ELSEVIER Computational Statistics & Data Analysis 30 (1999) 281–301

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Applications of quasi-periodic oscillation models to seasonal small count time series

Tomoyuki Higuchi *

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

Received 1 August 1997; received in revised form 1 August 1998; accepted 30 November 1998

Abstract

Quasi-periodic oscillation models for analysis of small count time series are considered within a framework of a generalized state space model (GSSM). In particular, we focus on the analysis of seasonal count data. The Monte Carlo filter (MCF) is fully employed in this study to handle a generalized state space model with higher state dimensions. To illustrate, we study three seasonal count data sets: polio incidence time series, the monthly number of drivers killed in road accidents, and the monthly number of the sun's spotless days. In addition, we demonstrate an application of the model proposed to the yearly occurrence of intense hurricanes with a quasi-periodic component associated with solar cycle activity.

Keywords: Small count data; Quasi-periodic oscillation; Generalized state space model; Monte Carlo filter; Sunspot number; Hurricane data

1. Introduction

In this study we are concerned with analysis of time series of small count data that are frequently obtained in many fields such as biomedical statistics and astrophysics. Our attention is focused on seasonal time series involving small counts, for example monthly numbers of Polio incidences (Zeger, 1988), because time-series involving relatively larger counts can be well analyzed by means of a time-series model without regarding the fact that the observed data follow a discrete distribution, and do not require a more sophisticated model in practice. Recent Bayesian

* E-mail: higuchi@ism.ac.jp.

0167-9473/99/$ – see front matter
PII: S 0167-9473(98)00097-8

approaches to time-series modeling have paid considerable attention to the analysis of small count data to illustrate their applications (e.g., West et al., 1985; Harvey and Fernandes, 1989; Fahrmeir, 1992; Kashiwagi and Yanagimoto, 1992; Frühwirth-Schnatter, 1994a; Chan and Ledolter, 1995; Grunwald et al., 1997; Durbin and Koopman, 1997; Shephard and Pitt, 1997).

Most of the Bayesian models proposed for dealing with small count data can be formulated to take a convenient form from a computational point of view, called the *generalized state space model* (GSSM) (Kitagawa, 1987). The GSSM is defined by a set of two models:

$$\text{system model } x_n = f_n(x_{n-1}, v_n) \tag{1}$$

and

$$\text{observation model } y_n \sim r(\cdot|x_n, \beta_o), \tag{2}$$

where x_n is a $k \times 1$ vector of unobserved state variables at discrete time of n, and y_n is a univariate observation. $\{v_n\}$ is independently and identically distributed (*i.i.d.*) with $v_n \sim q(\cdot|\beta_s)$, where $q(\cdot)$ denotes the l-dimensional non-Gaussian distribution, and $r(\cdot)$ is the conditional distribution of y_n given x_n. $f_n : \Re^k \to \Re^k$ is the system transition function and its form is assumed to be known. β_s and β_o are parameter vectors for describing q and r, respectively, and are called *hyperparameters* in Bayesian terminology (Lindley and Smith, 1972). For convenience, we combine β_s with β_o and denote them by $\beta^T = [\beta_s^T, \beta_o^T]$, where T is a transposition.

If the system model is a linear Gaussian transition equation given by

$$x_n = F_n x_{n-1} + v_n \tag{3}$$

and $r(y_n|x_n, \beta_o)$ depends on the state vector x_n through the linear predictor $\mu_n = H_n x_n$, where F_n and H_n are the $k \times k$ and $1 \times k$ matrices, respectively, then it is usually called a *dynamic generalized linear model* (DGLM) (West et al., 1985). Several examples for GSSM including DGLM can be seen in e.g., Kitagawa (1987, 1991), West et al. (1985), West and Harrison (1989), Fahrmeir (1992) and Frühwirth-Schnatter (1994a).

The GSSM approach which enables us to use recursive formulations together with an evaluation of the likelihood, has a unified framework. However, the GSSM still requires, for its practical application to data analysis, computationally extensive and difficult tasks due to the relatively high dimensionality of the state vector. In a case with a low dimension (for an example $k \leq 2$), a simple but flexible approach to approximate any conditional distribution by a first order spline or simple step functions is feasible. Several applications of this approach can be found in Kitagawa (1987, 1991).

For $2 < k$ an implementation using this simple approach is rarely practicable due to the inherent computational complexity and the large storage requirement (Fahrmeir (1992), Frühwirth-Schnatter (1994a)). To handle the high dimensionality the Monte Carlo method for filtering and smoothing has been proposed (Kitagawa, 1993, 1996; Gordon et al., 1993). While Gordon et al. (1993) called it *Bootstrap filter*, we refer to it as the *Monte Carlo Filter* according to a manner of Kitagawa (1993).

The treatment of the high dimensionality within the GSSM framework is not discussed in the present paper and can be found in West and Harrison (1989), Carlin et al. (1992), Fahrmeir (1992), Frühwirth-Schnatter (1994a, b), Durbin and Koopman (1997), and references therein.

As mentioned above, we are focusing on seasonal time series involving small counts. Usually, an analysis of the seasonal time series is carried out in terms of the procedure called *seasonal adjustment* which is designed to decompose a time series y_n into several possible components: a trend component t_n, seasonal component s_n, stationary component u_n, observation noise component ε_n, etc. Within the framework of the GSSM, its decomposition can be achieved by assuming the stochastically perturbed linear difference equation on each component. As for the seasonal component with a period of L, we usually adopt a simple representation such as $s_n = s_{n-L} + v_n$ or its modification $s_n = -(s_{n-L+1} + s_{n-L+2} + \cdots + s_{n-2} + s_{n-1}) + v_n$, where $v_n \sim \mathrm{N}(0, \tau^2)$. This representation for the seasonal component, which is usually called the standard basic structural model (BSM), has been already adopted in the DGLM to deal with monthly count data with low means (Durbin and Koopman, 1997). Although this representation of the seasonal component is usually adequate for economic time-series, it may give less accurate estimate for geophysical data because the seasonal frequency is often lower.

In this study, an alternative model for the seasonal component (e.g., West, 1995) is employed to satisfy this request from a geophysical point of view. In this model, the seasonal component is simply expressed as a sum of several pseudocyclical components. Each pseudocyclical component can be represented by a second order AR model (Higuchi et al., 1988; West, 1995). The presence of system noise makes the cycle stochastic rather than deterministic. As a result, this kind of AR process appears to have a quasi-periodic oscillation (QPO). This model has been adopted to represent the pseudocyclical behavior of an annual economic time-series within a framework of the structural time-series model (Harvey, 1985), but it takes a different form in the state-space representation. West (1995) called such a model a cyclical component model and demonstrated interesting examples of its application to real data. The performance of these models as a linear filter in a frequency domain has been numerically investigated (Higuchi, 1991).

The article is organized as follows. In Section 2 we propose a new model for analyzing the seasonal small count time series. In Section 3 we give a brief description of the Monte Carlo Filter (abbreviated to MCF henceforth) together with an explanation of the recursive formulation for an estimation of the conditional probability distribution. In Section 4, we describe the application of our methods to the seasonal count data set previously analyzed by Zeger, polio incidence time series (Zeger, 1988; Chan and Ledolter, 1995). We also illustrate our approach by demonstrating the application to the monthly numbers of car drivers killed in road accidents which is shown in pp. 519–523 of Harvey (1989), and the monthly numbers of the sun's spotless days. In Section 5 we modify the model to study binary time-series. An application to the yearly occurrence of intense hurricanes in the Atlantic basin will be given. Finally, Section 6 describes some computational aspects of the procedure.

2. Seasonal count data

We propose the GSSM for analysis of the seasonal small count data resulting in relatively higher state vector dimension. We give a description of the proposed model and a comparison with the model given by Chan and Ledolter (1995), denoted by CL model henceforth in this study, for dealing with the monthly number of cases of poliomyelitis from January 1970 to December 1983 ($N = 168$), as well as that examined originally by Zeger (1988). Here N is the total number of data points. The observations of polio incidences are indicated by crosses in Fig. 1.

2.1. Observation model

As in the previous publication (Chan and Ledolter, 1995), we also assume that the observation is generated from a Poisson distribution with time-varying mean λ_n: $y_n \sim \text{Poisson}(\lambda_n)$. Usually there exists seasonality in monthly data, and we model it by decomposing $\log \lambda_n$ into two factors: $\log \lambda_n = t_n + s_n$, where t_n and s_n are the trend and seasonal components, respectively. In other words, we deal with the non-stationary Poisson model in which the time-varying mean is expressed as the multiplicative form given by $\lambda_n = \exp(t_n)\exp(s_n)$.

