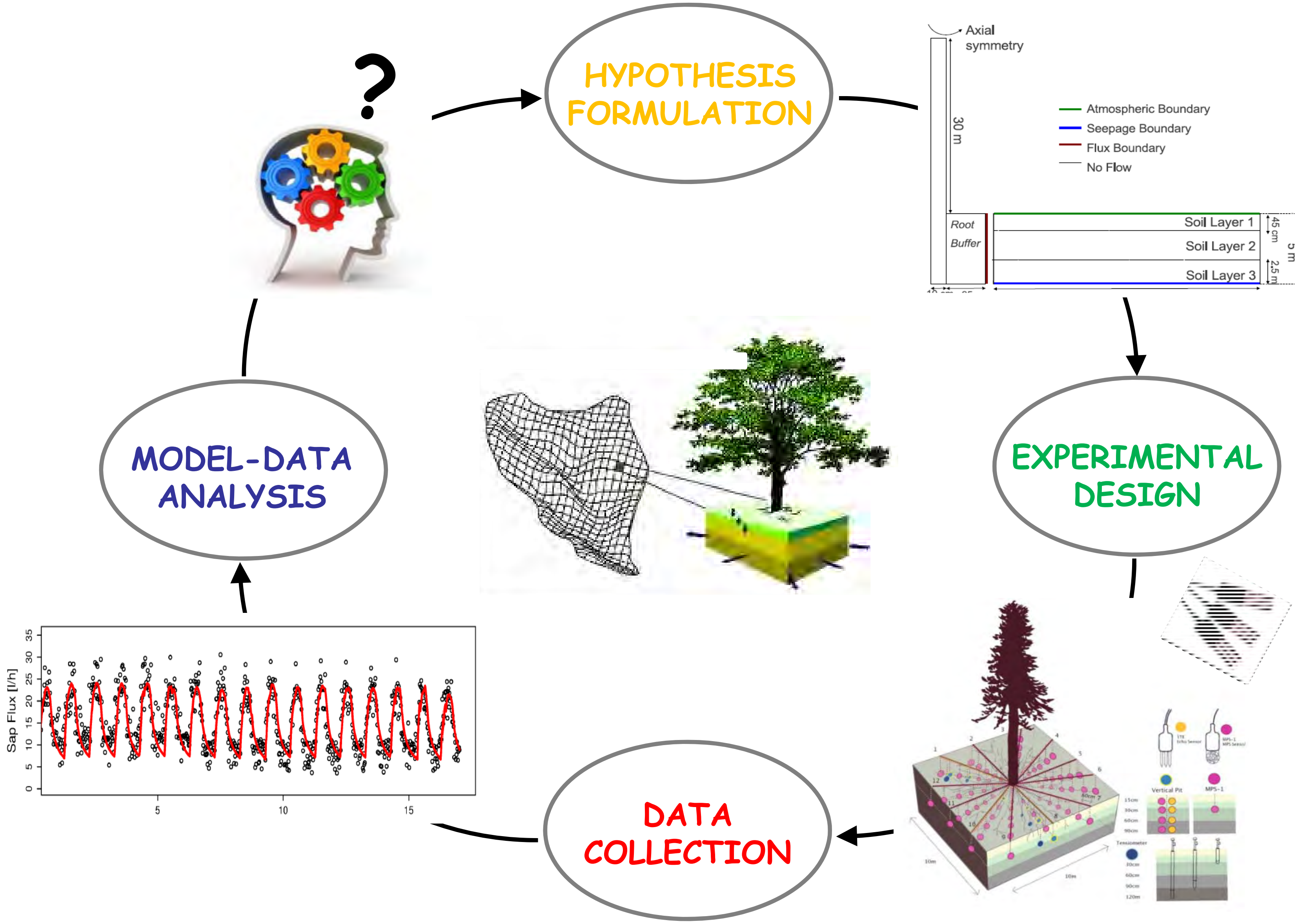
Hydrologic Data Assimilation Using Particle Markov Chain Monte Carlo Simulation

JASPER A. VRUGT

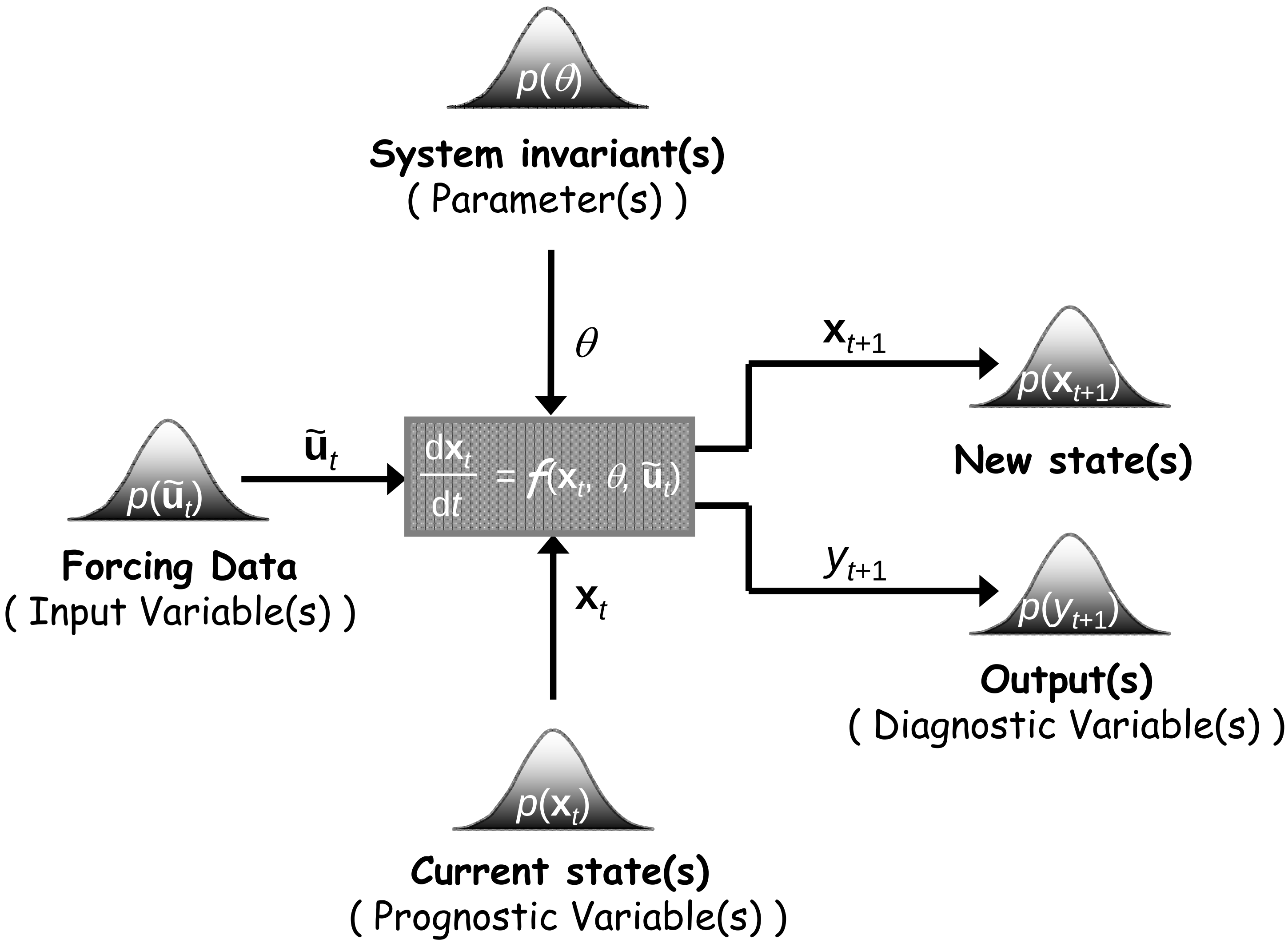
Department of Civil and Environmental Engineering
University of California Irvine, USA

Department of Earth System Science
University of California Irvine, USA

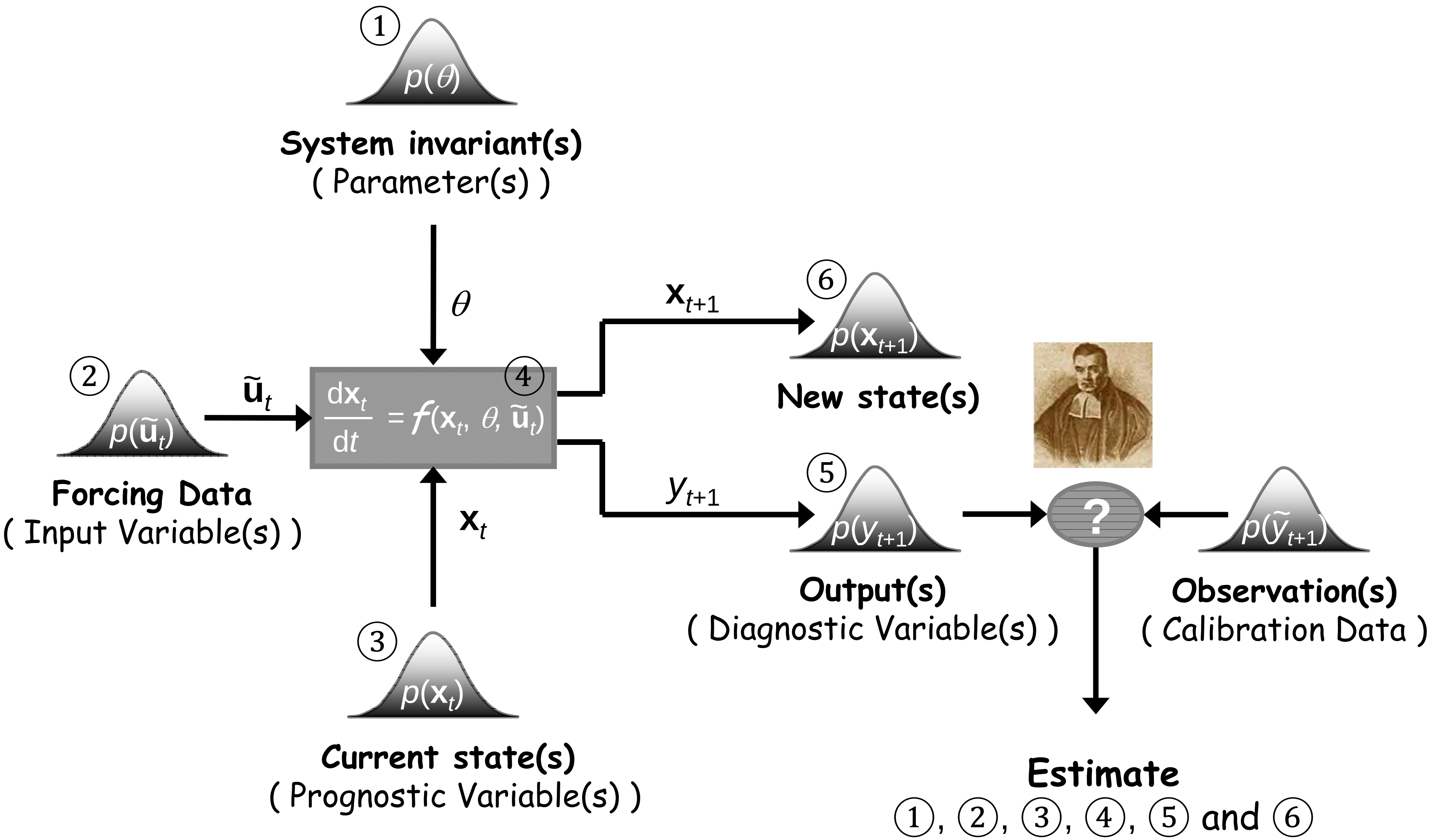
THE ITERATIVE RESEARCH CYCLE



REALITY COMPLEX: UNCERTAINTY



MODEL-DATA FUSION PROBLEM



BAYESIAN ANALYIS

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Thomas Bayes (1763). "**An Essay towards solving a Problem in the Doctrine of Chances**“, *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418.

