Inference and Filtering for Partially Observed Diffusion Processes via Sequential Monte Carlo

E. L. Ionides
Department of Statistics
The University of Michigan, Ann Arbor

Abstract

Diffusion processes observed partially, typically at discrete timepoints and possibly with observation error, arise when constructing stochastic models in continuous time. This paper introduces a novel Sequential Monte Carlo approach to inference for partially observed diffusion processes. The method of Sequential Monte Carlo provides an alternative to Markov Chain Monte Carlo methods, and has proven to be effective in complex models at the cutting edge of scientific research. The new methodology enables filtering, prediction, smoothing and parameter estimation in certain nonlinear models for which these are difficult or impossible using existing Monte Carlo methods. As a byproduct, a new method is presented for inference from discretely observed diffusion processes. A novel measure of filter accuracy is proposed, and used to highlight strengths and weaknesses of the methods.

Key Words: Discrete time sampling; Inference for stochastic processes; nonlinear diffusion; particle filter; stochastic differential equation.

1 Introduction

This paper develops new Sequential Monte Carlo techniques for partially observed diffusion processes. A partially observed Markov process is called a state space model. State space models have applications in many areas, including signal processing (Anderson and Moore, 1979), time series analysis (Shumway and Stoffer, 2000; Durbin and Koopman, 2001), economics (Harvey, 1989), finance (Shephard and Pitt, 1997), cell biology (Ionides et al., 2003), meteorology (Evensen and van Leeuwen, 1996), neuroscience (Brown et al., 1998), and various others (Doucet et al., 2001, Section IV). It appears that real world phenomena are often well modeled by some Markov process with sufficiently rich state space, perhaps constructed according to physical or chemical or economic principles, about which we can make only noisy or incomplete observations. In almost all the examples above, the Markov process can be considered as a continuous time process with continuous sample paths, and with observations occurring at discrete time points. We call such models partially observed diffusion processes. The widespread use of partially observed diffusion processes, and the desire to analyze ever more complex and general models, means that new methods for inference have potentially many important applications.

Introducing some notation, a state space model consists of an unobserved Markov process, x_t, called the *state process*, and an observed process, $y_t = g(x_t, t)$, called the *observation process*. Here, x_t takes values in a *state space*, $\mathcal{X}$, and y_t takes values in an *observation space*, $\mathcal{Y}$. To allow for noisy observations, we may wish to write $y_t = g(x_t, \eta_t, t)$ where η_t is a Markov process independent of x_t. Formally, we can then take (x_t, η_t) as the state process. We suppose that the observations occur

at discrete timepoints, $t = 1, \ldots, T$. When $\mathcal{X}$ is finite or countable, a state space model is called a hidden Markov model (Lander and Green, 1987; Rabiner, 1989). The case of linear Gaussian state space models was considered in the seminal work of Kalman (1960). We are concerned with the situation in which x_t is an an Itô diffusion in $\mathbb{R}^m$, defined as the solution to a stochastic differential equation (SDE)

$$dx_t = \mu(x_t, t, \theta)dt + \sigma(x_t, t, \theta)dW_t \tag{1}$$

with σ an $m \times m$ matrix and $\{W_t\}$ a Brownian motion in $\mathbb{R}^m$. Here, θ is a unknown parameter vector for which inference is required. For brevity and simplicity, dependence on θ and t will not always be written explicitly. We assume that μ and σ are sufficiently regular to guarantee that (1) has a unique solution (Oksendal, 1998, Theorem 5.2.1). We suppose that y_t lies in $\mathbb{R}^n$, with observations at discrete times $t = 1, \ldots, T$ given by

$$y_t = g\left(x_t, \eta_t, \theta, t\right), \tag{2}$$

where $\{\eta_t, t = 1, \ldots, T\}$, is an independent sequence that may be interpreted as measurement noise. This general setting includes observation processes taking a discrete set of possible values. As a concrete example, the reader may like to bear in mind a linear Gaussian observation equation,

$$y_t = Cx_t + D + \eta_t, \tag{3}$$

with $\eta_t \sim N(0, \tau\tau')$, C an $m \times n$ matrix and D a vector.

Section 2 discusses Sequential Monte Carlo methods for state space models. A new Sequential Monte Carlo technique, which we call a *conditional particle filter*, is proposed in Section 2.1. Section 2.4 introduces a new measure of filter accuracy, which is used to compare the effectiveness of competing filters. In Section 3 we show how to apply the conditional particle filter (CPF) in the context of partially observed diffusion processes. Consideration of numerical methods also leads us to propose a linearized conditional particle filter (LCPF) in Section 3.4.

Perhaps the main motivation for developing CPF and LCPF is that, as Section 3 reveals, they are well suited for some particular characteristics of partially observed diffusion processes. Transition densities of nonlinear diffusions and conditional nonlinear diffusions are hard to calculate, however diffusion processes are easy to simulate from. Likelihood ratios are also easy to calculate. In addition, nonlinear diffusions are similar to linear, Gaussian processes locally in space and time. The example of Section 3.5 demonstrates that CPF can compare favorably with existing filtering methods for partially observed diffusion processes, particularly with low observation noise. LCPF, while being less compuationally intensive than CPF, is found to be considerably more accurate than an extended Kalman filter. Remarkably, CPF can (with a little extra care) handle the important situation of singular noise. This is investigated in Section 3.6.

2 Sequential Monte Carlo Analysis of State Space Models

We first introduce some more notation, which is convenient for discussions of discrete time filtering. A vector of the observations is written as as $y_{1,T} = (y_1, \ldots, y_T)'$. We suppose that x_t has a conditional transition density $f(x_t|x_{t-1})$, and that y_t, conditional on $y_{1,t-1}$ and $x_{1,t}$, has density $f(y_t|y_{1,t-1}, x_{1,t}) = f(y_t|x_t)$. We adopt a convention that $f(\cdot \mid \cdot)$ is a generic density which is then specified by its arguments. One typically wishes to evaluate, or sample from, the conditional densities $f(x_t|y_{1,t})$, $f(x_{t+1}|y_{1,t})$ or $f(x_t|y_{1,T})$. These are known as the filtering, prediction and smoothing problems, respectively.

When trying to simulate a realization $x_{1,T} = (x_1, \ldots, x_T)'$ from a high dimensional density $f_T(x_{1,T})$, it may help to work sequentially by drawing $x_{1,t}$ from some density $f_t(x_{1,t})$ and then updating $x_{1,t}$ to generate a draw from $f_{t+1}(x_{1,t+1})$. This approach has been used for computer simulation of long-chain polymers since the 1950's and was surveyed by Liu (2001). We refer to such a technique as Sequential Monte Carlo (SMC). Sequential Monte Carlo provides an alternative to Markov Chain Monte Carlo methods, and has proven useful for many computationally challenging applications (Doucet et al., 2001). When SMC is used as a fundamental tool for hypothesis testing, parameter estimation, graphical representations and model diagnostics for a class of models, we describe this as an SMC analysis. Many researchers independently discovered that SMC is readily applicable to state space models (Crisan and Lyons, 1997; Del Moral, 1996; Gordon et al., 1993; Isard and Blake, 1996; Kitagawa, 1996).

