Multistart Newton-type optimisation methods for the calibration of conceptual hydrological models

Dmitri Kavetski, George Kuczera, Mark Thyer and Benjamin Renard

School of Engineering, University of Newcastle, New South Wales, Australia
Email: dmitri.kavetski@newcastle.edu.au

Keywords: *Calibration, multimodal objective functions, model smoothing, multistart Newton-type methods*

EXTENDED ABSTRACT

Conceptual hydrological models can parsimoniously capture dominant catchment dynamics. However, efficient calibration of their parameters is impeded by micro-scale roughness and complex macro-scale geometry (including multioptimality) of objective functions. These difficulties led to an abandonment of Newton-type optimisation methods (which rely on smoothness and converge to local optima) and motivated a shift towards global evolutionary searches.

However, recent work indicates that in many cases the optimisation problems listed above are numerical artefacts of the model implementation and can be removed using smoothing and more numerically stable model implementation. Importantly, the removal of micro-scale roughness of the objective function permits the application of Newton-type methods, which are the most efficient class of optimisation methods for smooth problems, especially as the number of dimensions (parameters) increases. In addition, the potential multioptimality of parameter distributions can be analysed using multistart optimisation strategies.

This paper makes several contributions to systematic analysis of parameter distributions using multistart Newton-type methods. It summarises and illustrates the problematic features of objective functions in hydrology (Figure 1), and how they can be alleviated using model re-formulation. The ability to achieve smooth objective functions enables the application of multistart Newton-type methods, and we supply case studies where these methods significantly outperform global SCE searches in terms of efficiency (Figure 2), as well as in diagnostic capabilities in identifying statistically significant multiple parameter optima. It is also argued that uniform seeding of the search region prior to the multistart analysis yields useful insight into the macro-structure of the distribution, including the relative sizes of attraction basins associated with multiple optima. Finally, it is discussed that Newton-type methods are the only computationally feasible way to carry out the optimisation analysis of high-dimensional objective functions such as those arising in Bayesian hierarchical models of data and model uncertainty, and that using analytical model derivatives significantly reduces the computational cost of the optimisation.

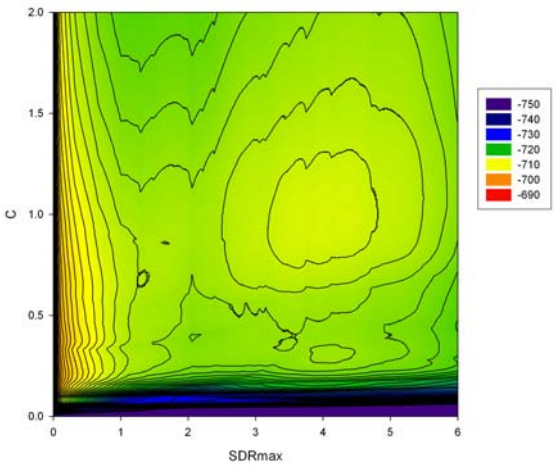

Figure 1. Multiple optima of the SFB model. The global mode lies on the bound SDRmax = 0.

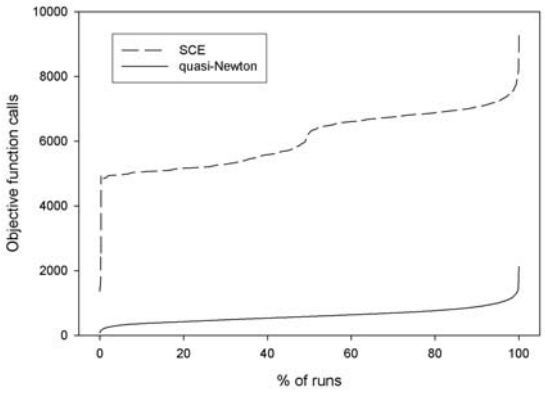

Figure 2. Computational cost of the multistart quasi-Newton and SCE methods.

