Bayesian approach to model-based signal processing: An overview.

Article *in* The Journal of the Acoustical Society of America · March 2010

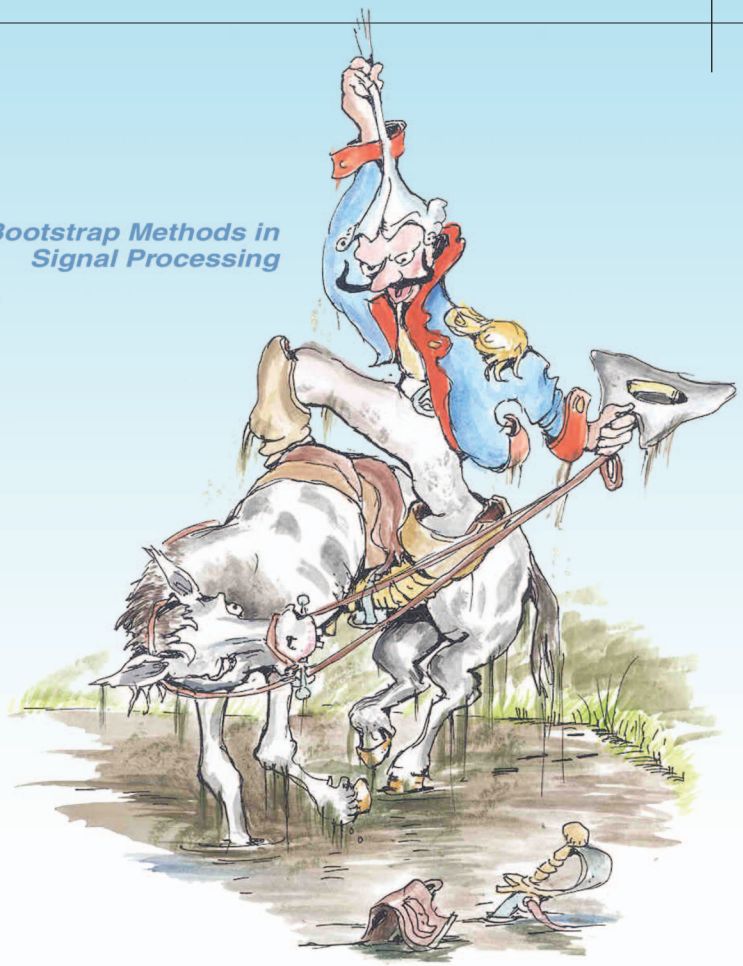
DOI: 10.1121/1.3385001 · Source: PubMed

1 author:

James Vincent Candy
Lawrence Livermore National Laboratory
251 PUBLICATIONS **2,087** CITATIONS

James V. Candy

Bootstrap Methods in Signal Processing



PROF. DR. KARL HEINRICH HOFMANN

Bootstrap Particle Filtering

Performance of a passive synthetic aperture in an uncertain ocean environment

In the real world, systems designed to extract signals from noisy measurements are plagued by errors evolving from constraints of the sensors employed, by random disturbances and noise, and, probably the most common, by the lack of precise knowledge of the underlying physical phenomenology generating the process in the first place. Thus, there is a strong need to incorporate any and all of the a priori knowledge available. Methods capable of extracting the desired signal from hostile environments require approaches that capture all of the information available and incorporate it into a processing scheme. This approach is model based [1], employing mathematical representations of the component processes involved, and the processor design evolves from the realm of statistical signal processing using a Bayesian approach. When signals are deeply buried in noise, especially as in ocean acoustics, then processing techniques utilizing these representations of the underlying phenomenology must be employed [2]. Here we develop the Bayesian model-based approach for processing highly uncertain ocean acoustic data.

The Bayesian approach to statistical signal processing includes the next generation of processors that have recently been realized with the advent of high-speed/high-throughput computers [3]. A brief tutorial of Bayesian nonlinear statistical signal processing techniques is presented commencing with an overview of Bayesian techniques and sequential processors [3]. Once the evolving Bayesian paradigm is established, simulation-based methods using sampling theory and Monte Carlo (MC) realizations are discussed [3]–[11]. Here the usual limitations of

1053-5888/07/$25.00©2007IEEE

nonlinear approximations and non-Gaussian processes prevalent in classical nonlinear processing algorithms (e.g., approximate Kalman filters) are no longer a restriction to perform Bayesian processing. It is shown how the underlying problem variables are easily assimilated into the Bayesian construct. Next, importance sampling methods are discussed and how they can be extended to sequential solutions is shown, implemented using Markovian models as their natural evolution [4]–[6]. With this in mind, the concept of a particle filter (PF), a discrete nonparametric representation of a probability distribution, is developed and shown how it can be implemented in a bootstrap manner using sequential importance sampling/resampling (SIR) methods to perform statistical inferences yielding a suite of popular estimators such as the conditional expectation, maximum a-posteriori, and median filters [7]–[15].

IN THE REAL WORLD, SYSTEMS DESIGNED TO EXTRACT SIGNALS FROM NOISY MEASUREMENTS ARE PLAGUED BY ERRORS EVOLVING FROM CONSTRAINTS OF THE SENSORS EMPLOYED.

With the generic bootstrap PF developed, we investigate the passive synthetic aperture problem from the signal processing perspective. We briefly define the problem and then develop a mathematical representation of a towed hydrophone sensor array in the ocean coupled to targets in random motion. Next, we develop a pragmatic simulation using this nonlinear space-time model discussing some of the important intricacies embedded within. We design the bootstrap PF for this problem and apply it to synthesized hydrophone data.

BAYESIAN APPROACH TO SIGNAL PROCESSING

Modern statistical signal processing techniques evolve directly from a Bayesian perspective, i.e., they are cast into a probabilistic framework using Bayes' theorem as the fundamental construct. More specifically, the information about a random signal, $x(t)$, required to solve a vast majority of estimation/processing problems is incorporated in the underlying probability distribution generating the process. For instance, the usual signal enhancement problem is concerned with providing the best (in some sense) estimate of the signal at time t based on all of the data available at that time. The corresponding distribution provides that information directly in terms of its underlying statistics. That is, by calculating the statistics of the process directly from its underlying distribution an enhanced signal can be extracted using a variety of estimators as well as evaluated by its set of performance statistics [1], [2].

Bayesian signal processing consists of three steps with the primary objective of estimating the posterior (after data is available) distribution governing the signal, $x(t)$, under investigation. The first step is gathering the underlying information (priors) governing evolution. The next step is to convert the underlying posterior distribution, $\hat{\Pr}(x(t)|Y_t)$, using Bayes' rule [3] and develop the algorithm for posterior estimation. It can be thought of as transforming the broad prior to the narrow posterior. The final step, after obtaining the posterior, is to decide on and extract meaningful statistics yielding the solution of the signal processing problem.

We define the unobserved signal or equivalently hidden variables as the set of N_x-vectors, $\{x(t)\}$, $t = 0, \ldots, N$ and the observables or equivalent measurements as the set of N_y-vectors, $\{y(t)\}$, $t = 0, \ldots, N$ considered to be conditionally independent of the signal variables. The goal in sequential Bayesian estimation is to sequentially estimate the conditional posterior distribution, $\Pr(x(0), \ldots, x(N)|y(0), \ldots, y(N))$. Once the posterior is estimated, then many of the interesting statistics characterizing the process under investigation can be exploited to extract meaningful information.

We start by defining two sets of random (vector) processes: $X_t := \{x(0), \ldots, x(t)\}$ and $Y_t := \{y(0), \ldots, y(t)\}$. Here we can consider X_t to be the set of dynamic random variables or parameters of interest and Y_t as the set of measurements or observations of the desired process. We start with Bayes' theorem for the conditional distribution as

$$\Pr(X_t|Y_t) = \frac{\Pr(Y_t|X_t) \times \Pr(X_t)}{\Pr(Y_t)}. \tag{1}$$

In Bayesian theory, the posterior defined by $\Pr(X_t|Y_t)$ is decomposed in terms of the prior $\Pr(X_t)$, its likelihood $\Pr(Y_t|X_t)$, and the evidence or normalizing factor, $\Pr(Y_t)$. Each has a particular significance in this construct.

It has been shown [3], [4] that under Markovian assumptions the distribution can be expressed, sequentially, as

$$\overbrace{\Pr(X_t|Y_t)}^{\text{New}} = \overbrace{\mathcal{W}(t, t-1)}^{\text{Weight}} \times \overbrace{\Pr(X_{t-1}|Y_{t-1})}^{\text{Old}}, \tag{2}$$

and the weight is defined by

$$\mathcal{W}(t, t-1) := \left[\frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{\Pr(y(t)|Y_{t-1})}\right].$$

This result is satisfying in the sense that we need only know the posterior distribution at the previous stage, $t-1$, scaled by a weighting function to sequentially propagate the posterior to the next stage. Even though this expression provides the full posterior solution, it is not physically realizable unless the distributions are known in closed form and the underlying multiple integrals or sums can be analytically determined. In fact, a more useful solution is the marginal posterior distribution [3], [4] given by the update recursion as

$$\overbrace{\Pr(x(t)|Y_t)}^{\text{Posterior}} = \frac{\overbrace{\Pr(y(t)|x(t))}^{\text{Likelihood}} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{\text{Prior}}}{\underbrace{\Pr(y(t)|Y_{t-1})}_{\text{Evidence}}}, \tag{3}$$

where we can consider the update or filtering distribution as a weighting of the prediction distribution as in the full case above, i.e.,

$$\overbrace{\Pr(x(t)|Y_t)}^{\text{Update}} = \overbrace{\mathcal{W}_u(t, t-1)}^{\text{Weight}} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{\text{Prediction}}, \tag{4}$$

where the weight in this case is defined by

$$\mathcal{W}_u(t, t-1) := \frac{\Pr(y(t)|x(t))}{\Pr(y(t)|Y_{t-1})}.$$

We summarize the sequential Bayesian processor in Table 1. These two sequential relations form the theoretical foundation of many of the sequential PF designs.