PRIOR, LIKELIHOOD, EVIDENCE, POSTERIOR

PRIOR CONDITIONAL PROBABILITY (LIKELIHOOD)

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

POSTERIOR EVIDENCE

NOTATION I USE IN THIS PRESENTATION

$$p(\theta|\tilde{\mathbf{y}}_{1:n}) = \frac{p(\theta)p(\tilde{\mathbf{y}}_{1:n}|\theta)}{p(\tilde{\mathbf{y}}_{1:n})}$$

$$p(\tilde{\mathbf{y}}_{1:n}) = \int_{\Theta} p(\theta)p(\tilde{\mathbf{y}}_{1:n}|\theta)\mathrm{d}\theta = \int_{\Theta} p(\theta, \tilde{\mathbf{y}}_{1:n})\mathrm{d}\theta$$

MONTE CARLO SIMULATION AND LIKELIHOOD

$$\underbrace{p(\theta, \tilde{\mathbf{y}}_{1:n})} = p(\theta)L(\theta|\tilde{\mathbf{y}}_{1:n})$$

TYPICALLY CANNOT BE ESTIMATED ANALYTICALLY
THUS, (MARKOV CHAIN) MONTE CARLO SAMPLING

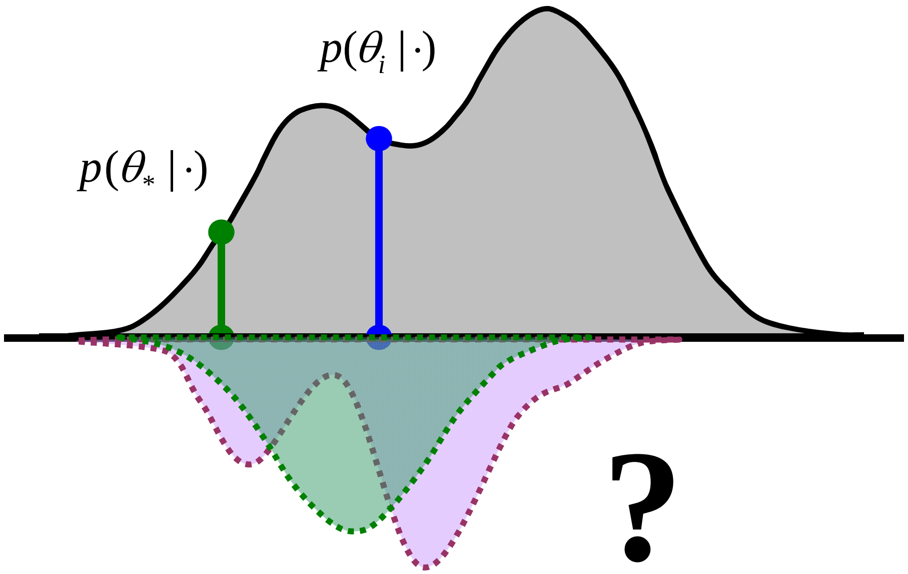
STANDARD GAUSSIAN LIKELIHOOD FUNCTION

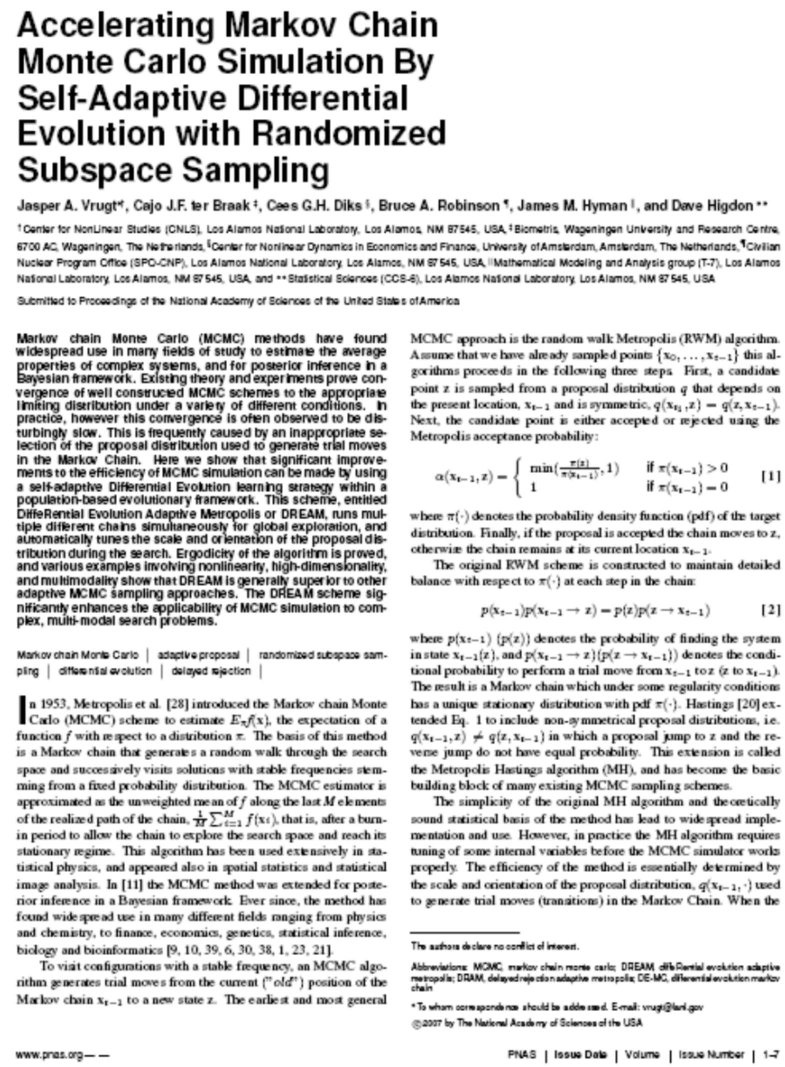
$$L(\theta|\tilde{\mathbf{y}}_{1:n}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_v^2}} \exp\left[-\frac{1}{2}\hat{\sigma}_v^{-2}(\tilde{y}_t - y_t(\theta))^2\right]$$

WHICH IS SIMILAR TO

$$L(\theta|\tilde{\mathbf{y}}_{1:n}) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_v^2}}\right)^n \exp\left[-\frac{1}{2}\hat{\sigma}_v^{-2}F_{\mathrm{SLS}}(\theta),\right]$$

DIFFERENTIAL EVOLUTION ADAPTIVE METROPOLIS





Accelerating Markov Chain Monte Carlo Simulation By Self-Adaptive Differential Evolution with Randomized Subspace Sampling

Jasper A. Vrugt*[†], Cajo J.F. ter Braak [‡], Cees G.H. Diks [§], Bruce A. Robinson [¶], James M. Hyman [‖], and Dave Higdon [**]

[†]Center for NonLinear Studies (CNLS), Los Alamos National Laboratory, Los Alamos, NM 87545, USA,[‡]Biometris, Wageningen University and Research Centre, 6700 AC, Wageningen, The Netherlands,[§]Center for Nonlinear Dynamics in Economics and Finance, University of Amsterdam, Amsterdam, The Netherlands,[¶]Civilian Nuclear Program Office (SPO-CNP), Los Alamos National Laboratory, Los Alamos, NM 87545, USA,[‖]Mathematical Modeling and Analysis group (T-7), Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [**]Statistical Sciences (CCS-6), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Markov chain Monte Carlo (MCMC) methods have found widespread use in many fields of study to estimate the average properties of complex systems, and for posterior inference in a Bayesian framework. Existing theory and experiments prove convergence of well constructed MCMC schemes to the appropriate limiting distribution under a variety of different conditions. In practice, however this convergence is often observed to be disturbingly slow. This is frequently caused by an inappropriate selection of the proposal distribution used to generate trial moves in the Markov Chain. Here we show that significant improvements to the efficiency of MCMC simulation can be made by using a self-adaptive Differential Evolution learning strategy within a population-based evolutionary framework. This scheme, entitled DiffeRential Evolution Adaptive Metropolis or DREAM, runs multiple different chains simultaneously for global exploration, and automatically tunes the scale and orientation of the proposal distribution during the search. Ergodicity of the algorithm is proved, and various examples involving nonlinearity, high-dimensionality, and multimodality show that DREAM is generally superior to other adaptive MCMC sampling approaches. The DREAM scheme significantly enhances the applicability of MCMC simulation to complex, multi-modal search problems.

Markov chain Monte Carlo | adaptive proposal | randomized subspace sampling | differential evolution | delayed rejection |

In 1953, Metropolis et al. [28] introduced the Markov chain Monte Carlo (MCMC) scheme to estimate $E_\pi f(\mathbf{x})$, the expectation of a function f with respect to a distribution π. The basis of this method is a Markov chain that generates a random walk through the search space and successively visits solutions with stable frequencies stemming from a fixed probability distribution. The MCMC estimator is approximated as the unweighted mean of f along the last M elements of the realized path of the chain, $\frac{1}{M}\sum_{t=1}^{M} f(\mathbf{x}_t)$, that is, after a burn-in period to allow the chain to explore the search space and reach its stationary regime. This algorithm has been used extensively in statistical physics, and appeared also in spatial statistics and statistical image analysis. In [11] the MCMC method was extended for posterior inference in a Bayesian framework. Ever since, the method has found widespread use in many different fields ranging from physics and chemistry, to finance, economics, genetics, statistical inference, biology and bioinformatics [9, 10, 39, 6, 30, 38, 1, 23, 21].

To visit configurations with a stable frequency, an MCMC algorithm generates trial moves from the current ("*old*") position of the Markov chain $\mathbf{x}_{t-1}$ to a new state z. The earliest and most general MCMC approach is the random walk Metropolis (RWM) algorithm. Assume that we have already sampled points $\{\mathbf{x}_0,\ldots,\mathbf{x}_{t-1}\}$ this algorithms proceeds in the following three steps. First, a candidate point z is sampled from a proposal distribution q that depends on the present location, $\mathbf{x}_{t-1}$ and is symmetric, $q(\mathbf{x}_{t_1}, z) = q(z, \mathbf{x}_{t-1})$. Next, the candidate point is either accepted or rejected using the Metropolis acceptance probability:

$$\alpha(\mathbf{x}_{t-1}, z) = \begin{cases} \min\left(\frac{\pi(z)}{\pi(\mathbf{x}_{t-1})}, 1\right) & \text{if } \pi(\mathbf{x}_{t-1}) > 0 \\ 1 & \text{if } \pi(\mathbf{x}_{t-1}) = 0 \end{cases} \qquad [1]$$

where $\pi(\cdot)$ denotes the probability density function (pdf) of the target distribution. Finally, if the proposal is accepted the chain moves to z, otherwise the chain remains at its current location $\mathbf{x}_{t-1}$.

