Monte Carlo Smoothing for Nonlinear Time Series

Simon J. GODSILL, Arnaud DOUCET, and Mike WEST

We develop methods for performing smoothing computations in general state-space models. The methods rely on a particle representation of the filtering distributions, and their evolution through time using sequential importance sampling and resampling ideas. In particular, novel techniques are presented for generation of sample realizations of historical state sequences. This is carried out in a forward-filtering backward-smoothing procedure that can be viewed as the nonlinear, non-Gaussian counterpart of standard Kalman filter-based simulation smoothers in the linear Gaussian case. Convergence in the mean squared error sense of the smoothed trajectories is proved, showing the validity of our proposed method. The methods are tested in a substantial application for the processing of speech signals represented by a time-varying autoregression and parameterized in terms of time-varying partial correlation coefficients, comparing the results of our algorithm with those from a simple smoother based on the filtered trajectories.

KEY WORDS: Bayesian inference; Non-Gaussian time series; Nonlinear time series; Particle filter; Sequential Monte Carlo; State-space model.

1. INTRODUCTION

In this article we develop Monte Carlo methods for smoothing in general state-space models. To fix notation, consider the standard Markovian state-space model (West and Harrison 1997)

$$x_{t+1} \sim f(x_{t+1}|x_t) \qquad \text{(state evolution density)},$$
$$y_{t+1} \sim g(y_{t+1}|x_{t+1}) \qquad \text{(observation density)},$$

where $\{x_t\}$ are unobserved states of the system, $\{y_t\}$ are observations made over some time interval $t \in \{1, 2, \ldots, T\}$, and $f(\cdot|\cdot)$ and $g(\cdot|\cdot)$ are prespecified state evolution and observation densities, which may be non-Gaussian and involve nonlinearity. It is assumed throughout that the required distributions can be represented by density functions, and that both $f(\cdot|\cdot)$ and $g(\cdot|\cdot)$ can be evaluated for any valid states and observations x_t and y_t; x_t and y_t may both be vectors. We assume that the process $\{x_t\}$ is Markov, generated according to the foregoing state evolution, and that the observation process $\{y_t\}$ is independent conditional on the state process $\{x_t\}$. Hence an expression for the joint distribution of states and observations can be obtained directly by the probability chain rule,

$$p(x_{1:t}, y_{1:t}) = f(x_1)\left(\prod_{i=2}^{t} f(x_i|x_{i-1})\right)\left(\prod_{i=1}^{t} g(y_i|x_i)\right),$$

where $f(x_1)$ is the distribution of the initial state. Here $x_{1:t} = (x_1, \ldots, x_t)$ and $y_{1:t} = (y_1, \ldots, y_t)$ denote collections of observations and states from time 1 through t. In proving the validity of our proposed smoothing algorithm, a more formal definition of the state-space model is needed; we present this in Appendix A.

A primary concern in many state-space inference problems is sequential estimation of the filtering distribution $p(x_t|y_{1:t})$. Updating the filtering density can be done in principle using the standard filtering recursions

$$p(x_{t+1}|y_{1:t}) = \int p(x_t|y_{1:t}) f(x_{t+1}|x_t)\, dx_t$$

and

$$p(x_{t+1}|y_{1:t+1}) = \frac{g(y_{t+1}|x_{t+1})\, p(x_{t+1}|y_{1:t})}{p(y_{t+1}|y_{1:t})}.$$

Similarly, smoothing can be performed recursively backward in time using the smoothing formula

$$p(x_t|y_{1:T}) = \int p(x_{t+1}|y_{1:T}) \frac{p(x_t|y_{1:t}) f(x_{t+1}|x_t)}{p(x_{t+1}|y_{1:t})}\, dx_{t+1}.$$

Inference in general state-space models has been revolutionized over the past decade by the introduction of cheap and massive computational resources and the consequent development and widespread application of Monte Carlo methods. In batch-based scenarios, Markov chain Monte Carlo (MCMC) methods have been widely used, and various powerful tools have been developed and proven in application (see, e.g., Carlin, Polson, and Stoffer 1992; Carter and Kohn 1994; Shephard 1994; Shephard and Pitt 1997; De Jong 1997; Aguilar, Huerta, Prado, and West 1999, Aguilar and West 1998, 2000; Pitt and Shephard 1999b). However, constructing an effective MCMC sampler in models with significant degrees of nonlinearity and non-Gaussianity is not always straightforward. Specifically, in these cases it can be hard to construct effective proposal distributions, either over collections of states simultaneously or even for single states conditional on all others. The danger then is that the MCMC will be slowly mixing and may never converge to the target distribution within realistic time scales.

Alternative Monte Carlo strategies based on sequential importance sampling, known generically as *particle filters*, have been rapidly emerging in areas such as target tracking for radar, communications, econometrics, and computer vision (West 1993; Gordon, Salmond, and Smith 1993; Kitagawa 1996; Liu and Chen 1998; Doucet, Godsill, and Andrieu 2000; Liu and West 2001; Pitt and Shephard 1999a; West and Harrison 1997; Doucet, De Freitas, and Gordon 2001). These methods allow propagation of completely general target filtering distributions through time using combinations of importance sampling, resampling and local MCMC moves. The methods have been proven for many examples, including highly nonlinear models that are not easily implemented using standard MCMC.

Simon J. Godsill is Reader in Statistical Signal Processing (E-mail: *sjg@eng.cam.ac.uk*) and Arnaud Doucet is University Lecturer (E-mail: *ad2@eng.cam.ac.uk*), Signal Processing Group, University of Cambridge, Cambridge CB2 1PZ, U.K.. Mike West is the Arts and Sciences Professor of Statistics and Decisions Sciences, Institute of Statistics and Decision Sciences, Duke University, Durham NC 27708-0251 (E-mail: *mw@isds.duke.edu*). This work was supported in part by grants from the Engineering and Physical Sciences Research Council and the National Science Foundation.

Journal of the American Statistical Association
March 2004, Vol. 99, No. 465, Theory and Methods
DOI 10.1198/016214504000000151

In particle filtering methods, the filtering density is approximated with an empirical distribution formed from point masses, or *particles*,

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{x_t^{(i)}}(x_t), \qquad \sum_{i=1}^{N} w_t^{(i)} = 1,\ w_t^{(i)} \geq 0, \tag{1}$$

where δ is the Dirac delta function and $w_t^{(i)}$ is a weight attached to particle $x_t^{(i)}$. Particles at time t can be updated efficiently to particles at time $t+1$ using importance sampling and resampling methods.

The theory of particle filtering is now quite well developed. For example, the empirical measure of (1) converges almost surely to the distribution associated with $p(x_t|y_{1:t})$ for all $t > 0$ as $N \to \infty$ under quite mild conditions on the state-space model. Moreover, rates of convergence to 0 have been established for expectations of mean squared error of functionals with respect to this filtering density. Hence particle filters are rigorously validated as a means for tracking the distribution of, and estimating the value of a hidden state over time. Some recent advances in convergence analysis have been given by Del Moral (1998), Crisan, Del Moral, and Lyons (1999), Crisan and Lyons (1999), Crisan and Doucet (2000), and Crisan (2001).

Although particle filtering theory and practice are now quite well established, smoothing aspects are less well established. Existing approaches to smoothing with particle filters have been aimed at approximating the individual *marginal* smoothing densities $p(x_t|y_{1:T})$, using either the two-filter formula (Kitagawa 1996) or forward filtering–backward smoothing (Doucet et al. 2000; Hürzeler and Künsch 1998). In many applications these marginal distributions are of limited interest, because investigations of historical states generally focus on *trajectories* and hence require consideration of collections of states together. For cases where a single "best" estimate for the smoothed trajectory is required, Godsill, Doucet, and West (2001) provided a sequential methodology for maximum a posteriori (MAP) sequence estimation based on dynamic programming and the Viterbi algorithm. However, a single best estimate is rarely appropriate in the Bayesian inference setting, especially when distributions are multimodal, and here we aim for random generation of state sequences.

The new methods provide a completion of particle filtering methodology that allows random generation of entire historical trajectories drawn from the joint smoothing density $p(x_{1:t}|y_{1:t})$. The method relies first on a forward-filtering pass that generates and stores a particle-based approximation to the filtering density at each time step. Then a backward "simulation smoothing" pass is carried out to generate sampled realizations from the smoothing density. The method can be seen as the nonlinear/non-Gaussian analog of the forward-filtering/backward-sampling algorithms developed for linear Gaussian models and hidden Markov models (Carter and Kohn 1994; Frühwirth-Schnatter 1994; De Jong and Shephard 1995). See also Hürzeler and Künsch (2001) for some recent relevant work in the particle filter setting. The proposed method is quite distinct from the MAP estimation procedure of Godsill, Doucet, and West (2001) in which the forward particle filter is used simply to generate a grid of possible state values at each time point, with the Viterbi algorithm used to trace out the most probable state trajectory through that grid of state values.

The article is organized as follows. Section 2 describes the basic particle filtering and smoothing framework. Section 3 introduces the proposed simulation smoother algorithm for general state-space models, with a proof of convergence for the new simulation smoother method deferred to the Appendix. Section 4 evaluates the method for a standard nonlinear model and Section 5 does so with an extensive application to speech data analysis. Finally, Section 6 closes the article with some discussion.

2. FILTERING AND SMOOTHING USING SEQUENTIAL IMPORTANCE SAMPLING

In this section we review the standard procedure for filtering and smoothing using sequential importance sampling. In practice this is found to be highly effective in the filtering mode, but, as demonstrated in our simulations, it can give very poor results in the smoothing mode. (See, e.g., Doucet et al. 2000, 2001, for a detailed overview of these standard methods.) A more formal description of the particle filter is given in the Appendix, including a statement of the theorem required to prove the smoother of the next section.

