GEOPHYSICS, VOL. 78, NO. 3 (MAY-JUNE 2013); P. V87–V100, 8 FIGS., 2 TABLES.
10.1190/GEO2012-0180.1

Geophysical signal processing using sequential Bayesian techniques

Caglar Yardim[1], Peter Gerstoft[1], and Zoi-Heleni Michalopoulou[2]

ABSTRACT

Sequential Bayesian techniques enable tracking of evolving geophysical parameters via sequential observations. They provide a formulation in which the geophysical parameters that characterize dynamic, nonstationary processes are continuously estimated as new data become available. This is done by using prediction from previous estimates of geophysical parameters, updates stemming from physical and statistical models that relate seismic measurements to the unknown geophysical parameters. In addition, these techniques provide the evolving uncertainty in the estimates in the form of posterior probability density functions. In addition to the particle filters (PFs), extended, unscented, and ensemble Kalman filters (EnKFs) were evaluated. The filters were compared via reflector and nonvolcanic tremor tracking examples. Because there are numerous geophysical problems in which the environmental model itself is not known or evolves with time, the concept of model selection and its filtering implementation were introduced. A multiple model PF was then used to track an unknown number of reflectors from seismic interferometry data. We found that when the equations that define the geophysical problem are strongly nonlinear, a PF was needed. The PF outperformed all Kalman filter variants, especially in low signal-to-noise ratio tremor cases. However, PFs are computationally expensive. The EnKF is most appropriate when the number of parameters is large. Because each technique is ideal under different conditions, they complement each other and provide a useful set of techniques for solving sequential geophysical inversion problems.

INTRODUCTION

There have been significant developments in sequential Bayesian techniques in the past decade due to advances in theoretical signal processing and a rapid increase in computational power. Sequential Bayesian filtering combines information on parameter evolution, a function that relates geophysical measurements to unknown quantities, and a statistical model for the random perturbations in the measurements.

These methods can roughly be classified as fast analytical methods using a Kalman framework that incorporates certain Gaussian and/or linearity assumptions and numerical sequential Monte Carlo (MC) methods commonly known as "particle filters" (PFs) used when the linear/Gaussian system restrictions of a Kalman filter (KF) are too restrictive for the problem (Ristic et al., 2004). In addition, there are hybrid methods such as the ensemble KF (EnKF) (Evensen, 2009) and its application in Gaussian mixtures (Dovera and Della Rossa, 2011) that incorporate different aspects of KFs and PFs.

The sequential estimation framework can be applied to many time-evolving geophysical observations such as microseismic activity or reservoir monitoring. Some geophysical applications that adopt a sequential approach are geodesy, in which a KF is used (Segall and Matthews, 1997), seismic strains (Llenos and McGuire, 2011) using an extended KF (EKF), and passive seismic monitoring (Baziw, 2005), in which PFs are used. Additionally, many problems in geoacoustics have been addressed mainly with the PF (Yardim et al., 2009, 2010, 2012). A good review of PFs and ensemble methods for meteorology data assimilation problems is given in van Leeuwen (2009). Reflector tracking can also be modeled as a sequential estimation problem by using range as the evolving index (Nicoli et al., 2002). An interesting application is earthquake forecasting based on data assimilation, in which sequential MC methods have been used for the renewal processes (Werner et al., 2011).

Manuscript received by the Editor 14 May 2012; revised manuscript received 27 September 2012; published online 12 April 2013.
[1]University of California San Diego, Scripps Institution of Oceanography, La Jolla, California, USA. E-mail: cyardim@ucsd.edu; gerstoft@ucsd.edu.
[2]New Jersey Institute of Technology, Department of Mathematical Sciences, Newark, New Jersey, USA. E-mail: michalop@njit.edu.

We demonstrate the application of EKFs, unscented KFs (UKFs), and PFs to nonvolcanic tremor (NVT) tracking (Obara, 2002; Rogers and Dragert, 2003; Peng and Gomberg, 2010) and multiple model PFs (MMPFs) to tracking reflectors in a seismic record section constructed using seismic interferometry (Curtis et al., 2006; Gerstoft et al., 2008; Wapenaar et al., 2008; Schuster, 2009; Siderius et al., 2010).

First observed a decade ago by Obara (2002) and Rogers and Dragert (2003), NVTs are characterized by low amplitudes, lack of high frequencies, emerging onsets, and duration of minutes to days. NVTs are mostly observed near megathrust earthquake source areas, suggesting that a detailed understanding of these may help characterize catastrophic events. The lack of strong impulsive phases across the array makes classic seismic methods less fruitful, but a frequency-domain approach with longer observation times combined with sequential estimation is a good alternative. Here we will track the NVT location using the 72-element Big Skidder seismic array located in Cascadia, WA.

In the second example, seismic interferometry constructs a seismic reflection sequence by interferometric processing of active or passive seismic data (Curtis et al., 2006; Wapenaar et al., 2008; Schuster, 2009). In the example used here, ambient noise on a vertical drifting array is used to build the reflection sequence of seabed reflectors (Gerstoft et al., 2008; Siderius et al., 2010). The interface depths and the reflection amplitudes are unknown and changing as the array drifts in the ocean. In addition, the number of interfaces is unknown and changes with time as well, as new reflectors emerge and old ones disappear. This requires MMPF implementation with a varying order of interfaces.

In this paper, the basics of sequential filters are summarized, starting from the KF and its variants and proceeding with PFs and multiple model formulations. First, the basic state-space, sequential Bayesian formulation, and Kalman framework are introduced. We then proceed with PFs, introducing sequential importance resampling (SIR). We examine and compare filters and present examples, illustrating practical challenges and solutions. Detailed descriptions of individual methods with full theoretical derivations can be found in signal-processing papers (Arulampalam et al., 2002; Gustafsson et al., 2002; Candy, 2007; Cappé et al., 2007; Djurić and Bugallo, 2010; Yardim et al., 2011) and texts on sequential Bayesian methods (Doucet et al., 2001a; Ristic et al., 2004; Candy, 2009; Evensen, 2009). These methods are subsequently used for NVT and interferometric tracking.

SEQUENTIAL BAYESIAN FORMULATION

Let $\mathbf{y}_t$ be the measurement vector (for example, a response along a seismic array) at step t, and let $\mathbf{x}_t$ represent the state vector (for example, the time-varying location of the source of the seismic signal $\mathbf{y}_t$), where $t = 1, \ldots, T$. The state vector dimension n_x could be known and fixed or unknown and varying with t. A major goal is to estimate parameters in $\mathbf{x}_t$ that evolve sequentially with time or space. As data $\mathbf{y}_t$ become available, the unknown parameters forming the state vector are estimated sequentially using the collective data history and prior knowledge on evolution of the state. Two equations define a state-space model:

$$\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{v}_t), \tag{1}$$

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{w}_t). \tag{2}$$

The state equation, equation 1, describes the evolution or transition of $\mathbf{x}_t$ with t. Function $\mathbf{f}_t$ is known and relates the state vector at step t to that at step $t-1$. Variable $\mathbf{v}_t$ is the process or state noise and has a known probability density function (PDF) $p(\mathbf{v}_t)$.

The measurement equation (or observation), equation 2, relates measurements $\mathbf{y}_t$ to state vector $\mathbf{x}_t$ through a known function $\mathbf{h}_t$. Quantity $\mathbf{w}_t$ is the measurement noise with a PDF $p(\mathbf{w}_t)$.

The states follow a first-order Markov process in which $\mathbf{x}_t$ is only dependent on state parameters at the previous time step, resulting in a probability $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. In addition, measurement $\mathbf{y}_t$ only depends on $\mathbf{x}_t$ and is independent of the state at other time steps for a given $\mathbf{x}_t$.

The state and measurement noise terms $\mathbf{v}_t$ and $\mathbf{w}_t$ can be additive, multiplicative, or incorporated in the state and measurement through complex functions of $\mathbf{f}_t$ and $\mathbf{h}_t$, respectively. The formulation includes fully dynamic, nonstationary cases, in which, in addition to the state vector $\mathbf{x}_t$ and data $\mathbf{y}_t$, functions $\mathbf{f}_t$ and $\mathbf{h}_t$ and noise components $\mathbf{v}_t$ and $\mathbf{w}_t$ can all change with t.

The objective of a sequential Bayesian technique is to track the evolution of the multidimensional (n_x-D) posterior PDF of $\mathbf{x}_t$. This enables any desired statistical quantity (such as mean, covariance, mode, maximum a posteriori estimate, credible intervals, and marginal posterior PDFs of any desired parameter) to be computed at will. In cases with non-Gaussian, high-dimensional PDFs, this computation is not easy and requires more advanced methods, as we will see. Assume data have been collected sequentially at $t = 1{:}T$. Defining $\mathbf{y}_{1:t} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t]$ as the set of data observed at the first t steps and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t]$ as the sequence of unknown state vectors, the desired posterior PDF is given by $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. Filtering enables all the previous and current measurements to be used in estimating $\mathbf{x}_t$. Compare this to a classical geophysical inversion problem in which the output is given as $p(\mathbf{x}_t|\mathbf{y}_t)$. The estimate of $\mathbf{x}_t$ does not use information from previous or future measurements. In a sequential problem, it is computationally inefficient to perform each inversion independently with t. Moreover, the estimates can fluctuate significantly from step to step, when all the information available at t is not used.

A complementary step to filtering is smoothing, where the desired output PDF is given by $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, $T \geq t$. A smoother is appropriate in applications in which future data have already been observed and are readily available. Therefore, past and future measurements can be exploited for the calculation of $p(\mathbf{x}_t|\mathbf{y}_{1:T})$. Smoothing adds a computational overhead to traditional forward filtering. Although inclusion of future data improves estimation in comparison to a one-way filtering approach, the increase in computational cost sometimes makes smoothing prohibitive. There are two main classes of smoothers: forward-backward and two-filter algorithms. The forward-backward method runs a normal filter and performs a correction starting from T and going backward in time. The two-filter method runs two independent filters, one normal filter and one starting from T and going backward. Then the results of these two filters are merged to obtain the smoothing density. Previous geophysical smoothing examples include lithology/fluid prediction (Larsen et al., 2006; Ulvmoen and Hammer, 2010; Rimstad and Omre, 2013), in which a forward-backward smoother is used in addition to a hidden Markov model that corresponds to the state-space formulation described by equations 1 and 2 (Scott, 2002). Forward-backward and two-filter smoothers are also used in NVT tracking (Yardim and Gerstoft, 2012). This paper focuses on the implementation of filtering approaches.

