XIX International Conference on Water Resources
CMWR 2012
University of Illinois at Urbana-Champain
June 17-22,2012

SEQUENTIAL MONTE CARLO METHODS FOR THE CALIBRATION OF STOCHASTIC RAINFALL-RUNOFF MODELS

Franz Konecny and Hans-Peter Nachtnebel

BOKU - University of Natural Resources and Life Sciences, Vienna
Muthgasse 18, A 1190 Vienna, Austria
e-mail: franz.konecny@boku.ac.at, web page: `http://www.boku.ac.at`

Key words: Stochastic rainfall-runoff models, calibration, Sequential Monte Carlo methods

Summary: The rainfall-runoff (R-R) model considered in this talk is a conceptual stochastic model, formulated in continuous-discrete state space form. We use a maximum likelihood (ML) approach based on an EM algorithm. In this approach, the key ingredient is the computation of smoothed additive functionals of hidden states. Sequential Monte Carlo methods (SMC), also called particle methods, are used for smoothing in the state space model. Instead of particle smoothers, unscented Kalman smoothers can be used.

1 INTRODUCTION

Conceptual rainfall-runoff models are frequently applied for real-time forecasting of discharge. The task of calibrating the model is an important problem faced by hydrologists and observed records of rainfall and runoff may be used to estimate the parameters of the model.

The model considered in this presentation is a conceptual stochastic model simulating the catchment response by a non-linear model built up from a cascade of non-linear reservoirs. This dynamic system can be represented by a system of stochastic differential equations

$$\begin{aligned} \dot{S}_1(t) &= I(t) - h(S_1(t)) + \dot{W}(t) \\ \dot{S}_2(t) &= h(S_1(t)) - h(S_2(t)) \\ &\dots \\ \dot{S}_m(t) &= h(S_{m-1}(t)) - h(S_m(t)) \end{aligned} \tag{1}$$

where I is the input (rainfall), S_i is the storage in the i-th reservoir. Due to input error an additive random term $\dot{W}$ was included, which may be white noise. The function $h(.)$ represents the outflow-storage relation. It is assumed to be of the form $h(s) = as^\beta$. The final output of the system is the runoff, given by $Y(t) = h(S_m(t))$. It is only observed at discrete times t_k, with a further measurement noise Z_k:

$$Y_k = h(S_m(t_k)) + Z_k. \tag{2}$$

For the calibration of the system parameters we use maximum likelihood (ML) estimation. Because only discrete observations of the output process (2) are available, the state process is hidden, it is an estimation problem with *partial observations.*

2 CONTINUOUS-DISCRETE STATE SPACE MODEL

The idea of state space models is that there is an unobserved state of interest that evolves through time, and that partial observations of the state are made at successive time-points. The dynamics of the the hydrologic model described by eqs.(1) Using a compact notation the evolutions is described by a stochastic differential equation

$$d\mathbf{S}(t) = \mathbf{b}_\theta(\mathbf{S}(t), I(t)) + \mathbf{G}_\theta \, d\mathbf{W}(t) \tag{3.1}$$

$$Y_k = h_\theta(\mathbf{S}(t_k)) + Z_k, \tag{3.2}$$

where t_k is the k-th observation time and Z_k is Gaussian white noise with the variance r^2, say. The subscript θ denotes the system parameter to be estimated.

3 TRANSITION DENSITY AND CONDITIONAL LIKELIHOOD

For likelihood based inference on $\mathbf{S}(t)$, it is necessary to have the transition density (t.d.)

$$p(\mathbf{s}_{k+1}|\mathbf{s}_k),$$

that is, the conditional density of $\mathbf{S}(t_{k+1})$ given $\mathbf{S}(t_k) = \mathbf{s}_k$.

Since $\mathbf{S}(t)$ is a Markov process, i.e. its conditional distribution of $\mathbf{S}(t_{k+1})$ given the past states $\mathbf{s}_{1:k}$ * depends only on $\mathbf{s}_k$ through the t.d. $p(\mathbf{s}_{k+1}|\mathbf{s}_k)$, and that the conditional distribution of Y_{t_k} given $\mathbf{s}_{1:k}$ and the past observations $y_{1:k-1}$ depends only upon $\mathbf{s}_k$ throught the *conditional likelihood* $p(y_k|\mathbf{s}_k)$. Because of the special form of the observation process (3) the conditional likelihood is of the form

$$p(y_k|\mathbf{s}_k) = \phi(y_k; h(\mathbf{s}_k), r^2),$$

where $\phi(.;\mu,\upsilon)$ is the density of normal distribution with the mean μ and the variance υ.

