UC Irvine
UC Irvine Previously Published Works

Title

Reply to comment by Chu et al. on "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing"

Permalink

https://escholarship.org/uc/item/3r9005vw

Journal

Water Resources Research, 50(3)

ISSN

0043-1397

Authors

Vrugt, Jasper A
Laloy, Eric

Publication Date

2014-03-01

DOI

10.1002/2013wr014425

Peer reviewed

COMMENTARY

Reply to comment by Chu et al. on "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing"

Jasper A. Vrugt[1,2,3] and Eric Laloy[4]

[1]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA, [2]Department of Earth System Science, University of California, Irvine, California, USA, [3]Institute for Biodiversity and Ecosystems Dynamics, University of Amsterdam, Amsterdam, Netherlands, [4]Institute for Environment, Health and Safety, Belgian Nuclear Research Centre, Mol, Belgium

1. Introduction

The comment of *Chu et al.* [2014], hereafter referred to as CYG, raises questions about some of the results presented in our paper (*Laloy and Vrugt* [2012], hereinafter LV, which is not to be confused with Las Vegas, although appropriate concerning the subject of this work on Monte Carlo simulation).

As a preamble, we would like to remark that the work presented in *Chu et al.* [2010] and LV (2012) concern two different fields of study. CYG view the model calibration as an optimization problem, and use common concepts to efficiently find a single realization of the parameter values that minimizes (or maximizes, if appropriate) some user-defined objective function. Our work, on the contrary, uses Bayesian principles coupled with MCMC simulation to derive a distribution of parameter values that honor the observed data. This distribution summarizes parameter and model predictive (simulation) uncertainty, a requirement for probabilistic analysis, operational forecasting, disentangling error sources, and decision making. The maximum a posteriori density (MAP) parameter values derived with MCMC simulation should reside in close vicinity of the "best" solution found with an optimization algorithm, if the exact same data set, prior distribution, and likelihood (objective) function are used. In this reply, we assume that CYG used a correct implementation of the MT-DREAM$_{(ZS)}$ algorithm and similar data set, prior and likelihood function as LV. Otherwise, the comparative analysis is meaningless. We emphasize this for three reasons. First, the results presented herein contradict CYG and are similar to those reported in LV but now with more trials plotted. Second, contributions in physics [*Horowitz et al.*, 2012; *Toyli et al.*, 2012; *Yale et al.*, 2013] and geophysics [*Linde and Vrugt*, 2013; *Laloy et al.*, 2012, *Rosas-Carbajal et al.*, 2014; T. Lochbuehler et al., Summary statistics from training images as model constraints in probabilistic inversion, *Geophysical Journal International*, in review, 2014] demonstrate proper convergence behavior of MT-DREAM$_{(ZS)}$ on complex and high-dimensional targets involving hundreds of parameters. Third, to justify their SP-UCI algorithm the original paper of *Chu et al.* [2010] portrays misleading results of the predecessor of DREAM, called SCEM-UA. Section 4 of this reply will address this latter issue in more detail.

We now respond to the comments of CYG. We use different sections with numbering corresponding to CYG.

2. Computational Time Unit

CYG find the Computational Time Unit (CTU) diagnostic to be a poor indicator of the performance of MT-DREAM$_{(ZS)}$. They suggest using the number of function evaluations or clock time instead. The CPU-time (s) scales linearly with CTU, or CPU = α·CTU, where α (s) denotes the average time it takes to complete a single function (model) evaluation. As α is dependent on the processor speed (hardware), LV purposely reported the CTU values. Note that we neglect the actual run time of MT-DREAM$_{(ZS)}$, in the determination of α, which is appropriate given the intended application of this algorithm to CPU-intensive forward models.

