Particle filters for partially observed diffusions

Paul Fearnhead
Lancaster University, UK

and Omiros Papaspiliopoulos and Gareth O. Roberts
University of Warwick, Coventry, UK

[Received February 2006. Final revision December 2007]

Summary. We introduce a novel particle filter scheme for a class of partially observed multivariate diffusions. We consider a variety of observation schemes, including diffusion observed with error, observation of a subset of the components of the multivariate diffusion and arrival times of a Poisson process whose intensity is a known function of the diffusion (Cox process). Unlike currently available methods, our particle filters do not require approximations of the transition and/or the observation density by using time discretizations. Instead, they build on recent methodology for the exact simulation of the diffusion process and the unbiased estimation of the transition density. We introduce the generalized Poisson estimator, which generalizes the Poisson estimator of Beskos and co-workers. A central limit theorem is given for our particle filter scheme.

Keywords: Auxiliary variables; Central limit theorem; Continuous time particle filtering; Cox process; Exact algorithm

1. Introduction

There is considerable interest in using diffusion processes to model continuous time phenomena in many diverse scientific disciplines. These processes can be used to model directly the observed data and/or to describe unobserved processes in a hierarchical model. This paper focuses on estimating the path of the diffusion given partial information about it. We develop novel particle filters for analysing a class of multivariate diffusions which are partially observed at a set of discrete time points.

Particle filtering methods are standard Monte Carlo methods for analysing partially observed discrete time dynamic models (Doucet et al., 2001). They involve estimating the filtering densities of interest by a swarm of weighted particles. The approximation error decreases as the number of particles, $N$, increases. However, filtering for diffusion processes is significantly more difficult than for discrete time Markov models since the transition density of the diffusion is unavailable in all except a few special cases. In many contexts even the observation density is intractable. Therefore, the standard propagation–weighting–resampling steps in the particle filter algorithm cannot be routinely applied.

To circumvent these complications, a further approximation, based on a time discretization of the diffusion, has been suggested (see for example Crisan et al. (1999) and Del Moral et al. (2001)). The propagation of each particle from one observation time to the next is done by
splitting the time increment into $M$, say, pieces and performing $M$ intermediate simulations according to an appropriate Gaussian distribution. As $M$ grows large this Gaussian approximation converges to the true diffusion dynamics. In this framework the computational cost of the algorithm is of order $M \times N$, and the true filtering distributions are obtained as both $M$ and $N$ increase.

Our approach does not rely on time discretization but builds on recent work on the exact algorithm (EA) for the simulation of diffusions (Beskos and Roberts, 2005; Beskos et al., 2006a, 2005a) and on the unbiased estimation of the diffusion transition density (Beskos et al., 2006b, 2005b). This algorithm can be used in a variety of ways to avoid time discretizations in the filtering problem. The potential of the EA in the filtering problem was brought up in the discussion of Beskos et al. (2006b); see the contributions by Chopin, Künsch and, in particular, Rousset and Doucet, who also suggested the use of a random-weight particle filter in this context.

One possibility is simply to use the EA to propagate the particles in the implementation of the bootstrap particle filter of Gordon et al. (1993), thus avoiding entirely the $M$ intermediate approximate simulations between each pair of observation times. We call this the exact propagation particle filter (EPPF). Where possible, a better approach is to adapt the EA to simulate directly from (a particle approximation to) the filtering density by using rejection sampling; we term this the exact simulation particle filter (ESPF).

However, our favoured method goes in a different direction. We work in the framework of the auxiliary particle filter of Pitt and Shephard (1999), where particles are propagated from each observation time to the next according to a user-specified density and then are appropriately weighted to provide a consistent estimator of the new filtering distribution. Because the transition density is unavailable, the weights that are associated with each particle are intractable. However, our approach is to assign to each particle a random positive weight which is an unbiased estimator of the true weight. We call this the random-weight particle filter (RWPF). Our algorithm yields consistent estimates of the filtering distributions. The replacement of the weights in a particle filter by positive unbiased estimators is an interesting possibility in more general contexts than the one that is considered in this paper. Indeed, in Section 3.2 we show that this approach amounts to a convenient augmentation of the state with auxiliary variables.

The construction of the unbiased estimators of the weights is one of the main contributions of this paper, and it is of independent interest. This is based on an extension of the Poisson estimator (PE) of Beskos et al. (2006b), which we call the generalized Poisson estimator (GPE). This estimator is guaranteed to return positive estimates (unlike the PE) and its efficiency (in terms of variance and computational cost) can be up to orders of magnitude better than the PE. Optimal implementation of the PE and the GPE is thoroughly investigated theoretically and via simulation.

All three time-discretization-free particle filters that we introduce are easy to implement, with the RWPF being the easiest and the most flexible to adapt to contexts that are more general than those considered here. A simulation study is carried out which shows that the RWPF is considerably more efficient than the ESPF, which is more efficient than the EPPF. We also provide a theoretical result which shows that our filters can have significant computational advantages over time discretization methods. We establish a central limit theorem (CLT) for the estimation of expectations of the filtering distributions using either of the EPPF, ESPF or RWPF. This is an extension of the results of Chopin (2004). The CLT shows that, for a fixed computational cost $K$, the errors in the particle approximation of the filtering distributions decrease as $K^{-1/2}$ in our methods, whereas it is known that the rate is $K^{-1/3}$ or slower in time discretization methods.
The main limitation of the methodology that is presented here is the requirement that the stochastic differential equation (SDE) specifying the underlying diffusion process can be transformed to one with an orthogonal diffusion matrix, and gradient drift. Although this framework excludes some important types of model (such as stochastic volatility models) it incorporates a wide range of processes which can model successfully many physical processes. However, our methods can handle a variety of discrete time observation schemes. In this paper we consider three schemes: noisy observations of a diffusion process, observation of a subset of the components of a multivariate diffusion and arrival times of a Poisson process whose intensity is stochastic and it is given by a known function of a diffusion.

The paper is organized as follows. Section 2 introduces the model for the underlying diffusion and the necessary notation, the observation schemes we consider and the simulated data sets on which we test our proposed methods. Section 3 introduces the RWPF and states the CLT. Section 4 introduces the main tool that is required in constructing the RWPF: the GPE. Several theoretical results are established for the GPE, and a simulation study is performed to assess its performance. Section 5 is devoted to the empirical investigation of the performance of the various particle filters that we introduce. Several implementation issues are also discussed. Section 6 closes with a discussion on extensions of the methodology and the appendices contain technical results and proofs.

2. Signal, data and assumptions

Let the signal be modelled by a d-dimensional diffusion process

\[ dX_s = \alpha(X_s) ds + dB_s, \quad s \in [0, t]. \] (1)

We assume throughout the paper that the drift is known. Our approach requires some assumptions which we summarize in this paragraph:

(a) \( \alpha \) is continuously differentiable in all its arguments,
(b) there is a function \( A : \mathbb{R}^d \to \mathbb{R} \) such that \( \alpha(u) = \nabla A(u) \) and
(c) there is an \( l > -\infty \) such that \( \phi(u) : = \{ \| \alpha(u) \|^2 + \nabla^2 A(u) \}/2 - l \geq 0 \).

Of these three conditions (a) and (c) are weak and the strictest is (b), which in the ergodic case corresponds to \( X \) being a time reversible diffusion.

The transition density of process (1) is typically intractable but a useful expression is available (see for example Beskos et al. (2006b) and Dacunha-Castelle and Florens-Zmirou (1986)):

\[ p_t(x_t|x_0) = \mathcal{N}_t(x_t - x_0) \exp \left\{ A(x_t) - A(x_0) - lt \right\} \mathbb{E} \left[ \exp \left\{ - \int_0^t \phi(W_s) ds \right\} \right]. \] (2)

In this expression \( \mathcal{N}_t(u) \) denotes the density of the d-dimensional normal distribution with mean 0 and variance \( tI_d \) evaluated at \( u \in \mathbb{R}^d \), and the expectation is taken with respect to a Brownian bridge, \( W_s \), \( s \in [0, t] \), with \( W_0 = x_0 \) and \( W_t = x_t \). The expectation in this formula typically cannot be evaluated.

The data consist of partial observations \( y_1, y_2, \ldots, y_n \), at discrete time points \( 0 \leq t_1 < t_2 < \ldots < t_n \). We consider three possible observation regimes.

(i) **Diffusion observed with error**: the observation \( y_i \) is related to the signal at time \( t_i \) via a known density function \( f(y_i|x_{t_i}) \). This model extends the general state space model by allowing the signal to evolve continuously in time. There is a wide range of applications which fit in this framework; see Doucet et al. (2001) for references.
(ii) Partial information: at time $t_i$ we observe $y_i = \zeta(X_{t_i})$ for some non-invertible known function $\zeta(\cdot)$. For example we may observe a single component of the $d$-dimensional diffusion. In this model type $f(y_i|X_{t_i}) = 1$ for all $X_{t_i}$ for which $\zeta(X_{t_i}) = y_i$.