2.2. Chan and Ledolter's model

Before we begin a description of our system model for the seasonal component, we refer to the CL model. In their model, t_n is decomposed into the deterministic and stochastic components: $t_n = \alpha_1 + \alpha_2 n/1000 + W_n$, where W_n is assumed to be a stationary Gaussian AR(1) process defined by $W_n = \rho W_{n-1} + v_{n,W}, v_{n,W} \sim \text{N}(0, \tau_W^2)$. s_n is given as the deterministic form by using trigonometric components involving the first two harmonics;

$$s_n = \alpha_3 \cos\left(\frac{2\pi n}{12}\right) + \alpha_4 \sin\left(\frac{2\pi n}{12}\right) + \alpha_5 \cos\left(\frac{2\pi n}{6}\right) + \alpha_6 \sin\left(\frac{2\pi n}{6}\right). \tag{4}$$

As seen in Fig. 1, the observation in November 1972 ($n = 35$) appears to be an outlier (Chan and Ledolter, 1995), and thus they dealt with it by introducing an additive component, $\alpha_7 I_n$, into t_n, where I_n is the indicator function which is equal to 1 in November 1972 ($n = 35$) and to zero elsewhere. The trend component in the CL model is summarized as $t_n = \alpha_1 + \alpha_2 n/1000 + \alpha_7 I_n + W_n$. $\{\alpha_j\}$ $(j = 1, \ldots, 7)$, ρ, and τ_W^2 are hyperparameters to be optimized. This GSSM they adopt is extremely simple due to the one dimensional state vector $x_n = [W_n]$.

2.3. Representation of seasonal pattern

We adopt a first-order trend model for t_n: $t_n = t_{n-1} + v_{n,t}, v_{n,t} \sim \text{N}(0, \tau_t^2)$. Our model for s_n is completely different from that in the CL model. We use the quasi-periodic oscillation (QPO) model (e.g., Higuchi et al., 1988) of the form $s_n = \sum_{j=1}^{J} s_{n,j}$, where

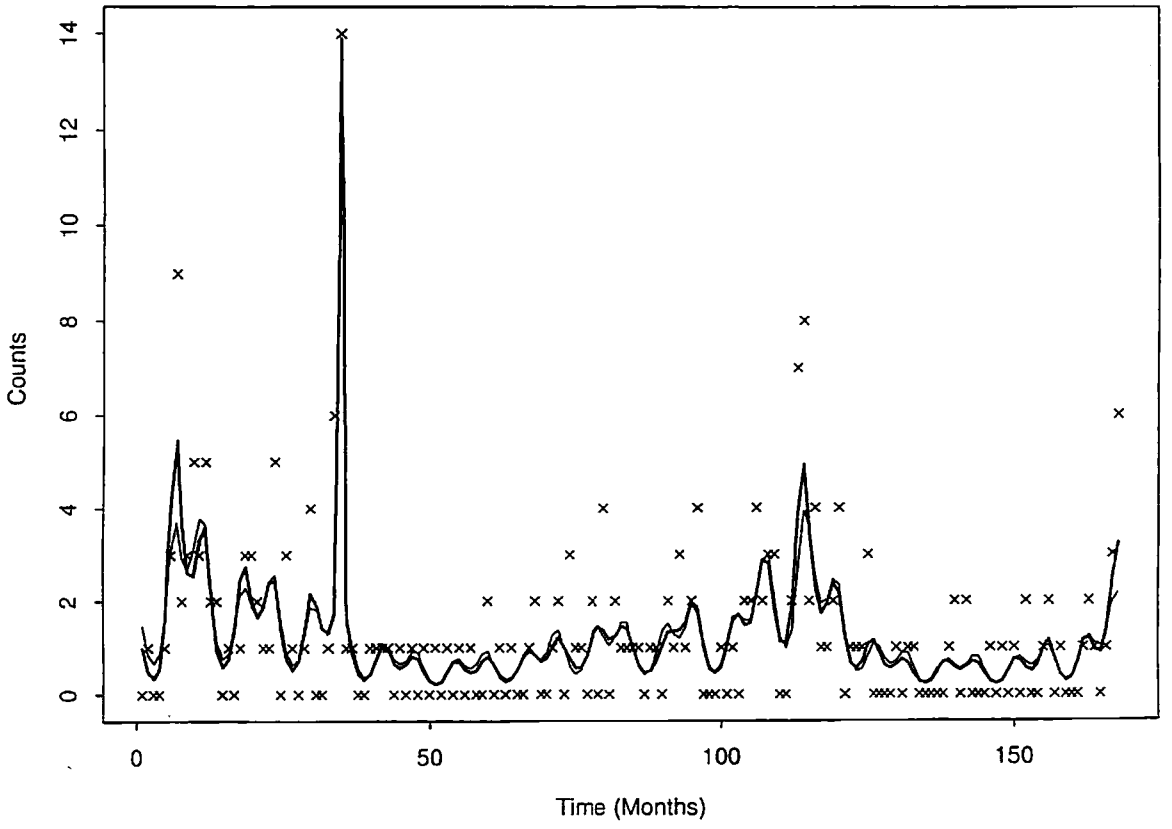

Fig. 1. Monthly number of cases of poliomyelitis from January 1970–December 1983, taken from Zeger (1988). The thin line connects the median of the posterior density of $p(\lambda_n|Y_N)$ corresponding to the smoothed time-varying Poisson mean in the Chan and Ledolter's model (1995). The thick line connects the median of the posterior density of $p(\lambda_n|Y_N)$, which is obtained by applying the QPO model we used in this study.

each component $s_{n,j}$ is represented by

$$s_{n,j} = 2\cos\left(\frac{2\pi}{T_j}\right) s_{n-1,j} - s_{n-2,j} + v_{n,s_j}, \quad v_{n,s_j} \sim \mathrm{N}(0, \tau_{s_j}^2) \tag{5}$$

with a fixed period T_j. This model allows us to represent a periodic component of distinct frequency with stochastically time-varying amplitude and phase (West, 1995). It therefore provides us with an opportunity to identify possible changes in amplitude and phase, in contrast to traditional parametric models such as harmonic regression, used in the CL model.

Similarly to the CL model, our interest is focused on the first two harmonic components, and s_n is expressed in this study by $s_n = s_{n,\mathrm{y}} + s_{n,\mathrm{h}}$ with

$$s_{n,\mathrm{y}} = 2\cos\left(\frac{2\pi}{12}\right) s_{n-1,\mathrm{y}} - s_{n-2,\mathrm{y}} + v_{n,\mathrm{y}}, \quad v_{n,\mathrm{y}} \sim \mathrm{N}(0, \tau_{\mathrm{y}}^2), \tag{6}$$

$$s_{n,\mathrm{h}} = 2\cos\left(\frac{2\pi}{6}\right) s_{n-1,\mathrm{h}} - s_{n-2,\mathrm{h}} + v_{n,\mathrm{h}}, \quad v_{n,\mathrm{h}} \sim \mathrm{N}(0, \tau_{\mathrm{h}}^2), \tag{7}$$

where s_y and s_h mean the QPO component with a yearly and half-yearly period, respectively. This model provides us with the GSSM

$$y_n \sim \text{Poisson}(\lambda_n = \exp(Hx_n)),$$
$$x_n = Fx_{n-1} + Gv_n \tag{8}$$

with a five-dimensional state vector, $x_n = [t_n, s_{n,y}, s_{n-1,y}, s_{n,h}, s_{n-1,h}]^T$ and with the following ingredients:

$$H = (1, 1, 0, 1, 0),$$

$$F = \begin{pmatrix} 1 & & & & \\ & 2\cos(\frac{2\pi}{12}) & -1 & & \\ & 1 & & & \\ & & & 2\cos(\frac{2\pi}{6}) & -1 \\ & & & 1 & \end{pmatrix},$$

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$v_n = [v_{n,t}, v_{n,y}, v_{n,h}]^T.$$

Here the empty entries of F are all zero and $v_n \sim N(0, R)$ with a diagonal variance matrix of $R = \text{diag}(\tau_t^2, \tau_y^2, \tau_h^2)$. Here τ_t^2, τ_y^2, and τ_h^2 are unknown hyperparameters to be optimized.

The combination of six harmonics to express the seasonal component has already been used to deal with the monthly economic time series (Ameen and Harrison, 1985). However, in the QPO model proposed here the system noise variance of each harmonic component composing the seasonal variation differs while it is common to all harmonic components in their model. Of course, to treat each variance independently requires computational tasks in terms of the optimization, but enables us to represent a wide class of time-varying seasonal patterns.

3. Monte Carlo filter (MCF)

In this section we give a brief explanation of the MCF that is adopted to estimate the conditional probability distribution of the state vector x_n. A detailed description of the MCF can be seen in Gordon et al. (1993) and Kitagawa (1996).

3.1. Recursive estimation

We begin by explaining the recursive formula underlying the MCF. (1) and (2) yield useful recursive formulas for the estimation of the conditional probability distribution of the state vector x_i given data $Y_j \equiv [y_1, y_2, \ldots, y_j]$, $p(x_i|Y_j)$, which are

formed by a set of the following two steps at each time n: *prediction* and *filtering* (e.g. Kitagawa, 1987; Harvey, 1989).

