A Particle Filtering Scheme for Processing Time Series Corrupted by Outliers

Cristina S. Maíz, Elisa M. Molanes-López, Joaquín Míguez, and Petar M. Djurić, Fellow, IEEE

Abstract—The literature in engineering and statistics is abounding in techniques for detecting and properly processing anomalous observations in the data. Most of these techniques have been developed in the framework of static models and it is only in recent years that we have seen attempts that address the presence of outliers in nonlinear time series. For a target tracking problem described by a nonlinear state-space model, we propose the online detection of outliers by including an outlier detection step within the standard particle filtering algorithm. The outlier detection step is implemented by a test involving a statistic of the predictive distribution of the observations, such as a concentration measure or an upper quantile. We also provide asymptotic results about the convergence of the particle approximations of the predictive distribution (and its statistics) and assess the performance of the resulting algorithms by computer simulations of target tracking problems with signal power observations.

Index Terms—Outlier detection, particle filtering, state-space models, target tracking.

I. INTRODUCTION

The detection of outliers, i.e., anomalous observations that deviate considerably from the majority of data, is a classical problem both in statistics and in engineering. If the outliers are not identified and processed properly, they may distort considerably the analysis of the data, especially if the latter is carried out using techniques that are not specifically designed to be robust. Outliers may occur for different reasons and they include human errors, instrument errors, fraudulent behavior and faults in systems. Additionally, the treatment given to an outlier, once detected, may differ depending on the application itself. It is therefore not surprising that there exist several definitions of outlier in the literature, none of them being universally accepted. We refer the reader to [1], where a survey of contemporary methodologies for outlier detection is presented and an exhaustive list of applications is given. One particular engineering application with a natural need for detecting outliers is the processing of data collected by wireless sensor networks. This is a relevant practical problem that has aroused interest lately [2].

In this paper, we are concerned with the online detection of outliers in dynamic systems described by state-space models. State-space modeling is widely used to describe dynamic systems in fields such as applied statistics [3], statistical signal processing [4], time series analysis [5], and econometrics [6]. It involves two equations, a (usually Markov) transition equation that describes the evolution over time of a hidden state vector of interest and an observation equation that relates the observations or measurements to the state vector. Although the most traditional state-space models assume linearity and normality, there exist many practical applications, including navigation and target tracking, where nonlinear and/or non-Gaussian state-space models provide a representation of reality [4].

In many applications where state-space models are used, the objective is often to estimate the so-called filtering distribution of the state, that is, the posterior marginal distribution of the state vector at time $t$ given all the observations up to that time. For a few special cases there exist closed forms for these posterior distributions, including the popular linear Gaussian state-space model. In this case, the Kalman filter [7] provides a set of recursive equations for determining the exact filtering distribution. However, exact closed form solutions are no longer available (in continuous state spaces) when the assumptions of linearity and normality are relaxed. Therefore, different approaches for approximating the filtering distributions under non-normality and non-linearity assumptions have been proposed, including the extended Kalman filter [7], [8], the unscented Kalman filter [9] or the particle filter [10]–[12].

In the framework of dynamic systems, it may happen that the observed data are contaminated with outliers generated by some other mechanism, very different from that assumed by the state-space model, producing a significant discrepancy with the pattern of regular observations. The presence of these anomalous observations may degrade the performance of both the Kalman and particle filters, thus yielding poor estimates of the signals of interest. Some authors have proposed to “robustify” the state-space model and/or the Kalman filter so that it operates properly in the presence of additive outliers [13]–[16]. However,
when using the particle filter, there is just a limited number of papers that have addressed the presence of outliers in the observations. Some authors have tried to make the particle filter more robust by modifying the observation model [17], by using a second order approximation of the log-likelihood to deal with nonlog-concave likelihoods [6] or by using an optimal resampling step in the sense of minimizing a squared error loss function [18]. In applications related to satellite positioning, alienation of outliers has been proposed by detection of multiple propagation effects combined with delayed sampling [19]. Recently, an application of particle filtering for data rectification in a time series framework has been proposed where one first detects the outliers and then replaces them by their expected values [20]. It is important to note that the latter approach is based on an explicit model for the outlier-generating mechanism, which is embedded into the state-space model and the particle filter algorithm. By contrast, our main objective here is to devise a particle filtering scheme that incorporates an outlier detection step without assuming any explicit model for the anomalous observations. It is also our intent to demonstrate that there exist certain advantages of the proposed scheme over the traditional particle filter algorithm.

Specifically, we propose to implement online detection of outliers (within the particle filtering scheme) by means of a test that involves some statistic of the predictive distribution of the observations (i.e., the probability distribution of the observation that involves some statistic of the predictive distribution of the outliers (within the particle time step)). It is important to note that the latter approach is based on an explicit model for the anomalous observations. Section II is devoted to describing the statistical criteria employed in the outlier detection test. In Section V we propose an extended algorithm that relaxes the hard decisions of the outlier detector from Section III. Section VI contains technical proofs regarding the asymptotic convergence of the approximations of the various distributions and statistics needed by the algorithm. In Section VII we present some computer simulation results for a target tracking problem. Different scenarios have been considered, including the intrusion of an interfering moving target in the monitored region, the presence of a fixed power source not accounted by the model and the displacement of one of the sensors from its original position. Finally, we provide some concluding remarks in Section VIII.

II. BACKGROUND

A. State-Space Models

Consider two stochastic processes \( \{ X_t \}_{t \geq 0} \) and \( \{ Y_t \}_{t \geq 1} \) that take values in \( \mathbb{R}^{d_x} \) and \( \mathbb{R}^{d_y} \), termed state process and observation process, respectively. For \( t = 0 \), the random variable (r.v.) \( X_0 \) has a probability density function (pdf) with respect to (w.r.t.) the Lebesgue measure, denoted as \( \tau_0(x_0) \). We often refer to this pdf as the prior density for the process \( \{ X_t \}_{t \geq 0} \). For \( t > 0 \), the process evolves according to the probability law

\[
\Pr\{ X_t \in A | X_{t-1} = x_{t-1} \} = \int_A \tau_t(x_t|x_{t-1})dx_t
\]

where \( \tau_t(\cdot|x_{t-1}) \) is a “transition” pdf w.r.t. the Lebesgue measure that determines the dynamics of the process, and \( A \) is a Borel subset of \( \mathbb{R}^{d_x} \). In the sequel, we denote the Borel \( \sigma \)-algebra on \( \mathbb{R}^{d_x} \) as \( \mathcal{B}(\mathbb{R}^{d_x}) \).

The observation process follows the conditional probability law

\[
\Pr\{ Y_t \in A' | X_t = x_t \} = \int_{A'} \lambda_t(y_t|x_t)dy_t
\]

where \( \lambda_t(\cdot|x_t) \) is the conditional pdf of \( Y_t \) given \( X_t = x_t \) (also w.r.t. the Lebesgue measure) and \( A' \in \mathcal{B}(\mathbb{R}^{d_y}) \). The function \( \lambda^p_t(x_t) \triangleq \lambda_t(y_t|x_t) \) is often termed the likelihood of the state \( X_t = x_t \) given the observation \( Y_t = y_t \). For fixed observations \( y_t, t = 1, 2, \ldots \), we assume that the functions \( \lambda^p_t(x_t) \) are bounded in \( x_t \) for every \( t \).

It is common to describe a state-space model directly by way of the three densities \( \tau_t(\cdot|x_0), \tau_t(\cdot|x_{t-1}) \) and \( \lambda_t(\cdot|x_t) \). Note that these pdf’s can possibly be non-Gaussian and the relationship between \( X_t \) and \( X_{t-1} \) (as well as between \( Y_t \) and \( X_t \)) be nonlinear.

When working with this class of models, the goal is often to estimate or predict the evolution of the state process given the
observations. In particular, we may be interested either in the a posteriori probability distribution
\[
\Pr\{X_t \in A \mid Y_{1:t} = y_{1:t}\} = \int_A \pi_t(x_t | y_{1:t}) \, dx_t
\]
where \(A \in \mathcal{B} (\mathbb{R}^d)\) and \(\pi_t(x_t | y_{1:t})\) is the filtering density of the state-space model at time \(t\), or in the predictive probability law for the state vector \(X_t\) given \(Y_{1:t-1} = y_{1:t-1}\), namely
\[
\Pr\{X_t \in A \mid Y_{1:t-1} = y_{1:t-1}\} = \int_A \pi_{t|t-1}(x_t | y_{1:t-1}) \, dx_t
\]
where \(A \in \mathcal{B} (\mathbb{R}^d)\) and the associated predictive pdf is
\[
\pi_{t|t-1}(x_t | y_{1:t-1}) = \int \pi_t(x_t, x_{t-1}) \pi_{t-1}(x_{t-1} | y_{1:t-1}) \, dx_{t-1}.
\]
Furthermore, we can also construct the predictive probability law for the r.v. \(Y_t\) given the previous record of observations \(Y_{1:t-1} = y_{1:t-1}\). Specifically
\[
\Pr\{Y_t \in A \mid Y_{1:t-1} = y_{1:t-1}\} = \int_A p_t(y_t | y_{1:t-1}) \, dy_t
\]
where \(A \in \mathcal{B} (\mathbb{R}^d)\) and the integral
\[
p_t(y_t | y_{1:t-1}) = \int \lambda_t(y_t | x_t) \pi_{t|t-1}(x_t | y_{1:t-1}) \, dx_t
\]
is the predictive pdf of \(Y_t\) given \(Y_{1:t-1} = y_{1:t-1}\).

In general, the filtering and predictive densities \((\pi_t, \pi_{t|t-1}, \pi_t)\) cannot be computed in closed-form (except for very special situations) and we need to resort to numerical techniques. We conclude this section with a brief example of a state-space model with nonlinear observations.

1) Example 1: Consider the problem of tracking a target that moves over a two dimensional area, monitored by a sensor network with \(N\) nodes. We assume that the target transmits a radio signal that can be received and measured by the nodes. Specifically, the state-space model that describes the dynamics and the observations of the target of interest is given by
\[
x_t = Q x_t + 1 + n_t
\]
\[
y_{k,t} = 10 \log_{10} \frac{P_0}{|r_t - s_k|^2} + \nu_{k,t}, \quad k = 1, \ldots, N
\]
where the state vector \(x_t = [x_{1:t}, x_{2:t}]^T \in \mathbb{R}^4\) contains the position, \(r_t = [x_{1:t}, x_{2:t}]^T\), and velocity, \(v_t = [\dot{x}_{1:t}, \dot{x}_{2:t}]^T\), of the target; \(Q = \begin{bmatrix} I_2 & T_s I_2 \\ 0_2 & I_2 \end{bmatrix}\) is a transition matrix; \(T_s\) denotes the observation period; \(I_2\) and \(0_2\) are 2 \times 2 identity and zero matrices, respectively; \(n_t\) is a 4 \times 1 zero-mean Gaussian-noise covariance matrix \(\Sigma_n\); \(y_{k,t}\) is a radio-signal power observation\(^1\) of the \(k\)th sensor and \(y_{.,.} = [y_{1:t}, \ldots, y_{N,t}] \in \mathbb{R}^{4N}\) is the observation vector; \(P_0\) is the power of the signal transmitted by the target; \(\gamma\) is a propagation factor that depends on the environment; the vector \(s_k = [s_{1:k}, s_{2:k}]^T\) denotes the position of the \(k\)th sensor; and \(\nu_{k,t}\), \(k = 1, \ldots, N\), are i.i.d. Gaussian r.v.’s with zero-mean and variance \(\sigma_{\nu}^2\). The prior pdf of \(X_t\) (at \(t = 0\)) is Gaussian, with a \(4 \times 1\) mean vector \(m_0\) and \(4 \times 4\) covariance matrix \(\Sigma_0\).