Suppose that we have at time t weighted particles $\{x_{1:t}^{(i)}, w_t^{(i)}; i = 1, 2, \ldots, N\}$ drawn from the smoothing density $p(x_{1:t}|y_{1:t})$. We can consider this an empirical approximation for the density made up of point masses,

$$p(x_{1:t}|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{x_{1:t}^{(i)}}(x_{1:t}), \qquad \sum_{i=1}^{N} w_t^{(i)} = 1,\ w_t^{(i)} \geq 0. \tag{2}$$

To update the smoothing density from time t to time $t+1$, factorize it as

$$p(x_{1:t+1}|y_{1:t+1}) = p(x_{1:t}|y_{1:t}) \times \frac{g(y_{t+1}|x_{t+1}) f(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})},$$

where the denominator is constant for a given dataset. To proceed from time t to $t+1$, we select trajectories from the approximation (2). In the simplest case—[the "bootstrap" filter (Gordon et al. 1993; Kitagawa 1996)]—N trajectories are drawn at random with replacement from $\{x_{1:t}^{(i)}; i = 1, 2, \ldots, N\}$ with probabilities $\{w_t^{(i)}\ i = 1, 2, \ldots, N\}$. In more sophisticated schemes, some part-deterministic variance reduction selection scheme is applied (Kitagawa 1996; Liu and Chen 1998; Doucet et al. 2000; Carpenter, Clifford, and Fearnhead 1999). A new state is then generated randomly from an importance distribution, $q(x_{t+1}|x_t, y_{t+1})$, and appended to the corresponding trajectory, x_t. The importance weight is updated to

$$w_{t+1} \propto \frac{g(y_{t+1}|x_{t+1}) f(x_{t+1}|x_t)}{q(x_{t+1}|x_t, y_{t+1})}.$$

Other selection schemes aim to improve performance at future time points by introducing a bias into the selection step. In these cases an additional term is included in the weight expression to allow for this bias. Examples of this include the most basic sequential imputation procedures of Liu and Chen (1995), where selection is carried out only rarely and weights are updated incrementally throughout the filtering pass, and the auxiliary particle filtering approach of Pitt and Shephard (1999a), in which a bias is intentionally introduced with the aim of boosting the number of particles in useful regions of the state space. (For further discussion of these issues, see Godsill and Clapp 2001.)

We call the smoothing methods described in this section the "standard trajectory-based smoothing" method. Filtering is obtained as a simple corollary of the smoothing technique by discarding the past trajectories $x_{1:t}$ once the update has been made to $t+1$. It is clear that the selection (or "resampling") procedure will lead to high levels of degeneracy in smoothed trajectories using this method. This is demonstrated in later simulations and motivates the development of novel smoothing methods based on backward simulation in the next section.

3. SMOOTHING USING BACKWARD SIMULATION

The new method proposed here assumes that Bayesian filtering has already been performed on the entire dataset, leading to an approximate representation of $p(x_t|y_{1:t})$ for each time step $t \in \{1, \dots, T\}$, consisting of weighted particles $\{x_t^{(i)}, w_t^{(i)}; i = 1, 2, \dots, N\}$. We note that the method is independent of the precise filtering algorithm and that any particle filtering scheme, whether deterministic or Monte Carlo, can be used. Recall that the primary goal here is to obtain sample realizations from the entire smoothing density to exemplify and generate insight into the structure of the smoothing distribution for collections of past states together. This can be based on the factorization

$$p(x_{1:T}|y_{1:T}) = p(x_T|y_{1:T}) \prod_{t=1}^{T-1} p(x_t|x_{t+1:T}, y_{1:T}), \tag{3}$$

where, using the Markovian assumptions of the model, we can write

$$\begin{aligned} p(x_t|x_{t+1:T}, y_{1:T}) &= p(x_t|x_{t+1}, y_{1:t}) \\ &= \frac{p(x_t|y_{1:t}) f(x_{t+1}|x_t)}{p(x_{t+1}|y_{1:t})} \\ &\propto p(x_t|y_{1:t}) f(x_{t+1}|x_t). \end{aligned} \tag{4}$$

Forward filtering generates a particulate approximation to $p(x_t|y_{1:t})$. Because we have from the foregoing that $p(x_t|x_{t+1}, y_{1:T}) \propto p(x_t|y_{1:t}) f(x_{t+1}|x_t)$, we immediately obtain the modified particle approximation

$$p(x_t|x_{t+1}, y_{1:T}) \approx \sum_{i=1}^{N} w_{t|t+1}^{(i)} \delta_{x_t^{(i)}}(x_t),$$

with modified weights

$$w_{t|t+1}^{(i)} = \frac{w_t^{(i)} f(x_{t+1}|x_t^{(i)})}{\sum_{j=1}^{N} w_t^{(j)} f(x_{t+1}|x_t^{(j)})}. \tag{5}$$

This revised particulate distribution can now be used to generate states successively in the reverse-time direction, conditioning on future states. Specifically, given a random sample $\widetilde{x}_{t+1:T}$ drawn approximately from $p(x_{t+1:T}|y_{1:T})$, take one step back in time and sample $\widetilde{x}_t$ from $p(x_t|\widetilde{x}_{t+1}, y_{1:T})$. The pair $(\widetilde{x}_t, \widetilde{x}_{t+1:T})$ is then approximately a random realization from $p(x_{t:T}|y_{1:T})$. Repeating this process sequentially back over time produces the following general "smoother-realization" algorithm:

Algorithm 1 (Sample realizations).

1. Choose $\widetilde{x}_T = x_T^{(i)}$ with probability $w_T^{(i)}$.
2. For $t = T-1$ to 1:
   - Calculate $w_{t|t+1}^{(i)} \propto w_t^{(i)} f(\widetilde{x}_{t+1}|x_t^{(i)})$ for each $i = 1, \dots, N$.
   - Choose $\widetilde{x}_t = x_t^{(i)}$ with probability $w_{t|t+1}^{(i)}$.
3. $\widetilde{\mathbf{x}}_{1:T} = (\widetilde{x}_1, \widetilde{x}_2, \dots, \widetilde{x}_T)$ is an approximate realization from $p(x_{1:T}|y_{1:T})$.

Further independent realizations are obtained by repeating this procedure as many times as needed. The computational complexity for each realization is $O(NT)$, in contrast with the $O(N^2T)$ required for marginal smoothing procedures (Kitagawa 1996; Doucet et al. 2000; Hürzeler and Künsch 1998); however, it should be realized that the computations in our method are then repeated for each realization drawn.

In the Appendix we prove the convergence of the smoothed realizations in terms of mean squared error for state estimation as the number of particles tends to infinity.

4. EXAMPLE 1: A NONLINEAR TIME SERIES MODEL

The new methods are first demonstrated for a standard nonlinear time series model (Kitagawa 1996; West 1993; Gordon et al. 1993). This model has been used extensively for testing of numerical filtering techniques, and here we use it to show the functionality and extended utility of the proposed smoother compared with the other available techniques (Kitagawa 1996; Doucet et al. 2000; Hürzeler and Künsch 1998).

The state-space equations are

$$x_t = \frac{x_{t-1}}{2} + 25\frac{x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + v_t$$

and

$$y_t = \frac{(x_t)^2}{20} + w_t,$$

where $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $w_t \sim \mathcal{N}(0, \sigma_w^2)$, and here $\sigma_v^2 = 10$ and $\sigma_w^2 = 1$ are considered fixed and known. The initial state distribution is $x_1 \sim \mathcal{N}(0, 10)$. The representation in terms of densities $f(x_t|x_{t-1})$ and $g(y_t|x_t)$ is straightforward.

A typical dataset simulated from this model is shown in Figure 1. Filtering is performed using a standard bootstrap par-

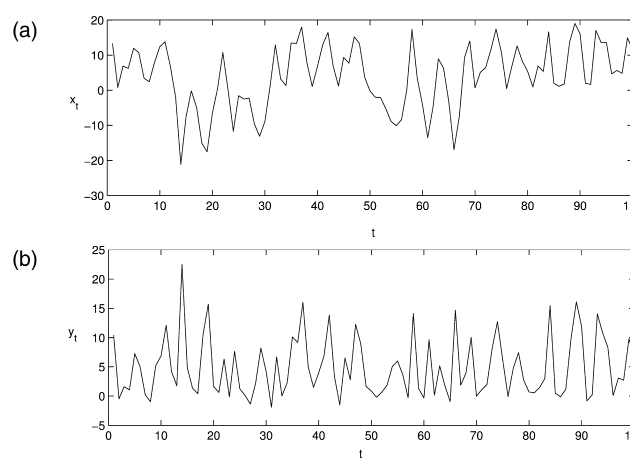

Figure 1. Simulated Data From the Nonlinear Time Series Model. (a) Hidden state sequence x_t; (b) observed data sequence y_t.

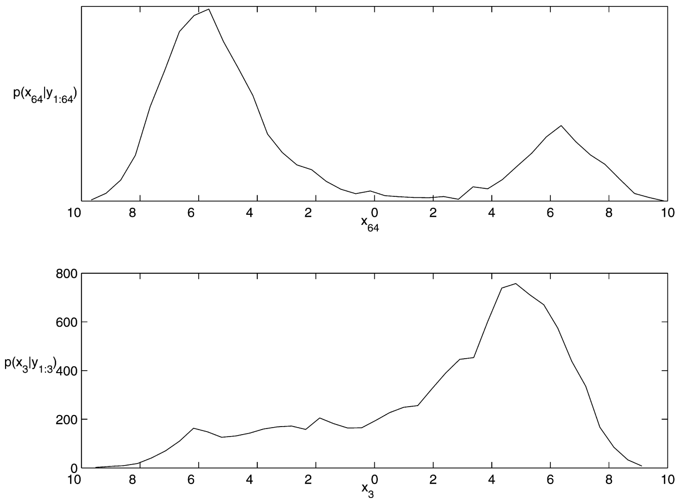

Figure 2. Filtering Density Estimates From the Particle Filter Output.

ticle filter (Gordon et al. 1993) with $N = 10{,}000$ particles. Applying filtering to the same dataset as in Figure 1 shows clear evidence for strong non-Gaussianity and multimodality in the filtering distributions; see, for example, the density estimates obtained from the particle filter output at times $t = 64$ and $t = 3$ in Figure 2.