1. INTRODUCTION

Conceptual models are important hydrological tools that can capture dominant catchment dynamics while remaining parsimonious and computationally efficient. However, their parameters $\boldsymbol{\theta}$ must be calibrated from observed data by minimising discrepancies between the model and the observed data, as measured by an objective function. The majority of objective functions in hydrology are related to the classical sum-of-squares (SS) function Φ.

$$\Phi(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(y_n - h_n(\boldsymbol{\theta}) \right)^2 \qquad (1)$$

where h_n is the prediction of the model at step n and y_n is the corresponding observed response.

However, model calibration in environmental modelling has been plagued by several numerical problems, including (PI) micro-scale roughness and (PII) complex macro-scale geometry of objective functions. The latter category includes strongly correlated parameters, especially if the correlation is nonlinear (yielding banana-shaped distributions) and multioptimality (multimodality) of the objective function. These difficulties proved intractable for earlier generations of optimisation methods (see Gupta and Sorooshian, 1985), resulting in a shift away from classical Newton-type optimisation towards global evolutionary searches. For example, the shuffled complex evolution (SCE) algorithm of Duan et al. (1992) is a widely used evolutionary search method.

However, since nonsmooth optimisation methods ignore gradient information about the objective function, they usually require many more function calls, with the loss of efficiency growing rapidly as the number of parameters increases. For simple lumped models, the loss of efficiency is often tolerable, yet for 2D/3D distributed models, the cost of model runs and the number of parameters grows significantly. Indeed, Tolson and Shoemaker (2007) show that evolutionary searches become computationally impractical with growing model complexity, while Skahill and Doherty (2006) show that with suitable enhancements, the Newton-type Gauss-Newton-Marquardt method can be far more efficient that the SCE method.

In addition, consider the computational requirements of calibration methods that directly account for input (rainfall) and model uncertainty in hydrological simulations. When this is accomplished using Bayesian Total Error Analysis (BATEA) (Kavetski *et al.*, 2006a; Kuczera *et al.*, 2006), the dimension of the objective function is proportional to the length of the calibration data. Consequently, long calibration periods with many storm events may require the optimisation of hundreds of latent variables. This task is intractable using nonsmooth methods, especially if multiple restarts are implemented.

Finally, a more subtle flaw of current global search methods is their limited diagnostic capability: knowledge of the global optimum is insufficient if there are several significant local optima.

There remains, therefore, a genuine need to develop more efficient and informative optimisation methods in hydrological model calibration. Tolson and Shoemaker (2007) pursue this objective using global evolutionary searches with dynamic dimension reduction. There are also promising advances in combining evolutionary optimisation and sampling, e.g., the hybrid SCE-Metropolis algorithm (Vrugt *et al.*, 2003).

Our approach in this paper is more direct. It is motivated by the fact that Newton-type methods represent the most efficient class of optimisation methods for smooth problems and, furthermore, are currently the only feasible methods for high-dimensional problems (Nocedal and Wright, 1999). Consequently, our preference is to revisit the reasons why Newton-type methods have lost favour in hydrologic modelling (problems PI and PII listed earlier) and to overcome these difficulties. Indeed, recent studies indicate that Newton-type optimisation can be used for hydrological models, provided (i) the model is sufficiently smooth with respect to its parameters, (ii) multiple initial points are used to explore the objective function and identify potential multiple optima, and (iii) robust implementations of Newton-type methods are used (Kavetski *et al.*, 2006c; Skahill and Doherty, 2006).

The objectives of this paper are as follows:

(1) Summarise and illustrate the problematic features PI (micro-scale roughness) and PII (complex macro-scale geometry) of objective functions and how they affect their optimisation;

(2) Build on the work of Kavetski et al. (2006c) and Skahill and Doherty (2006) and show that modern Newton-type optimisation strategies can calibrate hydrological parameters more efficiently and informatively than evolutionary search methods. This includes the ability to find the global optimum and efficiently diagnose multiple optima if they exist;

(3) Demonstrate that high-dimensional BATEA problems, with hundreds of latent variables, can be

handled using Newton-type methods, provided micro-scale roughness is reduced or eliminated. Moreover, it is shown that using analytical derivatives of the objective function can speed up the optimisation by orders of magnitude or more.