The original RWM scheme is constructed to maintain detailed balance with respect to $\pi(\cdot)$ at each step in the chain:

$$p(\mathbf{x}_{t-1})p(\mathbf{x}_{t-1} \rightarrow z) = p(z)p(z \rightarrow \mathbf{x}_{t-1}) \qquad [2]$$

where $p(\mathbf{x}_{t-1})$ ($p(z)$) denotes the probability of finding the system in state $\mathbf{x}_{t-1}(z)$, and $p(\mathbf{x}_{t-1} \rightarrow z)(p(z \rightarrow \mathbf{x}_{t-1}))$ denotes the conditional probability to perform a trial move from $\mathbf{x}_{t-1}$ to z (z to $\mathbf{x}_{t-1}$). The result is a Markov chain which under some regularity conditions has a unique stationary distribution with pdf $\pi(\cdot)$. Hastings [20] extended Eq. 1 to include non-symmetrical proposal distributions, i.e. $q(\mathbf{x}_{t-1}, z) \neq q(z, \mathbf{x}_{t-1})$ in which a proposal jump to z and the reverse jump do not have equal probability. This extension is called the Metropolis Hastings algorithm (MH), and has become the basic building block of many existing MCMC sampling schemes.

The simplicity of the original MH algorithm and theoretically sound statistical basis of the method has lead to widespread implementation and use. However, in practice the MH algorithm requires tuning of some internal variables before the MCMC simulator works properly. The efficiency of the method is essentially determined by the scale and orientation of the proposal distribution, $q(\mathbf{x}_{t-1}, \cdot)$ used to generate trial moves (transitions) in the Markov Chain. When the

The authors declare no conflict of interest.

Abbreviations: MCMC, markov chain monte carlo; DREAM, diffeRential evolution adaptive metropolis; DRAM, delayed rejection adaptive metropolis; DE-MC, differential evolution markov chain

*To whom correspondence should be addressed. E-mail: vrugt@lanl.gov

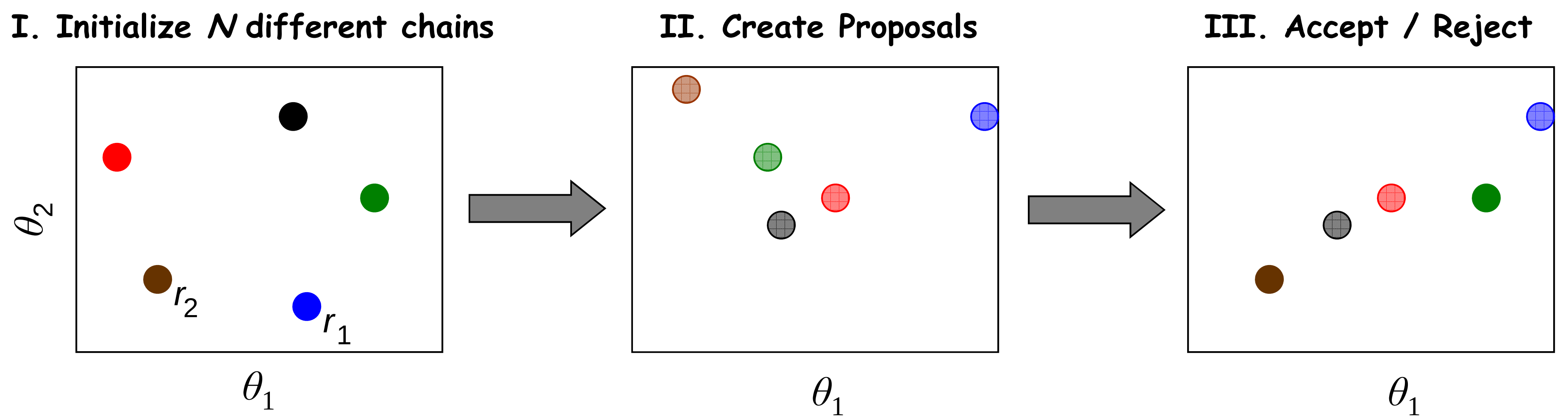
DREAM: Continuously Updates the Scale and Orientation of the Proposal Distribution

Maintains Detailed Balance and is Ergodic
Handles Multimodality Efficiently
High-dimensionality

ESPECIALLY DESIGNED FOR PARALLEL COMPUTING

Vrugt et al., *WRR*, (2008); Vrugt et al., *IJNSNS*, (2009)

DIFFERENTIAL EVOLUTION ADAPTIVE METROPOLIS



1. Initialize *N* different Markov Chains; $N \geq \frac{1}{2}n$
2. Create proposals in each chain *i = 1,…,N* according to:

$$\mathbf{z}^i = \theta^i_{t-1} + \gamma(\theta^{r_1}_{t-1} - \theta^{r_2}_{t-1}) + \mathbf{e}, \quad r_1 \neq r_2 \neq i \quad ; \quad \gamma = 2.4/\sqrt{2n}$$

3. Compute the Metropolis ratio, α^i in each chain *i = 1,…,N*:

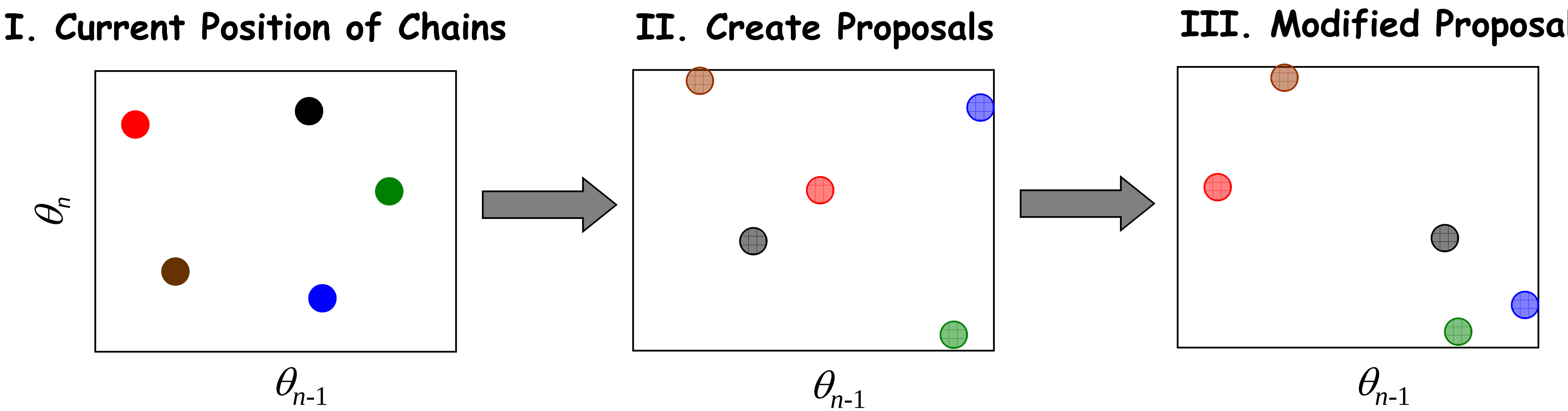
$$\alpha^i = \min(\pi(\mathbf{z}^i)/\pi(\theta^i_{t-1}),1)$$

4. If $\alpha^i \geq U[0,1]$ set $\theta^i_t = \mathbf{z}^i$ otherwise remain at current point, $\theta^i_t = \theta^i_{t-1}$

r_1 and r_2 are chains, e is drawn from uniform distribution with small support
The choice of γ should result in an acceptance probability of about 0.24

SUBSPACE (METROPOLIS-WITHIN-GIBBS) SAMPLING

IN HIGH-DIMENSIONS NOT OPTIMAL TO UPDATE ALL DIMENSIONS



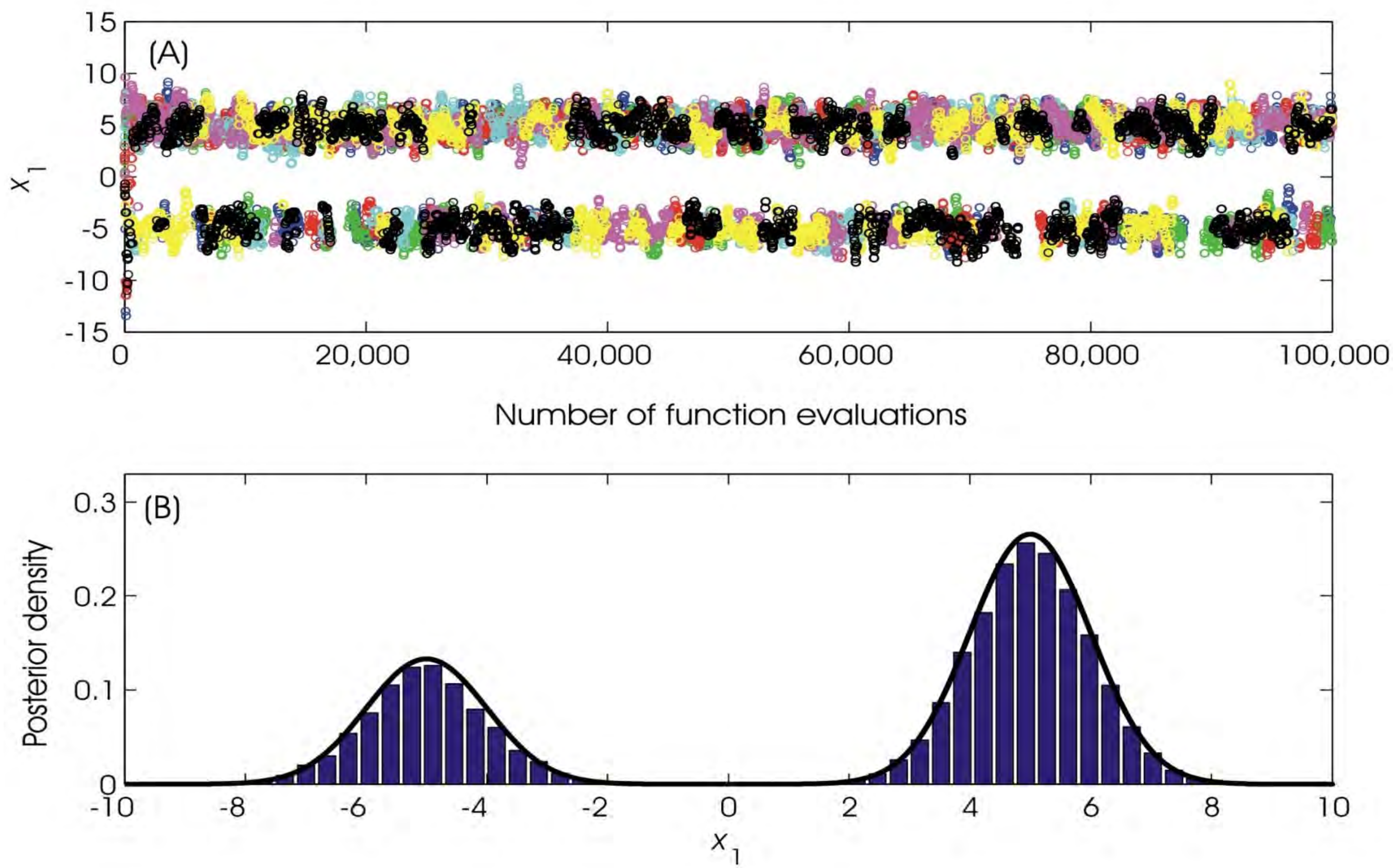
2. Create proposals in each chain *i = 1,…,N* according to:

$$\theta_*^i = \theta_{t-1}^i + \gamma(\theta_{t-1}^{r1} - \theta_{t-1}^{r2}) + \mathbf{e}, \quad r1 \neq r2 \neq i \quad ; \quad \gamma = 2.4/\sqrt{2n}$$