Smoothing is carried out using the proposed smoother, drawing 10,000 realizations from the smoothing density. A small random selection of the smoothed trajectories drawn from $p(x_{1:100}|y_{1:100})$ is shown in Figure 3. Now multimodality in the smoothing distribution can be seen, with separated clusters of trajectories visible in several parts of the time series. Histogram estimates of the individual smoothing distributions $p(x_t|y_{1:100})$ are shown in Figure 4. The sampled realizations are useful in themselves, but they can also be used to study and visualize in more detail the characteristics of the multivariate smoothing distribution. Note that this is a capability well beyond that of Kitagawa (1996), Doucet et al. (2000), and Hürzeler and Künsch (1998), which generate only the smoothed *marginals*, $p(x_t|y_{1:T})$. Figures 5–8 show a selection of bi-

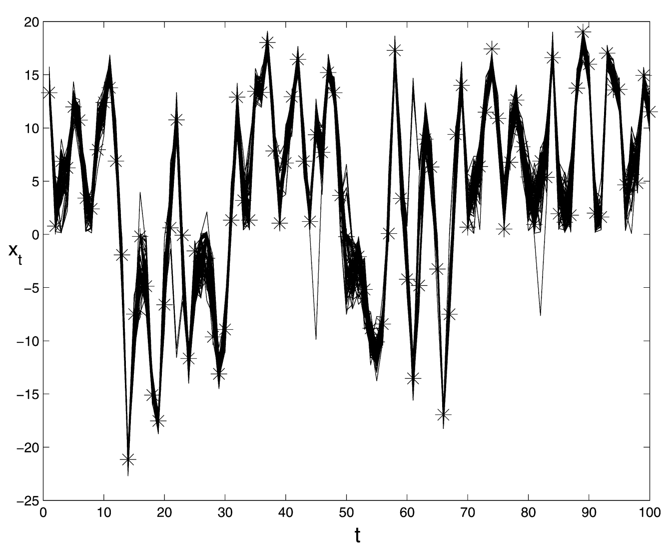

Figure 3. Smoothing Trajectories Drawn From $p(x_{1:100}|y_{1:100})$. The true simulated states shown as "$*$."

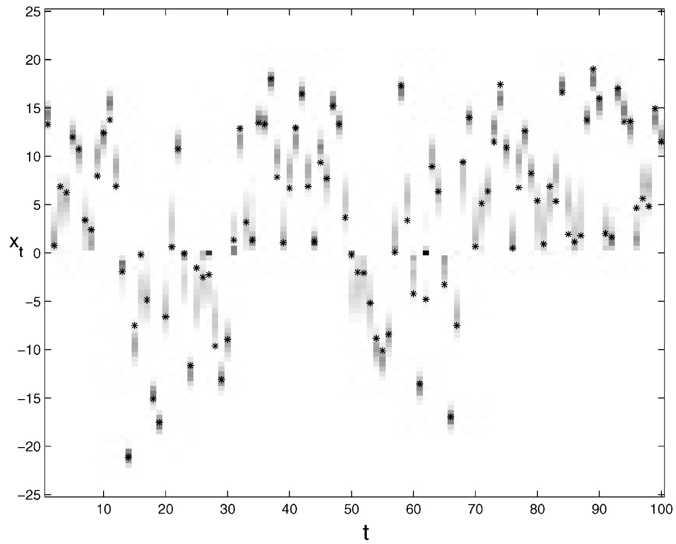

Figure 4. Histogram Estimates of Smoothing Densities, $p(x_t|y_{1:100})$, Shown as Gray-Scale Intensities in the Vertical Direction. The true simulated states are shown as "$*$."

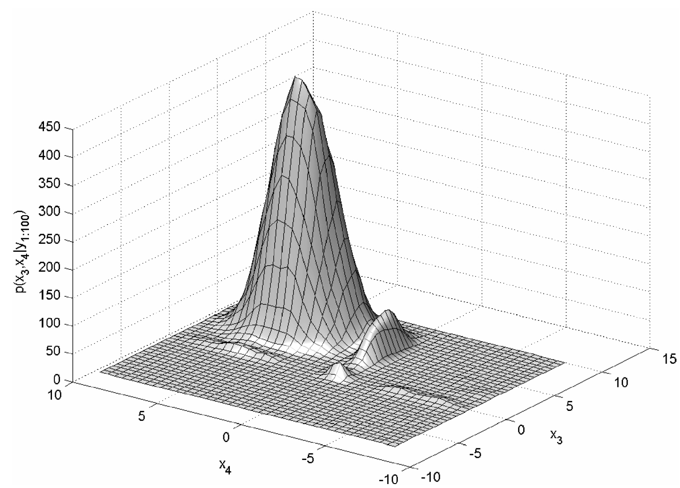

Figure 5. Kernel Density Estimate for $p(x_{3:4}|y_{1:100})$.

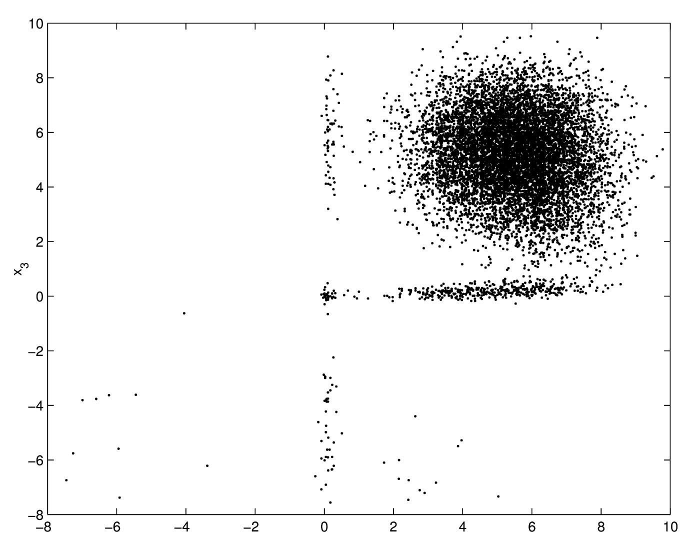

Figure 6. Scatterplot of Points Drawn From $p(x_{3:4}|y_{1:100})$.

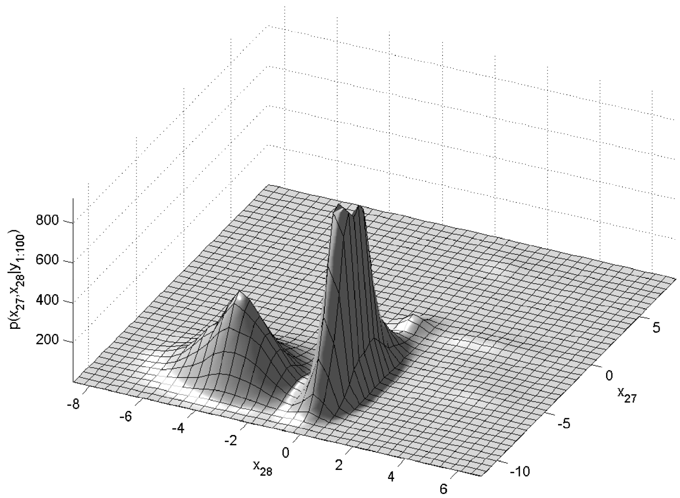

Figure 7. Kernel Density Estimate for $p(x_{27:28}|y_{1:100})$.

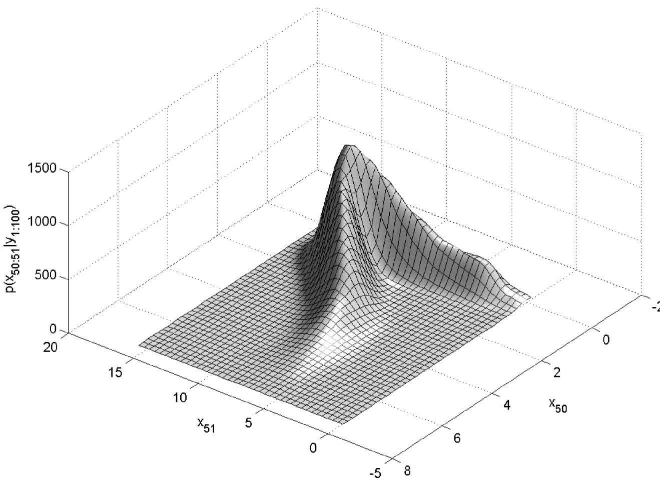

Figure 9. Kernel Density Estimate for $p(x_{50:51}|y_{1:100})$, Using $\sigma_v^2 = 1$ and $\sigma_w^2 = 9$.

variate marginals estimated from $p(x_{1:100}|y_{1:100})$ using two-dimensional scatterplots and kernel density estimates. The new smoothing method allows visualization of multimodality and complex interactions between states in a way that is not possible with the existing (univariate) marginal smoothing techniques. Repeated independent runs of the particle filter/smoother identified essentially identical features in the smoothing distribution, which gives us some confidence in the accuracy of the results. Different types of interactions become important in the smoothing density as the parameters of the model are changed. For example, with $\sigma_v^2 = 1$ and $\sigma_w^2 = 9$, we expect the dynamics of the state to play a more important role than in the previous simulation. This is borne out by the computed smoothing densities from a new set of simulations; see, for example, Figures 9 and 10, in which some diagonal structure is clearly present. Higher-dimensional smoothing marginals can be studied using three-dimensional visualization tools such as those available in Matlab.

5. EXAMPLE 2: APPLICATION TO SPEECH SIGNALS REPRESENTED BY TIME–VARYING AUTOREGRESSION MODELS

We now present a substantial application taken from the field of speech processing and analysis. Speech signals are inherently time-varying in nature, and any realistic representation thus should involve a model whose parameters evolve over time. One such model is the time-varying autoregression (TVAR) (Prado, West, and Krystal 1999; Kitagawa and Gersch 1996), in which the coefficients of a standard autoregression are allowed to vary according to some probability law. These models are of very wide utility and importance in engineering, scientific, and socioeconomic applications, but are typically applied subject to severe restrictions on the models for time variation in autoregressive coefficients for analytic reasons. More realistic models for patterns of variation over time in autoregressive parameters, and hence for the resulting "local" correlation and spectral structures, lead to intractable models and hence the need for simulation-based approximations, especially in sequential analysis contexts.

