Ann. Inst. Statist. Math.
Vol. 46, No. 4, 605–623 (1994)

THE TWO-FILTER FORMULA FOR SMOOTHING AND AN IMPLEMENTATION OF THE GAUSSIAN-SUM SMOOTHER

GENSHIRO KITAGAWA

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

(Received January 4, 1993; revised February 4, 1994)

Abstract. A Gaussian-sum smoother is developed based on the two filter formula for smoothing. This facilitates the application of non-Gaussian state space modeling to diverse problems in time series analysis. It is especially useful when a higher order state vector is required and the application of the non-Gaussian smoother based on direct numerical computation is impractical. In particular, applications to the non-Gaussian seasonal adjustment of economic time series and to the modeling of seasonal time series with several outliers are shown.

Key words and phrases: Non-Gaussian smoother, non-Gaussian filter, Gaussian mixture, nonstationary time series, outliers, seasonal adjustment.

1. Introduction

This paper is addressed to an implementation of the Gaussian-sum smoother. In recent years, the non-Gaussian state space modeling has become popular in time series analysis. See for example, West and Harrison (1989) and the references therein. Non-Gaussian modeling is especially useful in the analysis of nonstationary time series with abrupt changes in structure, in handling outliers and in the analysis of discrete or nonlinear processes.

Kitagawa (1987) showed a non-Gaussian smoothing algorithm implementation based on numerical approximations to the relevent probability distributions. That method has diverse applications. However, since that method involves computationally extensive numerical integration, the application of the method to problems with higher state dimension (say more than five dimensions) is impractical. On the other hand, there are some problems that require a higher dimensional state vector. For example, a standard analysis of seasonal adjustment of monthly data time series, requires a 13 dimensional state vector. Our previous experience of modeling seasonal time series (Kitagawa and Gersch (1984), Kitagawa (1989)) motivated the work reported here. Kitagawa (1989) used a state space model with Gaussian mixture system noise or observation noise density to handle abrupt changes in the trend or seasonal component or outliers in the seasonal data. Earlier work using

Gaussian mixtures (Sorenson and Alspach (1971), Alspach and Sorenson (1972), Harrison and Stevens (1976) and Anderson and Moore (1979)) did not address smoothing. Of course, the smoothing problem is critically important in statistical data analysis. In Kitagawa (1989), a high dimensional state vector (40 states) was used to get an approximate smoother. This achieved significant improvement of the seasonal and trend component estimates, at the expense of costly computations in both CPU time and memory.

On the basis of that experience, we were motivated to develop a Gaussian-sum smoother which directly yields the smoothing estimates without augmenting the state vector. In this paper, we show a Gaussian-sum smoother based on a non-Gaussian version of two-filter formula for smoothing. Consequently, it is now practical to use high dimensional non-Gaussian state space models.

The plan of the paper is as follows. In Section 2, several existing filtering and smoothing algorithms are briefly shown. Section 3 develops a non-Gaussian version of the two filter formula for smoothing. In Section 4, the Gaussian-sum smoother is developed from this two filter formula. Section 5 is devoted to some numerical problems related to the implementation of the Gaussian-sum smoother. Sections 6–8 are devoted to various numerical examples for the illustration of the proposed algorithm.

2. A brief review of the filtering and smoothing algorithms

2.1 *The state space model and the state estimation problems*

Assume that a time series y_n is expressed by a possibly non-Gaussian state space model

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n v_n, \\ y_n &= H_n x_n + w_n \end{aligned} \tag{2.1}$$

where x_n is an d_x-dimensional state vector, v_n and w_n are d_v-dimensional and 1-dimensional white noise sequences having density functions $q_n(v)$ and $r_n(w)$, respectively. The initial state vector x_0 is assumed to be distributed according to the density $p(x_0)$. The information from the observations up to time j is denoted by Y_j, namely, $Y_j \equiv \{y_1, \ldots, y_j\}$. The problem of state estimation is to evaluate $p(x_n \mid Y_j)$, the conditional density of x_n given the observations Y_j and the initial density $p(x_0 \mid Y_0) = p(x_0)$. For $n > j$, $n = j$ and $n < j$, these are the problems of prediction, filtering and smoothing, respectively.

2.2 *The Kalman filter and the smoother*

It is well known that if all of the noise densities $q_n(v)$ and $r_n(w)$ and the initial state density $p(x_0)$ are Gaussian, then the conditional density $p(x_n \mid Y_m)$ is also Gaussian and that the mean and the covariance may be obtained by the Kalman filter and the fixed interval smoothing algorithms (Anderson and Moore (1979)).

To be specific, if we put $q_n(v) \sim N(0, Q_n)$, $r_n(w) \sim N(0, R_n)$, $p(x_0 \mid Y_0) \sim N(x_{0|0}, V_{0|0})$ and $p(x_n \mid Y_m) \sim N(x_{n|m}, V_{n|m})$, then the Kalman filter is given as follows:

One-step ahead prediction:

$$
\begin{aligned}
x_{n|n-1} &= F_n x_{n-1|n-1}, \\
V_{n|n-1} &= F_n V_{n-1|n-1} F_n^t + G_n Q_n G_n^t.
\end{aligned} \tag{2.2}
$$

Filter

$$
\begin{aligned}
K_n &= V_{n|n-1} H_n^t (H_n V_{n|n-1} H_n^t + R_n)^{-1}, \\
x_{n|n} &= x_{n|n-1} + K_n (y_n - H_n x_{n|n-1}), \\
V_{n|n} &= (I - K_n H_n) V_{n|n-1}.
\end{aligned} \tag{2.3}
$$

Using these estimates, the smoothed density is obtained by the following, Fixed interval smoothing algorithm:

$$
\begin{aligned}
A_n &= V_{n|n} F_n^t V_{n+1|n}^{-1}, \\
x_{n|N} &= x_{n|n} + A_n (x_{n+1|N} - x_{n+1|n}), \\
V_{n|N} &= V_{n|n} + A_n (V_{n+1|N} - V_{n+1|n}) A_n^t.
\end{aligned} \tag{2.4}
$$

2.3 *The non-Gaussian filter and the smoother*

In Kitagawa (1987), it was shown that for the state space model (2.1) with non-Gaussian white noise v_n and w_n, the recursive formulas for obtaining the densities of the one step ahead predictor, the filter and the smoother are as follows:

One step ahead prediction:

$$
p(x_n \mid Y_{n-1}) = \int_{-\infty}^{\infty} p(x_n \mid x_{n-1}) p(x_{n-1} \mid Y_{n-1}) dx_{n-1}. \tag{2.5}
$$

Filtering:

$$
p(x_n \mid Y_n) = \frac{p(y_n \mid x_n) p(x_n \mid Y_{n-1})}{p(y_n \mid Y_{n-1})} \tag{2.6}
$$

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

Smoothing:

$$
p(x_n \mid Y_N) = p(x_n \mid Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1} \mid Y_N) p(x_{n+1} \mid x_n)}{p(x_{n+1} \mid Y_n)} dx_{n+1}. \tag{2.7}
$$

In Kitagawa (1987, 1988), an algorithm for implementing the non-Gaussian filter and smoother was developed by approximating each density function based on a continuous piecewise linear function and by performing numerical computations. This method was successfully applied to lower order systems (see Kitagawa (1987, 1988)).