[TABLE 1] SEQUENTIAL BAYESIAN PROCESSOR FOR FILTERING POSTERIOR.

PREDICTION

$$\Pr(x(t)|Y_{t-1}) = \int \Pr(x(t)|x(t-1)) \times \Pr(x(t-1)|Y_{t-1}) dx(t-1)$$

UPDATE/POSTERIOR

$$\Pr(x(t)|Y_t) = \Pr(y(t)|x(t)) \times \Pr(x(t)|Y_{t-1}) / \Pr(y(t)|Y_{t-1})$$

INITIAL CONDITIONS

$$\Pr(x(0)|Y_0)$$

MONTE CARLO APPROACH

MC methods are stochastic computational algorithms capable of efficiently simulating highly complex systems. Historically, motivated by games of chance and encouraged by the development of the first electronic computer, the MC approach was conceived by Ulam (1945), developed by Ulam and von Neumann (1947), and coined by Metropolis (1949). It evolved in the 1940s during the Manhattan project by scientists investigating calculations for atomic weapon designs [6]. The method evolved from such areas as computational physics, biology, chemistry, mathematics, engineering, materials, and finance, to name a few. MC methods offer an alternative approach to solving classical numerical integration and optimization problems. It utilizes Markov chain theory as its underlying foundation, establishing the concept that through random sampling, the resulting empirical distribution converges to the desired posterior called the stationary or invariant distribution of the chain. Markov chain MC techniques are based on sampling from probability distributions based on a Markov chain, which is a stochastic (state-space) system governed by a transition probability, having the desired posterior distribution as its invariant distribution. Thus, under certain assumptions, the chain converges to the desired posterior through proper random sampling as the number of samples become large (see [4] and [6] for details)—a crucial property.

In signal processing, we are interested in some statistical measure of a random signal or parameter usually expressed in terms of its moments. For example, suppose we have some signal function $f(X)$ with respect to some underlying probabilistic distribution $\Pr(X)$, then a typical measure to seek is its performance on the average, which is characterized by the expectation

$$E_X\{f(X)\} = \int f(X)\Pr(X)dX. \tag{5}$$

Instead of attempting to use numerical integration techniques, stochastic sampling techniques known as MC integration have evolved. The key idea embedded in the MC approach is to represent the required distribution as a set of random samples rather than a specific analytic function (e.g., Gaussian). As the number of samples becomes large, they provide an equivalent (empirical) representation of the distribution, enabling moments to be estimated directly (inference).

MC integration draws samples from the required distribution and then forms sample averages to approximate the sought-after distributions. Thus, MC integration evaluates (5) by drawing samples, $\{X(i)\}$ from $\Pr(X)$ with $\sim$ defined as drawn from the designated distribution. Assuming perfect sampling, this produces the estimated or empirical distribution given by

$$\hat{\Pr}(X) \approx \frac{1}{N}\sum_{i=1}^{N} \delta(X - X(i)),$$

which is a probability mass distribution with weights, $1/N$ and random variable or location $X(i)$. Substituting the empirical distribution into the integral gives

$$E_X\{f(X)\} = \int f(X)\hat{\Pr}(X)dX \approx \frac{1}{N}\sum_{i=1}^{N} f(X(i)) \equiv \overline{f}, \tag{6}$$

which follows directly from the sifting property of the delta function. Here $\overline{f}$ is said to be an MC estimate of $E_X\{f(X)\}$.

A generalization to the MC approach is known as importance sampling, which evolves from

$$I = \int_X p(x)dx = \int_X \left(\frac{p(x)}{q(x)}\right) \times q(x)\, dx \quad \text{for} \int q(x)dx = 1. \tag{7}$$

Here $q(x)$ is referred to as the sampling distribution or more appropriately, the importance sampling distribution, since it samples the target distribution, $p(x)$, nonuniformly giving more importance to some values of $p(x)$ than others. We say that the support of $q(x)$ covers that of $p(x)$, i.e., the samples drawn from $q(\cdot)$ overlap the same region (or more) corresponding to the samples of $p(\cdot)$. The integral in (7) can be estimated by the following:

- draw N-samples

$$X(i) \sim q(x) \text{ and } \hat{q}(x) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(x - X(i))$$

- compute the sample mean

$$I = E_q\left\{\frac{p(x)}{q(x)}\right\} \approx \int \left(\frac{p(x)}{q(x)}\right) \times \frac{1}{N}\sum_{i=1}^{N}\delta(x - X(i))dx$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{p(X(i))}{q(X(i))}.$$

The art in importance sampling is in choosing the importance distribution $q(\cdot)$, which approximates the target distribution $p(\cdot)$ as closely as possible. This is the principal factor effecting performance of this approach, since variates must be drawn from $q(x)$ that cover the target distribution. Using the concepts of importance sampling, we can approximate the posterior distribution with a function on a finite discrete support. Since it is usually not possible to sample directly from the posterior, we use importance sampling coupled with an easy to sample proposal distribution, say $q(X_t|Y_t)$—this is the crucial choice and design step required in Bayesian importance sampling methodology. Therefore, starting with a function of the set of variables, say $f(X_t)$, we would like to estimate its mean using the importance concept, i.e.,

$$E\{f(X_t)\} = \int f(X_t) \times \Pr(X_t|Y_t)\, dX_t, \tag{8}$$

where $\Pr(X_t|Y_t)$ is the posterior distribution.

Using the MC approach, we would like to sample from this posterior directly and then use sample statistics to perform the estimation. Therefore, we insert the proposal importance distribution $q(X_t|Y_t)$ as before

$$\hat{f}(t) := E\{f(X_t)\} = \int f(X_t)\left[\frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)}\right] \times q(X_t|Y_t)\, dX_t. \tag{9}$$

Now we apply Bayes' rule to the posterior distribution and define a weighting function as

$$\tilde{W}(t) := \frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)} = \frac{\Pr(Y_t|X_t) \times \Pr(X_t)}{\Pr(Y_t) \times q(X_t|Y_t)}. \tag{10}$$

Unfortunately, $\tilde{W}(t)$ is not useful because it requires knowledge of the evidence or normalizing constant $\Pr(Y_t)$ given by

$$\Pr(Y_t) = \int \Pr(Y_t|X_t) \times \Pr(X_t) dX_t, \tag{11}$$

which is usually not available. But by substituting (10) into (9) and defining a new weight $W(t)$, we obtain

$$\hat{f}(t) = \frac{1}{\Pr(Y_t)}\int f(X_t)\left[\frac{\Pr(Y_t|X_t)\Pr(X_t)}{q(X_t|Y_t)}\right] q(X_t|Y_t)dX_t$$
$$= \frac{1}{\Pr(Y_t)}\int W(t) f(X_t) q(X_t|Y_t)\, dX_t, \tag{12}$$

which is simply the expectation of the weighted function $E_q\{W(t)f(X_t)\}$ scaled by the normalizing constant. From this definition of the new weighting function, we have

$$W(t)q(X_t|Y_t) = \Pr(Y_t|X_t)\Pr(X_t). \tag{13}$$

Thus, we can now replace the troublesome normalizing constant of (12) by using (13) to replace the integrand of (11), i.e.,

$$\hat{f}(t) = \frac{E_q\{W(t)f(X_t)\}}{\Pr(Y_t)} = \frac{E_q\{W(t)f(X_t)\}}{\int W(t)q(X_t|Y_t)dX_t}$$
$$= \frac{E_q\{W(t)f(X_t)\}}{E_q\{W(t)\}}. \tag{14}$$

Now drawing samples from the proposal $X_t(i) \sim q(X_t|Y_t)$ and using the MC approach leads to the desired result. That is, from the perfect sampling distribution, we have that

$$\hat{q}(X_t|Y_t) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(X_t - X_t(i)), \tag{15}$$

and therefore substituting, applying the sifting property of the Dirac delta function, and defining the normalized weights

$$\mathcal{W}_i(t) := \frac{W_i(t)}{\sum_{i=1}^{N} W_i(t)} \quad \text{for}$$
$$W_i(t) = \frac{\Pr(Y_t|X_t(i)) \times \Pr(X_t(i))}{q(X_t(i)|Y_t)}, \tag{16}$$

we obtain the final estimate

$$\hat{f}(t) \approx \sum_{i=1}^{N}\mathcal{W}_i(t) \times f(X_t(i)). \tag{17}$$

The importance estimator is biased being the ratio of two sample estimators [as in (14)], but it can be shown that it asymptotically converges to the true statistic and the central limit theorem holds [5], [6]. Thus, as the number of samples increase ($N \to \infty$), an asymptotically optimal estimate of the posterior is

$$\hat{\Pr}(X_t|Y_t) \approx \sum_{i=1}^{N}\mathcal{W}_i(t) \times \delta(X_t - X_t(i)), \tag{18}$$

which is the goal of Bayesian estimation. Note that the new weight, $W(t) \propto \tilde{W}(t)$, where $\propto$ is defined as "proportional to" up to a normalizing constant.

