Advances in Water Resources 51 (2013) 457–478

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 21 April 2012

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 $\oplus$ 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

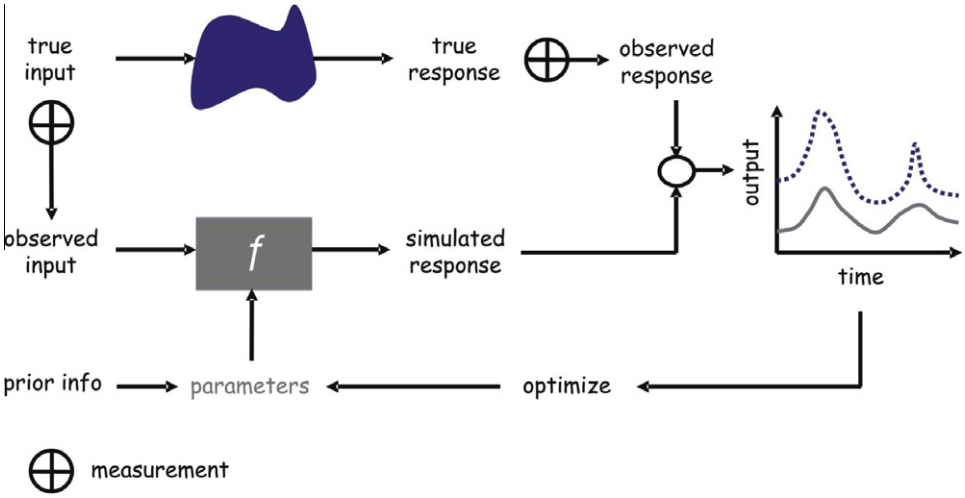

Fig. 1. Schematic overview of the model calibration problem: the model parameters are iteratively adjusted so that the predictions of the model, *f*, (represented with the solid line) approximate as closely and consistently as possible the observed response (indicated with the dotted line).

but demonstrate a significant bias towards lower streamflow values. A better compliance between model and data can be achieved by tuning the model parameters. If we write the dynamic nonlinear watershed model in a state-space formulation, the model calibration problem considered herein (Fig. 1) can be expressed as follows:

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

where $f(\cdot)$ is the (nonlinear) model operator expressing the watershed transition in response to forcing data $\mathbf{u}_t$ (rainfall, and potential evapotranspiration), model parameters, θ and state variables, $\mathbf{x}_t$. The variable ω_t represents errors in the model formulation, but is completely neglected in classical model calibration studies. In the remainder of this paper, we assume that the parameter space is bounded, $\theta \in \Theta \in \mathbb{R}^d$ and that the state space, Ω is of dimension $\mathbf{x}_t \in \Omega \in \mathbb{R}^\lambda$. Examples of system states within the context of watershed modeling include but are not limited to (spatially distributed) measurements of soil moisture content, pressure head, and groundwater table depth.

The measurement operator, $h(\cdot)$, defines the observation process and projects the model states, $\mathbf{x}_{t+1}$ to the model output, $\tilde{y}_{t+1}$ (observed response):

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

where v_{t+1} denotes the measurement error and ϕ stores any additional measurement parameters. This equation is quite popular and used in many textbooks and publications, but assumes that the system state at $t+1$ contains all necessary information to accurately predict the quantity of interest, y_{t+1}. On the contrary, if the calibration data consists of some time-averaged variable, then knowledge of the system (watershed) prior to $t+1$ is required, and thus an alternative formulation of Eq. (2) is warranted.

To establish whether $f(\cdot)$ provides an accurate description of the underlying system it is intended to represent, it is common practice to compare the simulated system behavior $\mathbf{y}_{1:n}(\theta)$ with respective observations $\tilde{\mathbf{y}}_{1:n}$. The difference between both is encapsulated in a residual vector $\mathbf{E}_{1:n}(\theta)$:

$$\mathbf{E}_{1:n}(\theta) = G[\tilde{\mathbf{y}}_{1:n}] - G[\mathbf{y}_{1:n}(\theta)] = \{e_1(\theta), \ldots, e_n(\theta)\}, \tag{3}$$

where $G[\,\cdot\,]$ allows for linear and nonlinear transformations of the model predictions and observational data. Examples include square root, Box-Cox [11], and normal quantile transformations [87,61], and the use of flow duration curves [116,10], spectral analysis [82,89] and wavelet spectral analysis [113,98,26]. Unfortunately, it is not particularly easy to work with the n-dimensional vector of error residuals, $\mathbf{E}_{1:n}$ directly and find the preferred parameter values. Instead, it is much easier to summarize the vector of residuals into a single measure of length, F, also called objective function. The goal of model calibration then simply becomes finding those values of θ that minimizes (maximizes, if appropriate) this criterion.

During the past 4 decades much progress has been made in the fitting of hydrologic models to data. That research has primarily focused on six different issues: (1) the development of specialized objective functions that appropriately represent and summarize the residual errors between model predictions and observations [53,106,66,5,57,99], (2) the search for efficient optimization algorithms that can reliably solve the watershed model calibration problem [133,31,12,75,136,102,119,77,62,123,112,125,88,130], (3) the determination of the appropriate quantity and most informative kind of data [67,107,44,135,124], (4) the selection and development of efficient and accurate numerical solvers for the partially structured differential and algebraic equation systems of hydrologic models [56,59,60,100], (5) the representation of uncertainty [7,9,65,94,111,134,34,45,69,117,118,132,16,35,80,85,86,8,51,57,58,64,76,103,122,1,36,37,126,127,20,81,109,97,105,128,129-,63,93,24], and (6) the development of inference techniques for refining the structural equations of hydrologic models [138,121,14,139].

Despite the progress made, increasing concern is surfacing in the hydrologic literature that the "classical" approach to model calibration introduced by *Carl Friedrich Gauss* (1794) has some serious deficiencies that necessitate the development of a more powerful paradigm. One of these deficiencies is that model structural and forcing (input) data errors are assumed to be either "negligibly small" or to be somehow "absorbed" into the error residuals, $\mathbf{E}_{1:n}(\theta)$. The residuals are then expected to behave statistically similar as the calibration data measurement error. Yet, many contributions to the hydrologic literature have demonstrated that this assumption is unrealistic. This is evidenced by error residuals that typically exhibit considerable variation in bias, variance, and correlation structures at different parts of the model response. Another deficiency is that the use of a single performance metric, F, no matter how carefully chosen, is inadequate to extract all information from the available calibration data. The use of such "insufficient statistic" promotes equifinality, and makes it unnecessarily difficult to find the preferred parameter values.

In this paper, we introduce a novel state-space filtering methodology that better preserves the actual information content of the calibration data. This Particle-DREAM method is inspired by recent developments in particle Markov chain Monte Carlo (MCMC) sampling [3] and combines the strengths of sequential Monte Carlo sampling and Markov chain Monte Carlo (MCMC) simulation with the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm [127,128]. We introduce two different variants of Particle-DREAM to satisfy assumptions regarding the temporal behavior of the model parameters. The standard P-DREAM$_{(IP)}$ filter assumes the parameters to be constant during the course of the simulation, and estimates $p(\theta|\tilde{\mathbf{y}}_{1:n})$ along with the evolving posterior state distribution. An alternative filter, entitled P-DREAM$_{(VP)}$ assumes the parameters to be time variant, and estimates a sequence of posterior parameter distributions, $p(\theta_1|\tilde{y}_1), p(\theta_2|\tilde{\mathbf{y}}_{1:2}), \ldots, p(\theta_n|\tilde{\mathbf{y}}_{1:n})$ jointly with the model states. We introduce the theory and concepts of both filters, and use numerical experiments to illustrate their usefulness and applicability to model-data synthesis studies. A subsequent paper [131] introduces an alternative modeling framework that disentangles the effect of input, output, model structural and parameter uncertainty in watershed hydrology.

The remainder of this paper is organized as follows. Section 2 introduces the rationale of our approach, followed by a detailed description and illustrative study of Particle-DREAM in Section 3 using the 40-dimensional Lorenz 96 model [72]. This study serves to demonstrate the advantages of particle resampling using MCMC simulation with DREAM. Section 4 demonstrates the performance of P-DREAM$_{(IP)}$ and P-DREAM$_{(VP)}$ using two different case studies involving the highly nonlinear Lorenz attractor [71], and a hydrologic watershed model. Here, we are especially concerned with parameter stationarity and sampling efficiency, and benchmark our results against the sequential importance resampling (SIR) filter of Moradkhani et al. [85] and Salamon and Feyen [97]. We highlight some additional developments that further improve filter efficiency when confronted with chaotic model dynamics and significant prediction bias. Finally, in Section 5 we summarize our results and discuss the main findings.

2. Standard Bayes law

The classical approach to model calibration considers the model parameters to be the only source of uncertainty and estimates their posterior probability density function (pdf) by maximizing $p(\theta|\tilde{\mathbf{y}}_{1:n})$ using Bayes theorem:

$$p(\theta|\tilde{\mathbf{y}}_{1:n}) = \frac{p(\theta)p(\tilde{\mathbf{y}}_{1:n}|\theta)}{p(\tilde{\mathbf{y}}_{1:n})}, \tag{4}$$

where $p(\theta)$ signifies the prior parameter distribution, and $L(\theta|\tilde{\mathbf{y}}_{1:n}) \equiv p(\tilde{\mathbf{y}}_{1:n}|\theta)$ denotes the likelihood function. The normalization constant or evidence, $p(\tilde{\mathbf{y}}_{1:n})$ is difficult to estimate directly in practice, and instead derived from integration

$$p(\tilde{\mathbf{y}}_{1:n}) = \int_{\Theta} p(\theta)p(\tilde{\mathbf{y}}_{1:n}|\theta)\mathrm{d}\theta = \int_{\Theta} p(\theta, \tilde{\mathbf{y}}_{1:n})\mathrm{d}\theta \tag{5}$$

over the parameter space so that $p(\theta|\tilde{\mathbf{y}}_{1:n})$ scales to unity. Explicit knowledge of $p(\tilde{\mathbf{y}}_{1:n})$ is desired for Bayesian model selection and averaging. A forthcoming paper by Schoups and Vrugt [101] uses the theory and concepts herein to develop an efficient path-sampling method [40] that numerically approximates, $p(\tilde{\mathbf{y}}_{1:n})$.

If our interest is only in the parameters of the model, then in practice we do not even need to calculate, $p(\tilde{\mathbf{y}}_{1:n})$ as all our statistical inferences (mean, standard deviation, etc.) can be made from the unnormalized density:

$$p(\theta, \tilde{\mathbf{y}}_{1:n}) = p(\theta)L(\theta|\tilde{\mathbf{y}}_{1:n}). \tag{6}$$

If we assume a noninformative (flat or uniform) prior, we are now left with a definition of the likelihood function. This function should adequately summarize the statistical properties of the residuals, $\mathbf{E}_{1:n}$. The choice of an adequate likelihood function, $L(\theta|\tilde{\mathbf{y}}_{1:n})$ has therefore been subject of considerable debate in the hydrologic and statistical literature. In its most simplistic case with scalar valued measurements ($\tilde{y}_t \in \mathbb{R}^1$) and uncorrelated, Gaussian distributed residuals with constant variance, σ_v^2, the likelihood function, $L(\theta|\tilde{\mathbf{y}}_{1:n})$, can be written as:

$$L(\theta|\tilde{\mathbf{y}}_{1:n}) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_v^2}} \exp\left[-\frac{1}{2}\hat{\sigma}_v^{-2}(\tilde{y}_t - y_t(\theta))^2\right], \tag{7}$$

where $\hat{\sigma}_v$ is an estimate of the standard deviation of the measurement error. The value of $\hat{\sigma}_v$ can be specified *a priori* based on knowledge of the measurements errors, or alternatively its value can be inferred simultaneously with the parameters in θ. In practice it is useful to remove the product sign from Eq. (7) and write:

$$L(\theta|\tilde{\mathbf{y}}_{1:n}) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_v^2}}\right)^n \exp\left[-\frac{1}{2}\hat{\sigma}_v^{-2}F_{\mathrm{SLS}}(\theta),\right] \tag{8}$$

in which $F_{\mathrm{SLS}}(\theta) = \sum_{t=1}^{n} e_t(\theta)^2$ denotes the sum of squared error and SLS stands for simple least squares.

This type of likelihood function was first introduced by Carl Friedrich Gauss (around 1794) within the context of (nonlinear) curve fitting, and these historical developments have led to current perspectives. For reasons of algebraic simplicity and numerical stability, it is often convenient to consider the log-likelihood function rather than $p(\tilde{\mathbf{y}}_{1:n}|\theta)$; the parameters or variables that maximize one also maximize the other. The log-likelihood, $\ell(\theta|\tilde{\mathbf{y}}_{1:n})$ of Eq. (7) is:

$$\ell(\theta|\tilde{\mathbf{y}}_{1:n}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_v^2\right) - \frac{1}{2}\sigma_v^{-2}\sum_{t=1}^{n}(\tilde{y}_t - y_t(\theta))^2. \tag{9}$$

The use of this likelihood function is convenient but not borne out of the statistical properties of the residuals that often violate assumptions of normality and independence.

In light of these considerations, Schoups and Vrugt [99] recently introduced a generalized likelihood function that is especially developed for nontraditional residual distributions with correlated, heteroscedastic, and non-Gaussian errors. When applied to daily streamflow data from a humid basin Schoups and Vrugt [99] found that (1) residuals 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 F_{SLS} approach, proper representation of the statistical distribution of residual errors yields tighter predictive uncertainty bands and different parameter uncertainty estimates that are far less sensitive to the particular time period used for calibration.

2.1. Practical implementation using Markov chain Monte Carlo simulation

A key task that remains is to summarize $p(\theta|\tilde{\mathbf{y}}_{1:n})$ in Eq. (4). Unfortunately, for most hydrologic models this task cannot be carried out by analytical means nor by analytical approximation, and MCMC methods are required to generate a sample of $p(\theta|\cdot)$. The basis of the MCMC method is a Markov chain that generates a random walk through the parameter space (also referred to as search space) and successively visits solutions with stable frequencies stemming from $p(\theta|\cdot)$. While initial applications of MCMC are primarily found in the fields of statistical physics, spatial statistics

and statistical image analysis, Gelfand and Smith [38] extended the method for posterior inference in a Bayesian framework. Nowadays, MCMC sampling is widely used for posterior exploration of multi-dimensional parameter spaces.

The earliest and general MCMC approach is the random walk Metropolis (RWM) algorithm [78]. Assume that we have already sampled points $\{\theta_0,\ldots,\theta_{k-1}\}$ this algorithms proceeds in the following three steps. First, a candidate point ϑ is sampled from a proposal distribution $q(\cdot|\cdot)$ that depends on the present location, θ_{k-1}. Next, the candidate point is either accepted or rejected using the Metropolis acceptance probability [78,50]:

$$\alpha(\vartheta,\theta_{k-1}) = 1 \wedge \frac{p(\vartheta|\tilde{\mathbf{y}}_{1:n})}{p(\theta_{k-1}|\tilde{\mathbf{y}}_{1:n})}\frac{q(\vartheta\rightarrow\theta_{k-1})}{q(\theta_{k-1}\rightarrow\vartheta)}, \tag{10}$$

where $p(\cdot)$ represents the posterior pdf, and $q(\theta_{k-1}\rightarrow\vartheta)$ ($q(\vartheta\rightarrow\theta_{k-1})$) denotes the conditional probability of the forward (backward) jump. This last ratio is unity if a symmetric proposal distribution is used and thus cancels out. Finally, if the candidate point is accepted the chain moves to ϑ, otherwise the chain remains at its current location θ_{k-1}. Following a so called burn-in period (of say, l steps), the chain approaches its stationary distribution and the vector $\{\theta_{l+1},\ldots,\theta_{l+m}\}$ contains samples from $p(\theta|\tilde{\mathbf{y}}_{1:n})$. The desired summary of the posterior distribution $p(\theta|\tilde{\mathbf{y}}_{1:n})$ is then obtained from this sample of m points.

3. Sequential Bayes law

Although the generalized likelihood function of Schoups and Vrugt [99] resolves practical problems with the choice of a suitable distribution for the residuals, all the information contained in $\mathbf{E}_{1:n}$ is lumped onto a single scalar, which makes it very difficult to diagnose, detect and resolve time-varying errors in the model structure, parameter nonstationarity and forcing data errors. Perhaps a more satisfactory approach would be to sequentially process the information contained in the data.

If we assume $\theta\in\Theta\in\mathbb{R}^d$ to be known, the evolving posterior distribution, $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ for the state-space formulation of Eqs. (1) and (2) can be written recursively as follows (see Appendix A):

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \frac{\overbrace{p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1})}^{\text{prior}}\overbrace{f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})}^{\text{model}}\overbrace{L_\theta(\tilde{y}_t|\mathbf{x}_t)}^{\text{likelihood function}}}{\underbrace{p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1})}_{\text{normalization constant}}}, \tag{11}$$

where $f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})$ signifies the transition probability density of the model states (model operator), and the likelihood function $L_\theta(\tilde{y}_t|\mathbf{x}_t)$ is the probability density of the observations given the state variables. This function uses the measurement operator of Eq. (2) to link $\mathbf{x}_t$ to the measured observations. If we integrate out $\mathbf{x}_{1:t-1}$ in Eq. (11), then we can derive the following expression of the marginal distribution, $p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t})_{t>1}$:

$$\overbrace{p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t})}^{\text{update step}} = \frac{L_\theta(\tilde{y}_t|\mathbf{x}_t)p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t-1})}{p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1})}, \tag{12}$$

where

$$\overbrace{p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t-1})}^{\text{prediction step}} = \int_\Omega f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})p_\theta(\mathbf{x}_{t-1}|\tilde{\mathbf{y}}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}. \tag{13}$$

The prediction and update step constitute the main part of particle filtering algorithms that numerically solve for the posterior state and/or parameter distribution given the information provided by all of the observations received up to the present time. Finally, the marginal likelihood, $p_\theta(\tilde{\mathbf{y}}_{1:n})$ in Eq. (4) can be computed from [30]:

$$p_\theta(\tilde{\mathbf{y}}_{1:n}) = p_\theta(\tilde{y}_1)\prod_{t=2}^{n}p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1}), \tag{14}$$

where

$$p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1}) = \int_\Omega p_\theta(\mathbf{x}_{t-1}|\tilde{\mathbf{y}}_{1:t-1})f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})L_\theta(\tilde{y}_t|\mathbf{x}_t)\ \mathrm{d}\mathbf{x}_{t-1:t}. \tag{15}$$

If the parameters are unknown, we use the prior density, $p(\theta)$ and infer the model parameters simultaneously with the model state trajectory using inference of the joint density, $p(\theta,\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$:

$$p(\theta,\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) \propto p(\theta)p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}). \tag{16}$$

This assumes the parameters to be time-invariant; a hypothesis that will be relaxed later on.