(1) *prediction*: Assuming knowledge of the posterior distribution for the state vector at time $n-1$, $p(x_{n-1}|Y_{n-1})$, compute the one-step-ahead predictive distribution at time n, $p(x_n|Y_{n-1})$, by

$$p(x_n|Y_{n-1}) = \int p(x_{n-1}|Y_{n-1}) \cdot p(x_n|x_{n-1})\,\mathrm{d}x_{n-1}. \tag{10}$$

(2) *filtering*: Based on the obtained distribution, $p(x_n|Y_{n-1})$, compute the posterior distribution at time n, $p(x_n|Y_n)$, by

$$p(x_n|Y_n) = \frac{p(y_n|x_n) \cdot p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})} = \frac{p(y_n|x_n) \cdot p(x_n|Y_{n-1})}{\int p(y_n|x_n) \cdot p(x_n|Y_{n-1})\,\mathrm{d}x_n}. \tag{11}$$

An initial distribution $p(x_0|Y_0)$ is defined a priori.

For the fixed (distribution) forms of $q(\cdot)$ and $r(\cdot)$, the optimal value of hyperparameters, β^*, is selected by assessing the log-likelihood, $\log p(Y_N|\beta)$ (Good, 1965),

$$\begin{aligned} l(\beta) = \log p(Y_N|\beta) &= \log \prod_{n=1}^{N} p(y_n|Y_{n-1}, \beta) \\ &= \sum_{n=1}^{N} \log p(y_n|Y_{n-1}, \beta), \end{aligned} \tag{12}$$

where $p(y_n|Y_{n-1}, \beta)$ is the conditional distribution of y_n, given data Y_{n-1}. We note that $p(y_n|Y_{n-1}, \beta)$ is the one-step-ahead predictive density which appears in Eq. (11). The best function forms among competing candidates are chosen so as to maximize Eq. (12) in the same manner. The conditional distribution, $p(x_n|Y_N)$, given the data Y_N, can be obtained by using the following recursive algorithm with help of all of $p(x_n|Y_{n-1})$ and $p(x_n|Y_n)$ stored on the pass of Eqs. (10) and (11).

smoothing:

$$p(x_n|Y_N) = p(x_n|Y_n) \int \frac{p(x_{n+1}|Y_N) \cdot p(x_{n+1}|x_n)}{p(x_{n+1}|Y_n)}\,\mathrm{d}x_{n+1}. \tag{13}$$

3.2. MCF algorithm

An essential idea of the MCF (Kitagawa, 1993, 1996; Gordon et al., 1993) is that we approximate an arbitrary conditional probability density function, $p(x_i|Y_j)$, by a set of realizations where the number of realizations, m, is fixed at each time n. For example, we express $p(x_n|Y_{n-1})$ by a set of m realizations: $Z_{n|n-1} \equiv \{z^{(i)}_{n|n-1} \mid i = 1, \ldots, m\}$. Similarly, $p(x_n|Y_n)$ is approximated by a set of m realizations: $Z_{n|n} \equiv \{z^{(i)}_{n|n} \mid i = 1, \ldots, m\}$.

The recursive calculations corresponding to Eqs. (10) and(11) are, respectively, defined in the following manner:

(MCF-1) *prediction*

Realize $Z_{n|n-1}$ of which each element is obtained by passing $z^{(i)}_{n-1|n-1}$ through the system model (1)

$$z^{(i)}_{n|n-1} = f(z^{(i)}_{n-1|n-1}, v^{(i)}_n), \tag{14}$$

where $v^{(i)}_n$ is a realization sampled from $q(\cdot)$.

(MCF-2) *filtering*

Given the observation y_n, evaluate the likelihood of each particle $z^{(i)}_{n|n-1}$, $r(y_n|z^{(i)}_{n|n-1})$, and resample $z^{(i)}_{n|n-1}$ with probability proportional to

$$\begin{aligned} p(z_n = z^{(i)}_{n|n-1}|Y_n) &= \frac{r(y_n|z^{(i)}_{n|n-1})p(z_n = z^{(i)}_{n|n-1}|Y_{n-1})}{\sum_{i=1}^m r(y_n|z^{(i)}_{n|n-1})p(z_n = z^{(i)}_{n|n-1}|Y_{n-1})} \\ &= \frac{r(y_n|z^{(i)}_{n|n-1})(1/m)}{\sum_{i=1}^m r(y_n|z^{(i)}_{n|n-1})(1/m)} \\ &= \frac{r(y_n|z^{(i)}_{n|n-1})}{\sum_{i=1}^m r(y_n|z^{(i)}_{n|n-1})}, \end{aligned} \tag{15}$$

to generate samples $Z_{n|n}$. In the MCF the value of $p(y_n|Y_{n-1})$ appearing in Eq. (12) is approximated by $(1/m)\sum_{i=1}^m r(y_n|z^{(i)}_{n|n-1})$.

In Kitagawa's numerical integration approach (1987) to GSSM, a final or smoothed estimation of the state vector is given using Eq. (13). In the MCF, Kitagawa (1993, 1996) proposed two alternative formulas for the smoothing algorithm: storing the state vector and a two-filter formula. It is not feasible to apply the former as it is, and so a modification is necessary for a workable algorithm. The easiest way is to use the fixed-lag smoothing (Anderson and Moore, 1979) which reduces to a filtering problem for extended state $x^{\mathrm{T}}_{n,E} = [x^{\mathrm{T}}_{n-L}, x^{\mathrm{T}}_{n-L+1}, \ldots, x^{\mathrm{T}}_n]$ (Doucet et al., 1995).

4. Real data application

4.1. Polio incidence time-series

We first applied the MCF to the CL model, explained in Section 2.2, with the estimates of hyperparameters given by Chan and Ledolter (1995). The thin line in Fig. 1 shows an estimated λ_n, $\widehat{\lambda_n}$, obtained by taking the median of the posterior density, $p(\lambda_n|Y_N)$. The number of particles used here is $m = 100\,000$. The fixed lag is set $L = 20$. The major difference between the result obtained using the MCF and that given by Chan and Ledolter (Fig. 1 of Chan and Ledolter) is the estimate of λ_n at the outlier, $\widehat{\lambda_{35}}$. The deterministic part in $\log \lambda_{35}$ of the CL model given by $t_{35} + s_{35} - W_{35}$ at the outlier point is approximately 2.28. The value of the smoothed means taken from Fig. 1 of Chan and Ledolter (1995) is, by visual inspection, about 6.5. This could imply an approximation of the stationary component, $\widehat{W_{35}} = \log(6.5) - 2.28 = -0.48$. This is quite different from our estimate, $\widehat{W_{35}} = 0.334$. We do not know why $\widehat{\lambda_{35}}$ presented by Chan and Ledolter takes such a small value, but $\widehat{\lambda_n}$ obtained

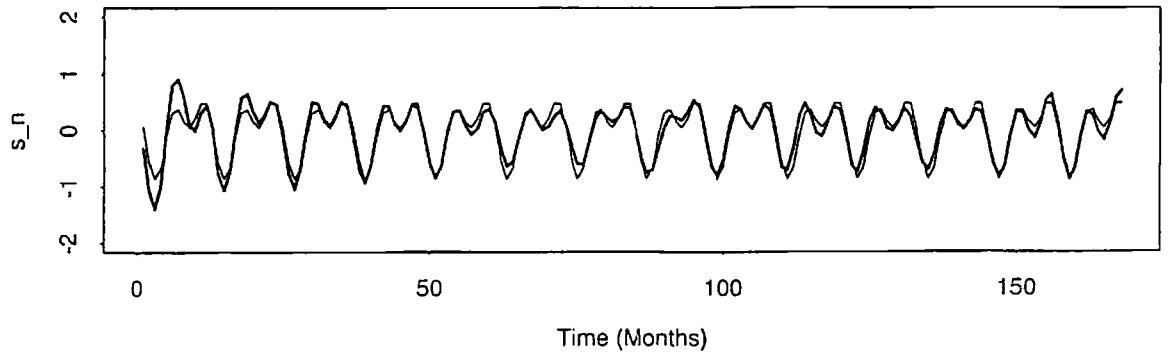

Fig. 2. The thin and thick lines correspond to the estimated seasonal component for the CL model and the QPO model, respectively.

by the MCF shows fairly good fit to the observation. Except for the estimation at $n = 35$, the MCF provides us $\widehat{\lambda}_n$ in good agreement with the result of Chan and Ledolter.

A thick line in Fig. 1 shows the median of the posterior density of $p(\lambda_n|Y_N)$, that is obtained by applying the QPO model explained in Section 2.3. The number of particles used here is $m = 100\,000$. The fixed lag is also set $L = 20$. For comparison with the result based on the CL model, a treatment of the outlier observed in November 1972 ($n = 35$) is also carried out, in the same manner mentioned above, by introducing an additive component $\alpha_7 I_n$. The value of α_7 is beforehand given by using the estimate obtained by Chan and Ledolter (1995) in order to make it easy to search for optimal hyperparameters giving the maximum likelihood. The estimated values for hyperparameters are $\widehat{\tau}_t^2 = 0.0068$, $\widehat{\tau}_y^2 = 0.0024$, and $\widehat{\tau}_h^2 = 0.00046$. The smaller estimate on the variance of the system noise for the trend component compared with that given by the CL model ($\widehat{\tau}_t^2 = 0.0336$) comes from the presence of the system noise for the seasonal component in our model, resulting in a gradual change in the variation of the seasonal pattern. In other words, the trend component in the CL model accounts for the stochastic behavior of the seasonal pattern in the QPO model.