The importance distribution can be modified to enable a sequential estimation of the desired posterior distribution, i.e., we estimate the posterior $\hat{\Pr}(X_{t-1}|Y_{t-1})$ using importance weights $\mathcal{W}(t-1)$. As a new sample becomes available, we

estimate the new weights $\mathcal{W}(t)$ leading to an updated estimate of the posterior $\hat{\Pr}(X_t|Y_t)$. This means that to obtain the new set of samples, $X_t(i) \sim q(X_t|Y_t)$ sequentially, we must use the previous set of samples, $X_{t-1}(i) \sim q(X_{t-1}|Y_{t-1})$. Thus, with this in mind, the importance distribution $q(X_t|Y_t)$ must admit a marginal distribution $q(X_{t-1}|Y_{t-1})$, implying a Bayesian factorization

$$\begin{aligned} q(X_t|Y_t) &= q(x(t), X_{t-1}|Y_t) \\ &= q(X_{t-1}|Y_{t-1}) \times q(x(t)|X_{t-1}, Y_t). \end{aligned} \tag{19}$$

This type of importance distribution leads to the desired sequential solution [7]. Recall the Bayesian solution of (2). Recognizing the denominator as just the evidence or normalizing distribution and not a function of X_t, we have

$$\begin{aligned} \Pr(X_t|Y_t) \propto \Pr(y(t)|x(t)) &\times \Pr(x(t)|x(t-1)) \\ &\times \Pr(X_{t-1}|Y_{t-1}). \end{aligned} \tag{20}$$

Substituting this expression for the posterior in the weight relation, we have

$$\begin{aligned} W(t) = \frac{\Pr(X_t|Y_t)}{q(X_t|Y_t)} &= \Pr(y(t)|x(t)) \\ &\times \frac{\Pr(x(t)|x(t-1))}{q(x(t)|X_{t-1}, Y_t)} \times \underbrace{\frac{\Pr(X_{t-1}|Y_{t-1})}{q(X_{t-1}|Y_{t-1})}}_{\text{Previous Weight}}, \end{aligned} \tag{21}$$

which can be written simply as

$$W(t) = W(t-1) \times \frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{q(x(t)|X_{t-1}, Y_t)}, \tag{22}$$

giving us the desired relationship—a sequential updating of the weight at each time-step. These results then enable us to formulate a generic Bayesian sequential importance sampling algorithm as follows:

- draw samples from the proposed importance distribution: $x_i(t) \sim q(x(t)|X_{t-1}, Y_t)$
- determine the required conditional distributions: $\Pr(x_i(t)|x(t-1))$, $\Pr(y(t)|x_i(t))$
- calculate the unnormalized weights: $W_i(t)$ with $x(t) = x_i(t)$
- normalize the weights: $\mathcal{W}_i(t)$
- estimate the posterior distribution: $\hat{\Pr}(X_t|Y_t) = \sum_{i=1}^{N} \mathcal{W}_i(t)\delta(x(t) - x_i(t))$.

Once the posterior is estimated, then desired statistics evolve directly.

MODERN STATISTICAL SIGNAL PROCESSING TECHNIQUES EVOLVE DIRECTLY FROM A BAYESIAN PERSPECTIVE.

BAYESIAN APPROACH TO THE STATE SPACE

Bayesian estimation relative to the state-space models is based on extracting the unobserved or hidden dynamic internal variables (states) from noisy measurement data. The Markovian state vector with initial distribution $\Pr(x(0))$ propagates temporally throughout the state space according to the conditional probabilistic transition distribution $\Pr(x(t)|x(t-1))$, while the conditionally independent measurements evolve from the likelihood distribution $\Pr(y(t)|x(t))$. We see that the dynamic state variable at time t is obtained through the transition probability based on the previous state (Markovian property), $x(t-1)$, and the knowledge of the underlying conditional probability. Once propagated to time t, the dynamic state variable is used to update or correct based on the likelihood probability and the new measurement, $y(t)$. Symbolically, this evolutionary process is $x(t-1) \sim \Pr(x(t)|x(t-1))$; $\{x(t), y(t)\} \sim \Pr(y(t)|x(t))$.

The usual model-based constructs of the dynamic state variables indicate that there is an equivalence between the probabilistic distributions and the underlying state/measurement transition models. The functional discrete Markovian (state) representation is given by the models: $x(t) = A(x(t-1), u(t-1), w(t-1))$ and $y(t) = C(x(t), u(t), v(t))$, where w and v are the respective process and measurement noise sources with u a known input. Here $A(\cdot)$ is the nonlinear (or linear) dynamic state transition function and $C(\cdot)$ the corresponding measurement function. Both conditional probabilistic distributions embedded within the Bayesian framework are completely specified by these functions and the underlying noise distributions $\Pr(w(t-1))$ and $\Pr(v(t))$. That is, we have the equivalence

$$\begin{aligned} A(x(t-1), u(t-1), w(t-1)) \Rightarrow \Pr(x(t)|x(t-1)) \\ \Leftrightarrow \mathcal{A}(x(t)|x(t-1)) \\ C(x(t), u(t), v(t)) \Rightarrow \Pr(y(t)|x(t)) \\ \Leftrightarrow \mathcal{C}(y(t)|x(t)). \end{aligned} \tag{23}$$

Thus, the state-space model along with the noise statistics and prior distributions define the required Bayesian representation or probabilistic propagation model, defining the evolution of the states and known inputs through the transition probabilities. Here, the dynamic state variables propagate throughout the state space specified by the transition probability ($\mathcal{A}(x(t)|x(t-1))$) using the embedded process model. That is, the unobserved state at time $t-1$ depends on the transition probability distribution to propagate to the state at time t. Once evolved, the state combines under the corresponding measurement at time t through the conditional likelihood distribution ($\mathcal{C}(y(t)|x(t))$) using the embedded measurement model to obtain the required likelihood distribution. These events continue to evolve throughout with the states propagating through the state transition probability using the process model and the measurements generated by the states and likelihood using the measurement model.

Analytically, to generate the model-based version of the sequential Bayesian processor, we replace the transition and

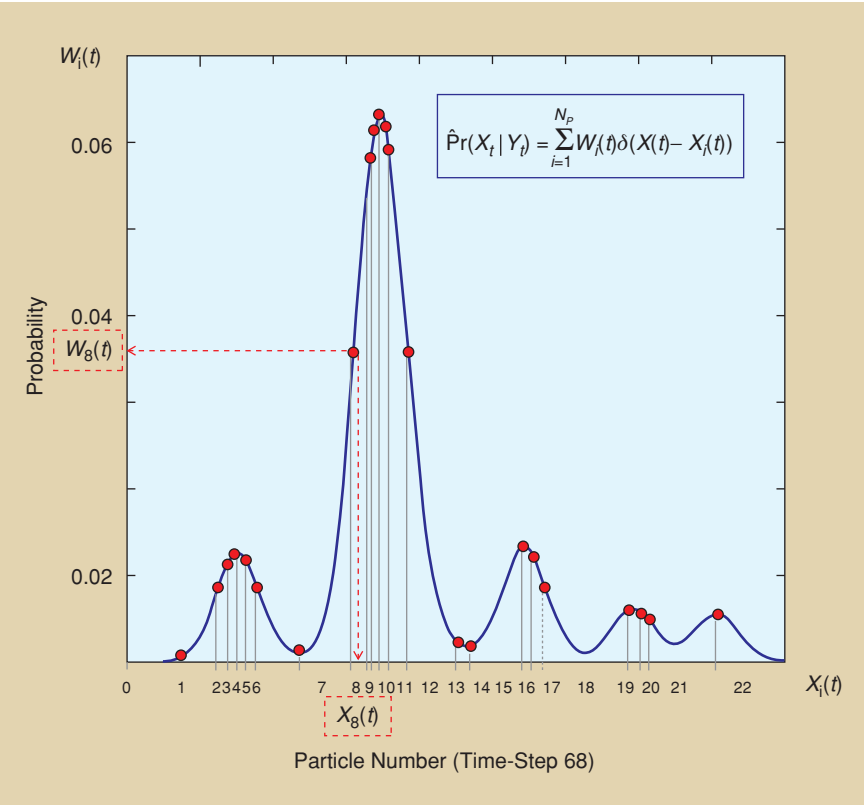
likelihood distributions with the conditionals of (23). The solution to the signal enhancement or equivalently state estimation problem is given by the filtering distribution $\Pr(x(t)|Y_t)$, which was solved previously (see Table 1). We start with the prediction recursion characterized by the Chapman-Kolmogorov equation [6], replacing transition probability with the implied model-based conditional, i.e.,

$$\Pr(x(t)|Y_{t-1}) = \int \overbrace{\mathcal{A}(x(t)|x(t-1))}^{\text{Embedded Process Model}} \times \overbrace{\Pr(x(t-1)|Y_{t-1})}^{\text{Prior}}\, dx(t-1). \quad (24)$$

Next, we incorporate the model-based likelihood into the posterior equation with the understanding that the process model has been incorporated into the prediction

$$\Pr(x(t)|Y_t) = \overbrace{\mathcal{C}(y(t)|x(t))}^{\text{Embedded Measurement Model}} \times \overbrace{\Pr(x(t)|Y_{t-1})}^{\text{Prediction}} \Big/ \Pr(y(t)|Y_{t-1}). \quad (25)$$