(iii) Cox process: in this regime the data consist of the observation times $t_i$ which are random and are assumed to be the arrivals of a Poisson process of rate $\nu(T_i)$, for some known function $\nu$. Such models are popular in insurance (Dassios and Jang, 2005) and finance (Engel, 2000; Duffie and Singleton, 1999), and they have recently been used to analyse data from single-molecule experiments (Kou et al., 2005). There is a significant difference between this observation regime and the two previous regimes. To have notation consistent with (i) and (ii) we let $y_i = t_i$ denote the time of the $i$th observation and define the likelihood $f(y_i|X_{t_i-1}, X_{t_i})$ to be the probability density that the next observation after $t_{i-1}$ is at time $t_i$. This density can be obtained by integrating

$$
\nu(X_s) \exp \left\{ - \int_{t_{i-1}}^{t_i} \nu(X_s) \, ds \right\},
$$

with respect to the distribution of $(X_s, s \in (t_{i-1}, t_i))$ conditionally on $X_{t_{i-1}} = x_{t_{i-1}}$ and $X_{t_i} = x_i$. The distribution of this conditioned process has a known density with respect to the Brownian bridge measure and it is given in lemma 1 of Beskos et al. (2006b). We can thus show that the density of interest is

$$
\nu(x_i) \frac{N_{t_i-t_{i-1}}(x_{t_i} - x_{t_{i-1}})}{p_{t_i-t_{i-1}}(x_{t_i}|x_{t_{i-1}})} \exp \left\{ A(x_{t_i}) - A(x_{t_{i-1}}) \right\} \left[ \exp \left\{ - \int_{t_{i-1}}^{t_i} \{ \phi(W_s) + \nu(W_s) \} \, ds \right\} \right],
$$

where the expectation is with respect to the law of a Brownian bridge from $x_{t_{i-1}}$ to $x_{t_i}$.

We take a Bayesian approach and assume a prior distribution for $X_0$. Our interest lies in the on-line calculation of the filtering densities: the posterior densities of the signal at time $t_i$ given the observations up to time $t_i$, for each $1 \leq i \leq n$. Although these densities are intractable, we propose a particle filter scheme to estimate recursively these densities at each observation time point. As we point out in Section 6, our approach allows the estimation of the filtering distribution of the continuous time path $(X_s, t_{i-1} < s < t_i)$.

A more flexible model for the signal is a diffusion process $Z$ which solves a more general SDE than the equation that we have assumed in process (1):

$$
dZ_s = b(Z_s) \, ds + \Sigma(Z_s) \, dB_s, \quad s \in [0, t].
$$

In contrast with process (1), process (5) allows the diffusion coefficient to be state dependent. Our methods directly apply to all such processes provided that there is an explicit transformation $Z_s \mapsto \eta(Z_s) = X_s$, where $X$ solves an SDE of the type (1); the implied drift $\alpha$ can be easily expressed in terms of $b$ and $\Sigma$ via Itô's formula and it will have to satisfy the conditions that we have already specified. In model (i) the likelihood becomes $f(y_i|\eta^{-1}(X_{t_i}))$, in model (ii) the data are $y_i = \zeta(\eta^{-1}(X_{t_i}))$ and in model (iii) the Poisson intensity is $\nu(\eta^{-1}(X_{t_i}))$, where $\eta^{-1}$ denotes the inverse transformation. Therefore, the extension of our methodology to general diffusions is straightforward when $d = 1$; under mild conditions process (5) can be transformed to process (1) by $\eta(Z_s) = \int_0^s \Sigma(z)^{-1} \, dz$, for some arbitrary $u^*$ in the state space of the diffusion. Moreover, the drift of the transformed process will typically satisfy the three conditions that we have specified. However, the extension is more difficult in higher dimensions. The necessary transformation is more complicated when $d > 1$ and it might be intractable or even impossible (Aït-Sahalia, 2004). Even when such a transformation is explicit it might imply a drift for $X$.
which violates condition (b). Nevertheless, many physical systems can be successfully modelled with diffusions which can be transformed to process (1).

Our particle filtering methods will be illustrated on two sets of simulated data, as follows.

2.1. Example 1: sine diffusion observed with error

The signal satisfies

\[ dX_s = \sin(X_s) \, ds + dB_s, \]  

and the data consist of noisy observations, \( y_i \sim N(X_{t_i}, \sigma^2) \). Fig. 1(a) in Section 5.1 shows a simulation of this model with \( \sigma = 0.2 \). In this case

\[ \phi(u) = \{\sin(u)^2 + \cos(u) + 1\}/2. \]  

This process is closely related to Brownian motion on a circle. It is convenient as an illustrative example since discrete time skeletons can be easily simulated from this process by using the most basic form of the EA (EA1 in Beskos et al. (2006a); R code is available on request from the authors).

2.2. Example 2: Ornstein–Uhlenbeck-driven Cox process

The second data set consists of the arrival times of a Poisson process, \( y_i = t_i \), whose intensity is given by \( \nu(X_s), s \geq 0 \), where

\[ \nu(x) = a + \beta|x|, \]

and \( X \) is an Ornstein–Uhlenbeck (OU) process,

\[ dX_s = -\rho X_s \, ds + dB_s. \]

The OU process is stationary with Gaussian marginal distribution, \( N(0, 1/2\rho) \). Thus, an interpretation for this model is that the excursions of \( X \) increase the Poisson intensity, whereas \( a \) corresponds to the intensity when \( X \) is at its mean level. An example data set is shown in Fig. 3 in Section 5.2, where we have taken \( a = 0, \beta = 20 \) and \( \rho = \frac{1}{2} \). Although the transition density of the OU process is well known,

\[ X_t | X_0 = x_0 \sim N\left[\exp(-\rho t)x_0, \frac{1}{2\rho}\left\{1 - \exp(-2\rho t)\right\}\right], \]

the observation density \( f(y_{i+1}|x_{t_i}, x_{t_{i+1}}) \) is intractable.

Examples 1 and 2 are examples of observation regimes (i) and (iii) respectively. We shall show that observation regime (ii) can be handled in a similar fashion to that of regime (i), so we have not included an accompanying example.

3. Random-weight particle filter

As in Section 2 we shall denote the observation at time \( t_i \) by \( y_i \), and \( p_{t_i}(\cdot|\cdot) \) will denote the system transition density over time \( t \) (see equation (2)). We shall write \( \Delta_i = t_{i+1} - t_i \), and the filtering densities \( p(x_{t_i}|y_{1:i}) \) will be denoted by \( \pi_i(x_{t_i}) \), where by standard convention \( y_{1:i} = (y_1, \ldots, y_i) \). To simplify the notation, when we introduce weighted particles below, we shall subscript both particles and weights by \( i \) rather than \( t_i \).
Our aim is to calculate the filtering densities \( \pi_i(x_t) \) recursively. Basic probability calculations yield the following standard filtering recursion for these densities:

\[
\pi_{i+1}(x_{t+1}) \propto \int f(y_{i+1}|x_t, x_{t+1}) \ p_{\Delta_i}(x_{t+1}|x_t) \ \pi_i(x_t) \ dx_t.
\]  

(8)

Particle filters approximate \( \pi_i(x_t) \) by a discrete distribution, denoted by \( \tilde{\pi}_i(x_t) \), whose support is a set of \( N \) particles, \( \{x_t^{(j)}\}_{j=1}^N \), with associated probability weight \( \{w_t^{(j)}\}_{j=1}^N \). Substituting \( \tilde{\pi}_i(x_t) \) for \( \pi_i(x_t) \) in expression (8) yields a (continuous density) approximation to \( \pi_{i+1}(x_{t+1}) \),

\[
\tilde{\pi}_{i+1}(x_{t+1}) \propto \sum_{j=1}^N w_t^{(j)} f(y_{i+1}|x_t^{(j)}, x_{t+1}) \ p_{\Delta_i}(x_{t+1}|x_t^{(j)}).
\]  

(9)

The aim of one iteration of the particle filter algorithm is to construct a further particle (discrete distribution) approximation to \( \tilde{\pi}_{i+1}(x_{t+1}) \).

We can obtain such a particle approximation via importance sampling, and a general framework for achieving this is given by the auxiliary particle filter of Pitt and Shephard (1999). We choose a proposal density of the form

\[
\sum_{j=1}^N \beta_t^{(j)} q(x_{t+1}|x_t^{(j)}, y_{i+1}).
\]  

(10)

The choice of suitable proposals, i.e. the choice of the \( \beta_t^{(j)} \)s and \( q \), is discussed in the analysis of our specific applications in Section 5.

To simulate a new particle at time \( t_{i+1} \) we

(a) simulate a particle \( x_t^{(k)} \) at time \( t_i \), where \( k \) is a realization of a discrete random variable which takes the value \( j \in \{1, 2, \ldots, N\} \) with probability \( \beta_t^{(j)} \), and

(b) simulate a new particle at time \( t_{i+1} \) from \( q(x_{t+1}|x_t^{(k)}, y_{i+1}) \).

The weight that is assigned to this pair of particles \( (x_t^{(k)}, x_{t+1}) \) is proportional to

\[
\frac{w_t^{(k)} f(y_{i+1}|x_t^{(k)}, x_{t+1}) \ p_{\Delta_i}(x_{t+1}|x_t^{(k)})}{\beta_t^{(k)} q(x_{t+1}|x_t^{(k)}, y_{i+1})}.
\]  

(11)

This is repeated \( N \) times to produce the set of weighted particles at time \( t_{i+1} \), \( \{(x_t^{(j)}, w_t^{(j)})\}_{j=1}^N \), which gives an importance sampling approximation to \( \pi_{i+1}(x_{t+1}) \). Renormalizing the weights is possible but does not materially affect the methodology or its accuracy. Improvements on independent sampling in step (a) can be made: see the stratified sampling ideas of Carpenter et al. (1999). The resulting particle filter has good theoretical properties including consistency (Crisan, 2001) and CLTs for estimates of posterior moments (Del Moral and Miclo, 2000; Chopin, 2004; Künsch, 2005), as \( N \to \infty \). Under conditions relating to exponential forgetting of initial conditions, particle filter errors stabilize as \( n \to \infty \) (Del Moral and Guionnet, 2001; Künsch, 2005).

