On the performance of particle filters with adaptive number of particles

Víctor Elvira[1,5] · Joaquín Miguez[2,3] · Petar M. Djurić[4]

Abstract

We investigate the performance of a class of particle filters (PFs) that can automatically tune their computational complexity by evaluating online certain predictive statistics which are invariant for a broad class of state-space models. To be specific, we propose a family of block-adaptive PFs based on the methodology of Elvira et al. (IEEE Trans Signal Process 65(7):1781–1794, 2017). In this class of algorithms, the number of Monte Carlo samples (known as *particles*) is adjusted periodically, and we prove that the theoretical error bounds of the PF actually adapt to the updates in the number of particles. The evaluation of the predictive statistics that lies at the core of the methodology is done by generating *fictitious observations*, i.e., particles in the observation space. We study, both analytically and numerically, the impact of the number K of these particles on the performance of the algorithm. In particular, we prove that if the predictive statistics with K fictitious observations converged exactly, then the particle approximation of the filtering distribution would match the first K elements in a series of moments of the true filter. This result can be understood as a converse to some convergence theorems for PFs. From this analysis, we deduce an alternative predictive statistic that can be computed (for some models) without sampling any fictitious observations at all. Finally, we conduct an extensive simulation study that illustrates the theoretical results and provides further insights into the complexity, performance and behavior of the new class of algorithms.

✉ Joaquín Miguez
joaquin.miguez@uc3m.es

Víctor Elvira
victor.elvira@ed.ac.uk

Petar M. Djurić
petar.djuric@stonybrook.edu

[1] School of Mathematics, University of Edinburgh, Edinburgh, UK

[2] Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Madrid, Spain

[3] IIS Gregorio Marañón, Madrid, Spain

[4] Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, USA

[5] The Alan Turing Institute, London, UK

1 Introduction

In science and engineering, there are many problems that are studied by way of dynamic probabilistic models. Some of these models describe mathematically the evolution of hidden states and their relations with observations, which are sequentially acquired. In many of these problems, the objective is to estimate sequentially the posterior probability distribution of the hidden model states. A methodology that has gained considerable popularity in the last two and a half decades is particle filtering (also known as sequential Monte Carlo) (Gordon et al. 1993; Liu et al. 1998; Doucet et al. 2001; Djurić et al. 2003; Künsch 2013). This is a Monte Carlo methodology that approximates the distributions of interest by means of random (weighted) samples.

Arguably, a key parameter of particle filters (PFs) is the number of generated Monte Carlos samples (usually termed *particles*). A larger number of particles improves the approximation of the filter but also increases the computational complexity. However, it is impossible to know a priori the appropriate number of particles to achieve a prescribed accu-

racy in the estimated parameters and distributions. So a question of great practical interest is how to determine the necessary number of particles to achieve a prescribed performance and, in particular, how to determine it automatically and in real time.

1.1 Particle filtering with time-varying number of particles

Until the publication of Elvira et al. (2017), not many papers had considered the selection/adaptation of the number of particles in a systematic and rigorous manner. In Elvira et al. (2017), a methodology was introduced to address this problem with the goal of adapting the number of particles in real time. The method is based on a rigorous mathematical analysis and we discuss it in more detail in Sect. 1.2.

Other efforts toward the same goal include the use of a Kullback–Leibler divergence-based approximation error by Fox (2003), where the divergence was defined between the distribution of the PF and a discrete approximation of the true distribution computed on a predefined grid. The ideas in Fox (2003) were further explored by Soto (2005). A heuristic approach based on the effective sample size was proposed by Straka and Šimandl (2006), while Cornebise (2009, Chapter 4) pursued similar ideas with a scheme that regenerated particles until a certain performance criterion was met. A disadvantage of using the effective sample size is that once a PF loses track of the hidden state, the effective sample size does not provide information for adjusting the number of particles. See other issues related to the effective sample size in Elvira et al. (2018).

A method for selecting the number of particles based on the particle approximation of the variance of the particle estimators was reported by Lee and Whiteley (2018), where the Feynman–Kac framework of Del Moral (2004) was used for the analysis. While well principled, this technique cannot be implemented online. Bhadra and Ionides (2016) suggested an autoregressive model for the variance of the estimators produced by the PF was employed, but the resulting method works offline as well. In a group of papers on alive PFs, the number of particles is adaptive and based on sampling schemes that ensure a predefined number of particles to have non-zero weights (LeGland and Oudjane 2005; Jasra et al. 2013; Moral et al. 2015). In Martino et al. (2017), a fixed number of particles is adaptively allocated to several candidate models according to their performances. In Hu et al. (2008), particle sets of the same size are generated until an estimation criterion for their acceptance is met.

1.2 Some background

In Elvira et al. (2017), we introduced a methodology for assessing the convergence of PFs that works online and

can be applied to a very broad class of state-space models and versions of the PF. The method is based on simulating fictitious observations from one-step-ahead predictive distributions approximated by the PF and comparing them with actual observations that are available at each time step. In the case of one-dimensional observations, a statistic is constructed that simply represents the number of fictitious observations which are smaller than the actual observation. It is proved in Elvira et al. (2017) that, as the PF converges, the predictive statistics become uniform on a discrete support and independent over time. From that realization, we proposed an algorithm for statistically testing the uniformity of the predictive statistic and, based on the test, update the number of particles in the PF.

The same type of predictive statistic has been used as a tool for evaluating ensemble forecasts in weather prediction, under the name of rank statistic—see, e.g., Anderson (1996), Hamill (2001) and Bröcker (2018). Recently, Talts et al. (2018) have also used the same statistic to validate both models and inferential methods in a Bayesian framework. The scheme in the latter paper is similar to the methodology we proposed in Elvira et al. (2017) for Bayesian inference in state-space models.

1.3 Contributions

In this paper, we propose a general block-adaptive PF where the number of particles is updated periodically, every W discrete time steps. It is a rather general scheme that provides a common framework for the procedures described in Elvira et al. (2017) and enables us to introduce different versions of the algorithm and to extend the analysis of the methodology.

In particular, we first tackle the problem of whether the updates in the number of particles carried out at the end of each block of length W translate into changes to the theoretical error bounds for the Monte Carlo estimators. While this is the kind of performance that one would *like* to have (e.g., we want to see smaller errors when we increase the number of particles), what the standard arguments for proving the convergence of the PF[1] yield directly are error bounds that depend on the minimum of the number of particles over time. Here, we use the approach of Del Moral (2004) to prove that, assuming that the state sequence is Markov and mixing, the approximation errors at the end of each block are bounded by the sum of two terms: one that depends on the number of particles in the current block and another one that decreases exponentially with the block length W.

Next, we turn our attention to the analysis of the impact of the number of fictitious observations, K, used by the algo-

---

[1] Either by induction as in Crisan (2001), Bain and Crisan (2008) or using the contraction properties of Markov kernels as in Del Moral (2004), Künsch (2005).

rithm to compute the predictive statistics. We first prove that if the predictive statistics with K fictitious observations are uniformly distributed, then the particle approximation of the filtering distribution match at least the first K elements in a series of moments that characterize the true filter completely. Let us remark that this result is (close to) a converse to Theorem 2 in Elvira et al. (2017): the latter says that when the PF converges, the predictive statistics become uniform, while the new result says that if the predictive statistics with K fictitious observations become uniform, then the PF necessarily converges to match at least K moments of the true filter. From this analysis, we deduce an alternative predictive statistic that can be computed (for some models) without sampling any fictitious observations at all and establish its connection with the original one.

Finally, we conduct an extensive simulation study that illustrates the theoretical results and provides further insights into the complexity, performance and behavior of the new class of algorithms. We show, for example, that choosing larger values of K leads to more accurate filters with a higher computational cost, while a smaller K yields a faster filter using less particles (but yielding rougher errors). We also illustrate how the approximation errors change with the number of particles (as predicted by the theory) or how the adaptive PF stabilizes around the same number of particles no matter the initial condition (i.e., whether started with many or few particles). Our simulations also show that the new predictive statistic (without fictitious observations) is effective but sensitive to prediction errors and hence it leads to higher computational loads.

1.4 Organization of the paper

In the next section, we briefly describe particle filtering as a sequential Monte Carlo methodology, then we introduce our notation and a general block-adaptive PF that updates the number of particles periodically and admits different implementations. In Sect. 3, we present our convergence analysis of block-adaptive PFs. In Sect. 4, we provide a detailed analysis of the number of generated fictitious particles and introduce a new predictive statistic that does not require generation of fictitious particles. In the last two sections, we present results of numerical experiments and our conclusions, respectively.

2 Background

2.1 State-space models and particle filtering

We investigate Markov state-space models described by the triplet of probability distributions

$$X_0 \sim p(x_0), \tag{2.1}$$

$$X_t \sim p(x_t|x_{t-1}), \tag{2.2}$$

$$Y_t \sim p(y_t|x_t), \tag{2.3}$$

where

- $t \in \mathbb{N}$ denotes discrete time;
- X_t is the system state at time t, i.e., a $d_x \times 1$-dimensional random vector taking values in a state space $\mathcal{X} \subseteq \mathbb{R}^{d_x}$,
- $p(x_0)$ is the a priori probability density function (pdf) of the state,
- $p(x_t|x_{t-1})$ is the conditional density of X_t given $X_{t-1} = x_{t-1}$;
- Y_t is a $d_y \times 1$-dimensional observation vector at time t, where $Y_t \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$ and is assumed conditionally independent of all the other observations given X_t,
- $p(y_t|x_t)$ is the conditional pdf of Y_t given $X_t = x_t$. It is often referred to as the *likelihood* of x_t, when it is viewed as a function of x_t for some fixed y_t.

Based on the model (2.1)–(2.3), we aim at estimating the sequence of posterior probability distributions $p(x_t|y_{1:t})$, $t = 1, 2, \ldots$, recursively. Many schemes addressing this task rely on the decomposition

$$p(x_t|y_{1:t}) \propto p(y_t|x_t) \int p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}) dx_{t-1},$$

that relates the so-called filtering pdf at time t, $p(x_t|y_{1:t})$, to the filtering density at time $t-1$, $p(x_{t-1}|y_{1:t-1})$.

Let us denote the filtering and the predictive posterior probability measures as

$$\pi_t(dx_t) := p(x_t|y_{1:t}) dx_t,$$
$$\xi_t(dx_t) := p(x_t|y_{1:t-1}) dx_t. \tag{2.4}$$

The measure π_t does not provide any further characterization of the probability distribution compared to the density $p(x_t|y_{1:t})$, however, Monte Carlo methods (including PFs) yield an approximation of π_t, rather than the pdf $p(x_t|y_{1:t})$. Another function that plays a central role in the methods investigated in this paper is the predictive pdf of the observations, $p(y_t|y_{1:t-1})$. We denote the associated probability measure as

$$\mu_t(dy_t) := p(y_t|y_{1:t-1}) dy_t.$$

It is well known that the predictive pdf is instrumental for model inference (Andrieu et al. 2010; Djurić and Míguez 2010; Chopin et al. 2012; Crisan and Miguez 2018).

The goal of particle filtering algorithms is to estimate sequentially the probability measures $\{\pi_t\}_{t \geq 1}$ as the observations $\{y_t\}_{t \geq 1}$ are collected. The basic method for accom-

plishing this is known as the bootstrap filter (BF) introduced by Gordon et al. (1993) (see also Doucet et al. 2000).

At time $t = 0$, the algorithm applies standard Monte Carlo to approximate the prior probability distribution, i.e., we generate M i.i.d. samples $\boldsymbol{x}_0^{(m)}$, $m = 1, \ldots, M$, from the pdf $p(\boldsymbol{x}_0)$. The samples $\boldsymbol{x}_0^{(m)}$ are often termed *particles*. Assume that the particles can be propagated over time, in such a way that we obtain a Monte Carlo approximation of the filtering distribution at time $t-1$ given by the particle set $\{\boldsymbol{x}_{t-1}^{(m)}\}_{m=1}^M$. At time t, the BF generates an estimate of π_t recursively, by taking three steps:

1. Draw new particles $\bar{\boldsymbol{x}}_t^{(m)}$, $m = 1, \ldots, M$, from the conditional pdf's, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(m)})$. Note at this step we essentially simulate the model dynamics by propagating the particles one step forward.
2. Compute normalized importance weights of the particles, denoted

$$w_t^{(m)} \propto p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{(m)}), \quad m = 1, \ldots, M.$$

   These weights are proportional to the likelihood and they satisfy $\sum_{m=1}^M w_t^{(m)} = 1$.
3. Resample the particles M times with replacement using the weights $\{w_t^{(m)}\}_{m=1}^M$ as probabilities (Li et al. 2015). This yields the new (and non-weighted) particle set $\{\boldsymbol{x}_t^{(m)}\}_{m=1}^M$.