We purposely do not use the number of function evaluations as performance diagnostic. This metric does not properly convey the CPU-time (CTU) of parallel algorithms such as DREAM$_{(ZS)}$ or MT-DREAM$_{(ZS)}$. These
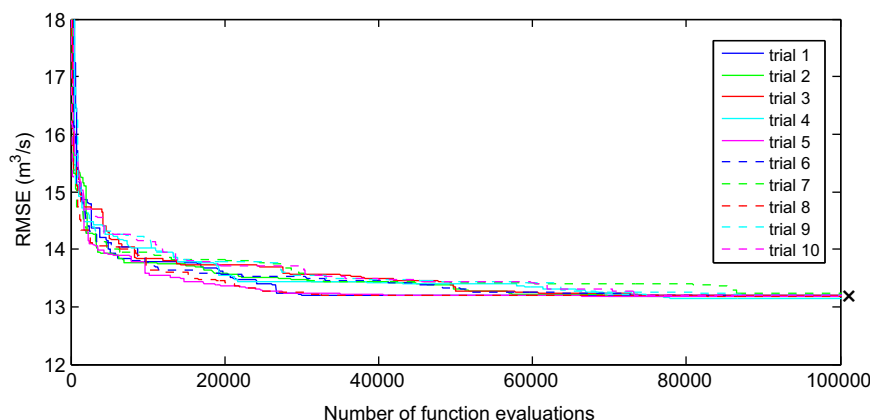
Figure 1. Trace plots of the best RMSE values sampled with MT-DREAM$_{(ZS)}$ for the SAC-SMA model and 2 years of daily discharge data from the Leaf River watershed in Mississippi. A Gaussian likelihood function is used with measurement error invoked jointly with the SAC-SMA model parameters. Each different trial is coded with a different color and/or line style. The "×" symbol at the right hand side plots a RMSE value of 13.2 m^3/s.

algorithms execute N function (model) evaluations in parallel at the same time a sequential method, such as SP-UCI performs a single model evaluation. Actually, the main contribution of LV is to introduce an efficient distributed MCMC algorithm, which achieves a much better computational efficiency than traditional sequential Metropolis (-Hastings) samplers. The authors argue that SP-UCI could be parallelized, yet this will require major modifications to the Nelder-Mead search engine. The reflection, expansion, contraction, and/or reduction (random) step of this Simplex-based heuristic are most effective when executed successively (one after another), and hence parallel implementation of SP-UCI will require important algorithmic concessions. We posit that this will deteriorate search efficiency, and favor use of an alternative method for population evolution. Indeed, distributed computing opens up completely new views on what constitutes an efficient optimization or sampling method. Search methods that exhibit superior search efficiency on a single processor might not be easily amenable to distributed implementation, and hence receive the desired performance.

CYG find the calculation of CTU confusing. We respectfully disagree with this assertion. In DREAM or DREAM$_{(ZS)}$, one generation (update of all of the N chains) simply equals one CTU. In MT-DREAM$_{(ZS)}$, a single update of all the chains requires a proposal and reference step. Thus, two CTUs are required for a single update of all N chains. If, for instance, 10,000 CTUs are required to achieve convergence to a limiting distribution, then the wall-time is approximately similar to 10,000 times the time it takes to complete a single function evaluation, pending the assumption that the model run time is independent of parameterization used. For most finite-element (volume) models, this latter assumption is, of course, not very realistic.

If one continues to focus on the number of function evaluations (as CYG inadvertently do), then one CTU in DREAM or DREAM$_{(ZS)}$ equals N function evaluations. In MT-DREAM$_{(ZS)}$, the proposal and evaluation step each require a different number of function evaluations. This simply follows from multitry Metropolis sampling. If we create k candidate points in each chain, then the proposal step (1 CTU) is equivalent to $N·k$ function evaluations. The reference step (1 CTU) requires $N·(k-1)$ function evaluations because the posterior densities of the current states of the chains are known. Thus, in MT-DREAM$_{(ZS)}$ one CTU amounts to on average, $N·(k-\frac{1}{2})$ function evaluations (see also paragraph 22 in LV). Thus, it is easy to derive one performance metric from the other. Yet, in practice the CTU values (and thus CPU-time) best portray the computational efficiency of a parallel algorithm.