A standard SMC approach, which we will call the particle filter (PF), involves recursively solving the prediction and filtering equations. These are given respectively by

$$f(x_t|y_{1,t-1}) = \int f(x_{t-1}|y_{1,t-1}) f(x_t|x_{t-1})\, dx_{t-1}, \tag{4}$$

$$f(x_t|y_{1,t}) = \frac{f(x_t|y_{1,t-1}) f(y_t|x_t)}{\int f(x_t|y_{1,t-1}) f(y_t|x_t)\, dx_t}. \tag{5}$$

We look for a solution in the Monte Carlo sense of producing a collection of random variables, or "particles," whose marginal densities approximate $f(x_t|y_{1,t-1})$ and $f(x_t|y_{1,t})$. Suppose $\{X^F_{t-1,j}, j = 1, \ldots, J\}$ has an empirical distribution approximating $f(x_{t-1}|y_{1,t-1})$. Using (4), each "particle" $X^F_{t-1,j}$ can be moved according to the transition density of the state space to give a particle $X^P_{t,j}$ whose marginal distribution approximates the prediction density, $f(x_t|y_{1,t-1})$. This Monte Carlo analogue to (4) is called the *prediction step*. The *filtering step* then involves setting $\{X^F_{t,j}\}$ to be a sample drawn with replacement from $\{X^P_{t,j}\}$ with probabilities proportional to

$$w_j = f(y_t|x_t{=}X^P_{t,j}).$$

This is an importance sampling procedure (Ripley, 1987) with sampling weights $\{\omega_j\}$. For initialization, $X^P_{1,j}$ is drawn from a density $f(x_1)$. Some variations on the standard PF algorithm are discussed in Section 2.2 and (Doucet et al., 2001). Particle filter algorithms have the difficulty that they work poorly when the observation error is small. Heuristically the weights, $\{w_j\}$, should then have high variance since most of the prediction steps will be inconsistent with the new observation. In the extreme case, where the density $f(x_t|y_{1,t})$ exists on a region of lower dimension than $f(x_t|y_{1,t-1})$, w_j is zero with probability one, and the algorithm breaks down. This will happen, for example, if $\tau\tau'$ in (3) is not invertible. Section 3.6 investigates this situation.

2.1 A New SMC Approach - the Conditional Particle Filter (CPF)

The prediction step in PF can be replaced by a more general proposal step (Liu, 2001). Pitt and Shepard (2001) allowed the proposal step to depend on the observation process and called such methods *adaptive particle filters*. Our approach to remedy the difficulties experienced by existing SMC algorithms for low or singular measurement error is to develop a new adaptive particle filter. Critically, in Section 3, we show how to evaluate this filter for systems such as (1) and (2). We replace (4) and (5) by a proposal equation

$$f_p(x_{t-1}, x_t|y_{1,t}) = f(x_{t-1}|y_{1,t-1}) f(x_t|x_{t-1}, y_t), \tag{6}$$

and a filtering equation

$$f(x_t|y_{1,t}) \propto \int f_p(x_{t-1}, x_t|y_{1,t-1}) f(y_t|x_{t-1})\, dx_{t-1}. \tag{7}$$

This algorithm is based on the identity

$$f(x_t|y_{1,t}) \;=\; \int \frac{f(x_{t-1}|y_{1,t-1}) f(y_t|x_{t-1})}{f(y_t|y_{1,t-1})} f(x_t|x_{t-1}, y_t)\, dx_{t-1}.$$

The proposal step involves constructing pairs $(X^F_{t-1,j}, X^P_{t,j})$ where $X^P_{t,j}$ is drawn from $f(x_t|y_t, x_{t-1}{=}X^F_{t-1,j})$. We then assign weight w_j to $X^P_{t,j}$ given by

$$w_j = f(y_t|x_{t-1}{=}X^F_{t-1,j}). \tag{8}$$

The filtering step, as for PF, is to construct $\{X^F_{t,j}\}$ by resampling with replacement from $\{X^P_{t,j}\}$ with weights w_j, or by following other schemes discussed in Section 2.2. The key to the algorithm is that, for the class of models including (1) and (2), these proposal and filtering steps can be solved (in a Monte Carlo sense).

We call the algorithm specified by (6), (7) and (8) a conditional particle filter (CPF). We postpone to Section 3 the issue of computing CPF for systems such as (1) and (2). In Sections 2.2 and 2.3 we discuss some topics arising in any SMC analysis. We discover that CPF, considered as a special case of SMC, has some unusual and useful properties. In Section 2.4 we consider the comparison and evaluation of filters, a discussion necessary for the examples of Secion 3.

2.2 Weights and Resampling

Manipulating weights in importance sampling type methods, such as SMC, can be done following the principle of properly weighted samples (Liu, 2001). A weighted sample $\{(X_i, w_i), i = 1, \ldots, N\}$ is said to be *properly weighted* for a density $f(x)$ if, for some constant C and any square integrable function $h(x)$,

$$E[h(X_i)w_i] = C \int h(x) f(x)\, dx.$$

For any set of positive resampling weights $\{a_i\}$, sampling with replacement from $\{(X_i, w_i/a_i)\}$ with weights a_i results in a new (approximately) proper sample for $f(x)$ (Rubin, 1987). The resampling weights originally used for SMC in state space models were $a_i = w_i$ (Kitagawa, 1996; Gordon et al., 1993). These weights prune heavily any X_i with small weight w_i, allowing computation at future steps to concentrate on "successful" particles. In practice, lighter pruning, say with $a_i = \sqrt{w_i}$, may be preferable (Liu, 2001). In fact, for the CPF and LCPF algorithms proposed here, resampling may be unnecessary. Heuristically, the reason for this is that for these algorithms the particles can be expected to have mixing properties. In the original SMC each particle $\tilde{x}_t$ moves around the state space according to the unconditional state Markov transition density $f(x_t|x_{t-1})$. Pruning is then required to remove particles whose location is unlikely given the observation process. In CPF, a particle $\hat{x}_t$ moves according to a conditional Markov transition density $f(x_t|x_{t-1}, y_t)$. If mixing occurs, then the distribution of $\hat{x}_t$ should not stray too far from the filtering distribution $f(x_t|y_{1,t})$. Mixing properties of similar conditional processes have been studied by (Bickel and Ritov, 1996; Bickel et al., 1998; Jensen and Petersen, 1999). The proper weight of $\hat{x}_t$ for $f(x_t|y_{1,t})$, the *filtering weight*, is $w_{f,t} = \prod_{i=1}^{t} v_i$ where $v_t = f(y_t|x_{t-1}{=}\hat{x}_{t-1})$. When t is not small, $w_{f,t}$ typically has high variance. However, if $\hat{x}_t$ is mixing then the recent weights will be more relevant, so *truncated*