The difficulty with implementing such a particle filter when the signal \( X \) is a diffusion process is that the transition density \( p_{\Delta_i}(x_{t+1}|x_t^{(k)}) \) which appears in expression (11) is intractable for most diffusions of interest, owing to the expectation term in equation (2). Furthermore, for observation model (iii) (but also for more general models), the likelihood term \( f(y_{i+1}|x_t^{(k)}, x_{t+1}) \) that is given in expression (4) cannot be calculated analytically.

We circumvent these problems by assigning each new particle a random weight which is a realization of a random variable whose mean is expression (11). The construction and simulation
of this random variable is developed in Section 4, and it is based on the particular expression for the transition density in equation (2). The replacement of the weights by positive unbiased estimators is an interesting possibility in more general contexts than that considered in this paper. Indeed, in Section 3.2 we show that this approach amounts to a convenient augmentation of the state with auxiliary variables.

3.1. Simulation of weights
In all models the weight that is associated with the pair \((x_i^{(k)}, x_{t+1})\) equals

\[
h_{i+1}(x_i^{(k)}, x_{t+1}, y_{i+1}) = \frac{w_i^{(k)} f(y_{i+1}|x_{t+1}) N_{\Delta t}(x_{t+1}^{(k)} - x_i^{(k)}) \exp\{A(x_{t+1}) - A(x_i^{(k)})\}}{\beta_i^{(k)} q(x_{t+1}^{(k)} | x_i^{(k)}, y_{i+1})},
\]

where \(h_{i+1}\) is a known function and, for \(0 < u < t\),

\[
\mu_g(x, z, u, t) := \mathbb{E} \left[ \exp \left\{ -\int_u^t g(W_s) \, ds \right\} \right],
\]

where the expectation is taken with respect to a \(d\)-dimensional Brownian bridge \(W\), starting at time \(u\) from \(W_u = x\) and finishing at time \(t\) at \(W_t = z\).

3.1.1. Models (i) and (ii)
For model types (i) and (ii)

\[
h_{i+1}(x_i^{(k)}, x_{t+1}, y_{i+1}) = \frac{w_i^{(k)} \nu(x_{t+1}) N_{\Delta t}(x_{t+1}^{(k)} - x_i^{(k)}) \exp\{A(x_{t+1}) - A(x_i^{(k)})\}}{\beta_i^{(k)} q(x_{t+1}^{(k)} | x_i^{(k)}, y_{i+1})},
\]

and \(g = \phi\). In model type (ii) the proposal distribution \(q(x_{t+1}^{(k)} | x_i^{(k)}, y_{i+1})\) should be chosen to propose only values of \(x_{t+1}\) such that \(\zeta(x_{t+1}) = y_{i+1}\); then \(f(y_{i+1}|x_{t+1}) = 1\).

3.1.2. Model (iii)
A synthesis of expressions (2), (4) and (11), with \(g = \phi + \nu\), gives

\[
h_{i+1}(x_i^{(k)}, x_{t+1}, y_{i+1}) = \frac{w_i^{(k)} \nu(x_{t+1}) N_{\Delta t}(x_{t+1}^{(k)} - x_i^{(k)}) \exp\{A(x_{t+1}) - A(x_i^{(k)})\}}{\beta_i^{(k)} q(x_{t+1}^{(k)} | x_i^{(k)}, y_{i+1})}.
\]

Section 4 shows how to construct for each pair of \((x, z)\) and times \((u, t)\), with \(u < t\), additional auxiliary variables \(V\), and a function \(r(y, x, z, u, t) \geq 0\), with the property that \(\mathbb{E}[r(V, x, z, u, t)|x, z] = \mu_g(x, z, u, t)\). The auxiliary variables are simulated according to an appropriate conditional distribution \(Q_g(.|x, z, u, t)\), and \(r\) is easy to evaluate. Our method replaces in the weight the intractable term \(\mu_g\) with its unbiased estimator \(r\).

3.1.3. Random-weight particle filter

Step 1: simulate a sample \(x_0^{(1)}, \ldots, x_0^{(N)}\) from \(p(x_0)\), and set \(w_0^{(j)} = 1/N\).

For \(i = 0, \ldots, n - 1\) and for \(j = 1, \ldots, N\) perform the following steps.

Step 2: calculate the effective sample size of the \(\{\beta_i^{(k)}\}\), \(\text{ESS} = \{\Sigma_{k=1}^N (\beta_i^{(k)})^2\}^{-1}\); if \(\text{ESS} < C\), for some fixed constant \(C\), simulate \(k_{i,j}\) from \(p(k) = \beta_i^{(k)}\), \(k = 1, \ldots, N\), and set \(\delta_i^{(j)} = 1\); otherwise set \(k_i, j = j\) and \(\delta_i^{(j)} = \beta_i^{(j)}\).

Step 3: simulate \(x_i^{(j)}\) from \(q(x_{t+1}^{(k)} | x_i^{(k)}, y_{i+1})\).
Step 4: simulate $v_{i+1} \sim Q_g(\cdot | x_i^{(k_i,)}, x_{i+1}, t_i, t_{i+1})$.

Step 5: assign particle $x_{i+1}^{(j)}$ a weight

$$w_{i+1}^{(j)} = \delta_{i+1}^{(j)} h_{i+1}(x_i^{(k_i,)}, x_{i+1}, y_{i+1}) r(v_{i+1}, x_i^{(k_i,)}, x_{i+1}, t_i, t_{i+1}).$$ (13)

Note that this algorithm contains a decision about whether or not to resample particles before propagation in step 2, with the decision being based on the effective sample size of the $\beta_i^{(j)}$. The constant $C$ can be interpreted as the minimum acceptable effective sample size. (See Liu and Chen (1998) for the rationale of basing resampling on such a condition.) Whether or not resampling occurs will affect the weight that is given to the new sets of particles, and this is accounted for by different values of $\delta_i^{(j)}$ in step 2. Optimally, the resampling for step 2 will incorporate dependence across the $N$ samples, e.g. the stratified sampling scheme of Carpenter et al. (1999) or the residual sampling of Liu and Chen (1998).

3.2. An equivalent formulation via an augmentation of the state

In the previous section we described a generic sequential Monte Carlo scheme where the exact weights in the importance sampling approximation of the filtering distributions are replaced by positive unbiased estimators. We now show that this scheme is equivalent to applying an ordinary auxiliary particle filter to a model with richer latent structure. We demonstrate this equivalent representation for model types (i) and (ii), since an obvious modification of the argument establishes the equivalence for model type (iii).

According to our construction, conditionally on $X_t, X_{t+1}, t_i$ and $t_{i+1}, V_{i+1}$ is independent of $V_j$ and $X_j$ for any $j$ that is different from $i$ or $i+1$. Additionally, it follows easily from the unbiasedness and positivity of $r$ that, conditionally on $X_t = x$, $r(v_{i+1}, x, x_{i+1}, t_i, t_{i+1})$ is a probability density function for $(X_{i+1}, V_{i+1})$ with respect to the product measure Leb(dz) × $Q_g$ (dv|x, z, t_i, t_{i+1}), where Leb denotes the Lebesgue measure.

Consider now an alternative discrete time model with unobserved states $(Z_i, V_i), i = 1, \ldots, n$, $Z_i \in \mathbb{R}^d$, with a non-homogeneous Markov transition density

$$p_{i+1}(Z_{i+1}, V_{i+1} | Z_i, V_i) = r(v_{i+1}, z_i, z_{i+1}, t_i, t_{i+1})$$

(this density is with respect to Leb × $Q_g$) and observed data $y_i$ with observation density $f(y_{i+1} | z_i, z_{i+1})$. By construction the marginal filtering distributions of $Z_i$ in this model are precisely $\pi_i(x_i)$, i.e. the filtering densities in expression (8). Consider an auxiliary particle filter that is applied to this model where we choose with probability $\beta_i^{(j)}$ each of the existing particles $(z_i^{(j)}, V_i^{(j)})$, and generate new particles according to the proposal

$$(z_{i+1}, v_{i+1}) \sim q(z_{i+1} | z_i^{(k)}, v_i), Q_g(dv_{i+1} | z_i^{(k)}, z_{i+1}, t_i, t_{i+1}) \text{Leb}(dz_{i+1}),$$

where $q$ is the same proposal density as in expression (10). The weights that are associated with each particle in this discrete time model are tractable and are given by equation (13). Therefore, the weighted sample $\{(z_{i+1}^{(j)}, w_{i+1}^{(j)})\}_{j=1}^n$ is precisely a particle approximation to $\pi_{i+1}(x_{i+1})$, and the RWPF is equivalent to an auxiliary particle filter on this discrete time model whose latent structure has been augmented with the auxiliary variables $V_i$.

This equivalent representation sheds light on many aspects of our method. Firstly, it makes it obvious that it is inefficient to average more than one realization of the positive unbiased estimator of $\mu_g$ per particle. Instead it is more efficient to generate more particles with only one realization of the estimator simulated for each pair of particles.

Secondly, it illustrates that the RWPF combines the advantages of the bootstrap and the auxiliary particle filter. Although it is easy to simulate from the probability distribution $Q_g$ (as
described in Section 4), it is very difficult to derive its density (with respect to an appropriate reference measure). Since the \( V_i \)s are propagated according to this measure, its calculation is avoided. This is an appealing feature of the bootstrap filter which propagates particles without requiring analytically the system transition density. However, the propagation of the \( Z_i \)s is done via a user-specified density which incorporates the information in the data.

Thirdly, it suggests that the RWPF will have similar theoretical properties with auxiliary particle filters applied to discrete time models. This is explored in Section 3.3.

### 3.3. Theoretical properties

Consider estimation of the posterior mean of some function \( \varphi \) of the state at time \( t_i \), \( E[\varphi(x_{t_i})|y_{1:i}] \). A natural approach to the investigation of particle filter effectiveness is to consider the limiting behaviour of the algorithm as \( N \to \infty \). For the standard auxiliary particle filter, Chopin (2004) introduced a CLT for estimation of this type of expectations. This CLT applies directly to both the EPPF and the ESPF.

