Statistica Sinica 1(1991), 371-388

A NONLINEAR SMOOTHING METHOD FOR TIME SERIES ANALYSIS

Genshiro Kitagawa

The Institute of Statistical Mathematics and University of Tokyo

Abstract: A nonlinear state space approach to the smoothing of time series is shown. The time series is expressed in state space model form where the system model or the observation model contains nonlinear functions of the state vector. Recursive formulas of prediction, filtering and smoothing for the nonlinear state space model are given. Numerical implementation of the formula is shown based on numerical approximation to the densities and numerical computation for the nonlinear transformation of variables, convolution of two densities, Bayes formula, and normalization. Significant merits of nonlinear state space modeling and of the proposed smoother are illustrated by two numerical examples. Empirical study on the numerical accuracy was also performed on one of the examples.

Key words and phrases: Filtering, smoothing, likelihood, AIC, nonlinear model, non-Gaussian distribution.

1. Introduction

A nonlinear smoothing methodology for time series analysis is shown here. The method is based on the general state space model and is particularly useful for time series that cannot be analysed satisfactorily by the standard linear time series models or by various linear approximation techniques.

In the analysis of nonstationary time series, the main issue has been the modeling of a time varying system. The use of a state space model is quite relevant for this purpose (e.g., Harrison and Stevens (1976), West and Harrison (1989)). In Kitagawa (1981), it was shown that the problem of modeling a nonstationary time series with drifting mean value, originally treated by a Bayesian linear model (Akaike (1980)), can be expressed in state space model form and that the convenient recursive filtering and smoothing methodology can be exploited. This method can be also applied to time-varying AR coefficient modeling and spectral estimation of nonstationary covariance time series (Kitagawa and Gersch (1985), Gersch and Kitagawa (1988) and the references therein).

All of these models can be expressed by linear Gaussian state space models and hence the conventional filtering and smoothing methods could have been successfully applied. However, there are various problems for which linear Gaussian modeling is inadequate. First, in the problem of trend estimation, the trend sometimes has jumps in addition to smooth and gradual changes. In this case, the estimate by a simple linear Gaussian model does not reflect the jumps or becomes extremely bumpy. Second, when the system has significant nonlinear characteristics, the recursive filtering algorithms based on linear approximation does not necessarily work well. In Kitagawa (1987), it was shown that filtering and smoothing formulas for non-Gaussian state space model can be realized by using a numerical method and are useful for the analysis of various types of nonstationary time series. It was also shown that this method can be easily extended to a wider class of general state space models (Kitagawa (1986)). Related papers on this subject are West et al. (1985) and Harvey and Fernandes (1989).

In this paper, we shall consider nonlinear models with the same setup. Obviously, such a situation is out of range of the well known Kalman filter and the recursive smoothing algorithms. There have been many attempts to develop filters for nonlinear systems. The extended Kalman filter, the second order filter and the Gaussian sum filter (Alspach and Sorenson (1972)) are well known approaches. All of these filters approximate the non-Gaussian distribution by one or several Gaussian distributions and are known to be satisfactory in various nonlinear problems (Anderson and Moore (1979)). However, these methods have several drawbacks. The methods based on single Gaussian density such as the extended Kalman filter may yield disastrous results when the true density is not unimodal. On the other hand, the Gaussian sum filter has various technical difficulties in actual implementation. By the numerical examples, we shall show typical phenomena related to these drawbacks.

It is thus desirable to develop a filtering and smoothing method that can handle general types of density. Here, we first show recursive filtering and smoothing formulas for a nonlinear state space model. They can be easily derived from the formulas for a general state space model (Kitagawa (1986)). We then approximate each of the probability density functions by a step function or a continuous piecewise linear function and realize the necessary operations on the density by numerical computations. This kind of direct method was attempted in an early stage of the development of nonlinear filters (Bucy and Senne (1971), de Figueiredo and Jan (1971)). Due to the recent development of fast computing facilities, however, it now becomes practical to rely on such direct numerical methods at least for lower order systems. In return for the intensive numerical computations, our method is free from Gaussian or linearity assumptions, and can also utilize a smoothing algorithm.

The objective of this article is to present a methodology for nonlinear smoothing. Specifically, we show a non-Gaussian version of filtering and smoothing formulas and also show numerical implementations of the formulas. Possible applications of our smoother are exemplified by two numerical results. In Section 2, we show a nonlinear state space model and derive formulas for recursive filtering and smoothing. Numerical implementation of these filter and smoother are shown in Section 3. The identification problem is briefly considered in Section 4. Section 5 is devoted to numerical examples where two nonlinear smoothing problems are considered and contrasted with conventional methods.

2. General State Space Model and State Estimation

Consider a system described by a nonlinear state space model

$$\begin{aligned} x_n &= g(x_{n-1}) + v_n \\ y_n &= h(x_n) + w_n, \end{aligned} \tag{1}$$

where y_n and x_n are ℓ-dimensional observation and m-dimensional state vector, respectively. v_n and w_n are m-dimensional and ℓ-dimensional white noise sequences having densities $q(v)$ and $r(w)$, respectively, which are independent of the past history of x_n and y_n. The initial state vector x_0 is assumed to be distributed according to the density $p(x_0)$. It should be noted that only (2) and (3) below are the essential requirements of the model for the following argument. Therefore, the additive error structure in (1) is not essential and the method presented here can be applied to a wider class of nonlinear models than the one formulated in (1).

The collections of the states and the observations up to time n are denoted by X_n and Y_n, namely, $X_n \equiv \{x_0, x_1, \ldots, x_n\}$ and $Y_n \equiv \{y_1, \ldots, y_n\}$. The conditional density of x_n given X_i and Y_j is denoted by $p(x_n|X_i, Y_j)$. The problem considered here is to evaluate $p(x_n|Y_j)$, the conditional density of x_n given observations Y_j. For $n > j$, $n = j$ and $n < j$, this formulates the problems of prediction, filtering and smoothing, respectively.