weights $w_{f,t,k} = \prod_{i=t-k}^{t} v_i$ can be used. A similar issue arises in smoothing, where $(\hat{x}_t, w_{s,t})$ is properly weighted for $f(x_t|y_{1,T})$ with *smoothing weights* and *truncated smoothing weights* given by $w_{s,t} = \prod_{i=1}^{T} v_i$, and $w_{s,t,k} = \prod_{i=t-k}^{t+k} v_i$. Although these truncated weights somewhat resemble previous methods using truncated smoothers of (Anderson and Moore, 1979, Sections 7.3 and 9.6) and Kitagawa and Sato (2001), this particular method appears new. Note that it is only the weights and not the particle process that is truncated. A reasonable value of k can be selected by bias/variance trade-off considerations. A large k increases the Monte Carlo variance of an estimated quantity (such as the mean or quantiles of the conditional state process, or the log likelihood of the observations). A small k introduces bias due to the truncation. There may be room for improvement on truncation as a way of taking advantage of mixing. Filtering weights of the form $w_{f,t} = \exp\left\{\sum_{i=1}^{t} \rho^{t-i} \log v_i\right\}$ might be considered, for example.

2.3 Parameter Estimation

The likelihood may be calculated via $f(y_{1,T}) = \prod_{t=1}^{T} f(y_t|y_{1,t-1})$ using a Monte Carlo approximation to the integral

$$f(y_t|y_{1,t-1}) = \int f(y_t|x_{t-1}) f(x_{t-1}|y_{1,t-1})\, dx_{t-1}.$$

Remembering that there is a parameter, θ, implicit in all the model equations, inference for θ may be carried out using the usual Bayesian or frequentist methods. When no resampling is carried out the Monte Carlo variation can be dealt with by fixing the seed of the random number generator, finding estimates $\hat{\theta}_1, \ldots, \hat{\theta}_L$ at L different seed values, and combining these estimates. In the presence of resampling, the branching of the particles depends on θ. One can then sample the likelihood with Monte Carlo error at different parameter values, a relatively simple approach which was given a theoretical justification by Ionides (2003b). Alternatively, one can adopt an importance sampling strategy, or use a Monte Carlo EM algorithms (Hürzeler and Künsch, 2001).

Parameter estimation may also be carried out by adding θ to the state space (Anderson and Moore, 1979). This method is more successful if θ is a dynamic parameter, i.e., $\theta = \theta_t$ where θ_t is itself a diffusion process. If θ does not change with time, say $\theta = \theta_t = \theta_0$ with some prior distribution $f(\theta_0)$, then the filtering estimate $f(\theta_t|y_{1,t})$ is formally a Bayesian posterior distribution for θ_0 given $y_{1,t}$. However, as pointed out in Section 2.2, CPF and all other SMC methods perform poorly when the conditional process $\hat{\theta}_t$ with transition density $f(\theta_t|\theta_{t-1} = \hat{\theta}_{t-1}, y_t)$ is slowly mixing. In the fixed parameter case, $\hat{\theta}_t = \hat{\theta}_0$ so $\hat{\theta}_t$ does not "mix" at all. The θ components of the particles are then stuck at their initial positions, and only their weights change with time.

2.4 Comparing Filters

Traditional criteria for comparing filters are the mean square errors of point estimates of the center (mean or median) of the prediction or filtering distributions (Doucet et al., 2001). For nonlinear models these criteria are not necessarily appropriate, particularly when the filtering or prediction densities are multimodal. It would clearly be desirable for a filter to be able to estimate well the whole of the conditional distribution. To this end, we propose a new method for evaluating filters via the *accuracy*, A, defined by

$$A^2 = E\left[\frac{1}{T}\sum_{t=1}^{T}(\hat{\lambda}_t - \lambda_t)^2\right] \tag{9}$$

where $\lambda_t = \log f(y_t|y_{1,t-1})$ and $\hat{\lambda}_t$ is the filter estimate of λ_t. Although an inaccurate filter might provide a reasonable point estimate of the state process or a parameter, an accurate filter (with

a small value of A) has to capture the whole conditional distribution of x_t given $y_{1,t-1}$. Accuracy concerns the ability to approximate the log likelihood, which is particularly relevant to likelihood based (Bayesian or frequentist) inference. Accuracy has a natural scale: Differences in the log likelihood of order 1 are of interest for inference, so a filter with $A < 1/\sqrt{T}$ may be considered satisfactory for likelihood based inference. Accuracy also provides a one dimensional measure, even for a vector valued process. The accuracy of a filter may be hard to evaluate, because λ_t is not known unless we already have a perfect filter, with $A = 0$. For the particular nonlinear models of Section 3.5, λ_t can be found exactly. In general, λ_t will be estimated from a filter known to be accurate on standard problems. For parameter estimation, a filter has to estimate the likelihood not only under the true model but also in a neighborhood around it. It seems reasonable to suppose that accuracy for a known model is indicative of accuracy for similar models. If the observation process is stationary or converges to stationarity then, for large t and T or more formally in a limit, we can replace (9) by

$$A^2 = E[(\hat{\lambda}_t - \lambda_t)^2].$$

This can be broken down by a bias/variance decomposition, as $A^2 = b^2 + V$, where $b = E[\hat{\lambda}_t - \lambda_t]$ and $V = \mathrm{Var}(\hat{\lambda}_t - \lambda_t)$. For a Monte Carlo filter, define

$$A^2 = E[E_{MC}[(\hat{\lambda}_t - \lambda_t)^2]].$$

where E_{MC} is the Monte Carlo expectation. Writing $\lambda_t^* = E_{MC}[\hat{\lambda}_t]$, the bias/variance decomposition is now $A^2 = b^2 + V + M$, where the bias is $b = E[\lambda_t^* - \lambda_t]$, the *filter variance* is $V = \mathrm{Var}(\lambda_t^* - \lambda_t)$ and the *Monte Carlo variance* is $M = \mathrm{Var}_{MC}(\hat{\lambda}_t - \lambda_t^*)$.