A key task that remains is to derive $p_\theta(\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ and $p(\theta,\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ (and their relevant marginal distributions) using the available observations. If the model operator $f_\theta(\cdot)$ and measurement operator $\hbar(\cdot)$ are linear, and ω_t and v_t are Gaussian, Kalman filter [54] type approaches find the exact posterior distribution. Nonlinear and non-Gaussian problems render an exact solution of $p_\theta(\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ and $p(\theta,\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ impossible, and therefore we need to resort to Monte Carlo based approximation methods. The next sections describe the underlying theory and concepts of this approach, and introduce a novel resampling methodology to adequately approximate the evolving posterior state and/or parameter distribution.

3.1. Practical implementation using sequential Monte Carlo sampling

In principle, we could approximate the sequence of posterior distributions, $\{p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})\}_{t\geqslant 1}$, using n different trials with DREAM by sequentially increasing the length of the calibration data set. Such an approach was used in [120] to demonstrate a difference in information content between throughfall and canopy water storage data for calibration of a single-layer rainfall interception model. This method works in practice, but is rather inefficient. Indeed, we can do much better than this by adopting an alternative framework using sequential Monte Carlo (SMC) sampling. These integration methods resolve many of the problems associated with traditional Kalman filter type approaches as they are not subject to assumptions of model linearity or Gaussianity. These methods were originally introduced in the early 1950s by physicists [48,96] and were used later [49,2,137] to track higher-order moments of the evolving target distribution. Nowadays, SMC methods are widely used in many different fields, including econometrics, signal processing, hydrology, and robotics, particularly because of tremendous increases in computational power.

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]

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

where $\delta_{(\cdot)}$ denotes the Dirac delta function and $\mathbf{W}_t = \left(W_t^1,\ldots,W_t^P\right)$ represent the normalized (importance) weights of the individual particles. This idea is similar to that in MCMC in which the posterior distribution is represented by a sample of draws. The differences are twofold. First, in SMC the particles carry weights, $\mathbf{W}_t$, whereas the draws in MCMC have equal probability. Secondly, SMC is designed to assimilate a new datum, that is, to evolve the particles in such a way that it represents the posterior at the next time point once the new datum has become available.

The SMC approach makes use of the following identity of Eq. (11)

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) \propto p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1}) f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t), \tag{18}$$

which suggests to reuse the samples at $t-1$ to approximate $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ at the subsequent time. Yet, in practice we cannot directly sample from $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ as this distribution is unknown and generally complex and high-dimensional even if we know $p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1})$. We therefore introduce an importance density, $q_\theta\left(\cdot|\tilde{y}_t, \left\{\mathbf{X}_{t-1}^{1:P}\right\}\right)$ with property $p_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_{1:t}) > 0 \Rightarrow q_\theta(\cdot) > 0$, and sample $\{\mathbf{X}_t^{1:P}\}$ from this distribution using the current datum, $\tilde{y}_t$ and particles, $\left\{\mathbf{X}_{t-1}^{1:P}\right\}$. The unnormalized importance weight, $\tilde{W}_t^i$ of the ith particle, $i = \{1,\ldots,P\}$ at time t then becomes:

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

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:

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

we derive the normalized importance weights which vary between 0 and 1.

This algorithm requires us to specify $\left\{q_\theta\left(\cdot|\tilde{y}_t, \left\{\mathbf{X}_{t-1}^{1:P}\right\}\right); t = 2,\ldots,n\right\}$. Many different techniques have been proposed in the statistical literature to design effective (efficient) importance distributions [27,115,90,91]. It is typically recommended to set $q_\theta(\mathbf{x}_t|\tilde{y}_t, \mathbf{x}_{t-1})$ as close as possible to $p_\theta(\mathbf{x}_t|\tilde{y}_t, \mathbf{x}_{t-1})$. In practice, however it remains typically difficult to select an adequate importance density. In response to this, Gordon et al. [43] have suggested to set $q_\theta(\mathbf{x}_t|\tilde{y}_t, \mathbf{x}_{t-1}) = f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})$ which results in the following equation for the incremental weights:

$$w_t\left(\left\{\mathbf{X}_{1:t}^i\right\}\right) = \frac{f_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right)}{f_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right)} = L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right). \tag{22}$$

This (default) strategy gives satisfactory performance, particularly when the model operator, $f_\theta(\cdot)$ adequately represents the "true" system dynamics, and/or the observations, $\tilde{\mathbf{y}}_{1:t}$ are not too informative. If the optimal importance distribution, $\{p_\theta(\mathbf{x}_t|\tilde{y}_t, \mathbf{x}_{n-1})\}_{t>1}$ is not used, repeated application of Eq. (19) causes particle impoverishment in which an increasing number of particles is exploring unproductive parts of the actual target distribution and assigned a negligible (zero) importance weight. This phenomenon is also referred to as particle drift or ensemble degeneracy. Not only does this result in a relative poor approximation of, $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ but also bad algorithmic efficiency as significant computational effort is devoted to carrying forward trajectories whose contribution to, $\hat{p}_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ is virtually zero.

The problem of particle degeneracy has received much attention in the statistical literature, and has inspired the development of a wide-variety of resampling methods [104], including residual [70], systematic [68,17], stratified [68,28] and multinomial [43] resampling. With high probability these methods discard "bad" samples with negligible importance and replace them with exact copies of more promising particles. This resampling step refocuses the thrust of the filter on the sampled points with the highest probability mass, but does not necessarily protect the particle filter from problems with weight degeneracy. A relatively large number of particles ($P \geqslant 500$) remains necessary for even the simplest inference problems to ensure that the evolving target distribution of interest is adequately sampled [6]. Here, we introduce an alternative resampling methodology that capitalizes on the advantages of MCMC simulation with DREAM [127,128]. This Particle-DREAM filter is theoretically coherent and simulation experiments presented in the forthcoming sections demonstrate that this approach requires far fewer particles than classical SIR filters to work well in practice. The details of Particle-DREAM will be outlined in the next section.

An unbiased estimate of the marginal likelihood, $p_\theta(\tilde{\mathbf{y}}_{1:t})$ in Eq. (14) is derived from the weights:

$$\hat{p}_\theta(\tilde{\mathbf{y}}_{1:t}) = \prod_{k=1}^t \left(\sum_{i=1}^P W_{k-1}^i w_k\left(\left\{\mathbf{X}_{1:k}^i\right\}\right)\right), \tag{23}$$

where the product term, $\sum_{i=1}^P W_{k-1}^i w_k\left(\left\{\mathbf{X}_{1:k}^i\right\}\right)$ is an approximation to $p_\theta(\tilde{y}_k|\tilde{\mathbf{y}}_{1:k-1})$ in Eq. (15) and $W_0^i = 1/P$. If $P \to \infty$, the empirical distribution of the particles converges asymptotically to $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ [25].

3.1.2. Particle resampling

Ideally, all P particles differ from each other and contribute equally to $\hat{p}_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$. In such situation, the variance of the weighs approaches zero, $\{\text{var}[\mathbf{W}_t] \Rightarrow 0\}_{t>1}$ and resampling is deemed unnecessary. On the contrary, if just a few particles receive considerable weights, then $\{\text{var}[\mathbf{W}_t]\}_{t>1}$ will increase, and resampling might seem appropriate. The effective sample size, ESS is a simple diagnostic that summarizes in a single number between 1 and P the variability of the normalized importance weights:

$$\text{ESS} = \left(\sum_{i=1}^P \left(W_t^i\right)^2\right)^{-1}. \tag{24}$$

In practice, resampling is performed whenever ESS drops below a pre-specified threshold, P^*. For convenience, we write $P^* = rP$, and consider different values of $r \in [0,1]$ in our numerical simulations. In general, the higher the value of r the more likely that resampling takes place. We now describe two resampling steps that are executed successively any time when ESS $< P^*$. For notational convenience, whenever the index i is used we mean 'for all $i \in \{1,\ldots,P\}$'.

Suppose that at some time t the value of ESS has sufficiently deteriorated and become much smaller than P, the total number of particles used to approximate $\hat{p}_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ in Eq. (17). Resampling is required to remove the bad particles, and rejuvenate the ensemble. The first step constitutes residual resampling, and replaces aberrant particles with low importance weights with exact copies of the most promising particles. We first compute a selection probability, $p_{(\{\mathbf{x}_t^i\})}$ of each individual particle:

$$p_{(\{\mathbf{x}_t^i\})} = \frac{PW_t^i - \left\lfloor PW_t^i \right\rfloor}{P - R}, \tag{25}$$

where the operator $\lfloor\cdot\rfloor$ rounds down to the nearest integer, and $R = \sum_{j=1}^P \lfloor PW_t^j \rfloor$. We retain those R particles whose value of $\lfloor P\mathbf{W}_t \rfloor \geqslant 1$, and fill the remaining $P - R$ spots by drawing particles from a multinomial distribution, as follows:

$$j = \arg\min_{j'}^P \left(\sum_{i=1}^{j'} p_{(\{\mathbf{x}_t^i\})} \geqslant u\right), \tag{26}$$

where j is an index from $\left\{\mathbf{X}_t^{1:P}\right\}$, and $u \sim U[0,1]$ is an auxiliary random label sampled from the standard uniform distribution. For each trial, we draw a different a new value for u. This resampling procedure leads to P equally-weighted particles, $\mathbf{W}_t = 1/P$ (hence $\{\text{var}[\mathbf{W}_t] \Rightarrow 0\}$) many of which are exact copies of one another. This resampling procedure is therefore prone to a rapid loss of particle diversity.

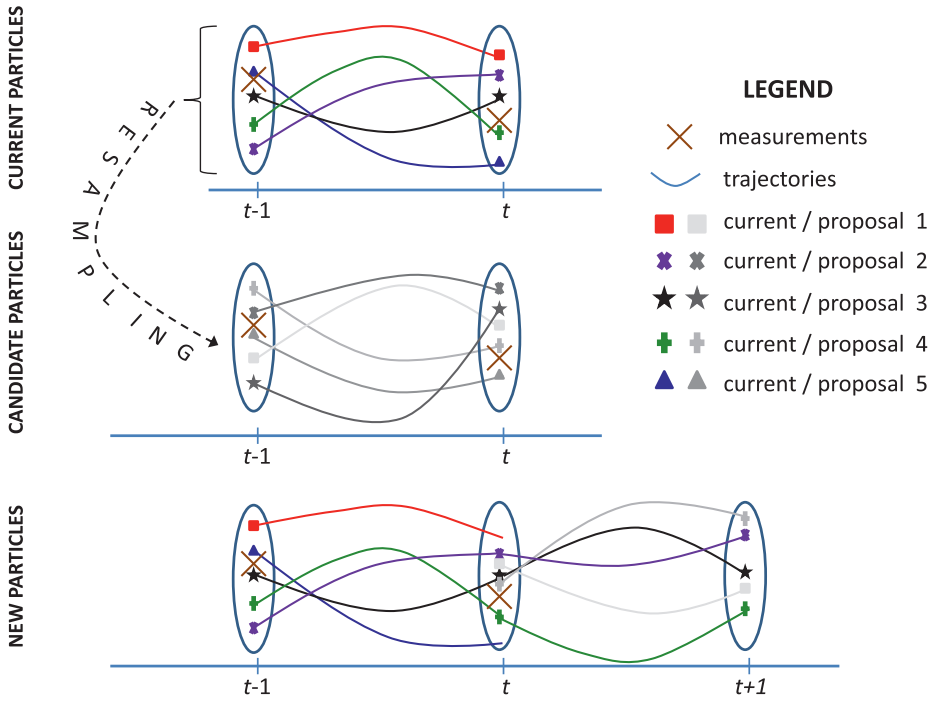

Fig. 2. Simplified representation of the resampling step with DREAM. Each different particle trajectory (solid lines) is coded with a different color and symbol. The original particles (top panel) are printed in color, whereas their corresponding candidate trajectories (after resampling with DREAM at time t) are illustrated with different gray tones (middle panel). The bottom panel illustrates the Metropolis acceptance/rejectance step. Proposal trajectories 1 and 5 are accepted (acceptance rate of 40%). Their corresponding state predictions at time t and/or parameter values (in case of P-DREAM$_{(VP)}$) replace their respective parents. The updated ensemble of particles at time t is then propagated forward and used to predict the observation at $t+1$. In this paper we focus primarily on prediction; hence, we update the ensemble at time t. Smoothing, on the contrary, is easily implemented by replacing the immediate past trajectories of the current particles with those of the accepted particles. In the present context this would further improve filter performance at $t-1$, yet assumes knowledge of future observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We now introduce a subsequent resampling step that promotes particle diversity and quality. We create new samples, $\left\{\mathbf{Z}_t^{1:P}\right\}$ using MCMC simulation with the DREAM algorithm [127,128]. This second resampling step is schematically illustrated in Fig. 2 and discussed in detail below.

We start by proposing a new particle, $\left\{\mathbf{Z}_{t-1}^i\right\}$, by resampling at time $t-1$ from the existing ensemble $\left\{\mathbf{X}_{t-1}^{1:P}\right\}$ using differential evolution [108]:

$$\left\{\mathbf{Z}_{t-1}^i\right\} = \left\{\mathbf{X}_{t-1}^i\right\} + (\mathbf{1}_\lambda + \mathbf{e}_\lambda)\gamma(\tau,\lambda')\left[\sum_{j=1}^{\tau}\left\{\mathbf{X}_{t-1}^{r_1(j)}\right\} - \sum_{h=1}^{\tau}\left\{\mathbf{X}_{t-1}^{r_2(h)}\right\}\right] + \boldsymbol{\varepsilon}_\lambda, \tag{27}$$

where λ is the dimensionality of the state space, $\gamma(\tau,\lambda') = 2.4/\sqrt{2\tau\lambda'}$ denotes a jump-rate that depends on the number of pairs, τ and number of coordinates (state variables), λ' that will be jointly updated, and $r_1(j) \neq r_2(h) \neq i$ for $j = 1,\ldots,\tau$ and $h = 1,\ldots,\tau$. The values of $\mathbf{e}_\lambda$ and $\boldsymbol{\varepsilon}_\lambda$ are drawn from $U_d(-b,b)$ and $N_d(0,b^*)$ with $|b| < 1$, and b^* small compared to the width of the proposal distribution. Interested readers are referred to Vrugt et al. [127,128] for further details.

We then use the model operator, f, in Eq. (1) to simulate $\mathbf{x}_t^i$ from $\mathbf{x}_{t-1}^i$, or $f_\theta(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$ and use the Metropolis ratio, $\alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\},\left\{\mathbf{X}_t^i\right\}\right)$ to decide whether to accept the proposal or not:

$$\alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\},\left\{\mathbf{X}_t^i\right\}\right) = 1 \wedge \frac{f_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right)L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right)L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{Z}_t^i\right\}\right)}{f_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right)L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right)L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right)}. \tag{28}$$

If the proposal is accepted, $\left\{\mathbf{Z}_t^i\right\}$ replaces $\left\{\mathbf{X}_t^i\right\}$, otherwise we retain the current particle. The acceptance probability in Eq. (28) is derived in Appendix B of this paper and leaves $\hat{p}_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ invariant. Proof of detailed balance of the transition kernel of DREAM is given in our previous work [110,127,128] and interested readers are referred to these publications for further details.

The main advantages to resampling and moving the particles at time $t-1$ rather than at t is that the particle ensemble at $t-1$ is more diverse as judged by the ESS diagnostic, and consequently the proposals created using Eq. (27) are likely to exhibit more diversity. Note that other MCMC filters have been documented in the (statistical) literature. For instance, RESAMPLE-MOVE filters [4,70,74,42] periodically rejuvenate the particle ensemble using MCMC simulation. These methods not only primarily focus on state estimation, but also use a rather inefficient reversible-jump RWM algorithm to approximate the moving target distribution.

3.2. Particle-DREAM: Pseudo-code

We now present a pseudo-code of Particle-DREAM, the filter that implements the theoretical and numerical developments outlined in the previous two sections. This filter merges the strengths of SMC sampling and MCMC simulation with DREAM to resolve problems with ensemble degeneracy, and ensure adequate particle diversity. The Particle-DREAM filter proceeds as follows. For notational convenience, whenever the index i is used we mean 'for all $i \in \{1,\ldots,P\}$'. The variable $\mathsf{F}(\cdot|\mathbf{p})$ signifies the discrete multinomial distribution on $\{1,\ldots,P\}$ of parameter $\mathbf{p} = (p_1,\ldots,p_P)$ with $p_j \geqslant 0$ and $\sum_{j=1}^{P} p_j = 1$.

AT TIME $t = 0$

STEP 1: INITIALIZATION

(a) Create $\left\{\mathbf{X}_0^i\right\}$ 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 $\left\{\mathbf{X}_t^i\right\} \sim f_\theta\left(\cdot\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right)$

STEP 3: UPDATE

(a) Calculate the incremental importance weight,

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

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

(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 $\mathrm{F}\left(\left\{\mathbf{X}_t^{1:P}\right\}\middle|p_{(\{\mathbf{X}_t^{1:P}\})}\right)$.