3. Modify only selected dimensions with crossover probability CR

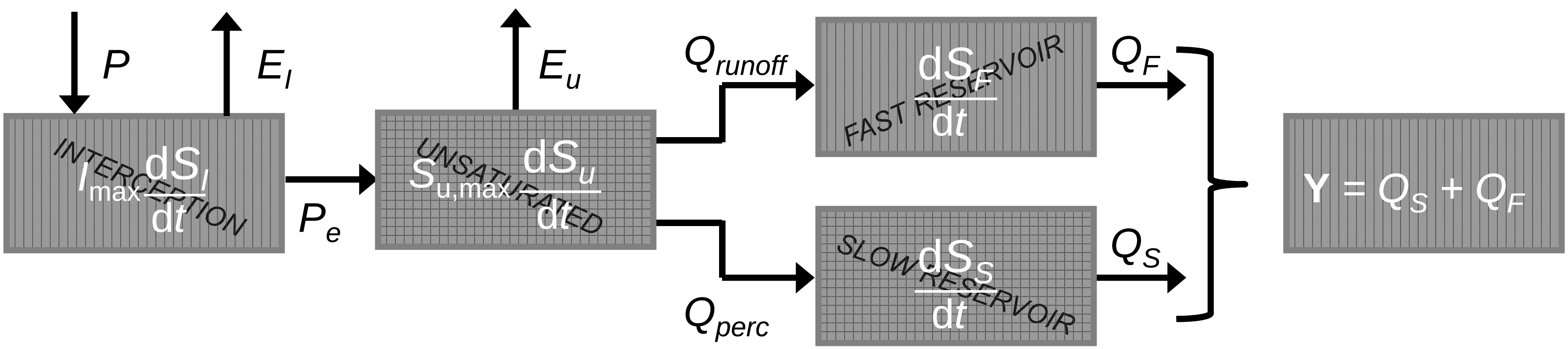
$$\theta_{j,*}^i = \begin{cases} \theta_j^i & if \;\; U \leq 1 - CR \\ \theta_{j,*}^i & \text{otherwise} \end{cases} \qquad j = 1,\ldots,d$$

EXAMPLE: 10-DIMENSIONAL BIMODALITY



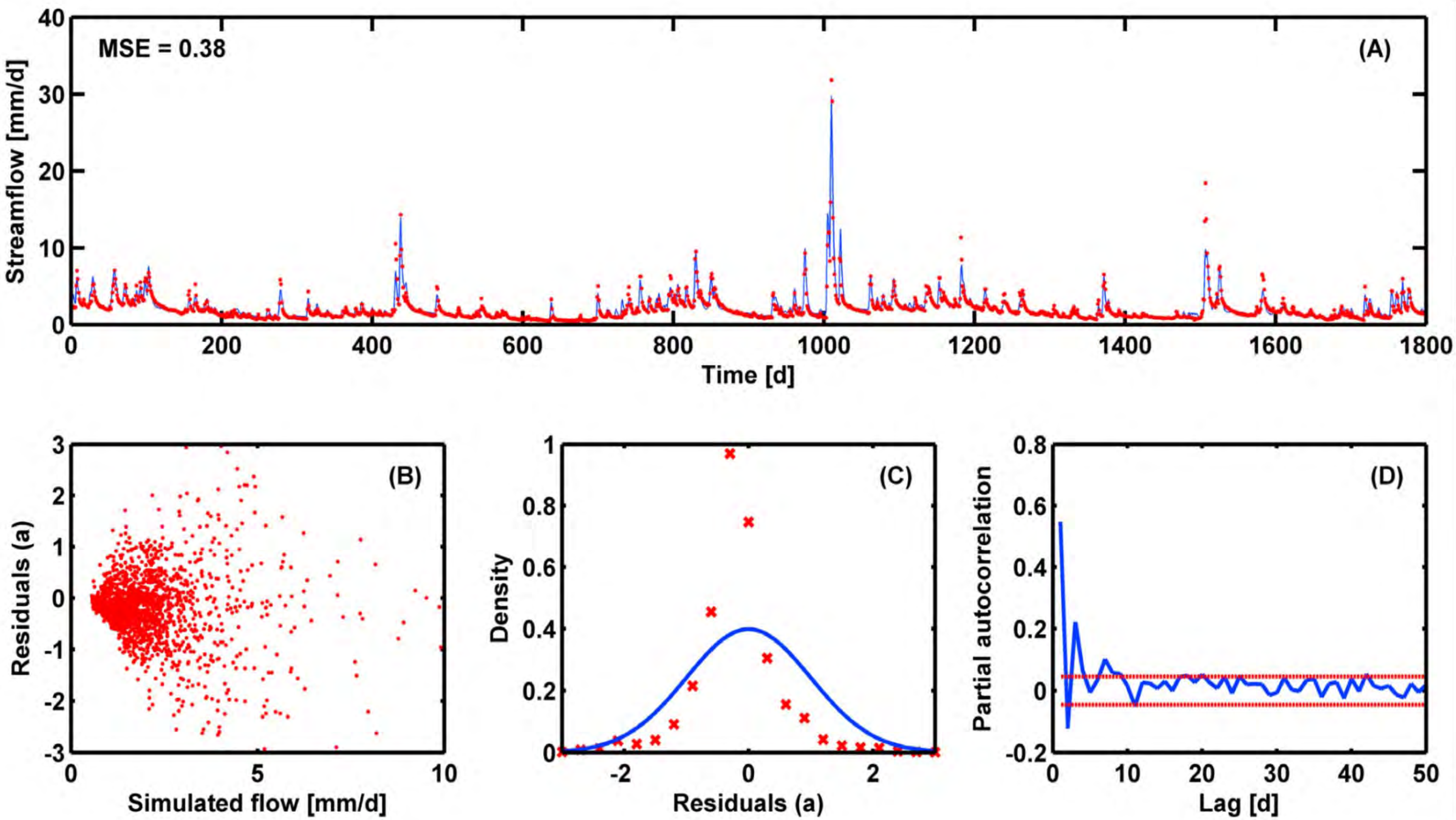
APPLICATION TO WATERSHED HYDROLOGY

Parameter	Symbol	Minimum	Maximum	Units
Maximum interception	$I_{\max}$	1	10	mm
Soil water storage capacity	$S_{\max}$	10	1000	mm
Maximum percolation rate	$Q_{\max}$	0	100	mm/d
Evaporation parameter	α_E	0	100	-
Runoff parameter	α_F	-10	10	-
Time constant, fast reservoir	K_F	0	10	days
Time constant, slow reservoir	K_S	0	150	days



Schoups and Vrugt, *WRR*, 2010

WATERSHED MODEL CALIBRATION



Schoups and Vrugt, *WRR* (2010)

GENERALIZED LIKELIHOOD FUNCTION

A LIKELIHOOD FUNCTION THAT TAKES BETTER CONSIDERATION OF NONTRADITIONAL ERROR RESIDUAL DISTRIBUTIONS

⇓

FOR PROPER TREATMENT OF UNCERTAINTY; A BETTER ALTERNATIVE TO LEAST SQUARES MODEL – DATA SYNTHESIS

WATER RESOURCES RESEARCH, VOL. 46, W10531, doi:10.1029/2009WR008933, 2010

A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors

Gerrit Schoups[1] and Jasper A. Vrugt[2,3,4]

Received 26 November 2009; revised 22 May 2010; accepted 10 June 2010; published 20 October 2010.

[1] Estimation of parameter and predictive uncertainty of hydrologic models has traditionally relied on several simplifying assumptions. Residual errors are often assumed to be independent and to be adequately described by a Gaussian probability distribution with a mean of zero and a constant variance. Here we investigate to what extent estimates of parameter and predictive uncertainty are affected when these assumptions are relaxed. A formal generalized likelihood function is presented, which extends the applicability of previously used likelihood functions to situations where residual errors are correlated, heteroscedastic, and non-Gaussian with varying degrees of kurtosis and skewness. The approach focuses on a correct statistical description of the data and the total model residuals, without separating out various error sources. Application to Bayesian uncertainty analysis of a conceptual rainfall-runoff model simultaneously identifies the hydrologic model parameters and the appropriate statistical distribution of the residual errors. When applied to daily rainfall-runoff data from a humid basin we find that (1) residual errors are much better described by a heteroscedastic, first-order, auto-correlated error model with a Laplacian distribution function characterized by heavier tails than a Gaussian distribution; and (2) compared to a standard least-squares approach, proper representation of the statistical distribution of residual errors yields tighter predictive uncertainty bands and different parameter uncertainty estimates that are less sensitive to the particular time period used for inference. Application to daily rainfall-runoff data from a semiarid basin with more significant residual errors and systematic underprediction of peak flows shows that (1) multiplicative bias factors can be used to compensate for some of the largest errors and (2) a skewed error distribution yields improved estimates of predictive uncertainty in this semiarid basin with near-zero flows. We conclude that the presented methodology provides improved estimates of parameter and total prediction uncertainty and should be useful for handling complex residual errors in other hydrologic regression models as well.

Citation: Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, *46*, W10531, doi:10.1029/2009WR008933.

1. Introduction

[2] Assessment of parameter and predictive uncertainty of hydrologic models is an essential part of any hydrologic study. Uncertainty analysis forms the basis for model comparison and selection [*Schoups et al.*, 2008], allows identification of robust water management strategies that take account of prediction uncertainties [*Ajami et al.*, 2008], and provides an impetus for targeted data collection aimed at improving hydrologic predictions and water management [*Feyen and Gorelick*, 2004]. Furthermore, accurate parameter uncertainty estimation is often required for regionalization and extrapolation of hydrologic parameters to ungauged basins [*Vrugt et al.*, 2002; *Zhang et al.*, 2008].

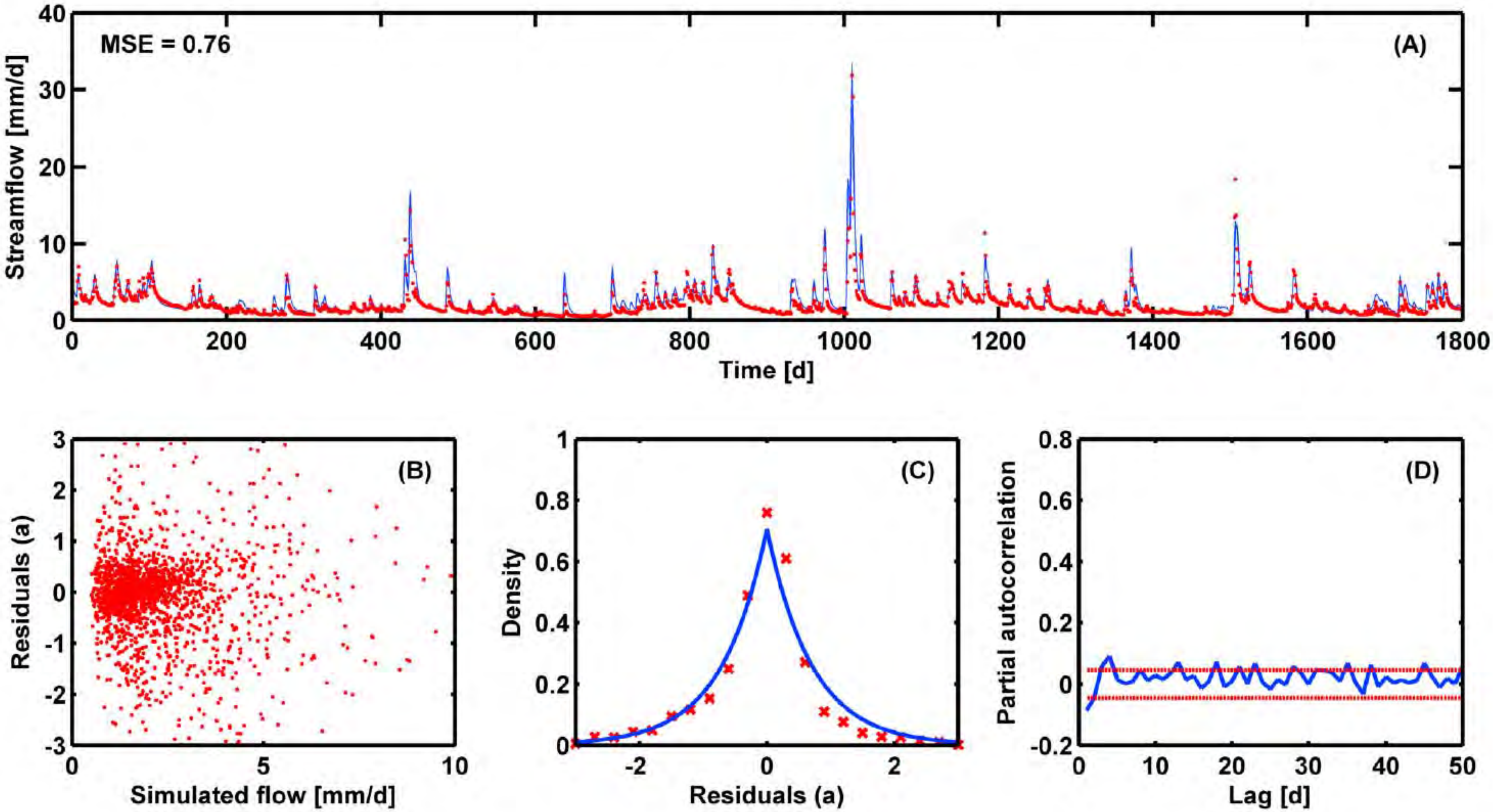
[3] Uncertainty analysis is commonly based on a regression model, whereby observations are represented by the sum of a deterministic component, i.e., the hydrologic model, and a random component describing remaining errors or residuals. These residual errors typically consist of a combination of input, model structural, output, and parameter errors. Model parameter inferences are then based on a likelihood function quantifying the probability that the observed data were generated by a particular parameter set [*Box and Tiao*, 1992]. The mapping from parameter space to likelihood space results in the identification of a range of plausible parameter sets given the data and allows estimation of parameter and predictive uncertainty.

