Bayesian Conditional Monte Carlo Algorithms for non linear time-series state estimation

Yohan Petetin*, François Desbouvries, *Senior Member, IEEE*

Abstract—Bayesian filtering aims at estimating sequentially a hidden process from an observed one. In particular, sequential Monte Carlo (SMC) techniques propagate in time weighted trajectories which represent the posterior probability density function (pdf) of the hidden process given the available observations. On the other hand, Conditional Monte Carlo (CMC) is a variance reduction technique which replaces the estimator of a moment of interest by its conditional expectation given another variable. In this paper we show that up to some adaptations, one can make use of the time recursive nature of SMC algorithms in order to propose natural temporal CMC estimators of some point estimates of the hidden process, which outperform the associated crude Monte Carlo (MC) estimator whatever the number of samples. We next show that our Bayesian CMC estimators can be computed exactly, or approximated efficiently, in some hidden Markov chain (HMC) models; in some jump Markov state-space systems (JMSS); as well as in multitarget filtering. Finally our algorithms are validated via simulations.

Index Terms—Conditional Monte Carlo, Bayesian filtering, Hidden Markov Models, jump Markov state space systems, Rao-Blackwell particle filters, multi-object filtering, probability hypothesis density

I. INTRODUCTION

LET us first fix notations. Bold letters denote vectors; $p(x)$, say, denotes the pdf of random variable (r.v.) X and $p(x|y)$, say, the conditional pdf of X given $Y = y$; if $i \leq j$ $p_{i:j|n}$ is a shorthand notation for $p(\mathbf{x}_{i:j}|\mathbf{y}_{0:n})$; if x^i are samples from $p(x)$ then the set $\{x^i\}_{i=1}^N$ can also be denoted $\mathbf{x}^{1:N}$; subscripts are reserved for labels or times indices and superscripts for realizations. In Bayesian filtering we consider two random processes $\{X_n\}_{n \geq 0}$ and $\{Y_n\}_{n \geq 0}$ with given joint probability law. Y_i is observed, i.e. we have realizations $\mathbf{y}_{0:n} = \{y_i\}_{i=0}^n$ of $\mathbf{Y}_{0:n} = \{Y_i\}_{i=0}^n$. Process $\{X_n\}$ is hidden, and our aim is to compute, for each time instant n, some moment of interest

$$\Theta_n = \int f(\mathbf{x}_{0:n}) p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n}) \mathrm{d}\mathbf{x}_{0:n} \qquad (1)$$

of the a posteriori pdf $p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n})$ of $\mathbf{X}_{0:n}$ given $\mathbf{y}_{0:n}$.

Unfortunately, in most models (1) cannot be computed exactly. Suboptimal solutions for computing Θ_n include SMC techniques [1] [2], which propagate over time weighted trajectories $\{\mathbf{x}_{0:n}^i, w_n^i\}_{i=1}^N$ with $\sum_{i=1}^N w_n^i = 1$. In other words, $\widehat{p}_{0:n|n} = \sum_{i=1}^N w_n^i \delta_{\mathbf{x}_{0:n}^i}$, in which δ is the Dirac mass, is a

Yohan Petetin is with CEA LIST, Route Nationale, 91400 Gif-sur-Yvette, France. François Desbouvries is with Mines Telecom Institute, Telecom SudParis, CITI Department, 9 rue Charles Fourier, 91011 Evry, France and with CNRS UMR 5157.

discrete (and random) approximation of $p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n})$, and Θ_n is finally estimated as

$$\widehat{\Theta}_n = \sum_{i=1}^N w_n^i f(\mathbf{x}_{0:n}^i). \qquad (2)$$

In this paper we do not discuss the choice of the SMC algorithm, but rather directly focus on $\widehat{\Theta}_n$, and see under which conditions one can improve this point estimator at a reasonable computational cost.

This leads us to variance reduction techniques and more specifically to Rao-Blackwellization. Roughly speaking, the rationale is as follows. Assume that the multiple integral (1) can be split in two parts, and the inner integral can be computed exactly, so that only the outer one needs to be approximated. Then one can reduce the dimension of the MC samples, and as a benefit the corresponding estimator $\widehat{\Theta}^{\mathrm{RB}}$ will have a reduced variance.

In Bayesian filtering the idea is either known as marginalized or Rao-Blackwellized Particle Filters (RB-PF) and has been applied so far to a spatial partition of the state vectors [3]–[7]; by spatial we mean that each vector $\mathbf{x}_k$ is split as $\mathbf{x}_k = [(\mathbf{x}_k^l)^T (\mathbf{x}_k^{nl})^T]^T$ (where $\mathbf{x}_k^l$ and $\mathbf{x}_k^{nl}$ are respectively the "linear" and "non-linear" components of vector $\mathbf{x}_k$), so that matrix $[\mathbf{x}_0, \cdots, \mathbf{x}_n]$ is partitioned into two block rows.

In this paper we propose another class of RB-PF; the main difference is that our partitioning of $\mathbf{x}_{0:n}$ is now temporal rather than spatial, i.e. that matrix $[\mathbf{x}_0, \cdots, \mathbf{x}_n]$ is split into two (block) columns, $[\mathbf{x}_0, \cdots, \mathbf{x}_{n-1}]$ and $\mathbf{x}_n$. Strangely enough, such a decomposition has not been considered before, whereas it arises naturally in the Bayesian filtering context: at time n we usually build $\widehat{\Theta}_n$ from $\widehat{p}_{0:n|n}$, but indeed $\widehat{p}_{0:n-1|n-1}$ was also available for free since, by nature, *sequential* MC algorithms construct $\widehat{p}_{0:n|n}$ *from* $\widehat{p}_{0:n-1|n-1}$. As we shall see in Section II-C, making use of $\widehat{p}_{0:n-1|n-1}$ enables us to propose a CMC estimator of moment (1). The rest of this paper is organized as follows. Section II describes our methodology. We first briefly recall the principles of variance reduction via conditioning and next discuss spatial vs. temporal RB estimators, as well as computational aspects; in particular, we recall the principle of local MC computations. We next identify Markovian models where our CMC estimator can be computed efficiently enough. So in section III we first consider HMC models, and we see that a Bayesian CMC estimator can be computed directly in some of them, or approximated otherwise. In section IV we develop our Bayesian CMC estimators for JMSS; particularly, the specificities of these models for CMC estimates are discussed in section IV-B. Finally in section V

we adapt Bayesian CMC from single- to multi-target scenarios. We end the paper with a conclusion.

II. SPATIAL VS. TEMPORAL RB-PF FOR BAYESIAN FILTERING

A. Rao-Blackwellization : tradeoff between variance reduction and computational complexity

Let us first recall the well known following result. Let X and Y be any two r.v. The law of iterated expectations and the law of total variance (sometimes called Rao-Blackwellization [8] [9], even though the conditioning is not always in terms of a sufficient statistic) respectively read

$$\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y), \quad (3)$$
$$\mathrm{var}(\mathrm{E}(Y|X)) = \mathrm{var}(Y) - \mathrm{E}(\mathrm{var}(Y|X)). \quad (4)$$

As a toy application of (3) (4) let us address the generic problem of computing via MC simulations the *partially computable* integral

$$\Theta = \int f(x_1, x_2) p(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 \quad (5)$$
$$= \int \left[\int f(x_1, x_2) p(x_2|x_1) \mathrm{d}x_2 \right] p(x_1) \mathrm{d}x_1; \quad (6)$$

by partially computable, we mean that neither (5) nor the outer integral in (6) can be computed exactly, but that the inner integral in (6) can. Let $\{(X_1^i, X_2^i)\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} p(x_1, x_2)$. From (3)-(4) (applied with $Y = f(X_1, X_2)$ and $X = X_1$) the so-called CMC estimator $\widetilde{\Theta} = \frac{1}{N} \sum_{i=1}^N \mathrm{E}(f(X_1^i, X_2)|X_1^i)$ has the same mean but lower variance than the crude MC one $\widehat{\Theta} = \frac{1}{N} \sum_{i=1}^N f(X_1^i, X_2^i)$. In other words, for estimating moment (5) it is preferable to compute whatever can be computed, and use MC simulations only when necessary.

Of course, the advantage of $\widetilde{\Theta}$ over $\widehat{\Theta}$ depends on X_1; ideally, one should easily sample from $p(x_1)$ (the choice of a "good" variable X_1 is not necessarily straightforward if $f(x_1, x_2)$ depends on x_2 only); the variance reduction in (4) should be as large as possible; but in the meantime function $g(x_1) = \mathrm{E}(f(X_1, X_2)|x_1)$ should remain computable at a reasonable computational cost. As far as these last two points are concerned, when partitioning as (X_1, X_2) a given set of variables $(\mathbf{X}_{0:n}$, say), the set X_1 should be as small as possible. More precisely, let $\Theta = \mathrm{E}(f(X_1, X_2, X_3))$ and let $\{(X_1^i, X_2^i)\}_{i=1}^N \overset{\text{i.i.d.}}{\sim} p(x_1, x_2)$. Then two Bayesian CMC estimators can be thought of: $\widetilde{\Theta}^{X_3}$ built from $\Theta = \mathrm{E}[\mathrm{E}(f(X_1, X_2, X_3)|X_1, X_2)]$, in which the inner expectation (w.r.t. X_3) is computed exactly; and $\widetilde{\Theta}^{(X_2, X_3)}$ built from $\Theta = \mathrm{E}[\mathrm{E}(f(X_1, X_2, X_3)|X_1)]$ and from $\widehat{p}(x_1)$. Estimator $\widetilde{\Theta}^{(X_2, X_3)}$ is preferable to $\widetilde{\Theta}^{X_3}$, but computing $\widetilde{\Theta}^{(X_2, X_3)}$ requires an additional expectation computation, since $\mathrm{E}(f(X_1, X_2, X_3)|X_1) = \mathrm{E}[\mathrm{E}(f(X_1, X_2, X_3)|X_1, X_2)]$.

B. Spatial RB-PF for Bayesian filtering

Variance reduction techniques based on (4) have been adapted to Bayesian filtering [3] [4] [5]. The aim is to compute (1), i.e. generically integral (5), but the difference with section II-A is that it is usually not possible to sample from $p(x_1, x_2)$,

and often $p(x_1, x_2) \propto p'(x_1, x_2)$ is only known up to a constant, whence the use of normalized importance sampling (IS) techniques [10]. So let now

$$\widehat{\Theta}(\mathbf{x}_1^{1:N}, \mathbf{x}_2^{1:N}) = \sum_{i=1}^N w_2^i(\mathbf{x}_1^{1:N}, \mathbf{x}_2^{1:N}) f(x_1^i, x_2^i) \text{ with } (x_1^i, x_2^i) \sim q_2, \quad (7)$$

$$\widetilde{\Theta}^{\mathrm{RB}}(\mathbf{x}_1^{1:N}) = \sum_{i=1}^N w_1^i(\mathbf{x}_1^{1:N}) \mathrm{E}(f(x_1^i, X_2)|x_1^i) \text{ with } x_1^i \sim q_1, \quad (8)$$

with $\sum_{i=1}^N w_1^i = \sum_{i=1}^N w_2^i = 1$. Estimator $\widetilde{\Theta}^{\mathrm{RB}}$ depends on samples $\{x_1^i\}_{i=1}^N$ only and is known as the RB estimator of Θ. However $\widetilde{\Theta}^{\mathrm{RB}}$ is known to outperform $\widehat{\Theta}$ only under specific assumptions on q_1, q_2, $\mathbf{w}_1^{1:N}$ and $\mathbf{w}_2^{1:N}$. In particular, if $w_1^i \propto w_1^{\mathrm{u},i} = p'(x_1^i)/q_1(x_1^i)$, $w_2^i \propto w_2^{\mathrm{u},i} = p'(x_1^i, x_2^i)/q_2(x_1^i, x_2^i)$ and $q_1(x_1) = \int q_2(x_1, x_2) \mathrm{d}x_2$, then the variance of $w_1^{\mathrm{u},i}$ can only be lower than that of $w_2^{\mathrm{u},i}$ [11]. If moreover (x_1^i, x_2^i) are independent, an asymptotic analysis based on (3) and (4) proves that $\widetilde{\Theta}^{\mathrm{RB}}$ indeed outperforms $\widehat{\Theta}$ [4]. However, independence never holds in the presence of resampling; in the general case, the comparison of both estimators depends on the choice of the importance distributions q_1 and q_2, and can be proved (asymptotically) only under specific sufficient conditions [12] [13].

RB-PF have been applied to the specific case where the state vectors $\mathbf{x}_{0:n}$ can be partitioned into a "linear" component $\mathbf{x}_2 = \mathbf{x}_{0:n}^l$ and a "non-linear" one $\mathbf{x}_1 = \mathbf{x}_{0:n}^{nl}$. Models in which $\widetilde{\Theta}^{\mathrm{RB}}$ can be computed exactly include linear and Gaussian JMSS [3] [4] or partially linear and Gaussian HMC [5]. In other models, it may be possible to approximate $\widetilde{\Theta}^{\mathrm{RB}}$ by using numerical approximations of $w_1(x)$ and of $\mathrm{E}(f(X_1, X_2)|x_1)$; however, due to the spatial structure of the decomposition of $\mathbf{x}_{0:n}$, approximating $\widetilde{\Theta}^{\mathrm{RB}}$ in (8) involves propagating numerical approximations over time. Finally, recent contributions [6] [7] propose to approximate the integral in (8) via a local MC method which also leads to an approximation of the importance weights $w_1^i(\mathbf{x}_1^{1:N})$ in (8).