STEP 4B: MCMC RESAMPLING

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

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

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

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

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

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

STEP 5: RESET WEIGHTS

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

In words, Particle-DREAM numerically approximates the evolving posterior state distribution, $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ using sequential Monte Carlo sampling and Markov chain Monte Carlo resampling. To start, an initial ensemble of particles is created from the prior state distribution, $p(\mathbf{x}_0)$. These particles are propagated forward by sampling from an importance density, and confronted with the next available observation using a likelihood function. This helps partition the state space into productive and unproductive regions. The importance weights of the individual particle trajectories are subsequently updated, and used to assess whether resampling is necessary. If the effective sample size has dropped below a user-defined threshold the particle ensemble is rejuvenated using residual resampling and MCMC simulation.

Particle-DREAM contains two algorithmic parameters whose values need to be specified by the user. This includes the number of particles, P, and threshold parameter, r that directly determines when to resample or not. Specific guidelines with respect to their values are given in the different case studies. Note that Particle-DREAM is embarrassingly parallel, and relatively easy to run on a distributed computing network. Each individual particle of the ensemble can be evolved on a different processor. This significantly speeds-up computational efficiency, and facilitates the use of CPU intensive simulation models.

To illustrate the advantages of particle resampling using MCMC simulation with DREAM, we compare the performance of Particle-DREAM against a classical SIR filter using the Lorenz 1996 model [72,73]. This highly nonlinear atmospheric "toy model", hereafter referred to as L96, simulates mid-latitude weather conditions using:

$$\frac{dx_j}{dt} = -x_{j-1}\overbrace{(x_{j-2} - x_{j+1})}^{(1)} - \underbrace{x_j}_{(2)} + \overbrace{F}^{(3)} + \omega_{x_j}, \tag{29}$$

where $\mathbf{x} = \{x_j; j = 1,\ldots,\lambda\}$ represents some atmospheric quantity of interest that is subject to (1) advection, (2) dissipation and (3) forcing. The state is discretized into λ sectors along the latitude circle, and assumed periodic, i.e. $x_j = x_{j+\lambda}$. A reference solution of $(x_{1t},\ldots,x_{\lambda t})$ with $\Delta t = 0.05$ and $\lambda = 40$ is computed for $t \in [0,10]$ by solving Eq. (29) from the initial condition $(x_1,\ldots,x_{19},x_{20},x_{21},\ldots,x_{40}) = (F,\ldots,F,\ F+\varsigma,F,\ldots,F)$ using a Runge–Kutta fourth-order scheme with an explicit integration time step of $dt = 0.005$ days and model error drawn from $N(0,\sqrt{\Delta t}\sum_\omega)$ with $\sum_\omega$ similar to a 40×40 identity matrix. To enforce chaotic model dynamics we set $F = 8$ and $\varsigma = 0.1$ [72]. Synthetic observations are subsequently created by perturbing each data point of the reference solution with a normally distributed measurement error with variance, $\sigma_v^2 = 0.1$. The final data set thus consists of the temporal dynamics of a 40-dimensional state variable space.

Table 1 summarizes the performance of Particle-DREAM with and without MCMC particle resampling. The latter results simply correspond to a classical SIR particle filter. Listed statistics include the root mean square error (RMSE) and % bias (BIAS) of the mean ensemble state prediction error, and the percentage of measurements contained in the 95% prediction interval. Each individual value is computed using the last 50% of the data, $t \in [5.05,10]$ and represents an average of 25 different trials. We separately list the results for a single and ten successive (in italic) MCMC resampling steps. In all our simulations, the initial particle ensemble is drawn from a multi-normal prior distribution, $p_\lambda(\mathbf{x}_0) \sim N(F,\sigma_v^2)$. We con-

Table 1
L96 model: summary statistics of the one-observation ahead state prediction error (RMSE and % BIAS) and 95% prediction uncertainty ranges (PERCT-percentage of observations within this interval) using particle DREAM with (left column) and without (right column) particle resampling using MCMC simulation with DREAM. We separately consider two different MCMC variants involving a single and ten consecutive (in italic) DREAM resampling steps. Listed values represent an average of 25 different trials.

	Particle DREAM					
	With MCMC simulation			Without MCMC simulation[a]		
	RMSE	PERCT, %	BIAS, %	RMSE	PERCT, %	BIAS, %
$P = 10$	3.60 (*1.78*)	22.8 (*50.9*)	4.19 (*1.60*)	4.05	18.6	4.56
$P = 25$	2.65 (*1.25*)	41.2 (*76.7*)	4.02 (*2.21*)	3.00	35.3	4.25
$P = 50$	1.76 (*1.02*)	57.0 (*84.8*)	2.71 (*0.79*)	2.17	55.0	3.27
$P = 100$	1.15 (*0.95*)	76.2 (*87.6*)	1.21 (*0.65*)	1.36	72.8	1.33
$P = 250$	0.92 (*0.68*)	85.2 (*90.3*)	0.98 (*0.45*)	1.05	83.8	1.14

[a] Similar to a SIR filter with residual resampling.

sider ensemble sizes of $P = \{10,25,50,100,250\}$ different particles, and a resampling threshold of $r = 0.7$. The most important results are as follows.

As expected, particle resampling using MCMC simulation with DREAM readily improves filter performance. This is observed for all three listed statistics, and provides strong support for our claim that MCMC resampling has desirable properties. The larger the number of successive MCMC resampling steps, the better the quality and diversity of the particle ensemble, and the closer Particle-DREAM tracks the evolving posterior state distribution. With ten consecutive DREAM resampling steps the RMSE of the one-observation-ahead mean ensemble forecast error has reduced to about 1/3–1/2 of the RMSE of a classical SIR filter. The largest (relative) improvements are found for the smallest ensemble sizes. Note that both filters exhibit difficulty in accurately characterizing the 95% prediction uncertainty ranges. The highly nonlinear dynamics of L96 are difficult to capture with a small particle ensemble, yet again MCMC simulation generates the most coherent results. Altogether these results demonstrate the advantages of particle resampling using MCMC simulation with DREAM.

Despite the enhanced performance, the use of MCMC simulation comes at the expense of an increased computational complexity and demand. Fortunately, the results in Table 1 and those of other studies presented herein illustrate that Particle-DREAM requires far fewer particles than a classical SIR filter to work well in practice. This constitutes an important advantage for real-time implementation. In the penultimate section of this paper, we highlight a more efficient MCMC proposal distribution that requires just a single DREAM resampling step to closely track the evolving posterior target distribution.

Up to now, we have almost solely focused our attention to state variable inference without recourse to parameter estimation. Indeed, Particle-DREAM assumes the parameters to be known *a priori*, an assumption that is particularly difficult to justify in hydrologic modeling. Many, if not all, watershed model parameters represent conceptual or effective properties that cannot be measured directly in the field at the scale of interest. The next two sections introduce two different variants of Particle-DREAM, also referred to as P-DREAM$_{(IP)}$ and P-DREAM$_{(VP)}$ that explicitly resolve the model parameters. The first filter, P-DREAM$_{(IP)}$ assumes the model parameters to be time-invariant and numerically approximates the joint target density, $p(\theta, \mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ of the model parameters and state trajectories. The other filter, P-DREAM$_{(VP)}$ allows for temporal parameter changes and resolves $p(\theta_{1:t}, \mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$. This approach is especially designed to detect temporal changes in the parameter values, yet violates the main assumption of the standard model calibration paradigm that parameters represent invariant system properties.

3.3. P-DREAM$_{(IP)}$: Time-invariant Parameters

If the parameters are unknown, we use the prior density, $p(\theta)$ and infer the model parameters simultaneously with the model state trajectory using inference of the joint density, $p(\theta, \mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) \propto p(\theta)p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$. This filter draws inspiration from our previous work [121], but implements some important changes to make sure that the filter converges adequately. The theoretical underpinning of P-DREAM$_{(IP)}$ can be found in Andrieu et al. [3].

The key idea is to run Particle-DREAM with a number of different Markov chains. These chains are generated with DREAM [127,128] and used to generate samples from the posterior parameter distribution. Each time a candidate point, θ^i, is created in the individual chains, Particle-DREAM is used to resolve the state estimation problem and approximate $p_{\theta^i}(\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$. The corresponding marginal likelihood, $p_{\theta^i}(\tilde{\mathbf{y}}_{1:n})$ is then used to decide whether to accept the proposal or not.

We now present a pseudo-code of P-DREAM$_{(IP)}$. Whenever the index i is used we mean 'for all $i \in \{1,\ldots,N\}$', where N denotes the number of (parallel) Markov chains created with DREAM.

STEP 1: INITIALIZATION
- (a) Set $k \leftarrow 1$.
- (b) Sample θ_k^i from the prior parameter distribution, $p(\theta)$.
- (c) Run Particle-DREAM to derive the marginal likelihood, $\hat{p}_{\theta_k^i}(\tilde{\mathbf{y}}_{1:n})$, $\{\mathbf{X}_{1:n}^{1:P}\}$ and the final importance weights, $\mathbf{W}_n$.
- (d) Randomly sample a single state trajectory, $\{\mathbf{X}_{1:n}^i\} \sim \mathsf{F}\left(\left\{\mathbf{X}_{1:n}^{1:P}\right\}\middle|\mathbf{W}_n\right)$ and store this path in $\mathbf{X}_{1:n}^i(k)$.

STEP 2: CREATE CANDIDATE POINTS
- (a) Create a candidate parameter vector, ϑ^i using differential evolution [108]

$$\vartheta^i = \theta_k^i + (\mathbf{1}_d + \mathbf{e}_d)\gamma(\tau, d')\left[\sum_{j=1}^{\tau}\theta_k^{r_1(j)} - \sum_{h=1}^{\tau}\theta_k^{r_2(h)}\right] + \boldsymbol{\varepsilon}_d$$

STEP 3: EVALUATE CANDIDATE POINTS
- (a) Compute $\hat{p}_{\vartheta^i}(\tilde{\mathbf{y}}_{1:n})$ with Particle-DREAM and store $\{\mathbf{X}_{1:n}^{1:P}\}$ and $\mathbf{W}_n$.
- (b) Sample $\left\{\mathbf{X}_{1:n}^{i*}\right\} \sim \mathsf{F}\left(\left\{\mathbf{X}_{1:n}^{1:P}\right\}\middle|\mathbf{W}_n\right)$.

STEP 4: METROPOLIS STEP
- (a) Update $k \leftarrow k + 1$
- (b) Set $\theta_k^i = \vartheta^i$, $\hat{p}_{\theta_k^i}(\tilde{\mathbf{y}}_{1:n}) = \hat{p}_{\vartheta^i}(\tilde{\mathbf{y}}_{1:n})$, and $\mathbf{X}_{1:n}^i(k) = \left\{\mathbf{X}_{1:n}^{i*}\right\}$ with probability

$$\alpha\left(\vartheta^i, \theta_{k-1}^i\right) = 1 \wedge \frac{p\left(\vartheta^i\right)\hat{p}_{\vartheta^i}(\tilde{\mathbf{y}}_{1:n})}{p\left(\theta_{k-1}^i\right)\hat{p}_{\theta_{k-1}^i}(\tilde{\mathbf{y}}_{1:n})}$$

- (c) Otherwise remain at the old position, $\theta_k^i = \theta_{k-1}^i$, and set $\hat{p}_{\theta_k^i}(\tilde{\mathbf{y}}_{1:n}) = \hat{p}_{\theta_{k-1}^i}(\tilde{\mathbf{y}}_{1:n})$ and $\mathbf{X}_{1:n}^i(k) = \mathbf{X}_{1:n}^i(k-1)$.

STEP 5: CHECK PARAMETER CONVERGENCE
- (a) For each individual parameter, $j = \{1,\ldots,d\}$ compute the R-statistic, $\hat{R}_j$ of Gelman and Rubin [39] using the last 50% of the samples stored in each of the N chains.
- (b) If $\hat{R}_j \leqslant 1.2$ for all j STOP, otherwise go to STEP 2.

The P-DREAM$_{(IP)}$ code developed herein is a particular version of the generic particle marginal Metropolis–Hastings sampler of Andrieu et al. [3] and consists of an inner SMC loop with Particle-DREAM to resolve $p_{\theta^i}(\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ and an outer MCMC simulation loop with DREAM to estimate the posterior parameter distribution. This hybrid framework is very similar to the SODA approach of Vrugt et al. [121], but includes two important changes to ensure that the filter converges to the exact posterior parameter and state distribution. In the first place, P-DREAM$_{(IP)}$ uses a particle filter rather than ensemble Kalman filter [32] to solve the sequential state estimation problem. This has several theoretical advantages, in particular the ability to handle nontraditional probability distributions. Secondly, P-DREAM$_{(IP)}$ estimates the evolving state posterior from the different trajectories stored in $\left\{\mathbf{X}_{1:n}^{1:P}\right\}$ (hence using adequate burn-in), rather than a separate EnKF run with SODA after the maximum likelihood values of the model parameters have been determined.

In the past decade other joint parameter and state estimation approaches have been introduced in the hydrologic literature. For instance, consider the SIR particle filter of Moradkhani et al. [85] (also used in [97]), hereafter conveniently referred to as *SIR*$_{(M)}$. This filter is trying to estimate $p(\theta, \mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ using simultaneous parameter and state estimation. The following transition prior is used for the parameters:

$$q(\theta_t|\tilde{\mathbf{y}}_t,\theta_{t-1}) = p(\theta_t|\theta_{t-1}). \quad (30)$$

To avoid sample impoverishment, the calibration parameters are randomly perturbed after each successive updating step using a multi-normal distribution:

$$\theta_t \sim N\left(\theta_{t-1}, s^2 \mathrm{Var}[\theta_{t-1}]\right), \quad (31)$$

where s is an algorithmic parameter that determines the "size" of the (hyper) volume around which each individual particle is explored.

Numerical results presented in [85] suggest convergence of the sampled parameters to a limiting distribution. Practical experience however suggests that the convergence properties and speed of $SIR_{(M)}$, are essentially dependent on the choice of the importance density, $q(\theta_t|\tilde{\mathbf{y}}_t,\theta_{t-1})$ and value of s used to corrupt the parameters after each successive time step. If s is taken too large, the resampling step will introduce parameter drift and corrupt the approximation of $p(\theta,\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$. On the contrary, if s is too small, then the resampled parameters exhibit insufficient dispersion, and underestimate the actual parameter uncertainty. In the absence of theoretical convergence proofs and clear guidelines on the selection of s in Eq. (31), the $SIR_{(M)}$ filter theoretically neither estimates exactly $p(\theta,\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ (time-invariant parameters) nor does it properly retrieve, $p(\theta_{1:n},\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$ (time variant parameters). The P-DREAM$_{(IP)}$ filter resolves this problem, but at the expense of using an outside MCMC simulation loop with DREAM. Note that the original paper of Moradkhani et al. [85] does not specify which value of s is actually being used in the simulation experiments. Another area of concern is that Eq. (30) does not preserve the actual parameter correlation. Each individual parameter is simply sampled in turn. This severely deteriorates sampling efficiency, an issue that becomes particularly important for multi-modal and high-dimensional parameter spaces. Application to a parameter estimation version of the nonlinear model of Kitagawa [68] discussed in Section 4.3 of Doucet and Tadić [29] will illustrate the importance of this deficiency.

3.4. P-DREAM$_{(VP)}$: Time-variant Parameters

The P-DREAM$_{(IP)}$ filter presented in the previous section, assumes the model parameters to be invariant. This hypothesis is convenient, but might not be justified when the system of interest experiences functional changes. For instance, urbanization and changes in land-use are known to affect the watershed response to rainfall. One way to account for this nonstationarity is to allow the parameters to vary dynamically with time. We therefore introduce an alternative variant of Particle-DREAM, called P-DREAM$_{(VP)}$ that considers the model parameters to be time-variant and approximates the joint density, $p(\theta_{1:t},\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ using state augmentation. The pseudo-code of P-DREAM$_{(VP)}$ is very similar to that of Particle-DREAM presented in Section 3.2, but contains two main changes to part (1) of the initialization and prediction step. We list these changes below.

STEP 1: INITIALIZATION

(1) Create $\left\{\mathbf{X}_0^i\right\}$ by sampling from the prior state distribution, $p_0(\mathbf{x}_0)$, sample $\{\boldsymbol{\theta}_0^{1:P}\}$ from $p_d(\theta)$ and define the augmented state vector, $\{\tilde{\mathbf{X}}_0^i\} = \left(\left\{\mathbf{X}_0^i\right\};\left\{\boldsymbol{\theta}_0^i\right\}\right)$.

STEP 2: PREDICTION

(1) Sample $\left\{\mathbf{X}_t^i\right\} \sim q_\theta(\cdot|\tilde{\mathbf{y}}_t,\left\{\mathbf{X}_{t-1}^i\right\})$ and sample $\left\{\boldsymbol{\theta}_t^i\right\} \sim q\left(\cdot|\tilde{\mathbf{y}}_t,\left\{\boldsymbol{\theta}_{t-1}^i\right\}\right)$.

The other three remaining steps are unaffected but now involve an extended state vector $\left\{\tilde{\mathbf{X}}_t^{1:P}\right\}_{t>1}$ rather than $\left\{\mathbf{X}_t^{1:P}\right\}_{t>1}$. The resampling step in P-DREAM$_{(VP)}$ thus jointly samples the state variables and associated parameter values, instead of Gibbs sampling of both groups of variables in turn [41]. This preserves the correlation structure. In practice, it is much more challenging to estimate $p(\theta_{1:t},\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$ than to derive $p(\theta,\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})_{t>1}$. Various reasons contribute to this, of which the added dimensionality of the parameter space, interactions between state variables and model parameters, and temporal changes in parameter sensitivity are most important. The numerical experiments presented below serve to illustrate this in more detail.