In Appendix F we extend the result of Chopin (2004) and give a further necessary condition on the random weights in RWPF under which a CLT still holds. This extra condition is condition 2. The expression for the variance of the estimator of \( E[\varphi(x_{t_i})|y_{1:i}] \) that is obtained with the RWPF differs from the expression in the standard case (i.e. when the weights are known) by an extra term that is caused by the randomness in the weights (see equations (27)–(29) and the comment on theorem 3 in Appendix F for further details). The ready adaptation of Chopin’s approach is facilitated by the observation that the RWPF can be re-expressed as a standard particle filter for an augmented state (see Section 3.2).

One important consequence of this CLT is that the errors in estimating \( E[\varphi(x_{t_i})|y_{1:i}] \) are of order \( N^{-1/2} \). Previous filtering methods for the diffusion problems that we consider are based on

(a) discretizing time and introducing \( M \) intermediate time points between each observation time,
(b) using an Euler, or higher order, approximation to the diffusion (1) and
(c) applying a particle, or other, filter to this approximate discrete time model.

See for example Crisan et al. (1999). Results giving the order of the errors in one such scheme were given by Del Moral et al. (2001). For models such as (i) and (ii) the errors are of order \( N^{-1/2} \) provided that the number of intermediate time steps \( M \) between each observation increases at a rate \( N^{1/2} \). Thus for fixed computational cost \( K \propto MN \) the errors decrease at a rate \( K^{-1/3} \). For models such as (iii), where the likelihood depends on the path of the state between two successive observations, the rate at which errors decrease will be slower, e.g. \( K^{-1/4} \) (Del Moral et al., 2001) or \( K^{-1/6} \) (theorem 1.1 of Crisan et al. (1999)).

### 4. Generalized Poisson estimators

We have already motivated the need for the simulation of a positive unbiased estimator of

\[
\mathbb{E}[E] \quad E = \exp \left\{ - \int_0^t g(W_s) \, ds \right\},
\]

where the expectation is taken with respect to a \( d \)-dimensional Brownian bridge \( W \). In this section we introduce a methodology for deriving such estimators and provide theoretical and simulation results regarding the variance of the estimators suggested. These results are of independent interest beyond particle filtering, so we present our methodology in a general way, where \( g \) is an arbitrary function that is assumed only to be continuous on \( \mathbb{R}^d \). We assume that
\(W_0 = x\) and \(W_t = z\), for arbitrary \(x, z \in \mathbb{R}^d\) and \(t > 0\). By the time homogeneity property of the Brownian bridge our methodology extends to the case where the integration limits change to \(u\) and \(u + t\), for any \(u > 0\).

Beskos et al. (2006b) proposed an unbiased estimator of expectation (14), the PE:

\[
\exp \{ (\lambda - c) t \} \lambda^{-\kappa} \prod_{j=1}^{\kappa} \{ c - g(W_{\psi_j}) \};
\]

(15)

\(\kappa\) is a Poisson random variable with mean \(\lambda t\), the \(\psi_j s\) are uniformly distributed on \([0, t]\), and \(c \in \mathbb{R}\) and \(\lambda > 0\) are arbitrary constants. (Here and below we assume that the empty product, i.e. when \(\kappa = 0\), takes the value 1.) The two main weaknesses of the PE are that it may return negative estimates and that its variance is not guaranteed to be finite. Both of these problems are alleviated when \(g\) is bounded. However, this is a very restrictive assumption in our context. Therefore, here, we introduce a collection of unbiased and positive estimators of \(E[U]\) (14) which generalize the PE. The methods that we consider allow \(c\) and \(\lambda\) to depend on \(W\) and permit \(\kappa\) to have a general discrete distribution. Firstly, we need to be able to simulate random variables \(L_W\) and \(U_W\) with

\[
L_W \leq g(W_s) \leq U_W, \quad \text{for all } s \in [0, t],
\]

(16)

and to be able to simulate \(W_s\) at any \(s\), given the condition that is implied by inequality (16). For unbounded \(g\) this is non-trivial. However, both of these simulations have become feasible since the introduction of an efficient algorithm in Beskos et al. (2005a). An outline of the construction is given in Appendix A.

Let \(U_W\) and \(L_W\) satisfy condition (16) and \(\psi_j, j \geq 1\), be a sequence of independent uniform random variables on \([0, t]\). Then, estimator (14) can be re-expressed as

\[
\mathbb{E} \left[ \exp(-U_W t) \exp \left( \int_0^t \{ U_W - g(W_s) \} \, ds \right) \right]
\]

\[
= \mathbb{E} \left[ \exp(-U_W t) \sum_{k=0}^{\infty} \frac{1}{k!} \left( \int_0^t \{ U_W - g(W_s) \} \, ds \right)^k \right]
\]

\[
= \mathbb{E} \left[ \exp(-U_W t) \mathbb{E} \left[ \sum_{k=0}^{\infty} \frac{k^k}{k!} \prod_{j=1}^{k} \{ U_W - g(W_{\psi_j}) \} | U_W, L_W \right] \right]
\]

\[
= \mathbb{E} \left[ \exp(-U_W t) \frac{t^\kappa}{\kappa! p(\kappa|U_W, L_W)} \prod_{j=1}^{\kappa} \{ U_W - g(W_{\psi_j}) \} \right],
\]

(17)

where \(\kappa\) is a discrete random variable with conditional probabilities \(P(\kappa = k|U_W, L_W) = p(k|U_W, L_W)\). The second equality in the above argument is obtained by using dominated convergence and Fubini’s theorem (which hold by positivity of the summands).

We can derive various estimators of \(\mathbb{E}[E](14)\) by specifying \(p(\cdot|U_W, L_W)\). The family of all such estimators will be called the GPE:

\[
\exp(-U_W t) \frac{t^\kappa}{\kappa! p(\kappa|U_W, L_W)} \prod_{j=1}^{\kappa} \{ U_W - g(W_{\psi_j}) \}.
\]

(18)

The following theorem (which is proved in Appendix B) gives the optimal choice for \(p(\cdot|U_W, L_W)\).

**Theorem 1.** The conditional second moment of the GPE given \(U_W\) and \(L_W\) is
\[
\exp(-2U_Wt) \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E} \left[ \left[ \int_0^t \{U_W - g(W_s)\}^2 \, ds \right]^k \bigg| U_W, L_W \right].
\]

If
\[
\sum_{k=0}^{\infty} \frac{k^{k/2}}{k!} \mathbb{E} \left[ \left[ \int_0^t \{U_W - g(W_s)\}^2 \, ds \right]^k \bigg| U_W, L_W \right]^{1/2} < \infty,
\]
then the second moment is minimized by the choice
\[
p(k|U_W, L_W) \propto \frac{t^{k/2}}{k!} \mathbb{E} \left[ \left[ \int_0^t \{U_W - g(W_s)\}^2 \, ds \right]^k \bigg| U_W, L_W \right]^{1/2},
\]
with minimum second moment given by
\[
\left( \exp(-U_Wt) \sum_{k=0}^{\infty} \frac{t^{k/2}}{k!} \mathbb{E} \left[ \left[ \int_0^t \{U_W - g(W_s)\}^2 \, ds \right]^k \bigg| U_W, L_W \right]^{1/2} \right)^2 < \infty,
\]
for almost all \( U_W \) and \( L_W \).

Although the right-hand side of expression (21) cannot be evaluated analytically, it can guide a suitable choice of \( p(\cdot|U_W, L_W) \). If \( W \) were known, the optimal proposal is Poisson with mean
\[
\lambda_W := \left[ t \int_0^t \{U_W - g(W_s)\}^2 \, ds \right]^{1/2}.
\]
We shall discuss two possible ways that expression (23) can be used to choose a good proposal.

A conservative approach takes \( p(\cdot|U_W, L_W) \) to be Poisson with mean \((U_W - L_W)t\) (an upper bound of \( \lambda_W \)). We call this estimator GPE-1. An advantage of GPE-1 is that its second moment is bounded above by \( \mathbb{E}[\exp(-2L_Wt)] \). Thus, under mild and explicit conditions on \( g \), which are contained in the following theorem (which is proved in Appendix C), the variance of the estimator is guaranteed to be finite.

**Theorem 2.** A sufficient condition for GPE-1 to have finite variance is that
\[
g(u_1, \ldots, u_d) \geq -\delta \sum_{i=1}^d (1 + |u_i|), \quad \text{for all } u_i \in \mathbb{R}, 1 \leq i \leq d, \delta \geq 0.
\]
Since \( \lambda_W \) is stochastic, an alternative approach is to introduce a (exogenous) random mean and to assume that \( p(\cdot|U_W, L_W) \) is Poisson with this random mean. For tractability we choose the random mean to have a gamma distribution, when \( p(\cdot|U_W, L_W) \) becomes a negative binomial distribution (estimator GPE-2):
\[
\exp(-U_Wt) \frac{t^\kappa \Gamma(\beta)(\beta + \gamma_W)^{\beta + \kappa}}{\Gamma(\beta + \kappa)\beta^{\beta + \kappa}} \prod_{j=1}^{\kappa} \{U_W - g(W_{\psi_j})\},
\]
where \( \gamma_W \) and \( \beta \) denote the mean and the dispersion parameter respectively of the negative binomial distribution. Since the negative binomial distribution has heavier tails than the PE, GPE-2 will have finite variance whenever there is a PE with finite variance. However, big gains in
efficiency can be achieved if $\gamma_W$ is chosen to be approximately $E[\lambda_W | U_W, L_W]$. There is a variety of ad hoc methods which can provide a rough estimation of this expectation. Applying Jensen's inequality to exchange the integration with the square power in equation (23), and subsequently approximating $E[g(W_s) | U_W, L_W]$ by $g(E[W_s])$, suggests taking

$$
\gamma_W = tU_W - \int_0^t g\left(\frac{t-s}{r} + \frac{y_s}{t}\right) \, ds > 0.
$$

A simulation study (part of which is presented in Section 4.1) reveals that this choice works very well in practice and GPE-2 has up to several orders of magnitude smaller variance than the PE or GPE-1. The integral can usually be easily evaluated; otherwise a crude approximation can be used.