2. NEWTON-TYPE METHODS FOR SMOOTH FUNCTIONS

Newton-type methods for a smooth function $\Phi(\boldsymbol{\theta})$ are based on a sequence of steps

$$\mathbf{s}_{(k+1)} = \boldsymbol{\theta}_{(k+1)} - \boldsymbol{\theta}_{(k)} = -\tilde{\mathbf{H}}_{(k)}^{-1}\tilde{\mathbf{g}}_{(k)} \qquad (2)$$

where $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{H}}$ approximate the gradient and Hessian of Φ and k is the iteration index.

To stabilise convergence to (at least) a local optimum, the correction $\mathbf{s}_{(k+1)}$ must be controlled using line searches (which ensure sufficient decrease of Φ along $\mathbf{s}_{(k+1)}$ at each step) or trust regions (which modify the length and direction of $\mathbf{s}_{(k+1)}$ based on a local model of Φ near $\boldsymbol{\theta}_{(k)}$).

A crucial requirement for Newton-type methods is the ability to meaningfully estimate the gradient of Φ. For sufficiently smooth functions, this can be accomplished analytically or using finite difference approximations. In addition, mature and reliable methods/software exist for constructing the approximations $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{H}}$, for carrying out line searches and trust region updates, and handling constrained variables (Nocedal and Wright, 1999).

If the objective function is highly nonsmooth, Newton-type methods break down. For such problems, direct search methods (e.g., simplex and SCE searches) are used. However, these methods are generally less efficient than Newton-type methods and their performance deteriorates rapidly as the dimension of the problem increases (Nocedal and Wright, 1999).

3. OPTIMISATION PROBLEMS: IMPACT OF MODEL PROPERTIES ON THE OBJECTIVE FUNCTION

We grouped calibration difficulties into categories PI and PII because their origins and strategies for overcoming them are distinctly different.

3.1. Micro-scale roughness (PI)

Since the (micro-scale) continuity of SS-based objective functions such as (1) is controlled by the continuity of the model h, micro-scale roughness of the objective function is always a consequence of the model structure: if the model contains non-smooth or discontinuous constitutive relations (e.g., storage-discharge functions), its objective function will also be discontinuous (Kavetski *et al.*, 2006b; Kavetski and Kuczera, 2007).

Figure 3 shows the impact of model thresholds on the micro-structure of the objective function. It shows a 1D slice through the Nash-Sutcliffe profile of the VIC model coupled with the degree-day snow model, calibrated to French Broad River data (Kavetski *et al.*, 2006b). The solid line is the profile for a model implementation with a step threshold of the form

$$P_{rain} = \begin{cases} 0 & if\ T \le T_0 \\ P & otherwise \end{cases} \qquad (3)$$

where P is the observed precipitation, P_{rain} is the rainfall input into the VIC model, T is the air temperature and T_0 is the melting-point parameter.

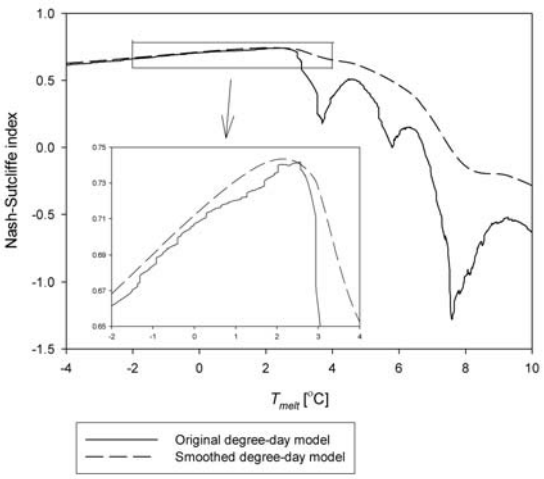

Figure 3. Impact of threshold smoothing on the objective function (from Kavetski *et al.*, 2006b).

The morass of discontinuities in the objective function precludes the application of gradient-based optimisation methods.

