US Army Corps of Engineers®
Engineer Research and
Development Center

Flood and Coastal Storm Damage Reduction

A Practical Two-Phase Approach to Improve the Reliability and Efficiency of Markov Chain Monte Carlo Directed Hydrologic Model Calibration

Brian E. Skahill and Jeffrey S. Baggett   March 2020

The US Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at http://acwc.sdp.sirsi.net/client/default.

A Practical Two-Phase Approach to Improve the Reliability and Efficiency of Markov Chain Monte Carlo Directed Hydrologic Model Calibration

Brian E. Skahill

Coastal and Hydraulics Laboratory
US Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Jeffrey S. Baggett

Department of Mathematics
University of Wisconsin – La Crosse
La Crosse, WI 54601

Final report

Abstract

Markov chain Monte Carlo (MCMC) methods are widely used in hydrology and other fields for posterior inference in a Bayesian framework. A properly constructed MCMC sampler is guaranteed to converge to the correct limiting distribution, but convergence can be very slow. While most research is focused on improving the proposal distribution used to generate trial moves in the Markov chain, this work instead focuses on efficiently finding an initial population for population-based MCMC samplers that will expedite convergence. Four case studies, including two hydrological models, are used to demonstrate that using multi-level single linkage implicit filtering stochastic global optimization to initialize the population both reduces the overall computational cost and significantly increases the chance of finding the correct limiting distribution within the constraint of a fixed computational budget.

Contents

Figures and Tables

Figures

Tables

Preface

1 Introduction

1.1 Background

Markov chain Monte Carlo (MCMC) simulation has been demonstrated to be a formal and flexible means by which to obtain robust estimates of hydrologic model uncertainty (Kuczera and Parent 1998; Campbell et al. 1999; Bates and Campbell 2001; Marshall et al. 2004; Engeland et al. 2005; Vrugt et al. 2006; Vrugt et al. 2008a,b). With MCMC, a Markov chain-directed random walk yields, upon equilibrium, random draws of the target posterior distribution. However, application of MCMC to characterize the posterior probability distribution can be computationally intensive.

A key element for MCMC samplers is the proposal distribution, which generates the candidate jumps for consideration as part of the Markov chain-directed random walk of the posterior. For a given problem, there are many possible acceptable proposal distributions. However, its specific choice can dramatically impact the overall efficiency of the sampler to the target equilibrium distribution. Proposal distributions that generate either small or large jumps yield low acceptance rates and slow convergence. The primary goal is to choose a proposal distribution that is easy to sample from, generates unbiased moves, and which results in optimal mixing of the chains.

An additional important and practical MCMC implementation issue involves specifying the initial values for the chain(s). Guidance for MCMC initialization is to start with a value as close to the center of the unknown posterior distribution or to employ multiple highly dispersed initial values (Gelman and Rubin 1992). With hydrologic models, however, there is little a priori knowledge of the high probability density region. Hence, chains are often initialized by uniform or latin hypercube sampling (LHS) from an uninformative uniform prior distribution. LHS is often used for reasons of computational efficiency. With LHS, the random samples of parameter values are generated in a controlled manner such that they are representative of the real variability (Xin 2014).

Within the hydrologic modeling community, the development of MCMC samplers has focused on the definition of the proposal distribution. Vrugt et al. (2003) introduced the Shuffled Complex Evolution Metropolis

algorithm (SCEM-UA), an efficient sampling strategy that is a modification to the Shuffled Complex Evolution algorithm (SCE-UA) global optimization algorithm (Duan et al. 1992, 1993). The SCEM-UA sampling strategy adapts the proposal distribution during sampler burn-in in an attempt to accelerate the rate of convergence. However, the SCEM-UA algorithm has subsequently been reported to lack detailed balance (Vrugt et al. 2008a). Smith and Marshall (2008) compared the effectiveness and efficiency of the adaptive Metropolis (Haario et al. 2001), delayed rejection adaptive Metropolis (DRAM) (Haario et al. 2006), and the differential evolution Markov chain (DE-MC) (ter Braak 2006) MCMC samplers in terms of their capacity to support hydrologic modeling. The DE-MC algorithm performed best with a known bimodal mixed normal target distribution, and all three samplers were reported to perform equally well with respect to convergence when they were initialized from a known high posterior density. Vrugt et al. (2008a, 2009) introduced differential evolution adaptive metropolis (DREAM) — an adaptive MCMC sampler that was designed to estimate the posterior probability density of hydrologic model parameters. DREAM is a modification to the DE-MC method (ter Braak 2006) that includes adaptations directed at improving overall algorithm efficiency and effectiveness. It outperformed the DE-MC and DRAM adaptive MCMC samplers when it was applied to infer the posterior parameter distribution of a Sacramento Soil Moisture Accounting (SAC-SMA) model (Vrugt et al. 2009). Kuczera et al. (2010) exploit the decaying memory of hydrologic systems to accelerate the convergence of their Bayesian Total Error Analysis multiblock MCMC sampling methodology for improved application with conceptual rainfall runoff models.

1.2   Objective

Markov chain initialization is another practical implementation issue that impacts the number of forward model calls for an MCMC sampler to converge. In their regional hydrologic modeling study, Engeland and Gottschalk (2002) initialized the MCMC sampler with parameters optimized with respect to the likelihood function rather than with random samples from the prior distribution. Kavetski et al. (2006a,b) initialized an adaptive population-based MCMC sampler using a multistart global optimization strategy that employed a Newton-type method for the local searches. Their method (Kavetski et al. 2006b,c) was limited to smooth problems. To improve upon burn-in efficiency for an application with the Soil and Water Assessment Tool hydrology model (Arnold et al. 1998),

Yang et al. (2007) initialized their MCMC sampler using results obtained from an application of the SCE-UA global optimization strategy. Gallagher and Doherty (2007) initialized their Model Independent Markov Chain Monte Carlo Analysis (MCMC sampler (Doherty 2003) with results obtained from application of the *trajectory repulsion* stochastic global optimization strategy, which uses the Levenberg-Marquardt method (Levenberg 1944; Marquardt 1963) for its local searches (Skahill et al. 2009; Skahill and Doherty 2006). They demonstrated resultant reduced MCMC burn-in run requirements, but, as with the solution methodology of Kavetski et al. (2006a,b), a smoothness assumption is required for application of their approach.

The knowledge gained by the application and development of computer-based calibration methodologies has provided the hydrologic modeling community with a better understanding of some of the complications associated with calibrating rainfall-runoff models. These complications include the existence of multiple local optima, non-smooth response surfaces, and long valleys in parameter space that are a result of excessive parameter correlation or insensitivity (Gupta et al. 2003; Kavetski et al. 2006c,d). For example, studies have reported upon these noted difficulties with calibrating the SAC-SMA model (Duan et al. 1992; Vrugt et al. 2009).

A method that has previously been demonstrated to support efficient stochastic global optimization for watershed model calibration (Regis and Shoemaker 2007) is the implicit filtering (IMFIL) algorithm (Kelley 2011; Kelley 2001; Gilmore and Kelley 1995). IMFIL is an iterative sampling algorithm designed to optimize complex geometric response surfaces that may be non-smooth, nonconvex, and possess multiple local optima. Central finite differences are applied, and the finite difference grid stencil is refined as the optimization unfolds. Shoemaker et al. (2007) used implicit filtering for Multi-Level Single Linkage (MLSL) (Rinnooy Kan and Timmer 1987a,b) stochastic global optimization supervised local searches, and the hybrid method (i.e., MLSL-IMFIL) demonstrated improved efficiency relative to the widely used SCE method when applied to calibrate two separate watershed model example problems with a fixed computational budget.

Previous efforts to improve MCMC initialization for hydrologic models that go beyond sampling from the prior distribution have used the SCE optimization method to estimate global maxima of the likelihood function.

However, none of the hydrologic modeling studies have examined the gains in efficiency and reliability for MCMC convergence when using other optimization approaches to initialize the Markov chains. This report attempts to fill that gap.

1.3 Approach

In this report, population-based MCMC samplers are initialized based on results derived from a prior application of the stochastic global optimization method MLSL, which uses the implicit filtering algorithm for its local searches. The improved efficiencies derived with the Markov chain initialization methodology are profiled by revisiting two case study problems previously reported upon using the population-based MCMC sampler DREAM (Vrugt et al. 2009). The first example involves application of the SAC-SMA hydrology model, and the second considers use of the Hydrologic Model (HYMOD). Two initial case studies involving known and difficult multimodal targets are comprehensively explored with respect to reliability and efficiency for different MCMC sampler configurations provide a basis for the subsequent Bayesian supervised hydrologic calibration examples with the SAC-SMA and HYMOD models.

2 Methods

This section summarizes the IMFIL, the stochastic global optimization method MLSL, and how information obtained from MLSL-IMFIL is used to initialize a population of chains for subsequent MCMC-directed hydrologic model calibration. The intent of using results obtained from MLSL-IMFIL is to yield a net improvement in overall MCMC sampler reliability and efficiency relative to a simple chain-initialization strategy based upon either uniform or LHS from an uninformative uniform prior distribution.

