Robust Inference by Particle Filtering

Robust inference by particle filtering

Víctor Elvira
IMT Lille Douai & CRIStAL (UMR CNRS 9189), Villeneuve d'Ascq, France - victor.elvira@imt-lille-douai.fr

Joaquín Míguez
Universidad Carlos III de Madrid, Leganés, Spain - joaquin.miguez@uc3m.es

Petar M. Djurić
Stony Brook University, Stony Brook, USA - petar.djuric@stonybrook.edu

Abstract

Particle filters (PFs) are recursive Monte Carlo methods for online tracking and forecasting in state-space systems. They are very general and, hence, can be used with a broad class of models, including ones that are nonlinear and/or non-Gaussian. PFs suffer from a number of drawbacks including their computational complexity and sensitivity to the choice of the state space model (i.e., its compatibility with the observed data). Indeed, modelling errors and sharp changes in the dynamics of the state or the observation processes that are not accounted for usually lead to a degradation of the performance of the PFs. In this paper we draw from recent results on online assessment of convergence of PFs to propose a simple scheme to (a) detect changes in a sate-space model from a series of observations that are described by the model and (b) re-estimate the model to make it compatible with the observed data. The detection stage is fully general, as it relies on a model-invariant statistic, while re-estimation can be done in several manners. Here, we discuss possible schemes and illustrate the theory with a simple example for a conditionally-linear Gaussian model.
Keywords: Particle filter, sequential Monte Carlo, state-space models, change-point detection, model error.

1 Introduction

State space models (SSMs) are broadly used for the formal representation of partially observed dynamical systems with uncertainty (see, e.g., Anderson & Moore (1979)). They involve two random sequences: the state of the system, $\{X_t\}_{t\geq 0}$, usually not directly observed, and the sequence of measurements or observations $\{Y_t\}_{t\geq 1}$. The system dynamics and the link between the state and the observations are often described probabilistically by the triplet

$$X_0 \sim K_0(x_0|\theta), \quad X_t \sim K(x_t|x_{t-1},\theta), \quad Y_t \sim g(y_t|x_t,\theta), \tag{1}$$

where θ is a vector of model parameters, $K_0(\cdot|\theta)$ is the probability density function (pdf) of the a priori distribution of the state at time $t = 0$, $K(x_t|x_{t-1},\theta)$ is the conditional density of a Markov kernel that describes the state dynamics and $g(y_t|x_t,\theta)$ is the conditional pdf of the observation y_t given the state x_t. Although several extensions are possible (e.g., allowing for K and g to change over time), model (1) captures the key ingredients of SSMs as they are used in a wide range of practical applications.
Given a sequence of observations $Y_t = y_t$, $t = 1, 2, \ldots$, the Bayesian filtering problem with an SSM consists in the computation of the posterior probability distribution of the state X_t, denoted Π_t, at each time step (Anderson & Moore, 1979; Bain & Crisan, 2008). This problem does not admit a closed form solution except in very particular cases, namely when model (1) is linear and Gaussian (then the sequence of Π_t's is Gaussian and can be computed via the Kalman filter; see, e.g., Anderson & Moore (1979)) or when the state space is discrete and finite.
Particle filters (PFs) are recursive Monte Carlo algorithms that approximate the sequence of posterior distributions π_t. Each approximation consists of a collection of Dirac delta measures located at random positions

in the state space, possibly with non-uniform weights. PFs can be used over a very broad class of SSMs, including nonlinear and non-Gaussian models (see Del Moral (2004); Bain & Crisan (2008) for a comprehensive treatment), with known and unknown parameters, as recently proved by Crisan & Miguez (2016) and Del Moral *et al.* (2016). While very flexible and conceptually simple, PFs suffer from a number of drawbacks related to their computational complexity and their sensitivity to mismatches in the specification of the SSM. Indeed, modelling errors and/or sharp changes in the dynamics of the state or the observation processes that are not accounted for can often cause a severe degradation in the performance of PFs. Such change points in the series of observations $\{Y_t\}_{t\geq 1}$ can occur, e.g., when any parameters in vector θ (assumed static in model (1)) change their value suddenly.

