The Rao-Blackwellised Variable Rate Particle Smoother for Conditionally Linear-Gaussian Models

Pete Bunch, Simon Godsill, *Member, IEEE,*

Abstract—The abstract goes here.

Index Terms—

I. INTRODUCTION

IN model-based schemes for probabilistic inference, some unknown quantity is treated as a random process which evolves over time according to a dynamic model. This latent state is observed at a discrete set of times via another random process modelling the measurement mechanism. Using the two models and applying Bayes' rule, inference of the hidden state can be made from the observations.

For simple models with linear dynamics and Gaussian-distributed random variables, optimal analytic inference algorithms exist, including the Kalman filter [1] and Rauch-Tung-Striebel (RTS) smoother [2]. For nonlinear, non-Gaussian models, such analytic solutions do not exist, and it is often necessary to employ numerical approximations, including the particle filter [3] and particle smoother [4], [5] (see [6], [7] for a thorough introduction to particle methods).

The simplest form of state-space model treats the quantity under consideration as a discrete-time random process synchronous with the observations, and uses a fixed or known Markovian model for the state transitions. This leads to the standard discrete-time hidden Markov model. Such "fixed rate" models are poorly suited to quantities with non-homogeneous state dynamics, such as occasional discontinuities in their evolution; for example, the price of a financial asset which may display large jumps at random times between periods of diffusion-like behaviour, or the kinematic state of a manoeuvring vehicle which may have sudden jumps in the acceleration when turns begin or end. In such cases, a "variable rate" model may be more appropriate, in which the state evolution is dependent upon a set of unknown changepoints.

In a variable rate model, the set of changepoints and associated parameters are modelled as a marked point process (MPP) [8], conditional upon which the state evolves according to some benign dynamics. In [9], [10], the conditional state evolution is treated as deterministic, while in [11], [12] a conditionally linear-Gaussian state model is considered.

The posterior distribution of the changepoint sequence is analytically intractable, but may be estimated using a particle filter. When the conditional state dynamics are linear-Gaussian, the efficiency of this filter may be greatly improved by using the method of Rao-Blackwellisation (see, e.g. [13], [14]). Rather than targeting the entire posterior distribution over changepoints and states, the particle filter is used to estimate only the nonlinear changepoint sequence, after which Kalman filtering and smoothing may be used to infer the linear states. This is the Rao-Blackwellised variable rate particle filter (RBVRPF) of [11], [12].

In this paper, a new algorithm is described for use with conditionally linear-Gaussian variable rate models for estimation of the smoothing distribution, i.e. the distribution over the sequence of changepoints and linear states given all the observations. This uses a similar derivation to that of the fixed rate Rao-Blackwellised particle smoother of [15]. The new algorithm is called the Rao-Blackwellised variable rate particle smoother (RBVRPS).

We review the structure of conditionally linear-Gaussian variable rate models in section II and revise the RBVRPF in section III. The new smoothing algorithm is presented in section IV with supporting simulations in section V.

II. CONDITIONALLY LINEAR-GAUSSIAN VARIABLE RATE MODELS

We consider a general model from time 0 to T, between which observations, $\{y_1 \ldots y_N\}$, are made at times $\{t_1 \ldots t_N = T\}$. The linear state at these times is written as $\{x_1 \ldots x_N\}$. During this period, an unknown number of changepoints, K, occur at times $\{\tau_1 \ldots \tau_K\}$, each with associated changepoint parameters, $\{u_1 \ldots u_K\}$. These are the marks of the MPP. Discrete sets containing multiple values over time will be written as, e.g. $y_{1:n} = \{y_1 \ldots y_n\}$.

The objective for inference will be to estimate the sequence of changepoints. This will be denoted as $\theta = \{\tau_{1:K}, u_{1:K}\}$. In addition, at a particular time t_n, the sequence will be divided into past $\theta_n = \{\tau_j, u_j \forall j : 0 \leq \tau_j < t_n\}$, and future $\theta_n^+ = \{\tau_j, u_j \forall j : t_n \leq \tau_j < T\}$. It will also be useful to define a variable for the changepoints which occur in the interval $[t_{n-1}, t_n)$, $\theta_{n \backslash n-1} = \{\tau_j, u_j \forall j : t_{n-1} \leq \tau_j < t_n\}$. The cardinality of these sets (i.e. the number of changepoints occurring in the appropriate range) is denoted K_n, K_n^+ and $K_{n \backslash n-1}$ respectively.

Note that θ_n not only conveys information about where changepoints occur, but also where they do not. i.e. it is implicit that $\tau_{K_n+1} > t_n$.

P. Bunch and S. Godsill are with the Department of Engineering, Cambridge University, UK. email: {pb404,sjg30}@cam.ac.uk

Assuming a Markovian changepoint sequence, a conditionally linear-Gaussian variable rate model can now be expressed by the following system equations:

$$\{\tau_k, u_k\} \sim p(u_k|\tau_k, \tau_{k-1}, u_{k-1})p(\tau_k|\tau_{k-1}) \quad (1)$$

$$x_n = A_n(\theta_n)x_{n-1} + w_n \quad (2)$$

$$y_n = C_n(\theta_n)x_n + v_n. \quad (3)$$

The random variables w_n and v_n have a zero-mean Gaussian distribution with covariance matrices Q_n and R_n respectively. The changepoint density will be constructed such that $P(\tau_k < \tau_{k-1}) = 0$. The factorisation of (1) may be relaxed, but is assumed throughout as it simplifies the subsequent derivations.

It will be useful to define the survivor function in the same manner as [10].

$$S(\tau_k, t) = P(\tau_{k+1} > t|\tau_k)$$
$$= 1 - \int_{\tau_k}^{t} p(\xi|\tau_k)d\xi \quad (4)$$

Using the convention that $\tau_0 = 0$ and u_0 is known a priori, the prior distribution over the changepoint sequence, θ_n, may now be expressed as,

$$p(\theta_n) = S(\tau_{K_n}, t_n)\prod_{k=1}^{K_n} p(\tau_k, u_k|\tau_{k-1}, u_{k-1}). \quad (5)$$

See [8] for a rigorous derivation of MPP probability measures.