C. Temporal RB-PF for Bayesian filtering

In this paper we propose another class of RB-PF; the main difference is that our partitioning (X_1, X_2) of $\mathbf{x}_{0:n}$ is now temporal rather than spatial (which, by contrast with spatially partitioned RB-PF, induces that the state space no longer needs to be multi-dimensional). Remember that the aim is to compute (5), or equivalently (6). Let us start from the following approximation of $p(x_1)$:

$$p(x_1) \approx \widehat{p}(x_1) = \sum_{i=1}^N w^i(\mathbf{x}_1^{1:N}) \delta_{x_1^i}. \quad (9)$$

For $1 \leq i \leq N$, let us assume that we can sample $x_2^i \sim p(x_2|x_1^i)$. Then

$$\widehat{p}(x_1, x_2) = \sum_{i=1}^N w^i(\mathbf{x}_1^{1:N}) \delta_{(x_1^i, x_2^i)} \quad (10)$$

(with unchanged weights) can be used as an approximation of $p(x_1, x_2)$. Next we have two options: computing the full expectation in (5) by using (10), or only the outer one in (6) by using (9). So let

$$
\widehat{\Theta}(\mathbf{x}_1^{1:N}, \mathbf{x}_2^{1:N}) = \sum_{i=1}^{N} w^i(\mathbf{x}_1^{1:N}) f(x_1^i, x_2^i), \tag{11}
$$

$$
\widetilde{\Theta}(\mathbf{x}_1^{1:N}) = \sum_{i=1}^{N} w^i(\mathbf{x}_1^{1:N}) \mathrm{E}(f(x_1^i, X_2)|x_1^i). \tag{12}
$$

In this paper, we shall call $\widehat{\Theta}(\mathbf{x}_1^{1:N}, \mathbf{x}_2^{1:N})$ (resp. $\widetilde{\Theta}(\mathbf{x}_1^{1:N})$) the Bayesian crude MC (resp. Bayesian CMC) estimator of Θ. Let us now compare $\widetilde{\Theta}$ to $\widehat{\Theta}$. Here each weight w^i may depend on $\{x_1^i\}_{i=1}^{N}$, but not on $\{x_2^i\}_{i=1}^{N}$; as a consequence $\widetilde{\Theta} = \mathrm{E}(\widehat{\Theta}|\mathbf{x}_1^{1:N})$, which induces that $\widetilde{\Theta}$ outperforms $\widehat{\Theta}$ (due to (3) and (4)), whatever the number N of particles, and whatever the MC approximation $\widehat{p}(x_1)$ in (9).

Let us comment this result. By contrast with section II-B, comparing $\widetilde{\Theta}$ to $\widehat{\Theta}$ is easy here, because weights w^i of the crude estimator (11) no longer depend on $\{x_2^i\}_{i=1}^{N}$ (compare (7)-(8) to (11)-(12)). This, in turn, is rendered possible because we assumed that one can sample from $p(x_2|x_1^i)$. Observe that this is not the case if the partitionning is spatial, at least if $\mathbf{x}_{0:n}$ are partitioned as a "linear" component $\mathbf{x}_2 = \mathbf{x}_{0:n}^l$ and a "non-linear" one $\mathbf{x}_1 = \mathbf{x}_{0:n}^{nl}$ with $\mathbf{x}_k = [(\mathbf{x}_k^l)^T (\mathbf{x}_k^{nl})^T]^T$. The reason why is that the generic conditional pdf $p(x_2|x_1)$ at time n reads $p(\mathbf{x}_{0:n}^l|\mathbf{x}_{0:n}^{nl}, \mathbf{y}_{0:n})$; so even in a model in which (theoretically) one could sample from $p(x_2|x_1)$, it is impossible in practice because each sample $x_2^i \sim p(x_2|x_1^i)$ is now a whole *new* trajectory $\mathbf{x}_{0:n}^{l,i}$, with dimension proportional to n (unfortunately sequentiality does not hold so we cannot just extend $\mathbf{x}_{0:n-1}^{l,i}$). As we shall see in sections III to V, we will consider the generic partitioning $(X_1, X_2) = (\mathbf{X}_{0:n-1}, \mathbf{X}_n)$; so the dimension of $x_2^i \sim p(x_2|x_1^i)$ is independent of n, and sampling from $p(x_2|x_1^i)$ is conceivable. It then remains to look for models where sampling from $p(x_2|x_1^i)$ is possible in practice.

D. Practical considerations

From section II-C, $\widetilde{\Theta}$ outperforms $\widehat{\Theta}$; so in the rest of this paper we look for Markovian models where $\widetilde{\Theta}$ can be computed efficiently. To that respect, a key point is the availability of an approximation $\widehat{p}(x_1)$ of $p(x_1)$ (which can be a problem in itself - see e.g. section III-B), as well as the integrability of $f(x_1, x_2)p(x_2|x_1)$. However as we shall see integral $\int f(x_1, x_2)p(x_2|x_1)\mathrm{d}x_2$ can be computed exactly only in a restrictive class of models. In order to widen the applicability of CMC estimators we thus need to consider models where (12) can be implemented approximately.

Implementations based on numerical approximations may rapidly prove unsatisfying, in particular in models with severe nonlinearities. Then one can use a local MC implementation of (12) The principle is as follows. Assume that neither the importance weights associated to $\widehat{p}(x_1)$ nor integral $\int f(x_1, x_2) p(x_2|x_1)\mathrm{d}x_2$ are computable. In such cases one can approximate the integral via a local MC method which also leads to an approximation of the importance

weights necessary in the computation of $\widehat{p}(x_1)$. In particular, if $p(x_1)p(x_2|x_1)=p(x_1, x_2)$ is known, then for a given x_1^i, a normalized IS technique which produces samples $\{x_2^{i,j}\}_{j=1}^{M}$ and which approximates a moment according to $p(x_2|x_1^i)$ (which is known up to a constant) also provides an unbiased estimator of $p(x_1^i)$. Thus weights $\{w_1^i\}_{i=1}^{N}$ associated to the approximation of $p(x_1)$ (which rely on $\{p(x_1^i)\}_{i=1}^{N}$) can be approximated by some weights $\{\widetilde{w}_1^i\}_{i=1}^{N}$, in which $\widetilde{w}_1^i$ depends on $\{x_2^{i,j}\}_{j=1}^{M}$. Now, in a sequential filtering context, $p(x_1, x_2)$ is typically only known up to a constant; however a sequential application of this technique ensures that the (marginal) target distribution deduced from the approximated weights is still $p(x_1)$. The idea has already been applied to spatial RB [6] [7] and, as we shall see (see in particular section III-D), can be adapted to the CMC problem discussed in this paper.

III. BAYESIAN CMC PF FOR SOME HMC MODELS

A. Deriving a Bayesian CMC estimator $\widetilde{\Theta}_n$

Let $\{\mathbf{X}_n\}_{n\geq 0}$ (resp. $\{\mathbf{Y}_n\}_{n\geq 0}$) be a p- (resp. q-) dimensional state vector (resp. observation). We assume that $(\mathbf{X}_n, \mathbf{Y}_n)$ follows the well known HMC model:

$$
p(\mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(\mathbf{x}_0) \prod_{i=1}^{n} f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}) \prod_{i=0}^{n} g_i(\mathbf{y}_i|\mathbf{x}_i),
$$
$$\tag{13}$$

in which $f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1})$ is the transition pdf of Markov chain $\{\mathbf{X}_n\}_{n\geq 0}$ and $g_i(\mathbf{y}_i|\mathbf{x}_i)$ the likelihood. We consider the problem of computing (1), and in particular moment $\Theta_n = \mathrm{E}_{p_{n|n}}(f(\mathbf{X}_n))$, which we rewrite as

$$
\Theta_n = \int \underbrace{f(\mathbf{x}_n)}_{f(x_2)} \underbrace{p(\mathbf{x}_{0:n-1}, \mathbf{x}_n|\mathbf{y}_{0:n})}_{p(x_1, x_2)} \underbrace{\mathrm{d}\,\mathbf{x}_{0:n-1}}_{x_1} \underbrace{\mathrm{d}\,\mathbf{x}_n}_{x_2} \tag{14}
$$

So (14) coincides with (5), with $X_1 = \mathbf{X}_{0:n-1}$, $X_2 = \mathbf{X}_n$, $f(x_1, x_2)$ depends on x_2 only, and $p(x_1, x_2)$ is the joint pdf

$$
p(\mathbf{x}_{0:n-1}, \mathbf{x}_n|\mathbf{y}_{0:n}) = \underbrace{p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})}_{p(x_1)} \underbrace{p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)}_{p(x_2|x_1)}.
$$

Let $\widehat{p}(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i \delta_{\mathbf{x}_{0:n-1}^i}$ be an approximation of $p(x_1) = p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ and $\widetilde{\mathbf{x}}_n^i \sim p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)$, for $1 \leq i \leq N$. From (11) and (12), the crude and CMC estimators of moment Θ_n defined in (14) are respectively

$$
\widehat{\Theta}_n(\mathbf{x}_{0:n-1}^{1,N}, \widetilde{\mathbf{x}}_n^{1:N}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i(\mathbf{x}_{0:n-1}^{1:N}) f(\widetilde{\mathbf{x}}_n^i), \tag{15}
$$

$$
\widetilde{\Theta}_n(\mathbf{x}_{0:n-1}^{1:N}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i(\mathbf{x}_{0:n-1}^{1:N}) \int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n.
$$
$$\tag{16}$$

B. SMC algorithms

Let us now address computational aspects. Contrary to most SMC algorithms which focus on an MC approximation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n-1})$, here we first need an approximation of the smoothing pdf $p(x_1) = p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$. Let us discuss the different solutions.

1) Smoothing techniques:

a) Smoothing via filtering: First, the propagation of an MC approximation $\{\mathbf{x}_{0:n-1}^i, w_{n-1}^i\}_{i=1}^N$ of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n-1})$ is a well known problem which has been studied in many contributions. SMC algorithms mainly include the sampling importance resampling (SIR) algorithms and the class of Auxiliary Particle Filtering (APF) ones [14], which rely on an importance distribution $q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}, \mathbf{y}_{n-1})$. Once an approximation $\{\mathbf{x}_{0:n-1}^i, w_{n-1}^i\}_{i=1}^N$ of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n-1})$ is obtained, an MC approximation $\{\mathbf{x}_{0:n-1}^i, \widetilde{w}_{n-1}^i\}_{i=1}^N$ of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ is deduced by setting

$$\widetilde{w}_{n-1}^i \propto w_{n-1}^i p(\mathbf{y}_n|\mathbf{x}_{n-1}^i); \tag{17}$$

b) Direct smoothing: On the other hand, one can directly focus on an MC approximation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ (actually, we do not need an approximation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n-1})$ in our formulation) via

$$p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n}) \propto p(\mathbf{x}_{0:n-2}|\mathbf{y}_{0:n-1}) \times$$
$$\frac{f_{n-1|n-2}(\mathbf{x}_{n-1}|\mathbf{x}_{n-2})g_{n-1}(\mathbf{y}_{n-1}|\mathbf{x}_{n-1})}{p(\mathbf{y}_{n-1}|\mathbf{x}_{n-2})}p(\mathbf{y}_n|\mathbf{x}_{n-1}). \tag{18}$$

Starting from an MC approximation $\{\mathbf{x}_{0:n-2}^i, \widetilde{w}_{n-2}^i\}_{i=1}^N$ of $p(\mathbf{x}_{0:n-2}|\mathbf{y}_{0:n-1})$, and given an importance distribution $q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})$, weight $\widetilde{w}_{n-1}^i$ associated to $\mathbf{x}_{n-1}^i \sim q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})$ reads

$$\widetilde{w}_{n-1}^i \propto \widetilde{w}_{n-2}^i \frac{f_{n-1|n-2}(\mathbf{x}_{n-1}^i|\mathbf{x}_{n-2}^i)g_{n-1}(\mathbf{y}_{n-1}|\mathbf{x}_{n-1}^i)}{q(\mathbf{x}_{n-1}^i|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})p(\mathbf{y}_{n-1}|\mathbf{x}_{n-2}^i)} \times$$
$$p(\mathbf{y}_n|\mathbf{x}_{n-1}^i). \tag{19}$$

Note that a key point of this formulation is that importance distribution $q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})$ can depend on the new observation $\mathbf{y}_n$, which leads to the consideration of an importance distribution more adapted to the smoothing problem;

c) Alternate smoothers: finally, note that several smoothing algorithms have been proposed to obtain robust MC approximations of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ [15] [16] [17]. Roughly speaking, most of these solutions start by getting an MC approximation $\{\mathbf{x}_{0:n}, w_n\}_{i=1}^N$ of $p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n})$ via a classical filtering SMC algorithm, then use an additional sampling mechanism in order to obtain a better MC approximation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$. Even though such strategies may give accurate approximations of a moment according to the smoothing distribution $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$, they will not be used in this paper; the reason why is discussed below (see §III-B2).