In this paper we draw from recent results on the validation of SSMs proposed by Djurić & Míguez (2010) and on the online assessment of convergence of PFs, as introduced by Elvira *et al.* (2017), and propose a simple scheme to (a) detect change points in the observed series $\{Y_t\}_{t\geq 1}$ and (b) re-calibrate the SSM to make it compatible with the observed data. The detection stage is fully general, as it relies on a certain predictive statistic which is invariant with respect (w.r.t.) to the SSM, while model re-calibration can be carried out in several manners. Here, we discuss possible schemes and illustrate the theory with a simple example for a conditionally-linear Gaussian model.

In the rest of the paper, we first recall some basic background material on the standard PF and explicitly describe the problem of detecting change points in the series of observations (before the performance of the PF deteriorates). The method for change-point detection is introduced in Section 3. It yields a general particle filtering scheme robust to sharp changes in the SSM. A simple, illustrative example for conditionally-linear Gaussian model is provided in Section 4 and, finally, conclusion are presented in Section 5.

2 Background and problem statement

We assume the SSM of Eq. (1), where the state sequence, X_t, takes values in $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ and the observation sequence, Y_t, takes values in $\mathcal{Y} \subseteq \mathbb{R}^{d_y}$, with $d_x, d_y \geq 1$. The parameter, θ, takes values in $\mathcal{D} \subseteq \mathbb{R}^{d_\theta}$, and $d_\theta \geq 1$ as well. The parameters are unknown in general and we model θ as a (possibly multidimensional) random variable (r.v.) with prior pdf $p_0(\theta)$. Given (1) and $\theta \sim p_0(\theta)$, we are implicitly assuming that all probability distributions of interest in the SSM have densities w.r.t. the Lebesgue measure.

The conditional pdf of the parameter vector θ given the states $X_{0:t} = x_{1:t}$ is denoted $p_t^x(\theta|x_{0:t})$ and the posterior density of θ given the state $X_t = x_t$ and the observations $Y_{1:t-1} = y_{1:t-1}$ is denoted $p_t^{xy}(\theta|x_t, y_{1:t-1})$. Here we use the notation $x_{t_1:t_2} = \{x_{t_1}, x_{t_1+1}, \ldots, x_{t_2}\}$ for sequences. We write the density of X_t conditional on the sequence $X_{0:t-1} = x_{0:t-1}$ as

$$\bar{K}_t(x_t|x_{0:t-1}) := \int K(x_t|x_{t-1}, \theta) p_{t-1}^x(\theta|x_{0:t-1}) d\theta,$$

while the likelihood of $X_t = x_t$ given the observations $Y_{1:t} = y_{1:t}$ is

$$\bar{g}_t(y_t|x_t, y_{1:t-1}) := \int g(y_t|x_t, \theta) p_{t-1}^{xy}(\theta|x_t, y_{1:t-1}) d\theta.$$

We assume the ability to sample from $\bar{K}_t$ and to evaluate $\bar{g}_t$, either exactly or with sufficient numerical accuracy. This means that we assume the ability to integrate out the parameter vector θ from the SSM, an action that is usually termed Rao-Blackwellisation in the particle filtering literature (e.g., in Doucet *et al.* (2000)). Note that in the above two equations we tacitly assume that the parameters that define the parameter vector θ are not shared by the conditional density of the Markov Kernel, $K(x_t|x_{t-1}, \theta)$ and the conditional pdf $g(y_t|x_t, \theta)$.

We aim at the recursive approximation of the posterior probability measure of the state X_t given the observations $Y_{1:t} = y_{1:t}$. This posterior measure is denoted $\Pi_t(A) := \mathsf{Prob}(X_t \in A|y_{1:t})$, for any Borel subset $A \subseteq \mathcal{X}$ and we assume it has an associated density $\pi_t(x_t|y_{1:t})$ w.r.t. the Lebesgue measure. The sequence of measures Π_t, $t \geq 1$, can be approximated using the classical sequential importance resampling (SIR) method of Doucet *et al.* (2000), that we outline as Algorithm 1. We use the common term *particles* for Monte Carlo samples in the state space $\mathcal{X}$, and denote them as x_t^i (for the i-th particle at time t).