We show the estimated seasonal component, $\widehat{s}_n$, in Fig. 2. The thin and thick lines represent the results based on the CL and QPO models, respectively. It should be reminded that the seasonal component for the CL model is deterministically given, as in Eq. (4), and then $\widehat{s}_n$ for the CL model repeats an identical cyclical pattern. In contrast, $\widehat{s}_n$ for the QPO model is based on a stochastical description and is capable of representing explicit time variation in amplitude and phase. $\widehat{s}_n$ for the QPO model is in good agreement with $\widehat{s}_n$ for the CL model, indicating that a parametric approach for describing the seasonal component is reasonable. However, further comparison makes it clear that the amplitude of $\widehat{s}_n$ for the QPO model gradually changes with time. In summary, it is pointed out that the QPO model provides us an opportunity to identify possibly time-varying amplitude characteristics. The components of $\widehat{s}_n$, $\widehat{s_{n,y}}$ and $\widehat{s_{n,h}}$, are shown in Fig. 3a and b, respectively. Detailed interpretation from a viewpoint of epidemiology is left open in this study.

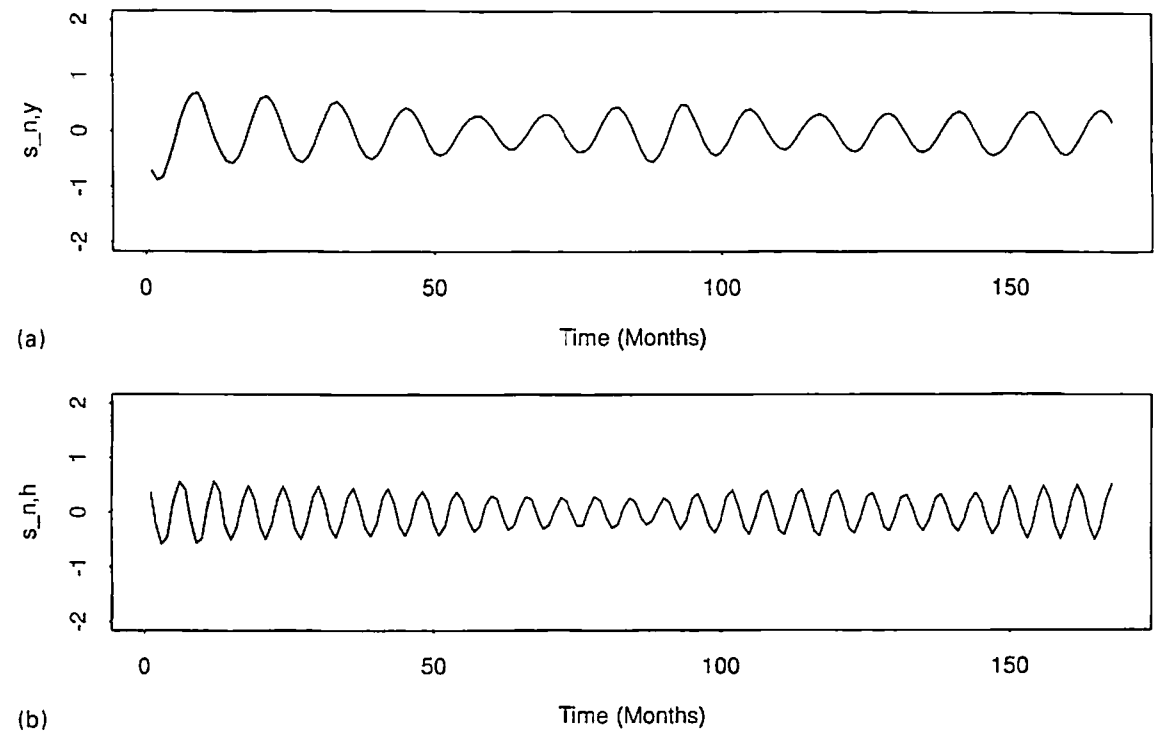

Fig. 3. Decomposition of seasonal component. (a) Yearly period component. (b) Half-yearly period component.

4.2. *Analysis of monthly numbers of drivers killed in road accidents*

We demonstrate an interesting feature of the QPO model for the seasonal time-series analysis by the example. The data we examine is the monthly number of light goods vehicle drivers killed in road accidents from 1969 to 1984 in Great Britain shown in pp. 519–523 of Harvey (1989). These data, shown with crosses in Fig. 4, have been recently analyzed with the GSSM given by Durbin and Koopman (1997). In their model (referred to as the DK model henceforth), the observations are assumed to have non-stationary Poisson distributions with a time-varying mean λ_n of which the logarithm is decomposed into three factors: trend, seasonal, and intervention factor. The intervention factor is simply expressed as αI_n, where I_n is the indicator function which is equal to 1 after the February 1983 ($n = 170$) when the seat belt legislation has been introduced. α, called the intervention parameter in Durbin and Koopman (1997), is a parameter to be optimized. The trend component is assumed to be a first-order trend model as adopted in Section 4.1. We also follow their GSSM except for the system model for the seasonal component. While they use the usual BSM model for the seasonal component expressed as $s_n = -\sum_{j=1}^{11} s_{n-j} + v_{n,s} v_{n,s} \sim \mathrm{N}(0, \tau_s^2)$, the QPO model is employed in this study. As in previous application, we will focus on the first two harmonics, resulting in a state vector of which element is given by $x_n = [t_n | s_{n,y}, s_{n-1,y} | s_{n,h}, s_{n-1,h}]^T$. If no significant variation with a period of less than half-yearly is expected to be observed, then this representation allows a more parsimonious parameterization for the state vector:

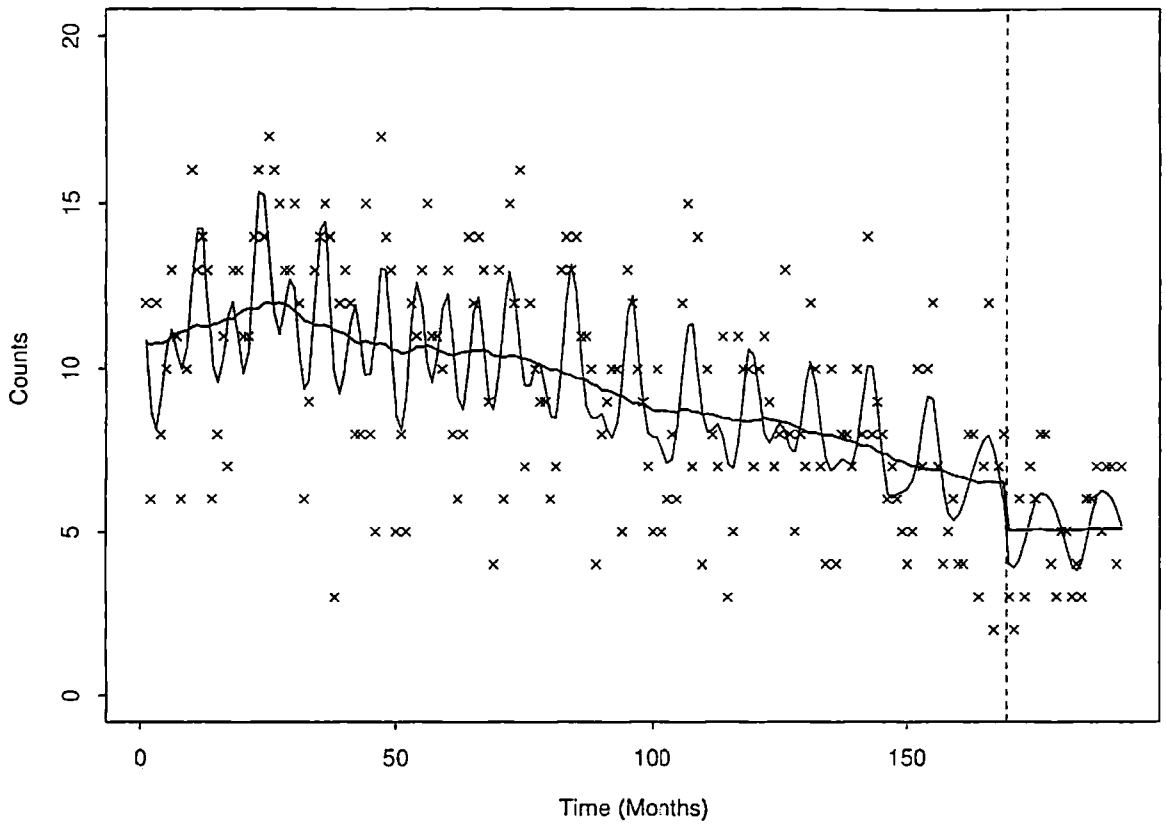

Fig. 4. Numbers of light goods vehicle drivers killed in road accidents in Great Britain from 1969 to 1984. The thick line indicates the exponent of the trend component given by $\exp(t_n + \alpha I_n)$, where I_n is an indicator function for the post legislation period and α is an intervention parameter.

for $J = 2$ the dimension of the state vector is 5 instead of 12 for the BSM. The hyperparameters in this GSSM are $\beta = [\tau_t^2, \tau_y^2, \tau_h^2, \alpha]^T$.

