Sequential Bayesian techniques applied to non-volcanic tremor

Caglar Yardim[1] and Peter Gerstoft[1]

Received 2 May 2012; revised 7 September 2012; accepted 12 September 2012; published 31 October 2012.

[1] This paper uses sequential Bayesian techniques such as particle filters and smoothers to track in time both the non-volcanic tremor (NVT) source location on the plate interface and the angle of arrival via horizontal phase slowness. Sequential Bayesian techniques enable tracking of evolving geophysical parameters via sequential tremor observations. These techniques provide a formulation where the geophysical parameters that characterize dynamic, non-stationary processes are continuously estimated as new data become available. In addition to the optimal solution, particle filters and smoothers can calculate the underlying probability densities for the desired parameters, providing the uncertainties in the estimates. The tremor tracking has been performed using array beamforming. Here it is demonstrated that the uncertainties both in the NVT source location estimates and phase slowness estimates are reduced using a particle filter compared to just using a beamformer based inversion. Particle smoothers further reduces the uncertainty, giving the best performance out of the three methods used here.

Citation: Yardim, C., and P. Gerstoft (2012), Sequential Bayesian techniques applied to non-volcanic tremor, *J. Geophys. Res.*, *117*, B10312, doi:10.1029/2012JB009420.

1. Introduction

[2] Non-volcanic tremor (NVT) was discovered a decade ago by *Obara* [2002] and *Rogers and Dragert* [2003] and is characterized by low amplitudes, lack of high frequencies, emerging onsets, and duration of minutes to days. NVT is mostly observed near megathrust earthquake source areas, suggesting that a detailed understanding of these may help characterizing catastrophic events [*Ghosh et al.*, 2012; *Obara*, 2011]. The lack of strong impulsive phases across the array makes classic seismic methods less fruitful, but a frequency domain approach with longer observation time combined with sequential estimation is a good alternative. Seismic non-volcanic tremor (NVT) is a continuous noise appearing at regular intervals of around 14 months in Cascadia, WA [*Obara*, 2002; *Rogers and Dragert*, 2003; *Peng and Gomberg*, 2010; *Ghosh et al.*, 2010a; *Vidale and Houston*, 2012]. In Cascadia, the Juan de Fuca plate is underthrusting below the North American plate, see Figure 1. Near the surface, the two plates are locked and, as the Juan de Fuca plate moves east, the North American plate is also dragged east. At regular intervals, seismic tremor is observed and simultaneously it can be observed with GPS 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.

[3] In addition to the currently obtained data, Bayesian sequential techniques enable us to use the previous data to infer the properties of NVT. This is done by combining the information extracted from the sequential data via an environmental evolution model. Here we use particle filters (PF) and smoothers to track the evolving horizontal phase speed and the uncertainty in the estimates. Particle methods are Bayesian sequential techniques that can deal with non-linear, non-Gaussian measurement and environmental evolution equations. The basic idea is that particles are vectors that represent the values of unknown parameters and each particle is a potential solution at a given time. The set of particles in the PF can be used to obtain the PDF/uncertainty in the estimates. This fluid set of particles flow through time and adapt to take into account the new likelihood obtained by the newly collected data. In addition, smoothers enable the algorithm to come back and correct the estimates at t_0 for data collected at $t > t_0$.

[4] This paper starts with an introduction to the PFs in section 2, including the state space formulation, Monte Carlo integration, and importance sampling. It then proceeds by explaining the PF used in this work; the auxiliary sequential importance sampling (ASIR) particle filter. In section 3, smoothing is introduced with derivations of three particle smoothers given. Section 4 deals with the application of these techniques on the NVT data from Cascadia. First, the NVT is tracked using a conventional beamformer [*Zhang et al.*, 2011], then using the ASIR PF, and finally by two different particle smoothers. Both horizontal slowness and source location are tracked.

2. Basics of Particle Filters

[5] Let $\mathbf{y}_t$ be the measurement vector (for example, signal along a seismic array) at step t and $\mathbf{x}_t$ represent the state

[1]Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA.

Corresponding author: C. Yardim, Marine Physical Laboratory 0238, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, USA. (cyardim@ucsd.edu)

©2012. American Geophysical Union. All Rights Reserved.
0148-0227/12/2012JB009420

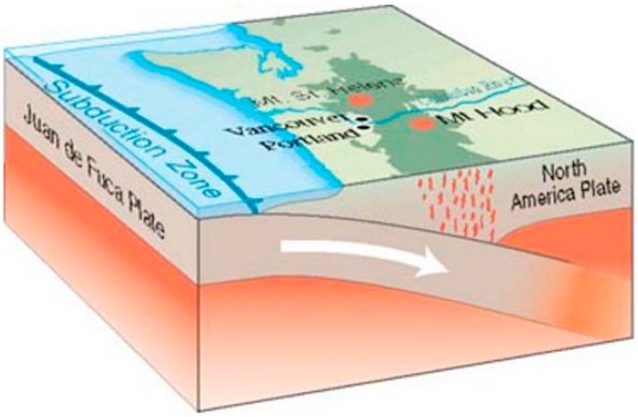

Figure 1. Non-volcanic tremor and structure of the Cascadia subduction zone.

vector (for example, time-varying slowness in x- and y-directions that can be used to give the angle of arrival or longitude and latitude of the tremor source on the slab), 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:

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

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

[6] The *state equation* (1), describes the evolution or transition of $\mathbf{x}_t$ with t and assumes that states follow a first order Markov process. 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 PDF $p(\mathbf{v}_t)$. This noise term represents the error between the modeled temporal evolution of the parameters and the real evolution. It dictates how much the PF should trust the evolution model and enables the PF to adapt to changing conditions. The larger the state noise is, the better the PF is at tracking sudden changes. The drawback is that the estimates become noisier since a larger noise is injected into the state-space model.

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

[8] 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 functions of $\mathbf{f}_t$ and $\mathbf{h}_t$, respectively. The formulation includes fully dynamic, non-stationary cases, where 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.

[9] 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, 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 definition for each problem type is given as:

[10] 1. $p(\mathbf{x}_t|\mathbf{y}_t)$; *inversion*. Used in non-sequential Bayesian estimation. 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 since all the information available at t is not utilized. This is typically used in seismic inversion [*Mosegaard and Tarantola*, 1995; *Sambridge and Mosegaard*, 2002; *Zollo et al.*, 2002; *Pasyanos et al.*, 2006; *Lancieri and Zollo*, 2008; *Bodin and Sambridge*, 2009; *Hauser et al.*, 2011]. A good review paper about Bayesian theory in seismics is given in *Mosegaard* [2011].

[11] 2. $p(\mathbf{x}_t|\mathbf{y}_{1:t})$; *filter*: The most frequently used form of sequential Bayesian estimation. Filtering enables all the previous and current measurements to be used in estimating $\mathbf{x}_t$. Some geophysical applications where a sequential approach has been adapted are geodesy, where a Kalman filter (KF) was used [*Segall and Matthews*, 1997], seismic strains [*Llenos and McGuire*, 2011] using an extended KF (EKF), passive seismic monitoring [*Baziw*, 2005] and geoacoustic inversion [*Yardim et al.*, 2010] where particle filters (PF) were employed. PF for meteorology data assimilation problems were recently reviewed in *van Leeuwen* [2009]. Reflector tracking can also be solved as a sequential estimation problem by using range as the evolving index [*Nicoli et al.*, 2002].

[12] 3. $p(\mathbf{x}_t|\mathbf{y}_{1:T})$; *smoother*. A smoother is appropriate in applications where all data have already been observed and are readily available. Therefore, both past and "future" measurements can be exploited. Smoothing is computationally more expensive than filtering. Although inclusion of future data improves estimation in comparison to a one-way filtering approach, the increase in computational cost sometimes makes smoothing less desirable than filtering. However, recently proposed particle smoothers have computational costs comparable to PFs, which may result in more extensive use of smoothers.

[13] Note that the state vector $\mathbf{x}_t$ at each t constitutes of new random variables whose statistics are different due to the non-stationary nature of the geophysical process. This means the full posterior PDF is $p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t})$. Hence, the posterior PDFs mentioned above are actually marginal PDFs of this full PDF. Using Bayes theorem and the Markov property gives

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

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 estimate $\mathbf{x}_t$. 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 $\mathbf{x}_{t-1}$:

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

Note that $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$ since, 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 (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 utilize this formulation whereby the evolving posterior PDF can be computed recursively as new data $\mathbf{y}_t$ become available. See *Ristic et al.* [2004] and *Yardim et al.* [2011] for a full derivation starting with the full posterior PDF $p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t})$ and prior $p(\mathbf{x}_0)$. The 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

$$\begin{aligned} \chi_t &: \{\mathbf{x}_t^i, w_t^i\}_{i=1}^{N_p} \\ p(\mathbf{x}_t|\mathbf{y}_{1:t}) &\cong \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i). \end{aligned} \tag{6}$$

[14] Sequential Bayesian techniques use importance sampling (IS) and Monte Carlo (MC) integration to compute and propagate the posterior PDF. The IS is a method employed to compute expectations with respect to one density using random samples drawn from another. Assume that we want to compute an integral $I = \int f(\mathbf{x})\mathrm{d}\mathbf{x}$. One way of computing I is using samplers [*Ó Ruanaidh and Fitzgerald*, 1996] after assuming $\mathbf{x}$ is a random variable with PDF $q(\mathbf{x})$. It is possible to rewrite I in the form of an expectation:

$$I \equiv \int_{\mathcal{X}} \left[\frac{f(\mathbf{x})}{q(\mathbf{x})}\right] q(\mathbf{x})\mathrm{d}\mathbf{x} = E_q\left\{\frac{f(\mathbf{x})}{q(\mathbf{x})}\right\}. \tag{7}$$

The integral can now be computed numerically via IS, by drawing N_p independent and identically distributed $\mathbf{x}$ samples from the sampling or proposal density $q(\mathbf{x})$:

$$\{\mathbf{x}^i\}_{i=1}^{N_p} \sim q(\mathbf{x}) \tag{8}$$

$$I \approx \frac{1}{N_p}\sum_{i=1}^{N_p} w(\mathbf{x}^i), \quad \text{where } w(\mathbf{x}^i) = \frac{f(\mathbf{x}^i)}{q(\mathbf{x}^i)} \tag{9}$$

is called the importance weight. The IS converges to the true value with diminishing error variance as the number of particles drawn from $q(\mathbf{x})$ increases. The variance is [*Ó Ruanaidh and Fitzgerald*, 1996]:

$$\mathrm{Var}_q\left(\widehat{I}\right) = \left\{E_q\left[\frac{f(\mathbf{x})}{q(\mathbf{x})}\right]^2 - I^2\right\}/N_p. \tag{10}$$

The variance in the estimate is minimum if $q(\mathbf{x})$ is proportional to $f(\mathbf{x})$ and increases as $q(\mathbf{x})$ deviates from the latter. Using IS requires the selection of a balanced $q(\mathbf{x})$: as easy to sample from as possible without sacrificing the accuracy of the method.