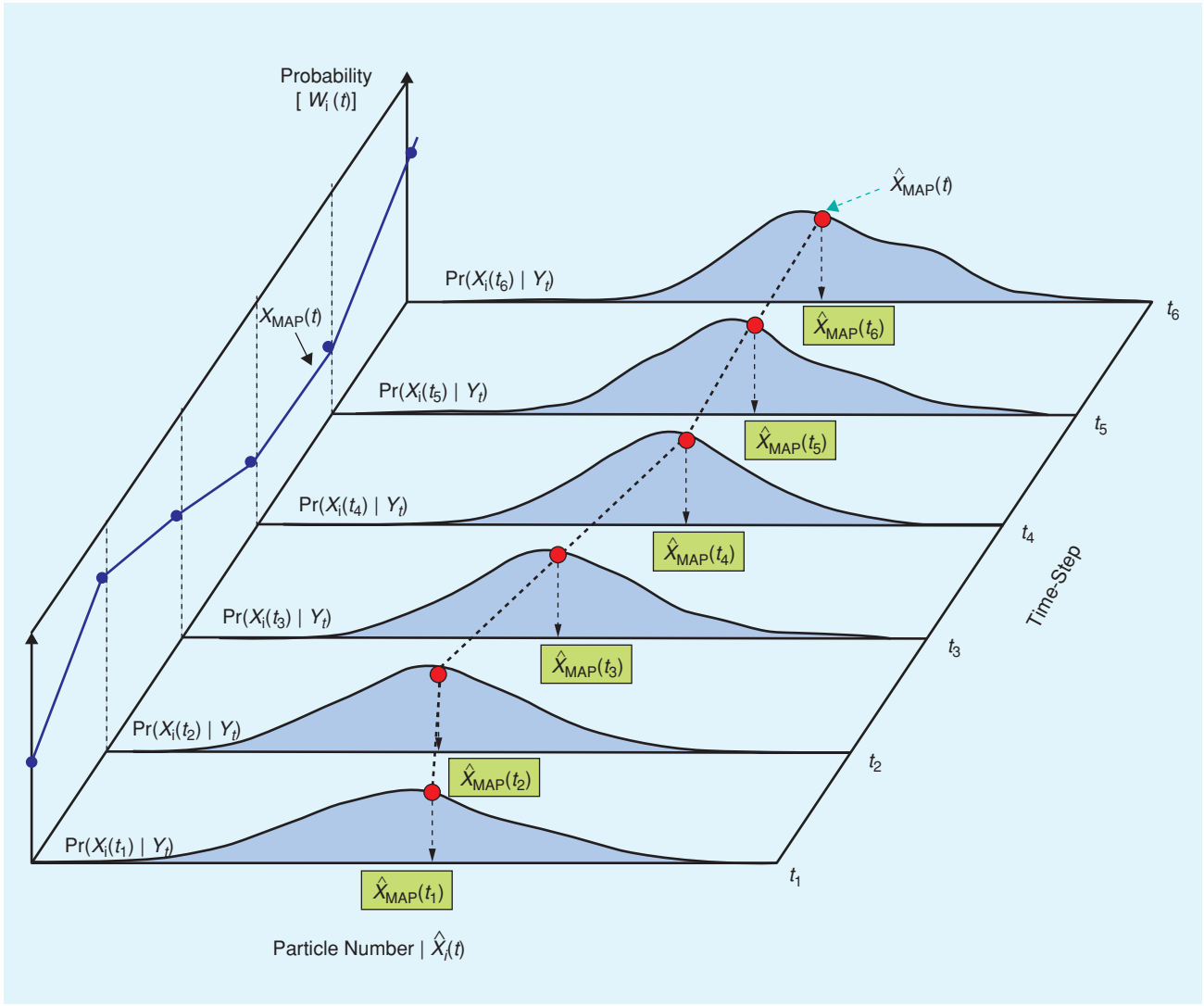
Thus, we see from the Bayesian perspective that the sequential Bayesian processor employing the state-space representation of (23) is straightforward.



[FIG1] PF representation of posterior probability distribution in terms of weights (probabilities) and particles (samples).

BAYESIAN PARTICLE FILTERS

Particle filtering is a sequential MC method employing the sequential estimation of relevant probability distributions using the concepts of importance sampling and the approximations of distributions with discrete random measures [7]–[11]. The key idea is to represent the required posterior distribution by a set of N_p-random samples, the particles, and associated weights, $\{x_i(t), \mathcal{W}_i(t)\}$; $i = 1, \cdots, N_p$ and compute the required MC estimates. Of course, as the number of samples become very large, the MC representation becomes an equivalent characterization of the posterior distribution. Thus, particle filtering is a technique to implement sequential Bayesian processors by MC simulation. It is an alternative to approximate Kalman filtering for nonlinear problems [1], [2], [7]. In particle filtering, continuous distributions are approximated by discrete random measures composed of these weighted particles where the particles are actually samples (realizations) of the unknown or hidden states from the Markovian state-space and the weights are the associated probability masses estimated using the Bayesian recursions as shown in Figure 1. From the figure, we see that associated with each particle $x_i(t)$ is a corresponding weight or (probability) mass $\mathcal{W}_i(t)$. Therefore, knowledge of this random measure $\{x_i(t), \mathcal{W}_i(t)\}$ characterizes an estimate of the empirical posterior distribution $\hat{\Pr}(x(t)|Y_t)$ at a particular instant of time t. Importance sampling plays a crucial role in state-space particle algorithm development. Particle filtering does not involve linearizations around current estimates but, rather, approximations of the desired distributions by these discrete measures. In comparison, the approximate Kalman filter recursively estimates the conditional mean and covariance that can be used to characterize the filtering posterior $\Pr(x(t)|Y_t)$ under Gaussian assumptions [1]. In summary, a PF is a sequential MC based point mass representation of probability distributions. It only requires a Markovian state-space representation of the underlying process to provide a set of particles that evolve at each time step leading to an instantaneous approximation of the target posterior distribution of the state at time t given all of the data up to that time. Figure 1 illustrates the evolution of the posterior at a particular time step. Here we see the estimated posterior based on 22 particles (nonuniformly spaced), and we select the eighth particle and weight to illustrate the instantaneous approximation at time t for x_i versus $\hat{\Pr}(x(t)|Y_t)$. The full posterior over six time steps is shown in Figure 2, where each individual time slice is displayed. The associated maximum a posteriori



[FIG2] Ensemble of instantaneous posterior probability distributions in terms of probabilities (weights), particles and time-steps ($\hat{\Pr}(x(t)|Y_t)$ versus $X_i(t)$ versus t). MAP inference annotated on ensemble along with the corresponding dynamic estimate, $\hat{X}_{MAP}(t)$ (left axis).

(MAP) estimates are illustrated across the ensemble along with the actual dynamic estimates extracted at each time-step (left side plot). Thus, statistics are calculated across the ensemble created over time to provide the inference estimates of the states. For example, the minimum mean-squared error (MMSE) estimate is easily determined by averaging over $x_i(t)$, since

$$\begin{aligned}\hat{x}_{\text{mmse}}(t) &= \int x(t)\Pr(x(t)|Y_t)dx \approx \int x(t)\hat{\Pr}(x(t)|Y_t)dx \\ &= \frac{1}{N_p}\sum_{i=1}^{N_p} x(t)\mathcal{W}_i(t)\delta(x(t)-x_i(t)) \\ &= \frac{1}{N_p}\sum_{i=1}^{N_p} \mathcal{W}_i(t)x_i(t),\end{aligned}$$

while the MAP estimate is simply determined by finding the sample corresponding to the maximum weight of $x_i(t)$ across the ensemble at each time step, i.e.,

$$\hat{x}_{\text{MAP}}(t) = \max_{x_i}\{\hat{\Pr}(x(t)|Y_t)\}. \tag{26}$$

The corresponding state-space PF (SSPF) evolving from this construct follows directly after sampling from the importance distribution, i.e.,

$$\begin{aligned}x_i(t) &\sim q(x(t)|x(t-1), y(t)) \\ W_i(t) &= W_i(t-1) \times \frac{\mathcal{C}(y(t)|x_i(t)) \times \mathcal{A}(x_i(t)|x_i(t-1))}{q(x_i(t)|x_i(t-1), y(t))} \\ \mathcal{W}_i(t) &= \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)},\end{aligned} \tag{27}$$

and the filtering posterior is estimated by

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t) \times \delta(x(t) - x_i(t)). \tag{28}$$

BOOTSTRAP PARTICLE FILTER

The basic design tool when developing the particle filtering algorithms is the choice of the importance sampling distribution, $q(\cdot)$. One of the most popular realizations of this approach was developed by using the transition prior as the importance proposal [3]. This prior is defined in terms of the state-space representation by $\mathcal{A}(x(t)|x(t-1)) \rightarrow A(x(t-1), u(t-1), w(t-1))$, which is dependent on the known excitation and process noise statistics. It is given by

$$\begin{aligned} q_{\text{prior}}(x(t)|x(t-1), Y_t) &\longrightarrow \Pr(x(t)|x(t-1)) \\ &\longrightarrow \mathcal{A}(x(t)|x(t-1)). \end{aligned}$$