In practice, the accuracy of a filter has to be estimated. Suppose K realizations of a Monte Carlo filter are available, for a stationary state space model, giving rise to estimated likelihood processes $\{\hat{\lambda}_{tk} : t = 1, \ldots, T, k = 1, \ldots, K\}$. Then we can write

$$\hat{\lambda}_{tk} = \lambda_t + b + \epsilon_t + \eta_{tk},$$

where $E[\epsilon_t] = E[\eta_{tk}] = 0$, and ϵ_t is independent of η_{tk}. The filter variance is $V = \mathrm{Var}(\epsilon_t)$ and the Monte Carlo variance is $M = \mathrm{Var}(\eta_{tk})$. A random effects model (Venables and Ripley, 2002) could be used to estimate b, M and V, though the standard errors will be too low unless one allows for serial correlation over time. Here we propose simpler unbiased estimators,

$$\begin{array}{rclcrcl} \hat{\lambda}_t^* & = & (1/K)\sum_k \hat{\lambda}_{tk}, & & \hat{M} & = & \frac{1}{T(K-1)}\sum_{t,k}(\hat{\lambda}_{tk} - \hat{\lambda}_t^*)^2, \\ \hat{b} & = & (1/T)\sum_t \hat{\lambda}_t^*, & & \hat{V} & = & \frac{1}{TK-1}\sum_{t,k}(\hat{\lambda}_{tk} - \hat{b})^2 - \hat{M}. \end{array}$$

These estimates are repeated for many independent realizations of the state space model, to get improved estimates with standard errors.

3 Inference for Partially Observed Diffusion Processes

Inference for discretely observed diffusion processes (when $y_t = x_t$ at discrete timepoints $t = 1, \ldots, T$) has been studied via likelihood based methods (Roberts and Stramer, 2001; Elerian et al., 2001), via estimating equations (Sorensen, 1997) and nonparametrically (Fan and Zhang, 2003; Jiang and Knight, 1997). Financial modeling has been a major motivating example. Partially but continuously observed nonlinear diffusions are the domain of the Wong-Zakai and Kallianpur-Striebel equations of filtering theory (Oksendal, 1998, Chapter 6). Numerical methods for continuous observation nonlinear filtering are poorly developed, and methods based on SMC may have a

role to play (Kloeden and Platen, 1999, Section 6.6). An SMC approach to the partially observed diffusion process (1) and (2), for the special case of a linear Gaussian state process, was developed in Shephard and Pitt (1997), motivated by a stochastic volatility financial model.

In this section, we show how to implement the CPF algorithm of Section 2.1 for the partial observation system (1) and (2). To implement CPF one has to simulate from $f(x_t|x_{t-1}, y_t)$, discussed in Section 3.1, and to evaluate $f(y_t|x_{t-1})$, discussed in Section 3.2. Section 3.3 discusses relevant numerical methods for SDEs. Carrying out CPF using a crude numerical solution gives rise to an algorithm of interest in its own right, presented in Section 3.4, which we call a linearized conditional particle filter (LCPF). Section 3.5 describes a test problem demonstrating the effectiveness of CPF. Section 3.6 shows that CPF also provides a novel approach for discretely observed diffusions.

3.1 Simulating Conditional Diffusions

Conditioning the diffusion $\{x_t\}$ on x_0 and y_1 results in a conditional diffusion $\{\hat{x}_t\} = \{x_t|x_0, y_1\}$. The SDE for $\hat{x}_t$ may be written as $d\hat{x}_t = \hat{\mu}(\hat{x}_t)dt + \sigma(\hat{x}_t)dW_t$, for $0 \le t \le 1$. Here $\sigma(\cdot)$ is the same as in (1) and $\hat{\mu}(\cdot)$ is not in general easy to calculate. Let $\tilde{x}_t$ be an approximation to $\hat{x}_t$, given by the SDE

$$d\tilde{x}_t = \tilde{\mu}(\tilde{x}_t)dt + \sigma(\tilde{x}_t)dW_t.$$

Write P, $\hat{P}$ and $\tilde{P}$ for the laws of $\{x_t|x_0\}$, $\{\hat{x}_t\}$ and $\{\tilde{x}_t\}$ respectively, for $0 \le t \le 1$. To carry out importance sampling from $\hat{P}$ using trial distribution $\tilde{P}$ it is sufficient to be able to calculate $d\tilde{P}/dP$, which is proportional to $d\tilde{P}/d\hat{P}$. Thus, although $\tilde{\mu}$ should be similar to $\hat{\mu}$ in order to have a computationally efficient simulation, it is not necessary to know $\hat{\mu}$. This property is the functional form of the familiar identity $f(x|y) = f(x, y)/f(y) \propto f(x, y)$. The constant of proportionality depends on y, but is not required for importance sampling.

Here, we use an approximation, $\tilde{x}_t$, which is constructed using a local linearization, following Ionides (2003a). Similar approximations were used by Roberts and Stramer (2001) and Ozaki (1992). The requirement on g in (2) is that one can make a local linear, Gaussian approximation to η_t and y_t. The better this approximation, the more computationally efficient the Monte Carlo scheme should be. The observation process does not have to be real valued, it could be categorical or integer valued. In some situations of practical interest, $d\hat{P}/dP$ does not exist, the most important case being where $\tau\tau'$ in (3) is singular or zero. This possibility is dealt with in Section 3.6.

3.2 Estimating the Conditional Density

An analogous situation to calculating the likelihood in a state space model via the methods of Section 2.1, for a bivariate random variable (X, Y) with density $f(x, y)$, is to estimate $f(y)$ using

$$f(y) \approx \hat{f}(y) = \frac{1}{K}\sum_{k=1}^{K} f(y|x{=}X_k) \tag{10}$$

where X_k is a sample from density $f(x)$. If X_k is instead drawn from another density $g(x)$ we have an estimate

$$f(y) \approx \hat{f}(y) = \frac{1}{K}\sum_{k=1}^{K} \frac{f(x{=}X_k)}{g(x{=}X_k)}. \tag{11}$$

The density $g(x)$ can be chosen to depend on the observation, y, and so the ideal choice would be $g(x) = f(x|y)$, in which case $\mathrm{Var}(\hat{f}(y)) = 0$. In the analogy with nonlinear state space models, $f(x|y)$ is not a density that can usually be calculated explicitly, though it can be sampled from

using the methods of Section 3.1. We then take $g(x)$ to be a linearized approximation to $f(x|y)$. Using the notation of Section 3.1, we can rewrite (11) in the form required in (8), as

$$f(y_1|x_0) = E_{\tilde{P}}\left[f(y_1|x_1{=}\tilde{x}_1)\frac{dP}{d\tilde{P}}\{\tilde{x}_t\}\right] \approx \frac{1}{K}\sum_{k=1}^{K} f(y_1|x_1{=}\tilde{x}_1^{(k)})\frac{dP}{d\tilde{P}}\{\tilde{x}_t^{(k)}\} \tag{12}$$

where $\{\tilde{x}_t^{(k)}\}$ is a sample from $\tilde{P}$. Conveniently, the same sample from $\tilde{P}$ can be used to compute (12) as is used for importance sampling from $\{\hat{x}_t\}$ in Section 3.1.

