Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Theory and Methods

Genshiro Kitagawa

To cite this article: Genshiro Kitagawa (1998) Theory and Methods, Journal of the American Statistical Association, 93:443, 1203-1215, DOI: 10.1080/01621459.1998.10473780

To link to this article: https://doi.org/10.1080/01621459.1998.10473780

Published online: 17 Feb 2012.

A Self-Organizing State-Space Model

Genshiro KITAGAWA

A self-organizing filter and smoother for the general nonlinear non-Gaussian state-space model is proposed. An expanded state-space model is defined by augmenting the state vector with the unknown parameters of the original state-space model. The state of the augmented state-space model, and hence the state and the parameters of the original state-space model, are estimated simultaneously by either a non-Gaussian filter/smoother or a Monte Carlo filter/smoother. In contrast to maximum likelihood estimation of model parameters in ordinary state-space modeling, for which the recursive filter computation has to be done many times, model parameter estimation in the proposed self-organizing filter/smoother is achieved with only two passes of the recursive filter and smoother operations. Examples such as automatic tuning of dispersion and the shape parameters, adaptation to changes of the amplitude of a signal in seismic data, state estimation for a nonlinear state space model with unknown parameters, and seasonal adjustment with a nonlinear model with changing variance parameters are shown to exemplify the usefulness of the proposed method.

KEY WORDS: Bayesian estimation; Filtering; Likelihood; Nonlinear model; Parameter estimation; Self-tuning; Smoothing.

1. INTRODUCTION

Since the development of the Kalman filter in 1960, the state-space model has been widely used in many fields of engineering. However, use of the state-space model became popular in the statistical community only after it was introduced for stochastic system identification by autoregressive moving average (ARMA) modeling (Akaike 1974) and in the dynamic linear model (Harrison and Stevens 1976). By the 1980s, the state-space model was a popular tool for handling nonstationary time series (Harvey 1989; Kitagawa and Gersch 1984).

In the latter half of that decade, many statisticians were interested in the analysis of various types of nonstandard time series, for which the ordinary linear Gaussian state-space model cannot yield reasonable results. To handle such problems, several types of nonlinear non-Gaussian state-space models and related recursive filtering and smoothing algorithms were developed: the dynamic generalized linear model (Fahrmeir 1992; Smith and Miller 1986; West and Harrison 1997; West, Harrison, and Migon 1985) and various extensions of the Kalman filter (Fahrmeir and Kaufmann 1991; Frühwirth-Schnatter 1994; Kitagawa 1994; Meinhold and Singpurwalla 1989; Schnatter 1992), the Gibbs sampler-based method (Carlin, Polson, and Stoffer 1992), and the sequential imputations of Kong, Liu, and Wong (1994).

On the other hand, in earlier work (Kitagawa 1987) I proposed an alternative non-Gaussian filter and smoother that can yield the exact marginal posterior density of the state for fairly general types of state-space models. Because that method is based on numerical integration, its application is limited only to the models with relatively low state dimension (say, less than or equal to 4). Despite the development of various refinements of the integration method (e.g., Hodges and Hale 1993; Tanizaki 1993), they did not yield an essential solution to the problem of modeling nonstandard time series.

A Monte Carlo filter and smoother was shown in earlier work (Kitagawa 1993, 1996) which is applicable to very general state-space models forms. [A similar "bootstrap filter" algorithm was proposed by Gordon, Salmond, and Smith (1993). The problem that Gordon et al. and I considered is in fact a signal extraction problem that is one-to-one with the smoothing problem (Kohn and Ansley 1988). Gordon et al. did not address smoothing.] In Monte Carlo type methods, arbitrary non-Gaussian densities are approximated by many particles that can be considered realizations from the distributions. With the development of these algorithms, it is now possible to use high-dimensional nonlinear non-Gaussian state-space models for the analysis of complex time series.

Nevertheless, a very important question remained: How to operate it without knowledge of system parameters? (Solo 1989). In the statistical community, the maximum likelihood method is commonly used to address that problem. But for nonlinear or non-Gaussian state-space modeling, two factors sometimes render the maximum likelihood method impractical. First, the non-Gaussian smoother is computationally intensive and the repeated application of a numerical optimization procedure for evaluating the likelihood may make it almost impractical. Second, and more importantly, the log-likelihood computed by the Monte Carlo filter is subject to a sampling error. Therefore, precise maximum likelihood parameter estimates can be obtained only by using a very large number of particles or by parallel application of many Monte Carlo filters.

The self-organizing state-space model proposed in this article was developed to mitigate this difficulty. In this approach the unknown parameters of the model are appended to the state vector, and both the state and the parameters

Genshiro Kitagawa is Professor, Department of Prediction and Control, The Institute of Statistical Mathematics, Tokyo 106-8569, Japan. The author thanks Will Gersch of the University of Hawaii and Tomoyuki Higuchi of the Institute of Statistical Mathematics for their useful comments. Careful reading and valuable comments and suggestions by the editor, an associate editor, and the referee—in particular, on the accuracy of the estimates—are gratefully acknowledged. This work was partially supported by Grants-in-Aid for Scientific Research 08680338, 09680318, and 08558021 from the Japanese Ministry of Education, Science, Sports and Culture.

Journal of the American Statistical Association
September 1998, Vol. 93, No. 443, Theory and Methods

are estimated simultaneously by the recursive filter and smoother. This kind of recursive estimation was tried earlier in the engineering literature (e.g., Ljung 1979; Mehra 1970; Solo 1980). Most of those efforts used the extended Kalman filter in which each non-Gaussian marginal prior and posterior density is approximated by a single Gaussian density. "Although this extended Kalman filter approach appears perfectly straightforward, experience has shown that with the usual state-space model, it does not work well in practice" (Andersen and Moore 1979, p. 284; see also Kitagawa 1991). Very likely, the main reason for the difficulty in recursive parameter estimation was the lack of a practical nonlinear non-Gaussian smoothing algorithm. In the method proposed in this article, accurate approximations to the marginal posterior distributions of the state and the parameters are obtained either by the non-Gaussian smoother (for low-dimensional state models) or the Monte Carlo smoother (for higher order state models).

In contrast to most of other self-tuning methods, which give on-line estimates of parameter values, the proposed method automatically yields the distribution of the parameters of the model; hence the name "self-organizing." In the examples shown herein, it is demonstrated that the self-organizing state-space model yields reasonable solutions to several problems, some of which have not previously been solved by other methods.

2. STATE-SPACE MODEL WITH AUGMENTED STATE VECTOR

Consider a non-Gaussian nonlinear state-space model for the time series y_n,

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

and

$$y_n = h(\mathbf{x}_n, w_n), \tag{2}$$

where $\mathbf{x}_n$ is an unknown state vector and $\mathbf{v}_n$ and w_n are the system noise and the observation noise with densities $q(\mathbf{v})$ and $r(w)$. Equation (1) is called the system model and (2) the observation model. The initial state $\mathbf{x}_0$ is assumed to be distributed according to the density $p_0(\mathbf{x})$. The possibly nonlinear functions $f(\mathbf{x}, \mathbf{v})$ and $h(\mathbf{x}, w)$ and the possibly non-Gaussian densities $q(\mathbf{v})$ and $r(w)$ may contain some parameters. The vector consisting of these unknown parameters is hereafter denoted by $\boldsymbol{\theta}$.

This nonlinear state-space model specifies the following two conditional density functions: $p(\mathbf{x}_n|\mathbf{x}_{n-1}, \boldsymbol{\theta})$, the density of the state $\mathbf{x}_n$ given the previous state $\mathbf{x}_{n-1}$, and $p(y_n|\mathbf{x}_n, \boldsymbol{\theta})$, the density of y_n given the state $\mathbf{x}_n$. The set of observations up to the time point j is denoted by $Y_j = \{y_1, \ldots, y_j\}$. Then $p(\mathbf{x}_n|Y_{n-1}, \boldsymbol{\theta})$ defines the density of the one-step-ahead predictor.

The likelihood of the model specified by the parameter $\boldsymbol{\theta}$ is obtained by

$$L(\boldsymbol{\theta}) = p(y_1, \ldots, y_N|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(y_n|Y_{n-1}, \boldsymbol{\theta}), \tag{3}$$

where $p(y_n|Y_{n-1}, \boldsymbol{\theta})$ is the conditional density of y_n given Y_{n-1} and is obtained by

$$\begin{aligned} p(y_n|Y_{n-1}, \boldsymbol{\theta}) &= \int p(y_n, \mathbf{x}_n|Y_{n-1}, \boldsymbol{\theta})\, d\mathbf{x}_n \\ &= \int p(y_n|\mathbf{x}_n, \boldsymbol{\theta}) p(\mathbf{x}_n|Y_{n-1}, \boldsymbol{\theta})\, d\mathbf{x}_n. \end{aligned} \tag{4}$$

The maximum likelihood estimator (MLE) of the parameter $\boldsymbol{\theta}$ is obtained by maximizing the log-likelihood $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$.