The tasks of filtering and smoothing may now be considered as estimation of the posterior distributions of θ and $x_{1:N}$.

III. THE RAO-BLACKWELLISED VARIABLE RATE PARTICLE FILTER

The Rao-Blackwellised Variable Rate Particle Filter (RB-VRPF) was first described in [11], and in [12] it was developed for use in a financial prediction algorithm. The objective of the algorithm is to sequentially estimate the distribution of the changepoint sequence, $p(\theta_n|y_{1:n})$, at each time t_n. The linear state filtering distribution, $p(x_n|\theta_n, y_{1:n})$, can then be estimated by a Kalman filter [1], [16].

The posterior changepoint distribution at time t_n may be expanded using Bayes' rule.

$$p(\theta_n|y_{1:n})$$
$$\propto p(y_n|\theta_n, y_{1:n-1})p(\theta_{n\backslash n-1}|\theta_{n-1})p(\theta_{n-1}|y_{1:n-1})(6)$$

The transition term, $p(\theta_{n\backslash n-1}|\theta_{n-1})$, can be constructed in a similar manner to (5) [8].

$$p(\theta_{n\backslash n-1}|\theta_{n-1})$$
$$= S(\tau_{K_n}, t_n)\prod_{j:t_{n-1}\leq\tau_j<t_n} p(\tau_j, u_j|\tau_{j-1}, u_{j-1}, \tau_j > t_{n-1})(7)$$

For all but the first changepoint in the interval, the density is given by the prior model of (1). For the first changepoint,

indexed by $k = K_{n-1} + 1$, we must account for the fact that a changepoint cannot occur before t_{n-1},

$$p(\tau_k, u_k|\tau_{k-1}, u_{k-1}, \tau_k > t_{n-1})$$
$$= \frac{1}{S(\tau_{k-1}, t_{n-1})}\begin{cases} p(\tau_k, u_k|\tau_{k-1}, u_{k-1}) & \tau_k > t_{n-1} \\ 0 & \tau_k < t_{n-1} \end{cases}.(8)$$

Practically, because changepoints will be relatively rare events, it is likely that $K_{n\backslash n-1} \leq 1$.

The likelihood term in (6), $p(y_n|\theta_n, y_{1:n-1})$ is the predictive distribution estimated by the Kalman filter, which will be a Gaussian distribution.

$$p(y_n|\theta_n, y_{1:n-1}) = \mathcal{N}(y_n|\mu_n, S_n) \quad (9)$$

The mean and variance are given by the following standard recursions (dependence on θ_n suppressed for clarity).

$$m_n^- = A_n m_{n-1} \quad (10)$$
$$P_n^- = A_n P_{n-1} A_n^T + Q_n \quad (11)$$
$$\mu_n = C_n m_n^- \quad (12)$$
$$S_n = C_n P_n^- C_n^T + R_n \quad (13)$$
$$K_n = P_n^- C_n^T S_n^{-1} \quad (14)$$
$$m_n = m_n^- + K_n(y_n - \mu_n) \quad (15)$$
$$P_n = P_n^- - K_n S_n K_n^T \quad (16)$$

Having defined the terms of the changepoint posterior, we now consider the inference algorithm. A particle filter is a numerical method used to approximate a filtering distribution with a discrete set of weighted samples drawn from that distribution using sequential importance sampling (IS). In this case, each particle will consist of a sequence of changepoints between 0 and t.

$$\hat{p}(\theta_n|y_{1:n}) = \sum_j w_n^{(j)}\delta_{\theta_n^{(j)}}(\theta_n) \quad (17)$$

where $\delta_x(X)$ is a dirac probability mass at $X = x$.

The particle filter is a recursive algorithm. At the nth step, a particle, $\theta_n^{(i)}$, is first sampled from an importance density.

$$q(\theta_n|y_{1:n}) = q(\theta_{n-1}|y_{1:n-1})q(\theta_{n\backslash n-1}|\theta_{n-1}, y_n) \quad (18)$$

The history proposal is constructed from the particles of the $(n-1)$th filtering distribution using an appropriately chosen set of proposal weights.

$$q(\theta_{n-1}|y_{1:n-1}) = \sum_j v_{n-1}^{(j)}\delta_{\theta_{n-1}^{(j)}}(\theta_{n-1}) \quad (19)$$

The particle is then extended from θ_{n-1} to θ_n by sampling from an importance density, $q(\theta_{n\backslash n-1}|\theta_{n-1}, y_n)$. Finally, the particle is weighted according to the ratio of the target and proposal densities.

$$
\begin{aligned}
w_n^{(i)} &= \frac{p(\theta_n^{(i)}|y_{1:n})}{q(\theta_n^{(i)}|y_{1:n})} \\
&\propto \frac{p(y_n|\theta_n^{(i)}, y_{1:n-1})p(\theta_{n\backslash n-1}^{(i)}|\theta_{n-1}^{(i)})p(\theta_{n-1}^{(i)}|y_{1:n-1})}{q(\theta_{n-1}^{(i)}|y_{1:n-1})q(\theta_{n\backslash n-1}^{(i)}|\theta_{n-1}^{(i)}, y_n)} \\
&= \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times \frac{p(y_n|\theta_n^{(i)}, y_{1:n-1})p(\theta_{n\backslash n-1}^{(i)}|\theta_{n-1}^{(i)})}{q(\theta_{n\backslash n-1}^{(i)}|\theta_{n-1}^{(i)})}
\end{aligned} \quad (20)
$$

The normalisation may be enforced by scaling the weights so that they sum to 1. This is the RBVRPF of [11], [12].