3.5. P-DREAM$_{(IP)}$ and P-DREAM$_{(VP)}$: proof of convergence

Whereas particle filters provide a general framework to generate samples from a moving target distribution, convergence properties of such methods are often not discussed. Unfortunately, classical limit theorems based on independent and identically distributed assumptions, do not apply because the individual trajectories (samples) of the particle filter interact, and thus, are statistically dependent. We refer the reader to Vrugt et al. [127,128] for a proof of detailed balance and ergodicity of DREAM, Crisan et al. [21], Crisan [22], Crisan and Doucet [23] for proof and concepts that Particle-DREAM leaves $p_\theta(\{\mathbf{X}_{1:n}\}|\tilde{\mathbf{y}}_{1:n})$ invariant, and Theorem 4 in Section 4.4 of Andrieu et al. [3] for a convergence proof of P-DREAM$_{(IP)}$. The asymptotic properties of P-DREAM$_{(VP)}$ are not yet very well understood. Some guidance is given in Kantas et al. [55].

4. Case studies

We conduct two different case studies to illustrate P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$. The first study considers the Lorenz oscillator [71], and illustrates the ability of both filters to jointly estimate model states and parameters when confronted with highly nonlinear and non-Gaussian model dynamics. The second case study explores the usefulness and applicability of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ by application to a parsimonious conceptual rainfall–runoff model using historical data from the Leaf River basin in the United States. Here, we are especially concerned with parameter nonstationarity and compare the posterior parameter and state distribution derived with both filters using synthetic and real-world daily discharge data. In both studies we assume $q_\theta(\mathbf{x}_t|\tilde{\mathbf{y}}_t,\mathbf{x}_{t-1}) = f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})$ and use Eq. (22) to calculate the values of the incremental weights. The identity, $\theta_t = \theta_{t-1}$ is used as importance density, $q(\theta_t|\tilde{\mathbf{y}}_t,\theta_{t-1})$ in P-DREAM$_{(VP)}$. This is rather simplistic but reasonable when augmented with a MCMC resampling step. In all our simulations with P-DREAM$_{(IP)}$ we use $N = 10$ different chains.

4.1. Case study I: the Lorenz attractor

The first case study considers the Lorenz attractor [71]. This highly nonlinear model consists of a system of three first-order coupled and nonlinear ordinary differential equations:

$$\begin{aligned}\frac{dx}{dt} &= \alpha(y-x) + \omega_x,\\ \frac{dy}{dt} &= x(\rho - z) - y + \omega_y,\\ \frac{dz}{dt} &= xy - \beta z + \omega_z\end{aligned} \quad (32)$$

with state variables $\mathbf{x} = \{x,y,z\}$, model parameters $\theta = \{\alpha,\beta,\rho\}$, and stochastic forcing $\omega_{(\cdot)} \sim N(0,\sqrt{\Delta t}\sum_\omega)$. Because of its highly nonlinear nature, the Lorenz model has served as a test bed for a wide variety of data assimilation methods. A reference solution of

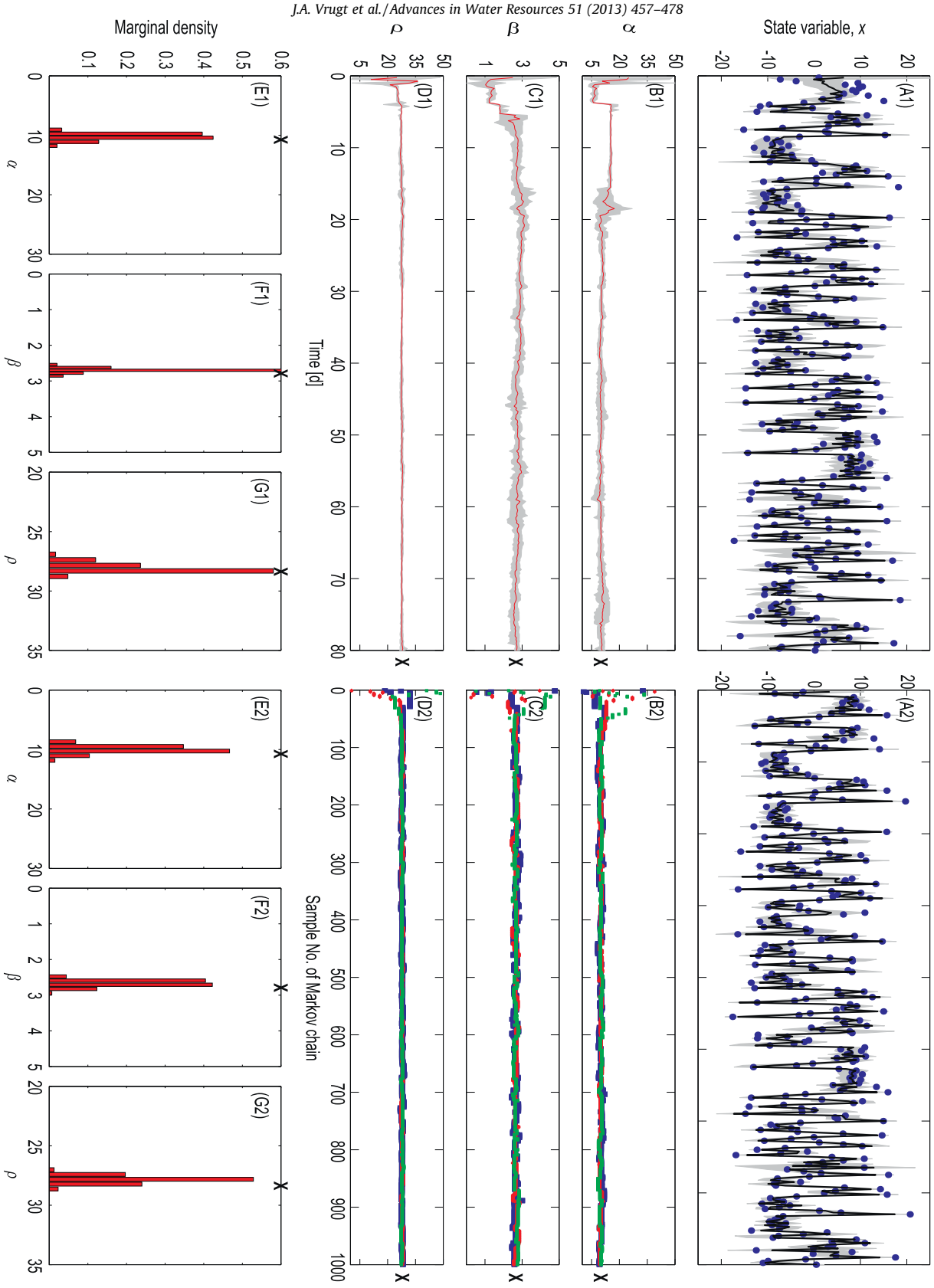

Fig. 3. Joint parameter and state estimation of the highly nonlinear Lorenz model: Results of P-DREAM$_{(VP)}$ (left column) and P-DREAM$_{(IP)}$ (right column) using $P = 100$ particles. The top panel summarizes the results of the state estimation problem, and plots the actual observations of state variable x (solid dots) and corresponding mean ensemble filter predictions (solid black line) and associated 95% prediction uncertainty intervals (gray region) derived with (A1) P-DREAM$_{(VP)}$ and (A2) P-DREAM$_{(IP)}$. The corresponding evolution of the sampled parameter estimates of α, β, and ρ is depicted in the middle three panels. For P-DREAM$_{(VP)}$ these plots (B1–D1) represent the evolving 95% uncertainty ranges of the marginal posterior distribution; the solid red line represents the ensemble mean. For P-DREAM$_{(IP)}$ these plots depict traces of the sampled parameter values in three randomly selected Markov chains, each coded with a different color and symbol. The bottom panel plots the marginal posterior parameter distributions derived with both filters. The cross symbol is used to denote the 'true' values of the parameters that were used to generate the synthetic state-variable observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$\{x_t, y_t, z_t\}$ with $\Delta t = 0.25$ is computed for $t \in [0,80]$ by solving Eq. (32) from the initial condition $(x_0, y_0, z_0) = (1.508870, -1.531271, 25.46091)$ using fixed values of the parameters $\{\alpha, \beta, \rho\} = (10, 28, 8/3)$, and a 3×3 error covariance matrix, $\sum_\omega$ with diagonal entries (variances) of 2.000, 12.13, and 12.31, respectively, and off-diagonal terms equal to zero. Synthetic observations were subsequently generated by imposing normally distributed noise with zero mean and variance equal to 2 to the reference solution. These settings of the initial solution, measurement and model error, and parameter values are exactly identical to those used previously in the literature [32,79,33]. The final data set thus consists of $\{x_{t1}, y_{t1}, z_{t1}\}, t \in \left[\frac{1}{4}, \frac{1}{4}, 80\right]$ and is used for joint parameter and state estimation. In each successive run of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ the prior state ensemble was generated by perturbing the initial conditions of the reference run with the measurement error of the observations. A uniform prior distribution, $p(\theta)$ of the model parameters was assumed with $\alpha \in [0,50]$, $\beta \in [0,5]$, and $\rho \in [0,50]$.

Fig. 3 compares the results of P-DREAM$_{(VP)}$ (left column) with time variant parameters against those separately derived with P-DREAM$_{(IP)}$ (right column) assuming time invariant values of the model parameters. The top panel (Fig. 3A1–2) summarizes the state estimation problem, and presents time series of observed (blue dots) and mean ensemble predicted (solid black line) values of state variable x of the Lorenz model. The results of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ appear very similar. The posterior mean prediction of both filters accurately tracks the observed state transitions with 95% uncertainty intervals that are statistically coherent and adequately bracket the observations. Similar results (not shown herein) are found for the other two state variables, y and z of the Lorenz model.

The middle three panels (Fig. 3B-D) present trace plots of the sampled α, β and ρ values. The left column depicts the evolution of the P-DREAM$_{(VP)}$ derived marginal posterior distributions. The solid red line tracks the posterior mean of the three different Lorenz parameters, and the gray region denotes the associated 95% uncertainty ranges. The marginal distributions sampled with P-DREAM$_{(VP)}$ rapidly converge to a limiting distribution with posterior uncertainty ranges that appear relatively tight and encompass the actual target parameter values ('x' symbol ar right hand side). The marginal distributions not only settle quickly, but also remain virtually unchanged during the subsequent time steps. This is perhaps to be expected, as a single realization of the parameter values was used to create the synthetic state observations. This finding not only inspires confidence in the ability of P-DREAM$_{(VP)}$ to resolve $p(\theta_{1:t}, \mathbf{x}_{1:t} | \tilde{\mathbf{y}}_{1:t})_{t>1}$ when confronted with highly nonlinear model dynamics, but also demonstrates that the observations contain sufficient information to jointly estimate parameter values and state variables.

The right column (Fig. 3B2-D2) presents the corresponding parameter estimates derived with P-DREAM$_{(IP)}$ assuming constant (time invariant) values of α, β and ρ between $t \in [0,80]$. Each Markov chain is coded with a different color and symbol. The sampled trajectories quickly converge to their target values (cross symbol at right hand side) with posterior uncertainty ranges that match closely to those separately obtained with P-DREAM$_{(VP)}$ in the corresponding panels on the left. The four randomly chosen chains demonstrate an excellent mixing, and a few thousands MCMC samples is sufficient to officially declare converge to a limiting distribution according to the $\widehat{R}$-statistic [39] (not shown herein).

The bottom panel shows histograms of the P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ derived marginal posterior parameter distributions. The histograms are very well defined (tight) and the posterior modes correspond closely with the actual parameter values (indicated with cross symbol) used to generate the calibration data. The marginal distributions sampled with both filters look very similar. This essentially explains the striking resemblance between the predictive uncertainty derived with P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ presented in the top panel.

We now demonstrate the advantages of P-DREAM$_{(VP)}$ over current state-of-the-art joint parameter and state estimation methods. Fig. 4 reports the outcome of this analysis, and compares the RMSE of the mean ensemble prediction error of the states (top panel) and parameters (bottom panel) of P-DREAM$_{(VP)}$ (blue line) against those separately derived with SIR$_{(M)}$ (red line) for five different ensemble sizes (x-axis), involving $P = \{10, 25, 50, 100, 250\}$ different particles, and five different values of the resampling threshold (columns). We use a value of $s = 0.1$ in Eq. (31), which gives better results than the values of $s \in [0.005, 0.025]$ used in recent work by Moradkhani and coworkers [24]. This figure highlights several important findings.

In the first place, note that P-DREAM$_{(VP)}$ consistently receives the best performance. The prediction errors of P-DREAM$_{(VP)}$ are substantially lower than their counterparts derived with $SIR_{(M)}$, irrespective of the ensemble size and value of the resampling threshold. The ability of MCMC simulation with DREAM to maintain adequate particle diversity not only allows for a much better tracking of the state variables, but simultaneously also improves the posterior inference of the model parameters.

A second interesting observation is that the performance of P-DREAM$_{(VP)}$ seems to be rather unaffected by the number of particles used to approximate $p(\theta_{1:t}, \mathbf{x}_{1:t} | \tilde{\mathbf{y}}_{1:t})_{t>1}$. An ensemble size of $P = 50$ particles receives very similar error statistics as those derived with much large values for P. Indeed, a spectacular reduction in prediction error is observed when going from $P = 10 \rightarrow P = 25$ particles, but beyond this point the improvement in skill appears rather marginal. This finding has important computational advantages. The performance of $SIR_{(M)}$, on the contrary, gradually improves with increasing ensemble size. This begs the question of how many particles to use in practice. About $P = 2{,}000$ particles are required with $SIR_{(M)}$ to receive similar error statistics as P-DREAM$_{(VP)}$ with $P = 50$ particles (not shown herein). This again highlights the advantages of particle resampling using MCMC simulation with DREAM.

A third significant and interesting observation is that the performance of both filters appears relatively unaffected by the choice of the threshold parameter r used to decide when to do resampling or not. A value of $r = 0.3$ is sufficient to maintain adequate particle diversity, and ensure a satisfactory performance of both filters. Based on these findings we set $r = 0.5$ or $P^* = P/2$. This threshold is also often used in the literature.

Table 2 investigates the statistical coherency of the prediction intervals derived with P-DREAM$_{(VP)}$ and $SIR_{(M)}$ and lists the coverage (PERCT, %) of the state observations within the one-observation-ahead 95% uncertainty ranges derived with both filters. We use five different ensemble sizes, involving $P = \{10, 25, 50, 100, 250\}$ particles. Listed values denote an average of 25 different trials. The 95% prediction intervals derived with P-DREAM$_{(VP)}$ are clearly superior to those estimated with $SIR_{(M)}$. An ensemble size of $P = 25$ particles with P-DREAM$_{(VP)}$ is sufficient to accurately estimate the model parameters and adequately capture the evolving posterior state distribution. On the contrary, $SIR_{(M)}$ exhibits a rather poor precision and coverage. An ensemble size of at least $P = 1{,}000$ particles is required with $SIR_{(M)}$ to obtain prediction uncertainty ranges that can be considered statistically coherent (not shown herein). This ensemble size is 40 times larger than needed for P-DREAM$_{(VP)}$. The variations in coverage among the different $SIR_{(M)}$ trials also appears rather large. This is obviously not desirable. Altogether, these findings again highlight the ability of MCMC simulation with DREAM to maintain adequate particle diversity, and significantly reduce the size of the ensemble required to jointly infer the model parameters and state variables.

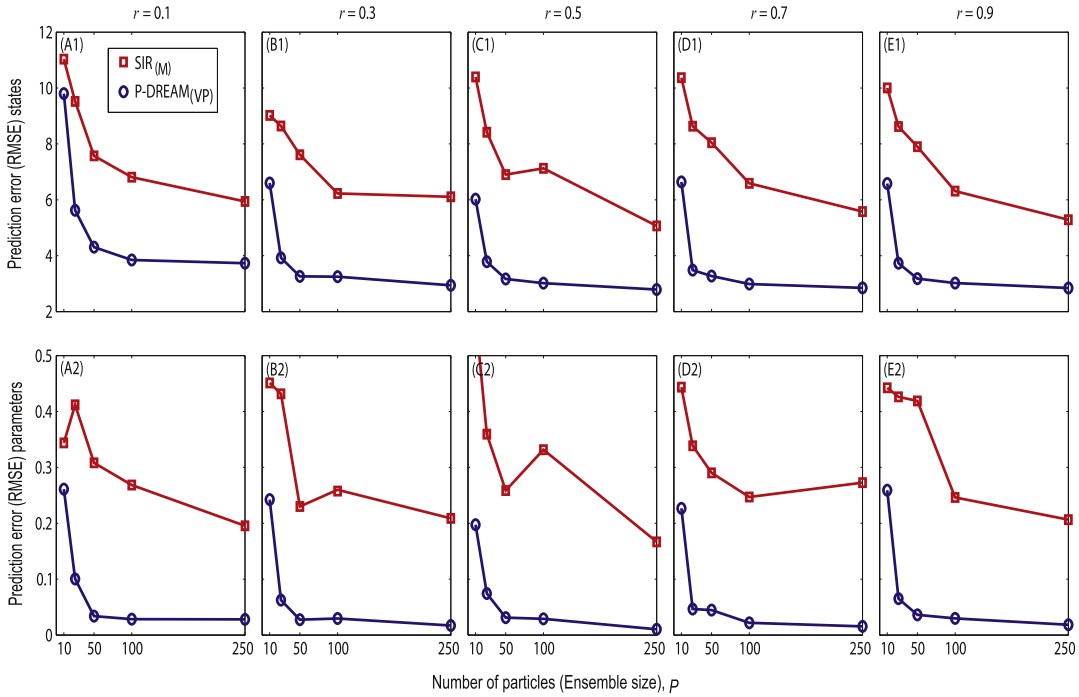

Fig. 4. Joint parameter and state estimation of the highly nonlinear Lorenz model: mean RMSE of the one-observation ahead forecast error of state variables x, y, and z (top panel) and parameters α, β and ρ (bottom panel) derived with P-DREAM$_{(VP)}$ (blue line) and SIR$_{(M)}$ (red line) using an ensemble size consisting of $P = \{10,25,50,100,250\}$ particles. Each column considers a different value of r, the threshold parameter that determines when to do resampling or not. Each individual value represents an average of 25 different trials. Note that P-DREAM$_{(VP)}$ receives much better performance than SIR$_{(M)}$ and requires far fewer particles to reach equivalent prediction errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Lorenz model: percentage of points contained in the joint 95% prediction uncertainty intervals of $\{x,y,z\}$ derived with P-DREAM$_{(VP)}$ and SIR$_{(M)}$. We separately consider five different values of the ensemble size and assume a resampling threshold of $r = 0.5$ or $P/2$. Listed values represent an average of 25 different trials. Standard deviations appear in italic.