From the particles and their weights, one can compute estimates of several probability measures and pdfs. The filtering measure π_t can be approximated as

$$\pi_t^M(d\boldsymbol{x}) = \sum_{m=1}^M w_t^{(m)} \delta_{\bar{\boldsymbol{x}}_t^{(m)}}(d\boldsymbol{x}),$$

where $\delta_{\bar{\boldsymbol{x}}_t^{(m)}}$ represents the Dirac delta measure located at $\bar{\boldsymbol{x}}_t^{(m)} \in \mathcal{X}$. Moreover, at time t, once $Y_{1:t-1} = y_{1:t-1}$ are available but $Y_t = y_t$ has not been observed yet, the predictive pdfs of X_t, denoted $\tilde{p}_t(\boldsymbol{x}_t) := p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$, and Y_t, denoted $p_t(\boldsymbol{y}_t) := p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})$, can be approximated as

$$\tilde{p}_t^M(\boldsymbol{x}_t) := \frac{1}{M} \sum_{m=1}^M p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(m)}), \quad \boldsymbol{x}_t \in \mathcal{X}, \quad \text{and} \quad (2.5)$$

$$p_t^M(\boldsymbol{y}_t) := \frac{1}{M} \sum_{m=1}^M p(\boldsymbol{y}_t|\bar{\boldsymbol{x}}_t^{(m)}), \quad \boldsymbol{y}_t \in \mathcal{Y}. \quad (2.6)$$

A key parameter in the standard BF is the number of particles M, which determines both the computational cost of the algorithm and also the accuracy of any estimators computed using the particles and weights (Del Moral, 2004; Bain

and Crisan, 2008). While M is fixed in conventional particle filtering methods, the focus of this paper is on algorithms where M can be updated sequentially (Elvira et al., 2017).

2.2 Block-adaptive selection of the number of particles

A generic block-adaptive method for selecting the number of particles is summarized in Table 1. Hereafter, we assume that the observations are one-dimensional (and hence we denote them as y_t instead of $\boldsymbol{y}_t$) unless explicitly indicated. The methods to be described can be adapted to systems with multidimensional observations in a number of ways—see Elvira et al. (2017, Section IV-E) for a discussion on this topic. Also note that we implement the algorithm based on the standard BF, but it is straightforward to extend the methodology to other PFs.

The block-adaptive method proceeds as follows. In Step 1(a) of Table 1, the filter is initialized with M_0 particles. The particle filter works at each time step in a standard manner with the current number of particles, as described in Step 2(a). The first modification with respect to (w.r.t.) the BF comes in Step 2(b), where K fictitious observations $\{\tilde{y}_t^{(k)}\}_{k=1}^K$ are simulated from the (approximate) predictive distribution of the observations $p_t^M(y_t)$ (see Elvira et al. 2017, Section IV-A for additional details). Assume that the fictitious observations are ordered, i.e., $\tilde{y}_t^{(1)} < \tilde{y}_t^{(2)} < \cdots < \tilde{y}_t^{(K)}$, and let y_t be the actual observation at time t. The statistic $A_{K,M_n,t}$ is a r.v. constructed as

$$A_{K,M_n,t} := \begin{cases} 0, & \text{if } y_t < \tilde{y}_t^{(1)}, \\ j, & \text{if } \tilde{y}_t^{(j)} < y_t < \tilde{y}_t^{(j+1)}, \text{ with } 1 \le j < K, \\ K, & \text{if } \tilde{y}_t^{(K)} < y_t. \end{cases}$$

Therefore, $A_{K,M_n,t}$ yields a non-negative integer (from 0 to K) that indicates how many fictitious observations are smaller than the actual observation y_t. We use $A_{K,M_n,t}$ to denote the r.v., while $a_{K,M_n,t}$ indicates a specific realization of it. The algorithm works with windows of varying size W_n. At the end of the nth window (Step 2(c)), the sequence

$$\begin{aligned} \mathcal{S}_n = \{ &a_{K,M_n,t-W_n+1}, a_{K,M_n,t-W_n+2}, \cdots \\ &\cdots, a_{K,M_n,t-2}, a_{K,M_n,t-1}, a_{K,M_n,t} \} \end{aligned} \quad (2.7)$$

is collected and processed for assessing the convergence of the filter. The number of particles is adapted (increased, decreased, or kept constant) based on the assessment.

When we assume that

– the fictitious observations $\{\tilde{y}_t^{(k)}\}_{k=1}^K$ are independently drawn from the same pdf as the actual observation y_t, i.e., the predictive pdf $p(y_t|y_{1:t-1})$, and

Table 1 General BF with block-adaptive number of particles, M_n

---

1. **[Initialization]**

   (a) Draw independent samples $\boldsymbol{x}_0^{(m)}$ from the prior $p(\boldsymbol{x}_0)$ and assign uniform weights, i.e.,

   $\boldsymbol{x}_0^{(m)} \sim p(\boldsymbol{x}_0), \quad m = 1, \ldots, M_0, \quad$ and

   $w_0^{(m)} = \frac{1}{M_0}, \quad m = 1, \ldots, M_0.$

   (b) Set $n = 0$ (block counter) and choose $W_0 > 0$ (size of the first block).

2. **[For $t = 1, 2, \ldots$]**

   (a) **Bootstrap particle filter:**

   – Sample $\bar{\boldsymbol{x}}_t^{(m)} \sim p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^{(m)}), m = 1, \ldots, M_n.$

   – Compute normalized weights, $\bar{w}_t^{(m)} \propto p(y_t | \bar{\boldsymbol{x}}_t^{(m)}), m = 1, \ldots, M_n.$

   (b) **Fictitious observations:**

   – Draw $\tilde{y}_t^{(k)} \sim p_t^M(y_t), \quad k = 1, \ldots, K.$

   – Compute $a_{K,M_n,t}$, i.e., the position of the actual observation y_t within the set of ordered fictitious observations $\{\tilde{y}_t^{(k)}\}_{k=1}^{K}$.

   (c)**Assessment of convergence:** If $t = \sum_{j=0}^{n} W_j - 1$ (end of the nth block) then:

   – Analyze the subsequence

   $\mathcal{S}_n = \{a_{K,M_n,t}, a_{K,M_n,t-1}, \ldots, a_{K,M_n,t-W_n+1}\}$ with some specific algorithm from Sect. 2.3.

   – Set $n = n + 1.$

   – Select the number of particles $M_n > 0.$

   – Select the block size $W_n > 0.$

   – Resample M_n particles with replacement, from the weighted set $\{\bar{\boldsymbol{x}}_t^{(m)}, w_t^{(m)}\}_{m=1}^{M_{n-1}}$, to obtain $\{\boldsymbol{x}_t^{(m)}\}_{m=1}^{M_n}.$

   Else:

   – Resample M_n particles with replacement, from the weighted set $\{\bar{\boldsymbol{x}}_t^{(m)}, w_t^{(m)}\}_{m=1}^{M_n}$, to obtain $\{\boldsymbol{x}_t^{(m)}\}_{m=1}^{M_n}.$

– the statistic $A_{K,M,t}$ becomes independent of M, i.e., $A_{K,M,t} = A_{K,t}$ is exact as well,

it is relatively straightforward to prove the two propositions below (Elvira et al. 2017).

Proposition 1 *If $y_t, \tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(K)}$ are i.i.d. samples from a common continuous (but otherwise arbitrary) probability distribution, then the pmf of the random variable (r.v.) $A_{K,t}$ is*

$$\mathbb{Q}_K(n) = \frac{1}{K+1}, \quad n = 0, \ldots, K. \quad (2.8)$$

Proposition 2 *If the r.v.'s $y_t, \tilde{y}_t^{(1)}, \ldots, \tilde{y}_t^{(K)}$ are i.i.d. with common pdf $p_t(y)$, then the r.v.'s in the sequence $\{A_{K,t}\}_{t\geq 1}$ are independent.*

In practical terms, Propositions 1 and 2 suggest that when the approximation errors in the PF are small, i.e., $p_t^M(dy_t) \approx p_t(dy_t)$, we can expect the statistics in the sequence $\mathcal{S}_n$ of Eq. (2.7) to be (nearly) independent and uniformly distributed. Therefore, testing whether the variates

$$a_{K,M_n,t-W_n+1}, a_{K,M_n,t-W_n+2}, \ldots, a_{K,M_n,t-1}, a_{K,M_n,t},$$

are independent and/or uniform is an indirect manner of assessing the convergence of the PF. The key advantage of

this approach is that Propositions 1 and 2 do not depend on the specific choice of the transition density $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$ and the likelihood $p(y_t | \boldsymbol{x}_t)$, and therefore the statistics can be applied to a very general class of state-space models. A detailed analysis of the approximation errors in the statistic $A_{K,M,t}$ and its pmf $\mathbb{Q}_{K,M,t}(n)$ is provided in Elvira et al. (2017). In particular, it is proved that $\lim_{M \to \infty} \mathbb{Q}_{K,M,t} = \mathbb{Q}_K(n)$ almost surely (a.s.) for every t.

2.3 Summary of algorithms for adapting the number of particles

The general block-adaptive framework of Table 1 allows for developing different algorithms. Here we outline two specific procedures that differ in the implementation of step 2(c) of Table 1—specifically, in the analysis of the subsequence of statistics denoted as $\mathcal{S}_n = a_{K,M_n,t-W_n+1:t}$. The first scheme tests whether the samples in $\mathcal{S}_n$ have a uniform distribution, while the second scheme assesses the correlation among the elements of $\mathcal{S}_n$.

2.3.1 Algorithm 1: Uniformity of $A_{K,M,t}$

This is the scheme originally proposed in Elvira et al. (2017), and it exploits Proposition 1. Under the null hypothesis of perfect convergence of the PF (i.e., $p_t^M(dy_t) = p_t(dy_t)$), the

r.v.'s $A_{K,M,t}$ are statistically uniform. Therefore, the algorithm tests if the variates in the subset $\mathcal{S}_n$ are i.i.d. uniform draws from the set $\{0, \ldots, K\}$. In Elvira et al. (2017), this is done by performing the Pearson's χ^2 test on the set $\mathcal{S}_n$ of Eq. (2.7).

To be specific, let $M_{\min}$ and $M_{\max}$ be the minimum and maximum number of particles, respectively, to be admitted by the PF. Further, let there be two threshold values $0 < p_\ell < p_h < 1$. At the end of the nth block, i.e., at time $t = \sum_{j=0}^{n} W_j - 1$, the Pearson's χ^2 test is run on $\mathcal{S}_n$ and it outputs a p value which we denote as $p^*_{K,n}$. The p value is compared against the thresholds p_ℓ and p_h. In particular,

- if $p^*_{K,n} \leq p_\ell$, then M_n is increased w.r.t. M_{n-1} by setting $M_n = \min\{cM_{n-1}, M_{\max}\}$, where $c > 1$ is a constant set by the user;
- if $p^*_{K,n} \geq p_h$, then M_n is decreased w.r.t. M_{n-1} by setting $M_n = \max\{c^{-1}M_{n-1}, M_{\min}\}$;
- otherwise, the algorithm sets $M_n = M_{n-1}$.

Intuitively, when the Pearson test yields a small p value this is interpreted as "poor performance" of the PF and the number of particles is increased. If the p value is large, this is interpreted as "good performance" and the computational effort is relaxed (by decreasing M_n). If the p value is interpreted as "average" (i.e., between the thresholds) then the computational effort is neither increased nor decreased. In the numerical examples in Elvira et al. (2017), the number of particles is increased or decreased by a factor of $c = 2$, while keeping the number M_n always between $M_{\min}$ and $M_{\max}$.