2) Discussion: Let us now discuss on the relevance of these three classes of smoothers to our problem. Even though approximation $\widehat{p}(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ impacts on estimator $\widetilde{\Theta}_n$, remember from section II-C that $\widetilde{\Theta}_n$ will outperform $\widehat{\Theta}_n$ whatever $\widehat{p}(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$; also keep in mind that our final goal is to compute estimator (16) in an efficient way. So let us first observe that computing estimator (16) involves pdf $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$, which happens to coincide with the so-called optimal (in the sense that it minimizes the variance of the importance weights) importance distribution of SIR *filtering* algorithms [18] [19] [11]. In addition, (17) and (19) in the first two points above highlight the role of $p(\mathbf{y}_n|\mathbf{x}_{n-1}) = \int f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1})\, g_n(\mathbf{y}_n|\mathbf{x}_n)\mathrm{d}\mathbf{x}_n$, which is related to $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n) = f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1})g_n(\mathbf{y}_n|\mathbf{x}_n)/$

$p(\mathbf{y}_n|\mathbf{x}_{n-1})$; it may thus be of practical interest to involve function $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$ in the sampling step of new particles. In addition, $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$ is the optimal conditional importance distribution in the filtering problem, but also plays a role in the one step backward smoothing one, since the optimal importance distribution for the smoothing problem (that which minimizes the conditional variance of weights $\widetilde{w}_{n-1}^i$ in (19)) is given by

$$q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}, \mathbf{y}_{n-1:n}) \propto p(\mathbf{x}_{n-1}|\mathbf{x}_{n-2}, \mathbf{y}_{n-1})p(\mathbf{y}_n|\mathbf{x}_{n-1}). \tag{20}$$

Finally, the algorithms [15] [16] [17] mentioned in point c) above focus on the computation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$, not on that of the integral in (16). Moreover, even if the integral in (16) can be computed exactly, the CMC estimate obtained from such a smoothing strategy does not necessarily outperform a crude estimate. As a simple illustration of this point, basic SMC algorithms provide an approximation of $p_{0:n|n}$ and so of $p_{0:n-1|n}$ given by $\widehat{p}_{0:n-1|n} = \sum w_n^i \delta_{\mathbf{x}_{0:n-1}^i}$. However weights w_n^i in general depend on the particles $\{\mathbf{x}_n^i\}_{i=1}^N$ which have just been sampled. In that case, the associated Bayesian CMC and crude estimators become

$$\widehat{\Theta}_n(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{x}_n^{1:N}) = \sum_{i=1}^N w_n^i(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{x}_n^{1:N})f(\mathbf{x}_n^i), \tag{21}$$

$$\widetilde{\Theta}_n(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{x}_n^{1:N}) = \sum_{i=1}^N w_n^i(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{x}_n^{1:N})\int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n, \tag{22}$$

which can no longer be compared easily (see the end of section II-C). If more sophisticated smoothers were used [15] [16] [17], $\widetilde{\Theta}_n$ in (22) could depend on other variables which would render the comparison more difficult. So from now on, we will only use the first two classes of smoothers for computing $p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n})$. We next focus on the implementation of $\widetilde{\Theta}_n$ in HMC models.

C. Computing $\widetilde{\Theta}_n$ exactly or via numerical approximations

Strictly speaking, $\widetilde{\Theta}_n$ can be computed only if $p(\mathbf{y}_n|\mathbf{x}_{n-1})$ is available (for the computation of weights $\widetilde{w}_{n-1}^i$, via (17) or (19)) and one can compute $\int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n$. As we now see, this is the case in some HMC models and for some functions $f(.)$. Let us e.g. consider the semi-linear stochastic models with additive Gaussian noise, given by $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) = \mathcal{N}(\mathbf{x}_n, \mathbf{f}_n(\mathbf{x}_{n-1}), \mathbf{Q}_n(\mathbf{x}_{n-1}))$ (where $\mathbf{f}_n$ is arbitrary) and $g_n(\mathbf{y}_n|\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_n, \mathbf{H}_n\mathbf{x}_n, \mathbf{R}_n^v)$. Such models can indeed be seen as the temporal equivalent of the partially linear and Gaussian model in which the spatial RB estimate [5] can be computed exactly. Then $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n) = \mathcal{N}(\mathbf{x}_n, \mathbf{m}_n(\mathbf{x}_{n-1}, \mathbf{y}_n), \mathbf{P}_n(\mathbf{x}_{n-1}))$ and $p(\mathbf{y}_n|\mathbf{x}_{n-1}) = \mathcal{N}(\mathbf{y}_n, \mathbf{H}_n\mathbf{f}_n(\mathbf{x}_{n-1}), \mathbf{L}_n(\mathbf{x}_{n-1}))$, in which parameters $\mathbf{m}_n(.)$, $\mathbf{P}_n(.)$ and $\mathbf{L}_n(.)$ can be computed exactly. So $\widetilde{\Theta}_n$ is computable for some functions $f(.)$. If $f(\mathbf{x})$ is a polynomial in $\mathbf{x}$, the problem reduces to computing the first moments of $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$. In the important particular case where $f(\mathbf{x}) = \mathbf{x}$, no further computation is indeed necessary; in this case the ith integral in (16)

is equal to $\mathbf{m}_n(\mathbf{x}_{n-1}^i, \mathbf{y}_n)$. Note also that in that case, computing $\widehat{\Theta}_n$ or $\widetilde{\Theta}_n$ indeed requires the same computational cost, because both estimators compute pdfs $p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)$, and use them to sample the new particles $\widetilde{\mathbf{x}}_n^i$ which, in both cases, are needed for the next time step. The only difference is that $\widehat{\Theta}_n = \sum_{i=1}^N \widetilde{w}_{n-1}^i \widetilde{\mathbf{x}}_n^i$, while $\widetilde{\Theta}_n = \sum_{i=1}^N \widetilde{w}_{n-1}^i \mathbf{m}_n(\mathbf{x}_{n-1}^i, \mathbf{y}_n)$. Finally, since $p(\mathbf{y}_n|\mathbf{x}_{n-1})$ and $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$ are available in these models, an approximation of the optimal distribution $p(\mathbf{x}_{n-1}|\mathbf{x}_{n-2}, \mathbf{y}_{n-1}, \mathbf{y}_n) \propto p(\mathbf{x}_{n-1}|\mathbf{x}_{n-2}, \mathbf{y}_{n-1})p(\mathbf{y}_n|\mathbf{x}_{n-1})$ for the smoothing problem can be obtained via an adaptation of classical linearizations/Unscented Transformation (UT) techniques [11] [20].

Now, if $p(\mathbf{y}_n|\mathbf{x}_{n-1})$ and/or moments of $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$ are not computable, local numerical approximations such as local linearizations [11], Taylor series expansion [20] or the UT [21] can be used in (16) and (17). By contrast with the spatial case, an interesting feature of temporal RB-PF is that such numerical approximations no longer need to be propagated over time. On the other hand, these techniques indeed compute an approximation of function $p(\mathbf{y}_n|\mathbf{x}_{n-1})$ (used to deduce an MC approximation of $p_{0:n-1|n}$) and can be very poor since it is built from an approximation of its first order moments.

D. A general implementation of $\widetilde{\Theta}_n$ based on local MC techniques

1) Application: We now discuss on the adaptation of the local MC method described in Section II-D. Applying this methodology, we approximate simultaneously the integral in (16) and the weights in (17) or (19). For a given trajectory $\mathbf{x}_{0:n-1}^i$, let us sample M particles $\mathbf{x}_n^{i,j} \sim q^{\text{CMC}}(\mathbf{x}_n|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n)$ (superscript CMC indicates that this importance distribution is used to perform the CMC step); we thus obtain an estimator of each integral in (16) and an unbiased estimator of each $p(\mathbf{y}_n|\mathbf{x}_{n-1}^i)$:

$$\int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n \approx \sum_{j=1}^M w_n^{\text{CMC},i,j} f(\mathbf{x}_n^{i,j}),$$

$$w_n^{\text{CMC},i,j} \propto \frac{f_{n|n-1}(\mathbf{x}_n^{i,j}|\mathbf{x}_{n-1}^i)g_n(\mathbf{y}_n|\mathbf{x}_n^{i,j})}{q^{\text{CMC}}(\mathbf{x}_n^{i,j}|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n)}, \sum_{j=1}^M w_n^{\text{CMC},i,j} = 1,$$

$$\widehat{p}(\mathbf{y}_n|\mathbf{x}_{n-1}^i) = \frac{1}{M} \sum_{j=1}^M \frac{f_{n|n-1}(\mathbf{x}_n^{i,j}|\mathbf{x}_{n-1}^i)g_n(\mathbf{y}_n|\mathbf{x}_n^{i,j})}{q^{\text{CMC}}(\mathbf{x}_n^{i,j}|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n)}. \quad (23)$$

Finally, weights in (17) are approximated by

$$\widetilde{w}_{n-1}^i \propto w_{n-1}^i \widehat{p}(\mathbf{y}_n|\mathbf{x}_{n-1}^i), \quad (24)$$

and weights in (19) by

$$\widetilde{w}_{n-1}^i \propto \widetilde{w}_{n-2}^i \frac{f_{n-1|n-2}(\mathbf{x}_{n-1}^i|\mathbf{x}_{n-2}^i)g_{n-1}(\mathbf{y}_{n-1}|\mathbf{x}_{n-1}^i)}{q(\mathbf{x}_{n-1}^i|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})\widehat{p}(\mathbf{y}_{n-1}|\mathbf{x}_{n-2}^i)} \times \widehat{p}(\mathbf{y}_n|\mathbf{x}_{n-1}^i), \quad (25)$$

where $q(\mathbf{x}_{n-1}|\mathbf{x}_{0:n-2}^i, \mathbf{y}_{n-1:n})$ is the importance distribution used to obtain samples $\{\mathbf{x}_{n-1}^i\}_{i=1}^N$. Our approximated Bayesian CMC estimator is now

$$\widetilde{\Theta}_n = \sum_{i=1}^N \widetilde{w}_{n-1}^i \left[\sum_{j=1}^M w_n^{\text{CMC},i,j} f(\mathbf{x}_n^{i,j}) \right]. \quad (26)$$

Note that one can check easily that the distribution targeted by weights $\widehat{w}(\mathbf{x}_{0:n-1}, \mathbf{x}_n^{\text{CMC},1:M})$ in (25) is $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ since

$$\widehat{p}(\mathbf{y}_n|\mathbf{x}_{n-1}) = \frac{1}{M} \sum_{j=1}^M \frac{f_{n|n-1}(\mathbf{x}_n^{:,j}|\mathbf{x}_{n-1})g_n(\mathbf{y}_n|\mathbf{x}_n^{:,j})}{q^{\text{CMC}}(\mathbf{x}_n^{:,j}|\mathbf{x}_{0:n-1}, \mathbf{y}_n)}$$

is an unbiased estimator of $p(\mathbf{y}_n|\mathbf{x}_{n-1})$.

2) Discussion: Of course, this local MC approximation requires an extra computational cost since for each particle $\mathbf{x}_{n-1}^i$, $i \in \{1, \cdots, N\}$, we sample an additional set of M particles $\mathbf{x}_n^{i,j}$, $j \in \{1, \cdots, M\}$. However, we expect that using a temporal RB approach will enable us to decrease the size of the MC approximation of $p_{0:n-1|n}$. This point will be particularly discussed in our simulations.

On the other hand, the computational cost due to the local sampling step can be controlled in another way, and indeed let us observe that the PF proposed in [22] can be seen as an approximated temporal RB estimate in which the local computations are done with N particles only. More precisely, instead of using a particular set $\{\mathbf{x}_n^{i,j}\}_{j=1}^M$ for each trajectory $\mathbf{x}_{0:n-1}^i$, one can first extend each trajectory with a set of particles $\{\mathbf{x}_n^i\}_{i=1}^N$ as in classical PF algorithms, then recycle these samples $\{\mathbf{x}_n^j\}_{j=1}^N$ in order to compute the local MC approximation of $p(\mathbf{y}_n|\mathbf{x}_{n-1}^i)$ and of $\int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n$. In other words, in this solution one samples a *unique* set of $M = N$ particles $\{\mathbf{x}_n^j\}_{j=1}^N$, rather than a set of $N \times M$ samples. Thus, the importance distribution $q^{\text{CMC}}(\mathbf{x}_n|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n)$ becomes

$$q^{\text{CMC}}(\mathbf{x}_n|\mathbf{x}_{0:n-1}^{1:N}, \mathbf{y}_n) = \sum_{i=1}^N w_{n-1}^i q(\mathbf{x}_n|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n). \quad (27)$$

Since the new set $\{\mathbf{x}_n^j\}_{j=1}^N$ is common to each sample $\mathbf{x}_{n-1}^i$, we easily check by plugging (24) in (26) that

$$\widetilde{\Theta}_n = \sum_{j=1}^N \overline{w}_{n-1}^j f(\mathbf{x}_n^j), \quad (28)$$

$$\overline{w}_{n-1}^j \propto g_n(\mathbf{y}_n|\mathbf{x}_n^j) \frac{\sum_{i=1}^N w_{n-1}^i f_{n|n-1}(\mathbf{x}_n^j|\mathbf{x}_{n-1}^i)}{\sum_{i=1}^N w_{n-1}^i q(\mathbf{x}_n^j|\mathbf{x}_{0:n-1}^i, \mathbf{y}_n)}. \quad (29)$$

Estimate (28) coincides with the estimate deduced from the Marginal PF proposed in [22]. The counterpart of this N samples technique is that it requires the computation of the N^2 weights $w_n^{\text{CMC},i,j}$. In addition, $\widetilde{\Theta}_n$ in (28) improves the classical SIR estimate only if the importance distribution $q^{\text{CMC}}(\mathbf{x}_n|\mathbf{x}_{0:n-1}^{1:N}, \mathbf{y}_n)$ is well chosen. For instance, in the simplest case where we set $q^{\text{CMC}}(\mathbf{x}_n|\mathbf{x}_{0:n-1}^{1:N}, \mathbf{y}_n) = \sum_{i=1}^N w_{n-1}^i f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}^i)$, weights $\overline{w}_{n-1}^j$ in (29) reduce to $\overline{w}_{n-1}^j \propto g_n(\mathbf{y}_n|\mathbf{x}_n^j)$, so $\widetilde{\Theta}_n$ in (28) reduces to the estimate based on the bootstrap filter [22].

