Received: 16 November 2016 | Revised: 25 July 2017 | Accepted: 18 August 2017

DOI: 10.1002/acs.2821

RESEARCH ARTICLE

WILEY

Marginalized approximate filtering of state-space models

K. Dedecius

Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 1143/4 Prague 8, 182 08, Czech Republic

Correspondence
K. Dedecius, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 1143/4, 182 08 Prague 8, Czech Republic.
Email: dedecius@utia.cas.cz

Funding information
Czech Science Foundation, Grant/Award Number: 16-09848S

Summary

The marginalized particle filtering (MPF) is a powerful technique reducing the number of particles necessary to effectively estimate hidden states of state-space models. This paper alleviates the assumption of a fully known and computationally tractable observation model. Exploiting the recent developments in the theory of approximate Bayesian computation (ABC) filtration, an ABC counterpart of MPF is proposed, applicable when the observation model is too complex to be evaluated analytically or even numerically, but it is still possible to sample from it by plugging in the state. The novelty is 2-fold. First, ABC methods have not been used in marginalized filtering yet. Second, a new multivariate robust method for evaluation of particle weights is proposed. The goal of this paper is to demonstrate the idea on the background of the MPF with a particular accent on exposition.

KEYWORDS

approximate filtering, filtering, marginalized filters, nonlinear filters, particle filtering, sequential Monte Carlo

1 | INTRODUCTION

State-space approach to modeling of time series is a well-established discipline popular in many engineering, financial, biological, and medicine applications (eg, the works of West and Harrison[1] and Durbin and Koopman[2]), and abundance of others. This paper focuses on a particular class of state-space models with the following specific structure:

$$\xi_{k+1} = f_k(\xi_k) + A_\xi x_k + G_\xi w_\xi \tag{1}$$

$$x_{k+1} = A_x x_k + G_x w_x \tag{2}$$

$$y_k \sim h_k(y_k|\xi_k), \tag{3}$$

where y_k are (possibly multivariate) observations; ξ_k and x_k are the hidden state variables; f_k is a real nonlinear multivariate function; $h_k(\cdot)$ is an observation model in the form of a probability density function; and A_x, A_ξ, G_x, and G_ξ are the matrices of compatible dimensions that may depend upon ξ_k (we omit the time indices for easy reading). The hidden process noise variable is normally distributed

$$w_k = \begin{bmatrix} w_\xi \\ w_x \end{bmatrix} \sim \mathcal{N}(0, Q) \tag{4}$$

with the (possibly time-varying) covariance matrix

$$Q = \begin{bmatrix} Q_\xi & Q_{\xi x} \\ (Q_{\xi x})^\mathsf{T} & Q_x \end{bmatrix}. \tag{5}$$

This model is particularly suited for navigation, integrating the inertial systems and terrain-aided positioning or Global Positioning System.[3,4]

Concerning the properties of f_k and h_k, 3 different cases may occur. First, if they are linear (or mildly nonlinear) functions, the celebrated Kalman filter (or its extended or unscented variants) dominates the class of possible solutions.[5-7] Second, if f_k is linear or nonlinear and h_k severely nonlinear but known and computationally tractable, the particle filters estimating x_k and ξ_k via a Monte Carlo sampling from the state space prevail, eg, the work of Cappé et al.[8] Moreover, the model structure allows marginalized (Rao-Blackwellized) particle filtering (MPF): ξ_k is sampled, whereas x_k is estimated via a Kalman filter.[9,10] The marginalization reduces the number of necessary particles and leads to a lower estimator variance.[3,9,10] The third case (*we focus on*) involves the frequent situations when h_k is too complex or even numerically intractable, or when it is a rough approximation of the true data-generating model (model misspecification), but the possibility to sample from the observation equation by plugging the state is preserved. Under these situations, the particle filters may require tedious adjustments (eg, chapter 3 in the work of Ristic[11]), special variants (eg, the works of Pitt and Shephard,[12] Maiz et al,[13] and Kotecha and Djurić[14]), or cannot be used at all.

In order to address these issues, the present paper adheres to the rapidly developing domain of the approximate Bayesian computation (ABC) methods, surveyed, eg, in the work of Green et al.[15] Essentially, these methods plug the Monte Carlo samples drawn from the state space directly into the observation model and evaluate their particle weights based on the proximity of the obtained pseudo-observations to the true observations y_k. The proximity is computed via a convenient kernel function.

Most of the ABC methods, ie, both static and sequential, adopt a uniform kernel assigning equal weights to particles yielding pseudo-observations within a predefined radius around the true observation and discarding the rest.[16-19] The generic ABC filter[20] is not an exception in this respect. As a result, it has attractive asymptotic properties ensuring convergence to the true state value, however, at the cost of rather impractical assumptions[21] requiring carefully designed adaptation of the kernel scale, eg, using a linear schedule[22] or effective sample size–based scale contraction.[23,24] An alternative method partially solving these issues is inspired by the kernel density estimation (KDE) theory.[25] It considers centered probability kernels with finite second-order moments and adapts the kernel scale exploiting the standard KDE criterion: minimization of the mean integrated square error (chapters 3.3 and 3.4 in the work of Silverman).[26] Recently, the author of this paper proposed yet another approach with *univariate* adaptive kernels that provides a good robustness and stability under much weaker assumptions.[27] Its idea is that the highest probability region (HPR) of the true observation model, ie, probability density $h_k(\cdot)$, would cover a corresponding region of possible observations, including the admissible pseudo-observations. As the true model is not available or usable, it is replaced by a convenient symmetric probability kernel centered at the true measurement y_k and with a scale parameter providing a required coverage of the pseudo-observations. This paper adopts this philosophy and extends it to multivariate cases in Section 3.3. This extension faces certain difficulties. For instance, the determination of the kernel scales in unidimensional cases is a straightforward use of the quantile function. However, there are no quantiles defined in the multivariate cases, and the determination needs to be based on the equiprobability curves. Section 3.4 presents an analytically tractable optimization of the bivariate normal kernel scale.