2.3.2 Algorithm 2: Correlation of $A_{K,M,t}$

Under the same null hypothesis of perfect convergence of the PF, the variates in $\mathcal{S}_n$ are i.i.d. (Proposition 2). Since independence implies absence of correlation, we can test if the samples of $\mathcal{S}_t$ are correlated, e.g., using the scheme in Elvira et al. (2016). Note that in estimating the autocorrelation of $A_{K,M,t}$, longer windows (larger values of W_n) may be needed to improve accuracy. However, larger block-sizes imply a loss of responsiveness in the adaptation of M_n.

The algorithm proposed in Elvira et al. (2016) follows the same general scheme described in the previous Sect. 2.3.1. At the end of the nth block (time $t = \sum_{j=0}^{n} W_j - 1$), a Pearson's correlation coefficient r_n is computed with the statistics in the set $\mathcal{S}_n$, and a statistical test using the Student's t-distribution is performed in order to obtain the p value $p^*_{K,n}$. This p value is used in the same way as in Sect. 2.3.1 in order to increase, decrease or maintain the number of particles M_n.

3 Error bounds for block-adaptive particle filters

We present an analysis of the class of block-adaptive filters outlined in Table 1, with either fixed ($W_n = W$ for all n) or adaptive (W_n updated together with M_n) block size from a viewpoint that was ignored in Elvira et al. (2017). To be specific, we prove that at the end of the nth window,[2] the error bounds for the estimators that are computed using the weighted particle set $\{w_t^{(m)}, \bar{x}_{t_n}^{(m)}\}_{m=1}^{M_n}$ can be written as a function of the *current* number of particles M_n—provided that the optimal filter π_t satisfies a stability condition (Del Moral, 2004). If one were to rely directly on classical convergence results for algorithms with fixed $M_n = M$ (see, e.g., Del Moral and Guionnet 2001; Crisan and Doucet 2002; Del Moral 2004; Künsch 2005; Míguez et al. 2013), the error bound at time t_n would be characterized as a function of the minimum of the number of particles employed up to that time, namely

$$M_n^{\min} := \min_{0 \leq j \leq n} M_j.$$

The current number of particles, M_n, can be considerably larger than $M_n^{\min}$ and, as a consequence, the error bound can be remarkably smaller.

3.1 Notation

For notational clarity and conciseness, let

$$\kappa_t(\mathrm{d}x_t | x_{t-1}) := p(x_t | x_{t-1})\mathrm{d}x_t$$

denote the Markov kernel that governs the dynamics of the state sequence $\{x_t\}_{t>0}$ and write

$$g_t^{y_t}(x_t) := p(y_t | x_t)$$

to indicate the conditional pdf of the observations.

We analyze the algorithm outlined in Table 1, which is essentially a BF with M_n particles in the nth time window, $\sum_{j=0}^{n-1} W_j \leq t < \sum_{j=1}^{n} W_j$, where W_j is the length of the jth window. As pointed out, the theoretical results we introduce are valid both for variable window lengths as well as for fixed $W_n = W$. Our analysis also holds independently of the update rule for M_n, as long as only positive values are permitted. Specifically, we assume that there is a positive lower bound $\underline{M}$ such that $M_n \geq \underline{M}$ for every $n \geq 0$. In practice, we usually have a finite upper bound $\overline{M} \geq M_n$ as well (but this plays no role in the analysis).

---

[2] Specifically, at time $t_n = \sum_{j=0}^{n} W_j - 1$.

For any integrable real function $f : \mathcal{X} \to \mathbb{R}$ and a probability measure α, we use the shorthand notation

$$(f, \alpha) = \int f(\boldsymbol{x})\alpha(\mathrm{d}\boldsymbol{x})$$

for integrals with respect to α. If α has a pdf $a(\boldsymbol{x})$, we also denote $(f, a) := (f, \alpha)$ when convenient. Intuitively, we aim at proving that the bounds for the approximation errors $|(f, \pi_t^{M_n}) - (f, \pi_t)|$, where

$$(f, \pi_t^{M_n}) = \int f(\boldsymbol{x})\pi_t^{M_n}(\mathrm{d}\boldsymbol{x}) = \sum_{m=1}^{M_n} w_t^{(m)} f(\bar{\boldsymbol{x}}_t^{(m)}),$$

effectively change when the number of particles M_n is updated. Since the measures $\pi_t^{M_n}$ are random, the approximation errors $(f, \pi_t^{M_n}) - (f, \pi_t)$ are real r.v.'s, and we can assess their L_p norms. We recall that for a real r.v. Z with probability measure α, the L_p norm of Z, with $p \geq 1$, is

$$\|Z\|_p := \left(\mathbb{E}\left[|Z|^p\right]\right)^{\frac{1}{p}} = \left(\int |z|^p \alpha(\mathrm{d}z)\right)^{\frac{1}{p}},$$

where $\mathbb{E}[\cdot]$ denotes the expected value w.r.t. the distribution of the r.v.

3.2 Error bounds

We show hereafter that by the end of the nth block of observations, the approximation error

$$\|(f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n})\|_p, \quad \text{where } t_n = \sum_{j=0}^{n} W_j - 1,$$

and f is a bounded real function, can be upper bounded by a function that depends on the current number of particles M_n and "forgets" past errors exponentially fast. This is true under certain regularity assumptions that we detail below.

Let us introduce the prediction-update (PU) operators Ψ_t that generate the sequence of filtering probability measures π_t given a prior measure π_0, the sequence of kernels κ_t and the likelihoods $g_t^{y_t}$.

Definition 1 Let $\mathcal{B}(\mathcal{X})$ denote the Borel σ-algebra of subsets of $\mathcal{X}$ and let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on the space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. We construct the sequence of PU operators $\Psi_t : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X}), t \geq 1$, that satisfy

$$(f, \Psi_t(\alpha)) = \frac{\left(f g_t^{y_t}, \kappa_t \alpha\right)}{(g_t, \kappa_t \alpha)}, \quad t = 1, 2, \ldots, \tag{3.1}$$

for any $\alpha \in \mathcal{P}(\mathcal{X})$ and any integrable real function f, where $\kappa_t \alpha(\mathrm{d}\boldsymbol{x}_t) = \int \kappa_t(\mathrm{d}\boldsymbol{x}_t|\boldsymbol{x}')\alpha(\mathrm{d}\boldsymbol{x}')$ is the result of applying the Markov kernel κ_t to the probability measure α.[3]

It is not hard to see that Definition 1 implies that $\pi_t = \Psi_t(\pi_{t-1})$. In order to represent the evolution of the sequence of filtering measures over several time steps, we introduce the composition of operators

$$\Psi_{t|t-r}(\alpha) := \left(\Psi_t \circ \Psi_{t-1} \circ \cdots \circ \Psi_{t-r+1}\right)(\alpha). \tag{3.2}$$

It is apparent that $\pi_t = \Psi_{t|t-r}(\pi_{t-r})$. The composition operator $\Psi_{t|t-r}$ is most useful for representing the filters obtained after r consecutive steps when we start from different probability measures at time $t - r$, i.e., for comparing $\Psi_{t|t-r}(\alpha)$ and $\Psi_{t|t-r}(\beta)$ for $\alpha, \beta \in \mathcal{P}(\mathcal{X})$.

In our analysis, we assume that the kernels $\kappa_t(\mathrm{d}\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ satisfy a *mixing* assumption (Del Moral 2004; Künsch 2005). While this can be stated in various ways, we follow the approach in Del Moral (2004), which relies on the composition of kernels to mix sufficiently over several time steps.

Assumption 1 (*Mixing kernel*) Let us write

$$\kappa_{t|t-m} = \kappa_t \circ \kappa_{t-1} \circ \cdots \circ \kappa_{t-m+1} \tag{3.3}$$

for the composition of m consecutive Markov kernels. For every $S \in \mathcal{B}(\mathcal{X})$ and integer $m \geq 1$, there exists a constant $\varepsilon_m > 0$, independent of t and S, such that

$$\inf_{\boldsymbol{x}_{t-m}, \boldsymbol{x}'_{t-m} \in \mathcal{X}} \frac{\kappa_{t|t-m}(S|\boldsymbol{x}_{t-m})}{\kappa_{t|t-m}(S|\boldsymbol{x}'_{t-m})} > \varepsilon_m.$$

Assumption 1 implies that the sequence of optimal filters generated by the operators $\Psi_t, t \geq 1$, is stable (Del Moral and Guionnet, 2001). To be specific, it can be proved (Del Moral and Guionnet, 2001; Del Moral, 2004) that

$$\lim_{r \to \infty} \sup_{\alpha, \beta \in \mathcal{P}(\mathcal{X})} \left|\left(f, \Psi_{t|t-r}(\alpha)\right) - \left(f, \Psi_{t|t-r}(\beta)\right)\right| = 0$$

exponentially fast. The intuitive meaning is that such sequences "forget" their initial condition over time. It also implies that approximation errors are also forgotten over time when propagated through the operators Ψ_t, a fact that is often exploited in the analysis of PFs.

The strongest assumption in our analysis is that the sequence of likelihoods is uniformly bounded away from zero (as well as upper bounded), as specified below.

Assumption 2 (*Bounds*) There exists a constant $\gamma > 0$ such that

$$0 < \gamma < g_t^{y_t}(\boldsymbol{x}) \leq 1 \tag{3.4}$$

---

[3] Note that if $\alpha = \pi_{t-1}$, then $\kappa_t \pi_{t-1}(\mathrm{d}\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\mathrm{d}\boldsymbol{x}_t$.

for every $t \geq 1$ and every $\boldsymbol{x} \in \mathcal{X}$.

Assumption 2 depends not only on the form of the likelihood $g_t^{y_t}(\boldsymbol{x}) = p(\boldsymbol{y}_t | \boldsymbol{x}_t)$ but also on the specific sequence of observations $y_1, y_2, \ldots$ While it may appear restrictive, this is rather typical in the analysis of PFs (see Del Moral 2004; Künsch 2005; Gupta et al. 2015; Crisan and Miguez 2017) and is expected to hold naturally when the state space $\mathcal{X}$ is compact (as well as in other typical scenarios[4]). Also note that any bounded likelihood function can be normalized to guarantee $g_t^{y_t} \leq 1$.

The error bounds for estimator $(f, \pi_{t_n}^{M_n})$ are made precise by the following statement.

Theorem 1 *Let $t_n = \sum_{j=0}^{n} W_j - 1$ and let $\pi_{t_n}^{M_n}$ be the particle approximation of the filtering measure π_{t_n} produced by the block-adaptive BF in Table 1. If Assumption 1 (mixing kernel) and Assumption 2 (bounds) hold, then for any $p \geq 1$,*

$$\sup_{|f| \leq 1} \left\| (f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n}) \right\|_p < \frac{mC}{\gamma^{2m-1} \varepsilon_m^3 \sqrt{M_n}}$$
$$+ \frac{2 \left(1 - \gamma^{m-1} \varepsilon_m^2 \right)^{\lfloor \frac{W_n}{m} \rfloor}}{\gamma^m \varepsilon_m} \sup_{|f| \leq 1} \left\| (f, \pi_{t_{n-1}}^{M_{n-1}}) - (f, \pi_{t_{n-1}}) \right\|_p,$$

where $\sup_{|f| < 1}$ denotes the supremum over all real functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_\infty \leq 1$. The real constants $C < \infty$, $\gamma > 0$ and $\varepsilon_m > 0$, as well as the integer $m \geq 1$, are independent of n, M_n and W_n.

See Appendix A for a proof.

Theorem 1 states that the error bound at the end of the current (nth) block depends on the error at the end of the previous (($n-1$)th) block plus an additional term that depends on the number of particles, M_n, used within the current block. Moreover, the block size W_n can be chosen in such a way that the "inherited error" due to, e.g., a lower number of particles M_{n-1} in the ($n-1$)th block can be forgotten when a sufficiently large window length W_n is selected in the nth block.