The exponent of $t_n + \hat{\alpha} I_n$ based on the median of the posterior distribution through MCF is plotted by a thick line in Fig. 4. The vertical dashed line indicates the time of the seat belt legislation being introduced. It is clearly seen that the seat belt legislation reduces the number of drivers killed in the accidents effectively. The estimated seasonal component based on the QPO model is shown in Fig. 5. It is obvious that the seasonal variation demonstrates a stochastic behavior that is generated by the presence of system noise. By visual inspection, the seasonal component can be divided into three periods (separated by two solid lines in figure) in terms of its pattern in time domain: 1969–1974 ($n = 1$–72), 1975–1980 ($n = 73$–144), and 1981–1984 ($n = 145$–192). However, a discussion on what factors would change the seasonal pattern is out of the scope of this study, we simply address that the QPO model is capable of representing the seasonal pattern sufficiently.

On the other hand, the seasonal variation in the DK model takes a deterministic form, because the variance of the system noise for their model was estimated to be zero; see details in Durbin and Koopman (1997). A comparison between the QPO and DK models is given in Table 1 using AIC (Akaike, 1974; Kitagawa, 1987), an alternative to BIC (see comment on the Kitagawa's paper, (1987), given by Martin

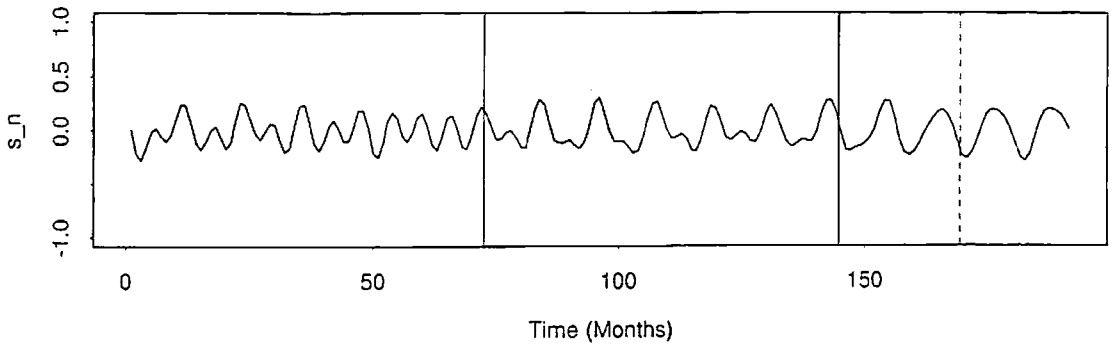

Fig. 5. Estimated seasonal component based on the QPO model.

Table 1
Comparison of two models

Model type	Mean of 100 AICs	Variance	Range of 100 AICs
DK	997.74	0.0083	997.53–997.96
QPO	991.53	0.0174	991.15–991.77

and Raftery, 1987, and the Kitagawa's rejoinder to it). The AIC value for the DK model is also dependent on the log-likelihood value calculated by the MCF. We used the hyperparameter values for the DK model shown in Durbin and Koopman (1997). We conduct 100 trials given the hyperparameter values for each model and show their simple statistic in Table 1. Table 1 suggests that the stochastic representation by the QPO model is favorable for this data set.

4.3. *Analysis of monthly spotless days*

We show one more example to demonstrate an application of the MCF to the QPO model designed to deal with a monthly time series. The data we examine is the monthly number of spotless days of the sun from January 1993 to July 1996 ($N=43$). The spotless days data taken from a publication given by US Department of Commerce, National Oceanic and Atmospheric Administration (NOAA) (1996) are collected and compiled from the USAF Solar Electro-Optical Network sites and Mt. Wilson Solar Observatory (`http://www.astro.ucla.edu/~ obs/150_srep.htm`). The observations are denoted by crosses in Fig. 6. Other spotless days data sets are available by compiling and editing the daily sunspot number data set that is obtained from the Sunspot Index Data Center (SIDC), Royal Observatory of Belgium, via the website (`http://www.oma.be/ KSB-ORB/SIDC/ index.html`), and sometimes takes slightly different values from the data examined here, as demonstrated in Table 2, due mainly to a different way of defining the sunspot number. The spotless data based on the SIDC data set are defined by the number of days with no sunspots and its difference from the NOAA data set is indicated by the value parenthesized in this table.

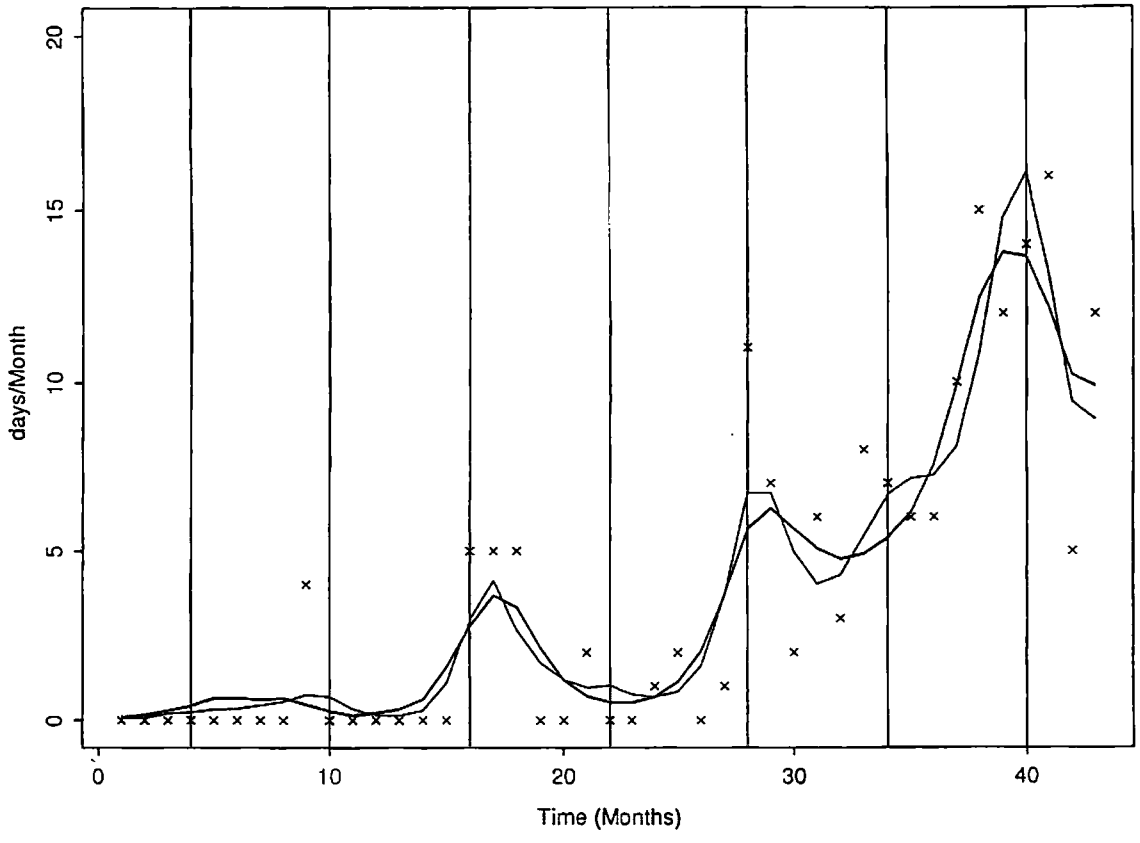

Fig. 6. Monthly sun spotless days (denoted by ×). The thin line is the estimated time-varying mean for the T+Y+H model. The thick line is for the T+Y model.