For nonlinear or non-Gaussian state-space model, the two factors mentioned in Section 1, sometimes render the maximum likelihood method impractical. Here, instead of estimating the parameter $\boldsymbol{\theta}$ by the maximum likelihood method, I consider a Bayesian estimation by augmenting the state vector as

$$\mathbf{z}_n = \begin{bmatrix} \mathbf{x}_n \\ \boldsymbol{\theta} \end{bmatrix}. \tag{5}$$

The state-space model for this augmented state vector $\mathbf{z}_n$ is given by

$$\mathbf{z}_n = F(\mathbf{z}_{n-1}, \mathbf{v}_n) \tag{6}$$

and

$$y_n = H(\mathbf{z}_n, w_n), \tag{7}$$

where

$$F(\mathbf{z}, \mathbf{v}) = \begin{bmatrix} f(\mathbf{x}, \mathbf{v}) \\ \boldsymbol{\theta} \end{bmatrix} \tag{8}$$

and

$$H(\mathbf{z}, w) = h(\mathbf{x}, w). \tag{9}$$

Under this setting, assume that the posterior distribution $p(\mathbf{z}_n|Y_N)$ is obtained given the entire observations $Y_N = \{y_1, \ldots, y_N\}$ and the prior distribution $p(\mathbf{x}_0, \boldsymbol{\theta})$. Because the original state vector $\mathbf{x}_n$ and the parameter vector $\boldsymbol{\theta}$ are included in the augmented state vector $\mathbf{z}_n$, it is possible to obtain the marginal posterior densities of the parameter and of the original state simultaneously without obtaining the MLE of $\boldsymbol{\theta}$ beforehand. That is, the marginal posterior densities of the state $\mathbf{x}_n$ and the parameter $\boldsymbol{\theta}$ are obtained by

$$p(\mathbf{x}_n|Y_N) = \int p(\mathbf{z}_n|Y_N)\, d\boldsymbol{\theta} = \int p(\mathbf{x}_n, \boldsymbol{\theta}|Y_N)\, d\boldsymbol{\theta}$$

and

$$p(\boldsymbol{\theta}|Y_N) = \int p(\mathbf{z}_n|Y_N)\, d\mathbf{x}_n = \int p(\mathbf{x}_n, \boldsymbol{\theta}|Y_N)\, d\mathbf{x}_n. \tag{10}$$

It should be emphasized here that to obtain the MLE of $\boldsymbol{\theta}$ in ordinary state-space modeling, the filtering algorithm needs to be applied many times for computation of the log-likelihood. On the other hand, the self-organizing smoother need be applied only once.

This method of Bayesian estimation of the parameter of the state-space model can be easily extended to a time-varying parameter situation in which the parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_n$ evolves with time n. For the estimation of the time-varying parameter, we shall define the augmented state vector $\mathbf{z}_n$ by

$$\mathbf{z}_n = \begin{bmatrix} \mathbf{x}_n \\ \boldsymbol{\theta}_n \end{bmatrix}. \quad (11)$$

To obtain a state-space model for this augmented state, a proper model for the time evolution of $\boldsymbol{\theta}_n$ is necessary. If we use a random walk model, $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} + \mathbf{v}_{n2}$, with $\mathbf{v}_{n2}$ a white noise sequence distributed with a density function $p(\mathbf{v})$, then the nonlinear function F is defined by

$$F(\mathbf{z}_{n-1}, \mathbf{v}_n) = \begin{bmatrix} f(\mathbf{x}_{n-1}, \mathbf{v}_{n1}) \\ \boldsymbol{\theta}_{n-1} + \mathbf{v}_{n2} \end{bmatrix}, \quad (12)$$

where the system noise is $\mathbf{v}_n = (\mathbf{v}_{n1}^t, \mathbf{v}_{n2}^t)^t$. Note that the random walk model is only one example of possible models.

3. IMPLEMENTATION OF THE SELF-ORGANIZING STATE-SPACE MODEL

For implementation of the self-organizing state-space model, a practical and precise method of recursive filtering and smoothing is required. In the engineering literature, the extended Kalman filter was used for the recursive estimation of the parameters (Anderson and Moore 1979; Ljung 1979; Solo 1980; Young 1984). But because the extended Kalman filter approximates the non-Gaussian state density by a single Gaussian density, it does not always work well in practice (Anderson and Moore 1979, p. 284; Kitagawa 1991).

In this article I use two different recursive filtering and smoothing methods that can yield much more accurate state estimation of nonlinear non-Gaussian state-space models than the extended Kalman filter. These methods were developed for the nonaugmented state-space model and are applicable to the augmented state-space model as well. The first method is the non-Gaussian filter/smoother based on the numerical representation of the arbitrary densities (Kitagawa 1987). The other is the Monte Carlo filter/smoother (Gordon et al. 1993; Kitagawa 1993, 1996) based on the approximation of the density by many particles, which can be considered the realizations of the density.

The first method can yield an arbitrarily precise posterior density. However, it is very computationally costly and can be applied only to lower order (say, up to four-dimensional state) models. The Monte Carlo filter/smoother method can be applied to higher-order, very complicated state-space models. This is achieved with a tradeoff of a substantial computational burden for accuracy in the approximation to the posterior density.

For the recursive estimation of the state, I consider the following three conditional distributions:

$p(\mathbf{z}_n|Y_{n-1})$ predictive density,

$p(\mathbf{z}_n|Y_n)$ filter density,

and

$p(\mathbf{z}_n|Y_N)$ smoother density.

Next, two algorithms to obtain these conditional densities are briefly explored.

3.1 Implementation by the Non-Gaussian Smoother

A non-Gaussian filter for the nonlinear, non-Gaussian state-space model (6) and (7) is given by (Kitagawa 1987, 1991)

$$p(\mathbf{z}_n|Y_{n-1}) = \int p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{z}_{n-1}|Y_{n-1})\, d\mathbf{z}_{n-1} \quad (13)$$

and

$$p(\mathbf{z}_n|Y_n) = \frac{p(y_n|\mathbf{z}_n)p(\mathbf{z}_n|Y_{n-1})}{p(y_n|Y_{n-1})}, \quad (14)$$

where $p(y_n|Y_{n-1})$ is obtained by $p(y_n|Y_{n-1}) = \int p(y_n|\mathbf{z}_n)$ $p(\mathbf{z}_n|Y_{n-1})\, d\mathbf{z}_n$.

The final estimate of the augmented state $\mathbf{z}_n$ is obtained by the smoothing algorithm

$$p(\mathbf{z}_n|Y_N) = p(\mathbf{z}_n|Y_n) \int \frac{p(\mathbf{z}_{n+1}|\mathbf{z}_n)p(\mathbf{z}_{n+1}|Y_N)}{p(\mathbf{z}_{n+1}|Y_n)}\, d\mathbf{z}_{n+1}. \quad (15)$$

(Applications of this non-Gaussian smoothing algorithm can be seen in Kitagawa 1987, 1988, 1991.) The problem with this non-Gaussian filtering and smoothing method is that it is realizable by computationally costly numerical integration and can be applied to only low-order state models.

3.2 Implementation by the Monte Carlo Smoother

(Kitagawa 1993, 1996), I used a Monte Carlo filter and smoother developed for higher-order state-space models. In this method each density function is approximated by many (say m = 10,000 or 100,000) particles, which can be regarded as independent realizations from that distribution. It can be shown that these particles can be obtained recursively by the following filtering algorithm:

1. Generate m realizations of the system noise $\mathbf{v}_n^{(j)} \sim p(\mathbf{v})$ for $j = 1, \ldots, m$.
2. Obtain m realizations $\mathbf{p}_n^{(j)}$, which approximate the predictive distribution $p(\mathbf{z}_n|Y_{n-1})$ by $\mathbf{p}_n^{(j)} = F(\mathbf{f}_{n-1}^{(j)}, \mathbf{v}_n^{(j)})$.
3. Compute the importance factor $\alpha_n^{(j)}$ by $\alpha_n^{(j)} = p\ (y_n|\mathbf{x}_n = \mathbf{p}_n^{(j)})$.
4. Obtain m particles, $\mathbf{f}_n^{(1)}, \ldots, \mathbf{f}_n^{(m)}$, which approximate the filter distribution $p(\mathbf{z}_n|Y_n)$ by the resampling (sampling with replacement) of $\mathbf{p}_n^{(1)}, \ldots, \mathbf{p}_n^{(m)}$ with sampling probabilities proportional to $\alpha_n^{(1)}, \ldots, \alpha_n^{(m)}$.

The "bootstrap filter" (Gordon et al. 1993) is a similar algorithm. Hürzeler and Künsch (1995) showed a similar rejection method–based filtering algorithm. One way to increase the accuracy of the approximation is to generate d different realizations of system noise for each particle $\mathbf{f}_{n-1}^{(j)}$. Then obtain $m \times d$ particles $\mathbf{p}_n^{(i)}, i = 1, \ldots, m \times d$, and at step 4 obtain m particles by resampling from $m \times d$ particles.