2.1 Implicit filtering (IMFIL)

The implicit IMFIL algorithm is a finite-difference-based sampling strategy that was designed to decrease the finite difference grid stencil as the search progresses in attempts to accommodate the optimization of complex geometric response surfaces, which may be non-smooth, nonconvex, and possess multiple local optima. Response surfaces with these attributes are often reported upon with the application of hydrology models. In particular, with IMFIL, a quasi-Newton local search is performed for a series of predetermined scales, and in each case the model Jacobian is estimated using finite differences with the given grid stencil. A given application of the method stops either when all of the predefined finite difference grid stencils are employed or a fixed computational budget is exhausted.

2.2 Multi-level single linkage (MLSL)

With a local search method, if there are different *regions of attraction* in parameter space, its solution will lead to just one of possibly many objective function minima; the particular one that is found is dependent upon the user-supplied set of initial parameter values. Stochastic global optimization (GO) can be employed as a remedy. Stochastic GO algorithms estimate the global minimum of the objective function by initiating local searches from global, randomly sampled points. The local and global phases can be iterated and/or the local searches may be initiated at some or all of the globally sampled points. Stochastic GO algorithms are guaranteed to converge, with probability one, to the global minimum as the sample size approaches infinity. Stronger convergence properties are possible for some stochastic algorithms, as mentioned below. Moreover, probabilistic-based stopping criteria can be developed for stochastic GO

methods (Rinnooy Kan and Timmer 1987a,b; Törn and Žilinskas 1987); however, an a priori computational budget may preclude any concern regarding termination criteria.

One would like to utilize stochastic GO methods that are not only reliable in finding the global minimum but also efficient in the sense that they minimize the return to previously visited local minima in parameter space. A modeler would possibly also like to receive some information on the locations of non-global minima, especially if these minima are little different in magnitude from the global minimum but are widely separate from it in parameter space. Ideally, a single local search within the region of attraction of each local minimum would be performed. This would not only ensure that each local minimum is identified just once but also that in fact all local minima is found. However, it is also desirable to employ a method that works well if one has a predetermined computational budget in that for a given effort it compares favorably with other methods. Clustering methods were designed to accommodate these requirements. They are variants of Multistart (the Multistart method samples points from a uniform distribution over the feasible parameter space and starts a local search from each of the sample points), and the basic idea behind them is to group close points, sampled from the feasible parameter space and for which the specified groups presumably relate to actual regions of attraction in parameter space, and to apply a single local search procedure within each identified cluster. Either reduction (wherein sampled points associated with the highest objective function values are temporarily removed) or concentration (wherein the sampled points are transformed through application of one or a few iterations of a local search procedure) is employed to identify a reduced sample as part of the clustering process to provide some assurance that in fact the specified groups correspond to regions of attraction of actual local minima. Clustering methods are often iterative in that the global and local phases are repeated sequentially until a stopping criterion is satisfied.

With clustering methods, it is possible that one cluster intersects multiple regions of attraction; hence, the global minimum could be missed or that one region of attraction contains more than one cluster, thus allowing for the same local minimum to possibly be identified more than once. MLSL is a clustering method that was developed to reduce the probability of not finding a local minimum or of finding a local minimum more than once (Rinnooy Kan and Timmer 1987a,b). MLSL mimics clustering by calculating

a critical distance r_k at each iteration, k. This critical distance can be used to build clusters, but instead, in MLSL, the decision as to whether a local search is to be initiated from a given reduced sample point is simply based on whether there exists another reduced sample point within the distance r_k of the given point with a corresponding lower objective function value. The critical distance r_k is reduced at each iteration. Under certain assumptions, MLSL has stronger convergence properties than simpler stochastic global optimization algorithms. First, if the algorithm continues forever, the number of local searches performed is finite. Second, if r_k tends to zero with increasing k, then every local minimum will be identified in finite time with probability one. The MLSL implementation follows that of Rinnooy Kan and Timmer (1987a,b) with a slight modification to sometimes avoid repeatedly finding the same local minima.

2.3   Markov chain Monte Carlo (MCMC)

MCMC is a formal Bayesian approach for estimating the posterior probability distribution of the adjustable model parameters. It treats the adjustable model parameters as random variables and relies upon Bayes' Theorem to compute their joint posterior probability distribution. Bayes' Theorem indicates that the posterior distribution is proportional to the product of the prior distribution, prescribed based on the modeler's best judgment, expert opinion, or literature estimates, among possible others, and the likelihood function (i.e., conditional distribution), which encapsulates the conditioning process with the observed dataset. The idea behind MCMC is that while one wants to compute a probability density, $p(\mathbf{p}|y)$, where $\mathbf{p}$ and y represent the vector of adjustable model parameters and the observed data, respectively, there is the understanding that such an endeavor may be impractical. Additionally, simply generating a large random sample from the probability density is as good as knowing its exact form. Hence, the problem then becomes one of efficiently generating a large number of random draws from $p(\mathbf{p}|y)$. It was discovered that an efficient means to this end is to construct a Markov chain, a stochastic process of values that unfold in time, with the following properties: (1) the state space (set of possible values) for the Markov chain is the same as that for $\mathbf{p}$; (2) the Markov chain is easy to simulate from; and (3) the Markov chain's equilibrium distribution is the desired probability density $p(\mathbf{p}|y)$. By constructing such a Markov chain, one could then simply run it to equilibrium (and this period is often referred to as the sampler *burn-in* period) and subsequently sample from its stationary distribution. The Gelman and Rubin quantitative diagnostic measure (Gelman and Rubin

1992) is commonly employed to assess chain convergence (Vrugt et al. 2003; Vrugt et al. 2008a; Vrugt et al. 2008b; Vrugt et al. 2009). A Markov chain with the above-mentioned properties can be constructed by choosing a symmetric proposal distribution and employing the Metropolis acceptance probability (Metropolis et al. 1953) to accept or reject candidate points. MCMC simulation is used for inference, search, and optimization with hydrologic models (Harmon and Challenor 1997; Kuczera and Parent 1998; Campbell et al. 1999; Bates and Campbell 2001; Makowski et al. 2002; Qian et al. 2003; Kanso et al. 2003; Vrugt et al. 2003; Vrugt et al. 2008a,b; Vrugt et al. 2009). This study employs for purposes of demonstration the population-based MCMC samplers DE-MC, DE-MC$_z$ (ter Braak and Vrugt 2008), and DREAM$_z$.

2.4 MCMC-MLSL-IMFIL

MLSL-IMFIL has been demonstrated to be an efficient way to calibrate a hydrologic model (Shoemaker et al. 2007). MLSL-IMFIL is applied with a predetermined computational budget specified not only for each IMFIL search but also for the MLSL supervised stochastic global optimization run-in aggregate, to determine a collection of local maxima of the posterior probability density function. An application of MLSL-IMFIL results in a set, of size w, of maxima of the posterior probability distribution. These locations are given by $\Omega = \{p_1, p_2, ..., p_w\}$ wherein each local maxima p_i has been identified n_i times. Each MLSL-IMFIL identified maximum from the set Ω is ranked based on its computed likelihood. The r largest ranked maxima from Ω are retained, and this reduced set is denoted by Ω_r. The maxima from Ω_r are used to initialize the population to subsequently evolve using Markov Chain Monte Carlo simulation. If the number of points needed to initialize the population is greater than of size r, then additional points are obtained as needed by slightly perturbing the points from Ω_r.

3   Examples and Results

Four case studies demonstrate a proof of concept of the population-based MCMC sampler initialization strategy, to improve sampler efficiency and reliability. The first two case studies mimic the proposed initialization strategy by considering applications to known target distributions wherein the MCMC population is initialized with random draws from the target. The final two cases studies apply the MLSL-IMFIL initialization strategy together with DE-MC$_z$ and DREAM$_z$ to calibrate the SAC-SMA and Hydrologic Model (HYMOD) hydrology model applications.

3.1   Bimodal normal target

Mixed normal distributions have served as example problems to profile the capacities of different MCMC samplers to fully characterize a known target. Liu et al. (2000) and also Liang and Wong (2001) considered the five dimensional mixed Gaussian

$$\pi(\mathbf{x}) = \frac{1}{3} N\big((0,0,0,0,0), I_5\big) + \frac{2}{3} N\big((5,5,5,5,5), I_5\big) \qquad (1)$$

For this distribution, the distance between the two modes is $5\sqrt{5} \sim 11.2$, and it is a documented difficult problem for Metropolis samplers. Ter Braak (2006) introduced the population-based DE-MC MCMC sampler and evaluated its performance considering the same five-dimensional mixed normal distribution as in Liu et al. (2000) and Liang and Wong (2001), but made more difficult by further separating the two modes:

$$\pi(\mathbf{x}) = \frac{1}{3} N\big((-5,-5,-5,-5,-5), I_5\big) + \frac{2}{3} N\big((5,5,5,5,5), I_5\big) \quad (2)$$

Not only based on this previous documented work, but also lessons learned from experience applying MCMC samplers to simultaneously optimize and infer conceptual hydrologic model parameters, defined here for the purposes of illustration is a mixture of two normal distributions on the square (-10,10) X (-10,10):

$$\pi(\mathbf{x}) = \frac{1}{2} N\big((-5,0), I_2\big) + \frac{1}{2} N\big((5,0), 0.0025 I_2\big) \qquad (3)$$

For this known target, one mode is wide while the other is narrow, and the distance between the two modes is specified (to be equal to 10) in a manner such that they are far apart from each other. Figure 1 is a plot characterizing the normal mixture distribution as defined in Equation 3, derived by application of an MCMC sampler, and presented for the reader to visualize the difficult nature of the problem.