[15] Equation (5) can be numerically calculated using IS [*Ristic et al.*, 2004]. The most popular PF implementation is called sequential importance resampling (SIR) [*Gordon et al.*, 1993; *Doucet et al.*, 2000; *Gilks and Berzuini*, 2001]. A single iteration of a SIR algorithm is illustrated in *Yardim et al.* [2011, Figure 5] for a PF with $N_p = 10$. A cylinder with height proportional to the weight represents each particle. At the prediction stage new particles are created from the particles representing the PDF of the previous step, $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$. The likelihood is calculated for each of these new particles using new data $\mathbf{y}_t$. Resampling follows this update stage, where 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.*, 2001]. A typical SIR type PF will consist of these three sections:

[16] *Predict:* This stage starts with the cloud of equal weight particles from the previous step $\{\mathbf{x}_{t-1}^i\}_{i=1}^{N_p}$, and creates a new set of predictions for the current step $\{\mathbf{x}_{t|t-1}^i\}_{i=1}^{N_p}$ by sampling from the transitional density $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. This is done by propagating each $\mathbf{x}_{t-1}^i$ through the state equation (1) together with a random realization from $\mathbf{v}_t$.

[17] *Update:* Note that $w_{t-1}^i = 1/N_p$ for all i because of resampling at step k-1. Having measured $\mathbf{y}_t$, the weight of each particle is computed and normalized:

$$w_t^i = \frac{p\left(\mathbf{y}_t|\mathbf{x}_{t|t-1}^i\right)}{\sum_{i=1}^{N_p} p\left(\mathbf{y}_t|\mathbf{x}_{t|t-1}^i\right)}, \tag{11}$$

where $p(\mathbf{y}_t|\mathbf{x}_{t|t-1}^i)$ is the likelihood function defined by the measurement equation (equation (2)), which includes statistical behavior of errors in the data. The weights of equation (11) are used in equation (6) for expressing the posterior PDF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$.

[18] *Resample:* New particles $\{\mathbf{x}_{t|t}^j, w_t^j = 1/N_p\}_{j=1}^{N_p}$ are drawn from a discrete approximation to density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ obtained at the *update* stage. All particle weights are now equal to $1/N_p$.

[19] The SIR algorithm has been successfully used to track similar environmental problems such as tracking of the atmospheric properties using electromagnetics [*Yardim et al.*, 2008] and tracking of geoacoustic and ocean properties using acoustics [*Yardim et al.*, 2010]. Here we use a slightly modified version of SIR called auxiliary SIR (ASIR) proposed by *Pitt and Shephard* [1999]. ASIR solves some of the weaknesses of SIR by using a better sampling density. The details and filter comparisons can be found in *Pitt and Shephard* [1999] and *Ristic et al.* [2004]. ASIR uses a function μ_t^i to represent $\mathbf{x}_t$. Commonly used functions are $\mu_t^i = \mathrm{E}[\mathbf{x}_t|\mathbf{x}_{t-1}^i]$ or a sample from $\mu_t^i \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}^i)$. The ASIR algorithm is given in Table 1.

[20] An advantage of sequential Bayesian methods over classical inversion is the reduction in the number of particles needed. In classical Bayesian inversion, the whole search space needs to be explored. However, in sequential Bayesian

Table 1. ASIR PF [*Pitt and Shephard*, 1999]

Step	Description
Resample	Calculate μ_t^i Compute and normalize β_t^i $\beta_t^i = \frac{p(\mathbf{y}_t\|\mu_t^i)w_{t-1}^i}{\sum_{i=1}^{N_p} p(\mathbf{y}_t\|\mu_t^i)w_{t-1}^i}$ Resample new particles $\mathbf{x}_{t-1}^{j_i}$ with indices j_i from the old particles $\mathbf{x}_{t-1}^i$ using β_t^i $\{\mu_t^i, \beta_t^i\}_{i=1}^{N_p} \mapsto \{j_i\}_{i=1}^{N_p}$
Predict	Sample new N_p particles at t $\{\mathbf{x}_t^i\}_{i=1}^{N_p} \sim p(\mathbf{x}_t\|\mathbf{x}_{t-1}^{j_i})$ given $\{\mathbf{x}_{t-1}^{j_i}\}_{i=1}^{N_p}$ using $\mathbf{x}_t^i = \mathbf{f}_t(\mathbf{x}_{t-1}^{j_i}, \mathbf{v}_t^i)$ $i = 1, \ldots, N_p$ where $\mathbf{v}_t^i$ are samples from the state noise PDF.
Update	Compute the likelihood $p(\mathbf{y}_t\|\mathbf{x}_t^i)$ for each $\mathbf{x}_t^i$. Normalize the weights: $w_t^i = \frac{p(\mathbf{y}_t\|\mathbf{x}_t^i)/p(\mathbf{y}_t\|\mu_t^{j_i})}{\sum_{i=1}^{N_p} p(\mathbf{y}_t\|\mathbf{x}_t^i)/p(\mathbf{y}_t\|\mu_t^{j_i})}$ The posterior PDF is approximated by $p(\mathbf{x}_t\|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i)$

filtering the state vector is assumed to follow the state equation. Thus, utilizing the information from the prior step and the known evolution of the states enables a focused search in the state space. Hence, the PF requires a smaller number of particles at each step.The number of particles for the filtering process to successfully estimate posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is problem-dependent. There are four main factors that determine N_p.

[21] The first one is the problem complexity. Nonlinearities in the state and measurement equations, complexity of the underlying posterior PDF, errors in the state and measurement equations, and desired filter output impact the selection of N_p. It takes fewer particles to track the median/MAP than the marginal PDF. Highly non-Gaussian, long tailed, peaked, multi-modal PDFs require a large number of particles. If the state $\mathbf{x}_t$ evolution differs from the evolution in equation (1), the actual state will deviate from predicted $\mathbf{x}_{t|t-1}$. A larger N_p is then needed for tracking.

[22] The second one is the accuracy in track estimates. Increasing the number of particles initially provides a large performance improvement in a PF. It has been shown in *Crisan and Doucet* [2002] that the error is of the order of $\mathcal{O}\left(N_p^{-1}\right)$. However, after a problem dependent N_p is reached, the performance stays relatively flat and increasing the computational cost provides only marginal benefits. An example in geoacoustic tracking is given in *Yardim et al.* [2009].

[23] The third one is the dimension n_x of the state vector. Ideal MC integration is independent of the state dimension for N_p statistically independent particles. However, for sequential filtering the resampling stage creates multiple copies of high likelihood particles, making N_p tightly related to the number of state variables [*Bengtsson et al.*, 2008]. The problem is further complicated by the fact that the importance density in SIR determines convergence. A good sampling density results in N_p increasing linearly with n_x [*Daum and Huang*, 2003]. The required N_p increases exponentially with n_x when the importance density is poorly chosen (often referred to as 'the curse of dimensionality'). Note that the NVT source tracking is a low-dimensional, nonlinear problem suitable for PFs. Many geophysics problems deal with a large number of unknown parameters. This will result in an increase in N_p. An alternative to the PF in these cases would be using an ensemble Kalman filter (EnKF) [*Evensen*, 2003] designed specifically for large dimensional tracking problems. The trade-off is that the PDFs propagated in EnKF have to be Gaussian due to the underlying Kalman framework, limiting their use in highly nonlinear problems.

[24] Finally, an upper limit for N_p is determined by the maximum number of particles that can be processed with limited computational resources, which is important especially for real-time filters. It is possible to have filters with adaptive N_p schemes [*Fox*, 2003]. Coupling PFs with error metrics, the sample size can be adapted depending on the error estimate. Such an implementation is particularly important during the initial steps, when there is a limited amount of information. To address the significant uncertainty, a large number of particles can be sampled, with the cloud size decreasing in subsequent steps.

3. Particle Smoothers

[25] Often in geophysics, the whole observation data set $\mathbf{y}_{1:T}$ is available for analysis. Then all the data can be used to extract the PDF $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ for $t = 1, \ldots, T$. In this approach, time is made to go both forward and backward while maintaining the physical relationships between measurements and model parameters. The ability to use "future" data $\mathbf{y}_{t+1:T}$ in addition to $\mathbf{y}_{1:t}$ improves the estimates and their uncertainty. The techniques that obtain $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ are termed smoothing algorithms.

[26] Under linear/Gaussian assumptions for the state and measurement equations (1) and (2) implementing a smoother is a straightforward extension to the Kalman filter. Just like in a KF, a Kalman smoother is an efficient, fast algorithm. For the nonlinear/non-Gaussian systems numerical particle smoothers are used. These work with the same basic philosophy as a PF; they represent the smoothing density using a set of N_p smoothing particles $\mathbf{x}_{t|T}^i$ and their weights $w_{t|T}^i$:

$$\chi_t^S = \left\{\mathbf{x}_{t|T}^i, w_{t|T}^i\right\}_{i=1}^{N_p} \tag{12}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = \sum_{i=1}^{N_p} w_{t|T}^i \delta\left(\mathbf{x}_t - \mathbf{x}_{t|T}^i\right) \tag{13}$$

A basic smoother formulation is developed by *Kitagawa* [1996]. This is a filtering formulation which estimates the joint PDF $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ from $p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})$. After the filter is run to time T, $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ is obtained as the marginal of the joint density. However, this method only gives accurate filtering results $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, since successive resampling creates a smoothing degeneracy for $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ when $t \ll T$, with an exponentially increasing error for large T-t [*Chopin*, 2004]. For large T-t, there is only one particle with significant weight representing the smoothing density. To prevent this, more accurate smoothing techniques have been developed. Almost all smoothers fall into one of these two categories:

[27] 1. A forward-backward smoother (FBS) runs a forward PF for $t = 1, \ldots, T$. This gives the set $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^{N_p}$

representing the filtering density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. The algorithm then runs a backward correction which updates the weights of the forward PF particles $w_t^i \rightarrow w_{t|T}^i$ so that they represent the smoothing PDF $p(\mathbf{x}_t|\mathbf{y}_{1:T})$. Note that the forward PF particles are used as smoothing particles in (12), $\mathbf{x}_{t|T}^i = \mathbf{x}_t^i$ [*Doucet et al.*, 2000].

[28] A full density sampler that uses a modified FBS algorithm is proposed by *Godsill et al.* [2004] to draw samples from $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T})$.

[29] 2. A two-filter smoother includes two independently running filters. The forward PF is initialized at $t = 0$ with $p(\mathbf{x}_0)$ and runs for $t = 1, \ldots, T$, resulting in $\{\mathbf{x}_t^i, w_t^i\}_{i=1}^{N_p}$. Similarly, the backward PF is initialized at $t = T + 1$ with $\tilde{p}(\mathbf{x}_{T+1})$, running backwards to $t = 0$ where $\tilde{p}(\cdot)$ represents PDFs related tothe backward PF giving $\{\tilde{\mathbf{x}}_t^i, \tilde{w}_t^i\}_{i=1}^{N_p}$. These two sets of particles and their weights are then used to compute new smoothing weights $w_{t|T}^i$. Note that the backward PF particles are used as smoothing particles in (12), $\mathbf{x}_{t|T}^i = \tilde{\mathbf{x}}_t^i$ [*Briers et al.*, 2010].

[30] Recently, a two-filter smoother that use ASIR as both forward and backward PFs is proposed. This two-ASIR smoother not only computes new smoothing weights $w_{t|T}^i$ but also new smoothing particles $\mathbf{x}_{t|T}^i$.