***Algorithm* 1.** *SIR algorithm with N particles.*

1. Initialisation: *At time $t = 0$, draw N independent particles $x_0^1, \ldots, x_0^N$ from the prior pdf $K_0(x_0)$.*
2. Recursive step: *Assume that we have generated sequences $x_{0:t-1}^i$, $i = 1, ..., N$, up to time $t-1$. At time t, proceed as follows:*
   (a) *Draw new particles $\bar{x}_t^i \sim \bar{K}_t(x_t|x_{0:t-1}^i)$, $i = 1, ..., N$.*
   (b) *Compute normalised importance weights $w_t^i \propto \bar{g}_t(y_t|x_t^i, y_{1:t-1})$, $i = 1, ..., N$.*
   (c) *Resample: for $i = 1, ..., N$, set $x_{0:t}^i = \{\bar{x}_t^j, x_{0:t-1}^j\}$, for some $j \in \{1, \ldots, N\}$, with probability w_t^j.*

Algorithm 1 describes a generic Rao-Blackwellised PF (RB-PF). When the parameter vector θ is known, the RB-PF reduces to the standard bootstrap filter of Gordon *et al.* (1993), where the i-th particle $\bar{x}_t^i$ is drawn from the Markov kernel $K(x_t|x_{t-1}^i)$ and its weight is, simply, $w_t^i \propto g(y_t|\bar{x}_t^i)$. At the (multinomial) resampling step, we simply set $x_t^i = \bar{x}_t^j$ with probability w_t^j. Note that, for known θ, there is no need to keep record of the complete sequences of states $x_{0:t}^i$, but only the last particle x_t^i. For the rest of this paper we assume θ is a r.v. (hence, unknown) and take Algorithm 1 as our basic building block.

The SIR method produces several Monte Carlo approximations. If δ_x denotes the unit Dirac delta measure located at $x \in \mathcal{X}$, the random measure $\Pi_t^N(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_t^i}(dx_t)$, is an estimate of the posterior measure $\Pi_t(dx) = \pi_t(x_t|y_{1:t})dx_t$. Moreover, the random measure $\Xi_t^N(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_t^i}(dx_t)$ is an approximation of the one-step-ahead predictive measure $\Xi_t(dx_t) = \xi_t(x_t|y_{1:t-1})dx_t$, where $\xi_t(x_t|y_{1:t-1})$ is the pdf of X_t conditional on $Y_{1:t-1} = y_{1:t-1}$. Finally, $\mu_t^N(y_t|y_{1:t-1}) := \frac{1}{N}\sum_{i=1}^N \bar{g}_t(y_t|\bar{x}_t^i, y_{1:t-1})$ is an estimate of the predictive pdf of the observation Y_t given the data record $Y_{1:t-1} = y_{1:t-1}$, namely,

$$\mu_t(y_t|y_{1:t-1}) = \int \bar{g}_t(y_t|x_t, y_{1:t-1})\xi_t(x_t|y_{1:t-1})dx_t \approx \frac{1}{N}\sum_{i=1}^N \bar{g}_t(y_t|\bar{x}_t^i, y_{1:t-1}) = \mu_t^N(y_t|y_{1:t-1}).$$

It is straightforward from the results in Míguez *et al.* (2013) to prove that Π_t^N and Ξ_t^N converge to Π_t and Ξ_t, respectively, in L_p as $N \to \infty$, while $\mu_t^N(y_t|y_{1:t-1}) \to \mu_t(y_t|y_{1:t-1})$ point-wise as $N \to \infty$.

We assume a scenario where the parameters of the SSM of Eq. (1) may undergo a significant change at an unknown time instant $t = T_{cp}$. This is a change-point where the dynamics of the state X_t or the conditional pdf $g(y_t|x_t, \theta)$ of the observations changes abruptly, possibly causing the SIR Algorithm 1 to loose track of the sequence of posterior distributions Π_t. With our description of a generic SSM, we associate a change-point to an instantaneous (and significant) change of the parameter vector θ, which is assumed unknown but static in the construction of Algorithm 1.