For the most basic "bootstrap" [3] form of the RBVRPF, $\theta_{n\backslash n-1}$ may be proposed from the prior transition density (7). This can be achieved by sampling new changepoints sequentially from the transition model (1) (apart from the first which is sampled from (8)) until one falls after the current time, t_n. This final future changepoint is discarded. (This process can be thought of as sampling the entire future changepoint sequence from t_{n-1} onwards, and then marginalising those which fall after t_n.) The bootstrap proposal leads to the usual simplification of the weight formula.

$$
w_n^{(i)} = \frac{w_{n-1}^{(i)}}{v_{n-1}^{(i)}} \times p(y_n|\theta_n^{(i)}, y_{1:n-1}) \quad (21)
$$

The choice of proposal weights, $\{v_{n-1}^{(i)}\}$, requires particular attention in the design of RBVRPFs. In some models a changepoint may not have an immediate effect on the observations, especially if a jump occurs in some quantity which is only observed via its integral, e.g. if there is a jump in the acceleration of a moving object, yet only the position is measured, the change will not be apparent until several more observations have been made. In the meantime, particles which contain a changepoint at the correct time may all have been removed by the resampling process. To avoid this loss of good particles, proposal weights should be chosen which preserve a significant number of low-weight particles. One scheme which has been found to work well is described in [9], in which proposal weights are given by:

$$
v_{n-1}^{(i)} \propto \max(1, N_F w_{n-1}^{(i)}) \quad (22)
$$

where N_F is the number of filtering particles. The RBVRPF is summarised in algorithm III.

A. Improving the Filter

The bootstrap RBVRPF as described above may perform poorly if changepoints are not obvious until significantly after they occur. In this case, the estimation may be improved by the introduction of resample-move (RM) steps [17]. In an RM scheme, optional Metropolis-Hastings (MH) moves are conducted to alter the particle states after the importance sampling has taken place. For variable rate models, any one of the previous changepoints, τ_k, or associated marks, u_k, could be adjusted. Because more observations are available

---

1: For each i, initialise particle sufficient statistics, $m_0^{(i)}$ and $P_0^{(i)}$ with prior values.
2: For each i, initialise particle changepoint sequence with $\theta_0^{(i)} \leftarrow \emptyset$.
3: **for** $n = 1 \ldots N$ **do**
4:     **for** $i = 1 \ldots N_F$ **do**
5:        Sample a changepoint history $\theta_{n-1}^{(i)} \sim \sum_j v_{n-1}^{(j)}\delta_{\theta_{n-1}^{(j)}}(\theta_{n-1})$.
6:        Propose $\theta_{n\backslash n-1}^{(i)} \sim q(\theta_{n\backslash n-1}|\theta_{n-1}^{(i)})$
7:        Add extension to changepoint sequence. $\theta_n^{(i)} \leftarrow \theta_{n-1}^{(i)} \cup \theta_{n\backslash n-1}^{(i)}$.
8:        Calculate $\mu_n^{(i)}$ and $S_n^{(i)}$ using (10) to (13).
9:        Update state mean and covariance $m_n^{(i)}$ and $P_n^{(i)}$ using (14) to (16).
10:        Calculate weight $w_n^{(i)}$ using (21).
11:     **end for**
12:     Scale weights such that $\sum_i w_n^{(i)} = 1$.
13: **end for**

Fig. 1. Rao-Blackwellised Variable Rate Particle Filter

than when the changepoint was first proposed, it may be possible to construct more informed proposals and so move the changepoints towards regions with higher posterior probability. It is even possible to retrospectively add or remove changepoints, using reversible jump MH moves [18]. Variable rate particle filters using RM with piecewise deterministic models are described in [10], [19].

Rather than conducting the IS and MH steps separately, it is possible to combine them using the framework of SMC samplers [20]. This was suggested in [10], again for piecewise deterministic dynamics, but the extension to conditionally linear-Gaussian models is straightforward.

Filtering schemes which alter past changepoints – whether using an SMC sampler or RM – are computationally expensive, because many likelihood calculations must be conducted, for each observation from the time of the change onwards. In some cases, it may be simpler to just use a bootstrap filter with more particles.

IV. THE RAO-BLACKWELLISED VARIABLE RATE PARTICLE SMOOTHER

Estimating changepoints online is a challenging task because the presence of a change may not be obvious until after it has happened. It is thus expected that a smoothing algorithm will provide significantly improved performance at changepoint estimation. In this section, the same Rao-Blackwellisation method is used to develop a particle smoother for variable rate models with linear-Gaussian state dynamics (RBVRPS). The derivation follows a similar course to that for the fixed rate Rao-Blackwellised Particle Smoother of [15].

The particles of the final filtering step approximate the distribution, $p(\theta|y_{1:N})$, which is desired smoothing distribution. However, in the same manner as the fixed rate filter-smoother of [21], this approximation is likely to lack path-

space diversity – the particles all share the same set of changepoints from early times, with variation only appearing for changepoints closer to T. For a good characterisation of the smoothing distribution, it is necessary to rejuvenate the set of particles. This is achieved with a backward pass through the observations in a similar manner to the forward-backward methods described in [5], [15].

The objective of the smoothing algorithm will be to generate a set of particles representing $p(\theta|y_{1:N})$. Each particle will consist of a list of changepoints. The target smoothing distribution may be expanded with Bayes rule.

$$p(\theta|y_{1:N}) = p(\theta_n^+|y_{1:N})p(\theta_n|\theta_n^+, y_{1:N}) \quad (23)$$

Thus, a particle representing $p(\theta_n^+|y_{1:N})$ may be extended backwards by sampling from the backwards conditional distribution, $p(\theta_n|\theta_n^+, y_{1:N})$, which may be approximated by reweighting the particle of the nth filtering distribution. The resulting particles are then marginalised by discarding the changepoints which come before t_{n-1} (which will still be suffering from low diversity) to leave a set of samples from $p(\theta_{n-1}^+|y_{1:N})$, and the procedure continues recursively. After a complete backwards pass through the observations, a single particle will have been generated. The procedure is then repeated until sufficient samples are obtained.