---

[4] For example, suppose that the observations are collected by sensors with limited sensitivity. To see this, consider a sensor located at $\boldsymbol{s}$ that measures the power transmitted by an object located at $\boldsymbol{x}$. Assuming free space, the received power (in dB's) can be modeled as $y = 10 \log_{10} \left(P_0 \|\boldsymbol{s} - \boldsymbol{x}\|^{-2} + \eta \right) + z$, where $z \sim N(0, \sigma^2)$ is Gaussian noise, P_0 is the transmitted power, and the parameter $\eta > 0$ determines the sensitivity of the sensor. The likelihood function is

$$g^y(\boldsymbol{x}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left(y - 10 \log_{10} \left(P_0 \|\boldsymbol{s} - \boldsymbol{x}\|^{-2} + \eta \right) \right)^2 \right\}.$$

As a consequence, when $\|\boldsymbol{s} - \boldsymbol{x}\| \rightarrow \infty$ the sensor observes $y = 10 \log_{10}(\eta) + z$ independently of the target position $\boldsymbol{x}$ and, in particular, for fixed $\boldsymbol{s}$ we have $\lim_{\|\boldsymbol{x}\| \rightarrow \infty} g^y(\boldsymbol{x}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left(y - 10 \log_{10} \eta \right)^2 \right\} > 0$. Intuitively, the sensor cannot "see" targets which are too far away.

This is due to the stability property of the PU operator Ψ_t (which is guaranteed under Assumption 1). In particular,

$$\lim_{W_n \rightarrow \infty} \sup_{|f| \leq 1} \left\| (f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n}) \right\|_p < \frac{mC}{\gamma^m \varepsilon_m^2 \sqrt{M_n}}.$$

4 Analysis of the number of fictitious observations, K

The performance analysis of Elvira et al. (2017) establishes the main results needed for a principled, online adaptation of the number of particles M, but leaves a number of questions unanswered. One of them, whether the error bounds of the particle estimators change as we update the number of particles online, has been addressed in Sect. 3. Two other major questions are

(i) whether the statistics $A_{K, M_n, t}$ becoming uniform is *sufficient* for the particle approximations $p_t^{M_n}$ and $\pi_t^{M_n}$ to converge toward p_t and π_t, respectively (the analysis in Elvira et al. (2017) only shows that this is *necessary*), and

(ii) how the choice of the number of fictitious observations K affects the performance of the adaptive algorithm (i.e., the approximation error of either $p_t^{M_n}$ or $\pi_t^{M_n}$).

We tackle these two issues in this section. To be precise, we prove that if the statistics $A_{K, M_n, t}$ are uniform r.v.'s for every $K \in \mathbb{N}$, then the approximate predictive pdf $p_t^{M_n}(y)$ becomes equal to the actual density $p_t(y)$ almost everywhere. This result serves as a converse for Theorem 2 in Elvira et al. (2017)—which states that $p_t^{M_n}(y) \overset{M_n \rightarrow \infty}{\Longrightarrow} p_t(y)$ implies that the statistics $A_{K, M_n, t}$ become uniform r.v.'s. Intuitively, the new result ensures that if the $A_{K, M, t}$'s are well-behaved then so are the particle approximations $p_t^{M_n}$. Our analysis also provides insight into the choice of $K < \infty$. Specifically, it yields a quantitative interpretation of how $p_t^{M_n}$ becomes closer to p_t when $A_{K, M_n, t}$ is uniform for larger and larger K.

As a by-product of this analysis, we identify an alternative statistic B_t (and its particle estimator $B_{M_n, t}$) that can be used for assessing the performance of the particle filter without generating fictitious observations. This alternative statistic admits an interpretation as the limit of the sequence $K^{-1} A_{K, M_n, t}$ when $K \rightarrow \infty$ and, therefore, it inherits the key theoretical properties of the statistics $A_{K, M_n, t}$.

4.1 A converse theorem

Let us consider the true predictive density $p_t(y)$ and an approximation, computed via particle filtering or otherwise, that we denote as $\hat{p}_t(y)$. In this subsection, we drop the num-

ber of particles M_n in the notation because the results to be presented are valid without regard for the type of approximation of the predictive distribution of the observations, that is, it is not important if it is approximated by a Monte Carlo-based method or is obtained via an analytical approach. The analysis in this section relies to a large extent on the properties of the cumulative distribution functions (cdf's) associated to $p_t(y)$ and $\hat{p}_t(y)$, which we denote as

$$F_t(a) = (\mathbf{1}_{(-\infty,a)}, p_t), \quad \text{and}$$
$$\hat{F}_t(a) = (\mathbf{1}_{(-\infty,a)}, \hat{p}_t),$$

respectively, where

$$\mathbf{1}_A(y) = \begin{cases} 1, & \text{if } y \in A \\ 0, & \text{otherwise} \end{cases}$$

denotes the set-indicator function.

The statistic $\hat{A}_{K,t}$ is computed by generating K i.i.d. fictitious observations from $\hat{p}_t$, denoted $y_t^{(1)}, \ldots, y_t^{(K)}$, and then computing the relative position of the actual observation y_t (distributed according to p_t) within the ordered fictitious observations. From Proposition 1, we know that if $p_t = \hat{p}_t$ then $\hat{A}_{K,t}$ is uniform for every $K \in \mathbb{N}$. Here, we pose the reverse question: if $\hat{A}_{K,t}$ is uniform for every $K \in \mathbb{N}$, can we claim that $p_t = \hat{p}_t$? Moreover, if only *some* statistic $\hat{A}_{K,t}$ is uniform (i.e., for *some* finite $K \in \mathbb{N}$), can we expect $\hat{p}_t$ to be close to p_t in some quantitative well-defined sense?

Our analysis relies on two basic results in probability theory.

Lemma 1 *Let Y be a continuous real r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let p denote a pdf, with cdf $F(\cdot) = (\mathbf{1}_{(-\infty,\cdot)}, p)$. The r.v. $F(Y)$ has uniform distribution $\mathcal{U}(0,1)$ if, and only if, Y is distributed according to p.*

Proof The inversion theorem (see, e.g., Theorem 2.1 in Martino et al. 2018) guarantees that if $Y \sim p$ then the r.v. $F(Y)$ is $\mathcal{U}(0,1)$. For the reverse implication, assume $F(Y) \sim \mathcal{U}(0,1)$, which implies that $\mathbb{P}(F(Y) < a) = a$ for any $a \in [0,1]$. As a consequence,

$$\mathbb{P}(Y < a) = \mathbb{P}(F(Y) < F(a)) = F(a),$$

hence $Y \sim p$.   □

Lemma 2 *Let p denote a pdf with associated cdf $F(\cdot) = (\mathbf{1}_{(-\infty,\cdot)}, p)$. For every $n \in \mathbb{N}$, we have $(F^n, p) = \frac{1}{n+1}$.*

Proof Let Y be a r.v. with pdf p; from Lemma 1 we have $F(Y) \sim \mathcal{U}(0,1)$, hence

$$(F^n, p) = \mathbb{E}\left[F(Y)^n\right] = \mathbb{E}[U^n],$$

where $U \sim \mathcal{U}(0,1)$. It is straightforward to verify that $\mathbb{E}[U^n] = \frac{1}{n+1}$.   □

Using the basic lemmas above, we establish the key result that relates the approximate cdf $\hat{F}_t$ to the true functions F_t and p_t.

Theorem 2 *Assume the observation Y_t is a continuous r.v. with a pdf p_t and cdf F_t. Let the pdf $\hat{p}_t$ and its associated cdf $\hat{F}_t(\cdot) = (\mathbf{1}_{(-\infty,\cdot)}, \hat{p}_t)$ be estimates of p_t and F_t, respectively. If the r.v. $\hat{A}_{K,t}$ constructed from $\hat{p}_t$ is uniform then*

$$(\hat{F}_t^n, p_t) = \frac{1}{n+1}, \quad \forall n \in \{0, 1, \ldots, K\}. \tag{4.1}$$

Proof See Appendix B.   □

Remark 1 Let Y_t be the actual observation with pdf p_t. Given the actual cdf F_t and its estimate $\hat{F}_t$, we can construct the r.v.'s $B_t = F_t(Y_t)$ and $\hat{B}_t = \hat{F}_t(Y_t)$. From Lemma 2, we readily obtain

$$\mathbb{E}[B_t^n] = (F_t^n, p_t) = \frac{1}{n+1} \quad \text{for every } n \geq 0.$$

However, Theorem 2 guarantees that if $\hat{A}_{K,t}$ is uniform, then

$$\mathbb{E}[\hat{B}_t^n] = (\hat{F}_t^n, p_t) = \frac{1}{n+1} \quad \text{for } n = 0, \ldots, K.$$

Therefore, if the statistic $\hat{A}_{K,t}$ is uniform, the r.v.'s $B_t = F_t(Y_t)$ and $\hat{B}_t = \hat{F}_t(Y_t)$ share their first K moments. This is a quantitative characterization of the similarity between F_t and $\hat{F}_t$. In particular, if $\hat{A}_{K,t}$ is uniform for every $K \in \mathbb{N}$, we have $\mathbb{E}[\hat{B}_t^n] = \mathbb{E}[B_t^n] = \frac{1}{n+1}$ for every $n \in \mathbb{N}$ and, as a consequence, $\hat{F}_t = F_t$ and $\hat{p}_t(y) = p_t(y)$ almost everywhere in the observation space.

4.2 Assessment without fictitious observations: the statistic B_t

If Y_t is a continuous r.v. with pdf p_t, then the sequence of statistics $B_t = F_t(Y_t)$, $t \in \mathbb{N}$, is i.i.d. with a common distribution $\mathcal{U}(0,1)$.[5] From Remark 1, it is apparent that we can use the particle filter to compute estimators $\hat{B}_t \equiv B_t^{M_n}$ over a window of observations (i.e., for $t_{n-1} < t \leq t_n$) and then use the estimates to assess the performance of the filter. Two straightforward approaches to performing this assessment are:

- testing for uniformity in (0,1) of the estimates $b_{t_{n-1}+1}^{M_n}, \ldots, b_{t_n}^{M_n}$ or

---

[5] The fact that every B_t is uniform is a consequence of Lemma 1. Independence can be proved by the same argument as in Proposition 3 of Elvira et al. (2017).

– evaluating the sample moments $\frac{1}{W_n} \sum_{t=t_{n-1}+1}^{t_n} \left(b_t^{M_n} \right)^m$, which should be close to $\frac{1}{m+1}$, according to Theorem 2, when the particle filter is "performing well."

Since the approximate cdf of Y_t computed via the particle filter $F_t^{M_n}(a) = (\mathbf{1}_{(-\infty,a)}, p_t^{M_n})$ is an integral w.r.t. $p_t^{M_n}$ and $B_t^{M_n} = F_t^{M_n}(Y_t)$, it follows that the estimates $B_t^{M_n} = b_t^{M_n}$ can be computed with $\mathcal{O}(M_n)$ operations, without generating any fictitious observations, as

$$b_t^{M_n} = (\mathbf{1}_{(-\infty,y_t)}, p_t^{M_n}) = \frac{1}{M_n} \sum_{m=1}^{M_n} \int_{-\infty}^{y_t} g_t^y(\bar{\mathbf{x}}_t^m) \mathrm{d}y.$$

Note, however, that calculating the $b_t^{M_n}$'s from the observation y_t demands the ability to integrate the conditional pdf of the observations $g_t^y(\bar{\mathbf{x}}_t) = p(y|\bar{\mathbf{x}}_t)$. This is a straightforward numerical task when the observation noise is additive and Gaussian, but it may not be possible for other models.

Provided it can be computed, the statistic $B_t^{M_n}$ converges to the actual r.v. B_t when $M_n \to \infty$ under the basic assumptions in Elvira et al. (2017), reproduced below for convenience (and restricted to the case of scalar observations).

(£) For each $t \geq 1$, the function g_t is positive and bounded, i.e., $g_t^y(\mathbf{x}) > 0$ for any $(y, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$ and $\|g_t\|_\infty = \sup_{(y,\mathbf{x}) \in \mathcal{Y} \times \mathcal{X}} |g_t^y(\mathbf{x})| < \infty$.

(𝔇) For each $t \geq 1$, the function $g_t^y(\mathbf{x})$ is Lipschitz-continuous w.r.t. y.

(ℭ) For any $0 < \beta < 1$ and any $p \geq 4$, the sequence of intervals

$$C_M := \left[-\frac{M^{\frac{\beta}{p}}}{2}, +\frac{M^{\frac{\beta}{p}}}{2} \right] \subset \mathbb{R}$$

satisfies the inequality $\mu_t(\overline{C_M}) \leq b M^{-\eta}$ for some constants $b > 0$ and $\eta > 0$ independent of M (yet possibly dependent on β and p), where $\overline{C_M} = \mathbb{R} \backslash C_M$ is the complement of C_M.