PI problems can be avoided by a careful numerical implementation of existing models, in particular, replacing thresholds and discontinuities in the model constitutive relationships by smooth transitions. The dashed line in Fig 1 shows the profile of a smoothed model, where the step discontinuity (3) is replaced by the smoothed step

$$P_{rain} = P \Big/ \left[1 + \exp\left(\frac{T - T_0}{m} \right) \right] \qquad (4)$$

where m is a smoothing parameter (e.g., 0.5-2 °C), normally kept constant during the calibration.

It can be seen that removing the discontinuity resulted in the objective profile becoming smooth and amenable to Newton-type optimisation. Indeed, a standard quasi-Newton code required 100-1000 function calls to optimise the model parameters (depending on the initial estimates), whereas the SCE search with comparable termination settings required 5,000-20,000 function calls. This is a significant gain in efficiency, especially when extrapolated to more complex models and high-dimensional BATEA-type calibration methodologies.

The selection of smoothing parameters is discussed in more detail by Kavetski and Kuczera (2007): moderately oversmoothing the model is preferable to undersmoothing it, because model parameters are only weakly afffected by smoothing (Figure 3) and Newton-type optimisation can be sensitive to micro-scale roughness. In addition, a sequence of calibrations gradually reducing the degree of smoothing can be implemented (but this did not appear necessary in our work thus far).

3.2. Macro-scale complexity (PII)

Macro-scale complexity of objective functions, in particular, multiple parameter optima are particularly challenging for all optimisation methods and are typically cited as the chief reason for using expensive global evolutionary searches (Duan *et al.*, 1992). Nevertheless, PII problems, including multioptimality, can also be overcome without forfeiting Newton-type methods.

Note that multioptimality is sometimes an artefact that can be removed by numerical reformulation and/or smoothing of the model. For instance, even episodic instabilities of time stepping methods can cause large distortions of parameter distributions (including secondary optima) that can be removed by more stable time stepping (Kavetski *et al.*, 2003). Even model thresholds can cause a morass of spurious local optima that disappear with smoothing (Kavetski and Kuczera, 2007).

Nevertheless, since hydrological models are intrinsically nonlinear in their parameters and the calibration data is highly uncertain (conditions that increase the likelihood of multiple parameter optima), a robust calibration strategy must be able to detect and deal with genuine parameter multioptimality. Moreover, it should have diagnostic capabilities beyond finding the global optimum and should identify and report statistically significant secondary optima, including optima lying near or on parameter bounds (Kavetski *et al.*, 2006c; Skahill and Doherty, 2006). Ideally, it should also estimate parameter uncertainty (including insensitive and correlated parameters) that can be used in its own right or to initialise a more thorough Monte Carlo analysis. It is our view that these goals are more readily and efficiently accomplished using Newton-type methods rather than evolutionary strategies.

4. MULTISTART METHODS FOR GLOBAL OPTIMISATION

The simplest strategy for global optimisation is to apply a local optimisation method to multiple seeds within the search region and assume the best result is the global optimum. One can distinguish between information-sharing methods (evolutionary schemes, eg, SCE search) versus individual-sequence methods (multistart Newton), depending on whether individual sequences use global information to adapt their local searches.

An important feature of individual-sequence methods is that they can report multiple optima found. In contrast, evolutionary schemes and standard simulated annealing schemes attempt to find the global mode directly. Early applications of multistart derivative-based methods in hydrology were problematic (Gupta and Sorooshian, 1985). Later, Duan et al. (1992) identified several causes of these difficulties, including nonsmoothness and multimodality, and advocated both (i) more sophisticated global strategies and (ii) derivative-free searches. Subsequent literature largely followed these guidelines.

Given the ability to remove micro-scale roughness from the surface of objective functions, and the advances in Newton-type optimisation algorithms, we revisit the global optimisation problem using Newton-type methods.