Let \(N(x; m, \Sigma)\) denote the normal density of the r.v. \(X\) with mean vector \(m\) and covariance matrix \(\Sigma\). Within this setup, the densities of interest are
\[
\tau_0(x_0) = N(x_0; m_0, \Sigma_0)
\]
\[
\tau_t(x_t | x_{t-1}) = N(x_t; Q x_{t-1}, \Sigma_n)
\]
\[
\lambda_{k,t}(y_{k,t} | x_t) = N(y_{k,t}; 10 \log_{10} \frac{P_0}{|r_t - s_k|^2}, \sigma_{\nu}^2).
\]

Assuming that the observations are conditionally independent, the likelihood can be computed as
\[
\lambda_{k,t}(y_{k,t} | x_t) = \prod_{k=1}^N \lambda_{k,t}(y_{k,t} | x_t)
\]
\[
\propto \prod_{k=1}^N \exp \left\{ -\frac{1}{2\sigma_{\nu}^2} \left( y_{k,t} - 10 \log_{10} \frac{P_0}{|r_t - s_k|^2} \right)^2 \right\}.
\]

B. Standard Particle Filter

Sequential Monte Carlo methods, also known as particle filters [22], have received considerable attention in the last years. The standard particle filter, or bootstrap filter [10], is a simple simulation-based algorithm that can be outlined as follows.

1) Initialization: draw \(M\) independent and identically distributed (i.i.d.) samples from \(\tau_0(x_0)\), and denote them as \(x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(M)}\). (2.1) Draw a new collection of random samples \(\hat{x}_t^{(m)}\), \(m = 1, \ldots, M\), from the transition pdf’s \(\tau_t(x_t | x_{t-1})\), \(m = 1, \ldots, M\). (2.2) Compute the importance weights \(w_t^{(m)} = \lambda_t(y_t, \hat{x}_t^{(m)}), m = 1, \ldots, M\), and normalize them, \(w_t^{(m)} = \frac{w_t^{(m)}}{\sum_{m=1}^M w_t^{(m)}}\).

2) Recursion: given the samples \(x_{t-1}^{(1)}, x_{t-1}^{(2)}, \ldots, x_{t-1}^{(M)}\) obtained at time \(t - 1\), take the following steps.

2.1) Draw a new collection of random samples \(\hat{x}_t^{(m)}\), from the transition pdf’s \(\tau_t(x_t | x_{t-1})\), \(m = 1, \ldots, M\).

2.2) Compute the importance weights \(w_t^{(m)} = \lambda_t(y_t, \hat{x}_t^{(m)}), m = 1, \ldots, M\), and normalize them, \(w_t^{(m)} = \frac{w_t^{(m)}}{\sum_{m=1}^M w_t^{(m)}}\).

2.3) Resample, i.e., generate new samples \(\hat{x}_t^{(1)}, \ldots, \hat{x}_t^{(M)}\) such that \(x_t^{(i)} = \hat{x}_t^{(m)}\) with probability \(w_t^{(m)}\), for \(i = 1, \ldots, M\) and \(m \in \{1, \ldots, M\}\).

We consider this basic algorithm in the rest of the paper for clarity, so as to place the focus specifically on the issue of outlier detection. Many variations of this standard filter exist (see, e.g., [23] for a recent tutorial or [24] for some advanced techniques) and it is straightforward to use the methodologies described in Sections III–V with any of them. Note, in particular, that the bootstrap filter uses the transition density \(\tau_t(x_t, x_{t-1})\) as a proposal function to draw new particles (step 2.1 of the algorithm). This can be substituted by a different proposal distribution as long as the importance weights are computed accordingly [11]. The convergence results in Section VI can be extended to algorithms that draw from an arbitrary proposal function, under mild regularity assumptions, by following arguments similar to those in [25].
The samples in the set \( \Omega_t^M \triangleq \{ x_t^{(m)} \}_{m=1}^M \) are usually termed “particles” and they yield a point-mass approximation of the filtering distribution. Specifically, let us define the random measure

\[
\pi^M_t(dx | y_{1:t}) \triangleq \frac{1}{M} \sum_{m=1}^M \delta_{x_t^{(m)}}(dx)
\]

where \( \delta_{x_t^{(m)}} \) denotes the unit delta measure centered at \( x_t^{(m)} \). Then, for any function \( h : \mathbb{R}^d \rightarrow \mathbb{R} \), integrable w.r.t. the filtering distribution, we can approximate the integral

\[
(h, \pi_t) \triangleq \int h(x_t) \pi_t(x_t, y_{1:t}) dx_t
\]

as

\[
(h, \pi_t) \approx (h, \pi^M_t) = \int h(\pi_{t-1}) \pi^M_t(dx_t | y_{1:t}) = \frac{1}{M} \sum_{m=1}^M \int h(x_t^{(m)}) dx_t^{(m)}.
\]

There are several results regarding the convergence of the approximation error \( \| (h, \pi_t) - (h, \pi^M_t) \| \) as \( M \rightarrow \infty \) (see, e.g., [26, Ch. 10] and references therein). Assume, for instance, that the function \( h \) is bounded. Then it can be proved [26, Corollary 10.28] that

\[
F \left[ \left\| (h, \pi_t) - (h, \pi^M_t) \right\|^2 \right] \leq \frac{\lambda}{M},
\]

where \( F[\cdot] \) denotes expectation (recall that \( \pi^M_t \) is random) and \( c_{t,h} \) is constant w.r.t. the number of particles \( M \).

C. Summary of Notation

To summarize, the densities of interest in the considered model are:

- \( \pi_t(x_t | x_{t-1}) \): the state transition pdf that determines the dynamics of the state process.
- \( \lambda_t(y_t | x_t) \triangleq \lambda_t^n(x_t) \): the likelihood function of the state \( X_t \) given the observation \( Y_t = y_t \).
- \( \pi_{t-1}(x_t | x_{t-1}) \): the prior predictive density of the state \( X_t \) at time \( t \) given the observations \( Y_{1:t-1} = y_{1:t-1} \).
- \( \pi_t(x_t | y_{1:t}) \): the filtering density function, i.e., the density of \( X_t \) given the observations \( Y_{1:t} = y_{1:t} \).
- \( p_t(y_{t+1} | x_t) \): the predictive density of \( Y_{t+1} \) given the previous observations \( Y_{1:t} = y_{1:t} \).

We will frequently use the notation \( \{ f, \pi \} = \int f(x) \pi(x) dx \) and \( \{ f, \mu \} = \int f(x) \mu(dx) \) for integrals w.r.t. either pdf’s, like \( \pi \), or measures, like \( \mu \).

III. PARTICLE FILTERING IN THE PRESENCE OF OUTLIERS

A. Outlier Detection

The state-space model underlying the PF algorithm involves a choice of the likelihood, \( \lambda_t(y_t | x_t) \), and the dynamics of the state, \( \pi_t(x_t | x_{t-1}) \), which, in turn, determine the computation of the importance weights. Therefore, if some observed vectors are outliers, i.e., they are generated by a random mechanism different from the one specified by the model, the resulting point-mass approximation of the filtering distribution, \( \pi^M_t \), computed under the assumptions of \( \pi_0(x_0), \pi_t(x_t | x_{t-1}) \)

and \( \lambda_t(y_t | x_t) \) may lead to poor inference about the state \( X_t \). In particular, the estimation error can be very large.

To alleviate this drawback, we propose to incorporate an outlier detection (OD) step into the PF algorithm. In this paper, we implement OD with the bootstrap filter reviewed in Section II, but the same procedure can be used with virtually any particle filter. As a general criterion for the detection of outliers in dynamical systems, we propose to assess the compatibility between the newly arrived observation \( Y_t = y_t \) and the predictive pdf \( p_t \). When the observation \( y_t \) is “surprising” according to \( p_t \), we declare it an outlier and discard it. For notational convenience, however, we do not suppress the element \( y_t \) from the sequence of observations but, instead, we keep it and treat it as uninformative (i.e., \( \lambda_t(y_t | x_t) \propto 1 \) for all \( x_t \)). As a consequence, if \( y_t \) is an outlier, we simply write \( \pi_t(x_t | x_{t-1}) = \pi_{t-1}(x_t | x_{t-1}) \). In the particle filter, the latter equality translates into setting \( x_t^{(m)} = x_t^{(m-1)} \) for \( m = 1, \ldots, M \).

Note that the density \( p_t \) depends both on the form of the state-space model (i.e., \( \pi_{t>0} \) and \( \lambda_t \)) and on the previous sequence of non-discarded observations \( Y_{1:t-1} = y_{1:t-1} \). To be specific, recall that \( Y_t \) is the r.v. that models the \( t \)th observation a priori, right before it is observed. Let us then consider a statistic that depends on the variable \( Y_t \) and the predictive density \( p_t \). We denote this statistic as \( D(Y_t, p_t) \). Given the previous record of observations \( Y_{1:t-1} = y_{1:t-1} \) and the assumed model, \( D(Y_t, p_t) \) is a r.v. taking values in \( \mathbb{R} \), whose distribution depends on the distribution of \( Y_t \). If we assume that the density of \( Y_t \) is exactly \( p_t \), then we can compute (either exactly or approximately) any desired moments of \( D(Y_t, p_t) \), such as its mean, its variance or its cumulative distribution function. Therefore, it is possible to use \( D(Y_t = y_t, p_t) \), where \( y_t \) is the actual observation, to perform an outlier detection test. The null hypothesis of the test is \( H_0 \) “the observation \( Y_t = y_t \) is a sample from \( p_t \).”

When the null hypothesis is rejected, the observation \( y_t \) is declared an outlier. In the engineering literature, the type I error of the test (the rejection of \( H_0 \) when \( H_0 \) is true) is termed a “false alarm,” while the type II error (the acceptance of \( H_0 \) when \( H_0 \) is not true) is referred to as a “missed detection.”