E. Simulations

Let us now discuss the performances of our (possibly approximated) Bayesian CMC estimators via simulations in several HMC models. Our results are averaged on $P = 200$ simulations and we compute several criterions of performance.

First, the mean square error (MSE) w.r.t. the filtering mean $\mathrm{E}(f(\mathbf{x}_n)|\mathbf{y}_{0:n})$ is computed,

$$\mathrm{MSE}_n^f = \frac{1}{P}\sum_{i=1}^{P}(\widetilde{\Theta}_n^i - \mathrm{E}(f(\mathbf{X}_n)|\mathbf{y}_{0:n}))^2, \qquad (30)$$

where $\widetilde{\Theta}_n^i$ is an estimate of $\mathrm{E}(f(\mathbf{X}_n)|\mathbf{y}_{0:n})$ for the i-th simulation, and $\mathrm{E}(f(\mathbf{X}_n)|\mathbf{y}_{0:n})$ is computed via a classical PF (for example a bootstrap one [23]) with $N = 10^5$ particles. We also compute the MSE at time n w.r.t. the true state $f(\mathbf{x}_n)$, $\mathrm{MSE}_n^t = \frac{1}{P}\sum_{i=1}^{P}(\widetilde{\Theta}_n^i - f(\mathbf{x}_n))^2$, which will be averaged over time via criterion

$$\mathcal{J} = \frac{1}{T}\sum_{n=1}^{T}\sqrt{\mathrm{MSE}_n^t}, \qquad (31)$$

where T is the length of the considered scenario. In all our simulations, we will take $f(\mathbf{x}_n) = \mathbf{x}_n$ (i.e. we look for estimating the hidden state $\mathbf{X}_n$).

1) Semi-linear model: We start with the HMC model $f_{n|n-1}(x_n|x_{n-1}) = \mathcal{N}(x_n, \mathrm{atan}(x_{n-1}), Q)$ and $g_n(y_n|x_n) = \mathcal{N}(y_n, x_n, R)$. In this semi-linear model, the Fully Adapted (FA) algorithm can be implemented [14] since it is possible to sample according to $p(x_n|x_{n-1}, y_n)$ and to compute $p(y_n|x_{n-1})$. So we compare a crude estimate based on the FA algorithm with $N = 1000$; a CMC estimate (CMC-FA) in which $p_{0:n-1|n}$ is deduced from the FA algorithm with $N = 1000$ particles; and a second CMC estimate in which we directly propagate an MC approximation of $p_{0:n-1|n}$ via IS (see the second paragraph of III-B1). The chosen importance distribution approximates the optimal importance distribution for the smoothing problem (20) which is not exactly computable here. However, since $p(\mathbf{x}_{n-1}|\mathbf{x}_{n-2}, \mathbf{y}_{n-1})$ and $p(\mathbf{y}_n|\mathbf{x}_{n-1})$ are computable (see III-C), (20) can be approximated via the UT described in [24]. The specific parameters for the UT are $\beta = 2$, $\alpha = 0.5$ and $\kappa = 2$ (see e.g. [24, p. 8] for the meaning of these parameters).

Fig. 1 and Table I display the filtered MSE in (30) for $Q = 10$ and $R = 1$ for $N = 1000$ and criterion $\mathcal{J}$ in (31) in function of Q and R for $N = 50$ and $N = 1000$ particles, respectively. These results show that the CMC estimates outperform the estimate based on the FA algorithm and also highlight the role of the approximation of $p_{0:n-1|n}$. In particular, the CMC estimate based on the approximation of importance distribution (20) is preferable here. Let us now analyze the results described in Table I, which displays criterion $\mathcal{J}$. First, the CMC estimates outperform that based on the FA algorithm whatever the parameters Q and R of the model and the number of samples N. Of course, the gap between the crude and the CMC estimates decreases when we take a large number of samples. Now, if we compare both CMC estimates, that based on an approximation of importance distribution (20) outperforms that based on the FA algorithm when the observation noise variance R is large as compared to the process noise variance Q; remember that particles x_n^i sampled from (20) take into account y_n and y_{n+1}, which is critical here, because y_n is not informative enough. By contrast, if R is small then y_n is very close to x_n and taking into account y_{n+1} is not essential.
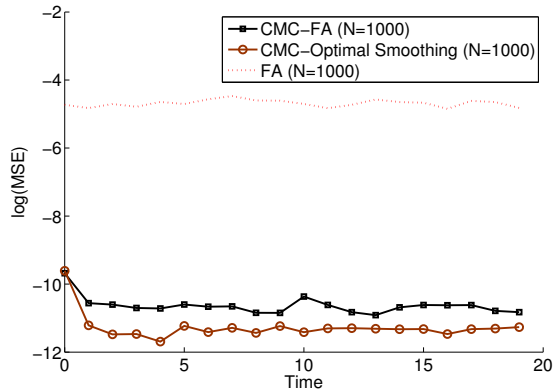

Fig. 1. MSE w.r.t to $\mathrm{E}(X_n|\mathbf{y}_{0:n})$. $f_{n|n-1}(x_n|x_{n-1}) = \mathcal{N}(x_n, \mathrm{atan}(x_{n-1}), Q)$, $g_n(y_n|x_n) = \mathcal{N}(y_n, x_n, R)$. CMC estimates are closer to the optimal estimate. Moreover using an importance distribution more adapted to the smoothing problem (i.e. approximating the optimal importance distribution $p(x_n|x_{n-1}, y_n, y_{n+1})$) can improve the natural temporal RB estimate deduced from the FA algorithm.

	FA	CMC-FA	CMC-opt
$Q = 0.1, R = 0.1$	0.2586 / 0.2423	0.2441 / 0.2411	0.2429 / 0.2411
$Q = 10, R = 0.1$	0.3200 / 0.3061	0.3039 / 0.3038	0.3039 / 0.3038
$Q = 0.1, R = 10$	0.6445 / 0.5775	0.6211 / 0.5732	0.5977 / 0.5707
$Q = 10, R = 1$	0.9833 / 0.9403	0.9362 / 0.9360	0.9360 / 0.9360
$Q = 1, R = 10$	1.1944 / 1.1294	1.1433 / 1.1273	1.1415 / 1.1226
$Q = 10, R = 10$	2.3960 / 2.2925	2.2856 / 2.2780	2.2846 / 2.2771

TABLE I
CRITERION $\mathcal{J}$ IN (31) FOR $f_{n|n-1}(x_n|x_{n-1}) = \mathcal{N}(x_n, \mathrm{atan}(x_{n-1}), Q)$, $g_n(y_n|x_n) = \mathcal{N}(y_n, x_n, R)$ IN FUNCTION OF Q AND R FOR $N = 50$ (LEFT) AND $N = 1000$ (RIGHT) PARTICLES. CMC ESTIMATES OUTPERFORM THE FA BASED ONE WHATEVER THE PARAMETERS, AND THE CHOICE OF THE IMPORTANCE DISTRIBUTION PLAYS A CRITICAL ROLE WHEN R IS LARGE.

2) Autoregressive Conditional Heteroscedasticity (ARCH) Model: We now consider the ARCH model used for financial datasets, $f_{n|n-1}(x_n|x_{n-1}) = \mathcal{N}(x_n, 0, \beta_0 + \beta_1 x_{n-1}^2)$, $g_n(y_n|x_n) = \mathcal{N}(y_n, x_n, R_n^v)$, and we set $R_n^v = 1$, $\beta_0 = 9$ and $\beta_1 = 3$. The realization we work on is displayed in Figure 2(a) and is characterized by calm and perturbed periods. Two crude and three CMC estimates are computed for this simulation. The crude ones are classical SMC estimates either based on the Bootstrap algorithm (particles are sampled according to $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1})$) or the FA one (particles are sampled according to $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$), and both use $N = 1000$ particles; the CMC estimates are based on the propagation of $p_{0:n-1|n-1}$ via the Bootstrap algorithm, with $N = 1000$ particles, or via the FA algorithm, with either $N = 1000$ or $N = 100$ particles.

Figure 2(b) displays the filtered MSE MSE_n^f for each estimate over time. First, the gain of the CMC estimates is larger in outlier periods (see for example times $n = 22$, $n = 40$, $n = 55$ and $n = 80$). Next, it is clear that the classical Bootstrap estimate is poor compared to the FA one. However, the CMC estimate which relies on a Bootstrap approximation of $p_{0:n-1|n-1}$ performs better than the classical FA-based estimate. It means that CMC estimates which are based on a filtering algorithm (see

the first paragraph of III-B1) can improve classical SMC ones, even when the approximation of $p_{0:n-1|0:n-1}$ is poor. This is confirmed by criterion $\mathcal{J}$ computed for each estimate: we get $\mathcal{J}(\text{Bootstrap}-1000) = 1.2965$, $\mathcal{J}(\text{FA}-1000) = 0.9965$, $\mathcal{J}(\text{CMC}-\text{Bootstrap}-1000) = 0.9707$, $\mathcal{J}(\text{CMC}-\text{FA}-100) = 0.9707$ and $\mathcal{J}(\text{CMC}-\text{FA}-100) = 0.9706$.

Finally, note that it is possible to decrease the size of the MC approximation used: the CMC-FA estimate with $N = 100$ particles outperforms the crude bootstrap and FA ones with $N = 1000$ particles, and within CMC algorithms, the CMC-FA estimate with $N = 100$ particles behaves similarly to the CMC-bootstrap one with $N = 1000$ particles. As we will see in the next section, this point is critical when we need to use additional MC approximations to compute similarly an MC approximation of $p_{0:n-1|n}$ and the integral in (16). Indeed, this simulation shows that it may be possible to decrease the size of the MC approximation $p_{0:n-1|n}$ without damaging the final estimate, when we use a CMC approach. Thus, in models where the additional local MC approximation developed in §III-D is needed, this saved computational cost can be used for the local MC approach .

3) Non linear and non gaussian model: We finally consider the non linear and non gaussian model $X_n = 1 + \sin(\pi\omega n) + \phi_1 X_{n-1} + U_n$, $Y_n = \phi_2 X_n^2 + V_n$ if $n \leq 30$ and $Y_n = \phi_3 X_n - 2 + V_n$ if $n > 30$, in which $\{U_n \sim \Gamma(3,2)\}$ ($\Gamma(.,.)$ stands for the Gamma distribution) and $\{V_n \sim \mathcal{N}(0, R)\}$ are i.i.d. and mutually independent, $R = 1$, $\omega = 4e-2$, $\phi_1 = 0.5$, $\phi_2 = 0.2$ and $\phi_3 = 0.5$ [24]. In this model $p(y_n|x_{n-1})$ is not available in closed form and it is not possible to compute exactly the integral in (16). So we have to resort to the local MC approximation of §III-D. Remember that we have to choose an importance distribution $q(x_n|\mathbf{x}_{0:n-1}, y_n)$ for getting an approximation of $p(\mathbf{x}_{0:n-1}|\mathbf{y}_{0:n})$ and another one $q^{\text{CMC}}(x_n|\mathbf{x}_{0:n-1}, y_n)$ to compute the importance weights and the integral in (16). We set $q(x_n|\mathbf{x}_{0:n-1}, \mathbf{y}_n) = q^{\text{CMC}}(x_n|x_{0:n-1}, \mathbf{y}_n) = f_{n|n-1}(x_n|x_{n-1})$ and (multinomial) resampling is done at each time step for all the algorithms. Of course, the use of such a local MC approximation increases the computational cost. So the objective of this simulation is to evaluate the role of N and M and the compromise between performances and computational cost. It is why we will take into account the computational time for the considered CMC estimates (simulations are executed with Matlab 2009). Note that the local MC approximation is adapted to parallel architectures, and that parallel implementations as those proposed in [6] would enable users to decrease the computational cost due to the local MC step. Our CMC estimates are compared with a classical SIR algorithm with $N = 1000$ particles which also use the transition pdf $f_{n|n-1}(x_n|x_{n-1})$ to obtain new samples. Several results are displayed:

First, Fig. 3(a) displays the filtered MSE (30) averaged over time w.r.t. the optimal estimate in function of N and M. Small values of M improve the classical SIR estimate, provided N is not too small. Indeed, if $N \geq 300$, $M = 10$ additional samples are sufficient to improve the MSE of the CMC estimate. By constrast, if N is small ($N = 100$), many local samples are necessary to approach the performances of the SIR estimate.
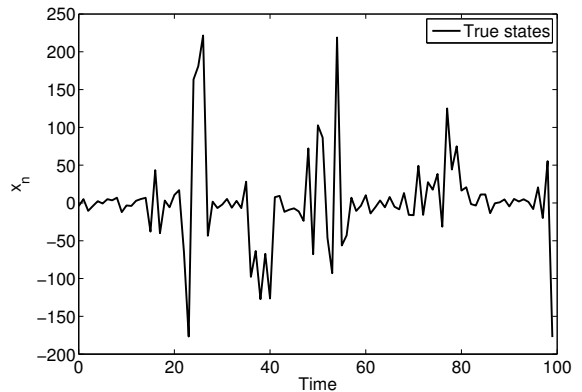
(a) ARCH scenario
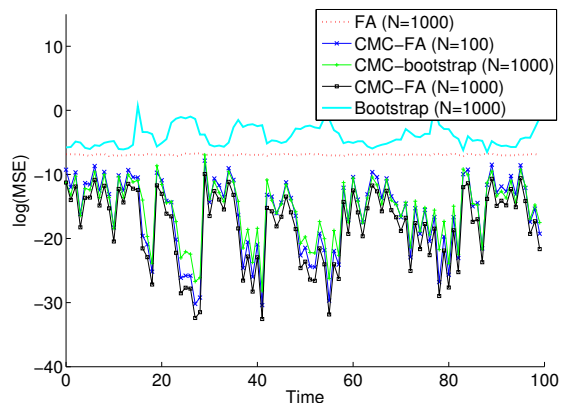
(b) MSE

Fig. 2. ARCH model - $\beta_0 = 9$, $\beta_1 = 3$ and $R_n^v = 1$ - (a) Considered scenario : alternance of calm and perturbed periods - (b) MSE w.r.t. $\text{E}(X_n|\mathbf{y}_{0:n})$ for two crude estimates (Bootstrap-1000 and FA-100) and three CMC estimates (CMC-Boostrap-1000, CMC-FA-100 and CMC-1000): CMC estimates improve the crude ones, particularly in outlier periods. Note that the CMC estimates present similar performances whatever the MC approximation of $p_{0:n-1|n-1}$.