[31] Often only the track of the maximum a posteriori (MAP) estimate is needed and then the much faster MAP smoother in *Godsill et al.* [2001] can be used. MAP smoother runs a forward PF and uses the particles $\mathbf{x}_t^i$ to construct a grid of possible values in time. Afterwards a Viterbi algorithm is run on top of this forward PF to obtain the most probable state trajectory along the grid, effectively giving the MAP of $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T})$.

[32] A problem with particle smoothers is their large computational cost, typically on the order of N_p^2, $O(N_p^2)$. However, there have been significant improvements in particle smoothing in the last decade both in terms of efficiency and accuracy. Recently proposed forward-backward and two-filter based smoothers such as *Klaas et al.* [2006] and *Fearnhead et al.* [2010] operate with an order of complexity of $O(N_p)$.

3.1. Forward-Backward Smoother

[33] The integral that needs to be solved sequentially in a FBS is given by expressing the smoothing density in terms of the filtering density [*Doucet et al.*, 2000]. Similar to (4) it is possible to express $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ as a function of $\mathbf{x}_{t+1}$ by starting from the joint PDF $p(\mathbf{x}_t, \mathbf{x}_{t+1}|\mathbf{y}_{1:T})$ and integrating out $\mathbf{x}_{t+1}$:

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = \int p(\mathbf{x}_t, \mathbf{x}_{t+1}|\mathbf{y}_{1:T}) d\mathbf{x}_{t+1}, \tag{14}$$

$$= \int p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:T}) \mathrm{d}\mathbf{x}_{t+1}. \tag{15}$$

Using Bayes' rule, the first term in (15) is given by

$$p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:t}) = \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{y}_{1:t}) p(\mathbf{x}_t|\mathbf{y}_{1:t})}{p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})}, \tag{16}$$

$$= \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t})}{\int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_t}. \tag{17}$$

Inserting this back into (15) and taking the $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ term out of the integral we arrive at

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = p(\mathbf{x}_t|\mathbf{y}_{1:t}) \int \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:T})}{\int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_t} \mathrm{d}\mathbf{x}_{t+1}. \tag{18}$$

The integral in the denominator of (18) is computed using IS equation (9) and the forward PF set $\{\mathbf{x}_t^k, w_t^k\}_{k=1}^{N_p}$:

$$\int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_t \cong \sum_{k=1}^{N_p} w_t^k p\left(\mathbf{x}_{t+1}|\mathbf{x}_t^k\right). \tag{19}$$

Then, the outer integral is computed using IS from $\{\mathbf{x}_{t+1}^j, w_{t+1}^j\}_{j=1}^{N_p}$ and (19) as

$$\int \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:T})}{\int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_t} \mathrm{d}\mathbf{x}_{t+1} \cong \sum_{j=1}^{N_p} \frac{w_{t+1|T}^j p\left(\mathbf{x}_{t+1}^j|\mathbf{x}_t\right)}{\sum_{k=1}^{N_p} w_t^k p\left(\mathbf{x}_{t+1}^j|\mathbf{x}_t^k\right)}. \tag{20}$$

Inserting (6) and (20) into (18), the smoothing density is obtained as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \cong \sum_{i=1}^{N_p} w_{t|T}^i \delta\left(\mathbf{x}_t - \mathbf{x}_t^i\right), \tag{21}$$

$$w_{t|T}^i = \sum_{j=1}^{N_p} \frac{w_t^i p\left(\mathbf{x}_{t+1}^j|\mathbf{x}_t^i\right)}{\sum_{k=1}^{N_p} w_t^k p\left(\mathbf{x}_{t+1}^j|\mathbf{x}_t^k\right)} w_{t+1|T}^j. \tag{22}$$

where $w_{T|T}^i = w_T^i$. This is computed using recursion from $t = T$ and going back in time, estimating the integral as an IS by correcting the filter weights using the "future" information. Hence, to compute the smoothing density, all we need to do is to compute the new smoothing weight designated by $w_{t|T}^i$ for each particle x_t^i by scaling the old weight w_t^i.

[34] The smoothing algorithm uses the particles of the forward PF as the smoothing particles, *i.e.* $\mathbf{x}_{t|T}^i = \mathbf{x}_t^i$, and only the weights are updated. Hence, if the forward PF particle cloud $\{\mathbf{x}_t^i\}_{i=1}^{N_p}$ poorly corresponds to the smoothing PDF, the forward-backward method will suffer.

3.2. Two-Filter Smoother

[35] Another way of writing the smoothing density is using Bayes' theorem instead of the integral formulation of the forward-backward filter:

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) = p(\mathbf{x}_t|\mathbf{y}_{1:t-1}, \mathbf{y}_{t:T}) \tag{23}$$

$$= \frac{p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) p(\mathbf{y}_{t:T}|\mathbf{x}_t)}{p(\mathbf{y}_{t:T}|\mathbf{y}_{1:t-1})} \tag{24}$$

$$\propto \underbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}_{\textit{forward PF}} \underbrace{p(\mathbf{y}_{t:T}|\mathbf{x}_t)}_{\textit{backward inf.filter}}. \tag{25}$$

The first part of the equation is given by a forward PF. The latter term is in an unconventional $p(\mathbf{y}|\mathbf{x})$ form. This

"density" is called the backward information filter [*Briers et al.*, 2010]. The term $p(\mathbf{y}_{t:T}|\mathbf{x}_t)$ is a likelihood function which is not a proper probability density; its value is not bounded and sampling methods cannot be used to compute it. The trick employed here replaces the backward information filter estimation problem with an equivalent backward PF.

[36] This is done by introducing an artificial prior $\tilde{p}(\mathbf{x}_t)$ and using Bayes' theorem

$$\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T}) \propto p(\mathbf{y}_{t:T}|\mathbf{x}_t)\tilde{p}(\mathbf{x}_t), \tag{26}$$

$$p(\mathbf{y}_{t:T}|\mathbf{x}_t) \propto \frac{\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T})}{\tilde{p}(\mathbf{x}_t)}, \tag{27}$$

to obtain the equivalent backward PF $\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T})$, ensuring a finite integral. Symbol $\tilde{p}(\cdot)$ is used to represent a backward PF. The prior $\tilde{p}(\mathbf{x}_t)$ can be selected as any function, however, as in all IS computations, poor selection of $\tilde{p}(\mathbf{x}_t)$ affects accuracy.

[37] Replacing the backward information filter with its backward PF equivalent form using (27) and writing the forward prediction density in its sequential form derived in (4), we get

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \propto \frac{\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T})}{\tilde{p}(\mathbf{x}_t)} \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1}. \tag{28}$$

The integral is computed using IS equation (9) and the forward PF particles $\{\mathbf{x}_{t-1}^j, w_{t-1}^j\}_{j=1}^{N_p}$ as

$$\int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})\mathrm{d}\mathbf{x}_{t-1} \cong \sum_{j=1}^{N_p} w_{t-1}^j p\left(\mathbf{x}_t|\mathbf{x}_{t-1}^j\right). \tag{29}$$

We now run a PF backwards in time starting from $t = T - 1$ with a prior density $\tilde{p}(\mathbf{x}_T)$. Since the PF runs backwards, the posterior PDF is of the form $\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T})$. The backward PF creates the backward particles and weights $\left\{\tilde{\mathbf{x}}_t^i, \tilde{w}_t^i\right\}_{i=1}^{N_p}$. The backward posterior can then be written as:

$$\tilde{p}(\mathbf{x}_t|\mathbf{y}_{t:T}) \cong \sum_{i=1}^{N_p} \tilde{w}_t^i \delta\left(\mathbf{x}_t - \tilde{\mathbf{x}}_t^i\right). \tag{30}$$

Inserting (29) and (30) into (28) and rearranging we get:

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \cong \frac{1}{\tilde{p}(\mathbf{x}_t)} \sum_{i=1}^{N_p} \tilde{w}_t^i \delta\left(\mathbf{x}_t - \tilde{\mathbf{x}}_t^i\right) \sum_{j=1}^{N_p} w_{t-1}^j p\left(\mathbf{x}_t|\mathbf{x}_{t-1}^j\right), \tag{31}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \cong \sum_{i=1}^{N_p} w_{t|T}^i \delta\left(\mathbf{x}_t - \tilde{\mathbf{x}}_t^i\right), \tag{32}$$

$$w_{t|T}^i = \frac{\tilde{w}_t^i}{\tilde{p}\left(\tilde{\mathbf{x}}_t^i\right)} \sum_{j=1}^{N_p} w_{t-1}^j p\left(\tilde{\mathbf{x}}_t^i|\mathbf{x}_{t-1}^j\right). \tag{33}$$

The two-filter smoother uses the particles obtained by the backward filter, *i.e.* $\mathbf{x}_{t|T}^i = \tilde{\mathbf{x}}_t^i$. Therefore, similarly to the forward-backward smoother, only the weights are recalculated. Thus, if the backward PF cloud $\left\{\tilde{\mathbf{x}}_t^i\right\}_{i=1}^{N_p}$ poorly corresponds to the smoothing PDF, the two-filter method becomes problematic.

3.2.1. Two-ASIR Smoother

[38] Both classical smoothers presented above use the filtering particles that may not be appropriate to represent the smoothing density. This problem can be avoided by using a modified two-filter smoother given by *Fearnhead et al.* [2010] that enables us to specifically sample new particles representing the smoothing density. Since the method requires forward and backward ASIR filters to run, it is referred to as the two-ASIR smoother. This smoother uses forward ASIR results $\{\mathbf{x}_{t-1}^i, w_{t-1}^i\}_{i=1}^{N_p}$ at $t-1$ and backward ASIR results at $\left\{\tilde{\mathbf{x}}_{t+1}^i, \tilde{w}_{t+1}^i\right\}_{i=1}^{N_p}$ at $t+1$ to sample a new set of smoothing particles $\left\{\bar{\mathbf{x}}_t^i\right\}_{i=1}^{N_p}$ that correspond to the underlying smoothing PDF. Let's start by rewriting the backward information filter as

$$p(\mathbf{y}_{t:T}|\mathbf{x}_t) \propto p(\mathbf{y}_t|\mathbf{x}_t)p\left(\mathbf{y}_{t+1:T}|\mathbf{x}_t\right), \tag{34}$$

Using almost an identical derivation given for (4) with the exception of the integral being on $\mathbf{x}_{t+1}$ instead of $\mathbf{x}_{t-1}$, the second term in (34) is expressed as:

$$p\left(\mathbf{y}_{t+1:T}|\mathbf{x}_t\right) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p\left(\mathbf{y}_{t+1:T}|\mathbf{x}_{t+1}\right)\mathrm{d}\mathbf{x}_{t+1}. \tag{35}$$