There are several global strategies that can be used in conjunction with Newton-type methods. The simplest is to initiate a single Newton-type sequence from uniformly distributed seeds. Rinnooy Kan and Timmer (1987) describe more sophisticated approaches that attempt to avoid starting more than a single chain in each region of attraction, which they estimate using cluster analysis. Skahill and Doherty (2006) use a different approach to achieve the same objective – each subsequent seed is picked as being maximally removed from previous search trajectories. However, it is our view that although single-seed-per-basin strategies may yield the highest efficiency in search for the global optimum, they forgo useful insights that can be obtained by uniformly seeding the search space. These insights are illustrated in the next section.

5. CASE STUDY 1: MULTISTART QUASI-NEWTON VS SCE SEARCHES

In an earlier study, Thyer et al. (1999) compared two probabilistic methods for calibrating the SFB model (a 6-parameter rainfall-runoff model): the simulated annealing method and the evolutionary-based SCE search, with the latter proving more efficient and consistent. Here we revisit the problem using a multistart quasi-Newton (QN) method, with the gradient estimated using finite differences. The multistart strategy consisted of sampling initial seeds from a uniform distribution over the feasible parameter space, and applying the QN method to each seed independently.

Figure 1 shows a 2D cross-section of the objective function of the SFB model calibrated to the Scott Creek data (Australia) (Thyer *et al.*, 1999). The global mode lies on the bound SDRmax = 0.0 and has the objective function value of Φ = -690, and there is a number of secondary optima in the interior of the parameter domain. Figure 4 shows a cross-section at a different location of the parameter space, showing additional local optima. These features make the calibration a challenging global optimisation problem.

Figure 5 shows a cumulative plot of the optima identified using multistart quasi-Newton and SCE searches. The results for the multistart SCE search agree with those previously obtained by Thyer et al., with the vast majority of the runs converging to optima with $\Phi \sim 700$. These locations correspond to the triplet of optima shown in Figure 4. It can also be seen that only ~ 0.3% of the SCE runs converged to the global optimum $\Phi \sim 692$ near the boundary of the parameter space.

The application of the multistart Newton method led to significantly different results. Due to the multioptimality of the problem, the Newton runs terminate in many different locations of the parameter space: 30% of the optima had $\Phi > 720$ (whereas all SCE runs terminated in better optima). However, 15% of the quasi-Newton sequences converged to the global mode located on the boundary, which proved very elusive for the SCE search. Another 40% of the runs converged to the $\Phi \sim 700$ optima.

Figure 2 compares the computational cost of the methods. It shows that the quasi-Newton sequences required far fewer function evaluations than the SCE search – most runs terminated after 500-1000 function calls, whereas the majority of SCE searches required 5000-8000 function calls. It is also noted that there was no special relation between the computational cost and the specific optima found – in some runs the global mode was found in fewer than average number of runs, in others it took more than average.

In addition to simply locating the global optimum, it is important that the optimisation method yield additional insight into the structure of the parameter distribution. The multistart Newton can generate at least two such insights: (i) local structure: the Hessian matrix computed during Newton-type optimisation can be used to approximate the covariance structure of the parameter distribution and (ii) global structure: the relative fractions of initial seeds yielding convergence to various local optima can be used to estimate the relative sizes of the regions of attraction, and hence the probability mass associated with each optimum.

The latter insight effectively uses the multistart framework as a hit-or-miss Monte Carlo method to estimate the volume of regions of attraction. Notably, such information can not be obtained from global methods such as the SCE search, since the information about the distinct regions of attraction is lost in the shuffling of simplex vertices (which, somewhat ironically, is one of the major strengths of the method as a global optimiser).

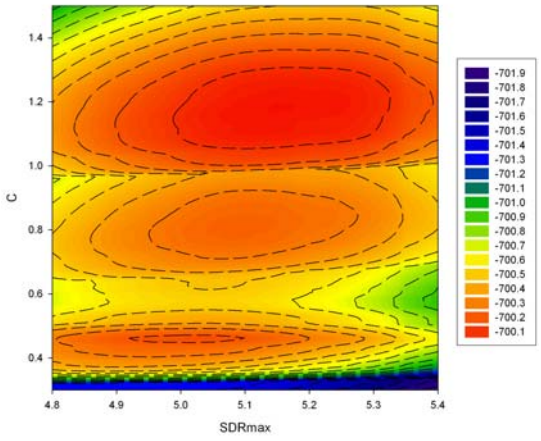

Figure 4. Triplet of local optima in a region of the parameter space adjacent to the interior mode (SDRmax ~ 4, C ~ 1.0) in Figure 1.