A significant merit of this Monte Carlo filter procedure is that it can be applied to almost any type of nonlinear and non-Gaussian state-space models with higher (say 10 or 20) dimensions. The observation equation (2) can be extended to a more general form so that discrete distribution models can be treated.

This filtering algorithm can be extended to fixed lag smoothing by storing the past particles and resampling the vector of particles $(\mathbf{p}_n^{(j)}, \mathbf{p}_{n-1}^{(j)}, \ldots, \mathbf{p}_{n-l}^{(j)})$ rather than the single particle $\mathbf{p}_n^{(j)}$. It can be shown that the particles can be considered realization from a smoothed posterior density (Kitagawa 1993, 1996).

Incidentally, a problem with this Monte Carlo filter is that the computed likelihood is subject to a sampling error, because each term in (3) is approximated by

$$p(y_n|Y_{n-1}) = \int p(y_n|\mathbf{x}_n)p(\mathbf{x}_n|Y_{n-1})\,d\mathbf{x}_n \cong \frac{1}{m}\sum_{j=1}^{m} \alpha_n^{(j)}. \tag{16}$$

Thus in estimating the parameter of the model, it is difficult to obtain arbitrary close approximations to the MLE unless a very large number of particles are used or an average of the approximated log-likelihoods is computed by the parallel use of many Monte Carlo filters. This fact motivated the development of the self-organizing state-space model.

4. EXAMPLES

This section considers four numerical examples. The first example demonstrates that the scale and/or shape parameters of the model can be automatically adjusted by the self-organizing state-space model. The second example shows that the self-organizing state-space model can be applied to state estimation with a time-varying variance parameter. The last two examples are concerned with the nonlinear state-space models. Empirical studies on the effect of the selection of the initial distribution and the number of particles are also performed with these examples.

4.1 Self-Tuning of the Scale and the Shape Parameters

Consider estimation of the trend of the data shown in Figure 1(a), which was artificially generated by a Gaussian distribution with time-dependent mean

$$y_n \sim \mathrm{N}(t_n, 1), \tag{17}$$

where the mean value function t_n is given by

$$t_n = \begin{cases} 0 & n = 1, \ldots, 100 \\ 1.5 & n = 101, \ldots, 200 \\ -1 & n = 201, \ldots, 300 \\ 0 & n = 301, \ldots, 400 \end{cases}. \tag{18}$$

For the estimation of the mean t_n, consider the simple first-order trend model

$$x_n = x_{n-1} + v_n \tag{19}$$

and

$$y_n = x_n + w_n, \tag{20}$$

where w_n is a Gaussian white noise with mean 0 and variance 1. For the system noise temporarily assume that $v_n \sim C(v; \tau^2)$, where $C(v; \tau^2)$ denotes the Cauchy distribution with the density function

$$p(v; \tau^2) = \frac{1}{\pi}\frac{\tau}{v^2 + \tau^2}. \tag{21}$$

The MLE of the dispersion parameter τ^2 is 3.53×10^{-5}.

For the simultaneous estimation of the state and the parameter $\tau_n^2 = \tau^2$, I used the two-dimensional state vector

$$\mathbf{z}_n = \begin{bmatrix} x_n \\ \log_{10}\tau_n^2 \end{bmatrix}. \tag{22}$$

In this state, to ensure positivity of the dispersion parameter τ^2, $\log_{10}\tau_n^2$ rather than τ_n^2 is included in the augmented state vector. The state-space model for $\mathbf{z}_n$ is then given by

$$\begin{bmatrix} x_n \\ \log_{10}\tau_n^2 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \log_{10}\tau_{n-1}^2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_n \tag{23}$$

and

$$y_n = [1, 0] \begin{bmatrix} x_n \\ \log_{10}\tau_n^2 \end{bmatrix} + w_n, \tag{24}$$

with $v_n \sim C(v;\, \tau_{n-1}^2)$. For the initial state $\mathbf{z}_0 = (x_0, \log_{10}\tau_0^2)^t$, it is assumed that x_0 and $\log_{10}\tau_0^2$ are distributed independently as a normal distribution N(0, 4) and a uniform distribution, $U(-8, -2)$. Here $U(a, b)$ denotes the uniform distribution over the interval $[a, b]$. The non-Gaussian smoother given in Section 3.1 was used for the computation. In numerical integration, 201×101 nodes were used.

The marginal filtered densities $p(x_n|Y_n)$ and $p(\log_{10}\tau_n^2|Y_n)$ are shown in Figure 1, (b) and (c). For $1 \le n \le 100$, the filtered density $p(\log_{10}\tau_n^2|Y_n)$ is very vague. Because the mean of the true model is a constant in this interval, the mode of the filtered density $p(\log_{10}\tau_n^2|Y_n)$ gradually shifts to smaller values. But at around $n = 100$ where the sudden changes of the mean value occurs, it moves from about -6 to a large value, about -4.5. The shift of the mode is also seen around at $n = 200$ and 300 where the sudden change of the mean value, μ_n, occurs. The dispersion of the distribution of $\log_{10}\tau_n^2$ decreases as the jump of the mode occurs and the marginal filtered distribution is gradually concentrated approximately at $\log_{10}\tau_n^2 = -4.5$.

Figure 1, (d) and (e), shows the marginal smoothed densities $p(x_n|Y_{400})$ and $p(\log_{10}\tau_n^2|Y_{400})$. Clearly, $p(\log_{10}\tau_n^2|Y_{400})$ is a constant over time, a natural consequence of the assumption that τ_n^2 is constant over time. The mode of $p(\log_{10}\tau_n^2|Y_{400})$ is quite close to the MLE, $\hat{\tau}^2 = 3.53 \times 10^{-5}$ ($\log\hat{\tau}^2 = -4.45$). On the other hand, the smoothed densities of the trend $p(x_n|Y_{400})$ shifted three times and clearly detected the sudden changes of the mean value. This marginal

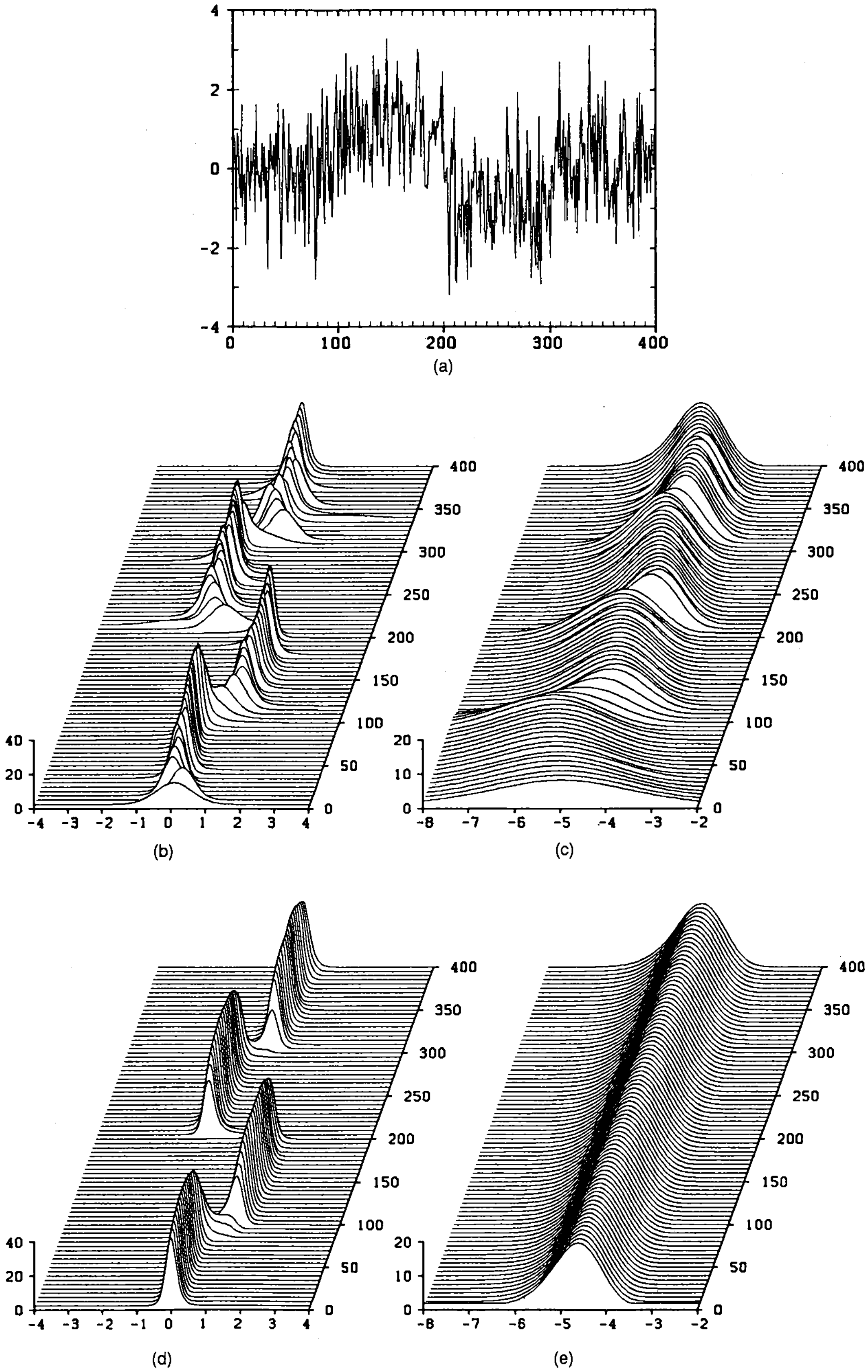

Figure 1. Trend Estimation by a Cauchy Noise Model for an Artificially Generated Data With Abruptly Changing Mean Value Function. (a) Artificially generated data; (b) filtered marginal posterior densities of the trend; (c) filtered marginal posterior densities of the dispersion parameter; (d) smoothed marginal posterior densities of the trend; (e) smoothed marginal posterior densities of the dispersion parameter.

smoothed density resembles that obtained using the MLE of the dispersion parameter, $\hat{\tau}^2$.