Table 2
Monthly number of spotless days of the sun

Year	1993	1994	1995	1996
January	0	0	2 (−2)	10 (+3)
February	0	0	0	15
March	0	0	1	12 (−2)
April	0	5	11 (+2)	14 (+2)
May	0	5 (+1)	7	16 (+1)
June	0	5	2	5
July	0	0	6	12
August	0	0	3 (+2)	
September	4 (−4)	2	8 (−1)	
October	0	0	7	
November	0	0	6 (+1)	
December	0	1	6	

We adopt the same observation model as used in the Polio data set. While we used a first order trend model in the previous example, here we assume that t_n follows a second order trend model given by $t_n = 2t_{n-1} - t_{n-2} + v_{n,t}, v_{n,t} \sim N(0, \tau_t^2)$. The system model for s_n is exactly the same as in Section 2.3. Accord-

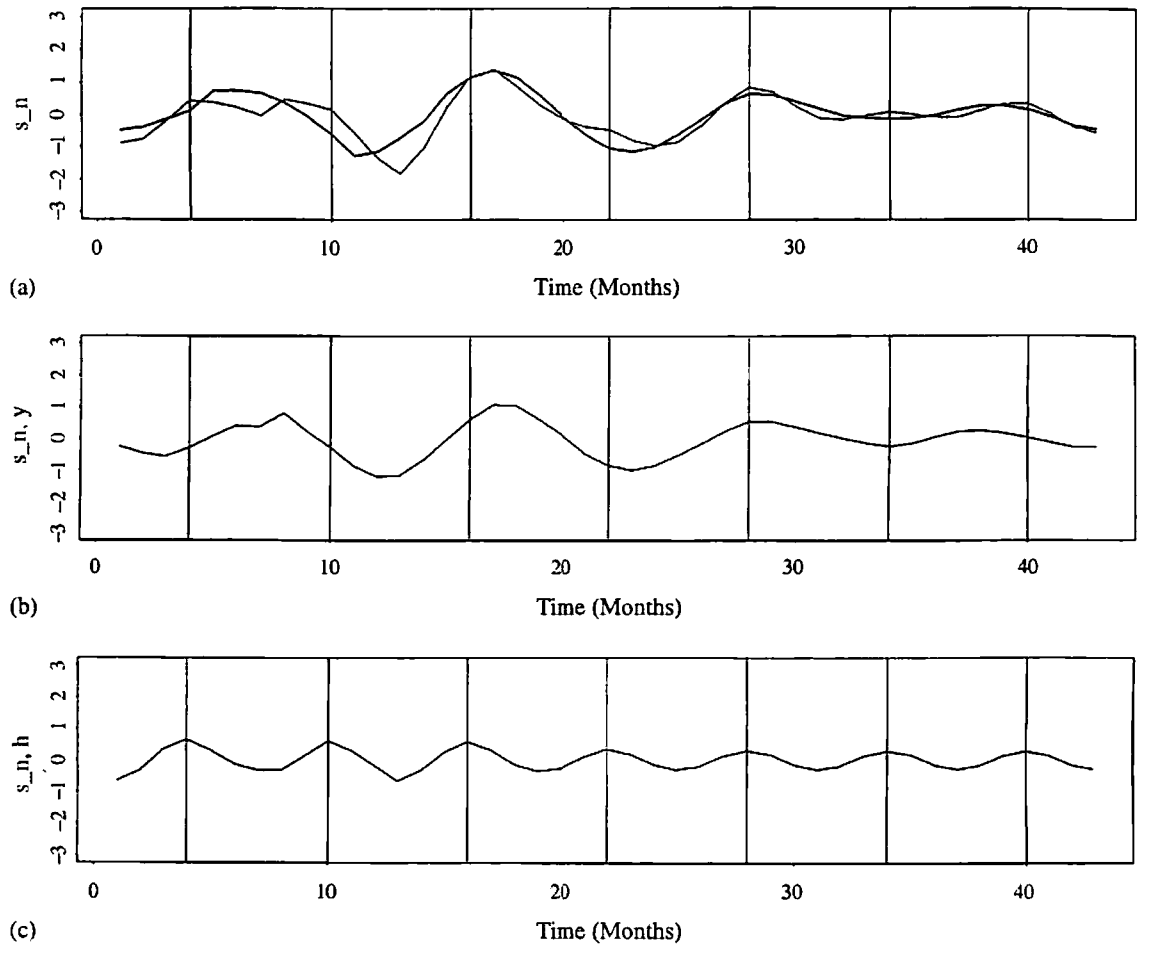

Fig. 7. Estimated seasonal component. (a) Seasonal component for the T+Y+H model (thin line) and T+Y model (thick line). (b) Yearly period component for the T+Y+H model. (c) Half-yearly period component for the T+Y+H model.

ingly, this modeling results in giving the GSSM with a six-dimensional state vector, $x_n = [t_n, t_{n-1} \mid s_{n,y}, s_{n-1,y} \mid s_{n,h}, s_{n-1,h}]^T$. The number of particles and fixed lag used here are $m = 100\,000$ and $L = 20$, respectively. The estimated hyperparameter values are $\widehat{\tau}_t^2 = 0.0019$, $\widehat{\tau}_y^2 = 0.015$, and $\widehat{\tau}_h^2 = 0.00043$.

The estimated time-varying mean is indicated by a thin line in Fig. 6. The thick line seen also in this figure is the estimate of time-varying mean based on the other model that will be explained below. The vertical lines indicate the data for April and October. Fig. 7a shows the seasonal component. It is clear that the seasonal variance decreases with time. Fig. 7b and c show the decomposition of Fig. 7a into yearly and half-yearly components, respectively. The half-yearly component appears to show a maximum phase in March and October. However, it is possible to consider simpler model without taking into account the seasonal variation, we therefore compare four models for describing $\log \lambda_n$ by using AIC: $\log \lambda_n = t_n$ model (T-model), $= t_n + s_{n,y}$ model (T+Y-model), $= t_n + s_{n,h}$ model (T+H-model), $= t_n + s_{n,y} + s_{n,h}$ (T+Y+H-model).

In the MCF, the log-likelihood is approximated by $(1/m)\sum_{i=1}^{m} r(y_n \mid z_{n|n-1}^{(i)})$, and intrinsically suffers from a sampling error. As a result, the AIC value is also subject

Table 3
Hyperparameter values

Model type	k	l	$\widehat{\tau_t^2}$	$\widehat{\tau_y^2}$	$\widehat{\tau_h^2}$	Mean of 100 AIC values
T	1	1	0.0098			204.56
T+Y	3	2	0.00093	0.029		200.83
T+H	3	2	0.0036		0.028	206.50
T+Y+H	5	3	0.0019	0.015	0.00043	205.70

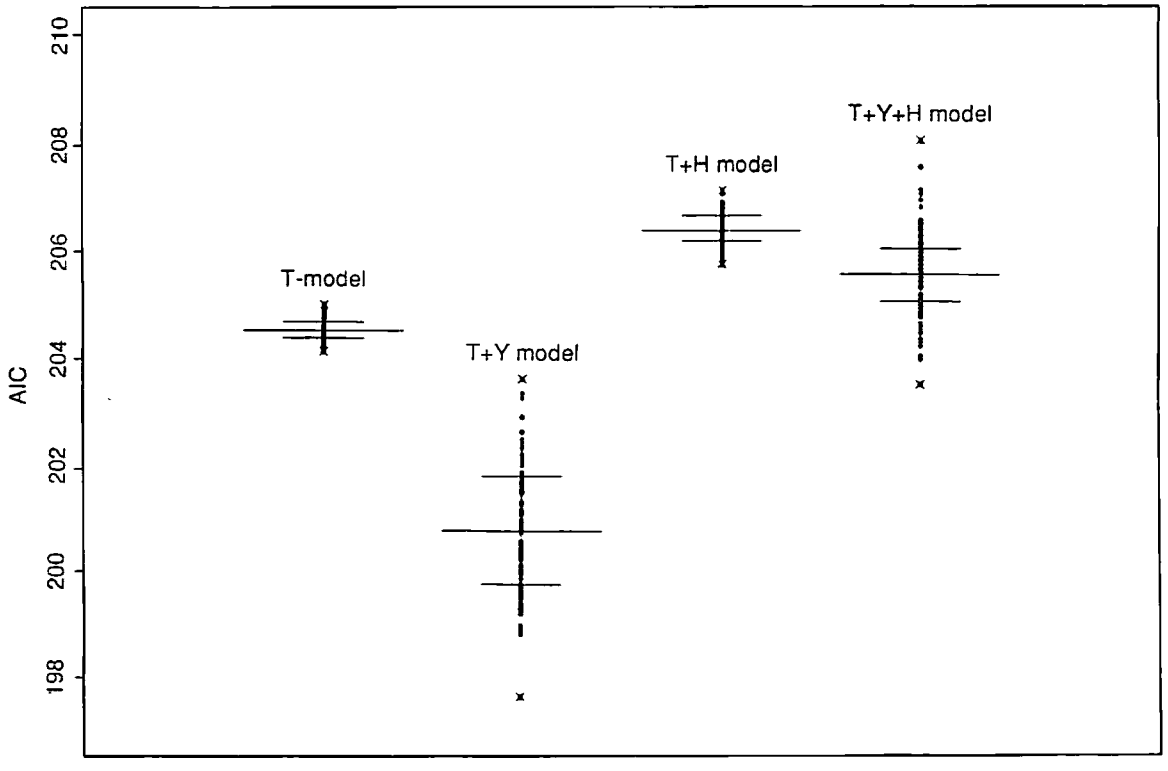

Fig. 8. Distribution of the AIC values for four different system models: T-model ($\log \lambda_n = t_n$), T+Y-model ($=t_n + s_{n,y}$), T+H-model ($=t_n + s_{n,h}$), T−Y+H-model ($=t_n + s_{n,y} + s_{n,h}$). Three horizontal lines for each model are the 25%, 50%, and 75% quantiles, respectively. The crosses are the maximum and minimum values.