Unfortunately, in all except a few special cases the t.d. is not analytically available and approximative methods are used. Pedersen[13] proposed approximations of the transition density function $p(\mathbf{s}_{k+1}|\mathbf{s}_k)$ via data augmentation. The interval $[t_k, t_{k+1}]$ is partitioned into M subintervals by grid points $t_k = \tau_0 < \tau_1 < ... < \tau_M = t_{k+1}$, each with equal length $h = \Delta t/M$. Let $\phi(.;\mathbf{m},\Sigma)$ denote a multivariate normal density with mean $\mathbf{m}$ and covariance Σ. The *Euler Approximation* of $p(\mathbf{s}_{k+1}|\mathbf{s}_k)$ is given by

$$p^{(M)}(\mathbf{s}_{k+1}|\mathbf{s}_k) = \int \prod_{m=0}^{M-1} p^{(1)}(h; \mathbf{u}_m, \mathbf{u}_{m+1})\, d(\mathbf{u}_1, ..., \mathbf{u}_{M-1}), \tag{4}$$

where $\mathbf{u}_0 = \mathbf{s}_k$ and $\mathbf{u}_M = \mathbf{s}_{k+1}$, and

$$p^{(1)}(h; \mathbf{u}_m, \mathbf{u}_{m+1}) = \phi(\mathbf{u}_{m+1}; \mathbf{u}_m + h\mathbf{b}(\mathbf{u}_m), h\mathbf{G}^T\mathbf{G}).$$

The integral on the r.h.s. is approximated using Monte Carlo methods. Several importance sampling techniques (cf. Durham and Gallant[5]) were suggested to make Pedersen's approximations more effective.

4 ML ESTIMATION AND EM ALGORITHM

For the calibration of a state space model the system parameters are to be estimated indirectly from the output measurements. So we are faced with a parameter estimation problem with partial observations. In this situation three approaches are possible:

4.1 The direct approach

The estimation problem from section 2 can formulated as an maximization problem of the log-likelihood function

$$\widehat{\theta} = \text{argmax}_\theta \ell_\theta(y_{1:n}), \quad \ell_\theta(y_{1:n}) = \log p_\theta(y_{1:n}), \tag{5}$$

where $p_\theta(y_{1:n})$ denotes the joint density function of n output measurements, specified by equation (2). It is given by

$$p_\theta(y_{1:n}) = \int ... \int p(\mathbf{s}_0) \prod_{k=1}^{n} p_\theta(\mathbf{s}_k|\mathbf{s}_{k-1}) \prod_{k=1}^{n} p_\theta(y_k|\mathbf{s}_k)\, d\mathbf{s}_{1:n} \tag{6}$$

It also admits the following recursive form

$$p_\theta(y_{1:n}) = p_\theta(y_n|y_{1:n-1}) p_\theta(y_{1:n-1}) = \prod_{k=1}^{n} p_\theta(y_k|y_{1:k-1}) \tag{7}$$

* We have adopted the following notation: for any sequence (x_k), we define $x_{n:m} = (x_n, x_{n+1}, ..., x_m)$.

with the convention $p_\theta(y_1|y_0) = \int p_\theta(y_1|\mathbf{s}_1)p_\theta(\mathbf{s}_1)\,d\mathbf{s}_1$. The log-likelihood is given by

$$\ell_\theta(y_{1:n}) := \log p_\theta(y_{1:n}) = \sum_{k=1}^{n} \log p_\theta(y_k|y_{1:k-1}).$$

One approach for solving this MLE problem is to use an iterative gradient-based procedure with a decreasing stepsize sequence (γ_j) so to average out the "noise" produced by MC estimates of the likelihood. at iteration step $j+1$ of the gradient algorithm, θ_j is updated by

$$\theta_{j+1} = \theta_j + \gamma_j \widehat{\nabla \ell_{\theta_j}}(y_{1:n})$$

where $\widehat{\nabla \ell_{\theta_j}}(y_{1:n})$ is a particle estimate of the likelihood gradient from time 1 to n at the current estimate θ_j as given in section 5 below.