Substituting this choice into the expression for the weights gives

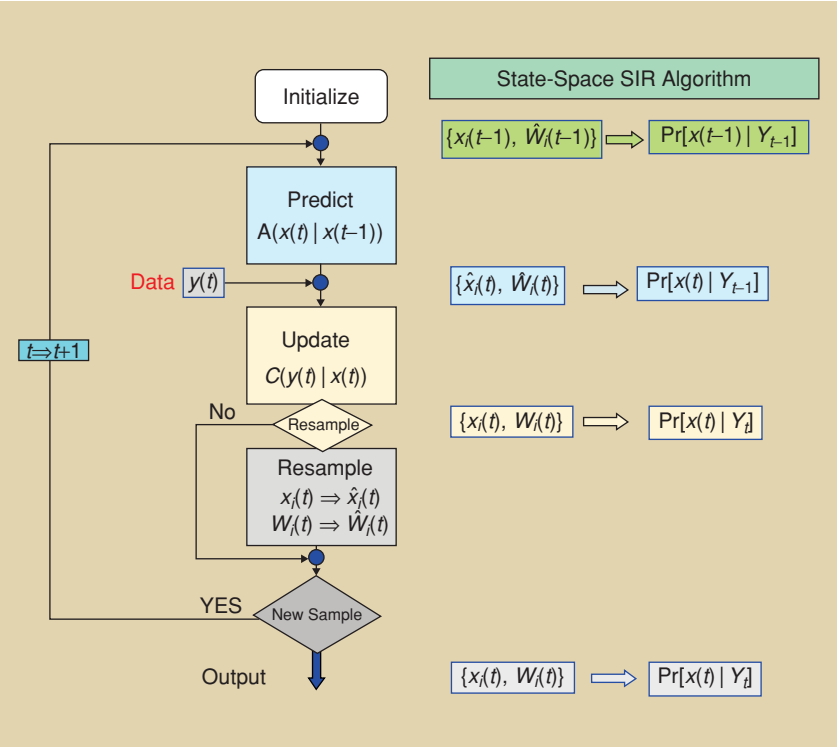
$$\begin{aligned} W_i(t) &= W_i(t-1) \times \frac{\mathcal{C}(y(t)|x_i(t)) \times \mathcal{A}(x(t)|x_i(t-1))}{q_{\text{prior}}(x(t)|x_i(t-1), Y_t)} \\ &= W_i(t-1) \times \mathcal{C}(y(t)|x_i(t)). \end{aligned}$$

Note two properties resulting from this choice of importance distribution. First, the proposed importance distribution does not use the most recent observation, $y(t)$, and second, this choice is easily implemented and updated by simply evaluating the measurement likelihood $\mathcal{C}(y(t)|x_i(t))$; $i = 1, \ldots, N_p$ for the sampled particle set. These weights require the particles to be propagated to time t before the weights can be calculated.

This choice can lead to problems, however, since the transition prior is not conditioned on the measurement data, especially the most recent, $y(t)$. Failing to incorporate the latest available information from the most recent measurement to propose new values for the states leads to only a few particles having significant weights when their likelihood is calculated. The transitional prior is a much broader distribution than the likelihood indicating that only a few particles will be assigned a large weight. The algorithm will degenerate rapidly. Thus, the SSPF algorithm takes the same generic form as before with the importance weights much simpler to evaluate with this approach. It has been called the bootstrap PF, the condensation PF, or the survival of the fittest algorithm [3], [12].

As mentioned, one of the major problems with the importance sampling algorithms is the depletion of the particles, i.e., they tend to increase in variance at each iteration. The degeneracy of the particle weights creates a problem that must be resolved before these particle algorithms can be of any pragmatic use in applications. The problem occurs because the variance of the importance weights can only increase in time [3] thereby making it impossible to avoid this weight degradation. Degeneracy implies that a large computational effort is devoted to updating particles whose contribution to the posterior is negligible. Thus, there is a need to somehow resolve this problem to make the simulation-based techniques viable. This requirement leads to the idea of resampling the particles.

The main objective in simulation-based sampling techniques is to generate independent identically distributed (i.i.d.) samples from the targeted posterior distribution to perform statistical inferences extracting the desired information. Thus, the importance weights are quite critical since they contain probabilistic information about each specific particle. In fact, they provide us with information about how probable a sample has been drawn from the target posterior [3], [4]. Therefore, the weights can be considered acceptance probabilities enabling us to generate independent (approximately) samples from the posterior $\Pr(x(t)|Y_t)$. The empirical distribution $\hat{\Pr}(x(t)|Y_t)$ is defined over a set of finite (N_p) random measures $\{x_i(t), W_i(t)\}$; $i = 1, \ldots, N_p$, approximating the posterior. Resampling, therefore, can be thought of as a realization of enhanced particles $\hat{x}_k(t)$ extracted from the original samples $x_i(t)$ based on their acceptance probability $W_i(t)$ at time t, i.e., statistically, we have



[FIG3] State-space SIR (bootstrap) particle filtering algorithm structure: initialization, prediction (state transition), updating (likelihood), resampling.

$$\Pr(\hat{x}_k(t) = x_i(t)) = W_i(t) \qquad \text{for } i = 1, \ldots, N_p \tag{29}$$

or symbolically, $\hat{x}_k(t) \Rightarrow x_i(t)$, with the set of new particles $\{\hat{x}_k(t)\}$ replacing the old set $\{x_i(t)\}$ and $\Rightarrow$ is defined as resampled from.

The fundamental concept in resampling theory is to preserve particles with large weights (large probabilities) while discarding those with small weights. Two steps must occur to resample effectively: 1) a decision, on a weight-by-weight basis, must be made to select the appropriate weights and reject the inappropriate and 2) resampling must be performed to minimize the degeneracy. The overall strategy, when coupled with importance sampling is termed SIR [3], [4]. A reasonable measure of degeneracy is the effective particle sample size [6] defined by

ONE OF THE MAJOR PROBLEMS WITH THE IMPORTANCE SAMPLING ALGORITHMS IS THE DEPLETION OF THE PARTICLES.

$$N_{\text{eff}}(t) := \frac{N_p}{E_q\{W^2(X_t)\}} \approx \frac{1}{\sum_{i=1}^{N_p} W_i^2(t)} \leq N_p, \tag{30}$$

and the decision is made by comparing it to a threshold, N_{thresh}, i.e., we use the rejection method [6] to decide whether or not to resample

$$\hat{N}_{\text{eff}}(t) = \begin{cases} \text{Resample} & \leq \ N_{\text{thresh}} \\ \text{Accept} & > \ N_{\text{thresh}}. \end{cases}$$

Once the decision is made to resample, a uniform sampling procedure can be applied removing samples with low importance weights and multiplying samples with high importance weights, generating a set of new samples so that $\Pr(\hat{x}_i(t) = x_j(t)) = \hat{W}_j(t)$. The resulting i.i.d. samples are uniform such that the sampling induces the mapping of $\{x_i(t), W_i(t)\} \rightarrow \{\hat{x}_i(t), \hat{W}_i(t)\}, \hat{W}_i(t) = 1/N_p \ \forall i$. Resampling decreases the degeneracy problem algorithmically but introduces its own set of problems. Theoretically, after one resampling step, the simulated trajectories are not statistically independent any longer; therefore, the simple convergence resulting under these assumptions lose their validity [3]. We summarize the bootstrap PF algorithm in Table 2 and illustrate the algorithm in Figure 3.

SYNTHETIC APERTURE PROCESSING FOR A TOWED ARRAY

Synthetic aperture processing is well known in airborne radar but not as familiar in sonar [16]–[21]. The underlying idea in creating a synthetic aperture is to increase the array length by motion, thereby increasing spatial resolution (bearing) and gain in signal-to-noise ratio (SNR). It has been shown that for stationary targets, the motion induced bearing estimates have smaller variance than that of a stationary array [2], [19]. Here we extend this to the case of both array and target motion, resulting in a further increase in SNR.

We define the acoustic array space-time processing problem as:

Given a set of noisy pressure-field measurements from a horizontally towed array of L-sensors in motion, find the best (minimum error variance) estimate of the target bearing.

We use the following nonlinear pressure-field measurement model for M monochromatic {plane wave} targets characterized by a corresponding set of temporal frequencies, bearings, and amplitudes, $[\{\omega_m\}, \{\theta_m\}, \{a_m\}]$; i.e.,

$$p(x, t_k) = \sum_{m=1}^{M} a_m e^{j\omega_m t_k - j\beta(x,t_k)\sin\theta_m} + n(t_k), \tag{31}$$

where $\beta(x, t_k) := k_o x(t_o) + v t_k$, $k_o = (2\pi/\lambda_o)$ is the wavenumber, $x(t_k)$ is the current spatial position along the x-axis in meters, v is the tow speed in m/s, and $n(t_k)$ is the additive random noise. The inclusion of motion in the generalized wave number β is critical to the improvement of the processing, since the synthetic aperture effect is actually created through the motion itself and not simply the displacement.

If we further assume that the single sensor equation above is expanded to include an array of L-sensors, $x \longrightarrow x_\ell$, $\ell = 1, \ldots, L$; then we obtain

$$\begin{aligned} p(x_\ell, t_k) &= \sum_{m=1}^{M} a_m e^{j\omega_m t_k - j\beta(x_\ell, t_k)\sin\theta_m} + n_\ell(t_k) \\ &\rightarrow \sum_{m=1}^{M} a_m \cos(\omega_m t_k - \beta(x_\ell, t_k)\sin\theta_m) + n_\ell(t_k) \end{aligned} \tag{32}$$

[TABLE 2] BOOTSTRAP SIR STATE-SPACE PARTICLE FILTERING ALGORITHM.