The test is designed to achieve a certain level of significance, \( \alpha > 0 \), which is an upper bound for the probability of a false alarm. To be precise, assume that \( D(Y_t, p_t) \) is a nonnegative r.v. and let \( d^*_t \) be a threshold value such that when \( Y_t \) is distributed according to \( p_t(y_t | y_{1:t-1}) \), then \( \Pr[D(Y_t, p_t) < d^*_t | Y_{1:t-1} = y_{1:t-1}] \leq \alpha \).

In that case and denoting the actual observation as \( Y_t = y_t \), the test consists of three steps:

1. Compute the threshold \( d^*_t \);
2. Compute the statistic \( D(Y_t, p_t) \); and
3. If \( D(Y_t, p_t) < d^*_t \) then declare \( y_t \) an outlier.

In general, \( p_t(y_t | y_{1:t-1}) \) and \( d^*_t \) may not be available in closed-form and we need to obtain numerical approximations, as described later.

Note that the pdf \( p_t \) depends on which observations have been accepted, and which have been discarded, as a result of the tests.

2This is the simplest version of the particle filter with outlier detection. More sophisticated algorithms are introduced in Section V.
at the time steps $t' < t$. Therefore, the outlier detection test at time $t$ depends on the outputs of the tests carried out at times $t' = 1, 2, \ldots, t - 1$. This is a sequential decision approach that fits naturally with online processing problems, where "reprocessing" of past data is not feasible.

Remark 1: We have assumed $D(Y, p_t) \geq 0$ for clarity of the exposition. There is nothing preventing us from developing the methodology in this section (and its extensions in Section V) for a r.v. $D(Y_t, p_t)$ taking values, e.g., on the whole real line. Assuming $D(Y_t, p_t) \geq 0$, however, makes some of the arguments and procedures described in the paper clearer. For example, if $D(Y_t, p_t) > 0$ it is enough to define a single threshold, while for “two sided” statistics (taking also negative values) we might need to define two or more thresholds.

Also, this constraint entails little loss of generality, since most meaningful statistics for outlier detection can be transformed into nonnegative form in a straightforward way. Assume, for instance, a statistic $D'(Y_t, p_t) \in \mathbb{R}$ such that

$$\Pr \left\{ a < D'(Y_t, p_t) < b | Y_{1:t-1} = y_{1:t-1} \right\} \leq \alpha$$

where $a, b \in \mathbb{R}$. We can easily define $D(Y_t, p_t) = |D'(Y_t, p_t)| = \frac{b-a}{2} \geq 0$ such that

$$\Pr \left\{ D(Y_t, p_t) < \frac{b-a}{2} | Y_{1:t-1} = y_{1:t-1} \right\} \leq \alpha.$$

B. Particle Approximations

Since the predictive density is given by the integral in (1), we can use the particle set $\Omega_t^{(M)}$ computed via the PF at time $t - 1$ to produce samples approximately distributed according to $p_t(y_t | y_{1:t-1})$. This is done by a two-step procedure. First, we draw $M$ predictive samples of the state, $x_t^{(m)}$, $m = 1, \ldots, M$, from the corresponding transition pdf’s $\tau_t(x_t | x_t^{(m)})$ (note that this task is already done in step 2.1 of the standard algorithm described in Section II-B. Then we draw $M$ samples, denoted $y_t^{(m)}$, $m = 1, \ldots, M$, from the conditional densities $\lambda_t(y_t | x_t^{(m)})$.

As a result, we obtain the point-mass approximation of the predictive distribution of the observations

$$p_t^{(M)}(dy | y_{1:t-1}) = \frac{1}{M} \sum_{m=1}^{M} \delta_{y_t^{(m)}}(dy).$$

and the approximation of integrals w.r.t. $p_t$ by means of sums becomes straightforward. In particular, let $f : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$ be a real function with domain in the observation space. Then

$$\langle f, p_t \rangle = \int f(y_t) p_t(y_t | y_{1:t-1}) dy_t$$

$$\approx \langle f, p_t^{(M)} \rangle = \frac{1}{M} \sum_{m=1}^{M} f(y_t^{(m)}),$$

and, under mild assumptions, the error $\langle f, p_t \rangle - \langle f, p_t^{(M)} \rangle$ can be proved to vanish as $M \to \infty$ (this result is explicitly obtained in Section VI, see Theorem 1). We will see that, often, the $D(y_t, p_t)$ statistics actually take the form of integrals w.r.t. $p_t$ and can be approximated in this way. We use the notation $D^{(M)}(y_t, p_t)$ to indicate the approximation of $D(y_t, p_t)$ constructed using $M$ particles. In Section VI we address the a.s. convergence of $D^{(M)}$ toward $D$ as $M \to \infty$.

Due to the difficulty of computing the threshold $d_t^{(a)}$ analytically, we also resort to particle approximations in order to select it. In particular, given the particles $y_t^{(m)}$, $m = 1, \ldots, M$, we can compute the transformed random sample

$$D_t^{(m)} \Delta D(y_t^{(m)} | p_t), \quad m = 1, \ldots, M$$

and approximate $d_t^{(a)}$ as the $\alpha$-quantile of the empirical distribution of $D_t^{(1)}, \ldots, D_t^{(M)}$, denoted $d_t^{\alpha^{(M)}}$.

C. Summary of the Basic Algorithm

The proposed bootstrap algorithm with OD steps (denoted as ODPF) is outlined in Table I. Note that it is convenient, from a notational point of view, to work with the density $p_t(y_t | y_{1:t-1})$ at each time step. However, if, e.g., $y_{t-1}$ had been declared an

<table>
<thead>
<tr>
<th>Initialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>for $m = 1, 2, \ldots, M$</td>
</tr>
<tr>
<td>Sample $x_t^{(m)}$ from $\tau_0(x_0)$ and set $y_t^{(m)} = \frac{1}{M}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recursions</th>
</tr>
</thead>
<tbody>
<tr>
<td>For $t = 1, 2, \ldots$</td>
</tr>
<tr>
<td>- draw $x_t^{(m)}$ from $\tau_t(x_t</td>
</tr>
<tr>
<td>- draw $y_t^{(m)}$ from $\lambda_t(y_t</td>
</tr>
<tr>
<td>- compute the threshold $d_t^{(a)}$</td>
</tr>
<tr>
<td>- compute the statistic $D^{(M)}(y_t, p_t)$</td>
</tr>
</tbody>
</table>

Perform the OD test:

If $D^{(M)}(y_t, p_t) < d_t^{(a)}$ then $y_t$ is an outlier

Set $x_t^{(m)} = \bar{x}_t^{(m)}$, $m = 1, 2, \ldots, M$.

Else

Compute the importance weight $w_t^{(m)} = \lambda_t(y_t | x_t^{(m)})$, $m = 1, 2, \ldots, M$.

Normalize the weights $w_t^{(m)} = \frac{w_t^{(m)}}{\sum_{m=1}^{M} w_t^{(m)}}$, $m = 1, 2, \ldots, M$.

Resample to obtain the particle set $\Omega_t^{(M)} = \{x_t^{(m)}\}_{m=1}^{M}$.
outlier, we should have excluded it from the sequence of collected observations. For the sake of simplicity, we keep the outlying observations in the conditional of the predictive density but treat them as dummy or uninformative data. For example, if \( y_{t-1} \) is an outlier, then \( p_t(y_t | y_{1:t-1}) = p_t(y_t | y_{1:t-2}) \).

IV. CRITERIA FOR OUTLIER DETECTION

The choice of the statistic \( D(Y_t, p_t) \) is key to the performance of the proposed ODPF algorithm, since it actually determines the specific criterion to decide whether an observation is an outlier or not. In this work, we have investigated the use of a couple of statistics, which we discuss in this section.

A. Concentration Measure

Let \( Y_t^M \) be the set of all possible i.i.d. samples of size \( M \) from the predictive distribution with density \( p_t \). According to [27], a “concentration measure” for \( p_t \) can be defined as a function \( F : \mathbb{R}^d \times Y_t^M \rightarrow \mathbb{R} \) such that \( \forall \xi^{(1)}, \xi^{(2)} \in \mathbb{R}^d, p_t(\xi^{(1)} | y_{1:t-1}) > p_t(\xi^{(2)} | y_{1:t-1}) \) implies

\[
\lim_{M \rightarrow \infty} P\{F(\xi^{(1)}, Y_t) > F(\xi^{(2)}, Y_t)\} = 1
\]

where \( Y_t = \{y_t^{(m)}\}_{m=1}^M \). If, instead, \( p_t(\xi^{(1)} | y_{1:t-1}) > p_t(\xi^{(2)} | y_{1:t-1}) \) implies

\[
\lim_{M \rightarrow \infty} P\{F(\xi^{(1)}, Y_t) < F(\xi^{(2)}, Y_t)\} = 0
\]

then \( F \) is a “sparsity measure.” When \( F \) is proportional to a consistent nonparametric density estimator, it can be shown that \( F \) is a concentration measure, while \( \frac{1}{F} \) is a sparsity measure (see [27, Example 1]).

Concentration measures can be applied to detect outliers. To be precise, an observation \( y_t \) attains a high measure of concentration when it lies within a region with high probability density. Conversely, a low measure of concentration indicates a low probability density around \( y_t \), which can, therefore, be considered an atypical realization for that distribution. Here, we specifically consider the concentration measure

\[
p_t(y_t, Y_t) = \frac{1}{M} \sum_{m=1}^M \exp \left( - \left( y_t - y_t^{(m)} \right) \Sigma^{-1} \left( y_t - y_t^{(m)} \right) \right).
\]  

and then, given a sample set \( Y_t = \{y_t^{(m)}\}_{m=1}^M \) generated via the PF, approximate \( D_{CM}(Y_t, p_t) \approx D_{CM}^M(Y_t, y_t) = F(y_t, Y_t) \), where \( F(y_t, Y_t) \) is given by (9).

The threshold \( d_t^{CM} \), needed for the outlier detection test, can be obtained as the \( \alpha \)-quantile of the transformed random sample \( F(y_t^{(1)}, Y_t), \ldots, F(y_t^{(M)}, Y_t) \).

B. Predictive Density

The concentration statistic described in Section IV-A discriminates the observations that lie in low density regions. However, this is done in an indirect way, by using an i.i.d. sample from \( p_t \) and the function \( F \). As suggested in [29], the predictive density \( p_t \) itself can be used to assess whether a realization \( Y_t = y_t \) is “surprising” (according to the terminology in [29]) or not.

In our work, this approach is appealing because the random variable \( Z_t = p_t(y_t | y_{1:t-1}) \) is an integral w.r.t. the predictive density of the state \( X_t, \pi_{t|t-1} \) and, hence, it can be accurately estimated using the PF. To be specific

\[
D_{PD}(Y_t, p_t) = p_t(y_t | y_{1:t-1})
\]

and

\[
p_t(y_t, y_{1:t-1}) = \left( \lambda_t^{Y_t, \pi_{t|t-1}} \right)
\]  

\[
\approx \left( \lambda_t^{Y_t}, \pi_{t|t-1} \right) = \frac{1}{M} \sum_{m=1}^M \lambda_t^{Y_t}(x_t^{(m)})
\]

where \( \pi_{t|t-1}(dx y_{1:t-1}) = \frac{1}{M} \sum_{m=1}^M \delta_{x_t^{(m)}}(dx) \) is the particle approximation of the predictive distribution of the state \( X_t \) given the observations \( Y_{1:t-1} = y_{1:t-1} \). The asymptotic convergence of this approximation follows from the results in, e.g., [26]. See also Section VI for a specific analysis.