By visual inspection, it can be seen that the relative speed of adjustment of the state estimate and parameter estimate in the filtering shown in Figure 1, (b) and (c) is almost the same. However, as can be seen in 1, (d) and (e), after smoothing the state estimate moves very quickly but the parameter estimate does not move at all. Actually, the relative speed of adjustment is controlled by the variance (or dispersion) of the system noise distribution for the parameters.

The effect of the initial distribution was checked with this model. Table 1 shows the likelihoods of eight distinct initial distributions: three normal distributions, N(0, 1), N(0, 4), N(0, 16); the Cauchy distribution; the two-sided exponential distribution, the uniform distribution over $(-4,4)$, and two δ functions concentrated on $x_0 = 0$ and 2. One can see that for the case of alternative continuous distributions, the log-likelihood values are rather insensitive to the selection of the particular initial distribution and the difference of the log-likelihoods are at most 1.2. On the other hand, the log-likelihood values are very sensitive to the location of the delta function. Even in this case, however, the smoothed marginal posterior distributions are visually indistinguishable except for the first several steps.

The self-organizing state-space model can also be applied to the self-tuning of the noise distribution, including the shape and the dispersion parameters. For the system noise v_n, consider the model $v_n \sim P(v;\tau^2,b)$, where $P(v;\tau^2,b)$ denotes the type VII Pearson family of distributions with density function

$$p(v;\tau^2,b) = \frac{\Gamma(b)\tau^{2b-1}}{\Gamma(1/2)\Gamma(b-1/2)}\frac{1}{(v^2+\tau^2)^b}. \tag{25}$$

The dispersion parameter $\tau^2(>0)$ and the shape parameter $b(>1/2)$ are unknown. This family of distributions includes the Cauchy distribution ($b = 1$), the t distribution ($b = (k+1)/2$, k: positive integer), and the Gaussian distribution ($b = \infty$).

For this model, I used the three-dimensional state vector

$$\mathbf{z}_n = \begin{bmatrix} x_n \\ \theta_{n1} \\ \theta_{n2} \end{bmatrix}, \tag{26}$$

where $\theta_{n1} = \log_{10}\tau_n^2 - 3\theta_{n2}$ and $\theta_{n2} = \log_{10}(b_n - 1/2)$. The transformation $\theta_{n2} = \log_{10}(b_n - 1/2)$ is used to ensure the condition $b_n > 1/2$. On the other hand, the transformation $\theta_{n1} = \log_{10}\tau_n^2 - 3\theta_{n2}$ is arbitrary selected so that it becomes not so sensitive to the value of b_n. The state-space model for $\mathbf{z}_n$ is then given by

$$\begin{bmatrix} x_n \\ \theta_{n1} \\ \theta_{n2} \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \theta_{n-1,1} \\ \theta_{n-1,2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} v_n \tag{27}$$

and

$$y_n = [1,0,0]\begin{bmatrix} x_n \\ \theta_{n1} \\ \theta_{n2} \end{bmatrix} + w_n, \tag{28}$$

where $v_n \sim P(v;\tau_n^2,b_n)$. For the initial state $\mathbf{z}_0$, it is assumed that x_0,θ_{01}, and θ_{02} are distributed independently as a normal distribution N(0, 4) and uniform distributions $U(-5,2)$ and $(-3,2)$.

The marginal filtered densities $p(x_n|Y_n), p(\theta_{n1}|Y_n)$ and $p(\theta_{n2}|Y_n)$ are shown in Figure 2, (a)–(c). For $1 \le n \le 100, p(\theta_{n1}|Y_n)$ and $p(\theta_{n2}|Y_n)$ are very broad. At around 101, where the sudden change of the mean value occurs, the mode of $p(\theta_{n1}|Y_n)$ moves to a larger value and that of $p(\theta_{n2}|Y_n)$ to a smaller value. Significant moves of the modes are also seen at around $n = 200$ and 300.

Figure 2, (d)–(f), shows the marginal smoothed densities $p(x_n|Y_{400}), p(\theta_{n1}|Y_{400})$, and $p(\theta_{n2}|Y_{400})$. $p(\theta_{n1}|Y_{400})$ and $p(\theta_{n2}|Y_{400})$ are constant over time, and the mode of the distributions are $\theta_{n1} \approx -1.5$ ($\tau_n^2 = 1.6\times10^{-5}$) and $\theta_{n2} \approx -1.1$ ($b_n = .58$). The marginal posterior density $p(x_n|Y_{400})$ resembles but is slightly more concentrated than the one obtained by assuming that the system noise is distributed as Cauchy distribution.

Table 1. Effect of the Initial Distribution to the Log-Likelihood

Initial distribution	*Log-likelihood*
N(0, 1)	−605.84
N(0, 4)	−606.03
N(0, 16)	−606.80
Cauchy	−605.93
Two-sided exponential	−605.78
$U(-4, 4)$	−606.95
δ_0	−604.62
δ_2	−623.82

4.2 Adaptation to the Arrival of a Signal

In the previous example, the parameter of the model was assumed to be constant over time, and accordingly the system noise for the parameter was assumed to be zero. By using nonzero white-noise sequence for system noise, the self-organizing smoother can be applied to a model with time-varying parameters.

Figure 3(a) shows a record of the north–south component of a micro-earthquake, with sampling interval $\Delta t = .02$ second and data length $N = 2{,}800$ (Takanami 1991). A seismic signal can be seen to arrive in the middle of the series. But due to the presence of the comparatively high background noise, the arrival time and the precise shape of the signal are not particularly clear.

Kitagawa and Takanami (1985) detected the seismic signal by the following model:

$$y_n = r_n + s_n + w_n, \qquad w_n \sim \mathrm{N}(0,\sigma^2), \tag{29}$$

$$r_n = \sum_{j=1}^{m} a_j r_{n-j} + u_n, \qquad u_n \sim \mathrm{N}(0,\tau_1^2), \tag{30}$$

and

$$s_n = \sum_{j=1}^{k} b_j s_{n-j} + v_n, \qquad v_n \sim \mathrm{N}(0,\tau_{2n}^2), \tag{31}$$