INITIALIZE	
$x_i(0) \sim \Pr(x(0)) \qquad W_i(0) = \frac{1}{N_p} \quad i = 1, \ldots, N_p$	[Sample]
IMPORTANCE SAMPLING	
$x_i(t) \sim \mathcal{A}(x(t)\|x_i(t-1)); \ w_i \sim \Pr(w_i(t))$	[State Transition]
WEIGHT UPDATE	
$W_i(t) = W_i(t-1) \times \mathcal{C}(y(t)\|x_i(t))$	[Weights]
WEIGHT NORMALIZATION	
$\mathcal{W}_i(t) = \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)}$	[Normalize]
RESAMPLING	
If $\hat{N}_{\text{eff}}(t) \leq N_{\text{thresh}}$, then $\hat{x}_i(t) \Rightarrow x_j(t)$	[Resample]

since our hydrophone sensors measure the real part of the complex pressure field, the final nonlinear measurement model of the system can be written in compact vector form as

$$\mathbf{p}(t_k) = \mathbf{c}[t_k; \Theta] + \mathbf{n}(t_k), \tag{33}$$

where $\mathbf{p}, \mathbf{c}, \mathbf{n} \in C^{L\times 1}$, are the respective pressure-field, measurement and noise vectors and $\Theta \in \mathcal{R}^{M\times 1}$ represents the target bearings and the corresponding vector measurement model

$$\mathbf{c}_\ell(t_k; \Theta) = \sum_{m=1}^{M} a_m \cos(\omega_m t_k - \beta(x_\ell, t_k) \sin\theta_m) \quad \text{for } \ell = 1, \ldots, L.$$

Since we model the bearings as a random walk (constants ($\dot{\Theta} = \mathbf{0}$) and noise) emulating random target motion, then an augmented Markovian state-space model evolves from first differences as

$$\Theta(t_k) = \Theta(t_{k-1}) + \mathbf{w}_\theta(t_{k-1}). \tag{34}$$

Thus, the state-space model is linear with no explicit dynamics, therefore, the process matrix $A = I$ (identity) and the relations are greatly simplified.

Now let us see how a PF using the bootstrap approach would be constructed according to the generic algorithm of Table 2. For this problem, we assume the additive noise sources are Gaussian, so we can compare results to the performance of the approximate processor. We will return to our original discrete notation $t_{k+1} \to t+1$ for the sampled-data representation.

BAYESIAN SEQUENTIAL PROCESSOR DESIGN

Let us cast this problem into the sequential Bayesian framework, i.e., we would like to estimate the instantaneous posterior filtering distribution $\hat{\Pr}(x(t)|Y_t)$ using the PF representation to be able to perform inferences and extract the target bearing estimates. Therefore, for the Bayesian processor and our problem, we have that the state transition probability is given by $(\Theta(t) \to x(t))$

$$\Pr(x(t)|x(t-1)) \longrightarrow \mathcal{A}(x(t)|x(t-1)) \sim \mathcal{N}(\Theta(t) : \mathbf{a}[\Theta(t-1)], \mathbf{R}_{w_\theta, w_\theta}),$$

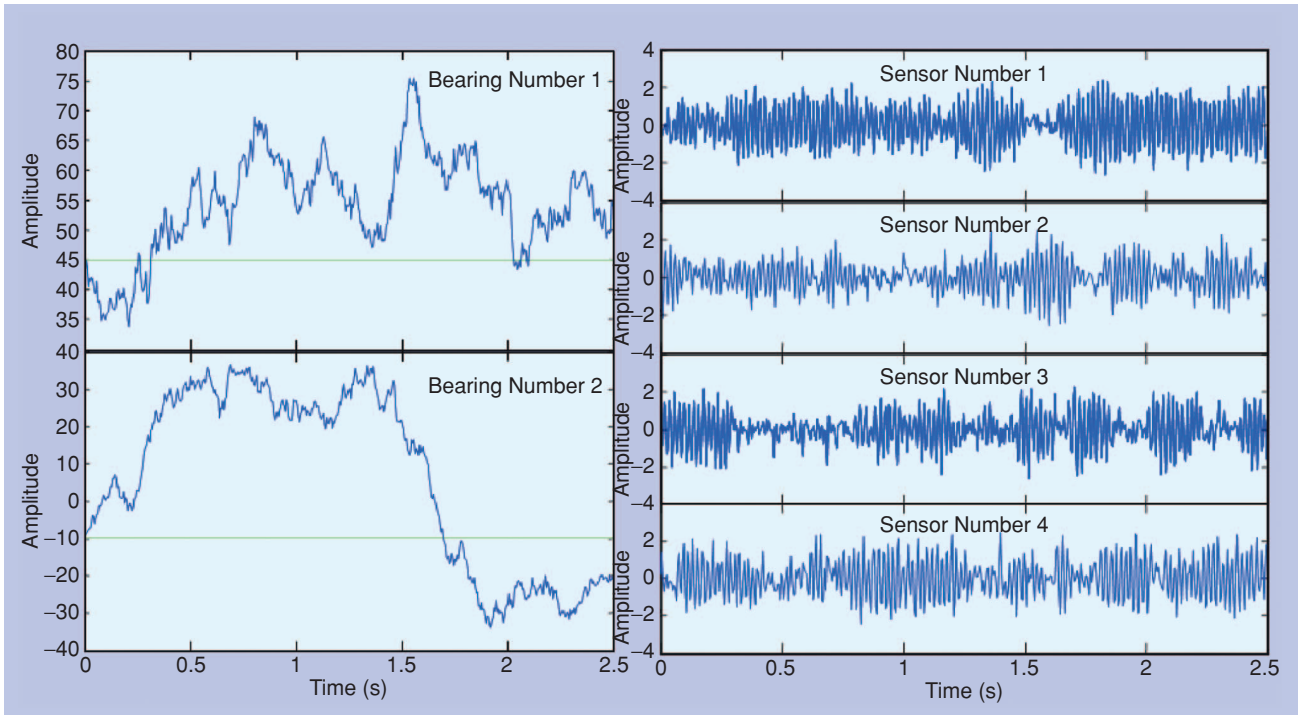
or in terms of our state transition (bearings) model, we have

$$\Theta(t) = \mathbf{a}[\Theta(t-1)] + w_\theta(t-1) = \Theta(t-1) + \mathbf{w}_\theta(t-1) \quad \text{for } \Pr(\mathbf{w}_\theta(t)) \sim \mathcal{N}(0, \mathbf{R}_{w_\theta w_\theta}).$$

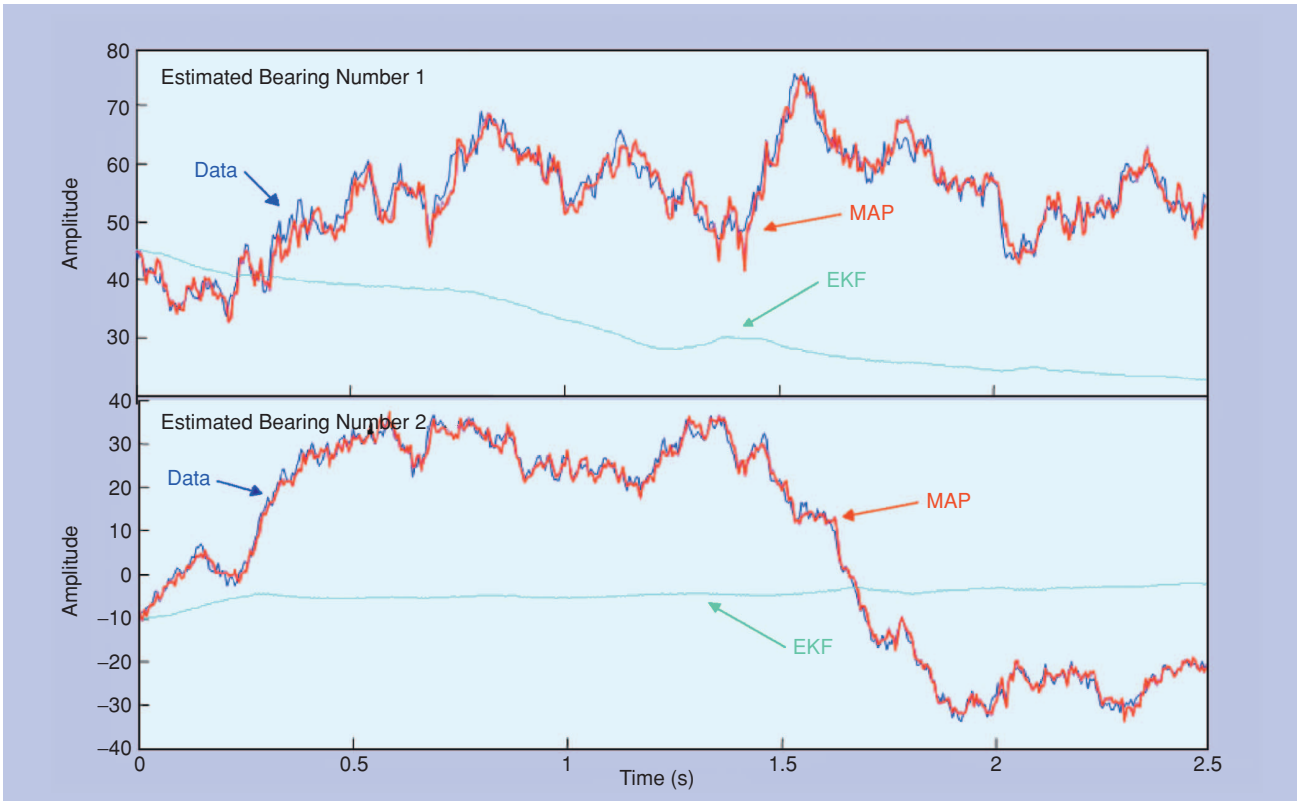
The corresponding likelihood is specified in terms of the measurement model $(y(t) \to p(t))$ as

$$\Pr(y(t)|x(t)) \longrightarrow \mathcal{C}(y(t)|x(t)) \sim \mathcal{N}(\mathbf{y}(t) : \mathbf{c}[\Theta(t)], \mathbf{R}_{vv}(t)),$$

where we have used the notation $z \sim \mathcal{N}(z : m_z, R_{zz})$ to specify the Gaussian distribution in random vector z. In terms of our



