An Efficient Sampling Scheme for Dynamic Generalized Models

By ROMY R. RAVINES, HELIO S. MIGON and ALEXANDRA M. SCHMIDT *

Instituto de Matemática, Universidade Federal do Rio de Janeiro,

Rio de Janeiro, Brazil

Abstract

A multimove sampling scheme for the states parameters of non-Gaussian and non-linear dynamic models for univariate time series is proposed. This procedure follows the Bayesian framework, within a hybrid Gibbs sampling algorithm with some steps of the Metropolis-Hastings. This sampling scheme combines the conjugate updating approach for generalized dynamic linear models, with the backward sampling of the state parameters used in normal dynamic linear models. Conjugate updating backward sampling (CUBS) significantly reduces the computing time needed to attain convergence of the chains, and is also simple to implement. A quite extensive Monte Carlo study is conducted in order to compare the results obtained using CUBS with those obtained using some algorithms previously proposed in the Bayesian literature. This method is also applied and, compared, on two different real datasets.

Keywords Bayesian paradigm; Linear Bayes; Non-linear models; Monte Carlo Markov chain.

* *Address for correspondence*: Helio S. Migon, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, Rio de Janeiro, RJ, Brazil. CEP 21945-970. *E-mail*: migon@im.ufrj.br.

1. Introduction

Non-Gaussian and non-linear state-space models are widely used in applied statistics (e.g. finance, environmental areas, etc.), in particular, under the Bayesian paradigm. In this paper, discrete time dynamic models defined in the exponential family are considered. A general representation of these models is

$$y_t|\eta_t, \phi \sim \exp[\phi\{y_t\eta_t - a(\eta_t)\}]b(y_t, \phi), \qquad t = 1, \ldots, T, \tag{1a}$$

$$g(\mu_t) = \boldsymbol{F}_t(\boldsymbol{\psi}_1)'\boldsymbol{\theta}_t \tag{1b}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t(\boldsymbol{\psi}_2)\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}) \tag{1c}$$

$$\boldsymbol{\theta}_0|\boldsymbol{Y}^0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0),$$

where the vectors $\boldsymbol{\theta}_t$ are known as state parameters and are related through time via (1c), the system equation; $\boldsymbol{\theta}_0|\boldsymbol{Y}^0$ is the initial information; η_t is the natural parameter and $E[y_t|\eta_t] = \mu_t = \dot{a}(\eta_t)$, is related to $\boldsymbol{\theta}_t$ via a known link function, $g(\cdot)$; and $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ denote all the unknown quantities involved in the definition of $\boldsymbol{F}_t(\boldsymbol{\psi}_1)$ and $\boldsymbol{G}_t(\boldsymbol{\psi}_2)$, respectively. Also, $\boldsymbol{w}_t$ is the system error with evolution variance $\boldsymbol{W}$. Commonly $\boldsymbol{W}$ is a diagonal matrix and can vary through time. Under the Bayesian framework, the model specification is completed after assigning prior distributions to all unknowns. The resultant posterior for the parameters in (1) does not have a closed form, therefore inference cannot be made exactly. Moreover, the $\boldsymbol{\theta}_t$'s are highly correlated, *a posteriori*. This paper aims to propose an *efficient* (computationally fast and easy to implement), independent Metropolis-Hastings algorithm to sample from the full conditional distributions of $\boldsymbol{\theta}_t$, $t = 1, \ldots, T$, when Markov chain Monte Carlo (MCMC) methods are used.

Before the spread of the use of MCMC techniques, West et al. (1985), Kitagawa (1987) and Fahrmeir (1992), among others, proposed different methods to obtain approximations to the pos-

terior distribution in this setting. In recent years, as can be seen in Migon et al. (2005), several MCMC algorithms to sample from the posterior distributions of the state parameters have been proposed. In particular, Gamerman (1998) and Geweke & Tanizaki (2001) considered Bayesian approaches using a Metropolis-Hastings step into the Gibbs sampler. Since the critical problem of the Metropolis-Hastings algorithm is the choice of the proposal density, they were concerned with the construction of efficient ones. For instance, Gamerman (1998) suggested the use of an adjusted normal dynamic linear model in order to build these proposal densities. His approximation can be used under three different schemes: single and block sampling of the state parameters and single sampling from the system disturbances. He concluded that single-move sampling performs better than the multimove one, and the disturbance sampling scheme exhibits a lower autocorrelation structure, leading to faster convergence of the chains. However, this algorithm requires high computational effort, since it is complicated to code and might need a long time to complete a single iteration. On the other hand, Geweke & Tanizaki (2001) proposed to sample the state parameters separately, that is, in single moves. They focused in the formulation of different candidates for the proposal density function and performed a Monte Carlo study to access their properties. They concluded that the results are quite robust to the choice of the proposal density.

The proposal here is to sample the state parameters, $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T\}$ of (1), in block, that is, using multimove sampling, in an analogous scheme to the forward filtering and backward sampling (FFBS) of Frühwirth-Schnater (1994) and Carter & Kohn (1994) for Gaussian dynamic linear models. Both papers concluded that sampling the states simultaneously produces faster convergence to the posterior distribution compared to single-state sampling. The key of FFBS is to use the decomposition of the posterior density as the product of the smoothing (retrospective) distributions. In the Gaussian case, the quantities needed for the smoothing conditional densities are supplied by

the Kalman filter. Therefore, to obtain a sample from the joint full conditional distribution, first one computes the forward (on-line) moments (the forward filtering step), and then samples from the smoothed normal distributions (the backward sampling step).

The main idea of the proposal introduced here is to replace the first step of the FFBS by conjugate updating, introduced by West et al. (1985), for dynamic models in the exponential family. That is, the forward distributions are approximated by introducing the conjugate prior for the canonical parameter, and a backward sampling step is used as in the Gaussian case. This algorithm is called conjugate updating backward sampling (CUBS). The strength of this method lies in using the information about the correlation structure amongst the components of $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T\}$. The main advantages of the proposed scheme are: it is easy to implement and leads to a reduction of the computational time needed. Conjugate updating can be used within a MCMC algorithm such that unknown quantities of (1) can be sampled by using standard techniques.

The remainder of this paper is organized as follows. In Section 2. the proposed MCMC sampling scheme is described and its advantages are pointed out. In Section 3. the results obtained from CUBS and from some other algorithms previously proposed in the Bayesian literature are compared. That section presents an extensive Monte Carlo study using artificial datasets and the application of these algorithms to two real datasets, which illustrate clearly the efficiency of the method. Finally, Section 4. presents some concluding remarks.

2. Proposed sampling scheme

In this section we propose a new MCMC sampling scheme based on the combination of conjugate updating and backward sampling. We also point out its main advantages in detail. To fix the notation, let $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T\}$ be the vector of states, $\boldsymbol{Y}^t = \{y_1, \ldots, y_t\}$ be the set of observations

up to time t, and $\Phi = (\phi, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{W})$ the vector of all remaining unknown quantities in (1). According to the specification in (1) and assuming independent priors for $\phi, \Psi_1, \Psi_2, \boldsymbol{W}$ and for the initial information about $\boldsymbol{\theta}_0$, the posterior distribution is

$$p(\boldsymbol{\Theta}, \boldsymbol{\theta}_0, \Phi | \boldsymbol{Y}^T) \propto \prod_{t=1}^{T} p(y_t|\boldsymbol{\theta}_t, \eta_t, \phi) p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{F}_t(\boldsymbol{\psi}_1), \boldsymbol{G}_t(\boldsymbol{\psi}_2), \boldsymbol{W}) p(\theta_0) p(\boldsymbol{\psi}_1) p(\boldsymbol{\psi}_2) p(\boldsymbol{W}) p(\phi). \quad (2)$$

Since the distribution in (2) does not have a known closed form, we propose the use of MCMC methods to obtain samples from it. The main concern is to sample from the full conditional posterior of $\boldsymbol{\Theta}$ which also does not have a closed form. In particular, we use a hybrid algorithm, a Gibbs sampling with some steps of the Metropolis-Hastings. Therefore, the focus is on the construction of an efficient proposal distribution in the Metropolis-Hastings step, when sampling the components of $\boldsymbol{\Theta}$.