	P-DREAM$_{(VP)}$ PERCT, %	SIR$_{(M)}$ PERCT, %
$P = 10$	66.5 (*5.7*)	39.2 (*21.1*)
$P = 25$	91.1 (*4.4*)	59.3 (*17.9*)
$P = 50$	96.1 (*2.2*)	65.9 (*17.6*)
$P = 100$	96.3 (*2.1*)	77.8 (*16.6*)
$P = 250$	95.7 (*1.8*)	84.0 (*13.4*)

The results of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ presented thus far consider a single MCMC resampling step each time particle diversity, ESS has dropped below the threshold value, P^*. This is computationally most efficient, but not necessarily optimal. Each additional and successive MCMC resampling steps might further rejuvenate the particle ensemble, and systematically improve the performance of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$. This would seem particularly useful in situations when the importance density, $q_\theta(\cdot|\tilde{y}_t, \{\mathbf{x}_t\})$ poorly describes the actual state transitions, and the observations are not too informative. Fig. 5 displays the relationship between the number of successive MCMC resampling steps with DREAM and the corresponding marginal likelihood, $p(\tilde{\mathbf{y}}_{1:n})$ or $p_\theta(\tilde{\mathbf{y}}_{1:n})$ derived with Particle-DREAM (left plot) and P-DREAM$_{(VP)}$ (right plot). We consider five different ensemble sizes, involving $P = \{10,25,50,100,250\}$ particles. For Particle-DREAM we assume *a priori* knowledge of the optimal parameter values. The most important results are as follows.

In the first place, notice that the predictive skill of P-DREAM$_{(VP)}$ significantly improves during the first resampling step with DREAM. The marginal likelihood has considerably increased, but additional MCMC steps appear counter productive. This is perhaps rather unexpected but highlights a subtle finding that has important ramifications. The marginal likelihood deteriorates because of parameter drift. This becomes more and more likely with increasing number of successive MCMC resampling steps. If the candidate states deviate too much from their current values then this is directly penalized for in the acceptance ratio as the proposed states will exhibit poor predictive performance. Unfortunately, this is not the case with the parameter values. A bad parameter proposal can still be accepted during MCMC resampling because the length of the trajectory used to compare the original (current) particles and their candidate points is simply too short (Fig. 2) to accurately differentiate between "good" and "bad" parameter values. This happens primarily at times when the model operator (transition prior) is insensitive to the parameters. The resulting parameter drift affects, sometimes dramatically, the predictive capabilities of the filter.

Parameter divergence is also observed from the statistical properties of the particle ensemble after each successive MCMC resampling step. The dispersion of the parameters remains quite similar during the first five consecutive DREAM trials, but then steadily increases with each additional resampling step. Unfortunately, this parameter drift cannot be resolved. This behavior is simply due to temporal variations in parameter sensitivity, and thus an artefact of the actual importance density (model operator) that is being used. A simple remedy to the problem of parameter drift would be to increase the lag L beyond one time step, and use a much larger block of historical observations during resampling. This should increase parameter sensitivity, and the resulting filter will become some joint variant of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$.

In the second place, note that the marginal likelihood derived with P-DREAM$_{(IP)}$ appears rather unaffected by particle resampling

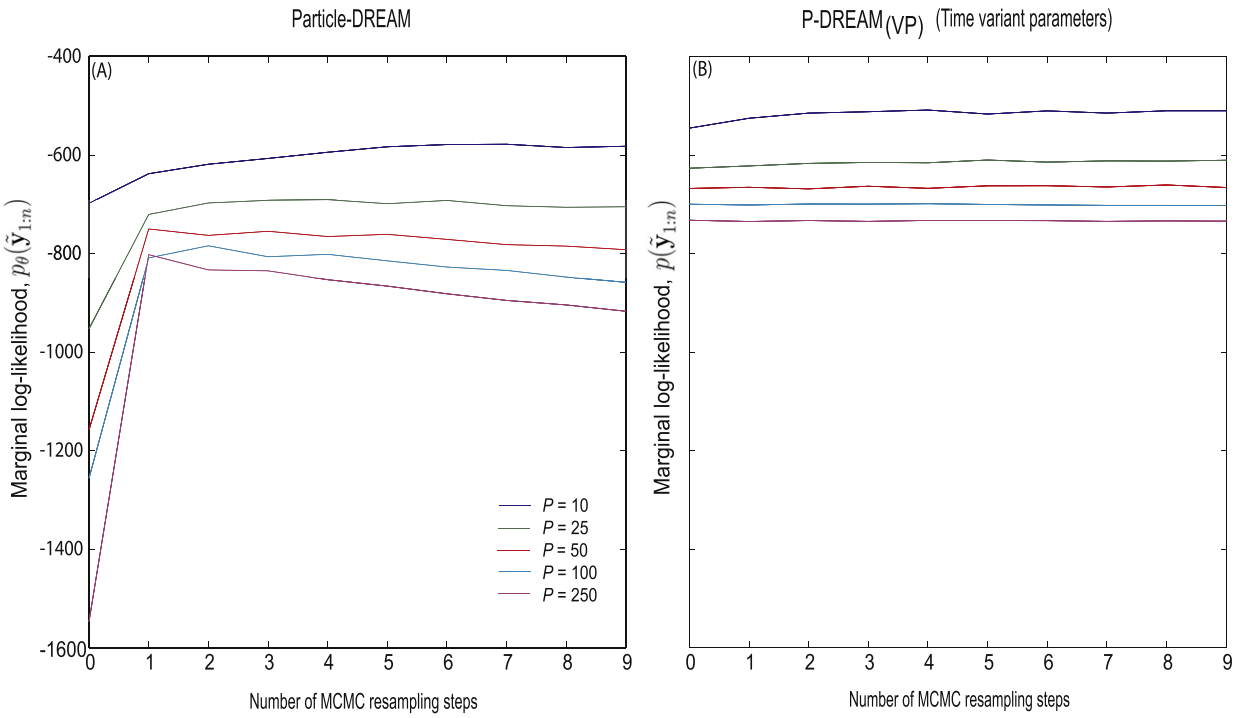

Fig. 5. Highly nonlinear Lorenz model: the effect of additional resampling steps with MCMC on the marginal likelihood derived with (A1) Particle-DREAM (left), and (A2) P-DREAM$_{(VP)}$ (right) using five different ensemble sizes (color coded), consisting of $P = \{10, 25, 50, 100, 250\}$ particles. Each individual line represents an average of 25 different trials. The results of Particle-DREAM were derived assuming a-prior knowledge of the optimal parameter values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with MCMC simulation. The model dynamics adequately track the observational data and do not require explicit resampling steps with DREAM to maintain adequate particle diversity. This finding provides strong support for our claim that parameter drift is the main cause for the decline in predictive performance of P-DREAM$_{(VP)}$ with increasing number of MCMC resampling steps. We will revisit this subject in the next case study using rainfall-discharge modeling.

Finally, the magnitude of the marginal likelihood depends on the ensemble size used to resolve the observed state dynamics. This directly follows from Eq. (23). In the limit of $P \to \infty$ the marginal likelihood will converge to its true value, yet in practice such large ensemble sizes are not required to obtain stable results. Altogether, the results presented herein suggest that a single MCMC step with DREAM is sufficient. Additional resampling steps beyond this point appear less productive, and only introduce potential problems with parameter drift.

4.2. Case study II: conceptual hydrologic modeling

We now move onto the second case study, and illustrate the usefulness and applicability of P-DREAM and P-DREAM$_{(IP)}$ for hydrologic prediction and uncertainty analysis using the HyMOD conceptual watershed model. This model, originally developed by Boyle [13], belongs to the class of probability-distributed soil-store models [83] and transforms rainfall and potential evaporation into discharge emanating from the catchment outlet. A schematic overview of this model appears in Fig. 6. The model consists of a relatively simple rainfall excess store that routes water to a groundwater reservoir (slow-flow tank), in parallel to a surface/subsurface reservoir consisting of a series of three quick-flow tanks with identical residence times. Water storage dynamics are based on mass balance principles using inflow from rainfall, and losses to evaporation, recharge of groundwater, interflow, and direct runoff. Computational details are given by Moore [84].

The model contains five parameters that require calibration against measured streamflow data: C_{max} (L) the maximum storage capacity of the catchment, b_{exp} (–) the degree of spatial variability of soil moisture capacity within the catchment, *Alpha* (–) the distribution factor of flow between the two series of reservoirs, and R_s (days) and R_q (days), the residence times of the linear slow and quick flow reservoirs, respectively. Table 3 lists the upper and lower bound of these respective parameters that together define the prior uncertainty ranges.

We consider two different streamflow time series in our analysis. The first study serves as a benchmark experiment of P-DREAM and P-DREAM$_{(IP)}$ and uses a three year record of artificial streamflow observations simulated with HyMOD using measured forcing data (28 July 1952–28 July 1955) from the Leaf River watershed in Mississippi (1944 km^2) and parameter values given by Moradkani et al. [85]. The synthetic discharge data are corrupted with a heteroscedastic measurement error, the standard deviation of which was taken to be one-tenth of the actual simulated values. The second study explores the performance of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ using the actual basin measured streamflow data.

In all of our numerical experiments, the initial states were created by sampling from a multi-normal prior distribution, $p(\mathbf{x}_0) \sim N(\mu_{\mathbf{x}_0}, \sum_{\mathbf{x}_0})$ with mean, $\mu_{\mathbf{x}_0}$ and covariance matrix $\sum_{\mathbf{x}_0}$, derived separately from MCMC simulation with DREAM using the first few discharge observations. This has some practical advantages as it removes the need for a substantial burn-in period and hence allows immediate tracking of the discharge observations. Note that Moradkhani et al. [85] and Salamon and Feyen [97] use a noninformative (uniform) prior state distribution. The P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ code developed herein provide similar results for such overdispersed prior, but this requires a longer burn-in period (not shown herein). The model error was assumed

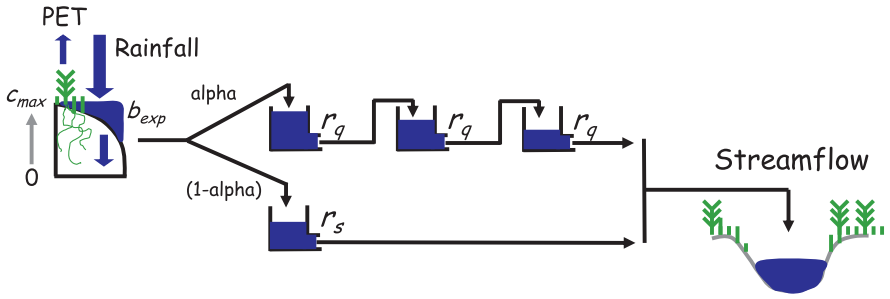

Fig. 6. Schematic representation of the HyMOD conceptual watershed model. The model consists of five different reservoirs (states) that jointly describe the rainfall–runoff transformation using five different parameters that require calibration against a historical record of discharge data.

to be heteroscedastic with standard deviation equal to one-fifth of the previous discharge observation, or in mathematical notation, $\omega_t^{\text{output}} \sim N(0, \frac{1}{5}\tilde{y}_{t-1})$. This model error in the discharge space was subsequently transformed to the state space, $\omega_t^{\text{states}} = \hbar^{-1}(\omega_t^{\text{output}})$ using analytical inversion of the measurement operator in Eq. (2). This state error serves as the stochastic forcing in Eq. (1). We deliberately use the observed discharge at $t-1$ to reflect real-time conditions.

Fig. 7 illustrates the results of P-DREAM$_{(VP)}$ for the artificial discharge observations. The top two graphs (Fig. 7A–B) summarize the streamflow dynamics and depict the measured (blue dots), and mean ensemble predicted (solid black line) streamflow values, the associated 95% prediction uncertainty intervals (gray region) and corresponding residuals. The bottom three graphs (Fig. 7C–E) depict the evolution of the sampled marginal posterior parameters distribution of the HyMOD parameters C_{max}, *Alpha*, and R_q. The red line depicts the trajectory of the posterior mean, and the gray region represents the associated 95% uncertainty intervals. The 'x' symbol at the right hand side of these three graphs depict the target values of the parameters used to create the synthetic streamflow data.

The mean ensemble streamflow prediction closely tracks the observed discharge data, with 95% uncertainty ranges that are statistically coherent (not shown herein) and adequately bracket the observations. The average quadratic forecast error of the posterior mean prediction corresponds closely with the size of the measurement and model error used to create the synthetic discharge data. This highlights the ability of P-DREAM$_{(VP)}$ to adequately track the observed discharge dynamics. The sampled marginal parameter distributions exhibit considerable uncertainty and move intermittently throughout the feasible parameter space. A few hundred discharge observations are deemed sufficient to estimate the quick flow recession parameter, R_q, but the other parameters remain poorly identified. The meandering of C_{max} and *Alpha* is perhaps rather unexpected, but an immediate consequence of the model operator used to transition the states from one time step to the next. During periods with low parameter sensitivity, the width of the marginal distributions is allowed to grow significantly. This is perhaps best illustrated between days 400–500 of the discharge record. These long recession periods are essentially driven by the baseflow parameter R_s, and the other HyMOD parameters diverge from their actual values used to create the synthetic streamflow data. Unfortunately, there is little that we can do about this. The length of the resampled trajectory is simply insufficient to ensure adequate parameter sensitivity, and avoid problems with parameter drift. Block resampling with a larger historical trajectory ($L \gg 1$) is required to increase parameter sensitivity and resolve divergence problems. Yet, this is beyond the scope of the present paper. Altogether, the findings presented herein caution against the use of a single discharge observation for sequential parameter and state estimation, and raise concern about the conclusions of some previous contributions [85,97].

Table 3
Prior uncertainty ranges and description of the HyMOD parameters.

Parameter	Description	Minimum	Maximum	Unit
C_{max}	Maximum water storage of the watershed	10.00	1000.00	mm
b_{exp}	Spatial variability of the soil moisture capacity	0.10	2.00	
Alpha	Distribution factor between fast and slow flow tank	0.01	0.99	
R_s	Residence time of slow flow reservoir	0.001	0.10	days^{-1}
R_q	Residence time of quick flow reservoir	0.10	1.00	days^{-1}

We now illustrate and discuss the results of P-DREAM$_{rm(IP)}$ for the same model and data set. Fig. 8 summarizes the outcome of this analysis. The top two panels (Fig. 8A–B) summarize the model performance and plot the measured (red dots) and filter predicted (black line) discharge observations and error residuals. The gray region represents the 95% prediction uncertainty ranges derived from the particle ensemble. The bottom three graphs depict trace plots of the DREAM generated parameter samples in three randomly selected Markov chains, each coded with a different color and symbol. The cross symbols at the right hand side of these three different panels denote the target parameter values.

The mean ensemble discharge prediction closely matches the synthetic streamflow observations with associated prediction intervals and residuals that are consistent with the assumed model and measurement error and similar in size as those previously derived with P-DREAM$_{(IP)}$ using time-variant parameters. The error of the mean discharge prediction is consistent with the assumed measurement and model error, and highlights that the filter has converged appropriately. The different Markov chains closely group around the actual parameter values used to create the artificial streamflow data, and exhibit an excellent mixing. A few hundred samples in each individual chain are sufficient to converge to a limiting distribution. This highlights the efficiency of MCMC simulation with DREAM. Note that the different parameters are very well defined, and display very little uncertainty. Altogether these results inspire confidence in the ability of P-DREAM$_{(IP)}$ to resolve the evolving state distribution and model parameters, and successfully approximate $p_{\theta^j}(\mathbf{x}_{1:n}|\tilde{\mathbf{y}}_{1:n})$.

