then the test problem can be used as a surrogate in evaluating the performance of a range of meta-heuristics under several configurations to determine which is most suited to calibrating the model. This will then improve the selection process of a meta-heuristic/configuration for subsequent calibrations.

The aim of this work was to compare the SCE with Seibert's GA for calibrating rainfall runoff models. A secondary aim was to determine whether similarities in optimisation strategies for a model calibration and a test function could be detected: this could enable fast and optimal algorithm configuration for model

calibration. The SCE and the GA were compared under different configurations on two test problems. The best performing configurations of each algorithm on each problem were then applied to the calibration of a rainfall runoff model. The results of the final generation of algorithm evolution were then compared.

2. METHODS AND MATERIALS

The two algorithms have been applied under their various configurations on the test problems and the rainfall runoff model using a population size of 50 and have been evolved over 50 generations. This resulted in the SCE having approximately 20% more objective function evaluations than the GA. Results are compared by generation rather than the number of evaluations in order to have a consistent independent variable. Though there has been much attention focused on multi-parameter calibration when applied to calibration of hydrologic models in recent years (Gill et al. 2006; Khu & Madsen 2005; Seibert 2000; Tang et al. 2006), this paper focuses on single objective calibration to reduce complexity.

2.1. Shuffled Complex Evolution algorithm

The Shuffled Complex Evolution (SCE) algorithm was introduced by Duan et al. (1993) and has proven to be effective at calibrating hydrologic models. It is a form of differential evolution, which uses geometric operations on candidate solutions in parameter space to search for optimal solutions. The SCE algorithm has been tested under four different configurations, outlined in Table 1. The only difference between the implementation used in this study and the description of SCE given by Duan et al. (1993) is that in this study there is no early termination. This is so a direct comparison between performances can be made.

Table 1. Configurations used for the SCE algorithm

Configuration	Population size	No. complexes	Complex size	Simplex size
25 Cmplx	50	25	2	2
10 Cmplx	50	10	5	5
5 Cmplx (S:10)	50	5	10	10
5 Cmplx (S:5)	50	5	10	5

2.2. Real coded genetic algorithm

Seibert's genetic algorithm has had more considerable modifications made to it for this study: it has been adapted to use single objective optimisation for these experiments. The selection component of the algorithm was modified so that its level of elitism can be selected through a parameter. This was done by fitting a Boltzmann distribution to the fitness scores of the candidate solutions, as discussed by Maza and Tidor (1991). This consists of an additional variable to the algorithm, 'temperature', which is used to configure how elitist the selection process is. High temperature values indicate an 'egalitarian' approach, which evens the chances of individuals being selected; low values indicate an 'elitist' strategy, which highly favours individuals with good fitness scores. Though Maza and Tidor recommend a changing temperature variable as the algorithm evolves, fixed values have been used in this study. This is the only parameter used to configure the GA. Temperature values are on a log scale (which varies depending upon the function to be solved) and have been compared to the algorithm without fitness scaling. Seibert's GA also has a post-evolution local search; this has been removed to enable a more direct comparison with the SCE.

2.3. Test functions

The test functions used are the Rosenbrock two-parameter problem and the Hartman six parameter problem. Both functions are given by Duan et al. (1993). The same configurations of SCE were used on each problem (see Table 1), however it has been necessary to use different temperature ranges for the Boltzmann fitness scaler on the GA; the Hartman function had ranges between 0.2 and 100.0 on a log scale (ie 0.2, 0.5, 1.0, 2.0...20.0, 50.0, 100.0) and the Rosenbrock function had values between 20,000 and 50,000,000. Both functions have also been tested without Boltzmann scaling: as the test functions are minimisation functions and the GA uses roulette selection, the inverse of the objective function value has been used in selection here.

2.4. Rainfall runoff model

The four best algorithms/configurations (ie best GA configuration on the Hartman function, the best GA on the Rosenbrock function, the best SCE on the Hartman function, and the best SCE on the Rosenbrock function) were applied to the calibration of a rainfall runoff model of the Florentine River in Tasmania. This

is a two-tap AWBM (Parkyn & Wilson 1997) that consists of eight parameters to solve for. The objective function for the calibration is the Nash Sutcliffe efficiency score (Nash & Sutcliffe 1970) which represents the ability of the modeled flow to describe the observed flow at a given point: one indicates a perfect match and negative values are possible. This function is a maximising function, as opposed to the test problems which are minimising.