2.4 *The Gaussian-sum filter*

For higher order state space models such as the one for the seasonal adjustment of monthly data however, the application of this direct numerical method is impractical due to the huge amount of computation involved in numerical integration.

One practical way to mitigate this computational burden is the use of a Gaussian-sum filter (Sorenson and Alspach (1971), Alspach and Sorenson (1972), Harrison and Stevens (1976) and Anderson and Moore (1979)). In Kitagawa (1989), it was shown that such a Gaussian-sum filter can be easily derived from the non-Gaussian filter by using Gaussian mixture approximations to the related densities. Specifically, the following approximations were used:

$$(2.8)\qquad \begin{aligned} p(v_n) &= \sum_{i=1}^{K_v} \alpha_i \varphi_i(v_n), & p(w_n) &= \sum_{j=1}^{K_w} \beta_j \varphi_j(w_n), \\ p(x_n \mid Y_{n-1}) &= \sum_{k=1}^{L_n} \gamma_{kn} \varphi_k(x_n \mid Y_{n-1}), & p(x_n \mid Y_n) &= \sum_{\ell=1}^{M_n} \delta_{\ell n} \varphi_\ell(x_n \mid Y_n). \end{aligned}$$

Here φ_i denotes a properly defined Gaussian density. Substituting these into (2.5) and (2.6), we immediately obtain the following algorithm for the Gaussian-sum filtering.

One step ahead prediction:

$$(2.9)\qquad \begin{aligned} p(x_n \mid Y_{n-1}) &= \sum_{i=1}^{K_v} \sum_{\ell=1}^{M_{n-1}} \gamma_{i\ell,n} \varphi_{i\ell}(x_n \mid Y_{n-1}) \\ &\equiv \sum_{k=1}^{L_n} \gamma_{k,n} \varphi_k(x_n \mid Y_{n-1}). \end{aligned}$$

Here $\gamma_{i\ell,n} = \alpha_i \delta_{\ell,n-1}$ and $\varphi_{i\ell}(x_n \mid Y_{n-1})$ is the one-step-ahead predictor of x_n obtained under the assumptions that the filter of x_{n-1} is $\varphi_\ell(x_{n-1} \mid Y_{n-1}) \sim N(x^\ell_{n-1|n-1}, V^\ell_{n-1|n-1})$ and that $p(v_n) = \varphi_i(v_n) \sim N(0, Q_i)$. Therefore, $\varphi_{i\ell}(x_n \mid Y_{n-1})$ is also Gaussian and its mean and covariance can be obtained by the ordinary Kalman predictor:

$$(2.10)\qquad \begin{aligned} x^{i\ell}_{n|n-1} &= F_n x^\ell_{n-1|n-1}, \\ V^{i\ell}_{n|n-1} &= F_n V^\ell_{n-1|n-1} F^t + G_n Q_i G^t_n. \end{aligned}$$

In (2.9), $\equiv$ means renumbering the double summation by a single summation, and therefore $L_n = K_v M_{n-1}$.

Filtering:

$$(2.11)\qquad \begin{aligned} p(x_n \mid Y_n) &\propto \sum_{j=1}^{K_w} \sum_{k=1}^{L_n} \delta_{jk,n} \varphi_{jk}(x_n \mid Y_n) \\ &\equiv \sum_{\ell=1}^{M_n} \delta_{\ell n} \varphi_\ell(x_n \mid Y_n). \end{aligned}$$

Here, $\delta_{jk,n} = \beta_j \gamma_{kn} \varphi_{jk}(y_n \mid Y_{n-1})$, $M_n = K_w L_n$ and $\varphi_{jk}(x_n \mid Y_n) \sim N(x^{jk}_{n|n}, V^{jk}_{n|n})$ is the filter of x_n obtained under the assumption that the one-step-ahead predictor of x_n is $N(x^k_{n|n-1}, V^k_{n|n-1})$ and $p(w_n) = \varphi_j(w_n) \sim N(0, \sigma_j^2)$. Therefore, it can be obtained by the following Kalman filter:

$$\begin{aligned} K_n^{jk} &= V^k_{n|n-1} H_n^t (H_n V^k_{n|n-1} H_n^t + \sigma_j^2)^{-1}, \\ x^{jk}_{n|n} &= x^k_{n|n-1} + K_n^{jk}(y_n - H_n x^k_{n|n-1}), \\ V^{jk}_{n|n} &= (I - K_n^{jk} H_n) V^k_{n|n-1}. \end{aligned} \tag{2.12}$$

3. The two-filter formula for smoothing

As shown briefly in the previous section, the non-Gaussian filter naturally yields the Gaussian-sum filter which can be applied to higher order state space models. To date however, the corresponding Gaussian-sum smoother has not been developed. The problem, in the derivation of the Gaussian-sum smoother, is due to the presence of division by non-Gaussian densities in (2.7). That spoils the effectiveness of the Gaussian-sum approximation. Therefore, another formula that does not explicitly contain division by a non-Gaussian density is motivated. In this section, such a formula is derived.

First, we note that the smoothed density $p(x_n \mid Y_N)$ can be expressed as follows:

$$\begin{aligned} p(x_n \mid Y_N) &= p(x_n \mid Y_{n-1}, Y^n) \\ &= p(x_n, Y^n \mid Y_{n-1}) p(Y^n \mid Y_{n-1})^{-1} \\ &= p(x_n \mid Y_{n-1}) p(Y^n \mid x_n, Y_{n-1}) p(Y^n \mid Y_{n-1})^{-1} \\ &= p(x_n \mid Y_{n-1}) p(Y^n \mid x_n) p(Y^n \mid Y_{n-1})^{-1}. \end{aligned} \tag{3.1}$$

Here $Y^n \equiv \{y_n, \dots, y_N\}$ is the information from the present and the future observations. Since $p(x_n \mid Y_{n-1})$ has been already given by filtering and $p(Y^n \mid Y_{n-1})$ is a constant which does not depend on x_n, the smoothed density can be obtained if $p(Y^n \mid x_n)$ is given. Here this term can be evaluated by the following backward filtering.