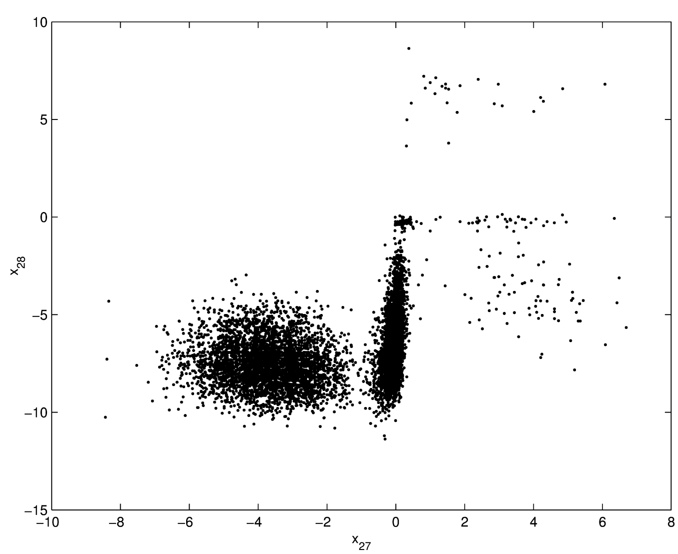

Figure 8. Scatterplot of Points Drawn From $p(x_{27:28}|y_{1:100})$.

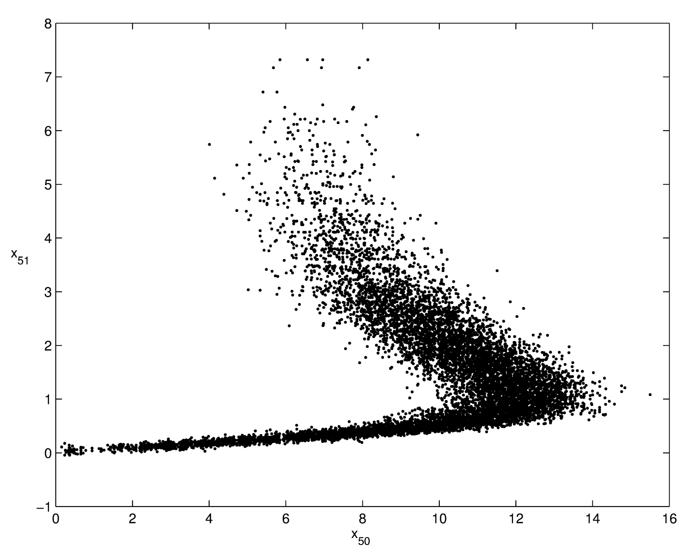

Figure 10. Scatterplot of Points Drawn From $p(x_{50:51}|y_{1:100})$.

Here we consider a nonlinear parameterization of the TVAR model in terms of partial correlation (PARCOR) coefficients (Friedlander 1982). This is especially relevant in acoustical contexts, such as speech processing, because the PARCOR coefficients can be regarded as the parameters of a linear acoustical tube whose characteristics are time-varying. This acoustical tube model can be regarded as an approximation to the characteristics of the vocal tract (Proakis, Deller, and Hansen 1993). By allowing the width parameters of the acoustical tube model, and hence the instantaneous PARCOR coefficients, to vary over time, we can allow for the physical changes that occur in the vocal tract shape as the speaker utters different sounds. The nonlinear model implied by such a scheme is not readily estimated using standard optimal techniques, such as MCMC, or approximation methods, such as the extended Kalman filter, owing to the strongly nonlinear nature of the transformation between TVAR coefficients and instantaneous PARCOR coefficients. Thus we see sequential Monte Carlo filtering and smoothing as the method of choice in this case. A related, although not identical, TVAR model has been considered in a different application by Kitagawa and Gersch (1996).

5.1 Model Specifications

A signal process $\{z_t\}$ is generated in the standard fashion from a Gaussian distribution centered at the linear prediction from the previous time step,

$$f\left(z_t|z_{t-1:t-P}, a_t, \sigma_{e_t}\right) = \mathcal{N}\left(\sum_{i=1}^{P} a_{t,i} z_{t-i}, \sigma_{e_t}^2\right).$$

Here $a_t = (a_{t,1}, a_{t,2}, \ldots, a_{t,P})'$ is the time-varying AR(P) coefficient vector and $\sigma_{e_t}^2$ is the innovation variance at time t. Note that both the AR coefficient vector and the innovation variance are assumed to be time-varying, so we specify evolution models for these as well. As noted earlier, the changes in AR coefficients over time will model changes in the vocal tract shape as the speaker makes different utterances. Time variation in innovation variance will model changes in the strength of the excitation signal in the glottis. The speech signal $\{z_t\}$ is assumed to be partially observed in additive independent Gaussian background noise, so that the observation process $\{y_t\}$ is generated according to

$$g(y_t|z_t, \sigma_v) = \mathcal{N}(z_t, \sigma_v^2),$$

where σ_v^2 is here assumed to be constant and known, corresponding to a fixed level of background noise in the environment that has been measured during a silent period. Although some applications consider the environment to be noise-free, we argue that in any speech setting there will always be a certain degree of noise, which should be modeled to capture the true character of the unobserved signal $\{z_t\}$.

Furthermore, this framework allows us to perform noise reduction for noisy speech signals, a task frequently required for applications in mobile telephony and speech recognition, for example. Note that the extension of our methods to the case of time-varying and correlated observation noise is immediate.

For our simulations, a Gaussian first-order autoregression is assumed for the log-standard deviation $\phi_{e_t} = \log(\sigma_{e_t})$, namely

$$f\left(\phi_{e_t}|\phi_{e_{t-1}}, \sigma_{\phi_e}^2\right) = \mathcal{N}\left(\alpha\phi_{e_{t-1}}, \sigma_{\phi_e}^2\right),$$

where α is a positive constant just less than unity and $\sigma_{\phi_e}^2$ is a fixed hyperparameter.

The model now requires specification of the time variation in the TVAR coefficient vector a_t itself. This is the main interest in our application, because we wish to find a model that is physically meaningful for the application and easily interpretable. Possibly the simplest choice of all, and that most common in previous work (Prado et al. 1999; Kitagawa and Gersch 1996), is a first-order Gaussian autoregression directly on the coefficients a_t,

$$f(a_t|a_{t-1}, \sigma_a^2) = \mathcal{N}(\beta a_{t-1}, \sigma_a^2 I).$$

More elaborate schemes of this sort are possible, such as a higher-order autoregression involving AR coefficients from further in the past or a nonidentity covariance matrix, the latter being a key feature of such models for some authors and applications, as in the work of (and references cited by) Prado et al. (1999) and West and Harrison (1997). However, models of this form do not have a particularly strong physical interpretation for acoustical systems such as speech. Moreover, an unconstrained TVAR exhibits large ("unstable") oscillations inconsistent with real speech data. Improved stability can be achieved by ensuring that the instantaneous poles, that is, the roots of the polynomial $(1 - \sum_{i=1}^{P} a_{t,i} L^{-i})$, lie strictly within the unit circle. This can be achieved by constraining the autoregression appropriately to have zero probability in the unstable region for the coefficients. However, simulating such a condition is very expensive using rejection sampling methods. A possible way to achieve greater stability in the model would be to model the roots directly (see Huerta and West (1999a, b) for the time-invariant case). However, there are unresolved issues here of dealing with complex roots that evolve into real roots and vice versa (see Prado et al. 1999). These issues do not arise if one works in the reflection coefficient or, equivalently, the partial correlation (PARCOR) coefficient domain (Friedlander 1982). Here the equivalent condition is that each reflection coefficient must simply be constrained to the interval $(-1, +1)$. The standard Levinson recursion is used to transform between a_t and the reflection coefficients ρ_t. Many models are possible for the time variation of ρ_t, including random walks based on beta distributions and inverse logit-transformed normal, and all would be feasible in our sequential Monte Carlo framework. We have chosen a simple truncated normal autoregression for discussion here:

Random PARCOR model:

$$f_t(\rho_t|\rho_{t-1}, \sigma_a^2) \propto \begin{cases} \mathcal{N}(\beta\rho_{t-1}, \sigma_a^2 I) & \text{if } \max\{|\rho_{t,i}|\} < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where β is a coefficient just less than 1. All of the simulations reported in this article have used this model. As we have already hinted, the TV-PARCOR framework is appealing not simply because of the ease of checking stability and evaluating the transition density, but also because a reflection coefficient model has a strong physical interpretation in certain systems, notably speech and other acoustic sounds, generated through tube-like mechanisms.

The state-space model is now fully specified. The state vector is

$$x_t = \left(z_{t:t-P+1}, \rho_t, \phi_{e_t}\right)'.$$

The hyperparameters σ_a^2, α, β, $\sigma_{\phi_e}^2$, and σ_v^2 are assumed to be prespecified and fixed in all of the simulations. The initial state probability is truncated Gaussian over the "stable" parameter region for the model.

5.2 Filtering and Smoothing

The first step in analyzing the data is to perform a complete forward sweep of a Monte Carlo filtering algorithm to produce weighted particles $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ for $t = 1, 2, \ldots, T$, drawn approximately according to $P(dx_t|y_{1:t})$. Because our smoothing method is quite general and not tied to any particular filtering method, we do not consider all of the possibilities in detail. We have experimented with various types of Monte Carlo filter, adapted to our specific TVAR model, using the standard sequential importance sampling filter (Doucet et al. 2000; Gordon et al. 1993; Liu and Chen 1998), the *auxiliary particle filter* (APF) (Pitt and Shephard 1999a), and schemes that incorporate local MCMC moves to improve the filtering approximation (MacEachern, Clyde, and Liu 1999; Gilks and Berzuini 2000). We observe empirically a moderate improvement in performance when the APF is used and note that for some challenging cases, incorporating MCMC moves is also worthwhile. The importance function that we use for the unobserved state is the state transition density, modified such that the current z_t is simulated exactly from its full conditional density, which is Gaussian, that is,

$$q(x_t|x_{t-1}, y_t) = p\big(z_t|z_{t-1:t-P}, a_t, \sigma_{e_t}, y_t\big) \times f(\rho_t|\rho_{t-1}, \sigma_a^2) f\big(\phi_{e_t}\big|\phi_{e_{t-1}}, \sigma_{\phi_e}^2\big).$$

(A full discussion of the relative merits of the various possible filtering schemes and importance functions applied to a related TVAR model can be found in Godsill and Clapp 2001.) After the filtering pass, smoothing is then carried out using the new method.