Figure 1. Bimodal normal mixture distribution of Equation 3.
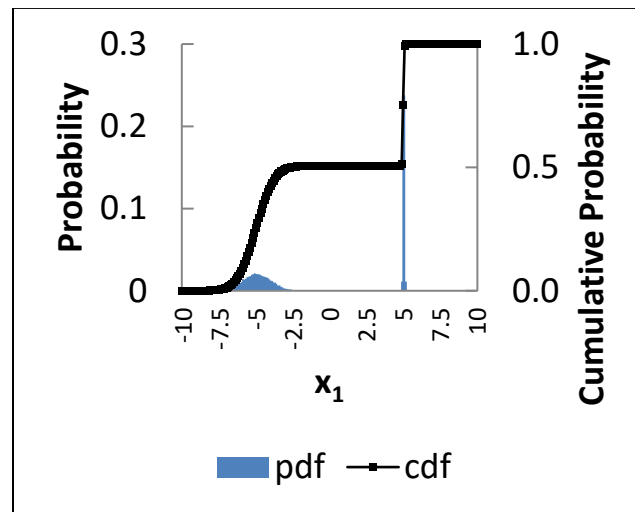
A total of 18 distinct numerical experiments were performed for the known target as defined in Equation 3 using either the population-based MCMC sampler DE-MC (ter Braak 2006) or its adaption DE-MC$_z$ (ter Braak and Vrugt 2008). For each of the experiments, 30 random trials were performed, the sampler input parameter, which weights the two unique vectors from either the current (DE-MC) or past states (DE-MC$_z$) in the definition of the jump proposal, excluding the current chain index and randomly selected without replacement, was defined to be equal to 1 approximately every fifth generation to mitigate against the potential of becoming trapped in a single mode within a multimodal distribution (Vrugt et al. 2009), and 1,000,000 evolutions were arbitrarily specified to define sampler burn-in, followed by an arbitrarily specified number of post burn-in monitoring period runs. The first 13 experiments involved different configurations with the MCMC sampler DE-MC, including the consideration of population size, initialization of the population using either LHS from the uninformative uniform prior distribution defined for the interval [-10,10], or by taking two random draws from each mode of the target distribution, and employment, or not, of an outlier chain removal and replacement functionality implemented as defined in Vrugt et

al. (2008), and if operative, by considering different levels of outlier detection. The final five experiments employed the adaptive sampler DE-MC$_z$, with the initial row size, M_o, of the matrix $\mathbf{Z}$ that contains the current and past states of the chains specified to be equal to either 20 (for experiments 14, 17, and 18) or 40 (for experiments 15 and 16), the matrix $\mathbf{Z}$ initially populated using LHS from the uninformative uniform prior distribution defined for the interval [-10,10], and the first N rows of $\mathbf{Z}$ used to define the initial population. In the final two experiments, the first four rows of $\mathbf{Z}$ were replaced with two random draws from each mode of the target distribution as defined in Equation 3. The fourteenth, fifteenth, and seventeenth experiments considered a population size of 4 whereas the sixteenth and eighteenth experiments specified N to be equal to 2, the dimension size of the problem. For each trial in a given experiment, a thinned history of the chains for the first random variable was plotted to observe if the chains did or did not see both modes during and upon completion of sampler burn-in. The configurations for each of the 18 experiments and their related results are summarized in Table 1.

Table 1. Summary of the 18 numerical experiments for the bimodal normal mixture distribution of Equation 3 (N = No; Y = Yes; LHS = latin hypercube sampling; U[-10,10] = uniform distribution defined for the interval [-10,10]).

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See Both Modes at the End of Sampler Burn-In
1	DE-MC	4	LHS from U[-10,10]	N	3/30
2	DE-MC	4	LHS from U[-10,10]	Y; 1.5 IQR[1]	0/30
3	DE-MC	4	Two random draws from each mode of the target distribution	N	30/30
4	DE-MC	4	Two random draws from each mode of the target distribution	Y; 1.5 IQR	11/30
5	DE-MC	4	Two random draws from each mode of the target distribution	Y; 2.0 IQR	14/30
6	DE-MC	4	Two random draws from each mode of the target distribution	Y; 3.0 IQR	16/30
7	DE-MC	4	Two random draws from each mode of the target distribution	Y; 5.0 IQR	25/30
8	DE-MC	8	LHS from U[-10,10]	N	3/30
9	DE-MC	8	LHS from U[-10,10]	Y; 1.5 IQR	5/30
10	DE-MC	20	LHS from U[-10,10]	N	10/30
11	DE-MC	20	LHS from U[-10,10]	Y; 1.5 IQR	7/30
12	DE-MC	40	LHS from U[-10,10]	N	18/30
13	DE-MC	40	LHS from U[-10,10]	Y; 1.5 IQR	20/30
14	DE-MC$_z$	4	Z, of row size 20, is initialized using LHS from U[-10,10]; X is initialized using the first N rows of Z	N	26/30

---

[1] interquartile range

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See Both Modes at the End of Sampler Burn-In
15	DE-MC$_z$	4	Z, of row size 40, is initialized using LHS from U[-10,10]; X is initialized using the first N rows of Z	N	29/30
16	DE-MC$_z$	2	Z, of row size 40, is initialized using LHS from U[-10,10]; X is initialized using the first N rows of Z	N	30/30
17	DE-MC$_z$	4	Z, of row size 20, is initialized using LHS from U[-10,10]; The first N rows are replaced with two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
18	DE-MC$_z$	2	Z, of row size 20, is initialized using LHS from U[-10,10]; The first 2N rows are replaced with random draws from the target distribution; X is initialized using the first N rows of Z	N	30/30

The population-based MCMC sampler DE-MC, initialized using LHS from the uniform prior, saw both modes of the known target distribution for only 3 of the 30 trials for population sizes equal to 4 and 8. The number of successes increased to 10 and 18 when the population size was increased to 20 and 40, respectively. With the DE-MC sampler, when the population of size 4 was initialized using two random draws from each mode of the target, and outlier chain removal and replacement was inactive, all 30 trials were successful. For the same DE-MC sampler configuration, but with outlier chain removal and replacement active, the success rate

progressively decreased as the IQR-based outlier detection level increased. At a detection level of 1.5 IQR, the number of successes was 11. At the suggested detection level of 2.0 IQR (Vrugt et al. 2008, 2009), the number of observed successes was 14. Regardless of population size for DE-MC simulations whose populations were initialized from the uninformative uniform prior, when outlier chain removal and replacement was operative at the 1.5 IQR detection level, the results do not indicate that the functionality positively or negatively impacted sampler reliability. Application of the adaptive sampler DE-MC$_z$, whose jump proposals are a function not only of the current but also a thinned history of past states and whose initialization is more exhaustive for an equivalent population size, by a factor of 5 as it was configured in the fourteenth experiment when compared with the DE-MC sampler profiled in the first experiment, both with N set equal to 4, dramatically improved the overall measured success rate from 10% to 86.66% when initialization involved LHS of the uniform prior. In experiment 15, the population size remained constant, but the initial row size of the matrix $\mathbf{Z}$ was doubled to equal 40, and the success rate further improved. In experiment 16, M_0 remained fixed at 40, but the population size was reduced to equal 2; however, the sampler success rate did not decrease. For the adaptive sampler DE-MC$_z$ at N equal to 4, when the initial M_0 rows of $\mathbf{Z}$ are filled using LHS from the uninformative uniform prior, and the initial N rows, characterizing the initial state for $\mathbf{X}$, are replaced with two random draws from each mode of the target distribution, the success rate improved to 100%. The success rate for the sampler DE-MC$_z$ remained at 100% when the population size was reduced to 2, equal in size to the dimensionality of the problem, even though while the matrix $\mathbf{Z}$ did, there was no assurance that for each trial the initial population $\mathbf{X}$ contained information about both modes of the target distribution.

3.2 Twenty-component mixture normal target

Liang and Wong (2001) considered simulation from the two-dimensional normal mixture distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{20} w_i exp\left(-\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i)\right) \tag{4}$$

with $\sigma = 0.1$, $w_i = 0.05$, and the mean vectors uniformly drawn from the square [0,10] x [0,10] as listed in Table 2.

Table 2. The mean vectors of the 20 components of the mixture normal distribution.

i	μ_{i1}	μ_{i2}	i	μ_{i1}	μ_{i2}	i	μ_{i1}	μ_{i2}	i	μ_{i1}	μ_{i2}
1	2.18	5.76	6	3.25	3.47	11	5.41	2.65	16	4.93	1.50
2	8.67	9.59	7	1.70	0.50	12	2.70	7.88	17	1.83	0.09
3	4.24	8.48	8	4.59	5.60	13	4.98	3.70	18	2.26	0.31
4	8.41	1.68	9	6.91	5.81	14	1.14	2.39	19	5.54	6.86
5	3.93	8.82	10	6.87	5.40	15	8.33	9.50	20	1.69	8.11

A total of 27 distinct numerical experiments were performed using either the population-based MCMC sampler DE-MC (ter Braak 2006) or its adaption DE-MC$_z$ (ter Braak and Vrugt 2008) for the challenging known target as defined in Equation 4 but made more difficult by assigning the standard deviation equal in value to 0.01 rather than 0.1. The distances between component 4 and its nearest neighboring component and component 15 and its nearest neighboring component (excluding component two) are equal in value to 315 and 384 times the standard deviation, respectively. For all but 1 of the 27 numerical experiments, 30 random trials were performed, the sampler input parameter which weights the two unique vectors from either the current (DE-MC) or past states (DE-MC$_z$) in the definition of the jump proposal, excluding the current chain index and randomly selected without replacement, was defined to be equal to 1 approximately every fifth generation to mitigate against the potential of becoming trapped in a single mode within a multimodal distribution (Vrugt et al. 2009), and 1,000,000 evolutions were arbitrarily specified to define sampler burn-in, followed by an arbitrarily specified 5,000,000 post burn-in monitoring period runs.