The above nonlinear system (1) can be expressed in general form as

$$\begin{aligned} x_n &\sim p(x_n|x_{n-1}) \\ y_n &\sim p(y_n|x_n), \end{aligned} \tag{2}$$

for which the conditional densities $p(x_n|x_{n-1})$ and $p(y_n|x_n)$ are given by

$$p(x_n|x_{n-1}) = q(x_n - g(x_{n-1})), \quad p(y_n|x_n) = r(y_n - h(x_n)).$$

Our general state space model implies that the conditional distributions satisfy the following Markov properties:

$$\begin{aligned} p(x_n|X_{n-1}, Y_{n-1}) &= p(x_n|x_{n-1}) \\ p(y_n|X_n, Y_{n-1}) &= p(y_n|x_n). \end{aligned} \tag{3}$$

Under these properties, it can be shown that the density of x_n conditional on x_{n+1} and the entire observations Y_N is reduced to

$$\begin{aligned} p(x_n|x_{n+1}, Y_N) &= p(x_n|x_{n+1}, Y_n, Y^{n+1}) \\ &= \frac{p(x_n|x_{n+1}, Y_n)p(Y^{n+1}|x_n, x_{n+1}, Y_n)}{p(Y^{n+1}|x_{n+1}, Y_n)} \\ &= p(x_n|x_{n+1}, Y_n). \end{aligned} \tag{4}$$

Here $Y^{n+1} \equiv \{y_{n+1}, \cdots, y_N\}$ and the last equality follows from the fact that given x_{n+1}, Y^{n+1} and x_n are conditionally independent.

In Kitagawa (1986), it was shown that for the general state space model with (3) and (4), the recursive formulas for obtaining one step ahead prediction, filtering and smoothing densities are given as follows:

One step ahead prediction

$$\begin{aligned} p(x_n|Y_{n-1}) &= \int_{-\infty}^{\infty} p(x_n, x_{n-1}|Y_{n-1})dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n|x_{n-1}, Y_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1}. \end{aligned} \tag{5}$$

Filtering

$$p(x_n|Y_n) = p(x_n|y_n, Y_{n-1}) = \frac{p(y_n|x_n, Y_{n-1})p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} = \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} \tag{6}$$

where $p(y_n|Y_{n-1})$ is obtained by $\int p(y_n|x_n)p(x_n|Y_{n-1})dx_n$.

Smoothing

$$\begin{aligned} p(x_n|Y_N) &= \int_{-\infty}^{\infty} p(x_n, x_{n+1}|Y_N)dx_{n+1} \\ &= \int_{-\infty}^{\infty} p(x_{n+1}|Y_N)p(x_n|x_{n+1}, Y_N)dx_{n+1} \\ &= \int_{-\infty}^{\infty} p(x_{n+1}|Y_N)p(x_n|x_{n+1}, Y_n)dx_{n+1} \\ &= p(x_n|Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1}|Y_N)p(x_{n+1}|x_n, Y_n)}{p(x_{n+1}|Y_n)} dx_{n+1} \\ &= p(x_n|Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1}|Y_N)p(x_{n+1}|x_n)}{p(x_{n+1}|Y_n)} dx_{n+1}. \end{aligned} \tag{7}$$

These formulas (5), (6) and (7) show a recursive relation between the state densities. A generalization of the filtering and the smoothing formulas for linear systems with degenerate system noise density $p(x_n|x_{n-1})$ is given in Kohn and Ansley (1987). Intuitively, they might be interpreted as follows. For the additive noise case such as the model (1), the predictor is given by the (nonlinear) convolution of two densities. The filter is proportional to the product of two densities. The smoother is much more complicated but has a similar interpretation.

For linear Gaussian systems, the conditional densities $p(x_n|Y_{n-1})$, $p(x_n|Y_n)$ and $p(x_n|Y_N)$ are characterized by the mean vectors and the covariance matrices and hence (5), (6) and (7) are equivalent to the well known Kalman filter and the fixed interval smoothing algorithms (Kalman (1960), Anderson and Moore (1979)). For nonlinear state space models, however, due to the nonlinear transformation of the state variables, the conditional density $p(x_n|Y_j)$ becomes non-Gaussian even when both v_n and w_n are Gaussian and cannot be specified by using the first two moments. In the following section, we will present a numerical method for handling non-Gaussian state densities.

3. Numerical Implementation of the Nonlinear Smoothing Formulas

A primitive but flexible method of expressing an arbitrary density function is to use a numerical approximation. In Kitagawa (1987), each density function was approximated by a continuous piecewise linear (first order spline) function. A more efficient numerical integration techniques are developed by Pole and West (1988). See also West and Harrison (1989). We shall use here a simple step function approximation, which is specified by the number of segments, k, location of nodes, z_i $(i = 0, \ldots, k)$, and the value of the density at each segment, p_i $(i = 1, \ldots, k)$. The use of this simpler approximation was motivated by the following two reasons. Firstly, according to the author's experience, the simple step function approximation is numerically stable and has sufficient accuracy. Secondly, it is quite easy to implement a nonlinear transformation of the state densities. We shall express the approximated step function by the triple $\{k, z_i, p_i\}$. Specifically, we use the following notations: $p(x_n|Y_{n-1}) \sim p_n(x_n) \equiv \{k, z_i, p_i\}$, $p(x_n|Y_n) \sim f_n(x_n) \equiv \{k, z_i, f_i\}$, $p(x_n|Y_N) \sim s_n(x_n) \equiv \{k, z_i, s_i\}$. Similarly, the system noise density $q(x)$ is discretized by using kq segments, i.e., $q(x) \sim \tilde{q}(x) \equiv \{kq, zq_i, q_i\}$. In the simplest implementation, $z_i = z_0 + (z_k - z_0)i/k$, $kq = 2k$, $zq_0 = z_0 - z_k$, $zq_{2k} = z_k - z_0$, $zq_i = zq_0 + (z_k - z_0)i/k$.

To realize the recursive formulas shown in the preceding section, it is necessary to develop a numerical method for the nonlinear transformation of variables, the convolution of densities, Bayes theorem, and the normalization.

One step ahead prediction

From (5), p_i $(i = 1, \dots, k)$ is obtained by

$$\begin{aligned} p_i = p_n(z_i) &\simeq \int_{-\infty}^{\infty} p(z_i|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1} \\ &= \int_{-\infty}^{\infty} \hat{p}(z_i|g(x_{n-1}))\hat{p}(g(x_{n-1})|Y_{n-1})dg(x_{n-1}) \\ &= \int_{-\infty}^{\infty} q(z_i - y)\hat{f}_{n-1}(y)dy = \sum_{j=1}^{k} \int_{y_{j-1}}^{y_j} \tilde{q}(z_i - y)\hat{f}_{n-1}(y)dy, \quad (8) \end{aligned}$$

with $y \equiv g(x_{n-1})$. The density for the transformed state y, $\hat{f}_{n-1}(y)$, can be obtained in various ways. If $g(x)$ is a monotone function, the density of the transformed state y is obtained by $f(g^{-1}(x_n))\frac{dg^{-1}}{dx_n}$. However, the $\hat{f}_{n-1}(y) \equiv \{k, z_i, \hat{p}_i\}$ can be evaluated numerically by the following algorithm. For simplicity, we assume that the nodes $\{z_i\}$ are equally spaced and that $\Delta z = (z_k - z_0)/k$.