5.3 Results

A long section of noisy speech data is presented in Figure 11. The time-varying characteristics of the signal are clearly evident, with data points 1–6,800 (approximately) representing the word "rewarded" and points 6,800–10,000 representing

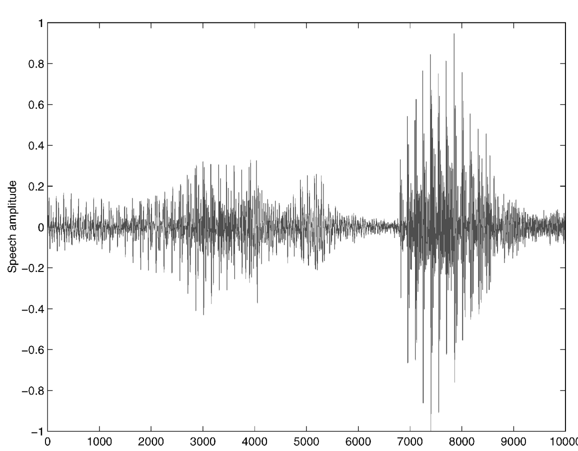

Figure 11. Speech Data. .62 second of a U.S. male speaker saying the words "rewarded by." Sample rate, 16 kHz, resolution, 16-bit, from the TIMIT speech database.

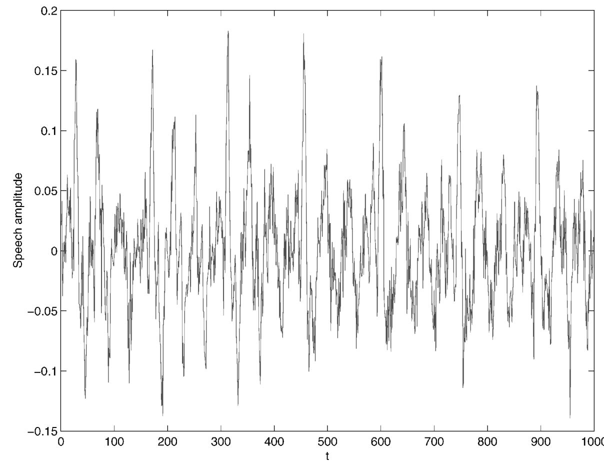

Figure 12. Noisy Speech Data, First 1,000 Data Points.

the word "by." A filtering and smoothing analysis is performed for the first 1,000 data points. A white Gaussian noise signal with known standard deviation $\sigma_v = .02$ is added to the signal. The observed noisy data are shown in Figure 12. The other fixed hyperparameters used were $\sigma_a = .01$, $\alpha = .99$, $\beta = 1$, and $\sigma_{\phi_e} = .001$. These were determined empirically by trial and error and by past experience with similar audio datasets (Godsill and Rayner 1998). The TVAR model order was $P = 4$; this was chosen in accordance with the findings of Vermaak, Andrieu, Doucet, and Godsill (2002) and also gave the best results in subjective listening tests performed on the smoothed signal estimates. The number of particles required will depend on the dimensionality of the state vector, which can be quite high in a model such as this, and also on the degree of posterior uncertainty about those parameters. Again, for this model the posterior distributions of parameters such as the TV-PARCOR coefficients can be quite diffuse, requiring a large number of particles for accurate representation. Filtering and smoothing were carried out using a wide variety of particle numbers, ranging from $N = 100$ up to $N = 20{,}000$. Filtered estimates of quantities such as posterior mean parameter values were found to become quite robust from one realization of the filter to another provided that the number of particles exceeded 1,000.

The last 200 observed data points are shown in Figure 13. Filtered means and 5/95 percentiles for the TV-PARCOR co-

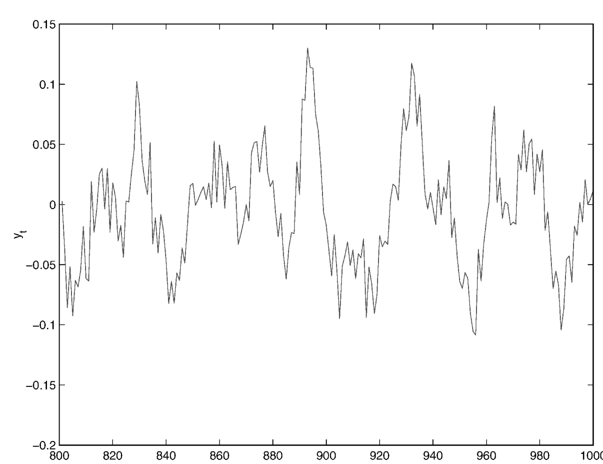

Figure 13. Noisy Speech, $t = 801, \ldots, 1{,}000$.

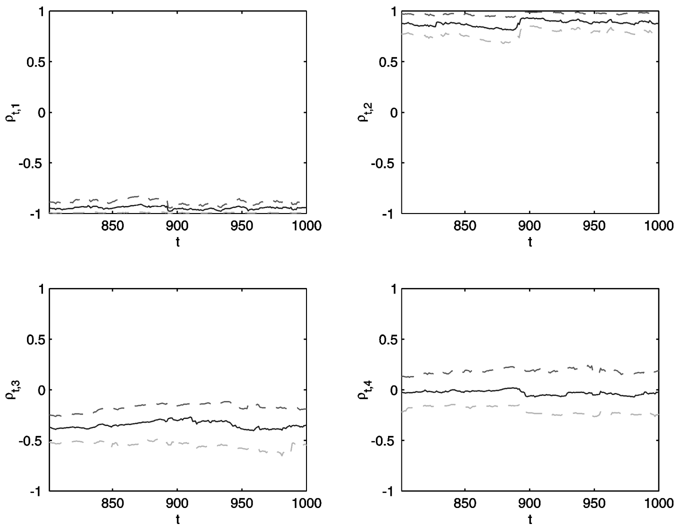

Figure 14. Posterior Mean and 5/95 Percentiles for Filtered TV-PARCOR Coefficient Estimates.

efficients are shown in Figure 14 for the last 200 data points, by which time it is assumed that the effects of the Gaussian initial state prior are negligible. The estimated trajectories are quite random in nature, reflecting the uncertainty about their value without the aid of retrospective smoothing. Similarly, the filtered mean estimate for the signal process is shown in Figure 15. Although some smoothing has occurred, it is fairly clear that the resulting estimate has followed the shape of the noisy signal too closely as a result of filtering without any retrospective sequence analysis.

In comparison, smoothing computations were then performed, using either our proposed backward sampling algorithm from Section 3 or the standard filtered trajectory approach outlined in Section 2. Figure 16 shows 10 realizations from the smoothing density using our proposed smoother. Figure 17 shows 4 separate realizations plotted on top of the observed noisy data, and Figure 18 shows the mean of 10 realizations plotted with the observed data. These results demonstrate that

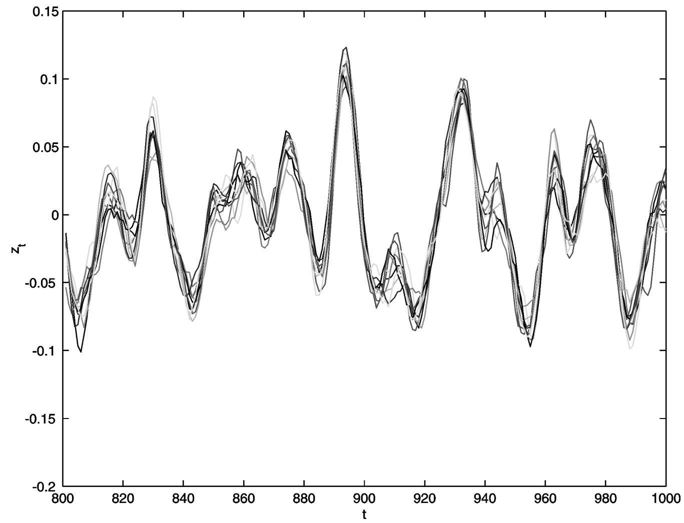

Figure 16. Ten Realizations From the Smoothed Signal Process.

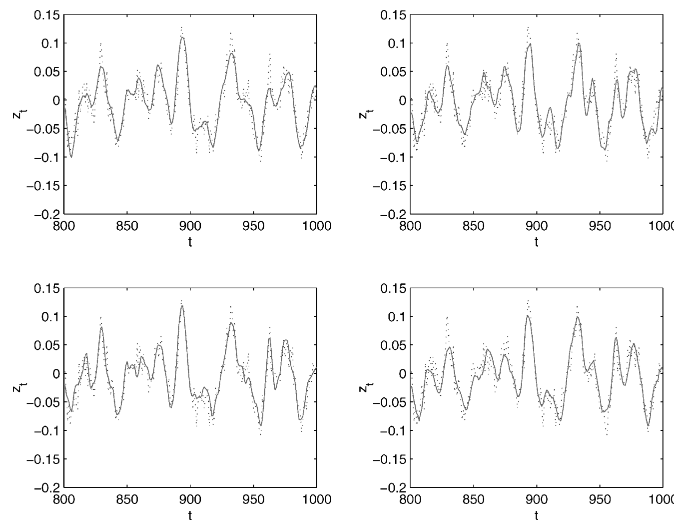

Figure 17. Four Realizations From the Smoothing Density for the Signal Process, With Noisy Data Shown by the Dotted Lines.

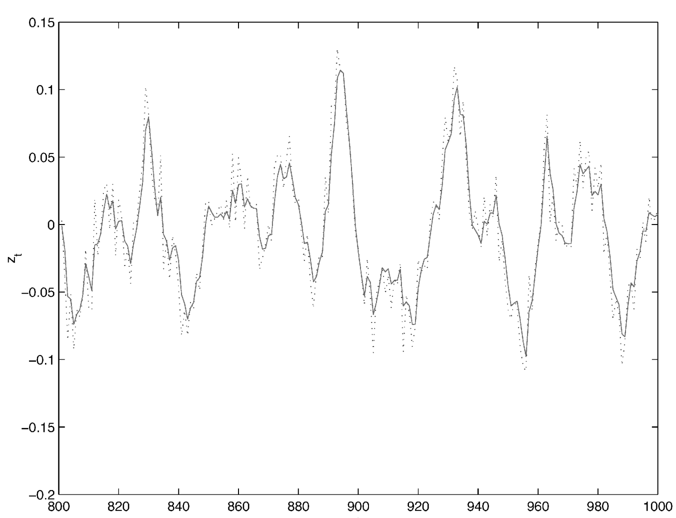

Figure 15. Filtered Speech, Estimated Posterior Mean (——), Noisy Data (······).