So the difficulty is to find a compromise between the quality of the approximation of $p_{0:n-1|n}$ and the approximation of the quantities needed to compute CMC estimates. It can be seen that $N = 300$ particles are here sufficient for CMC estimates.

Next computational cost is of course a critical point for CMC estimates. So Fig. 3(b) displays the averaged computational time for one iteration and for each CMC estimate. Note that using local sets of M samples enables us to reduce N and, as an additional side effect, to also reduce the computational time due to the propagation of the main set of N particles (in particular during the resampling step), see Fig. 3(b). Moreover, in order to take into account both the MSE and the computational time, we compute the efficiency defined as $\text{Eff}(n) = 1/(\text{MSE}_n^f \times \text{E}(C(n)))$ where $C(n)$ is the CPU time to compute an estimate at time n [25], see Fig. 3(c). It appears that the most efficient settings of the CMC estimate are $N = 300$ and $M = 20$.

Finally Table II displays criterion $\mathcal{J}$ in (31) in the case where $R = 0.00001$; it is indeed a challenging scenario for classical SIR algorithms since small values of R make it

difficult to guide particles into regions with high likelihood. Of course, in the first column the value is independent of M. Here, for a small value of N (e.g. $N = 100$), the CMC estimates improve the SIR based one even when the number M of additional samples is weak. In addition, Fig. 3(b) remains valid for this simulation since the computational time is independent of the chosen parameters.
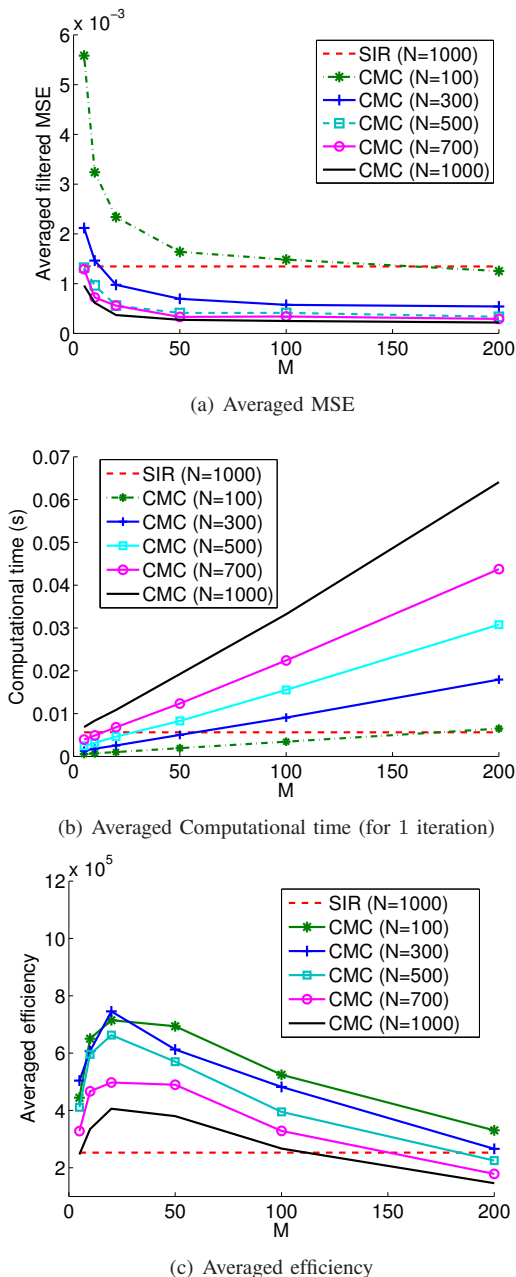

(a) Averaged MSE

(b) Averaged Computational time (for 1 iteration)

(c) Averaged efficiency

Fig. 3. Non linear and non gaussian model - (a) - MSE w.r.t. $\mathrm{E}(\mathbf{X}_n|\mathbf{y}_{0:n})$ for CMC estimates in function of N and M - (b) Averaged computational time of CMC estimates for one iteration - (c) Averaged efficiency. CMC estimates improve the SIR estimate for small values of N and M (see Fig. (a)), and at a reasonable computational cost (see Fig. (b)). Note that for a given N, the maximum of efficiency is achieved for $M = 20$ (see Fig. (c)).

IV. BAYESIAN CMC ALGORITHMS FOR JMSS MODELS

The CMC estimates have been discussed in detail in previous section. We now show that they can be extended to JMSS

models. A JMSS model reads

$$p(\mathbf{r}_{0:n}, \mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(r_0)\prod_{i=1}^{n} p(r_i|r_{i-1})\times$$

$$p(\mathbf{x}_0)\prod_{i=1}^{n} f_{i|i-1}(\mathbf{x}_i|\mathbf{x}_{i-1}, r_i)\prod_{i=0}^{n} g_i(\mathbf{y}_i|\mathbf{x}_i, r_i). \qquad (32)$$

Model (32) can be thought of as an HMC model (13), in which $f_{i|i-1}$ and g_i now depend on the realization of a discrete Markov Chain $\{R_n\}_{n\geq 0}$ where each R_n takes its values in $\{1, \cdots, K\}$. So now both $\mathbf{X}_n$ and R_n are hidden, and as in section III, we focus on $\mathrm{E}(f(\mathbf{X}_n)|\mathbf{y}_{0:n})$. As is well known [26] [27] [4], in a JMSS exact Bayesian filtering is either impossible (in the general case) or an NP-hard problem (in the linear and Gaussian case), so one has to use suboptimal techniques and in particular SMC methods. The goal of this section is twofold. First, we show that the RB estimates obtained in the previous section can be adapted for general JMSS model. Next, we discuss on how specificities of these models (in the sense that a part of the hidden state is now discrete) lead to other RB estimates.

A. Bayesian CMC algorithms for non linear JMSS models

Let us first rewrite Θ_n as

$$\Theta_n = \sum_{\mathbf{r}_{0:n-1}, r_n} \int f(\mathbf{x}_n) p(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}, \mathbf{x}_n, r_n|\mathbf{y}_{0:n})\mathrm{d}\mathbf{x}_{0:n-1}\mathrm{d}\mathbf{x}_n.$$

In model (32), $(\mathbf{X}_n, R_n)$ is a Markov chain and $((\mathbf{X}_n, R_n), \mathbf{Y}_n)$ is an HMC in which transitions and likelihoods respectively read $p(r_n|r_{n-1})\times f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}, r_n)$ and $g_n(\mathbf{y}_n|\mathbf{x}_n, r_n)$; so the methodology described in Section III can be directly applied when we partition $(\mathbf{X}_{0:n}, \mathbf{R}_{0:n})$ as (X_1, X_2) with $X_1 = (\mathbf{X}_{0:n-1}, \mathbf{R}_{0:n-1})$ and $X_2 = (\mathbf{X}_n, R_n)$. Let us assume that an MC approximation of $p(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$ is given by $\widehat{p}(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i \delta_{\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i}$ and $(\mathbf{x}_n^i, r_n^i) \sim p(\mathbf{x}_n, r_n|\mathbf{x}_{n-1}^i, r_{n-1}^i, \mathbf{y}_n)$. The crude estimate $\widehat{\Theta}_n$ and the CMC one $\widetilde{\Theta}_n^{(\mathbf{X}_n, R_n)}$ read

$$\widehat{\Theta}_n(\mathbf{x}_{0:n}^{1:N}, \mathbf{r}_{0:n}^{1:N}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i f(\mathbf{x}_n^i), \qquad (33)$$

$$\widetilde{\Theta}_n^{(\mathbf{X}_n, R_n)}(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{r}_{0:n-1}^{1:N}) = \sum_{i=1}^{N} \widetilde{w}_{n-1}^i \times$$

$$\int f(\mathbf{x}_n) p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, , r_{n-1}^i, \mathbf{y}_n)\mathrm{d}\mathbf{x}_n \qquad (34)$$

(for simplicity f depends on $\mathbf{x}_n$, not on r_n, but the extension is straightforward). Let us now adress computational aspects. Since $((\mathbf{X}_n, R_n), \mathbf{Y}_n)$ is an HMC with augmented state, the techniques described in Section III to obtain an approximation of $p(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$ and to compute the integral in (34) can still be used, except that the key quantities of Section III,

	SIR-1000	CMC-100	CMC-300	CMC-500	CMC-700	CMC-1000
$M = 5$	0.2546	0.5077	0.2859	0.2054	0.1615	0.1300
$M = 10$	0.2546	0.3689	0.1865	0.1138	0.0813	0.0643
$M = 20$	0.2546	0.2481	0.1242	0.0748	0.0574	0.0452
$M = 50$	0.2546	0.1270	0.0560	0.0392	0.0273	0.0195
$M = 100$	0.2546	0.0865	0.0253	0.0166	0.0111	0.0111
$M = 200$	0.2546	0.0487	0.0172	0.0154	0.0110	0.0041

TABLE II

CRITERION $\mathcal{J}$ FOR NON LINEAR AND NON GAUSSIAN MODEL - $R = 0.00001$. IN THIS CHALLENGING SCENARIO, A FEW NUMBER OF SAMPLES M ARE SUFFICIENT TO IMPROVE THE SIR BASED ESTIMATE, EVEN WHEN N IS SMALL. THE COMPUTATIONAL TIME OF OUR ESTIMATES IS ALSO GIVEN IN FIG. 3(B).

$p(\mathbf{y}_n|\mathbf{x}_{n-1})$ and $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n)$, are now replaced by

$$p(\mathbf{y}_n|\mathbf{x}_{n-1}, r_{n-1}) =$$
$$\int \sum_{r_n} p(r_n|r_{n-1}) f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}, r_n) g_n(\mathbf{y}_n|\mathbf{x}_n, r_n) d\mathbf{x}_n, \quad (35)$$

$$p(\mathbf{x}_n|\mathbf{x}_{n-1}, r_{n-1}, \mathbf{y}_n) \propto$$
$$\sum_{r_n} p(r_n|r_{n-1}) f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}, r_n) g_n(\mathbf{y}_n|\mathbf{x}_n, r_n) \quad (36)$$

respectively. All the computational aspects described in paragraphs III-A up to III-D remain valid (up to the adaptations given above) so are not further discussed in this section.

B. Specific CMC estimates for JMSS

Throughout section IV the hidden state $(\mathbf{X}_n, R_n)$ in augmented dimension contains a discrete component R_n; as we now see, this feature enables us to propose a number of CMC estimates which all differ from each other through a partition (X_1, X_2) of $(\mathbf{X}_{0:n}, \mathbf{R}_{0:n})$. Section IV-B1 discusses partition $X_1 = (\mathbf{X}_{0:n-1}, \mathbf{R}_{0:n})$, $X_2 = \mathbf{X}_n$; Section IV-B2 discusses partition $X_1 = (\mathbf{X}_{0:n}, \mathbf{R}_{0:n-1})$, $X_2 = R_n$; finally Section IV-B3 focuses on partition $X_1 = \mathbf{R}_{0:n-1}$, $X_2 = R_n$, in the case where a spatial RB decomposition has already been applied (in particular in linear and Gaussian JMSS).