Using the Bayes theorem and the Markov property results in

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \tag{3}$$

where the likelihood $\mathcal{L}(\mathbf{x}_t) = p(\mathbf{y}_t|\mathbf{x}_t)$ obtained from the new data $\mathbf{y}_t$ is combined with the prior knowledge $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ to provide an estimate of $\mathbf{x}_t$. It is possible to compute $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ as a function of $\mathbf{x}_{t-1}$ by starting from the joint PDF $p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ and integrating out $\mathbf{x}_{t-1}$:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}, \\ &= \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}. \end{aligned} \tag{4}$$

Note that $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$ because, given $\mathbf{x}_{t-1}$, any data from $1:t-1$ become irrelevant. The posterior PDF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is then expressed as a function of the posterior at the previous step $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ by inserting the integral form of $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ back into equation 3:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}. \tag{5}$$

Sequential Bayesian techniques use this formulation, through which the evolving posterior PDF can be computed recursively as new data $\mathbf{y}_t$ become available.

When state and measurement equations 1 and 2 are linear and the underlying PDFs are Gaussian, equation 5 can be computed analytically. This optimum sequential estimator is the KF (Kalman, 1960). In this case, one only needs to propagate the mean $\hat{\mathbf{x}}_t$ and covariance $\hat{\mathbf{P}}_t$ estimates at each step. The output of the filter is

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\hat{\mathbf{x}}_t, \hat{\mathbf{P}}_t). \tag{6}$$

As long as the problem is not strongly nonlinear and/or non-Gaussian, it might be possible to approximate the problem in a Kalman framework and, hence, the output is approximately in the form of equation 6. For example, if the problem is "mildly" nonlinear, then equations 1 and 2 can be approximated by linear functions using their Jacobians. Alternatively, statistical linearization with UKFs (van der Merwe et al., 2001) may be considered. Those are shown to provide improved results, as will be discussed later in more detail.

However, for strongly nonlinear/non-Gaussian problems, a PF is needed. The non-Gaussian posterior PDFs are approximated by creating a set of $i = 1, \ldots N_p$ particles $\mathbf{x}_t^i$, each with weight w_t^i, where

$$\chi_t : \{\mathbf{x}_t^i, w_t^i\}_{i=1}^{N_p} \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) \cong \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i). \tag{7}$$

The purpose of a PF is tracking in time t these particles and their weights as they pass through nonlinear equations in equations 1 and 2. The posterior PDF can be constructed at any time from the current locations and weights of the particles using equation 7. This enables us to compute any desired quantity such as the minimum mean square error estimate, variance, and marginal distributions by taking the integral of the posterior

$$I = \int g(\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t})\mathrm{d}\mathbf{x}_t, \tag{8}$$

where $g(\mathbf{x}_t) = \mathbf{x}_t$ for the mean, $g(\mathbf{x}_t) = (\mathbf{x}_t - \mu_{\mathbf{x}})^2$ for the variance, and $g(\mathbf{x}_t) = \delta(\mathbf{x}_t - \mathbf{x}_t(i))$ for the ith element marginal distribution.

KALMAN FILTERS

Kalman (1960) realized that the sequential Bayesian problem can be solved analytically if the evolving PDF is Gaussian. He came up with his famous KF, which is shown to be an optimal estimator in the minimum mean squared error sense. Having a Gaussian posterior for all steps t implies that state $\mathbf{v}_t$ and measurement $\mathbf{w}_t$ noise components are additive and Gaussian, $\mathbf{f}_t$ and $\mathbf{h}_t$ are known linear functions of the state and measurement vectors, and the prior $p(\mathbf{x}_0)$ is Gaussian. This reduces the system in equations 1 and 2 to

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{v}_t, \tag{9}$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t, \tag{10}$$

where $\mathbf{F}_t$ and $\mathbf{H}_t$ are, respectively, the matrices associated to the linear operators $\mathbf{f}_t$ and $\mathbf{h}_t$, with state and measurement noise covariances denoted as $\mathbf{Q}_t$ and $\mathbf{R}_t$. Because a Gaussian density is uniquely and completely defined in equation 6 by its first two moments, mean and covariance, the KF only needs to recompute $\{\hat{\mathbf{x}}_t, \hat{\mathbf{P}}_t\}$ from $\hat{\mathbf{x}}_{t-1}$, $\hat{\mathbf{P}}_{t-1}$ and take into account the new data $\mathbf{y}_t$. This gives us the two-step procedure that forms the KF given in Table 1, which is repeated at each t:

1) Predict: This stage predicts the current value of the state given its previous value using equation 9. This is represented by $\hat{\mathbf{x}}_{t|t-1}$ and $\hat{\mathbf{P}}_{t|t-1}$.
2) Update: This stage updates the estimate by checking how well the predicted value fits with the new data $\mathbf{y}_t$ and correcting if there is any difference. This is done by computing the error term $\mathbf{y}_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}$, called the innovation. Then the KF corrects the predicted state vector by this error multiplied by the Kalman gain $\mathbf{K}_t$, as shown in Table 1. This provides the final posterior PDF represented by $\hat{\mathbf{x}}_{t|t}$ and $\hat{\mathbf{P}}_{t|t}$.

Table 1. Kalman and extended Kalman filters.

Predict:
$\hat{\mathbf{x}}_{t\|t-1} = \mathbf{f}_t(\hat{\mathbf{x}}_{t-1\|t-1})$
$\hat{\mathbf{P}}_{t\|t-1} = \mathbf{F}_t\hat{\mathbf{P}}_{t-1\|t-1}\mathbf{F}_t^{\mathrm{T}} + \mathbf{Q}_t$
Update:
$\hat{\mathbf{x}}_{t\|t} = \hat{\mathbf{x}}_{t\|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t\|t-1}))$
$\hat{\mathbf{P}}_{t\|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\hat{\mathbf{P}}_{t\|t-1}$
Kalman gain:
$\mathbf{K}_t = \hat{\mathbf{P}}_{t\|t-1}\mathbf{H}_t^{\mathrm{T}}(\mathbf{R}_t + \mathbf{H}_t\hat{\mathbf{P}}_{t\|t-1}\mathbf{H}_t^{\mathrm{T}})^{-1}$
For KF:
$\mathbf{f}_t = \mathbf{F}_t$, $\mathbf{h}_t = \mathbf{H}_t$
For EKF:
$\mathbf{F}_t = \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}_{t-1}}\big\|_{\hat{\mathbf{x}}_{t-1\|t-1}}$, $\mathbf{H}_t = \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}_t}\big\|_{\hat{\mathbf{x}}_{t\|t-1}}$

The KF algorithm is described in Table 1. Although it is easy to implement, the KF has strict linearity and Gaussian PDF requirements that make it unsuitable for many dynamic systems.

An obvious way to extend the KF framework is by linearizing functions $\mathbf{f}_t$ and $\mathbf{h}_t$. Another filter, the EKF (Cox, 1964), locally linearizes the state and measurement equations using the first terms in the Taylor series expansions (Table 1). For the EKF to perform well, nonlinearities should be small and the underlying densities should be close to Gaussian. Unlike the KF, the EKF is no longer an optimal sequential estimator due to the linearization approximations. Nevertheless, the EKF has been implemented successfully in a large number of applications in areas such as radar and sonar target tracking, among others.

Unscented Kalman filter

The elegance of the Kalman framework rests on its Gaussian assumption, where the entire information content of $\mathbf{x}_t$ can be captured by propagating the mean and covariance through time. A nonlinear system disrupts the flow of evolving Gaussian PDFs because a Gaussian random variable passing through a nonlinear transformation loses its Gaussian form. Hence, the Kalman framework will still work properly if we use a filter that enforces a Gaussian distribution. This leads to the derivative-free UKF (Wan and van der Merwe, 2001), which uses sigma points coupled with unscented transforms (UTs) (Julier et al., 2000) to propagate means and covariances through nonlinear functions.

The UT algorithm uses a deterministic algorithm to select a set of samples called sigma points around the mean of the $\mathbf{x}_t$. These sigma points and weights are used to obtain the mean and covariance of the random variable $\mathbf{x}_t$. When the random variable undergoes a nonlinear transformation, the sigma points are propagated through this nonlinear function and are used to reconstruct the new mean and covariance using the UT weights (Julier et al., 2000). Unlike the EKF, the nonlinearity is maintained and, hence, the mean and covariance estimates can be computed accurately to at least the second order of the nonlinearity (third, if the initial PDF is Gaussian).

Whereas the EKF enforces linearity through analytic linearization, the UKF enforces a Gaussian distribution while keeping the functional nonlinearity. Because, under this assumption, a Gaussian input to a nonlinear transformation results in a Gaussian output, the system is effectively linearized in a statistical rather than analytical sense. This still enables the filter to carry all necessary information by propagating only the mean and covariance as required by the KF. The UKFs have two disadvantages. First, if nonlinearity is severe, the UKF cannot accurately compute the mean and covariance. Second, the densities may be strongly non-Gaussian, in which case just two moments are inadequate in describing them, even if those can be calculated correctly.

Ensemble Kalman filter

One of the most commonly used filters in geophysics is the EnKF (Evensen, 1994). These have been explicitly designed for high-dimensional problems and are of interest here exactly because of that. Many inversion problems are high dimensional, particularly the ones frequently encountered in geophysical and hydrological fields, ocean and atmospheric modeling, and data analysis (Houtekamer and Mitchell, 2005; Evensen, 2009; Myrseth and Omre, 2010; Seiler et al., 2010).

As discussed in van Leeuwen (2009), the PF (which will be analytically introduced in the next section) is cost-prohibitive in high-dimensional problems and using insufficient number of particles to mitigate cost degrades the PF performance. Thus, it should be reserved for problems nonlinear and non-Gaussian enough in which a KF and its extension fail. The EnKF is a Kalman variant for systems in which the state is composed of a large number of variables. It is possible to argue that the EnKF is a hybrid between a KF and a PF, but it relies on fewer assumptions than the former and is less computationally onerous than the latter. The EnKF still assumes that the updated PDFs are Gaussian, hence, the problem can be based on a Kalman framework. However, the EnKF still can be considered as an approximate sequential MC method because the ensemble members are exactly forwarded and approximately updated.

Working with a large covariance matrix in a high-dimensional system is computationally inefficient for the classical KF. This drawback is addressed by using an ensemble of points, similar to the particle set in a PF, to replace the covariance matrix with the ensemble covariance matrix using MC sampling.

When the problems are high dimensional and non-Gaussian, an EnKF using a Gaussian mixture (Grana and Della Rossa, 2010) model can be used. Gaussian mixture models represent a non-Gaussian PDF as a sum of numerous Gaussian PDFs, enabling the implementation of a KF.

Because the EnKF is a widely used method in geophysics, we focus in this paper on sequential Bayesian techniques that have not been extensively used to date but are powerful in inversion problems in geophysics. For detailed theory and practical implementation of EnKFs, we refer the readers to Evensen (2003, 2009).

PARTICLE FILTERS

An alternative to the restrictive Kalman framework is to use sequential MC techniques also known as PFs. The basic PF recursively computes integral equation 5 using importance sampling (IS) (Ó Ruanaidh and Fitzgerald, 1996). The fundamental working principles of IS are given in Figure 1.