[FIG4] **Synthetic aperture sonar tracking problem: simulated target motion from initial bearings of 45° and −10° and array measurements (−10 dB SNR).**



[FIG5] **PF bearing estimates: 45° PF bearing (state) estimates and simulated target tracks (EKF, MAP).**

problem, we have that

$$\ln \mathcal{C}(y(t)|x(t)) = \kappa - \frac{1}{2}(\mathrm{y}(t) - \sum_{m=1}^{M} a_m \cos(\omega_m t - \beta(t)\sin\theta_m))' \mathrm{R}_{vv}^{-1} \cdot \left(\mathrm{y}(t) - \sum_{m=1}^{M} a_m \cos(\omega_m t - \beta(t)\sin\theta_m)\right)$$

with κ a constant, $\beta \in \mathcal{R}^{L\times 1}$, and $\beta(t) := [\beta(x_1, t)| \ldots |\beta(x_L, t)]'$, the dynamic wavenumber expanded over the array. Thus, the SIR algorithm becomes

- draw samples (particles) from the state transition distribution:

$$\Theta_i(t) \sim \mathcal{N}(\Theta(t) : \mathrm{a}[\Theta(t-1)], \mathrm{R}_{w_\theta, w_\theta})$$
$$w_{\theta_i}(t) \sim \Pr(w(t)) \sim \mathcal{N}(0, R_{w_{\theta_i} w_{\theta_i}}),$$
$$\Theta_i(t) = \Theta_i(t-1) + w_{\theta_i}(t-1)$$

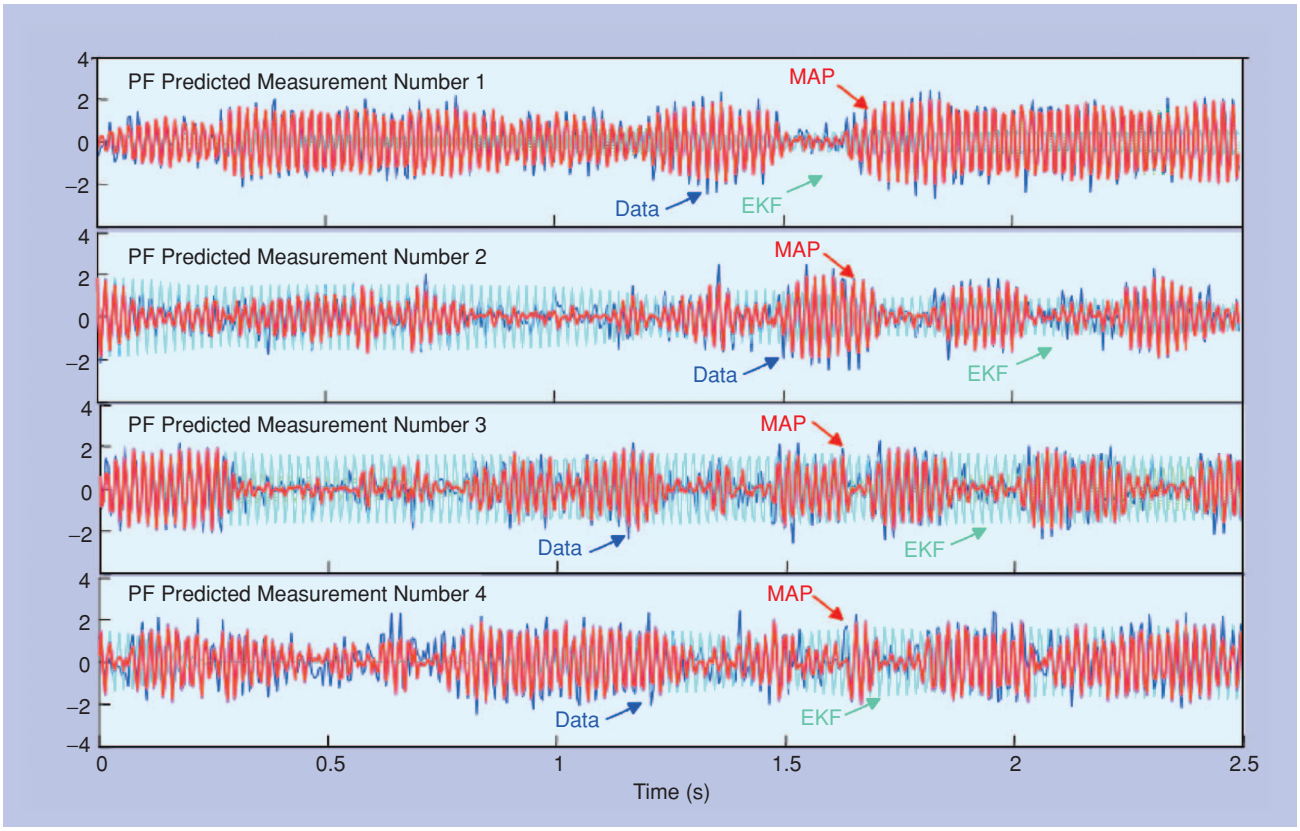
- estimate the likelihood, $\mathcal{C}(\mathrm{y}(t)|\Theta(t)) \sim \mathcal{N}(\mathrm{y}(t) : \mathrm{c}[\Theta(t)], \mathrm{R}_{vv}(t))$ with $c_\ell(t; \Theta_i) = \sum_{m=1}^{M} a_m \cos(\omega_m t_k - \beta(x_\ell, t)\sin\Theta_{m,i}(t))$ for $\ell = 1, \ldots, L$ and $\Theta_{m,i}$ is the ith-particle at the mth-bearing angle
- update and normalize the weight: $\mathcal{W}_i(t) = W_i(t)/\sum_{i=1}^{N_p} W_i(t)$
- resample: $\hat{N}_{\mathrm{eff}}(t) \le N_{\mathrm{thresh}}$
- estimate the instantaneous posterior: $\hat{\Pr}(\Theta(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t)\delta(\Theta(t) - \Theta_i(t))$
- perform the inference by estimating the corresponding statistics: $\hat{\Theta}_{\mathrm{map}}(t) = \arg\max \hat{\Pr}(\Theta(t)|Y_t)$; $\hat{\Theta}_{\mathrm{mmse}}(t) = E\{\Theta(t)|Y_t\} = \sum_{i=1}^{N_p} \Theta_i(t)\hat{\Pr}(\Theta(t)|Y_t)$; $\hat{\Theta}_{\mathrm{median}}(t) = \mathrm{median}(\hat{\Pr}(\Theta(t)|Y_t)))$.

NARROWBAND SYNTHETIC APERTURE SIMULATION

Consider the following simulation of the synthetic aperture using a four-element, linear towed array with two moving targets with the following parameters:

- *Targets*: Unity amplitudes with temporal frequency is 50 Hz, wavelength = 30 m, speed = 5 m/s
- *Array*: four-element linear towed array with 15 m pitch
- *PF*: $N_\theta = 2$ states (bearings), $N_y = 4$ sensors, $N = 500$ samples, $N_p = 1,000$ weights
- SNR: −10 dB
- *Noise:* White Gaussian with $R_{ww} = \mathrm{diag}\ (2.5)$, $R_{vv} = \mathrm{diag}\ (0.1414)$; sampling interval 0.005 s
- *Initial Conditions*: (bearings and uncertainty) $\Theta_o = [45° - 10°]'$, $P_o = \mathrm{diag}\ (10^{-10})$.