To summarize, this paper focuses on inference of partially linear state-space models given by Equations 1 to 3 with too complex, analytically or numerically intractable, or misspecified observation Equation 3. The *key novelty* is 2-fold. First, it lies in the estimation of the nonlinear part using an ABC filter. The linear part is estimated via standard linear algorithms. Second, a new robust multivariate kernel–based method for evaluation of particle weights is proposed. The resulting filter is very universal and can be used in a much wider range of problems and generalizations of the originally considered state-space model.

For expository purposes, the marginalized ABC filter is developed on the background of the marginalized particle filter of Nordlund and Gustafsson.[3] However, it should be stressed that the estimation of the nonlinear part is completely different.

2 | MARGINALIZATION AND IDENTIFICATION OF THE LINEAR AND NONLINEAR PARTS

The state-space model (1) to (3) can be partitioned into a nonlinear and a conditional linear part, involving ξ_k and $x_k|\xi_k$, respectively. Therefore, the joint filtering probability density for x_k and $\xi_{1:k} = \{\xi_1, \dots, \xi_k\}$ given observations

$y_{1:k} = \{y_1, \dots, y_k\}$ can be factorized as

$$\pi(x_k, \xi_{1:k}|y_{1:k}) = \pi(x_k|\xi_{1:k}, y_{1:k})\,\pi(\xi_{1:k}|y_{1:k}), \tag{6}$$

$$= \pi(x_k|\xi_{1:k})\,\pi(\xi_{1:k}|y_{1:k}), \tag{7}$$

which means that that the observations directly carry information about $\xi_{1:k}$ and, in turn, information about x_k can be inferred from $\xi_{1:k}$. The filtering density for $\xi_{1:k}$ arises from marginalization of the linear states from the joint filtering density

$$\pi(\xi_{1:k}|y_{1:k}) = \int \pi(x_k, \xi_{1:k}|y_{1:k})\,dx_k. \tag{8}$$

Let us now characterize the linear and nonlinear parts. The underlying factorization procedure, including decorrelation, is due to Nordlund and Gustafsson,[3] where the details can be found.

The linear part of the state-space model is given by

$$x_{k+1} \sim \mathcal{N}\left(\bar{A}_x x_k + G_x Q_{\xi x}\left(G_\xi Q_\xi\right)^{-1} z_k, G_x\left[Q_x - Q_{\xi x}^{\mathsf{T}}\left(Q_\xi\right)^{-1} Q_{\xi x}\right] G_x^{\mathsf{T}}\right), \tag{9}$$

$$z_k \sim \mathcal{N}\left(A_\xi x_k, G_\xi Q_\xi G_\xi^{\mathsf{T}}\right), \tag{10}$$

where

$$z_k = \xi_{k+1} - f_k(\xi_k) = A_\xi x_k + G_\xi w_\xi, \tag{11}$$

$$\bar{A}_x = A_x - G_x Q_{\xi x}\left(G_\xi Q_\xi\right)^{-1} A_\xi. \tag{12}$$

In order to avoid confusion with indices, we stick with z_k in the rest of this paper. This means the equivalence

$$\pi(x_k|z_{1:k}) \equiv \pi(x_k|\xi_{1:k}). \tag{13}$$

If the prior filtering density $\pi(x_k|z_{1:k-1}) \equiv \mathcal{N}(x_k^-, P_k^-)$, the recursive Bayesian filtering yields the Kalman filter. In the following text, the marks $^-$ and $^+$ denote variables before and after the incorporation of z_k, respectively.

The nonlinear part describing the evolution $\xi_k \to \xi_{k+1}$ requires the statistical knowledge of x_k carried by the filtering density $\pi(x_k|z_{1:k})$. The normality of $x_k|z_{1:k-1}$, the chain rule, and marginalization yield

$$\xi_{k+1} \sim \mathcal{N}\left(f_k(\xi_k) + A_\xi x_k^-, G_\xi Q_\xi G_\xi^{\mathsf{T}} + A_\xi P_k^- (A_\xi)^{\mathsf{T}}\right), \tag{14}$$

$$y_k \sim h_k(\,y_k|\xi_k). \tag{15}$$

Depending on the properties of h_k, this model can be sequentially estimated by a convenient Monte Carlo filter, eg, the particle or the ABC filter.

Remark 1. In many cases, the intrinsically Bayesian approach to deriving the density of $\xi_{k+1}|\xi_k$ via the chain rule and marginalization is connected with (often computationally demanding) approximations. A simple straightforward way around this issue is the use of the *plug-in distribution*, obtained by plugging in the point estimate of x_k into Equation 1

$$\xi_{k+1} \sim \mathcal{N}\left(f_k(\xi_k) + A_\xi x_k^-, G_\xi Q_\xi G_\xi^{\mathsf{T}}\right) \tag{16}$$

$$y_k \sim h_k(\,y_k|\xi_k). \tag{17}$$

This approach is asymptotically equivalent but may exhibit a discrepancy for small sample sizes as the plug-in distribution ignores the uncertainty about the inferred state.[28]

3 | ESTIMATION OF THE NONLINEAR PART

3.1 | Particle filtering

Assume a standard nonlinear state-space model given by

$$\xi_{k+1} \sim p(\xi_{k+1}|\xi_k), \tag{18}$$

$$y_k \sim h_k(\,y_k|\xi_k), \tag{19}$$

where, consistently with the previous section, ξ_k stands for a hidden state variable with a distribution $p(\xi_{k+1}|\xi_k)$, and y_k is an observation with a distribution $h_k(y_k|\xi_k)$.*

Under usual conditions where h_k is a tractable probability density, the particle filtration[8,29] of the nonlinear states ξ_k is achieved by the Bayesian update in the form

$$\pi(\xi_{1:k}|y_{1:k}) \propto \pi(\xi_{1:k-1}|y_{1:k-1})\, p(\xi_k|\xi_{k-1})\, h_k(y_k|\xi_k), \tag{20}$$

approximated in the point-mass sense by a set of I Monte Carlo samples $\xi_k^{(i)}$ drawn from the state space and assigned with weights $w_k^{(i)}$ summing to unity,

$$\pi(\xi_{1:k}|y_{1:k}) \approx \sum_{i=1}^{I} w_k^{(i)} \delta_{\xi_k^{(i)}}(\xi_k). \tag{21}$$