The IS method is used to compute integrals by using random samples drawn from a density referred to as the proposal or sampling density $q(\mathbf{x})$. To numerically estimate an integral of $p(\mathbf{x})$, we select an appropriate $q(\mathbf{x})$ easy to sample from, draw N_p samples $\{\mathbf{x}_t^i\}_{i=1}^{N_p}$, and

	(Nonsequential) Importance sampling IS	(Sequential) Sequential importance sampling SIS
To compute the integral,	$I = \int_{\mathcal{X}} p(\mathbf{x})\mathrm{d}\mathbf{x}$	$p(\mathbf{x}_t\|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t\|\mathbf{x}_t) \int p(\mathbf{x}_t\|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}\|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}$
Sample particles from a sampling density $q(.)$,	$\{\mathbf{x}^i\}_{i=1}^{N_p} \sim q(\mathbf{x})$	$\{\mathbf{x}_t^i\}_{i=1}^{N_p} \sim q(\mathbf{x}_t\|\mathbf{x}_{t-1}^i, \mathbf{y}_t)$
Add the result of each particle according to its importance,	$\widehat{I} \cong \frac{1}{N_p}\sum_{i=1}^{N_p} w^i$	$p(\mathbf{x}_t\|\mathbf{y}_{1:t}) \propto \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i)$
Determined by the weight of that particle.	$w^i = \frac{p(\mathbf{x}^i)}{q(\mathbf{x}^i)}$	$w_t^i = \frac{p(\mathbf{y}_t\|\mathbf{x}_t^i) p(\mathbf{x}_t^i\|\mathbf{x}_{t-1}^i) p(\mathbf{x}_{t-1}^i\|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_t^i\|\mathbf{x}_{t-1}^i, \mathbf{y}_t)}$

Figure 1. Basic philosophy of IS and derivation of sequential IS (SIS) starting from equation 5.

add the results together, after weighting the result of each particle according to its calculated weight as given in Figure 1.

As a numerical method, the IS introduces a certain amount of error in the integral estimation. The variance in the estimate is minimum if $q(\mathbf{x})$ is proportional to $p(\mathbf{x})$ and increases as $q(\mathbf{x})$ deviates from the latter. This means that the sampling density should be selected as close as possible to the integrand. If it is possible to sample from the integrand itself $p(\mathbf{x}) = q(\mathbf{x})$, each particle will have a unit weight (Figure 1) and the error introduced by IS is minimum. On the other extreme, where $q(\mathbf{x})$ does not even intersect with $p(\mathbf{x})$, all the particles sampled from $q(\mathbf{x})$ will have zero weights and the IS will fail. In nonlinear/non-Gaussian problems, it is rarely easy to sample from $p(\mathbf{x})$ and simpler PDFs such as Gaussian are typically used as densities $q(\mathbf{x})$. In summary, a good sampling density $q(\mathbf{x})$ must be easy to sample from and, at the same time, have enough particles with nonzero weights intersecting $p(\mathbf{x})$.

Bayesian filtering requires performing successive IS runs to compute the integral in equation 5 sequentially with t. The output of each IS run forms the prior for the next one. This process is referred to as SIS. The relationship between IS and SIS can be seen in Figure 1. Just as in IS, the integration starts with the selection of a suitable sampling density. It can be shown (Doucet et al., 2000; Ristic et al., 2004) that the selection of a sampling density in the form of $q(\mathbf{x}_t|\mathbf{x}^i_{t-1}, \mathbf{y}_t)$ enables recursive evaluation of the importance weights in time. Note that this sampling density ideally uses all the information we currently have: the value of a particle at the previous step $\mathbf{x}^i_{t-1}$ and current data $\mathbf{y}_t$. For each particle, the weight is given as the ratio of the integrand to the sampling density computed at $\mathbf{x}^i_t$. The integral is then a weighted sum of these particles.

Starting with a set of particles and weights $\{\mathbf{x}^i_{t-1}, w^i_{t-1}\}^{N_p}_{i=1}$ and using equation 7, we can approximate the posterior density at $t-1$ as $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \sum_{i=1}^{N_p} w^i_{t-1}\delta(\mathbf{x}_{t-1} - \mathbf{x}^i_{t-1})$. Hence, the term $p(\mathbf{x}^i_{t-1}|\mathbf{y}_{1:t-1})$ in the weight w^i_t in Figure 1 is equal to w^i_{t-1}. The new weights become

$$w^i_t \propto \frac{p(\mathbf{y}_t|\mathbf{x}^i_t)\,p(\mathbf{x}^i_t|\mathbf{x}^i_{t-1})}{q(\mathbf{x}^i_t|\mathbf{x}^i_{t-1}, \mathbf{y}_t)}\, w^i_{t-1}. \tag{11}$$

To represent a true PDF, $\sum_i w^i_t$ needs to add up to one. Hence, all the weights are normalized once they are computed using equation 11. A formal derivation is provided in Appendix A.

As in IS, a key issue in SIS design is choosing a good proposal density. The optimal choice that allows minimum IS error is $q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)$ (see Doucet et al. [2000] for proof). However, this is not easy to implement in most cases. A simple choice is

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t) = p(\mathbf{x}_t|\mathbf{x}_{t-1}). \tag{12}$$

Note that this suboptimal IS proposal density only depends on the state equation and does not take into account data $\mathbf{y}_t$. Hence, it is referred to as a blind proposal by Pitt and Shephard (1999). This selection reduces equation 11 to

$$w^i_t \propto p(\mathbf{y}_t|\mathbf{x}^i_t) w^i_{t-1}. \tag{13}$$

Although SIS provides a complete framework for performing sequential Bayesian estimation, its implementation quickly runs into the problem of sampling degeneracy. After a few iterations of successive SIS cycles, the process leads to a cloud containing few particles with large weights and numerous particles with negligible weights. In the extreme case, there is only one particle left with large w^i_t, resulting in poor filter performance.

Sequential importance resampling

A second sampling stage is proposed by Gordon et al. (1993) at the end of the integral calculation at each t to mitigate degeneracy. The purpose of this resampling stage (Smith and Gelfand, 1992) is to create more high-weight particles from the original set of particles $\{\mathbf{x}^i_t\}^{N_p}_{i=1}$. The modified filter is called SIR (Gordon et al., 1993; Gilks and Berzuini, 2001) and is the most popular PF implementation. Resampling is either performed at the end of each time step or when the effective number of particles N_p^{eff} needed to prevent the degeneracy problem encountered in SIS drops below a threshold $N_p^{\text{eff}} = 1/\sum_{i=1}^{N_p}(w^i_t)^2$ (Kong et al., 1994).

The resampling process takes the posterior filtering PDF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ represented by the particle set $\{\mathbf{x}^i_t, w^i_t\}^{N_p}_{i=1}$ of SIS at the end of each step and redistributes samples so that all weights of the new particles are the same; that is, $\{\mathbf{x}^j_t, w^j_t = 1/N_p\}^{N_p}_{j=1}$. This results in a larger number of particles in the high likelihood regions, preventing degeneracy.

The process at step t is summarized in Table 2. At the SIS prediction stage, new particles are created from the particles representing the PDF of the previous step, $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$. This first stage starts with the cloud of equal-weight particles from the previous step $\{\mathbf{x}^i_{t-1}\}^{N_p}_{i=1}$ and creates a new set of predictions for the current step $\{\mathbf{x}^i_t\}^{N_p}_{i=1}$ by sampling from the transitional density $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. This is done by propagating each $\mathbf{x}^i_{t-1}$ through the state equation 1 together with a random realization from $\mathbf{v}_t$. The weight of each particle is evaluated through the likelihood and normalized, where $p(\mathbf{y}_t|\mathbf{x}^i_t)$ is the likelihood function defined by the measurement

Table 2. Sequential importance resampling PF (Gordon et al., 1993).

Predict:

Sample new N_p particles at t

$\{\mathbf{x}^i_t\}^{N_p}_{i=1} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1})$ given $\{\mathbf{x}^i_{t-1}\}^{N_p}_{i=1}$

using $\mathbf{x}^i_t = \mathbf{f}_t(\mathbf{x}^i_{t-1}, \mathbf{v}^i_t)$ $i = 1, \ldots, N_p$

where $\mathbf{v}^i_t$ are samples from the state noise PDF.

Update:

- Compute the likelihood $p(\mathbf{y}_t|\mathbf{x}^i_t)$ for each $\mathbf{x}^i_t$.
- Normalize the weights:

$w^i_t = \frac{p(\mathbf{y}_t|\mathbf{x}^i_t)}{\sum_{i=1}^{N_p} p(\mathbf{y}_t|\mathbf{x}^i_t)}$

- The posterior PDF is approximated by

$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N_p} w^i_t \delta(\mathbf{x}_t - \mathbf{x}^i_t)$

Resample:

Resample N_p particles $\mathbf{x}^j_t$ with equal weights

$\{\mathbf{x}^i_t, w^i_t\}^{N_p}_{i=1} \mapsto \{\mathbf{x}^j_t, \frac{1}{N_p}\}^{N_p}_{j=1}$ s.t.

$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \frac{1}{N_p}\sum_{j=1}^{N_p} \delta(\mathbf{x}_t - \mathbf{x}^j_t)$

equation, equation 2. The newly computed particles and weights are used for expressing the posterior PDF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. Resampling follows this update stage, in which a new set of particles is formed from the previous one. The larger the weight of a particle, the more new particles it generates during resampling (Doucet et al., 2001a). These new particles $\{\mathbf{x}_t^j, w_k^j = 1/N_p\}_{j=1}^{N_p}$ will now have equal weights. The filter is initiated by a prior probability density $p(\mathbf{x}_0)$. The prior can be an initial guess or it can be obtained from an independent inversion if data $\mathbf{y}_0$ are available.

A potential drawback of resampling is the generation of multiple copies of high-likelihood particles. In the extreme case, in which the state noise is very small, the particle diversity is lost with all particles being identical. This is called sample impoverishment (Ristic et al., 2004). Although SIR provides robust tracking performance in many nonlinear/non-Gaussian tracking problems, it has certain disadvantages such as sample impoverishment, blind proposal density in the prediction step that does not depend on data, and sensitivity to outliers. To address these problems, appropriate PF variants have been developed (Gustafsson et al., 2002; Ristic et al., 2004; Cappé et al., 2007). This paper uses the basic SIR algorithm to compare the PF with KFs.

The multiple model particle filter

In dynamic problems, the model may be unknown or uncertain for each step. Therefore, the model connecting the unknown parameters and the data may need to be corrected or updated as new data become available. Examples include a variation in the measurement equation at different regimes (e.g., near-field versus far-field calculations, planar versus curved wavefront beamforming), variation in nonstationary noise statistics, or a varying number of parameters to be estimated with state. This problem is typically referred to as the varying model and is often the case in seismic applications, such as tracking an unknown number of reflectors using interferometry. Successful tracking under such circumstances requires sequential filtering algorithms that are capable of jumping or switching between models (Musso et al., 2001; Vermaak et al., 2005). These filters are called multiple model PFs (MMPFs) (Ristic et al., 2004).