In the eighties, before the advent of the use of MCMC schemes, West et al. (1985), proposed a sequential algorithm to assess the posterior distribution of the states of dynamic generalized linear models (DGLM). The conjugate updating algorithm is based on an approximation performed at the prior level using linear Bayes estimation. In DGLM, the prior on the states is partially specified through the first and second moments, say $\boldsymbol{a}_t$ and $\boldsymbol{R}_t$, and is used to determine r_t and s_t, the parameters of the conjugate prior (CP) for η_t, the natural parameter of $p(y_t|\eta_t, \Phi)$ at time t, that is: $\eta_t|\boldsymbol{Y}^{t-1} \sim CP(r_t, s_t)$. This is done by solving a relative simple system of two equations as shown in the appendix (see Table 6). The parameters of the conjugate posterior for the η_t are easily obtained after y_t is observed. The information in this posterior is then propagated to the posterior distribution of the states. Since this distribution is not fully specified, West et al. (1985) proposed the use of linear Bayes estimation with regard to a quadratic loss function to derive an approximation of the posterior distribution of the state vector. More specifically, the approximation

is as follows:

$$p(\boldsymbol{\theta}_t \mid \boldsymbol{Y}^t, \Phi) \propto p(\boldsymbol{\theta}_t \mid \boldsymbol{Y}^{t-1}, \Phi) p(y_t \mid \boldsymbol{\theta}_t, \Phi)$$
$$= \int p(\boldsymbol{\theta}_t \mid \eta_t, \boldsymbol{Y}^{t-1}, \Phi) \underbrace{p(\eta_t \mid \boldsymbol{Y}^{t-1}, \Phi) p(y_t \mid \eta_t, \Phi)}_{\text{Conjugate Analysis}} d\eta_t.$$

Since the first term in the integrand is partially available, the linear Bayes principle is used to obtain

$$p(\boldsymbol{\theta}_t \mid \boldsymbol{Y}^t, \Phi) \propto \int \underbrace{p(\boldsymbol{\theta}_t \mid \eta_t, \boldsymbol{Y}^{t-1}, \Phi)}_{\text{Linear Bayes' estimation}} p(\eta_t \mid \boldsymbol{Y}^t, \Phi) d\eta_t$$
$$\cong [\boldsymbol{m}_t, \boldsymbol{C}_t],$$

where $[\boldsymbol{m}_t, \boldsymbol{C}_t]$ indicates that $p(\boldsymbol{\theta}_t \mid \boldsymbol{Y}^t, \Phi)$ is partially specified by the first and second moments given by

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t \boldsymbol{F}_t (f_t^* - f_t)/q_t$$
$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{F}_t' \boldsymbol{R}_t (1 - q_t^*/q_t)/q_t,$$

where $\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}, \boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t$ and (f_t, q_t) and (f_t^*, q_t^*) are, respectively, the prior and posterior, mean and variance of η_t. It is worth pointing out that the well-known Kalman filter algorithm is recovered under the normal assumption for y_t and the identity link function.

In the normal DLM context, it is known that the single move implementation of the MCMC is strongly inefficient due to the high correlation among the state parameters. To overcome this situation, Frühwirth-Schnater (1994) and Carter & Kohn (1994) proposed, only for the Gaussian case, an algorithm to sample the whole $\boldsymbol{\Theta}$ in (1), in a multimove step. Their approach is based in the well known decomposition of the state's joint full conditional distribution

$$p(\boldsymbol{\Theta} | \boldsymbol{Y}^T, \Phi) = p(\boldsymbol{\theta}_T | \boldsymbol{Y}^T, \Phi) \prod_{t=1}^{T-1} p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^t, \Phi). \tag{3}$$

Since $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^t, \Phi) \propto p(\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t, \boldsymbol{Y}^t, \Phi)p(\boldsymbol{\theta}_t \mid \boldsymbol{Y}^t, \Phi)$, it is straightforward to show that $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \boldsymbol{Y}^t, \Phi)$ is $N(\boldsymbol{m}_t^s, \boldsymbol{C}_t^s)$, where

$$\boldsymbol{m}_t^s = \boldsymbol{m}_t + \boldsymbol{C}_t\boldsymbol{G}_t'(\boldsymbol{G}_t\boldsymbol{C}_t\boldsymbol{G}_t' + \boldsymbol{W}_t)^{-1}(\boldsymbol{\theta}_{t+1} - \boldsymbol{G}_t\boldsymbol{m}_t) \tag{4a}$$

$$\boldsymbol{C}_t^s = \boldsymbol{C}_t - \boldsymbol{C}_t\boldsymbol{G}_t'(\boldsymbol{G}_t\boldsymbol{C}_t\boldsymbol{G}_t' + \boldsymbol{W}_t)^{-1}\boldsymbol{G}\boldsymbol{C}_t, \tag{4b}$$

and where $(\boldsymbol{m}_t, \boldsymbol{C}_t)$ are respectively the first and second moments obtained from the Kalman filter. Therefore, the FFBS algorithm consists mainly of sampling sequentially (backwardly) the elements of $\boldsymbol{\Theta}$.

Based on the two ideas briefly explained above, when sampling from the full conditional distribution of $\boldsymbol{\Theta}$, $p(\boldsymbol{\Theta}|\cdot)$, described in (3), here we propose that a candidate value, $\boldsymbol{\Theta}^p = \{\boldsymbol{\theta}_1^p, \ldots, \boldsymbol{\theta}_T^p\}$, can be sampled from a multivariate normal, $N(\boldsymbol{m}^s, \boldsymbol{C}^s)$, with $\boldsymbol{m}^s$ and $\boldsymbol{C}^s$ as given in (4b). More specifically, in a first step, the filtering densities moments, $\boldsymbol{m}, \boldsymbol{C}$, are determined by a conjugate updating running from $t = 1$ to $t = T$, instead of using the Kalman filter used in the FFBS. After that, each $\boldsymbol{\theta}_t^p$ is sequentially sampled, from time T to time 1, from its smoothing distribution, as in the Gaussian case. So basically, in order to implement this strategy, one needs to complete the model by introducing a conjugate prior (CP) for the natural parameter in the exponential family.

Therefore, the main idea of this scheme is as follows. At the $i-$th iteration of the MCMC scheme, let $\boldsymbol{m}_t^{(i)}$ and $\boldsymbol{C}_t^{(i)}, t = 1, \ldots, T$ be the first two moments of the filtering distributions obtained via the conjugate updating and $\boldsymbol{m}_t^{s(i)}$ and $\boldsymbol{C}_t^{s(i)}$ be the first two moments of the smoothing distributions. Then a sample of $\boldsymbol{\Theta}$ is obtained as follows:

1. Sample $\boldsymbol{\theta}_T^p$ from $N(\boldsymbol{m}_T^{(i)}, \boldsymbol{C}_T^{(i)})$;

2. Sample $\boldsymbol{\theta}_t^p$ from $N(\boldsymbol{m}_t^{s(i)}, \boldsymbol{C}_t^{s(i)})$, $\quad t = T-1, \ldots, 1$; where

$$\boldsymbol{m}_t^{s(i)} = \boldsymbol{m}_t^{(i)} + \boldsymbol{C}_t^{(i)}\boldsymbol{G}_t'(\boldsymbol{G}_t\boldsymbol{C}_t^{(i)}\boldsymbol{G}_t' + \boldsymbol{W})^{-1}(\boldsymbol{\theta}_{t+1}^p - \boldsymbol{G}_t\boldsymbol{m}_t^{(i)})$$

$$\boldsymbol{C}_t^{s(i)} = \boldsymbol{C}_t^{(i)} - \boldsymbol{C}_t^{(i)}\boldsymbol{G}_t'(\boldsymbol{G}_t\boldsymbol{C}_t^{(i)}\boldsymbol{G}_t' + \boldsymbol{W})^{-1}\boldsymbol{G}_t\boldsymbol{C}_t^{(i)}.$$

3. Set $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^p$ with probability π and $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^{(i-1)}$ with probability $1-\pi$, where $\pi = \min(1, A)$ and

$$A = \frac{\omega(\boldsymbol{\Theta}^p)}{\omega(\boldsymbol{\Theta})}, \tag{5}$$

where $\omega(\boldsymbol{\Theta}^p) = \dfrac{p(\boldsymbol{\Theta}^p|\boldsymbol{Y}^T, \Phi)}{q(\boldsymbol{\Theta}^p|\boldsymbol{m}^{s(i)}, \boldsymbol{C}^{s(i)})}$, and $q(\cdot)$ is a proposal density with moments $\boldsymbol{m}$ and $\boldsymbol{C}$.

The CUBS procedure is in fact a kind of data augmentation algorithm where the natural parameters $\{\eta_1, \ldots, \eta_T\}$ play the role of a set of latent variables that help efficient simulation of the state parameters. However, since several approximations are made to attempt to use a block sampler, this result is actually an alternative procedure to construct an independent proposal density for a Metropolis-Hastings step. Additionally, although in this paper the use of a normal proposal distribution is emphasized, one could explore many others distributions, e.g. a t-Student proposal distribution.

The proposal distribution obtained via CUBS provides a good approximation to the target distribution $p(\boldsymbol{\Theta}|\cdot)$, due to the fact that conjugate updating is generally very accurate, so the acceptance rates obtained are always reasonable. Another advantage is that CUBS is very easy to implement, diminishing the chances of involuntary mistakes in programming the routines. In some cases, numerical methods are needed to compute the exact values of r_t and s_t. However, one can use approximations that work well and contribute to reduce the computational time burden.