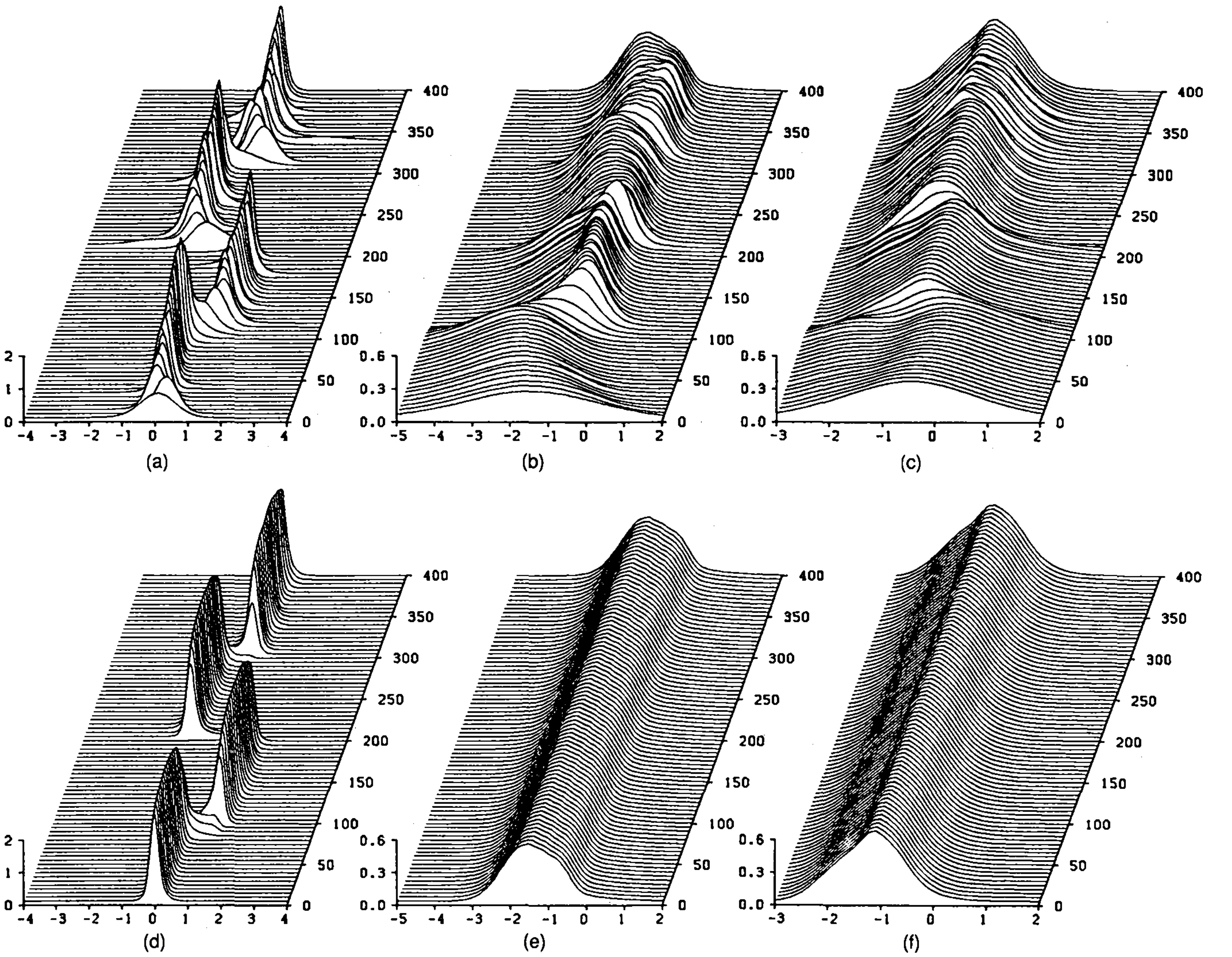

Figure 2. Trend Estimation by a Pearson-Type Noise Model for an Artificially Generated Data With Abruptly Changing Mean Value Function. (a) Filtered marginal posterior densities of the trend; (b) filtered marginal posterior densities of the scale parameter; (c) filtered marginal posterior densities of the shape parameter; (d) smoothed marginal posterior densities of the trend; (e) smoothed marginal posterior densities of the scale parameter; (f) smoothed marginal posterior densities of the shape parameter.

where r_n, s_n, and w_n are the background noise, the seismic signal, and the observation noise components. The autoregressive models for the background noise and seismic signal can be estimated by the maximum likelihood method using properly defined state-space models. The foregoing models expressed in state-space model form are

$$\left[\begin{array}{c} r_n \\ r_{n-1} \\ \vdots \\ r_{n-m+1} \\ \hline s_n \\ s_{n-1} \\ \vdots \\ s_{n-k+1} \end{array}\right] = \left[\begin{array}{cccc|cccc} a_1 & a_2 & \cdots & a_m & & & & \\ 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ \hline & & & & b_1 & b_2 & \cdots & b_k \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & \end{array}\right] \times \left[\begin{array}{c} r_{n-1} \\ r_{n-2} \\ \vdots \\ r_{n-m} \\ \hline s_{n-1} \\ s_{n-2} \\ \vdots \\ s_{n-k} \end{array}\right] + \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \hline 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{array}\right] \left[\begin{array}{c} u_n \\ v_n \end{array}\right]$$

and

$$y_n = [1 \quad 0 \quad \cdots \quad 0 | 1 \quad 0 \quad \cdots \quad 0]\mathbf{x}_n + w_n. \tag{32}$$

Because this is an ordinary state-space model, the state vector $\mathbf{x}_n$ and hence the background noise r_n and the signal s_n, can be estimated by the Kalman filter and the fixed interval smoother if all of the parameters of the model are given.

The difficulty in this approach is that the variance of the innovation of the seismic signal, τ_{2n}^2, changes significantly with time. In the aforementioned work of Kitagawa and

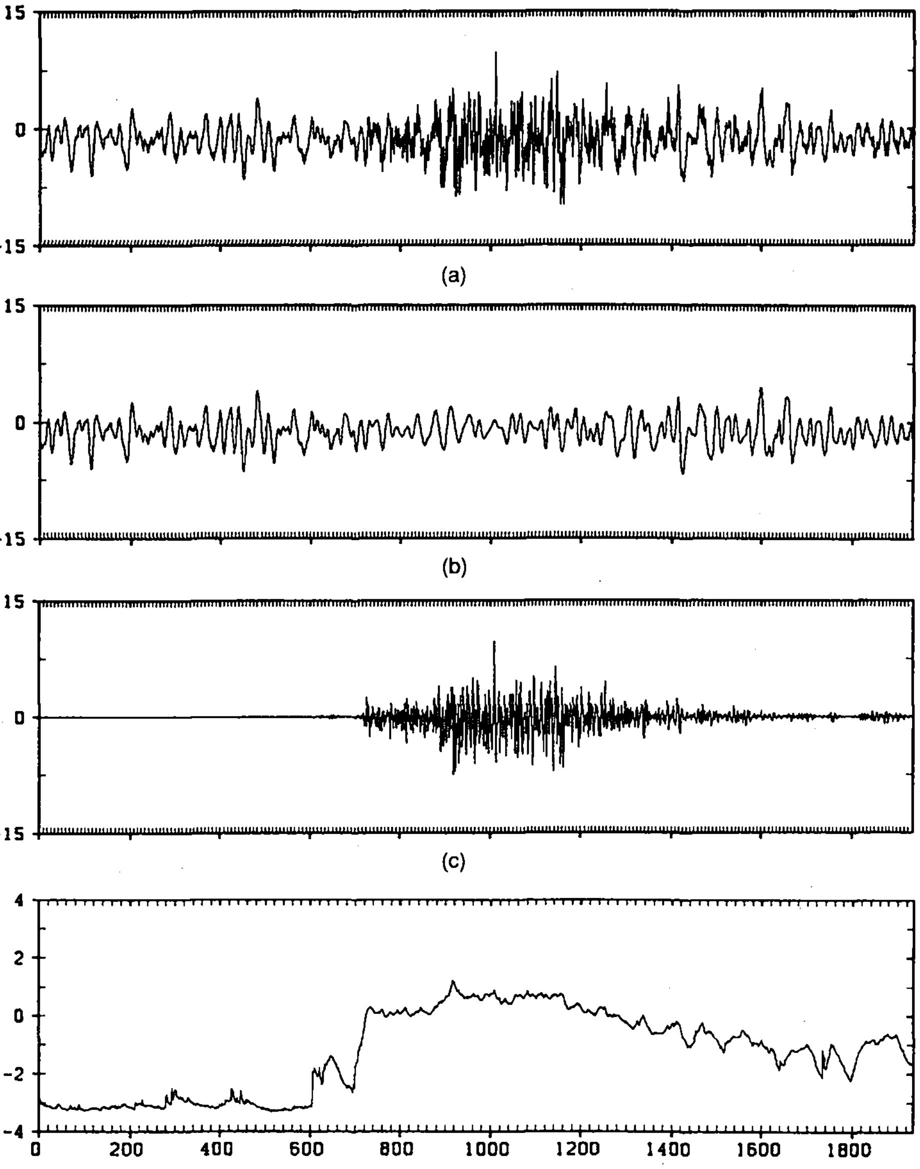

Figure 3. Decomposition of Noisy Seismic Data. (a) Observed time series; (b) posterior median of the background noise; (c) posterior median of the seismic wave; (d) posterior median of the time-varying log-variance of the system noise for the seismic component.

Takanami, a local likelihood was introduced and used to select the optimal value of the variance from a finite number of possible values. In this article, to construct the self-organizing smoother I define the augmented state vector $\mathbf{z}_n$ by

$$\mathbf{z}_n = \begin{bmatrix} \mathbf{x}_n \\ \log_{10} \tau_{2n}^2 \end{bmatrix}. \tag{33}$$

In the analysis that follows, because both of the orders m and k selected by the Akaike information criterion (AIC) were 6, the dimension of the augmented state vector was 13. Then the original state-space model (32) and the random walk model for the evolution of the logarithm of τ_{2n}^2,

$$\begin{aligned} \log_{10} \tau_{2n}^2 &= \log_{10} \tau_{2,n-1}^2 + \zeta_n, \\ \zeta_n &\sim \mathrm{N}(0, \xi^2), \end{aligned} \tag{34}$$

constitute the state-space model for the augmented state vector, $\mathbf{z}_n$.