The classical PF assumes a fixed and known model and tracks a constant number of model parameters. On the contrary, the MMPF includes an extra parameter in the state vector that denotes which model each particle uses. The filter tracks not only the model parameters but also the most suitable model itself. A critical element in MMPF is the selection of a model transition matrix Π_m. The entries of Π_m are probabilities, determining how particles belonging to a model at time $t-1$ can transition to a different model at time t. Therefore, an MMPF SIR has a transitioning stage preceding the prediction.

The classical MMPF is based on SIR (Ristic et al., 2004). More advanced MMPFs use a technique called reversible jump Markov chain MC (Doucet et al., 2001b; Khan et al., 2005). A typical MMPF step involves the following:

- Model switching: Start with $\{m_{t-1}^i, \mathbf{x}_{t-1}^i\}_{i=1}^{N_p}$, where particle i follows model m_{t-1}^i. Predict the models of new particles m_t^i at t using the model transition matrix Π_m. This implies that some state variables $\mathbf{x}_k$ can exit or more state variables can be "born," with the model number decreasing or increasing, respectively. The dimension of the state vector is, thus, appropriately modified.
- Model-conditioned SIR: Using the newly predicted model m_t^i for each particle, implement a model-conditioned SIR exactly as described for the single model PFs. It should be noted here that, in problems in which higher order models represent increasing model complexity such as an increased number of interfaces, the model with the highest order is favored. This is because it is natural that a more complicated model with a larger number of parameters provides a better fit to the data. However, a better fit may imply that we are overfitting the data rather than introducing a better underlying model. The MMPF tries to achieve a good balance between overfitting and oversimplification. This is done by penalizing the higher order models. Within the Bayesian framework, the prior distributions on the unknown parameters serve as "penalizing" terms, similar to the penalization in the Akaike and Schwarz-Rissanen criteria (Schwarz, 1978), which appropriately weigh each model number.

EXAMPLES

Two geophysical examples are illustrated below for comparison of KF variants and PFs. In the first example, NVT is used to compare the EKF, UKF, and PF under different tremor signal-to-noise ratios (S/Ns). The NVT is usually of low strength, and that means the filters have to operate under rapidly fluctuating, low S/Ns. This significantly amplifies the penalty for using the linear/Gaussian approximation used in a Kalman framework. The NVT is tracked first as an angle of arrival, and then the NVT source location on the North American Plate boundary is tracked through an alternate formulation. The source location tracking process is used to compare sequential and nonsequential Bayesian inversion methods used to show the advantages of sequential Bayesian techniques.

The second example is the spatial tracking of seabed sedimentary reflectors in shallow water using active and passive measurements. The active method uses synthetic acoustic data from a ship-towed array and an active source. This example explores the effects of state noise on KFs and PFs. The passive method uses a vertical array along the water column drifting with the current. Here time step t is effectively a spatial step depending on the speed of the water current. As the array drifts, it records ambient noise, which provides information about the seabed properties under it. This allows us to construct a range-dependent seabed profile with multiple reflectors and use sequential methods to track the spatial variation. This example demonstrates the advantages of using a MMPF.

Example I: Seismic tremor tracking

Seismic NVT is continuous noise appearing at regular intervals (about 14 months in Cascadia, WA) (Obara, 2002; Rogers and Dragert, 2003; Peng and Gomberg, 2010). In Cascadia, the Juan de Fuca Plate is underthrusting below the North American Plate. Near the surface, the two plates are locked and as the Juan de Fuca Plate moves east, the North American Plate is also dragged in the same direction. At regular intervals, a seismic tremor is observed and, simultaneously, it can be observed with a global positioning system that the North American Plate is moving west. It is believed that the tremor originates from the plate boundary some 30 km below surface, where the plates are not strongly locked.

We demonstrate particle filtering on 2D beamforming with a temporary array given in Figure 2a and 2b installed in the Cascadia

subduction zone, WA (Figure 2c), to detect seismic tremors (Ghosh et al., 2009, 2012; Zhang et al., 2011). The results are shown in the slowness domain (slowness is here the horizontal phase slowness or inverse phase speed) and in source location on the plate boundary. At the Cascadia array, tremors typically travel with horizontal phase velocities of 10 km/s (slowness 0.1 s/km). The array is quite dense, with 72 sensors placed within a 1.2 km^2 area.

Tremor state-space model

The NVT state and measurement equations are given by

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t, \tag{14}$$

$$\mathbf{y}_t(f_j) = a_t(f_j)\mathbf{d}(\mathbf{x}_t, f_j) + \mathbf{w}_t(f_j), \tag{15}$$

where $\mathbf{w}_t(f_j)$ is a vector of additive nonstationary complex-valued Gaussian noise processes along the seismic array, each with a variance of $\nu_t(f_j)$ for each frequency f_j and step t, $a_t(f_j)$ is the complex-valued source strength, and $\mathbf{d}(\mathbf{x}_t, f_j)$ is the function relating the state parameters to the measurement.

There are two possible state vector selections depending on the required output. In the first one, we are interested in tracking the direction of the tremor signal that arrives at the array. Hence, $\mathbf{d}(\mathbf{x}_t, f_j)$ is a beamformer composed of plane wave phase delays $e^{i\omega\Delta\mathbf{r}\mathbf{x}_t}$, where $\mathbf{x}_t = [s_x s_y]_t^T$ is the horizontal slowness and $\Delta\mathbf{r}$ describes the coordinates of the array sensors relative to the mean coordinates. This formulation will be referred to as tremor arrival angle tracking.

For tracking the NVT location on the slab, the state vector is $\mathbf{x}_t = [\text{lon lat}]_t^T$, the longitude and latitude of the moving source. In this case, $\mathbf{d}(\mathbf{x}_t, f_j)$ incorporates a ray tracer code that propagates the seismic signal from any particle (prospective source location on the slab) to the seismic sensors, effectively using back propagation. This process will be referred to as tremor source location tracking.

Beamforming can be done either in time domain by time-delaying and stacking the time series recorded by the seismic sensors or in the frequency domain, where the time delays are replaced with appropriate phase shifts. One advantage of frequency-domain beamforming is the ability to use only the desired frequencies. For example, this enables us to take out the frequencies with significant anthropogenic noise. It also reduces the computation time because only a small subset of frequencies representative of the frequency band is used in the inversion. An initial analysis shows that there is little difference between using 8 or 80 frequencies for a

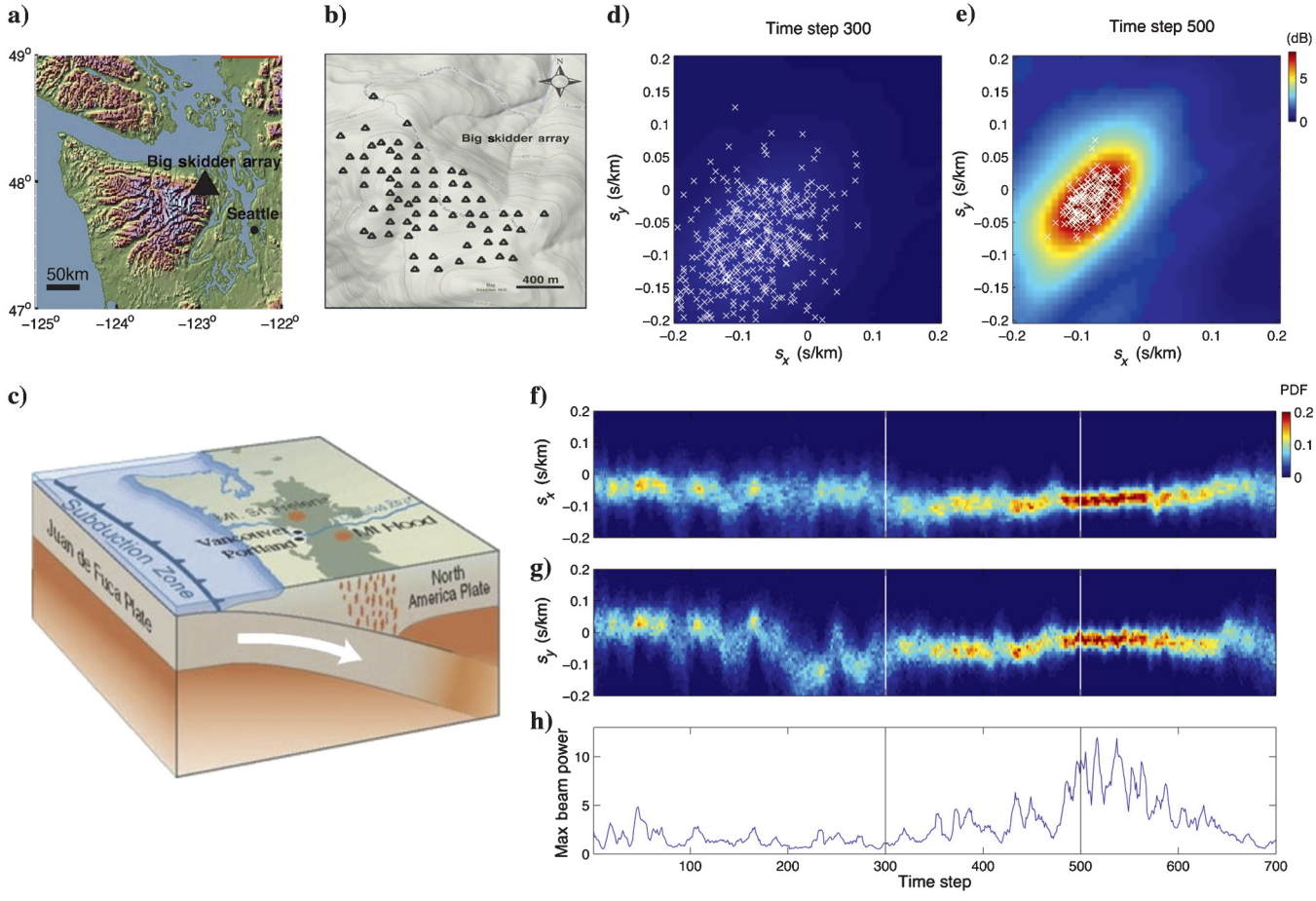

Figure 2. NVT: (a) location and (b) details of the array deployed in Cascadia along with (c) the structure of the Cascadia subduction zone. The PF tracking results of peak slowness (s_x, s_y) for 1 h. Two-dimensional objective function, equation 19, with PF particles (×) superimposed at time step (d) $t = 300$ and (e) $t = 500$, indicated by vertical lines in panels (f)-(h). Time-evolving posterior PDF of (f) s_x and (g) s_y obtained from PF. Each vertical slice corresponds to a normalized histogram of particles at that time, effectively giving the posterior PDF. The PDF is sharper for $t = 500$. (h) Maximum beamformer power.

4–17 Hz beamformer. Here a frequency-domain beamformer is selected with eight frequencies.