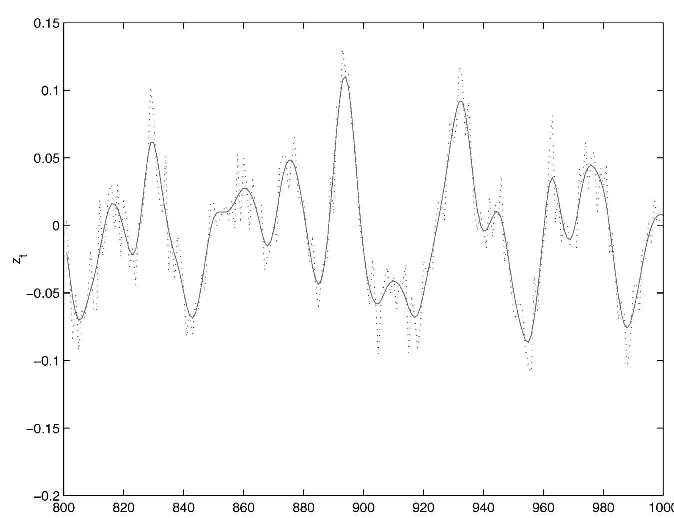

Figure 18. Average of 10 Realizations From the Smoothing Density With Noisy Data Shown by the Dotted Line.

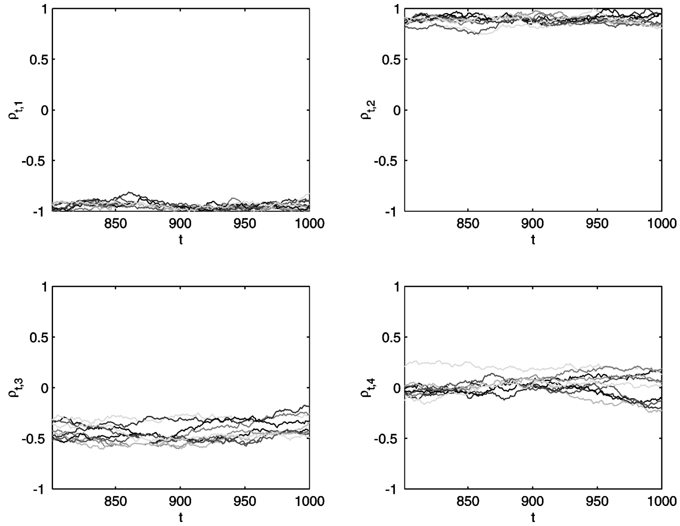

Figure 19. Ten Realizations From the Smoothing Density for the TV-PARCOR Coefficients.

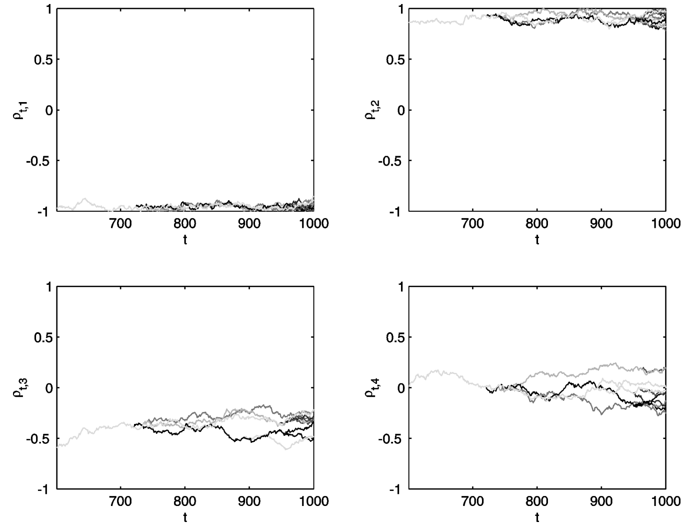

Figure 21. Ten Realizations From the Smoothing Density for the TV-PARCOR Coefficients Using the Standard Trajectory-Based Method (N = 2,000).

good diversity is achieved by the backward sampling smoother and that (visually) much-improved sequences are generated when compared with the filtered mean estimates of Figure 15.

Realizations from the smoothing density for the TV-PARCOR coefficients are given in Figure 19. A plausible degree of uncertainty is indicated by this result, and this may be compared with results obtained using the standard trajectory-based smoother. To highlight the differences between our proposed method, and the standard method, we smooth retrospectively back to $t = 600$. Graphs of 10 realizations from the smoothing density are presented in Figures 20 and 21, this time comparing our proposed method and the standard method. It is clear that the standard method degenerates to a single trajectory quite rapidly, whereas our proposed method achieves a plausible degree of variability between sampled realizations right back to $t = 600$. This is typical of the results that we have observed over many different sections of the data. The proposed method is always observed to improve on the standard method when equal numbers of particles are used for the filtering pass. Of course, our method can also degrade if the number of filtered particles is too small to contain plausible smoothed trajectories. An example of this is shown in Figure 22, which clearly indicates that the trajectories generated are too tightly clustered around a particular "modal" trajectory, and hence a misleading inference about posterior uncertainty could be obtained. Nevertheless, the result is still visually improved compared with the standard method for the same number of particles (Figure 23).

Finally, a filtering/smoothing pass was carried out on an entire sentence of speech, lasting some 2.5 seconds. Results can be auditioned by listening to the simulated signal trajectories. An example of this can be found *http://www-sigproc.eng.cam.ac.uk/˜sjg/TV-PARCOR*, where noisy speech and extracted speech can be compared. The results are satisfactory, eliminating some of the high-frequency frame-based artefacts ("musical

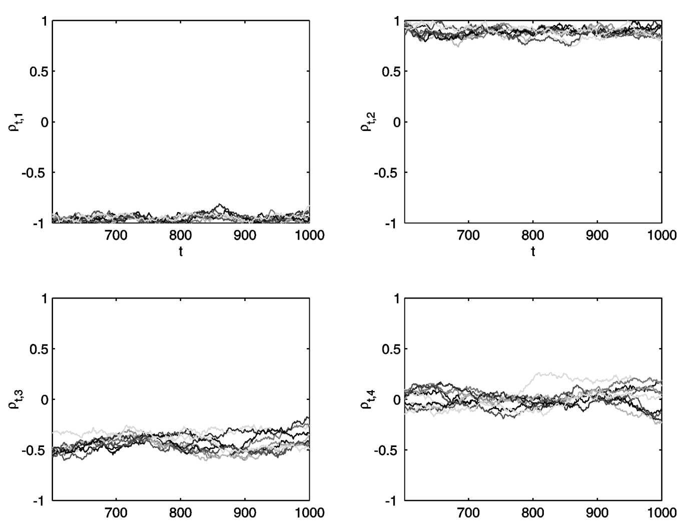

Figure 20. Ten Realizations From the Smoothing Density for the TV-PARCOR Coefficients Using the Proposed Simulation Smoothing Method (N = 2,000).

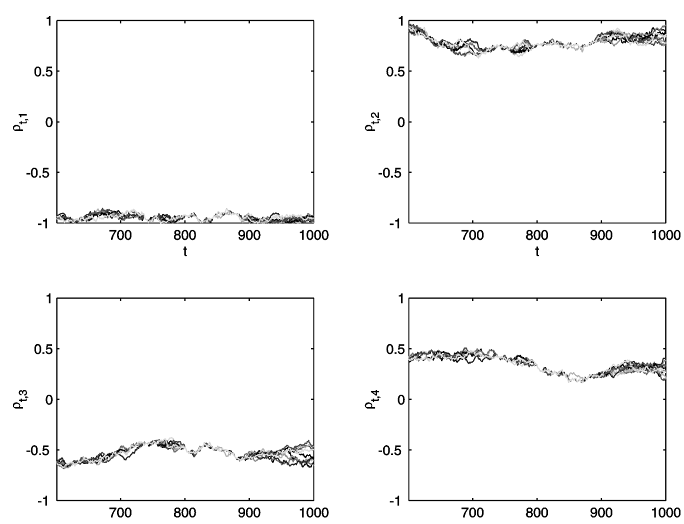

Figure 22. Ten Realizations From the Smoothing Density for the TV-PARCOR Coefficients Using the Proposed Simulation Smoothing Method (N = 500).

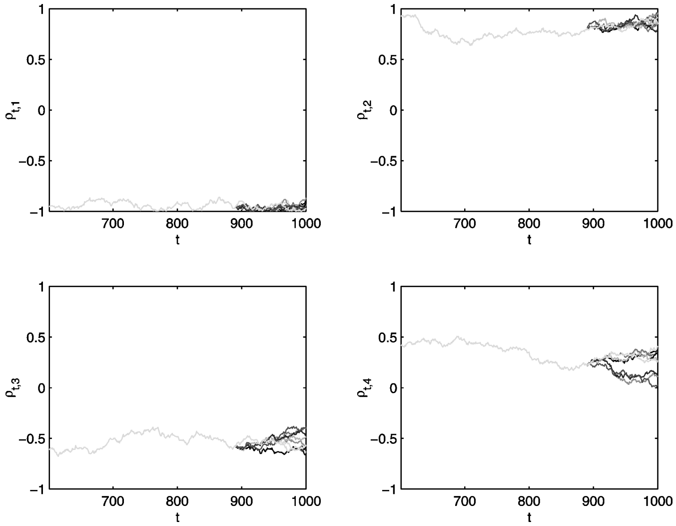

Figure 23. Ten Realizations From the Smoothing Density for the TV-PARCOR Coefficients Using the Standard Trajectory-Based Method (N = 500).

noise") observed with standard speech-enhancement algorithms, although the TVAR model causes some sibilance during unvoiced consonant sounds, such as "ss."

6. DISCUSSION

Recent years have seen a huge surge of interest in particle filtering, motivated by practical problems of sequential analysis in dynamic models in many areas of engineering, the natural sciences, and socioeconomics (Doucet et al. 2001). Our work here is not specific to any one algorithm and takes the established notion of sequentially updated particulate representations of posterior distributions in state-space models as the starting point for smoothing. In speech processing as in other applications, it is often critically important to "look back over time" for several or many time steps to assess and evaluate how new data revise the view of the recent past. Hence smoothing algorithms are key, and our work here develops effective approaches that apply regardless of the filtering method adopted.