The first six experiments involved different configurations with the MCMC sampler DE-MC, including the consideration of population size, initialization of the population using either LHS from the uninformative uniform prior distribution defined for the box [0,10] X [0,10] or by taking one or two random draws from each mode of the target distribution, and employment, or not, of an outlier chain removal and replacement functionality implemented as defined in Vrugt et al. (2008), and if operative, by considering different levels of outlier detection. The following 21 experiments employed different configurations of the adaptive sampler DE-MC$_z$, considering different population sizes, different values for the initial row size, M_o, of the matrix **Z** that contains the current

and past states of the chains, and how the matrix **Z** was initially populated, including use of LHS from the box [0,10] X [0,10], random draws from each mode of the target distribution as defined in Equation 4, or a combination thereof. For each trial in a given experiment, a thinned history of the evolving population, post burn-in, was plotted to observe if the chains did or did not correctly see all 20 modes during the monitoring period. The configurations for each of the 27 experiments and their related results are summarized in Table 3.

Table 3. Summary of the 27 numerical experiments for the 20 component normal mixture distribution of Equation 1 (N = No; Y = Yes; LHS = latin hypercube sampling; U[0,10] = uniform distribution defined for the interval [0,10]).

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
1	DE-MC	20	LHS from U[0,10]	N	0/30
2	DE-MC	20	One random draw from each mode of the target distribution	N	11/30
3	DE-MC	40	Two random draws from each mode of the target distribution	N	18/30
4	DE-MC	40	Two random draws from each mode of the target distribution	Y; 1.5 IQR	3/30
5	DE-MC	40	Two random draws from each mode of the target distribution	Y; 2.0 IQR	10/30
6	DE-MC	40	Two random draws from each mode of the target distribution	Y; 3.0 IQR	17/30
7	DE-MC$_z$	20	**Z**, of row size 200, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	15/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
8	DE-MC$_z$	20	$\mathbf{Z}$, of row size 400, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	24/30
9	DE-MC$_z$	20	$\mathbf{Z}$, of row size 1000, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	29/30
10	DE-MC$_z$	20	$\mathbf{Z}$, of row size 1500, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	30/30
11	DE-MC$_z$	10	$\mathbf{Z}$, of row size 1000, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	29/30
12	DE-MC$_z$	10	$\mathbf{Z}$, of row size 1500, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	30/30
13	DE-MC$_z$	4	$\mathbf{Z}$, of row size 1500, is initialized using LHS from U[0,10]; $\mathbf{X}$ is initialized using the first N rows of $\mathbf{Z}$	N	27/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
14	DE-MC$_z$	2	**Z**, of row size 1500, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	26/30
15	DE-MC$_z$	4	**Z**, of row size 2000, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	29/30
16	DE-MC$_z$	2	**Z**, of row size 2000, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	27/30
17	DE-MC$_z$	2	**Z**, of row size 2500, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	9/10
18	DE-MC$_z$	2	**Z**, of row size 3000, is initialized using LHS from U[0,10]; **X** is initialized using the first N rows of **Z**	N	30/30

| Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
19	DE-MC$_z$	20	Z, of row size 200, is initialized using LHS from U[0,10]; The first N rows are replaced with one random draw from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
20	DE-MC$_z$	20	Z, of row size 200, is initialized using LHS from U[0,10]; The first 2N rows are replaced with two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
21	DE-MC$_z$	10	Z, of row size 200, is initialized using LHS from U[0,10]; The first 2N rows are replaced with one random draw from each mode of the target distribution; X is initialized using the first N rows of Z	N	28/30
22	DE-MC$_z$	10	Z, of row size 200, is initialized using LHS from U[0,10]; The first 4N rows are replaced with two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
23	DE-MC$_z$	4	**Z**, of row size 200, is initialized using LHS from U[0,10]; The first 10N rows are replaced with two random draws from each mode of the target distribution; **X** is initialized using the first N rows of **Z**	N	29/30
24	DE-MC$_z$	4	**Z**, of row size 200, is initialized using LHS from U[0,10]; The first 20N rows are replaced with two copies of two random draws from each mode of the target distribution; **X** is initialized using the first N rows of **Z**	N	30/30
25	DE-MC$_z$	2	**Z**, of row size 200, is initialized using LHS from U[0,10]; The first 40N rows are replaced with two copies of two random draws from each mode of the target distribution; **X** is initialized using the first N rows of **Z**	N	30/30
26	DE-MC$_z$	4	**Z**, of row size 80, populated with two copies of two random draws from each mode of the target distribution; **X** is initialized using the first N rows of **Z**	N	30/30

| Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
27	DE-MC$_z$	2	Z, of row size 80, populated with two copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30

The first experiment, which involved application of the MCMC sampler DE-MC with N equal to 20, and the population initialized using LHS from the box [0,10] X [0,10] did not yield a single success across all 30 trials. The second experiment resulted in 11 successes when only the initialization of the population was changed to involve a single random draw from each component of the distribution. In the third experiment, when the population size was doubled to 40, and two random draws from each component of the distribution were used to initialize the population, the number of successes further increased to 18. The next three experiments were configured in the same way as the third experiment, but to also include application of outlier chain removal and replacement at three different outlier detection levels, viz., 1.5 IQR in experiment 4, 2.0 IQR in experiment 5, and 3.0 IQR in the experiment 6. The fourth experiment resulted in a drop in the number of successes from 18 to 3. The fifth experiment yielded 10 successes, and application of the sixth experiment resulted in 17 successes. Figure 2(a) and Figure 2(b) are plots of a thinned history of the saved post burn-in random draws associated with trials 1 and 15 from experiment 6, respectively. Examining Figure 2 and the contents of Table 3, it is clear that in the first trial, the thirteenth component of the distribution was completely missed and that in the fifteenth trial, the first component was missed as the sampler was configured and applied.

Figure 2. Scatter plots of post burn-in random draws for two unsuccessful trials from experiment 6 and one successful trial from experiment 18.
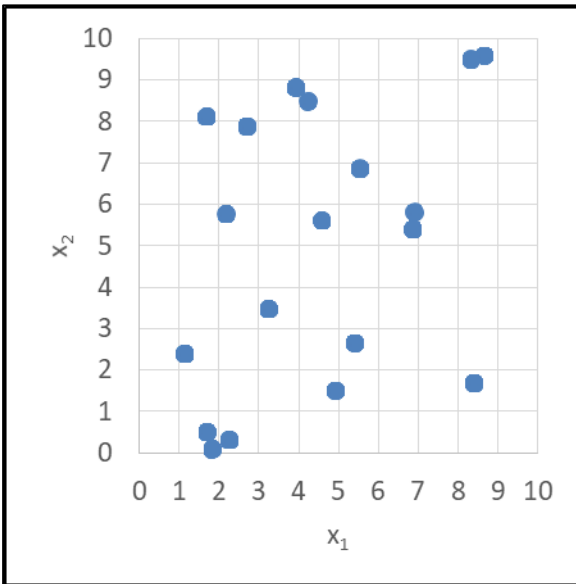
(a)Experiment 6, trial 1.
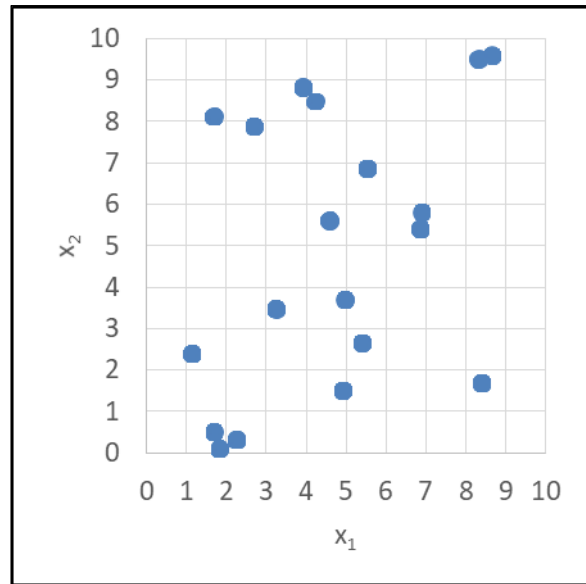
(b)Experiment 6, trial 15.
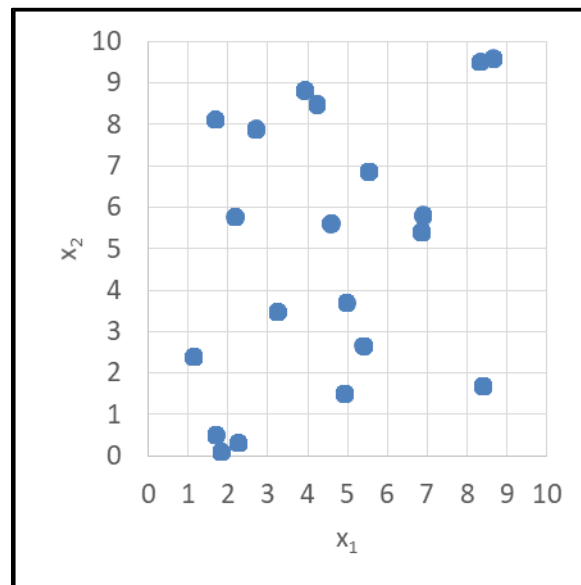
(c)Experiment 18, trial 15.