1) Reducing the computational cost of (34): Instead of setting $X_1 = (\mathbf{X}_{0:n-1}, \mathbf{R}_{0:n-1})$, $X_2 = (\mathbf{X}_n, R_n)$, let us now set $X_1 = (\mathbf{X}_{0:n-1}, \mathbf{R}_{0:n})$, $X_2 = \mathbf{X}_n$. The CMC estimate $\widetilde{\Theta}_n^{\mathbf{X}_n}$ corresponding to this new partition reads

$$\widetilde{\Theta}_n^{\mathbf{X}_n}(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{r}_{0:n}^{1:N}) = \sum_{i=1}^N \widetilde{w}_{n-1}^i \int f(\mathbf{x}_n) p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n, r_n^i) d\mathbf{x}_n. \quad (37)$$

The difference between (37) and (34) comes from the integral which is computed. In (34), and for a given sample $(\mathbf{x}_{n-1}^i, r_{n-1}^i)$, one needs to compute a moment of $p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, r_{n-1}^i, \mathbf{y}_n) \propto \sum_{r_n} p(r_n|r_{n-1}) p(\mathbf{y}_n|\mathbf{x}_{n-1}, r_n) p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{y}_n, r_n)$ and so moments of $p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n, r_n)$ for each $r_n \in \{1, \cdots, K\}$; by contrast, in (37), we only need to compute a moment according to $p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{y}_n, r_n^i)$, where r_n^i is the particle which has just been sampled. In summary, due to partitions and the discussion in §II-A, $\widetilde{\Theta}_n^{(\mathbf{X}_n, R_n)}$ is preferable to $\widetilde{\Theta}_n^{\mathbf{X}_n}$ but requires an extra computational cost.

2) Reducing the variance of weights in a general SMC approach: In this section, we show that it is always possible to reduce the variance of the weights of

any SMC method in non-linear JMSS by using partition $X_1 = (\mathbf{X}_{0:n}, \mathbf{R}_{0:n-1})$ and $X_2 = R_n$. Let us briefly recall the principle of SMC methods based on the SIR algorithm. Starting from an MC approximation $\{(\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i), w_n^{i}\}_{i=1}^N$ of $p(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n-1})$, an MC approximation of $p(\mathbf{x}_{0:n}, \mathbf{r}_{0:n}|\mathbf{y}_{0:n})$ is obtained by sampling $(\mathbf{x}_n^i, r_n^i) \sim q^{\mathrm{SIR}}(\mathbf{x}_n, r_n|\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{y}_n)$; the associated estimate $\widehat{\Theta}_n^{\mathrm{SIR}}$ reads

$$\widehat{\Theta}_n^{\mathrm{SIR}}(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{r}_{0:n}^{1:N}) = \sum_{i=1}^N \frac{w_n^{u,i}}{\sum_{j=1}^N w_n^{u,j}} f(\mathbf{x}_n^i),$$

$$w_n^{u,i} = w_{n-1}^i \frac{p(r_n^i|r_{n-1}^i) f_{n|n-1}(\mathbf{x}_n^i|r_n^i, \mathbf{x}_{n-1}^i) g_n(\mathbf{y}_n|\mathbf{x}_n^i, r_n^i)}{q^{\mathrm{SIR}}(\mathbf{x}_n^i, r_n^i|\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{y}_n)}.$$

Now if we want to build the CMC estimate associated to partition $X_1 = (\mathbf{X}_{0:n}, \mathbf{R}_{0:n-1})$ and $X_2 = R_n$, we need an approximation of $p(\mathbf{x}_{0:n}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$. This approximation can be also obtained via importance sampling; starting again from the MC approximation $\{(\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i), w_{n-1}^i\}$ of $p(\mathbf{x}_{0:n-1}, \mathbf{r}_{0:n-1}|\mathbf{y}_{0:n-1})$, and sampling $\mathbf{x}_n^i$ according to $q(\mathbf{x}_n|\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{y}_n) = \sum_{r_n} q^{\mathrm{SIR}}(\mathbf{x}_n, r_n| \mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{y}_n)$, the CMC estimate $\widetilde{\Theta}_n^{R_n}$ associated to this partition reads

$$\widetilde{\Theta}_n^{R_n}(\mathbf{x}_{0:n-1}^{1:N}, \mathbf{r}_{0:n-1}^{1:N}) = \sum_{i=1}^n \frac{w_n^{\mathrm{RB},u,i}}{\sum_{j=1}^N w_n^{\mathrm{RB},u,j}} f(\mathbf{x}_n^i),$$

$$w_n^{\mathrm{RB},u,i} = w_{n-1}^i \times$$
$$\frac{\sum_{r_n} p(r_n|r_{n-1}^i) f_{n|n-1}(\mathbf{x}_n^i|r_n, \mathbf{x}_{n-1}^i) g_n(\mathbf{y}_n|\mathbf{x}_n^i, r_n)}{q(\mathbf{x}_n|\mathbf{x}_{0:n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{y}_n)}.$$

It is easy to check that $\mathrm{var}(w_n^{\mathrm{RB},u,i}) \leq \mathrm{var}(w_n^{u,i})$; in addition, using the arguments described in §II-B, $\widetilde{\Theta}_n^{R_n}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$ from an asymptotical point of view.

3) Specific CMC estimates for linear and Gaussian JMSS models: Let us finally focus on linear and Gaussian JMSS models, i.e. models in which $f_{n|n-1}$ and g_n in (32) read

$$f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}, r_n) = \mathcal{N}(\mathbf{x}_n, \mathbf{F}_n(r_n)\mathbf{x}_{n-1}, \mathbf{Q}(r_n)), \quad (38)$$
$$g_n(\mathbf{y}_n|\mathbf{x}_n, r_n) = \mathcal{N}(\mathbf{y}_n, \mathbf{H}_n(r_n)\mathbf{x}_n, \mathbf{R}_n^v(r_n)). \quad (39)$$

In this model, it is well known that it is possible to apply a spatial decomposition (see e.g [4]) by writing

$$\Theta_n = \sum_{\mathbf{r}_{0:n}} \underbrace{p(\mathbf{r}_{0:n}|\mathbf{y}_{0:n})}_{\mathrm{PF}} \int f(\mathbf{x}_n) \underbrace{p(\mathbf{x}_n|\mathbf{r}_{0:n}, \mathbf{y}_{0:n})}_{\mathrm{KF}} d\mathbf{x}_n, \quad (40)$$

in which $p(\mathbf{r}_{0:n}|\mathbf{y}_{0:n})$ is computed via PF, while the integral is computed exactly for each sample $\mathbf{r}_{0:n}^i$ via Kalman Filtering (KF).

Now we show that a further variance reduction of this estimate is possible. Indeed (40) coincides with (5), up to the identification: $X_1 = \mathbf{R}_{0:n-1}$, $X_2 = R_n$, $f(x_1, x_2) = \phi(\mathbf{r}_{0:n}) = \int f(\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{r}_{0:n-1}, r_n, \mathbf{y}_{0:n})\mathrm{d}\mathbf{x}_n$, and $p(x_1, x_2)$ is the joint pdf

$$p(\mathbf{r}_{0:n}|\mathbf{y}_{0:n}) = \underbrace{p(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})}_{p(x_1)}\underbrace{p(r_n|\mathbf{r}_{0:n-1}, \mathbf{y}_{0:n})}_{p(x_2|x_1)}.$$

We need to compute both factors (we cannot simply apply the results of §III-A, because in (32) the marginal chain $(R_n, \mathbf{Y}_n)$ is not an HMC, as was $(\mathbf{X}_n, \mathbf{Y}_n)$ in (13)). Let $\widehat{p}(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n}) = \sum_{i=1}^{N}\widetilde{w}_{n-1}^i\delta_{\mathbf{r}_{0:n-1}^i}$, and let $r_n^i \sim p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$. Then the Bayesian crude and CMC estimators of Θ_n respectively read

$$\widehat{\Theta}_n(\mathbf{r}_{0:n-1}^{1:N}, \mathbf{r}_n^{1:N}) = \sum_{i=1}^{N}\widetilde{w}_{n-1}^i(\mathbf{r}_{0:n-1}^{1:N})\phi(\mathbf{r}_{0:n-1}^i, r_n^i), \quad (41)$$

$$\widetilde{\Theta}_n(\mathbf{r}_{0:n-1}^{1:N}) = \sum_{i=1}^{N}\widetilde{w}_{n-1}^i(\mathbf{r}_{0:n-1}^{1:N})\times$$
$$\sum_{r_n}\phi(\mathbf{r}_{0:n-1}^i, r_n)p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n}). \quad (42)$$

Consequently, the computation of $\widetilde{\Theta}_n$ in (42) relies on the computation of $p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$ ($\phi(\mathbf{r}_{0:n-1}^i, r_n)$ can be computed by KF for all r_n, $r_n \in \{1, \cdots, K\}$) and on an approximation of $p(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$. Let us address this last point. Starting from an MC approximation of $p(\mathbf{r}_{0:n-2}|\mathbf{y}_{0:n-1})$, an MC approximation of $p(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$ can be deduced from

$$p(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n}) \propto p(\mathbf{r}_{0:n-2}|\mathbf{y}_{0:n-1})\times$$
$$p(r_{n-1}|\mathbf{r}_{0:n-2}, \mathbf{y}_{0:n-1})p(\mathbf{y}_n|\mathbf{r}_{0:n-1}, \mathbf{y}_{0:n-1})$$

via IS. Since the model is linear and Gaussian, $p(\mathbf{y}_n|\mathbf{r}_{0:n-1}, \mathbf{y}_{0:n-1})$ can be computed via KF from the Gaussian pdf $p(\mathbf{x}_{n-1}|\mathbf{r}_{0:n-1}, \mathbf{y}_{0:n-1})$ and so

$$p(r_{n-1}|\mathbf{r}_{0:n-2}, \mathbf{y}_{0:n-1}) =$$
$$\frac{p(\mathbf{y}_{n-1}|\mathbf{y}_{0:n-2}, \mathbf{r}_{0:n-2}, r_{n-1})p(r_{n-1}|r_{n-2})}{p(\mathbf{y}_n|\mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}) = \sum_{r_n}\mathcal{N}}. \quad (43)$$

(here $\mathcal{N}$ stands for numerator) is also computable. Consequently, except the MC approximation of $p(\mathbf{r}_{0:n-1}|\mathbf{y}_{0:n})$, no additional approximations are needed to compute $\widetilde{\Theta}_n$ in (42). Note also that the optimal importance distribution (for the smoothing problem) proportional to $p(r_{n-1}|\mathbf{r}_{0:n-2}, \mathbf{y}_{0:n-1})p(\mathbf{y}_n|\mathbf{r}_{0:n-1}, \mathbf{y}_{0:n-1})$, is computable too. For this importance distribution weights $\widetilde{w}_{n-1}^i$ in (42) become

$$\widetilde{w}_{n-1}^i \propto \widetilde{w}_{n-2}^i\times$$
$$\sum_{r_{n-1}}p(r_{n-1}|\mathbf{r}_{0:n-2}^i, \mathbf{y}_{0:n-1})p(\mathbf{y}_n|\mathbf{r}_{0:n-2}^i, r_{n-1}, \mathbf{y}_{0:n-1})$$

and thus do not depend on $\{r_{n-1}^i\}_{i=1}^{N}$.

Estimator $\widetilde{\Theta}_n$ in (42) can be seen as a (temporal) further RB step (i.e., $X_1 = \mathbf{R}_{0:n-1}$, $X_2 = R_n$) of the already (spatially partitionned) RB estimator (41) (i.e., $X_1 = \mathbf{R}_{0:n}$, $X_2 = \mathbf{X}_{0:n}$). From a computational point of view, $\widetilde{\Theta}_n$ can be computed under the same assumptions as those needed for computing

$\widehat{\Theta}_n$, at the price of a slight extra computational effort, as we see from (41) and (42). Indeed, $\{p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})\}_{r_n=1}^{K}$ in (43) have to be computed for both estimators. The difference is that in the CMC algorithm we directly compute means $\sum_{r_n}\phi(\mathbf{r}_{0:n-1}^i, r_n)p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$, which requires running K KF updating steps per trajectory $\mathbf{r}_{0:n-1}^i$, while the crude estimator first extends each trajectory before computing conditional expectations.

V. BAYESIAN CMC ALGORITHMS FOR MULTI-TARGET FILTERING

In this final section we adapt CMC from single- to multi-target filtering. As we shall see in section V-A, *multi*-object filtering (see e.g. [28]) essentially reduces to computing $\Theta_n = \int f(\mathbf{x}_n)v_n(\mathbf{x}_n)\mathrm{d}\mathbf{x}_n$ in which $v_n(\mathbf{x}_n)$ is now the so-called Probability Hypothesis Density (PHD), i.e. the a posteriori spatial density of the expected number of targets given all measurements. The same difficulties hold as in single-object filtering: in general, neither PHD v_n nor moment Θ_n can be computed exactly; so again SMC techniques propagate a weighted samples approximation $\{\mathbf{x}_n^i, w_n^i\}_{i=1}^{N}$ of v_n, and Θ_n is estimated by $\sum_{i=1}^{N}w_n^i f(\mathbf{x}_n^i)$. Though in this multi-object context we do not necessary deal with classical pdf (here $\sum_{i=1}^{N}w_n^i$, which in general is different from 1, is an estimator of the number of targets), the discussion in section II-C still holds, up to some adaptations. We begin with a brief review of multi-object filtering and next adapt CMC to the multi-target framework.