If the future changepoints and their parameters, $\tilde{\theta}_n^+$, have already been sampled (and may thus be considered fixed), then the backward conditional distribution may be expressed in terms of the filtering distribution.

$$
\begin{aligned}
p(\theta_n|&\tilde{\theta}_n^+, y_{1:N}) \\
&\propto p(\theta_n, \tilde{\theta}_n^+|y_{1:N}) \\
&= \int p(x_n, \theta_n, \tilde{\theta}_n^+|y_{1:N})dx_n \\
&\propto \int p(y_{n+1:N}|x_n, \theta_n, \tilde{\theta}_n^+, y_{1:n})p(x_n, \theta_n, \tilde{\theta}_n^+|y_{1:n})dx_n \\
&= \int p(y_{n+1:N}|x_n, \tilde{\theta}_n^+)p(x_n|\theta_n, y_{1:n})dx_n \\
&\quad \times p(\tilde{\theta}_n^+|\theta_n)p(\theta_n|y_{1:n}) \quad (24)
\end{aligned}
$$

Finally, the RBVRPF approximation is substituted for the filtering distribution.

$$\hat{p}(\theta_n|\tilde{\theta}_n^+, y_{1:N}) = \sum_i \tilde{w}_n^{(i)}\delta_{\theta_n^{(i)}}(\theta_n) \quad (25)$$

where the backwards conditional weights are given by

$$
\begin{aligned}
\tilde{w}_n \propto \int &p(y_{n+1:N}|x_n, \tilde{\theta}_n^+) \\
&\times p(x_n|\theta_n^{(i)}, y_{1:n})dx_n p(\tilde{\theta}_n^+|\theta_n^{(i)}) \quad (26)
\end{aligned}
$$

As before, normalisation is enforced by scaling the weights so that they sum to 1.

The changepoint transition term $p(\tilde{\theta}_n^+|\theta_n)$ may be expressed as:

1: Run Rao-Blackwellised particle filter to approximate $p(\theta_n|y_{1:n})$ with particles $\{\theta_n^{(i)}\}$ and $\{p(x_n|\theta_n^{(i)}, y_{1:n})\}$ as normal distributions with moments $\{m_n^{(i)}\}$ and $\{P_n^{(i)}\}$. Store all results.
2: **for** $i = 1 \ldots N_S$ **do**
3:    Initialise particle using $\tilde{\theta}^{(i)} \sim \sum_j w_N^{(j)}\delta_{\theta^{(j)}}(\theta)$.
4:    **for** $n = N \ldots 1$ **do**
5:       Backwards Kalman filter: Calculate $\tilde{m}_n^{-(i)}$ and $\tilde{P}_n^{-(i)}$ using 29 to 35.
6:       **for** $j = 1 \ldots N_P$ **do**
7:          Calculate weight $\tilde{w}_n^{(j)}$ using (36).
8:       **end for**
9:       Sample $\tilde{\theta}_n^{(i)} \sim \sum_j \tilde{w}_n^{(j)}\delta_{\theta_n^{(j)}}(\theta_n)$.
10:       Discard $\tilde{\theta}_{n-1}^{(i)}$.
11:    **end for**
12: **end for**

Fig. 2. Rao-Blackwellised Variable Rate Particle Smoother

$$
\begin{aligned}
p(\tilde{\theta}_n^+|\theta_n) &= p(\tilde{\tau}_{K_n+1:K}, \tilde{u}_{K_n+1:K}|\tau_{K_n}, u_{K_n}, \tau_{K_n+1} > t_n) \\
&\propto p(\tilde{u}_{K_n+1}|\tilde{\tau}_{K_n+1}, \tau_{K_n}, u_{K_n})p(\tilde{\tau}_{K_n+1}|\tau_{K_n}). \quad (27)
\end{aligned}
$$

The state density in (26) is the familiar Kalman filter estimate, given by $p(x_n|\theta_n^{(i)}, y_{1:n}) = \mathcal{N}(x_n|m_n^{(i)}, P_n^{(i)})$. The likelihood term, $p(y_{n+1:N}|x_n, \tilde{\theta}_n^+)$, is an improper density over x_n, and may be calculated analytically using a backwards Kalman filter, in a similar manner to that used in the two-filter smoother [15], [16], [22]. Such a backwards Kalman filter uses the following recursions. Details are provided in appendix A, and in the aforesaid references.

$$p(y_{n+1:N}|x_n, \tilde{\theta}_n^+) = Z_n\mathcal{N}(x_n|\tilde{m}_n^-, \tilde{P}_n^-) \quad (28)$$

$$\tilde{m}_n^- = A_{n+1}^{-1}\tilde{m}_{n+1} \quad (29)$$
$$\tilde{P}_n^- = A_{n+1}^{-1}(\tilde{P}_{n+1} + Q_{n+1})A_{n+1}^{-T} \quad (30)$$
$$\tilde{\mu}_n = C_n\tilde{m}_n^- \quad (31)$$
$$\tilde{S}_n = C_n\tilde{P}_n^- C_n^T + R_n \quad (32)$$
$$\tilde{K}_n = \tilde{P}_n^- C_n^T \tilde{S}_n^{-1} \quad (33)$$
$$\tilde{m}_n = \tilde{m}_n^- + \tilde{K}_n(y_n - \tilde{\mu}_n) \quad (34)$$
$$\tilde{P}_n = \tilde{P}_n^- - \tilde{K}_n\tilde{S}_n\tilde{K}_n^T \quad (35)$$

Substituting into (26), the backwards conditional weights are given by:

$$\tilde{w}_n \propto p(\theta_n^+|\theta_n^{(i)})\mathcal{N}(\tilde{m}_n^-|m_n, \tilde{P}_n^- + P_n) \quad (36)$$

Using these weights, a sample of θ_n may be drawn from the particle distribution of (25). Once sampling has progressed backwards from $n = N \ldots 1$, a complete particle from the smoothing distribution will have been generated. This procedure may then be repeated until sufficient particles have been obtained. The procedure is summarised in algorithm 2.

V. SIMULATIONS

A. Finance Application

The RBVRPS algorithm was tested on the financial time series model of [11], [12], in which prices of an asset are treated as noisy observations of a latent state, which evolves according to a drift-diffusion with occasional jumps.