Finally, given the state parameters, $\boldsymbol{\Theta}$, samples from the posterior distribution of other parameters in the model, Φ, can be obtained using standard techniques, like a Metropolis-Hastings step,

slice sampling (Neal, 2003), adaptive rejection Metropolis sampling, ARMS (Gilks & Wild, 1992; Gilks et al., 1995), etc. This step will not affect the convergence of the state sampler proposed here, but it can significatively affect the computational time needed to reach convergence of all the chains (the whole algorithm). However, in most cases, this computational time will be shorter than the MCMC algorithms proposed by Gamerman (1998) and Geweke & Tanizaki (2001). For ease of understanding, the MCMC scheme is summarized as follows.

Algorithm

1. **Initialization**: set initial values $\boldsymbol{\Theta}^{(0)}, \Phi^{(0)}$ and iteration counter $i = 1$;

2. **Sample** $\boldsymbol{\Theta}^{(i)}$ using CUBS:

   (a) Compute the moments of $p(\boldsymbol{\theta}_t|\boldsymbol{Y}^t, \Phi^{(i-1)})$, $\boldsymbol{m}_t^{(i)}$ and $\boldsymbol{C}_t^{(i)}$, $t = 1, \ldots, T$; with conjugate updating;

   (b) Sample $\boldsymbol{\Theta}^p$ with the backward sampling.

      i. Sample $\boldsymbol{\theta}_T^p$ from $N(\boldsymbol{m}_T^{(i)}, \boldsymbol{C}_T^{(i)})$

      ii. Sample $\boldsymbol{\theta}_t^p, t = T-1, \ldots, 1$, from $N(\boldsymbol{m}_t^{s(i)}, \boldsymbol{C}_t^{s(i)})$;

   (c) Set $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^p$ with probability π and $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^{(i-1)}$ with probability $1 - \pi$, where $\pi = \min(1, A)$ and A as defined in (5).

3. **Sample** $\Phi^{(i)}$ using, in general, a Metropolis-Hastings step;

4. **Update**: set $i = i + 1$ and return to 2 once convergence is attained and the needed sample size is reached.

3. Comparing different sampling schemes

3·1. A Monte Carlo study

This section describes a Monte Carlo study aiming to compare the use of CUBS within a MCMC algorithm with some single move sampling schemes proposed by Gamerman (1998) and Geweke & Tanizaki (2001).

3·1..1 Simulation of the artificial data

For the simulation procedure, we used a fairly simple model just to focus the comparisons in the samples of the state parameters generated by several known methods. The first order dynamic Poisson model shown in (6), with $\theta_0 = 0.50$ and $W = 0.01$, was used to generate three groups of 100 time series each. The groups differed just in the length of the time series. That is, we assumed that

$$y_t \sim \text{Poisson}(\lambda_t), \qquad t = 1, \ldots, T, \tag{6a}$$

$$\eta_t = \log(\lambda_t) = \theta_t \tag{6b}$$

$$\theta_t = \theta_{t-1} + w_t, \qquad w_t \sim N(0, W) \tag{6c}$$

$$\theta_0 | \boldsymbol{Y}^0 \sim N(m_0, C_0), \tag{6d}$$

and generated random numbers of θ_t based on (6c), obtaining $\boldsymbol{Y}^T$ based on (6a) for the cases $T = 50, 100, 300$. After that, we performed the MCMC routines to obtain samples from the posterior distributions of $\boldsymbol{\Theta}$ and W. In order to complete the model, we assigned an inverse gamma prior distribution with both parameters equal to 0.001, that is, $IG(0.001, 0.001)$, for W, and a zero mean normal with some large fixed variance for θ_0.

3·1..2 Sampling schemes under study

The unknown quantities in (6) are $(\theta_0, \theta_1 \ldots, \theta_T, W)$. Since the full conditional distribution of W is an inverse gamma, one can sample directly from it, while for the state parameters one can use a Metropolis-Hastings step with different proposed algorithms.

⋆ From Gamerman (1998)

As was mentioned in Section 1., Gamerman (1998) suggested the use of an adjusted dynamic Gaussian model such as

$$\tilde{y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, \tilde{V}_t) \tag{7a}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t), \tag{7b}$$

where the adjusted observations $\tilde{y}_t$ and its associated variances $\tilde{V}_t$ are

$$\tilde{y}_t = \eta_t + (y_t - \mu_t)g'(\mu_t)$$

$$\tilde{V}_t = \ddot{a}(\eta_t)[g'(\mu_t)]^2.$$

In the case of (6) $\tilde{y}_t = \log(\lambda_t) + \lambda_t^{-1}(y_t - \lambda_t)$ and $\tilde{V} = \lambda_t^{-1}$. Based on (7), Gamerman (1998) proposed two different sampling schemes, as described below:

- **Proposal I**. Sample w_t^p, a candidate value for the system disturbance w_t in (6c), from the full conditional distribution obtained considering the following re-parametrization of (7):

$$\tilde{y}_t = \boldsymbol{F}_t \sum_{j=1}^{t} \boldsymbol{G}^{t-j}\boldsymbol{w}_j + v_t, \quad \boldsymbol{G}_t = \boldsymbol{G}\,\forall t, \quad t = 2, \ldots, T,$$

$$v_t \sim N(0, \tilde{V}_t), \quad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}), \quad \boldsymbol{w}_1 \sim N(\boldsymbol{m}_1, \boldsymbol{C}_1)$$

- **Proposal II**. Sample $\boldsymbol{\theta}_t^p$, a candidate value for θ_t in (6), from the full conditional distribution of $\boldsymbol{\theta}_t$ in (7) (with a normal proposal density).

Actually a third alternative was proposed, which consists of applying the FFBS algorithm in (7). However, this proposal is not recommended by the author due the low acceptance rates it produces, therefore, it is not investigated in this paper.

⋆ From Geweke & Tanizaki (2001)

Geweke & Tanizaki (2001) tested different proposal densities and recommended sampling the state parameters separately, that is, sampling the state parameters one at a time. Three of the schemes they proposed are:

- **Proposal I**. Sample $\boldsymbol{\theta}_t^p$ from the density function obtained from the system equation. Therefore, $\boldsymbol{\theta}_t^p$ is sampled from $N(\boldsymbol{G}_t(\boldsymbol{\psi}_2)\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_t)$, in this case, from (6c), $N(\theta_{t-1}, W_t)$.

- **Proposal II**. Sample $\boldsymbol{\theta}_t^p$ from a normal density with mean and variance based on the extended Kalman smoothed estimates at time t. The extended Kalman filter and smoother is an algorithm for exponential family state space models. It consists of applying the Kalman filter and smoother in an approximating Gaussian dynamic model, as defined in (7). Specifically, in order to approximate the posterior mode of the states, the response function (the inverse of the link function) is linearized and the observation model (1) is approximated with a Gaussian observation model. In the case of (6), the approximating normal model is constructed with $\tilde{y}_t = \exp(-\eta_t)y_t + \eta_t - 1$ and $\tilde{V} = \exp(-\eta_t)$.

- **Proposal III**. Sample $\boldsymbol{\theta}_t^p$ from a normal density with mean based in a random walk and variance based on the extended Kalman smoothed estimates at time t.

In this paper, we compare the two proposed single moves alternatives from Gamerman (1998), and the three alternatives from Geweke & Tanizaki (2001) described above. We also considered the smoothed moments obtained after the use of conjugate updating to construct individual proposal

densities to θ_t. This was done to evaluate the single move sampling, which derives straightforwardly from the proposed scheme. Also this might be considered as an analogous alternative to proposal II of Geweke & Tanizaki (2001). Therefore, the following seven alternatives to construct the proposal density are compared:

I. Conjugate updating and backward sampling (CUBS): Multimove sampling, as explained in Section 2..

II. Conjugate updating - single-move: The proposal is a normal density with mean and variance based on the smoothed moments of the conjugate updating at time t;

III. From Gamerman (1998) - Proposal I: Single move, sampling from the system disturbances;

IV. From Gamerman (1998) - Proposal II: Single move, sampling from the state parameters;

V. From Geweke & Tanizaki (2001) - Proposal I: The proposal is the density function obtained from the system equation.

VI. From Geweke & Tanizaki (2001) - Proposal II: The proposal is a normal density with mean and variance based on the extended Kalman smoothed estimates at time t.

VII. From Geweke & Tanizaki (2001) - Proposal III: The proposal is a normal density with mean based on a random walk (mean equals the current value of the chain) and variance based on the extended Kalman smoothed estimates at time t.