Fig. 9 presents histograms of the posterior marginal parameter distributions derived with P-DREAM$_{(VP)}$ (top panel: green) and P-DREAM$_{(IP)}$ (bottom panel: red) after processing all the discharge data. Each column across the graph considers a different HyMOD parameter. For convenience, both panels use a common *x*-and *y*-axis. This enables a direct comparison between the marginal distri-

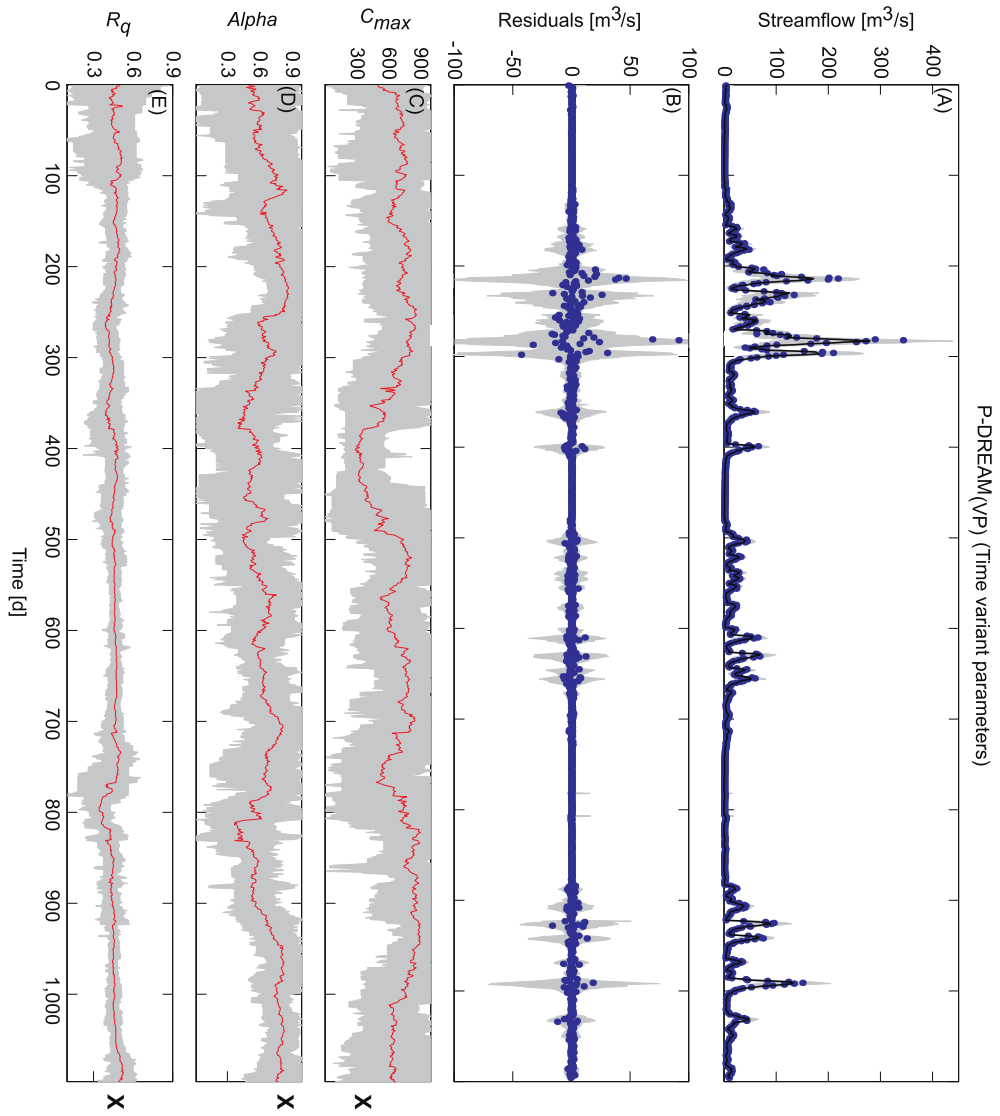

Fig. 7. Joint parameter and state estimation of the HyMOD conceptual watershed model using synthetic streamflow observations: Results of P-DREAM$_{(VP)}$ assuming time variant values of the model parameters. The top two panels (A–B) summarize the calibration data, and compare the measured discharge dynamics (blue dots) against the mean ensemble predictions (solid black line) of the filter. The gray region signifies the 95% prediction uncertainty ranges. The bottom three panels (C–E) depict the evolution of the 95% uncertainty ranges of the sampled marginal distributions of C_{max}, *Alpha*, and R_q. The red line represents the mean ensemble prediction, and the cross symbols at the right hand side depict the actual parameter values used to generate the synthetic streamflow data. The error deviation of the measurement and model error was taken to be 10% and 20% of the simulated discharge data, respectively. The marginal posterior parameter distributions express considerable uncertainty and continue to meander intermittently around the actual parameter values used to generate the synthetic data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

bution derived with both filters. The most important results are as follows.

Most of the histograms deviate considerable from a normal distribution. The posterior distributions sampled with P-DREAM$_{(VP)}$ extend over a large part of the prior defined parameter ranges (e.g. Table 3). Temporal variations in parameter sensitivity causes the marginal distributions to disperse from their actual values used to generate the artificial discharge data. This finding essentially demonstrates that individual discharge observations contain insufficient information to warrant joint inference of the state variables and model parameters. Indeed, if we assume the parameter values to be time invariant (bottom panel), then the marginal distributions become much better defined, and center closely around their respective target values.

To provide further insights into the performance of P-DREAM$_{(VP)}$, please consider Fig. 10 that presents trace plots of the effective sample size, ESS (top graph) and acceptance rate of candidate particles during MCMC resampling. Each time ESS drops below the threshold value of $P^* = P/2 = 125$ (dashed black line) particle resampling is necessary. Both diagnostics exhibit considerable temporal variation, and vary dynamically from one time step (discharge observation) to the next. The effective sample size exhibits the largest dynamics and oscillates between values of ESS $\in$ [30,200]. This diversity is sufficient to guarantee adequate filter performance. The acceptance rate, on the contrary, varies less dynamically and displays a more cyclic behavior that mimics the actual streamflow dynamics. On average, about one-sixth of the proposals is being accepted, which is theoretically nearly optimal [95].

We next run P-DREAM and P-DREAM$_{(IP)}$ using historical streamflow observations from the Leaf River watershed during the same data period (28 July 1952 – 28 July 1955). Fig. 11 summarizes the outcome of this analysis using an ensemble size of $P = 250$ particles. The results are very similar to those obtained previously dis-

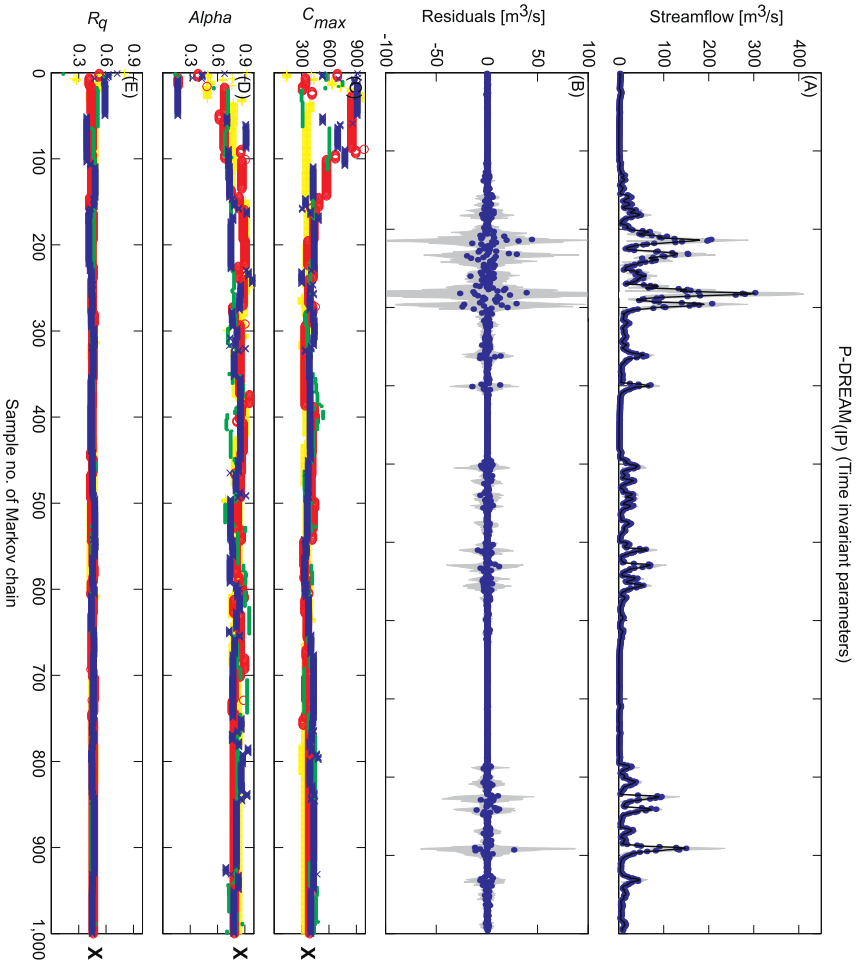

Fig. 8. Joint parameter and state estimation of the HyMOD conceptual watershed model using synthetic streamflow data: results of P-DREAM$_{(IP)}$ assuming time invariant values of the model parameters. The top two panels (A-B) summarize the calibration data, and compare the measured discharge dynamics (blue dots) against the mean ensemble predictions (solid black line) of the filter. The gray region signifies the 95% prediction uncertainty ranges. The bottom three panels (C–E) present trace plots of the sampled parameter values in three randomly selected Markov chains, each coded with a different symbol and color. The cross symbol at the right hand side indicate the actual parameter values used to generate the synthetic discharge data. The different chains mix very well and quickly locate the target parameter values. A few hundred samples appears sufficient to officially declare convergence to a limiting distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cussed for the synthetic discharge data. The top panels illustrates that P-DREAM$_{(VP)}$ (left column) and P-DREAM$_{(IP)}$ (right column) do a good job in tracking the streamflow dynamics, with 95% prediction uncertainty ranges (gray regions) that are statistically adequate and encompass about 96% of the discharge observations. The RMSE of the one-day-ahead forecast error of both filters is about 16 m^3/s, significantly lower than its value of approximately 22.5 m^3/s separately derived with HyMOD using a conventional batch calibration approach. This performance improvement is due to state variable updating.

The evolving marginal distributions sampled with P-DREAM$_{(VP)}$ exhibit considerable nonstationarity, and continue to move up and down the feasible parameter space. Posterior parameter uncertainty especially increases during periods with low parameter sensitivity. This is most evident between days 700–900 of the historical record, when the hydrograph is essentially driven by the baseflow parameter, R_s. It is comforting to observe that the modes of the marginal parameter distribution sampled with P-DREAM$_{(VP)}$ at $t = 1096$ match closely with the respective counterparts estimated with P-DREAM$_{(IP)}$ using the marginal likelihood of the entire historical record. Note that a much smaller ensemble size of $P = 50$ was also deemed sufficient for the present case study. This concludes the simulation experiments of our paper.

5. Some final remarks

The P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ approaches developed herein exhibit difficulty to recuperate quickly if the particle ensemble consistently over or underestimates the respective observation. This bias is very difficult to remove in practice, unless a prohibitively large number of successive MCMC resampling steps is used (see L96 problem). This is rather inefficient and also significantly increases computational costs. To resolve this problem we draw inspiration from the original Kalman Filter [54] and introduce a second and alternative proposal distribution in P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ that uses explicitly the distance of the ith particle to the current observation:

$$\left\{\mathbf{Z}_t^i\right\} = \left\{\mathbf{X}_t^i\right\} \pm \hbar^{-1}\left[\tilde{y}_t - \left(h(\left\{\mathbf{X}_t^i\right\}, \phi) + v_t^i\right)\right] + \omega_t^i. \tag{33}$$

This analysis step updates the particles in the direction of the respective observations. The ±sign is used to enforce jump symmetry and satisfy detailed balance. Future work will consider this alternative proposal distribution.

We decided not to compare the results of P-DREAM$_{(IP)}$ and P-DREAM$_{(VP)}$ against those of a classical least-squares fitting method. Such juxtaposition has already been reported in a previous publi-

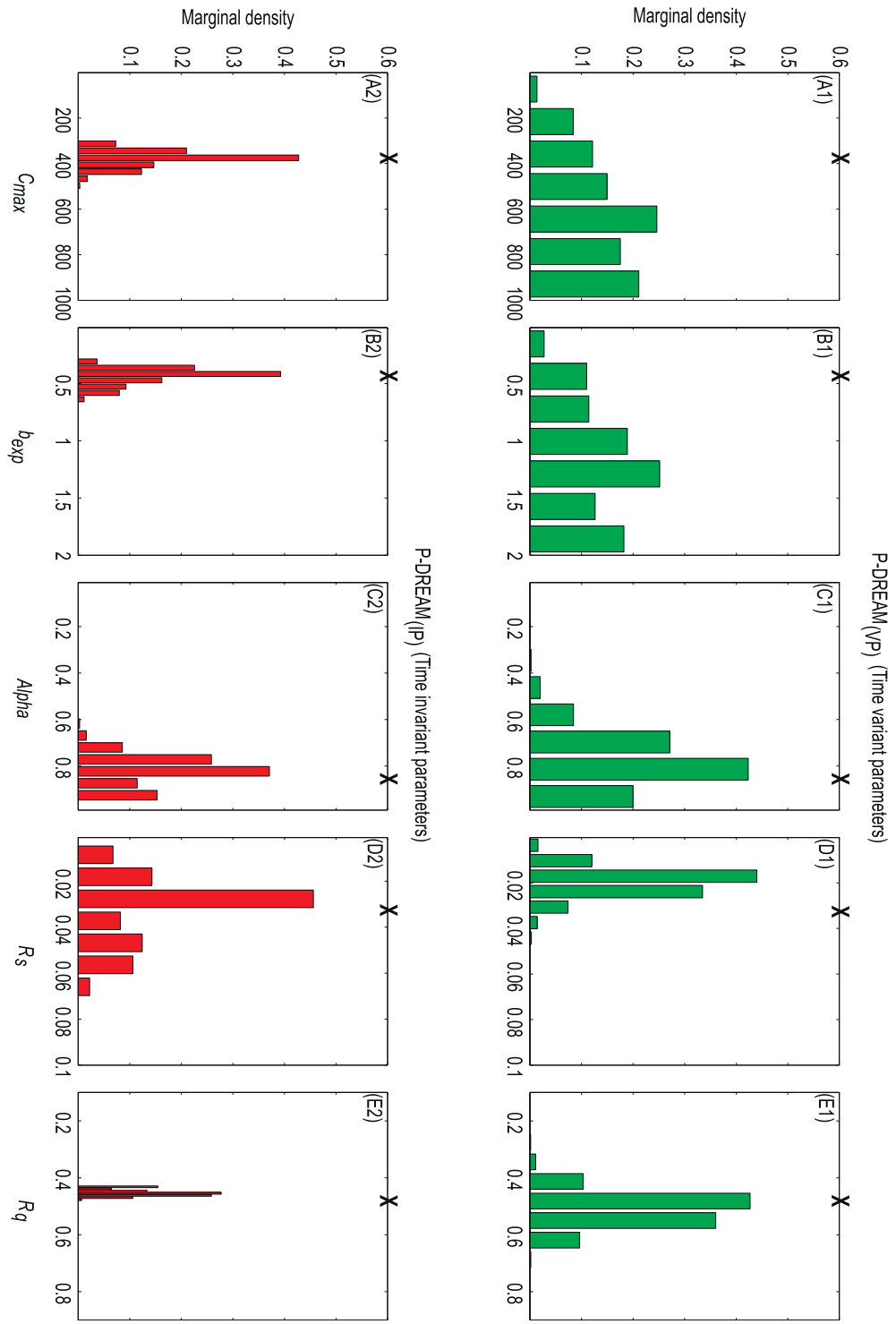

Fig. 9. Joint parameter and state estimation of the HyMOD conceptual watershed model using synthetic streamflow data: Histograms of the marginal posterior parameter distributions of (A1–A2) C_{max}, (B1–B2), b_{exp}, (C1–C2), *Alpha*, (D1–D2) R_s and (E1–E2), R_q derived with P-DREAM$_{(VP)}$ (1: top row) and P-DREAM$_{(IP)}$ (2: bottom row) at the end of simulation. The actual parameter values used to create the artificial streamflow data are separately indicated in each individual graph using the cross symbol.

cation [121] and has convincingly demonstrated that explicit treatment of model and input data error via joint parameter and state estimation significantly alters the posterior distribution of the HyMOD parameters. Findings presented in Clark and Vrugt [19] using a parsimonious snow model and data from the Lake Eldora SNOTEL site in Colorado, USA actually suggest that the posterior parameter distributions derived with methods such as SODA (and thus P-DREAM$_{(IP)}$ are physically more realistic. This is because the parameters are no longer compensating for model structural deficiencies and errors in the forcing data.

Finally, the jump-factor, $\gamma(\tau, \lambda') = 2.4/\sqrt{2\tau\lambda'}$ in Eq. (27) is considered theoretically optimal for batch sampling methods [95]. Unfortunately, this choice for γ might not necessarily be optimal for sequential sampling schemes. Our future work will investigate whether a 23% acceptance rate can still be considered optimal for the performance of Particle-DREAM, P-DREAM$_{(IP)}$ and P-DREAM$_{(VP)}$.