3. Computational Results

We are puzzled why CYG reports such different outcomes than those presented in our paper. The MATLAB code of MT-DREAM$_{(ZS)}$ we shared with them (multiple different versions that differ in user-friendliness) and many others gives the exact same statistics, convergence diagnostics, and trace plots as those presented in LV. For instance, consider Figure 1 that presents the evolution of the best RMSE values sampled with MT-DREAM$_{(ZS)}$ for 10 consecutive trials with the SAC-SMA model (case study 3 in LV). To facilitate comparison
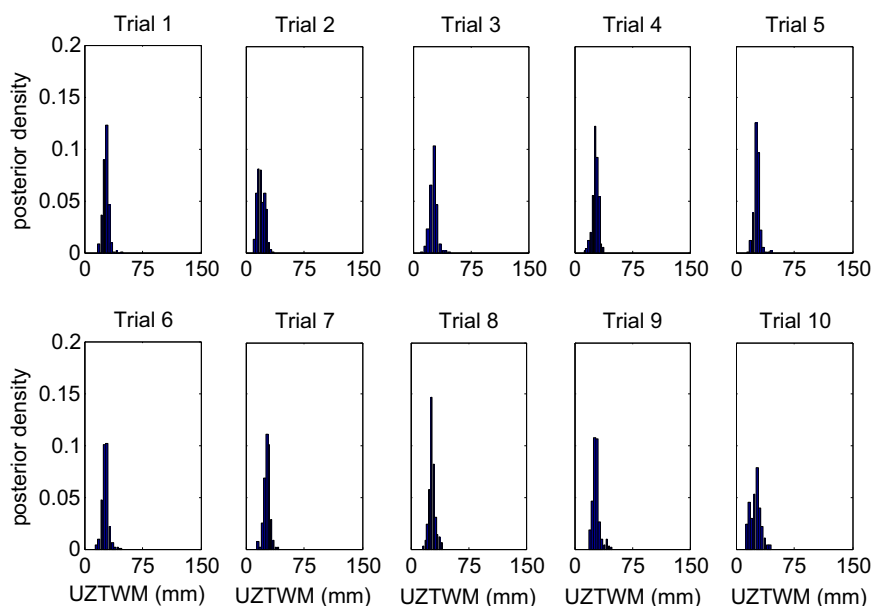
Figure 2. Histograms of the marginal posterior distribution of UZTWM (mm) derived from the last 10% of the sampled Markov chains. Each different subplot displays the outcome of a different trial with MT-DREAM$_{(ZS)}$. The marginal distributions of UZTWM appear very similar and range between 10 and 50 with a MAP value of about 25. Note that the similarity of the marginal distributions would be further enhanced if more posterior samples are used/generated.

with the results of CYG, we plot on the *x* axis the number of function evaluations rather than CTU. Each trial of MT-DREAM$_{(ZS)}$ converges to RMSE values of about 13.2 m^3/s (posterior range of 13.1–13.45) which is substantially lower than values of about 13.8 m^3/s shown in Figure 2c of CYG. We suspect that CYG incorrectly treated the upper bound of the measurement error of the discharge data, a latent variable that LV estimates jointly with the model parameters, and whose MAP value is similar to the posterior mean RMSE value (not shown herein). Note that Figure 2a of CYG confirms the findings of LV that SP-UCI converges prematurely to RMSE values of about 14.0–14.3 m^3/s. The inferences CYG make in their comment about the optimized value of UZTWM are therefore rather meaningless. Indeed, Figure 2 of CYG confirms that SP-UCI converges prematurely, and it is therefore not surprising that some of the calibrated parameter values will deviate from their actual MAP values. The prospects for regionalization would of course be diminished if the parameter values are calibrated toward wrong values.