The remaining experiments involved application of the adaptive sampler DE-MC$_z$ (ter Braak and Vrugt 2008). In experiments 7 through 25, the matrix $\mathbf{Z}$ was initially populated using LHS from the square [0,10] X [0,10] and the population $\mathbf{X}$ was initialized using the first N rows of $\mathbf{Z}$. However, for experiments 19 through 25, the first 20 to 80 rows of Z were

subsequently replaced with one, two, or two copies of two random draws from each component of the target distribution. In the final two experiments, the entire initial row space of **Z** was filled with two copies of two random draws from each component of the target distribution.

Experiments 7 through 10 involved application of the adaptive sampler DE-MC$_z$ (ter Braak and Vrugt 2008) with N equal to 20 and increasing values for the initial row space size, M_o, of the matrix **Z**. The number of successful trials progressively increased as the value for M_o increased. With M_o equal to 200, only one-half of the trials were successful. When M_o was increased to equal to 1,000 in experiment 9, 29 of the 30 trials were successful. With M_o equal to 1,500, all 30 trials were successful in experiment 10. Experiments 11 and 12, which examined sampler performance when the population size was reduced to 10, demonstrated the same success rates as their similarly configured counterparts, viz., experiments 9 and 10, respectively. Experiments 13 through 18 examined sampler performance for population sizes equal in value to either 4 or 2. The initial row space size of the matrix **Z** needs to be approximately 3,000 to yield a 100% success rate across all 30 trials when the population size is equal in value to 2. For the same initial row space size for the matrix **Z**, a larger population size yields a higher success rate, as observed when one compares the results obtained across experiments 15 and 16 and also 12 through 14. The results shown in Figure 2(c), which show the empirically derived distributions from the monitoring period random draws for the first random variable, were obtained from the fifteenth trial associated with the eighteenth experiment.

Regardless of whether one or two random draws from each component of the distribution are used to replace the initial 20 or 40 rows of **Z**, when the adaptive sampler DE-MC$_z$ is configured with a population size of 20 and the initial row space size for **Z** is equal to 200, all 30 trials successfully characterize the target distribution. When the population size is reduced to 10, all 30 trials are successful when two random draws from each component replace the first 40 rows of **Z**; however, two of the trials are unsuccessful when only one random draw from each component of the distribution replaces the first 20 rows of **Z**. When the population size is further reduced to be equal in value to 4, one of the trials is unsuccessful when two random draws from each component of the target replace the first 40 rows of **Z**. Regardless of whether the population size is equal in value to 4 or 2, or the initial row space size for **Z** is equal to 200 or 80,

when the first 80 rows of **Z** are replaced or populated with two copies of two random draws from each component of the target distribution all 30 trials are successful. A limited series of additional numerical experiments, each involving 30 trials, were performed to quantify, for two separate population sizes, the computational cost to achieve the completion of sampler burn-in for different configurations of the adaptive sampler DE-MC$_z$ with the known target distribution defined in Equation 4. Due to the problematic nature of simply using the Gelman and Rubin quantitative convergence diagnostic to evaluate sampler convergence, for each trial in a given experiment, the completion of sampler burn-in was defined to occur when the percent error was less than 1 for both components of the mean vector and also the main diagonal and off-diagonal components of the covariance matrix associated with the two-dimensional 20-component normal mixture distribution defined in Equation 4. Experiments 18, 27, 10, and 20 summarized in Table 3 were repeated with the previously noted sampler convergence criteria specified as operative, as were also the additional related sampler configurations summarized in Table 4. Table 5 summarizes the results associated with each experiment.

Table 4. Additional experiments performed to evaluate sampler efficiency (N = No; Y = Yes; LHS = latin hypercube sampling; U[0,10] = uniform distribution defined for the interval [0,10]; MLSL-N = Multi Level Single Linkage global optimization method with Newton's method used for local searches).

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
28	DE-MC$_z$	2	**Z**, of row size 160, populated with four copies of two random draws from each mode of the target distribution; **X** is initialized using the first *N* rows of **Z**	N	30/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
29	DE-MC$_z$	2	Z, of row size 320, populated with eight copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
30	DE-MC$_z$	2	Z, of row size 400, populated with 10 copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
31	DE-MC$_z$	2	Z, of row size 600, populated with 15 copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
32	DE-MC$_z$	2	Z, of row size 800, populated with 20 copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
33	DE-MC$_z$	20	Z, of row size 200, is initialized using LHS from U[0,10]; The first 4N rows are replaced with two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
34	DE-MC$_z$	20	Z, of row size 160, populated with four copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
35	DE-MC$_z$	20	Z, of row size 200, populated with five copies of two random draws from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30
36	DE-MC$_z$	2	Z, of row size 160, populated, using MLSL-N, with four copies of two upgrade vectors obtained from each mode of the target distribution; X is initialized using the first N rows of Z	N	30/30

Experiment	MCMC Sampler	Population Size	Population Initialization	Outlier Chain Removal and Replacement; Outlier Detection	Number of Trials for which the Chains Did See All Modes
37	DE-MC$_Z$	2	**Z**, of row size 320, populated, using MLSL-N, with eight copies of two upgrade vectors obtained from each mode of the target distribution; **X** is initialized using the first *N* rows of **Z**	N	30/30

With the exception of experiments 36 and 37, for a given experiment, the mean number of total model calls reported in Table 5 is the computed average based on all 30 trials, excluding the cost of initialization. The reported total model calls excluded the cost associated with initialization into the overall accounting since 10 of these 12 experiments involved the use of draws that were randomly generated from the known target rather than identified by using an actual search method. However, for experiments 36 and 37, the matrix **Z** and also the initial population **X** were initialized using results obtained from application of the stochastic global optimization method MLSL, which employed Newton's method for its local searches. For these two experiments, the total model calls reported in Table 5 do account for the cost of initialization. Also reported in Table 5 for each population size are the average percent reductions, based on all 30 trials, achieved when intelligence is injected into the initialization of the MCMC supervised sampling process relative to simply using LHS of the box [0,10] X [0,10] with a density previously identified to be necessary to ensure sampler reliability across all trials. Among the limited sampler configurations that explored the impact of incorporating intelligence into MCMC sampler initialization rather than reliance upon LHS from flat priors, for population sizes of 2 and 20, the greatest mean percent reductions in total model calls achieved, excluding consideration of initialization costs, were approximately 60% and 43%, respectively. The lowest mean percent reduction in total model calls achieved for either population size was approximately 30%. For each of the experiments that employed random draws from the posterior itself for initialization, relative to using LHS of the box [0,10] X [0,10], comparable percent reductions

were also observed in the variability of the total number of model calls. For the two experiments that utilized results obtained from a prior application of MLSL to initialize the MCMC sampler, the mean percent reductions relative to the cost of experiment 18 were approximately 40%.

Table 5. The mean number of total model calls required to achieve burn-in requirements (except for experiments 36 and 37, excluding the cost of initialization), and the mean percent reductions relative to an initialization using LHS at a density level previously identified to maintain sampler reliability across all thirty trials. (*=includes cost of initialization; LHS = latin hypercube sampling)

Experiment	Population Size	Mean Number of Total Model Calls	Mean Percent Reduction
18	2	2,907,909 2,910,909*	
27	2	1,973,700	32.1
28	2	1,614,828	44.5
29	2	1,175,211	59.6
30	2	1,227,842	57.8
31	2	1,758,285	39.5
32	2	1,396,735	52.0
36	2	1,763,994	39.4
37	2	1,830,834	37.1
10	20	3,209,160	
20	20	2,004,868	37.5
33	20	2,203,707	31.3
34	20	1,838,475	42.7
35	20	1,946,834	39.3

3.3 Hydrologic model applications

Vrugt et al. (2009) demonstrated application of the DREAM sampler to infer the posterior distribution of 13 (d=13) SAC-SMA hydrologic model parameters. Their case study is revisited herein, and the five-parameter (d=5) HYMOD hydrology model is also considered; however, instead, DE-MC$_z$ and DREAM$_z$ are used, initialized either by application of LHS or using results obtained from application of MLSL-IMFIL.

Two experiments were performed for each hydrology model and each sampler wherein 30 trials were performed for each experiment. For the first experiment, the populations were initialized with LHS while for the second experiment MLSL-IMFIL was used to initialize both **X** and **Z**. For the HYMOD experiments, MLSL-IMFIL cost 1946 model evaluations on average to converge. For each SAC-SMA, trial a budget of 200,000 model calls was used for MLSL-IMFIL. For the MLSL-IMFIL experiments, the reported number of model evaluations includes the model evaluations performed by MLSL-IMFIL and by the MCMC sampler. (See Table 6 for the experimental results for the hydrology models.)