Initialization

$$p(Y^N \mid x_N) = p(y_N \mid x_N). \tag{3.2}$$

Backward filtering

$$\begin{aligned} p(Y^{n+1} \mid x_n) &= \int_{-\infty}^{\infty} p(Y^{n+1}, x_{n+1} \mid x_n) dx_{n+1} \\ &= \int_{-\infty}^{\infty} p(Y^{n+1} \mid x_{n+1}, x_n) p(x_{n+1} \mid x_n) dx_{n+1} \\ &= \int_{-\infty}^{\infty} p(Y^{n+1} \mid x_{n+1}) p(x_{n+1} \mid x_n) dx_{n+1}, \end{aligned} \tag{3.3}$$

$$\begin{aligned} p(Y^n \mid x_n) &= p(Y^{n+1}, y_n \mid x_n) \\ &= p(Y^{n+1} \mid x_n) p(y_n \mid Y^{n+1}, x_n) \\ &= p(Y^{n+1} \mid x_n) p(y_n \mid x_n). \end{aligned} \tag{3.4}$$

The quantity $p(Y_n \mid x_n)$ is not a state conditional density. Rather, it is an entity that is similar to the likelihood, and with care, it can be evaluated formally by the backward filtering algorithm. If $p(Y^n \mid x_n)$ as a function of x_n can be considered as a non-degenerate density function of x_n, then the backward filter is identical to the ordinary filter. Unfortunately, this does apply for the initial several points. For example, for many important applications, the observation model

$$y_N = Hx_N + w_N \tag{3.5}$$

cannot be directly inverted to get a distribution for x_N since the rank of H is generally less than the dimension of the state vector x_n. This exhibits the difficulty in initializing the backward filter. One practical way to avoid this problem is to use $\tau_0^2 I$ for some large τ_0^2 as the initial covariance of x_{N+1} and that actually works for many problems. However, starting from x_{N-m+1}, the repeated use of (2.1) yields

$$\begin{bmatrix} y_N \\ \vdots \\ y_{N-m+2} \\ y_{N-m+1} \end{bmatrix} = \begin{bmatrix} HF^{m-1} \\ \vdots \\ HF \\ H \end{bmatrix} x_{N-m+1} + \begin{bmatrix} G & \cdots & F^{m-2}G \\ & \ddots & \vdots \\ & & G \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_{N-1} \\ \vdots \\ v_{N-m+1} \end{bmatrix} + \begin{bmatrix} w_N \\ \vdots \\ w_{N-m+2} \\ w_{N-m+1} \end{bmatrix}. \tag{3.6}$$

That can be expressed as

$$y_{N-m+1}^N = H_{(m)} x_{N-m+1} + G_{(m)} v_{N-m+1}^{N-1} + w_{N-m+1}^N, \tag{3.7}$$

using an obvious notation. Therefore, by taking m so that $H_{(m)}$ is non-singular, we can invert this model to get the expression for x_{N-m+1} as

$$x_{N-m+1} = H_{(m)}^{-1} y_{N-m+1}^N - H_{(m)}^{-1} G_{(m)} v_{N-m+1}^{N-1} - H_{(m)}^{-1} w_{N-m+1}^N. \tag{3.8}$$

Here denoting the j-th column of $H_{(m)}^{-1} G_{(m)}$ and $H_{(m)}^{-1}$ by g_{mj} and h_{mj}, respectively, x_{N-m+1} can be expressed as

$$x_{N-m+1} = H_{(m)}^{-1} y_{N-m+1}^N - \sum_{j=1}^{m-1} g_{mj} v_{N-j} - \sum_{j=1}^{m} h_{mj} w_{N-j+1}. \tag{3.9}$$

It should be noted that in this case the backward filter is started from x_{N-m+1} rather than x_N. A more practical way, although not exact but is sufficiently acculate, is to use the initial distribution $p(x_0)$ even for x_N.

4. Gaussian-sum smoother

We now consider the development of a Gaussian-sum smoother. Since the Gaussian-sum version of the filter has been already given in (2.9)–(2.12), it suffices to show the implementation of smoothing (3.1).

Assume that $p(x_n \mid Y_{n-1})$ and $p(Y^n \mid x_n)$ are expressed by

$$\begin{aligned} p(x_n \mid Y_{n-1}) &= \sum_{k=1}^{L_n} \gamma_{kn} \varphi_k(x_n \mid Y_{n-1}), \\ p(Y^n \mid x_n) &= \sum_{\ell=1}^{M'_n} \delta'_{\ell n} \varphi_\ell(Y^n \mid x_n), \end{aligned} \tag{4.1}$$

where $\varphi_k(x_n \mid Y_{n-1}) \equiv N(x^k_{n|n-1}, V^k_{n|n-1})$ and $\varphi_\ell(Y_n \mid x_n) \equiv N(z^\ell_{n|n}, U^\ell_{n|n})$. Then by analogy to the derivation of the Gaussian-sum filter (2.11), for the state space model

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n v_n, \\ z_n &= x_n + w_n, \end{aligned} \tag{4.2}$$

the Gaussian-sum smoother is obtained by

$$\begin{aligned} p(x_n \mid Y_N) &\propto p(Y^n \mid x_n) p(x_n \mid Y_{n-1}) \\ &= \sum_{\ell=1}^{M'_n} \sum_{k=1}^{L_n} \delta'_{\ell n} \gamma_{kn} \varphi_\ell(Y^n \mid x_n) \varphi_k(x_n \mid Y^{n-1}) \\ &\equiv \sum_{\ell=1}^{M'_n} \sum_{k=1}^{L_n} \delta'_{\ell n} \gamma_{kn} \varphi_{\ell k}(x_n \mid Y_N). \end{aligned} \tag{4.3}$$