Replacing the information filter $p(\mathbf{y}_{t+1:T}|\mathbf{x}_{t+1})$ at $t+1$ using (27) gives

$$p\left(\mathbf{y}_{t+1:T}|\mathbf{x}_t\right) \propto \int p(\mathbf{x}_{t+1}|\mathbf{x}_t) \frac{\tilde{p}\left(\mathbf{x}_{t+1}|\mathbf{y}_{t+1:T}\right)}{\tilde{p}(\mathbf{x}_{t+1})} \mathrm{d}\mathbf{x}_{t+1}, \tag{36}$$

$$p(\mathbf{y}_{t:T}|\mathbf{x}_t) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_{t+1}|\mathbf{x}_t) \frac{\tilde{p}\left(\mathbf{x}_{t+1}|\mathbf{y}_{t+1:T}\right)}{\tilde{p}(\mathbf{x}_{t+1})} \mathrm{d}\mathbf{x}_{t+1}, \tag{37}$$

Inserting these back into (25) and rearranging the terms in the form of an integral for than can be computed by the IS equation (9), we have:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:T}) \propto & \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_{t+1}|\mathbf{x}_t) \\ & \times p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \frac{\tilde{p}\left(\mathbf{x}_{t+1}|\mathbf{y}_{t+1:T}\right)}{\tilde{p}(\mathbf{x}_{t+1})} \mathrm{d}\mathbf{x}_{t-1}\mathrm{d}\mathbf{x}_{t+1}. \end{aligned} \tag{38}$$

[39] To compute this integral we need to define a set of ASIR parameters. Following *Fearnhead et al.* [2010] and defining $\bar{(\cdot)}$ as the parameters related to smoothing, we need to select:

[40] 1. an ASIR weight $\bar{\beta}_t^{(j,k)}$,

[41] 2. some ASIR function $\bar{\mu}_t^i = f\left(\mathbf{x}_{t-1}^{j_i}, \tilde{\mathbf{x}}_{t+1}^{k_i}\right)$ for the jth particle of forward PF and kth particle of backward PF,

[42] 3. prediction stage IS density $\bar{q}\left(\bar{\mu}_t^i|\mathbf{x}_{t-1}^{j_i}, \mathbf{y}_t, \tilde{\mathbf{x}}_{t+1}^{k_i}\right)$.

[43] Note that the set $\bar{\mu}_t^i$ forms a good representation for the smoothing density at t; we use those as the smoothing particles $\bar{\mathbf{x}}_t^i = \bar{\mu}_t^i$. Moreover, having both forward and backward filters as ASIR allows to approximate the smoothing $\bar{\beta}_t^{(j,k)} \cong \beta_t^{j_i}\tilde{\beta}_t^{k_i}$ [*Fearnhead et al.*, 2010]. This approximation reduces the method complexity to $O(N_p)$.

Table 2. Two-ASIR Particle Smoother

Step	Description
Forward ASIR	Compute and store $\{\mathbf{x}_t^i, w_t^i, \beta_t^i\}_{i=1}^{N_p}$ for $t = 1, \ldots, T$
	$p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \approx \sum_{i=1}^{N_p} w_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i)$
Backward ASIR	Compute and store $\left\{\tilde{\mathbf{x}}_t^i, \tilde{w}_t^i, \tilde{\beta}_t^i\right\}_{i=1}^{N_p}$ for $t = T, \ldots, 1$
	$\tilde{p}(\mathbf{x}_t \mid \mathbf{y}_{t:T}) \approx \sum_{i=1}^{N_p} \tilde{w}_t^i \delta(\mathbf{x}_t - \tilde{\mathbf{x}}_t^i)$
Resample	To sample a smoothing particle $\bar{\mathbf{x}}_t^i$ at time t
	Resample new forward PF particles $\mathbf{x}_{t-1}^i$ with indices j_i from the old particles $\mathbf{x}_{t-1}^i$ using β_t^i
	$\{\mathbf{x}_{t-1}^i, \beta_t^i\}_{i=1}^{N_p} \mapsto \{j_i\}_{i=1}^{N_p}$
	Resample new backward particles $\tilde{\mathbf{x}}_{t-1}^{k_i}$ with indices k_i from the old particles $\tilde{\mathbf{x}}_{t+1}^i$ using $\tilde{\beta}_t^i$
	$\left\{\tilde{\mathbf{x}}_{t+1}^i, \tilde{\beta}_t^i\right\}_{i=1}^{N_p} \mapsto \{k_i\}_{i=1}^{N_p}$
Predict	Using the forward particle at *t-1*, backward particle at $t + 1$, and the current measurement $\mathbf{y}_t$ obtain smoothing particle
	$\{\bar{\mathbf{x}}_t^i\}_{i=1}^{N_p} \sim \bar{q}(\bar{\mathbf{x}}_t^i \mid \mathbf{x}_{t-1}^{j_i}, \mathbf{y}_t, \tilde{\mathbf{x}}_{t+1}^{k_i})$
Update	Compute $\bar{w}_{t\mid T}^i$ from (40) $\mapsto$ normalize $\bar{w}_{t\mid T}^j = \frac{\bar{w}_{t\mid T}^j}{\sum_{i=1}^{N_p} \bar{w}_{t\mid T}^j}$
	$p(\mathbf{x}_t \mid \mathbf{y}_{1:T}) \approx \sum_{i=1}^{N_p} \bar{w}_{t\mid T}^i \delta(\mathbf{x}_t - \bar{\mathbf{x}}_t^i)$

With these definitions, we can now compute the integral in (38) using IS:

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:T}) \cong \sum_{i=1}^{N_p} \bar{w}_{t\mid T}^i \delta(\mathbf{x}_t - \bar{\mathbf{x}}_t^i), \tag{39}$$

$$\bar{w}_{t\mid T}^i \propto \frac{p\left(\bar{\mathbf{x}}_t^i \mid \mathbf{x}_{t-1}^{j_i}\right) p(\mathbf{y}_t \mid \bar{\mathbf{x}}_t^i) p(\tilde{\mathbf{x}}_{t+1}^{k_i} \mid \bar{\mathbf{x}}_t^i) w_{t-1}^{j_i} \tilde{w}_{t+1}^{k_i} / \tilde{p}(\tilde{\mathbf{x}}_{t+1}^{k_i})}{\bar{q}\left(\bar{\mathbf{x}}_t^i \mid \mathbf{x}_{t-1}^{j_i}, \mathbf{y}_t, \tilde{\mathbf{x}}_{t+1}^{k_i}\right) \beta_t^{j_i} \tilde{\beta}_t^{k_i}}. \tag{40}$$

The two-ASIR smoother is summarized in Table 2.

4. Tremor State Space Model

[44] The NVT state and measurement equations are given by:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t \tag{41}$$

$$\mathbf{y}_t(f_j) = a_t(f_j)\mathbf{d}(\mathbf{x}_t, f_j) + \mathbf{w}_t(f_j), \tag{42}$$

where $\mathbf{w}_t(f_j)$ is a vector of additive non-stationary complex 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 forward model.

[45] There are two possible state vector and forward model selections depending on the required output. In the first one, we are interested in tracking the direction the tremor signal that arrives at the array. Hence, the forward model 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. Here, this formulation will be referred to as tremor arrival angle tracking.

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

[47] Beamforming can be done either in time domain by time-delaying and adding the time-series recorded by the seismic sensors or in 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 since only a small subset of frequencies representative of the frequency band is used in the inversion. Initial analyses show that there is little difference between using 8 or 80 frequencies for a 4–17 Hz beamformer. Hence a frequency domain beamformer is selected here with 8 frequencies.

[48] Since PFs and smoothers are sequential Bayesian techniques capable of giving the PDF of NVT source location at each time step, a Bayesian beamformer is used here for comparison purposes. In fact all these methods will use the exact same beamformer based likelihood function derived below for a fair comparison.

[49] The classical Bartlett power objective function is obtained from the assumption of additive complex Gaussian noise $\mathbf{w}_t(f_j)$ 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 (42), with a corresponding Bartlett processor based, multifrequency likelihood function [*Gerstoft and Mecklenbräuker*, 1998]:

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_\mathrm{f}} \frac{1}{(\pi\nu_j)^{n_\mathrm{h}}} \exp\left[-\frac{||\mathbf{y}_t(f_j) - a_t(f_j)\mathbf{d}(\mathbf{x}_t, f_j)||^2}{\nu_j}\right], \tag{43}$$

where n_h is 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 both time and frequency. In a Bayesian seismic inversion, we first estimate the unknown source term $a_t(f_j)$ for all frequencies at each time step using a maximum likelihood (ML) estimator and insert the result analytically into the likelihood formulation.

[50] The unknown source is estimated analytically via a ML by solving $\partial\mathcal{L}/\partial a_t = 0$:

$$\hat{a}_t(f_j) = \frac{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H} \mathbf{y}_t(f_j)}{||\mathbf{d}(\mathbf{x}_t, f_j)||^2}. \tag{44}$$

Inserting the source estimate back into (43), the likelihood becomes [*Huang et al.*, 2006]

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_\mathrm{f}} \frac{1}{(\pi\nu_t(f_j))^{n_\mathrm{h}}} \exp\left[-\frac{\phi_j(\mathbf{x}_t)}{\nu_t(f_j)}\right], \tag{45}$$

$$\phi_j(\mathbf{x}_t) = \mathbf{y}_t^\mathrm{H}(f_j)\mathbf{y}_t(f_j) - \frac{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H} \mathbf{y}_t(f_j)\mathbf{y}_t(f_j)^\mathrm{H} \mathbf{d}(\mathbf{x}_t, f_j)}{\mathbf{d}(\mathbf{x}_t, f_j)^\mathrm{H} \mathbf{d}(\mathbf{x}_t, f_j)}, \tag{46}$$

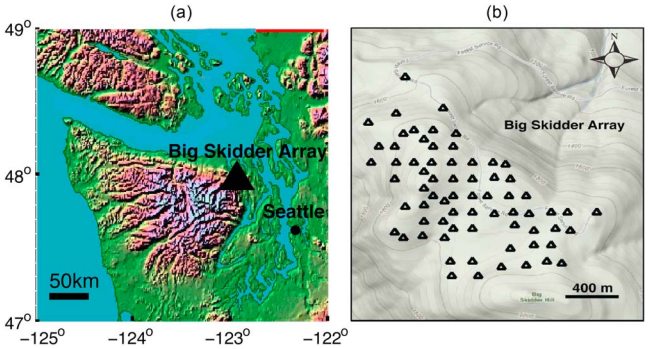

Figure 2. Tremor array deployed in Cascadia.

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 since a classical frequency-domain beamformer is given by $\mathbf{y}_t^{\mathrm{H}}\mathbf{d}(\mathbf{x}_t)$. Hence the likelihood formulation in (45) and (46) enables the conversion of the beamformer output into a Bayesian PDF capable of providing the uncertainty in the estimates and will be shared by the Bayesian beamformer, PF, and both smoothers used in this paper.