Model runtime is considerably longer than the runtime of the test functions, so only 20 trials of each of these tests were performed. Each objective function evaluation consisted of a model run over a 10 year period with a six month warm up and was on a daily time-step. Data were split into two categories, train (nine years) and test (one year); the test year was selected randomly for each trial. The best individual parameter set (as measured on the training period) was evaluated on the test data range at the end of each generation.

3. RESULTS

The best configurations of the meta-heuristics on the test functions are shown in Table 2. These configurations were then applied to the calibration of the rainfall runoff model.

3.1. Results of the test functions

Table 2. Best meta-heuristic configurations on test functions

Function	Best configuration	
	GA (temperature)	SCE (no. complexes)
Hartman	0.5	10
Rosenbrock	50,000,000	25

The by-generation performance of these configurations is plotted in Figure 1. The GA trials that used the Boltzmann fitness scaler were found to converge better than the GA which used inverse score weighting.

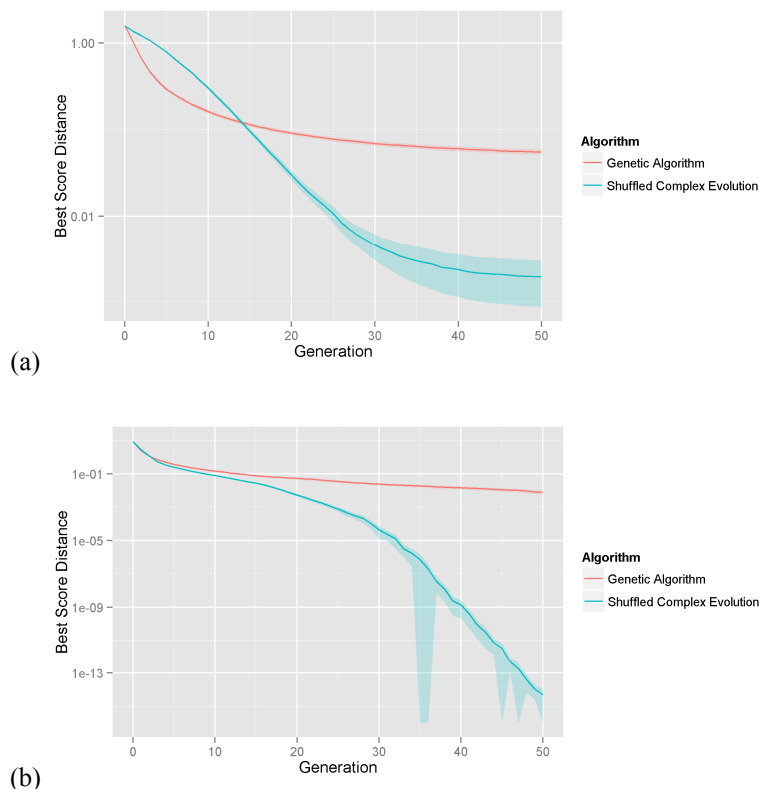


Figure 1. Performance of the best algorithm configurations on the (a) Hartman and (b) Rosenbrock functions, 1000 trials. Population size: 50, Fitness Score is ‘distance’ from the optimal value of the objective function. Error bars indicate 95% confidence interval. ‘Best score distance’ is the distance to global objective function value. The distortion of the lower confidence interval in (b) is due to the very low values (which have become negative in some generations) and that the Y axis is on a log scale.

It can be seen in Figure 1a that the best configuration of the SCE outperforms the best configuration of the GA on the Hartman function; the same is true for the Rosenbrock function shown in Figure 1b. It appears that the SCE algorithm approaches the global optimum within a very fine tolerance.

It is apparent from Table 2 that the Hartman function is solved with a more elitist strategy (in both algorithms) than the Rosenbrock function. This runs counter to Duan et al.'s (1993) description of the functions where the Rosenbrock function has a single optimum, while the Hartman function has four: it would seem that more optima would require a more diverse approach. When visualizing the surface of the Rosenbrock function, however, it is apparent that the surface consists of a large valley: locating the single optimum in this valley may be a difficult procedure and require a strategy that maintains a diverse population.

3.2. Results of the rainfall runoff model

The performance of the SCE algorithm with 25 complexes is plotted in Figure 2. It can be seen that the Nash Sutcliffe score in the initial generation (which represents the best score of 50 randomly selected parameter sets) is already above 0.9, indicating that the model is quite easy to calibrate. The goal of optimisation in this case then is to improve this score as much as possible.