We assume that the initial density $p(\mathbf{s}_0)$ does not depend on the parameter θ. In this case the score vector (gradient of the log-likelihood with respect of θ takes the following form (Cappe et al.[2]):

$$\begin{aligned}\nabla \ell_\theta(y_{1:n}) = E_\theta\left[\nabla_\theta \log p_\theta(\mathbf{s}_{1:n}, y_{1:n})\middle| y_{1:n}\right] &= E_\theta\left[\sum_{k=1}^{n} \nabla_\theta \log p_\theta(\mathbf{s}_k|\mathbf{s}_{k-1})\middle| y_{1:n}\right] \\ &+ E_\theta\left[\sum_{k=1}^{n} \nabla_\theta \log p_\theta(y_k|\mathbf{s}_k)\middle| y_{1:n}\right] \end{aligned} \tag{8}$$

4.2 Prediction error estimation

The prediction error (PE) method (Ljung[10], Schön et al.[15]) is based on the minimization of the discrepancy between the measured outputs of the system and the predicted outputs according to the state space model. For linear state and observation equations with additive noise the PE estimators coincides with ML estimators. In the nonlinear case, the PE estimator is a *quasi-maximum likelihood estimator*.

A general solution to this problem is given as

$$\widehat{\theta}_{PE} = \text{argmin}_\theta V_n(\theta) \tag{9}$$

with a cost function $V_n(\theta)$ of the form

$$V_n(\theta) = \frac{1}{n}\sum_{k=1}^{n} |\nu_k(\theta)|^2, \;\; \nu_k(\theta) = y_k - \widehat{y}_{k|k-1}(\theta), \tag{10}$$

where

$$\widehat{y}_{k|k-1}(\theta) := E_\theta[y_k|y_{1:k-1}]. \tag{11}$$

is the mean square optimal one-step ahead predictor of y_k. Computing the esimate $\widehat{\theta}_{PE}$ requires the solution of a minimization problem. The cost function $V_n(.)$ is, generally spoken, non-convex, and its minimizer $\widehat{\theta}_{PE}$ will not be expressible in closed form.

4.3 The Expectation Maximization Algorithm

A very popular alternative method for parameter estimation is the expectation-maximization (EM) algorithm (Dempster et al.[3], Kantas et al.[8]). It is an iterative method which increases the log-likelihood at each iteration. Each iteration of the EM algorithm consists of the following two steps.

1. *E-Step:* Set $\theta' = \widehat{\theta}_j$ and compute

$$\mathcal{Q}(\theta, \theta') = E_{\theta'}\left[p_\theta(\mathbf{s}_{1:n}, y_{1:n})|y_{1:n}\right].$$

2. *M-Step:* Find
$$\widehat{\theta}_{j+1} = \text{argmax}_\theta \mathcal{Q}(\theta, \theta').$$

If a stopping rule is satisfied, then set final estimate $\widehat{\theta} = \widehat{\theta}_{j+1}$, else repeat steps 1 and 2 with $j <= j+1$.
The employment of the EM algorithm requires routines for computing the expectation involved in $\mathcal{Q}(\theta, \theta')$, and also tools for maximizing $\mathcal{Q}(\theta, \theta')$ with respect to θ.

The Expectation Step

The expectation step corresponds to the calculation of $\mathcal{Q}(\theta, \theta_k)$. Making use of the Markov property and Bayes' rule we obtain

$$\begin{aligned} p_\theta(\mathbf{s}_{1:n}, y_{1:n}) &= p_\theta(y_{1:n}|\mathbf{s}_{1:n}) p_\theta(\mathbf{s}_{1:n}) \\ &= p(\mathbf{s}_0) \prod_{k=1}^{n} p_\theta(\mathbf{s}_k|\mathbf{s}_{k-1}) p_\theta(y_k|\mathbf{s}_k). \end{aligned} \tag{12}$$