The threshold for a test with significance level \( \alpha \) can be (approximately) found from an i.i.d. sample of the r.v. \( D_{PD}(Y_t, p_t) \) by computing the empirical \( \alpha \)-quantile. The realizations of the r.v. \( D_{PD}(Y_t, p_t) \) are generated by first drawing \( y_t^{(1)}, \ldots, y_t^{(M)} \) via the PF and then evaluating, for \( M = 1, \ldots, M \)

\[
D_{PD}(y_t^{(m)}, p_t) \approx D_{CM}^M(y_t^{(m)}, p_t) = \left( \lambda_t^{y_t^{(m)}}, \pi_{t|t-1} \right)
\]

C. Other Detection Criteria

Alternative forms of the statistic \( D(Y_t, p_t) \) may be successfully exploited for outlier detection. A common criterion for outlier detection, that can be easily adapted to the dynamic setup in this paper, involves statistical depth functions [30]. One example is the spatial depth [31]

\[
D_{SD}(Y_t, p_t) = 1 - \int S(y_t - Y_t)p_t(y_t, y_{1:t-1})dy_t
\]

where \( \|z_T y\| = \sqrt{z_T y} \) denotes the Euclidean norm of a column vector \( z \in \mathbb{R}^d \), \( z_T u \) is the scalar product of the column vectors

\[
\text{and then, given a sample set } Y_t = \{y_t^{(m)}\}_{m=1}^M \text{ generated via the PF, approximate } D_{CM}(Y_t, p_t) \approx D_{CM}^M(Y_t, y_t) = F(y_t, Y_t), \text{ where } F(y_t, Y_t) \text{ is given by (9).}
\]

The threshold \( d_t^{CM} \), needed for the outlier detection test, can be obtained as the \( \alpha \)-quantile of the transformed random sample \( F(y_t^{(1)}, Y_t), \ldots, F(y_t^{(M)}, Y_t) \).

B. Predictive Density

The concentration statistic described in Section IV-A discriminates the observations that lie in low density regions. However, this is done in an indirect way, by using an i.i.d. sample from \( p_t \) and the function \( F \). As suggested in [29], the predictive density \( p_t \) itself can be used to assess whether a realization \( Y_t = y_t \) is “surprising” (according to the terminology in [29]) or not.

In our work, this approach is appealing because the random variable \( Z_t = p_t(y_t | y_{1:t-1}) \) is an integral w.r.t. the predictive density of the state \( X_t, \pi_{t|t-1} \) and, hence, it can be accurately estimated using the PF. To be specific

\[
D_{PD}(Y_t, p_t) = p_t(y_t | y_{1:t-1})
\]

and

\[
\Rightarrow \left( \lambda_t^{Y_t}, \pi_{t|t-1} \right) = \frac{1}{M} \sum_{m=1}^M \lambda_t^{Y_t}(x_t^{(m)})
\]

where \( \pi_{t|t-1}(dx y_{1:t-1}) = \frac{1}{M} \sum_{m=1}^M \delta_{x_t^{(m)}}(dx) \) is the particle approximation of the predictive distribution of the state \( X_t \) given the observations \( Y_{1:t-1} = y_{1:t-1} \). The asymptotic convergence of this approximation follows from the results in, e.g., [26]. See also Section VI for a specific analysis.

The threshold for a test with significance level \( \alpha \) can be (approximately) found from an i.i.d. sample of the r.v. \( D_{PD}(Y_t, p_t) \) by computing the empirical \( \alpha \)-quantile. The realizations of the r.v. \( D_{PD}(Y_t, p_t) \) are generated by first drawing \( y_t^{(1)}, \ldots, y_t^{(M)} \) via the PF and then evaluating, for \( M = 1, \ldots, M \)

\[
D_{PD}(y_t^{(m)}, p_t) \approx D_{CM}^M(y_t^{(m)}, p_t) = \left( \lambda_t^{y_t^{(m)}}, \pi_{t|t-1} \right)
\]

C. Other Detection Criteria

Alternative forms of the statistic \( D(Y_t, p_t) \) may be successfully exploited for outlier detection. A common criterion for outlier detection, that can be easily adapted to the dynamic setup in this paper, involves statistical depth functions [30]. One example is the spatial depth [31]

\[
D_{SD}(Y_t, p_t) = 1 - \int S(y_t - Y_t)p_t(y_t, y_{1:t-1})dy_t
\]
that pass the outlier detection test. This largest valid subvector of $y_t$ can actually be found using the following procedure.

1) Draw $y_t^{(1)}, \ldots, y_t^{(M)}$ using the PF. Set $L = d_y$.
2) Find all $L$ possible $L$-dimensional subvectors of $y_t$ and denote them as $y_t^L(1), \ldots, y_t^L(n)$, where $n = \binom{d_y}{d_y - L}$.
   Let $\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n$ denote the indices that yield the $L$-dimensional subvector $y_t^L(r)$ for $r = 1, \ldots, n$.
3) Set $\mathcal{I} = \emptyset$. For $r = 1, \ldots, n$:
   a) Compute the threshold $d_t^L(r)$ as the $\alpha$-quantile of the empirical distribution of
      \[
      y_t^L(r)_m = \left(\begin{array}{c}
      y_t^{(m)}_{\tilde{r}_1}, \ldots, y_t^{(m)}_{\tilde{r}_L}
      \end{array}\right)
      \]
      for $m = 1, \ldots, M$.
   b) Compute the statistic $D(y_t^L(r), p_t)$ for the subvector $y_t^L(r)$ (obtained from the actually observed $y_t$).
   c) If $D(y_t^L(r)) > d_t^L(r)$ then set $\mathcal{I} = \mathcal{I} \cup \{r\}$, else $y_t^L(r)$ is declared an outlier.
4) If $\mathcal{I} = \emptyset$ then set $L = L - 1$ and go back to step 2. Else, let $r_{\text{max}} = \arg \max_{r \in \mathcal{I}} D(y_t^L(r), p_t)$ and declare $y_t^L(r_{\text{max}})$ a valid vector of observations.

This algorithm outputs the largest subvector that passes the proposed outlier detection test. If there are several subvectors with maximum dimension that pass the test, then we choose the subvector with the largest value of the test statistic (this is done in step 4 above). This selection is coherent with our method because, for all the considered criteria, a large value of the statistic indicates compatibility between the observation and the assumed model.

This method is optimal in the sense that we find the biggest collection of observations that passes the outlier detection test. However, it is also a computationally demanding procedure that cannot be applied in many practical problems. Indeed, if $d_y$ is large (say $d_y = 20$), then there are, e.g., 15 504 subvectors of dimension 15.

In the examples of Section VII we will use a simplified algorithm that applies the outlier detection test to the scalar observations $y_{t,1}, \ldots, y_{t,d_y}$.

The set $\mathcal{I}$ is composed by indices

\[ r_{\text{max}} = \arg \max_{r \in \mathcal{I}} D(y_t^L(r), p_t) \]

and declare $y_t^L(r_{\text{max}})$ a valid vector of observations.

V. EXTENDED ALGORITHMS

The outlier detection test described in Section III makes a “hard” decision on the nature of the observation $y_t \in \mathbb{R}^{d_y}$, meaning that either its $d_y$ components are accepted and used as data or all of them are discarded together. However, it is easy to imagine situations in which a column vector of observations $y_t = [y_{t,1}, \ldots, y_{t,d_y}]^T \in \mathbb{R}^{d_y}$ is declared an outlier because just a few (or even just one) of its components are atypical, while the other ones are valid data that should not be discarded. Ideally, we would like to find the largest subvector of the form

\[ y_t^L = [y_{t,1}, y_{t,2}, \ldots, y_{t,L}]^T \in \mathbb{R}^{d_y}, i_1 < i_2 < \ldots < i_L, L \leq d_y, \]

consisting of elements of the original vector $y_t \in \mathbb{R}^{d_y}$.
of the $L$ scalar observations that pass the test. Then, we build the $L$-dimensional observation vector $y^L_t = \left[ y_{1t}, \ldots, y_{L,t} \right]^T$ and the $M$ $L$-dimensional particles $y^{L(m)}_t = \left[ y^{(m)}_{1t}, \ldots, y^{(m)}_{L,t} \right]^T$ with $i_1 < \cdots < i_L$ and $i_j \in \mathcal{I}$, $j = 1, \ldots, L$ and $m = 1, \ldots, M$. Finally, we test whether $y^L_t$ is an outlier, then is discarded and no data are used at time $t$. The details of this simplified version of the extended algorithm (denoted by E-ODPF from here on) are given in Table II. Note that the statistic $D$ can be substituted by $\hat{D}$, $\tilde{D}$, or others.

VI. CONVERGENCE OF PARTICLE APPROXIMATIONS

The proposed outlier detection criteria are largely based on the ability of generating samples from the predictive density of the observations, $p_t(y_t|y_{1:t-1})$ and using them to approximate integrals of the form $\int f(y_t) \mathbf{d}y_t$, where $f : \mathbb{R}^d \to \mathbb{R}$ is a real function in the observation space. In this section, we formally extend the classical results of [26], [35] on the asymptotic convergence of the particle filter to account also for integrals w.r.t. the density $p_t$. The latter theoretical result is applied to prove that the statistics proposed in Section IV ($D_{CM}$, $D_{FP}$) can be accurately approximated, under mild assumptions, using the samples generated by the particle filter. Finally, we also prove that the threshold $d^T$ needed by the outlier detection test can also be approximated with arbitrary small error whenever it has a continuous and invertible cdf. For the specific cases of the statistics $D_{CM}$ and $D_{FP}$, this condition is satisfied by systems where the density $p_t$ is positive and continuous.

A. Approximation of the Test Statistics

Assume that the sequence of observations $Y_{1:T} = y_{1:T}$ is arbitrary but we keep it fixed (hence, non-random) for a sufficiently large time horizon $T$. Then, for any $0 \leq t \leq T$, let $\{\pi^{M}_{t}\}_{M \in \mathcal{N}}$, $\{\pi^{M}_{t-1}\}_{M \in \mathcal{N}}$, and $\{\{p_{t}\}_{m}^{M}\}_{M \in \mathcal{N}}$ be sequences of approximating measures for the filtering distribution, the predictive distribution of the state and the predictive distribution of the observations, respectively. Typically, these approximations are random, but such randomness is only due to the mechanism for the generation of particles (not because of the observations). Recall that we can construct $p_{t}^{M}$ according to (7), using samples drawn from the conditional pdf of the observation given the particles, $\lambda_t(y_t|x_t(m))$, $m = 1, \ldots, M$.