[51] As for the noise variance term $\nu_t(f_j)$, one way is to estimate the noise from the data [*Zhang et al.*, 2011] and insert it into (45). An alternative way is to treat the unknown $\nu_t(f_j)$ as a nuisance parameter and use a ML estimator, similar to the one for source estimation. Solving $\partial\mathcal{L}/\partial\nu_t(f_j) = 0$ results in:

$$\widehat{\nu}_t(f_j) = \frac{\phi_j(\mathbf{x}_t)}{n_{\mathrm{h}}}, \tag{47}$$

$$\mathcal{L}(\mathbf{x}_t) = \prod_{j=1}^{n_{\mathrm{f}}} \left(\frac{n_{\mathrm{h}}}{\mathrm{e}\pi\phi_j(\mathbf{x}_t)} \right)^{n_{\mathrm{h}}}. \tag{48}$$

4.1. Cascadia Tremor Analysis

[52] The Bayesian beamformer, PF, FBS, and two-ASIR smoother are used to track the NVT using a 2-D beamforming with a temporary array installed in Cascadia, WA (see Figure 2), to detect seismic tremor [*Ghosh et al.*, 2009, 2010a, 2010b; *Zhang et al.*, 2011]. 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. All data analyses are done for May 7, 2008. In all of the data analyses, the time step is 5 s, providing 720 observations in an hour. Flat priors are used in all methods. The outputs of the Bayesian beamformer, PF, and the smoothers represent $p(\mathbf{x}_t|\mathbf{y}_t)$, $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, respectively. Since the methods respectively use an increasing amount of data, it is natural for the smoother to outperform both the PF and the beamformer, and the PF to outperform the beamformer.

[53] Four cases are analyzed. The first three use tremor source location tracking state space formulation whereas the last one uses tremor arrival angle tracking state-space formulation:

[54] 1. Case I: The effects of the selection of frequency band on tremor source location tracking are investigated. A 400-particle PF is used to track data from 11:40–12:30 UTC using 3–17 Hz, 3–8 Hz, and 10–17 Hz frequency bands.

[55] 2. Case II: A beamformer and PF are compared in tremor source location tracking. Following *Ghosh et al.* [2010a, Figure 6], data from 11:40–12:30 UTC is used. The frequency band is 3.9–17.6 Hz.

[56] 3. Case III: The benefits of using smoothers (FBS and two-ASIR) in tremor source location tracking are explored. Again data from 11:40–12:30 UTC with 3.9–17.6 Hz is used. All filters and smoothers use 400 particles.

[57] 4. Case IV: PF and the smoothers are compared in the alternate tremor arrival angle tracking formulation where we track horizontal slowness using data from 0:00–1:00 UTC. A frequency band of 3.9–17.6 Hz and 400 particles are used.

4.1.1. Tremor Source Location Tracking

[58] From *Zhang et al.* [2011, Figure 3] the frequencies with acceptable tremor signal to noise ratio (SNR) lie between 1.2–18 Hz. The frequency band used in the inversion is expected to make a difference in the source location estimates. Low frequencies such as 3–8 Hz used in *Ghosh et al.* [2010a] have the highest SNR, roughly SNR $\propto f^{-2}$ [*Zhang et al.*, 2011]. However, low frequencies mean small array aperture/wavelength ratio. Assuming 10 km/s phase speed and 1 km array diameter give an array aperture of 0.3, 1, and 1.8λ at 3, 10, and 18 Hz, respectively. This results in fat beamformer lobes with lower angular resolution at low frequencies. Thus, a small amount of error can result in a large shift in the source location estimate. On the other hand, the high frequency range (10–18 Hz) has low SNR but a much sharper beamforming resolution which may compensate for the low SNR. In case I, we tracked the source using the low, high and all usable frequency bands.

[59] The results are given as evolving 1-D probability density functions (PDFs) for the source longitude and latitude in Figure 3. Note that the high frequency PF (Figure 3c) has the sharpest PDFs. The high frequency track is also more stable with smaller variations in NVT source location relative to the low frequency PF result (Figure 3b). The two tracks differ from each other in the first 15 minutes and at around 12:15 UTC, where the NVT signal is weak (Figure 4c). They give similar results between 11:15–12:10 UTC and at the end of the track around 12:30 UTC.

[60] An important aspect of the 2-D beamformer is that it provides an angle of arrival at the array location and an undetermined range. Determining the source range requires an extra piece of information. This is provided by the assumption that the tremor originates on the slab. This enables mapping of the angle of arrival by back-projection onto the slab and this produces an estimate for range.

[61] There are two main factors that can affect the range estimation using back-projection: small arrival angles and low SNR. When the array is far from the tremor source, the propagation path is almost parallel to the surface with a very narrow angle of arrival. Thus, a small uncertainty in the angle estimates (i.e. a narrow angular cone of signals launched

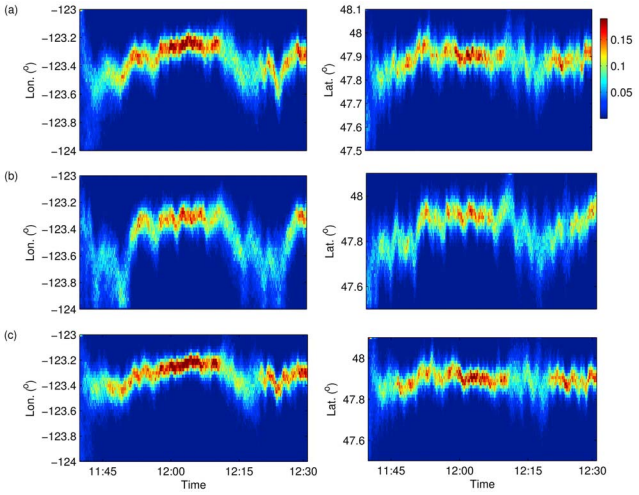

Figure 3. Case I: Effects of the frequency band used for inversion. Evolving PDF of source location obtained by PFs for (a) f = 3–17 Hz, (b) 3–8 Hz, and (c) 10–17 Hz.

back towards the slab) will be back-projected into a large area on the slab. Similarly low SNR will increase the uncertainty in the range estimate. This large range uncertainty results in the typical ellipsoid shaped results with a radially smeared beamformer output (Figures 4a and 4b). Therefore, the back-projection beamformer source location uncertainty can be reduced by using an array as close as possible to the tremor source, which will result in both increased arrival angle and SNR.

[62] Therefore, the rapid variations at low frequencies (Figure 3b) may be due to the weak NVT signal strength coupled with a fat beamforming mainlobe. Other possible scenarios include source longitude and latitude changing faster suggested by the low-frequency case or having different sections of the slab generating NVT at different frequencies. The PF that uses both the low and high frequencies (Figure 3a) gives values in between the two extremes. Determination of an appropriate frequency band that will give the most accurate source localization requires a more detailed analysis and comparisons with results from tremor locations other than Cascadia. For the reminder of this paper we will be using the entire frequency band (3–18 Hz).

[63] Results of case II where beamforming is compared to a PF are given in Figures 4–6. The beamformer is used to statistically estimate the source location by inverting the data at each time step and combining the results of all beamformer inversions to obtain an evolving PDF that can be compared to the PF result. Then a 400-particle PF is used to track the source and the uncertainty in the source estimates. At each time step these 400 particles are first predicted using the state equation (41) and then their weights are updated using the beamformer based likelihood (45) that use the data $\mathbf{y}_t$ arriving at that time step as shown in Table 1. The beamformer and PF are compared at different SNR

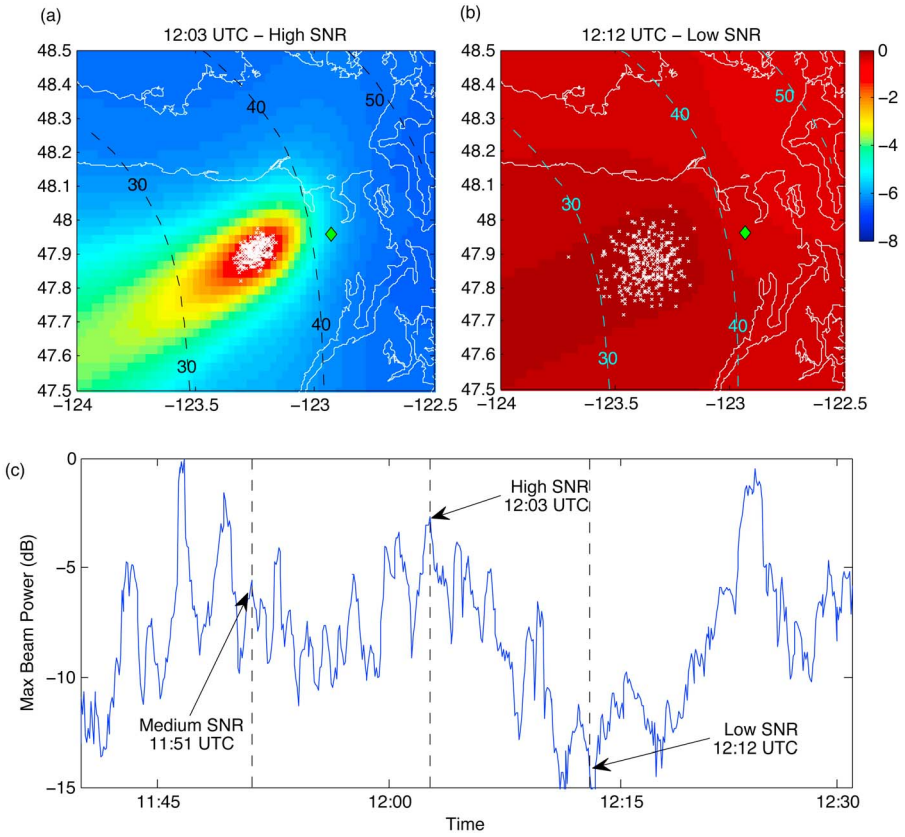

Figure 4. Case II: Beamformer vs. PF comparison. PF particles (×) plotted over beamformer output in dB for (a) high SNR and (b) low SNR. (◇) shows the array location. Dashed lines give the slab depth at that location. (c) Maximum beam power showing the fluctuating NVT source strength.

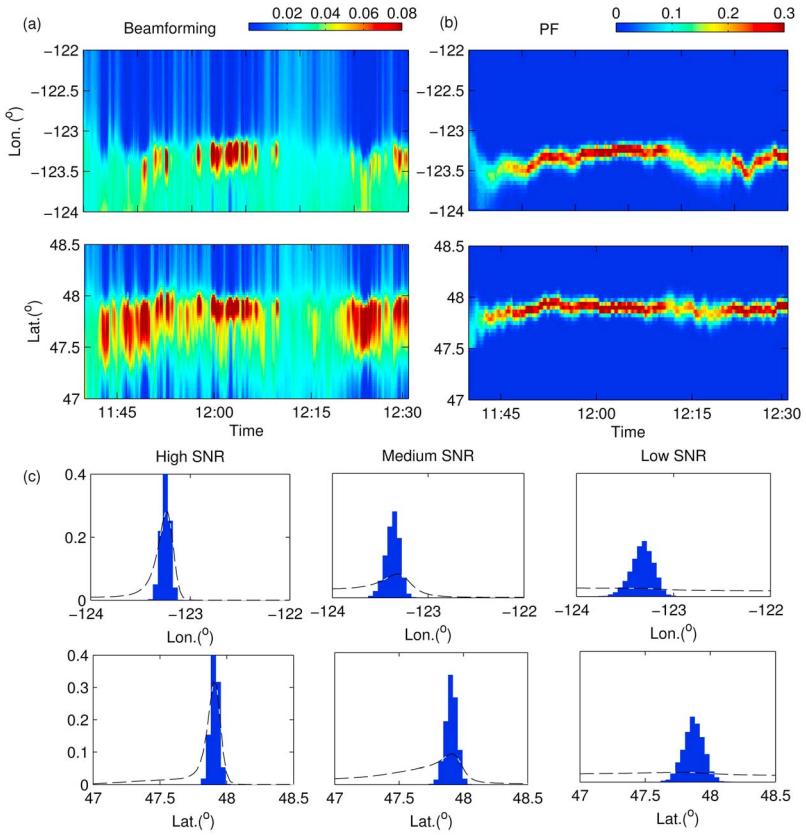

Figure 5. Case II: Beamformer vs. PF comparison. Evolving 1-D PDFs for source longitude and latitude obtained using (a) Bayesian beamforming and (b) a PF. (c) 1-D PDF comparison for beamforming (dashed) and PF (solid histograms) at strong (12:03 UTC), medium (11:51 UTC), and weak (12:12 UTC) NVT source strength.

values (Figure 4c). First, the beamformer objective function ϕ_j in (46) is given in dB together with the PF particles for high and low SNR cases in Figures 4a and 4b. Both methods agree well with high SNR, however, the beamformer objective function at low SNR is flat. Although the particle spread is increased, the PF tracks the source location even in low SNR, due to its sequential formulation given in Table 1.