Upon taking the logarithm and taking conditional expectation we get

$$\begin{aligned} \mathcal{Q}(\theta, \theta') &= c + \sum_{k=1}^{n} E_{\theta'} \left[\log p_\theta(\mathbf{s}_k|\mathbf{s}_{k-1})|y_{1:n}\right] \\ &\quad + \sum_{k=1}^{n} E_{\theta'} \left[\log p_\theta(y_k|\mathbf{s}_k)|y_{1:n}\right]. \\ &= \sum_{k=1}^{n} \int \log p_\theta(\mathbf{s}_k|\mathbf{s}_{k-1}) p_\theta(\mathbf{s}_{k-1:k}|y_{1:n})\, d\mathbf{s}_{k-1:k} \\ &\quad + \sum_{k=1}^{n} \int \log p_\theta(y_k|\mathbf{s}_k) p_\theta(\mathbf{s}_k|y_{1:n})\, d\mathbf{s}_k \end{aligned} \tag{13}$$

The additive structure of (13) establishes that for the evaluation of $\mathcal{Q}(\theta, \theta')$ we require the marginal smoothing densities $p_{\theta'}(\mathbf{s}_k|y_{1:n})$ and $p_{\theta'}(\mathbf{s}_{k-1:k}|y_{1:n})$, rather than the complete joint density $p_{\theta'}(\mathbf{s}_{1:n}|y_{1:n})$. These marginal smoothing densities will be approximated using SMC methods, as discussed in section 5.

Besides the linear case where explicit solutions are available, the M-step is carried out by use of a general purpose maximizer.

5. SEQUENTIAL MONTE CARLO

Sequential Monte Carlo (SMC), also known as particle filtering and particle smoothing, approximates the exact filtering and smoothing relations by propagating particle trajectories in the state space of the latent process. For more details we refer to Doucet et al.[4] or Cappe et al.[2] and simply present below the sequential importance sampling (SIS) with resampling algorithm. In order to keep notation simple, we fix the model parameters and omit θ in the following.

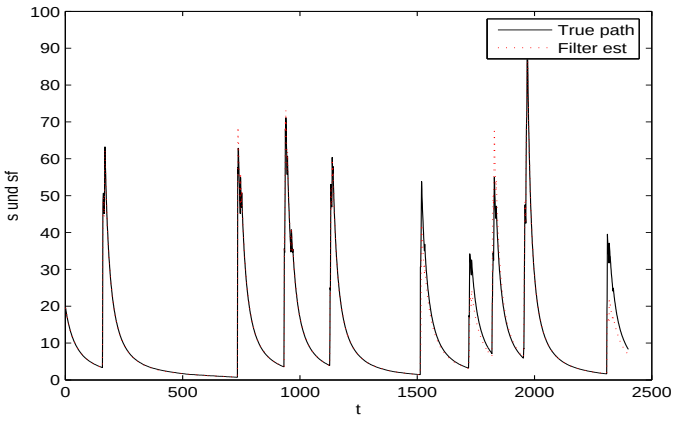
5.1 Sequential Importance Sampling (SIS)

At time zero, a number of N particles $\mathbf{s}_0^{(i)}, i = 1, .., N$, are drawn from the density $p(\mathbf{s}_0)$. These initial particles are assigned with the importance weights $w_0^{(i)} = 1/N$. Henceforth, the particle paths $\mathbf{s}_{0:m}^{(i)} = (\mathbf{s}_0^{(i)}, .., \mathbf{s}_m^{(i)}), i = 1, .., N$, are recursively updated according to the following procedure.

Given a set of N particles $(\mathbf{s}_{1:k-1}^{(i)}, i = 1, .., N)$, we obtain a set of particles $(\mathbf{s}_{1:k}^{(i)}, i = 1, .., N)$ by simulating the dynamic of (3.1). The particles are weighted by

$$w_k^{(i)} := w(\mathbf{s}_k^{(i)}) = \frac{p(y_k|\mathbf{s}_k^{(i)})}{\sum_{j=1}^{N} p(y_k|\mathbf{s}_k^{(j)})}. \tag{14}$$

This iterative scheme produces a weighted measure $(\mathbf{s}_{1:k}^{(i)}, w_k^{(i)}, i = 1, .., N)$. A resampling (selection) step may be included at this point that chooses the fittest particles (this is the SIR algorithm of Gordon et al.[9]).