Let $B(\mathbb{R}^d)$ denote the set of bounded real functions over $\mathbb{R}^d$, i.e.

$$B(\mathbb{R}^d) = \left\{ g : \mathbb{R}^d \to \mathbb{R} \text{ such that } \|g\|_{\infty} = \sup_{x \in \mathbb{R}^d} g(x) < \infty \right\}. $$

As a first step in our analysis, we aim at showing that, for any bounded function $f \in B(\mathbb{R}^d)$, the (random) absolute approximation error $(f, p_{t}^{M}) - (f, p_{t})$ converges to 0 almost surely (a.s.) when $M \to \infty$. For this purpose, we take advantage of a classical result.

Lemma 1: If $Y_{1:T} = y_{1:T}$ is arbitrary but fixed, then $\forall h \in B(\mathbb{R}^d)$ and $0 \leq t \leq T$

$$\lim_{M \to \infty} \{h, \pi^{M}_{t-1}\} - \{h, \pi_{t}^{M}\} = 0 \text{ a.s.} \quad (12)$$

Proof: See [26, Th. 10.29].
The following theorem, whose proof follows closely the arguments in [35] and [26], establishes the convergence of the particle approximations of integrals w.r.t. the density $p_t$.

**Theorem 1:** If $Y_{1:t-1} = y_{1:t-1}$ is arbitrary but fixed, then $\forall f \in B(\mathbb{R}^d)$ and $0 < t < T$

$$\lim_{M \to \infty} \int f \left( p_t^M \right) - \int f \left( p_t \right) = 0 \quad \text{a.s.} \quad (13)$$

**Proof:** See Appendix A.

In the sequel we apply Theorem 1 to show that the approximation error of the test statistics converges toward 0. Recall that we write $D^M(y_t, p_t)$ to denote the approximation of the statistic $D(y_t, p_t)$, using $M$ particles, in such a way that $D^M(y_t, p_t)$ is given by the approximation in (10), while $D^M_{\text{CM}}(y_t, p_t) = F(y_t, y_t)$, where $y_t = \{ y_t^{(m)} \}_{m=1,...,M}$, as shown in (9).

Lemma 1 and Theorem 1 enable a straightforward assessment of the approximations of the test statistics. In particular, a direct application of Lemma 1 yields

$$\lim_{M \to \infty} \left| D^M_P(y_t, p_t) - D^M_{\text{CM}}(y_t, p_t) \right| = 0 \quad \text{a.s.} \quad (13)$$

provided that $\lambda^v \in B(\mathbb{R}^d)$ for $1 \leq t \leq T$. Hence, we can achieve an arbitrarily accurate approximation of the statistic $D^M_P(y_t, p_t)$. Similarly, (13) yields

$$\lim_{M \to \infty} \left| D^M_{\text{CM}}(y_t, p_t) - D^M_{\text{CM}}(y_t, p_t) \right| = 0 \quad \text{a.s.}$$

since, for fixed $y_t$, $h^v(y_t) = \exp \left\{ -(y_t - y)^\top \Sigma^{-1} (y_t - y) \right\}$ is a bounded real function in the space of the observations, i.e., $h^v \in B(\mathbb{R}^d)$.

### B. Approximation of the Thresholds

In order to implement the proposed outlier detection test, we not only have to approximate the statistic $D(y_t, p_t)$ but also the threshold $d^\alpha$ using particles. Specifically, we use the artificial observations $\{ y_t^{(m)} \}_{m=1,...,M}$ to generate a sample of size $M$, $D^{(1)}, \ldots, D^{(M)}$, where $D^{(m)} = D(y_t^{(m)}, p_t)$, and then choose the $\alpha$-quantile of the empirical distribution of the sample, denoted $d^\alpha_{\text{CM}}$, as an approximation of the true threshold $d^\alpha$. Therefore, the convergence of this approximation needs to be established. Note that this is not trivial because the samples $D^{(m)}$ are not necessarily i.i.d. and cannot be computed exactly (as discussed in Section VI-A, but they are approximated using particles and this fact has to be taken into account in the analysis).

We first prove a preliminary result for the approximation of quantiles from size $M$ samples and then use it to show that $|d^\alpha_{\text{CM}} - d^\alpha| = 0$ a.s. when $M \to \infty$ for the statistic $D^M_{\text{CM}}(y_t, p_t)$. The argument for $D^M_P(y_t, p_t)$ is very similar and hence we skip it.

1) Preliminary Results: Let $D \geq 0$ be a nonnegative, scalar r.v. with an associated probability measure $\mu$ and define the cumulative distribution function (cdf) of $D$ as

$$G(d) = \int I_d(x) \mu(dx)$$

where $I_d(x)$ is the indicator function that yields 1 when $x \in [a, b]$ and zero otherwise. Given a sample of size $M$, $D^{(1)}, D^{(2)}, \ldots, D^{(M)}$, we approximate the cdf as

$$G^M(d) = \frac{1}{M} \sum_{m=1}^M I_d(D^{(m)}) \quad (14)$$

Consider now the following assumptions on the function $G$ and the sample $D^{(1)}, \ldots, D^{(M)}$.

**Assumption 1:** $G$ is continuous and invertible.

**Assumption 2:** For every $d > 0$,

$$\lim_{M \to \infty} G(d) - G^M(d) = 0 \quad \text{a.s.}$$

Additionally, for any $\alpha \in (0, 1)$, define $d^\alpha > 0$ as the value such that $G(d^\alpha) = \alpha$. Because of Assumption 1, $d^\alpha$ is unique.

Similarly, $d^\alpha_{\text{CM}}$ is an approximation of $d^\alpha$ constructed from $G^M$. Specifically

$$d^\alpha_{\text{CM}} \triangleq \inf_{m \in \{1,...,M\}} \left\{ D^{(m)} : G^M(D^{(m)}) \geq \alpha \right\}$$

Then we can state the following result regarding the approximation error.

**Theorem 2:** Under Assumptions 1 and 2, $\lim_{M \to \infty} |d^\alpha - d^\alpha_{\text{CM}}| = 0 \quad \text{a.s.}$ for any $\alpha \in (0, 1)$.

**Proof:** See [36, Theorem 2.3.1].

**Remark 2:** Assumptions 1 and 2 are slightly stronger than those in [36, Theorem 2.3.1]. Indeed, Theorem 2 can actually be proved assuming that $d^\alpha$ is unique, but $G$ not necessarily continuous and invertible. We will show, however, that Assumption 1 and 2 hold true for the particle approximations used in the outlier detection step of the particle filter.

2) Convergence of $d^\alpha_{\text{CM}}$: We focus on the case of $D_{\text{CM}}(y_t, p_t)$. In order to take advantage of Theorem 2 and guarantee that the approximate threshold $d^\alpha_{\text{CM}}$ converges to the correct value $d^\alpha$, we need to prove that Assumptions 1 and 2 actually hold true in our case. We provide such proofs for the case in which the density $p_t$ is continuous and positive. Specificaly note that if $p_t$ is continuous, then the transformation $D_{\text{CM}}(y_t, p_t)$ is itself a continuous mapping of $y_t \in \mathbb{R}^d$ into $(0,1)$ and the cdf $G^M(d)$ is also continuous. In particular

$$D_{\text{CM}}(\cdot, p_t) : \mathbb{R}^d \to (0,1),$$

where $a = \inf_{y \in \mathbb{R}^d} D_{\text{CM}}(y, p_t)$, $b = \sup_{y \in \mathbb{R}^d} D_{\text{CM}}(y, p_t)$ and $G : [a, b] \to [0,1]$. Also, if $p_t > 0$, then $\int_a^b p_t(y)dy > 0$ for every non-null Borel set $A$. The two lemmas below establish that $G(d)$ is invertible (hence, Assumption 1 is true) and that the sample $D^{(m)} = D_{\text{CM}}(y_t^{(m)}, p_t)$, $m = 1, \ldots, M$, can be used to obtain a convergent approximation of $G(d)$ (hence, Assumption 2 also holds).

**Lemma 2:** Assume that the density $p_t > 0$ is continuous and let $G$ denote the cdf of the r.v. $D_{\text{CM}}(Y_t, p_t)$.

If we choose $d$
and \( d' \) such that \( a < d < d' < b \), then \( G(d') - G(d) > 0 \). In particular, \( G \) is strictly increasing and, hence, invertible.

**Proof:** See Appendix B.

**Lemma 3:** Assume that the density \( p_t > 0 \) is continuous, let \( G \) denote the cdf of the r.v. \( D_{CM}(Y_t, p_t) \) and let \( G'(d) \) be like in (14), where \( D_{CM}^{(m)} \) is replaced by \( D_{CM}^{(m)}(y_t^{(m)}, p_k) \). Then

\[
\lim_{M \to \infty} G(d) - G'(d) = 0 \quad \text{a.s.} \quad (15)
\]

**Proof:** See Appendix C.

Lemmas 2 and 3 guarantee that Assumptions 1 and 2 hold true for the case of \( D_{CM} \{y_t, p_t\} \), provided \( p_t \) is continuous and positive. Therefore, Theorem 2 implies that \( \lim_{M \to \infty} |a_t^{x,M} - d_t^{x,M}| = 0 \) a.s.

VII. COMPUTER SIMULATIONS

We illustrate the validity of the ODPF by applying it to the target tracking problem of Example 1. This is a typical problem where particle filters can be useful, but in our simulations we have assumed that different types of outliers may appear within the sequence of observations. We have obtained results for three distinct scenarios, that differ in the way the outlying observations are generated. Specifically, we have carried out three sets of simulations under the following scenarios:

- Scenario A: an interfering moving target appears within the region monitored by the sensor network.
- Scenario B: besides the target to be tracked, there is a fixed power source in the field, not accounted by the model.
- Scenario C: one of the sensors in the network is moved to an unknown position.

We remark that the mechanism by which the outliers are generated is not included in the state space model from which the particle filter is derived (this model describes only the typical evolution of the state and the observations). Therefore, the role of the outlier detection step is to identify any atypical observations and remove them to avoid estimation errors.

We test the performance of the algorithms described in previous sections over these scenarios. To be precise, we present results for

- the algorithm outlined in Table I with the CM statistic (denoted as OPC in the sequel) and the PD statistic (denoted as OPP);
- the extended algorithm shown in Table II with the CM statistic (labeled E-OPC) and the PD statistic (labeled E-OPP).

We compare these algorithms to the data-rectification particle filter (DRPF) of [20]. This technique assumes an additive model for outliers in the observation equation.

All the data used in the simulations (both “regular” and outlying observations) are synthetic.