The classical Bartlett power objective function is obtained from the assumption of additive complex-valued Gaussian noise $\mathbf{w}_t(f_j)$ independent for each frequency f_j and step t, which is written as $\mathbf{w}_t(f_j) = \mathbf{y}_t(f_j) - a_t(f_j)\mathbf{d}(\mathbf{x}_t, f_j)$ following equation 15, with a corresponding Bartlett processor based, multifrequency likelihood function composed of n_f multiplicative complex Gaussian PDFs (Gerstoft and Mecklenbräuker, 1998):

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_\mathrm{f}} \frac{1}{(\pi\nu_j)^{n_\mathrm{h}}} \exp\left[-\frac{\|\mathbf{y}_t(f_j) - a_t(f_j)\mathbf{d}(\mathbf{x}_t, f_j)\|^2}{\nu_j}\right], \tag{16}$$

where n_h and n_f are the numbers of seismic sensors and the frequencies used in tracking and ν_j is the noise variance at frequency f_j. Note that the noise variance and the source strength evolve with time and frequency. We are not interested in the values of these parameters (nuisance parameters), but to track the source, these parameters need to be known. A straightforward way is to incorporate these into the state equation and track them together with the parameters of interest. However, this significantly expands the state-space dimension and can degrade the track performance. In a Bayesian seismic inversion, an alternative is to first estimate the unknown source term $a_t(f_j)$ for all frequencies at each time step using a maximum likelihood (ML) estimator and get rid of the nuisance parameter by replacing it with its ML estimate in the likelihood formulation.

The unknown source is estimated analytically via an ML calculation by solving $\partial\mathcal{L}/\partial a_t = 0$,

$$\hat{a}_t(f_j) = \frac{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H}\mathbf{y}_t(f_j)}{\|\mathbf{d}(\mathbf{x}_t, f_j)\|^2}. \tag{17}$$

Inserting the source estimate back into equation 16, the likelihood becomes (Huang et al., 2006)

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_\mathrm{f}} \frac{1}{(\pi\nu_t(f_j))^{n_\mathrm{h}}} \exp\left[-\frac{\phi_j(\mathbf{x}_t)}{\nu_t(f_j)}\right], \tag{18}$$

$$\phi_j(\mathbf{x}_t) = \mathbf{y}_t^\mathrm{H}(f_j)\mathbf{y}_t(f_j) - \frac{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H}\mathbf{y}_t(f_j)\mathbf{y}_t(f_j)^\mathrm{H}\mathbf{d}(\mathbf{x}_t, f_j)}{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H}\mathbf{d}(\mathbf{x}_t, f_j)}, \tag{19}$$

where ϕ_j is the Bartlett objective function. The first term in ϕ_j is the total seismic signal power and is a constant (not a function of $\mathbf{x}_t$). The second term in ϕ_j is a normalized beamformer output because a classical frequency-domain beamformer is given by $\mathbf{y}_t^\mathrm{H}\mathbf{d}(\mathbf{x}_t)$. Hence the likelihood formulation in equations 18 and 19 enables the conversion of the beamformer output into a PDF capable of providing the uncertainty in the estimates.

As for the noise variance term $\nu_t(f_j)$, one way to estimate it is from the data (Zhang et al., 2011). It can then be inserted into equation 18. An alternative way is to treat the unknown $\nu_t(f_j)$ as a nuisance parameter, as mentioned before. We can then use an ML estimator, similar to the one for source estimation. Solving $\partial\mathcal{L}/\partial\nu_t(f_j) = 0$ results in

$$\hat{\nu}_t(f_j) = \frac{\phi_j(\mathbf{x}_t)}{n_\mathrm{h}}, \tag{20}$$

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_\mathrm{f}} \left(\frac{n_\mathrm{h}}{\mathrm{e}\pi\phi_j(\mathbf{x}_t)}\right)^{n_\mathrm{h}}. \tag{21}$$

Observed tremor

A simple tracking result is shown in Figure 2, where we are tracking peak slowness s_x and s_y across the array for 1 h starting at 0:00 Coordinated Universal Time (UTC) on 7 May 2008. Each snapshot is 5 s, providing 720 observations in an hour. From the data, we extract the 2D slowness vector (s_x, s_y) based on total beam power at eight frequencies from 3.9 to 17.6 Hz. The PF is initialized ($t = 0$) with 400 uniformly distributed particles across (s_x, s_y). For each time step t, we first predict the values based on time step $t - 1$ using the measurement equation, equation 14. Then we update the values based on the observed data, that is, based on the likelihood function given in equation 21.

The beamformer outputs for s_x and s_y are given in Figure 2a and 2b along with the set of particles $\{\mathbf{x}_t^i\}_{i=1}^{400}$ obtained by the PF at $t =$ 300 and 500. From these particles, the PDFs for s_x and s_y are estimated at each time step as indicated in Figure 2c and 2d. Note that around $t = 500$, the beamformer output in Figure 2e becomes stronger and, hence, the estimates have low uncertainty with a small particle spread (Figure 2b).

Filter comparison is performed by running the EKF, UKF, and the SIR PF with the same initial settings, state-space equations with identical noise characteristics, and the same starting prior PDFs. Because the EKF and UKF can only operate in a Gaussian environment, initial PDF $p(\mathbf{x}_0)$ is selected as a Gaussian density for the PF as well. The results are given in Figures 3 and 4. The mean for the EKF and UKF is plotted in Figure 3 on top of the evolving PDF of slowness parameters from the PF. The EKF and UKF can generally track the changing parameter. However, starting at $t = 180$, the s_y rapidly drops, followed by a sudden increase at $t = 250$. Notice how both KFs were too slow to adapt to the rapid changes in parameters (an issue called slow convergence, which will be explored in the next example). The UKF performs better than the EKF due to its ability to handle the nonlinearities more effectively. For example, the UKF is capable of tracking s_x for $t = 620 - 720$, resulting in a pattern similar to that of the PF. On the other hand, the EKF is unable to adapt to the changing values of the slowness parameter.

Slowness-azimuth plots for the filters are also provided at three different time steps in Figure 4. The EKF and UKF have Gaussian outputs and are represented by covariance ellipsoids, in which the major axes correspond to multiples of standard deviation for s_x and s_y. This is compared to the spread of the resampled (equal weight) particles of the PF at the corresponding time steps. The density of the particles does not need to be Gaussian. These results demonstrate:

- The EKF ellipsoids fail to capture the correlation between s_x and s_y, whereas the UKF ellipsoids generally correspond well to the particle plot.
- The EKF and UKF mean values are relatively biased (particularly s_y in Figure 3b) with respect to the particles for $t =$ 150–380 due to the slow convergence (Figure 4a).

- As shown in Figure 4a, the spread of the particles and the standard deviation of the KFs are largest at $t = 300$. The distributions get tight with reduced uncertainty and similar mean values in all three filters when the tremor signal is strong (Figure 4b).
- For $t = 620$–680, the EKF is unable to track s_x (Figure 3a), whereas the UKF follows the PF results closely, as can also be seen in Figure 4c.
- There are times (e.g., $t = 20$–90 in Figure 3a for the UKF and $t = 30$–100 in Figure 3b for EKF) when the EKF or UKF diverges from the other two filters.

In an alternative formulation, the longitude and latitude of the tremor source on the slab (~30 km deep) are tracked as state parameters instead of slowness. An important aspect of the 2D beamformer is that it provides an angle of arrival at the array location and an undetermined range. Determining the source range requires an additional piece of information. This is provided by the assumption that the tremor originates on the plate interface. This enables mapping of the angle of arrival by back projection onto the slab, and this produces an estimate for range. This results in good angle of arrival tracking but poor localization because the radial uncertainty is large, creating a long-tailed nonsymmetric PDF in a Bayesian beamformer inversion.

The PF also has its largest uncertainty radially because the beamformer and the PF use the same beamformer-based likelihood function (equations 18 and 21). However, the state equation, equation 14, used by the PF mitigates the degradation in range at a low S/N. The state equation forces the PF to ignore the rapid range fluctuations due to the poor likelihood function of the low S/N data. The PF will only "believe" the new range inferred from the current data when the S/N increases or when the likelihood function consistently keeps pointing to the new range at consecutive steps, indicating that the range change is likely not a random fluctuation.

Sequential (PF) and beamformer methods are compared using the 2D PDF for the back-projected tremor location on the slab in Figure 5 for low, medium, and high S/N values. Both methods use the exact same likelihood formulation. Data are from the same day as before, starting at 11:40 UTC. Notice that the PF location PDF is sharper than beamformer PDF at all S/N levels. Also note that the radial uncertainty is significantly less in PF.

Example II: Reflector tracking

The first part of this example demonstrates how the KF and PF differ in a nonlinear geophysical problem. It also examines the effects of state noise $\mathbf{v}_t$ on the estimated track quality. Assume that we use acoustic data from a towed array to invert for the sediment parameters below the seafloor. The state vector is composed of the ocean, sediment, and bottom sound speed profiles, sediment thickness, attenuation, and density. Through an acoustic propagation code, it is possible to compute the acoustic field across the array for a given environment. Along the array trajectory, the sediment is stable with a fixed thickness and slowly decreasing sound speed. However, at $t = 140$ min, a geologic formation with a sudden increase in thickness and drop in sound speed appears. The environment is tracked using a PF and an EKF.

As shown in Figure 6a, the environment changes slowly at the beginning of the track and both filters are able to track the sediment parameters correctly. However, at the sudden jump, the EKF fails and diverges, whereas the particles in the PF successfully follow the trajectory. The state noise $\mathbf{v}_t$ alters significantly the filter behavior. When $\mathbf{v}_t$ is reduced in the state equation, equation 1, the filter trusts the state equation more than the measurement. When $\mathbf{v}_t$ is increased, the filter assumes that the state equation is less reliable. For an infinite $\mathbf{v}_t$, the filter ignores entirely the state equation and becomes a set of successive independent inversions in time.

Because a filter with high noise $\mathbf{v}_t$ trusts less the previous sediment values, it can easily adapt to the sudden jump at $t = 140$ min in Figure 6a. On the contrary, the track is less noisy with smaller $\mathbf{v}_t$ but there is a risk of track divergence if the state equation fails. This

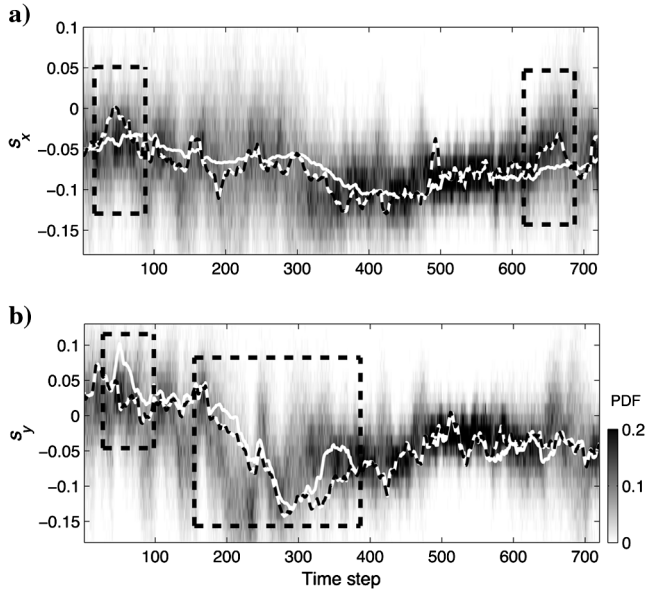

Figure 3. One-dimensional PDF of slowness for 1 hour corresponding to 720 filter time steps in the form of normalized (maximum at 1) histograms of particles. The background image is the PDF of the PF for (a) s_x and (b) s_y. Solid and dashed lines represent the EKF and UKF estimates, respectively. Rectangles show the features discussed in the text.