1. For $i = 1$ to k
   - $\hat{f}_i = 0$
2. For $i = 1$ to k
   - (a) $y_0 = \min\{g(z_{i-1}), g(z_i)\}$
   - (b) $y_3 = \max\{g(z_{i-1}), g(z_i)\}$
   - (c) $i_0 = [\frac{y_0 - z_0}{\Delta z}]$, $i_1 = [\frac{y_3 - z_0}{\Delta z}] + 1$
   - (d) for $j = i_0 + 1$ to i_1
     - $y_1 = \max\{y_0, z_0 + (j-1)\Delta z\}$
     - $y_2 = \min\{y_3, z_0 + j\Delta z\}$
     - $\hat{f}_j = \hat{f}_j + \frac{y_2 - y_1}{y_3 - y_0} f_i$

Using this $\hat{f}_j$, (8) can be evaluated approximately by $p_i \simeq \sum_{j=1}^{k} \tilde{q}_{i-j}\hat{f}_j$.

Filtering

f_i $(i = 1, \dots, k)$ is obtained by

$$f_i = f_n(z_i) = \frac{p_n(z_i)r(y - h(z_i))}{C} = \frac{p_i r_i}{C}. \quad (9)$$

Here y is the given observation at that time and $r_i = r(y - h(z_i))$ can be evaluated directly from the function $r(w)$. In (9), C is the normalizing constant given by

$$C = \int_{-\infty}^{\infty} p_n(x)r(y-h(x))dx \simeq \sum_{i=1}^{k} \int_{z_{i-1}}^{z_i} p_n(x)r(y-h(x))dx \simeq \sum_{i=1}^{k} p_i r_i \Delta z. \quad (10)$$

Smoothing

s_i $(i = 1, \ldots, k)$ is obtained by

$$\begin{aligned} s_i = s_n(z_i) &= f_n(z_i) \int_{-\infty}^{\infty} \frac{s_{n+1}(y)\hat{q}(y - g(z_i))}{f_{n+1}(y)} dy \\ &= f_n(z_i) \sum_{j=1}^{k} \int_{z_{j-1}}^{z_j} \frac{s_{n+1}(y)\hat{q}(y - g(z_i))}{f_{n+1}(y)} dy. \end{aligned}$$

The integral in the summation is given approximately by $s_j \hat{q}_j / \hat{f}_j$, where $s_j = s_{n+1}(z_j)$, $\hat{f}_j = \hat{f}_{n+1}(z_j)$ and $\hat{q}_j$ can be evaluated either directly or numerically by a similar way as the one for the prediction.

4. Identification

The nonlinear model presented in the previous section usually has several unknown parameters. The maximum likelihood estimates of the parameters can be obtained by maximizing the log likelihood defined by

$$\ell(\theta) = \log p(y_1, \ldots, y_N) = \sum_{n=1}^{N} \log p(y_n | y_1, \ldots, y_{n-1}) = \sum_{n=1}^{N} \log p(y_n | Y_{n-1}).$$

It is interesting to note that each $p(y_n | Y_{n-1})$ is the quantity appearing in (6) and can be evaluated by (10). Therefore the log likelihood of the nonlinear model is obtained as a by-product of the nonlinear filter.

If we have several candidate models, the goodness of fit of the model can be evaluated by the value of AIC defined by (Sakamoto et al. (1986))

$$\text{AIC} = -2 \max \ell(\hat{\theta}) + 2(\text{number of free parameters}).$$

Thus the best choice of the model can be found by simply picking out the one with the smallest value of AIC. A large number of nonlinear models can be expressed in the nonlinear state space model form (1). Therefore, the log-likelihood and the AIC shown above provide a unified tool for parameter estimation and model identification.

5. Numerical Examples and Discussion

5.1. Comparison with the extended Kalman filter

We consider the data artificially generated by the following nonlinear model which was originally considered by Andrade Netto et al. (1978) and discussed in the rejoinder of Kitagawa (1987):

$$x_n = \frac{1}{2}x_{n-1} + \frac{25x_{n-1}}{1+x_{n-1}^2} + 8\cos(1.2n) + v_n$$
$$y_n = \frac{x_n^2}{20} + w_n. \tag{11}$$

The x_n and y_m shown in Figure 1 are generated by Gaussian random numbers $x_0 \sim N(0,5)$, $v_n \sim N(0,1)$ and $w_n \sim N(0,10)$. The problem is to estimate the true signal x_n from the sequence of observations $\{y_n\}$ assuming that the model (11) is known. Our nonlinear filter and smoother were applied to the problem. For comparison, the well-known extended Kalman filter, the second order filter and the linearized fixed interval smoother associated with these filters were also applied (Sage and Melsa (1971)). In the filtering and the smoothing, the following discretization is arbitrarily used: $k = 400$, $z_0 = -30$, $z_k = 30$ and $p(x_0) = N((z_0 + z_k)/2, (z_k - z_0)^2/16)$. Figure 2 shows the posterior densities $p(x_{17}|Y_m)$, $m = 16, \dots, 20$ and 100. From the left to the right each column of the figure show the results obtained by the extended Kalman filter, by the second order filter and by our nonlinear filter and smoother, respectively. This figure shows a typical situation where these algorithms yield quite different results. By our nonlinear filter, the one step ahead predictive density $p(x_{17}|Y_{16})$ is very broad and bimodal, and this bimodality extends to the filtered density $p(x_{17}|Y_{17})$ and to the smoothed density $p(x_{17}|Y_{18})$. On the other hand, the extended Kalman filter approximates each density $p(x_{17}|Y_m)$ by a single Gaussian density. Although for $m \geq 19$ the smoothed density obtained by our nonlinear smoother, $p(x_{17}|Y_m)$, also becomes unimodal and resembles a Gaussian density, its location is completely different from the one of the linearized smoother and is actually on the other side of the origin. The second order filter shown in the middle column also approximates the posterior density by a single Gaussian density. However, the estimates by the filter are very conservative and have large variances.

