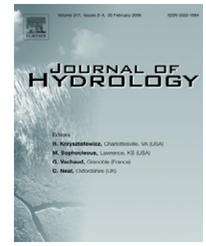
Journal of Hydrology (2008) **353**, 109–128

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol




# Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search

**Vadim Kuzmin** [a,*], **Dong-Jun Seo** [b,c], **Victor Koren** [b]

[a] *Department of Civil and Environmental Engineering, University of Melbourne, Victoria 3010, Australia*
[b] *Hydrology Laboratory, Office of Hydrologic Development, National Weather Service, 1325 East-West Highway, Silver Spring, MD 20910, USA*
[c] *University Corporation for Atmospheric Research, Visiting Scientist Programs, 3300 Mitchell Lane, Boulder, CO 80301, USA*

Received 22 March 2007; received in revised form 4 October 2007; accepted 4 February 2008

**KEYWORDS**
Automatic calibration;
Hydrologic model;
A priori estimates;
Parametric uncertainty;
Stepwise line search;
Shuffled complex evolution

**Summary** Routine availability of hydrological, geological, and other physiographic data today allows us to obtain a priori estimates of hydrologic model parameters prior to explicit model calibration. When informative a priori estimates of model parameters are available, the problem of hydrologic model calibration becomes one of filtering, i.e. improving the a priori estimates based on observations of input and output to and from the hydrologic system, respectively, rather than one of bounded global optimization based solely on the input and output data as in traditional model calibration. Given that global optimization is computationally very expensive and does not, in general, transfer the spatial patterns of soil and land surface characteristics to the model parameters, the filtering approach is particularly appealing for automatic calibration of distributed hydrologic models. Toward that ultimate goal, we explore in this work calibration of a lumped hydrologic model via limited optimization of a priori estimates of the model parameters. The technique developed for the purpose is a simple yet effective and efficient pattern search algorithm called the Stepwise Line Search (SLS). To evaluate the methodology, calibration and validation experiments were performed for 20 basins in the US National Weather Service West Gulf River Forecast Center's (NWS/WGRFC) service area in Texas. We show that SLS locates the posterior parameter estimates very efficiently in the vicinity of the a priori estimates that are comparable, in terms of reducing the objective function value, to those from global minimization. A cross validation experiment indicates that, when parametric

* Corresponding author. Tel.: +61 4 8344 5628; fax: +61 4 8344 4616.
*E-mail address:* v.kuzmin@civenv.unimelb.edu.au (V. Kuzmin).


doi:10.1016/j.jhydrol.2008.02.001

uncertainty due to lack of calibration data is considered, limited optimization of a priori parameters using SLS may be preferred to global optimization.


## Introduction

With the widespread availability of spatial data sets of soil, land cover, land use, and others, it is now possible in many parts of the world to obtain a priori estimates of hydrologic model parameters following some form of re-parameterization of them in terms of pedologic and physiographic data (Koren et al., 2000, 2003; Leavesley et al., 2003; Mertens et al., 2004; Vieux and Moreda, 2003). While such a priori estimates are undoubtedly subject to significant uncertainties due to various sources of error and scale differences in observations and modeling, they reflect, to a varying degree of accuracy, the spatial variability of the parameters, and can provide an informative first-guess estimate for the hydrologic model parameters prior to or in the absence of explicit model calibration. When such a priori estimates are available, the problem of automatic calibration becomes one of filtering, i.e., improving the a priori estimates based on observed data (typically precipitation and streamflow), rather than one of bounded global optimization as in traditional model calibration. While tremendous advances have been made in recent years in estimation of parameters in lumped hydrologic models and assessment of their uncertainty (Beven and Kirkby, 1979; Duan, 2003; Gupta et al., 2003; Koren et al., 2003; Kuczera and Parent, 1998; Schaake et al., 2001), the currently available automatic calibration techniques are generally based on global optimization which requires a very large number of function evaluations. As such, they are not very amenable to estimation of distributed parameters in fine-scale hydrologic modeling. Also, while Monte Carlo-type automatic calibration techniques may be reasonable for models that operate at daily or 6-hourly time steps, they may be computationally too expensive to be operationally viable at 1-hour time step even for lumped models. In addition to computational considerations, automatic calibration based on global optimization as practiced today in lumped modeling does not transfer the spatial patterns of pedologic and physiographic characteristics observed in the spatial data to the hydrologic model parameters very well. For example, automatically-calibrated parameters from global optimization for two adjacent basins in the same area with similar pedologic and physiographic properties may not share similarities that are duly expected from physical considerations. As such, automatic calibration via global optimization is not very conducive to hydrologic modeling over a large area where interdependence of hydrologic model parameters among adjacent basins may be important (Koren et al., 2003). Given the above observations, we argue that some combination of estimation of a priori parameters and 'limited' optimization of them offers a more effective and practical path to estimation of distributed parameters than simply extending global optimization to a distributed parameter setting (Refsgaard, 1997). The overarching objective of this work is to explore such a filtering approach toward development of an operationally viable methodology for routine automatic calibration of distributed hydrologic models. As a first step toward that ultimate goal, we focus in this work on calibration of a lumped hydrologic model. Specifically, we seek a local (as opposed to global) parameter optimization technique that offers the kind of performance and computational efficiency for calibration of lumped models to bring optimization of distributed parameters within the realm of possibility for operational hydrologic forecasting. We do recognize that successful application of such a technique to lumped models may not necessarily translate to that to distributed models, for which different strategies may be pursued (see, e.g., Eckhardt and Arnold, 2001; Madsen, 2003; Vieux et al., 2004; Heuvelmans et al., 2006; Francés et al., 2007). We note here that application of the local optimization technique described in this work to distributed hydrologic models is ongoing and the results will be reported in the near future.

In classical estimation theory, a priori information about the model parameters is treated typically as a penalty term added to the (typically quadratic) objective function (Jazwinski, 1970; Schweppe, 1973):

$$\min \quad J = [Q_o - Q_s(X)]^T P_Q^{-1} [Q_o - Q_s(X)] + [X - X_a]^T R_X^{-1} [X - X_a], \tag{1}$$

$$\text{subject to} \quad X_L \leqslant X \leqslant X_U. \tag{2}$$

In the above, denotes the vector of the observed streamflow, $X$ denotes the vector of the model parameters being estimated, $X = (x_1, \ldots, x_N)$, $Q_s(X)$ denotes the vector of the model-simulated flow corresponding to $Q_o$, $P_Q$ denotes the error covariance matrix associated with the model-simulated flow, $X_a$ denotes the vector of the a priori estimates of the model parameters, $R_X$ denotes the error covariance matrix associated with $X_a$, and $X_L$ and $X_U$ denote the vectors of the lower and upper bounds of $X$, respectively. Throughout this paper, the upper- and lowercase letters denote vector and scalar variables, respectively. Though cast as a least squares problem for simplicity and intuitiveness, the above formulation, or its variant, is essentially equivalent to Bayesian estimation (Misirli et al., 2003) or Kalman filtering (Jazwinski, 1970) under a varying set of assumptions and interpretations. The details of their relationships are not central to the development of this paper and are not given here. The above type of formulation has been widely used in groundwater modeling with gradient-based optimization techniques.

Because $Q_s(X)$ is generally a highly nonlinear function of $X$ and the a priori estimates of $X$ are subject to potentially significant biases due to errors and scale differences in observations and modeling, the above formulation faces at least two serious difficulties in estimation of hydrologic model parameters. The first is that, because of large nonlinearity of $Q_s(X)$, even a small systematic bias or random error in the a priori estimates of $X$ can lead to a very poor solution.