We have developed and presented fairly simple and efficient methods for generating sample realizations of joint smoothing densities in a general model context. Smoothing has not been stressed by earlier authors in the sequential simulation literature, and where it has been studied, approaches have been limited to approximating the time-specific marginal smoothing distributions for individual states. We reiterate that this narrow focus is limiting and potentially misleading in many applications. Investigations of patterns of changes in historical states should focus on the joint trajectories of past states and hence necessarily involve consideration of joint smoothing densities, not simply the collection of marginals. Generating sample realizations is the most efficient, effective, and intuitive approach to studying complicated multivariate joint distributions; hence our focus on sampling algorithms for smoothing. This concept parallels that in forecasting, where studies of "sampled futures" are critical exercises in any serious modeling-for-forecasting activity.

There are research challenges in many aspects of the sequential simulation arena, including real needs for improved particle filtering algorithms, and reconciliation of the several variants of sequential importance sampling, resampling, and auxiliary particle methods. The present article ignores issues of learning on fixed model parameters in addition to time-varying states, a broader problem also ignored by most other authors in the field but critical in many applications, such as the challenging multifactor models of Pitt and Shephard (1999b) and Aguilar and West (2000). In our current work with TVAR models, we are developing analyses for both parameters and states using the auxiliary particle plus methods of Liu and West (2001). It should be noted that the smoothing methods developed and illustrated here apply directly in this context as well, providing a comprehensive smoothing algorithm.

APPENDIX: PROOF OF CONVERGENCE FOR THE BACKWARD SIMULATION SMOOTHER

Here we specify a more formal measure-theoretic framework, given as by Crisan (2001) and Crisan and Doucet (2000), for proof of convergence. Let (Ω, F, P) be a probability space on which two vector real-valued stochastic processes $X = \{X_t, t \in \mathbb{N}^*\}$ and $Y = \{Y_t, t \in \mathbb{N}^*\}$ are defined; let n_x and n_y be the dimensions of the state space of X and Y. The process X is unobserved, whereas the process Y is observed.

Let X be a Markov process with respect to the filtration $\mathcal{F}_t \triangleq \sigma(X_s, Y_s, s \in \{1, \dots, t\})$ with initial distribution $X_1 \sim f(dx_1)$ and transition kernel

$$f(dx_t|x_{t-1}) \triangleq P(X_t \in dx_t|X_{t-1} = x_{t-1}). \tag{A.1}$$

We assume that the observations are statistically independent given the signal and that they satisfy

$$g(dy_t|x_t) \triangleq P(Y_t \in dy_t|X_t = x_t). \tag{A.2}$$

For the sake of simplicity, we assume that $f(dx_t|x_{t-1})$ and $g(dy_t|x_t)$ admit densities with respect to the Lebesgue measure; that is, $f(dx_t|x_{t-1}) = f(x_t|x_{t-1})\,dx_t$ and $g(dy_t|x_t) = g(y_t|x_t)\,dy_t$, corresponding to the densities $f(\cdot|\cdot)$ and $g(\cdot|\cdot)$ in the main text.

We use the following standard shorthand. If μ is a measure, then φ is a function and K is a Markov kernel,

$$(\mu, \varphi) \triangleq \int \varphi\, d\mu,$$

$$\mu K(A) \triangleq \int \mu(dx) K(A|x),$$

and

$$K\varphi(x) \triangleq \int K(dz|x)\varphi(z).$$

We also use the following notation:

$$\pi_{t|t-1}(dx_t) \triangleq P(dx_t|Y_{1:t-1} = y_{1:t-1})$$

and

$$\pi_{t|t}(dx_t) \triangleq P(dx_t|Y_{1:t} = y_{1:t}) \qquad \text{(the "filtering" measure).}$$

The general particle filtering method described in Section 2 seeks successively to approximate these two measures using randomly propagated empirical measures.

More generally, we use the following shorthand for smoothing distributions:

$$\pi_{t_1:t_2|t_3}(dx_{t_1:t_2}) \triangleq P(dx_{t_1:t_2}|Y_{1:t_3} = y_{1:t_3}).$$

We define the weighted approximation to the filtering measure as

$$\pi_{t|t}^N(dx_t) \triangleq \sum_{i=1}^N w_t^{(i)} \delta_{x_t^{(i)}}(dx_t), \tag{A.3}$$

and the unweighted measure following resampling (selection) as

$$\pi_{t|t}'^{\,N}(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_t'^{(i)}}(dx_t). \tag{A.4}$$

Similarly, P smoothed realizations generated using Algorithm 1 can be used to form an unweighted approximation to the joint density. These realizations are sampled according to the discrete approximation $\pi_{1:t}^N$ of the true smoothing measure $\pi_{1:T}$. In Theorem A.2 we prove the convergence of this approximating measure to the true smoothing measure $\pi_{t:T|1:T}$ for any $t \in \{1, 2, \dots, T\}$.

Define a (joint) measure $\rho_t(dx_{t-1:t}) \triangleq \pi_{t-1|t-1}(dx_{t-1})q(dx_t| y_t, x_{t-1})$, chosen such that it is absolutely continuous with respect to $\pi_{t-1:t|t-1}(dx_{t-1:t})$ and such that h_t is the following (strictly positive) Radon–Nykodym derivative:

$$h_t(y_t, x_{t-1}, x_t) \propto \frac{\pi_{t-1:t|t-1}(dx_{t-1:t})}{\rho_t(dx_{t-1:t})}.$$

We now state sufficient conditions for convergence of the particle filter. Let $B(\mathbb{R}^n)$ be the space of bounded, Borel-measurable functions on $\mathbb{R}^n$ and denote, for any $\varphi \in B(\mathbb{R}^n)$,

$$\|\varphi\| \triangleq \sup_{x\in\mathbb{R}^n} \varphi(x).$$

Assume from now on that the observation process is fixed to a given observation record, $Y_{1:T} = y_{1:T}$. All subsequent convergence results are presented on this basis.

Consider the following assumption.

Assumption A.1. $\pi_{t-1:t|t}$ is absolutely continuous with respect to ρ_t. Moreover, $g(y_t|x_t)h_t(y_t, x_{t-1}, x_t)$ is positive and bounded in argument $(x_{t-1}, x_t) \in (\mathbb{R}^{n_x})^2$.

The following theorem for the particle filter is then a direct consequence of work of Crisan (2001) and Crisan and Doucet (2000).

Theorem A.1 (Crisan and Doucet 2000). Under Assumption A.1, for all $t > 0$, there exists $c_{t|t}$ independent of N such that for any $\phi \in B(\mathbb{R}^{n_x})$,

$$E\big[\big((\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi)\big)^2\big] \le c_{t|t}\frac{\|\phi\|^2}{N},$$

where the expectation is over all realizations of the random particle method.

Now define

$$\overline{f}(x) \triangleq \|f(x|\cdot)\|$$

and consider the following assumption.

Assumption A.2. For any $t \in (1, \dots, T)$, we have

$$\left(\pi_{t|T}, \left(\frac{\overline{f}}{\pi_{t|t-1}}\right)^2\right) < \infty.$$

We can now state the main theorem concerning convergence of the proposed simulation smoothing algorithm.

Theorem A.2. Under Assumptions A.1 and A.2, for all $t \in (1, \dots, T)$, there exists $c_{t|T}$ independent of N such that for any $\phi \in B(\mathbb{R}^{n_x T})$

$$E\big[\big((\pi_{1:T|T}^N, \phi) - (\pi_{1:T|T}, \phi)\big)^2\big] \le c_{1|T}\frac{\|\phi\|^2}{N},$$

where $c_{1|T}$ can be computed using the backward recursion,

$$c_{t|T} = \left((c_{t+1|T})^{1/2} + 2(c_{t|t})^{1/2}\left(\pi_{t+1|T}, \left(\frac{\overline{f}}{\pi_{t+1|t}}\right)^2\right)^{1/2}\right)^2, \tag{A.5}$$

and $c_{t|t}$ is given by Theorem A.1.

Proof. We prove here that for any $\phi_t \in B(\mathbb{R}^{n_x(T-t+1)})$, there exists $c_{t|T}$ independent of N such that

$$E\big[\big((\pi_{t:T|T}^N, \phi_t) - (\pi_{t:T|T}, \phi_t)\big)^2\big] \le c_{t|T}\frac{\|\phi_t\|^2}{N}.$$

The proof proceeds by induction. Theorem A.1 ensures that the result is true for $t = T$. Now, rewrite $(\pi_{t:T|T}, \phi_t)$ as

$$(\pi_{t:T|T}, \phi_t) = \left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}, \phi_t f)}{\pi_{t|t} f}\right).$$

Then decompose the error term $(\pi_{t:T|T}^N, \phi_t) - (\pi_{t:T|T}, \phi_t)$ as

$$\begin{aligned}
&(\pi_{t:T|T}^N, \phi_t) - (\pi_{t:T|T}, \phi_t)\\
&\quad = \left((\pi_{t+1:T|T}^N - \pi_{t+1:T|T}), \frac{(\pi_{t|t}^N, \phi_t f)}{\pi_{t|t}^N f}\right)\\
&\qquad + \left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N, \phi_t f)(\pi_{t|t} - \pi_{t|t}^N, f)}{(\pi_{t|t}^N, f)\pi_{t+1|t}}\right)\\
&\qquad + \left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N - \pi_{t|t}, \phi_t f)}{\pi_{t+1|t}}\right).
\end{aligned} \tag{A.6}$$

Then, by Minkowski's inequality, we have

$$\begin{aligned}
&E\big[(\pi_{t:T|T}^N, \phi_t) - (\pi_{t:T|T}, \phi_t)\big]^{1/2}\\
&\quad \le E\left[\left((\pi_{t+1:T|T}^N - \pi_{t+1:T|T}), \frac{(\pi_{t|t}^N, \phi_t f)}{\pi_{t|t}^N f}\right)^2\right]^{1/2}\\
&\qquad + E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N, \phi_t f)(\pi_{t|t} - \pi_{t|t}^N, f)}{(\pi_{t|t}^N, f)\pi_{t+1|t}}\right)^2\right]^{1/2}\\
&\qquad + E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N - \pi_{t|t}, \phi_t f)}{\pi_{t+1|t}}\right)^2\right]^{1/2}.
\end{aligned} \tag{A.7}$$