3·1..3 Performance comparison criteria

In order to compare the sampling schemes, we computed the inefficiency factor and the root mean square error (RMSE) for $\boldsymbol{\Theta}$ and $\boldsymbol{Y}$. The inefficiency factor is given by

$$\text{INEF} = 1 + 2\sum_{j=1}^{n-1} \frac{n-j}{n}\rho_j,$$

where ρ_j is the autocorrelation of lag j for the values of a specific chain of length n ($n \to \infty$). INEF aims to evaluate the efficiency of the MCMC estimation. The larger INEF is, the less efficient the sampling scheme is (see Gamerman & Lopes (2006) for further details). Let RMSE_{θ_t} be the root square error for $\theta_{rep,t}$ defined by

$$\text{RMSE}_{\theta_t} = \Big(\frac{1}{K}\sum_{k=1}^{K}\big(\hat{\theta}_{rep,t}^{(k)} - \theta_t^{(k)}\big)^2\Big)^{1/2},$$

where $\hat{\theta}_{rep,t}^{(k)}$ is the posterior sample mean and $\theta_t^{(k)}$ is the true value of the $t-$th latent variable (state) of the $k-$th artificial dataset ($K = 100$). The mean and standard deviation of RMSE_{θ_t}, are given, respectively, by

$$mean(\text{RMSE}_{\boldsymbol{\Theta}}) = \frac{1}{T}\sum_{t=1}^{T}\text{RMSE}_{\theta_t}$$

and

$$sd(\text{RMSE}_{\boldsymbol{\Theta}}) = \Big(\frac{1}{T}\sum_{t=1}^{T}\big(\text{RMSE}_{\theta_t} - mean(\text{RMSE}_{\boldsymbol{\Theta}})\big)^2\Big)^{1/2}.$$

In other words, $mean(\text{RMSE}_{\boldsymbol{\Theta}})$ and $sd(\text{RMSE}_{\boldsymbol{\Theta}})$ are summaries of the RMSE_{θ_t} obtained considering all the replications of all instants for each dataset. The values for $\text{RMSE}_{\boldsymbol{Y}}$ are obtained in an analogous way.

Other complementary criteria reported are: the computational time elapsed and the acceptance rates. The time series were generated in `R` (see R Development Core Team (2005)) and the MCMC routines were written in `Ox` version 3.20 (see Doornik (2002))[1].

[1]The computational routines are available for free from http:/www.dme.ufrj.br/romy.

3·1..4 Results

The results showed in this section are based on samples of size 10000 obtained from chains which were run for 50000 iterations, where the first 40000 were discarded as the burn-in period. The only exception is Scheme **III**, whose chains were run for 15000 iterations and the burn-in was 10000. We adopted this strategy following the results reported by Gamerman (1998) about its fast convergence and the high time needed to complete each iteration. A thinning interval was not used to produce the summaries. This is because one of the goals of this study was to measure the autocorrelation each method produces, one iteration after the other.

An interesting result is that under Scheme **I**, inefficiencies exhibited by the state parameters' chains are independent of the position in time, that is, the inefficiencies of the chains of θ_t for t at the beginning, middle or end of the time series are almost the same, unlike in the other sampling schemes. Also, inefficiencies do not depend on the time series' size. Although this is a multimove step, the acceptance rates were always reasonable, on average, between 26% and 46%. The single-move schemes (**II**, **V**, **VII** and **VII**) were the least efficient ones. Although the empirical autocorrelation functions are not shown here, in most of the cases, the ones from Scheme **I** are as good as those of Scheme **III**.

From Table 1 it can be noticed that the RMSE's of all the schemes are very similar, as expected, for both, the response ($\boldsymbol{Y}$) and state ($\boldsymbol{\Theta}$) variables. The acceptances rates of Schemes **III** and **IV** are the highest, more than 90% in all cases, while the acceptance rates of **V**, **VI** and **VII** are always between 30% and 50%. The acceptance rates of Scheme **I** are, on average, similar to the other single move schemes. In some series its acceptance rate is the smallest, as expected, since it corresponds to a multimove step. However, it is *far different from zero*, contrary to the result reported by Gamerman (1998) when using his proposal in a multimove step. In addition, as also expected, it

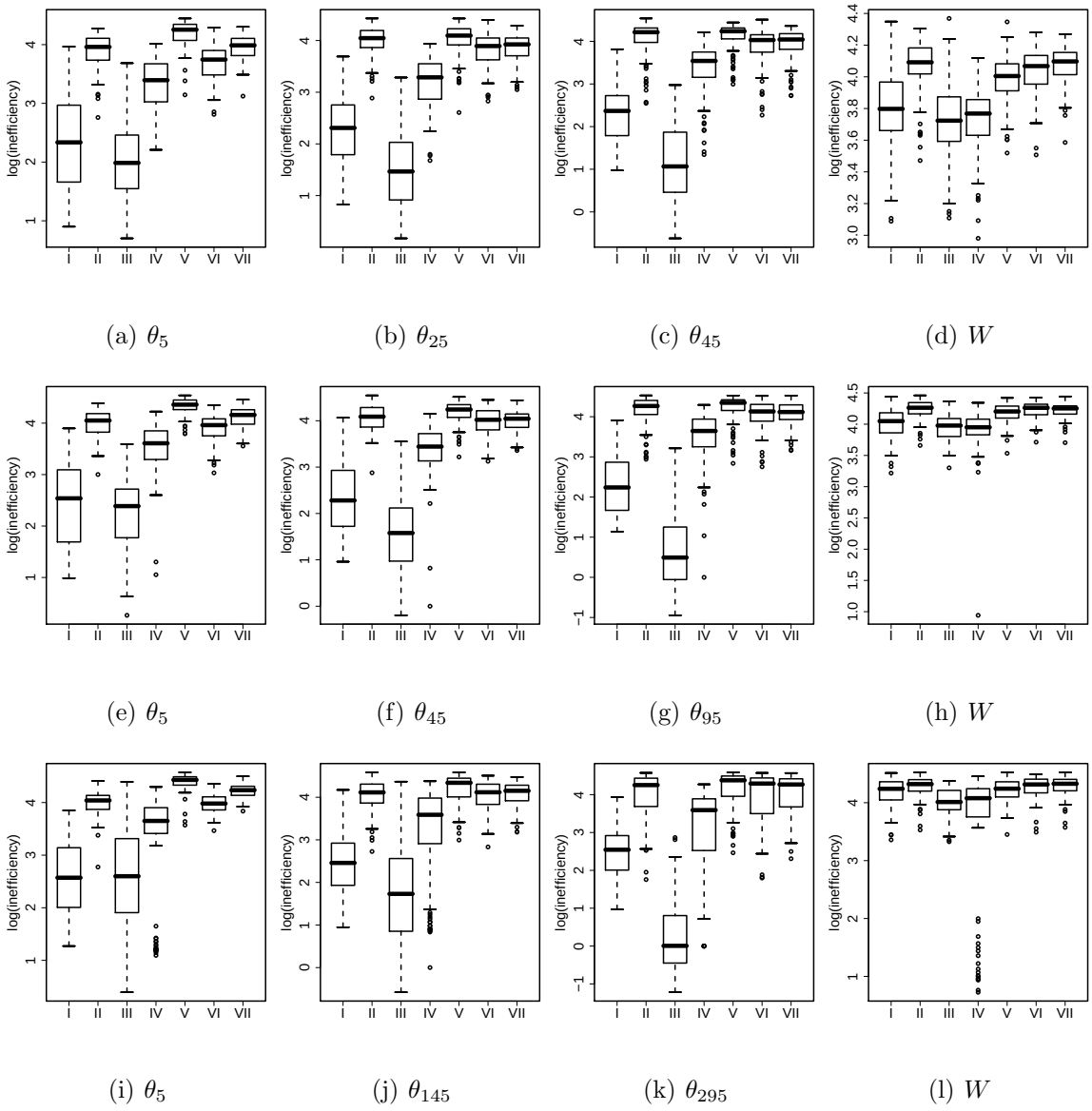
can be noted that the global acceptance rate decreases as the length of the time series increases. The computational times elapsed for Schemes **I**, **II**, **V**, **VI** and **VII** are similar, however, Scheme **III** needed almost the same time to complete *only* 10% of the number of iterations completed by the other schemes. Attention must be paid to the results obtained under Scheme **IV** for the time series of size 300. The $\text{RMSE}_{\boldsymbol{Y}}$'s were much greater than the errors under other schemes. This can be partially explained by the acceptance rate obtained with this method. Although its mean was 95.5%, its standard deviation was 14.12, indicating that for some observations in some replications the acceptance rate was small (almost zero), leading to a poor adjustment. The latter is reflected in the $\text{RMSE}_{\boldsymbol{\Theta}}$'s, which are also the greatest ones.