A. A brief review of Random Finite Sets (RFS) based multi-target filtering

In multi-object filtering we now look for estimating an unknown number of targets from a set of observations which are either due to detected targets or are false alarms measurements. Early solutions [29] [30] include a costly matching mechanism between targets and observations. Alternately, solutions based on RFS, which are sets of r.v. with random and time-varying cardinal (see e.g. [31]), no longer require such a matching mechanism. The RFS formulation was first used to derive the multi-object Bayesian filter, which generalizes the classical single object one [28]. However this filter involves the computation of set integrals of multi-object densities, and cannot in general be computed in practice (unless the number of targets is small [32]). Later on, Mahler proposed to propagate a first order moment of the multi-object density, the so-called PHD or intensity [28]. Let $|X \cap S|$ be the number of objects in RFS X which belong to region S; then the PHD $v(\mathbf{x})$ is defined as the spatial density of the expected number of targets, i.e.

$$\int_{S \subset \mathbb{R}^p} v(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathrm{E}(|X \cap S|). \quad (44)$$

Let now $v_n(\mathbf{x})$ be the a posteriori PHD given past measurements $Z_{0:n} = \{Z_0, \cdots, Z_n\}$ (be they due to detected targets or to false alarms), where Z_k is the set of measurements available at time k. Let $p_{s,n}(.)$ (resp. $p_{d,n}(.)$) be the probability of survival (resp. of detection) at time n which can depend on state $\mathbf{x}_{n-1}$ (resp. on $\mathbf{x}_n$); $\kappa_n(.)$ (resp. $\gamma_n(.)$) the

intensity of the false alarms measurements (resp. of the birth targets) at time n; and let us assume that the predicted and clutter processes are Poisson, and that each target evolves and generates observations independently of one another. Then v_n is propagated as follows (we assume for simplicity that there is no spawning) [28] [31]:

$$v_{n|n-1}(\mathbf{x}_n) = \int p_{s,n}(\mathbf{x}_{n-1}) f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1}) \mathrm{d}\mathbf{x}_{n-1}$$
$$+ \gamma_n(\mathbf{x}_n), \tag{45}$$
$$v_n(\mathbf{x}_n) = [1 - p_{d,n}(\mathbf{x}_n)] v_{n|n-1}(\mathbf{x}_n)$$
$$+ \sum_{\mathbf{z} \in Z_n} \frac{p_{d,n}(\mathbf{x}_n) g_n(\mathbf{z}|\mathbf{x}_n) v_{n|n-1}(\mathbf{x}_n)}{\kappa_n(\mathbf{z}) + \int p_{d,n}(\mathbf{x}_n) g_n(\mathbf{z}|\mathbf{x}_n) v_{n|n-1}(\mathbf{x}_n) \, \mathrm{d}\mathbf{x}_n}. \tag{46}$$

B. Deriving a Bayesian CMC PHD estimator

The problem we address is to compute moment $\Theta_n = \int f(\mathbf{x}_n) v_n(\mathbf{x}_n) \mathrm{d}\mathbf{x}_n$ (typically, we shall take either $f(\mathbf{x}_n) = 1$, in order to compute an estimate of the number of targets, or $f(\mathbf{x}_n) = \mathbf{1}_S(\mathbf{x}_n)$, where S is some region of interest). From now on we assume that $p_{d,n}$ does not depend on $\mathbf{x}_n$. Plugging (45) in (46), the PHD at time n can be written as

$$v_n(\mathbf{x}_n) = \sum_{i=1}^4 v_n^i(\mathbf{x}_n) \tag{47}$$

where

$$v_n^1(\mathbf{x}_n) = [1 - p_{d,n}] \int p_{s,n}(\mathbf{x}_{n-1}) \times$$
$$f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1}) \mathrm{d}\mathbf{x}_{n-1}, \tag{48}$$
$$v_n^2(\mathbf{x}_n) = [1 - p_{d,n}] \gamma_n(\mathbf{x}_n),$$
$$v_n^3(\mathbf{x}_n) = \sum_{\mathbf{z} \in Z_n} \frac{p_{d,n} g_n(\mathbf{z}|\mathbf{x}_n) \int p_{s,n}(\mathbf{x}_{n-1})}{B_n(\mathbf{z})} \times$$
$$f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1}) \mathrm{d}\mathbf{x}_{n-1}, \tag{49}$$
$$v_n^4(\mathbf{x}_n) = \sum_{\mathbf{z} \in Z_n} \frac{p_{d,n} g_n(\mathbf{z}|\mathbf{x}_n) \gamma_n(\mathbf{x}_n)}{B_n(\mathbf{z})},$$

and where $B_n(\mathbf{z}) = \kappa_n(\mathbf{z}) + B_n^1(\mathbf{z}) + B_n^2(\mathbf{z})$, with

$$B_n^1(\mathbf{z}) = \int p_{d,n} g_n(\mathbf{z}|\mathbf{x}_n) \int p_{s,n}(\mathbf{x}_{n-1}) \times$$
$$f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1}) \mathrm{d}\mathbf{x}_{n-1} \mathrm{d}\mathbf{x}_n$$
$$= p_{d,n} \int p_{s,n}(\mathbf{x}_{n-1}) p(\mathbf{z}|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1}) \mathrm{d}\mathbf{x}_{n-1},$$

and $B_n^2(\mathbf{z}) = \int p_{d,n} g_n(\mathbf{z}|\mathbf{x}_n) \gamma_n(\mathbf{x}_n) \mathrm{d}\mathbf{x}_n$. Term v_n^1 (resp. v_n^2) is due to non-detected persistent (resp. birth) targets, while v_n^3 (resp. v_n^4) is due to detected persistent (resp. birth) targets. From (47) we see that

$$\Theta_n = \sum_{i=1}^4 \underbrace{\int f(\mathbf{x}_n) v_n^i(\mathbf{x}_n) \mathrm{d}\mathbf{x}_n}_{\Theta_n^i}, \tag{50}$$

and we now consider whether one can adapt the CMC methodology of section II-C to any of the moments Θ_n^i.

First, $v_n^2(\mathbf{x}_n)$ and $v_n^4(\mathbf{x}_n)$ do not depend on $v_{n-1}(\mathbf{x}_{n-1})$ so we compute (or estimate via crude MC) Θ_n^2 and Θ_n^4, whence

$\widehat{\Theta}_n^2$ and $\widehat{\Theta}_n^4$. On the other hand, the computation of $v_n^1(\mathbf{x}_n)$ and of $v_n^3(\mathbf{x}_n)$ depends on $v_{n-1}(\mathbf{x}_{n-1})$. Now the PHD is not a pdf (it is a positive function, but remember from (44) that it does not sum to 1), and weights $\{w_{n-1}^i\}_{i=1}^{L_{n-1}}$ may depend on variables different from $\mathbf{x}_{n-1}^{1:L_{n-1}}$ (but which are known at time $n-1$). But these differences do not impact the discussion of section II-C which can still be used in this context. Indeed, we have $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) g_n(\mathbf{z}|\mathbf{x}_n) = p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}_{n-1})$, so Θ_n^1 and Θ_n^3 can be rewritten as

$$\Theta_n^1 = [1 - p_{d,n}] \int \mathrm{E}(f(\mathbf{X}_n)|\mathbf{x}_{n-1}) \times$$
$$[p_{s,n}(\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1})] \mathrm{d}\mathbf{x}_{n-1},$$
$$\Theta_n^3 = \sum_{\mathbf{z} \in Z_n} \int \mathrm{E}(f(\mathbf{X}_n)|\mathbf{x}_{n-1}, \mathbf{z}) \times$$
$$\left[\frac{p_{d,n} p_{s,n}(\mathbf{x}_{n-1}) p(\mathbf{z}|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1})}{B_n(\mathbf{z})} \right] \mathrm{d}\mathbf{x}_{n-1}.$$

Let $\widehat{v}_{n-1} = \sum_{i=1}^{L_{n-1}} w_{n-1}^i \delta_{\mathbf{x}_{n-1}^i}$ and $\widehat{\gamma}_n = \sum_{i=1}^{L_{\gamma_n}} w_{\gamma_n}^i \delta_{\mathbf{x}_{\gamma_n}^i}$ be MC approximations of $v_{n-1}(\mathbf{x}_{n-1})$ and of $\gamma_n(\mathbf{x}_n)$, respectively. Let us start with Θ_n^1. Even if it is not a pdf, factor $p_{s,n}(\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1})$ within brackets plays the role of $p(x_1)$ in (6), and can be approximated by $\sum_{i=1}^{L_{n-1}} w_n^{1,i} \delta_{\mathbf{x}_{n-1}^i}$ where $w_n^{1,i} = [1 - p_{d,n}] p_{s,n}(\mathbf{x}_{n-1}^i) w_{n-1}^i$. So the crude MC and Bayesian CMC estimators of Θ_n^1 are respectively $\widehat{\Theta}_{1,n} = \sum_{i=1}^{L_{n-1}} w_n^{1,i} f(\mathbf{x}_n^i)$ and $\widetilde{\Theta}_{1,n} = \sum_{i=1}^{L_{n-1}} w_n^{1,i} \mathrm{E}(f(\mathbf{x}_n)|\mathbf{x}_{n-1}^i)$ in which $\mathbf{x}_n^i \sim f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}^i)$. Let us next consider Θ_n^3. For each measurement $\mathbf{z} \in Z_n$, factor $\frac{p_{d,n} p_{s,n}(\mathbf{x}_{n-1}) p(\mathbf{z}|\mathbf{x}_{n-1}) v_{n-1}(\mathbf{x}_{n-1})}{B_n(\mathbf{z})}$ within brackets plays the role of $p(x_1)$ in (6), and can be approximated by $\sum_{i=1}^{L_{n-1}} w_n^{3,i}(\mathbf{z}) \delta_{\mathbf{x}_{n-1}^i}$ where $w_n^{3,i}(\mathbf{z}) = p_{d,n} p_{s,n}(\mathbf{x}_{n-1}^i) p(\mathbf{z}|\mathbf{x}_{n-1}^i) w_{n-1}^i / \widetilde{B}_n(\mathbf{z})$ and $\widetilde{B}_n(\mathbf{z})$ is an MC estimator of $B(z)$. So the crude MC and Bayesian CMC estimators of Θ_n^3 are respectively $\widehat{\Theta}_n^3 = \sum_{\mathbf{z}} \sum_{i=1}^{L_{n-1}} w_n^{3,i}(\mathbf{z}) f(\mathbf{x}_n^{\mathbf{z},i})$ and $\widetilde{\Theta}_n^3 = \sum_{\mathbf{z} \in Z_n} \sum_{i=1}^{L_{n-1}} w_n^{3,i}(\mathbf{z}) \mathrm{E}(f(\mathbf{X}_n)|\mathbf{x}_{n-1}^i, \mathbf{z})$ in which $\mathbf{x}_n^{\mathbf{z},i} \sim p(\mathbf{x}_n|\mathbf{x}_{n-1}^i, \mathbf{z})$.

In summary, the crude MC PHD estimator $\widehat{\Theta}_n$ of Θ_n is $\widehat{\Theta}_n = \sum_{i=1}^4 \widehat{\Theta}_n^i$, while our Bayesian CMC PHD estimator $\widetilde{\Theta}_n$ is a sum of two crude MC and two Bayesian CMC estimators: $\widetilde{\Theta}_n = \widetilde{\Theta}_n^1 + \widehat{\Theta}_n^2 + \widetilde{\Theta}_n^3 + \widehat{\Theta}_n^4$. Since $\widetilde{\Theta}_n^1$ and $\widetilde{\Theta}_n^3$ are computed from the same MC approximation of $v_{n-1}(\mathbf{x}_{n-1})$, $\widetilde{\Theta}_n = \mathrm{E}(\widehat{\Theta}_n|\{\mathbf{x}_{n-1}^i\}_{i=1}^{L_{n-1}}, \{\mathbf{x}_{\gamma_n}^i\}_{i=1}^{L_{\gamma_n}}, Z_n)$, so $\widetilde{\Theta}_n$ outperforms $\widehat{\Theta}_n$.

C. Computing the CMC PHD filter $\widetilde{\Theta}_n$ in practice

In the multi-target filter problem, we look for computing an estimator of the number of targets and of multi-target states. From (44), an estimator of the number of targets is given by

$$\widetilde{N}_n = \sum_{i=1}^{L_{n-1}} w_n^{1,i} + \sum_{\mathbf{z} \in Z_n} \sum_{i=1}^{L_{n-1}} w_n^{3,i}(\mathbf{z}) + \sum_{i=1}^{L_{\gamma_n}} w_n^{2,i} + \sum_{\mathbf{z} \in Z_n} \sum_{i=1}^{L_{\gamma_n}} w_n^{4,i}(\mathbf{z}). \tag{51}$$

The procedure to extract persistent targets consists in looking for local maxima of $\sum_{i=1}^{L_{n-1}} w_n^{1,i} p(\mathbf{x}_n|\mathbf{x}_{n-1}^i) +$

$\sum_{\mathbf{z}\in Z_n}\sum_{i=1}^{L_{n-1}} w_n^{3,i}(\mathbf{z})p(\mathbf{x}_n|\mathbf{x}_{n-1}^i,\mathbf{z})$. For birth targets, this procedure cannot be used if the PHD due to birth targets was computed via an MC approximation. One can use clustering techniques [32], or the procedure described in [33], which consists in looking for measurements $\mathbf{z}$ such that $\sum_{i=1}^{L_{\gamma_n}} w_n^{4,i}(\mathbf{z})$ is above a given threshold (typically 0.5) (this procedure can also be used for persistent targets); then an estimator of the state associated to $\mathbf{z}$ is given by $\sum_{i=1}^{L_{\gamma_n}} w_n^{4,i}(\mathbf{z})\mathbf{x}_{\gamma_n}^i$. However, birth targets become persistent targets at the next time step; so their extraction becomes easy at the next iteration since an SMC extraction procedure can be avoided.