Particle Filter

5.2 Marginal Particle Filter

The marginal particle filter (MPF) make use of the marginal density $p(\mathbf{s}_k|y_{1:k})$ rather than the joint density $p(\mathbf{s}_{1:k}|y_{1:k})$. Approaches based on the joint density will inevitably run into problems as the sample size increases. The sequential nature of the algorithm means that the variance is high, leading to most paths having vanishing small probability. this problem is known as *degeneracy* of the weights whose variance tends to increase to infinity.

The predictive density is obtained by marginalizing

$$p(\mathbf{s}_k|y_{1:k-1}) = \int p(\mathbf{s}_k|\mathbf{s}_{k-1})p(\mathbf{s}_{k-1}|y_{1:k-1})\, d\mathbf{s}_{k-1} \tag{15}$$

hence

$$\begin{aligned} p(\mathbf{s}_k|y_{1:k}) &\propto p(y_k|\mathbf{s}_k)p(\mathbf{s}_k|y_{1:k-1}) \\ &= p(y_k|\mathbf{s}_k) \int p(\mathbf{s}_k|\mathbf{s}_{k-1})p(\mathbf{s}_{k-1}|y_{1:k-1})\, d\mathbf{s}_{k-1}. \end{aligned}$$

The integral in equation (15) is typically not evaluable analytically. By SMC we have the particle approximation $(\mathbf{s}_{k-1}^{(i)}, w_{k-1}^{(i)}, i = 1, .., N)$ of $p(\mathbf{s}_{k-1}|y_{1:k-1})$ and can approximate the r.h.s. of (15) by the weighted kernel estimate

$$\sum_{j=1}^{N} w_{k-1}^N p(\mathbf{s}_k|\mathbf{s}_{k-1}^{(j)}).$$

While we are free to choose any proposal distribution that has appropriate support, it is convenient to assume that the proposal takes a similar form, namely

$$q(\mathbf{s}_k|y_{1:k}) = \sum_{j=1}^{N} w_{k-1}^N q(\mathbf{s}_k|y_k, \mathbf{s}_{k-1}^{(j)}). \tag{16}$$

The importance weights are now on the marginal space

$$\widetilde{w}_k = \frac{p(\mathbf{s}_k|y_{1:k})}{q(\mathbf{s}_k|y_{1:k})}. \tag{17}$$

5.3 Particle Smoother

The basic version of the particle filter actually provides us with an approximation of the smoothing distribution $p(\mathbf{s}_{1:n}|y_{1:n})$ at no extra cost. The stored particle paths $(\mathbf{s}_{1:k}^{(i)}; k = 1, ..n)$ and their associated weights $(w_k^{(i)}; k = 1, .., n)$ can be considered as a weighted sample from the joint

smoothing distribution. By marginalization we obtain an approximation of the marginal smoothing densities $p(\mathbf{s}_k|y_{1:n}), k = 1, ..n$. However these approximations are poor when n is large because of the degeneracy problem. Several alternative schemes have been proposed: Fixed-lag approximation (Kitagawa and Sato[9] and Olsson et al.[12]), forward filtering-backward smoothing and two-filter smoothing (cf. Briers et al.[1]).

The fixed-lag approximation is the simplest approach and relies on the fact that for state space models with good forgetting properties, we have

$$p(\mathbf{s}_{1:k}|y_{1:n}) \approx p(\mathbf{s}_{1:k}|y_{1:\min(k+\Delta,n)}) \tag{18}$$

for Δ large enough; that is observations collected at times $l > k + \Delta$ do not bring additional information about $\mathbf{S}(t), t \leq t_k$. This suggests a very simple scheme - simply don't update the estimate of $\mathbf{S}(t_k)$ after the $l = k + \Delta$

In practice, it has been observed (cf. Olsson et al.[12]) that the lag Δ controls a bias/variance tradeoff: when Δ is to small a biased estimate results; as Δ is enlarged, the bias disappears but the estimation variance raises. At time $k + \Delta$ The particle filter yields the following approximation of $p(\mathbf{s}_{1:k+\Delta}|y_{1:k+\Delta})$:

$$\widehat{p}(\mathbf{s}_{1:k+\Delta}|y_{1:k+\Delta}) = \sum_{i=1}^{N} w_{1:k+\Delta}^{(i)} \delta(\mathbf{s}_{1:k+\Delta} - \mathbf{s}_{1:k+\Delta}^{(i)}) \tag{19}$$

By marginalization, we obtain an estimate of the fixed-lag smoothing density:

$$\widehat{p}(\mathbf{s}_k|y_{1:k+\Delta}) = \sum_{i=1}^{N} w_{1:k+\Delta}^{(i)} \delta(\mathbf{s}_k - \mathbf{s}_k^{(i)}) \tag{20}$$

When Δ is high, the approximation of the smoothing distribution will be poor.