Our aim is to extend Algorithm 1 in such a way that such change-points can be detected and a suitable action can be taken to keep the sequence of approximate measures Π_t^N locked to the true posteriors Π_t.

3 Online change-point detection

We propose an online change-point detection scheme based on the SSM validation method introduced in Djurić & Míguez (2010) and the technique for online assessment of the performance of PFs introduced by Elvira *et al.* (2017).

The method is based on assessing the observation $Y_t = y_t$ by means of the predictive pdf $\mu_t(y_t|y_{1:t-1})$ at each time step. As discussed in Section 2 above, $\mu_t(y_t|y_{1:t-1})$ cannot be obtained in a closed form, but it can be approximated by the random pdf $\mu_t^N(y_t|y_{1:t-1})$. Here we assume that it is easy to generate independent and identically distributed (i.i.d.) samples from the density $\mu_t^N(y_t|y_{1:t-1})$. Under regularity assumptions, and for sufficiently large N, it can be proved (see Elvira *et al.* (2017)) that the actual observation $Y_t = y_t$ and the samples from $\mu_t^N(y_t|y_{1:t-1})$ are (approximately) i.i.d. *under the (null) hypothesis that the observation $Y_t = y_t$ is indeed generated by the SSM of* (1). Otherwise, if there is a mismatch or modelling error in the SSM, the samples from $\mu_t^N(y_t|y_{1:t-1})$ and $Y_t = y_t$ are not identically distributed in general.

Algorithm 2. *Online change-point detection with non-overlapping windows of length W.*

1. *At every time t,* ***[computation of the statistic]:***
   (a) *Draw $\tilde{y}_t^{(k)} \sim \mu_t^N(y_t|y_{t-1})$, for $k = 1, \ldots, K$.*
   (b) *Compute the realisation $A_{K,t} = a_{K,t}$, i.e., the position of y_t within the set of ordered fictitious observations $\{\tilde{y}_t^{(k)}\}_{k=1}^K$.*
2. *If $t = nW$ for $n\mathbb{N}$,* ***[at the end of each window of length*** *W]:*
   (a) *Compute the χ_t^2 statistic over the empirical distribution of $\mathcal{S}_t = \{a_{K,t-W+1}, a_{K,t-W+2}, ..., a_{K,t}\}$ as*

$$\chi_t^2 = \sum_{j=0}^{K} \frac{(O_j - E_j)^2}{E_j}, \tag{2}$$

   where $O_j = |\{a_{K,\tau} \in \mathcal{S}_t : a_{K,\tau} = j\}|$ is the frequency of $a_{K,\tau} = j$ and $E_j = W\mathbb{Q}_K(j) = \frac{W}{K+1}$ is the expected frequency under the null hypothesis.
   (b) *Calculate the p-value $p_{K,t}^*$ by comparing χ_t^2 with the χ^2-distribution with K degrees of freedom.*
   (c) *If $p_{K,t}^* \leq \gamma$:*
   - *Declare a change-point.*
   - *Fit the SSM using $y_{t-W+1}, \ldots, y_t$ (see the example in Section 4).*

To be specific, we simulate K "fictitious observations", denoted $\tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(K)}$, from the approximate predictive pdf $\mu_t^N(y_t|y_{1:t-1})$. Then, we define the set $\mathcal{A}_{K,t} := \{y \in \{\tilde{y}_t^{(k)}\}_{k=1}^K : y < y_t\}$, where y_t is the actual observation at time t, and the r.v. $A_{K,t} := |\mathcal{A}_{K,t}| \in \{0, 1, ..., K\}$. Note that $\mathcal{A}_{K,t}$ is the set of fictitious observations which are smaller than the actual one, while $A_{K,t}$ is the number of such observations. If we let $\mathbb{Q}_K$ denote the probability mass function (pmf) of A_K, it is not hard to show that $\mathbb{Q}_K$ is uniform in $\{0, 1, ..., K\}$, independently of the value and distribution of y_t (hence independently of the underlying SSM). See Elvira *et al.* (2017) for a proof of this result.