Considering all the 100 time series, Figure 1 shows the summary, through a box plot, of the log inefficiencies of the sampled chains of the states, θ_t, at times $t = 5, 25, 45$ for $T = 50$, $t = 5, 45, 95$ for $T = 100$, $t = 5, 145, 295$ for $T = 300$ and W for all the cases. From Figure 1, it can be noticed that for series of sizes $T = 50$ and $T = 100$, Schemes **V**, **VI** and **VII** exhibit the highest inefficiencies, while for Scheme **III** the inefficiency decreases through time. The box plots under Scheme **I** show that with this method the best inefficiencies were obtained at $t = 5$ and $t = 25$, and the second best at $t = 45$. In the case of $T = 300$, Scheme **III** performed slightly better than Scheme **I** in the middle of the time series, e.g., $t = 145$, and better at the end, e.g., $t = 295$. Nevertheless, Scheme **I** was better than the rest of the single-move sampling methods used.

Figure 2 presents the box plots of the logarithm of the medians of the samples from the posterior distributions of W. The estimates are reasonable under all the schemes. Nevertheless, under Scheme **V** it seems that the state's variance is always underestimated. This could be due to a poor mixing of the chains. Additionally, it is worth commenting that results not shown here revealed that when the value of W is larger, for example, $W = 0.1$, Schemes **III** and **IV** could have problems to

Table 1: **Root Mean Square Error (RMSE), Acceptance Rate (mean of $T \times K$ rates) and Computational Time (mean of $K = 100$ elapsed times) for time series of size** $T = 50, 100, 300$.

	$\text{RMSE}_{\boldsymbol{Y}}$		$\text{RMSE}_{\boldsymbol{\Theta}}$		Acceptance rate		Time
Scheme	mean	sd	mean	sd	mean	sd	
			$T = 50$				
I	1.2525	0.1095	0.2366	0.0361	42.6297	–	280.30
II	1.2766	0.1169	0.2342	0.0263	33.5253	7.1575	214.98
III	1.2599	0.1107	0.2492	0.0513	97.2529	1.2287	629.08
IV	1.2443	0.1148	0.2583	0.0290	98.1449	0.6814	120.79
V	1.3155	0.1322	0.2376	0.0345	51.4230	4.5316	89.04
VI	1.2403	0.1137	0.2611	0.0280	37.7968	6.3019	248.25
VII	1.2365	0.1109	0.2596	0.0303	44.6136	5.8161	219.65
			$T = 100$				
I	1.3099	0.1334	0.2244	0.0277	38.4543	–	334.25
II	1.3272	0.1435	0.2266	0.0243	32.3509	6.9867	257.16
III	1.3137	0.1347	0.2335	0.0409	97.9422	0.8513	1533.09
IV	1.4342	0.3821	0.2542	0.0491	98.7127	3.2285	147.62
V	1.3831	0.1526	0.2861	0.0388	51.4612	4.0418	104.54
VI	1.3128	0.1420	0.2357	0.0199	34.2718	6.8640	301.94
VII	1.3078	0.1405	0.2304	0.0216	42.0167	6.0428	274.08
			$T = 300$				
I	1.6085	0.3039	0.2281	0.0352	31.3526	–	247.76
II	1.6158	0.3065	0.3798	0.2213	32.1778	9.4706	314.09
III	1.6107	0.2980	0.2273	0.0376	98.4285	0.6872	2158.93
IV	4.5229	3.7538	0.5459	0.2526	95.5152	14.1218	163.87
V	1.7110	0.3295	0.4718	0.1330	51.3656	4.1081	74.44
VI	1.6076	0.3012	0.2365	0.0344	32.8817	8.9858	371.31
VII	1.6066	0.3025	0.2739	0.0876	40.3615	7.8025	338.89



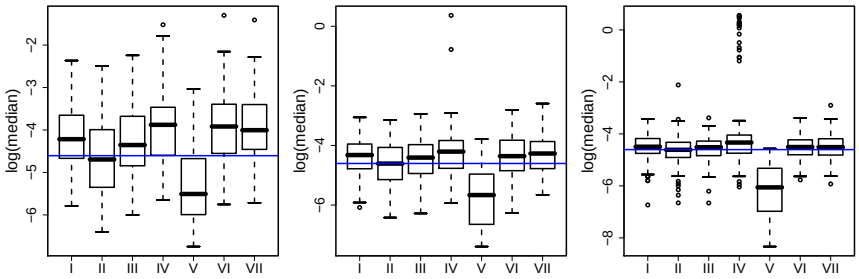
(a) θ_5 (b) θ_{25} (c) θ_{45} (d) W

(e) θ_5 (f) θ_{45} (g) θ_{95} (h) W

(i) θ_5 (j) θ_{145} (k) θ_{295} (l) W

Figure 1: **Box plots of the (log)inefficiencies of θ_t, at three different instants, and W: (a)–(d) $T = 50$, (e)–(h) $T = 100$ and (i)–(l) $T = 300$.**

approximate the target posterior distribution, causing zero acceptance rates.



(a) $T = 50$ (b) $T = 100$ (c) $T = 300$

Figure 2: **Box plots of the (log)medians of the samples from the posterior distribution of W.**

3·2. Modeling rainfall in Tokyo

We used all the sampling schemes described in the previous section to fit a dynamic binomial model to the number of occurrences of rainfall over 1mm in Tokyo for each day during 1983-1984. This example was analyzed by Kitagawa (1987) and Gamerman (1998), among others. The goal here is to estimate the probability of success (occurrence of rainfall) for each calendar day. The model considered follows a binomial response with success probability π_t, such that $logit(\pi_t)$ follows a first-order dynamic trend, that is

$$y_t|n, \pi_t \sim \text{Binomial}(n, \pi_t), \quad t = 1, \ldots, 366; \; n = 1 \text{ if } t = 60 \text{ and } n = 2 \text{ if } t \neq 60,$$

$$\text{logit}(\pi_t) = \theta_t \tag{8a}$$

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim N(0, W). \tag{8b}$$

One chain of 10^6 iterations was run and after a burn-in period of 90000, every 50th iteration sample was stored (except for Scheme **III**). In this particular case, due to the small values of the response variable, we used numerical methods to compute the moments of the conjugate prior of π_t under CUBS.

Table 2 shows the $\text{RMSE}_{\boldsymbol{Y}}$, the mean and standard deviation of the acceptance rates, considering all θ_t. It is clear, from these figures that the goodness of fit obtained with all the schemes was very similar. The lowest acceptance rate corresponds to the multimove sampling schemes, while the highest ones correspond to Schemes **III** and **IV**. Figure 3 shows the autocorrelation functions of θ_t at $t = 50, 150, 250, 350$. It can be seen that the samples with the lowest autocorrelations were obtained under Schemes **I** and **III**, while the samples under Schemes **IV**, **V**, and **VI** exhibit a higher autocorrelation.

The box plots of the log inefficiencies computed for the whole state vector are shown in Figure 4(a). From this, it can be noticed that the best results were also obtained under Schemes **I** and **III** (chains with the lowest inefficiencies). However, Scheme **III** exhibits a high dispersion associated with this measure. As observed from our Monte Carlo study, this sampling scheme tends to be more efficient when sampling those state parameters which are closer to the end of the time series than those which are on the beginning.

Table 3 shows some summaries of the samples from the posterior distribution of W and Figure 4(b) shows the box plot of the posterior samples of the variance W. The posterior medians obtained are between 0.04 and 0.10 (except for scheme **V**, which, from our Monte Carlo study, is expected to underestimate the state variance).

3·3. Modeling runoff as a function of rainfall

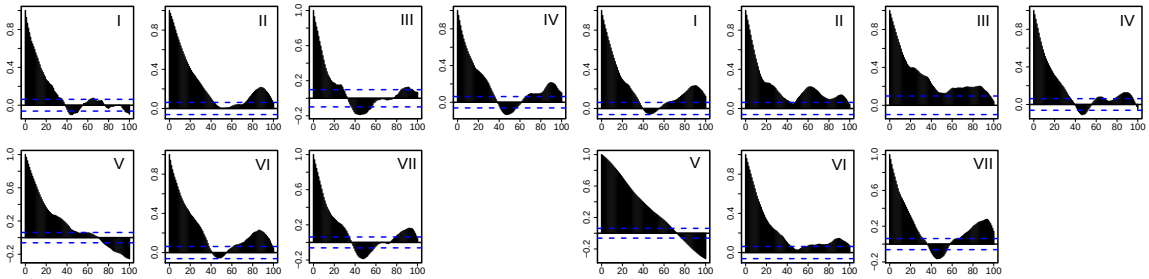
Based on the previous results, it can be concluded that the two best schemes were **I** and **III**. In this section we compared the performance of both schemes in a non-linear dynamic model. Specifically, we fitted a non-linear dynamic model for runoff as a function of rainfall for data from the Fartura River in São Paulo, Brazil. The time series are from January 1960 to December 1964. The dataset