The latent state is a vector with two elements, the underlying value of the asset, and the trend followed by this value.

$$\mathbf{x}_n = [x_n, \dot{x}_n]^T \tag{37}$$

This evolves continuously according to a jump-diffusion model:

$$d\mathbf{x}_t = \begin{bmatrix} 0 & 1 \\ 0 & -\lambda \end{bmatrix} \mathbf{x}_t dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} d\beta(t) + d\mathbf{J}_t \tag{38}$$

where λ introduces a mean regression effect on the trend and $\beta(t)$ is standard Brownian motion (with unit diffusion constant). The jump term, $d\mathbf{J}_t$, is zero everywhere except where jumps occur.

$$d\mathbf{J}_t = \begin{cases} \mathbf{J}_k & t \in \{\tau_k\} \\ 0 & \text{elsewhere} \end{cases} \tag{39}$$

$$\mathbf{J}_k \sim \mathcal{N}(\mathbf{J}_k | \mathbf{0}, Q_{J,u_k}) \tag{40}$$

Jumps occur at a random set of times, $\{\tau_k\}$, and are one of two types: value jumps, indicated by $u_k = 1$, and trend jumps, indicated by $u_k = 2$.

The jump covariance matrices are,

$$Q_{J,u_k} = \begin{cases} \begin{bmatrix} \sigma_{J1}^2 & 0 \\ 0 & 0 \end{bmatrix} & u_k = 1 \\ \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{J2}^2 \end{bmatrix} & u_k = 2 \end{cases} . \tag{41}$$

This model may be discretised at the observation times using standard methods (see e.g. [23], [24]*Not sure what to cite here. Simo's thesis? Or the references therein which I've never seen?*). Assuming observations of value only and Gaussian observation noise with standard deviation σ_y^2, the resulting discrete time dynamics are described by the following equations (see appendix **??**):

$$\mathbf{x}_n = A\mathbf{x}_{n-1} + \mathbf{w}_n \tag{42}$$

$$y_n = C\mathbf{x}_n + v_n \tag{43}$$

where the $\mathbf{w}_b$ and v_n are Gaussian random variables with covariance matrixes Q_n and R respectively. The time between observations times is denotes $\Delta t = t_n - t_{n-1}$.

$$A = \begin{bmatrix} 1 & \frac{1}{\lambda}(1 - e^{(-\lambda \Delta t)}) \\ 0 & e^{(-\lambda \Delta t)} \end{bmatrix} \tag{44}$$
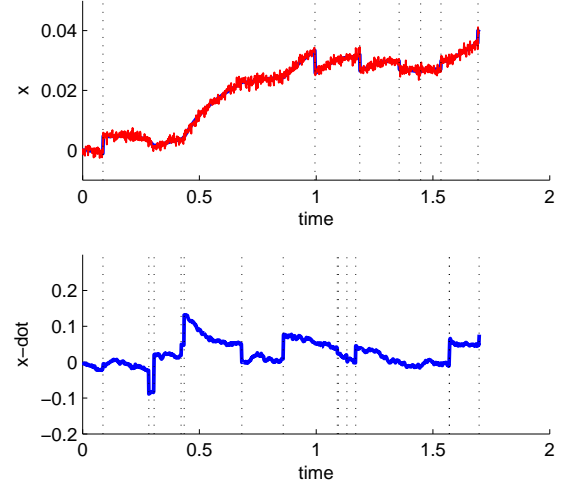
$$C = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{45}$$

Fig. 3. An example simulated data set, showing value (top) and trend (bottom). Value observations are overlaid. Jump times are shown as dotted vertical lines.

$$Q_n = \begin{cases} Q_D + Q_{J,u_k} & \exists k : \tau_k \in [t_{n-1}, t_n] \\ Q_D & \text{otherwise} \end{cases} \tag{46}$$

$$Q_D = \frac{\sigma^2}{2\lambda} \begin{bmatrix} q_1 & q_2 \\ q_2 & q_3 \end{bmatrix} \tag{47}$$

$$q_1 = \frac{1}{\lambda^2}(2\lambda \Delta t - (3 - e^{(-\lambda \Delta t)})(1 - e^{(-\lambda \Delta t)}) \tag{48}$$

$$q_2 = \frac{1}{\lambda}(1 - e^{(-\lambda \Delta t)})^2 \tag{49}$$

$$q_3 = 1 - e^{(-2\lambda \Delta t)} \tag{50}$$

$$R = [\sigma_y^2] \tag{51}$$

The times between changepoints were assumed to be exponentially distributed, with equal probability of value and trend jumps.

$$p(\tau_k, u_k | \tau_{k-1}, u_{k-1}) = P(u_k)p(\tau_k | \tau_{k-1}) \tag{52}$$

$$p(\tau_k | \tau_{k-1}) = \begin{cases} \frac{1}{\alpha}\exp(\alpha(\tau_k - \tau_{k-1})) & \tau_k > \tau_{k-1} \\ 0 & \tau_k < \tau_{k-1} \end{cases} \tag{53}$$

$$P(u_k) = \begin{cases} 0.5 & u_k = 1 \\ 0.5 & u_k = 2 \end{cases} \tag{54}$$

The algorithms were first tested on artificial data simulated from this model. The following parameters were used: $\Delta t = 0.0017$, $N = 1000$, $\alpha = 20$, $\lambda = 5$, $\sigma = 0.05$, $\sigma_{J1} = 0.005$, $\sigma_{J2} = 0.05$, $\sigma_y = 0.001$.

The filter used $N_P = 100$ particles, and the smoother resampled $N_S = 100$ sequences. Bootstrap proposals were used for the filter.

An example realisation simulated from the model is shown in figure 3.

