Particle filters in state space models with the presence of unknown static parameters

Geir Storvik*

Email: geirs@math.uio.no

Abstract

In this paper a Monte Carlo filter for dynamic state space models handling unknown static parameters is introduced. The approach is based on marginalizing the static parameters out of the posterior distribution such that only the state vector needs to be considered. Although such an marginalization always can be applied, real-time applications are only possible when the distribution of the unknown parameters given both observations *and* the hidden state vector only depend on some low-dimensional sufficient statistics. Such sufficient statistics are present in many of the commonly used state space models. Marginalizing out the static parameters avoids the problem of impoverishment which typically occur when static parameters are included as part of the state vector. The new filter is tested on several different models and shows promising results.

Keywords

State space models, sequential updating, particle filters, global parameters, marginalization, sufficient statistics

*Geir Storvik is an Associate Professor at the University of Oslo and a part time researcher at the Norwegian Computing Center.

1 Introduction

Dynamic state space models [1, 2, 3] are useful for describing data in many different areas, for instance engineering [4], finance mathematics [5], environmental data [6], geophysical science [7] and disease data [8].

Using $p(\cdot|\cdot)$ for a generic conditional distribution, the general (discrete time) state space model is given by

$$\mathbf{x}_k \sim p(\mathbf{x}_k|\mathbf{x}_0, ..., \mathbf{x}_{k-1}; \boldsymbol{\theta}), \quad \text{system} \tag{1a}$$

$$\mathbf{z}_k \sim p(\mathbf{z}_k|\mathbf{x}_k; \boldsymbol{\theta}), \quad \text{observations} \tag{1b}$$

where $\mathbf{z}_k$ contain the observations at time t_k, while $\{\mathbf{x}_k\}$ is an underlying stochastic process which in some cases may have a physical meaning while in other cases merely is included in order to describe the distribution of the observation process properly. $\boldsymbol{\theta}$ is a vector containing static parameters which in some cases can be specified but in many cases are unknown. Typically, some prior distribution is placed on $\mathbf{x}_0$.

An important task when analyzing data by state space models is estimation of the underlying state process based on measurements from the observation process. The interest might be on $\mathbf{x}_k$ itself (or a function of $\mathbf{x}_k$), or merely is a tool for making prediction on $\mathbf{z}_k$. Given data $\mathbf{z}_{1:N} = (\mathbf{z}_1, ..., \mathbf{z}_N)$, estimation of $\mathbf{x}_{1:N}$ is usually referred to as *off-line* estimation, while *on-line* estimation is sequential estimation of $\mathbf{x}_k$ based on $\mathbf{z}_{1:k}$ for $k = 1, 2, 3, ...$ In this paper we will concentrate on the latter problem. In particular the focus will be on problems where new observations arrive frequently (hours/minutes/seconds) and real-time estimation/prediction is essential.

In a few cases, including linear Gaussian models and hidden Markov chains, the distribution of $\mathbf{x}_k$ given the observations $\mathbf{z}_{1:k}$ can be computed exactly using recursive formula's. For situations where analytical solutions are impossible to obtain, stochastic simulation can be applied. Numerous papers have been written on construction of algorithms based on Markov Chain Monte Carlo (MCMC) dealing with general state space models (see Gamerman [9] and the references therein). Although such procedures may be effective for off-line estimation, there are problems with full MCMC in the case of on-line estimation. The MCMC algorithm needs to be restarted at each time point. Further the dimension of the vector to be simulated increase with time.

An alternative to full MCMC at each time step is construction of simulation algorithms for sequential updating the posterior distributions. Such (slightly different) algorithms have been developed independently in many fields [10, 11, 12, 13, 14] with different names (Bootstrap filter, Monte Carlo filter, Particle filter). The excellent review by Doucet [15] even contain some references back to late 60's. A collection of papers describing the state of the art in this field can be found in [16]. In this paper, such algorithms will be denoted *particle filters*.

The main idea behind particle filters is as follows: Suppose a sample (possibly weighted) $\mathbf{S}_{k-1} = \{(\mathbf{x}_{1:k-1}^{(i)}, w_{k-1}^{(i)}), i = 1, ..., N\}$ is simulated from $p(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k-1})$. When a new observation $\mathbf{z}_k$ arrives, $\mathbf{S}_{k-1}$ is updated to a new sample $\mathbf{S}_k = \{(\mathbf{x}_{1:k}^{(i)}, w_k^{(i)}), i = 1, ..., N\}$

from $p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})$ using the model (1) and some appropriate simulation technique. Techniques for obtaining $\mathbf{S}_k$ include rejection sampling [10], sampling/importance resampling [11] and MCMC [14, 17]. Assuming $f(\mathbf{x}_{1:k})$ is of interest, the posterior expectation I_k^f of this quantity is approximated by

$$\widehat{I}_k^f = \sum_{i=1}^{N} w_k^{(i)} f(\mathbf{x}_{1:k}^{(i)}) / \sum_{i=1}^{N} w_k^{(i)}. \tag{2}$$

Although particle filters have proven successful in many simulation experiments and in analysis of real data, a main problem with such an approach is how to handle the presence of unknown *static* parameters. A common trick in engineering is to include the parameters as part of the state space vector $\mathbf{x}_k$. Berzuini et al. [14] put this approach in a formal Bayesian setting. However, the non-dynamics in the parameters makes the parameter samples degenerate into one or a few different values when k increases.

Gordon et al. [11], West [18], Bølviken et al. [19] and Liu and West [20] introduced diversity in the set of particles by adding random noise to the particles, which in this context is similar to approximate the non-dynamic parameters by some slowly changing dynamic ones. In addition to the problem of choosing "diversity" procedure, this result in that old observations are down-weighted and the parameter estimates obtained at a given time is mainly depending on the most recent observations.

This paper considers an alternative approach. Suppose that for *given* $\mathbf{x}_{1:k}$ the distribution of $\boldsymbol{\theta}$ is analytically tractable. In particular, the distribution of $\boldsymbol{\theta}$ is assumed to depend on $\mathbf{x}_{1:k}$ only through some low-dimensional sufficient statistics. In such cases, only samples of $\mathbf{x}_{1:k}$ given $\mathbf{z}_{1:k}$ is needed, since estimates of the posterior distribution for the static parameters can be obtained either through Rao-Blackwellization or by a simple additional simulation step. Furthermore, updating the sample to a new sample one time-step further on can be performed by simulation of the state vector *and* the parameters simultaneously. The approach can be considered as a sort of marginalization of $\boldsymbol{\theta}$ from the posterior. Such marginalizations are however usually applied only when the marginal posterior can be calculated analytically [21]. In this paper, the marginalization is performed through simulations.