Table 2: **RMSE$_{\boldsymbol{Y}}$ and acceptance rates for the Tokyo rainfall data.**

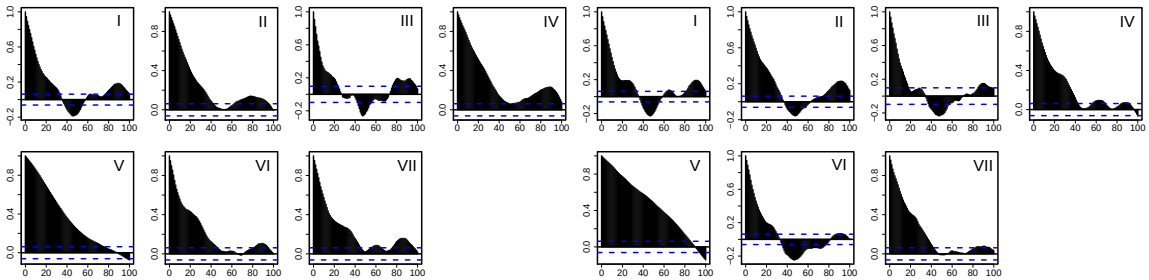
	RMSE$_{\boldsymbol{Y}}$		Acceptance rate	
Scheme	mean	sd	mean	sd
I	0.4698	0.2814	9.6400	–
II	0.4928	0.2825	29.9859	3.3905
III	0.4714	0.2836	98.0405	1.0536
IV	0.4772	0.2831	99.6911	4.9041
V	0.5771	0.2958	52.0549	1.8006
VI	0.4686	0.2891	26.6605	8.1579
VII	0.4793	0.2835	43.5408	4.0982

Table 3: **Summaries from the posterior samples of W for the Tokyo rainfall data.**

Scheme	mean	sd	2.50%	median	97.50%	log(ineff)
I	0.0919	0.0477	0.0235	0.0841	0.2187	3.5850
II	0.0492	0.0392	0.0079	0.0407	0.1496	3.7911
III	0.0903	0.0489	0.0254	0.0830	0.1984	2.8077
IV	0.0716	0.0443	0.0179	0.0618	0.1777	2.4307
V	0.0004	0.0003	0.0001	0.0003	0.0012	4.1382
VI	0.0838	0.0392	0.0326	0.0756	0.1787	4.1091
VII	0.0662	0.0380	0.0192	0.0570	0.1646	3.1724

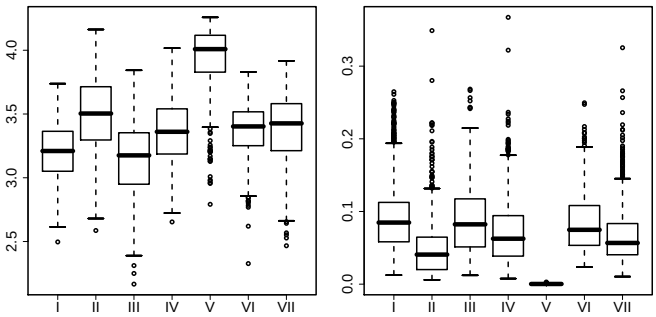


(a) $t = 50$ (b) $t = 150$



(c) $t = 250$ (d) $t = 350$

Figure 3: **Autocorrelations of $\theta_t, t = 50, 150, 250, 350$. for the Tokyo rainfall data.**



(a) (log)inefficiency of $\boldsymbol{\theta}$ (b) W

Figure 4: **Box plots of the (log)inefficiencies of $\theta_t, t = 1, \ldots, 366$ and posterior samples of W for the Tokyo rainfall data.**

is available in Monteiro (1992).

Let y_t and x_t denote the runoff and rainfall at month t. Since y_t is a continuous positive variable, it can be assumed that it follows a gamma distribution, so the proposed model is

$$y_t \sim \text{Gamma}(\mu_t, \nu), \quad t = 1, \ldots, T, \tag{9a}$$

$$\log(\mu_t) = \alpha + E_t \tag{9b}$$

$$E_t = \rho E_{t-1} + \gamma x_t + \omega_t; \quad \omega_t \sim N(0, W) \tag{9c}$$

where μ_t is the mean and ν is the shape parameter of the gamma distribution. α represents the basic level or stream flow, ρ represents the recharge factor or permanence rate of the rainfall effect, γ is the velocity of response to precipitation, related to soil saturation and ω_t is the system error that characterizes the stochastic transfer function (Migon & Monteiro, 1997).

The full conditional distributions of γ and W in (9) are normal and inverse gamma, respectively, so they can be sampled directly. The ones of α and ρ do not have a known closed form, therefore one can use a Metropolis-Hastings step to generate samples of α and the slice sampling method to generate samples of ρ. Given ρ and γ, the previous generalized dynamic linear model setting is obtained. Therefore, in order to obtain samples from the posterior distribution of $\boldsymbol{E}$, one can use either scheme **I** or **III**, in independent routines.

Aiming to do a fair comparison, we ran two parallel chains under each scheme for a period of 10 minutes, that is, we ran the routines for a fixed period of time and observed the length of the generated chains. Table 4 shows the number of iterations performed under each scheme. It can be observed that under algorithm **I**, on average, 77477 iterations were performed, while under **III** only 3946 were obtained. However, the minimum size of burn-in needed for the convergence of all the chains was of 2000 iterations for Scheme **I**, while it was just 500 iterations for Scheme **III**. These numbers were defined after a visual inspection of all the resultant trace plots.

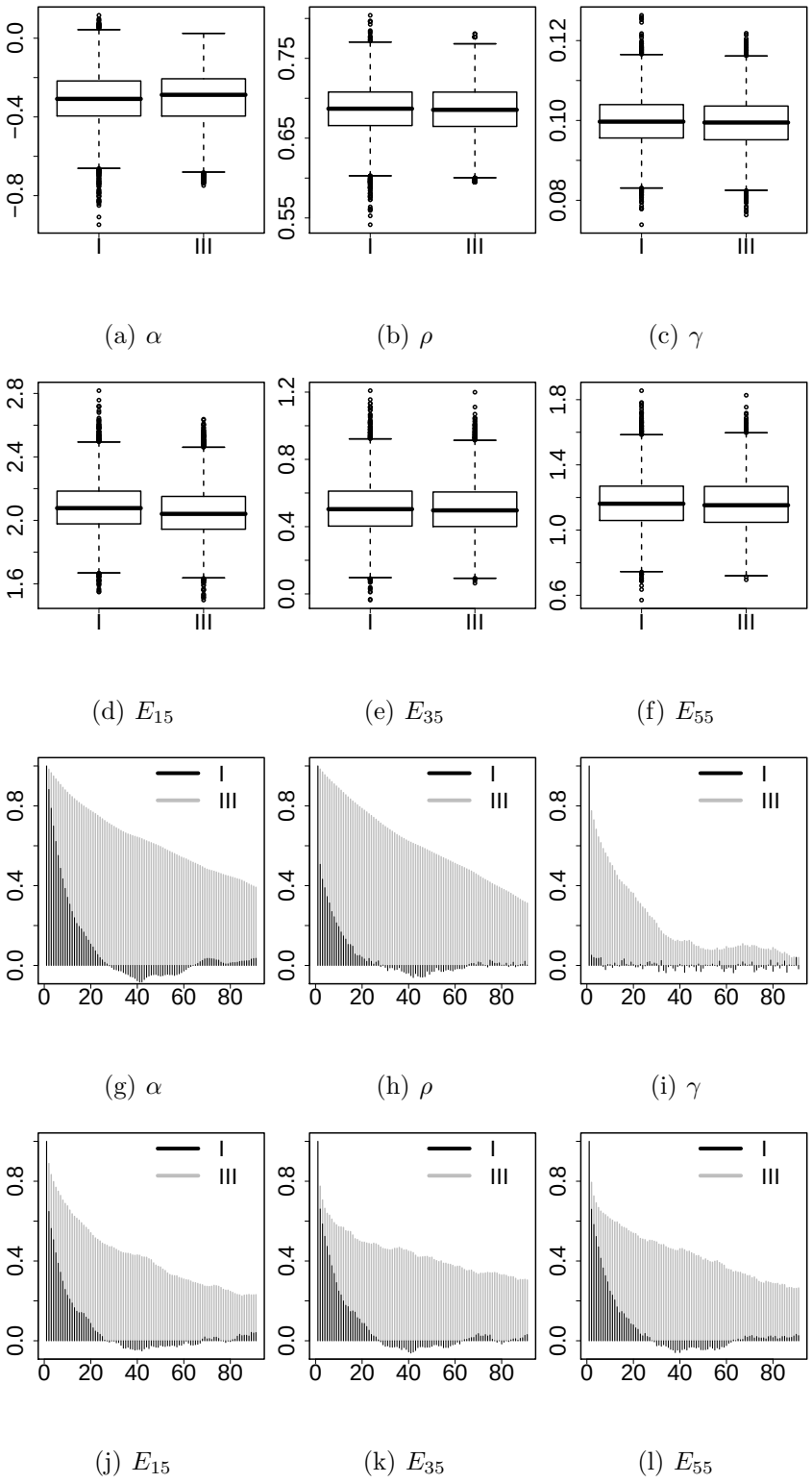
In order to compare the autocorrelations and inefficiency factors of the chains obtained, samples of size 3363 were used. This number was computed as follows: the number of iterations, after the burn-in period, performed under Scheme **I** was divided by the number of iterations performed under Scheme **III** and the result was taken as the thinning interval to be used in the first one. According to Table 4, the thinning interval in this case was 22 iterations long. The motivation for this exercise is the known fact that a practical solution to overcome the high autocorrelation exhibited by the outputs of a MCMC routine is to run a large number of iterations and to store samples after every k iterations.