The term $\delta_{\xi_k^{(i)}}(\xi_k)$ stands for the Dirac delta function located at $\xi_k^{(i)}$. The particle filters update the weights $w_k^{(i)}$ via the Bayes' theorem

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{h_k\left(y_k|\xi_k^{(i)}\right) p\left(\xi_k^{(i)}|\xi_{k-1}^{(i)}\right)}{q\left(\xi_k^{(i)}|\xi_{0:k-1}^{(i)}, y_k\right)}, \tag{22}$$

where $q(\xi_k^{(i)}|\xi_{0:k-1}^{(i)}, y_k)$is a convenient proposal distribution. Identifying it with the state function (18) leads to the bootstrap filter,[29] where the particle weights are straightforwardly updated according to

$$w_k^{(i)} \propto w_{k-1}^{(i)} h_k\left(y_k|\xi_k^{(i)}\right). \tag{23}$$

A common drawback of most particle filters is the degeneracy of particle weights due to their gradual concentration to a few particles. A way around this problem is to use a resampling procedure, surrogating the original set of particles by a new one, more dense in regions of high probability.[8]

3.2 | Approximate filtering

Now, let us assume that h_k is *not a tractable or fully known* density, but only an approximate (ie, misspecified) density, a stochastic function describing the true noise-corrupted process, or any other convenient mathematical model approximating the true observation-generating system. To remedy this situation, the ABC filters[20,21,25,30] provide the point-mass approximation (21) of the target density $\pi(\xi_{1:k}|y_{1:k})$ using a different strategy. The Monte Carlo samples $\xi_k^{(i)}, i = 1, \ldots, I$ drawn from the state space are directly plugged into the observation model (19). It produces pseudo-observations $u_k^{(i)}$, and the weights $w_k^{(i)}$ of the state-space particles $\xi_k^{(i)}$ are calculated as proportional to the closeness of these pseudo-observations to the true observation y_k. This closeness is measured by a user-selected probability kernel function $\tilde{g}_{\varepsilon_k}(y_k, u_k)$ with a scaling parameter ε_k. That is, the ABC counterpart of Equation 20 takes the form

$$\tilde{\pi}(\xi_{1:k}|y_{1:k}) \propto \tilde{\pi}(\xi_{1:k-1}|y_{1:k-1})\, p(\xi_k|\xi_{k-1}) \int \tilde{g}_{\varepsilon_k}(y_k, u_k) h_k(y_k|\xi_k)\, du_k, \tag{24}$$

and the ABC counterpart of the particle weights update (23) under the bootstrap-type proposal distribution (14) is

$$\omega_k^{(i)} \propto w_{k-1}^{(i)} \tilde{g}_{\varepsilon_k}\left(y_k, u_k^{(i)}\right). \tag{25}$$

The choice of $\tilde{g}_{\varepsilon_k}$ is still an open problem. The first ABC-filtration paper of Jasra et al[20] adopted consistently with the static ABC framework the uniform kernel

$$\tilde{g}_{\varepsilon_k}\left(y_k, u_k^{(i)}\right) \propto \mathbb{1}_{A_{\varepsilon_k, y_k}}\left(u_k^{(i)}\right), \tag{26}$$

where

$$A_{\varepsilon_k, y_k} = \{u : \rho(u, y_k) \leq \varepsilon_k\}, \varepsilon_k > 0. \tag{27}$$

is a closed set determined by metric ρ. This *accept-reject* algorithm produces degenerate (uniform) posterior weights. Moreover, it is potentially unstable, eg, under outliers, where the kernel needs to be appropriately tuned to prevent a situation where all particles have zero posterior weights. The authors show that under stable modeling conditions, ie,

*Later, we will identify this model with the nonlinear submodel given by (14) and (15).

fixed ε_k and $I \to \infty$, the filter converges to a biased estimator, whose bias tends to zero if $\varepsilon_k \to 0$. Although this result is attractive in a long run, a nonuniform weighting strategy may lead to better estimates in shorter time-series scenarios.

The nonuniform kernels with adaptive scale parameters were proposed by Calvet and Czellar[25] to solve the double-convergence issues. Their plug-in–based method was inspired by the traditional KDE and naturally inherited its advantages and disadvantages discussed, eg, in Silverman's KDE monograph.[26] Recently, the author of this paper proposed an alternative method with an adaptive kernel tuning procedure,[27] where the main idea is that if the true observation model were known, its $p\%$ HPR would cover a related set of possible pseudo-observations. Since the true model is unknown, the adaptive kernel serves as its substitute. The idea comes from the comparison of Equations 20 and 24. The following section extends this method to multivariate cases.

3.3 | Adaptive multivariate kernels

The motivation of the proposed method is as follows: assume that there is a true probabilistic model, ie, probability density $h_k(y_k|\xi_k)$, which generates the observations y_k. Then, its $100p\%$ HPR would cover the corresponding region of possible observations, including the pseudo-observations $u_k^{(i)}$ generated by admissible samples $\xi_k^{(i)}$ from the corresponding state space. Since the true model is unknown or intractable, it is surrogated by a convenient symmetric probability kernel $\tilde{g}_{\varepsilon_k}(y_k, u_k^{(i)})$ centered at y_k with a scale ε_k assuring the coverage of a preset number of pseudo-observations $\alpha \in \{1, \dots, I\}$ (or their fraction, simplifying asymptotic analysis for $I \to \infty$) by a p-HPR of this kernel. Naturally, the resulting kernel is not equivalent to the true observation-generating model, but the asymptotic convergence of the filter is assured by the preservation of a high proportion of admissible state particles. Unlike in the work of Calvet and Czellar,[25] the existence of the first 2 moments is not necessary. The only requirement is that the kernel has the location and scale parameters, and it is symmetric.