We have confined our presentation to the case where the expectation in expression (14) is with respect to the Brownian bridge measure. Nevertheless, as pointed out in Beskos et al. (2006b) the PE can be constructed in exactly the same way when the expectation is taken with respect to an arbitrary diffusion bridge measure, as long as exact skeletons can be simulated from this measure. The GPE can also be implemented in this wider framework, provided that the process $W$ can be constructed to satisfy condition (16).

4.1. Simulation study

We consider a smooth bounded test function $g(u) = \{\sin(u)^2 + \cos(u) + 1\}/2$. This has been chosen in view of example 1. The function $g$ is periodic, with period $2\pi$. In $[0, 2\pi]$ it has local minima at 0 and $2\pi$, global minimum at $\pi$ and maxima at $\pi/3$ and $5\pi/3$. Since $g$ is bounded by $9/8$ we can construct a PE which returns positive estimates by setting $c \geq 9/8$. Under this constraint, Beskos et al. (2006b) argued that a good choice is $c = \lambda = 9/8$. Simulation experiments suggested that the performance of GPE-2 is quite robust to the choice of the dispersion parameter $\beta$. We have fixed it in our examples to $\beta = 10$. Table 1 summarizes estimates of the variance of the estimators based on $10^4$ simulated values. We see that GPE-2 can be significantly more efficient than the PE, in particular when taking into account $E[\kappa]$. In general, the performance of PE is sensitive to the choice of $c$ and $\lambda$. GPE-1 is typically less efficient than GPE-2. Table 1 also gives the value of $\text{var}(E)$ which takes significantly smaller values (by a couple of orders

| Estimator | Results for the following pairs of $(x, z)$: |
|-----------|-------------------------------------------|
|           | $x = 0, z = 0$                           | $x = 0, z = \pi$ | $x = \pi, z = \pi$ |
|---|---|---|---|
| Variance | PE | 0.202 | 0.200 | 0.027 |
|        | GPE-1 | $4.21 \times 10^{-3}$ | 0.208 | 0.034 |
|        | GPE-2 | $2.08 \times 10^{-3}$ | 0.220 | 0.033 |
|        | $\text{var}(E)$ | $3.74 \times 10^{-5}$ | $3.27 \times 10^{-3}$ | $4.72 \times 10^{-3}$ |
|        | PE | 1.118 | 1.126 | 1.121 |
|        | GPE-1 | 0.130 | 1.091 | 0.744 |
|        | GPE-2 | 0.119 | 0.329 | 0.735 |

†For comparison we give also $\text{var}(E)$. We also report an estimate of $E[\kappa]$. We consider three different pairs of starting and ending points $(x, z)$ and time increment $t = 1$. The estimates were obtained from a sample of $10^4$ realizations.
Table 2. Monte Carlo estimates based on $10^4$ realizations of the root-mean-square error divided by the true value of four estimators of $p_t(z|x)$, of expression (6) for $t=1$ and various $x$ and $z$†

| Estimator | Results for the following pairs of $(x, z)$: |
|-----------|---------------------------------------------|
|           | $x=0, z=0$ | $x=0, z=\pi$ | $x=\pi, z=\pi$ |
| PE        | 1.25       | 0.93         | 0.17           |
| GPE-2     | 0.13       | 0.78         | 0.2            |
| DG-1      | 0.5        | 0.45         | 0.3            |
| DG-5      | 0.28       | 0.19         | 0.22           |

†As true value we take the estimate that is produced by averaging the estimations given by GPE-2. The numbers of intermediate points that were used for each estimator are 1 and 5 for DG-1 and DG-5 respectively; the numbers of Brownian bridge simulations for the PE and GPE-2 are given in Table 1 ($E[\kappa]$).

of magnitude) than any of the PE, GPE-1 or GPE-2, illustrating the efficiency cost of these auxiliary variable constructions in absolute terms.

We have also investigated how the efficiency of the PE and GPE-2 varies with the time increment $t$ and in particular for small $t$ (the results are not shown). These empirical results suggest that the coefficient of variation of the errors of both the PE and GPE-2 are $O(t^\delta)$ for some $\delta > 0$, but that the value of $\delta$ differs for the two estimators. In the cases that we investigated, GPE-2 appears to have a faster rate of convergence than the PE.

The results of this simulation study have been verified for other functions $g$ (the results are not shown). We have experimented with differentiable (e.g. $g(u) = u$) and non-differentiable (e.g. $g(u) = |u|$) unbounded functions. In these cases it is impossible to design a PE which returns positive estimates with probability 1. Again, we have found that GPE-2 performs significantly better than the PE.

It is important to mention that alternative Monte Carlo methods exist which yield consistent but biased estimates of expression (14). One such estimator is obtained by replacing the time integral in expression (14) with a Riemann approximation that is based on a number, $M$ say, of intermediate points. This technique was used to construct a transition density estimator in Nicolau (2002) and effectively underlies the transition density estimator of Durham and Gallant (2002) (when the diffusion process has constant diffusion coefficient). The approach of Durham and Gallant (2002) has been used in Markov chain Monte Carlo and filtering applications (Golightly and Wilkinson, 2006; Chib et al., 2006; Ionides, 2003). In the filtering context it provides an alternative to RWPF, where the weights are approximated. It is not the purpose of this paper to carry out a careful comparison of RWPF with such variants. However, as an illustration we present a very small scale comparison in the context of estimating the transition density, $p_t(z|x)$, of expression (6) for $t=1$ and $x$ and $z$ as in Table 1. We compare four methods. Two are based on equation (2) and use the PE and GPE-2 to generate estimators of the expectation. The other two, DG-1 and DG-5, are two implementations of the Durham and Gallant (2002) estimator, with one and five intermediate points respectively. We compare the methods in terms of their root-mean-square error divided by the true value (i.e. the coefficient of variation). As the true value we used the estimate of GPE-2. The results of the comparison are presented in Table 2. Note that DG-1 and DG-5 simulate many more variables than GPE-2 to construct their estimates.
5. Comparison of particle filters on the simulated data

We now demonstrate the performance of the various particle filters that we have presented on the two examples that were introduced in Section 2.

5.1. Analysis of the sine diffusion

We first consider analysing the sine diffusion of example 1. The simulated data are shown in Fig. 1(a). We compare four implementations of the particle filter each of which avoids time discretizations by using methodology that is based on the EA for simulating diffusions:

(a) the EPPF, which uses the EA for implementing a bootstrap filter;
(b) the ESPF, which adapts the EA to simulate by rejection sampling from the filtering densities;
(c) RWPF1, an implementation of RWPF using the PE (see Table 1) to simulate the weights;
(d) RWPF2, an implementation of RWPF using GPE-2 to simulate the weights.

Details on the implementation of the EPPF and ESPF are given in Appendix D.

In this simple example the ESPF is more efficient than the EPPF, since it has the same computational cost, but it is proposing from the optimal proposal distribution. However, we have efficiently implemented the ESPF exploiting several niceties of this simple model, in particular the Gaussian likelihood and the fact that the drift is bounded. In more general models implementation of the ESPF can be considerably more difficult and its comparison with the EPPF less favourable owing to smaller acceptance probabilities.

In this context where φ is bounded we can speed up the implementation of GPE-2 with practically no loss of efficiency by replacing $U_W$ in expressions (24) and (25) by 9/8, which is the

![Fig. 1.](image-url)
upper bound of $\phi$. In this case, there is no need to simulate $U_W$ and $L_W$. We have implemented this simplification in RWPF2.

Algorithms EPPF, RWPF1 and RWPF2 used the stratified resampling algorithm of Carpenter et al. (1999), with resampling at every iteration. For RWPF1 and RWPF2 we chose the proposal distribution for the new particles on the basis of the optimal proposal distribution that is obtained if the sine diffusion is approximated by the Ozaki discretization scheme (the details are in Appendix E). For the EPPF we chose the $\beta_{i}^{(k)}$’s to be those obtained from this approximation.

The number of particles that were used in each algorithm was set so that each filter had comparable central processor unit (CPU) cost, which resulted in 500, 500, 910 and 1000 particles used for each algorithm. For these numbers of particles, the EPPF and ESPF on average required the proposal of 1360 particles and required 675 Brownian bridge simulations within the accept–reject step (c) at each iteration of the algorithm. By comparison RWPF1 and RWPF2 simulated respectively 910 and 1000 particles and required on average 1025 and 850 Brownian bridge simulations to generate the random weights at each iteration.

Note that the comparative CPU cost of the four algorithms, and in particular that of the EPPF and ESPF compared with RWPF1 and RWPF2, depends on the underlying diffusion path. The acceptance probabilities within the EPPF and ESPF depend on the values of $x_{i,j}^{k_{i,j}}$ and $x_{i+1,j}$, and grow small when both these values are close to 0 (mod $2\pi$). (In the long run the diffusion will visit these regions infrequently and will stay there for short periods.) Thus, simulated paths which spent more or less time in this region of the state space would result in the EPPF and ESPF having a larger or smaller CPU cost respectively.