[64] The beamforming and PF are more easily compared in terms of their posterior PDFs for the source location. Figure 5 shows the evolving PDFs for source longitude and latitude obtained using both methods. The beamforming PDF is much more sensitive to the variations in the tremor SNR and frequently loses the track of tremor location whereas the PF tracks the tremor the entire time even though the uncertainty in Figure 5b increases at low SNR at around 11:40 and 12:15 UTC. Vertical cuts from these plots at three different times representing low, medium, and high SNR are given in Figure 5c. In all cases the PF provides a sharper PDF with low uncertainty in the source location. In addition, for low SNR the beamformer PDF is flat, unable to localize the NVT source. Even though the peak location is around [−123.25°, 47.9°] there is significant uncertainty towards [−124°, 47°] (towards southwest) in the beamformer results, creating an asymmetric long tailed PDF, an indication of the large radial uncertainty that is described earlier.

[65] The PF also has its largest uncertainty radially since both the beamformer and the PF use the same beamformer based likelihood function. However, the state equation (41) used by the PF mitigates the degradation in range at low SNR. The state equation forces the PF to ignore the rapid range fluctuations due to the poor likelihood function of the low SNR data. The PF will only believe the new range inferred from the current data when SNR 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.

[66] Finally, the two methods are compared using the 2-D uncertainty for the back-projected tremor location on the slab in Figure 6 for different SNR values. Again notice that the PF location uncertainty is less than the beamforming results at all SNR levels. Note that the radial uncertainty in the beamforming results have significantly been mitigated in the PF PDFs.

[67] Case III involves comparing the smoothers with the PF results. All methods use 400 particles and all methods start with a flat prior corresponding to a uniformly distributed initial particles. For each time step t the PF first *predicts* the values based on time step t-1 using the state equation (41). Then, it *updates* the values based on the observed data, that is, based on the likelihood function given in (45). The FBS results are obtained by running a backward correction given by (21) and (22) on the PF results. The two-ASIR smoother requires running a second PF backward in time, one that starts at 12:30 and goes back to 11:40 UTC.

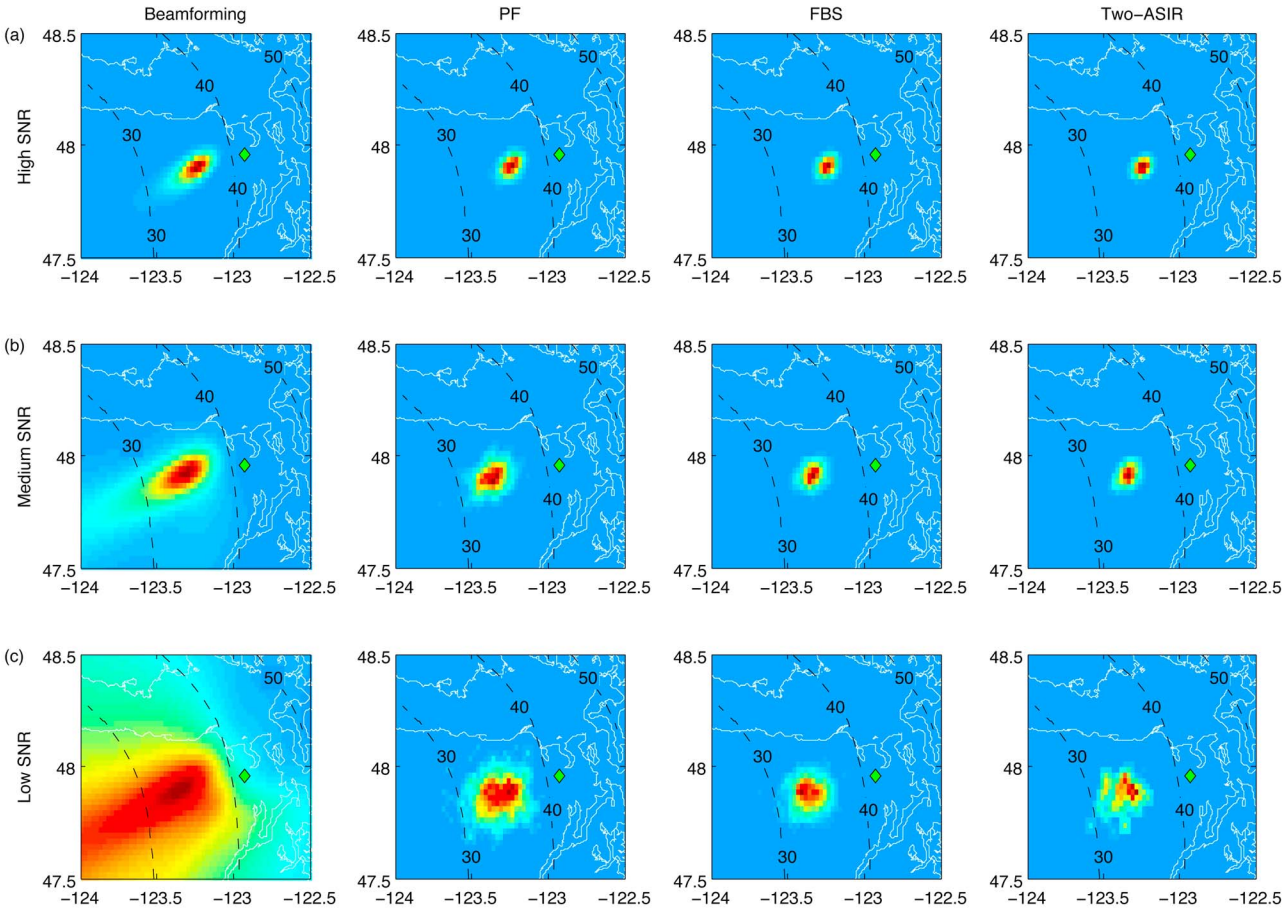

Figure 6. Case II and III: Comparison of Bayesian beamformer, PF, FBS, and two-ASIR smoother. 2-D PDFs for the source location on the slab for (a) strong (12:03 UTC), (b) medium (11:51 UTC), and (c) weak (12:12 UTC) NVT source strength. For comparison purposes, the PDFs are normalized individually to make the peak values 1. (◇) shows the array location. Dashed lines give the slab depth at that location.

Then the forward and backward PFs are combined together to obtain a smoothing distribution using Table 2.

[68] The results for case III are given in terms of source location PDFs on the slab (Figure 6) and 1-D PDFs of the evolving tremor longitude and latitude (Figure 7) for all three algorithms. Note that both smoothers outperform both the beamformer and the PF, resulting in a further reduction in the source location uncertainty. Even at low SNR (Figure 6c) the smoother PDFs are sharp. The smoothers are comparable to each other. Significant reduction in the localization uncertainty enables the PF and the smoothers to keep track of the NVT at all times, whereas the beamformer requires a high SNR to accurately localize the NVT.

[69] The NVT source migration direction reverses several times. As shown in *Ghosh et al.* [2010a, Figure 6] the tremor moves up and down multiple times within an hour. Figure 8 is obtained by plotting the source location PDF at each time step where a reversal occurs in the tremor movement direction and superimposing them to clearly display the beginning and ends of each migration. This is similar to *Ghosh et al.* [2010a, Figure 6] but also provides the underlying uncertainty in the estimates in addition to the point estimate. Notice that every time the source migrates southwest, it

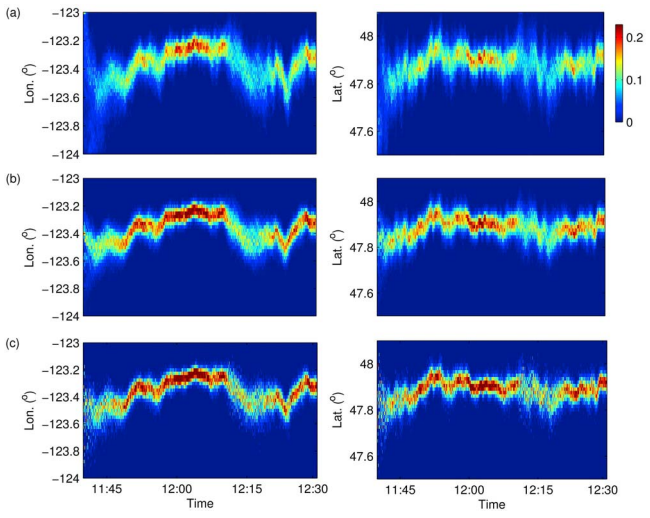

Figure 7. Case III: PF vs. FBS and two-ASIR smoother comparison. Evolving 1-D PDFs for source longitude and latitude obtained using (a) PF, (b) FBS, and (c) two-ASIR smoother.

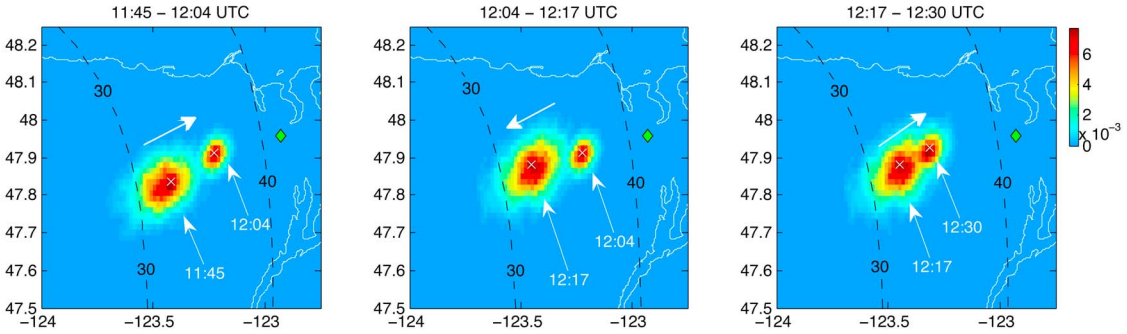

Figure 8. Cascadia NVT source migration between 11:40–12:30 UTC: Rapid back and forth source migration within an hour. The 2-D smoother PDFs at the beginning and end of each trend artificially superimposed for visual display. The arrows indicate the direction of NVT source migration. (×) represents the MAP solution at each time for visual comparison. (◇) shows the array location. Dashed lines give the slab depth at that location.

moves away from the array and the range ambiguity increases as discussed before.