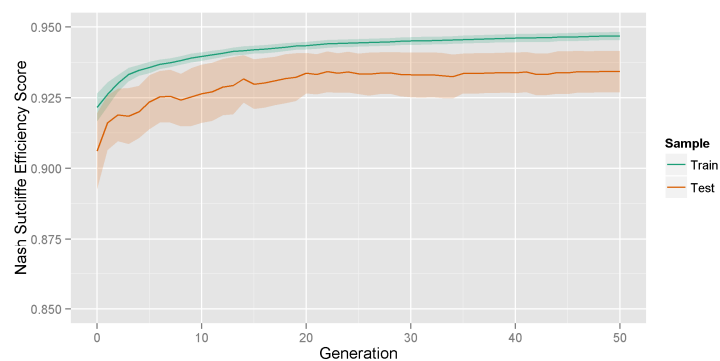


Figure 2. Performance of the SCE algorithm with 25 complexes on the rainfall runoff model; 20 trials; population size: 50. Error bars indicate 95% confidence intervals.

The results of each meta-heuristic/configuration in the final generation on both the test data and training data are shown in Figure 3. It can be seen that all configurations appear to perform similarly. On the training data, the SCE performs better with 25 complexes; this difference appears to be more significant when comparing the test data, however the 95% confidence intervals are wide and more trials would be necessary to determine if this is a trend.

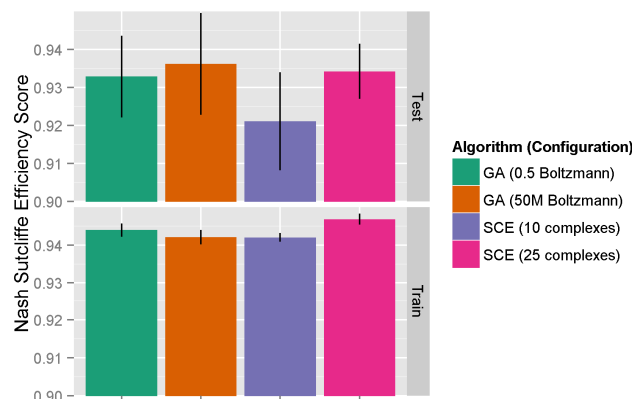


Figure 3. Final generation performance of the algorithms on the rainfall runoff model; 20 trials; population size: 50. Error bars indicate 95% confidence intervals.

4. DISCUSSION

The difference in the relative performance of the algorithms and their configurations on the test functions appears to be significant, however this performance difference may be largely academic unless a convergence criteria is used, such as a tolerance to the objective function value or distance to the global optimum in parameter space. In Figure 1a it can be seen that the GA initially performs well, however by generation 15 the SCE algorithm shows better convergence. This supports Kuczera's (1997) finding that a GA flounders near the global optimum, even though a different variant of GA was used in that study. Kuczera attributes the superior convergence of the SCE to its global shuffling, however it is also likely that its reflection and contraction operators allow it to scale its search when honing in on areas of interest. It was initially thought that Seibert's GA could match this scaling by use of its cross-over operator and its local search mutation. To illustrate, if two parents straddle the global optimum in one or more parameters, a traditional GA will not be able to locate the optimum from the information contained in the parents; SCE and Seibert's GA would be able to. Seibert's GA is limited to searching within the bounds of both parents (with the exception of mutation), while SCE is able to search outside of this area with its reflection operator. It may be possible to improve the performance of this GA through a targeted selection of the probabilities of each cross-over operator.

The performance of the algorithms on applied problems such as the rainfall runoff model is more difficult to evaluate. Here, a professional hydrologist can judge whether one algorithm has significantly better convergence than another. The results plotted in Figure 3 indicate that though the SCE with 25 complexes performs better on the training data than the other algorithms, this difference is quite small and that the algorithm selection may not be important. The results on the test data are more interesting: the mean score of the final generation of the SCE with 10 complexes is poorer, however as only 20 trials have been used the confidence intervals are wide. Additionally, the test scores are evaluated on only a single year's data, and the test data are randomly selected for each trial. It may be that some years of data are more difficult to calibrate: this would impact the test scores more than the training scores. As such, the performance on the test data may reflect more on which years of data were selected for test rather than the algorithm performance. Many more trials of each algorithm will be necessary to determine if this is a trend.