Let us now detail some applications of the CMC-PHD filter. We first assume that $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) = \mathcal{N}(\mathbf{x}_n;\mathbf{F}_n\mathbf{x}_{n-1};\mathbf{Q}_n)$, $g_n(\mathbf{z}|\mathbf{x}_n) = \mathcal{N}(\mathbf{z};\mathbf{H}_n\mathbf{x}_n;\mathbf{R}_n)$, and that γ_n is a Gaussian mixture (GM), i.e. that $\gamma_n(\mathbf{x}_n) = \sum_{i=1}^{N_{\gamma_n}} w_{\gamma_n}^i \mathcal{N}(\mathbf{x}_n;\mathbf{m}_{\gamma_n}^i;\mathbf{P}_{\gamma_n}^i)$. For such models a GM implementation has been proposed [34], which consists in propagating a GM approximation of PHD v_n via (45)-(46). The mixture grows exponentially due to the summation on the set of measurements in (46), so pruning and merging approximations are necessary. In addition, this implementation requires that $p_{d,n}$ and $p_{s,n}$ are constant (or possibly GM [34]). In our algorithm we do not need to make any assumption about $p_s(\mathbf{x}_{n-1})$. For this model $B_n^2(\mathbf{z})$ is directly computable, and the Bayesian CMC procedure for estimating the number of targets and extracting the states is valid since $p(\mathbf{x}_n|\mathbf{x}_{n-1},\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x}_{n-1})$ are computable. Finally, in the case where $p_s(\mathbf{x}_{n-1})$ is constant, we have at our disposal three implementations of the PHD filter: the GM one [34], the SMC one [32] and our Bayesian CMC one. These three implementations will be compared in section V-D below.

Now, if γ_n is not a GM our method remains valid since $\Theta_{2,n}$, $\Theta_{4,n}$ and $B_n^2(\mathbf{z})$ can be computed via an MC approximation; the GM structure for persistent targets is kept. Finally in non linear models, the approximation techniques developed in paragraphs III-C and III-D can still be used in this context. In particular, the local linearization and MC methods can be used but require to compute quantities of interest for all measurements $\mathbf{z}$ in $\mathcal{Z}_n$.

D. Simulations

We now compare our Bayesian CMC PHD estimator to alternative implementations of the PHD filter. The MSE criterion used previously is not appropriate in the multi-target context: since the number of targets evolves, a performances criterion should take into account an estimator of the number of targets and one of their states. So in this section we will use the optimal subpattern assignment (OSPA) distance [35], which is a classical tool for comparing multi-target filtering algorithms. Let $X = \{x_1,...,x_m\}$ and $Y = \{y_1,...,y_n\}$ be two finite sets, which respectively represent the estimated and true sets of targets. For $1 \le p < +\infty$ and $c > 0$, let $d^{(c)}(x,y) = min(c,||x-y||)$ ($||.||$ is the Euclidean norm) and let Π_n be the set of permutations on $\{1,2,...,n\}$. The OSPA metric is defined by

$$\overline{d}_p^c(X,Y) \triangleq \left(\frac{1}{n}\left(\min_{\pi\in\Pi_n}\sum_{i=1}^m d^{(c)}(x_i,y_{\pi(i)})^p + c^p(n-m) \right)\right)^{\frac{1}{p}}$$

if $m \le n$, and $\overline{d}_p^c(X,Y) \triangleq \overline{d}_p^c(Y,X)$ if $m > n$. The term $\min_{\pi\in\Pi_n}\sum_{i=1}^m d^{(c)}(x_i,y_{\pi(i)})^p$ represents the localization error, while the second term represents the cardinality error. We set $p = 2$ and $c = 100$.

We focus on a linear and Gaussian model, in which the GM-PHD is used as a benchmark solution, and a Gaussian model with range and bearing measurements. So we compare the GM-PHD, the SMC-PHD and our Bayesian CMC-PHD filters in the first scenario and the SMC-PHD and the approximated CMC-PHD filters in the second one. We track the position and velocity of the targets so $\mathbf{x}_n = [p_x,\dot p_x,p_y,\dot p_y]_n^T$. Let also $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}) = \mathcal{N}(\mathbf{x}_n;\mathbf{F}_n\mathbf{x}_{n-1},\mathbf{Q}_n)$ and $g_n(\mathbf{z}_n|\mathbf{x}_n) = \mathcal{N}(\mathbf{z}_n;\mathbf{H}_n\mathbf{x}_n,\mathbf{R}_n)$ in the linear case or $g_n(\mathbf{z}_n|\mathbf{x}_n) = \mathcal{N}(\mathbf{z}_n;[\mathrm{atan}(p_y/p_x),\sqrt{p_x^2+p_y^2}]^T,\mathbf{R}_n^{rb})$ in the non linear one. We set $\mathbf{F}_n = \mathbf{I}_2 \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$, $\mathbf{R}_n^{rb} = \begin{bmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix}$, $\mathbf{Q}_n = \sigma_v^2\mathbf{I}_2 \otimes \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}$, $\mathbf{H}_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, $\mathbf{R}_n = \sigma_z^2\mathbf{I}_2$.

We compare the SMC-PHD estimator to our Bayesian CMC one in the case where both algorithms use the transition pdf $f_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1})$ (remember that in our approach, we need to propagate a discrete approximation of the PHD, even if it not used for computing an estimator of the number of targets). We take $T = 2$, $\sigma_v = 3$ but $\sigma_z = 0.3$ or $\sigma_b = 0.002$ and $\sigma_r = 0.06$, which means that likelihood $g_n(\mathbf{z}|\mathbf{x}_n)$ is sharp; since the transition pdf does not take into account available measurements, it is difficult to guide particles into promising regions, so this experimental scenario is challenging for the SMC-PHD implementation. Particles are initialized around the measurements [33]. In the pure SMC implementation, we use $N_b = 50$ particles per newborn target and $N = 200$ particles per persistent target, while for the Bayesian CMC we only use $N_b = 20$ particles per newborn target, $N = 20$ particles per persistent target and possibly $M = 10$ if the local MC method is required. The probability of detection is $p_{d,n} = 0.95$ and that of survival $p_{s,n} = 0.98$, for all n, $1 \le n \le 100$, and we generate 10 false alarm measurements (in mean). We consider a scenario with 6 targets which appear either at $n = 0$, $n = 20$ or $n = 50$. We also test the GM implementation in which $T_p = 10^{-5}$ for the pruning threshold, $T_m = 4m$ for the merging threshold and we keep at most $N_{\max} = 100$ Gaussians.

The OSPA distances for both scenarios are displayed in Figure 4. The Bayesian CMC approach outperforms the SMC one and copes with the issue of guiding particles in promising regions. Even if we use the transition density for getting a discrete approximation of v_{n-1}, the Bayesian CMC approach provides a correct estimate of the number of targets, by contrast to the SMC one in which the new set $\{\mathbf{x}_n^i,w_n^i\}_{i=1}^{L_n}$ is used to deduce a discrete approximation of v_n, then an estimate of the number of targets. In other words, even if

$\widehat{v}_{n-1}$ is poor, the CMC estimator performs well and is close to the GM based one.
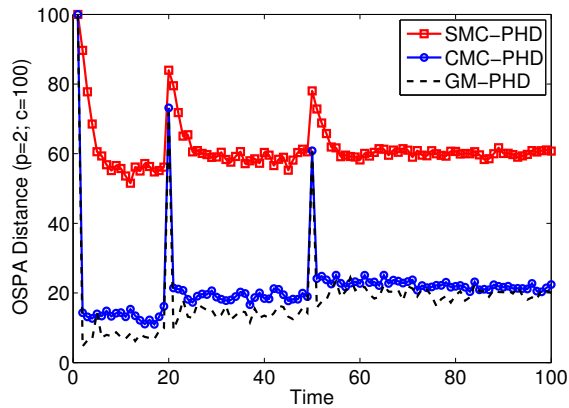
(a) linear scenario
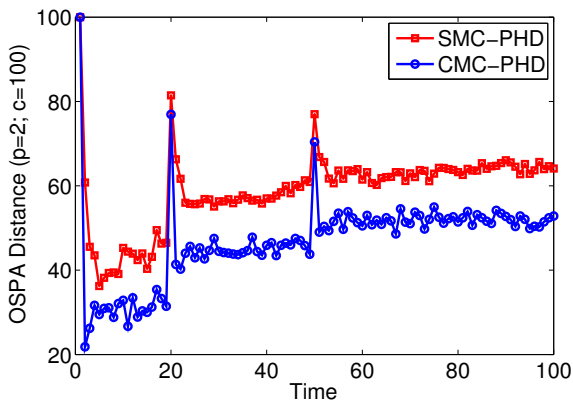
(b) range bearing scenario

Fig. 4. Comparison of the Bayesian CMC and SMC estimates for multi-object filtering. Linear and Gaussian scenario (a): the CMC estimate is close to the GM one - (b) range bearing tracking scenario. The Bayesian CMC (resp. SMC) implementation uses $N_b = 20$ (resp. $N_b = 50$) particles for birth objects, $N = 20$ (resp. $N_b = 200$) for persistent ones. The CMC estimate relies on a local approximation for scenario (b): we set $M = 10$ (which gives a total of 200 particles for persistent objects), which is sufficient to improve the classical PF.

VI. Conclusion

In this paper we proposed a CMC estimator of a moment of interest in a Bayesian filtering context. Our method relies on the recursive nature of SMC algorithms and can be seen as a temporal, rather than spatial, RB-PF procedure. We next showed that a CMC estimator can indeed be computed, or approximated efficiently, in a variety of Markovian models, including HMC or JMSS. In particular, local MC implementations have been discussed and evaluated in terms of variance reduction vs. computational cost. Finally we adapted Bayesian CMC to multi-target filtering, and showed that our CMC PHD estimator has interesting practical features as compared to alternate (SMC or GM) implementations of the PHD filter.

References

[1] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer Verlag, 2001.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking," *IEEE Tr. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[3] R. Chen and J. S. Liu, "Mixture Kalman filters," *J. R. Statist. Soc. B*, vol. 62, pp. 493–508, 2000.

[4] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 613–24, March 2001.

[5] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear nonlinear state-space models," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2279–2289, 2005.

[6] T. Chen, T. Schön, H. Ohlsson, and L. Ljung, "Decentralized Particle Filter With Arbitrary State Decomposition," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 465–478, February 2011.

[7] A. Johansen, N. Whiteley, and A. Doucet, "Exact approximation of Rao-Blackwellised particle filters," in *16th IFAC Symposium on System Identification*, 11-12 July 2012.

[8] A. E. Gelfand and A. F. M. Smith, "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[9] G. Casella and C. Robert, "Rao-blackwellization of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996.

[10] J. Gewecke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–1339, November 1989.

[11] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[12] N. Chopin, "Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference," *The Annals of Statistics*, vol. 32, no. 6, pp. 2385–2411, 2004.

[13] F. Lindsten, T. Schön, and J. Olsson, "An explicit variance reduction expression for the Rao-Blackwellized particle filter," in *18th World Congress of the Int. Federation of Automatic Control (IFAC)*, August 2011.

[14] M. K. Pitt and N. Shephard, "Filtering via simulation : Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–99, June 1999.

[15] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[16] A. Doucet, M. Briers, and S. Sénécal, "Efficient block sampling strategies for sequential Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 1–19, 2006.

[17] P. Fearnhead, D. Wyncoll, and J. Tawn, "A sequential smoothing algorithm with linear computational cost," *Biometrika*, vol. 97, pp. 447–464, 2010.

[18] V. Zaritskii, V. Svetnik, and L. Shimelevich, "Monte Carlo techniques in problems of optimal data processing," *Automation and remote control*, pp. 95–103, 1975.

[19] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–88, March 1994.

[20] S. Saha, P. K. Manda, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian proposal density using moment matching in SMC methods." *Statistics and Computing*, vol. 19-2, pp. 203–208, 2009.

[21] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, vol. 92, March 2004, pp. 401–422.

[22] M. Klaas, N. de Freitas, and A. Doucet, "Toward practical N^2 Monte Carlo: the Marginal Particle Filter," in *Uncertainty in Artificial Intelligence (UAI)*, Arlington, Virginia, 2005, pp. 308–315.

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

[24] R. v. d. Merwe, A. Doucet, N. De Freitas, and E. Wan, "The Unscented Particle Filter," *Advances in Neural Information Processing Systems*, 2000.

[25] P. L'Ecuyer, "Efficiency improvement and variance reduction," in *Winter Simulation Conference 1994*, 1994, pp. 122–132.

[26] J. K. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Transactions on Automatic Control*, vol. 27, no. 5, pp. 1054–65, October 1982.

[27] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New-York: John Wiley and sons, 2001.

[28] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, October 2003.

[29] Y. Bar-Shalom, *Tracking and data association*. San Diego, CA: Academic Press Professional, Inc., 1987.

[30] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 2009.

[31] R. Mahler, *Statistical Multisource Multitarget Information Fusion*. Artech House, 2007.

[32] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, 2005.

[33] B. Ristic, D. Clark, and B. Vo, "Improved SMC implementation of the PHD filter," in *Proceedings of the 13th International Conference on Information Fusion*, 2010.

[34] B.-N. Vo and W. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4091–4104, November 2006.

[35] D. Schuhmacher, B. T. Vo, and B. N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, August 2008.