While in unidimensional problems, the credibility region is easily found via the quantile function (inverse cumulative distribution function); in the multidimensional problems, we need to adopt the equiprobability curves, which makes the HPR calculation more elaborate. Below, in Section 3.4, a particularly appealing analytically tractable example of a bivariate normal kernel is given. It is also used in the simulation examples (see Section 4).

Assume now that a particular kernel function, the number α, and the coverage p are chosen a priori. Then, the kernel scale adaptation proceeds at each time instant k as follows:

1. The kernel with yet unknown scale parameter ε_k is centered at y_k by translation.
2. The pseudo-observations $u_k^{(i)}$ are simulated from the model, and the αth least distant pseudo-observation u_k^{α} is found.
3. Then, scale ε_k is calculated via constructing the equicoordinate p-credibility region with u_k^{α} lying on its boundary. The resulting ε_k is plugged into the kernel to calculate the posterior particle weights via Equation 25.

This approach to the evaluation of weights of $\xi_k^{(i)}$ is the core of the proposed marginalized ABC filter summarized in Algorithm 1. The subsequent section shows the derivation of the multivariate normal kernel scale.

Remark 2. A natural and important extension of the method postponed to further research is the construction of asymmetric kernels. Another investigated alternative to the current method is the approximation of multivariate kernels by independent marginals.

3.4 | Example: bivariate normal kernel

Let $y_k \in \mathbb{R}^2$. The bivariate normal kernel is the function

$$\tilde{g}_{\varepsilon_k}\left(y_k, u_k^{(i)}\right) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left[-\frac{1}{2}\left(u_k^{(i)} - y_k\right)^{\mathsf{T}} \Sigma^{-1}\left(u_k^{(i)} - y_k\right)\right], \tag{28}$$

where $\Sigma \in \mathbb{R}^{2\times 2}$ is a covariance matrix with $\varepsilon_k^2 > 0$ on the main diagonal and zeros elsewhere. Inspection of Equation 28 reveals the Mahalanobis distance

$$\left(r^{(i)}\right)^2 = \left(u_k^{(i)} - y_k\right)^{\mathsf{T}} \Sigma^{-1}\left(u_k^{(i)} - y_k\right) \tag{29}$$

$$= \varepsilon_k^{-2}\left\|u_k^{(i)} - y_k\right\|^2. \tag{30}$$

Now, the goal is to compute the scale ε_k such that $u_k^{(i)}$ lies on the equicoordinate delimiting the volume p of the HPR of the kernel density $\tilde{g}_{\varepsilon_k}(y_k, u_k^{(i)})$. For this purpose, we need to derive the inverse of the cumulative distribution function. Note that the equicoordinates given by Equation 30 are circles, whose polar parameterization is

$$u_k^{(i)} = r^{(i)}\varepsilon_k[\cos\phi, \sin\phi]^\mathsf{T} + y_k, \quad \phi \in [0, 2\pi]. \tag{31}$$

Its Jacobian is $r^{(i)}\sqrt{\det\Sigma}$. Let us substitute $\rho = r^{(i)}$ for the subsequent integration. The volume of kernel (28) at the point $u_k^{(i)}$ in polar parameterization is

$$p = \int_0^{2\pi}\int_0^{r^{(i)}} \rho\sqrt{\det\Sigma}\frac{1}{2\pi\sqrt{\det\Sigma}}e^{-\frac{1}{2}\rho^2}d\phi d\rho = \frac{1}{2\pi}\int_0^{2\pi} d\phi \int_0^{r^{(i)}} \rho e^{-\frac{1}{2}\rho^2}d\rho \tag{32}$$

$$= \int_0^{\frac{(r^{(i)})^2}{2}} e^{-s}ds \tag{33}$$

$$= 1 - e^{-\frac{(r^{(i)})^2}{2}}. \tag{34}$$

By taking inverse function and back-substitution for $r^{(i)}$,

$$\varepsilon_k^2 = -\frac{\left\|u_k^{(i)} - y_k\right\|^2}{2\ln(1-p)}. \tag{35}$$

The resulting ε_k^2 corresponding to $u_k^{(i)} \equiv u_k^\alpha$ is now used as the (squared) scaling parameter in Equation 28.

Algorithm 1 Marginalized ABC filter

Sample initial particles $\xi_1^{(i)}, i = 1, \dots, I$ from a suitable prior distribution. Assign uniform initial particle weights $\omega_0^{(i)} = 1/I$. Choose a kernel function $\tilde{g}_{\varepsilon_k}$, set credibility region level p and the associated number of particles α to be covered by it. For $k = 1, 2, \dots$ do:

Nonlinear filter update:

1. Obtain observation y_k.
2. Update particle weights $\omega_k^{(i)}$:
   - Simulate pseudoobservations $u_k^{(i)} \sim h_k(y_k|\xi_k^{(i)})$.
   - Calculate distances $||u_k^{(i)} - y_k||$.
   - Find u_k^α, the αth least distant pseudoobservation.
   - Calculate kernel scale ε_k.
   - New weights are given by $\omega_k^{(i)} \propto \omega_{k-1}^{(i)}\tilde{g}_{\varepsilon_k}(y_k, u_k^{(i)})$.
3. Resample according to a selected criterion.
4. Propagate particles using the proposal distribution.

Linear filter update:

1. Calculate Kalman gain.
2. Perform Kalman data update (for each particle).
3. Perform Kalman time update (for each particle).

4 | EXAMPLES

The purpose of the following 2 simulation examples is 2-fold. First, Example 1 shows that in the well-specified model case, where the generic particle filter performs well, the performance of the marginalized approximate filter is only a little worse, although it does not know the noise properties. Then, Example 2 demonstrates that under heavy-tailed (Cauchy) noise, the misspecification issue may significantly influence the tracking ability of the particle filter, whereas the approximate filter retains its stability.