To be specific, we have the following result.

Proposition 3 *Let Y_t be a r.v. with pdf $p_t(y_t)$, and let the observations $y_{1:t-1}$ be fixed. If the Assumptions (£), (𝔇) and (ℭ) hold, then there exists a sequence of non-negative r.v.'s $\{\varepsilon_t^{M_n}\}_{M_n \in \mathbb{N}}$ such that $\lim_{M_n \to \infty} \varepsilon_t^{M_n} = 0$ a.s. and*

$$B_t - \varepsilon_t^{M_n} \leq B_{M_n,t} \leq B_t + \varepsilon_t^{M_n}. \tag{4.2}$$

In particular, $\lim_{M_n \to \infty} B_{M_n,t} = B_t$ a.s. and the distribution of $B_t^{M_n}$ converges to $\mathcal{U}(0, 1)$ when $M_n \to \infty$.

Proof Recall that $B_t = (\mathbf{1}_{(-\infty,Y_t)}, p_t)$ and $B_t^{M_n} = (\mathbf{1}_{(-\infty,Y_t)}, p_t^{M_n})$. Therefore, Proposition 3 is a straightforward consequence of Theorem 1 in Elvira et al. (2017), provided that Assumptions (£), (𝔇), and (ℭ) hold. □

Finally, we note the strong connection between the statistics $B_t^{M_n}$ and $A_{K,M_n,t}$. Recall $A_{K,M_n,t}$ represents the number of fictitious observations that are smaller than y_t, while $B_t^{M_n}$ represents the probability $\int_{-\infty}^{Y_t} p_t^{M_n}(y) \mathrm{d}y$. Intuitively, when $K \to \infty$, the empirical rate of observations smaller than y_t should converge to the probability of a fictitious observation being smaller than Y_t. More precisely, we can state the proposition below.

Proposition 4 *If $Y_t \sim p_t$ is a continuous r.v., then*

$$\lim_{K \to \infty} \frac{A_{K,M_n,t}}{K} = B_t^{M_n}.$$

Proof Recall that $B_t^{M_n} = (\mathbf{1}_{(-\infty,Y_t)}, p_t^{M_n}) \leq 1$. It is possible to estimate this integral by drawing K samples $\tilde{y}_t^{(k)}$ from μ_t^M and building the standard Monte Carlo estimator $\frac{1}{K} \sum_{k=1}^K \mathbf{1}_{(-\infty,y_t)}(\tilde{y}_t^{(k)}) = \frac{A_{K,M_n,t}}{K}$. Note that this estimator is unbiased, i.e., according to the strong law of large numbers,

$$\lim_{K \to \infty} \frac{A_{K,M_n,t}}{K} \equiv \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^K \mathbf{1}_{(-\infty,y_t)}(\tilde{y}_t^{(k)}) = B_t^{M_n}.$$

□

5 Numerical experiments

In the first experiment, we show the relation between the correlation coefficient of $A_{K,M,t}$ and the MSE of an estimator obtained from the particle approximation in a nonlinear state-space model. Then, we complement the results of Sect. 4.1, showing numerically some properties of $A_{K,M,t}$ for different values of K and M, and their connection to the statistic $B_{M,t}$. Third, we illustrate numerically the convergence of the block-adaptive BF.

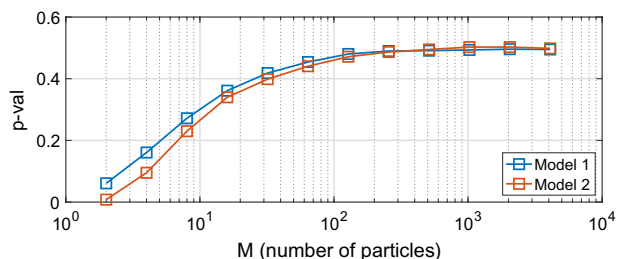## 5.1 Assessing convergence from the correlation of $A_{K,M,t}$.

Consider the stochastic growth model (see, e.g., Djurić and Míguez 2010),

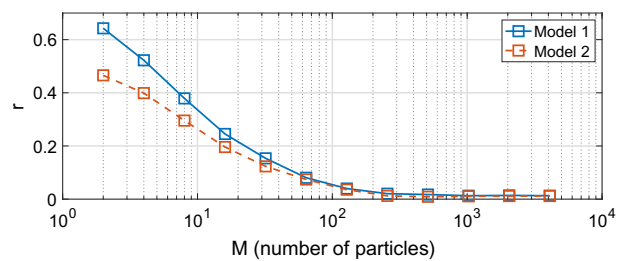$$x_t = \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1 + x_{t-1}^2} + 8\cos(\phi t) + u_t, \tag{5.1}$$

$$y_t = \frac{x_t^2}{20} + v_t, \tag{5.2}$$
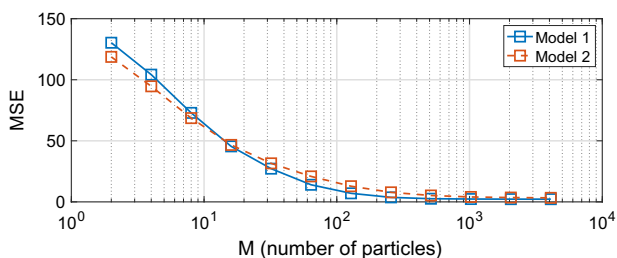
(a) MSE in the estimate of the posterior mean.

(b) Algorithm 1 of Section 2.3. We show the p-value of the Pearson's χ^2 test for assessing the uniformity of the statistic $A_{K,M,t}$.

(c) Algorithm 2 of Section 2.3. The computed Pearson's correlation coefficient r as a function of M.

Fig. 1 Stochastic growth model: MSE, p value of the Pearson's χ^2, and Pearson's correlation coefficient r

where $\phi = 0.4$, and u_t and v_t are independent Gaussian r.v.'s with zero mean, and variance σ_u^2 and σ_v^2, respectively. At this point, we define two models:

– Model 1: $\sigma_u = 1$ and $\sigma_v = 0.5$,
– Model 2: $\sigma_u = 2$ and $\sigma_v = 0.1$.

In this example, we ran the BF for $T = 5,000$ time steps, always with a fixed number of particles M. We tested different values, namely $M \in \{2, 2^2, 2^3, \ldots, 2^{12}\}$. In order to assess the behavior of $A_{K,M,t}$, we set $K = 7$ fictitious observations.

Figure 1a shows the mean squared error (MSE) of the estimate of the posterior mean for each value of M, which obviously decreases as M increases. Figure 1b displays the p value of the Pearson's χ^2 test for assessing the uniformity of $A_{K,M,t}$ (in the domain $A_{K,M,t} \in \{0, \ldots, K+1\}$) in windows of length $W = 20$ (Algorithm 1 of Sect. 2.3; see more details in Elvira et al. 2017). Clearly, increasing the number of particles also increases the p value, i.e., the distribution of the

statistic becomes closer to the uniform distribution. Figure 1c is related to Algorithm 2 of Sect. 2.3.2. We show the sample Pearson's correlation coefficient r, using the whole sequence of statistics $\{a_{K,M,t}\}_{t=1}^T$, computed with a lag $\tau = 1$. All results are averaged over 200 independent runs.

We observe that when we increase M, the correlation between consecutive statistics decreases. It is interesting to note that the curve of the correlation coefficient r has a very similar shape to the MSE curve and, hence, can be used as a proxy. While r can be easily computed, the MSE is always unknown. This shows the utility of Algorithm 2.

It can be seen that both algorithms can identify a malfunctioning of the filter when the number of particles is insufficient. We note that Algorithm 2 works better for Model 1 than for Model 2 because the autocorrelation of the statistics is more sensitive for detecting the malfunctioning for low M. However, Algorithm 1 works better for Model 2 because the p value of the uniformity test is always smaller than in Model 1, i.e., it is more discriminative. Therefore, there is no clear superiority of one algorithm over the other.

5.2 Effect of the choice of the number of fictitious observations K

In this experiment, we evaluate the effect of the value K in the performance of the uniformity test of the statistic $A_{K,M,t}$. We use the same model parameters as in the previous example. First, we fix the number of particles $M = \{2, 4, 16, 64, 256, 1024, 4096\}$ for each run during the T time steps (i.e., no adaptation is performed). Then, with $W = 15$ and $K \in \{2, 3, 5, 7, 10, 20\}$, we compute the p value of the Pearson's χ^2 test for assessing the uniformity of the statistic $A_{K,M,t}$. In Table 2, we show the average of the p value over 1000 independent runs for all combinations of K and M. We can see that the misbehavior of the filter with a low number of particles M can be detected regardless the number of fictitious observations K. For a larger K, in general the p value decreases but it does not make a significant difference, which confirms our hypotheses in Sect. 4: (a) the framework is robust to the selection of K; (b) increasing K increases the detection power of the algorithms; (c) for reasonable values of K, when the filter misbehaves, the assessment of the uniformity of $A_{K,M,t}$ detects the misbehavior (and when the filter works well, there are not false alarms with large K); and (d) a small K can be selected, which implies a low extra computational complexity of the proposed methodology.

In a second experiment, we implement Algorithm 1 described in Sect. 2.3 on the same model, now using $T = 10^4$ as the length of the time series. We set the algorithms parameters as $p_\ell = 0.2$, $p_h = 0.6$, $W \in \{50, 200\}$, and an initial number of particles $M_0 = \{16, 1024\}$. In Table 3, we show the resulting number of particles averaged over the last 50 windows of the adaptive algorithm

Table 2 (Ex. of Section 5.2) p value of the uniformity test for different values of K and M, averaged over 1000 independent runs

M / K	2	3	5	7	10	20
2	0.0006547	5.571e−06	1.891e−08	2.466e−10	7.979e−13	0
4	0.0358	0.009322	0.001157	0.0003728	4.382e−05	1.473e−07
16	0.3775	0.3153	0.2684	0.2396	0.2293	0.1996
64	0.4681	0.4354	0.4288	0.4285	0.4256	0.4117
256	0.5511	0.5495	0.5667	0.5801	0.5759	0.5781
1024	0.5793	0.578	0.5738	0.5863	0.5845	0.6074
4096	0.5472	0.5686	0.5778	0.5913	0.5949	0.593

For each run, the number of particles M is fixed

Table 3 (Ex. of Section 5.2) Averaged number of particles in the last 50 windows with $T = 10^4$, averaged over 100 independent runs for different values of the window size $W \in \{50, 200\}$ and initial number of particles $M_0 = \{16, 128, 1024\}$

| T | M_0 | W | K | | | | | | | | |
			2	3	5	7	9	10	11	15	20
2×10^4	16	50	153.47	207.93	231.86	251.93	247.18	248.27	251.18	258.31	272.79
2×10^4	128	50	151.82	204.25	231.2	252.19	246.66	248.01	255.4	255.1	279.91
2×10^4	1024	50	152.38	209.92	231.68	248.69	251.2	246.17	254.79	251.57	267.49
2×10^4	16	200	303.02	371.48	417.23	414.44	434.48	434.94	433.03	440.29	438.54
2×10^4	128	200	304.23	382.7	399.5	418.79	421.21	438.79	436.32	437.64	439.61
2×10^4	1024	200	301.2	366.35	400.22	432.99	447.4	434.83	431.76	444.38	445.81

implemented for each value of fictitious observations, $K \in \{2, 3, 5, 7, 9, 10, 11, 15, 20\}$. The results are also averaged over 100 independent runs. Figure 2 shows the same results of the averaged number of particles as a function of K. We see again that the selected number of particles does not depend much on K, as observed in the previous experiment. We also see that the window length does have an effect, requiring a higher number of particles when W is larger. The reason is that a larger W implies that more realizations of the statistic are observed, so in cases where the filter is tracking but with some non-negligible errors, it is more likely that the statistical test rejects the null hypothesis whenever more evidence is accumulated.