Both the RBVRPS and the final processing step of the RBVRPF produce a particle approximation to the changepoint sequence posterior, $p(\theta|y_{1:N})$. It is actually no simple task to compare the quality of these approximations. Each particle
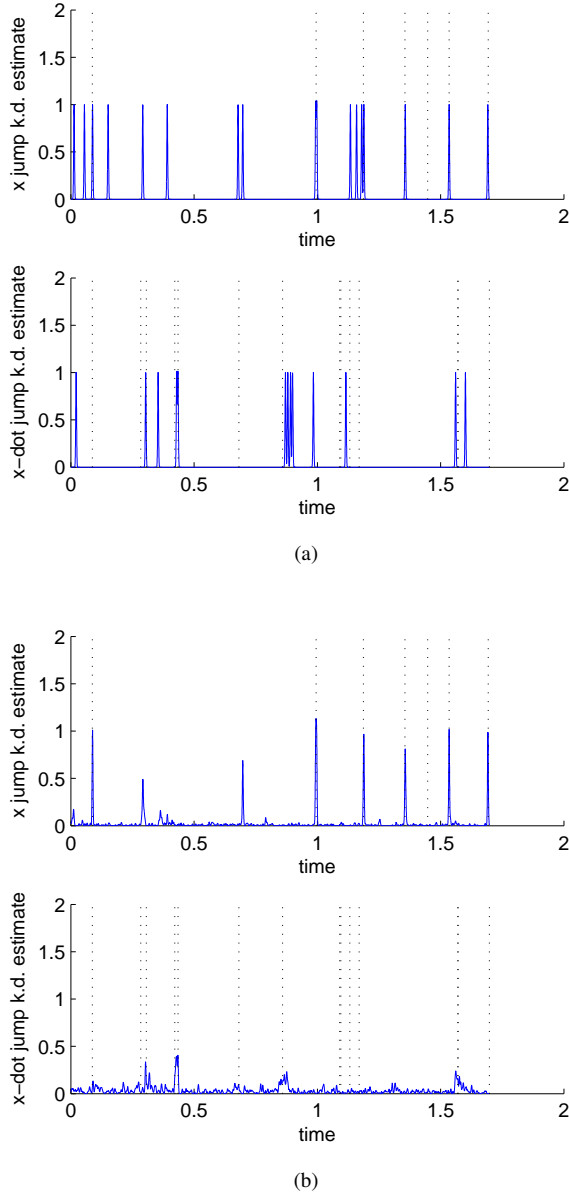
Fig. 4. Filter (a) and smoother (b) kernel density estimates for value (top) and trend (bottom) jump times. Correct times overlayed as dotted vertical lines.

Fig. 5. Trend estimates using RBVRPF + Kalman filter (a) RBVRPF + RTS smoother (b) and RBVRPS + RTS smoother. Solid lines show means of each particle. Dashed lines show mean ± 2 standard deviations.

consists of a list of changepoint times and types, and there is no obvious way to take an average over the set to produce a single estimate. Furthermore, once a best estimate has been calculated, it is not obvious how the error should be assessed, as it is not trivial to define a distance metric between lists of varying lengths.

For a first, qualitative comparison of the changepoint sequence distributions, it is possible to generate kernel density estimates for the changepoint times. These are constructed by simply adding a (unit amplitude) Gaussian kernel for each changepoint present in each kernel and scaling by the number of particles. The resulting function provides an approximate measure of the likelihood of finding a jump at a given time. These are shown in figure 4 for the example data set.

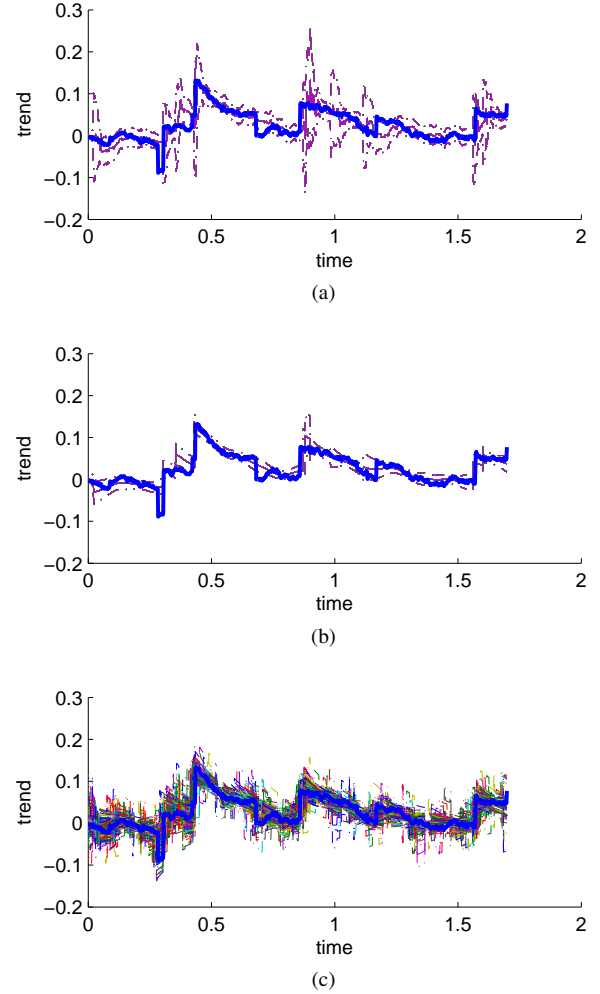It is immediately apparent that the approximation produced

by the filter lacks diversity. The jump time kernel density estimate comprises peaks with magnitude 1, with regions of 0 between. This is a result of every particle containing an almost identical list of changepoints, as a result of resampling. In contrast, the smoother is clearly providing a less degenerate particle representation, and generally contains larger kernel density peaks at the times of larger jumps. Furthermore, the kernel density for the trend jump times contains broader peaks, capturing the fact that because the trend is not observed directly, the jumps are harder to localise precisely.

A further comparison of the algorithms is possible via the state estimates they generate. Here we can compare three options: the filtering results, using the RBVRPF and a Kalman filter for the state estimates; the filter-smoother results, using the final RBVRPF approximation for the changepoint sequence and a Rauch-Tung-Striebel (RTS) smoother for the state estimates; and the full smoothing results, using the RBVRPS followed by an RTS smoother. For the example data set, the trend estimates are shown in figure 5 (The estimates of value are less informative, as this quantity is observed). Again, the improved particle diversity of the smoother is apparent.

For a quantitative comparison, the algorithms were tested

TABLE I
CHANGEPOINT SEQUENCE ESTIMATION PERFORMANCE.

	RBVRPF	RBVRPS
mean unique no. sequences	7.8	100
mean unique no. jump times	47.8	1076.7

TABLE II
STATE ESTIMATION PERFORMANCE.