to this sampling error. A discussion on how to deal with this problem will be given briefly in the final section. The values of the hyperparameters involved in each model are optimized by maximizing the log-likelihood based on the MCF, and are listed in Table 3. Each calculation for the fixed value of hyperparameters is performed at least 100 times, and its value is defined by the mean among these trials. For simplicity, we fix the number of particles $m = 100\,000$ without paying attention to the difference in both the state vector dimension, k, and the system noise vector dimension, l, for each model. Fig. 8 shows the distribution of 100 AIC values with the best hyperparameters for each model. For each model, the three horizontal lines indicate the 25%, 50%, and 75% quantiles, respectively, of 100 AIC values. The crosses denote the maximum and minimum AIC values. It is clear in this figure

Table 4
Yearly number of cyclones and intense hurricanes

	0	1	2	3	4	5	6	7	8	9
1940s										
Total cyclone number					11	11	6	9	9	13
Intense hurricane number					3	2	1	2	4	3
1950s	13	10	7	14	11	12	8	8	10	11
	6	2	3	3	2	5	2	2	4	2
1960s	7	11	5	9	12	6	11	8	8	18
	2	6	0	2	5	1	3	1	0	3
1970s	10	13	7	8	11	9	10	6	12	9
	2	1	0	1	2	3	2	1	2	2
1980s	11	12	6	4	13	11	6	7	12	11
	2	3	1	1	1	3	0	1	3	2
1990s	14	8	7	8	7	19				
	1	2	1	1	0	5				

that the spotless days number data set favors the T+Y model. The inclusion of the half-yearly component in $\log \lambda_n$ appears to be unnecessary for this data set. The estimated time-varying mean and seasonal component for the T+Y model are indicated in Fig. 6 and Fig. 7a, respectively, by the thick line.

5. Analysis of occurrences of intense hurricanes

In this section we demonstrate the possibility of developing the observation model to handle a small count binary time series. The data set examined is the yearly data for the total number of Atlantic tropical cyclones and numbers of intense hurricanes during the years 1944–1995, shown in Table 4. A description of this data set such as a definition of an intense hurricane can be seen in Landsea et al. (1996).

Landsea et al. (1996) investigated each trend by applying a linear fit to each time series and concluded there was a significant downward trend in intense hurricanes in contrast to an insignificant decrease in the total number of tropical cyclones. Their result is important geophysically because it is contrary to the expectation that globally tropical cyclone activity may be enhanced due to increasing concentration of greenhouse gases (Landsea et al., 1996). In this study, we focus only on the time series of the possibility of having an intense hurricane given by the ratio of y_n to C_n, where C_n and y_n are the numbers of total cyclones and intense hurricanes, respectively. Specifically, we deal with a conditional probability $p(y_n|C_n)$ instead of considering a joint probability $p(y_n, C_n)$.

For an analysis of this data set, we consider the following dynamic binary logit model (in Fahrmeir notation, 1992):

$$y_n \sim \binom{C_n}{y_n} \alpha_n^{y_n}(1-\alpha_n)^{C_n-y_n}, \tag{16}$$

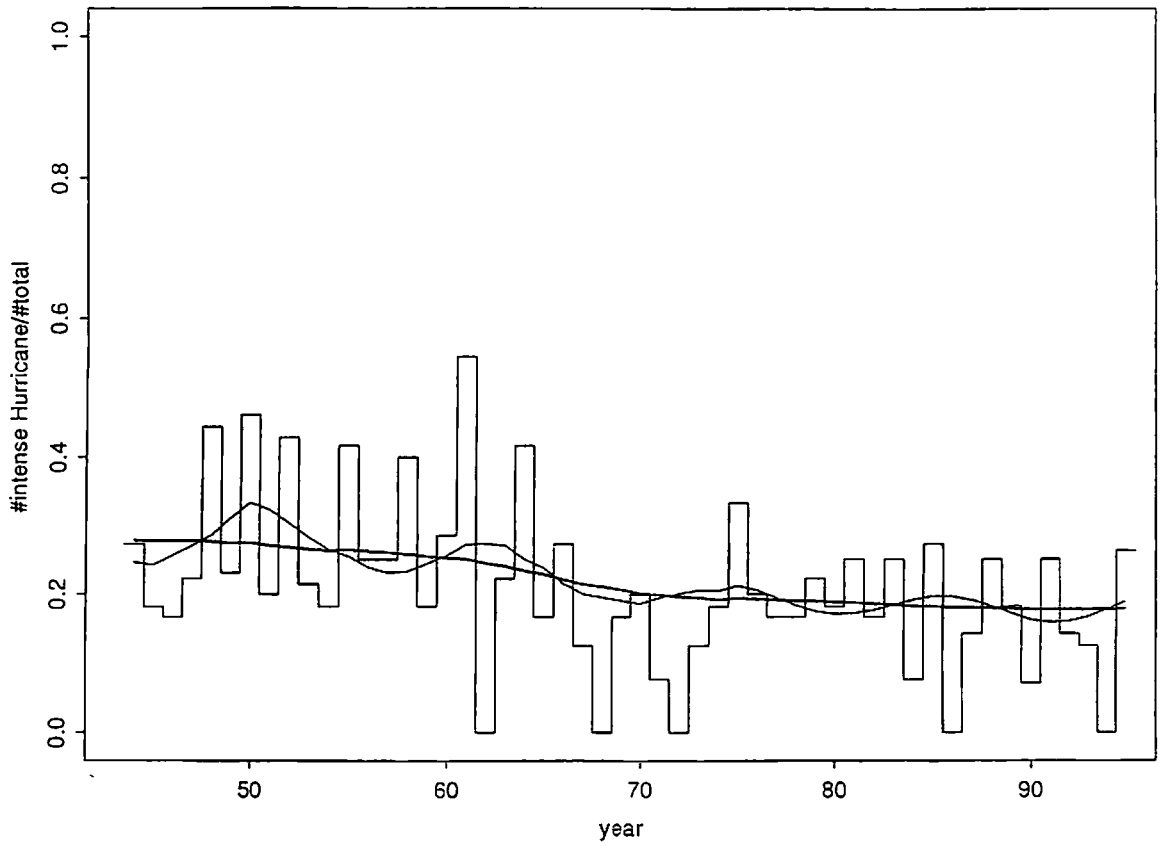

Fig. 9. The estimate of the intense hurricane probability in Atlantic basin, $\widehat{\alpha_n}$. A ratio of the intense hurricane number to the total cyclone number, y_n/C_n, is given by a thin line. The thin and thick lines are a median point of posterior densities of α_n for the T+S and T models, respectively.

where the probability of an occurrence of the intense hurricane, α_n, is related to q_n by $\alpha_n = \exp(q_n)/(1+\exp(q_n))$ through a logit transformation. q_n is in this study decomposed into two factors: a trend component t_n, and solar cycle activity component s_n, $q_n = t_n + s_n$, in an attempt to investigate the effect of solar cycle activity on the occurrence of intense hurricanes. The solar cycle activity component, s_n, is described by the QPO model with a period of 11 yr;

$$s_n = 2\cos\left(\frac{2\pi}{11}\right) s_{n-1} - s_{n-2} + v_{n,s}, \quad v_{n,s} \sim \mathrm{N}(0, \tau_s^2). \tag{17}$$

We assume that t_n follows a first-order trend model given by $t_n = t_{n-1} + v_{n,t}, v_{n,t} \sim \mathrm{N}(0, \tau_t^2)$. This modeling results in giving the GSSM with a three-dimensional state vector $x_n = [t_n | s_n, s_{n-1}]^{\mathrm{T}}$.

Fig. 9 shows a median of $\widehat{\alpha_n}$ based on the posterior distribution $p(\alpha_n|Y_N)$. The number of particles and fixed lag for a smoothing are set to be $m = 100\,000$ and $L = 20$, respectively. The optimized hyperparameter values, τ_t^2 and τ_s^2, are $\widehat{\tau_t^2} = 0.0053$ and $\widehat{\tau_s^2} = 0.00091$, respectively.

We show in Fig. 10a and b the estimated component, $\widehat{t_n}$ and $\widehat{s_n}$, respectively. In Fig. 10b, 16% and 84% points (the so-called $\pm\sigma$ points) for $p(s_n|Y_N)$ are also shown. A significant trend of fewer intense hurricanes is clearly seen in Fig. 10a.