A. Scenario A: Dynamic Interferer

In this first scenario, we assume that the outliers are caused by an interfering moving object that appears in the monitored region. For the simulations, we have used as a prior for the state the Gaussian distribution \( \tau_0 \{x_0\} = \mathcal{N}(x_0;[1,1,1]^T,\frac{1}{3}I_3) \), the covariance matrix for the state transition density is \( \Sigma_n = \text{diag}[2,2,3,3] \), the power of the signal transmitted by the target is \( P_0 = 1 \), the power decay exponent is \( \gamma = 2 \), the observational noise terms are i.i.d. Gaussian variables with zero mean and variance \( \sigma_o^2 = 1 \) and the observation period is \( T_o = 1 \) s. The sensor network has \( N = 8 \) nodes deployed over a two-dimensional field, with coordinates \((-400, -400), (400, -400), (0, -400), (-200, 0), (400, 400), (0, 400), (-400, 400), (200, 0)\), respectively. From \( t = 51 \) to \( t = 60 \), the sensors are “captured,” meaning that they measure the power of the signal transmitted by the interferer and not that of the intended target. Thus, during this period, the observations have the form (different from the typical model)

\[
y_{k+1} = 10 \log_{10} \left( \frac{P_0^i}{\|r_{t}^i - s_k^i\|^2} \right) + \nu_{k,t}
\]

for \( k = 1,2, \ldots , 8 \) and \( t = 51, \ldots , 60 \), where \( P_0^i \neq P_0 \) and \( r_{t}^i \neq r_{t}^j \) are the power of the signal transmitted by the interferer and its position at time \( t \), respectively [see (2) and (3) of Example 1 for the typical model]. The power of the interferer is \( P_0^i = 1.5P_0 \), it has a constant velocity \( v_i \equiv [x_{i,1},1,4x_{i,2},5,1] \) and its position evolves according to \( r_{t}^i = r_{t-1}^i + v_iT_o \) or \([n_1^i,n_2^i]^T\), where \( r_{50}^i = r_{50} \) (it starts at the same position as the true target) and \( n_1^i, n_2^i \) are i.i.d. Gaussian variables with zero mean and variance 0.5. Each simulation has a maximum of 100 time steps and it is stopped if the moving target exceeds the boundary of the square region delimited by the coordinates \([\pm 400, \pm 400]\). In case the target reaches the boundary before \( t = 51 \), the simulation is restarted so that there is at least one outlier.

This scenario is well suited to test the basic algorithm presented in Section III-C. Note that, in this case, if a vector \( y_t \) is an outlier then every component of the vector is an outlier (it has been generated by the interferer) hence we would not gain anything by applying the extended algorithm of Section V to test every scalar observation separately. For comparison, we have applied the proposed OPP and OPC algorithms, as well as the DRPF of [20] to track the target. The latter method relies on the assumption that the outliers are additive and the observational noise can be accurately modeled by a mixture of Gaussian variables with an impulsive component, i.e., the noise contribution at the \( k \)th sensor is the r.v.

\[
\nu_{k,t} \sim (1 - \epsilon)N(0,\sigma_o^2) + \epsilon N(0,b^2\sigma_o^2)
\]

where \( \epsilon \) is the prior probability that an observation is an outlier and \( b = 10 \). The DRPF estimates the probability, say \( P_{imp} \), that the noise in \( y_{k,t} \) is due to the impulsive component. If \( P_{imp} > 0.5 \), then \( y_{k,t} \) is declared an outlier (but is used by the PF as a regular observation, since the impulsiveness of the noise is built into the space-state model, unlike in the proposed ODPFs). Also note that the DRPF tests each component separately, while our method gives a result for the whole vector \( y_t \).

Each detection method has parameters that need to be selected. The DRPF has the parameter \( \epsilon \) and the ODPFs require to choose the significance level \( \alpha \). Furthermore, depending on the test statistic of choice, the ODPF may need additional parameters. In particular, the OPC necessitates to select the covariance matrix \( \Sigma \) [see (9)] while the OPP with the PD
First, we have investigated how the value of these parameters affect the false alarm rate (FAR) and the missed detection rate (MDR) of the outliers as well as the root mean square error (RMSE) of the target position estimates. To do this, we have kept fixed the value of the parameter and have run 1000 simulations of this scenario with \( M = 2000 \) particles. The generation of the trajectories of the true and interfering targets, as well as the sensor observations, is independent across different computer simulations (only the time instant where the sensors are captured is kept fixed at \( t = 51 \), as described above). We have applied the OPP, OPC, and DRPF methods to detect the outliers and track the target trajectory.

The results are shown in Fig. 1. Fig. 1(a) shows the FAR, the MDR and RMSE attained by the DRPF with different values of the parameter \( \epsilon \). The DRPF has a poor performance in this scenario, with high false alarm and missed detection probabilities and a large RMSE. The DRPF assumes an additive model for outliers and this does not match the actual mechanism by which outliers are generated. Note that in this scenario the outliers are not additive, since they are due to the presence of an interfering target.

Fig. 1(b) depicts the FAR, the MDR and the RMSE achieved by the OPC for several values of \( \alpha \) and the scale factor \( s \). As it was expected (since we have proved that \( \lim_M \to \infty \frac{d_M}{M} = \frac{d_R}{\alpha} \) a.s.), the FAR in the simulations is very close to the theoretical significance level \( \alpha \). Furthermore, there is a tradeoff between the FAR and the MDR, in such a way that as the FAR increases, the MDR decreases, and vice versa. Fig. 1(b) (middle row) also illustrates the relevance of the factor \( s \). A value of \( s = 1 \) leads to poor results (both the FAR and the RMSE are relatively large) while for \( s = 10, s = 100, \) and \( s = 1000 \) we attain a trade off between the FAR and the MDR. The RMSE ranges between 20 and 25, which is always less than the RMSE achieved using the DRPF.

We have also applied the OPP method in this scenario and plotted the FAR, the MDR and RMSE as functions of the significance level \( \alpha \) in Fig. 1(c). The results are very similar to those yielded by the OPC technique. We note, however, that for \( \alpha > 0.04 \), the RMSE attained by the OPP method is always below the RMSE of the OPC algorithm for all the selected values of the bandwidth parameter \( s \). In particular, this means...
that the OPP algorithm delivers a (slightly) better performance without the need to tune the extra parameter $s$ (which does not exist in the OPP technique).

Fig. 2 shows the results of a typical simulation for this scenario. We have plotted the trajectories estimated by the DRPF and by the OPP algorithms. The DRPF gets locked to the interferer and, when the interference disappears and regular observations are received again, it takes more than 40 time steps to get locked to the desired target again. The OPP, instead, detects the outliers, discards them and keeps predicting the desired target trajectory until regular observations are received again. As a consequence, the estimation error is much lower and the algorithm almost immediately gets locked to the correct trajectory when typical observations become available.

As a final remark for this example, we note the different effect of false alarms and missed detections in the performance of the tracking algorithms. While false alarms lead to the disposal of valid data, the effect of a missed detection is to track the target using an observation coming from an interferer. While both of them degrade the performance, occasional false alarms may not have a noticeable effect on the RMSE performance because the target track can usually be accurately predicted over a few time steps. Thus, in Fig. 1, the higher FAR for the OPC trackers with scale factor $s > 1$ attained as $a$ is increased does not actually lead to a comparably higher RMSE. A similar observation can be made for the OPP tracker in Fig. 1(c).

**B. Scenario B: Fixed Interfering Power Source**

In this scenario, the tracking is carried out in a sensor network affected by an interfering power source. We assume that the interfering source is active for the first 25 time steps and, from $t = 26$ to $t = 100$, it starts working intermittently following a Markov model with transition probability matrix $P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$, where $P_{ij}$ is the probability of transition from state $i \in \{1, 2\}$ to state $j \in \{1, 2\}$. The state 1 indicates that the power source is off while the state 2 indicates that the source is active. The target is allowed to move over a region delimited by the coordinates $(±250, ±550)$. The simulation is stopped if the target exceeds the boundary of this region before $t = 100$ and it is restarted if it does so before the source starts working at $t = 26$.

The sensor network is composed by $N = 8$ nodes deployed in a two-dimensional area with coordinates $(-200, 300)$, $(200, -300)$, $(0, -300)$, $(-160, 0)$, $(200, 500)$, $(0, 500)$, $(-200, 500)$, and $(100, 0)$, respectively. The interfering power source is located at the point $r^i = (0, 900)$ and generates a fixed power of $P_0^i = 1.5P_i$ when it is active. As a prior distribution for the state, we have used the Gaussian pdf $\mathcal{N}(x_0; [-5, -5, 50, 150]^T, \frac{1}{2}I_4)$, the variance of the observation noise is set to $\sigma_w^2 = 0.5$ and the rest of the parameters keep the same values as in scenario A.

When the source is inactive, the observations are generated following (3) and when it is working they have the form

$$y_{k,t} = 10 \log_{10} \left( \frac{P_0^i}{|r_t - x_k|^5} + \frac{P_0^i}{|r_t - x_k|^7} \right) + \nu_{k,t}$$

$16$ $k = 1, \ldots, N$, i.e., they depend on the sum of the power due to the desired target plus the power of the interfering source, but this knowledge is not available to the trackers.

The network is designed so that sensors 5, 6, and 7 are more affected by the interfering source than the other nodes, since they are closer, and so the received power due to that source is higher than the power in nodes 1, 2, 3, 4, and 8. However, all the sensors are influenced by the interference to some extent. Besides, since the observations are nonlinear, the difference between (3) and (16) is not constant (even though the interference source generates a fixed power) and depends strongly on the value of $\frac{P_0^i}{|r_t - x_k|^5}$. Consequently, when the power source is working we cannot guarantee either that the components 5, 6,
and 7 of \( y_t \) are always outliers or that the rest of the components are safe from being outliers. In this scenario, it is not possible to determine precisely when a scalar observation is an outlier and when it is not, but what is true is that when the interferer is active, the whole vector \( y_t \) is an outlier. For this reason, in the simulations presented here, we have computed the FAR and the MDR for each observation vector as well as the overall RMSE using the extended algorithms (E-OPP and E-OPC) of Section V in the scenario B. We have run 1000 independent simulations using \( M = 2000 \) particles and \( 1000 \) independent simulations.

The results are shown in Table III. The E-OPC and E-OPP algorithms obtain similar results while the DRPF gets the best performance in this scenario. This is because the outliers caused by the interference source can be seen as additive noise with an enhanced power, which is exactly the model that the DRPF method assumes for them. Therefore, in this case the DRPF matches the mechanism by which outliers have been generated and the performance is excellent. The E-ODPFs still work correctly, the RMSE is not high and the FAR is close to the significance level \( \alpha \) chosen beforehand.

When we introduced the E-ODPF in Section V, we stated that when \( y_t \) is declared an outlier, the estimation can be improved if we do not discard the whole vector \( y_t \) and use those components of \( y_t \) that are not atypical. To illustrate this idea, we have studied the improvement of the RMSE by using the E-ODPFs in this scenario. We have run 1000 independent simulations using \( M = 2000 \) particles and different significance levels, \( \alpha = 0.01, 0.02, \ldots, 0.12 \) and have computed the RMSE of the target position averaged over the 1000 runs with the E-OPP, E-OPC, OPP, and OPC algorithms. The results are shown in Fig. 3.