What is also puzzling are the results plotted in Figure 2c of CYG. Although the final sampled RMSE values with the MT-DREAM$_{(ZS)}$ algorithm are clearly lower than their values of SP-UCI, most of the different trials shown in the trace plot exhibit erratic downward spikes. We have never observed this rather awkward sampling behavior. Yet, it does demonstrate the presence of solutions with much lower RMSE values than those reported in CYG for SP-UCI.

CYG argue that SP-UCI is two orders of magnitude more efficient than MT-DREAM$_{(ZS)}$. This is another puzzling comment. First, how can SP-UCI be more efficient if it converges to the wrong solution? Second, Figure 1 presented herein illustrates that the slowest trial of MT-DREAM$_{(ZS)}$ needs about 14,000 SAC-SMA model evaluations to find RMSE values lower than 14 m^3/s. The worst trial of SP-UCI (pink line in Figure 2 of CYG) terminates prematurely after about 7500 model evaluations with a RMSE of approximately 14.1 m^3/s. We hence fail to understand how CYG conclude that SP-UCI is two orders of magnitude more efficient than MT-DREAM(ZS)? Note that the convergence rate of MT-DREAM$_{(ZS)}$ for the SAC-SMA model is easily enhanced by changing some of the values of the algorithmic parameters. However, settings that promote rapid convergence for relatively simple sampling and/or calibration problems, such as the SAC-SMA model, might not necessarily provide the explorative capabilities needed to solve high-dimensional search problems.

We invite others to verify the numerical results of DREAM, DREAM$_{(ZS)}$, and MT-DREAM$_{(ZS)}$ by downloading the different codes from our website: http://faculty.sites.uci.edu/jasper/sample/. The toolbox contains 17 example problems that have been reported in the hydrologic, vadose zone, (atmospheric) physics,

statistical, and pedometrics literature. These papers demonstrate excellent convergence results on a wide range of parameter dimensionalities and model complexities. Other contributions in physics [*Horowitz et al.*, 2012; *Toyli et al.*, 2012; *Yale et al.*, 2013] and geophysics [*Linde and Vrugt*, 2013; *Laloy et al.*, 2012, *Rosas-Carbajal et al.*, 2014] demonstrate similar findings.

The results of SP-UCI reported in our paper originate from the code that was kindly emailed to us by Dr. Chu in November 2010. We used default values for the algorithmic variables, but did additional numerical testing to maximize SP-UCIs performance. We also added a routine for boundary handling, which is used in the DREAM package to honor the feasible parameter space. Note that our version of SP-UCI provides actually somewhat better results for the SAC-SMA study than those reported in Figure 2a of CYG. One of the 25 trials of SP-UCI reported in LV was able to find RMSE values of around 13.2 m³/s—which is substantially lower than the minimum value of about 14 m³/s presented in CYG.

3.1. SAC-SMA Model (13 Parameters)
The SAC-SMA study considered in LV (2012) was inspired by the study of *Tang et al.* [2006] who demonstrated that SP-UCI's predecessor, the well-known SCE-UA algorithm [*Duan et al.*, 1992] converged inadequately for the 2 year Leaf River data set under consideration.

Again, we are surprised that CYG report a quite different behavior of our algorithms. Figure 1 has demonstrated that MT-DREAM$_{(ZS)}$ converges consistently to RMSE values of 13.2–13.25 m³/s for the SAC-SMA model and 2 year calibration data set. Similar results as in LV are observed for DREAM and DREAM$_{(ZS)}$ and we omit these findings from our reply as they have been discussed extensively in other publications. We did compile a table that summarizes the performance of the DREAM, DREAM$_{(ZS)}$, and MT-DREAM$_{(ZS)}$ algorithms for 25 consecutive trials with the SAC-SMA model. This rather large table can be obtained from the first author upon request along with the software from our website.