The online change-point detection scheme proposed here is a Pearson's χ^2 test performed over the sequence $\mathcal{S}_t = \{A_{K,t-W+1}, \ldots, A_{K,t}\}$ of the last W samples of the statistic $A_{K,t}$. The subsequence $\mathcal{S}_t$ can be constructed either as a sliding window or as non-overlapping subsequences of length W. The null hypothesis is that the W samples in $\mathcal{S}_t$ are i.i.d. draws from the uniform distribution in $\{0, 1, ..., K\}$. When the p-value output by the test lies below a predefined significance level γ, we declare a change-point. Algorithm 2 outlines the proposed online change-point detection scheme. This procedure can be readily combined with the SIR method, outlined as Algorithm 1, to yield a PF that automatically detects changes in the underlying SSM, provided that N is large enough to guarantee $\mu_t^N(y_t|y_{1:t-1}) \approx \mu_t(y_t|y_{1:t-1})$.

The re-calibration of the SSM can be carried out in different ways. For the RB-PF, a natural option is to reset $p_t^x(\theta|x_{0:t}) = p_t^{xy}(\theta|x_t, y_{1:t-1}) = p_0(\theta)$ for $t = (n-1)W + 1$ when a change-point is detected. In this way, we 'drop' the information conveyed by the old observations $y_{1:(n-1)W}$ (as it does not appear compatible with the latest window $y_{(n-1)W+1}, ..., y_{nW}$), and re-start the computation of the posterior distributions of θ. Alternatively, if we prefer to operate a standard PF with an estimate of the parameter, say $\hat{\theta}_t$, when a change-point is declared at time $t = nW$, we can drop the parameter estimate $\hat{\theta}_{nW}$ and re-start the parameter estimation algorithm from time $t = (n-1)W + 1$ onwards.

4 Example: a conditionally-linear Gaussian SSM

As a simple illustration of the method, let us assume the conditionally-linear and Gaussian SSM given by

$$X_t = \theta X_{t-1} + U_t \quad \text{and} \quad Y_t = X_t + V_t, \tag{3}$$

Algorithm 3. *SIR algorithm with N particles for the SSM of Eq.* (3)

1. Initialisation: *At time $t = 0$, draw N independent particles $x_0^1, \ldots, x_0^N$ from the prior pdf $N(x_0|\sigma_0^2)$. Let $\hat{\theta}_0^i = 0$ and $\sigma_{\theta,0}^{i,2} = \sigma_\theta^2$ for every $i = 1, ..., N$.*

2. Recursive step: *Assume that we have generated the triplets $\left\{x_{0:t-1}^i, \hat{\theta}_{t-1}^i, \sigma_{\theta,t-1}^{i,2}\right\}_{i=1}^N$, up to time $t-1$. At time t, proceed as follows:*

   (a) *For $i = 1, ..., N$,*

   a.1) *compute $\sigma_{x,t}^{i,2} = (x_{t-1}^i)^2 \sigma_{\theta,t-1}^{i,2} + \sigma_u^2$ and $\hat{x}_t^i = \hat{\theta}_{t-1}^i x_{t-1}^i$, then draw $x_t^i \sim N(x_t|\hat{x}_t^i, \sigma_{x,t}^{i,2})$;*

   a.2) *update $\hat{\theta}_t^i = \hat{\theta}_{t-1}^i + \frac{\sigma_{\theta,t-1}^{i,2} x_{t-1}^i (x_t^i - \hat{x}_t^i)}{\sigma_{x,t}^{i,2}}$ and $\sigma_{\theta,t}^{i,2} = \sigma_{\theta,t-1}^{i,2}\left[1 - \frac{(x_{t-1}^i)^2 \sigma_{\theta,t-1}^{i,2}}{\sigma_{x,t}^{i,2}}\right]$.*