Table 6. Experimental results for the hydrology models. (For experiments denoted by LHS+, the first M_+ rows of Z are filled with information derived from MLSL-IMFIL or random draws from the known target itself ($N \leq M_+ \leq Mo$). The first N rows of Z define X. When Z is initialized solely using LHS, these experiments are denoted by LHS.) (* Four of the thirty trials did not converge within the limit of 500,000 model calls; thus, the average number of model calls and percent reduction should both be larger.)

Hydrologic Model	Sampler	Initialization	Average Number of Model Evaluations	% Reduction by LHS+
HYMOD	DE-MC$_z$	LHS M_o=50, N=10	94651*	
HYMOD	DE-MC$_z$	LHS+ M_o=50, N=10, M_+=10	17314	82%*
HYMOD	DREAM$_z$	LHS M_o=50, N=5	15118	
HYMOD	DREAM$_z$	LHS+ M_o=50, N=5, M_+=5	5987	60%
SAC-SMA	DE-MC$_z$	LHS M_o=130, N=26	753098	
SAC-SMA	DE-MC$_z$	LHS+ M_o=130, N=26, M_+=26	276619	63%
SAC-SMA	DREAM$_z$	LHS M_o=130, N=13	486086	
SAC-SMA	DREAM$_z$	LHS+ M_o=130, N=13, M_+=13	256329	47%

4   Discussion and Conclusions

The primary intent of the numerical experiments involving the known bimodal normal mixture target was to communicate the impact of population initialization on sampler reliability. LHS from uniform priors is commonly applied when MCMC is used for hydrologic model calibration. This form of population initialization, for N equal to 4, yielded success rates of 10% and 87% for the DE-MC and DE-MC$_z$ samplers, respectively. The results of additional experiments with the bimodal normal mixture target suggest that injecting more intelligence into the initialization process has clear potential to improve sampler reliability, dramatically in the case for the profiled population-based MCMC sampler DE-MC and less so for its profiled adaption, DE-MC$_z$. Both profiled population-based MCMC samplers exhibited 100% success rates when the population to evolve was initialized using random draws from the target distribution.

Observed strengths of the profiled adaptive sampler DE-MC$_z$ are that its jump proposals are a function not only of the current population but also a thinned history of its past states and that its initialization involves filling the matrix $\mathbf{Z}$, whose initial row space is greater than the population size, N. With the initial row space size for $\mathbf{Z}$ greater than N, it permits for more exploration to occur prior to population evolution, and it is the explanation for the previously observed 77% increase in the success rate for the profiled adaptive sampler DE-MC$_z$ relative to DE-MC when comparing the results from experiments 1 and 14. The previously mentioned strengths of the profiled adaptive sampler DE-MC$_z$ also permit the use of fewer chains relative to its native form DE-MC (ter Braak and Vrugt 2008). Its success rate increased from 87% to 100%, in experiment 16, when the initial row size was doubled, to 40, and the population size was decreased by a factor of 2, to 2, equal in value to the dimensionality of the problem. The results obtained in experiment 16 and also in the final experiment, which also employed a population size of 2, emphasize that the initial population alone must not necessarily contain all of the modes, as long as they are also seen in $\mathbf{Z}$, which is a basis for jump proposal generation.

MCMC simulation can be computationally expensive when it is measured in terms of the total number of forward model calls that are required before one can be confident that sampling is occurring with stable frequency from the posterior distribution. It is not uncommon for one or more chains to

become and remain entrapped for prolonged periods, resulting in reduced sampler efficiency. Moreover and related, another real and potentially problematical situation involves full characterization of multimodal posteriors, particularly when the modes are distinct and far apart, as is the case with the bimodal mixed normal target of Equation 3. Vrugt et al. (2008a) introduced the outlier chain removal and replacement functionality as a means by which to improve MCMC sampler efficiency. A secondary intent with this case study was to clearly communicate the knowledge gained through experience with application of the outlier chain removal and replacement functionality. This known bimodal normal mixture target definition of Equation 3 was informed based on the experience with MCMC supervised hydrologic model calibration. It is clear, based on the results from the first 13 experiments, that the outlier chain removal and replacement functionality degrades sampler reliability, particularly when the population is initialized intelligently rather than arbitrarily and the posterior is multimodal with modes well separated from each other. Moreover, the results further suggest that the functionality is tunable; hence, its use introduces an additional input parameter that one must adjust to optimize sampler performance for a given problem. Its use effectively homogenizes the population, and if it is not tuned properly for the given problem, it can negatively impact sampler performance.

The experiments performed with the bimodal normal mixture target of Equation 3 demonstrated that intelligent sampler initialization has the potential to increase reliability with respect to a population-based MCMC sampler's capacity to fully characterize the equilibrium distribution. Later case studies will investigate the impact of intelligent initialization on sampler efficiency.

In this example, a previously documented challenging multimodal normal mixture problem was revisited but made even more difficult by reducing the standard deviation by one order of magnitude. The capacities of the population-based MCMC samplers DE-MC and its adaption DE-MC$_z$ to fully characterize the known target after a predetermined rather exhaustive burn-in period were measured through a series of experiments, each consisting of 30 random trials. The experiments involved different sampler configurations, primarily focusing on population size, the initial row space size of the matrix $\mathbf{Z}$ for the sampler DE-MC$_z$, and the means by which $\mathbf{X}$ and $\mathbf{Z}$ were initialized. As with the first example problem, the

impact of using outlier chain removal and replacement was also explored with this 20-component normal mixture target.

The MCMC sampler DE-MC was unable to successfully characterize the known multimodal target a single time, across all 30 trials, when its population size was specified to be equal in value to 10 times the dimension of the problem and the population was initialized by using LHS from the box [0,10] X [0,10]. When a single random draw from each component of the target was used to initialize the population of size 20, the sampler DE-MC was able to successfully characterize the distribution for 11 of the 30 random trials. By contrast, its adaption, the sampler DE-MC$_z$, using the same population size and initialization strategies, was able to fully characterize the known multimodal target with 100% success rates across all trials. When LHS from the box [0,10] X [0,10] was used to initialize the matrix $\mathbf{Z}$, a value for M_o equal in value to 1,500 was required to ensure the 100% success rate. Lower values specified for M_o yielded lower success rates across all 30 trials. For example, specified values for M_o equal to 200, 400, and 1,000 yielded 50%, 80%, and 97% success rates, respectively. When the matrix $\mathbf{Z}$ was initialized using LHS from the box [0,10] X [0,10] and then a single random draw from each component of the distribution was used to overwrite the first N equal to 20 rows of $\mathbf{Z}$, and initialize $\mathbf{X}$, then a value for M_o specified equal to 200 was sufficient to ensure the 100% success rate. Moreover, the sampler DE-MC$_z$ was able to successfully characterize the known multimodal target across all trials using a population size of two, either with M_o equal to 3,000 when $\mathbf{Z}$ was initialized using LHS from the box [0,10] X [0,10], or 80 when the matrix $\mathbf{Z}$ was populated with two copies of two random draws from each mode of the target distribution.

The MCMC sampler DE-MC did exhibit improved success with increasing population size. With a population of size 40 initialized using two random draws from each component of the target, the number of successes further increased to 18 out of 30. However, for this problem, the impact of intelligent population initialization was less dramatic for the DE-MC sampler than it was with the first example. In the first example, two random draws from each of the two modes yielded a 100% success rate across the 30 trials; whereas, using a similar initialization strategy with the 20-component normal mixture target, the sampler DE-MC failed to fully characterize the distribution for 12 of the 30 trials.

The success rate of the adaptive sampler DE-MC$_z$ with the population size fixed, consistently increased as the initial row space size for **Z**, initialized using LHS from the box [0,10] X [0,10], was increased. It was also observed for the sampler DE-MC$_z$ that when the population size is reduced, and the matrix **Z** is initialized using LHS from the box [0,10] X [0,10], the value for the initial row space size for **Z** must be increased to maintain an equivalent success rate. For the sampler DE-MC$_z$, the experiments also demonstrated that it was possible to successfully characterize the multimodal distribution across all trials while using a population size equal in value to the dimension of the problem, as long as the matrix **Z** was either initialized with information about the posterior modes of the distribution itself or by simply sampling the entire box with sufficient density prior to evolution.

Regardless of whether the sampler DE-MC or its adaption DE-MC$_z$ was used, there was no observed consistency across those trials which failed to fully characterize the known distribution. In particular, the one or many components missed were not consistent from trial to trial within a given experiment or across the set of experiments. For an equivalent population size, the DE-MC sampler possessed a higher failure rate. Its adaption, DE-MC$_z$, was more successful by virtue of its capacity not only to specify an initial row space size for **Z** greater than *N*, but also in that its jump proposals are a function of **Z**, which grows as the population evolves.

The results associated with this multimodal example further illustrated the problematic nature of the outlier chain removal and replacement functionality. With this 20-component normal mixture problem, it was demonstrated that use of the outlier chain removal and replacement functionality clearly decreased the reliability of the MCMC sampler DE-MC to fully characterize the target distribution. Moreover, the experiments demonstrated that outlier chain removal and replacement is tunable; hence, it introduces an additional parameter that one must properly configure to optimize sampler performance for a given problem. For this problem, when the outlier detection level was specified to be consistent with a previously reported suggested value, the DE-MC sampler exhibited reduced reliability relative to simply not using the functionality.