Here $\varphi_{\ell k}(x_n \mid Y_N)$ is the Gaussian density whose mean and the covariance are obtained by replacing H_n by I and y_n by $z^\ell_{n|n}$,

$$\begin{aligned} J^{\ell k}_n &= V^k_{n|n-1}(V^k_{n|n-1} + U^\ell_{n|n})^{-1}, \\ x^{\ell k}_{n|N} &= x^k_{n|n-1} + J^{\ell k}_n (z^\ell_{n|n} - x^k_{n|n-1}), \\ V^{\ell k}_{n|N} &= (I - J^{\ell k}_n) V^k_{n|n-1}. \end{aligned} \tag{4.4}$$

5. Remarks on the Gaussian mixture approximation

Some remarks on the Gaussian mixture approximation are in order here. In this Gaussian mixture approximation, each Gaussian component can be evaluated by the Kalman filter. This is the most appealing feature of this method. However, this method has a severe drawback in that the number of necessary Gaussian components increases exponentially with time. If the number of Gaussian components used for the noise densities are K_v and K_w, respectively, then the number of necessary Gaussian components for the state density at the n-th step is $(K_v \cdot K_w)^n$ times that of the initial density. In this section, we will show a method to overcome this problem.

5.1 *Reduction of the number of Gaussian components*

To avoid the explosion of the number of Gaussian components, we re-approximate the densities by a reduced (collapsed) number of Gaussian components at each time step (Harrison and Stevens (1976)). The re-approximation is motivated by the observation that a relatively small number of Gaussian densities can approximate a large class of distributions and by the anticipation that the complexity of the density will not increases very significantly with the evolution of the time step.

A practical way is to approximate the prediction and filtering densities by a fixed number of components at each step of the recursion such as:

$$\begin{aligned} p(x_n \mid Y_{n-1}) &= \sum_{k=1}^{m_{pn}} \gamma_{kn}\varphi_k(x_n \mid Y_{n-1}) \longrightarrow \sum_{k=1}^{m} \gamma'_{kn}\varphi'_k(x_n \mid Y_{n-1}), \\ p(x_n \mid Y_n) &= \sum_{\ell=1}^{m_{fn}} \delta_{\ell n}\varphi_\ell(x_n \mid Y_n) \longrightarrow \sum_{\ell=1}^{m} \delta'_{\ell n}\varphi'_\ell(x_n \mid Y_n). \end{aligned} \tag{5.1}$$

In principle, this can be realized by finding the minimizer of a measure of the dissimilarity between the true and approximated densities. For that purpose, we use the Kullback-Leibler information number,

$$I(p(\cdot);q(\cdot)) = \int \log \frac{p(x)}{q(x)} p(x)dx. \tag{5.2}$$

Here $p(x)$ and $q(x)$ are the true and the approximated densities, respectively. This evaluation generally involves computationally costly numerical integration and nonlinear optimization with respect to many parameters. It is not practical, especially for higher order models. A seemingly appealing method of adopting Gaussian terms with largest m probabilities γ_{kn} or δ_{ln} was found to be inefficient. This is mainly due to the fact that even if the weight of a Gaussian component is very small at a certain time point, it may become large at the next time step. Typically that can be seen for a Gaussian component with large variance when an outlier or the jump of the parameter occurs. Ignoring such a component will be disasterous. From our experience, a practical way is to measure the similarity of two Gaussian components of filtered densities by

$$\begin{aligned} D(\varphi_k, \varphi_l) &= -2\delta_k\delta_\ell\{I(\varphi_k;\varphi_l) + I(\varphi_l;\varphi_k)\} \\ &= \delta_k\delta_\ell\{V_k^{-1}V_l + V_l^{-1}V_k + (\mu_k - \mu_l)^t(V_k^{-1} + V_l^{-1})(\mu_k - \mu_l)\} \end{aligned} \tag{5.3}$$

and pool the pair of Gaussian densities, φ_k and φ_l, that minimizes $D(\varphi_k, \varphi_l)$. Here μ_k and V_k are the mean and the covariance of the Gaussian component φ_k.

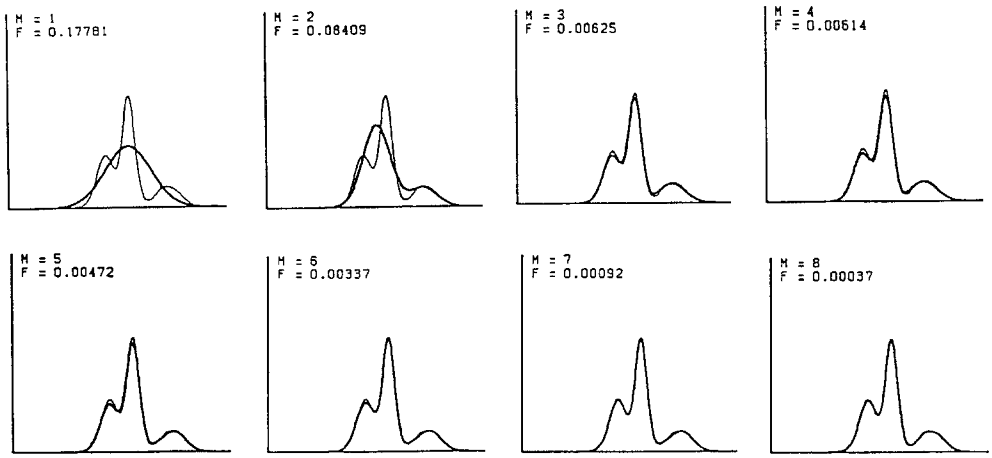

Fig. 1. Gaussian-sum approximations with reduced number of Gaussian components. Bold curves show approximated and the fine curves show the true density.

The pooling of two Gaussian densities, φ_k and φ_l, can be realized by the following calculations:

$$\begin{aligned}
\mu_{kl} &= \frac{\delta_{kn}\mu_k + \delta_{ln}\mu_l}{\delta_k + \delta_l}, \\
V_{kl} &= \frac{\delta_k\{V_k + (\mu_{kl} - \mu_k)(\mu_{kl} - \mu_k)^t\} + \delta_l\{V_l + (\mu_{kl} - \mu_l)(\mu_{kl} - \mu_l)^t\}}{\delta_k + \delta_l}.
\end{aligned} \tag{5.4}$$

By repeating this procedure, we can get an approximation to the conditional state density with a fixed number of Gaussian components.

In implementing this procedure, we systematically examine all possible pairs of Gaussian components. The use of the symmetrized Kullback-Leibler measure, (5.3), is of course, ad-hoc. A different measure might be optimal. In what follows however, we do show that it is a very effective approach.