The array simulation was executed as the target moved according to a random walk specified by the process noise and sensor array measurements with −10 dB SNR. The results are shown in Figure 4. In Figure 4(a) we see the noisy synthesized bearings and in Figure 4(b) we see four noisy sensor measurements at the moving array. The bearing (state) estimates are shown in Figure 5, where we observe the targets making a variety of course alterations. The PF is able to track the target motions quite well while we observe the extended Kalman filter



[FIG6] **PF measurement estimates: four channel hydrophone estimates.**

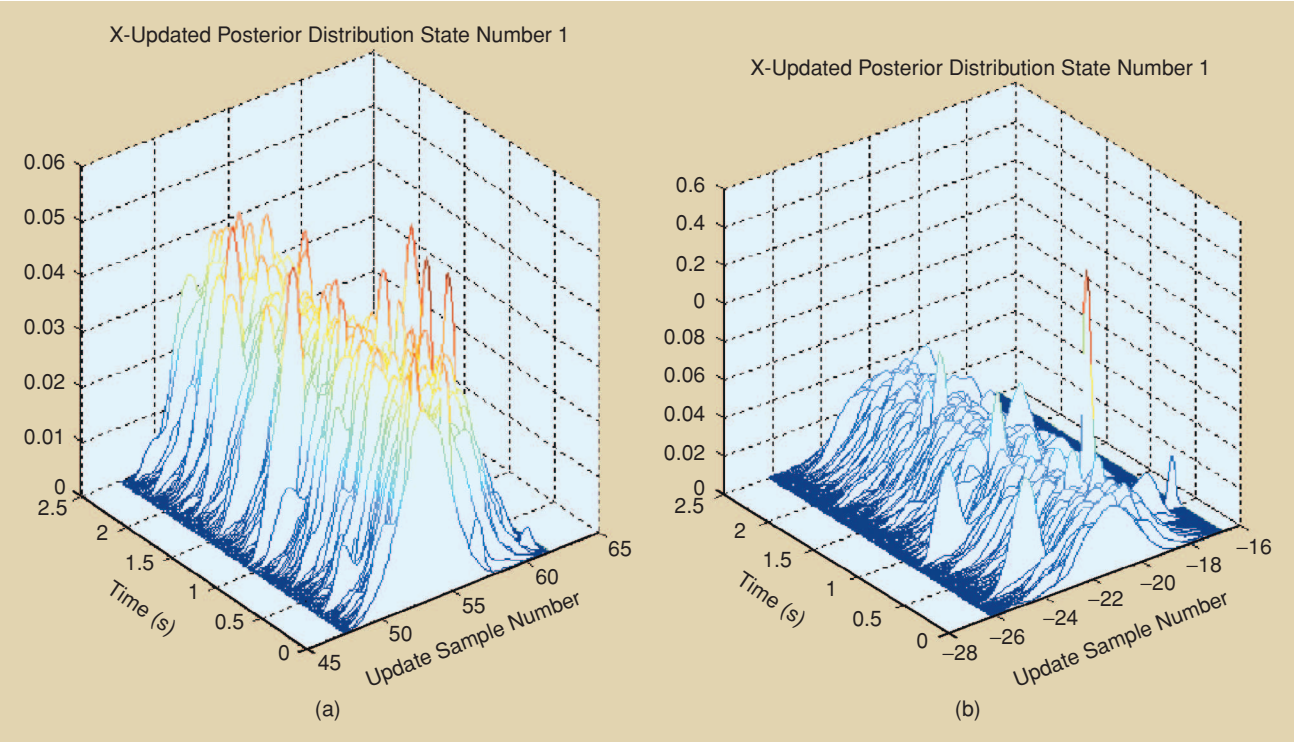
(EKF) [1] unable to respond quickly enough and finally losing track completely. Both the MAP and MMSE estimates using the estimated posterior provide excellent tracking. Note that these bearing inputs would provide the raw data for an XY-tracker [1]. The PF estimated or filter measurements are shown in Figure 6. As expected, the PF tracks the measurement data quite well while the EKF is again in error. Using the usual optimality tests for performance demonstrates that the PF processor works well, since each measurement channel is zero-mean and white with the weighted sum square residual (WSSR) statistic lying below the threshold. This indicates a white innovations sequence, demonstrating the tracking ability of the PF processor, at least in a classical sense [1]. The instantaneous posterior distributions for both bearing estimates are shown in Figure 7. Here we see the Gaussian nature of the bearing estimates generated by the random walk. Clearly, the PF performs quite well for this problem.

It is also interesting to note how the performance of this processor changes by changing the number of particles selected as well as the number of resampling steps required. These steps are controlled by the threshold (N_{thresh}), i.e., if $N_{\text{eff}} < N_{\text{thresh}}$, resampling occurs; otherwise, it does not. The performance is quantified by a measure of accuracy (RMS error) and computational cost (number resampling steps/run). We selected the number of particles as $N_p = 1{,}000,\ 500,$ and 250 with corresponding threshold, $N_{\text{thresh}} = N_p$. As expected, in all three cases, $(N_p = 1{,}000,\ 500,\ 250)$, the minimum root mean-squared error (RMS) occurred by resampling at every time-step ($N_{\text{thresh}} = N_p$) with RMS state-error pairs for each respective N_p given by: $(\text{RMS}_1, \text{RMS}_2) = (8\%, 6\%)$; $(8.2\%, 6.1\%)$; $(8.3\%, 6.2\%)$, indicating that the RMS state-error increases by using fewer and fewer particles and resampling at every time-step.

Next, using $N_p = 1{,}000$, we varied N_{thresh} in 100 particle increments from 500 down to 100 particles, demanding fewer resampling steps. The resulting RMS errors for each selected threshold were, respectively, $(N_{\text{thresh}} = 500, 400, 300, 200, 100) \longrightarrow (\text{RMS}_1, \text{RMS}_2) = (8\%, 6\%)$; $(9\%, 6.5\%)$; $(10\%, 7.8\%)$; $(11\%, 8.2\%)$; $(14.6\%, 14.6\%)$, with corresponding resampling steps of $(N_{\text{resample}} = 500, 370, 187, 119, 67)$, clearly demontrating the tradeoff between accuracy and resampling. So we see by this analysis that as the number of resampling steps is diminished (threshold setting), the accuracy decreases but so does computational cost, as expected. In all cases, except the $N_{\text{thresh}} = 100$, the innovations sequence produced by the PF remained white and zero-mean, classical properties indicating successful performance [1]. Finally, for the 100 particle threshold, the performance deteriorated, as indicated by the nonwhite innovations and the increased RMS errors.

SUMMARY

In this article, we have provided an overview of nonlinear statistical signal processing based on the Bayesian paradigm. We showed that the next-generation processors are well founded on MC simulation-based sampling techniques. We reviewed the



[FIG7] **PF instantaneous posterior bearing estimates: 45° and −10° targets.**

development of the sequential Bayesian processor using the state-space models. The popular bootstrap algorithm was outlined and applied to an ocean acoustic synthetic aperture towed array target tracking problem to test the performance of a particle filtering technique.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Eneregy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48 while Dr. Candy was on sabbatical leave at the University of Cambridge, U.K., collaborating with Prof. Simon Godsill of the Signal Processing Group. The author would also like to thank the helpful and constructive comments of the reviewers who helped to significantly improve this manuscript.

AUTHOR

James V. Candy (tsoftware@aol.com) is the chief scientist for engineering at Lawrence Livermore National Laboratory and an adjunct professor at the University of California, Santa Barbara. He is a Fellow of the IEEE and the Acoustical Society of America and Life Member (Fellow) at the University of Cambridge (Clare Hall College). He received the IEEE Distinguished Technical Achievement Award for the development of model-based signal processing in ocean acoustics and is an IEEE Distinguished Lecturer for oceanic processing. He authored *Signal Processing: The Model-Based Approach* and *Signal Processing: The Modern Approach* (McGraw-Hill, 1986 and 1988, respectively) and *Model-Based Signal Processing* (Wiley/IEEE Press, 2006).

REFERENCES

[1] J.V. Candy, *Model-Based Signal Processing*. Hoboken, NJ: Wiley, 2006.
[2] E.J. Sullivan and J.V. Candy, "Space-time array processing: The model-based approach," *J. Acoust. Soc. Amer.*, vol. 102, no. 5, pp. 2809–2820, 1997.
[3] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.
[4] O. Cappe, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*. New York: Springer-Verlag, 2005.
[5] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer-Verlag, 1995.
[6] J. Liu, *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag, 2001.
[7] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA: Artech House, 2004.
[8] S. Godsill and P. Djuric, "Monte Carlo methods for statistical signal processing," *IEEE Trans. Signal Processing (Special Issue)*, vol. 50, no. 2, pp. 173–499, 2002.
[9] P. Djuric, J. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Mag.* vol. 20, no. 5, pp. 19–38, 2003.
[10] S. Haykin and N. de Freitas, "From Kalman filters to particle filters," *Proc. IEEE (Special Issue on Sequential State Estimation)*, vol. 92, no. 3, pp. 399–574, 2004.
[11] A. Doucet and X. Wang, "Monte Carlo methods for signal processing," *IEEE Signal Processing Mag.*, vol. 24, no. 5, pp. 152–170, 2005.
[12] N. Gordon, D. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Proc. Inst. Elect. Eng.*, vol. 140, no. 2, pt. F, pp. 107–113, 1993.
[13] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*. New York: Springer-Verlag, 1996.
[14] M. Isard and A. Blake, "Condensation–conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.
[15] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
[16] R.E. Williams, "Creating an acoustic synthetic aperture in the ocean," *J. Acoust. Soc. Amer.*, vol. 60, no. 1, pp. 60–73, 1976.
[17] N. Yen and W.M. Carey, "Applications of synthetic aperture processing to towed array data," *J. Acoust. Soc. Amer.*, vol. 86, no. 2, pp. 754–765, 1989.
[18] S. Stergiopoulus and E.J. Sullivan, "Extended towed array processing by an overlap correlator," *J. Acoust. Soc. Amer.*, vol. 86, no. 1, pp. 158–171, 1989.
[19] E.J. Sullivan, W.M. Carey, and S. Stergiopoulus, "Editorial in special issue on acoustic synthetic aperture processing," *IEEE J. Ocean. Eng.*, vol. 17, no. 1, pp. 1–7, 1992.
[20] D.B. Ward, E.A. Lehmann, and R.C. Willianson, "Particle filtering algorithm for tracking and acoustic source in a reverberant environment," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.
[21] M. Orton and W.C. Fitzgerald, "Bayesian approach to tracking multiple targets using sensor arrays and particle filters," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 216–223, 2002. [SP]