[1]Department of Water Management, Delft University of Technology, Delft, Netherlands.
[2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
[3]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands.
[4]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA.

0043-1397/10/2009WR008933

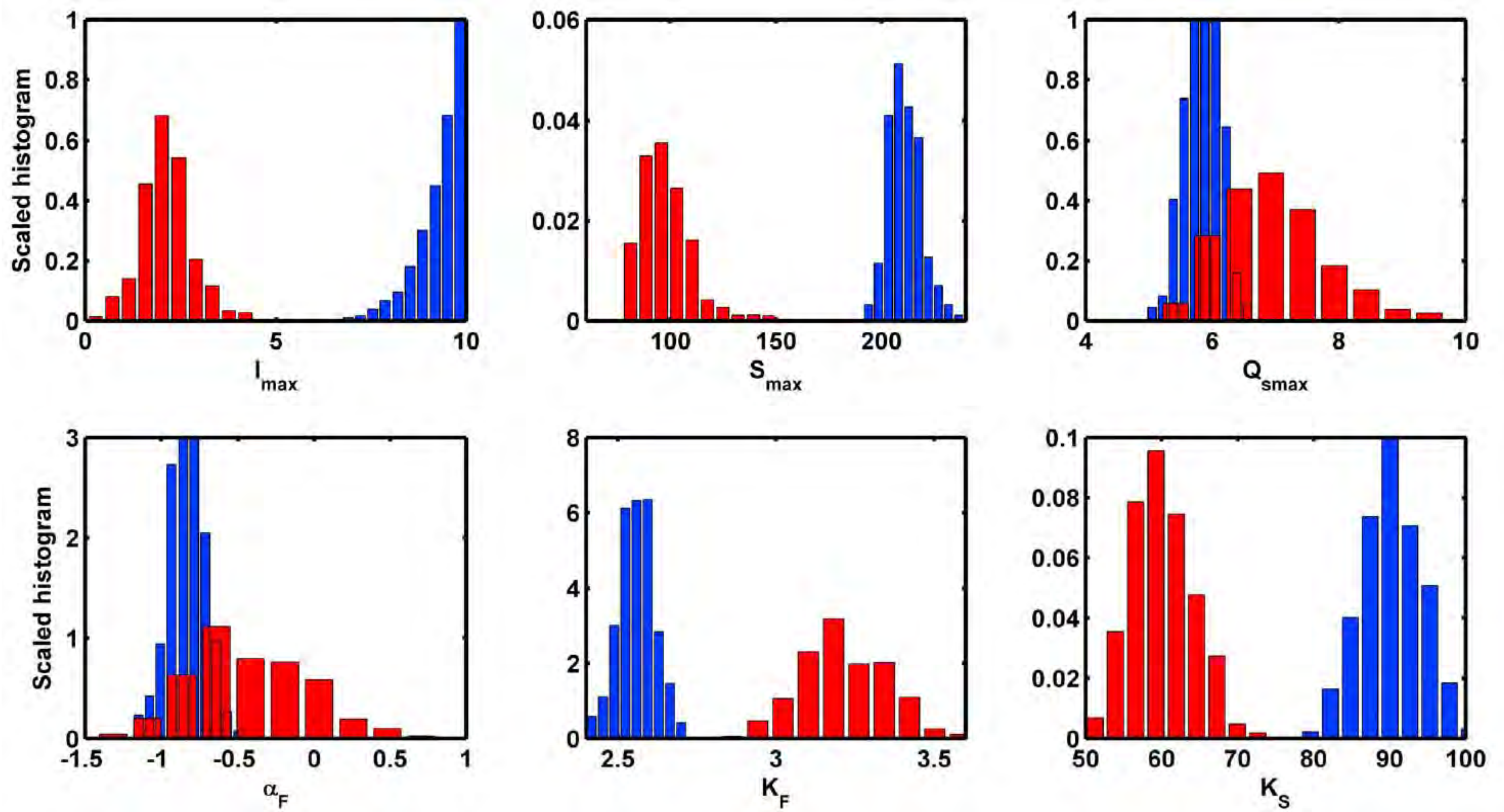
Schoups and Vrugt, *WRR* (2010)

GENERALIZED LIKELIHOOD FUNCTION



Schoups and Vrugt, *WRR* (2010)

POSTERIOR PARAMETER DISTRIBUTIONS

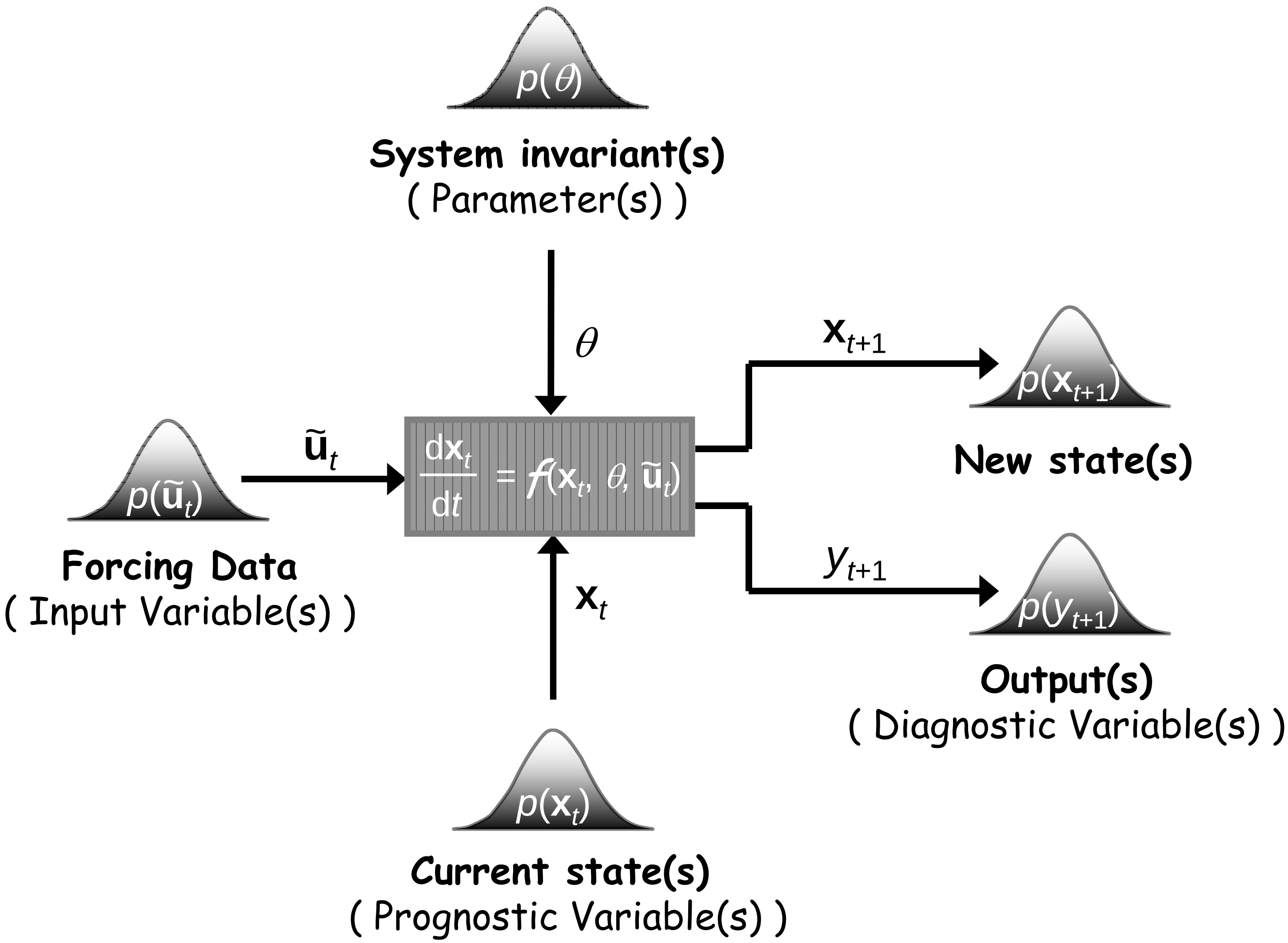


BLUE: STANDARD LEAST SQUARES

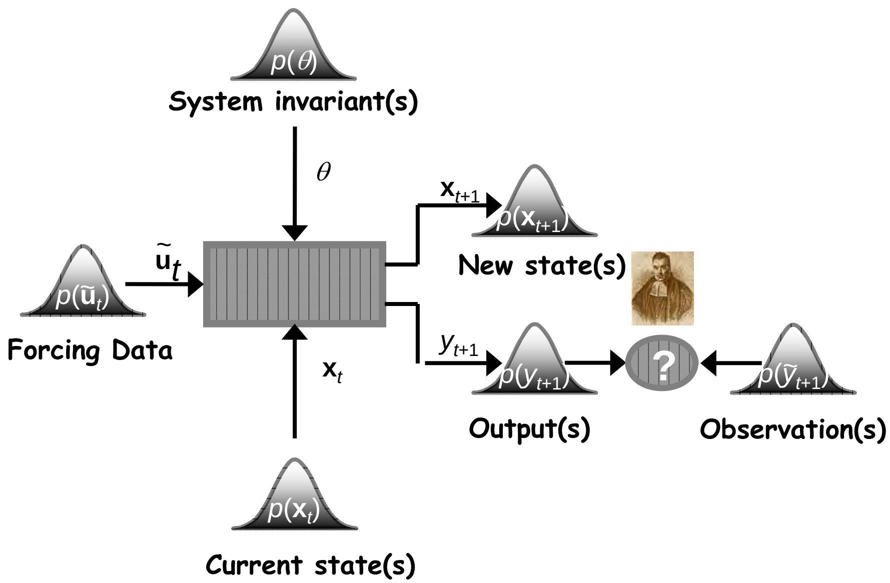
RED: GENERALIZED LIKELIHOOD FUnCTION

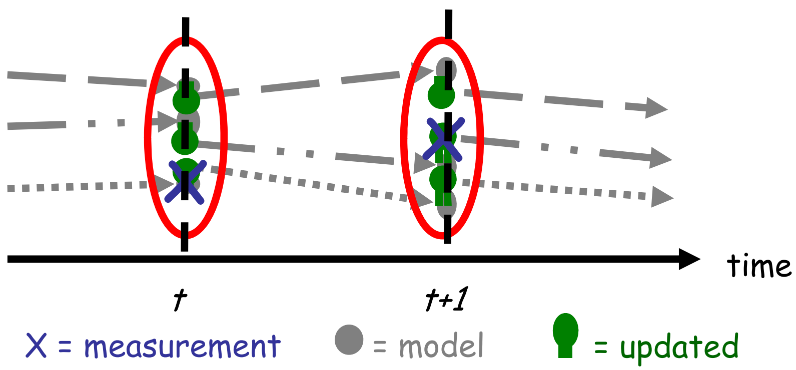
Schoups and Vrugt, *WRR* (2010)