By the Monte Carlo smoothing with the number of particles $m = 10{,}000$ and the lag length $l = 20$, the observed time series can be decomposed to the background noise r_n, the seismic signal s_n, and the observation noise w_n. The variance of the system noise for $\log_{10} \tau_{2n}^2, \xi^2 = .008$ was chosen by maximizing the log-likelihood on a coarse grid (see Fig. 4). Figure 3, (b) and (c), shows the median of the smoothed marginal posterior distributions of the background noise and the seismic signal. The background noise is almost homoscedastic in the whole interval, and also the seismic signal is clearly identified. The observation noise component is minuscule, looks homoscedastic, and is not shown here. Figure 3(d) shows the median of the marginal posterior distribution of $\log_{10} \tau_{2n}^2$. It can be seen that the variance of the innovation of the signal is mostly less than 10^{-3} before $n = 630$, increases to $1 \sim 10$ when the seis-

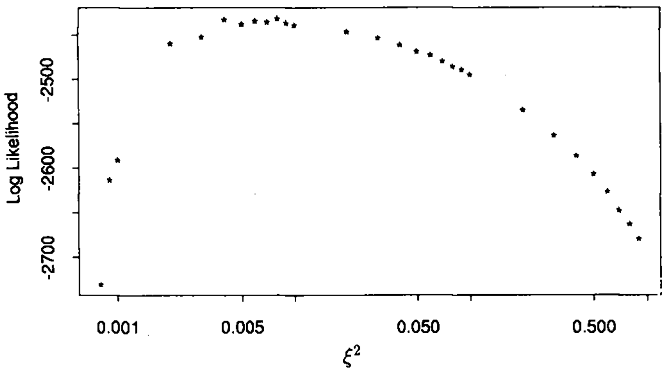

Figure 4. Log-Likelihood Versus the Variance of the System Noise ξ^2; m = 100,000.

mic signal arrived, and then gradually decreases toward the original level.

If the seismic signal is larger than that in the present example, by including individual models for the P and S waves, it is possible to decompose the observed series in a more complicated model including the background noise, the P wave, the S wave, and the observation noise.

4.3 Nonlinear Smoothing With Unknown Model Parameters

Assume that the data y_n are generated by nonlinear models,

$$x_n = \frac{1}{2}\, x_{n-1} + \frac{25x_{n-1}}{1+x_{n-1}^2} + 8\, \cos(1.2n) + v_n \tag{35}$$

and

$$y_n = \frac{x_n^2}{20} + w_n, \tag{36}$$

with $x_0 \sim \mathrm{N}(0,5), v_n \sim \mathrm{N}(0,1)$, and $w_n \sim \mathrm{N}(0,10)$. The problem is to estimate the unknown signal $x_1, \ldots, x_N$ from the observations $y_1, \ldots, y_N$ shown in Figure 5, (a) and (b). The signal process generated by (35) shifts from positive side to negative side and vice versa randomly, and the unobserved signal process x_n has two phases. But the nonlinear transformation in the observation model (36) makes it difficult to distinguish between the positive phase and the negative phase. This nonlinear smoothing problem was first considered by Andrade et al. (1979) and was reanalyzed by Carlin et al. (1992), Gordon et al. (1993), Hürzeler and Künsch (1995), Kitagawa (1991), and Kitagawa and Gersch (1996). In these papers it was shown that by a proper nonlinear smoothing algorithm, the signal x_n can be recovered. In these analyses, the parameters of the model were assumed to be known. [Carlin et al. (1992) considered the case where some of the coefficients are unknown.]

In this section I consider the situation where the most of the coefficients are unknown and use the following model:

$$x_n = \theta_{n3}x_{n-1} + \frac{\theta_{n4}x_{n-1}}{1+x_{n-1}^2} + \theta_{n5}\, \cos(1.2n) + v_n, \tag{37}$$

$$y_n = \theta_{n6}x_n^2 + w_n, \tag{38}$$

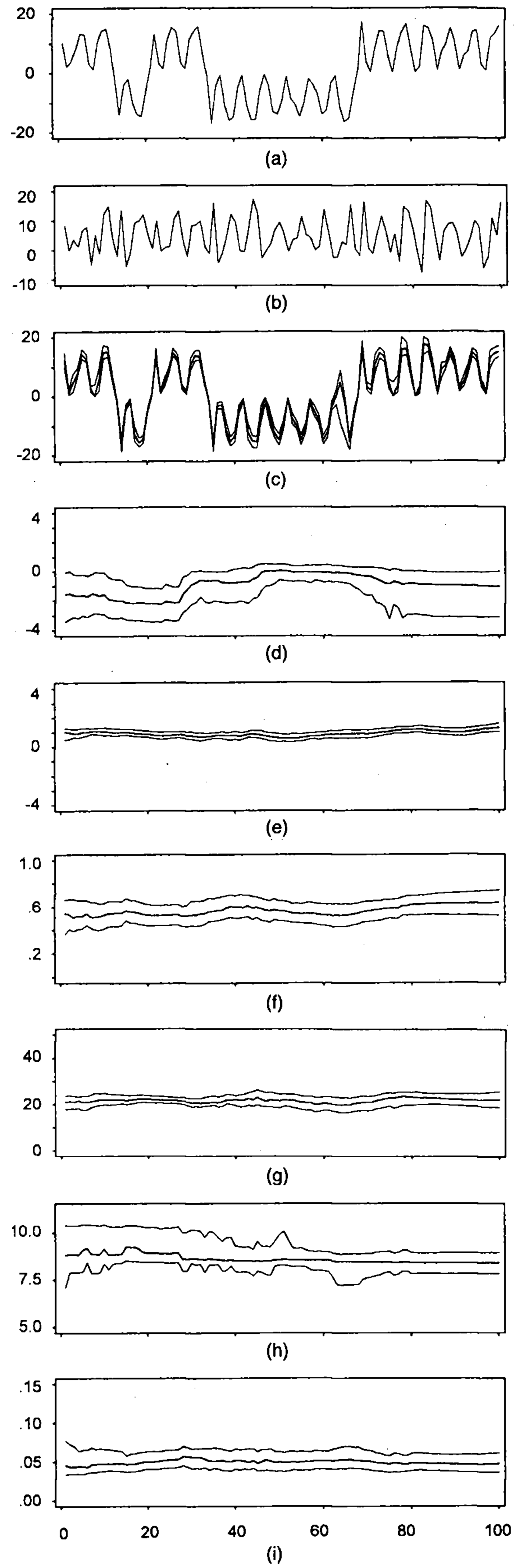

Figure 5. Nonlinear Smoothing. (a) Unobserved signal; (b) generated time series; (c) smoothed distribution of the signal; (d)–(j) smoothed distribution of the parameters $\theta_{n1} - \theta_{n6}$.

Table 2. The Variances of the Log-Likelihood for Different System Noise Dimensions, k, and the Number of Particles, m

	m			
k	1,000	10,000	100,000	1,000,000
1	.094	.010	.001	.001
3	.515	.047	.005	.001
7	1.273	.596	.051	.005

and

$$v_n \sim \mathrm{N}(0, \exp\{\theta_{n1}\}), \qquad w_n \sim \mathrm{N}(0, \exp\{\theta_{n2}\}). \quad (39)$$

For the estimation of these six unknown parameters θ_{nj}, $j = 1, \ldots, 6$, we assume the following simple random walk models

$$\theta_{nj} = \theta_{n-1,j} + u_{nj}, \qquad u_{nj} \sim \mathrm{N}(0, \xi_j^2), \qquad j = 1, \ldots, 6. \quad (40)$$

Then the self-organizing state-space model is obtained by defining the seven-dimensional state vector as

$$\mathbf{z}_n = (x_n, \theta_{n1}, \ldots, \theta_{n6})^T. \quad (41)$$

Figure 5(c) shows the median and 16% and 84% percentile points of the marginal posterior density $p(\mathbf{x}_n|Y_N)$. The details of the Monte Carlo smoother are: the number of particles m = 10,000, the lag length for smoothing $l = 20, x_0 \sim \mathrm{N}(0,5), \theta_{01} \sim U(-4,4), \theta_{02} \sim U(-4,4), \theta_{03} \sim U(0,1), \theta_{04} \sim U(0,50), \theta_{05} \sim U(5,10)$, and $\theta_{06} \sim U(0,.15)$.

It can be seen that, even with the current setting in which the six parameters are unknown, the true signal shown in Figure 5(a) is mostly recovered. The obtained posterior distribution is quite similar to the one obtained by assuming that the parameters are exactly known (Kitagawa 1991).

The marginal posterior distributions of the parameters are shown in Figure 5, (d)–(i). The variance of v_n, namely that of θ_{n1}, increased at around $n = 30$. But the distributions of the other coefficients are almost constant over time, and the medians are close to the true values given in (35) and (36).

The effect of the number of particles m to the accuracy in computing the log-likelihood is investigated with this example. In Figure 3, four boxplots on the left side show the log-likelihoods with m = 1,000, 10,000, 100,000 and 1,000,000 obtained by repeating the Monte Carlo filtering 100 times using different random numbers. It can be seen that the 50% confidence width are 2.36, .94, .33, and .09 for m = 1,000, 10,000, 100,000, and 1,000,000. Compared with the result obtained by 1,000,000 particles, the log-likelihoods computed with m = 1,000 have downward bias by about 2.36. The one with m = 10,000 and 100,000 are about .28 and .004. The variances of the log-likelihoods are shown in the last row of Table 2.