We like to suggest that the SAC-SMA convergence problems may in large part be due to poor numerics [*Kavetski et al.*, 2006; *Kavetski and Clark*, 2010; *Schoups et al.*, 2010]. The use of an implicit rather than explicit numerical solver for the partially structured differential and algebraic equation systems of the SAC-SMA model would significantly enhance the smoothness of the response surface and resolve many of the optimization difficulties reported in the literature.

3.2. Groundwater Model Calibration (241 Parameters)
CYG claim that SP-UCI actually exhibits better behavior than reported in our paper. Again, our results are based on the code that was emailed to us by Dr. Chu in 2010. We conducted a few trials, and reported their outcome in Figure 6 of LV. The original analysis in LV and the results presented in the top panel of Figure 3 of CYG, illustrate that SP-UCI converges prematurely. Indeed, the problem of population degeneration is very difficult to resolve, which contradicts statements made in the original paper of *Chu et al.* [2010]. For instance, consider the first sentence of their paper, which reads "*An innovative algorithm, shuffled complexes with principal components analysis (SP-UCI), is developed to overcome a critical deficiency of the shuffled complex evolution scheme: population degeneration.*" In paragraph 8, the authors explicate that this critical deficiency is particularly evident in "*. . .high-dimensional or complex cases.*"

CYG argue that DREAM(ZS) and MT-DREAM(ZS) converge inadequately for the groundwater calibration problem, and present a figure to illustrate this in more detail. Unfortunately, CYG terminated their DREAM$_{(ZS)}$ and MT-DREAM$_{(ZS)}$ trials prematurely using a number of function evaluations (FEs) that is much smaller than originally used in LV (details in next section). Our results are based on CTUs rather than the number of FEs (see our first response). Note that our paper reports the outcome of a single trial. For any 241-dimensional problem, variations are expected in the convergence rate between different trials. Indeed, additional runs demonstrate that MT-DREAM$_{(ZS)}$ needs between 100,000– 500,000 CTUs for one of the Markov chains to find WSSR values of around 200. Thus each of the additional trials converges to the approximate same WSSR values as reported in *Keating et al.* (2010) and derived after several months of calculations with a variety of different search approaches, including singular value decomposition, and manual intervention. The SP-UCI algorithm on the other hand again converges prematurely (see Figure 3 of CYG and Figure 6 of LV), and MT-DREAM$_{(ZS)}$ finds the approximate same WSSR values as reported in *Keating et al.* [2010] and derived from several months of calculations with a variety of different search approaches, including singular value decomposition, and manual intervention.

Our experience suggests that population degeneration is very difficult, if not impossible, to avoid with any optimization method. If the response surface is nonideal (and very high dimensional), optimization methods will often exhibit difficulty to maintain an adequate population diversity, a requirement to find the global optimum. The word adequate is important here. The diversity of the population should carefully balance exploration (ability to traverse the search space; large trial moves) with convergence (ability to refine existing solutions; small trial moves). A too large diversity of the population complicates convergence, whereas insufficient diversity diminishes the explorative capabilities of the algorithm to traverse the search space in pursuit of the global optimum.

Multichain MCMC methods are less prone to population degeneration, a requirement for successful convergence. To explore the target distribution, MCMC methods can accept trial moves with lower density than the current state of the chain(s). This helps to escape from local areas of attraction. Moreover, because of detailed balance requires the chains are forced to always jump from their last position. This promulgates sample diversity, and avoids collapse of the population to a single point.

4. Effectiveness of MT-DREAM$_{(ZS)}$

Determining when Markov chains have converged to a limiting distribution is a very difficult problem that has received a lot of interest in the statistical literature. Various diagnostics have been developed, some of which are reported in Tables 1, 2, and 4 of LV. In practice, visual inspection of the sampled chain trajectories can help to determine when convergence has been achieved. Indeed, Figure 5 of LV demonstrates that the chains sample the approximate same WSSR values, and mix relatively well. This, along with trace plots of the sampled parameters, inspires confidence that the MT-DREAM$_{(ZS)}$ algorithm has converged to a limiting distribution after about 1.5 million CTUs (see Figure 6 of LV).