In particular, Fig. 3(a) shows the results using the OPC-like algorithms. It may appear surprising, since the RMSE achieved by the E-OPC method is only slightly better than the RMSE obtained with the simpler OPC algorithm. In Fig. 3(b) we have plotted the comparison between the OPP and the E-OPP techniques, and in this case the RMSE is nearly identical. As discussed earlier, the interference source affects all sensors in the network to some extent, while the extended algorithms are advantageous only when a “clean” subvector of observations \( y^L \) can be identified. In conclusion, in this scenario, the gain of applying the extended algorithm is very low due to the fact that outliers are present in all the components of vector \( y_t \) when the interfering source is active.

### C. Scenario C: Displacement of A Sensor

This scenario is obtained by a slight modification of the setup in Section VII-B. Indeed, the sensor network we consider now is the same as in scenario B, except that one sensor is displaced from its original position for some reason, e.g., due to environmental factors [37]. It keeps operating properly but, due to the incorrect information about its location, its measurements appear as outliers. The state space model is also the same as in scenario B. In the simulations, we assume that at time instant \( t = 51 \), sensor 4 is moved from its original location \((-160, 0)\) to the new coordinates \((-260, 300)\) and it remains in that position until the end of the simulation at \( t = 100 \), or until the moment when the target exceeds the boundary of the region \((\pm250, \pm550)\), as in scenario B.

We have run 1000 independent simulations using \( M = 2000 \) particles (both all PF’s) and \( \alpha = 0.05 \), and then we have computed the FAR and the MDR of each component and the RMSE of the target position estimates, averaged over the 1000 trials. With the aim of removing the dependence on the test performed over the whole vector, the FAR and the MDR are obtained by always carrying out an OD test over each component, regardless of the output of the preceding OD test on the full vector \( y_t \). The RMSE is calculated for the extended algorithms E-OPP and E-OPC as described in Table II.

The results are shown in Table IV. Note that outliers can only exist for \( y_{4,t} \) and, therefore, the MDR in the rest of dimensions is always 0. Hence, we only include in Table IV the estimated MDR for sensor 4. The DRPF works worse than the ODPFs, which achieve similar results with both statistics. In this scenario, outliers cannot be modeled as impulsive noise, since they are nonadditive. The DRPF fails because the model assumed by that algorithm does not match the true mechanism that generates the outliers.

We have carried out the same experiment as in scenario B (see Fig. 3) to compare the E-OPP and E-OPC algorithms with the simpler OPP and OPC techniques. In Fig. 4(a) we have

<table>
<thead>
<tr>
<th>Method</th>
<th>FAR</th>
<th>MDR</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRPF</td>
<td>0.01469</td>
<td>0.00591</td>
<td>12.26</td>
</tr>
<tr>
<td>E-OPC</td>
<td>0.06568</td>
<td>0.00015</td>
<td>13.41</td>
</tr>
<tr>
<td>E-OPP</td>
<td>0.06360</td>
<td>0.00015</td>
<td>13.53</td>
</tr>
</tbody>
</table>

---

**Table III**  
FAR and MDR of the Test Over the Whole Vector \( y_t \), and RMSE in Scenario B. In the E-ODPFs, the Significance Level was \( \alpha = 0.05 \) and in the DRPF, the Parameter \( \epsilon \approx 0.1 \)
Table IV
FAR, MDR, AND RMSE ACHIEVED BY THE THREE METHODS IN SCENARIO C WITH SIGNIFICANCE LEVEL $\alpha = 0.05$

<table>
<thead>
<tr>
<th>Component</th>
<th>DRPF</th>
<th>E-OPC</th>
<th>E-OPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAR</td>
<td>MDR</td>
<td>RMSE</td>
</tr>
<tr>
<td>1</td>
<td>0.0556</td>
<td>0.0547</td>
<td>0.0543</td>
</tr>
<tr>
<td>2</td>
<td>0.0364</td>
<td>0.0555</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>0.0548</td>
<td>0.0534</td>
<td>0.0547</td>
</tr>
<tr>
<td>4</td>
<td>0.0012</td>
<td>0.0534</td>
<td>0.0576</td>
</tr>
<tr>
<td>5</td>
<td>0.0211</td>
<td>0.0538</td>
<td>0.0517</td>
</tr>
<tr>
<td>6</td>
<td>0.0337</td>
<td>0.0536</td>
<td>0.0515</td>
</tr>
<tr>
<td>7</td>
<td>0.0383</td>
<td>0.0538</td>
<td>0.0515</td>
</tr>
<tr>
<td>8</td>
<td>0.0512</td>
<td>0.0564</td>
<td>0.0622</td>
</tr>
</tbody>
</table>

Fig. 4. Comparison between the ODPF and the E-ODPF in Scenario C with (a) the CM statistic and (b) the PD statistic.

VIII. CONCLUSION

We have proposed an algorithm that includes an online outlier detection stage within the traditional particle filtering steps. The outlier detection is carried out by means of a test that involves some statistic of the predictive distribution of the observations (i.e., the probability distribution of the observation at time $t$ given the data collected up to time $t-1$). The first approach tests each observation vector at time $t$ and accepts or discards all of its components. We have also considered an extended version that keeps the valid components and only discards the ones that are outliers. As for the test statistics, we have investigated the use of concentration measures and the predictive density of the observations. The algorithms need to implement numerical approximations of various statistics of the predictive distribution using the particle filter, including the predictive density itself, integrals and quantiles. We have proved that all these approximations converge almost surely as the number of particles grows. We have also provided some numerical examples over three different scenarios in a tracking problem framework and we have compared the proposed algorithms to the one introduced in [20].

The particularity of the latter technique is that it assumes an explicit model for outlying observations. We have shown that if the outliers are not generated by the model assumed for them, the performance deteriorates severely. On the contrary, our method does not require the assumption of any model for the outliers and, therefore, exhibits a considerable versatility that enables us to successfully apply it in a broad class of scenarios.

APPENDIX A

PROOF OF THEOREM 1

Note first that $p_t$ can be written in terms of the predictive density of the state $\pi_{t|t-1}$ and the likelihood $\lambda_i^t$. Specifically, for any $y \in \mathbb{R}^d$, $p_t(y|\{y_{l|l-1}\}) = \{\lambda_i^y, \pi_{t|t-1}\}$. Therefore, any integral with respect to $p_t$ can be rewritten as $(\tilde{f}, \pi_{t|t-1})$, where $\tilde{f}(x) = \{f, \lambda_i\} = \int f(y)\lambda_i^y(x|y)dy$. Using this property, a simple triangle inequality yields

$$\|\tilde{f}\|_\infty \leq \|f\|_\infty = \sup_{y \in \mathbb{R}^d} |f(y)|.$$  

Since $f \in B(\mathbb{R}^d)$, also $\tilde{f} \in B(\mathbb{R}^d)$ and, moreover

$$\|\tilde{f}\|_\infty \leq \|f\|_\infty = \sup_{y \in \mathbb{R}^d} |f(y)|.$$  

Therefore, the second term on the right-hand side (RHS) of (17) can be controlled using Lemma 1, with $h = \tilde{f}$, i.e.

$$\lim_{N \rightarrow \infty} \|\tilde{f}\|_\infty (N) = 0$$  

(19)
In order to prove the convergence of the first term on the RHS of (17), let us define
\[ \mathcal{E}_t = \sigma \left( x_{r}^{(m)}, \bar{x}_{t}^{(m)} \right); \quad m = 1, \ldots, M, 0 \leq r < t, \quad 0 < s \leq t \]
the \( \sigma \)-algebra generated by the random particles \( x_{r}^{(m)} \) and \( \bar{x}_{t}^{(m)} \), \( m = 1, \ldots, M \). Also let \( E_1 \mid \mathcal{E}_t \) denote expectation conditional on \( \mathcal{E}_t \). It is straightforward to see that
\[
E_1 \left[ f(y_{t}^{(m)}) \right] = \int f(y) \lambda_t^r \left( x_t^{(m)} \right) dy - \tilde{f}(\bar{x}_t^{(m)}) \tag{20}
\]
hence the random variables \( \varepsilon_t^{(m)} = f(y_{t}^{(m)}) - \tilde{f}(\bar{x}_t^{(m)}) \) are independent and zero-mean conditional on \( \mathcal{E}_t \). As a consequence\(^7\),
\[
E \left[ \left( f(p_t^M) - \left( \tilde{f}, \pi_t^M \right) \right)^4 \mid \mathcal{F}_t \right] = \frac{3}{M^2} \sum_{k=1}^{M} \left( \int f(y_{t}^{(k)}) - \tilde{f}(\bar{x}_t^{(k)}) \right)^2 \mid \mathcal{F}_t \times E \left[ \left( f(y_{t}^{(i)}) - \tilde{f}(\bar{x}_t^{(i)}) \right)^2 \mid \mathcal{F}_t \right] + \frac{1}{M^2} \sum_{i=1}^{M} b_\varepsilon \left( f(y_{t}^{(i)}) - \tilde{f}(\bar{x}_t^{(i)}) \right)^4 \mid \mathcal{F}_t \tag{21}
\]
and, using the fact that \( \| \tilde{f} \|_\infty \leq \| f \|_\infty < \infty \) [see (18)], we obtain\(^8\),
\[
E \left[ \left( f(p_t^M) - \left( \tilde{f}, \pi_t^M \right) \right)^4 \mid \mathcal{F}_t \right] \leq \frac{c \| f \|_\infty^4}{M^2} \tag{22}
\]
where \( c \) is a constant independent of \( M \). Since the bound in (22) is independent of \( \mathcal{F}_t \), it follows that we can take the unconditional expectation of the error and obtain
\[
E \left[ \left( f(p_t^M) - \left( \tilde{f}, \pi_t^M \right) \right)^3 \mid \mathcal{F}_t \right] \leq \frac{c \| f \|_\infty^4}{M^2} \tag{23}
\]
From (23), a standard Borel-Cantelli argument yields
\[
\lim_{M \to \infty} \left( f(p_t^M) - \left( \tilde{f}, \pi_t^M \right) \right) = 0 \quad \text{a.s.} \tag{24}
\]
Taking together (17), (19), and (24) we conclude that (13) holds true.

Remark 3: Lemma 1 and Theorem 1 are still valid when either \( h = h^M \) or \( f = f^M \) (i.e., when the test functions depend on the number of particles \( M \)) as long as \( h^M \in B(\mathbf{R}^d) \), \( f^M \in \mathcal{H}(\mathbf{R}^d) \) and the corresponding upper bounds do not depend on \( M \), i.e., there exist \( \| h \|_\infty < \infty \) and \( \| f \|_\infty < \infty \) such that \( \| h^M \|_\infty = \| h \|_\infty \) and \( \| f^M \|_\infty = \| f \|_\infty \). In the case of Theorem 1, this is apparent from (23). See [38, Lemma 1] for a generalization of Lemma 1 with explicit convergence rates in terms of \( \| h \|_\infty \).