5.2 *An illustrative example*

Here, we illustrate how our procedure works in reducing the number of Gaussian components. A worked example is illustrated in Fig. 1. The true density $g(y)$ was arbitrarily defined using 16 Gaussian components. In Fig. 1, the fine curves show the true density $g(y)$. The bold curves show the Gaussian-sum approximations with m components, obtained by successively applying the indicated procedure. It can be seen that fairly reasonable approximations are obtained for $m \geq 3$. For $m > 8$, two curves are visually indistinguishable and thus they are not shown.

The second column of Table 1 shows the Kullback-Leibler information numbers of the Gaussian-sum approximations with m components, obtained by the indicated method. The third column shows the K-L information numbers of the best Gaussian-sum approximations. They are obtained by using the computationally costly numerical integration method and numerical optimization in $(3m-1)$-dimensional parameter space. It is quite time consuming and is impractical to

Table 1. Kullback-Leibler numbers of the Gaussian-sum approximation with smaller number of components. The second column shows the ones for the approximation by the proposed method. The third column shows the ones for the best approximations.

m	approximation	best
15	0.00000	0.00000
14	0.00000	0.00000
13	0.00011	0.00000
12	0.00012	0.00000
11	0.00013	0.00000
10	0.00013	0.00000
9	0.00013	0.00000
8	0.00037	0.00001
7	0.00092	0.00012
6	0.00337	0.00012
5	0.00472	0.00013
4	0.00614	0.00013
3	0.00625	0.00258
2	0.08409	0.06967
1	0.17781	0.17781

use in filtering and smoothing algorithms. The table shows that the increase of the K-L information number with the decreasing number of Gaussian components induced by the approximations can be considerable. The example also reveals the merit of using the Gaussian mixture as compared with the single approximating Gaussian density.

6. Comparison with direct numerical approximations

In this section, we will check the validity of our implementation of the Gaussian-sum smoother by comparing with the estimates obtained by the non-Gaussian smoother for the state space model (Kitagawa (1987))

$$x_n = x_{n-1} + v_n, \qquad y_n = x_n + w_n. \tag{6.1}$$

In that paper, it was assumed that $w_n \sim N(0, \sigma^2)$ and $q(v) \propto (v^2+\tau^2)^{-b}$. However, for comparison with the Gaussian-sum smoother, we assume here that $q(v)$ is a mixture of two Gaussian densities defined by

$$q(v) = \alpha N(0, \tau_1^2) + (1-\alpha)N(0, \tau_2^2). \tag{6.2}$$

The maximum likelihood estimates of the parameters are $\hat{\alpha} = 0.991$, $\hat{\tau}_1^2 = 0.00013$ and $\hat{\sigma}^2 = 1.03$. τ_2^2 is arbitrarily fixed to 4. Figure 2 shows the marginal posterior densities obtained by the (almost) exact non-Gaussian smoother. The bold curve

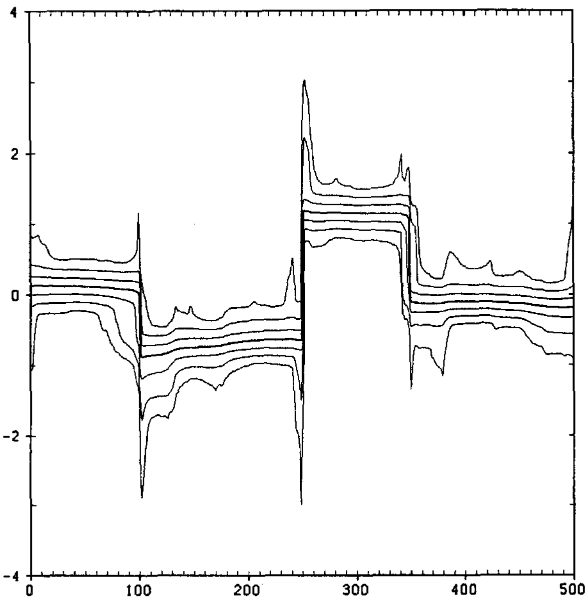

Fig. 2. The marginal posterior densities obtained by the non-Gaussian smoother.

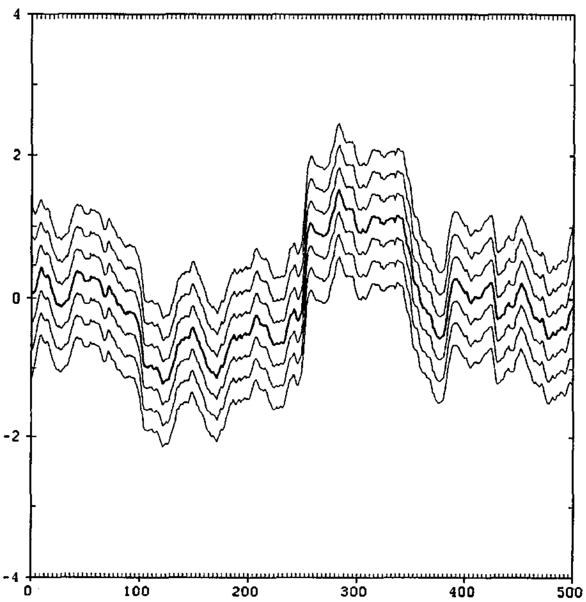

Fig. 3. The marginal posterior densities obtained by an equivalent Gaussian smoother.

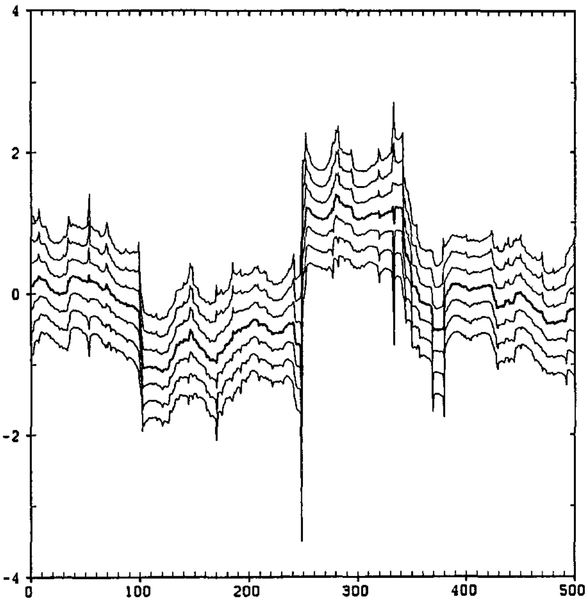

Fig. 4. The marginal posterior densities obtained by the Gaussian-sum smoother ($m = 1$).