Figures 3 and 4 show the averaged number of particles (over 100 independent runs) as a function of time. In Fig. 3, each subplot is obtained by fixing $M_0 \in \{16, 128, 1024\}$ and each line represents the evolution of the number of particles for each $K \in \{2, 3, 5, 7, 9, 15, 20\}$. We see that for most values K, the averaged number of particles is very similar. It is interesting to note that at some stages, the required number of particles is larger, and this is better detected with slightly larger values of K. In Fig. 4, we show the same information, but now each subplot is obtained by fixing $K \in \{5, 9\}$, and each line represents the evolution of the number of particles for each initialization $M_0 \in \{16, 128, 1024\}$. We note that regardless the initial number of particles, after around
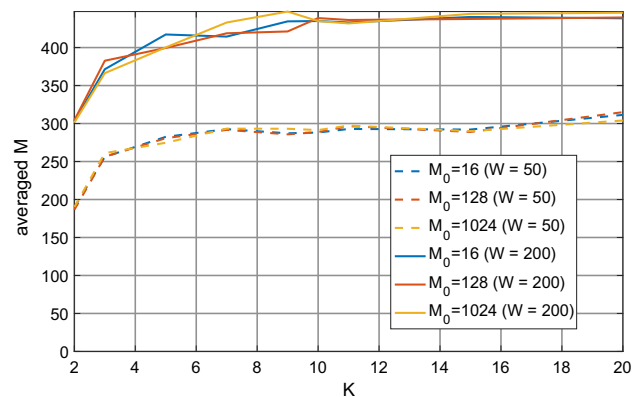

Fig. 2 Number of particles averaged over the last 50 windows of the adaptive algorithm implemented as a function of the number fictitious observations $K \in \{2, 3, 5, 7, 10, 20, \}$. Each curve represents different values of the window size $W \in \{50, 200\}$ and initial number of particles $M_0 = \{16, 128, 1024\}$

3, 000 time steps, the averaged number of particles is the same.

5.3 The three-dimensional Lorenz system

Table 4 shows results of the Lorenz example described in Elvira et al. (2017, Section V-A) with fixed number of particles M. We show the MSE in the approximation of the
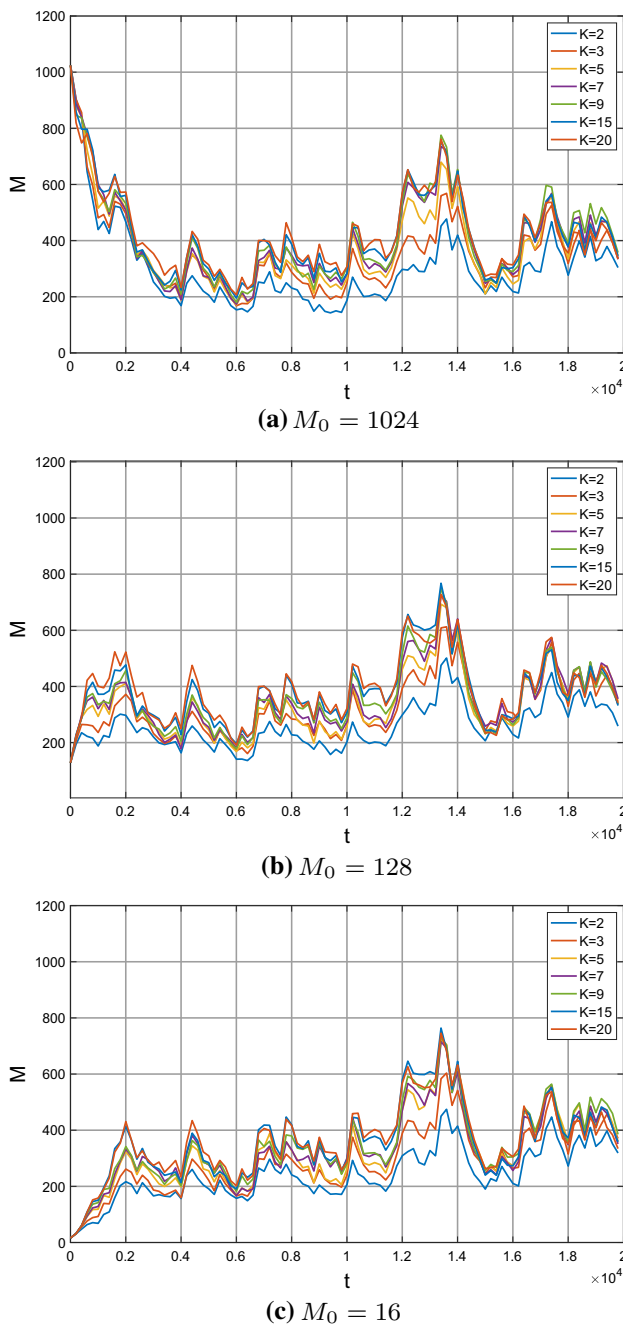
(a) $M_0 = 1024$

(b) $M_0 = 128$

(c) $M_0 = 16$

Fig. 3 Averaged number of particles as a function of time for a fixed number of initial number of particles $M_0 = \{16, 128, 1024\}$
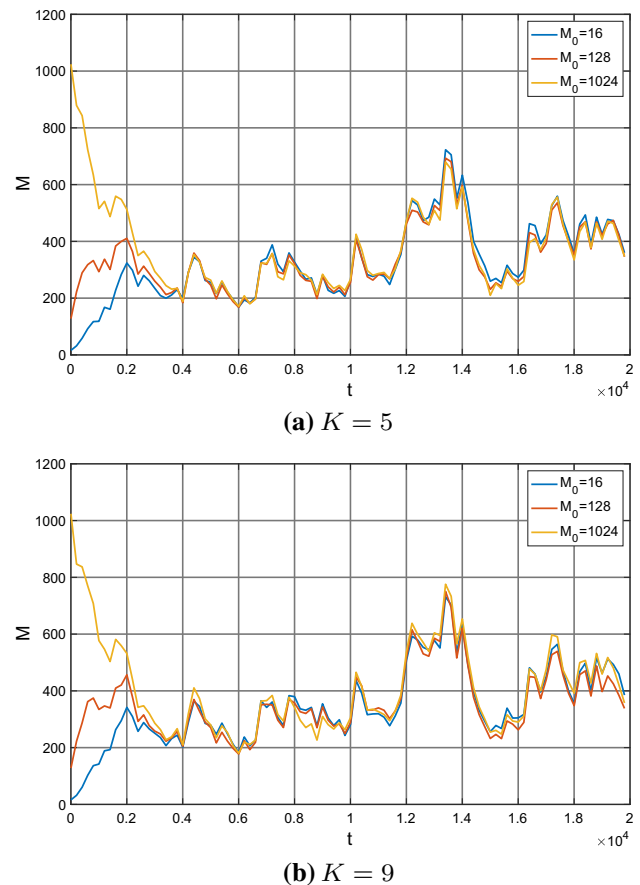

(a) $K = 5$

(b) $K = 9$

Fig. 4 Averaged number of particles as a function of time for a fixed number of fictitious observations $K = \{5, 9\}$

5.4 Behavior of $A_{K,M,t}$ and its relation with $B_{M,t}$

In Fig. 5, we show the histograms of $A_{K,M,t}$ and $B_{M,t}$ for the stochastic growth model described in (5.1) and (5.2). We set $K \in \{3, 5, 7, 10, 20, 50, 100, 1000, 5000\}$. The BF is with fixed $M = 2^{14}$. When K grows, the pmf seems to converge to the pdf of $B_{M,t}$.

In Table, 5 we show the averaged absolute error (distance) between the realizations of r.v.'s $A_{K,M,t}/K$ and $B_{M,t}$ for the stochastic growth model with fixed $M = 2^{14}$. The results are averaged over $T = 100$ time steps in 100 independent runs. It is clear that when K grows, the deviation between both r.v.'s, which take values in $(0, 1)$, decreases. Thus, for $K = 5000$, the difference is on average 0.43%.

5.5 Forgetting property in the block-adaptive bootstrap particle filter

In this section, we assess the approximation errors when the block-adaptive BF increases the number of particles. To that end, we run two specific state-space models where, in the first half of time steps, the number of particles is set to M_1 while,

posterior mean, averaged over 200 runs. Again r is the sample Pearson's correlation coefficient, using the whole sequence of statistics $\{a_{K,M,t}\}_{t=1}^{T}$ with a lag $\tau = 1$, and p val is the p value of the Pearson's χ^2 test for assessing the uniformity of the same set. Similar conclusions as in the previous example can be obtained.

Table 4 Lorenz Model: $\Delta = 10^{-3}$, $T_{\mathrm{obs}} = 200\Delta$, $\sigma^2 = 0.5$

Fixed M	8	16	32	64	128	256	512	1024	2048	4096	8192	16,384
MSE	105.63	75.56	40.19	15.69	5.90	2.90	1.77	1.55	1.53	1.52	1.52	1.52
$\hat{R}(1)$	0.6927	0.4939	0.2595	0.1132	0.0463	0.0273	0.0210	0.0190	0.0195	0.0151	0.0151	0.0192
p value	0.0393	0.1276	0.2923	0.4279	0.4823	0.5016	0.5117	0.5106	0.4998	0.5141	0.5040	0.5181

Algorithm details: $W = 20$, $K = 7$. MSE in the approximation of the posterior mean, the averaged $\hat{R}(1)$, and the averaged p value of the Pearson's chi-square test on the uniformity on $\mathcal{S}_t$
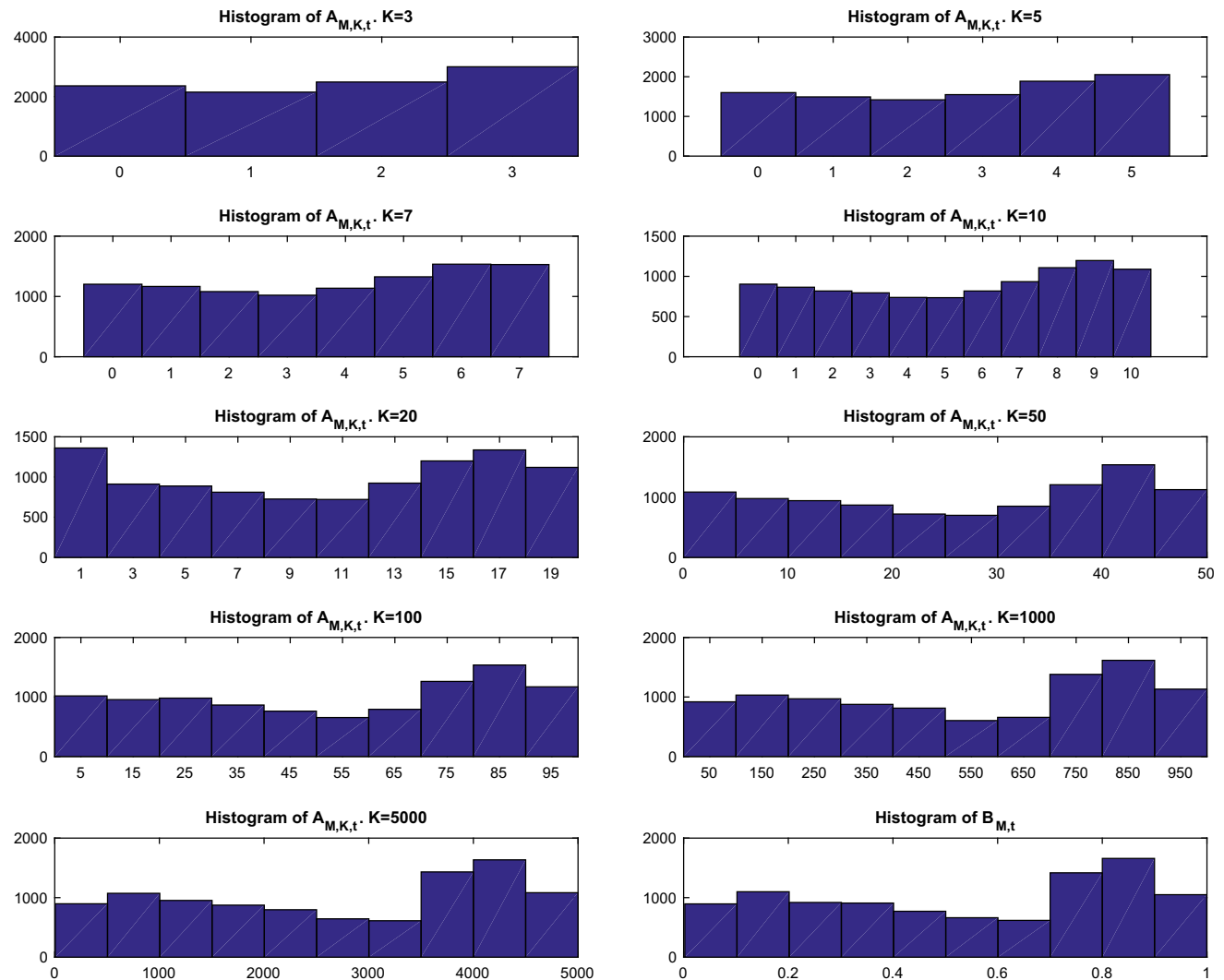
Fig. 5 (Ex. of Section 5.5) Histograms of the r.v. $A_{K,M,t}$ and $B_{M,t}$ in the stochastic growth model with fixed $M = 2^{14}$. The histograms accumulate $T = 100$ time steps in 100 independent runs

in the second half, the number of particles is $M_2 > M_1$. We then compute the MSE of predicted observations (in the last quarter of the time steps), and we compare it with the standard BF with M_2 particles used from the beginning.