Figure 3 shows the trace of time n versus the smoothed posterior density $p(x_n|Y_N)$ obtained by our nonlinear smoother. In the figure the bold curve shows the 50% point of the posterior density and two fine curves express the 2.3% and 97.7% points which correspond to the two standard error interval of the Gaussian densities. + indicates the true value of x_n. It can be seen that remarkably good results were obtained by our smoother. Figure 4 shows the plot of the smoothed median of x_n minus true x_n with two standard error interval. Comparing this with Figure 5, which shows the results by the extended Kalman filter based linearized smoother, the significant merit of the nonlinear smoother can be seen.

An empirical study on the effect of the selection of the number of nodes was also performed with the same example. Table 1 shows the effect of the number of nodes, k, on the computing time and on the accuracy of the obtained posterior densities. The first column of the table shows the CPU-time in seconds spent for the computation by a main frame computer, HITAC-M682H with internal array processors (about 60MFLOPS). From this table, it can be seen that the necessary CPU-time is less than the order of k^2. The convergence of the nonlinear filter and the smoother as the number of nodes increases was checked by four criteria defined as follows:

$$\Delta F(k) = \sum_{n=1}^{N} \log p_k(y_n|Y_{n-1}) - \sum_{n=1}^{N} \log p_{4096}(y_n|Y_{n-1})$$

$$E_1(k) = \sum_{n=1}^{N} I(p_{4096}(x_n|Y_n); p_k(x_n|Y_n))$$

$$E_2(k) = \sum_{n=1}^{N} I(p_{4096}(x_n|Y_N); p_k(x_n|Y_N))$$

$$E_3(k) = \sum_{n=1}^{N} I(p_{4096}(x_n|Y_N); p_k(x_n|Y_n)).$$

Here $p_k(x_n|Y_j)$ denotes the posterior density of x_n given the observation Y_j obtained by using the approximation with k nodes, and $I(p(x); q(x))$ is the Kullback-Leibler information number of the density $p(x)$ with respect to the density $q(x)$ defined by $I(p(x); q(x)) = \int \log \frac{p(x)}{q(x)} p(x) dx$. Since the true density is unknown, the one obtained by the finest mesh ($k = 4096$) is used as the "true" density. $\Delta F(k)$ is the difference of the log-likelihood values between the approximation and the "true" density. $E_1(k)$ and $E_2(k)$ are the summation over time interval of the information numbers of the filtered and the smoothed densities, respectively. $E_3(k)$ measures the difference between the "true" smoother and the approximated filter.

From the table, it can be seen that the difference of the log-likelihood is less than 0.1 for $k \geq 512$. However, in view of the experience that in comparing several models, the approximated log likelihood fluctuates similarly, a coarser mesh with $k = 128$ might be sufficient for model identification. $E_1(k)$ and $E_2(k)$ both converge to zero apparently with order $O(k^2)$. On the other hand, the $E_3(k)$ seems to converge to a constant. This value shows a significant advantage of the smoothing over the filtering formula. It is worth mentioning that the estimates by the smoothing formula with coarse mesh ($k = 64$) has an accuracy equivalent to that of estimation by the filtering formula. In the table, the extended Kalman filter is also evaluated. It can be seen that the extended Kalman filer is by far worse than our nonlinear filter or smoother with the coarsest mesh.

In summary, this example reveals two important points in the nonlinear filtering problem:

1. The extended Kalman filter and any other filter that approximate the density by a single Gaussian density may produce disastrous results when the true density is not unimodal.
2. The information from future observations is quite important to single out the location of the state. The difference of $p(x_{17}|Y_{19})$ and $p(x_{17}|Y_{17})$ and the value of $E_3(k)$ clearly demonstrate this. Thus the use of smoother is essential to get a good estimate of the state.

5.2. Two-dimensional problem

The second example is a passive receiver problem. A similar problem was considered by Bucy and Senne (1971) and Alspach and Sorenson (1972). In this example, the problem is to locate a target in two-dimensional space which is gradually moving.

This target is observed according to the scalar nonlinear measurement model

$$y_n = h(x_n^1, x_n^2) + w_n \tag{12}$$

where

$$h(x_n^1, x_n^2) = \tan^{-1}\left\{\frac{x_n^2 - \sin\beta_n}{x_n^1 - \cos\beta_n}\right\}, \quad \beta_n = \beta_{n-1} + \Delta\beta. \tag{13}$$

Here β_0 and $\Delta\beta$ are given constants and w_n is a Gaussian white noise with known variance σ^2. This is a simple example of the vector tracking problem of locating a moving object by observing the relative angle observed on a rotating observatory. Figure 6 shows two examples of the trajectory. (For Case 1, $x_n^1 = 22\cos t_n - 9$, $x_n^2 = 30\sin t_n - 11$, and for Case 2, $x_n^1 = 1.5\cos s_n + 22\cos t_n - 9$, $x_n^2 = 1.5\sin s_n + 30\sin t_n - 11$ with $t_n = (30 + 0.08n)/180\pi$, $s_n = (30 + 6n)/180\pi$). Figure 7 shows two artificially generated series y_n which are obtained by observing these trajectories according to the measurement model (12) and (13) with $\beta_0 = 0$, $\Delta\beta = 1$, $w_n \sim N(0, 0.02^2)$ for Case 1 and $w_n \sim N(0, 0.01^2)$ for Case 2. For the estimation of this moving object, we consider the following smoothness prior model (Kitagawa and Gersch (1984), Gersch and Kitagawa (1988)):

$$\begin{aligned} \Delta^d x_n^1 &= v_n^1 \\ \Delta^d x_n^2 &= v_n^2. \end{aligned} \tag{14}$$

Here the difference order d is either 1 or 2 and v_n^1 and v_n^2 are mutually independent Gaussian white noise sequence with variances, τ_1^2 and τ_2^2, respectively. The smoothness prior model (14) with the observation model (12) constitutes our nonlinear state space model for estimating the location of the object. It should be noted that the Gaussianity of neither v_n nor w_n are essential in our model. The value of τ_1^2 and τ_2^2 are estimated by maximizing the log-likelihood defined by (11). For simplicity in maximizing likelihood, we assumed that $\tau_1^2 = \tau_2^2 = \tau^2$ for the second order model. Table 2 shows the maximum likelihood estimates of τ^2, τ_1^2 and τ_2^2 and the associated log-likelihoods and the AICs. The discretization used in the computation are summarized in Table 3. The initial state density, $p(x_0)$, was arbitrarily set to the two- or four-dimensional Gaussian density with the ith mean $(z_0^i + z_{k_i}^i)/2$ and the diagonal covariance matrix with the ith diagonal element given by $(z_{k_i}^i - z_0^i)^2/16$. Figure 8 shows the contour of the posterior density $p(x_n^1, x_n^2|Y_N)$ for $n = 20$, 40, 60 and 80 for Case 2. Figure 9 shows the trace of (p_{1n}^j, p_{2n}^j), where p_{1n}^j and p_{2n}^j $(j = 1, \dots, 7)$ are the 0.13%, 2.27%, 15.87%, 50%, 84.13%, 97.73%, 99.87% points of the marginal posterior density of $p(x_n^1|Y_N)$ and $p(x_n^2|Y_N)$, respectively. In non-Gaussian case, except for the 50% point (namely $j = 4$), these points do not have particular meaning. However, having the contour lines in Figure 8 in mind, we can imagine the move of the posterior density on 2-D space.