Although the required assumption will restrict the set of models possible to process by this approach, many important (and widely used) models are included. In particular, models for which the underlying process is Gaussian and linear in the parameters involved (but not necessarily linear in the $\mathbf{x}$-process) can be handled by this approach. Further, the assumption about Gaussian noise in the system process can be relaxed to include "partial non-Gaussian" processes as defined by Shephard [22]. Both T-distributions and mixture of Gaussians fall into this group of models. These system processes can be combined with any observational distribution which do not contain any additional unknown parameters. Also many distributions *with* unknown parameters can be handled by this approach, but typically special treatment is needed in each case.

In Section 2 the general particle filters are reviewed. Section 3 introduces the new particle filter for situations with unknown static parameters. Section 4 considers some

particular classes of models which fit into this framework, while the new filter is applied on different types of models in Section 5. Finally, a summary and discussion is given in Section 6.

2 Particle filters in general

This section discusses particle filters in situations where the static parameters are known. For the time being $\boldsymbol{\theta}$ will be suppressed in the notation.

Today many different versions of particle filters exists (see Doucet et al. [16]). Two different motivations are typically used in construction of a filter. One approach is based on importance sampling. In this case $\mathbf{x}_k$ is simulated sequentially from some importance distribution $f_k(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k})$ and the whole trajectory $\mathbf{x}_{1:k}$ is given importance weight

$$w_k = \frac{p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})}{\prod_{l=1}^{k} f_l(\mathbf{x}_l|\mathbf{x}_{1:l-1}, \mathbf{z}_{1:l})}.$$

N such sequences are simulated in parallel, giving a weighted sample at each time point t_k. Restrictions on the importance distributions are needed both for ease in simulation and in order to make the computation of the weights possible. See [23] for further discussion on this approach, which usually is named *sequential importance sampling* (SIS). A problem with this approach is that when time evolves the variance of the weights will increase ([16]) making the estimate (2) unstable. A common trick to avoid this is to *resample* from $\mathbf{S}_k$ with probabilities proportional to $w_k^{(i)}$. Liu and Chen [23] give some heuristics on when to resample.

An alternative approach is based on the approximation

$$\widehat{p}(\mathbf{x}_{1:k}|\mathbf{z}_{1:k-1}) = \frac{\sum_{i=1}^{N} w_k^{(i)} I(\mathbf{x}_{1:k-1} = \mathbf{x}_{1:k-1}^{(i)}) p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})}{\sum_{i=1}^{N} w_k^{(i)}}. \tag{3}$$

By Bayes rule,

$$\begin{aligned} p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k}) \propto & p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k-1}) p(\mathbf{z}_k|\mathbf{x}_k) \\ \approx & \widehat{p}(\mathbf{x}_{1:k}|\mathbf{z}_{1:k-1}) p(\mathbf{z}_k|\mathbf{x}_k). \end{aligned} \tag{4}$$

A new sample $\mathbf{S}_k$ can now be obtained by simulating from this approximative distribution.

A possible rejection sampling procedure [10] for simulation from (4) is to sample $\mathbf{x}_{1:k}$ from $\widehat{p}(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and accept the sample with probability proportional to $p(\mathbf{z}_k|\mathbf{x}_k)$. This procedure can be repeated N times in order to obtain a new (unweighted) sample $\mathbf{S}_k = \{\mathbf{x}_k^1, ..., \mathbf{x}_k^B\}$. In practice the acceptance probability for this simple algorithm will be far to low, making the need for other approaches. In [17] constructions of more efficient proposal distributions are given and some other sampling approaches including sampling/importance resampling and MCMC are discussed. Not that the use of MCMC here is on a much smaller dimension than if a full MCMC scheme was to be applied.

A main problem with the particle filters described above is that when simulating $\mathbf{x}_{1:k}$ at time k, the first $k-1$ components can only take the values given in $\mathbf{S}_{k-1}$. Gilks and Berzuini [24] and Carpenter et al. [25] introduced the possibilities of changing the whole vector $\mathbf{x}_{1:k}$ according to some Markov transition kernel having $p(\mathbf{x}_{1:k}|\mathbf{z}_{1:k})$ as stationary distribution. Such an approach *do* solve some of the degereracy problems that can occur for the more standard particle filters, but in general the complexity in simulating from the Markov transition kernel will increase in time, giving much the same problems as for full MCMC simulation.

Given the main framework, much freedom is available to the user on how to specify the algorithm. Care has to be taken however in order to make the algorithm work properly both at fixed k for large enough N and for fixed N with time increasing. Storvik [26] demonstrated through some simulation experiments that the numerical errors introduced can accumulate linearly for some particle filters. Recently, some theoretical results on particle filters have appeared. Berzuini et al. [14] established a central limit theorem for the estimator (2) for the importance sampling approach with resampling at each stage. More general results on convergence is given in [27], which shows that most algorithms proposed will converge properly as $N \to \infty$. These results are however based on increasing the number of particles N to infinity. In [3] it is proven that if N grows like k^2 as k increases, the error in the approximative distribution remains fixed during time. The order k^2 can probably be improved by introducing additional conditions or by constructing more efficient filters.

3 Particle filters including non-dynamic parameters

In this section an approach for particle filtering in the presence of unknown parameters will be introduced. The usual approach is to include the parameters as part of the state vector $(\mathbf{x}_k, \boldsymbol{\theta})$. Because of the non-dynamic feature of the parameters, samples of $\boldsymbol{\theta}$ at time t_k can only take the values given at time t_{k-1}. Since some of these values become very unlikely when new observations arrive, this will result in an impoverishment of the set of distinct $\boldsymbol{\theta}$ values.

The approach taken in this paper is based on a different idea. Assume that the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{z}_{1:k}$ *and* $\mathbf{x}_{1:k}$ depend on some sufficient statistics $\mathbf{T}_k = \mathbf{T}_k(\mathbf{x}_{1:k}, \mathbf{z}_{1:k})$, where $\mathbf{T}_k$ is easy to update recursively.