CYG draw conclusions that MT-DREAM$_{(ZS)}$ has not adequately converged for the 241-dimensional groundwater case study. Yet their analysis in Figure 3 has stopped after about 2 million function evaluations, which is roughly equivalent to about 150,000 CTUs (with $N = 3$ chains and $k = 5$ parallel trials). This is substantially smaller than the 300,000 CTUs and 1,500,000 CTUs that are plotted in Figures 5 and 6 of LV! The same holds for Figure 4 in CYG that shows the evolution of the R-statistic. Note that if one limits attention to the first 150,000 CTUs plotted in Figures 5 and 6 of LV, then a very similar pattern of sampled WSSR values emerges as presented in CYG. CYG just terminated their run prematurely.

CYG also argues (Figure 5) that different trials of MT-DREAM$_{(ZS)}$ lead to different posterior distributions of the SAC-SMA parameters. This finding is incorrect. Figure 1 presented herein has clearly demonstrated that all consecutive trials with MT-DREAM$_{(ZS)}$ converge to the approximate same posterior RMSE values of about 13.2 m^3/s. This corresponds to a stable marginal distribution of UZTWM (Figure 2). These findings yet again contradict the findings of CYG. The convergence behavior of MT-DREAM$_{(ZS)}$ has been studied extensively in different case studies reported in the literature, and has been verified numerically by independently participants of our short courses. Again, we invite others to confirm our findings by downloading the respective algorithms from our website.

5. Final Comments

The original paper of *Chu et al.* [2010] misleads the WRR readership by presenting an inappropriate comparison between SP-UCI and SCEM-UA, a predecessor of DREAM. Their paper lists RMSE values of SCEM-UA for the SAC-SMA model and Leaf River watershed of 21.6 m^3/s (Table 1), which is substantially higher than the value of 17.98 (m^3/s) derived with SP-UCI. This comparison with SCEM-UA appears to be an important theme of the paper (Table 1, Figures 7 and 11, and main text), but fails to mention that the SCEM-UA-derived MAP parameters, listed in Table 1 of *Chu et al.* [2010], are derived using a 3 year (1952–1954) calibration data period. This is clearly stated in caption of Table 2 of *Vrugt et al.* [2006] from which these values have been taken. Their SP-UCI counterparts, on the contrary, correspond to an 11 year (1953–1963) calibration data set. Hence, it is no surprise that the results of SP-UCI are superior. If we use the same 11 year data set with SCEM-UA (DREAM), then equivalent MAP values of the RMSE of about 18 m^3/s are found along with a sample of the posterior parameter distribution! Many statements made in *Chu et al.* [2010], for example in the abstract "... *SP-UCI ...retrieves the optimal parameter values with the lowest recorded root-mean-squared*

error of simulated daily runoff against the observation" or paragraph 6 "*. . .Therefore, experiments in this study demonstrate that SCE-UA achieves much better parameter sets than that achieved by SCEM-UA*" are consequently wrong and misleading.

We believe that the problem of interest is not so much to identify the optimum values of some model parameters. The hydrologic community has moved on to quantification of uncertainty, and in particular diagnosis of model inadequacy. This necessitates the use of sampling rather than optimization methods. The SAC-SMA study in LV was simply used to demonstrate the convergence behavior of MT-DREAM$_{(ZS)}$ for a widely studied model and data set. A comparison with SP-UCI was deemed appropriate to benchmark the validity of the MAP values sampled with MT-DREAM$_{(ZS)}$.

Finally, we believe that emphasis should shift to process-based calibration. The very construction of the likelihood function—as a summary variable of the (usually averaged) properties of the error residuals—dilutes and mixes the available information into an index having little remaining correspondence to specific behaviors of the system [*Gupta et al.*, 2008]. This has inspired *Vrugt and Sadegh* [2013] to introduce "likelihood-free" inference as vehicle for diagnostic model evaluation. This class of methods is also referred to as Approximate Bayesian Computation (ABC) and relaxes the need for an explicit likelihood function in favor of one or multiple different summary statistics rooted in environmental theory that together have a more compelling diagnostic power than some aggregated measure of the size of the error residuals.