APPENDIX B

PROOF OF LEMMA 2

Let \( J_A : \mathbf{R}^d \to \{0, 1\} \) denote an indicator function for the set \( A \), i.e., \( J_A(y) = 1 \) if \( y \in A \) and \( J_A(y) = 0 \) otherwise. For a fixed sequence of observations \( y_{1:t-1} \), we can express the cdf as
\[
G(d') = \int J_{A_d}(y)p_t(y|y_{1:t-1})dy = \left( J_{A_e}, p_t \right)
\]
where \( A_d = \{ y \in \mathbf{R}^d : D_{CM}(y, p_t) \leq d \} \). (25)

Therefore, for \( a < d < d' < b \), the difference \( G(d') - G(d) \) is nonnegative, since
\[
G(d') - G(d) = \left( J_{A_{d'}} - J_{A_d} \right)(y)p_t(y|y_{1:t-1})dy = \left( J_{A_{d'} \setminus A_d} \right)(y)p_t(y|y_{1:t-1})dy \geq 0 \tag{26}
\]
where the equality in (26) follows from the fact that \( A_d \subseteq A_{d'} \).

In order to prove that \( G(d') - G(d) > 0 \) (strictly positive), it is sufficient to show that \( \ell(A_{d'} \setminus A_d) > 0 \), where \( \ell(A) \) denotes the Lebesgue measure of a Borel set \( A \). Indeed, if \( R \) is a Borel set with \( \ell(R) > 0 \), then \( J_R(p_t) = \int_R p_t(y|y_{1:t-1})dy > 0 \), since \( p_t > 0 \).

To see that \( \ell(A_{d'} \setminus A_d) > 0 \), choose any value \( d'' \in (d, d') \). Since \( D_{CM}(\cdot, p_t) \) is continuous,
(a) there exists \( y'' \in \mathbf{R}^d \) such that \( D_{CM}(y'', p_t) = d'' \), hence \( y'' \in A_{d'} \setminus A_d \), and
(b) there exists a small enough \( \epsilon > 0 \) such that for every \( y' \in O_{\epsilon}(y'') \),
\[
O_{\epsilon}(y'') \triangleq \{ y \in \mathbf{R}^d : \| y - y'' \| < \epsilon \}
\]
is the open ball centered at \( y'' \) and with radius \( \epsilon > 0 \), it is true that
\[
d < D_{CM}(y, p_t) < d'.
\]
As a consequence, \( O_{\epsilon}(y'') \subseteq A_{d'} \setminus A_d \) and, since \( \epsilon > 0 \), it follows that \( \ell(A_{d'} \setminus A_d) > \ell(O_{\epsilon}(y'')) > 0 \).

APPENDIX C

PROOF OF LEMMA 3

Let us define the set
\[
A_d^M \triangleq \{ y \in \mathbf{R}^d : D_{CM}(y, p_t) \leq d \}
\]
and, for \( y \in \mathbf{R}^d \),
\[
A_d^M(y, p_t) - D_{CM}(y, p_t) \leq D_{CM}(y, p_t) - D_{CM}^M(y, p_t)
\]
From Section VI-A we know that \( \lim_{M \to \infty} D_{CM}^M(y, p_t) = D_{CM}(y, p_t) \) a.s. This means that for any \( \epsilon > 0 \) there exists \( M_\epsilon \) such that for all \( M > M_\epsilon \), we obtain \( |D_{CM}(y, p_t) - D_{CM}^M(y, p_t)| < \epsilon \) or, equivalently

\[
-c < D_{CM}(y, p_t) - D_{CM}^M(y, p_t) < \epsilon. \tag{27}
\]

If we define the sets

\[
A_{d^*}^t = \{ y \in \mathbb{R}^{d_0} : D_{CM}(y, p_t) - d < \epsilon^* \} \\
A_{d^-\epsilon} = \{ y \in \mathbb{R}^{d_0} : D_{CM}(y, p_t) - d < -\epsilon \}
\]

then from (27) we can deduce that \( A_{d^-\epsilon} \subseteq A_{d^*}^t \) for every \( M > M_\epsilon \). Also, from (25) we obtain the (similiar) relationship \( A_{d^-\epsilon} \subseteq A_{d^*}^t \). Moreover, since \( D_{CM}(y, p_t) \) is a continuous mapping of \( y \in \mathbb{R}^{d_0} \) onto \( \{a, b\} \) and \( p_t \) is a density w.r.t. the Lebesgue measure, it follows that for every \( \delta > 0 \) there exists \( \epsilon_\delta > 0 \) such that \( \mu_t(A_{d^*+\epsilon} \setminus A_{d_-}) < \delta \), where \( \mu_t(A) \defeq (A, p_t) \). Similarly, for every \( \delta > 0 \) there exists \( \epsilon_\delta > 0 \) such that \( \mu_t(A_{d^-\epsilon} \setminus A_{d^*}) < \delta \).

Next, define the symmetric set difference

\[
A_d^M \triangle A_d = (A_d^M \setminus A_d) \cup (A_d \setminus A_d^M).
\]

Again, there exists \( M_\epsilon \) such that \( A_d^M \subseteq A_d^\epsilon \) for every \( M > M_\epsilon \) and, as a consequence, \( A_d^M \setminus A_d \subseteq A_d^\epsilon \setminus A_d \). There also exists \( M_\epsilon \) such that \( A_d^\epsilon \subseteq A_d^{\epsilon_\delta} \) for every \( M > M_\epsilon \), which implies that \( A_d^M \setminus A_d \subseteq A_d^\epsilon \setminus A_d^{\epsilon_\delta} \). Therefore, for every \( \delta > 0 \)

\[
A_d^M \triangle A_d \subseteq (A_d^\epsilon \setminus A_d) \cup (A_d \setminus A_d^{\epsilon_\delta})
\]

and

\[
\mu_t(A_d^M \triangle A_d) \leq \mu_t(A_d^\epsilon \setminus A_d) + \mu_t(A_d \setminus A_d^{\epsilon_\delta}) < 2\delta. \tag{28}
\]

Equation (28) is the basic result we need in order to prove that (15) holds true. Let us first note that \( G(d) = (A_d, p_t) \), while \( G^M(d) = (A_d^{\epsilon_\delta}, p_t) \). To establish the limit (15), we need to show that the absolute error \( |G^M(d) - G(d)| \) can be made arbitrarily small by taking \( M \) large enough. With that aim, we apply the triangle inequality to obtain

\[
|G^M(d) - G(d)| = |(J_{A_d^{M+\epsilon_\delta}}, p_t^M) - (J_{A_d}, p_t)|
\]

\[
\leq |(J_{A_d^{M+\epsilon_\delta}}, p_t^M) - (J_{A_d^M}, p_t)| + |(J_{A_d^M}, p_t) - (J_{A_d}, p_t)|. \tag{29}
\]

From Theorem 1 and Remark 3 (note that \( \|J_{A_d^M}\| = 1 \) for every \( \delta > 0 \) there exists \( M_\delta \) such that, for all \( M > M_\delta \)

\[
(J_{A_d^M}, p_t^M) - (J_{A_d}, p_t) < \epsilon
\]

hence the first term in (29) can be bounded. As for the second term, note that

\[
|(J_{A_d^M}, p_t^M) - (J_{A_d}, p_t)| \leq \mu_t(A_d^M \triangle A_d) \leq 2\delta. \tag{31}
\]

for every \( M > \max\{M_\epsilon, M_\delta, M_\delta\} \), where the last inequality follows from (28).

If we choose \( M > \max\{M_\epsilon, M_\epsilon, M_\delta, M_\delta\} \) and take together (29),(30) and (31), we obtain

\[
|G^M(d) - G(d)| \leq 3\epsilon
\]

where \( \delta > 0 \) can be made as small as needed.
Cristina S. Maíz received the M.Sc. degree in telecommunication engineering from University Carlos III of Madrid, Spain, in 2007. She joined the Department of Signal Theory and Communications of the same university, where she is also pursuing the Ph.D. degree in statistical signal processing, dynamical systems, and applications of Monte Carlo methods. In 2011, she joined the Aerospace and Defense Division of the Spanish company, GMV.

Elisa M. Molanes-López received the M.Sc. degree in mathematics in 2000 from the University of Santiago de Compostela, Spain. She received the Ph.D. degree (European mention) in statistics in 2007 from the University of A Coruña, Spain. During her Ph.D. studies, she had the opportunity to visit several universities: Universiteit Hasselt (Belgium), Université Catholique de Louvain (Belgium), and The University of Texas. Late in 2007, she joined the Department of Statistics, Universidad Carlos III, as a Visiting Professor. Her research interests are in the fields of nonparametric statistics, survival analysis, and Receiver Operating Characteristic (ROC) curves. She has coauthored five international journal papers in the area of statistics.

Joaquín Míguez received the M.Sc. and Ph.D. degrees in computer engineering from the University of A Coruña, Spain, in 1997 and 2000, respectively. Late in 2000, he joined the Department of Electronics and Systems, University of A Coruña, where he became an Associate Professor in July 2003. From April through December 2001, he was a Visiting Scholar with the Department of Electrical and Computer Engineering of the State University of New York at Stony Brook. In September 2004, he moved to Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, also as an Associate Professor. His research interests are in the fields of statistical signal processing, Bayesian analysis, dynamical systems and the theory and applications of Monte Carlo methods. He has coauthored more than 30 international journal papers in the fields of signal processing, communications, mathematical physics, and statistics.

Dr. Míguez has delivered lectures on various European universities and research centers. He is a co-recipient of the IEEE Signal Processing Magazine Best Paper Award 2007.

Petar M. Djurič (S’86–M’90–SM’99–F’06) received the B.S. and M.S. degrees in electrical engineering from the University of Belgrade, in 1981 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Rhode Island, in 1990. From 1981 to 1986, he was a Research Associate with the Institute of Nuclear Sciences, Belgrade. Since 1990, he has been with Stony Brook University, where he is a Professor in the Department of Electrical and Computer Engineering. He works in the area of statistical signal processing, and his primary interests are in the theory of signal modeling, detection, and estimation and application of the theory to a wide range of disciplines.

Prof. Djurič has been invited to lecture at many universities in the United States and overseas. In 2007, he received the IEEE Signal Processing Magazine Best Paper, and in 2008 he was elected Chair of Excellence of Universidad Carlos III de Madrid-Banco de Santander. During 2008–2009 he was a Distinguished Lecturer of the IEEE Signal Processing Society. In 2012, he received the EURASIP Technical Achievement Award. He has served on numerous committees for the IEEE, and is currently a Member-at-Large of the Board of Governors of the Signal Processing Society.