Three boxplots on the right side of Figure 6 show the log-likelihoods obtained by putting $d = 10$ and m = 1,000, 10,000 and 100,000. Here d is the number of different realization of the system noise used in prediction step given in section 3.2.

It can be seen that with this modification, 50% confidence bounds are reduced to 2.32, .74, and .20, and the bias

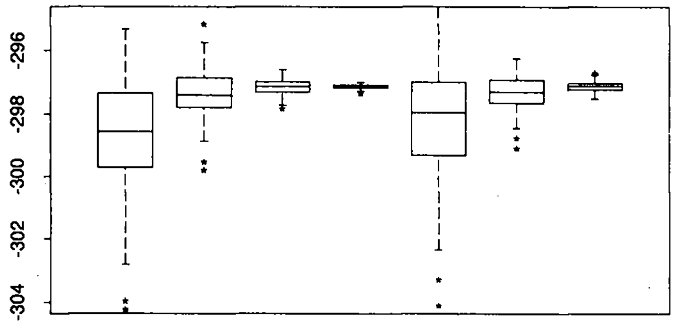

Figure 6. Boxplots of the Log-Likelihoods. (From left to right) m = 1000, 10,000, 100,000 and 1,000,000; and m = 1000, 10,000, and 100,000 with d = 10.

of the medians are .83, .19 and .003. Obviously, the log-likelihoods computed with $d = 10$ are more accurate than the ones computed with $d = 1$. But it is less accurate in variance and in bias than the ones obtained from $d = 1$ and the number of particles given by $10 \times m$. Therefore, it is better to simply increase the number of particles with $d = 1$. It is also possible to reduce the variance by repeating the Monte Carlo filtering with m particles many times and obtain the mean (or trimmed mean, etc.) of the log-likelihoods. The bias cannot be reduced by this method, however.

Table 2 shows the effect of the number of particles, m, and the dimension of the system noise, k, to the accuracy in computing the log-likelihood. The one-dimensional model assumes that all of the parameters of the model are known and only the state is unknown. In the three-dimensional model, only the variances of the noises, θ_{n1} and θ_{n2}, are assumed unknown. From Table 2 it can be seen that, as expected, the variance is reduced on the order of m^{-1}. On the other hand, the variance is, very roughly, proportional to the square of the dimension of the system noise. In this particular example, the dimension of the system noise is identical to that of the state. But, for many important applications (e.g., the second and forth examples in this article), the dimension of the state is much higher than that of the system noise. According to my experiences, in such a case the accuracy is much more dependent on the dimension of the system noise than on the state dimension.

So far the input signal to the system, $\cos(1.2n)$, was assumed to be known. Here I consider the possibility of estimating the input signal simultaneously. Figure 7 shows the change in the log-likelihood when the frequency, 1.2, is considered as a parameter. The log-likelihood becomes the

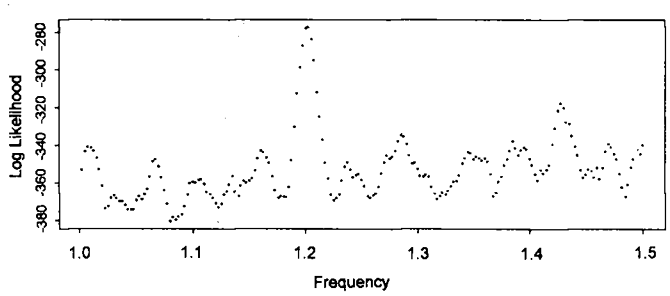

Figure 7. Log-Likelihood Versus the Frequency Parameter From Left to Right for the Input Signal, m = 100,000.

maximum at around $\theta = 1.2$, but it has many other local maxima. This indicates that the simultaneous estimation of the frequency is very difficult unless the parameter space is limited to a very small interval. It probably would be better to estimate the frequency by maximizing the log-likelihood and then apply the self-organizing state-space model for the simultaneous estimation of the state and other unknown parameters.

4.4 Nonlinear Seasonal Adjustment

Consider a multiplicative seasonal adjustment model for the monthly data,

$$y_n = T_n \times S_n \times W_n, \tag{42}$$

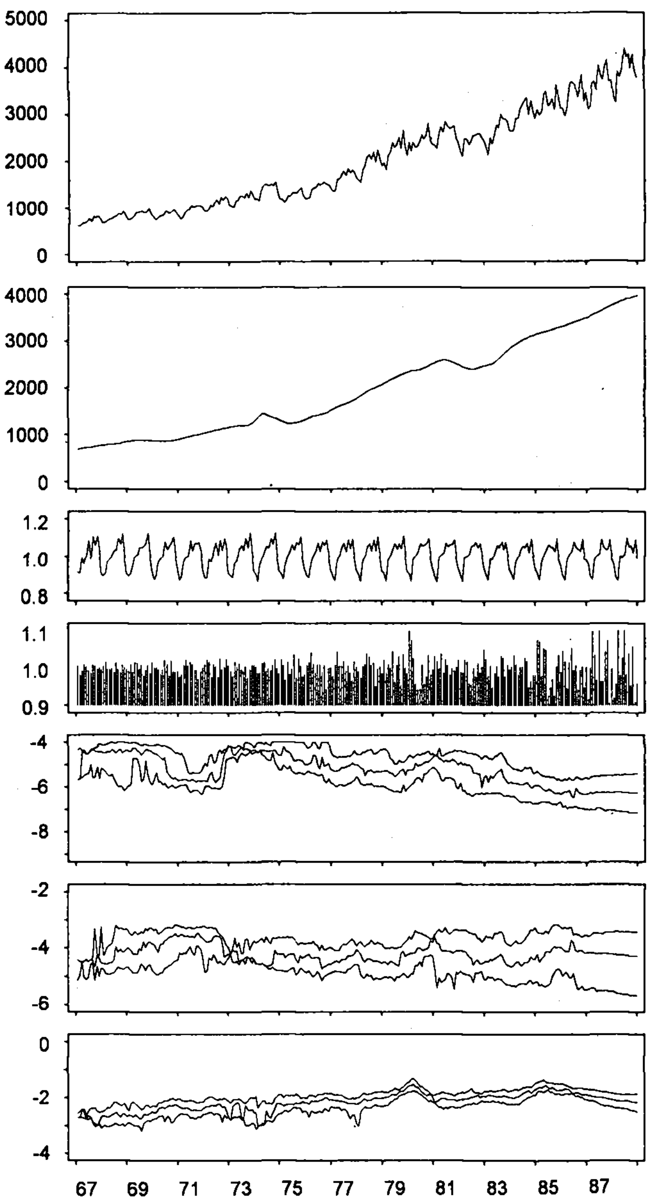

Figure 8. Seasonal Adjustment of BHDWWS (wholesale hardware sales) Data by a Nonlinear Model. (a) Original unadjusted data; (b) posterior median of the trend; (c) posterior median of the seasonal factor; (d) residual of the model; (e) median and 2σ interval of the variance of the system noise for the trend; (f) median and 2σ interval of the variance of the system noise for the seasonal factor; (g) median and 2σ interval of the variance of the observation noise.

where T_n is the trend, S_n is the seasonal factor, and W_n is the irregular component. The main purpose of this example is to show that the Monte Carlo smoother can handle nonlinear models directly. Note that if it is assumed that W_n is distributed as log-normal distribution, an almost equivalent model can be estimated by fitting the ordinary additive seasonal adjustment models after taking logarithm of the data, for example,

$$\log y_n = t_n + s_n + w_n,$$

$$t_n = 2t_{n-1} - t_{n-2} + v_n,$$

and

$$s_n = -(s_{n-1} + \cdots + s_{n-11}) + u_n, \tag{43}$$

with $w_n \sim \mathrm{N}(0, \sigma^2), v_n \sim \mathrm{N}(0, \tau_1^2)$, and $u_n \sim \mathrm{N}(0, \tau_2^2)$. This model can also be fit directly to the data y_n rather than to $\log y_n$ by a nonlinear observation model,

$$y_n = \exp\{t_n\} \times \exp\{s_n\} \times \exp\{w_n\}, \tag{44}$$

with obvious relations, $T_n = \exp\{t_n\}, S_n = \exp\{s_n\}$, and $W_n = \exp\{w_n\}$. The Monte Carlo smoother can be applied to these models with unknown σ^2, τ_1^2, and τ_2^2.