Consider now the three terms on the right side:

- First term. For any $x_{t+1:T}$, we have

$$\left\|\frac{(\pi_{t|t}^N, \phi_t f)}{\pi_{t|t}^N f}\right\| \le \|\phi_t\|.$$

  Thus, using the induction hypothesis, we obtain

$$E\left[\left((\pi_{t+1:T|T}^N - \pi_{t+1:T|T}), \frac{(\pi_{t|t}^N, \phi_t f)}{\pi_{t|t}^N f}\right)^2\right] \le c_{t+1|T}\frac{\|\phi_t\|^2}{N}. \tag{A.8}$$

- Second term. For any $x_{t+1:T}$, we have

$$\left|\frac{(\pi_{t|t}^N, \phi_t f)(\pi_{t|t} - \pi_{t|t}^N, \phi_t f)}{(\pi_{t|t}^N, f)\pi_{t+1|t}}\right| \le \frac{|(\pi_{t|t} - \pi_{t|t}^N, f)|}{\pi_{t+1|t}}\|\phi_t\|.$$

Thus

$$E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N, \phi_t f)(\pi_{t|t} - \pi_{t|t}^N, f)}{(\pi_{t|t}^N, f)\pi_{t+1|t}}\right)^2\right]$$

$$\leq E\left[\left(\pi_{t+1:T|T}, \frac{|(\pi_{t|t} - \pi_{t|t}^N, f)|}{\pi_{t+1|t}}\right)^2\right]\|\phi_t\|^2$$

$$= \left(\pi_{t+1:T|T}, \pi_{t+1|t}^{-2} E\left((\pi_{t|t} - \pi_{t|t}^N, f)^2\right)\right)\|\phi_t\|^2$$

(Jensen's inequality).

By Theorem A.1 and Assumption A.2,

$$E\left((\pi_{t|t} - \pi_{t|t}^N, f)^2\right) \leq c_{t|t}\frac{\overline{f}^2}{N},$$

and hence

$$E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N, \phi_t f)(\pi_{t|t} - \pi_{t|t}^N, f)}{(\pi_{t|t}^N, f)\pi_{t+1|t}}\right)^2\right]$$

$$\leq \frac{c_{t|t}}{N}\left(\pi_{t+1|T}, \frac{\overline{f}^2}{\pi_{t+1|t}^2}\right)\|\phi_t\|^2. \quad \text{(A.9)}$$

- Third term. We have

$$E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N - \pi_{t|t}, \phi_t f)}{\pi_{t+1|t}}\right)^2\right]$$

$$\leq E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N - \pi_{t|t}, \phi_t f)^2}{\pi_{t+1|t}^2}\right)\right]$$

(Jensen's inequality)

$$= \left(\pi_{t+1:T|T}, \pi_{t+1|t}^{-2} E\left((\pi_{t|t}^N - \pi_{t|t}, \phi_t f)^2\right)\right).$$

By Theorem A.1 and Assumption A.2, we obtain

$$E\left((\pi_{t|t}^N - \pi_{t|t}, \phi_t f)^2\right) \leq \frac{c_{t|t}}{N}\overline{f}^2\|\phi_t\|^2,$$

so that

$$E\left[\left(\pi_{t+1:T|T}, \frac{(\pi_{t|t}^N - \pi_{t|t}, \phi_t f)}{\pi_{t+1|t}}\right)^2\right]$$

$$\leq \frac{c_{t|t}}{N}\left(\pi_{t+1|T}, \frac{\overline{f}^2}{\pi_{t+1|t}^2}\right)\|\phi_t\|. \quad \text{(A.10)}$$

The result follows by combining (A.6), (A.8), (A.9), and (A.10).

[Received ????. Revised ????.]

REFERENCES

Aguilar, O., Huerta, G., Prado, R., and West, M. (1999), "Bayesian Inference on Latent Structure in Time Series" (with discussion), in *Bayesian Statistics 6*, eds. J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 3–26.

Aguilar, O., and West, M. (1998), "Analysis of Hospital Quality Monitors Using Hierarchical Time Series Models," in *Case Studies in Bayesian Statistics in Science and Technology*, Vol. 4, eds. C. Gatsonis, R. E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, New York: Springer-Verlag, pp. 287–302.

——— (2000), "Bayesian Dynamic Factor Models and Variance Matrix Discounting for Portfolio Allocation," *Journal of Business and Economic Statistics*, 18, 338–357.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "Building Robust Simulation-Based Filters for Evolving Datasets," technical report, University of Oxford.

Carter, C. K., and Kohn, R. (1994), "On Gibbs Sampling for State-Space Models," *Biometrika*, 81, 541–553.

Crisan, D. (2001), "Parficle Filters: A Theoretical Perspective," in *Sequential Monte Carlo Methods in Practice*, eds. A. Doucet, J. F. G. De Freitas, and N. J. Gordon, New York: Springer-Verlag.

Crisan, D., Del Moral, P., and Lyons, T. (1999), "Discrete Filtering Using Branching and Interacting Particle Systems," *Markov Processes and Related Fields*, 5, 293–318.

Crisan, D., and Doucet, A. (2000), "Convergence of Sequential Monte Carlo Methods," Technical Report CUED/F-INFENG/TR.381, Cambridge University, Engineering Dept.

Crisan, D., and Lyons, T. (1999), "A Particle Approximation of the Solution of the Kushner–Stratonovich Equation," *Probability Theory and Related Fields*, 5, 293–318.

De Jong, P. (1997), "The Scan Sampler for Time Series Models," *Biometrika*, 84, 929–937.

De Jong, P., and Shephard, N. (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350.

Del Moral, P. (1998), "Measure-Valued Processes and Interacting Particle Systems," *The Annals of Applied Probability*, 8, 438–495.

Doucet, A., De Freitas, J. F. G., and Gordon, N. J. (eds.) (2001), *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag.

Doucet, A., Godsill, S. J., and Andrieu, C. (2000), "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, 10, 197–208.

Friedlander, B. (1982), "Lattice Filters for Adaptive Processing," *Proceedings of the IEEE*, 70(8), 829–867.

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

Gilks, W. R., and Berzuini, C. (2000), "Following a Moving Target—Monte Carlo Inference for Dynamic Bayesian Models," *Journal of the Royal Statistical Society*, Ser. B, 63, 1–20.

Godsill, S. J., and Clapp, T. C. (2001), "Improvement Strategies for Monte Carlo Particle Filters," in *Sequential Monte Carlo Methods in Practice*, eds. A. Doucet, J. F. G. De Freitas, and N. J. Gordon, New York: Springer-Verlag.

Godsill, S. J., Doucet, A., and West, M. (2001), "Maximum a posteriori Sequence Estimation Using Monte Carlo Particle Filters," *Annals of the Institute of Statistical Mathematics*, 53, 82–96.

Godsill, S. J., and Rayner, P. J. W. (1998), *Digital Audio Restoration: A Statistical Model-Based Approach*, Berlin: Springer-Verlag.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, 140, 107–113.

Huerta, G., and West, M. (1999a), "Bayesian Inference on Periodicities and Component Spectral Structure in Time Series," *Journal of Time Series Analysis*, 20, 401–416.

——— (1999b), "Priors and Component Structures in Autoregressive Time Series Models," *Journal of the Royal Statistical Society*, Ser. B, 61, 881–899.

Hürzeler, M., and Künsch, H. R. (1998), "Monte Carlo Approximations for General State Space Models," *Journal of Computational and Graphical Statistics*, 7, 175–193.

——— (2001), "Approximating and Maximising the Likelihood for a General State Space Model," in *Sequential Monte Carlo in Practice*, eds. A. Doucet, J. F. G. De Freitas, and N. J. Gordon, New York: Springer-Verlag, pp. 159–176.

Kitagawa, G. (1996), "Monte Carlo Filter and Smoother for Nonlinear Non-Gaussian State-Space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

Kitagawa, G., and Gersch, W. (1996), *Smoothness Priors Analysis of Time Series*, New York: Springer-Verlag.

Liu, J., and West, M. (2001), "Combined Parameter and State Estimation in Simulation-Based Filtering," in *Sequential Monte Carlo Methods in Practice*, eds. A. Doucet, J. F. G. De Freitas, and N. J. Gordon, New York: Springer-Verlag, pp. 197–217.

Liu, J. S., and Chen, R. (1995), "Blind Deconvolution via Sequential Imputations," *Journal of the American Statistical Association*, 90, 567–576.

——— (1998), "Sequential Monte Carlo Methods for Dynamical Systems," *Journal of the American Statistical Association*, 93, 1032–1044.

MacEachern, S. N., Clyde, M., and Liu, J. S. (1999), "Sequential Importance Sampling for Nonparametric Bayes Models: The Next Generation," *Canadian Journal of Statistics*, 27, 251–267.

Pitt, M. K., and Shephard, N. (1999a), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599.

——— (1999b), "Time-Varying Covariances: A Factor Stochastic Volatility Approach" (with discussion), in *Bayesian Statistics 6*, eds. J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 547–570.

Prado, R., West, M., and Krystal, A. D. (1999), "Evaluation and Comparison of EEG Traces: Latent Structure in Nonstationary Time Series," *Journal of the American Statistical Association*, 94, 1083–1095.

Proakis, J., Deller, J., and Hansen, J. (1993), *Discrete-Time Processing of Speech Signals*, New York: Macmillan.

Shephard, N. (1994), "Partial Non-Gaussian State Space," *Biometrika*, 81, 115–131.

Shephard, N., and Pitt, M. K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–667.

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. J. (2002), "Particle Methods for Bayesian Modeling and Enhancement of Speech Signals," *IEEE Transactions on Speech and Audio Processing*, 10, 173–185.

West, M. (1993), "Mixture Models, Monte Carlo, Bayesian Updating and Dynamic Models," in *Computing Science and Statistics: Proceedings of the 24th Symposium on the Interface*, ed. J. H. Newton, Fairfax Station, VA: Interface Foundation of North America, pp. 325–333.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.