Assume that an approximate sample $\mathbf{S}_{k-1}$ is available from the simultaneous distribution $p(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k-1})$. Again the sample $\mathbf{S}_{k-1}$ is to be updated to a new sample $\mathbf{S}_k$ at time t_k. Even though only the $\{\mathbf{x}_k\}$ process will be stored, simulation simplifies if $\boldsymbol{\theta}$ is included as an ancillary variable in the simulation step. The approach is based on that

$$\begin{aligned} p(\mathbf{x}_{1:k}, \boldsymbol{\theta}|\mathbf{z}_{1:k}) \propto & p(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k-1}) p(\boldsymbol{\theta}|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k-1}) \times \\ & p(\mathbf{x}_k|\mathbf{x}_{1:k-1}; \boldsymbol{\theta}) p(\mathbf{z}_k|\mathbf{x}_k; \boldsymbol{\theta}) \\ = & p(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k-1}) p(\boldsymbol{\theta}|\mathbf{T}_{k-1}) \times \\ & p(\mathbf{x}_k|\mathbf{x}_{1:k-1}; \boldsymbol{\theta}) p(\mathbf{z}_k|\mathbf{x}_k; \boldsymbol{\theta}). \end{aligned} \tag{5}$$

Using the approximation in (3), simulation from (5) can be performed as before, but with the additional step that also $\boldsymbol{\theta}$ needs to be simulated. The simplest approach would be to simulate $\mathbf{x}_{1:k-1}$ from $\widehat{p}(\mathbf{x}_{1:k-1}|\mathbf{z}_{1:k-1})$, $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\mathbf{T}_{k-1})$, $\mathbf{x}_k$ from $p(\mathbf{x}_k|\mathbf{x}_{1:k-1};\boldsymbol{\theta})$ and accept with probability proportional to $p(\mathbf{z}_k|\mathbf{x}_k;\boldsymbol{\theta})$. However, any simulation technique such as sampling/importance resampling or MCMC can be applied. Also the SIS approach can be used.

The important part about this approach is that the parameter $\boldsymbol{\theta}$ simulated at time t_k do not depend on values simulated at previous time points. This avoids the problem with impoverishment.

In principle, the existence of a low-dimensional sufficient statistic for $\boldsymbol{\theta}$ is not necessary, because we only need to be able to evaluate or simulate from $p(\boldsymbol{\theta}|\mathbf{x}_{1:k},\mathbf{z}_{1:k})$. However, in order to make the filter run fast and not have increasing complexity as time evolves, the need for $p(\boldsymbol{\theta}|\mathbf{x}_{1:k},\mathbf{z}_{1:k})$ only to depend on $(\mathbf{x}_{1:k},\mathbf{z}_{1:k})$ through $\mathbf{T}_k$ becomes appearant.

Following [16], a SIS with resampling (SISR) algorithm which includes static parameters is defined as follows:

Importance sampling For $i = 1, ..., N$,

- sample $\boldsymbol{\theta} \sim f_{k,1}(\boldsymbol{\theta}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{z}_{0:k})$,
- sample $\tilde{\mathbf{x}}_k^{(i)} \sim f_{k,2}(\mathbf{x}_k|\boldsymbol{\theta},\mathbf{x}_{0:k-1}^{(i)},\mathbf{z}_{0:k})$ and define $\tilde{\mathbf{x}}_{1:k}^{(i)} \stackrel{\Delta}{=} (\mathbf{x}_{0:k-1}^{(i)},\tilde{\mathbf{x}}_k^{(i)})$,
- evaluate the importance weights

$$\widetilde{w}_k^{(i)} = w_{k-1}^{(i)} \frac{p(\mathbf{z}_k|\boldsymbol{\theta},\tilde{\mathbf{x}}_k^{(i)})p(\boldsymbol{\theta}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{z}_{0:k})p(\tilde{\mathbf{x}}_k^{(i)}|\boldsymbol{\theta},\mathbf{x}_{k-1}^{(i)})}{f_{k,1}(\boldsymbol{\theta}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{z}_{0:k})f_{k,2}(\tilde{\mathbf{x}}_k^{(i)}|\boldsymbol{\theta},\mathbf{x}_{k-1}^{(i)})}.$$

Resampling For $i = 1, ..., N$,

- sample an index j_i from $\{1, ..., N\}$ with probabilities proportional to $\widetilde{w}_k^{(j_i)}$,
- put $\mathbf{x}_{1:k}^{(i)} = \tilde{\mathbf{x}}_{1:k}^{(j_i)}$ and $w_k^{(i)} = N^{-1}$.

As for the ordinary SISR algorithm, resampling can be performed at each time step or according to some specific rules [12]. Stratified sampling [25] and other variance reduction methods can also easily be incorporated.

The dependence on $\boldsymbol{\theta}$ in the weights increase the variability in the weights. This variability can be decreased in situations where both

$$p(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{z}_{0:k})p(\tilde{\mathbf{x}}_k^{(i)}|\boldsymbol{\theta},\mathbf{x}_{k-1}^{(i)})d\boldsymbol{\theta}$$

and

$$f_k(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}) = \int_{\boldsymbol{\theta}} f_{k,1}(\boldsymbol{\theta}|\mathbf{x}_{0:k-1}^{(i)}, \mathbf{z}_{0:k}) f_{k,2}(\tilde{\mathbf{x}}_k^{(i)}|\boldsymbol{\theta}, \mathbf{x}_{k-1}^{(i)}) d\boldsymbol{\theta}$$

can be evaluated, in which case the weights can be modified to

$$\tilde{w}_k^{(i)} = \tilde{w}_{k-1}^{(i)} \frac{p(\mathbf{z}_k|\boldsymbol{\theta}, \tilde{\mathbf{x}}_k^{(i)}) p(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}{f(\tilde{\mathbf{x}}_k^{(i)}|\mathbf{x}_{k-1}^{(i)})}.$$

In most situations, however, these integrals are not attainable.

Samples of $\boldsymbol{\theta}$ based on $\mathbf{z}_{1:k}$ are directly available through the algorithm. In order to estimate $\boldsymbol{\theta}$, a better approach is to use Rao-Blackwellization, as described in [16].

4 Gaussian-based system processes

In this section some particular classes of models that can be used in the framework developed in Section 3 will be discussed. Only models for system processes will be considered, because these can be discussed in general terms. When unknown parameters are present in the observational distribution, special treatment usually is needed, and will therefore be discussed through specific examples in Section 5. In many cases, the parameters involved in the observation process are given from other sources. In such cases, no extra treatment is needed.

We will concentrate on Gaussian-based models. This is both because such models are commonly applied as system processes and because sufficient statistics are easily calculated for this class of models.

4.1 Gaussian system processes

A particular useful class of models is obtained when the underlying state process is Gaussian, but where the observation distribution is arbitrary. Assume

$$\mathbf{x}_k = \mathbf{F}_k'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_k, \quad \boldsymbol{\varepsilon}_k \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{Q}), \tag{6}$$

where $\mathbf{F}_k = \mathbf{F}(\mathbf{x}_{k-1})$ is a matrix with elements possibly non-linear functions of $\mathbf{x}_{k-1}$. The unknown parameters are in this case $\boldsymbol{\beta}$ and σ ($\mathbf{Q}$ is assumed known, the general case can also be handled but becomes much more complex).