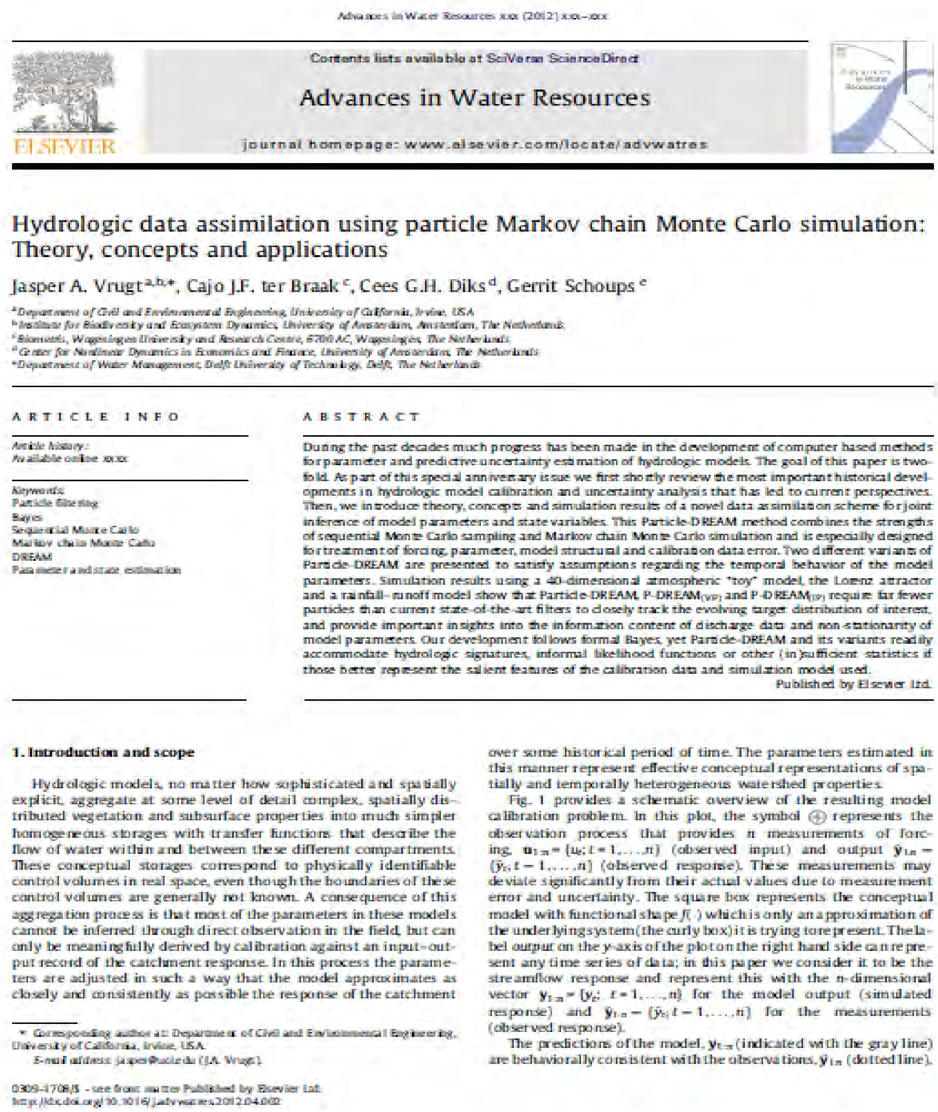
TREATMENT OF STATE UNCERTAINTY



DATA ASSIMILATION USING PARTICLE-DREAM







Advances in Water Resources xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

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

Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications

Jasper A. Vrugt[a,b,*], Cajo J.F. ter Braak[c], Cees G.H. Diks[d], Gerrit Schoups[e]

[a] Department of Civil and Environmental Engineering, University of California, Irvine, USA
[b] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands
[c] Biometris, Wageningen University and Research Centre, 6700 AC, Wageningen, The Netherlands
[d] Center for Nonlinear Dynamics in Economics and Finance, University of Amsterdam, The Netherlands
[e] Department of Water Management, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Particle filtering
Bayes
Sequential Monte Carlo
Markov chain Monte Carlo
DREAM
Parameter and state estimation

ABSTRACT

During the past decades much progress has been made in the development of computer based methods for parameter and predictive uncertainty estimation of hydrologic models. The goal of this paper is two-fold. As part of this special anniversary issue we first shortly review the most important historical developments in hydrologic model calibration and uncertainty analysis that has led to current perspectives. Then, we introduce theory, concepts and simulation results of a novel data assimilation scheme for joint inference of model parameters and state variables. This Particle-DREAM method combines the strengths of sequential Monte Carlo sampling and Markov chain Monte Carlo simulation and is especially designed for treatment of forcing, parameter, model structural and calibration data error. Two different variants of Particle-DREAM are presented to satisfy assumptions regarding the temporal behavior of the model parameters. Simulation results using a 40-dimensional atmospheric "toy" model, the Lorenz attractor and a rainfall–runoff model show that Particle-DREAM, P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ require far fewer particles than current state-of-the-art filters to closely track the evolving target distribution of interest, and provide important insights into the information content of discharge data and non-stationarity of model parameters. Our development follows formal Bayes, yet Particle-DREAM and its variants readily accommodate hydrologic signatures, informal likelihood functions or other (in)sufficient statistics if those better represent the salient features of the calibration data and simulation model used.

Published by Elsevier Ltd.

1. Introduction and scope

Hydrologic models, no matter how sophisticated and spatially explicit, aggregate at some level of detail complex, spatially distributed vegetation and subsurface properties into much simpler homogeneous storages with transfer functions that describe the flow of water within and between these different compartments. These conceptual storages correspond to physically identifiable control volumes in real space, even though the boundaries of these control volumes are generally not known. A consequence of this aggregation process is that most of the parameters in these models cannot be inferred through direct observation in the field, but can only be meaningfully derived by calibration against an input–output record of the catchment response. In this process the parameters are adjusted in such a way that the model approximates as closely and consistently as possible the response of the catchment over some historical period of time. The parameters estimated in this manner represent effective conceptual representations of spatially and temporally heterogeneous watershed properties.

Fig. 1 provides a schematic overview of the resulting model calibration problem. In this plot, the symbol ⊕ represents the observation process that provides n measurements of forcing, $\mathbf{u}_{1:n} = \{u_t; t = 1, \ldots, n\}$ (observed input) and output $\tilde{\mathbf{y}}_{1:n} = \{\tilde{y}_t; t = 1, \ldots, n\}$ (observed response). These measurements may deviate significantly from their actual values due to measurement error and uncertainty. The square box represents the conceptual model with functional shape $f(\cdot)$ which is only an approximation of the underlying system (the curly box) it is trying to represent. The label output on the y-axis of the plot on the right hand side can represent any time series of data; in this paper we consider it to be the streamflow response and represent this with the n-dimensional vector $\mathbf{y}_{1:n} = \{y_t; t = 1, \ldots, n\}$ for the model output (simulated response) and $\tilde{\mathbf{y}}_{1:n} = \{\tilde{y}_t; t = 1, \ldots, n\}$ for the measurements (observed response).

The predictions of the model, $\mathbf{y}_{1:n}$ (indicated with the gray line) are behaviorally consistent with the observations, $\tilde{\mathbf{y}}_{1:n}$ (dotted line).

* Corresponding author at: Department of Civil and Environmental Engineering, University of California, Irvine, USA.
E-mail address: jasper@uci.edu (J.A. Vrugt).

0309-1708/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.advwatres.2012.04.002

PARTICLE-DREAM

Joint Parameter and State Estimation

Vrugt et al., *AWR*, (2012)

SEQUENTIAL BAYES LAW

MODEL FORMULATION (STATE-SPACE)

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \theta, \mathbf{u}_t) + \omega_t, \quad (1)$$

MEASUREMENT OPERATOR

$$\tilde{y}_{t+1} = h(\mathbf{x}_{t+1}, \phi) + v_{t+1}, \quad (2)$$

SEQUENTIAL BAYES LAW (PARAMETERS ASSUMED KNOWN!)

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \overbrace{p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1})}^{\text{prior}} \frac{\overbrace{f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})}^{\text{model}} \quad \overbrace{L_\theta(\tilde{y}_t|\mathbf{x}_t)}^{\text{likelihood function}}}{\underbrace{p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1})}_{\text{normalization constant}}}$$

$$\overbrace{p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t})}^{\text{update step}} = \frac{L_\theta(\tilde{y}_t|\mathbf{x}_t)p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t-1})}{p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1})}$$

$$\overbrace{p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t-1})}^{\text{prediction step}} = \int_\Omega f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})p_\theta(\mathbf{x}_{t-1}|\tilde{\mathbf{y}}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}$$

Doucet and Johansen, 2011; Vrugt et al., *AWR*, (2012)

SEQUENTIAL MONTE CARLO (SMC) METHODS

3.1.1. Theory

Sequential Monte Carlo (SMC) methods sequentially approximate the evolving posterior state distribution, $\{p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})\}_{t>1}$ and corresponding sequence of marginal distributions, $\{p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t})\}_{t>1}$ for any given $\theta \in \Theta$ using a set of P random samples, $\left\{\mathbf{X}_{1:t}^{1:P}\right\}$ also called particles [30]

$$\hat{p}_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \sum_{i=1}^{P} W_t^i \delta_{(\{\mathbf{X}_{1:t}^i\})} \text{ and } \hat{p}_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t}) = \sum_{i=1}^{P} W_t^i \delta_{(\{\mathbf{X}_t^i\})}, \tag{17}$$

$$\tilde{W}_t^i \propto W_{t-1}^i w_t\left(\left\{\mathbf{X}_{1:t}^i\right\}\right), \tag{19}$$

where $w_t\left(\left\{\mathbf{X}_{1:t}^i\right\}\right)$ denote the incremental importance weights [30]:

$$w_t\left(\left\{\mathbf{X}_{1:t}^i\right\}\right) = \frac{f_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right)}{q_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\tilde{y}_t, \left\{\mathbf{X}_{t-1}^i\right\}\right)}. \tag{20}$$

After normalization:

$$W_t^i = \frac{\tilde{W}_t^i}{\sum_{j=1}^{P} \tilde{W}_t^j}, \tag{21}$$

Doucet and Johansen, 2011; Vrugt et al., *AWR*, (2012)

CRUX OF SMC: RESAMPLING

The SMC approach makes use of the following identity

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) \propto p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1}) f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t),$$