For the first order trend model $(m=2, k_1=k_2=200, n=100)$, the necessary CPU-time was 66 seconds. On the other hand, for the second order trend model $(m = 4)$, the filtering with a coarse mesh $(k_1 = k_3 = 50, k_2 = k_4 = 7)$, $(k_1 = k_3 = 70, k_2 = k_4 = 11)$, and $(k_1 = k_3 = 100, k_2 = k_4 = 15)$, took 98, 590 and 2617 seconds, respectively.

Acknowledgment

The author is grateful to the referee and the associate editor for their valuable comments which improved the presentation of the paper. In particular, the simple proof of the equation (4) given by the referee is gratefully acknowledged. This research was supported in part by a grant from the Japanese Ministry of Education, Science and Culture.

Table 1. Monte Carlo study on the effect of the number of nodes (k), to the CPU time and the accuracy

Table 2. Summary statistics of the fitted models to the 2-D tracking problem

Table 1

k	CPU	ΔF	E_1	E_2	E_3
32	0.10	58.4	80.7	144.0	483.9
64	0.20	6.21	16.1	46.8	64.6
128	0.42	3.59	4.26	12.5	51.7
256	0.94	1.42	1.02	3.30	47.5
512	2.33	0.39	.234	.758	43.9
1024	6.55	0.21	.048	.152	42.3
2048	20.69	0.02	.007	.022	41.7
4096	71.80	—	—	—	—
EK	.005	650.5	2114.7	2572.9	2238.1

Table 2

Case	2D model	4D model
1	LL=223.36 AIC=−442.71 $\tau_1^2 = 0.032$ $\tau_2^2 = 0.014$	LL=231.69 AIC=−461.39 $\tau^2 = 0.12 \times 10^{-5}$
2	LL=239.39 AIC=−474.78 $\tau_1^2 = 0.041$ $\tau_2^2 = 0.031$	LL=260.65 AIC=−518.30 $\tau^2 = 0.0013$

Table 3. Summary of the discretizing parameters used the nonlinear filter and smoother

Case	2D model	4D model
1	$k_1 = k_2 = 200$ $z_0^1 = 6, \quad z_{k_1}^1 = 14$ $z_0^2 = 2, \quad z_{k_2}^2 = 10$	$k_1 = k_3 = 100$ $k_2 = k_4 = 15$ $z_0^1 = 6, \quad z_{k_1}^1 = 14$ $z_0^2 = -0.013, \quad z_{k_2}^2 = -0.021$ $z_0^3 = 2, \quad z_{k_3}^3 = 10$ $z_0^4 = 0.031, \quad z_{k_4}^4 = 0.039$
2	$k_1 = k_2 = 200$ $z_0^1 = 5, \quad z_{k_1}^1 = 15$ $z_0^2 = 2, \quad z_{k_2}^2 = 12$	$k_1 = k_3 = 100$ $k_2 = k_4 = 15$ $z_0^1 = 5, \quad z_{k_1}^1 = 15$ $z_0^2 = -0.2, \quad z_{k_2}^2 = 0.2$ $z_0^3 = 2, \quad z_{k_3}^3 = 12$ $z_0^4 = -0.2, \quad z_{k_4}^4 = 0.2$

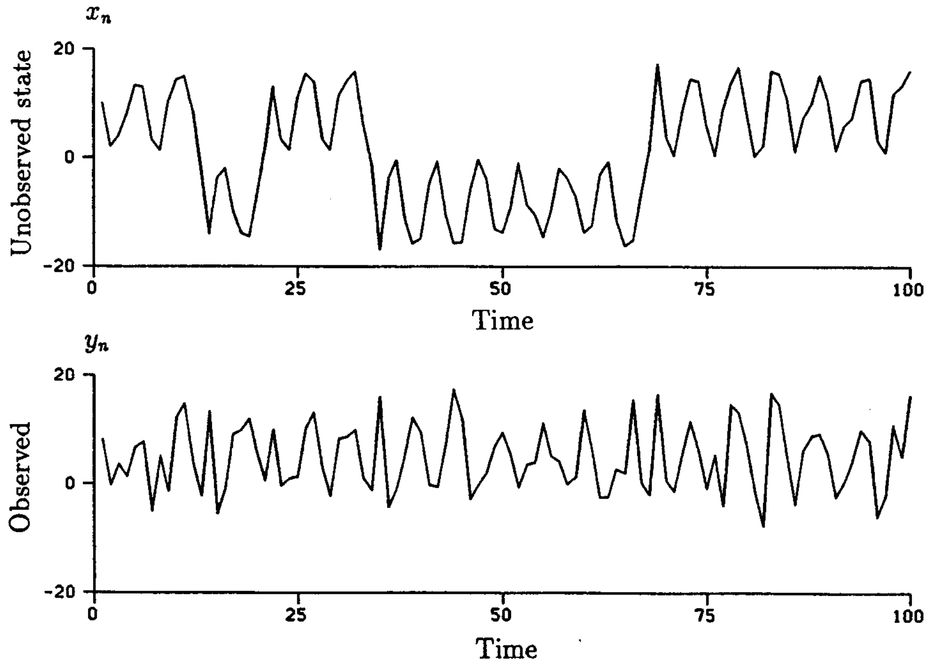

Figure 1. True signal x_n, $n = 1, \ldots, 100$, and the observations y_n, $n = 1, \ldots, 100$.