3.3 Numerical Methods

For a practical implementation of CPF the necessary sample paths and stochastic integrals must be discretely approximated. Appropriate numerical methods are discussed by Kloeden and Platen (1999). Different algorithms arise for weak and strong approximations. Here, we use strong approximations in order to guarantee the correct joint distributions of solutions to varying SDEs driven by the same Brownian motion. The Euler scheme, or order 0.5 strong Taylor approximation, to

$$dx_t = \mu(x_t, t)dt + \sigma(x_t, t)dW_t$$

is given by $x_{t+\epsilon} = x_t + \mu(x_t, t)\epsilon + \sigma(x_t, t)\Delta_\epsilon W_t$, where $\Delta_\epsilon W_t = W_{t+\epsilon} - W_t$. Set P to be the law of $\{x_t, 0 \le t \le 1\}$ and $\tilde{P}$ to be the law of the solution $\{\tilde{x}_t, 0 \le t \le 1\}$ to

$$d\tilde{x}_t = \tilde{\mu}(\tilde{x}_t, t)dt + \sigma(\tilde{x}_t, t)dW_t.$$

The likelihood ratio evaluated at $\{\xi_t, 0 \le t \le 1\}$, denoted $\frac{d\tilde{P}}{dP}\{\xi_t\}$, can be approximated by

$$\exp\left\{\frac{1}{2\epsilon}\sum_{n=0}^{N-1}[2\left(\tilde{\mu}_n - \mu_n\right)'\left(\sigma_n\sigma_n'\right)^{-1}\Delta_{(1/N)}\xi_{(n/N)} + \mu_n'(\sigma_n\sigma_n')^{-1}\mu_n - \tilde{\mu}_n'(\sigma_n\sigma_n')^{-1}\tilde{\mu}_n]\right\}, \tag{13}$$

where $\mu_n = \mu(\xi_{(n/N)}, n/N)$, *etc.* Note that (13) is both a Riemann sum approximation to the integral given by the Girsanov theorem and the exact likelihood ratio of the Euler approximation. For higher order Taylor approximations, these two quantities will usually be different. The exact likelihood ratio of the approximation appears more reasonable to use, though this would be hard to find for more complex schemes. The exception to this is the Milstein scheme (order 1.0 strong Taylor approximation) in the special case where W_t and x_t are real-valued or when $\sigma(\cdot)$ satisfies the commutativity property of Kloeden and Platen (1999, Section 10.3).

3.4 The Linearized Conditional Particle Filter (LCPF)

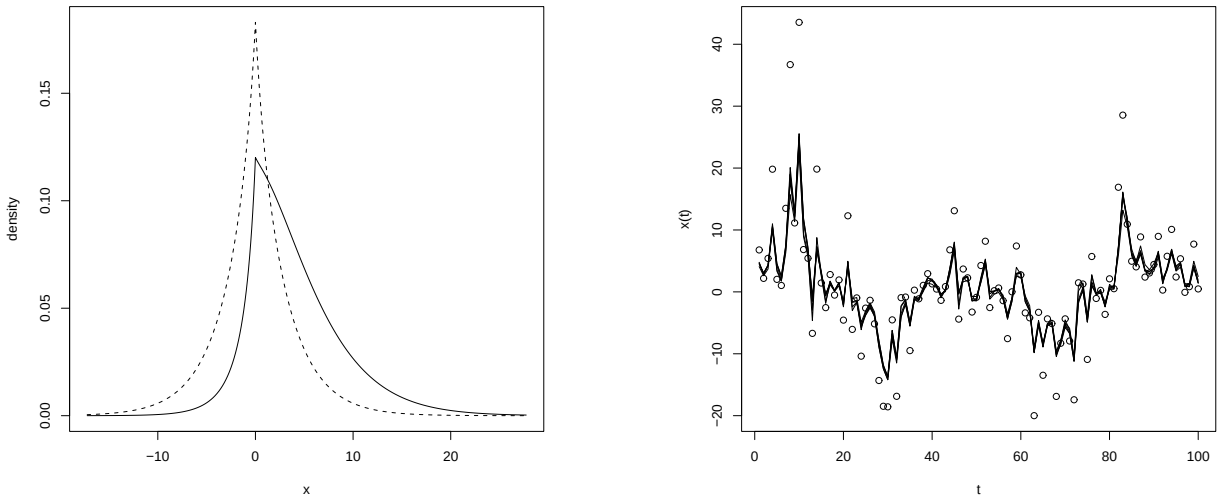
The Euler scheme of Section 3.3 with $\epsilon = 1$ corresponds to a local Gaussian approximation which may be written as

$$x_t|x_{t-1} \sim N\left(A(x_{t-1}), \Phi(x_{t-1})\right). \tag{14}$$

A corresponding local Gaussian observation equation is

$$y_t|x_t, x_{t-1} \sim N\left(C(x_{t-1})x_t + D(x_{t-1}), \Psi(x_{t-1})\right). \tag{15}$$

The LCPF consists of applying the CPF recursions in (6), (7) and (8) to the system specified by (14) and (15). We can not sample from the proposal density (6) and evaluate w_j in (8) using standard methods for the multivariate normal distribution. The model given by (14) and (15)



(a) **(b)**

Figure 1: **(a)** Prediction density, $f(x_t|y_{1,t-1})$, for (16) and (17) when the prediction median is $x_t^p = 3$ (solid line) and $x_t^p = 0$ (dashed line), when $\tau = 1$ and t is large. This is calculated by transforming the analytically tractable model given by (18) and (19). **(b)** A realization from (16) and (17), with $T = 100$. Transformed observations, $h(y_t)$, are shown as points. The filtering means, estimating the unobserved process x_t, are shown superimposed for EKF, PF, LCPF and CPF. The transformed Kalman filter for estimating z_t is also shown. All the methods agree closely on their point estimates of the unobserved process (the lines are indistinguishable).

somewhat resembles the linearization employed in the Extended Kalman filter (EKF), the widely used and fairly successful technique of linearizing a nonlinear model and then applying the Kalman filter (Anderson and Moore, 1979). The EKF is not an exact filter for the linearized model, in the sense that it does not find the true conditional distribution of the state process. In particular, the EKF estimate of the conditional distribution is necessarily unimodal. The LCPF, on the other hand, gives an exact solution, apart from Monte Carlo error, for the locally Gaussian system of (14) and (15). This system can be viewed as an approximation to a diffusion process of interest, but also has some relevance in its own right: In phenomenological modeling of financial or engineering systems, the underlying continuous time model may be a heuristic that can be discarded once it has been used to motivate a discrete time model.

The observation equation (15) has the interesting feature that C, D and Ψ can depend on x_{t-1}. This useful fact arises from (6) and (7), and has the consequence that the filtering weights are $f(y_t|x_{t-1})$ rather than the usual PF weights of $f(y_t|x_t)$. This is convenient since $f(y_t|x_t)$ may be poorly behaved when some linear combination of components of x_t can be observed with little or no error, whereas $f(y_t|x_{t-1})$ includes extra variation from the evolution of the state space which may push the distribution away from singularity. The author is not aware of previous work on methods closely resembling LCPF.