RESAMPLING WITH DREAM AT *t*-1

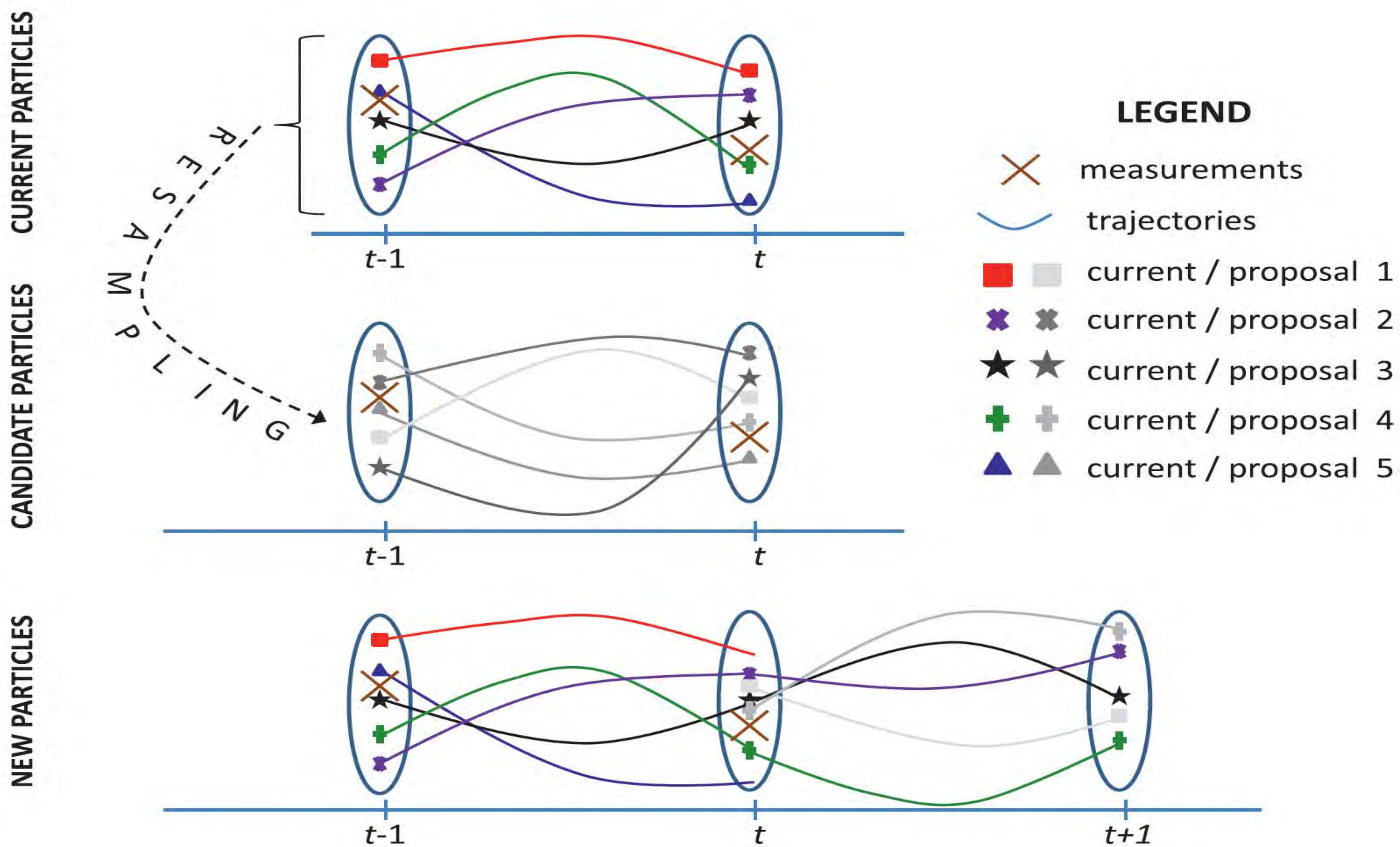
$$\left\{\mathbf{Z}_{t-1}^i\right\} = \left\{\mathbf{X}_{t-1}^i\right\} + (\mathbf{1}_\lambda + \mathbf{e}_\lambda)\gamma(\tau, \lambda')\left[\sum_{j=1}^{\tau}\left\{\mathbf{X}_{t-1}^{r_1(j)}\right\} - \sum_{h=1}^{\tau}\left\{\mathbf{X}_{t-1}^{r_2(h)}\right\}\right] + \boldsymbol{\varepsilon}_\lambda,$$

WITH METROPOLIS ACCEPTANCE PROBABILITY

$$\alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\}, \left\{\mathbf{X}_t^i\right\}\right) = 1 \wedge \frac{f_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{Z}_t^i\right\}\right)}{f_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right)}$$

Vrugt et al., *AWR,* (2012)

RESAMPLING WITH DREAM



Vrugt et al., *AWR,* (2012)

PSEUDO-CODE OF PARTICLE-DREAM

AT TIME $t = 0$

STEP 1: INITIALIZATION

(a) Create $\{\mathbf{X}_0^i\}$ by sampling from the prior state distribution, $p_\theta(\mathbf{x}_0)$.

(b) Assign a uniform (normalized) importance weight, $W_0^i = 1/P$.

(c) Allocate the marginal likelihood, $p_\theta(\cdot) = 1$.

AT TIME $t = 1, \ldots, n$,

STEP 2: PREDICTION

(a) Sample $\{\mathbf{X}_t^i\} \sim f_\theta\left(\cdot \middle| \{\mathbf{X}_{t-1}^i\}\right)$

STEP 3: UPDATE

(a) Calculate the incremental importance weight,

$$w_t\left(\{\mathbf{X}_t^i\}\right) = L_\theta\left(\tilde{y}_t \middle| \{\mathbf{X}_t^i\}\right)$$

(b) Compute the (unnormalized) importance weight, $\tilde{W}_t^i = W_{t-1}^i w_t\left(\{\mathbf{X}_t^i\}\right)$.

(c) Update the marginal likelihood, $p_\theta(\tilde{\mathbf{y}}_{1:t}) = p_\theta(\tilde{\mathbf{y}}_{1:t-1}) \sum_{j=1}^{P} \tilde{W}_t^j$

(d) Normalize the importance weights, $W_t^i = \left(\tilde{W}_t^i\right) / \left(\sum_{j=1}^{P} \tilde{W}_t^j\right)$.

STEP 4: RESAMPLING?

(a) Calculate the effective sample size, $\mathrm{ESS} = \sum_{j=1}^{P} \left(W_t^j\right)^{-2}$.

(b) if $\mathrm{ESS} \leqslant P^*$ go to next step, otherwise return to Step 2.

STEP 4A: RESIDUAL RESAMPLING

(a) Compute $p_{(\{\mathbf{x}_t^i\})} = (PW_t^i - \lfloor PW_t^i \rfloor)/(P - R)$.

(b) Retain those R particles with $\lfloor P\mathbf{W}_t \rfloor \geqslant 1$.

(c) Fill the remaining $P - R$ spots by sampling from $\mathrm{F}\left(\{\mathbf{X}_t^{1:P}\} \middle| p_{(\{\mathbf{x}_t^{1:P}\})}\right)$.

STEP 4B: MCMC RESAMPLING

(a) Create a candidate point, $\{\mathbf{Z}_{t-1}^i\}$ using

$$\{\mathbf{Z}_{t-1}^i\} = \{\mathbf{X}_{t-1}^i\} + (\mathbf{1}_\lambda + \mathbf{e}_\lambda)\gamma(\tau, \lambda') \left[\sum_{j=1}^{\tau} \{\mathbf{X}_{t-1}^{r_1(j)}\} - \sum_{h=1}^{\tau} \{\mathbf{X}_{t-1}^{r_2(h)}\}\right] + \boldsymbol{\varepsilon}_\lambda$$

(b) Compute the Metropolis ratio, $\alpha_\theta\left(\{\mathbf{Z}_t^i\}, \{\mathbf{X}_t^i\}\right)$

$$\alpha_\theta\left(\{\mathbf{Z}_t^i\}, \{\mathbf{X}_t^i\}\right) = 1 \wedge \frac{f_\theta\left(\{\mathbf{Z}_{t-1}^i\} \middle| \{\mathbf{X}_{t-2}^i\}\right) L_\theta\left(\tilde{y}_{t-1} \middle| \{\mathbf{Z}_{t-1}^i\}\right) L_\theta\left(\tilde{y}_t \middle| \{\mathbf{Z}_t^i\}\right)}{f_\theta\left(\{\mathbf{X}_{t-1}^i\} \middle| \{\mathbf{X}_{t-2}^i\}\right) L_\theta\left(\tilde{y}_{t-1} \middle| \{\mathbf{X}_{t-1}^i\}\right) L_\theta\left(\tilde{y}_t \middle| \{\mathbf{X}_t^i\}\right)}$$

(c) Accept the proposal with probability $\alpha_\theta\left(\{\mathbf{Z}_t^i\}, \{\mathbf{X}_t^i\}\right)$.

(d) If accepted, set $\{\mathbf{X}_t^i\} = \{\mathbf{Z}_t^i\}$, otherwise maintain the current particle.

STEP 5: RESET WEIGHTS

(a) Set $W_t^i = 1/P$, (and return to step 2).

BUT WHAT TO DO WITH PARAMETERS?

TWO DIFFERENT POSSIBILITIES

P-DREAM$_{(VP)}$ $\Rightarrow$ STATE AUGMENTATION

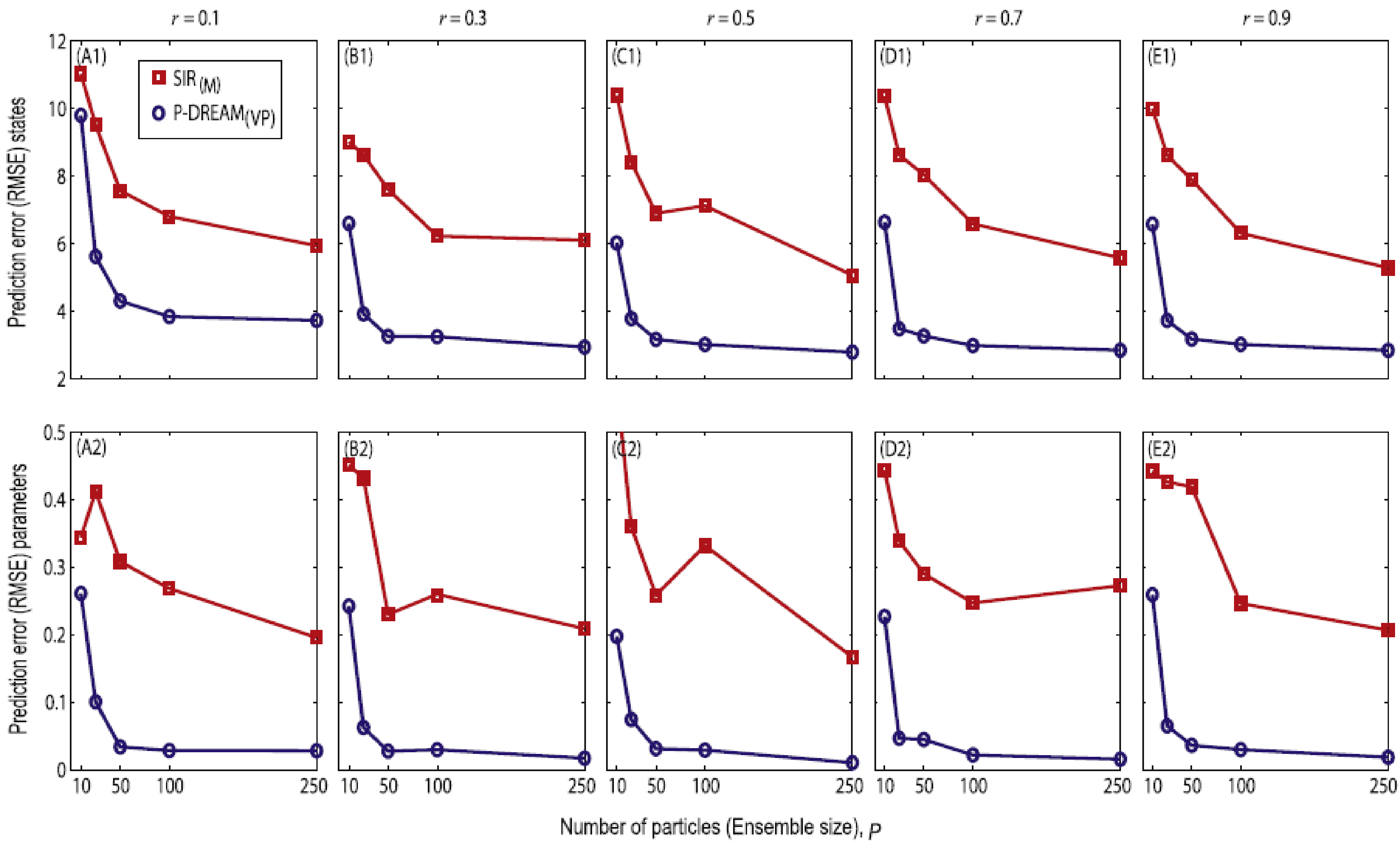
P-DREAM$_{(IP)}$ $\Rightarrow$ OUTSIDE DREAM LOOP