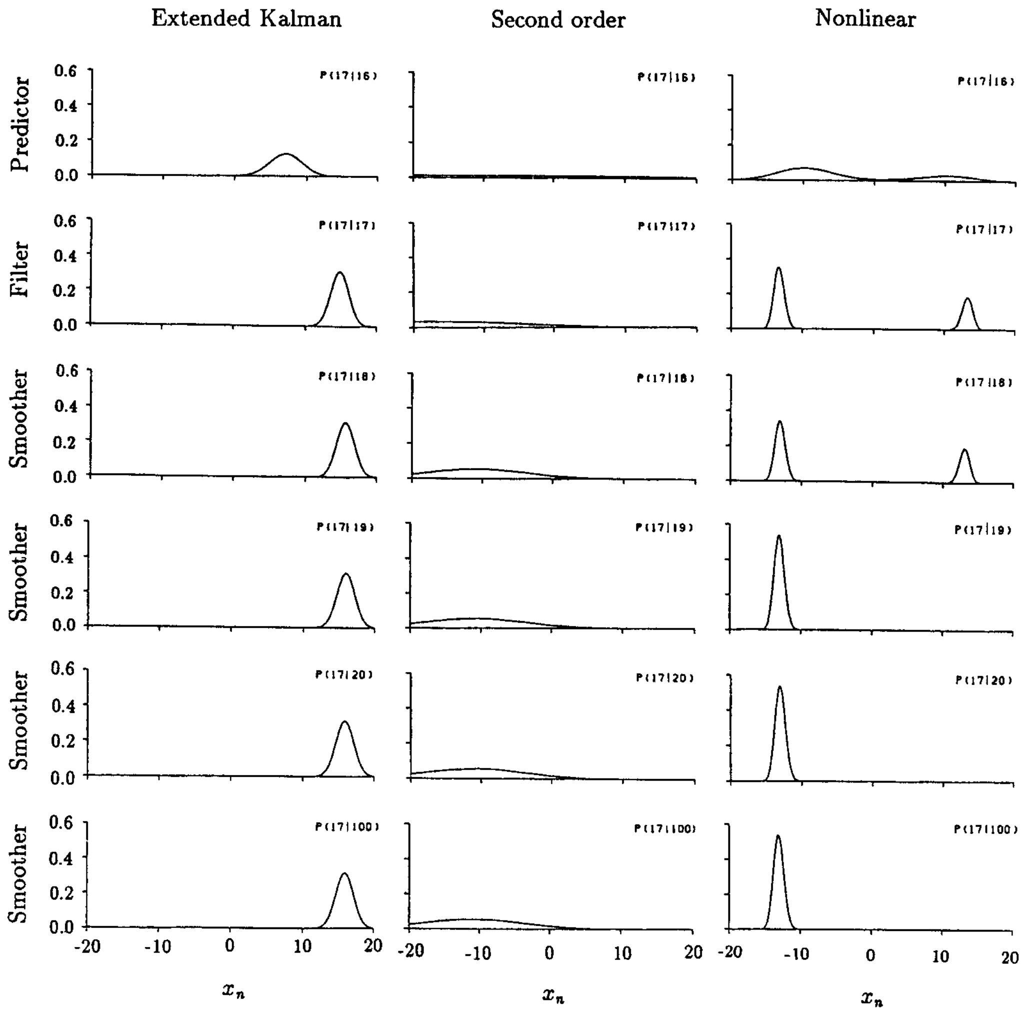

Figure 2. Posterior densities $p(x_{17}|Y_m)$, $m = 16, \ldots, 20$ and 100 obtained by the extended Kalman filter based smoother (left), the second order filter based smoother (middle) and our nonlinear smoother (right).

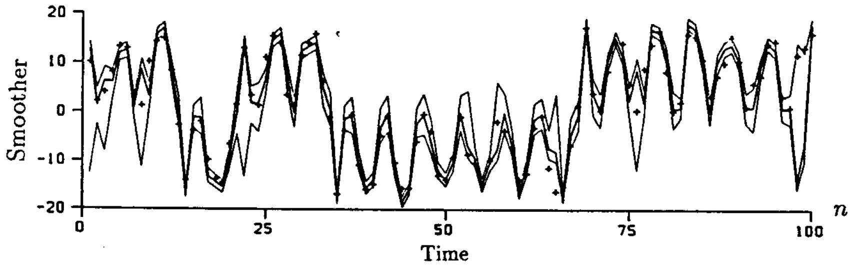

Figure 3. Posterior density $p(x_n|Y_N)$ obtained by our nonlinear smoother. The bold curve shows the median and the fine curves show two standard error interval. + indicates the true value.

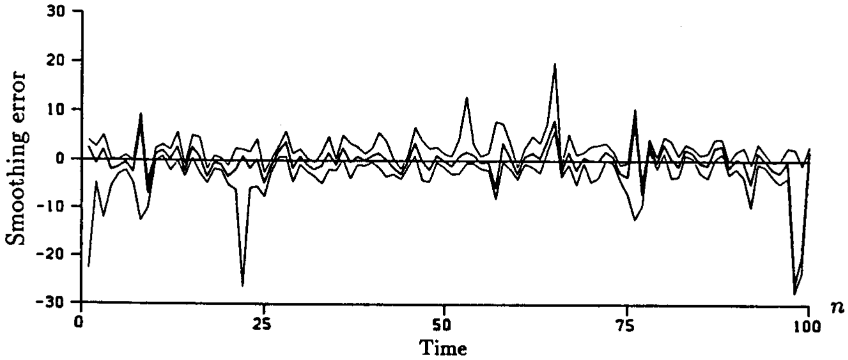

Figure 4. Posterior medians minus true values with two standard error intervals for the nonlinear smoother.

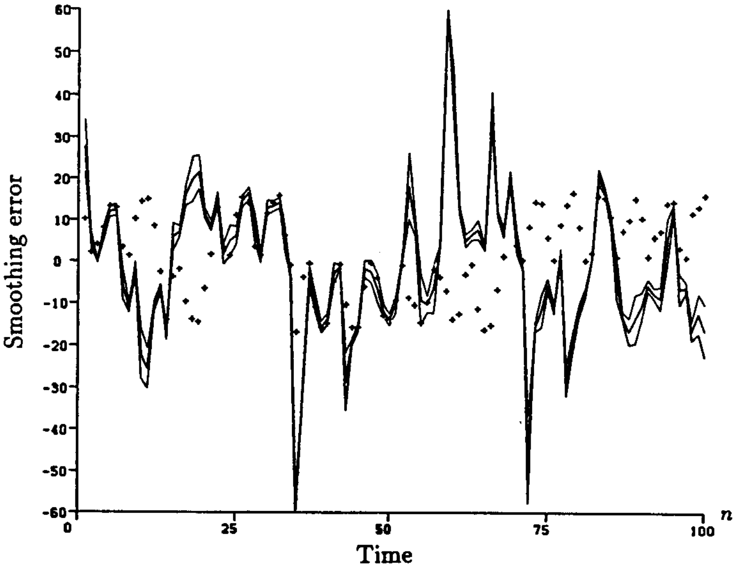

Figure 5. Posterior medians minus true values with two standard error intervals for the extended Kalman filter/smoother.