3.5 An Example

For a simple numerical example, consider a one dimensional nonlinear filtering problem for which an exact closed form filter exists, namely

$$dx_t = [-\alpha h^{-1}(x_t)h'(h^{-1}(x_t)) + \frac{1}{2}h''(h^{-1}(x_t))]dt + h'(h^{-1}(x_t))dW_t \quad (16)$$

$$y_t = h^{-1}(x_t) + \eta_t, \qquad \eta_t \sim N(0, \tau^2) \quad (17)$$

where $h(\cdot)$ has an inverse $h^{-1}(\cdot)$, a derivative $h'(x) = dh/dx$ and a second derivative $h''(x)$. In this section only, h' does not represent the transpose of h. If $x_t = h(z_t)$ then z_t and y_t form a linear Gaussian system

$$dz_t = -\alpha z_t dt + dW_t \quad (18)$$

$$y_t = z_t + \eta_t \quad (19)$$

The likelihood of $y_{1,T}$ can be found using the Kalman filter, if the linear representation in (18) and (19) is known. We demonstrate approximating this likelihood using (16) and (17) with an extended Kalman filter (EKF), particle filter (PF), particle Kalman filter (LCPF) and conditional particle filter (CPF). Here we consider the transformation

$$x = h(z) = [(|z| + 1)^2 - 1]sgn(z)$$

where $sgn(z) = \{1 \text{ if } z > 0, -1 \text{ if } z < 0\}$. This transform was chosen as it is a relatively simple differentiable, unbounded, nonlinear function with a differentiable, unbounded inverse. It leads to long tailed distributions and to a peak in the conditional density $f(x_t|y_{1,t})$ at $x_t = 0$ (see Fig. 1(a)). Using the inverse of this transform would give short tails and a bimodal stationary distribution for x_t.

Table 1 compares four filters, using the accuracy measure developed in Section 2.4. It shows that CPF is the most accurate filter in this situation, followed by PF, LCPF and finally EKF. The differences between the filters detected by the accuracy score are primarily in the tails of the conditional distributions. This is displayed graphically in Fig. 1(b), where the conditional means are seen to be indistinguishable for the different filters. The filter variance of EKF is large, because the conditional state distribution is not well approximated by a Gaussian (see Fig. 1(a)). If, say, $T = 100$ then $A\sqrt{T} > 1$ for EKF, and so EKF does not produce a reliable estimate of the log likelihood of the observations. The particle filter, PF, has a relatively high Monte Carlo error, $\sqrt{M}$, especially when the observation error, τ, becomes small. On the other hand, $\sqrt{M}$ decreases with smaller τ for LCPF and CPF. The filter error, $\sqrt{V}$, which is interpreted as the error after averaging over many replications of a Monte Carlo filter, is larger for each filter when $\tau = 0.25$. The techniques introduced in Section 3.6 for $\tau = 0$ can, in principle, be adapted improve the accuracy of CPF for small τ.

3.6 When There is No Measurement Error

With no measurement error, $\eta_t = 0$, the system (1) and (3) becomes

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t \quad (20)$$

$$y_t = Cx_t, \ \ t = 1, 2, \ldots T. \quad (21)$$

This case requires special attention because the law of x_t given $y_{1,t}$ will not have a density with respect to Lebesgue measure on $\mathbb{R}^m$, but instead on the linear space $\{x : Cx = y_t\}$ (recalling that

	τ	$b \times 10^2$	$\sqrt{V} \times 10^2$	$\sqrt{M} \times 10^2$	$A \times 10^2$
EKF	1	-3.32	25.1	n/a	25.4
LCPF	1	-1.62	15.5	4.67	16.4
PF	1	-0.45	3.26	8.54	9.27
CPF	1	-0.16	3.94	3.99	5.67
EKF	0.25	-8.1	42.7	n/a	43.9
LCPF	0.25	-5.3	30.7	2.94	31.8
PF	0.25	-2.52	13.7	21.7	29.8
CPF	0.25	-0.82	9.78	3.56	10.5

Table 1: Bias, b, filter error, $\sqrt{V}$, Monte Carlo error, $\sqrt{M}$, and accuracy, A. Calculated by simulation for (16) and (17), with parameter settings given below, for the conditional particle filter (CPF), linearized conditional particle filter (LCPF), particle filter (PF) and extended Kalman filter (EKF). Simulation errors are in the last digit presented.

Parameter	Description	Required for	Value
N_p	# of Particles	PF, LCPF, CPF	200
N_t	# of Steps per Observation for Numerical Solution of SDE	PF, CPF	5
N_f	# of lags used for filtering (Sec. 2.2)	LCPF, CPF	3
N_r	# of Trials for Importance Resampling	CPF	10

$x_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^n$, and supposing that $\sigma\sigma'$ is invertible). A situation of particular interest is $C = I$, the identity matrix, in which case the model is called a discretely observed diffusion. Many methods have been suggested for discretely observed diffusions (Elerian et al., 2001, and references theirin), motivated by applications in economics and finance. The framework considered here provides some fresh insights and new methodology. An algorithm that can handle no measurement error may also be expected to deal with small measurement error, so studying this case helps to explain why CPF can out-perform PF. It also leads to a technique for further improvement of CPF for low or singular noise.

The standard particle filter (PF), described in Section 2.1, does not work for $\eta_t = 0$ since $f(y_t|x_t = X^P_{t,j}) = 0$ with probability 1. The extended Kalman filter (EKF) and linearized conditional particle filter (LCPF) function as usual. The conditional particle filter (CPF) can be applied, but some extra care is required. Let P be the law of $\{x_t|x_0\}$, for $t \in [0,1]$, and $\hat{P}$ the law of $\{x_t|x_0, Cx_1 = y_1\}$, for $t \in [0,1]$. In this case, $dP/d\hat{P}$ does not exist, since Novikov's condition for Girsanov's theorem fails to apply (Oksendal, 1998, Theorem 8.6.5). A way around this problem is to define P_δ and $\hat{P}_\delta$ as the laws of x_t and $\hat{x}_t$ restricted to $t \in [0, 1-\delta]$. Now $dP_\delta/d\hat{P}_\delta$ does exist, and so

$$f(y_1|x_0) = E_{\hat{P}_\delta}\left[f(y_1|x_{1-\delta}{=}\hat{x}_{1-\delta})\frac{dP_\delta}{d\hat{P}_\delta}\{\hat{x}_t\}\right]. \tag{22}$$