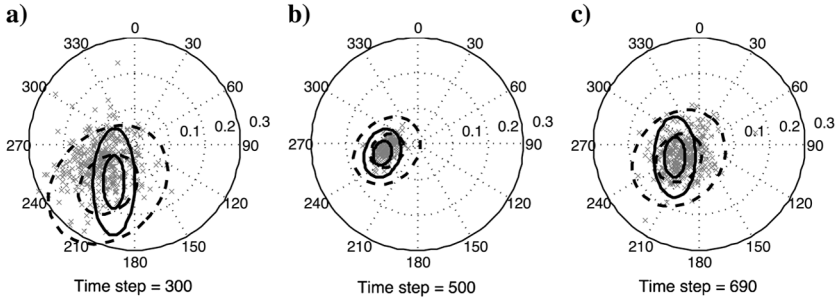

Figure 4. Slowness-azimuth filter outputs at (a) $t = 300$ (b) $t = 500$, and (c) $t = 790$. The symbols (×) represent the particles in the PF. Ellipses represent σ and 2σ covariance ellipsoids (corresponding to 39% and 86% of the total probability mass, respectively) for the EKF (red) and the UKF (black).

can be seen at $t = 140$ min in Figure 6b. It is possible to entirely lose the track (sediment thickness) or have slow convergence (sediment sound speed). The selection of good state noise is referred to as filter tuning.

The second part of this example demonstrates the feasibility of seismic reflector (horizon) tracking using MMPFs. The experimental setup is changed from an active acoustic source setup to seabed reflection tracking with noise interferometry (Curtis et al., 2006; Wapenaar et al., 2008; Schuster, 2009) in the ocean (Gerstoft et al., 2008; Siderius et al., 2010). The sea surface generated noise travels downward, passing through the drifting array, and some of that noise is reflected from the sea bottom and deeper reflectors. This upward traveling reflected noise passes through the array again. Hence, by steering the array looking direction upward first and then downward by adjusting the beamformer weights (shading factor), it is possible to capture the incident and reflected noise from which the reflector depths and strengths can be inferred.

Reflection sequences are produced via noise crosscorrelation interferometry and stacking the results (Siderius et al., 2010). In the example used here, ambient noise 50–4000 Hz bandpass filtered on a 32-element vertical drifting array is used to construct the seabed reflection sequence (Gerstoft et al., 2008; Siderius et al., 2010). Minimum variance distortionless response (MVDR) adaptive beamforming is used to steer the array. The fathometer output provides strong correlation of upward and downward traveling noise at certain time delays, corresponding to reflections from sediment interfaces. An obtained seismic record section (Siderius et al., 2010) is shown in Figure 7a. For the prior information, we assume equal probability that there are six, seven, or eight reflectors in the seabed structure. We select a transition matrix Π_m:

$$\Pi_m = \begin{pmatrix} 0.7 & 0.3 & 0.0 \\ 0.15 & 0.7 & 0.15 \\ 0.0 & 0.3 & 0.7 \end{pmatrix}. \tag{22}$$

This 3×3 matrix allows the number of reflectors at the next time step to remain the same or change by one for each particle. Each particle retains its model number with probability of 0.7.

The state and measurement equations for fathometer tracking (Michalopoulou et al., 2012) are given as

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t, \tag{23}$$

$$\mathbf{y}_t = \mathbf{h}(\mathbf{x}_t) + \mathbf{w}_t, \tag{24}$$

where $\mathbf{v}_t$ and $\mathbf{w}_t$ are state and measurement noise and $\mathbf{h}(\cdot)$ is the fathometer (either conventional or MVDR) processor. For the PF, the state vector is composed of the reflector depths z and the reflection amplitudes a at each reflector for a fixed number of reflectors giving $\mathbf{x}_t = [\mathbf{z}_t^{\mathrm{T}} \mathbf{a}_t^{\mathrm{T}}]^{\mathrm{T}}$. For the MMPF, an extra model parameter m_t is also included with $\mathbf{x}_t = [m_t \mathbf{z}_t^{\mathrm{T}} \mathbf{a}_t^{\mathrm{T}}]^{\mathrm{T}}$, where m_t is

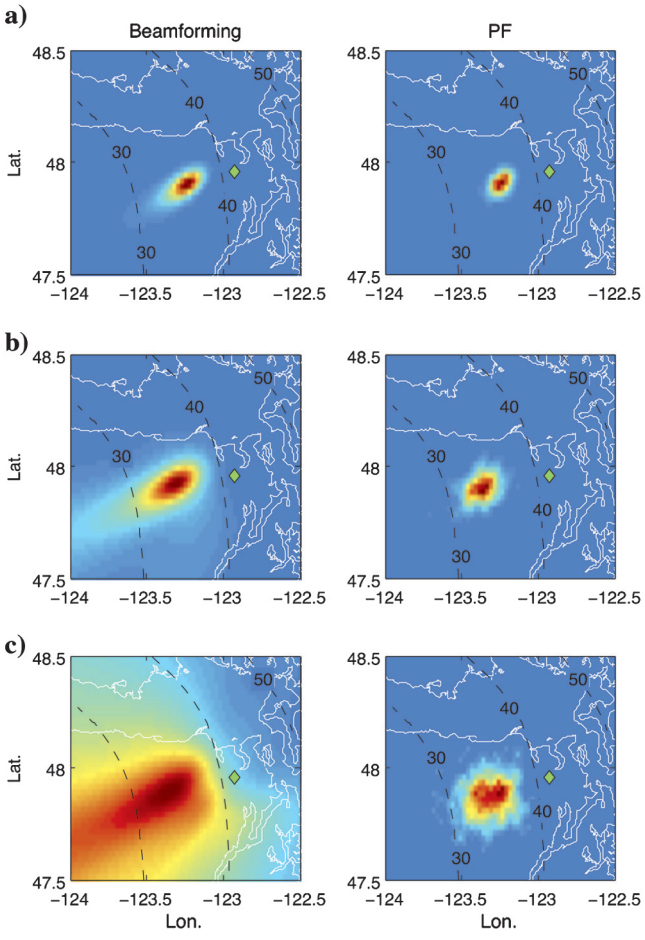

Figure 5. Tremor PDF from beamforming and PF on the map given in the form of normalized (maximum at 1) histogram for the PF. Two-dimensional PDFs for the source location on the slab for (a) strong (12:03 UTC), (b) medium (11:51 UTC), and (c) weak (12:12 UTC) NVT source strength. The symbol (◇) shows the array location. Dashed lines give the slab depth at that location.

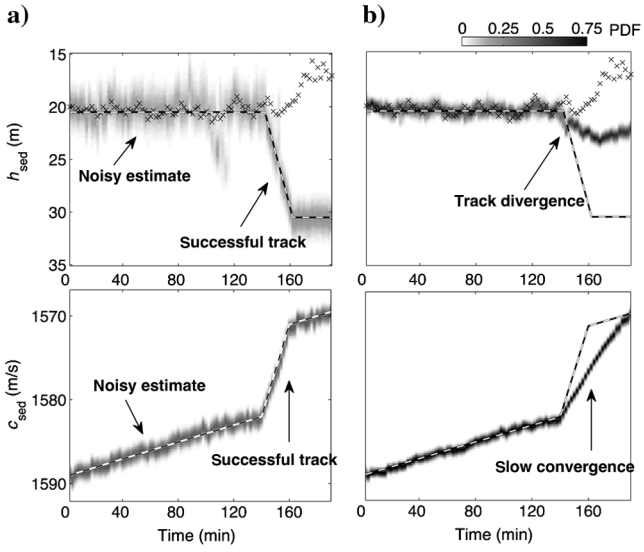

Figure 6. Effects of state noise $\mathbf{v}_t$ on a filter tracking sediment thickness and sound speed using a moving acoustic array. The environment is slowly varying until $t = 140$ min, at which point the sediment rapidly thickens with a reduced sound speed. Distribution of the particles of a PF for (a) large $\mathbf{v}_t$ and (b) small $\mathbf{v}_t$ together with the mean EKF trajectory (dotted) and true environmental parameters (dashed). Each vertical slice is a normalized histogram. To be able to compare all four cases, all subplots are scaled with the same number.

the model number that corresponds to models with an increasing number of reflectors. The lengths of vectors $\mathbf{z}_t$ and $\mathbf{a}_t$ are determined by m_t. The MMPF implementation solves for $\mathbf{a}_t$ at each step using ML estimation (Andrieu and Doucet, 1999; Larocque et al., 2002), similar to the ML estimator given in equation 17 used in the seismic tremor tracking example.

Using an MMPF, each reflector sequence is modeled as a sum of weighted and shifted reflections and tracked the location of reflection peaks along with the number of reflectors (Jain and Michalopoulou, 2011). Thus, the ith particle in the PF uses a reflector model corresponding to m_t^i, which has $2 \times m_t^i$ unknown environmental parameters (depth and strength of each reflection) in addition to m_t^i. Posterior PDFs for reflector depths and model number were calculated via the MMPF at each record t. The layer depths superimposed on the record of Figure 7b are mode values calculated from the reflection location posterior PDFs in depth, identifying distinct reflectors. The PF closely follows seabed tracks; it allows existing ones to exit when the reflectors become weak and new ones to emerge when new, consistent structure appears. Amplitudes a_t are given in Figure 7c with significant reflections at 130–135 and 150–155 m.

Figure 8 illustrates the evolution of the probability mass function (PMF) of the model number (number of reflectors) for the MVDR results, demonstrating that the number of reflectors is consistently estimated as seven or eight most of the time. The figure focuses on records between 1 and 30. Note how the number of reflectors changes between records 10 and 20. A large probability is assigned for eight reflectors at record 10. As weak reflector tracks are lost, the PMF of $m_t = 7$ is increased, and when the reflector appears again, it decreases.