Assume $\boldsymbol{\beta}$ and σ^2 to have priors $\boldsymbol{\beta} \sim \mathsf{N}(\boldsymbol{\beta}_0, \sigma^2\mathbf{C}_0)$ and $\sigma^2 \sim \mathsf{IG}(\frac{\nu_0}{2}, \frac{d_0}{2})$ where IG is the inverse Gamma distribution. Then standard theory [2] gives

$$[\boldsymbol{\beta}|\mathbf{x}_{1:k}, \mathbf{z}_{1:k}, \sigma^2] \sim \mathsf{N}(\mathbf{m}_k, \sigma^2\mathbf{C}_k) \tag{7a}$$

$$[\sigma^2|\mathbf{x}_{1:k}, \mathbf{z}_{1:k}] \sim \mathsf{IG}(\frac{\nu_k}{2}, \frac{d_k}{2}) \tag{7b}$$

where the sufficient statistics $\mathbf{m}_k, \mathbf{C}_k, d_k$ and ν_k are updated according to the equations

$$\begin{aligned}
\mathbf{D}_k &= \mathbf{F}_k' \mathbf{C}_{k-1} \mathbf{F}_k + \mathbf{Q}, && \text{(8a)}\\
\mathbf{C}_k &= \mathbf{C}_{k-1} - \mathbf{C}_{k-1}\mathbf{F}_k \mathbf{D}_k^{-1} \mathbf{F}_k' \mathbf{C}_{k-1}, && \text{(8b)}\\
\mathbf{m}_k &= \mathbf{m}_{k-1} + \mathbf{C}_{k-1}\mathbf{F}_k \mathbf{D}_k^{-1} [\mathbf{x}_k - \mathbf{F}_k' \mathbf{m}_{k-1}], && \text{(8c)}\\
d_k &= d_{k-1} + (\mathbf{x}_k - \mathbf{F}_k' \mathbf{m}_{k-1})' \mathbf{D}_k^{-1} (\mathbf{x}_k - \mathbf{F}_k' \mathbf{m}_{k-1}), && \text{(8d)}\\
\nu_k &= \nu_{k-1} + 1. && \text{(8e)}
\end{aligned}$$

The particle filter approach described in Section 3 can therefore easily be applied.

4.2 Partial non-Gaussian state space

Shephard [22] introduced a class of non-Gaussian time-series models allowing the noise process to be T-distributed or a mixture of Gaussians. The particle filter approach presented in this paper is also applicable for this type of models. For illustration purpose, only T-distributed noise will be considered.

Rewrite model (6) to

$$\mathbf{x}_k = \mathbf{F}_k' \boldsymbol{\beta} + \check{\boldsymbol{\varepsilon}}_k / \omega_k, \quad \check{\boldsymbol{\varepsilon}}_k \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{Q}), \omega_k \sim \chi^2_\nu, \tag{9}$$

with $\check{\boldsymbol{\varepsilon}}_k$ and ω_k being independent. The posterior distributions of $\boldsymbol{\beta}$ and σ^2 given both $\mathbf{x}_{1:k}$ *and* $\boldsymbol{\omega}_{1:k}$ become equal to (7), but the updating formula's (8) are slightly changed by substituting $\mathbf{Q}$ in (8a) with $\mathbf{Q}/\omega_k$.

The state vector needs in this case to be extended to include ω_k. Simulation *conditional on* ω_k can be performed as in the Gaussian case with small modifications. Simulation of ω_k given all the other variables is also easy,

$$\omega_k \sim \mathsf{G}\left(\frac{\nu + q}{2}, \frac{1}{2}(1 + \frac{1}{\sigma^2}(\mathbf{x}_k - \mathbf{F}_k' \boldsymbol{\beta})' \mathbf{Q}^{-1} (\mathbf{x}_k - \mathbf{F}_k' \boldsymbol{\beta})) \right),$$

where q is the dimension of $\mathbf{x}_k$. Direct simulation of all variables involved is however not longer possible, but a blocked Gibbs sampler approach switching between sampling ω_k and a block containing all the other variables can be applied.

5 Experiments

In this section some examples of dynamic models will be considered in order to evaluate the performance of the particle filter when some static parameters are unknown. For each example, a SISR filter including static parameters using $f_{k,1}(\boldsymbol{\theta}|\mathbf{x}^{(i)}_{0:k-1}, \mathbf{z}_{0:k}) = p(\boldsymbol{\theta}|\mathbf{T}_{k-1})$ and $f_{k,2}(\mathbf{x}_k|\boldsymbol{\theta}, \mathbf{x}_{0:k-1}, \mathbf{z}_{0:k}) = p(\mathbf{x}_k|\boldsymbol{\theta}, \mathbf{x}_{k-1})$ is applied. Resampling is performed at each time step, resulting in that the weights $w_k^{(i)}$ reduces to $p(\mathbf{z}_k|\boldsymbol{\theta}, \tilde{\mathbf{x}}_k^{(i)})$. More efficient filters, where the proposal distribution depend on the new observation $\mathbf{z}_k$, can however be constructed similar to the standard SISR filters.

In each case, the estimated posterior distributions is compared with those obtained from a full MCMC run at each time step using a huge number of iterations. The free software BUGS [28] is used for the MCMC runs. The current version of BUGS can however not handle the model considered in Section 5.2 and therefore a comparison is missing in this case.

5.1 A dynamic generalized linear model

West et al. [29] considered a general class of dynamic Bayesian models. They studied the case where the underlying system process is linear and the distribution for the observation $\mathbf{z}_k$ conditioned on the underlying state vector $\mathbf{x}_k$ is in the exponential family. We will consider applications where the observed data are (possibly multivariate) binary, making the logistic model an obvious choice. Such models have been used in i.e. ecology [30] where a number of observers indicate whether the population at the current time is either high or low. Here, only a simplified version will be considered:

$$x_k \sim \mathrm{N}(ax_{k-1}, \sigma^2), \tag{10a}$$
$$z_k \sim \mathrm{Binom}(r, \mathrm{logit}\{\alpha + \beta x_k\}), \tag{10b}$$

for $k = 1, 2, ...$ The unknown static parameters are $\boldsymbol{\theta} = (a, \sigma^2, \alpha, \beta)$.

Data was simulated according to the model with $a = 0.9, \sigma^2 = 1$ and $\alpha = \beta = 0.5$. Assume first the parameters in the observation process, α and β, are known, while a priori $a \sim \mathrm{N}(0, \sigma_a^2)$ and $\sigma^2 \sim IG(\nu_0/2, d_0/2)$. In this particular case the recursions given in (8) can be applied to update the sufficient statistics. Figure 1 shows posterior means and quantiles obtained from the SISR filter using $N = 2000$ particles. The same quantities calculated using full MCMC at each time step are also plotted for comparision. For the $\{\mathbf{x}_k\}$ process and the estimates of a at each time step, both posterior means and quantiles are almost identical. Some discrepancy is present for the σ estimates, but the results are still acceptable.