For CPF we simulate from a law $\tilde{P}$ close to $\hat{P}$ and estimate $f(y_1|x_0)$ using

$$f(y_1|x_0) = E_{\tilde{P}_\delta}\left[f(y_1|x_{1-\delta}{=}\tilde{x}_{1-\delta})\frac{dP_\delta}{d\tilde{P}_\delta}\{\tilde{x}_t\}\right]. \tag{23}$$

In the examples below we see that as long as $\tilde{P}$ is close to $\hat{P}$ the product of the two terms $f(y_1|x_{1-\delta}{=}\tilde{x}_{1-\delta})$ and $dP_\delta/\tilde{P}_\delta$ in (23) is of order $(1/\sqrt{\delta}) \times \sqrt{\delta}$ for typical sample paths. A small value of δ contributes numerical instability, at a slow rate of $1/\sqrt{\delta}$, while fortunately adding relatively little Monte Carlo variation since the numerical instabilities cancel. The identification of this phenomenon (which the author found surprising), opens up the possibility of using CPF for singular measurement error distributions, and in particular for discretely observed diffusions. We will see, in Examples 1 and 2 below, that δ can be chosen sufficiently small that $f(y_t|x_{t-\delta}{=}\tilde{x}_{t-\delta})$ has a good linear Gaussian approximation, without causing numerical instabilities or high Monte Carlo variance.

Example 1. (Brownian bridge). Consider the one dimensional system $x_t = W_t$ and $y_1 = x_1$, which is a case of (20) and (21). Supposing that $x_0 = y_1 = 0$, the conditional process of $\hat{x}_t$ is a Brownian bridge, with SDE $d\hat{x}_t = \hat{x}_t(1-t)^{-1}dt + dW_t$. Then,

$$\frac{dP_\delta}{d\hat{P}_\delta}(\xi) = \exp\left\{2\int_0^{1-\delta}\frac{\xi_t\, d\xi_t}{1-t} + \int_0^{1-\delta}\frac{\xi_t^2}{(1-t)^2}\, dt\right\}. \tag{24}$$

As δ approaches 0, there are numerical difficulties evaluating the right hand side of (22), since $f(y_1|x_{1-\delta})$ is approximately the density of a $N(x_{1-\delta}, \delta)$ variable. Thus for typical sample paths under $\hat{P}_\delta$, $f(y_1|x_{1-\delta}{=}\hat{x}_{1-\delta})$ is of order $1/\sqrt{\delta}$. Novikov's condition, in this example, concerns the random variable

$$\nu_\delta(\hat{x}) = \frac{1}{2}\int_0^{1-\delta}\frac{\hat{x}_t^2}{(1-t)^2}\, dt.$$

The condition fails because $\lim_{\delta\to 0} E_{\hat{P}}[\exp\{\nu_\delta(\hat{x})\}] = \infty$. Conditional on ν_δ, in this example, $dP_\delta/d\tilde{P}_\delta$ has the same distribution as $\exp\{\sqrt{2\nu_\delta}Z - \nu_\delta\}$, where $Z \sim N(0,1)$. Keeping track of how ν_δ increases as δ approaches 0 may help to choose a reasonable value of δ.

N_t	$\hat{f}(y_1\|x_0)$	$dP_\delta/d\tilde{P}_\delta$	$f(y_1\|\tilde{x}_{1-\delta})$	$\nu_\delta(\tilde{x})$	δ
5	0.399(0.216)	1.00(2.63)	0.57(0.28)	0.32(0.37)	0.2
10	0.399(0.235)	0.92(3.68)	0.79(0.40)	0.64(0.67)	0.1
25	0.396(0.248)	0.83(8.60)	1.24(0.64)	1.09(1.05)	0.04
50	0.398(0.251)	0.68(18.0)	1.75(0.91)	1.41(1.26)	0.02

Table 2: Monte Carlo estimates, $\hat{f}(y_1|x_0)$, of $f(y_1|x_0)$ using CPF for a Brownian bridge with $\delta = 1/N_t$. The corresponding Monte Carlo standard deviations are given in parentheses. Based on a simulation study with sample size 10^4.

Method	N_t	δ	$\sqrt{V} \times 10^2$	$\sqrt{M} \times 10^2$	$A \times 10^2$
EKF	n/a	n/a	61	n/a	62
CPF	5	0.2	13.8	4.51	14.6
CPF	10	0.1	7.62	5.05	9.21
CPF	10	0.2	10.2	3.76	10.9
CPF	20	0.05	4.12	5.48	6.90
CPF	20	0.1	5.07	4.23	6.63
CPF	20	0.2	8.63	3.33	9.32

Table 3: Comparing EKF with CPF for the nonlinear diffusion in (16), observed discretely with no measurement error. The bias, b, was always small and negative, and is not reported. For CPF, $N_r = 200$. Setting $N_r = 800$ results in indistinguishable b and $\sqrt{V}$, and halves $\sqrt{M}$ (not reported). The computational effort is proportional to $N_t \times N_r$. CPF provides a marked improvement on EKF (which here is simply a local linearization of the diffusion process), even for modest values of N_t and N_r. Results are by simulation, with simulation error in the last digit presented.

A test problem is to calculate $f(y_1{=}0|x_0{=}0)$ by using (23), the correct answer being $1/\sqrt{2\pi} = 0.3989$. Here $\tilde{x}_t$ is taken to be an Euler approximation to the Brownian bridge $\hat{x}_t$, based on discretizing $[0, 1]$ into N_t equal intervals and setting $\delta = 1/N_t$ (formally, $\tilde{x}_t$ is the diffusion process with piecewise constant coefficients corresponding to this Euler approximation). Table 2 shows that the SD of the estimate of $f(y_1|x_0)$ increases remarkably slowly as δ becomes small. The importance weights, $dP_\delta/d\tilde{P}_\delta$ have a very long tailed distribution for small δ, but the large values of $dP_\delta/d\tilde{P}_\delta$ match small values of $f(y_1{=}0|x_{1-\delta}{=}\tilde{x}_{1-\delta})$. Since the observed mean of $dP_\delta/d\tilde{P}_\delta$ is less than one for small δ, the extreme tail of the distribution is not being sampled here. However, the estimate $\hat{f}(y_1|x_0)$ of $f(y_1|x_0)$ still performs well for small δ.

Note that in this example $\hat{f}(y_1|x_0)$ is unbiased, since the numerical solution of the SDE for x_t is exact in the sense that the Euler method for Brownian motion gives correct finite dimensional distributions. This reminds us that the bias is due to the error in the numerical solution for x_t, not in the approximation of $\hat{x}_t$ by $\tilde{x}_t$. For the Brownian bridge, $\hat{x}_t$, the Euler solution is not exact.