Using a standard PF with a fixed number of reflectors that is different than the true number of reflectors results in one of two possible scenarios: If the model used has fewer interfaces than the true number, some reflectors are not tracked. If more interfaces are considered, the PF is forced to track reflectors that do not exist and will include incorrect tracks. Even with the eight-reflector model, there

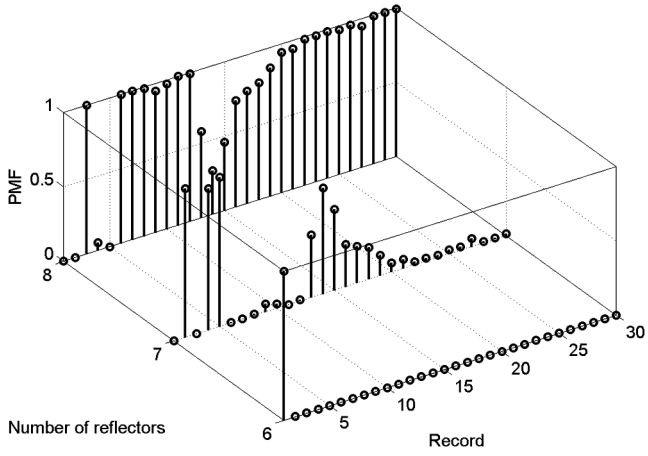

Figure 8. Multiple seismic reflector tracking results: the PMF of the number of reflectors for the MVDR results for records 1 through 30.

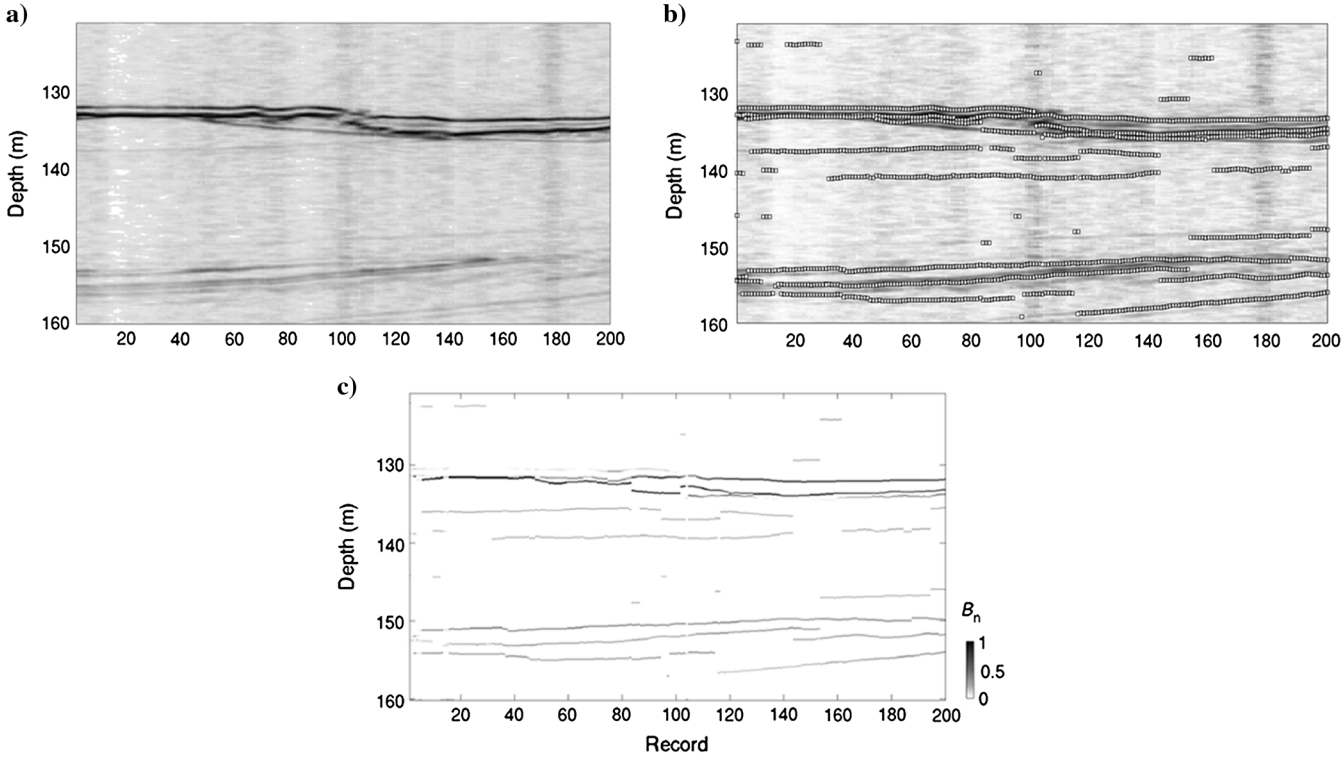

Figure 7. Multiple seismic reflector tracking results: (a) output of ocean interferometry (Siderius et al., 2010) and (b) MMPF reflector tracks (squares) for an unknown number of reflectors as a function of record number, plotted on top of the interferometry results and (c) amplitude tracks a_t. All interferometry outputs ($\mathbf{B}_n$) are normalized.

are times when a short reflector (horizon) appears (which may or may not be due to noise) and the filter jumps from the low amplitude long reflectors at larger depths to the short reflector, creating artificial gaps in the long reflector. However, one must be careful with the range of possible model numbers. As the model number increases, so does the state dimension, and the track quality may suffer.

CONCLUSIONS

Sequential Bayesian filters were introduced with an emphasis on applications to geophysical problems. The theoretical background for sequential Bayesian methods was summarized. Filters were grouped into two main categories: those that use the Kalman philosophy and those termed sequential MC techniques, also known as PFs.

The first category assumed the underlying parameter statistics were Gaussian and the nonlinearity was not strong enough to significantly disrupt the Gaussian statistics. Four types of KF were discussed: the classical KF for strictly linear/Gaussian problems, the EKF that can handle mild nonlinearities by locally linearizing the equations, the UKF that keeps the nonlinearity while still assuming a Gaussian density, and the EnKF designed for problems with large state vectors using an ensemble of particles to compute means and covariances using MC analysis. The second category consisted of PFs, which can handle nonlinear/non-Gaussian cases. They use an evolving set of particles, with each particle representing a possible estimate for the set of parameters that are being tracked. The PF equations were derived starting from classic IS, moving to the SIS, and, finally, obtaining the most commonly used PF, the SIR filter. Model order selection in a sequential Bayesian framework was examined with the MMPF. Here the model was unknown or uncertain and possibly changed for each step. Therefore, the model connecting the unknown parameters and the data need to be updated as new data became available. This was shown to be relevant in many geophysical tracking problems such as tracking a varying number of sedimentary reflectors.

The NVT tracking example with a nonlinear measurement equation showed that a PF is needed to successfully track the tremor. Whereas the EKF had the worst tracking performance and the UKF failed in the presence of low S/N NVT, the PF was able to maintain a successful track through the entire time series. The tracking capabilities of the EKF and PF were compared in a reflector tracking example. This showed that the PF was robust to sudden jumps in state parameter values in nonlinear systems whereas KFs diverged. The MMPF result showed with real data that it was possible to track the sediment reflectors even when the number of reflectors was unknown and changing. Sequential Bayesian filtering is a good addition to the already existing inversion methods. As the examples showed, they are suitable for a variety of geophysical problems and there is a diverse set of KFs and PFs that can be tailored according to each problem.

ACKNOWLEDGMENTS

This work has been supported by National Science Foundation grants EAR-0944109 and OCE-1030022 and Office of Naval Research grants N00014-10-1-0073 and N00014-11-1-0320.

APPENDIX A

SEQUENTIAL IMPORTANCE SAMPLING DERIVATION

As a sequential Bayesian method, a formal PF derivation starts with the full posterior density $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$. To obtain the filtering PDF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, we need to integrate the full posterior:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2\ldots\mathrm{d}\mathbf{x}_{t-1}. \quad \text{(A-1)}$$

This can be rewritten in a form similar to equation 8 and solved using an IS as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \int \delta(\mathbf{x}_t - \mathbf{x}'_t)p(\mathbf{x}'_{1:t}|\mathbf{y}_{1:t})\mathrm{d}\mathbf{x}'_{1:t}, \quad \text{(A-2)}$$

$$\approx \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i), \quad \text{(A-3)}$$

$$w_t^i \propto \frac{p(\mathbf{x}_{1:t}^i|\mathbf{y}_{1:t})}{q(\mathbf{x}_{1:t}^i|\mathbf{y}_{1:t})}, \quad \text{(A-4)}$$

for some sampling density $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$. By itself, this formulation is not in a sequential form. To sequentially implement IS, one needs to express $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ and w_t as functions of their values at the previous step $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$ and w_t, respectively. This allows us to use the cloud of particles and their associated weights at the previous time step $\{\mathbf{x}_{t-1}^i, w_{t-1}^i\}_{i=1}^{N_p}$ to compute $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^{N_p}$. This is done first by selecting a sampling distribution $q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ of the form

$$q(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t})q(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}). \quad \text{(A-5)}$$

Then the full posterior PDF is rewritten using the Bayes rule as (Ristic et al., 2004)

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}). \quad \text{(A-6)}$$

Inserting equations A-5 and A-6 into equation A-4 and dropping the constant term $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$,

$$w_t^i \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^i)p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i, \mathbf{y}_{1:t})} \times \frac{p(\mathbf{x}_{1:t-1}^i|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_{1:t-1}^i|\mathbf{y}_{1:t-1})}, \quad \text{(A-7)}$$

which simplifies to the recursive filter form in equation 11.

REFERENCES

Andrieu, C., and A. Doucet, 1999, Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC: IEEE Transactions on Signal Processing, **47**, 2667–2676, doi: 10.1109/78.790649.

Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp, 2002, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking: IEEE Transactions on Signal Processing, **50**, 174–188, doi: 10.1109/78.978374.

Baziw, E., 2005, Real-time seismic signal enhancement utilizing a hybrid Rao-Blackwellized particle filter and hidden Markov model filter: IEEE Geoscience and Remote Sensing Letters, **2**, 418–422, doi: 10.1109/LGRS.2005.852711.

Candy, J. V., 2007, Bootstrap particle filtering: IEEE Signal Processing Magazine, **24**, no. 4, 73–85, doi: 10.1109/MSP.2007.4286566.

Candy, J. V., 2009, Bayesian signal processing: Classical, modern and particle filtering methods: John Wiley & Sons.

Cappé, O., S. Godsill, and E. Moulines, 2007, An overview of existing methods and recent advances in sequential Monte Carlo: Proceedings of the IEEE, **95**, 899–924, doi: 10.1109/JPROC.2007.893250.

Cox, H., 1964, On the estimation of state variables and parameters for noisy dynamic systems: IEEE Transactions on Automatic Control, **9**, 5–12, doi: 10.1109/TAC.1964.1105635.

Curtis, A., P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, 2006, Seismic interferometry — Turning noise into signal: The Leading Edge, **25**, 1082–1092, doi: 10.1190/1.2349814.

Djurić, P. M., and M. F. Bugallo, 2010, Particle filtering, *in* T. Adali, and S. Haykin, eds., Advances in adaptive filtering: Wiley, 271–332.

Doucet, A., N. de Freitas, and N. Gordon, 2001a, Sequential Monte Carlo methods in practice: Springer.

Doucet, A., S. Godsill, and C. Andrieu, 2000, On sequential Monte Carlo sampling methods for Bayesian filtering: Statistics and Computing, **10**, 197–208, doi: 10.1023/A:1008935410038.