Turn now to the case when also α and β are unknown. In this case it will be advantageous to reparametrize the model such that all the parameters becomes part of the system process. This can be done by defining $\check{x}_k = \alpha + \beta x_k$. Then the model can be written as

$$\check{x}_k \sim \mathrm{N}(\alpha + a(\check{x}_{k-1} - \alpha), \check{\sigma}^2), \tag{11a}$$
$$z_k \sim \mathrm{Binom}(\mathrm{logit}\{\check{x}_k\}), \tag{11b}$$

where $\check{\sigma} = \beta\sigma$. Note that β and σ are only identifiable through their product. This means that the $\{x_k\}$ process can only be recovered up to an unknown scale factor. In the ecology example referred to above this is not a serious concern since interest is primarily in the possibly seasonality patterns of the process.

We will assume a priori $a \sim \mathrm{N}(0, \sigma_a^2)$, $\check{\sigma}^2 \sim IG(\frac{\nu_0}{2}, \frac{d_0}{2})$ and $\alpha \sim \mathrm{N}(0, \sigma_\alpha^2)$. In this case, the model do not fit into (6) because a and α appear in the model through their product.

Also direct simulation from $p(a, \check{\sigma}^2, \alpha | T_k)$ is not possible. It is however easy to show that

$$
\begin{aligned}
[\check{\sigma}^2 | a, \alpha, \check{x}_{1:k}, z_{1:k}] \sim & IG(\nu_k/2, d_k/2) \\
[a | \check{\sigma}^2, \alpha, \check{x}_{1:k}, z_{1:k}] \sim & \mathrm{N}(m_k^a, c_k^a) \\
[\alpha | \check{\sigma}^2, a, \check{x}_{1:k}, z_{1:k}] \sim & \mathrm{N}(m_k^\alpha, c_k^\alpha)
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_k =& \nu_0 + k \\
d_k =& \sum_{l=1}^{k} \check{x}_l^2 - 2a(1-a)\alpha \sum_{l=1}^{k} \check{x}_l \check{x}_{l-1} + a^2 \sum_{l=1}^{k} \check{x}_{l-1}^2 - \\
& 2(1-a)\alpha [\sum_{l=1}^{k} \check{x}_l - a \sum_{l=1}^{k} \check{x}_{l-1}] + (k-1)(1-a)^2\alpha^2 \\
m_k^a =& \frac{\sigma_a^2 (\sum_{l=1}^{k} \check{x}_l \check{x}_{l-1} - \alpha \sum_{l=1}^{k} (\check{x}_l + \check{x}_{l-1}) + (k-1)\alpha^2)}{\check{\sigma}^2 + \sigma_a^2 (\sum_{l=1}^{k} \check{x}_{l-1}^2 - 2\alpha \sum_{l=1}^{k} \check{x}_{l-1} + (k-1)\alpha^2)} \\
c_k^a =& \frac{\sigma_a^2 \check{\sigma}^2}{\check{\sigma}^2 + \sigma_a^2 (\sum_{l=1}^{k} \check{x}_{l-1}^2 - 2\alpha \sum_{l=1}^{k} \check{x}_{l-1} + (k-1)\alpha^2)} \\
m_k^\alpha =& \frac{\sigma_\alpha^2 (1-a)(\sum_{l=1}^{k} \check{x}_l - a \sum_{l=1}^{k} \check{x}_{l-1})}{\check{\sigma}^2 + (k-1)\sigma_\alpha^2 (1-a)^2} \\
c_k^\alpha =& \frac{\sigma_\alpha^2 \check{\sigma}^2}{\check{\sigma}^2 + (k-1)\sigma_\alpha^2 (1-a)^2}
\end{aligned}
$$

showing that

$$
\mathbf{T}_k = (\sum_{l=1}^{k} \check{x}_l, \sum_{l=1}^{k} \check{x}_{l-1}, \sum_{l=1}^{k} \check{x}_l^2, \sum_{l=1}^{k} \check{x}_{l-1}^2, \sum_{l=1}^{k} \check{x}_l \check{x}_{l-1})
$$

is a sufficient statistic for $(a, \check{\sigma}^2, \alpha)$ given $(\check{x}_{1:k}, z_{1:k})$. A SISR algorithm which samples $(a, \check{\sigma}^2, \alpha)$ approximately from $p(a, \check{\sigma}^2, \alpha | T_k)$ using a few Gibbs sampling steps was therefore applied. $\check{x}_k$ was simulated from $p(\check{x}_k | a, \check{\sigma}^2, \alpha, \check{x}_{k-1})$. The weights $\tilde{w}_k^{(i)}$ were put to $p(z_k | \check{x}_k)$, ignoring the error introduced by using the approximative Gibbs sampling algorithm for simulating the static parameters.

In Figure 2, posterior means and quantiles obtained from the SISR filter using $N = 2000$ particles and 5 Gibbs sampler steps is plotted together with the same quantities calculated using full MCMC at each time step. Also in this case, there is a nice agreement between the estimates obtained by the particle filter and the ones given by the full MCMC runs.

5.2 A Gamma-Poisson model

Consider the model

$$\begin{aligned} x_k &\sim \mathrm{N}(ax_{k-1}, \sigma^2) \\ y_k &\sim \mathrm{Poisson}(x_k^2) \end{aligned}$$

Note that $x_k^2 \sim \mathrm{Gamma}(\frac{1}{2}, \frac{2\sigma^2}{1-a^2})$, so this can also be described as a Gamma-Poisson process where a controls the autocorrelation of the Gamma-process, while $\beta = \frac{2\sigma^2}{1-a^2}$ defines the scale of the Gamma variable.

Figure 3 shows the results obtained by using a SISR filter on this example where data was simulated from the true model using $a = 0.9$ and $\sigma = 1$. The filter was applied with $N = 2000$ particles. Because BUGS had problems handling this model, no comparison with full MCMC was performed in this case. Instead, comparison with a filter using $N = 50000$ particles was performed. Again the results are quite satisfactory.

5.3 A linear partial Gaussian process

The final example considered here is a simple linear model where the observation noise is assumed to follow a T-distribution. The model can be written as

$$x_k = ax_{k-1} + \sigma\varepsilon_k, \tag{13a}$$

$$z_k = x_k + \tau v_k, \tag{13b}$$

where $\{\varepsilon_k\}$ and $\{v_k\}$ are independent zero-mean white noise processes, the first being Gaussian with variance equal to one, while the other follows a T-distribution with ν degrees of freedom. Such models can be used to allow for outlier in the observations [22]. The unknown static parameters are in this case $\boldsymbol{\theta} = (a, \sigma^2, \tau^2)$ (ν is assumed known).