4.1 | Example 1: well-specified model

The goal of the illustrative example is to demonstrate that the proposed marginalized ABC filter performance is close to the MPF. We consider the popular task of target tracking, where $T = [x_{1,k}, x_{2,k}]^\mathsf{T}$ is a position on a 2-dimensional Cartesian plane; 2 tracking sensors have positions $S_1 = [20, 50]^\mathsf{T}$ and $S_2 = [0, 10]^\mathsf{T}$. The state $x_k = [x_{1,k}, x_{2,k}, x_{3,k}, x_{4,k}]^\mathsf{T}$ consists of planar positions and associated velocities. This popular model (eg, the work of Crassidis and Junkins[4]) has the form

$$x_{k+1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + w_k \tag{36}$$

$$y_k = \begin{bmatrix} d(S_1, T) \\ d(S_2, T) \end{bmatrix} + v_k, \tag{37}$$

where $d(S_\cdot, T)$ denotes the Euclidean distance of the object from the respective sensor. For simulation purposes $\Delta t = 0.1$, and the noise variables are defined as

$$w_k \sim \mathcal{N}(0, Q) \qquad \text{with} \qquad Q = 0.2 \begin{bmatrix} \frac{\Delta t^3}{3} I_{2\times 2} & \frac{\Delta t^2}{2} I_{2\times 2} \\ \frac{\Delta t^2}{2} I_{2\times 2} & \Delta t I_{2\times 2} \end{bmatrix}, \tag{38}$$

$$v_k \sim \mathcal{N}(0, R) \qquad \text{with} \qquad R = 0.3 I_{2\times 2}, \tag{39}$$

where $I_{2\times 2}$ denotes the identity matrix of rank 2. Two hundred data samples are simulated from the model with the initial $x_0 = [15, 10, 0, 0]$. The proposed marginalized ABC filter is compared with its direct MPF counterpart; both use 1000 particles and systematic resampling[4] at each time step. The marginalized ABC filter lacks the full knowledge of the observation model and proceeds with the distances simulations. The prior state values for both filters are $x_1^- = [0, 0]^\mathsf{T}$ and $P_1^- = 100 I_{2\times 2}$, and the initial particles for nonlinear states are sampled from $\mathcal{N}([10, 10]^\mathsf{T}, 10 I_{2\times 2})$. The marginalized ABC filter setting is $\alpha = 300$ and $p = 0.6$.

The final root-mean-square error (RMSE) values averaged over 100 runs are [0.765, 0.561, 0.857, 0.613] in the case of the marginalized ABC filter and [0.476, 0.369, 0.691, 0.671] in the case of MPF. The marginalized ABC filter yields slightly worse results as expected, but the tracking is still very good. Moreover, the quality of estimation of the linear part is comparable to MPF. This can be seen from Figure 1, where the box plots depict the final RMSE values of the repeated

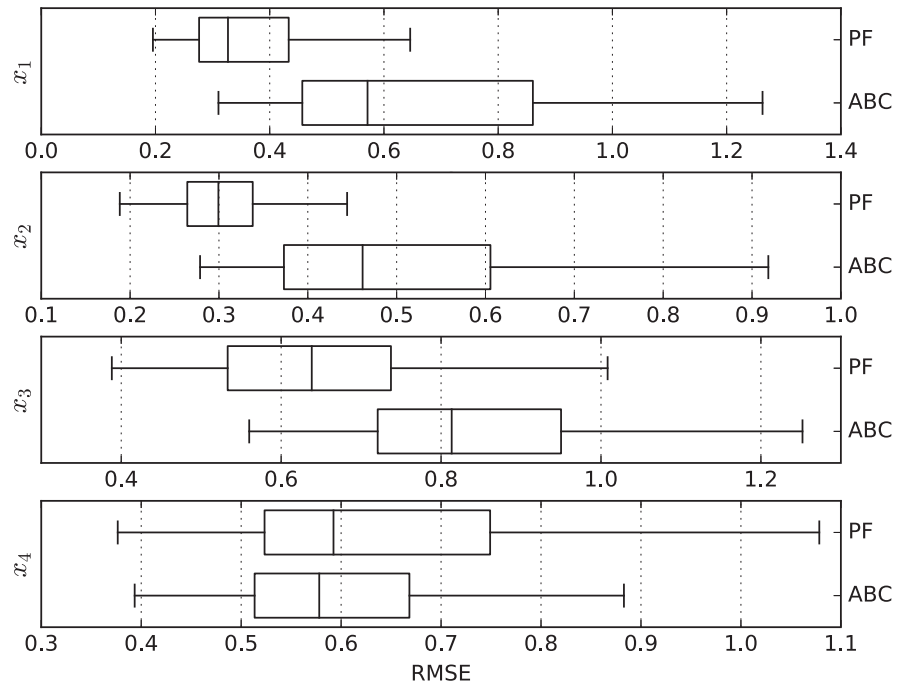

FIGURE 1 (Normal noise) Box plots of final root-mean-square error (RMSE) values of 100 repeated experiments. The boxes depict the median values and the upper and lower quartiles. The length of the whiskers is 1.5 times the interquartile range. Outliers are not shown. ABC, approximate Bayesian computation; PF, particle filter