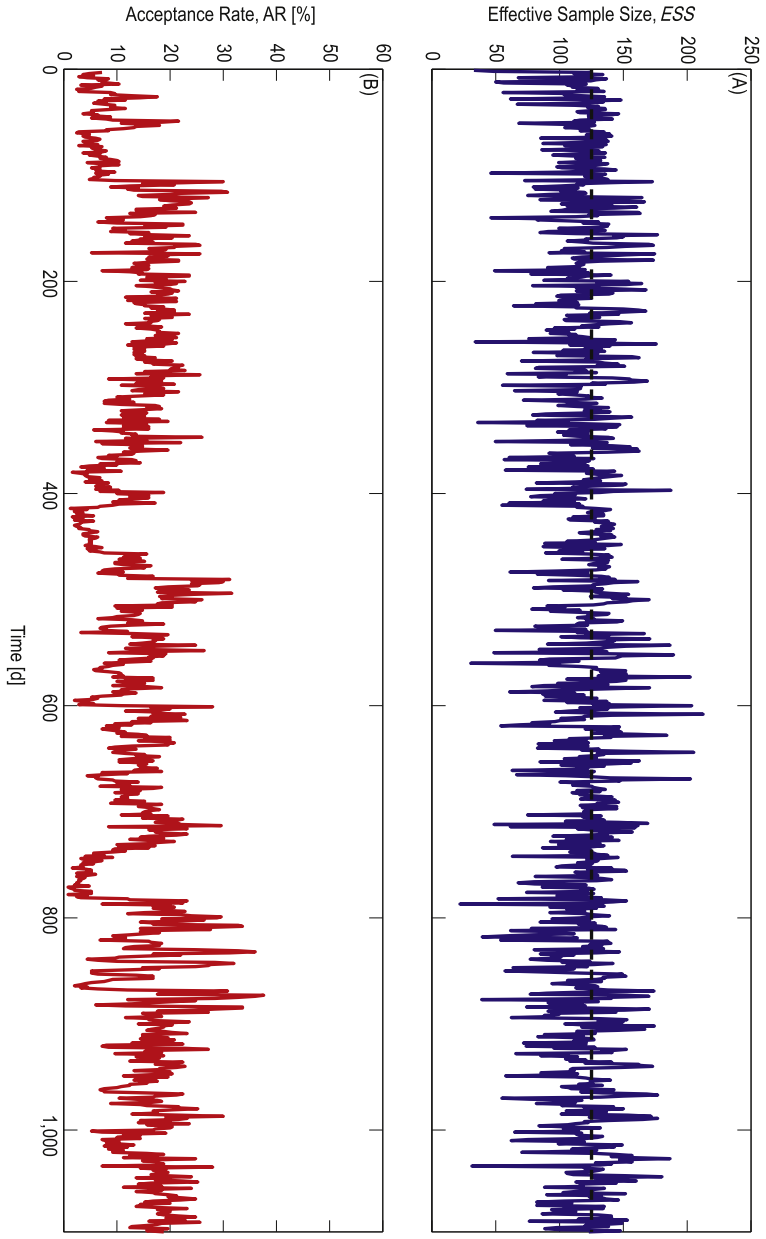

Fig. 10. Joint parameter and state estimation of the HyMOD conceptual watershed model using synthetic streamflow data: Traces of the effective sample size (top row, blue) and acceptance rate (bottom row, red) during filtering with P-DREAM$_{(VP)}$. The dotted black line in the top panel represents the threshold value, $r = 0.5$ used to determine when to do particle resampling or not. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Results and conclusions

This study has introduced the theory and concepts of Particle-DREAM, a novel data assimilation methodology for inference of state variables and/or model parameters. This new particle filter combines the strengths of sequential Monte Carlo sampling and Markov chain Monte Carlo simulation to resolve well-documented problems with sample impoverishment. Two different variants of Particle-DREAM were presented to satisfy assumptions regarding the temporal behavior of the model parameters.

Three preliminary case studies were used to demonstrate the usefulness and applicability of particle MCMC simulation. The first study used the Lorenz 96 atmospheric "toy" model and convincingly demonstrated that MCMC resampling with DREAM ensures adequate particle diversity. Particle-DREAM closely tracked the highly nonlinear dynamics of the 40-dimensional state space using just $P = 50$ particles. A classical SIR filter requires at least five times as many particles to achieve a similar level of performance.

The second case study considered the Lorenz attractor and demonstrated the ability of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ to jointly infer the model parameter and state variables when confronted with highly nonlinear model dynamics. This study also demonstrated that both filters require far fewer particles than classical SIR particle filters to work well in practice. The MCMC resampling step with DREAM promotes particle diversity, and reduces problems with filter degeneracy.

The third case study explored the usefulness of P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ by application to rainfall-discharge modeling using the HyMOD conceptual watershed model and historical streamflow data from the Leaf River watershed in Mississippi. Both filters considerably improve the consistency between the model predictions and calibration data with prediction intervals

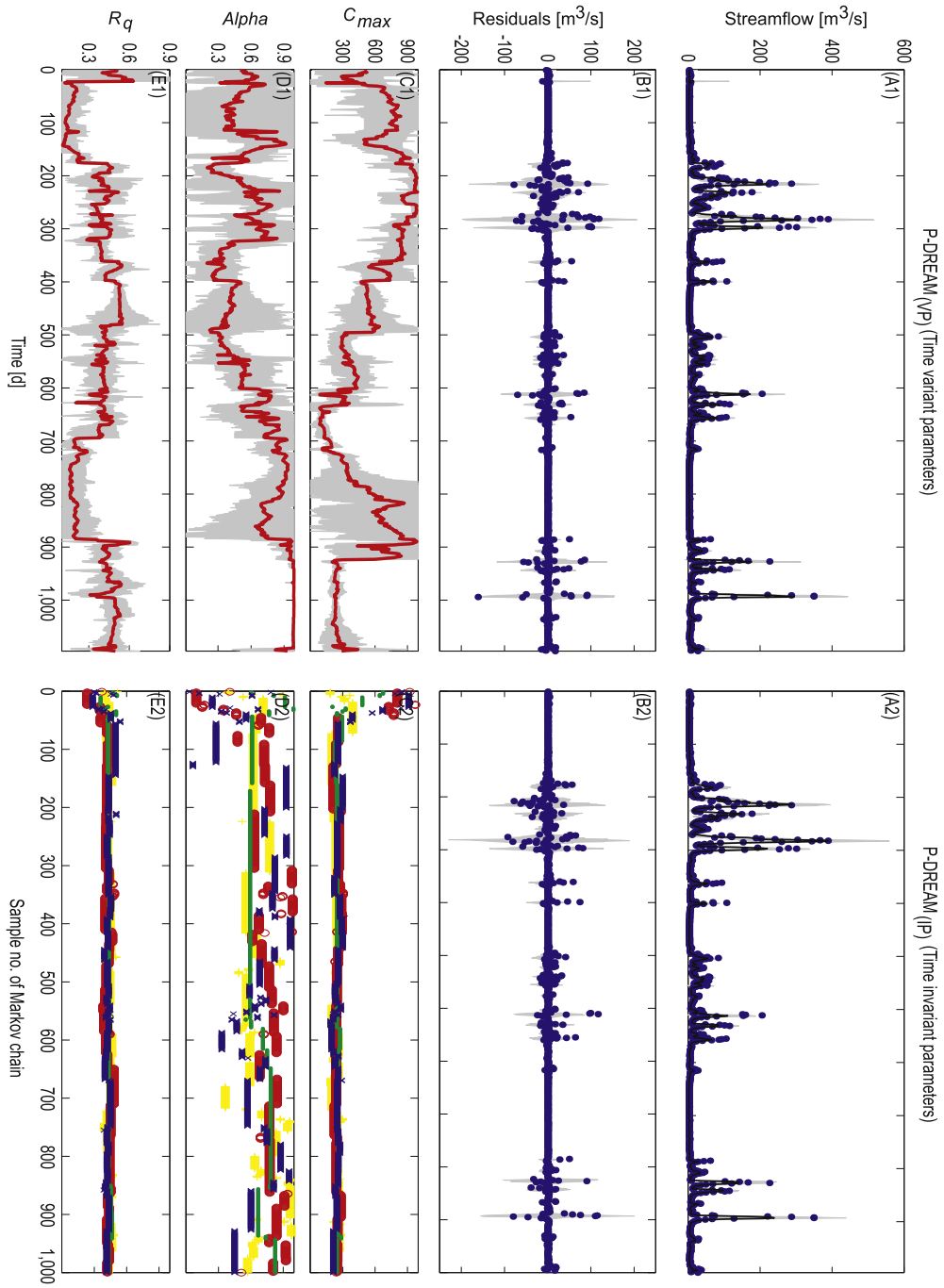

Fig. 11. Joint parameter and state estimation for the parsimonious HyMOD conceptual watershed model using discharge data from the Leaf River watershed in Mississippi: Results of P-DREAM$_{(VP)}$ (left column) and P-DREAM$_{(IP)}$ (right column) using $P = 250$ particles. The top panel plots the actual discharge data (solid dots) and corresponding mean ensemble model predictions (solid black line) and associated 95% prediction uncertainty intervals (gray region). To provide better insights into the model-data mismatch, the second panel (B1–B2) presents the error residuals of the mean ensemble filter prediction derived with both filter. Finally, the bottom three panels, depict the evolution of the sampled parameter estimates of C_{max}, *Alpha*, and R_q. For P-DREAM$_{(VP)}$ these plots (C1–E1) represent the evolving 95% uncertainty ranges of the marginal posterior distribution; the solid red line represents the ensemble mean. For P-DREAM$_{(IP)}$ these plots (C2–E2) depict traces of the sampled parameter values in three randomly selected Markov chains, each coded with a different color and symbol. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

that are approximately similar and statistically coherent. The P-DREAM$_{(VP)}$ sampled marginal posterior distributions converge adequately to the appropriate target values but demonstrate considerable non-stationarity. This finding, caused by temporal changes in parameter sensitivity, contradicts results from previous studies. Future work will provide remedies to resolve problems with parameter drift by using a much longer historical trajectory of the observations. This method is very similar to smoothing, but maintains the advantages of particle resampling with MCMC simulation.

In this paper, we have selected a formal Bayes implementation with a relatively simple Gaussian error criterion, yet Particle-DREAM, P-DREAM$_{(VP)}$ and P-DREAM$_{(IP)}$ readily accommodate informal likelihood functions or signature indices if those better represent the salient features of the data and simulation model used.

Acknowledgments

The work of Jasper A. Vrugt has been made possible through funding from the Department of Energy, Office of Biological and Environmental Research, and a J. Robert Oppenheimer Fellowship of the Los Alamos National Laboratory postdoctoral program. Computer support, provided by the SARA center for parallel computing at the University of Amsterdam, The Netherlands, is highly appreciated.

Appendix A. Derivation of the recursive form of Bayes law

The unnormalized posterior distribution $p_\theta(\mathbf{x}_{1:t}, \tilde{\mathbf{y}}_{1:t})$ at any time $t > 1$ satisfies

$$p_\theta(\mathbf{x}_{1:t}, \tilde{\mathbf{y}}_{1:t}) = p_\theta(\mathbf{x}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t) \tag{34}$$

with normalized posterior distribution, $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ that can be written as follows:

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \frac{p_\theta(\mathbf{x}_{1:t}, \tilde{\mathbf{y}}_{1:t})}{p_\theta(\tilde{\mathbf{y}}_{1:t})}, \tag{35}$$

where $p_\theta(\tilde{\mathbf{y}}_{1:t})$ signifies the normalization constant. If we combine Eqs. (34) and (35) we derive:

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \frac{p_\theta(\mathbf{x}_{1:t-1}, \tilde{\mathbf{y}}_{1:t-1}) f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t)}{p_\theta(\tilde{\mathbf{y}}_{1:t})} \tag{36}$$

which is equal to:

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = \frac{p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1}) p_\theta(\tilde{\mathbf{y}}_{1:t-1}) f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t)}{p_\theta(\tilde{\mathbf{y}}_{1:t})}, \tag{37}$$

and with

$$p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1}) = \frac{p_\theta(\tilde{\mathbf{y}}_{1:t})}{p_\theta(\tilde{\mathbf{y}}_{1:t-1})}, \tag{38}$$

results in the following recursion of $p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t})$ presented in Eq. (11) of the paper [30]:

$$p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) = p_\theta(\mathbf{x}_{1:t-1}|\tilde{\mathbf{y}}_{1:t-1}) \frac{f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) L_\theta(\tilde{y}_t|\mathbf{x}_t)}{p_\theta(\tilde{y}_t|\tilde{\mathbf{y}}_{1:t-1})}. \tag{39}$$

Appendix B. Derivation of Metropolis acceptance probability

The jump probability from $\left\{\mathbf{X}_t^i\right\}$ to $\left\{\mathbf{Z}_t^i\right\}$ is:

$$J_\theta\left(\left\{\mathbf{X}_t^i\right\} \to \left\{\mathbf{Z}_t^i\right\}\right) = J_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\} \to \left\{\mathbf{Z}_{t-1}^i\right\}\right) f_\theta\left(\left\{\mathbf{Z}_t^i\right\} \middle| \left\{\mathbf{Z}_{t-1}^i\right\}\right) \tag{40}$$

with probability of the reverse jump:

$$J_\theta\left(\left\{\mathbf{Z}_t^i\right\} \to \left\{\mathbf{X}_t^i\right\}\right) = J_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\} \to \left\{\mathbf{X}_{t-1}^i\right\}\right) f_\theta\left(\left\{\mathbf{X}_t^i\right\} \middle| \left\{\mathbf{X}_{t-1}^i\right\}\right). \tag{41}$$

The transition kernel of DREAM is symmetric, and thus $J_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\} \to \left\{\mathbf{Z}_{t-1}^i\right\}\right) = J_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\} \to \left\{\mathbf{X}_{t-1}^i\right\}\right)$.

The target distribution of the filter is [30]:

$$\begin{aligned} p_\theta(\mathbf{x}_{1:t}|\tilde{\mathbf{y}}_{1:t}) &\propto p(\mathbf{x}_{1:t}) L_\theta(\tilde{\mathbf{y}}_{1:t}|\mathbf{x}_{1:t}) \\ &\propto p(\mathbf{x}_0) p(\mathbf{x}_1|\mathbf{x}_0) \cdots p(\mathbf{x}_t|\mathbf{x}_{t-1}) \prod_{k=1}^{t} L_\theta(\tilde{y}_k|\mathbf{x}_k). \end{aligned} \tag{42}$$

The accept ratio, $\alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\}, \left\{\mathbf{X}_t^i\right\}\right)$ therefore becomes:

$$\begin{aligned} \alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\}, \left\{\mathbf{X}_t^i\right\}\right) &= 1 \wedge \frac{p_\theta\left(\left\{\mathbf{Z}_t^i\right\}|\tilde{\mathbf{y}}_{1:t}\right) J_\theta\left(\left\{\mathbf{Z}_t^i\right\} \to \left\{\mathbf{X}_t^i\right\}\right)}{p_\theta\left(\left\{\mathbf{X}_t^i\right\}|\tilde{\mathbf{y}}_{1:t}\right) J_\theta\left(\left\{\mathbf{X}_t^i\right\} \to \left\{\mathbf{Z}_t^i\right\}\right)} \\ &= 1 \wedge \frac{f_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) f_\theta\left(\left\{\mathbf{Z}_t^i\right\}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{Z}_t^i\right\}\right) f_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right)}{f_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) f_\theta\left(\left\{\mathbf{X}_t^i\right\}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right) f_\theta\left(\left\{\mathbf{Z}_t^i\right\}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right)} \end{aligned} \tag{43}$$

which simplifies to:

$$\alpha_\theta\left(\left\{\mathbf{Z}_t^i\right\}, \left\{\mathbf{X}_t^i\right\}\right) = 1 \wedge \frac{f_\theta\left(\left\{\mathbf{Z}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{Z}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{Z}_t^i\right\}\right)}{f_\theta\left(\left\{\mathbf{X}_{t-1}^i\right\}\middle|\left\{\mathbf{X}_{t-2}^i\right\}\right) L_\theta\left(\tilde{y}_{t-1}\middle|\left\{\mathbf{X}_{t-1}^i\right\}\right) L_\theta\left(\tilde{y}_t\middle|\left\{\mathbf{X}_t^i\right\}\right)}, \tag{44}$$

which is similar to Eq. (28) of our paper. This concludes the derivation.

References

[1] Ajami NK, Duan Q, Sorooshian S. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour Res 2007;43:W01403. http://dx.doi.org/10.1029/2005WR004745.
[2] Akashi H, Kumamoto H. Construction of discrete-time nonlinear filter by Monte Carlo methods with variance-reducing techniques. Syst Control 1975;19:211–21.
[3] Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. J R Stat Soc Ser B 2010;72(3):269–342.
[4] Berzuini C, Best N, Gilks W, Larizza C. Dynamic conditional independence models and Markov chain Monte Carlo methods. J Am Stat Assoc 1997;92:1403–12.
[5] Bates BC, Campbell EP. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modeling. Water Resour Res 2001;37(4):937–47.
[6] Bengtsson T, Bickel P, Li B. Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. IMS collections probability and statistics: essays in honor of David A. Freedman, vol. 2; 2008. p. 316–34, doi:10.1214/193940307000000518.
[7] Beven KJ, Binley AM. The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 1992;6:279–98.
[8] Beven K. A manifesto for the equifinality thesis. J Hydrol 2006;320:18–36. http://dx.doi.org/10.1016/j.jhydrol.2005.07.007.
[9] Beven K, Freer J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. J Hydrol 2001;249:11–29.
[10] Blazkova S, Beven K. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. Water Resour Res 2009;45:W00B16. http://dx.doi.org/10.1029/2007WR006726.
[11] Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Ser B 1964;26:211–46.
[12] Boyle DP, Gupta HV, Sorooshian S. Toward improved calibration of hydrologic models: combining the strenghts of manual and automatic methods. Water Resour Res 2000;36(12):3663–74.
[13] Boyle DP. Multicriteria calibration of hydrologic models. Ph.D. Dissertation. Tucson: Department of Hydrology and Water Resources, University of Arizona; 2000.
[14] Bulygina N, Gupta H. Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. Water Resour Res 2011;47:W05514. http://dx.doi.org/10.1029/2010WR009614.
[16] Butts MB, Payne JT, Kristensen M, Madsen H. An evaluation of the impact of model structure on hydrological modeling uncertainty for streamflow simulation. J Hydrol 2004;298:242–66. http://dx.doi.org/10.1016/j.jhydrol.2004.03.042.
[17] Carpenter J, Clifford P, Fearnhead P. Improved particle filter for nonlinear problems. IEEE Proc Radar Son Nav 1999;146(1):2–7.
[19] Clark MP, Vrugt JA. Unraveling uncertainties in hydrologic model calibration: addressing the problem of compensatory parameters. Geophys Res Lett 2006;33:L06406. http://dx.doi.org/10.1029/2005GL025604.
[20] Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, et al. Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models. Water Resour Res 2008;44:W00B02. http://dx.doi.org/10.1029/2007WR006735.
[21] Crisan D, Del Moral P, Lyons T. Discrete filtering using branching and interacting particle systems. Markov Process Rel Fields 1999;5:293–318.
[22] Crisan D. Particle filters: a theoretical perspective. In: Doucet A, de Freitas JFG, Gordon NJ, editors. Sequential Monte Carlo methods in practice. New York: Springer-Verlag; 2001.
[23] Crisan D, Doucet A. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans Signal Process 2002;50(3):736–46.
[24] DeChant CM, Moradkhani H. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resour Res 2012. http://dx.doi.org/10.1029/2011WR011011.
[25] Del Moral P, Doucet A, Jasra A. Sequential Monte Carlo samplers. J R Stat Soc B 2006;68(3):411–36.
[26] Dhanya CT, Kumar DN. Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. Water Resour Res 2011;47:W06507. http://dx.doi.org/10.1029/2010WR010173.
[27] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat Comput 2000;10:197–208.