The good performance of the model calibration before optimisation began is also interesting. That the model obtains a Nash Sutcliffe score of greater than 0.9 from the best of a randomly selected population of 50 indicates that the model is quite easy to calibrate. More complex models might produce a different result.

5. CONCLUSIONS AND RECOMMENDATIONS

Significant differences between the algorithms and their configurations are evident when applied to the test problems. There is no measure or meaning associated with the raw scores however, and it is difficult to determine how important these differences are. The performance differences are less noticeable when the algorithms are used to calibrate the rainfall runoff model however.

The best configuration of an algorithm is dependent on the objective function it is attempting to solve, and these differences can be significant. The Rosenbrock function has been found to be solved with a more egalitarian strategy with both SCE and GA, and the Hartman function requires a more elitist approach. The SCE algorithm outperforms the GA on both test functions under the experiment conditions. Similarities between the strategies of the test functions and the model calibration, however, were not detected; more results would be required.

The addition of Boltzmann fitness scaling has been found to improve the performance of Seibert's algorithm on the problems tested. The cost of this fitness scaling is that it is not known what the best temperature is for a given objective function; this needs to be determined prior to performing a calibration of a model. To complicate this further, the original Boltzmann fitness scaler as proposed by Maza and Tidor (1991) uses a variable temperature; it starts with an egalitarian strategy and becomes more elitist as evolution proceeds. This strategy may improve the GA performance further; however this requires three algorithm parameters: initial temperature, final temperature and step temperature. SCE will not benefit from fitness scaling, as it uses rankings of fitness scores for its selection rather than weightings of the fitness scores directly.

5.1. Further work

There are other factors aside from the ones presented in this paper that could affect algorithm performance. All arrangements presented here consisted of a population of size 50 evolved over 50 generations; however an algorithm, or a problem, may be more sensitive to this arrangement than its parameter configuration.

It would be of interest to the authors to see a catalogue built up of several objective functions (such as Rosenbrock and Hartman) and the best configurations chosen to solve them. This could potentially be used to find patterns in these objective functions and others (such as model calibration) to quickly determine the best strategy to solving them. This catalogue should be coupled with a range of model calibrations in order to maximise its benefit.

ACKNOWLEDGEMENTS

The authors acknowledge Greg Carson and Angus Swindon of Hydro Tasmania for permission to use data and models of the Florentine River catchment.

REFERENCES

- Bessaou, M. & Siarry, P., 2001. A genetic algorithm with real-value coding to optimize multimodal continuous functions. *Structural and Multidisciplinary Optimization*, 23(1), pp.63–74.
- Dandy, G.C., Simpson, A.R. & Murphy, L.J., 1996. An improved genetic algorithm for pipe network optimization. *Water resources research*, 32(2), pp.449–458.
- Duan, Q.Y., Gupta, V.K. & Sorooshian, S., 1993. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and Applications*, 76(3), pp.501–521.
- Gill, M.K. et al., 2006. Multiobjective particle swarm optimization for parameter estimation in hydrology. *Water Resources Research*, 42(7), pp.1–14.
- Janikow, C.Z. & Michalewicz, Z., 1991. An experimental comparison of binary and floating point representations in genetic algorithms. In *Proceedings of the fourth international conference on genetic algorithms*. California, USA, pp. 31–36.
- Khu, S.T. & Madsen, H., 2005. Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research*, 41(3), pp.1–14.
- Kuczera, G., 1997. Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research*, 33(1), pp.177–185.
- Maza, M. de la & Tidor, B., 1991. *Boltzmann Weighted Selection Improves Performance of Genetic Algorithms*, Cambridge, MA.
- Nash, J.E. & Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), pp.282–290.
- Parkyn, R. & Wilson, D., 1997. Real-Time Modelling of the Tributary Inflows to ECNZ's Waikato Storages. In *24th Hydrology & Water Resources Symposium Proceedings IEAust*. Auckland, NZ, pp. 445–449.
- Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Science*, 4(2), pp.215–224.
- Tang, Y., Reed, P. & Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrology and Earth System Sciences*, 10(2), pp.289–307.
- Wang, Q., Chiew, F.H. & McMahon, T.A., 1995. Calibration of environmental models by genetic algorithms. In *Proc. of MODSIM 95, International Congress on Modelling and Simulation*.
- Wang, Y.-C., Yu, P.-S. & Yang, T.-C., 2010. Comparison of genetic algorithms and shuffled complex evolution approach for calibrating distributed rainfall-runoff model. *Hydrological Processes*, 24(8), pp.1015–1026.