[70] Another interesting observation is the appearance of multiple tremor source locations at various times (Figure 9). These multiple sources are usually short events typically lasting on the order of minutes. As seen in the figure, most of the time a second peak appears in the objective function. However, in rare cases such as at 11:01 UTC three peaks are also observed.

[71] A PF is run to observe the appearance of a second tremor source. The results (Figure 10) show that the transition from a single to a double source happens under a minute in all cases analyzed. The migration rate is also faster then the average NTV migration speed. Once formed, both sources evolve, sometimes alternating in strength, where eventually one dies out.

4.1.2. Tremor Arrival Angle Tracking

[72] In case IV, the Bayesian beamformer, PF and the smoothers are also used to track the peak slowness (slowness is here the horizontal phase slowness or inverse phase speed) s_x and s_y across the array for 1 h starting at midnight of 7 May 2008. From the data, we extract the 2-D 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 = 1) with 400 uniformly distributed particles across (s_x, s_y).

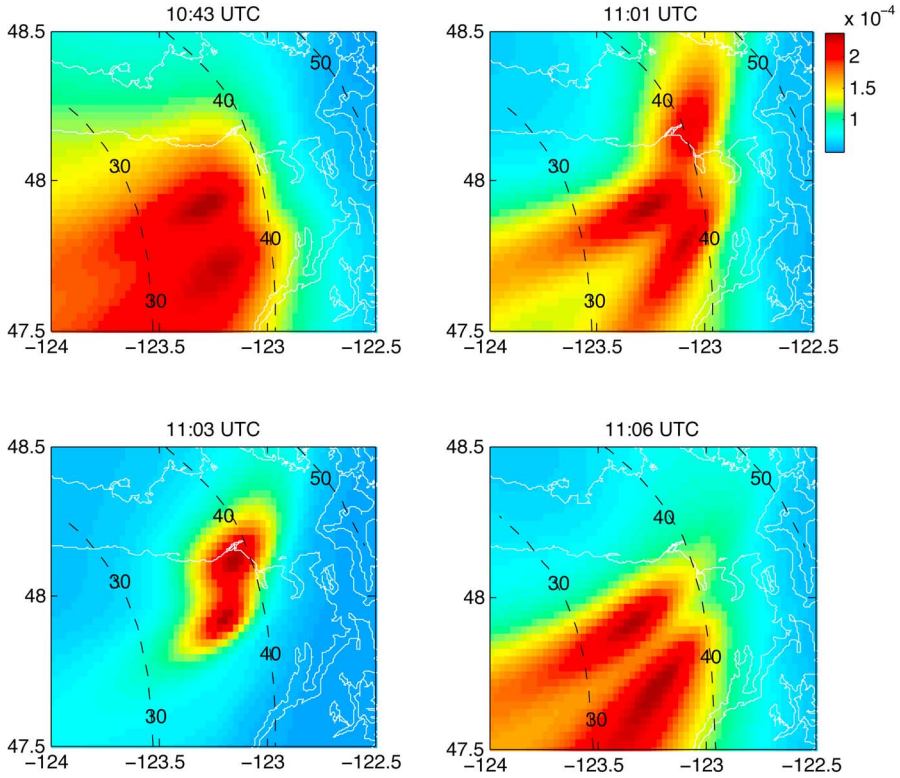

Figure 9. Beamformer output showing multiple sources at different times along the track. Dashed lines give the slab depth at that location.

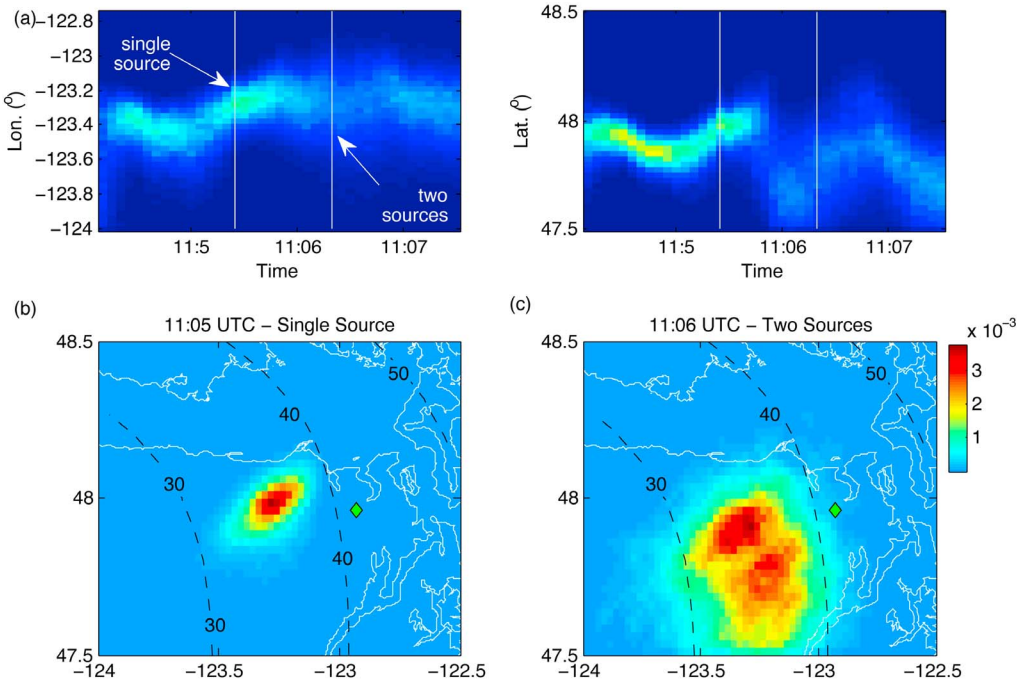

Figure 10. Evolution of source location obtained by a PF as a second peak appears. (a) 1-D PDFs of tremor longitude and latitude. 2-D PDF of tremor location on the slab (b) before, and (c) after a second peak appears.

[73] The beamformer $p(\mathbf{x}_t|\mathbf{y}_t)$, PF $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and smoothing $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ results for horizontal slowness parameters are given in Figure 11. The ability to incorporate the dataset up to t allows the PF to perform better than the beamformer and the ability to use the entire data set in parameter estimation allows the smoothers to have a smaller uncertainty in the PDFs than the PF.

[74] The distribution of the particles for the three methods used here can be seen in Figure 12. Figure 12a is the result at a low SNR. The spread of the particles in the two smoothers is similar and significantly less than the spread in the PF. As the SNR increases the distributions of the particles get narrower indicating lower uncertainty in the estimates. In angle tracking, the two-ASIR smoother outperforms the FBS at medium to high SNR.

[75] The root time averaged mean squared error (RTAMS) given in *Ristic et al.* [2004] and *Yardim et al.* [2008] calculated for the interval $t = [t_1, t_2]$ is the performance metric that we use:

$$\mathrm{RTAMS} = \left[\sum_{t=t_1}^{t_2} \sum_{i=1}^{N_p} \frac{\left(\mathbf{x}_t^i - \widehat{\mathbf{x}}_t\right)^2}{(t_2 - t_1 + 1)N_p} \right]^{1/2}, \qquad (49)$$

where $\widehat{\mathbf{x}}_t$ is the mean state vector at t, computed using the particle cloud. The RTAMS error for s_x calculated for the entire track gives 45×10^{-3} s/km for the PF. This uncertainty is reduced by 36% for the FBS, and 38% for the two-ASIR smoother. Even though the smoothers have almost identical RTAMS errors for the entire track, the FBS performs better for the first 30 min with RTAMS values of 54, 33, and 34×10^{-3} s/km for the PF, FBS, and the two-ASIR smoother, respectively. After the first 30 min, the two-ASIR smoother outperforms the FBS with RTAMS values of 37, 26, and 23×10^{-3} s/km, respectively.

5. Conclusions

[76] Non-volcanic tremors (NVT) contain important information about the processes at the deep roots of the faults

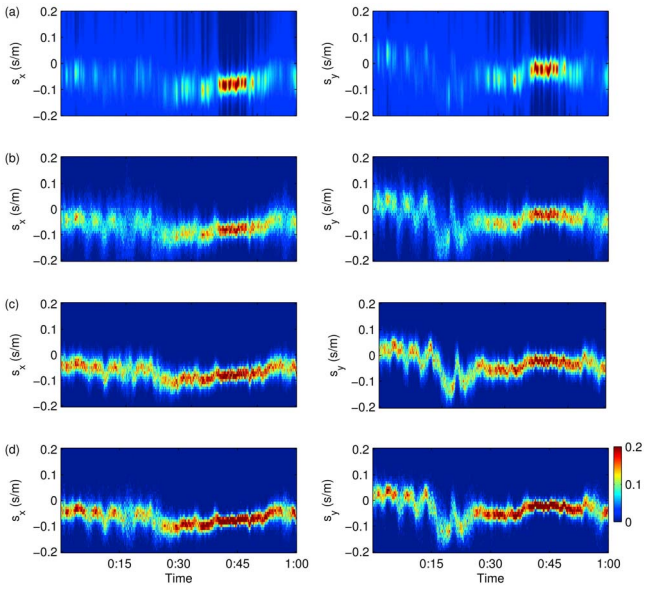

Figure 11. Case IV: Time evolution of the posterior probability of peak slowness (s/km) for one hour for (a) beamformer, (b) PF, (c) FBS, and (d) two-ASIR smoother.

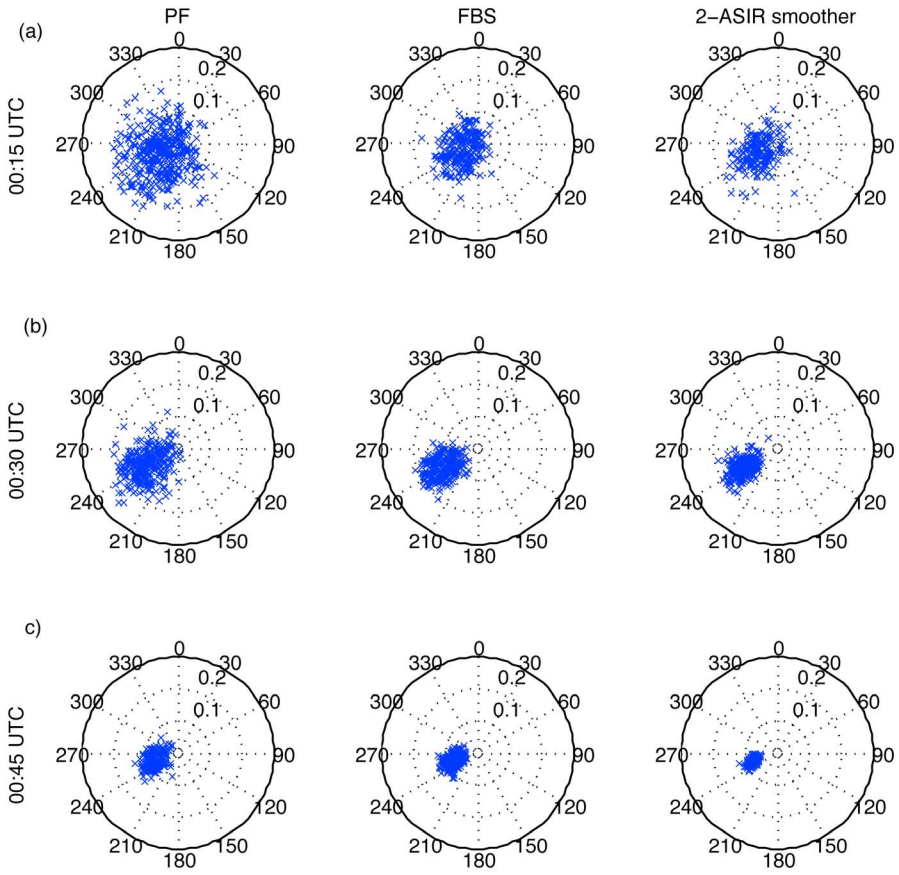

Figure 12. Case IV: Slowness-azimuth outputs for an ASIR particle filter, a forward-backward smoother, and the two-ASIR smoother at time three different time steps. Each (×) is a particle with identical weight so the histogram of the particles represent the underlying posterior probability density.