We compared the four filters on the basis of the variability of estimates of the mean of the

---

**Fig. 2.** Relative efficiency of the four particle filter algorithms at estimating the filtering mean $E[X_{t_{i}}|Y_{1:t}]$ (each line gives the relative efficiency of one algorithm compared with RWPF2 (-----); see the text for details):

- , RWPF1;
- , ESPF;
- , EPPF

---
filtering distribution of the state across 500 independent runs of each filter. Results are given in Fig. 2, and output from one run of RWPF2 is shown in Fig. 1. The comparative results in Fig. 2 are for estimating the mean of the filtering distribution at each iteration (similar results were obtained for various quantiles of the filtering distribution). They show RWPF2 performing best with an average gain in efficiency of 15% over RWPF1, 50% over the ESPF and 200% over the EPPF. Interpretation of these results suggests that (for example) the ESPF would be required to run with \( N = 750 \) (taking 1.5 times the CPU cost for this data set) to obtain accuracy comparable with that of RWPF2.

Varying the parameters of the model and implementation of the algorithms will affect the relative performance of the algorithms. In particular increasing or decreasing \( \sigma^2 \), the variance of the measurement error, will increase or decrease the relative efficiency of the EPPF relative to the other filters respectively. Similar results occur as \( \Delta \) is decreased or increased respectively. The relative performance of the other three algorithms appears to be more robust to such changes. We considered implementing the EPPF with \( \beta_i^{(k)} = w_i^{(k)} \), and also using an Euler rather than an Ozaki approximation of the sine diffusion to construct the proposal distribution for RWPF1 and RWPF2, but neither of these changes had any noticeable effect on the performance of the methods. We also considered resampling less often, setting \( C = N/4 \) in step 2 of the RWPF algorithm (so resampling when the effective sample size of the \( \beta_i^{(j)} \)'s was less than \( N/4 \)) and this reduced the performance of the algorithms substantially (by a factor of 2 for RWPF1 and RWPF2).

We also investigated the effect of increasing the amount of time, \( \Delta \), between observations. To do this we used the above data taking

(a) every 10th or
(b) every 20th

time point.

To measure the performance of the filter for these different scenarios we used the effective sample size ESS of Carpenter et al. (1999). ESS is calculated on the basis of the variance of estimates of posterior means across independent runs of the filter, but this variance is compared with the posterior variance to give some measure of how many independent draws from the posterior would produce estimators of the same level of accuracy. We focus on estimates of the posterior mean of the state at observation times; and if \( \hat{s}^2 \) is the sample variance of the particle filter’s estimate of \( \mathbb{E}[X_{t_i}|y_{1:i}] \) across 100 independent runs, and \( \hat{\sigma}^2 \) is an estimate of \( \text{var}(X_{t_i}|y_{1:i}) \), then ESS is \( \hat{\sigma}^2 / \hat{s}^2 \). Note that comparing filters by their ESS is equivalent to comparing filters on the basis of the variance of the estimators.

Table 3 gives ESS-values for various values of \( \Delta \). We see that the ESS-values drop dramatically as \( \Delta \) increases, and the filter is inefficient for \( \Delta = 20 \). This drop in performance is due to the large variability of the random weights in this case. The variability of these weights is due to

(a) the variability of

\[
\exp\left\{ - \int_{t_i}^{t_{i+1}} g(W_s) \, ds \right\},
\]

across different diffusion paths, and

(b) the Monte Carlo variability in estimating this for a given path.

To evaluate what amount is due to (a), we tried a particle filter that estimates expression (26) numerically by simulating the Brownian bridge at a set of discrete time points (for this example we sampled values every half time unit) and then using these to evaluate the integral numerically.
Table 3. Comparison of the filter’s mean ESS-values for different time intervals between observations ($\Delta$)†

| Filter            | Results for the following values of $\Delta$: |
|-------------------|-----------------------------------------------|
|                   | $\Delta = 10$ | $\Delta = 20$ |
| RWPF2             | 73           | 5              |
| Discretization    | 80           | 12             |
| pseudoRWPF2       | 923          | 933            |

†Results are for the RWPF using GPE-2 (RWPF2), a filter that numerically approximates the weight through discretizing the diffusion process (discretization), and the RWPF after introducing uninformative observations at unit time intervals (pseudo-RWPF2).

This approach is closely related to the importance sampling approach of Durham and Gallant (2002) and Nicolau (2002); see Section 4.1. The results for this filter are also given in Table 3 (note that the ESS-values ignore any bias that is introduced through this numerical approximation), and we again see small ESS-values, particularly for $\Delta = 20$. This filter’s performance is very similar to the RWPF, which suggests that the Monte Carlo variability in (b) is a small contributor to the poor performance of the RWPF in this case.

Finally we tried introducing pseudo-observations at all integer time intervals where currently no observation is made. The RWPF is then run as above, but with no likelihood contribution to the weight at the time points where there are these uninformative observations. The idea is that now $\Delta = 1$, so that the variance of the random weights is well behaved, but we still have adaptation of the path of the diffusion in the unit time interval prior to an observation to take account of the information in that observation. Results are again shown in Table 3, and the ESS-values are very high (and close to the optimal value, that of the number of particles, 1000). The computational cost is only roughly doubled by adding these extra pseudo-observations, as the total computational cost for the simulation of the Brownian bridge is unchanged. These results are reasonably robust to the choice of how frequently to introduce these uninformative observations (the results are not shown).

5.2. Analysis of the Cox process

We now consider applying the RWPF to example 2 from Section 2, the OU-driven Cox process. The data that we analysed are given in Fig. 3(a). It is either impossible or difficult to adapt the other two EA-based particle filters (the EPPF and the ESPF) to this problem. For instance we cannot implement the EPPF as the likelihood function is not tractable. So we just focus on the efficiency of the RWPF in estimating the filtering distribution of $|X_t|$.

Our implementation of the RWPF was based on proposing particles from the prior distribution, so $\beta_i^{(k)} = w_i^{(k)}$ and $q(x_{t+1}^j | x_t^j, y_{t+1})$ is just the OU transition density $p(x_{t+1}^j | x_t^j)$. We simulated the random weights by GPE-2. We calculated the filtering density at each observation time, and also at 56 pseudo-observation times that were chosen so that the maximum time difference between two consecutive times for which we calculated the filtering density was 0.1. This was
Fig. 3. (a) Simulation from the Cox process of example 2 and results from analysis by the RWPF (---, path of the absolute of the underlying diffusion; *, observed arrival times; o, filter estimates from the RWPF; ——, 90% credible interval for the absolute of the diffusion (although for clarity they are shown for all times, the credible intervals were calculated only at and apply for times where filter estimates are shown)) and (b) ESS of the RWPFs weights, defined as \( \sum_{j=1}^{N} w_{ij}^2 / \sum_{j=1}^{N} w_{ij} \), over time (the dramatic increases in the ESSs correspond to resampling times)

necessary to avoid the number of Brownian bridge simulations that were required to simulate the weights being too large for long interobservation times, and also to control the variance of the random weights (see above). The likelihood function for these non-observation times is obtained by removing \( \nu(x_{ti}) \) from expression (4).

We set the number of particles to 1000 and resampled when the ESS of the \( \beta_j \)s was less than 100 (\( C = N/10 \) in step 2 of the algorithm in Section 3). Although results for the sine diffusion suggest that this will result in an algorithm that resamples too infrequently, we chose to have a low threshold so that we could monitor the performance of the particle filter by how the ESS of the particle filter weights decay over time. The results of one run of this filter are shown in Fig. 3(a). The computational efficiency of this method can be gauged by Fig. 3(b) where the ESS of the \( w_{ij} \)s is plotted over time.

6. Discussion

We have described how recent methods for the exact simulation of diffusions and the unbiased estimation of diffusion exponential functionals can be used within particle filters, so that the resulting particle filters avoid the need for time discretization. Among the approaches that we have introduced special attention was given to the RWPF which implements an auxiliary particle filter but simulates the weights that are allocated to each particle. We showed that this methodology is equivalent to an auxiliary particle filter applied to an appropriately expanded model. We expect that this methodology will have interesting applications to models that are different from those considered in this paper, which, however, involve intractable dynamics or likelihoods.
We have focused on the filtering problem, estimating the current state given observations to date. However, extensions to prediction are trivial—merely requiring the ability to simulate from the state equation, which is possible via the EA algorithms. It is also straightforward to use the idea of Kitagawa (1996), where each particle stores the history of its trajectory, to obtain approximations of the smoothing density (the density of the state at some time in the past given the observations to date).

Although particles store values of the state only for each observation time, it is straightforward to fill in the diffusion paths between these times to produce inferences about the state at any time. A particle approximation to the distribution of \((X_t, t_i-1 < s < t_i)\) conditionally on the data \(y_{1:i}\) can be constructed by using the current set of weighted particles \(\{(x_{i-1}^{(j)}, x_i^{(j)})\}_{j=1}^N\) with weights \(\{w_i^{(j)}\}\), as follows. Firstly we need to introduce some notation; we denote by \(x_{i-1}^{(j)}\) the value of the particle at time \(t_{i-1}\) from which the \(j\)th particle at time \(t_i\) is descended. The particle approximation is given by a set of weighted paths \(\{(x_s, t_{i-1} < s < t_i)^{(j)}\}_{j=1}^N\) with weights \(\{w_i^{(j)}\}\). Each path is a diffusion bridge starting from \(x_{i-1}^{(j)}\) and finishing at \(x_i^{(j)}\) and it can be simulated by using EA, as described in Beskos et al. (2006a, 2005a). In observation regimes (i) and (ii) the EA is applied to simulate a diffusion bridge with density with respect to the Brownian bridge measure given by \(\exp\{-\int_{t_{i-1}}^{t_i} \phi(X_s) \, ds\}\), whereas in regime (iii) the corresponding density is \(\exp\{-\int_{t_{i-1}}^{t_i} \{\phi(X_s) + \nu(X_s)\} \, ds\}\). This representation can be directly exploited to draw inferences for any function of a finite skeleton of \(X\) in between observation times.