Table 6 shows the MSE of a BF run on the linear Gaussian model described by

$$x_t = ax_{t-1} + u_t, \tag{5.3}$$

$$y_t = x_t + v_t, \tag{5.4}$$

with $T = 1000$, $\sigma_u = \sqrt{0.5}$, $\sigma_v = 1$, and $a = 0.9$. We simulate one example with $M_1 = 100$ and $M_2 = 1000$ (left side of the table), and another one with $M_1 = 1000$ and $M_2 = 10,000$ (right side of the table). In both cases, we are able to show that the BF achieves in the last quarter of time

Table 5 (Ex. of Section 5.5) Averaged absolute error (distance) between the realizations of the r.v.'s $A_{K,M,t}$ and $B_{M,t}$ for the stochastic growth model $M = 2^{14}$

K	2	3	5	7	10	20	50	100	1000	5000
$\left| B_{M,t} - \frac{A_{M,K,t}}{K} \right|$	0.2254	0.1836	0.1409	0.1183	0.0987	0.0696	0.0435	0.0305	0.0097	0.0043

The results are averaged over $T = 100$ time steps in 100 independent runs

Table 6 Linear Gaussian Model: $T = 1000$, $\sigma_x = \sqrt{0.5}$, $\sigma_y = 1$, $a = 0.9$

M_1	100	1000	100	1000	10,000	1000
M_2			1000			10,000
MSE (last $T/4$)	$8.90\dot{\text{i}}0^{-3}$	$9.02\dot{\text{i}}0^{-4}$	$8.99\dot{\text{i}}0^{-4}$	$9.02\dot{\text{i}}0^{-4}$	$8.93\dot{\text{i}}0^{-5}$	$8.69\dot{\text{i}}0^{-5}$

M_1 particles for $t \in \{1, \ldots, \frac{T}{2}\}$ and M_2 particles for $t \in \{\frac{T}{2} + 1, \ldots, T\}$

Table 7 Stochastic growth model: $T = 1000$, $\sigma_x = 1$, $\sigma_y = 0.1$, $\phi = 0.4$

M_1	50	1000	50	200	4000	200	1000	20,000	1000
M_2			1000			4000			20,000
MSE (last $T/4$)	16.69	1.493	1.564	4.815	1.386	1.374	1.494	1.348	1.335

M_1 particles for $t \in \{1, \ldots, \frac{T}{2}\}$ and M_2 particles for $t \in \{\frac{T}{2} + 1, \ldots, T\}$

steps (from $t = 750$ to $t = T = 1000$) the same MSE as if the largest number of particles was set at the very beginning.

Table 7 presents analogous results for the stochastic growth model described in the first experiment, with $T = 1000$, $\sigma_x = 1$, $\sigma_y^2 = 0.1$, and $\phi = 0.4$. Now we simulate the BF with the following pairs of number of particles $(M_1, M_2) \in \{(50, 1000), (200, 4000), (1000, 20,000)\}$. We arrive at the same conclusions.

6 Summary and conclusions

In this paper, we have provided new methodological, theoretical and numerical results on the performance of particle filtering algorithms with an adaptive number of particles. We have looked into a class of PFs that update the number of particles periodically, at the end of observations blocks of a prescribed length. Decisions on whether to increase or decrease the computational effort are automatically made based on predictive statistics which are computed by generating *fictitious observations*, i.e., particles in the observation space. For this type of algorithms, we have proved that:

(a) The error bounds for the adaptive PF depend on the current number of particles (say, M_n) and the dependence on former values (say $M_{n-1}, M_{n-2}, \ldots$) decays exponentially with the block length. This result holds under standard assumptions on the Markov kernel of the state-space model (as discussed in Del Moral (2004), Künsch (2005) and others). This result, which does not follow from classical convergence theorems for Monte Carlo

filters, implies that one can effectively tune the performance of the PF by adapting the computational effort.

(b) Convergence of the predictive statistics used for making decisions on the adaptation of the computational effort implies convergence of the PF itself. To be specific, we have proved that if the predictive statistics computed with K fictitious observations attain a uniform distribution then the true filtering distribution and its particle approximation have K common moments. This result can be understood as a converse to the convergence theorem introduced in Elvira et al. (2017). It guarantees that assessing the convergence of the PF using the proposed predictive statistics is a sound approach (the "more uniform" the predictive statistics, the better the PF general performance).

In addition to the theoretical analysis, we have carried out an extensive computer simulation study. On one hand, the numerical results have corroborated the theoretical results, e.g., by showing how increasing the number of particles directly improves the performance (and past errors are forgotten), or how increasing the number of fictitious observations K (or the block size) leads to a higher computational effort and more accurate estimators. We have also shown that the proposed block-adaptive algorithms are stable w.r.t. the initial number of particles.

Overall, the proposed class of algorithms is easy to implement and can be used with different versions of the PF. It also enables the automatic, online tuning of the computational

complexity without time-consuming (and often unreliable) trial-and-error procedures.

A Proof of Theorem 1

The argument of the proof relies on the following lemma:

Lemma 3 *If Assumptions 1 and 2 hold, the composition of $r \geq 1$ consecutive PU operators satisfies the inequality*

$$\sup_{|f| \leq 1} \left| \left(f, \Psi_{t|t-r}(\alpha) \right) - \left(f, \Psi_{t|t-r}(\beta) \right) \right|$$
$$\leq \frac{2 \left(1 - \varepsilon_m^2 \gamma^{m-1} \right)^{\lfloor \frac{r}{m} \rfloor}}{\gamma^m \varepsilon_m} \sup_{|\tilde{f}| \leq 1} \left| (\tilde{f}, \alpha) - (\tilde{f}, \beta) \right|$$

for any probability measures $\alpha, \beta \in \mathcal{P}(\mathcal{X})$.

The proof of Lemma 3 follows immediately from Propositions 4.3.3 and 4.3.7 in Del Moral (2004). This is now a classical result that has been exploited for many particle-based algorithms (see, e.g., Gupta et al. (2015) or Crisan and Miguez (2017)).

Recall that $t_n = \sum_{j^0}^{n} W_j$ is the last time instant of the nth block of observations. For any bounded real test function $f : \mathcal{X} \mapsto \mathbb{R}$, the approximation error $(f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n})$ can be written in terms of one-step-ahead differences using the "telescope" decomposition

$$(f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n}) = \sum_{k=0}^{W_n-2} D_k(M_n)$$
$$+ D(M_n, M_{n-1}) + E(M_{n-1}), \text{ (A.1)}$$

where

$$D_k(M_n) := \left(f, \Psi_{t_n|t_n-k} \left(\pi_{t_n-k}^{M_n} \right) \right)$$
$$- \left(f, \Psi_{t_n|t_n-k} \left(\Psi_{t_n-k} \left(\pi_{t_n-k-1}^{M_n} \right) \right) \right)$$

are the local (one step) differences in the nth window,

$$D(M_n, M_{n-1}) = \left(f, \Psi_{t_n|t_n-W_n+1} \left(\pi_{t_n-W_n+1}^{M_n} \right) \right)$$
$$- \left(f, \Psi_{t_n|t_n-W_n+1} \left(\Psi_{t_n-W_n+1} \left(\pi_{t_n-W_n}^{M_{n-1}} \right) \right) \right)$$

is the one-step difference between the last time instant of window $n-1$ and the first time instant of window n, and

$$E(M_{n-1}) = \left(f, \Psi_{t_n|t_n-W_n} \left(\pi_{t_n-W_n}^{M_{n-1}} \right) \right)$$
$$- \left(f, \Psi_{t_n|t_n-W_n} \left(\pi_{t_n-W_n} \right) \right)$$

is the approximation error inherited from window $n-1$. Using Lemma 3 and writing $t_{n-1} = t_n - W_n$ for conciseness, we readily find that for any test function f, with $\|f\|_\infty \leq 1$, the different terms on the rhs of (A.1) can be upper bounded as

$$|D_k(M_n)|$$
$$\leq \frac{2 \left(1 - \varepsilon_m^2 \gamma^{m-1} \right)^{\lfloor \frac{k}{m} \rfloor}}{\gamma^m \varepsilon_m}$$
$$\times \sup_{|\tilde{f}| \leq 1} \left| (\tilde{f}, \pi_{t_n-k}^{M_n}) - \left(\tilde{f}, \Psi_{t_n-k} \left(\pi_{t_n-k-1}^{M_n} \right) \right) \right|, \quad \text{(A.2)}$$

$$|D(M_n, M_{n-1})| \leq \frac{2 \left(1 - \varepsilon_m^2 \gamma^{m-1} \right)^{\lfloor \frac{W_n-1}{m} \rfloor}}{\gamma^m \varepsilon_m}$$
$$\times \sup_{|\tilde{f}| \leq 1} \left| (\tilde{f}, \pi_{t_{n-1}+1}^{M_n}) - \left(\tilde{f}, \Psi_{t_{n-1}+1} \left(\pi_{t_{n-1}}^{M_{n-1}} \right) \right) \right|, \text{(A.3)}$$

and

$$|E(M_{n-1})| \leq \frac{2 \left(1 - \varepsilon_m^2 \gamma^{m-1} \right)^{\lfloor \frac{W_n}{m} \rfloor}}{\gamma^m \varepsilon_m}$$
$$\times \sup_{|\tilde{f}| \leq 1} \left| (\tilde{f}, \pi_{t_{n-1}}^{M_{n-1}}) - \left(\tilde{f}, \pi_{t_{n-1}}^{M_{n-1}} \right) \right|. \quad \text{(A.4)}$$

Moreover, the measures $\pi_{t_n-k}^{M_n}$ are importance sampling approximations of $\Psi_{t_n-k} \left(\pi_{t_n-k-1}^{M_n} \right)$, for $k = 0, \ldots, W_n - 2$. As a consequence, it can easily be shown that that there is a constant $0 < c < \infty$, independent of M_n, t_n and k, such that

$$\left\| (f, \pi_{t_n-k}^{M_n}) - \left(f, \Psi_{t_n-k} \left(\pi_{t_n-k-1}^{M_n} \right) \right) \right\|_p \leq \frac{\|f\|_\infty c}{\sqrt{M_n}} \quad \text{(A.5)}$$

for any bounded test function f and any $p > 1$. The same result also holds for $\pi_{t_{n-1}+1}^{M_n}$ and $\Psi_{t_{n-1}+1} \left(\pi_{t_{n-1}}^{M_{n-1}} \right)$, namely,

$$\left\| (f, \pi_{t_{n-1}+1}^{M_n}) - \left(f, \Psi_{t_{n-1}+1} \left(\pi_{t_{n-1}}^{M_{n-1}} \right) \right) \right\|_p \leq \frac{\|f\|_\infty c}{\sqrt{M_n}}.$$
$$\text{(A.6)}$$

Finally, if we apply Minkowsky's inequality in Eq. (A.1) and then combine the bounds (A.2) and (A.3) with (A.5) and (A.6), respectively, we obtain, for any $p \geq 1$

$$
\sup_{|f| \leq 1} \left| (f, \pi_{t_n}^{M_n}) - (f, \pi_{t_n}) \right|_p
$$
$$
\leq \sum_{k=0}^{W_n-1} \frac{2c \left(1 - \varepsilon_m^2 \gamma^{m-1}\right)^{\lfloor \frac{k}{m} \rfloor}}{\gamma^m \varepsilon_m \sqrt{M_n}}
$$
$$
+ \frac{2 \left(1 - \varepsilon_m^2 \gamma^{m-1}\right)^{\lfloor \frac{W_n}{m} \rfloor}}{\gamma^m \varepsilon_m} \sup_{|f| \leq 1} \left\| \left(f, \pi_{t_{n-1}}^{M_{n-1}}\right) - (f, \pi_{t_{n-1}}) \right\|_p \text{(A.7)}
$$