References

Chu, W., X. Gao, and S. Sorooshian (2010), Improving the shuffled complex evolution scheme for optimization of complex nonlinear hydrological systems: Application to the calibration of the Sacramento soil moisture accounting model, *Water Resour. Res.*, *46*, W09530, doi:10.1029/2010WR009224.

Chu, W., T. Yang, and X. Gao (2014), Comment on "High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing" by Eric Laloy and Jasper A. Vrugt, *Water Resour. Res.*, *50*, doi:10.1002/2012WR013341.

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

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes.*, *22*(18), 3802–3813.

Horowitz, V. R., B. J. Aleman, D. J. Christle, A. N. Cleland, and D. D. Awschalom (2012), Electron spin resonance of nitrogen-vacancy centers in optically trapped nanodiamonds, *Proc. Natl. Acad. Sci. U.S.A*, *109*(34), 13,493–13,497, doi:10.1073/pnas.1211311109.

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*, W10511, doi:10.1029/2009WR008896.

Kavetski, D., G. Kuczera, and S. W. Franks (2006), Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, *J. Hydrol.*, *320*(1-2), 173–186.

Keating, E, J. Doherty, J. A. Vrugt, and Q. Kang (2010), Optimization and uncertainty assessment of strongly non-linear groundwater models with high parameter dimensionality, *Water Resour. Res.*, *46*, W10517, doi:10.1029/2009WR008584.

Laloy, E., and J. A. Vrugt, (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resour. Res.*, *48*, W01526, doi:10.1029/2011WR010608.

Laloy, E., N. Linde, and J. A. Vrugt (2012), Mass conservative three-dimensional water tracer distribution from Markov chain Monte Carlo inversion of time-lapse ground-penetrating radar data, *Water Resour. Res.*, *48*, W07510, doi:10.1029/2011WR011238.

Linde, N., and J.A. Vrugt (2013), Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion, *Vadose Zone J.*, *12*(1), doi:10.2136/vzj2012.0101.

Rosas-Carbajal, M., N. Linde, T. Kalscheuer, and J. A. Vrugt (2014), Two-dimensional probabilistic inversion of plane-wave electromagnetic data: Methodology, model constraints and joint inversion with electrical resistivity data, *Geophys. J. Int.*, *196*(3), 1508–1524, doi:10.1093/gji/ggt482.

Schoups, G., J. A. Vrugt, F. Fenicia, and N. C. van de Giesen (2010), Inaccurate numerical implementation of conceptual hydrologic models corrupts accuracy and efficiency of MCMC simulation, *Water Resour. Res.*, *46*, W10530, doi:10.1029/2009WR008648.

Tang, Y., P. Reed, and T. Wagener (2006), How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sci.*, *10*, 289–307.

Toyli, D. M., D. J. Christle, A. Alkauskas, B. B. Buckley, C. G. van de Walle, and D. D. Awschalom (2012), Measurement and control of single nitrogen-vacancy center spins above 600 K, *Phys. Rev. X*, *2*, 031001, doi:10.1103/PhysRevX.2.031001.

Vrugt, J. A., and M. Sadegh (2013), Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resour. Res.*, *49*, 4335–4345, doi:10.1002/wrcr.20354.

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, *J. Hydrol.*, *325*, 288–307.

Yale, C. G., B. B. Buckley, D. J. Christle, G. Burkard, F. J. Heremans, L. C. Bassett, and D. D. Awschalom (2013), All-optical control of a solid-state spin using coherent dark states, *Proc. Natl. Acad. Sci. U. S. A*, *110*(19), 7595–7600, doi:10.1073/pnas.1305920110.