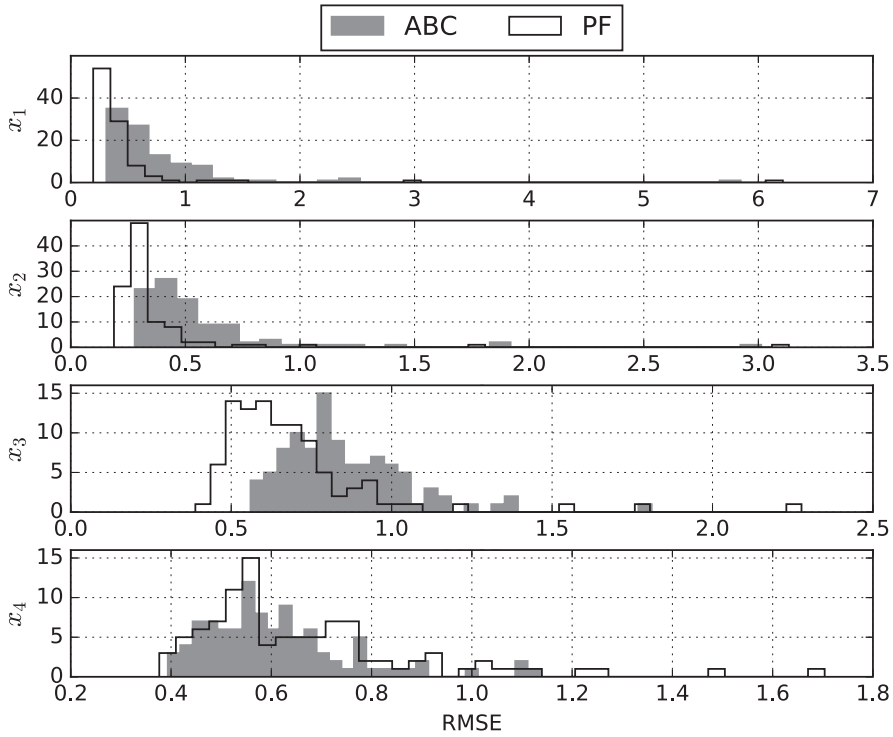

FIGURE 2 (Normal noise) Histograms of final root-mean-square error (RMSE) values of estimates of state variables $x_1, \ldots, x_4$. ABC, approximate Bayesian computation; PF, particle filter

experiments. The boxes depict the median values and the upper and lower quartiles. The length of the whiskers is 1.5 times the interquartile range. Outliers are filtered out. The distribution of the RMSE values shows Figure 2. The dominance of the particle filter is obvious; however, the results of the approximate filter are still very satisfactory, even the estimation of x_4 is slightly better than in the particle filter. Finally, Figure 3 depicts the averaged evolution of RMSE for both methods.

From the experimental viewpoint, the proposed marginalized ABC filter provides a good filtering quality under an observation model whose stochastic nature is not specified. Our experience shows robustness with respect to α and p; thus, there is no tedious tunning of parameters typical for many Monte Carlo methods.

4.2 | Example 2: misspecified model with heavy-tailed noise

The aim of the second example is to demonstrate that if the observation model noise is misspecified, then the tracking performance of the generic marginalized particle filter can be significantly worse than the performance of the approximate Bayesian filter that completely disregards the noise properties. For this purpose, the same model structure as in the previous example is adopted with the exception that the observation noise is Cauchy-distributed

$$v_k \sim Cauchy(0, R) \qquad \text{with} \qquad R = 0.3I_{2\times 2}, \tag{40}$$

where R denotes the scaling matrix. The noise realizations are uncorrelated. Otherwise, the setting of the example is identical to the previous example. Again, the results are averaged over 100 independent runs.

The final RMSE values averaged over the runs are [0.991, 0.813, 0.992, 0.723] in the case of the marginalized ABC filter and [14.28, 11.194, 3.608, 3.169] in the case of the marginalized particle filter. The latter clearly lost its tracking ability. This becomes apparent also from Figure 4 that shows box plots of the final RMSE values (outliers are filtered out). While the marginalized ABC filter retains its stable tracking performance, the marginalized particle filter has a significant dispersion of the final RMSE, indicating its instability and sensitivity to the heavy-tailed noise. Figure 5 compares the distribution of the final RMSE values by means of histograms. Obviously, the values of the marginalized ABC filter are highly concentrated, which is not the case of MPF. Finally, Figure 6 depicts the evolution of RMSE values averaged over the 100 experiment repetitions.

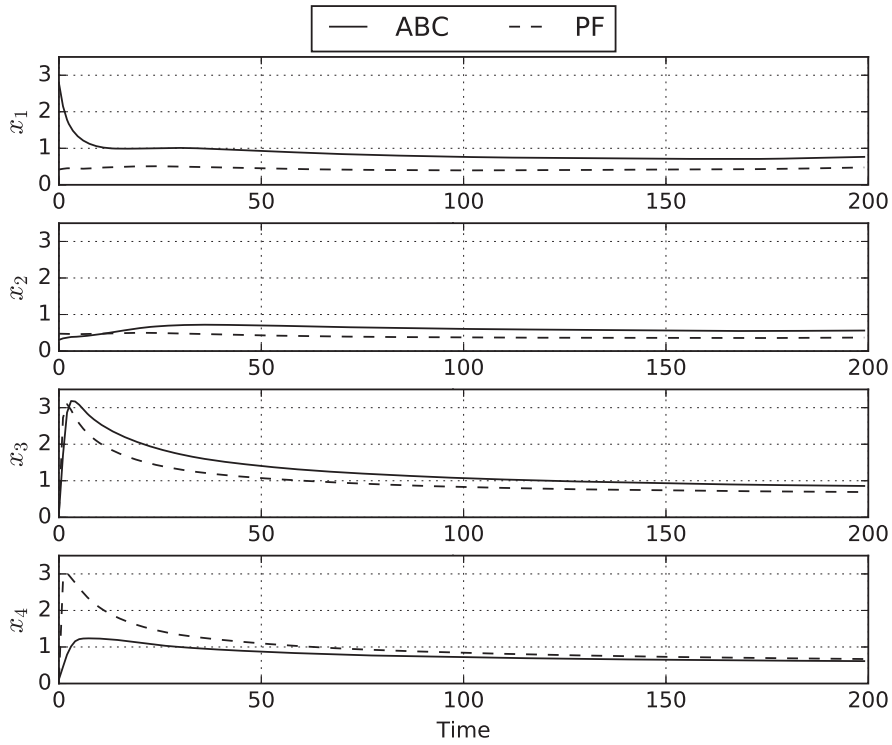

FIGURE 3 (Normal noise) Root-mean-square error (RMSE) evolution of estimates of x_1 (top) to x_4 (bottom) averaged over 100 independent experiment runs. ABC, approximate Bayesian computation; PF, particle filter