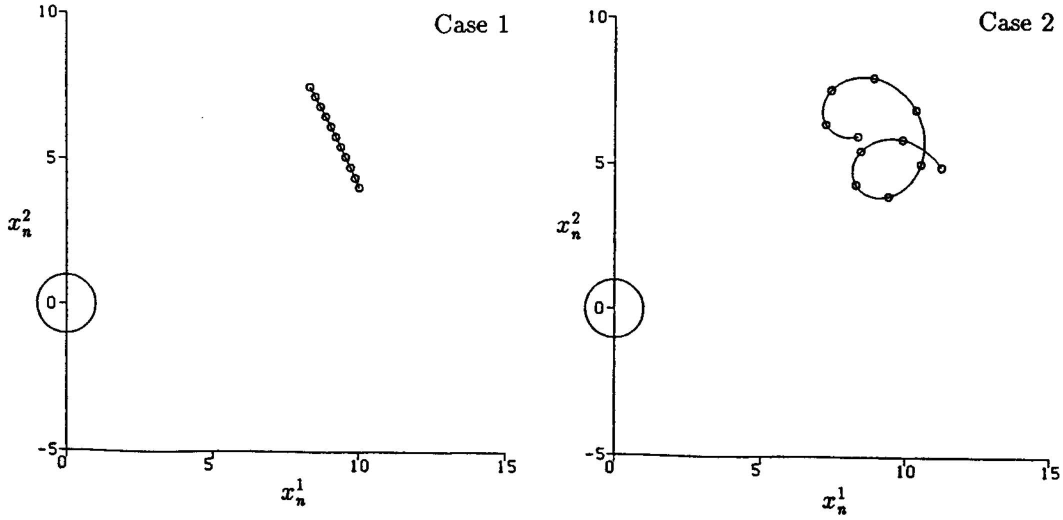

Figure 6. Trajectory of the moving object and the receiver: Case 1 and Case 2.

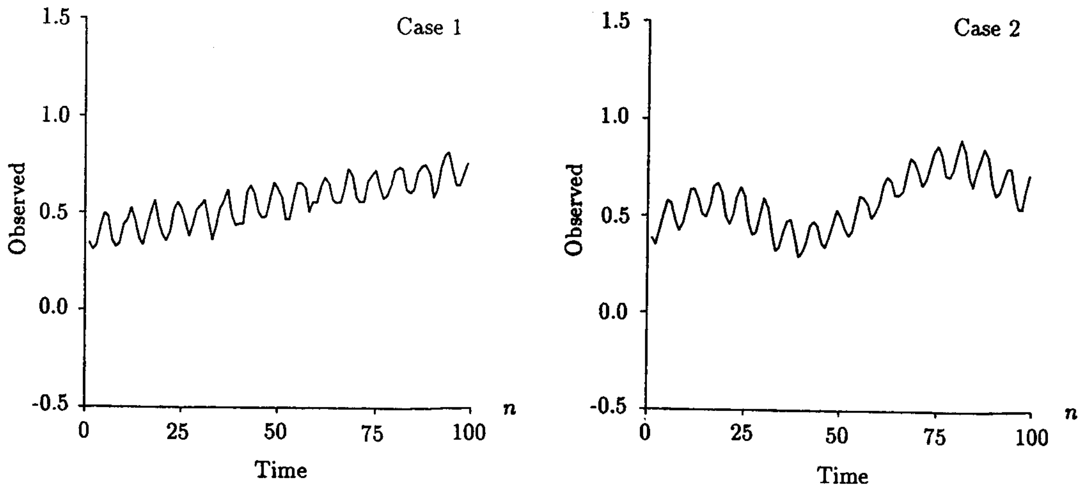

Figure 7. Observed angle y_n: Case 1 and Case 2.