Since the system process follows the model discussed in Section 4.1, the sufficient statistics for (a, σ^2) can be updated according to the equations in (8). Given the model formulation above, no sufficient statistic for τ is available. Rewrite, however, the observation model as

$$z_k = x_k + \tau\check{v}_k/\sqrt{\delta_k},$$

where now $\check{v}_k$ is a Gaussian variable, while δ_k is an independent χ^2_ν-variable. The distribution for z_k is of course unaltered, but the point is that a sufficient statistic for τ given $z_{1:k}, x_{1:k}$ *and* $\delta_{1:k}$ is available. In particular,

$$[\tau^2 | x_{1:k}, z_{1:k}, \delta_{1:k}] \sim \mathsf{IG}(\frac{\eta_k}{2}, \frac{e_k}{2})$$

where

$$\begin{aligned} e_k &= e_{k-1} + \delta_k(z_k - x_k)^2, \\ \eta_k &= \eta_{k-1} + 1. \end{aligned}$$

Similar to the approach discussed in Section 4.2, (x_k, δ_k) is used as state vector in the particle filter. Note that simulation of δ_k from the prior is simply to draw from a χ^2_ν distribution, since there is no dynamic structure in this variable.

Data was simulated from this model with $a = 0.9, \sigma^2 = 1, \tau^2 = 1$ and $\nu = 5$. In Figure 4, posterior means and quantiles obtained from the SISR filter using $N = 2000$ particles is plotted together with the same quantities calculated using full MCMC at each time step. For the $\{x_k\}$ process and the estimates of a at each time step, both posterior means and quantiles are almost identical. Some discrepancy is present for the σ estimates, but the results are still acceptable. Also the estimates for τ are very good.

6 Discussion

The particle filter is a powerful method in processing a huge range of dynamic models. This paper discussed an approach based on the particle filter for tackling unknown parameters.

The approach has been tested on several different models, all giving estimates almost identical to the ones obtained by running full MCMC at each time point. Running full MCMC is not practical in real time processing, because the number of variables to be simulated increase in time. In contrast, the particle filters constructed in this paper only need a fixed amount of computation at each time point, making real time processing possible.

In order to make this approach work in real time, a crucial assumption is that the posterior distribution for the parameters depend on the underlying system process only through some sufficient statistics that can be updated recursively. For many models commonly applied, such sufficient statistics exists. In some other cases, the state vector can be extended by an additional variable, for which the extended model fit into the framework. When sufficient statistics are not available, the approach can still be applied, but the computational complexity will increase with time. In this paper, only Gaussian based system processes have been considered. The approach should however be possible to apply also for many other types of models.

Acknowledgement

This work was finalized during the author's sabatical at the Department of Statistics and Demography at the University of Odense. The author would like to thank the department for being able to use all their facilities during this period. I am also grateful to Bård Storvik for valuable comments on an earlier draft of this paper.

References

[1] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press, 1989.

[2] M. West and J. Harrison, *Bayesian forecasting and dynamic models*, Springer Series in Statistics. Springer-Verlag, New York, second edition, 1997.

[3] Künsch, H. R., "State space and hidden Markov models," in *Complex Stochastic Systems*, O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Eds., number 87 in Monographs on Statistics and Applied Probability, pp. 109–173. Chapman and Hall, London, 2001.

[4] B. D. O. Anderson and J. B. Moore, *Optimal filtering*, Electrical Engineering Series. Prentice Hall, 1979.

[5] N. Shephard, "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models with Econometric, Finance and Other Applications*, D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, Eds., pp. 1–67. Chapman and Hall, London, 1996.

[6] C. K. Wikle, M. Berliner, and N. Cressie, "Hierarchical Bayesian space-time models," *Environmental and Ecological Statistics*, vol. 5, no. 2, 1998.

[7] M. Ghil and K. Ide, "Data assimilation in meteorology and oceanography: Theory and practice," *J. Meteor. Soc. Japan, Special Issue*, vol. 75, no. 1B, pp. 111–496, 1997.

[8] B. Jørgensen, S. Lundbye-Christensen, P. X. K. Song, and L. Sun, "A longitudal study of emergence room visits and air pollution for Prince George British Columbia," *Statist. Med*, vol. 15, pp. 823–836, 1996.

[9] D. Gamerman, "Markov chain Monte Carlo for dynamic generalised models," *Biometrika*, vol. 85, no. 1, pp. 215–227, 1998.

[10] P. Müller, "Monte Carlo integration in general dynamic models," *Contemporary Mathematics*, vol. 115, pp. 145–163, 1991.

[11] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[12] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–288, 1994.

[13] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Computational and Graphical Statistics*, vol. 5, pp. 1–25, 1996.

[14] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza, "Dynamic conditional independence models and Markov chain Monte Carlo methods," *Journal of the American Statistical Association*, vol. 92, pp. 1403–141, 1997.

[15] A. Doucet, "On sequential simulation-based method for Bayesian filtering," Tech. Rep. TR.310, CUED/F-INFENG, 1998.

[16] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," in *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer-Verlag, 2001, To appear.

[17] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.

[18] M. West, "Mixture models, Monte Carlo, Bayesian updating and dynamic models," *Computing Science and Statistics*, vol. 24, pp. 325–333, 1993.

[19] E. Bølviken, P. J. Acklam, N. Christophersen, and J. M. Størdal, "Monte Carlo filters for non-linear state estimation," Tech. Rep. 14, University of Oslo, department of mathematics, section of statistics, 1997.

[20] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer-Verlag, 2001, To appear.

[21] J. S. Liu, R. Chen, and T. Logvinenko, "A theoretical framework for sequential importance sampling and resampling," in *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer-Verlag, 2001, To appear.

[22] N. Shephard, "Partial non-Gaussian state space," *Biometrika*, vol. 81, no. 1, pp. 115–31, 1994.

[23] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[24] W. R. Gilks and C.. Berzuini, "Following a moving target – Monte Carlo inference for dynamic Bayesian models," *Journal of Royal Statistical Society, Series B*, vol. 63, pp. 1–20, 2001.

[25] T. R. Carpenter, P. Clifford, and P. Fearnhead, "Building robust simulation-based filters for evolving data sets," Tech. Rep., Department of Statistics, University of Oxford, UK, 1998.