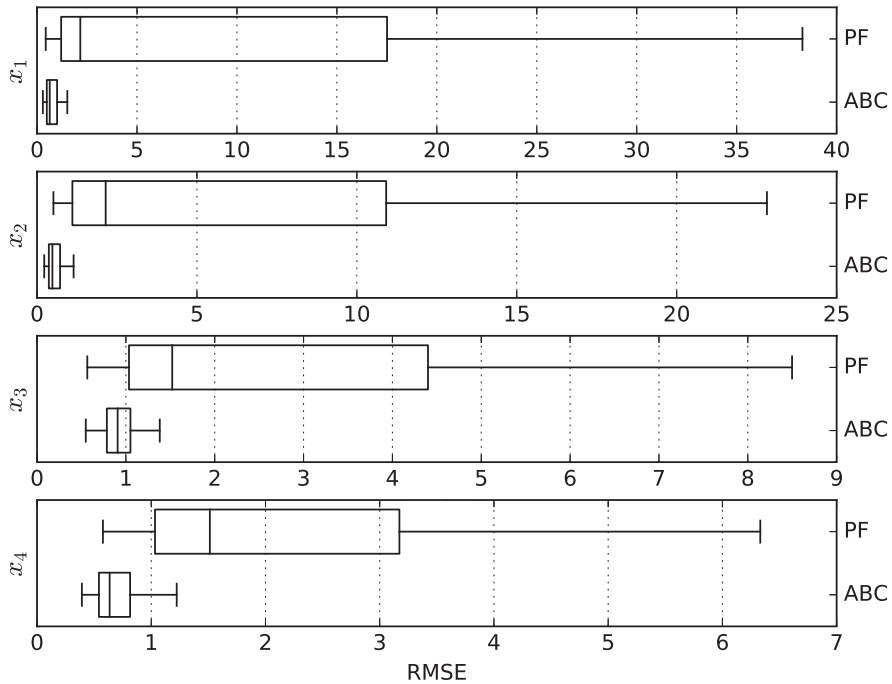

FIGURE 4 (Cauchy noise) Box plots of final root-mean-square error (RMSE) values of 100 repeated experiments. The boxes depict the median values and the upper and lower quartiles. The length of the whiskers is 1.5 times the interquartile range. Outliers are not shown. ABC, approximate Bayesian computation; PF, particle filter

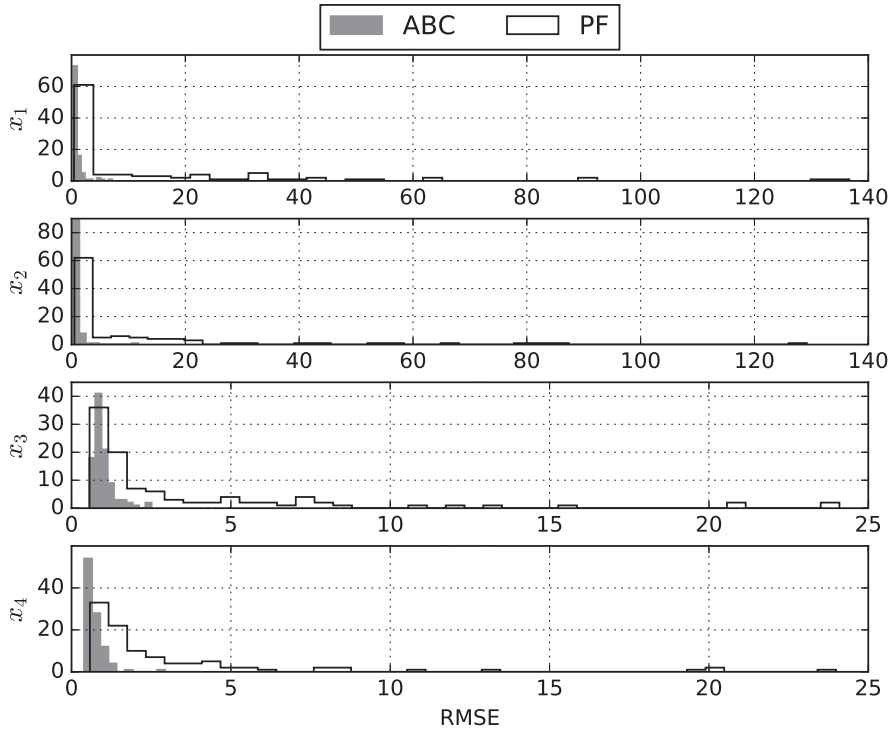

FIGURE 5 (Cauchy noise) Histograms of final root-mean-square error (RMSE) values of estimates of state variables $x_1, \ldots, x_4$. ABC, approximate Bayesian computation; PF, particle filter

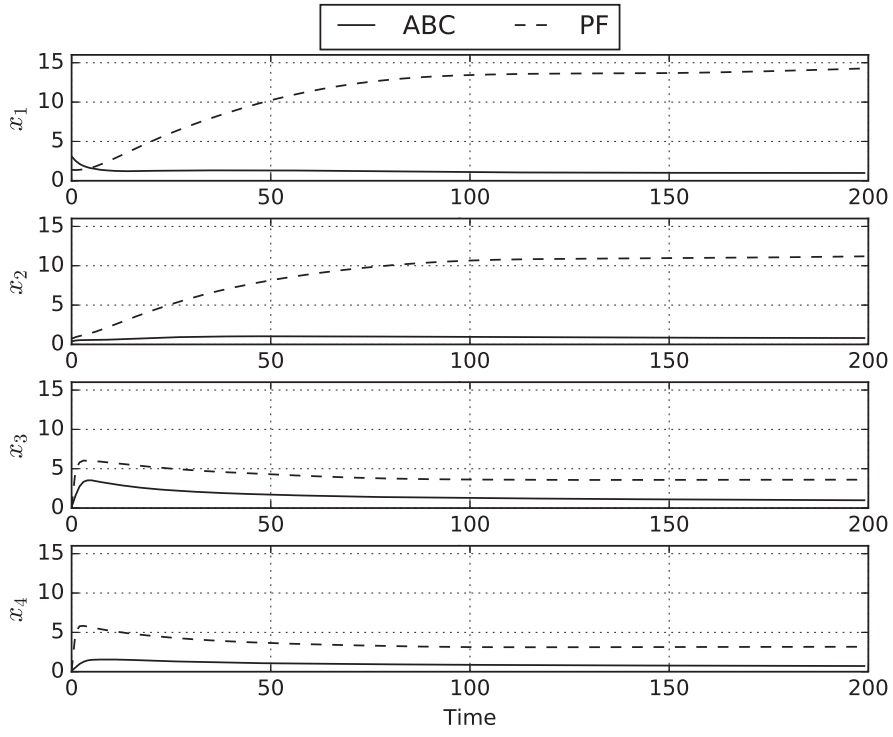

FIGURE 6 (Cauchy noise) Root-mean-square error (RMSE) evolution of estimates of x_1 (top) to x_4 (bottom) averaged over 100 independent experiment runs. ABC, approximate Bayesian computation; PF, particle filter