	RBVRPF and KF	RBVRPF and RTS	RBVRPS and RTS
mean value estimate RMSE	5.31×10^{-4}	4.48×10^{-4}	4.16×10^{-4}
mean trend estimate RMSE	2.56×10^{-2}	1.79×10^{-2}	1.49×10^{-2}

on 10 realisations from the model, each of 1000 time steps, and the following statistics were calculated for each:

- The number of unique changepoint sequences. This is a measure of particle diversity of the approximation.
- The number of unique changepoint times. Another measure of particle diversity.
- The root-mean-square error (RMSE) of the mean state estimate. The mean state estimate is the average of the Gaussian means from all the particles.

The results are shown in tables I and II.

The new RBVRPS algorithm outperforms the filter in terms of both state estimation accuracy and particle diversity.

Finally, we demonstrate the algorithms on a real financial data set, 1000 data points representing 1.7s of USD-GBP foreign exchange rate data from April 2008, using the same parameters as for the previous simulations. The data shown in Fig. 6, along with jump time kernel density estimates from the filter and smoother. In this example, there is no ground truth against which to judge the results. However, the smoother does a pleasing job of estimating jump times in the same areas that one would if analysing the data "by eye". We submit this simply to illustrate the practical applications of the model and algorithm.

VI. CONCLUSION

A new smoothing algorithm has been introduced for variable rate models which have linear-Gaussian dynamics conditional on a set of unknown changepoints. The algorithm employs Rao-Blackwellisation, using particle methods to estimate the nonlinear changepoint sequence, and Kalman filtering/smoothing to estimate the linear state components. The smoothing algorithm has been shown to improve particle diversity and accuracy when compared to the filter.

APPENDIX A
DERIVATION OF THE BACKWARDS KALMAN FILTER

The RBVRPS weights require the calculation of the "backwards filter" term, $p(y_{n+1:N}|x_n, \tilde{\theta}_n^+)$, which may considered as an unnormalised density. For linear-Gaussian state dynamics, this may be calculated using the backwards Kalman filter
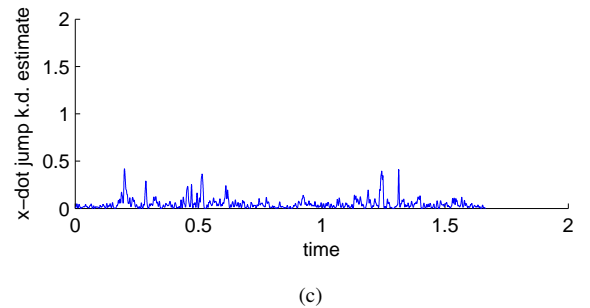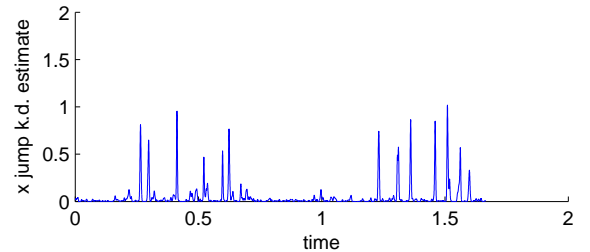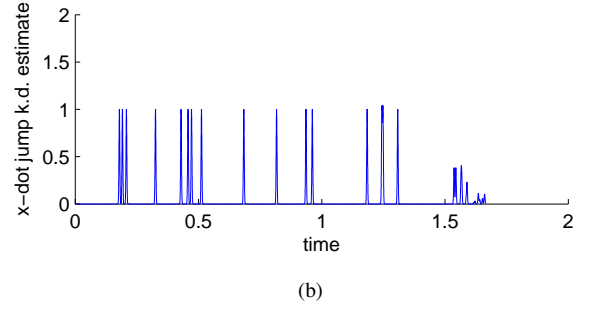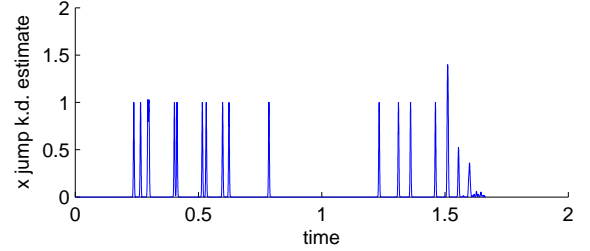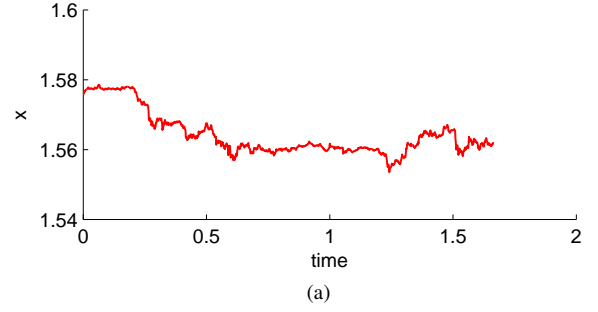

Fig. 6. Foreign exchange data (a), along with filter (a) and smoother (b) kernel density estimates for the state and trend jump times.

recursions [15], [22]. Assume that at time t_{n+1} there is a Gaussian density such that,

$$p(y_{n+1:N}|x_{n+1}, \tilde{\theta}_n^+) = \tilde{Z}_{n+1}\mathcal{N}(x_{n+1}|\tilde{m}_{n+1}, \tilde{P}_{n+1}). \quad (55)$$