The knowledge of the approximate distribution of optima and their local covariance structure can be used to construct more efficient Monte Carlo samplers. For example, more efficient "hierarchical" importance samplers can be developed by first sampling a mode according to the approximate probability mass of each modal region, and then using the covariance structure for local Gaussian importance sampling.

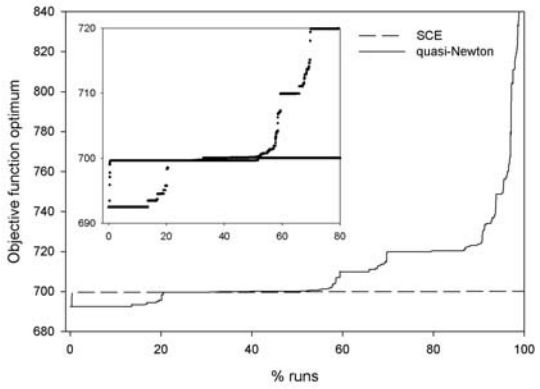

Figure 5. Distribution of optima computed using multistart quasi-Newton and SCE methods.

6. BATEA OPTIMISATION USING NEWTON-TYPE METHODS

Bayesian total error analysis (BATEA) is a calibration methodology that seeks to explicitly formulate and parameterise models of data uncertainty and model structural error (Kavetski *et al.*, 2006a; Kuczera *et al.*, 2006). For example, rainfall uncertainty can be parameterised using storm multipliers (latent variables). A Bayesian hierarchical strategy is then used to infer both the model parameters and latent variables. As a result, BATEA calibrations are computational expensive, since, eg., for a 100-storm time series, at least 100 storm multipliers need to be calibrated, as well as the model parameters themselves. In addition, BATEA assumes that the structural error of hydrological models can be described by stochastic variation of one or more model parameters over storm-event time-scales. This further increases the dimensionality of the calibration problem.

Despite the high-dimensionality of BATEA calibrations, the parameter distributions can be sampled using Markov Chain Monte Carlo methods such as Gibbs or Metropolis samplers (see Kuczera *et al.*, 2007). However, it is beneficial to pre-optimise the model parameters and latent variables to prevent the Markov chains from being trapped far from high-probability regions of the parameter distribution.

The high-dimensionality of BATEA objective functions precludes its optimisation using nonsmooth methods such as the SCE search. Consequently, we use quasi-Newton methods to pre-optimise the BATEA latent variables and model parameters. In addition, it remains important to diagnose whether the objective function is multioptimal, mandating the use of the multistart framework.

An additional challenge of high-dimensional optimisation using quasi-Newton methods is the estimation of the gradient of the objective function. If this is accomplished using finite difference methods, each component of the gradient requires at least one function call, making the optimisation progressively expensive.

The alternative is analytical differentiation of the objective function. While this can be tedious for complex models, the efficiency gains when optimising using analytical gradients can be spectacular. Table 1 shows a comparison of the cost of optimising BATEA objective functions using quasi-Newton methods with finite difference derivatives vs analytical derivatives. The hydrological model in this case study is a 2-store model with linear storage-discharge relationship. 1 year of data was used in the analysis, requiring the calibration of 71 latent variables and parameters.

Table 1. Computational cost of optimisng BATEA objective functions using quasi-Newton methods with/without analytical derivatives.

Method	Function calls	Scaled CPU time
FD-gradient	154027	15.8
Analytical	1085	1

Table 1 shows that (i) Newton-type methods can sucesfully optimise high-dimensional BATEA-type objective functions with more than a hundred latent variables, (ii) Analytical derivatives lead to a dramatic reduction in the number of function calls for optimisation, especially as the dimenson of the problem increases and (iii) the gains in CPU time are balanced by the cost of evaluating analytical derivatives, but remain significantly lower than with finite difference gradients.