CAUTIONARY NOTE

VARIABLE PARAMETERS (STATE AUGMENTATION) NOT RECOMMENDED!!!!

Vrugt et al., *AWR,* (2012)

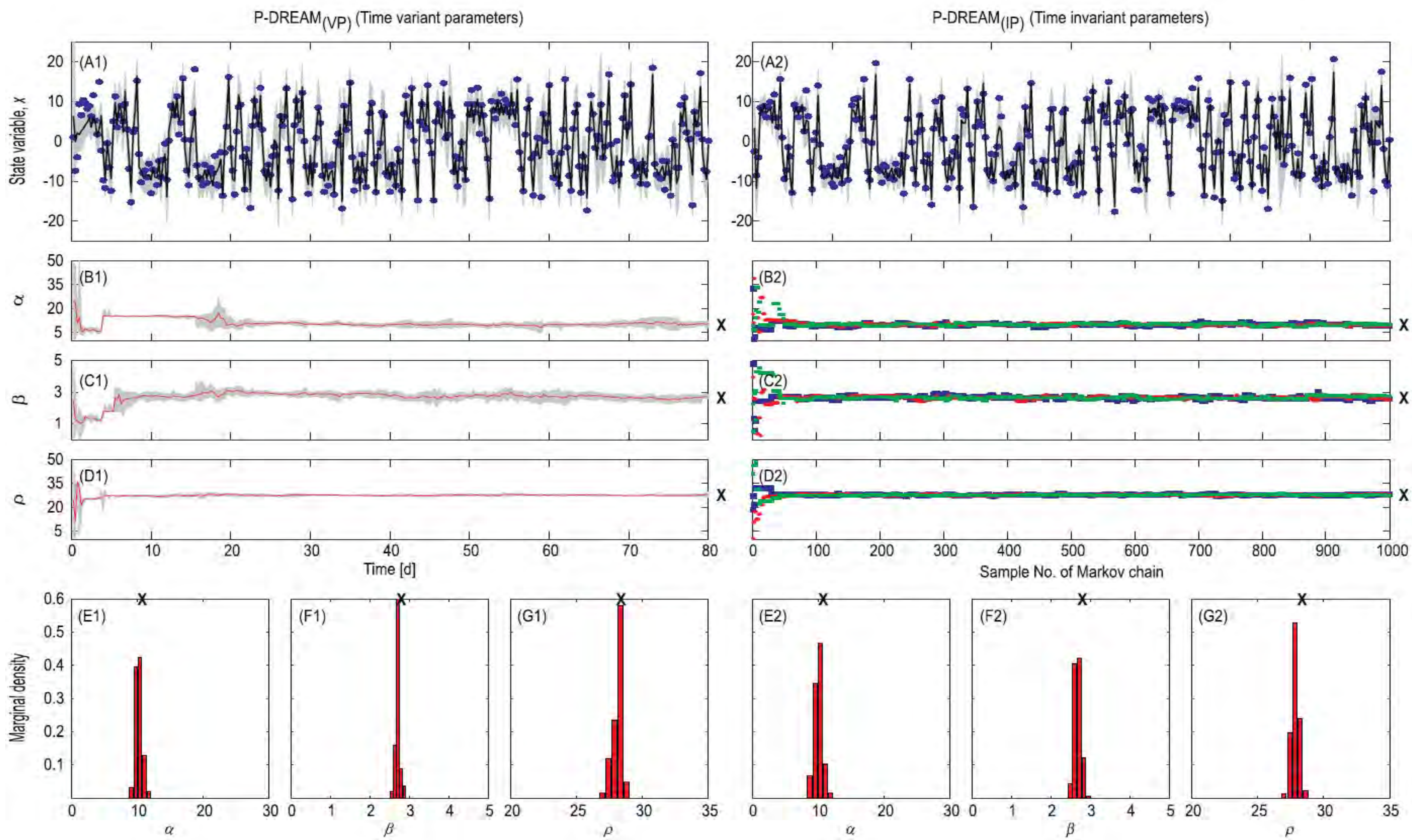
BENCHMARK STUDY: LORENZ MODEL



PARAMETERS ADDED TO STATE VECTOR – NOT RECOMMENDED!!

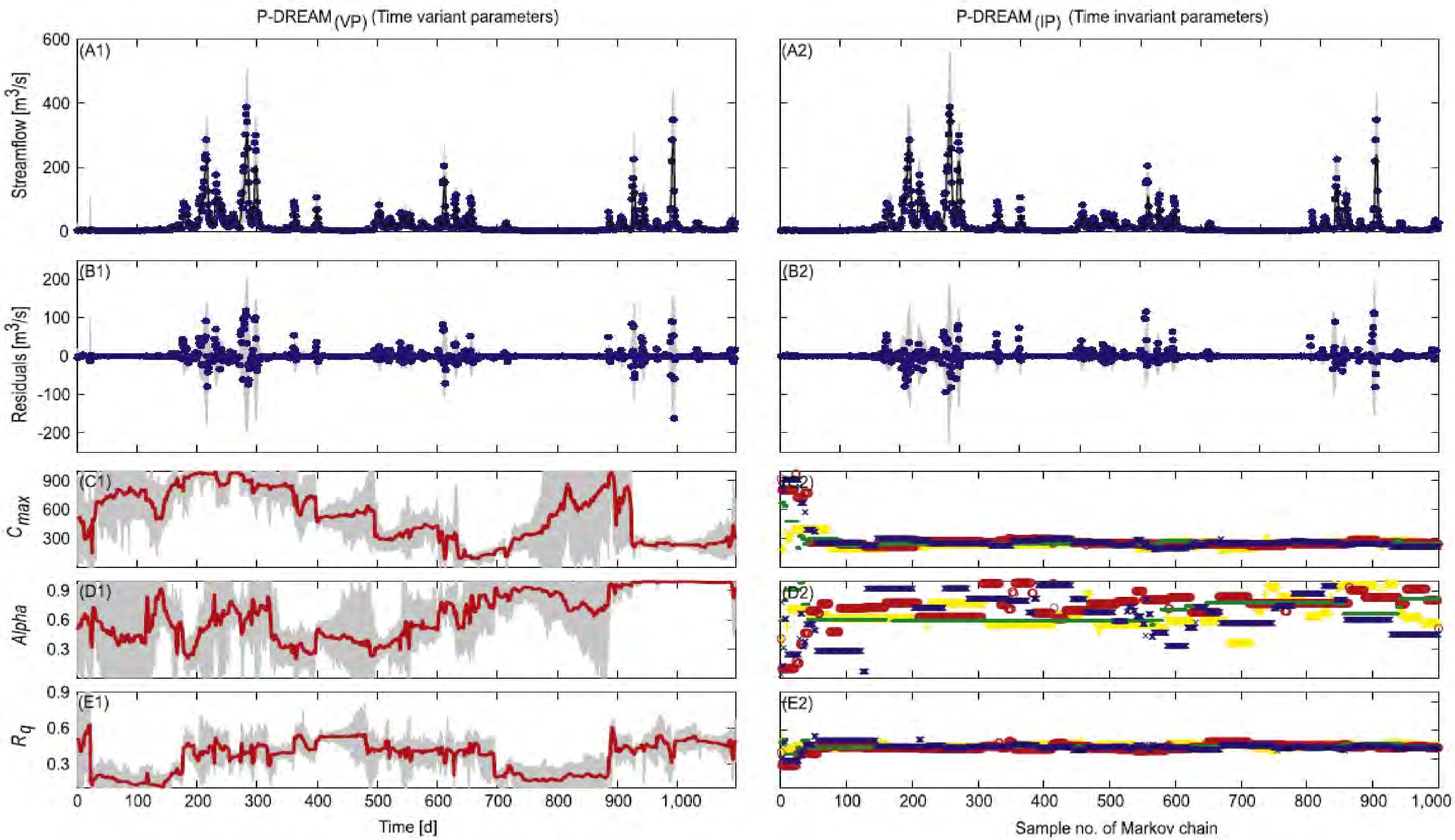
Vrugt et al., *AWR*, (2012)

LORENZ MODEL: TRACE PLOTS AND MARGINAL DISTRIBUTIONS



Vrugt et al., *AWR,* (2012)

CASE STUDY: HYDROLOGIC MODEL



Vrugt et al., *AWR*, (2012)

SOME OBSERVATIONS

PARTICLE FILTERS ARE STATISTICALLY SOUND (IF IMPLEMENTED CORRECTLY!) BUT VERY INEFFICIENT

MANY PARTICLES ARE REQUIRED TO DERIVE STABLE STATE PDF. NOT VIABLE FOR LARGE-SCALE AND REAL-TIME APPLICATION

RESAMPLING WITH MCMC AVOIDS PARTICLE COLLAPSE – BUT USE OF LIKELIHOOD RATHER THAN DATA ITSELF LEADS TO SLOW CONVERGENCE TO TARGET DISTRIBUTION.

STATE AUGMENTION NOT RECOMMENDED. PARAMETERS SHOW WONDERFUL TIME VARIATIONS – BUT THESE RESULTS ARE MEANINGLESS AND DUE TO INSUFFICIENT SAMPLE SIZE / POOR USE OF STATISTICS

MANY STUDIES (ALSO FOR ENKF) USE SUBJECTIVE TUNING FACTORS TO FORCE CONVERGENCE PARAMETER ENSEMBLE. THERE IS NO STATISTICAL BASIS FOR THIS APPROACH. MCMC IS THE ONLY FORMAL APPROACH

PARTICLE FILTER WITH ENKF UPDATE

DIRECTED UPDATED TOWARDS OBSERVATIONS

$$\left\{\mathbf{Z}_t^i\right\} = \left\{\mathbf{X}_t^i\right\} \pm \hbar^{-1}\left[\tilde{y}_t - \left(\hbar(\left\{\mathbf{X}_t^i\right\}, \phi) + v_t^i\right)\right] + \omega_t^i.$$

CAUTION: DETAILED BALANCE!!!

SOFTWARE: FACULTY.SITES.UCI.EDU/JASPER