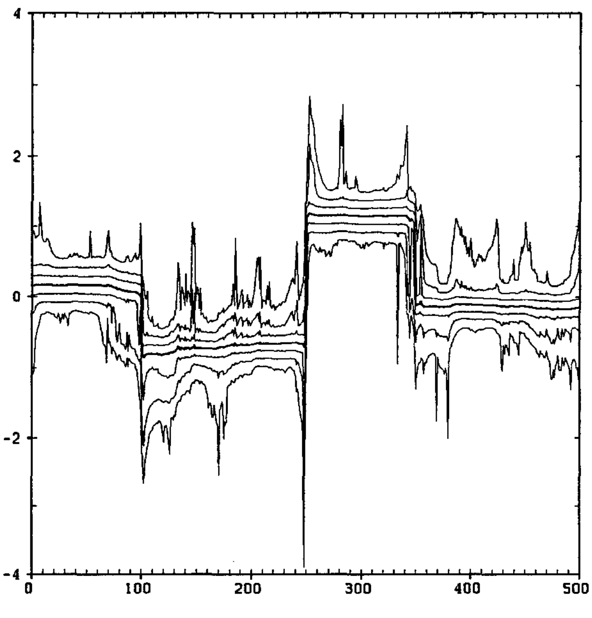

Fig. 5. The marginal posterior densities obtained by the Gaussian-sum smoother ($m = 2$).

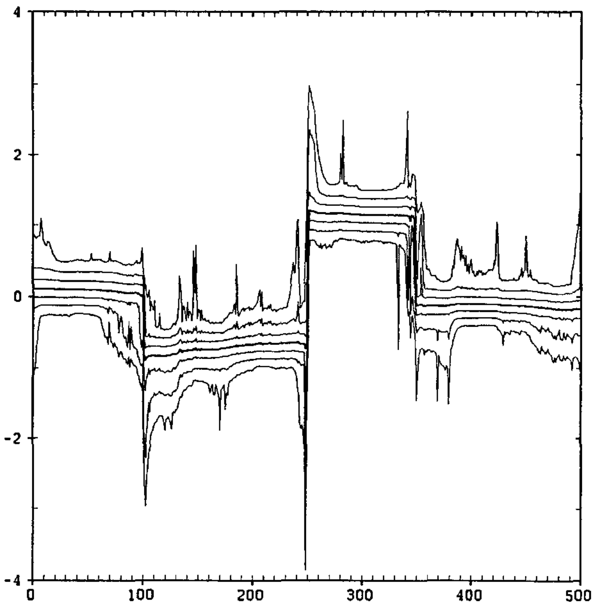

Fig. 6. The marginal posterior densities obtained by the Gaussian-sum smoother ($m = 4$).

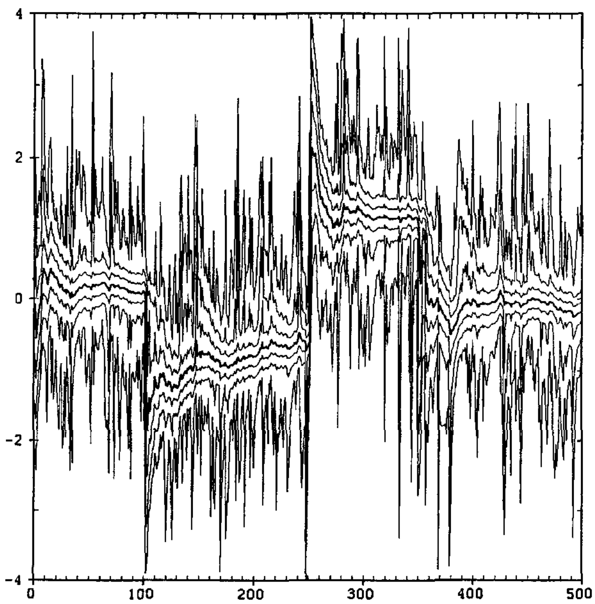

Fig. 7. The marginal posterior densities obtained by the non-Gaussian filter.

shows the trace of the posterior median. The other six fine curves show the 0.13, 2.27, 15.87, 84.13, 97.73 and 99.87 percentile points of the marginal posterior density which correspond to the $\pm 1 \sim \pm 3$ standard deviations of the Gaussian density. To compare this result, we applied the Gaussian-sum smoother using the same parameters. Figure 3 is obtained by an equivalent Gaussian model with the system noise variance $\tau^2 = 0.991 \times 0.00013 + 0.009 \times 4 = 0.03613$. The estimated trend is very wiggly and does not clearly show the jump of the trend. Figures 4–6 show the marginal posterior densities obtained by the Gaussian-sum smoothers with $m = 1$, 2 and 4, respectively. From the model assumption, we put $K_v = 2$ and $K_w = 1$. We also tested for $m = 8$, 16, etc. However, the difference from Fig. 6 is not apparent. From these illustrations, it is clear that even with small number of Gaussian components (even $m = 2$), we can get fairly good approximations to the one obtained by the non-Gaussian smoother. Finally, Fig. 7 shows the result obtained using the non-Gaussian filter. Comparison with this figure clearly shows the potential of using the smoothing algorithm in estimating the trend component.

7. Non-Gaussian seasonal adjustment

Kitagawa (1989) introduced a non-Gaussian version of the seasonal adjustment model. Detection of the abrupt changes of the trend and the seasonal components caused by the energy crisis or strikes was shown in an economic time series example. In that paper, non-Gaussian seasonal adjustment was realized using a Gaussian sum filter. The augmented state vector state space model used was,

$$(7.1)\qquad \left[\begin{array}{c} t_n \\ t_{n-1} \\ \vdots \\ t_{n-19} \\ \hline s_n \\ s_{n-1} \\ \vdots \\ s_{n-19} \end{array}\right] = \left[\begin{array}{cccc|cccc} 2 & -1 & & & & & & \\ 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & 0 & & & & \\ \hline & & & & -1 & \cdots & -1 & 0 \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & 0 \end{array}\right] \left[\begin{array}{c} t_{n-1} \\ t_{n-2} \\ \vdots \\ t_{n-20} \\ \hline s_{n-1} \\ s_{n-2} \\ \vdots \\ s_{n-20} \end{array}\right] + \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \hline 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{array}\right] v_n .$$