The results of the numerical experiments indicate that for this challenging multimodal problem a population size of two can be specified while using the adaptive sampler DE-MC$_z$ and yield success across all 30 trials, where

success is measured in terms of the ability to characterize the complete distribution upon completion of an arbitrarily specified comprehensive burn-in period of 1,000,000 evolutions. The explorations profiled herein either involved employment of LHS using a large value for the initial row space size for $\mathbf{Z}$ and/or by populating $\mathbf{Z}$ with random draws from the target itself. What has not been measured and compared so far is the efficiency of one particular sampler configuration relative to another. Of additional interest is whether an initialization that involves random draws from the target itself results in increased overall sampler efficiency, measured in terms of the total number of model calls required to achieve the completion of sampler burn-in, when compared with a sampler configuration that is in line with how MCMC is used with hydrologic simulation, viz., LHS of uninformative uniform priors.

By initializing the population to be evolved with information about the target itself, rather than by using LHS of the box [0,10] X [0,10], it was observed that an approximate 37% increase in reliability for the population-based MCMC sampler DE-MC when it was configured with a population of size 20. For the adaptive MCMC sampler DE-MC$_z$, the initial experiments that focused on sampler reliability underscored that DE-MC$_z$ can effectively see all 20 components of the posterior distribution, with a population size as small as two, as long as sufficient effort is directed at filling the matrix $\mathbf{Z}$ prior to MCMC sampling. However, the additional experiments clearly indicated that if one can inject learned intelligence about the target itself into the matrix $\mathbf{Z}$, then one can yield substantially improved sampler efficiencies relative to simply sampling the box [0,10] X [0,10] with a sufficient density to ensure reliability across all of the trials.

Clearly, intelligent initialization reduces the overall computational cost of MCMC supervised sampling while maintaining sampler reliability. The results, particularly those associated with experiments 18 and 27, motivate one to give consideration to employment of an effective and efficient search method prior to MCMC sampling, to find draws from the target distribution to initialize $\mathbf{X}$ and $\mathbf{Z}$, not only to ensure sampler performance to fully characterize the equilibrium distribution but also to investigate if there are efficiency gains to be derived with such an approach relative to the conventional application of MCMC in the hydrologic modeling literature, which is to simply sample from uninformative uniform priors. The results associated with experiments 36 and 37 clearly show that all 20 modes can

be found with a limited computational budget, and that information can be used to initialize MCMC sampling of the posterior distribution.

This study was motivated by the experiences with excessively long burn-in run requirements when using MCMC samplers to reliably calibrate hydrologic models and also by the observation that the development and application of Bayesian samplers for hydrologic simulation has primarily focused on the MCMC proposal distribution. Herein, the focus is on a proof of concept assessment of the potential merits to be derived from a directed rather than arbitrary initialization strategy that attempts to align itself with conventional guidance for application of MCMC. The experimental results from the four profiled case studies suggest that when using population-based MCMC that it pays, in terms of improved reliability and sampler efficiency to complete burn-in, to optimize first.

The experimental results from the four profiled case studies suggest that initializing a MCMC sampler's population with local minima can dramatically increase the chance of converging to the correct distribution within a predetermined computational budget and significantly decrease the computational time until sampler convergence. The first two case studies, sampling from known multimodal posterior distributions, demonstrate that initializing the chains by randomly sampling from the high posterior density regions increases the reliability of the sampler. In the first case study, the reliability increases from 86% to 100% when replacing LHS initialization with initialization from the distribution itself. In the second case study, the reliability increased from 50% to 100%. In the second case study it is shown that sampler initialization using MLSL-N stochastic global optimization results can be equally effective in increasing reliability. Moreover, the MLSL-N initialized sampler runs resulted in a mean 40% reduction in total model calls to complete burn-in relative to simply using LHS (with M_0=3000) to initialize the population.

For both hydrologic model applications, there is a comparison of the total number of model calls that are required for the sampler to complete burn-in when using MLSL-IMFIL for initialization versus that of simply using LHS. For the HYMOD model, the average run length when using MLSL-IMFIL initialization decreased by at least 82% compared to LHS initialization when using the DE-MC$_z$ sampler and by 60% when using the DREAM$_z$ sampler. For the SAC-SMA model using MLSL-IMFIL,

initialization decreased the average run length by 63% for the DE-MC$_z$ sampler and by 47% for the DREAM$_z$ sampler.

These improved efficiencies are notable at first glance but do not tell the complete story. For the 13-parameter SAC-SMA hydrology model application, the reported total number of model calls to complete burn-in includes the predetermined budget prescribed for the initialization; hence, it takes the DE-MC$_z$/DREAM$_z$ samplers, after use of MLSL-IMFIL, approximately 76619/56329 calls to sample with stable frequency from the equilibrium distribution. Therefore, the conclusion is that it takes approximately 338/215% more model calls to optimize simply using the DE-MC$_z$/DREAM$_z$ samplers relative to the proposed blended algorithm approach. With the five-parameter HYMOD model application, it takes approximately 4074/569% more model calls to optimize simply using the DE-MC$_z$/DREAM$_z$ samplers. The conjecture is that the requirements necessary for the proper design of MCMC jump distributions limit their capacities relative to other methods, which have been demonstrated to be efficient and reliable for global optimization.

With the case studies profiled herein, a likelihood formulation consistent with IMFIL for Least Squares was employed, which was used for the local searches within MLSL for initialization of the MCMC sampler. Alternate likelihood formulations would require use of the more general form of IMFIL. A proof-of-concept assessment of an initialization strategy using two samplers with documented use within the hydrologic modeling community was demonstrated. The hypothesis is that comparable results would result with additional applications involving other available population-based samplers.

A well-constructed MCMC sampler will eventually converge to the correct posterior distribution, but the number of model calls until convergence can be prohibitively expensive. The proof-of-concept approach was demonstrated to reduce the length of burn-in by leveraging the merits of a well-documented efficient and reliable global optimization clustering method to initialize a population-based sampler. It requires no modifications to existing MCMC samplers. Other optimization approaches to initialize a sampler for hydrologic simulation are possible; however, the recommendation is the use of methods that share the known merits of MLSL-IMFIL for further investigation. MLSL-IMFIL identifies the high probability density regions of the posterior distribution, and its results

could be used to build and refine a surrogate response surface while sampling. Opportunities also exist to provide broader application of this proof of concept by cohesively blending MLSL-IMFIL with a given sampler into a single algorithm.

References

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. "Large Area Hydrologic Modeling and Assessment—Part 1: Model Development." *J. Am. Water Resour. Assoc.* 34(1): 73–89.

Bates, B. C., and E. P. Campbell. 2001. "A Markov Chain Monte Carlo Scheme for Parameter Estimation and Inference in Conceptual Rainfall Runoff Modeling." *Water Resour Res.* 37(4): 937–948.

Campbell, E. P., D. R. Fox, and B. C. Bates. 1999. "A Bayesian Approach to Parameter Estimation and Pooling in Nonlinear Flood Event Models." *Water Resourc. Res.* 35(1): 211–220.

Doherty, J. 2003. *MICA: Model-Independent Markov Chain Monte Carlo Analysis.* Brisbane, Australia: Watermark Numerical Computing.

Duan, Q. S., S. Sorooshian, and V. K. Gupta.1992. "Effective and Efficient Global Optimization for Conceptual Rainfall Runoff Models." *Water Resour. Res.* 28(4): 1015–1031.

Duan, Q. S, V. K. Gupta, and S. Sorooshian. 1993. "A Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization." *J. Optim. Theory Appl.* 76(3): 501–521.

Engeland, K., and L. Gottschalk. 2002. "Bayesian Estimation of Parameters in a Regional Hydrological Model." *Hydrol Earth Syst Sci.* 6(5): 883–898.

Engeland, K., C. Y. Xu, and L. Gottschalk. 2005. "Assessing Uncertainties in a Conceptual Water Balance Model Using Bayesian Methodology." *Hydrol. Sci. J.* 50(1): 45–63.

Gallagher, M. R., and J. Doherty. 2007. "Parameter Estimation and Uncertainty Analysis for a Watershed Model." *Environ. Model. Software* 22: 1000–1020. doi:10.1016/j.envsoft.2006.06.007

Gelman, A., and D. B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Stat. Sci.* 7(4): 457–472.

Gilmore, P., and C. T. Kelley. 1995, "An Implicit Filtering Algorithm for Optimization of Functions with Many Local Minima." *SIAM J. Optim.* 5: 269–285.

Gupta, H. V., S. Sorooshian, T. S. Hogue, and D. P. Boyle. 2003. "Advances in Automatic Calibration of Watershed Models." *Calibration of Watershed Modes*, Vol. 6. *Water Science and Application.* Edited by Q. Duan, H. Gupta, S. Sorooshian, A. Rousseau, and R. Turcotte. Washington, DC: AGU.

Haario, H., E. Saksman, and J. Tamminen. 2001., "An Adaptive Metropolis Algorithm." *Bernoulli* 7(2): 223–242. doi:10.2307/3318737.