Table 4: **Comparison of Schemes I and III with Rio Fartura data**

Criterion	I	III
Time elapsed by chain (in seconds)	600	600
Iterations completed by chain 1	78949	3961
Iterations completed by chain 2	76005	3931
Average of iterations by chain	77477	3946
Time to perform 100 iterations (in seconds)	0.77	15.20
Burn-in (from trace plots)	2000	500
Thinning interval	22	1
Final sample size by chain	3363	3363

Figure 5 shows the box plots and autocorrelations of the samples from the posterior distributions of some of the parameters in (9). Notice that, as expected, the resultant posterior samples are very similar, under both schemes. Also, the autocorrelations obtained from samples under Scheme **I** are lower than those under Scheme **III**.

Table 5 presents the potential scale reduction factors, PSRF or $\hat{R}$ (Gelman & Rubin, 1992), and the inefficiency factors, INEF, for some of the parameters in (9). From this we observe that all the



(a) α (b) ρ (c) γ

(d) E_{15} (e) E_{35} (f) E_{55}

(g) α (h) ρ (i) γ

(j) E_{15} (k) E_{35} (l) E_{55}

Figure 5: **Box plots and autocorrelations of some parameters' chains of the model in (9) for Rio Fartura data.**

values of PSRF, are close to 1, indicating that all the chains converged. The INEF obtained when the whole sample is used, that is, when the thinning interval is 1, shows that samples generated under Scheme **I** are more autocorrelated than those under Scheme **III**. However, taking advantage of the larger number of iterations performed under Scheme **I**, that is, using the thinning interval equal to 22, the INEF's are smaller for **I**. This result reinforces the idea that the fact of performing thousands of iterations under CUBS in a few seconds is an advantage over other sampling schemes.

Table 5: **Potential scale reduction factors and Inefficiency factors for some of the parameters in (9)**

	I						III			
	PSRF[a]		INEF by chain				PSRF[a]		INEF by chain	
	$\hat{R}$	97.5%	1 [b]	2 [b]	1 [c]	2 [c]	$\hat{R}$	97.5%	1 [b]	2 [b]
α	1.01	1.03	5.12	5.11	2.87	2.48	1.02	1.08	4.77	4.89
ρ	1.00	1.01	4.58	4.57	2.36	2.14	1.14	1.52	4.68	4.97
γ	1.00	1.00	2.75	2.70	1.13	1.22	1.07	1.29	3.59	3.49
ν	1.00	1.00	3.91	3.85	1.43	1.47	1.08	1.31	3.61	3.70
W	1.00	1.00	3.40	3.40	1.29	1.34	1.04	1.18	3.06	3.32
E_0	1.00	1.00	3.91	3.86	1.88	1.55	1.00	1.01	3.53	3.86
E_5	1.00	1.02	4.83	4.81	2.67	2.27	1.02	1.10	4.28	4.55
E_{15}	1.00	1.01	4.80	4.77	2.65	2.31	1.00	1.00	4.37	4.56
E_{35}	1.01	1.03	4.85	4.83	2.63	2.37	1.01	1.04	4.41	4.70
E_{55}	1.00	1.01	4.86	4.84	2.63	2.42	1.01	1.07	4.41	4.61

[a] Potential scale reduction factors (Gelman & Rubin, 1992).

[b] Every iteration.

[c] Every 25th iteration.

4. Concluding remarks and current research

In this paper we showed that the algorithm proposed by West et al. (1985), conjugate updating, can be used with satisfactory results to construct a proposal density in a Metropolis-Hastings step to sample in block all the state parameters of a generalized dynamic model. A Monte Carlo study was performed to compare this proposal with other sampling schemes previously established in the literature. We observe that CUBS works well, brings satisfactory results and has as main advantages: the ease of implementation and its low computational demand, resulting in fewer autocorrelated posterior samples.

Although the model used in the Monte Carlo study is a simple one, it can be stated that CUBS works well even with complex models like transfer functions where, for example, Scheme **IV** should not be used and Scheme **III** is hard to implement and is very computationally demanding.

Similar methods were independently developed by Yuhong Wu, Huiyan Sang and Mike West, in an unpublished work presented at the Valencia/ISBA 8th World Meeting on Bayesian Statistics, 2006. Their sampling scheme is also based on linear Bayes' idea. They used their approach in analyzing a collection of long, multiple time series of monthly rainfall patterns over a collection of geographical locations. They obtained really satisfactory results and encourage the use of the method in complex models (applications). This Duke's experience further verifies the utility and accuracy of the proposed sampling scheme.

As was illustrated here, the response variable can be a time series, where the parameters in (1) vary over time, but also the response can be an i.i.d. sample from a model with fixed parameters. In that case, posterior samples can be obtained using a Metropolis-Hastings step where the proposal density is constructed using CUBS. The idea of this approach is that the fixed parameters can be estimated sequentially: as new observations are introduced to the algorithm, more information is

gained about the fixed parameters, so at the end of the sample (when the last observation is used in the conjugate updating step), one must be really close to the mode of its posterior distribution.

Although the emphasis of CUBS applications resides in the exponential family distributions, this algorithm is more general because what is really needed is the existence of a conjugate prior for the location parameter of the response distribution. For example, CUBS can be used with the log-normal distribution.

The application of the scheme proposed to make inference in k-parameter distributions is one of our current topics of research. The idea is to use a multivariate conjugate prior for exponential family distributions (Bernardo & Smith, 1994, Chapter 5) to approximate the first moments of the state parameter's filtering distributions.

Acknowledgments

This work is part of Romy R. Ravines' PhD program under the supervision of Helio S. Migon and Alexandra M. Schmidt. The work of Romy R. Ravines was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES-Brazil. Helio S. Migon and Alexandra M. Schmidt were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-Brazil. The authors thank Mike West and Yuhong Wu for their helpful comments and suggestions.

References

Abramovitch, M. & Stegun, A. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*. New York: Dover Publications.

Bernardo, J. & Smith, A. (1994). *Bayesian Theory*. Chichester: John Wiley & Sons.

Carter, C. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81 541–553.

Doornik, J. (2002). *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press and Oxford, 3rd ed. URL `www.nuff.ox.ac.uk/Users/Doornik`.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic linear models. *Journal of the American Statistical Association* 87 501–509.

Frühwirth-Schnater, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15 183–202.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* 85 215–227.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. New York: Chapman & Hall / CRC.

Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7 457–511.

Geweke, J. & Tanizaki, H. (2001). Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. *Computacional Statistics & Data Analysis* 37 151–170.

Gilks, W., Best, N. & Tan, K. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 44 455–472.

Gilks, W. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41 337–348.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of non-stationary time series. *Journal of the American Statistical Association* 82 1032–1041.

Migon, H., Gamerman, D., Lopes, H. & Ferreira, M. (2005). Bayesian dynamic models. In D. Dey & C. Rao, eds., *Handbook of Statistics, 25*.

Migon, H. & Monteiro, A. B. (1997). Rain-fall modelling: An application of Bayesian forecasting. *Stochastic Hydrology and Hydraulics* 11 115–127.

Monteiro, A. B. (1992). *Modelos dinâmicos aplicados a modelagem chuva-vazão*. Master's thesis, Instituto de Matemática, UFRJ, Rio de Janeiro. Brazil. (in Portuguese).

Neal, R. (2003). Slice sampling (*with discussion*). *Annals of Statistics* 31 705–767.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

West, M., Harrison, J. & Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* 80 73–83.