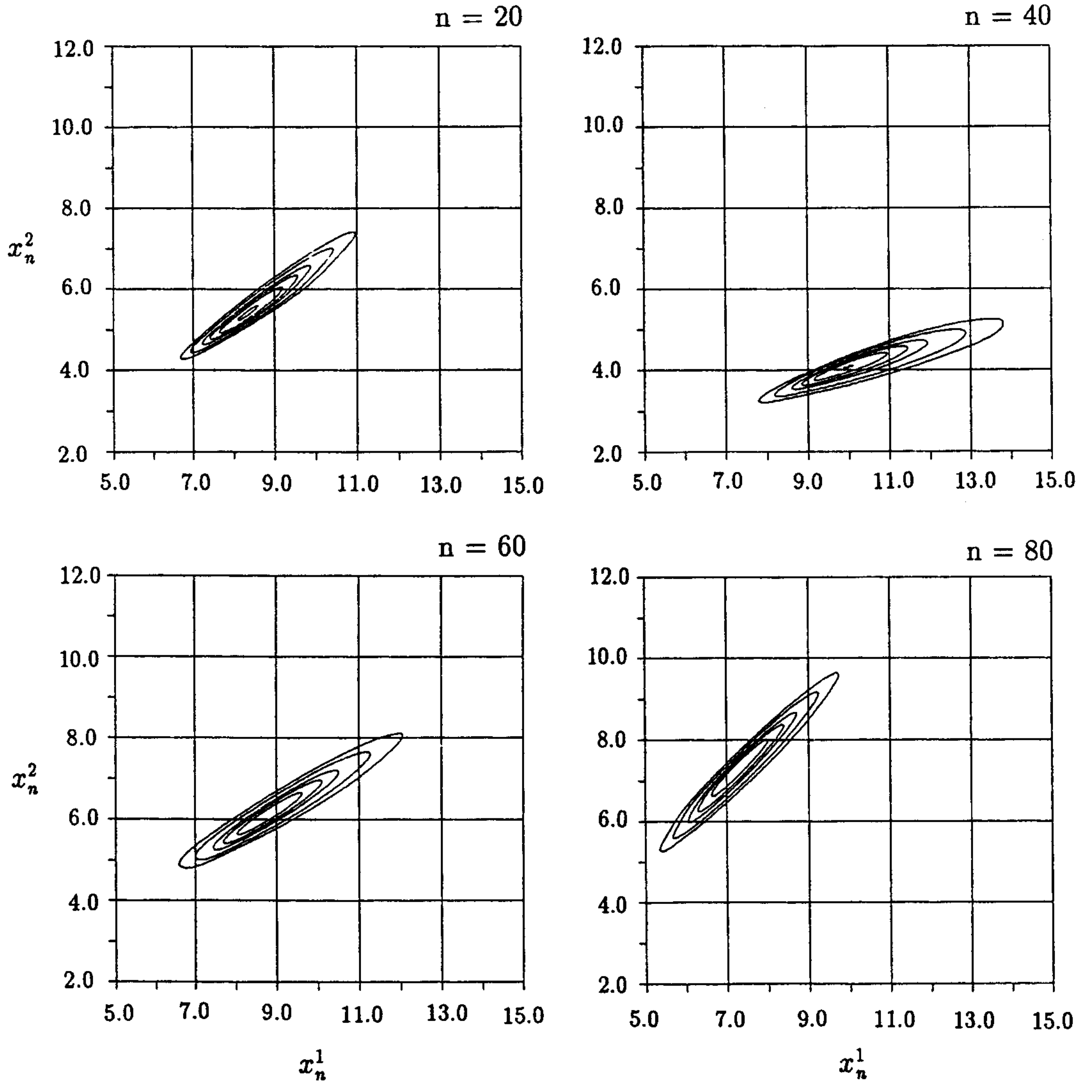

Figure 8. Contour of the posterior densities, $p(x_n^1, x_n^2|Y_N)$, $n = 20, 40, 60$ and 80 for Case 2.

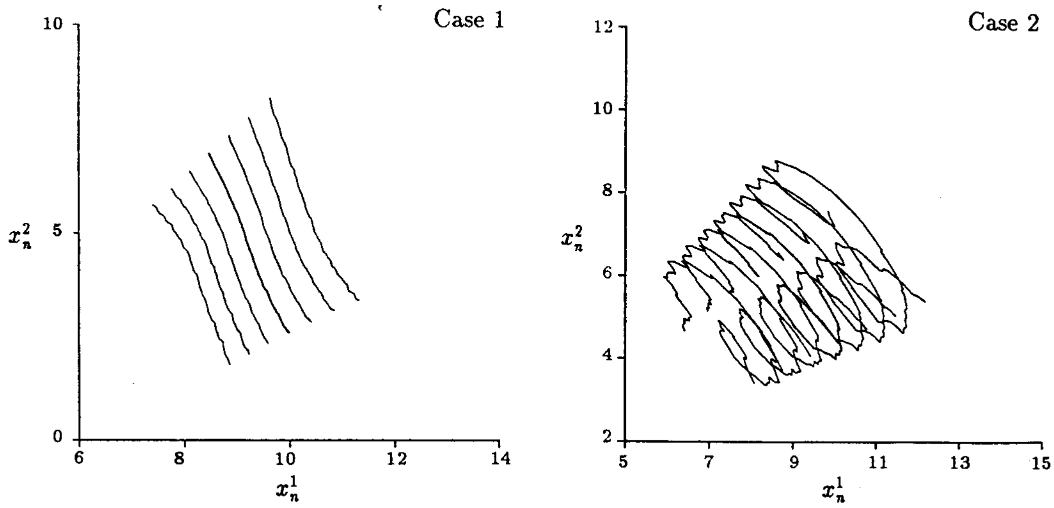

Figure 9. Trace of (p^j_{n1}, p^j_{n2}) where p^j_{n1} and p^j_{n2} are 0.13, 2.27, 15.87, 50, 84.13, 97.73 and 99.87 percentile points of the marginal posterior density of $p(x_n^1|Y_N)$ and $p(x_n^2|Y_N)$.

References

Akaike, H. (1980). Likelihood and Bayes procedure. In *Bayesian Statistics* (Edited by J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith), 143–166. University Press, Valencia, Spain.

Alspach, D. L. and Sorenson, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Automat. Control* **17**, 439–448.

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

Andrade Netto, M. L., Gimeno, L. and Mendes, M. J. (1978). On the optimal and suboptimal nonlinear filtering problem for discrete-time systems. *IEEE Trans. Automat. Control* **23**, 1062–1067.

Bucy, R. S. and Senne, K. D. (1971). Digital synthesis of nonlinear filters. *Automatica* **7**, 287-289.

de Figueiredo, R. J. P. and Jan, Y. G. (1971). Spline filters. *Proceedings of the 2nd Symposium on Nonlinear Estimation Theory and its applications*, 127–141, San Diego.

Gersch, W. and Kitagawa, G. (1988). Smoothness priors in time series. In *Bayesian Analysis of Time Series and Dynamic Models* (Edited by J. C. Spall), 431–476. Marcel Dekker, New York.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser.B* **38**, 205–247.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations (with discussion). *J. Bus. & Econom. Statist.* **7**, 407–422.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME* (*J. Basic Engineerings*) **82D**, 35–45.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *J. Time Ser. Anal.* **2**, 103–116.

Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality. *J. Amer. Statist. Assoc.* **79**, 378–389.

Kitagawa, G. and Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. *IEEE Trans. Automat. Control* **30**, 48–56.

Kitagawa, G. (1986). Non-Gaussian smoothness prior approach to irregular time series analysis. *Adaptive Systems in Control and Signal Processing 1986* (Edited by K. J. Åström and B. Wittenmark), 303–308. Pergamon Press, Oxford.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion). *J. Amer. Statist. Assoc.* **82**, 1032–1063.

Kohn, R. and Ansley, C. F. (1987). Discussion of the paper by Kitagawa. *J. Amer. Statist. Assoc.* **82**, 1041–1044.

Pole, A. and West, M. (1988). Efficient Numerical Integration in Dynamic Models. Warwick Research Report, No. 136, Department of Statistics, University of Warwick.

Sage, A. P. and Melsa, J. L. (1971). *Estimation Theory with Applications to Communications and Control.* McGraw-Hill Series in Systems Science, McGraw-Hill, New York.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics.* D. Reidel Publishing Company, Tokyo.

West, M. and Harrison, P. J. (1989). *Bayesian Forecasting and Dynamic Models.* Springer Series in Statistics, Springer-Verlag, New York.

West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Amer. Statist. Assoc.* **80**, 73–97.

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku Tokyo, Japan 106.
and
Department of Economics, the University of Tokyo, Japan.

(Received December 1989; accepted January 1991)