5 | CONCLUSION

The approximate Bayesian filters provided robust estimation of state-space models in situations where the standard Bayesian methods suffer from an observation model misspecification. This paper proposed a marginalized filter suitable

for simultaneous Monte Carlo-based estimation of nonlinear states and analytically tractable estimation of linear states. The Monte Carlo part, ie, the approximate Bayesian filter, additionally involves an adaptive kernel tuning procedure, providing robust evaluation of particle weights. Two simulation examples demonstrate that the tracking performance of the proposed filter is close to the generic marginalized particle filter under a well-specified model scenario and that the proposed filter provides superior results if the observation model is misspecified.

ACKNOWLEDGEMENTS

This work was supported by the Czech Science Foundation under grant 16-09848S. The author acknowledges the collaboration with M. Berka in the early stage of the research works.

ORCID

K. Dedecius http://orcid.org/0000-0002-2233-3530

REFERENCES

1. West M, Harrison J. *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag; 1997.
2. Durbin J, Koopman SJ. *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press; 2012.
3. Nordlund P-J, Gustafsson F. Sequential Monte Carlo filtering techniques applied to integrated navigation systems. Paper presented at: Proceedings of the 2001 American Control Conference, vol. 6; 2001; Arlington, WA, USA.
4. Crassidis JL, Junkins JL. *Optimal Estimation of Dynamic Systems*. Boca Raton, FL, USA: CRC Press; 2011.
5. Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng*. 1960;82:35-45.
6. Julier S, Uhlmann J, Durrant-Whyte HF. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans Autom Control*. 2000;45(3):477-482.
7. Simon D. *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley-Interscience; 2006.
8. Cappé O, Godsill SJ, Moulines E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proc IEEE*. 2007;95(5):899-924.
9. Andrieu C, Doucet A. Particle filtering for partially observed Gaussian state space models. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):827-836.
10. Schön T, Gustafsson F, Nordlund P-J. Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Trans Signal Process*. 2005;53(7):2279-2289.
11. Ristic B. *Particle Filters for Random Set Models*. New York: Springer; 2013.
12. Pitt MK, Shephard N. Filtering via simulation: auxiliary particle filters. *J Am Stat Assoc*. 1999;94(446):590-599.
13. Maiz CS, Molanes-Lopez EM, Miguez J, Djurić PM. A particle filtering scheme for processing time series corrupted by outliers. *IEEE Trans Signal Process*. 2012;60(9):4611-4627.
14. Kotecha J, Djurić PM. Gaussian sum particle filtering. *IEEE Trans Signal Process*. 2003;51(10):2602-2612.
15. Green PJ, Łatuszyski K, Pereyra M, Robert CP. Bayesian computation: a perspective on the current state, and sampling backwards and forwards. *Stat Comput*. 2015;25(4):835-862.
16. Jasra A. Approximate Bayesian computation for a class of time series models. *Int Stat Rev*. 2015;83(3):405-435.
17. Ehrlich E, Jasra A, Kantas N. Gradient free parameter estimation for hidden Markov models with intractable likelihoods. *Methodol Comput Appl Probab*. 2015;17(2):315-349.
18. Martin JS, Jasra A, Singh SS, Whiteley N, Del Moral P, McCoy E. Approximate Bayesian computation for smoothing. *Stoch Anal Appl*. 2014;32(3):397-420.
19. Yildirim S, Singh SS, Dean T, Jasra A. Parameter estimation in hidden Markov models with intractable likelihoods using sequential Monte Carlo. *J Comput Graph Stat*. 2015;24(3):846-865.
20. Jasra A, Singh SS, Martin JS, McCoy E. Filtering via approximate Bayesian computation. *Stat Comput*. 2012;22(6):1223-1237.
21. Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Stat Comput*. 2012;22(6):1167-1180.
22. Beskos A, Crisan D, Jasra A, Whiteley N. Error bounds and normalizing constants for sequential Monte Carlo samplers in high dimensions. *Adv Appl Prob*. 2014;46(1):279-306.
23. Biau G, Cérou F, Guyader A. New insights into approximate Bayesian computation. *Annales de lInstitut Henri poincare, Probabilites et Statistiques*. 2015;51(1):376-403.
24. Del Moral P, Doucet A, Jasra A. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat Comput*. 2012;22(5):1009-1020.
25. Calvet LE, Czellar V. Accurate methods for approximate Bayesian computation filtering. *J Financ Econ*. 2015;13(4):798-838.
26. Silverman B. *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman and Hall; 1986.
27. Dedecius K. Adaptive kernels in approximate filtering of state-space models. *Int J Adapt Control Signal Process*. 2017;31(6):938-952.

28. Smith RL. Bayesian and frequentist approaches to parametric predictive inference. *Bayesian Stat*. 1999;6:589-612.
29. Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc Radar Signal Process*. 1993;140(2):107.
30. Dean T, Singh S, Jasra A, Peters G. Parameter estimation for hidden Markov models with intractable likelihoods. *Scand J Stat*. 2014;41(4):970-987.

How to cite this article: Dedecius K. Marginalized approximate filtering of state-space models. *Int J Adapt Control Signal Process*. 2017;1–12. https://doi.org/10.1002/acs.2821