   (b) *Compute normalised importance weights $w_t^i \propto g_t(y_t|x_t^i)$, $i = 1, ..., N$.*

   (c) *Resample the triplets $\{x_{0:t}^i, \hat{\theta}_t^i, \sigma_{\theta,t}^{i,2}\}_{i=1}^N$ with replacement according to the weights $\{w_t^i\}_{i=1}^N$.*

where U_t and V_t are independent, zero-mean white Gaussian sequences with variance σ_u^2 and σ_v^2, respectively, i.e., $U_t \sim N(u_t|0, \sigma_u^2)$ and $V_t \sim N(v_t|0, \sigma_v^2)$. The static parameter θ is unknown; we assume a Gaussian prior pdf with zero mean and variance σ_θ^2, i.e., $p_0(\theta) = N(\theta|0, \sigma_\theta^2)$. The prior pdf for the state is also Gaussian, specifically, $K_0(x_0) = N(x_0|0, \sigma_0^2)$.
For this SSM, the RB-PF is relatively simple. The non-Markov kernel associated to the i-th sequence of particles, $\bar{K}_t(x_t|x_{0:t-1}^i)$, can be written as

$$\bar{K}_t(x_t|x_{0:t-1}^i) = \int K(x_t|x_{t-1}^i, \theta) p_{t-1}^x(\theta|x_{0:t-1}^i) d\theta, \tag{4}$$

where $K(x_t|x_{t-1}^i, \theta) = N(x_t|\theta x_{t-1}^i, \sigma_u^2)$ is linear-Gaussian and $p_{t-1}^x(\theta|x_{0:t-1}^i)$ can easily be shown to be Gaussian as well. Indeed, given the sequence $X_{0:t-1} = x_{0:t-1}^i$ we can write down the linear-Gaussian model

$$p_0(\theta) = N(\theta|0, \sigma_\theta^2), \quad \theta_n = \theta_{n-1}, \quad x_n^i = x_{n-1}^i \theta + v_n,$$

and compute the posterior pdf of θ conditional on $x_{0:n}^i$, $p_n^x(\theta|x_{0:n}^i)$, exactly using a Kalman filter. In particular,

$$p_{t-1}^x(\theta|x_{0:t-1}^i) = N(\theta|\hat{\theta}_{t-1}^i, \sigma_{\theta,t-1}^{i,2}) \tag{5}$$

where $\hat{\theta}_{t-1}^i$ and $\sigma_{\theta,t-1}^{i,2}$ are computed recursively for each particle. Substituting (5) into (4) yields a Gaussian kernel for the r.v. $X_t|x_{0:t-1}^i$, namely

$$\bar{K}_t(x_t|x_{0:t-1}^i) = \int N(x_t|\theta x_{t-1}^i, \sigma_u^2) N(\theta|\hat{\theta}_{t-1}^i, \sigma_{\theta,t-1}^{i,2}) d\theta = N(x_t|\hat{x}_t^i, \hat{\sigma}_{x,t}^{i,2}).$$

The computation of the weights is simple, as the pdf $g(y_t|x_t, \theta) = g(y_t|x_t) = N(y_t|x_t, \sigma_v^2)$ is independent of θ in the SSM of Eq. (3). Algorithm 3 summarises the procedure.
We have applied Algorithm 3 to track the SSM over $2,000$ time steps, with $\sigma_u^2 = \sigma_v^2 = 1$, $N = 1000$ particles and $W = 50$ time steps. From time $t = 0$ to $t = 999$ the parameter is fixed as $\theta = 0.2$ but at time $t = T_c = 1,000$ there is a change point and the parameter switches to $\theta = 0.999$. Figure 1 (middle) shows that the this change point is easily detected as the p-value of the test drops sharply below the significance threshold $\gamma = 10^{-2}$. At that time, the N Kalman filters are re-started, i.e., we set $\hat{\theta}_{T_c}^i = 0$ and $\sigma_{\theta,T_c}^{i,2} = \sigma_{\theta,0}^2$ for every $i = 1, ..., N$. As a result, Figure 1 (left panel) shows that the RB-PF keeps track of the states x_t and it quickly adapts to the change in the parameter value (right panel).