Haario, H., M. Laine, A. Mira, and E. Saksman. 2006., "DRAM: Efficient Adaptive MCMC. *Stat. Comput.* 16(4): 339–354. doi:10.1007/s11222-006-9438-0.

Harmon, Robin, and Peter Challenor. 1997. "A Markov chain Monte Carlo Method for Estimation and Assimilation into Models." *Ecological Modelling* 101(1): 41–59, https://doi.org/10.1016/S0304-3800(97)01947-9

Kanso, A., M.-C. Gromaire, E. Gaume, B. Tassin, and G. Chebbo. 2003. "Bayesian Approach for the Calibration of Models: Application to an Urban Stormwater Pollution Model." *Water Science and Technology* 47(4): 77–84.

Kavetski, D., G. Kuczera, and S. W. Franks. 2006a. "Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 1. Theory." *Water Resour. Res.* 42: W03407. doi:10.1029/2005WR004368

Kavetski, D., G. Kuczera, and S. W. Franks. 2006b. "Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 2. Application." *Water Resour. Res.* 42: W03408. doi:10.1029/2005WR004376

Kavetski, D., G. Kuczera, and S. W. Franks. 2006c. "Calibration of Conceptual Hydrological Models Revisited: 2. Improving Optimization and Analysis." *J. Hydrol.* 320: 187–201.

Kavetski, D., G. Kuczera, and S. W. Franks. 2006d. "Calibration of Conceptual Hydrological Models Revisited: 1. Overcoming Numerical Artefacts." *J. Hydrol.* 320(1-2): 173–186.

Kelley, C. T. 2001. *Implicit Filtering and Nonlinear Least Squares Problems.* CRSC-TR01-17. Raleigh, NC: Center for Research in Scientific Computation at North Carolina State University.

Kelley, C. T. 2011. *Implicit Filtering*. No. 23 in Software Environments and Tools. SIAM, Philadelphia.

Kuczera, G., and E. Parent. 1998. "Monte Carlo Assessment of Parameter Uncertainty in Conceptual Catchment Models: The Metropolis Algorithm." *J. Hydrol.* 211: 69–85.

Kuczera, G., D. Kavetski, B. Renard, and M. Thyer. 2010. "A Limited-Memory Acceleration Strategy for MCMC Sampling in Hierarchical Bayesian Calibration of Hydrological Models." *Water Resour. Res.* 46: W07602. doi:10.1029/2009WR008985

Levenberg, K. 1944. "A Method for the Solution of Certain Problems in Least Squares." *Q. Appl. Math.* 2: 164–168.

Liang, F. M., and W. H. Wong. 2001. "Real-Parameter Evolutionary Monte Carlo with Applications to Bayesian Mixture Models." *Journal of the American Statistical Association* 96: 653–666.

Liu, J. S., F. Liang, and W. H. Wong. 2000. "The Multiple-Try Method and Local Optimization in Metropolis Sampling." *Journal of American Statistical Association* 95: 121–134.

Makowski, David, Daniel Wallach, and Marie Tremblay. 2002. *Using a Bayesian Approach to Parameter Estimation; Comparison of the GLUE and MCMC Methods*. http://dx.doi.org/10.1051/agro:2002007. 22. 10.1051/agro:2002007

Marquardt, D. 1963. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." *SIAM Journal on Applied Mathematics* 11: 431–441. http://dx.doi.org/10.1137/0111030

Marshall, L., D. Nott, and A. Sharma. 2004. "A Comparative Study of Markov Chain Monte Carlo Methods for Conceptual Rainfall-Runoff Modeling." *Water Resour Res.* 40: W02501. doi:10.1029/2003WR002378

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *J Chem Phys* 21: 1087–1092.

Qian, S. S., C. A. Stow, and M. E. Borsuk. 2003. "On Monte Carlo Methods for Bayesian Inference." *Ecol. Modell.* 159(2–3): 269–277. doi:10.1016/S0304-3800(02)00299-5

Regis, Rommel, and Christine Shoemaker. 2007. "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions." *INFORMS Journal on Computing* 19: 497–509. 10.1287/ijoc.1060.0182

Rinnooy Kan, A. H. G., and G. Timmer. 1987a. "Stochastic Global Optimization Methods, Part I: Clustering Methods." *Math. Program.* 39: 27–56.

Rinnooy Kan, A. H. G., and G. Timmer. 1987b. "Stochastic Global Optimization Methods, Part II: Multi Level Methods." *Math. Program.* 39: 57–78.

Skahill, B., J. Baggett, S. Frankenstein, and C. W. Downer. 2009. "More Efficient PEST Compatible Model Independent Model Calibration." *Environmental Modelling & Software* (24): 517–529.

Skahill, B., and J. Doherty. 2006. "Efficient Accommodation of Local Minima in Watershed Model Calibration." *J. Hydrol.* 329: 122–139.

Smith, T. J., and L. A. Marshall. 2008. "Bayesian Methods in Hydrologic Modeling: A Study of Recent Advancements in Markov Chain Monte Carlo Techniques." *Water Resour. Res.* 44: W00B05. doi:10.1029/2007WR006705

ter Braak, C. J. F. 2006. "A Markov Chain Monte Carlo Version of the Genetic Algorithm Differential Evolution: Easy Bayesian Computing for Real Parameter Spaces." *Stat. Comput.* 16(3): 239–249. doi:10.1007/s11222-006-8769-1

ter Braak, C. J. F., and J. A. Vrugt. 2008. *Statistics and Computing* 18: 435–436. https://doi.org/10.1007/s11222-008-9104-9

Törn, Aimo, and Antanas Zilinskas. 1987. *Global Optimization.* Vol. 350 of *Lecture Notes in Computer Science.* Berlin Heidelberg: Springer Verlag.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Soorooshian. 2003. "A Shuffled Complex Evolution Metrolopolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters." *Water Resourc. Res.* 39(8): 1201–1215.

Vrugt, J. A., H. V. Gupta, S. C. Dekker, S. Sorooshian, T. Wagener, and W. Bouten. 2006. "Application of Stochastic Parameter Optimization to the Sacramento Soil Moisture Accounting Model." *Journal of Hydrology* 325: 288–307.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson. 2008a. "Treatment of Input Uncertainty in Hydrologic Modeling: Doing Hydrology Backward with Markov Chain Monte Carlo Simulation." *Water Resources Research* 44: W00B09. doi:10.1029/2007WR006720

Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson. 2008b. "Equifinality of Formal (DREAM) and Informal (GLUE) Bayesian Approaches in Hydrologic Modeling?" *Stochastic Environ. Res. Risk Assess*. 23(7): 1011–1026.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. 2009. "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling." *International Journal of Nonlinear Sciences and Numerical Simulation* 10(3): 271–288.

Xin, L. 2014. *Numerical Methods for Engineering Design and Optimization: Latin Hypercube Sampling (LHS)*. https://users.ece.cmu.edu/~xinli/classes/cmu_18660/Lec25.pdf

Yang, J., P. Reichert, and K. C. Abbaspour. 2007. "Bayesian Uncertainty Analysis in Distributed Hydrologic Modeling: A Case Study in the Thur River Basin (Switzerland)." *Water Resour. Res*. 43:W10401. doi:10.1029/2006WR005497

Acronyms and Abbreviations

DE-MC	differential evolution Markov chain
DRAM	delayed rejection adaptive Metropolis
DREAM	differential evolution adaptive metropolis
GO	global optimization
HYMOD	hydrologic model
IMFIL	implicit filtering
LHS	latin hypercube sampling
MCMC	Markov chain Monte Carlo
MLSL	multi-level single linkage
SAC-SMA	Sacramento Soil Moisture Accounting
SCEM-UA	Shuffled Complex Evolution Metropolis Algorithm

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED *(From - To)*
March 2020	Final Report	

4. TITLE AND SUBTITLE

A Practical Two-Phase Approach to Improve the Reliability and Efficiency of Markov Chain Monte Carlo Directed Hydrologic Model Calibration

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Brian E. Skahill and Jeffrey S. Baggett

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Coastal and Hydraulics Laboratory
US Army Engineer Research and
Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Department of Mathematics
University of Wisconsin – La Crosse
La Crosse, WI 54601

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CHL TR-20-2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
441 G. Street, NW
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)
USACE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Funding Acct Code C10102, CF160J; AMSCO Code 031398

14. ABSTRACT
Markov chain Monte Carlo (MCMC) methods are widely used in hydrology and other fields for posterior inference in a Bayesian framework. A properly constructed MCMC sampler is guaranteed to converge to the correct limiting distribution, but convergence can be very slow. While most research is focused on improving the proposal distribution used to generate trial moves in the Markov chain, this work instead focuses on efficiently finding an initial population for population-based MCMC samplers that will expedite convergence. Four case studies, including two hydrological models, are used to demonstrate that using multi-level single linkage implicit filtering stochastic global optimization to initialize the population both reduces the overall computational cost and significantly increases the chance of finding the correct limiting distribution within the constraint of a fixed computational budget.

15. SUBJECT TERMS
Hydrologic models, Markov processes, Mathematical models, Monte Carlo method, Numerical analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Brian Skahill
Unclassified	Unclassified	Unclassified	SAR	50	**19b. TELEPHONE NUMBER** *(Include area code)* 971-804-0373

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18