[28] Doucet A, De Freitas N, Gordon NJ. Sequential Monte Carlo methods in practice. New York: Springer; 2001. 581p.
[29] Doucet A, Tadić VB. Parameter estimation in general state-space models using particle methods. Ann Inst Stat Math 2003;55(2):409–22.
[30] Doucet A, Johansen AM. A tutorial on particle filtering and smoothing: fifteen years later. In: Crisan et D, Rozovsky B, editors. Handbook of Nonlinear Filtering. Oxford University Press; 2011.
[31] Duan Q, Gupta VK, Sorooshian S. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resour Res 1992;28(4):1015–31.
[32] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. J Geophys Res 1994;99(C5):10143–62.
[33] Evensen G, van Leeuwen PJ. An ensemble Kalman smoother for nonlinear dynamics. Mon Weather Rev 2000;128:1852–67.
[34] Freer J, Beven K, Ambroise B. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resour Res 1996;32(7):2161–73.
[35] Georgakakos KP, Seo DJ, Gupta H, Schaake J, Butts MB. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. J Hydrol 2004;298:222–41.
[36] Jacquin AP, Shamseldin AY. Development of a possibilistic method for the evaluation of predictive uncertainty in rainfall–runoff modeling. Water Resour Res 2007;43:W04425. http://dx.doi.org/10.1029/2006WR005072.
[37] Fenicia F, Savenije HHG, Matgen P, Pfister L. A comparison of alternative multiobjective calibration strategies for hydrological modeling. Water Resour Res 2007;43:W03434. http://dx.doi.org/10.1029/2006WR005098.
[38] Gelfand AE, Smith AF. Sampling based approaches to calculating marginal densities. J Am Stat Assoc 1990;85:398–409.
[39] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7:457–72.
[40] Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Stat Sci 1998;13(2):163–85.
[41] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 1984;6(6):721–41. http://dx.doi.org/10.1109/TPAMI.1984.4767596.
[42] Gilks WR, Berzuini C. Following a moving target-Monte Carlo inference for dynamic Bayesian models. J R Stat Soc Ser B 2001;63(1):127–46.
[43] Gordon N, Salmond D, Smith AFM. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. Proc IEEE Inst Electr Electron Eng F 1993;140(2):107–13.
[44] Gupta VK, Sorooshian S. The relationship between data and the precision of parameter estimates of hydrological models. J Hydrol 1985;81:57–77.
[45] Gupta HV, Sorooshian S, Yapo PO. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour Res 1998;34(4):751–63.
[48] Hammersley JM, Morton KW. Poor man's Monte Carlo. J R Stat Soc Ser B 1954;16(1):23–38.
[49] Handschin JE. Monte Carlo techniques for prediction and filtering of nonlinear stochastic processes. Automatica 1970;6:555–63.
[50] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;57:97–109.
[51] Huard D, Mailhot A. A Bayesian perspective on input uncertainty in model calibration: application to hydrological model ABC. Water Resour Res 2006;42:W07416. http://dx.doi.org/10.1029/2005WR004661.
[53] Ibbitt RP, OŠDonnell T. Designing conceptual catchment models for automatic fitting methods. In: Mathematical models in hydrology symposium, IAHS–AISH. Publication No. 2; 1974. p. 461–75.
[54] Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng Ser D Trans ASME 1960;82:35–45.
[55] Kantas N, Doucet A, Singh SS, Maciejowski JM, An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In 15th IFAC symposium on system identification, SYSID, Saint-Malo, France, 2009.
[56] Kavetski D, Kuczera G, Franks SW. Semi-distributed hydrological modeling: a "saturation path" perspective on TOPMODEL and VIC. Water Resour Res 2003;39(9):1246. http://dx.doi.org/10.1029/2003WR002122.
[57] Kavetski D, Kuczera G, Franks SW. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resour Res 2006;42:W03407. http://dx.doi.org/10.1029/2005WR004368.
[58] Kavetski D, Kuczera G, Franks SW. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resour Res 2006;42:W03408. http://dx.doi.org/10.1029/2005WR004376.
[59] Kavetski D, Kuczera G, Franks SW. Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artifacts. J Hydrol 2006;320(1–2):173–86. http://dx.doi.org/10.1016/j.jhydrol.2005.07.012.
[60] Kavetski D, Clark MP. Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resour Res 2010;46:W10511. http://dx.doi.org/10.1029/2009WR008896.
[61] Kelly KS, Krzysztofowicz R. A bivariate meta-Gaussian density for use in hydrology. Stoch Hydrol Hydraul 1997;11(1):17–31. http://dx.doi.org/10.1007/BF02428423.
[62] Khu ST, Madsen H. Multiobjective calibration with Pareto preference ordering: an application to rainfall–runoff model calibration. Water Resour Res 2005;41:W03004. http://dx.doi.org/10.1029/2004WR003041.
[63] Kuczera G, Kavetski D, Renard B, Thyer M. A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. Water Resour Res 2010;46:W07602. http://dx.doi.org/10.1029/2009WR008985.
[64] Kuczera G, Kavetski D, Franks S, Thyer M. Towards a Bayesian total error analysis of conceptual rainfall–runoff models: characterizing model error using storm-dependent parameters. J Hydrol 2006;331:161–77. http://dx.doi.org/10.1016/j.jhydrol.2006.05.010.
[65] Kuczera G, Parent E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J Hydrol 1998;211(1–4):69–85.
[66] Kuczera G. Improved parameter inference in catchment models, 1. Evaluating parameter uncertainty. Water Resour Res 1983;19(5):1151–62. http://dx.doi.org/10.1029/WR019i005p01151.
[67] Kuczera G. On the relationship between the reliability of parameter estimates and hydrologic time series used in calibration. Water Resour Res 1982;18(1):146–54.
[68] Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J Comput Graph Stat 1996;5(1):1–25.
[69] Kuczera G, Parent E. Monte Carlo assessment of parameter uncertainty in conceptual catchments model: the Metropolis algorithm. J Hydrol 1998;211(1–4):69–85.
[70] Liu JS, Chen R. Sequential Monte Carlo methods for dynamic systems. J Am Stat Assoc 1998;93(443):1032–44.
[71] Lorenz EN. Deterministic non-periodic flow. J Atmos Sci 1963;20:130–41.
[72] Lorenz EN. Predictability: a problem partly solved. Proceedings seminar on predictability, vol. 1. Reading, Berkshire, UK: ECMWF; 1996. p. 1–18.
[73] Lorenz EN, Emanuel KA. Optimal sites for supplementary weather observations: simulation with a small model. J Atmos Sci 1998;55:399–414.
[74] MacEachern SN, Muller P. Estimating mixture of dirichlet process models. J Comput Graph Stat 1998;7:223–38.
[75] Madsen H. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. J Hydrol 2000;235:276–88.
[76] Marshall LA, Nott DJ, Sharma A. Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework. Hydrol Process 2006;21:847–61.
[77] Mazi K, Koussis AD, Restrepo PJ, Koutsoyiannis D. A groundwater-based, objective-heuristic parameter optimization method for a precipitation-runoff model and its application to a semi-arid basin. J Hydrol 2004;290:243–58.
[78] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–92.
[79] Miller RN, Ghil M, Ghautiez F. Advanced data assimilation in strongly nonlinear dynamical system. J Atmos Sci 1994;51:1037–55.
[80] Montanari A, Brath A. A stocastic approach for assessing the uncertainty of rainfall–runoff simulations. Water Resour Res 2004;40:W01106. http://dx.doi.org/10.1029/ 2003WR002540.
[81] Montanari A, Grossi G. Estimating the uncertainty of hydrological forecasts: a statistical approach. Water Resour Res 2008;44:W00B08. http://dx.doi.org/10.1029/ 2008WR006897.
[82] Montanari A, Toth E. Calibration of hydrological models in the spectral domain: an opportunity for scarcely gauged basins? Water Resour Res 2007;43:W05434. http://dx.doi.org/10.1029/2006WR005184.
[83] Moore RJ. The probability-distributed principle and runoff production at point and basin scales. Hydrol Sci J 1985;30(2):273–97.
[84] Moore RJ. The PDM rainfall–runoff model. Hydrol Earth Syst Sci 2007;11(1):483–99.
[85] Moradkani H, Hsu K, Gupta H, Sorooshian S. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour Res 2005;41:W05012. http://dx.doi.org/10.1029/2004WR003604.
[86] Moradkhani H, Sorooshian S, Gupta HV, Houser PR. Dual state parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 2005;28:135–47.
[87] Moran PAP. Simulation and evaluation of complex water system operations. Water Resour Res 1970;6:1737–42.
[88] Pauwels VRN. A multi-start weight-adaptive recursive parameter estimation method. Water Resour Res 2008;44(4):W044156. http://dx.doi.org/10.1029/2007WR005866.
[89] Pauwels VRN, De Lannoy GJM. Multivariate calibration of a water and energy balance model in the spectral domain. Water Resour Res 2011;47:W07523. http://dx.doi.org/10.1029/2010WR010292.
[90] Pitt MK, Shephard N. Filtering via simulation: auxiliary particle filters. J Am Stat Assoc 1999;94:590–9.
[91] Pitt MK, Shephard N. Auxiliary variable based particle filters. In: Doucet A, de Freitas N, Gordon NJ, editors. Sequential Monte Carlo methods in practice. New York: Springer-Verlag; 2001.
[93] Renard B, Kavetski D, Leblois E, Thyer M, Kuczera G, Franks SW. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: characterizing rainfall errors using conditional simulation. Water Resour Res 2011;47:W11516. http://dx.doi.org/10.1029/2011WR010643.
[94] Rings J, Vrugt JA, Schoups G, Huisman JA, Vereecken H. Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. Water Resour Res 2012. http://dx.doi.org/10.1029/2011WR011607.

[95] Roberts GO, Rosenthal JS. Optimal scaling for various Metropolis–Hastings algorithms. Stat Sci 2001;16:351–67.
[96] Rosenbluth MN, Rosenbluth AW. Monte Carlo calculation of the average extension of molecular chains. J Chem Phys 1956;23(2):356–9.
[97] Salamon P, Feyen L. Assessing parameter, precipitation, and predictive uncertainty in a distibuted hydrological model using sequential data assimilatiom with the particle filter. J Hydrol 2009;376:428–42. http://dx.doi.org/10.1016/j.jhydrol.2009.07.051.
[98] Schaefli B, Maraun D, Holschneider W. What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. Adv Water Resour 2007;30:2511–25. http://dx.doi.org/10.1016/j.advwatres.2007.06.004.
[99] Schoups G, Vrugt JA. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. Water Resour Res 2010;46:W10531. http://dx.doi.org/10.1029/2009WR008933.
[100] Schoups G, Vrugt JA, Fenicia F, van de Giesen NC. Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. Water Resour Res 2010;46:W10530. http://dx.doi.org/10.1029/2009WR008648.
[101] Schoups G, Vrugt JA. Bayesian inference of the marginal density of hydrologic models: path-sampling using DREAM. Water Resour Res; submitted for publication.
[102] Seibert J. Multi-criteria calibration of a conceptual rainfall–runoff model using a genetic algorithm. Hydrol Earth Syst Sci 2000;4(2):215–24.
[103] Slater AG, Clark MP. Snow data assimilation via an ensemble Kalman filter. J Hydrometeorol 2006;7(3):478–93.
[104] Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling–resampling perspective. J Am Stat Assoc 1992;46(2):84–8.
[105] Solomatine DP, Shrestha DL. A novel method to estimate model uncertainty using machine learning techniques. Water Resour Res 2009;45:W00B11. http://dx.doi.org/10.1029/2008WR006839.
[106] Sorooshian S, Dracup JA. Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: correlated and heteroscedastic error cases. Water Resour Res 1980;16(2):430–42.
[107] Sorooshian S, Gupta VK, Fulton JL. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall–runoff models: influence of calibration data variability and length on model credibility. Water Resour Res 1983;19(1):251–9.
[108] Storn R, Price K. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 1997;11:341–59.
[109] Reichert P, Mieleitner J. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resour Res 2009;45:W10402. http://dx.doi.org/10.1029/2009WR007814.
[110] ter Braak CJF, Vrugt JA. Differential evolution Markov chain with snooker updater and fewer chains. Stat Comput 2007;18(4):435–46. http://dx.doi.org/10.1007/s11222-008-9104-9.
[111] Thiemann M, Trosset M, Gupta HV, Sorooshian S. Bayesian recursive parameter estimation for hydrologic models. Water Resour Res 2001;37(10):2521–35. http://dx.doi.org/10.1029/2000WR900405.
[112] Tolson BA, Shoemaker CA. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resour Res 2007;43:W01413. http://dx.doi.org/10.1029/2005WR004723.
[113] Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Am Meteorol Soc 1998;79:61–78.
[115] van der Merwe R, Doucet A, De Freitas N, Wan E. The unscented particle filter. In: Leen TK, Dietterich TG, Tresp V, editors. Advances in Neural Information Processing Systems (NIPS13). MIT Press; 2000.
[116] Vogel RM, Fennessey NM. Flow-duration curves. I: new interpretations and confidence interval. J Water Resour Planning Manage 1994;120(4):485–504.
[117] Vrugt JA, Bouten W, Gupta HV, Sorooshian S. Toward improved identifiability of hydrologic model parameters: the information content of experimental data. Water Resour Res 2002;38(12). http://dx.doi.org/10.1029/2001WR001118. Article No. 1312.
[118] Vrugt JA, Gupta HV, Bouten W, Sorooshian S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour Res 2003;39(8):1201. http://dx.doi.org/10.1029/2002WR001642.
[119] Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour Res 2003;39(8). http://dx.doi.org/10.1029/2002WR001746. Article No. 1214.
[120] Vrugt JA, Dekker SC, Bouten W. Identification of rainfall interception model parameters from measurements of throughfall and forest canopy storage. Water Resour Res 2003;39(9). http://dx.doi.org/10.1029/2003WR002013. Article No. 1251.
[121] Vrugt JA, Diks CGH, Bouten W, Gupta HV, Verstraten JM. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resour Res 2005;41(1):W01017. http://dx.doi.org/10.1029/2004WR003059.
[122] Vrugt JA, Gupta HV, Nualláin BÓ, Bouten W. Realtime data assimilation for operational ensemble streamflow forecasting. J Hydrometeorol 2006;7(3):548–65. http://dx.doi.org/10.1175/JHM504.1.
[123] Vrugt JA, Nualláin BÓ, Robinson BA, Bouten W, Dekker SC, Sloot PMA. Application of parallel computing to stochastic parameter estimation in environmental models. Comput Geosci 2006;32(8):1139–55. http://dx.doi.org/10.1016/j.cageo.2005.10.015.
[124] Vrugt JA, Gupta HV, Sorooshian S, Wagener T, Bouten W. Application of stochastic parameter optimization to the Sacramento soil moisture accounting model. J Hydrol 2006;325(1–4):288–307. http://dx.doi.org/10.1016/j.jhydrol.2005.10.041.
[125] Vrugt JA, Robinson BA. Improved evolutionary optimization from genetically adaptive multimethod search. Proc Natl Acad Sci USA 2007;104:708–11. http://dx.doi.org/10.1073/pnas.0610471104.
[126] Vrugt JA, Robinson BA. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. Water Resour Res 2007;43:W01411. http://dx.doi.org/10.1029/2005WR004838.
[127] Vrugt JA, ter Braak CJF, Clark MP, Hyman JM, Robinson BA. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour Res 2008;44:W00B09. http://dx.doi.org/10.1029/2007WR006720.
[128] Vrugt JA, ter Braak CJF, Diks CGH, Higdon D, Robinson BA, Hyman JM. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 2009;10(3):273–90.
[129] Vrugt JA, ter Braak CJF, Gupta HV, Robinson BA. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch Environ Res Risk Assess 2009;23(7):1011–26. http://dx.doi.org/10.1007/s00477-008-0274-y.
[130] Vrugt JA, Robinson BA, Hyman JM. Self-adaptive multimethod search for global optimization in real-parameter spaces. IEEE Trans Evolut Comput 2009;13(2):243–59. http://dx.doi.org/10.1109/TEVC.2008.924428.
[131] Vrugt JA. Resolving uncertainty in hydrologic modeling: Better statistics or measurements. Water Resour Res, Submitted for publication.
[132] Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. Hydrol Process 2003;17:455–76.
[133] Wang Q. The genetic algorithm and its application for calibrating conceptual rainfall–runoff models. Water Resour Res 1991;27(9):2467–71.
[134] Weerts AH, El Serafy GYH. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall–runoff models. Water Resour Res 2006;42:W09403. http://dx.doi.org/10.1029/2005WR004093.
[135] Yapo PO, Gupta HV, Sorooshian S. Calibration of conceptual rainfall–runoff models: sensitivity to calibration data. J Hydrol 1996;181:23–48.
[136] Yapo PO, Gupta HV, Sorooshian S. Multi-objective global optimization for hydrologic models. J Hydrol 1998;204:83–97.
[137] Zaristkii VS, Svetnik VB, Shimelevich LI. Monte Carlo technique in problems of optimnal data processing. Automat Remote Control 1975(12):95–103.
[138] Young PC. Advances in real-time flood-forecasting forecasting. Philos Trans R Soc Phys Eng Sci 2002;360(9):1433–50.
[139] Young PC. Hypothetico-inductive data-based mechanistic modeling of hydrological systems. Water Resour Res; submitted for publication.