5 Conclusions

We have introduced a simple scheme for automatic detection of change points in a series of observations produced by a dynamical system modelled as an SSM. The scheme is based on a predictive statistic that is

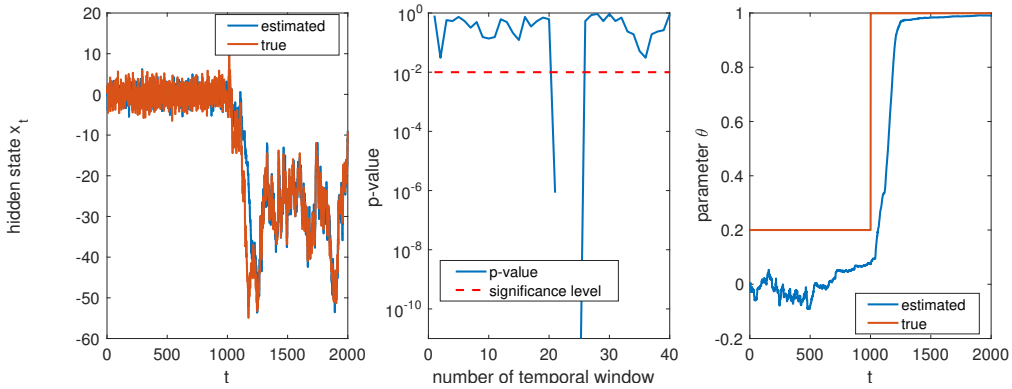

Figure 1: *Left:* Evolution of the true hidden state and the mean of the filtered distribution. *Middle:* p-value of the Pearson's χ^2 test to assess the uniformity of the statistics $A_{K,t}$. *Right:* Evolution of the true parameter θ and the mean of the posterior $p(\theta|y_{1:t})$ approximated by the RB-PF.

invariant w.r.t. the assumed SSM and can be computed naturally using a PF. Modelling errors, or simply mismatches between the observations and the assumed model, can be detected by testing the empirical distribution of the predictive statistics. If the model is coherent with the observations, the predictive statistics are uniform. If they are not, then we detect a change point and start a procedure to re-estimate the SSM. An illustrative example for a Rao-Blackwellised PF has been presented.

References

Anderson, B. D. O., & Moore, J. B. 1979. *Optimal Filtering*. Englewood Cliffs.

Bain, A., & Crisan, D. 2008. *Fundamentals of Stochastic Filtering*. Springer.

Crisan, D., & Miguez, J. 2016. Nested particle filters for online parameter estimation in discrete-time state-space Markov models. *to appear in Bernoulli*, **arXiv:1308.1883v3 [stat.CO]**.

Del Moral, P. 2004. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

Del Moral, P., Jasra, A., & Zhou, Y. 2016. Biased Online Parameter Inference for State-Space Models. *Methodology and Computing in Applied Probability*, 1–23.

Djurić, P. M., & Míguez, J. 2010. Assessment of Nonlinear Dynamic Models by Kolmogorov-Smirnov Statistics. *IEEE Transactions Signal Processing*, **58**(10), 5069–5079.

Doucet, A., Godsill, S., & Andrieu, C. 2000. On sequential Monte Carlo Sampling methods for Bayesian filtering. *Statistics and Computing*, **10**(3), 197–208.

Elvira, V., Miguez, J., & Djurić, P. M. 2017. Adapting the Number of Particles in Sequential Monte Carlo Methods Through an Online Scheme for Convergence Assessment. *IEEE Transactions on Signal Processing*, **65**(7), 1781–1794.

Gordon, N., Salmond, D., & Smith, A. F. M. 1993. Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation. *IEE Proceedings-F*, **140**(2), 107–113.

Míguez, J., Crisan, D., & Djurić, P. M. 2013. On the convergence of two sequential Monte Carlo methods for maximum a posteriori sequence estimation and stochastic global optimization. *Statistics and Computing*, **23**(1), 91–107.