But because

$$\begin{aligned} T_n &= \exp\{t_n\} = \exp\{2t_{n-1} - t_{n-2} + v_n\} \\ &= T_{n-1}\left(\frac{T_{n-1}}{T_{n-2}}\right) V_n \end{aligned} \tag{45}$$

and

$$\begin{aligned} S_n &= \exp\{s_n\} \\ &= \exp\{u_n - (s_{n-1} + \cdots + s_{n-11})\} \\ &= \frac{U_n}{S_{n-1} \times \cdots \times S_{n-11}}, \end{aligned} \tag{46}$$

a state vector

$$\mathbf{x}_n = [\alpha_n(1), \alpha_n(2), \beta_n(1), \beta_n(2), \ldots, \beta_n(11)]^T \tag{47}$$

with

$$\alpha_n(1) = T_n, \qquad \alpha_n(2) = \frac{T_{n-1}}{T_{n-2}},$$

and

$$\beta_n(1) = S_n,$$

$$\beta_n(j) = S_{n-1} \times S_{n-2} \times \cdots \times S_{n-12+j}, \qquad j = 2, \ldots, 10, \tag{48}$$

may be considered. The relation between $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$ is nonlinear and is given by

$$\alpha_n(1) = \alpha_{n-1}(1) \times \alpha_{n-1}(2) \times V_n,$$

$$\alpha_n(2) = \alpha_{n-1}(2) \times V_n,$$

$$\beta_n(1) = \frac{U_n}{\beta_{n-1}(1) \times \beta_{n-1}(2)},$$

$$\beta_n(j) = \beta_{n-1}(1) \times \beta_{n-1}(j+1), \qquad j = 2, \ldots, 9,$$

and

$$\beta_n(11) = \beta_{n-1}(1). \tag{49}$$

To estimate these states and the unknown parameters simultaneously by the self-organizing smoother, define the augmented 16-dimensional state vector $\mathbf{z}_n$ by

$$\mathbf{z}_n = \begin{bmatrix} \mathbf{x}_n \\ \tau_{1n}^2 \\ \tau_{2n}^2 \\ \sigma_n^2 \end{bmatrix}. \tag{50}$$

The Monte Carlo filter and smoother can handle such a nonlinear non-Gaussian state-space model.

For the time evolution of the parameters, τ_{1n}^2, τ_{2n}^2, and σ_n^2, assume the same random walk model as in (40). Figure 8(a) shows the BHDWWS (wholesale hardware sales) for 1967–1988 from U.S. Bureau of the Census. Figure 8, (b)–(d), shows the median of the smoothed marginal posterior densities of the trend, the seasonal factor, and the noise components obtained by the Monte Carlo smoother with $m = 100{,}000$ and $l = 25$. The estimated trend is very smooth but clearly detects the drops in 1974 and 1981. The seasonal factor is gradually changing with time. This change of the seasonal factor can be explained as a trading day effect (Kitagawa and Gersch 1996). The variance of the observation noise component increased after 1980.

Figure 4, (e)–(g), shows the median and the corresponding effective plus and minus 2σ (about 2% and 98%) points of the smoothed marginal distribution of the logarithm of the variances, $\log_{10}\tau_{1n}^2, \log_{10}\tau_{2n}^2$, and $\log_{10}\sigma_n^2$. The variance of the system noise for the trend, τ_{n1}^2, is small in 1971–1972, 1977–1980, and 1983–1988, indicating that in these periods the trend component increased at an almost constant rate. The variance in 1985–1988 is almost $\frac{1}{100}$ of that in 1967–1970 and 1973–1975. The variance τ_{n2}^2 reflects the rate of the changes of the seasonal patterns but is not so significant as that of the trend component. The variance of the observation noise in the 1980s is almost 10 times greater than the one before 1975.

5. CONCLUSION AND DISCUSSION

A self-organizing state-space model has been proposed. Using this model, it is possible to operate the smoothing algorithms without exact knowledge of the parameters of the models. Using either the non-Gaussian filter/smoother or the Monte Carlo filter/smoother, fairly accurate approximations to the marginal posterior densities of both the state and parameters of very complicated nonlinear non-Gaussian state-space models can be obtained. The proposed method is computationally much more efficient than ordinary state-space modeling in which it is necessary to apply the filtering algorithm many times to obtain MLEs of the unknown model parameters.

The numerical examples suggest that this method yields reasonable estimates for several interesting time series problems. The effect of the selection of the initial distribution and of the number of particles are checked with these numerical examples. In these examples, the simplest random walk–type model was used for the time changes of the parameters of the models. It is possible to use other models such as higher-order trend models or autoregressive-type models. It may also possible to consider a composite model that considers component model simultaneously. This subject is left for the future study, however.

The third example suggests that the variance in computing the log-likelihood of the model is roughly of the order of k^2m^{-1}, where k is the dimension of the system noise and m is the number of particles. Thus an ad hoc way of determining m is to evaluate the variance of the log-likelihood using a small m, say 1,000, and estimate an m with which the variance becomes smaller than a certain value, say .1. Note that I usually consider the difference of AIC values less than 1; equivalently, that of the log-likelihood, less than .5, is insignificant. But a general rule for deciding m is very difficult, and a practical rule similar to that of Cowles and Carlin (1996) and Kong et al. (1994) may be possible and useful.

[Received February 1996. Revised September 1997.]

REFERENCES

Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall.

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, 439–500.

Cowless, M. K., and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Farmeir, L., and Kaufmann, H. (1991), "On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression," *Metrika*, 38, 37–60.

Frühwirth-Schnatter, S. (1994a), "Applied State-Space Modelling of Non-Gaussian Time Series Using Integration-Based Kalman Filtering," *Statistics and Computing*, 4, 259–269.

——— (1994b), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

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

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

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

Hodges, P. E., and Hale, D. F. (1993), "A Computational Method for Estimating Densities of Non-Gaussian Nonstationary Univariate Time Series," *Journal of Time Series Analysis*, 14, 163–178.

Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series" (with discussion), *Journal of the American Statistical Association*, 82, 1032–1063.

——— (1988), "Numerical Approach to Non-Gaussian Smoothing and Its Applications," in *Computing Science and Statistics, Proceedings of the 20th Symposium on the Interface*, eds. E. J. Wegman, D. T. Gantz, and J. J. Miller, pp. 379–388.

——— (1991), "A Nonlinear Smoothing Method for Time Series Analysis," *Statistica Sinica*, 1, 371–388.

——— (1993), "A Monte Carlo Filtering and Smoothing Method for Non-Gaussian Nonlinear State-Space Models," *Proceedings of the 2nd U.S.–Japan Joint Seminar on Statistical Time Series Analysis*, Honolulu, January 25–29, pp. 110–131.

——— (1994), "Two-Filter Formula for Smoothing and an Implementation of the Gaussian-Sum Smoother," *Annals of the Institute of Statistical Mathematics*, 46, 605–623.

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

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors–State-Space Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association*, 79, 378–389.

——— (1996), *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics 116, New York: Springer-Verlag.

Kitagawa, G., and Takanami, T. (1985), "Extraction of Signal by a Time Series Model and Screening Out Micro-Earthquakes," *Signal Processing*, 8, 303–314.

Kohn, R., and Ansley, C. F. (1988), "Equivalence Between Bayesian Smoothness Priors and Optimal Smoothing for Function Estimation," in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J. C. Spall, New York: Marcel Dekker, pp. 393–430.

Kong, A., Liu, J. S., and Wong, W. H. (1994), "Sequential Imputations and Bayesian Missing-Data Problems," *Journal of the American Statistical Association*, 89, 278–288.

Hürzeler, and Künsch, H. R. (1995), "Monte Carlo Approximations for General State-Space Models," Research Report 73, Seminar für Statistik, Eidgenössische Technische Hochschule (ETH).

Ljung, L. (1979), "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," *IEEE Transactions on Automatic Control*, AC-24, 36–50.

Mehra, R. K., "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, AC-15, 175–184.

Meinhold, R. J., and Singpurwalla, N. D. (1989), "Robustification of Kalman Filter Models," *Journal of the American Statistical Association*, 84, 479–486.

Schnatter, S. (1992), "Integration-Based Kalman Filtering for a Dynamic Generalized Linear Trend Model," *Computational Statistics and Data Analysis*, 13, 447–459.

Smith, R. L., and Miller, J. E. (1986), "A Non-Gaussian State-Space Model and Application to Prediction of Records," *Journal of the Royal Statistical Society*, 48, 79–88.

Solo, V. (1980), "Some Aspect of Recursive Parameter Estimation," *International Journal of Control*, 32, 395–410.

——— (1989), "Adaptive Spectral Factorization," *IEEE Transactions on Automatic Control*, AC-34, 1047–1051.

Takanami, T. (1991), "Seismograms of Foreshocks of 1982 Urakawa-Oki Earthquakes," AISM Data 43-3-01; *AISM Data Library*.

Tanizaki, H. (1993), *Nonlinear Filters*, Lecture Notes in Economics and Mathematical Systems, 400, New York: Springer-Verlag.

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

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting" (with discussion), *Journal of the American Statistical Association*, 80, 73–97.

Young, P. (1984), *Recursive Estimation and Time-Series Analysis*, Berlin, Springer-Verlag.