A Some equations for prior and posterior parameters

Let r_t and s_t be the parameters of the conjugate prior of η_t. Consider f_t and q_t the prior moments obtained from the linear model and f_t^* and q_t^* the posterior ones. Table 6 shows the equations to be solved in order to compute f_t^* and q_t^* in the most frequently used distributions. In Table 6, $\gamma(\cdot)$ and $\gamma'(\cdot)$ denote the digamma and trigamma functions, respectively. In the last column the following approximations are used: $\gamma(x) \approx \log(x)$ and $\gamma'(x) \approx 1/x$ (see Abramovitch & Stegun (1965) for details).

B Gamma likelihood

In this section we show the CUBS for a dynamic gamma model:

$$y_t|\mu_t, \nu \sim \text{Gamma}(\mu_t, \nu), \quad t = 1, \ldots, N \tag{10a}$$

$$\log(\mu_t) = F_t'\theta_t \tag{10b}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N[0, W] \tag{10c}$$

$$\theta_0|D_0 \sim N[m_0, C_0], \tag{10d}$$

where $\text{Gamma}(\mu_t, \nu)$ refers to a gamma distribution with expected value μ_t and shape parameter ν.

Moments for $\log(\mu_t)$

Assuming ν known, from the conjugate analysis we have $(-\eta_t|r_t, s_t) \sim \text{Gamma}(s+1, r)$. Since $\log(\mu_t) = -\log(-\eta_t) = g(\eta_t)$, we have

$$\begin{aligned} E[g(\eta_t)] &= E[-\log(-\eta_t)] &&= -E[\log(-\eta_t)] &&= \log r_t - \gamma(s_t+1). \\ Var[g(\eta_t)] &= Var[-\log(-\eta_t)] &&= Var[\log(-\eta_t)] &&= \gamma'(s_t+1), \end{aligned}$$

Table 6: Example of equations to be solved to compute the prior and posterior parameters involved in the updating step of CUBS.

Distribution	Link	Prior and posterior moments		Approximations
Binomial	logit	$f_t^{(1)}$	$= \gamma(r_t) - \gamma(s_t)$	$\approx \log(r_t) - \log(s_t)$
		$g_t^{(1)}$	$= \gamma'(r_t) - \gamma'(s_t)$	$\approx 1/r_t - 1/s_t$
		f_t^*	$= \gamma(r_t + y_t) - \gamma(s_t + n_t - y_t)$	
		g_t^*	$= \gamma'(r_t + y_t) + \gamma'(s_t + n_t - y_t)$	
Binomial	identity	r_t	$= f_t/n_t[f_t/q_t(n_t - f_t) - 1]$	
		s_t	$= (1 - f_t/n_t)[f_t/q_t(n_t - f_t) - 1]$	
		f_t^*	$= n_t(r_t + y_t)/(r_t + s_t + n_t)$	
		g_t^*	$= [f_t^*(n_t - f_t^*)]/(r_t + s_t + n_t + 1)$	
Poisson	log	$f_t^{(1)}$	$= \gamma(r_t) - \log(s_t)$	$\approx \log(r_t) - \log(s_t)$
		$g_t^{(1)}$	$= \gamma'(r_t)$	$\approx 1/r_t$
		f_t^*	$= \gamma(r_t + y_t) - \log(s_t + 1)$	
		g_t^*	$= \gamma'(r_t + y_t)$	
Poisson	identity	r_t	$= f_t^2/q_t$	
		s_t	$= f_t/q_t$	
		f_t^*	$= (r_t + y_t)/(s_t + 1)$	
		g_t^*	$= (\alpha + y_t)/(s_t + 1)^2$	
Gamma	log	$f_t^{(1)}$	$= \log(r_t) - \gamma(s_t + 1)$	$\approx \log(r_t) - \log(s_t + 1)$
		$g_t^{(1)}$	$= \gamma'(s_t + 1)$	$\approx 1/(s_t + 1)$
		f_t^*	$= \log(r_t + \nu y_t) - \gamma(s_t + \nu + 1)$	
		g_t^*	$= \gamma'(s_t + \nu + y_t)$	

$^{(1)}$ numerical methods or approximations can be used to obtain r_t and s_t.

where $\gamma(x) = d \log \Gamma(x)/dx = \Gamma'(x)/\Gamma(x) \approx \log(x)$ is the *digamma function* and $\gamma'(x) \approx 1/x$ is the *trigamma function*.

CUBS for the Gamma Dynamic Model in (10)

I. Set $t = 1$;

II. Compute the posterior moments for the states, $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$:

(a) Compute from the model the moments of the prior on $\boldsymbol{\theta}_t$ and $g(\eta_t)$:

$$\boldsymbol{\theta}_t|D_{t-1} \sim [\boldsymbol{a}_t, \boldsymbol{R}_t], \qquad \text{where } \boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1} \text{ and } \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t \tag{11}$$

$$g(\eta_t)|D_{t-1} \sim [f_t, q_t], \qquad \text{where } f_t = \boldsymbol{F}_t'\boldsymbol{a}_t \text{ and } q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t; \tag{12}$$

(b) Determine from the conjugate analysis the moments of the prior of $g(\eta_t)$:

$$E[g(\eta_t)|D_{t-1}] = f_t = \quad \log r_t - \gamma(s_t + 1) \quad \approx \log r_t - \log(s_t + 1) \tag{13}$$

$$Var[g(\eta_t)|D_{t-1}] = q_t = \quad \gamma'(s_t + 1) \quad \approx 1/(s_t + 1); \tag{14}$$

(c) Obtain from the conjugate analysis the moments of the posterior for $g(\eta_t)$:

$$E[g(\eta_t)|D_t] = f_t^* = \log(r_t + \nu y_t) - \gamma(s_t + \nu + 1) \tag{15}$$

$$Var[g(\eta_t)|D_t] = q_t^* = \gamma'(s_t + \nu + 1); \tag{16}$$

(d) Compute from the model the moments of the posterior distribution for $\boldsymbol{\theta}_t$:

$$\boldsymbol{\theta}_t|D_t \sim [\boldsymbol{m}_t, \boldsymbol{C}_t], \qquad \text{where:}$$

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(f_t^* - f_t)\frac{1}{q_t} \tag{17}$$

$$\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t\big(1 - \frac{q_t^*}{q_t}\big)\frac{1}{q_t}. \tag{18}$$

III. Set $t = t + 1$ and return to **II** if $t < N$;

IV. Sample $\boldsymbol{\theta}_N$ from $N(m_N, C_N)$;

V. Set $t = N - 1$ and sample $\boldsymbol{\theta}_t$ from $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t+1}, D_t, \boldsymbol{\theta}) = N(\boldsymbol{m}_t^s, \boldsymbol{C}_t^s)$;

VI. Set $t = t - 1$ and return to **V** if $t > 1$.

Steps IIb and IId above are specific for each conjugate family. Steps IV-VI form the Backward Sampling. The moments of $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t+1}, D_t, \boldsymbol{\theta})$ are obtained from:

$$\boldsymbol{\theta}_t | D_t \sim N\big[\boldsymbol{m}_t, \boldsymbol{C}_t\big]$$

$$\boldsymbol{\theta}_{t+1} | \boldsymbol{\theta}_t, D_t \sim N\big[\boldsymbol{G}_{t+1}\boldsymbol{\theta}_t, \boldsymbol{G}_{t+1}\boldsymbol{C}_t\boldsymbol{G}'_{t+1} + \boldsymbol{W}_{t+1}\big]$$

$$\begin{pmatrix} \boldsymbol{\theta}_{t+1} \\ \boldsymbol{\theta}_t \end{pmatrix} \Big| D_t \sim N\left[\begin{pmatrix} \boldsymbol{G}_{t+1}\boldsymbol{m}_t \\ m_t \end{pmatrix}, \begin{pmatrix} \boldsymbol{G}_{t+1}\boldsymbol{C}_t\boldsymbol{G}'_{t+1} + \boldsymbol{W}_{t+1} & \boldsymbol{G}_{t+1}\boldsymbol{C}_t \\ \boldsymbol{C}_t\boldsymbol{G}_{t+1} & \boldsymbol{C}_t \end{pmatrix}\right].$$

Let $\boldsymbol{R}_{t+1} = \boldsymbol{G}_{t+1}\boldsymbol{C}_t\boldsymbol{G}'_{t+1} + \boldsymbol{W}_{t+1}$ and $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{G}'_{t+1}\boldsymbol{R}^{-1}_{t+1}$, then

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t \sim N\big(\boldsymbol{m}^s_t, \boldsymbol{C}^s_t\big),$$

where

$$\boldsymbol{m}^s_t = \boldsymbol{m}_t + \boldsymbol{B}_t(\boldsymbol{\theta}_{t+1} - \boldsymbol{G}_{t+1}\boldsymbol{m}_t)$$

$$\boldsymbol{C}^s_t = \boldsymbol{C}_t - \boldsymbol{B}_t\boldsymbol{R}_{t+1}\boldsymbol{B}'_t.$$