controlling episodic tremor and slip. In this paper, the Cascadia tremor was analyzed using sequential Bayesian techniques. The aim is to use better signal processing techniques to characterize the source of the tremor. The theoretical background for sequential Bayesian methods was summarized along with the concept of filtering and smoothing. The difference between a particle filter (PF) that uses previous data obtained up to and including the current time and a smoother that uses data from the entire dataset that includes future values were explained. Two main smoothing frameworks, the forward-backward and the two-filter smoothers, were compared. Particular emphasis was given to a recently developed two-ASIR smoother.

[77] As demonstrated, geophysical data offer a rich environment for implementing sequential Bayesian filtering. The Cascadia seismic tremor tracking example showed how to form the state and measurement equations, compute the likelihood, and implement the geophysical PF. The PF was shown to outperform classical beamformer and continue tracking the NVT even in low SNR conditions. Smoothing and filtering were compared in the seismic tremor smoothing example. It was shown that using future data in addition to past data improves the quality of the estimates and reduces the uncertainty.

[78] Both Cascadia NVT source location on the slab and angle of arrival in terms of horizontal slowness were tracked. The effects of the frequency band used in tracking was analyzed. Rapid back and forth source migrations were explored. Multiple NTV source location events were discussed. A particle filter was used to track to track a transition from a single source to a double one.

[79] **Acknowledgments.** This work has been supported by NSF grants EAR-0944109, OCE-1030022, and ONR grant N00014-11-1-0320.

References

Baziw, E. (2005), Real-time seismic signal enhancement utilizing a hybrid Rao-Blackwellized particle filter and hidden Markov model filter, *IEEE Geosci. Remote Sens. Lett.*, *2*(4), 418–422, doi:10.1109/LGRS.2005.852711.

Bengtsson, T., P. Bickel, and B. Li (2008), Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in *Probability and Statistics*, *IMS Collect.*, vol. 2, pp. 316–334, Inst. of Math. Stat., Bethesda, Md., doi:10.1214/193940307000000518.

Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, *178*(3), 1411–1436, doi:10.1111/j.1365-246X.2009.04226.x.

Briers, M., A. Doucet, and S. R. Maskell (2010), Smoothing algorithms for state-space models, *Ann. Inst. Stat. Math.*, *62*(1), 61–89, doi:10.1007/s10463-009-0236-2.

Chopin, N. (2004), Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference, *Ann. Stat.*, *32*(6), 2385–2411, doi:10.1214/009053604000000698.

Crisan, D., and A. Doucet (2002), A survey of convergence results on particle filtering methods for practitioners, *IEEE Trans. Signal Process.*, *50*(3), 736–746, doi:10.1109/78.984773.

Daum, F., and J. Huang (2003), Curse of dimensionality and particle filters, paper presented at Aerospace Conference, Inst. of Electr. and Electron. Eng., Big Sky, Mont., doi:10.1109/AERO.2003.1235126.

Doucet, A., S. Godsill, and C. Andrieu (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.*, *10*(3), 197–208, doi:10.1023/A:1008935410038.

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

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*(4), 343–367, doi:10.1007/s10236-003-0036-9.

Fearnhead, P., D. Wyncoll, and J. Tawn (2010), A sequential smoothing algorithm with linear computational cost, *Biometrika*, *97*(2), 447–464, doi:10.1093/biomet/asq013.

Fox, D. (2003), Adapting the sample size in particle filters through KLD-sampling, *Int. J. Rob. Res.*, *22*(12), 985–1003, doi:10.1177/0278364903022012001.

Gerstoft, P., and C. F. Mecklenbräuker (1998), Ocean acoustic inversion with estimation of *a posteriori* probability distributions, *J. Acoust. Soc. Am.*, *104*(2), 808–819, doi:10.1121/1.423355.

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, *Geophys. Res. Lett.*, *36*, L17316, doi:10.1029/2009GL039080.

Ghosh, A., J. E. Vidale, J. R. Sweet, K. C. Creager, A. G. Wech, H. Houston, and E. E. Brodsky (2010a), Rapid, continuous streaking of tremor in Cascadia, *Geochem. Geophys. Geosyst.*, *11*, Q12010, doi:10.1029/2010GC003305.

Ghosh, A., J. E. Vidale, J. R. Sweet, K. C. Creager, A. G. Wech, and H. Houston (2010b), Tremor bands sweep Cascadia, *Geophys. Res. Lett.*, *37*, L08301, doi:10.1029/2009GL042301.

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

Gilks, W. R., and C. Berzuini (2001), Following a moving target—Monte Carlo inference for dynamic Bayesian models, *J. R. Stat. Soc. B*, *63*, 127–146, doi:10.1111/1467-9868.00280.

Godsill, S., A. Doucet, and M. West (2001), Maximum a posteriori sequence estimation using Monte Carlo particle filter, *Ann. Inst. Stat. Math.*, *53*, 82–96, doi:10.1023/A:1017968404964.

Godsill, S. J., A. Doucet, and M. West (2004), Monte Carlo smoothing for nonlinear time series, *J. Am. Stat. Assoc.*, *99*, 156–168, doi:10.1198/016214504000000151.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993), Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proc. F Radar Signal Process.*, *140*(2), 107–113.

Hauser, J., K. M. Dyer, M. E. Pasyanos, H. Bungum, J. I. Faleide, S. A. Clark, and J. Schweitzer (2011), A probabilistic seismic model for the European Arctic, *J. Geophys. Res.*, *116*, B01303, doi:10.1029/2010JB007889.

Huang, C.-F., P. Gerstoft, and W. S. Hodgkiss (2006), Uncertainty analysis in matched-field geoacoustic inversions, *J. Acoust. Soc. Am.*, *119*(1), 197–207, doi:10.1121/1.2139075.

Kitagawa, G. (1996), Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graph. Stat.*, *5*, 1–25.

Klaas, M., M. Briers, N. De Freitas, A. Doucet, S. Maskell, and D. Lang (2006), Fast particle smoothing: If I had a million particles, paper presented at International Conference on Machine Learning, Assoc. for Comput. Mach., Pittsburgh, Pa., doi:10.1145/1143844.1143905.

Lancieri, M., and A. Zollo (2008), A Bayesian approach to the real-time estimation of magnitude from the early P and S wave displacement peaks, *J. Geophys. Res.*, *113*, B12302, doi:10.1029/2007JB005386.

Llenos, A. L., and J. J. McGuire (2011), Detecting aseismic strain transients from seismicity data, *J. Geophys. Res.*, *116*, B06305, doi:10.1029/2010JB007537.

Mosegaard, K. (2011), Quest for consistency, symmetry, and simplicity—The legacy of Albert Tarantola, *Geophysics*, *76*(5), W51–W61, doi:10.1190/geo2010-0328.1.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447, doi:10.1029/94JB03097.

Nicoli, M., V. Rampa, and U. Spagnolini (2002), Hidden Markov model for multidimensional wavefront tracking, *IEEE Trans. Geosci. Remote Sens.*, *40*(3), 651–662, doi:10.1109/TGRS.2002.1000324.

Obara, K. (2002), Nonvolcanic deep tremor associated with subduction in southwest Japan, *Science*, *296*(5573), 1679–1681, doi:10.1126/science.1070378.

Obara, K. (2011), Characteristics of and interactions between non-volcanic tremor and related slow earthquakes in the Nankai subduction zone, southwest Japan, *J. Geodyn.*, *52*(3–4), 229–248, doi:10.1016/j.jog.2011.04.002.

Ó Ruanaidh, J. J. K., and W. J. Fitzgerald (1996), *Numerical Bayesian Methods Applied to Signal Processing*, Springer, New York.

Pasyanos, M. E., G. A. Franz, and A. L. Ramirez (2006), Reconciling a geophysical model to data using a Markov chain Monte Carlo algorithm: An application to the Yellow Sea–Korean Peninsula region, *J. Geophys. Res.*, *111*, B03313, doi:10.1029/2005JB003851.

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

Pitt, M., and N. Shephard (1999), Filtering via simulation: Auxiliary particle filter, *J. Am. Stat. Assoc.*, *94*, 590–599, doi:10.1080/01621459.1999.10474153.

Ristic, B., S. Arulampalam, and N. Gordon (2004), *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, Boston, Mass.

Rogers, G., and H. Dragert (2003), Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip, *Science*, *300*(5627), 1942–1943, doi:10.1126/science.1084783.

Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, *40*(3), 1009, doi:10.1029/2000RG000089.

Segall, P., and M. Matthews (1997), Time dependent inversion of geodetic data, *J. Geophys. Res.*, *102*(B12), 22,391–22,409, doi:10.1029/97JB01795.

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

Vidale, J. E., and H. Houston (2012), Slow slip: A new kind of earthquake, *Phys. Today*, *65*(1), 38–43, doi:10.1063/PT.3.1399.

Yardim, C., P. Gerstoft, and W. S. Hodgkiss (2008), Tracking refractivity from clutter using Kalman and particle filters, *IEEE Trans. Antennas Propagat.*, *56*(4), 1058–1070, doi:10.1109/TAP.2008.919205.

Yardim, C., P. Gerstoft, and W. S. Hodgkiss (2009), Tracking of geoacoustic parameters using Kalman and particle filters, *J. Acoust. Soc. Am.*, *125*(2), 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, *J. Acoust. Soc. Am.*, *128*(1), 75–87, doi:10.1121/1.3438475.

Yardim, C., Z.-H. Michalopoulou, and P. Gerstoft (2011), An overview of sequential Bayesian filtering in ocean acoustics, *IEEE J. Oceanic Eng.*, *36*(1), 73–91, 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, *Geochem. Geophys. Geosyst.*, *12*, Q10007, doi:10.1029/2011GC003759.

Zollo, A., L. D'auria, R. De Matteis, A. Herrero, J. Virieux, and P. Gasparini (2002), Bayesian estimation of 2-D P-velocity models from active seismic arrival time data: Imaging of the shallow structure of Mt Vesuvius (southern Italy), *Geophys. J. Int.*, *151*(2), 566–582, doi:10.1046/j.1365-246X.2002.01795.x.