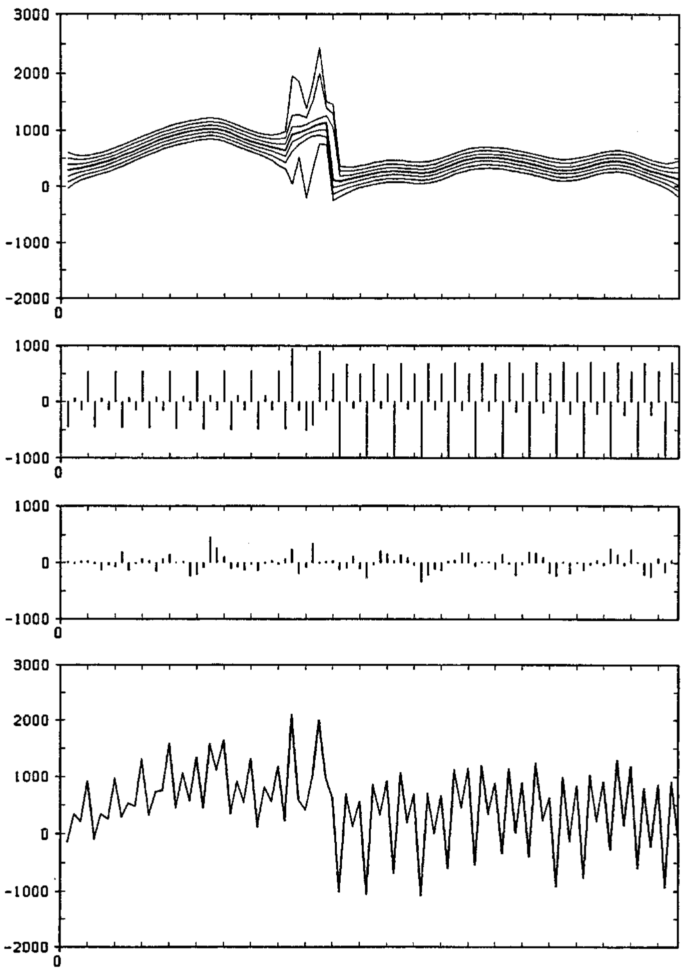

Fig. 8. Non-Gaussian seasonal adjustment of the increase of inventory of private company in Japan data. From top to the bottom, they show the estimated trend, seasonal and noise components and the original series.

Using this state space model, we obtained the fixed-lag smoother $x_{n-19|n}$.

In contrast with that model, by using the Gaussian-sum smoother developed in this paper, we can obtain the fixed interval smoother $x_{n|N}$ directly from the state space model with the state vector $x_n = (t_n, t_{n-1}, s_n, \ldots, s_{n-3})^t$ for quarterly data and $x_n = (t_n, t_{n-1}, s_n, \ldots, s_{n-11})^t$ for monthly data. The results of the analysis of an economic series (quarterly series of increase of inventories of private companies in Japan, 1965–1983) is shown in Fig. 8. The estimates obtained by the Gaussian-sum filter is shown in Kitagawa (1989). The differences of the estimated trend and seasonal components by the present fixed interval smoother and the fixed-lag smoother are not apparent in this example. However, due to the reduction of the dimension of the state vector (from 40 to 5 in the present case), the computing time and the necessary storage are significantly reduced (to about 1/64). That

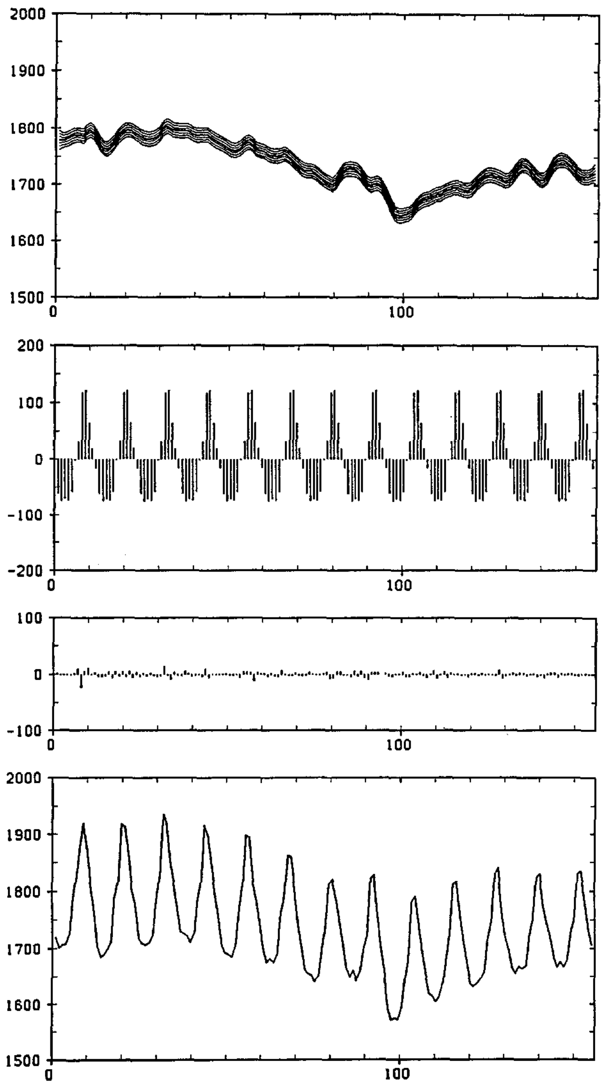

Fig. 9. Estimated trend, seasonal and residual components for BLSALLFOOD data obtained by the standard Gaussian model.

simplification will enable us to apply the non-Gaussian smoothing method to longer data or to more sophisticated and complex models.