Acknowledgements

The authors acknowledge helpful comments from the Joint Editor and referees.

Appendix A: The layered Brownian motion

The algorithm that was proposed in Beskos et al. (2005a) starts by creating a partition of the sample space of \(W\) for the given \(W_0 = x\) and \(W_i = y\). Writing \(x = (x_1, \ldots, x_d)\), for a user-specified constant \(a > \sqrt{\kappa/3}\), a sequence of subsets of \(\mathbb{R}^d\) is formed as \(A_j = \{u = (u_1, \ldots, u_d) : \min(x_{i_1}, y_{i_1}) - ja < u_i \leq \max(x_{i_1}, y_{i_1}) + ja, j \geq 0\}\). This sequence defines a partition of the sample space of the form \(\cup_{j=1}^\infty D_j\), where a path belongs to \(D_j\) if and only if the path has exceeded the bounds that are determined by \(A_{j-1}\) but not the bounds that are determined by \(A_j\). In Beskos et al. (2005a) it was shown how to simulate the random variable which determines which of the \(D_j\)’s \(W\) belongs to, and how to simulate \(W\) at any collection of times conditional on this random variable: the layered Brownian bridge construction. Since \(g\) is assumed continuous, knowing \(W \in D_j\) is sufficient to determine \(U_W\) and \(L_W\) which satisfy condition (16). In fact, in the simplified setting where \(g\) is bounded, as in the sine diffusion of example 1, the layered Brownian bridge construction can be avoided since it is easy to choose \(U_W\) and \(L_W\) independently of \(W\).

Appendix B: Proof of theorem 1

\[ I := \frac{t}{\kappa!} \prod_{j=1}^\infty \{U_W - g(W_i)\}. \]

Then, expression (19) is established as follows:

\[
\mathbb{E}[I^2 | U_W, L_W] = \mathbb{E}[\mathbb{E}[I^2 | \kappa, W]] = \mathbb{E}\left[ \frac{t^{2\kappa}}{\{\kappa! \, p(k|U_W, L_W)\}^2} \left[ \int_0^t \left\{ U_W - g(W_i) \right\}^2 ds \right]^{\kappa} \right]
\]

\[
= \mathbb{E}\left[ \frac{t^{\kappa}}{\{\kappa! \, p(k|U_W, L_W)\}^2} \mathbb{E}\left[ \left[ \int_0^t \left\{ U_W - g(W_i) \right\}^2 ds \right]^{\kappa} | U_W, L_W, \kappa \right] \right]
\]
\[
\sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E} \left[ \int_0^t (U_W - g(W_s))^2 \, ds \right] |U_W, L_W|.
\]

Fubini's theorem and dominated convergence are used above (which are valid since the integrands are positive almost surely). Expression (22) is obtained by using the following result (which can be easily proved by using Jensen's inequality). Let \( f_i > 0 \) for \( i = 1, 2, \ldots \). Then the sequence of \( p_i \), which minimize \( \sum_{i=0}^{\infty} f_i / p_i \) under the constraint \( \sum p_i = 1 \) is given by \( p_i = \sqrt{f_i / \sum \sqrt{f}_i} \).

Appendix C: Proof of theorem 2

The GPE-1 estimator is less than or equal to \( \exp(-L_W t) \) so the result holds if \( \mathbb{E}[\exp(-L_W t)] < \infty \), where the expectation is with respect to a \( d \)-dimensional Brownian bridge from \( x \) at time 0 to \( y \) at time \( t \). However,

\[
\mathbb{E}[\exp(-L_W t)] = \int_0^\infty P\{\exp(-L_W t) > w\} \, dw
\]

\[
= \int_0^\infty P\{L_W < -\log(w)\} \, dw \leq \int_0^\infty P\left\{ \delta \sum_{i=1}^d (1 + M_i) > \log(w) \right\} \, dw
\]

where \( M_i = \sup_{0 \leq s \leq t} |W_s| \) by using the growth bound in theorem 2. Furthermore,

\[
\int_0^\infty P\left\{ \delta \sum_{i=1}^d (1 + M_i) > \log(w) \right\} \, dw \leq \int_0^\infty \sum_{i=1}^d P\{\delta (1 + M_i) > d^{-1} \log(w)\} \, dw
\]

\[
= \int_0^\infty \sum_{i=1}^d P\{M_i > (d\delta)^{-1} \log(w) - 1\} \, dw.
\]

It remains therefore to bound the \( d \) integrals on the right-hand side of this expression. However, from the Bachelier–Levy formula for hitting times for Brownian motion and bridges,

\[
P(M_i > v) \leq \exp[-2(v - \max\{x_i, y_i\})^2 / t] + \exp[-2(\min\{x_i, y_i + v\})^2 / t]
\]

and so

\[
P\{M_i > (d\delta)^{-1} \log(w) - 1\} \leq \exp[-2((d\delta)^{-1} \log(w) - 1) - \max\{x_i, y_i\})^2 / t] + \exp[-2[\min\{x_i, y_i
\]

\[
+ (d\delta)^{-1} \log(w) - 1\}]^2 / t
\]

which recedes like \( w^{-d \log(w)} \) as \( w \to \infty \), thus concluding the proof.

Appendix D: Exact propagation particle filter and exact simulation particle filter for example 1

The EPPF generates the new particles according to the following procedure:

(a) choose one of the current particles \( x^{(k_i, i)}_t \), where particle \( j \) is chosen with probability \( \beta_j^{(i)} \);
(b) propose \( x_{t+1} \) from the normal distribution with mean \( x^{(k_{i+1}, i)}_t \), and variance \( \Delta_i \);
(c) accept this proposal with probability \( \exp\{-\cos(x_{t+1}) - 1\} \); if the proposal is rejected return to (a);
(d) accept this proposal with probability

\[
\mathbb{E} \left[ \exp\left\{ -\int_0^{\Delta_i} \phi(W_s) \, ds \right\} \right],
\]

where the expectation is with respect to the law of a Brownian bridge from \( W_0 = x^{(k_{i+1}, i)}_t \) and \( W_{\Delta_i} = x_{t+1} \), and \( \phi \) is given in equation (7). If the proposal is rejected return to (a); otherwise \( x_{t+1} \) is the new particle at time \( t_{i+1} \) with weight \( w_{t+1} = f(y_{t+1} | x_{t+i+1}) \).

Step (d) is performed by using retrospective sampling as described in Beskos et al. (2006a). The ESPF proceeds as above but with steps (a) and (b) replaced by the step

(a') propose \( (x^{(k_{i+1}, i)}, x_{t+1}) \) according to the density proportional to
and denote by \( E \). Also let \( \tilde{E} \) is based on a first-order Taylor series expansion of the drift about some value \( \theta \). Consider a function \( \Phi \) to be the set of measurable functions \( \Phi \) and we assume independent and identically distributed sampling of \( \tau = \sigma^2 \Delta_i / (\sigma^2 + \Delta_i) \).

The algorithm is repeated until \( N \) values for \( x_{t_i} \) have been accepted, each with weight \( 1/N \).

**Appendix E: Proposal distribution for example 1**

Consider a diffusion satisfying SDE (1), with \( d = 1 \) for simplicity. The Ozaki approximation of this SDE is based on a first-order Taylor series expansion of the drift about some value \( x \). For the sine diffusion of example 1, we obtain the approximating SDE

\[
d\tilde{X}_t = -\cos(x) \{ x - \tan(x) \} ds + dB_t.
\]

So \( \tilde{X}_t = \{ x - \tan(x) \} \) is an OU process as defined in example 2 with \( \rho = \cos(x) \) and \( \sigma = 1 \). To calculate \( q(x_{t_i}^1 | x_{t_i}^j, y_{t_i+1}) \) we compute the product of the transition density that is given by the Ozaki approximation about \( x = x_{t_i}^j \) and the likelihood function \( f(y_{t_i+1} | x_{t_i+1}) \). Defining

\[
\tau^2 = [1 - \exp \{ -2 \cos(x_j^j) \Delta_i \}] / 2 \cos(x_j^j)
\]

and

\[
\eta = x_j^j - \tan(x_j^j) [1 - \exp \{ -2 \cos(x_j^j) \Delta_i \}]
\]

we obtain that \( q(x_{t_i+1}^1 | x_{t_i}^j, y_{t_i+1}) \) is normal with mean \( (\eta \sigma^2 + y_{t_i+1} \tau^2) / \tau^2 \sigma^2 \) and variance \( \eta^2 \tau^2 / (\eta^2 + \tau^2) \). Furthermore, we calculate \( \beta_j^j \sim w_j^j N_{\tau^2 \sigma^2}(y_{t_i+1} - \eta) \).

**Appendix F: Central limit theorem**

For notational simplicity, we consider a special case of our particle filter, which is chosen to resemble those which were considered in Chopin (2004). We choose our proposal density for time \( t_i \) to have \( \beta_i = w_i^j \), and we assume independent and identically distributed sampling of \( X_{t_i}^j \) in step 2. The particle filter of Chopin (2004) splits up simulating particles at time \( t_i \) into

(a) a resampling of particles at time \( t_i \) and
(b) a propagation of each of these particles to time \( t_{i+1} \).

Our assumption of independent and identically distributed sampling is equivalent to the multinomial resampling case of Chopin (2004). (The conditions for the CLT are the same if the residual sampling methods of Liu and Chen (1998), but the variances differ.) For simplicity we consider observation model (i) or (ii), though the result extends easily to observation model (iii).