Example 2. (A Nonlinear Diffusion). For a slightly more challenging example, we consider the nonlinear diffusion (16) observed discretely with no measurement error. Some results are presented in Table 3. The filter error, $\sqrt{V}$, decreases as N_t increases and as δ decreases. The Monte Carlo error, $\sqrt{M}$, increases as δ decreases, but decreases proportional to $(N_r)^{-1/2}$ as the number of Monte Carlo replications, N_r, increases. The possible values of δ for this numerical implementation are integer multiples of $1/N_t$. Notice that taking $\delta = 1/N_t$ is not necessarily the best choice.

4 Discussion

We have introduced two new filtering algorithms for the important class of models which we call partially observed diffusion processes. The explanation for why the conditional particle filter (CPF) can be effective for partially observed diffusions is that we have developed good methods to simulate from $f(x_t|x_{t-1}, y_t)$ and to evaluate $f(y_t|x_{t-1})$. These methods, presented in Section 3, are based on a local Gaussian approximation to the state process, guaranteed if the state process is a diffusion process. The observation process can be non-Gaussian. In situations where CPF algorithms can be constructed, they have some desirable properties. In Sections 3.5 and 3.6 we saw that CPF algorithms can handle small or singular observation noise. In Section 2.2 we saw that little or no resampling may be required for CPF. Resampling is undesirable since it adds Monte Carlo variability. Also, an algorithm with little or no resampling is particularly suitable for parallel computation. Future work will attempt to increase out understanding of how to take advantage of mixing properties of the CPF weights, following the discussion in Section 2.2. The simple method implemented here, truncation at a reasonable lag, was effective on the test problems presented.

Since CPF may be considered as a special case of SMC, many strategies invented for fine-tuning other SMC algorithms (Doucet et al., 2001) may also be applicable to CPF. A considerable reduction in Monte Carlo variation may be possible by applying smoothing techniques, if $f(x_t|y_{1,t})$ and/or $f(y_{t+1}|x_t)$ are sufficiently smooth as functions of x_t. This idea was investigated for unconditional particle filters by Stavropoulos and Titterington (2001). Considerable computational gains may also be available for partially linear/Gaussian systems. This occurs where the state process can be split into two components, $x_t = (x_t^{(1)}, x_t^{(2)})$, and $x_t^{(1)}$ given x_{t-1} and $x_t^{(2)}$ is conditionally linear and Gaussian. Investigations of unconditional particle filtering in this situation include (Liu et al., 2001) and (Murphy and Russell, 2001).

For modeling of turbulence, financial markets and motion of bacterial cells there is some interest in removing the requirement that x_t has continuous sample paths, by allowing dW_t in (1) to be the increment process for a Lévy process rather than Brownian motion (Barndorff-Nielsen et al., 2001). The methods developed here are not immediately applicable to SDEs driven by Lévy processes, but may be a step in the right direction.

A possible generalization is that the state evolution equation and observation equation can be allowed to depend on previous observations, as in the conditional Gaussian filter (Anderson and Moore, 1979). State space methods also extend readily to unequally spaced observations, and similar methods apply for continuous time observations (Oksendal, 1998, Chapter 6).

We have outlined many possible refinements of CPF. However, the basic algorithm of (6) and (7), together with the techniques developed in Section 3, are already enough to provide a fairly simple approach that can be effective for nonlinear filtering problems that are difficult to solve by previous Monte Carlo strategies.

References

Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering.* Prentice-Hall, New Jersey.

Barndorff-Nielsen, O. E., Mikosch, T. M., and Resnick, S., editors (2001). *Lévy Processes, Theory and Applications.* Birkhäuser, New York.

Bickel, P. J. and Ritov, Y. (1996). Inference in hidden Markov models, I: Local asymptotic normality in the stationary case. *Bernoulli*, 2:199–228.

Bickel, P. J., Ritov, Y., and Ryden, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Annals of Statistics*, 26:1614–1635.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18:7411–7425.

Crisan, D. and Lyons, T. (1997). Nonlinear filtering and measure-valued processes. *Probability Theory and Related Fields*, 109(2):217–244.

Del Moral, P. (1996). Nonlinear filtering using random particles. *Theory of Probability and Its Applications*, 40:690–701.

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

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, 69:959–993.

Evensen, G. and van Leeuwen, P. J. (1996). Assimilation of geostat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124:58–96.

Fan, J. and Zhang, C. (2003). A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association*, 98:118–134.

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

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hürzeler, M. and Künsch, H. R. (2001). Approximating and maximising the likelihood for a general state-space model. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 159–175. Springer, New York.

Ionides, E. L. (2003a). Inference and filtering for partially observed diffusion processes via sequential Monte Carlo. *Tech. report #405, Univ. of Michigan Dept. of Statistics*.

Ionides, E. L. (2003b). Maximum smoothed likelihood estimation. *submitted to Statistica Sinica*.

Ionides, E. L., Fang, K. S., Isseroff, R. R., and Oster, G. F. (2003). Stochastic models for cell motion and taxis. *Journal of Mathematical Biology*.

Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density. In *European Conference on Computer Vision*, pages 343–356. Cambridge, UK.

Jensen, J. L. and Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Annals of Statistics*, 27:514–535.

Jiang, G. J. and Knight, J. L. (1997). A nonparametric approach to the estimation of diffusion processes, with an application to short-term interest rate models. *Econometric Theory*, 13:615–645.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25.

Kitagawa, G. and Sato, S. (2001). Monte carlo smoothing and self-organising state-space model. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 177–195. Springer, New York.

Kloeden, P. E. and Platen, E. (1999). *Numerical Soluion of Stochastic Differential Equations*. Springer, New York, 3rd edition.

Lander, E. S. and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Science, USA*, 84:2363–2367.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.

Liu, J. S., Chen, R., and Logvinenko, T. (2001). A theoretical framework for sequential importance sampling with resampling. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 225–242. Springer, New York.

Murphy, K. and Russell, S. (2001). Rao-Blackwellised particle filtering for dynamic Bayesian networks. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 499–515. Springer, New York.

Oksendal, B. (1998). *Stochastic Differential Equations*. Springer, New York, 5th edition.

Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamic systems: a local linearization approach. *Statistica Sinica*, 2:113–135.

Pitt, M. K. and Shepard, N. (2001). Auxillary variable based particle filters. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 273–293. Springer, New York.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257–285.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley, New York.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88:603–621.

Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the sir algorithm. comment on the calculation of posterior distributions by data augmentation, by M. A. Tanner and W. H. Wong. *Journal of the American Statistical Assocication*, 82:543–546.

Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667.

Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*. Springer, New York.

Sorensen, M. (1997). Estimating functions for discretely observed diffusions: A review. In Basawa, I. V., Godambe, V. P., and Taylor, R. L., editors, *Selected Proceedings of the Symposium on Estimating Functions*, pages 305–325. IMS Lecture Notes.

Stavropoulos, P. and Titterington, D. M. (2001). Improved particle filters and smoothing. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 295–317. Springer, New York.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S-Plus*. Springer, New York, 4th edition.