The state transition model is linear-Gaussian, so the backwards predictive density is given by,

$$
\begin{aligned}
& p(y_{n+1:N}|x_n, \tilde{\theta}_n^+) \\
&= \int p(y_{n+1:N}|x_{n+1}, \tilde{\theta}_n^+) p(x_{n+1}|x_n, \tilde{\theta}_n^+) dx_{n+1} \\
&= \int \tilde{Z}_{n+1} \mathcal{N}(x_{n+1}|\tilde{m}_{n+1}, \tilde{P}_{n+1}) \\
& \qquad \mathcal{N}(x_{n+1}|A_{n+1}x_n, Q_{n+1}) dx_{n+1} \\
&= \int \tilde{Z}_{n+1} ||A_{n+1}^{-1}|| \mathcal{N}(x_{n+1}|\tilde{m}_{n+1}, \tilde{P}_{n+1}) \\
& \qquad \mathcal{N}(x_n|A_{n+1}^{-1}x_{n+1}, A_{n+1}^{-1}Q_{n+1}A_{n+1}^{-T}) dx_{n+1} \\
&= \tilde{Z}_n^- \mathcal{N}(x_n|\tilde{m}_n^-, \tilde{P}_n^-)
\end{aligned}
\tag{56}
$$

where,

$$
\tilde{m}_n^- = A_{n+1}^{-1} \tilde{m}_{n+1} \tag{57}
$$

$$
\tilde{P}_n^- = A_{n+1}^{-1}(\tilde{P}_{n+1} + Q_{n+1})A_{n+1}^{-T} \tag{58}
$$

$$
\tilde{Z}_n^- = \tilde{Z}_{n+1}||A_{n+1}||^{-1}, \tag{59}
$$

where $||A||$ is used to denote the magnitude of the determinant of A.

The update step simply involves multiplying by the likelihood of the observation at time t_n.

$$
\begin{aligned}
& p(y_{n:N}|x_n, \tilde{\theta}_n^+) \\
&= p(y_{n+1:N}|x_n, \tilde{\theta}_n^+) p(y_n|x_n, \tilde{\theta}_n^+) \\
&= \tilde{Z}_n^- \mathcal{N}(x_n|\tilde{m}_n^-, \tilde{P}_n^-) \mathcal{N}(y_n|C_n x_n, R_n) \\
&= \tilde{Z}_n \mathcal{N}(x_n|\tilde{m}_n, \tilde{P}_n)
\end{aligned}
\tag{60}
$$

$$
\tag{61}
$$

where,

$$
\tilde{\mu}_n = C_n \tilde{m}_n^- \tag{62}
$$

$$
\tilde{S}_n = C_n \tilde{P}_n^- C_n^T + R_n \tag{63}
$$

$$
\tilde{K}_n = \tilde{P}_n^- C_n^T \tilde{S}_n^{-1} \tag{64}
$$

$$
\tilde{m}_n = \tilde{m}_n^- + \tilde{K}_n(y_n - \tilde{\mu}_n) \tag{65}
$$

$$
\tilde{P}_n = \tilde{P}_n^- - \tilde{K}_n \tilde{S}_n \tilde{K}_n^T \tag{66}
$$

$$
\tilde{Z}_n = \tilde{Z}_n^- \mathcal{N}(y_n|\tilde{\mu}_n, \tilde{S}_n). \tag{67}
$$

All steps use only standard Gaussian identities.

The backwards filter is initialised using the likelihood of the final step, which is given by the observation model.

$$
p(y_N|x_N) = \mathcal{N}(y_N|C_N x_N, R_N) \tag{68}
$$

If the state is not fully observed, i.e. the dimensionality of the observations is less than that of the state, then it will be necessary to use observations from several steps for the initialisation.

ACKNOWLEDGMENT

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal Of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.6247&rep=rep1&type=pdf

[2] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, 1965.

[3] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2, pp. 107 –113, April 1993.

[4] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000, 10.1023/A:1008935410038. [Online]. Available: http://dx.doi.org/10.1023/A:1008935410038

[5] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/016214504000000151

[6] O. Cappé, S. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[7] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *The Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.157.772&rep=rep1&type=pdf

[8] M. Jacobsen, *Point process theory and applications: marked point and piecewise deterministic processes*, C. Newman and S. Resnick, Eds. Birkhauser, 2006.

[9] S. Godsill, J. Vermaak, W. Ng, and J. Li, "Models and algorithms for tracking of maneuvering objects using variable rate particle filters," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 925 –952, May 2007.

[10] N. Whiteley, A. M. Johansen, and S. Godsill, "Monte carlo filtering of piecewise deterministic processes," *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, pp. 119–139, 2011. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/jcgs.2009.08052

[11] S. Godsill, "Particle filters for continuous-time jump models in tracking applications," *ESAIM: Proc.*, vol. 19, pp. 39–52, 2007. [Online]. Available: http://dx.doi.org/10.1051/proc:071907

[12] H. L. Christensen, J. Murphy, and S. J. Godsill, "Forecasting high-frequency futures returns using online langevin dynamics," *submitted to IEEE journal of selected topics in signal processing*, vol. under review, p. under review, 2012.

[13] G. Casella and C. P. Robert, "Rao-blackwellisation of sampling schemes," *Biometrika*, vol. 83, no. 1, pp. 81–94, 1996. [Online]. Available: http://biomet.oxfordjournals.org/content/83/1/81.abstract

[14] A. Doucet, N. De Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. Citeseer, 2000, pp. 176–183.

[15] S. Sarkka, P. Bunch, and S. Godsill, "A backward-simulation based rao-blackwellized particle smoother for conditionally linear gaussian models," in *Preprint submitted to 16th IFAC Symposium on System Identification.*, 2012.

[16] B. Anderson and J. Moore, *Optimal filtering*. Prentice-Hall (Englewood Cliffs, NJ), 1979.

[17] W. R. Gilks and C. Berzuini, "Following a moving target-Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001. [Online]. Available: http://dx.doi.org/10.1111/1467-9868.00280

[18] P. Green, "Reversible jump Markov chain Monte Carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, p. 711, 1995.

[19] K. Gilholm and N. O. Everett, "A novel algorithm for tracking targets using manoeuvre models," in *Target Tracking and Data Fusion: Algorithms and Applications, 2008 IET Seminar on*, april 2008, pp. 143 –148.

[20] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[21] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. pp. 1–25, 1996. [Online]. Available: http://www.jstor.org/stable/1390750

[22] D. Fraser and J. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *Automatic Control, IEEE Transactions on*, vol. 14, no. 4, pp. 387 – 390, aug 1969.

[23] B. Øksendal, *Stochastic differential equations: an introduction with applications*. Springer Verlag, 2003.

[24] S. Sarkka, "Recursive bayesian inference on stochastic differential equations," Ph.D. dissertation, Helsinki University of Technology Laboratory of Computational Engineering, 2006.

Pete Bunch Biography text here.

Simon Godsill Biography text here.