8. Outliers in time series

The mixture model for observation noise is especially useful for handling outliers in time series (see for example, Tsay (1986)). Figure 9 shows the results of standard seasonal adjustment of the BLSALLFOOD data (the Bureau of Labor Statistics, all employees in food industries, January 1966–December 1979, $N = 156$, Kitagawa and Gersch (1984)). The original data, the estimated trend

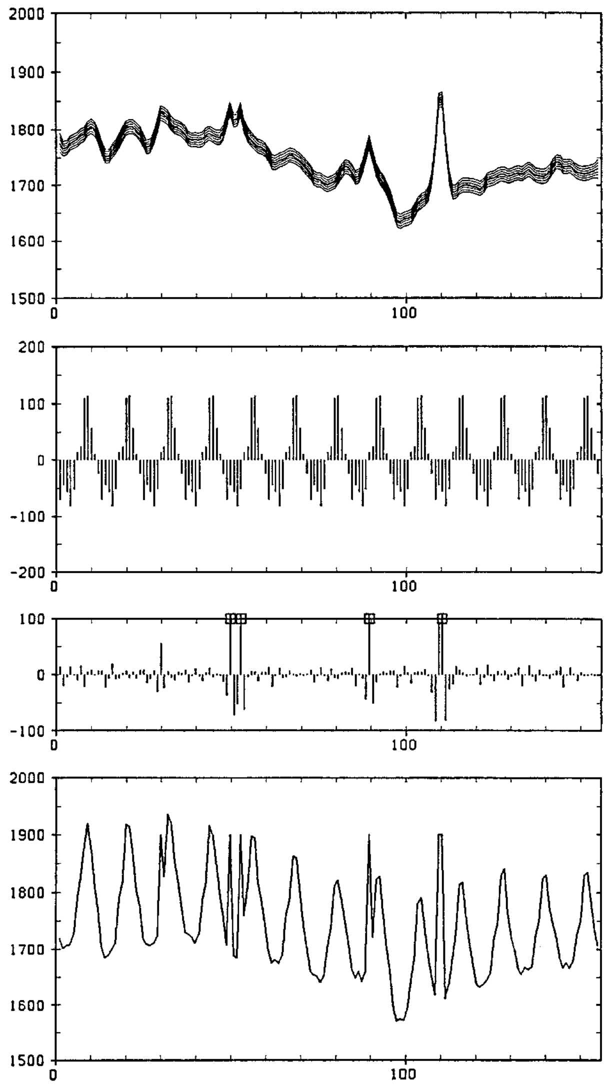

Fig. 10. Estimated trend, seasonal and residual components for BLSALLFOOD data with six outliers obtained by the standard Gaussian model.

with $\pm 1 \sim \pm 3$ standard error intervals, the estimated seasonal component and the residuals are shown in the figure.

To check the robustness of the Gaussian mixture observational noise model, we consider an artificially contaminated series of the BLSALLFOOD data. The data shown in Fig. 10 is obtained by replacing six observations, y_{29}, y_{50}, y_{53}, y_{90}, y_{110}, y_{111} by the number 1900. That data are considered as outliers. The effect of these outliers are evaluated by comparing Fig. 10 with Fig. 9. The estimated trend by the standard seasonal adjustment model for this data is strongly affected by these outliers and is wiggly. The estimated seasonal pattern is also different from the one shown in Fig. 9. The log-likelihood of the model is -2795.2.

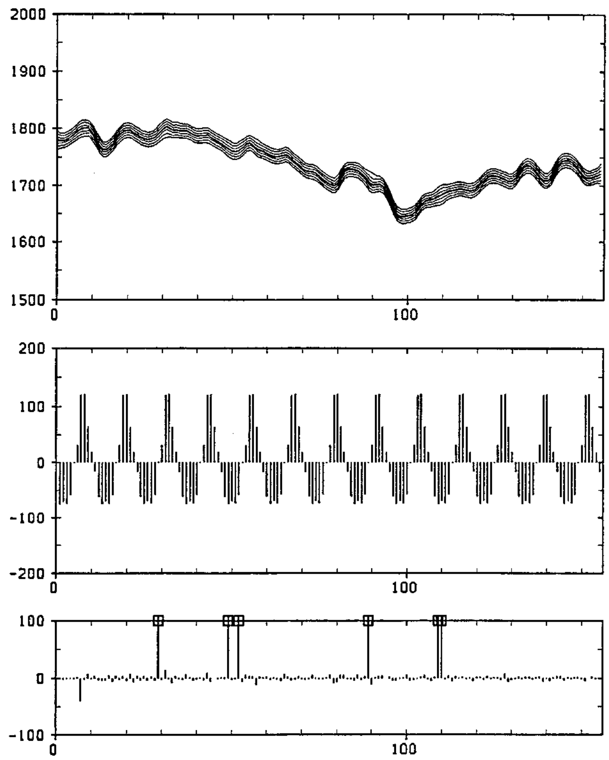

Fig. 11. Estimated trend, seasonal and residual components for BLSALLFOOD data with six outliers obtained by the Gaussian-mixture observational noise model.

Figure 11 shows the results obtained by assuming that the observational noise density is a mixture of two Gaussian densities

$$r(w) \sim \alpha N(0, \sigma^2) + (1 - \alpha) N(\mu, \tau^2). \tag{8.1}$$

Approximate maximum likelihood estimates of the parameters are $\hat{\alpha} = 0.96$, $\hat{\sigma}^2 = 30.3$, $\hat{\tau}^2 = 4 \times 10^4$. The estimated trend and the seasonal components are indistinguishable from the ones in Fig. 9. The residuals take large values corresponding to the presence of the outliers. The AIC value of this model is -690.2 indicating that this model is significantly better than the standard Gaussian model.

Incidentally, by applying the Gaussian mixture model to the original uncontaminated data shown in Fig. 9, we obtain almost identical estimates of trend and seasonal components. The log-likelihood of the model is -650.8. Compared with the log-likelihood value of -649.4 of the best Gaussian model, it is clear that the approximation made using this model is excellent. This example suggests that the Gaussian mixture model is robust to outliers and also that it yields an excellent approximation when applied to data without outliers.

Acknowledgements

The author is grateful to Professor Will Gersch of University of Hawaii for his careful reading of and various suggestions on the manuscript. He also thanks the referees for their valuable suggestions.

References

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

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

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **34**, 1–41.

Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series (with discussion), *J. Amer. Statist. Assoc.*, **76**, 1032–1064.

Kitagawa, G. (1988). Numerical approach to non-Gaussian smoothing and its applications, *Computing Science and Statistics: Proceedings of the 20th Symposium on the Interface* (eds. E. J. Wegman, D. T. Gantz and J. J. Miller), 379–388.

Kitagawa, G. (1989). Non-Gaussian seasonal adjustment, *Comput. Math. Appl.*, **18**(6/7), 503–514.

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

Sorenson, H. W. and Alspach, D. L. (1971). Recursive Bayesian estimation using Gaussian sums, *Automatica*, **7**, 465–479.

Tsay, R. (1986). Time series modeling in the presence of outliers, *J. Amer. Statist. Assoc.*, **81**, 132–141.

West, M. and Harrison, J. (1989). Bayesian forecasting and dynamic models, Springer Series in statistics, Springer, New York.