Doucet, A., N. J. Gordon, and V. Krishnamurthy, 2001b, Particle filters for state estimation of jump Markov linear systems: IEEE Transactions on Signal Processing, **49**, 613–624, doi: 10.1109/78.905890.

Dovera, L. Della Rossa, E., 2011, Multimodal ensemble Kalman filtering using Gaussian mixture models: Computational Geosciences, **15**, 307–323, doi: 10.1007/s10596-010-9205-3.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics: Journal of Geophysical Research, **99**, 10143–10162, doi: 10.1029/94JC00572.

Evensen, G., 2003, The ensemble Kalman filter: Theoretical formulation and practical implementation: Ocean Dynamics, **53**, 343–367, doi: 10.1007/s10236-003-0036-9.

Evensen, G., 2009, Data assimilation: The ensemble Kalman filter: Springer.

Gerstoft, P., W. S. Hodgkiss, M. Siderius, C.-F. Huang, and C. H. Harrison, 2008, Passive fathometer processing: Journal of the Acoustical Society of America, **123**, 1297–1305, doi: 10.1121/1.2831930.

Gerstoft, P., and C. F. Mecklenbräuker, 1998, Ocean acoustic inversion with estimation of *a posteriori* probability distributions: Journal of the Acoustical Society of America, **104**, 808–819, doi: 10.1121/1.423355.

Ghosh, A., J. E. Vidale, and K. C. Creager, 2012, Tremor asperities in the transition zone control evolution of slow earthquakes: Journal of Geophysical Research, **117**, B10301, doi: 10.1029/2012JB009249.

Ghosh, A., J. E. Vidale, J. R. Sweet, K. C. Creager, and A. G. Wech, 2009, Tremor patches in Cascadia revealed by seismic array analysis: Geophysical Research Letters, **36**, L17316, doi: 10.1029/2009GL039080.

Gilks, W. R., and C. Berzuini, 2001, Following a moving target — Monte Carlo inference for dynamic Bayesian models: Journal of the Royal Statistical Society Series B, **63**, 127–146, doi: 10.1111/1467-9868.00280.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation: IEEE Proceedings F, Radar and Signal Processing, **140**, 107–113, doi: 10.1049/ip-f-2.1993.0015.

Grana, D., and E. Della Rossa, 2010, Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion: Geophysics, **75**, no. 3, O21–O37, doi: 10.1190/1.3386676.

Gustafsson, F., F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. J. Nordlund, 2002, Particle filters for positioning, navigation, and tracking: IEEE Transactions on Signal Processing, **50**, 425–437, doi: 10.1109/78.978396.

Houtekamer, P. L., and H. L. Mitchell, 2005, Ensemble Kalman filtering: Quarterly Journal of the Royal Meteorological Society, **131**, 3269–3289, doi: 10.1256/qj.05.135.

Huang, C.-F., P. Gerstoft, and W. S. Hodgkiss, 2006, Uncertainty analysis in matched-field geoacoustic inversions: Journal of the Acoustical Society of America, **119**, 197–207, doi: 10.1121/1.2139075.

Jain, R., and Z.-H. Michalopoulou, 2011, A particle filtering approach for multipath arrival time estimation from acoustic time series: Journal of the Acoustical Society of America, **129**, EL236–EL241, doi: 10.1121/1.3574766.

Julier, S., J. Uhlmann, and H. F. Durrant-White, 2000, A new method for nonlinear transformation of means and covariances in filters and estimators: IEEE Transactions on Automatic Control, **45**, 477–482., doi: 10.1109/9.847726.

Kalman, R. E., 1960, A new approach to linear filtering and prediction problems: Transactions of the ASME Journal of Basic Engineering, **82**, 35–45, doi: 10.1115/1.3662552.

Khan, Z., T. Balch, and F. Dellaert, 2005, MCMC-based particle filtering for tracking a variable number of interacting targets: IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 1805–1819, doi: 10.1109/TPAMI.2005.223.

Kong, A., J. S. Liu, and W. H. Wong, 1994, Sequential imputations and Bayesian missing data problems: Journal of the American Statistical Association, **89**, 278–288, doi: 10.1080/01621459.1994.10476469.

Larocque, J. R., J. P. Reilly, and W. Ng, 2002, Particle filters for tracking an unknown number of sources: IEEE Transactions on Signal Processing, **50**, 2926–2937, doi: 10.1109/TSP.2002.805251.

Larsen, A. L., M. Ulvmoen, H. Omre, and A. Buland, 2006, Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model: Geophysics, **71**, no. 5, R69–R78, doi: 10.1190/1.2245469.

Llenos, A. L., and J. J. McGuire, 2011, Detecting aseismic strain transients from seismicity data: Journal of Geophysical Research, **116**, B06305, doi: 10.1029/2010JB007537.

Michalopoulou, Z.-H., C. Yardim, and P. Gerstoft, 2012, Particle filtering for passive fathometer tracking: Journal of the Acoustical Society of America, **131**, EL74–EL80, doi: 10.1121/1.3670004.

Musso, C., N. Oudjane, and F. LeGland, 2001, Improving regularised particle filters, *in* A. Doucet, N. de Freitas, and N. Gordon, eds., Sequential Monte Carlo methods in practice: Springer, 247–271.

Myrseth, I., and H. Omre, 2010, The ensemble Kalman filter and related filters, *in* L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, eds., Large-scale inverse problems and quantification of uncertainty: John Wiley & Sons, 217–246.

Nicoli, M., V. Rampa, and U. Spagnolini, 2002, Hidden Markov model for multidimensional wavefront tracking: IEEE Transactions on Geoscience and Remote Sensing, **40**, 651–662, doi: 10.1109/TGRS.2002.1000324.

Obara, K., 2002, Nonvolcanic deep tremor associated with subduction in southwest Japan: Science, **296**, 1679–1681, doi: 10.1126/science.1070378.

Ó Ruanaidh, J. J. K., and W. J. Fitzgerald, 1996, Numerical Bayesian methods applied to signal processing: Springer-Verlag.

Peng, Z., and J. Gomberg, 2010, An integrated perspective of the continuum between earthquakes and slow-slip phenomena: Nature Geoscience, **3**, 599–607, doi: 10.1038/ngeo940.

Pitt, M., and N. Shephard, 1999, Filtering via simulation: Auxiliary particle filter: Journal of the American Statistical Association, **94**, 590–599, doi: 10.1080/01621459.1999.10474153.

Rimstad, K., and H. Omre, 2013, Approximate posterior distributions for convolutional two-level hidden Markov models: Computational Statistics and Data Analysis, **58**, 187–200, doi: 10.1016/j.csda.2012.09.001.

Ristic, B., S. Arulampalam, and N. Gordon, 2004, Beyond the Kalman filter: Particle filters for tracking applications: Artech House.

Rogers, G., and H. Dragert, 2003, Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip: Science, **300**, 1942–1943, doi: 10.1126/science.1084783.

Schuster, G. T., 2009, Seismic interferometry: Cambridge University Press.

Schwarz, G., 1978, Estimating the dimension of a model: Annals of Statistics, **6**, 461–464, doi: 10.1214/aos/1176344136.

Scott, S. L., 2002, Bayesian methods for hidden Markov models: Journal of the American Statistical Association, **97**, 337–351, doi: 10.1198/016214502753479464.

Segall, P., and M. Matthews, 1997, Time dependent inversion of geodetic data: Journal of Geophysical Research, **102**, 22391–22409, doi: 10.1029/97JB01795.

Seiler, A., G. Evensen, J. A. Skjervheim, J. Hove, and J. G. Vabø, 2010, Using the ensemble Kalman filter for history matching and uncertainty quantification of complex reservoir models, *in* L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, eds., Large-scale inverse problems and quantification of uncertainty: John Wiley & Sons, 247–271.

Siderius, M., H. Song, P. Gerstoft, W. S. Hodgkiss, P. Hursky, and C. Harrison, 2010, Adaptive passive fathometer processing: Journal of the Acoustical Society of America, **127**, 2193–2200, doi: 10.1121/1.3303985.

Smith, A. F. M., and A. E. Gelfand, 1992, Bayesian statistics without tears: A sampling-resampling perspective: American Statistician, **46**, 84–88.

Ulvmoen, M., and H. Hammer, 2010, Bayesian lithology/fluid inversion — Comparison of two algorithms: Computational Geosciences, **14**, 357–367, doi: 10.1007/s10596-009-9155-9.

van der Merwe, R., A. Doucet, N. De Freitas, and E. Wan, 2001, The unscented particle filter: Advances in Neural Information Processing Systems, **13**, 584–590.

van Leeuwen, P. J., 2009, Particle filtering in geophysical systems: Monthly Weather Review, **137**, 4089–4114, doi: 10.1175/2009MWR2835.1.

Vermaak, J., S. J. Godsill, and P. Perez, 2005, Monte Carlo filtering for multi-target tracking and data association: IEEE Transactions on Aerospace and Electronic Systems, **41**, 309–332, doi: 10.1109/TAES.2005.1413764.

Wan, E. A., and R. van der Merwe, 2001, The unscented Kalman filter, *in* S. Haykin, ed., Kalman filtering and neural networks: John Wiley & Sons.

Wapenaar, C., D. Draganov, and J. Robertsson, 2008, Seismic interferometry: History and present status: SEG.

Werner, M. J., K. Ide, and D. Sornette, 2011, Earthquake forecasting based on data assimilation: Sequential Monte Carlo methods for renewal point

processes: Nonlinear Processes in Geophysics, **18**, 49–70, doi: 10.5194/npg-18-49-2011.
Yardim, C., and P. Gerstoft, 2012, Sequential Bayesian techniques applied to non-volcanic tremor: Journal of Geophysical Research, **117**, B10312, doi: 10.1029/2012JB009420.
Yardim, C., P. Gerstoft, and W. S. Hodgkiss, 2009, Tracking of geoacoustic parameters using Kalman and particle filters: Journal of the Acoustical Society of America, **125**, 746–760, doi: 10.1121/1.3050280.
Yardim, C., P. Gerstoft, and W. S. Hodgkiss, 2010, Geoacoustic and source tracking using particle filtering: Experimental results: Journal of the Acoustical Society of America, **128**, 75–87, doi: 10.1121/1.3438475.
Yardim, C., P. Gerstoft, and W. S. Hodgkiss, 2012, Sequential geoacoustic inversion at the continental shelfbreak: Journal of the Acoustical Society of America, **131**, 1722–1732, doi: 10.1121/1.3666012.
Yardim, C., Z.-H. Michalopoulou, and P. Gerstoft, 2011, An overview of sequential Bayesian filtering in ocean acoustics: IEEE Journal of Oceanic Engineering, **36**, 71–89, doi: 10.1109/JOE.2010.2098810.
Zhang, J., P. Gerstoft, P. M. Shearer, H. Yao, J. E. Vidale, H. Houston, and A. Ghosh, 2011, Cascadia tremor spectra: Low corner frequencies and earthquake-like high frequency falloff: Geochemistry, Geophysics, Geosystems, **12**, Q10007, doi: 10.1029/2011GC003759.