5.4 Particle Implementation of the EM algorithm

The particle smoother gives a sample based approximation of the marginal smoothing densities which can be written as

$$p(\mathbf{s}_k|y_{1:n}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{s}_k - \mathbf{s}_k^{(i)}) \tag{21}$$

$$p(\mathbf{s}_{k-1:k}|y_{1:n}) \approx \sum_{i=1}^{N} w_{k-1:k}^{(i)} \delta(\mathbf{s}_{k-1} - \mathbf{s}_{k-1}^{(i)}) \delta(\mathbf{s}_k - \mathbf{s}_k^{(i)}) \tag{22}$$

where N is the number of particles and $w_k^{(i)}$ is the weight associated with the $i-$th particle computed by the smoothing procedure. The approximation of the auxiliary function, obtained by substituting the approximations (21) and (22) into (13) is given by

$$\begin{aligned} \widehat{\mathcal{Q}}(\theta, \theta') = & \sum_{k=1}^{n} \sum_{i=1}^{N} w_{k-1:k}^{(i)} \log p_\theta(\mathbf{s}_k^{(i)}|\mathbf{s}_{k-1}^{(i)}) \\ & \sum_{k=1}^{n} \sum_{i=1}^{N} w_k^{(i)} \log p_\theta(y_k|\mathbf{s}_k^{(i)}) \end{aligned} \tag{23}$$

6 UNSCENTED PARTICLE METHODS

The *extended Kalman filter* (EKF) is a suboptimal filter for nonlinear systems. It is based on a simple approximation to the nonlinear equations that describe the rainfall-runoff process, whereas the unscented Kalman filter (UKF) uses the *unscented transformation* (UT) of Julier et al.[7]. This is a method for calculating the moments of a random variable which undergoes a nonlinear transformation. It is based on the principle that is easier to approximate a probability distribution than a general nonlinear transformation. The UT is similar to Gauss-Hermite

quadrature for numerical approximation of multidimensional integrals. The UKF assumes that the pdf of the state vector is Gaussian, and this distribution is sampled at a number of carefully chosen points to approximate the multidimensional integrals required. The UKF approximations is generally much better than the EKF, and the computational complexity of the UKF and EKF are roughly the same.

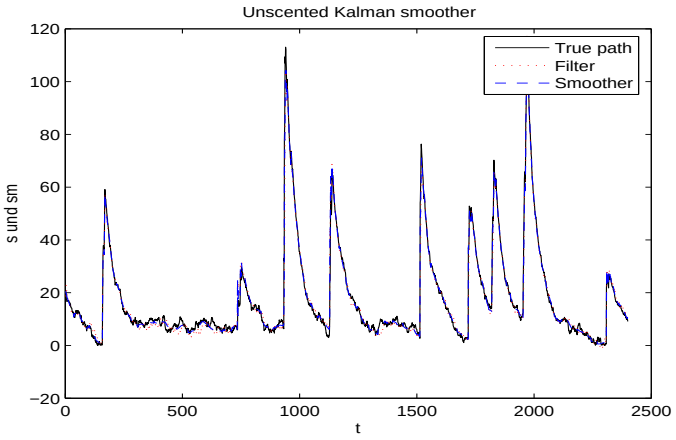
Unscented Kalman smoothing (UKS) procedures for continuous-discrete time state space models were formulated recently in a paper of Särkkä[14]. The normal approximations of the marginal smoothing densities of the state are

$$p_{\theta'}(\mathbf{s}_k|y_{1:n}) \approx \phi(\mathbf{s}_k; \mathbf{m}_{k|n}, P_{k|n})$$
$$p_{\theta'}(\mathbf{s}_k, \mathbf{s}_{k+1}|y_{1:n}) \approx \phi(\mathbf{s}_{k:k+1}; \mathbf{m}_{k,k+1|n}, P_{k,k+1|n})$$