A limitation of using analytical derivatives is the difficulty in differentiating the model equations. Indeed, for many models this is too tedius and may require significant code modifications. However, for cases where this can be accomplished, Newton optimisation with analytical gradients permits hitherto prohibitive analysis of parameter distributions accounting for data and model error.

7. CONCLUSIONS

Efficient calibration of conceptual hydrological models has traditionally been impeded by complex micro- and macro- structure of objective functions. Although earlier approaches favoured nonsmooth global optimisation such as the SCE method, recent work began addressing these difficulties using multistart Newton-type methods. This paper

overviews the origins of complexity of objective functions and how they can be tackled using numerical smoothing/reformulation of the model. This opens several avenues for more systematic analysis of parameter distributions, in particular, using multistart Newton-type methods. This paper makes several contributions in this direction. Firstly, we supply case studies where Newton-type methods comfortably and consistently outperform global SCE searches. Secondly, it is argued that designing multistart strategies to avoid seeding a single attraction region with multiple seeds (to prevent convergence to the same optimum and thus boost efficiency) can yield less insight into the macro-structure of the distribution (the relative sizes of attraction basins associated with multiple optima). Finally, it is shown that analytical model derivatives, in combination with numerically smooth model formulations, can significantly improve the computational efficiency and reliability of the model calibration, especially for high-dimensional calibration problems such as those arising in Bayesian hierarchical models of data and model uncertainty.

8. REFERENCES

Duan, Q., S. Sorooshian and V. K. Gupta (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4):1015-1031.

Gupta, V. K. and S. Sorooshian (1985) The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. *Water Resources Research*, *21*(4):473-485.

Kavetski, D. and G. Kuczera (2007) Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. *Water Resources Research*, *43*:W03411, 03411-03419.

Kavetski, D., G. Kuczera and S. W. Franks (2003) Semi-distributed hydrological modelling: A "saturation path" perspective on TOPMODEL and VIC. *Water Resources Research*, *39*(9):8.1-8.8.

Kavetski, D., G. Kuczera and S. W. Franks (2006a) Bayesian analysis of input uncertainty in hydrological modeling. I. Theory. *Water Resources Research*, *42*(W03407):1-9.

Kavetski, D., G. Kuczera and S. W. Franks (2006b) Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology (Special Issue on The Model Parameter Estimation Experiment - MOPEX)*, *32*(1-2):173-186.

Kavetski, D., G. Kuczera and S. W. Franks (2006c) Calibration of conceptual hydrological models revisited: 2. Improving optimisation and analysis. *Journal of Hydrology (Special Issue on The Model Parameter Estimation Experiment - MOPEX)*, *32*(1-2):187-201.

Kuczera, G., D. Kavetski, S. W. Franks and M. Thyer (2006) Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, *331*(1-2):161-177.

Kuczera, G., D. Kavetski, B. Renard and M. Thyer (2007). Bayesian Total Error Analysis For Hydrologic Models: Markov Chain Monte Carlo Methods To Evaluate The Posterior Distribution. MODSIM-2007 (this volume).

Nocedal, J. and S. J. Wright (1999) *Numerical Optimization*. Springer-Verlag, New York.

Rinnooy Kan, A. H. G. and G. T. Timmer (1987) Stochastic global optimization methods. Part I: Clustering methods. *Mathematical Programming*, *39*:27-56.

Skahill, B., E. and J. Doherty (2006) Efficient accommodation of local minima in model calibration. *Journal of Hydrology*, *in press*.

Thyer, M., G. Kuczera and B. C. Bates (1999) Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing methods. *Water Resources Research*, *35*(3):767-773.

Tolson, B. A. and C. A. Shoemaker (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*:W01413.

Vrugt, J. A., H. V. Gupta, W. Bouten and S. Sorooshian (2003) A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8):1.1-1.16.