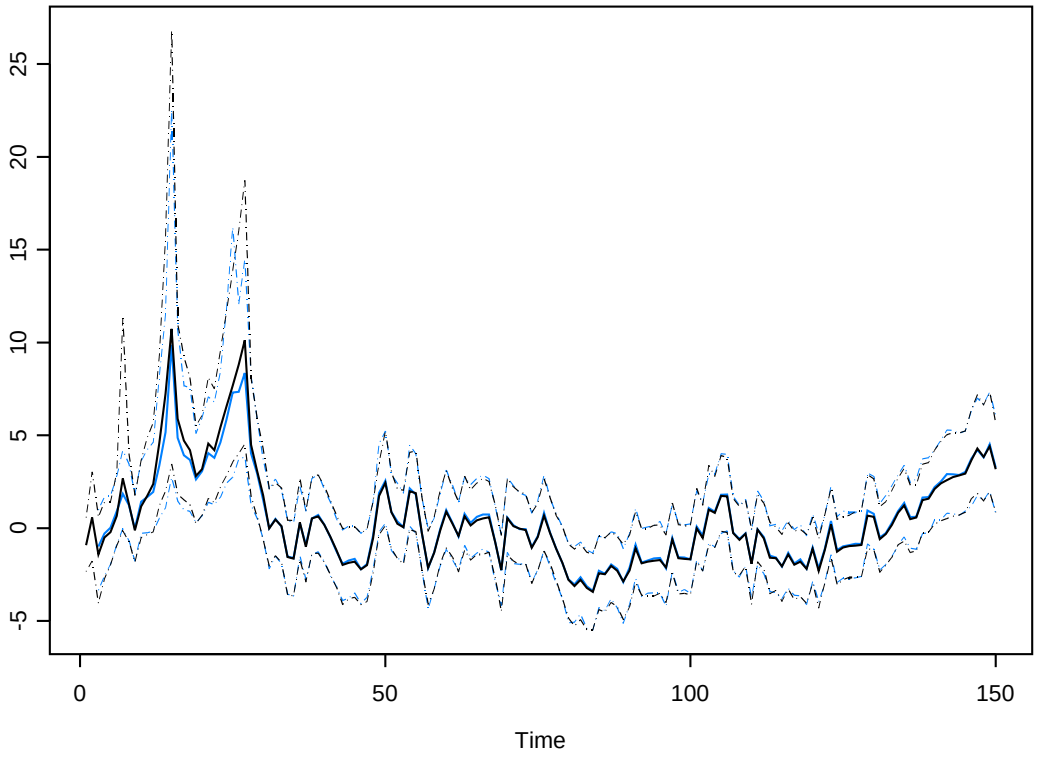
[26] G. Storvik, "Discussion on *Monte Carlo methods for dynamic Bayesian problems* by Berzuini and Gilks," Discussion contribution to the HSSS workshop on *Models and inference in HSSS: Recent developments and perspectives*, 2001.

[27] D. Crisan and A. Doucet, "Convergence of sequential Monte Carlo methods," Submitted for publication, 2000.

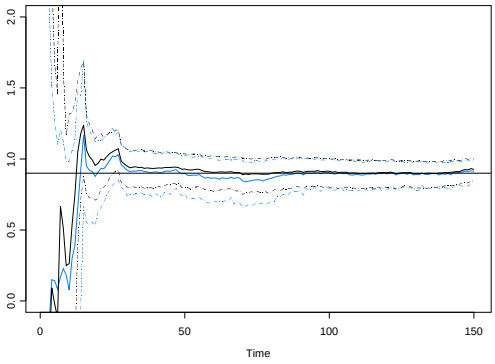
[28] D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks, "Bugs 0.5: Bayesian inference using Gibbs sampling, manual (version ii)," MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, 1996.

[29] M. West, J. Harrison, and H. S. Migon, "Dynamic generalized linear models and Bayesian forecasting," *Journal of the American Statistical Association*, vol. 80, pp. 73–97, 1985.

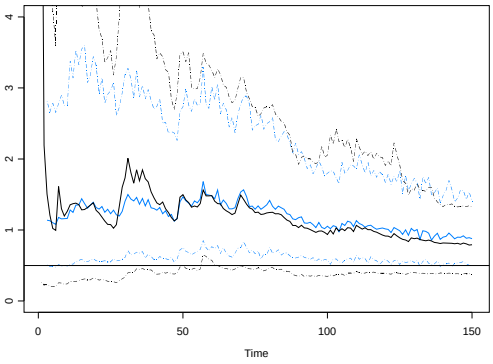
[30] E. Bølviken and G. Storvik, "Deterministic and stochastic particle filters in state space models," in *Sequential Monte Carlo in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer-Verlag, 2001, To appear.



Estimates of x_k

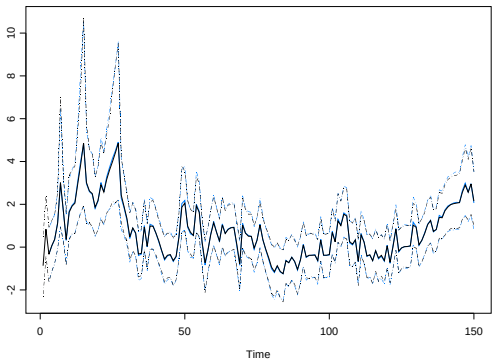


Estimates of a

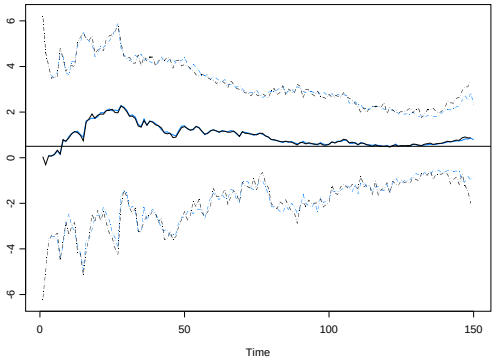


Estimates of σ

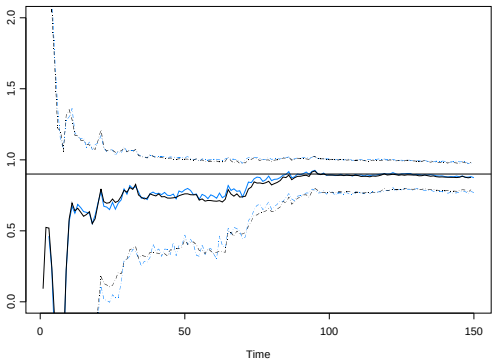
Figure 1: *Model* (10) *with* α *and* β *known. Posterior expectations (solid lines) and 0.025 and 0.975 quantiles (dashed lines) based on a particle filter (black) and full MCMC (grey).*