and

$$\begin{aligned}\widehat{\mathcal{Q}}(\theta, \theta') = &\int \log p_\theta(\mathbf{s}_0)|y_{1:n})\phi(\mathbf{s}_0; \mathbf{m}_{0|n}, P_{0|n})\, d\mathbf{s}_0 \\ &+ \sum_{k=0}^{n-1} \int \log p_\theta(\mathbf{s}_{k+1}|\mathbf{s}_k)|y_{1:n})\phi(\mathbf{s}_{k:k+1}; \mathbf{m}_{k,k+1|n}, P_{k,k+1|n})\, d\mathbf{s}_{k:k+1} \\ &+ \sum_{k=1}^{n} \int \log p_\theta(y_k|\mathbf{s}_k)\phi(\mathbf{s}_k; \mathbf{m}_{k|n}, P_{k|n})\, d\mathbf{s}_k.\end{aligned} \tag{24}$$

In van der Merwe et al.[16] the UKF was used to approximate the proposal distribution of the particle filter, which results in the so-called *unscented particle filter* (UPF). It was observed, that the UKF generates proposal distributions that match the true filter more closely than the EKF.



7 CONCLUSIONS

In this paper methods for the calibration of conceptual R-R models are discussed. The estimation method is ML with partial observations. Direct maximization of the log-likelihood function and scaling the gradients is difficult especially in the case of a high-dimensional parameter. The EM algorithm is more stable and typically computationally cheaper for high-dimensional problems. The E-step in the EM algorithm is performed by using particle smoothing or UKS. Compared to the particle smoother, the UKS is computationally far less expensive. In some numerical experiments we found that the improvement of the UPF over the UKF is marginal or nil. The M-step is carried out by a general purpose optimization procedure.

REFERENCES

[1] M. Briers, A. Doucet and S. Maskell. Smoothing Algorithms for State-Space Models. *Annals of the Institute of Statistical Mathematics* 62(1), 61-89, (2010).

[2] O. Cappe, E. Moulines and T. Ryden. *Inference in Hidden Markov Models.* Springer Science, New York, (2005).

[3] A.P. Dempster, N.M. Laird and D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *JRSS Ser. B* 39, 1-38, (1977).

[4] A. Doucet, N. de Freitas and N. Gordon. *Sequential Monte Carlo Methods in Practice.* Springer Science, New York, (2001).

[5] G.B. Durham and A.R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics* 20, 297-338, (2002).

[6] N.J. Gordon D.J. Salmond and A.F.M. Smith. Novel approach to nonlinear/non-aussian Bayesian estimation. *IEE Proc.-F* 140(2), 107-113, (1993).

[7] S.J. Julier, J.K. Uhlmann. Unscented Filtering and Nonlinear Estimation. In: *Proc. IEEE* 92(3), 401-422, (2004).

[8] N. Kantas, A. Doucet, S.S. Singh, J.M. Maciejowski. An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models. In:15th IFAC Symposium on System Identification, Saint-Malo, France, (2009).

[9] G. Kitagawa and S. Sato. Monte Carlo Smoothing and Self-Organising State-Space Model. In Ref. [4].

[10] L. Ljung. *System identification, theory for the user..* 2nd ed., NJ: Prentice Hall, (1999).

[11] H. Moradkhani and S. Sorooshian. General Review of Rainfall-Runoff Modeling. In Sorooshian, S. et al.(eds.) *Hydrologic Modelling and the Water Cycle.* Springer Science, New York, (2008).

[12] J. Olsson, O. Cappe, R. Douc, and E. Moulines. Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli* 14(1), 155-179, (2008).

[13] A.R. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22, 55-71, (1995).

[14] S. Särkkä. Continuous-Time and Continuous-Discrete-Time Unscented Rauch-Tung-Striebel Smoothers. *Signal Processing* 90(1), 225-235, (2010).

[15] T.B. Schön, A.,Wills and B. Ninness. System identification of nonlinear state-space models. *Automatica* 47, 39-49, (2011).

[16] R. van der Merve,A. Doucet,N. de Freitas, E. Wan. The unscented particle filter. Technical Report CUED/F-INFENG/TR 380, Cambridge Univ. Engineering Dept., (2000).