The proof is complete by pointing out that, taking $W_n \to \infty$ in the sum $\sum_{k=0}^{W_n-1}(\cdot)$ on the rhs of (A.7), we arrive at

$$
\sum_{k=0}^{W_n-1} \frac{2c \left(1 - \varepsilon_m^2 \gamma^{m-1}\right)^{\lfloor \frac{k}{m} \rfloor}}{\gamma^m \varepsilon_m \sqrt{M_n}} \leq \frac{mC}{\varepsilon_m^3 \gamma^{2m-1} \sqrt{M_n}}, \tag{A.8}
$$

where $C = 2c$. □

B Proof of Theorem 2

First, we express the pmf $\hat{\mathbb{Q}}_{K,t}$ of the r.v. $\hat{A}_{K,t}$ as a function of the predictive pdf of the observations. Specifically, we note that $\hat{\mathbb{Q}}_{K,t}(n)$ is the probability that exactly n fictitious observations are smaller than the actual y_t. Hence, $\forall n \in \{0, \ldots, K\}$,

$$
\hat{\mathbb{Q}}_{K,t}(n) = \int_{-\infty}^{\infty} \mathbb{P}(\hat{A}_{K,t} = n | y_t = z) p_t(z) \mathrm{d}z
$$
$$
= \int_{-\infty}^{\infty} \binom{K}{n} \hat{F}_t(z)^n \left(1 - \hat{F}_t(z)\right)^{K-n} p_t(z) \mathrm{d}z
$$
$$
= \binom{K}{n} \int_{-\infty}^{\infty} \hat{F}_t(z)^n \left(\sum_{i=0}^{K-n} \binom{K-n}{i} (-1)^i \hat{F}_t(z)^i \right) p_t(z) \mathrm{d}z
$$
$$
= \binom{K}{n} \sum_{i=0}^{K-n} \int_{-\infty}^{\infty} \binom{K-n}{i} (-1)^i \hat{F}_t(z)^{n+i} p_t(z) \mathrm{d}z,
$$
$$
\tag{B.1}
$$

where $\hat{F}_t(z)$ is the approximation of the cdf of the predictive observation evaluated at z. Note that, as a consequence, $\hat{F}_t(z)$ is also the probability of a single fictitious observation being smaller than z.

Recall that we want to prove that (4.1) when $\hat{A}_{K,t}$ is uniform (and, as a consequence, $\hat{\mathbb{Q}}_{K,t}(n) = \frac{1}{K+1}$ for every $n \in \{0, \ldots, K\}$). We apply an induction argument in n. The case for $n = K$ is obvious by rewriting (B.1) as

$$
\hat{\mathbb{Q}}_{K,t}(K) = \int_{-\infty}^{\infty} \hat{F}_t(z)^K p_t(z) \mathrm{d}z = (\hat{F}_t^K, p_t),
$$

hence $(\hat{F}_t^K, p_t) = \frac{1}{K+1}$. Next, we assume that for a specific $n \in \{1, \ldots, K\}$, the identity (4.1) holds for all $i \in \{n, \ldots, K\}$ and then aim at proving that it also holds for $n - 1$. Let us write the pmf of (B.1) at $n - 1$ as

$$
\hat{\mathbb{Q}}_{K,t}(n-1)
$$
$$
= \binom{K}{n-1} \sum_{i=0}^{K-n+1} \int_{-\infty}^{\infty} \binom{K-n+1}{i} (-1)^i \hat{F}_t(z)^{n+i-1} p_t(z) \mathrm{d}z
$$
$$
= \binom{K}{n-1} \left[\sum_{i=1}^{K-n+1} \binom{K-n+1}{i} \frac{(-1)^i}{n+i} + \int_{-\infty}^{\infty} \hat{F}_t(z)^{n-1} p_t(z) \mathrm{d}z \right]
$$
$$
= \binom{K}{n-1} \frac{1}{n} \left(\sum_{i=0}^{K-n+1} \binom{K-n+1}{i} (-1)^i \frac{n}{n+i} - 1 \right)
$$
$$
+ \binom{K}{n-1} \int_{-\infty}^{\infty} \hat{F}_t(z)^{n-1} p_t(z) \mathrm{d}z
$$
$$
= \binom{K}{n-1} \left[\frac{1}{n} \left(\frac{1}{\binom{K+1}{K-n+1}} - 1 \right) + \int_{-\infty}^{\infty} \hat{F}_t(z)^{n-1} p_t(z) \mathrm{d}z \right],
$$
$$
\tag{B.2}
$$

where the second identity is obtained replacing all the integrals (except the one corresponding to $n - 1$) using the induction hypothesis; for the third equality, we split the series between the terms $i > 0$ and the term $i = 1$, and in the fourth equation, we substitute the series using identity (1.41) in Gould (1972). Once again, since $\hat{A}_{K,t}$ is uniform, we have $\hat{\mathbb{Q}}_{K,t}(n-1) = \frac{1}{K+1}$, hence we rewrite (B.2) as

$$
\frac{1}{K+1} = \binom{K}{n-1} \frac{1}{n} \left(\frac{1}{\binom{K+1}{K-n+1}} - 1 \right)
$$
$$
+ \binom{K}{n-1} \int_{-\infty}^{\infty} \hat{F}_t(z)^{n-1} p_t(z) \mathrm{d}z,
$$

where precisely, $\int_{-\infty}^{\infty} \hat{F}_t(z)^{n-1} p_t(z) \mathrm{d}z = (\hat{F}_t^{n-1}, p_t)$. If we simply solve for the latter integral and simplify, we arrive at

$$
(\hat{F}_t^{n-1}, p_t) = \frac{1}{\binom{K}{n-1}} \left(\frac{1}{K+1} - \binom{K}{n-1} \frac{1}{n} \left(\frac{1}{\binom{K+1}{K-n+1}} - 1 \right) \right)
$$
$$
= \frac{(n-1)!(K-n+1)!}{K!} \frac{1}{K+1} - \frac{1}{n} \left(\frac{(K-n+1)!n!}{(K+1)!} - 1 \right)
$$
$$
= \frac{(n-1)!(K-n+1)!}{(K+1)!} - \frac{(K-n+1)!(n-1)!}{(K+1)!} + \frac{1}{n}
$$
$$
= \frac{1}{n}. \tag{B.3}
$$

References

Anderson, J.L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J. Clim. **9**(7), 1518–1530 (1996)

Andrieu, C., Doucet, A., Holenstein, R.: Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. B **72**(3), 269–342 (2010)

Bain, A., Crisan, D.: Fundamentals of Stochastic Filtering. Springer, Berlin (2008)

Bhadra, A., Ionides, E.L.: Adaptive particle allocation in iterated sequential Monte Carlo via approximating meta-models. Stat. Comput. **26**(1–2), 393–407 (2016)

Bröcker, J.: Assessing the reliability of ensemble forecasting systems under serial dependence. Q. J. R. Meteorol. Soc. **144**(717), 2666–2675 (2018)

Chopin, N., Jacob, P.E., Papaspiliopoulos, O.: SMC2: an efficient algorithm for sequential analysis of state space models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **75**, 397–426 (2012)

Cornebise, J.: Adaptive sequential Monte Carlo methods. Ph.D. thesis, Télécom ParisTech, 2010. 38, p 49 (2009)

Crisan, D.: Particle filters—a theoretical perspective, chap 2. In: Doucet, A., de Freitas, N., Gordon, N. (eds.) Sequential Monte Carlo Methods in Practice, pp. 17–42. Springer, Berlin (2001)

Crisan, D., Doucet, A.: A survey of convergence results on particle filtering. IEEE Trans. Signal Process. **50**(3), 736–746 (2002)

Crisan, D., Miguez, J.: Uniform convergence over time of a nested particle filtering scheme for recursive parameter estimation in state-space Markov models. Adv. Appl. Probab. **49**(4), 1170–1200 (2017)

Crisan, D., Miguez, J.: Nested particle filters for online parameter estimation in discrete-time state-space Markov models. Bernoulli **24**(4A), 3039–3086 (2018)

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

Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Annales de l'Institut Henri Poincaré (B) Probab. Stat. **37**(2), 155–194 (2001)

Djurić, P.M., Míguez, J.: Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics. IEEE Trans. Signal Process. **58**(10), 5069–5079 (2010)

Djurić, P.M., Kotecha, J.H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M.F., Míguez, J.: Particle filtering. IEEE Signal Process. Mag. **20**(5), 19–38 (2003)

Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. **10**(3), 197–208 (2000)

Doucet, A., de Freitas, N., Gordon, N. (eds.): Sequential Monte Carlo Methods in Practice. Springer, New York (2001)

Elvira, V., Míguez, J., Djurić, P.M.: A novel algorithm for adapting the number of particles in particle filtering. In: 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, pp 1–5 (2016)

Elvira, V., Míguez, J., Djurić, P.M.: Adapting the number of particles in sequential Monte Carlo methods through an online scheme for convergence assessment. IEEE Trans. Signal Process. **65**(7), 1781–1794 (2017)

Elvira, V., Martino, L., Robert, C.P.: Rethinking the effective sample size. arXiv:1809.04129 (2018)

Fox, D.: Adapting the sample size in particle filters through KLD-sampling. Int. J. Robot. Res. **22**(12), 985–1003 (2003)

Gordon, N., Salmond, D., Smith, A.F.M.: Novel approach to nonlinear and non-Gaussian Bayesian state estimation. IEE Proc. F Radar Signal Process. **140**, 107–113 (1993)

Gould, H.W.: Combinatorial Identities: A Standardized Set of Tables Listing 500 Binomial Coefficient Summations. Morgantown, VA (1972)

Gupta, S.D., Coates, M., Rabbat, M.: Error propagation in gossip-based distributed particle filters. IEEE Trans. Signal Inf. Proces. Netw. **1**(3), 148–163 (2015)

Hamill, T.M.: Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. **129**(3), 550–560 (2001)

Hu, X., Schon, T., Ljung, L.: A basic convergence result for particle filtering. IEEE Trans. Signal Proces. **56**(4), 1337–1348 (2008)

Jasra, A., Lee, A., Yau, C., Zhang, X.: The alive particle filter. arXiv:1304.0151 (2013)

Künsch, H.R.: Recursive Monte Carlo filters: algorithms and theoretical bounds. Ann. Stat. **33**(5), 1983–2021 (2005)

Künsch, H.R.: Particle filters. Bernoulli **19**(4), 1391–1403 (2013)

Lee, A., Whiteley, N.: Variance estimation in the particle filter. Biometrika **105**(3), 609–625 (2018)

LeGland, F., Oudjane, N.: A sequential particle algorithm that keeps the particle system alive. In: 13th European Signal Processing Conference. IEEE, pp 1–4 (2005)

Li, T., Bolić, M., Djurić, P.M.: Resampling methods for particle filtering. IEEE Signal Process. Mag. **32**(3), 70–86 (2015)

Liu, J.S., Chen, R., Wong, W.H.: Rejection control and sequential importance sampling. J. Am. Stat. Assoc. **93**(443), 1022–1031 (1998)

Martino, L., Read, J., Elvira, V., Louzada, F.: Cooperative parallel particle filters for online model selection and applications to urban mobility. Digit. Signal Process. **60**, 172–185 (2017)

Martino, L., Luengo, D., Miguez, J.: Independent Random Sampling Methods. Springer, Berlin (2018)

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

Moral, P.D., Jasra, A., Lee, A., Yau, C., Zhang, X.: The alive particle filter and its use in particle Markov chain Monte Carlo. Stoch. Anal. Appl. **33**(6), 943–974 (2015)

Soto, A.: Self adaptive particle filter. In: IJCAI, pp. 1398–1406 (2005)

Straka, O., Šimandl, M.: Particle filter adaptation based on efficient sample size. In: 14th IFAC Symposium on System Identification (2006)

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., Gelman, A.: Validating Bayesian inference algorithms with simulation-based calibration. arXiv:1804.06788 (2018)