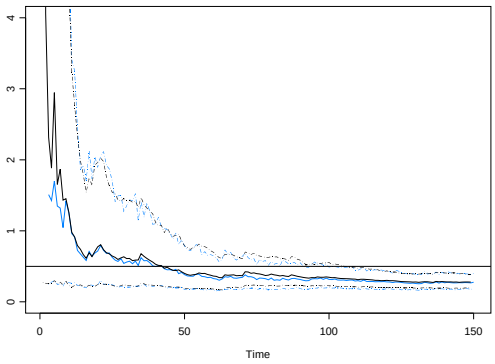
Estimates of x_k



Estimates of α

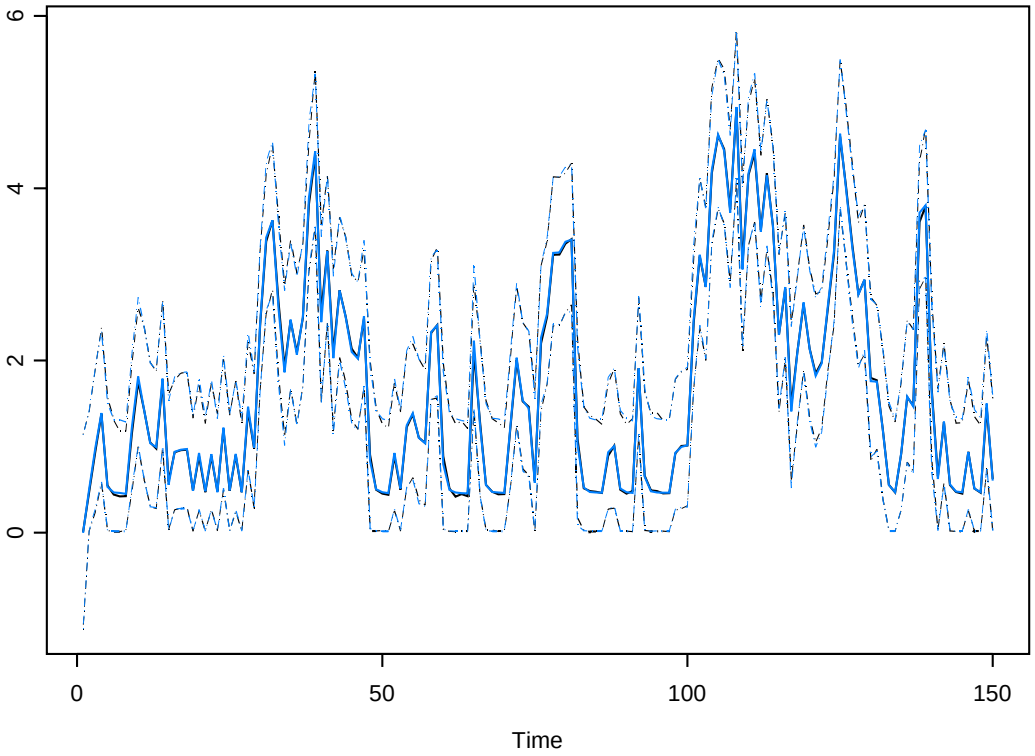


Estimates of a

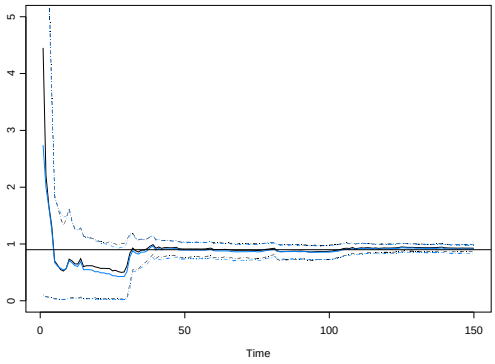


Estimates of σ

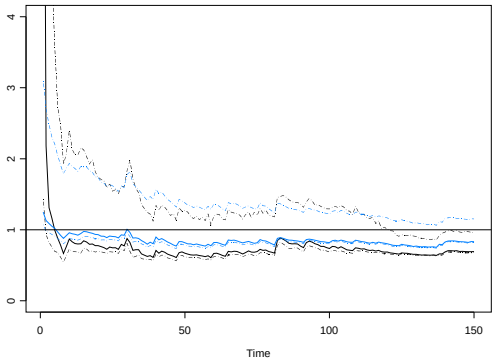
Figure 2: *Model* (10) *with* α *and* β *unknown. Posterior expectations (solid lines) and 0.025 and 0.975 quantiles (dashed lines) based on a particle filter (black) and full MCMC (grey).*



Estimates of x_k

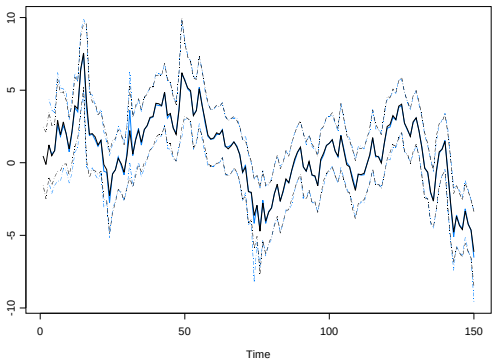


Estimates of a

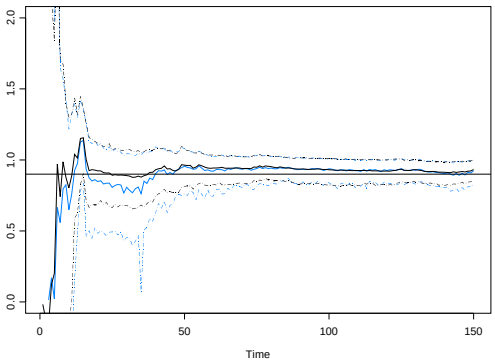


Estimates of σ

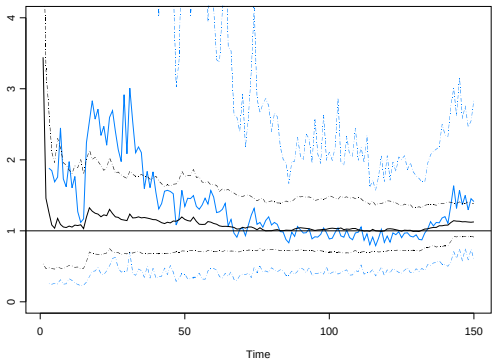
Figure 3: *Model* (12). *Posterior expectations (solid lines) and 0.025 and 0.975 quantiles (dashed lines) based on a particle filter using* $N = 2000$ *particles (black) and* $N = 50000$ *particles (grey).*



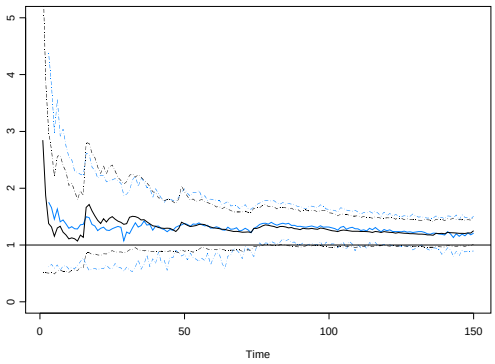
Estimates of x_k



Estimates of σ



Estimates of a



Estimates of τ

Figure 4: *Model* (13). *Posterior expectations (solid lines) and 0.025 and 0.975 quantiles (dashed lines) based on a particle filter (black) and full MCMC (grey).*