Let \( \theta_i^j = (x_i^j, k_{t_i+1}^j) \), where \( k_{t_i+1}^j \) is the index that is sampled in step 2 when simulating the \( j \)-particle at time \( t_i \) and \( \theta_i^j \) is the \( j \)-th particle at time \( t_i \) together with the particle at time \( t_i \) from which it is descended. Also let \( E_{\theta_i} \) denote conditional expectation given \( \theta_i \). Similarly, let \( \mu_i(\theta_i) = \mu_{\theta_i}(x_{t_i-1}, x_t, t_{i-1}, t_i) \), and denote by \( R_i \) the unbiased estimator of \( \mu_i(\theta_i) \), i.e. \( E[R_i] = \mu_i(\theta_i) \). An important quantity is \( \sigma_i^2(\theta_i) = \text{var}(R_i) \).

We define \( \tilde{E}_i[\varphi] \) and \( \text{var}_i(\varphi) \) to be the posterior mean and variance of an arbitrary function \( \varphi(\theta) \) at time \( i \), and consider particle filter estimates of \( \tilde{E}_i[\varphi] \). Let \( \tilde{\pi}_i(\theta_i) \) be the density \( p(x_{t_i} | y_{t_i-1}) q(x_t | x_{t_i}) \). Finally define \( E_{\theta_i}[\varphi] \) and \( \text{var}_i(\varphi) \) to be shorthand for the conditional expectation and variance of \( \varphi(\theta_i) \) with respect to \( q(x_{t_i} | x_{t_i-1}) \) (which are functions of \( x_{t_i-1} \)). We denote \( \| \| \) to be the Euclidean norm and define recursively \( \Phi \) to be the set of measurable functions \( \varphi \) such that, for some \( \delta > 0 \), \( E_{\theta_i}[\| h \varphi(\theta_i) \|^{2+\delta}] < \infty \), and that the function \( x_{t_i-1} \mapsto E_{\theta_i}[h \varphi(\theta_i)] \) is in \( \Phi_i \).

**Theorem 3.** Consider a function \( \varphi \); define \( \tilde{V}_0 = \text{var}_0(\varphi) \) and, by induction,

\[
\tilde{V}_i(\varphi) = \tilde{V}_{i-1}(\tilde{E}_i[\varphi]) + \text{var}_{i-1}(\text{var}_i(\varphi)), \quad \text{for } i > 0,
\]

\[
V_i(\varphi) = \frac{\tilde{V}_i \{ \mu_i h_i(\varphi - \tilde{E}_i[\varphi]) \} + E_{\theta_i}[(\varphi - \tilde{E}_i[\varphi])^2 \sigma_i^2 h_i^2]}{E_{\theta_i}[\mu_i h_i]^2}, \quad \text{for } i \geq 0,
\]

and the result extends easily to observation model (iii).
Then if, for all $i$, 
(a) $x_i \mapsto 1$ belongs to $\Phi_i$, 
(b) $\mathbb{E}[h_i^2 \sigma_i^2] < \infty$ and 
(c) $\mathbb{E}[\sigma_i \varphi h_i]^{2+\delta} < \infty$ for some $\delta > 0$

for any $\varphi \in \Phi_i$, $\mathbb{E}[\varphi]$ and $V_i(\varphi)$ are finite and we have the following convergence in distribution as the number of particles, $N$, tends to $\infty$: 

$$
N^{1/2} \left\{ \sum_{j=1}^{N} w_j^{(i)} \varphi(x_i^{(j)}) - \mathbb{E}[\varphi] \right\} \to N\{0, V_i(\varphi)\}.
$$

Remark 1. Equations (27)–(29) refer to the changes in variance of the weighted particles due to the propagation, weighting and resampling stages at iteration $i$. Only equation (28) differs from the respective result in Chopin (2004), and this is due to the second term on the right-hand side, which represents the increase in variance due to the randomness of the weights. Condition (a) is taken from Chopin (2004) and applies to standard particle filters; conditions (b) and (c) are new and are conditions bounding the variance of the random weights which ensure that $V_i(\varphi)$ is finite.

Proof. We adapt the induction proof in Chopin (2004), considering in turn the propagation, weighting and resampling steps. Our filter differs from the standard particle filter only in terms of the weighting step, and therefore we need only to adapt the result of lemma A2 in Chopin (2004). In fact, equations (27) and (29) are identical to the corresponding quantities in Chopin (2004); therefore it remains to show equation (28). We define the constant $K = \mathbb{E}[R_i h_i]$ and $\mathbb{E}[\varphi] = R_i h_i(\varphi - E_i[\varphi]) / K$. Within the enlarged signal space framework, we can apply equation (4) of Chopin (2004), to give 

$$
V_i(\varphi) = \bar{V}_i(\varphi^*) = \bar{V}_{i-1}\{\mathbb{E}_\theta[\varphi^*]\} + \mathbb{E}_\theta[\text{var}_\theta(\varphi^*)].
$$

Now we can calculate $\mathbb{E}_\theta[\varphi^*]$ by first taking expectation over the auxiliary variables (conditional on $\theta_i$). This gives $\mathbb{E}_\theta[\varphi^*] = \mathbb{E}_\theta[\mu_i h_i(\varphi - E_i[\varphi]) / K]$. Similarly we obtain 

$$
\text{var}_\theta(\varphi^*) = \text{var}_\theta\{\mathbb{E}[R_i h_i(\varphi - E_i[\varphi]) / K]\} + \mathbb{E}_\theta[\text{var}\{R_i h_i(\varphi - E_i[\varphi]) / K\}]
$$

$$
= \text{var}_\theta\{\mu_i h_i(\varphi - E_i[\varphi]) / K\} + \mathbb{E}_\theta[\sigma_i^2 h_i^2 \{\varphi - E_i[\varphi]\}^2 / K^2].
$$

(Here the expectation and variance in equation (30) are with respect to the auxiliary variables.) Combining these results gives equation (28). The regularity conditions (a)–(c) translate directly also.

References

Aït-Sahalia, Y. (2004) Closed-form likelihood expansions for multivariate diffusions. Working Paper. Bendheim Center for Finance, Princeton University, Princeton. (Available from \url{http://www.princeton.edu/~yacine/research.htm}.)
Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005a) A new factorisation of diffusion measure and finite sample path contractions. To be published.
Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005b) Monte Carlo maximum likelihood estimation for discretely observed diffusion process. To be published.
Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006a) Retrospective exact simulation of diffusion sample paths with applications. Bernoulli, \textbf{12}, 1077–1098.
Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006b) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). J. R. Statist. Soc. B, \textbf{68}, 333–382.
Beskos, A. and Roberts, G. O. (2005) Exact simulation of diffusions. Ann. Appl. Probab., \textbf{15}, 2422–2444.
Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for non-linear problems. IEE Proc. Radar, Sonar Navig., \textbf{146}, 2–7.
Chib, S., Pitt, M. K. and Shephard, N. (2006) Likelihood based inference for diffusion driven state space models. Working Paper. Nuffield College, Oxford. (Available from \url{www.nuff.ox.ac.uk/users/shephard/}.)
Chopin, N. (2004) Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32, 2385–2411.

Crisan, D. (2001) Particle filters—a theoretical perspective. In *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, N. de Freitas and N. Gordon), pp. 17–41. New York: Springer.

Crisan, D., Del Moral, P. and Lyons, T. J. (1999) Interacting particle systems approximations of the Kushner-Stratonovich equation. *Adv. Appl. Probab.*, 31, 819–838.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19, 263–284.

Dassios, A. and Jang, H.-W. (2005) Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *J. Appl. Probab.*, 42, 93–107.

Del Moral, P. and Guionnet, A. (2001) On the stability of interaction processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincare Probab. Statist.*, 37, 155–194.

Del Moral, P., Jacod, J. and Protter, P. (2001) The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Reltd Flds*, 120, 346–368.

Del Moral, P. and Miclo, L. (2000) Branching and interacting particle systems: approximations of Feymann-Kac formulae with applications to non-linear filtering. *Sem. Probab. Stras.*, 34, 1–145.

Doucet, A., de Freitas, N. and Gordon, N. (2001) An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*, pp. 3–14. New York: Springer.

Duffie, D. and Singleton, K. J. (1999) Modeling term structures of defaultable bonds. *Rev. Finan. Stud.*, 12, 687–720.

Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with comments). *J. Bus. Econ. Statist.*, 20, 297–338.

Engel, R. F. (2000) The econometrics of ultra-high-frequency data. *Econometrica*, 68, 1–22.

Golightly, A. and Wilkinson, D. J. (2006) Bayesian sequential inference for nonlinear multivariate diffusions. *Statist. Comput.*, 16, 323–338.

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

Ionides, E. (2003) Inference and filtering for partially observed diffusion processes via sequential Monte Carlo. *Working Paper*. University of Michigan, Ann Arbor. (Available from http://www.stat.lsa.umich.edu/~ionides/pubs/WorkingPaper-filters.pdf.)

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

Kou, S. C., Xie, X. S. and Liu, J. S. (2005) Bayesian analysis of single-molecule experimental data (with discussion). *Appl. Statist.*, 54, 469–506.

Künsch, H. R. (2005) Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.*, 33, 1983–2021.

Liu, J. S. and Chen, R. (1998) Sequential Monte Carlo methods for dynamic systems. *J. Am. Statist. Ass.*, 93, 1032–1044.

Nicolau, J. (2002) A new technique for simulating the likelihood of stochastic differential equations. *Econometr. J.*, 5, 91–103.

Pitt, M. K. and Shephard, N. (1999) Filtering via simulation: auxiliary particle filters. *J. Am. Statist. Ass.*, 94, 590–599.