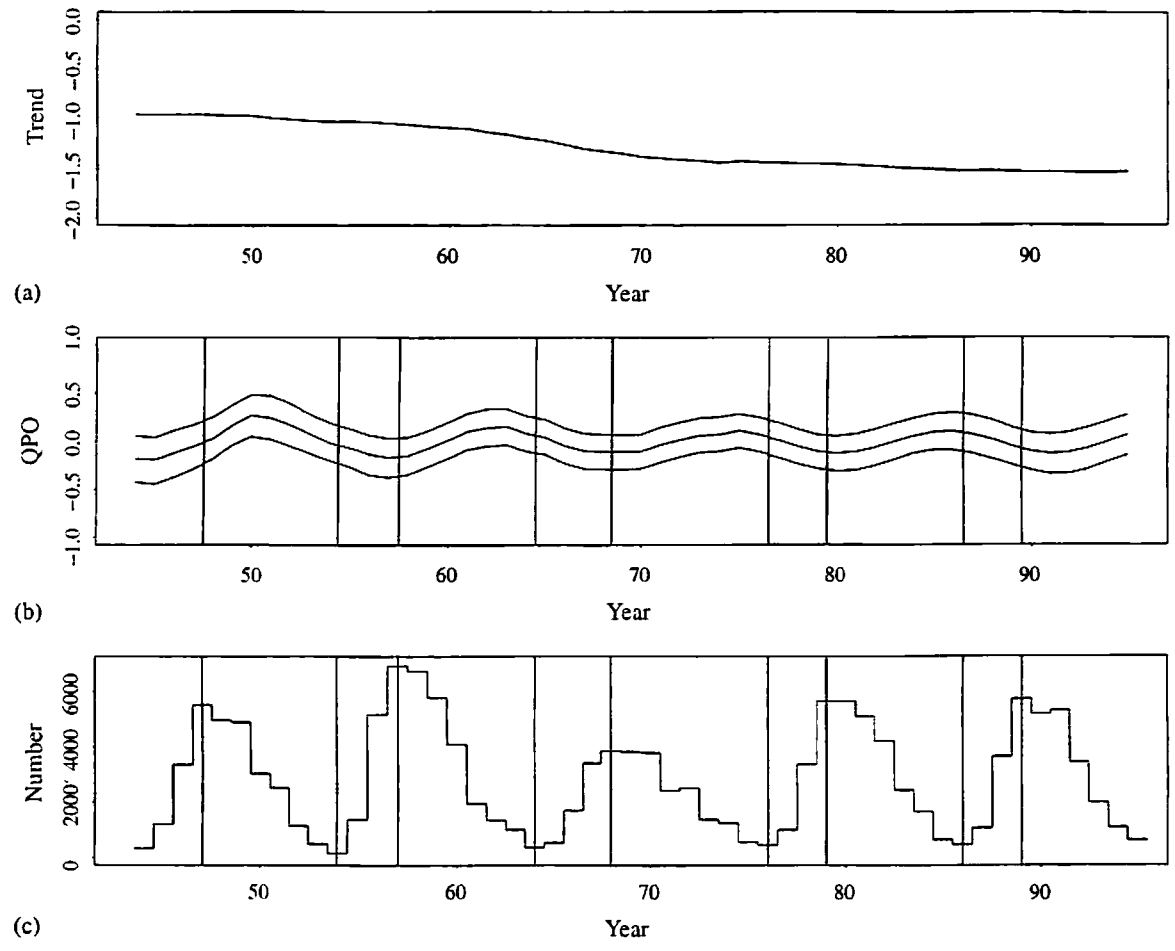

Fig. 10. (a) Estimated trend components $\widehat{t_n}$ for the T+S model (thin line) and T model (thick line). The results of two models are visually indistinguishable. (b) Solar cycle activity component $\widehat{s_n}$ with $\pm\sigma$ points for the T+S model. (c) Yearly sunspot number.

Fig. 10b shows that the QPO component does not show a consistent 11 yr period associated with solar cycle activity. In Fig. 10c, we show a time series of yearly sunspot numbers, calculated using the SIDC data set to show the association of the QPO component to the solar cycle activity. The maximum and minimum points in each solar cycle are indicated by the vertical thick and thin lines, respectively, in Fig. 10b and c.

It is apparent that the phase of the QPO component does not have a good accordance with that of the sunspot number, due mainly to a gradual change in the period of the QPO component. We therefore compare this model (T+S model) with a model (T model) with the simplest description of q_n: $q_n = t_n$. The values of the hyperparameters involved in each model are optimized by maximizing the log-likelihood based on the MCF, and are listed in Table 5. As in the previous attempt to compare four models in Section 4.3, each calculation for the fixed value of hyperparameters is performed at 100 times, and each mean and range of 100 AIC values are shown in Table 5. For simplicity, we fix the particle number $m = 100\,000$ for each case.

Table 5
Comparison of two models

Model type	k	l	$\widehat{\tau_t^2}$	$\widehat{\tau_s^2}$	Mean of 100 AICs	Range of 100 AICs
T	1	1	0.0050		164.16	164.11–164.22
T+S	3	2	0.0053	0.00091	178.74	178.56–178.89

A significantly smaller AIC value for the T model in comparison with that for the T+S model suggests that a trend component alone is sufficient for describing the fluctuation in the intense hurricane probability. It appears that there is no QPO component associated with solar activity. The estimated $\widehat{\alpha_n}$ and $\widehat{t_n}$ for the T model with the best hyperparameter value are superposed in Figs. 9 and 10a by a thick line, respectively. A difference between two estimated trend components can not be seen visually in Fig. 10a.

6. Computational considerations

The log-likelihood value computed through the MCF is intrinsically subject to a sampling error. The simplest way to solve it is to increase the particle number m as much as possible within the limits of computer memory. Another simple technique for reducing the sampling error is to define the log-likelihood value for a fixed value of hyperparameter as a summary statistic based on multiple evaluations of the log-likelihood. In this study, we use both ways together. Actually, because the models adopted in this study belong to the DGLM that is a special case of the GSSM, an other approach based on the importance sampling method is more efficient for estimating the log-likelihood of DGLM (Durbin and Koopman, 1997).

The final estimate of the state vector is given by the result with the maximum likelihood value among trials. Of course, there are other possibilities to define the final estimate by making use of information gathered by a multiple evaluation of the log-likelihood. As for the smoothing, we can take another approach which relies on Markov chain Monte Carlo (MCMC) (e.g., see Shephard and Pitt, 1997 and references therein). If we are simply interested in the smoothing, estimation and testing of model parameters, the method based on MCMC is favorable because an exact calculation of smoothing is difficult within a framework of the MCF. In contrast, if we are concerned with model diagnostics through recursive residuals and functional evaluation of the likelihood for computing, e.g., model choice criteria, the MCF is superior.

In this study we do not pay serious attention to fine tuning of initial-state values to avoid the computational time necessary for optimizing them, because the effect of initial state values on the log-likelihood value is not a significant factor in optimization, in particular for a case with relatively larger N. Actually an ad hoc treatment to control the initial-state values has been applied so as to maximize the log-likelihood.

Acknowledgements

I am grateful to Prof. Kitagawa at the Institute of Statistical Mathematics for his helpful and useful comments. I also thank Prof. C.T. Russell for his hospitality during my visit to University of California, Los Angeles. I wish to thank the referee for his/her valuable comments to improve the manuscript of this paper.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control AC-19, 716–723.

Ameen, J.R.M., Harrison, P.J., 1985. Normal discount Bayesian models. In: Bernard, J.M., DeGroot, M.H., Smith, A.F.M. (Eds.), Bayesian statistics, vol. 2. Elsevier Science Publishers, North-Holland, pp. 271–298.

Anderson, B.D.O., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.

Chan, K.S., Ledolter, J., 1995. Monte Carlo EM estimation for time series models involving counts. J. Am. Statist. Assoc. 90 (429), 242–252.

Carlin, B.P., Polson, N.G., Stoffer, D.S., 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. J. Am. Statist. Assoc. 87 (418), 493–500.

Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.

Doucet, A., Barat, E., Duvaut, P., 1995. A Monte Carlo approach to recursive Bayesian state estimation. Proc. IEEE Signal Processing/Athos Workshop on Higher Order Statistics, 12–14 June, Girona, Spain.

Fahrmeir, L., 1992. Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear model. J. Am. Statist. Assoc. 87 (418), 501–509.

Frühwirth-Schnatter, S., 1994a. Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. Statist. Comput. 4, 259–269.

Frühwirth-Schnatter, S., 1994b. Data augmentation and dynamic linear models. J. Time Series Anal. 15 (2), 183–202.

Good, I.J., 1965. The Estimation of Probabilities. MIT Press, Cambridge, MA.

Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F 140 (2), 107–113.

Grunwald, G.K., Hamza, K., Hyndman, R.J., 1997. Some properties and generalizations of non-negative Bayesian time series models. J. Roy. Statist. Soc. B (3) 615–626.

Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations (with discussions). J. Business Econom. Statist. 7 (4), 407–422.

Harvey, A.C., 1985. Trends and cycles in macroeconomic time series. J. Business Econom. Statist. 3 (3), 216–227.

Harvey, A.C., 1989. Forcasting, Structural Time Series Models, and Kalman Filter. Cambridge University Press, Cambridge.

Higuchi, T., 1991. Frequency domain characteristics of linear operator to decompose a time series into the multi-components. Ann. Inst. Statist. Math. 43, 469–492.

Higuchi, T., Kita, K., Ogawa, T., 1988. Bayesian statistical inference to remove periodic noise in the optical observations aboard a spacecraft. Appl. Opt. 27, 4514–4519.

Kashiwagi, N., Yanagimoto, T., 1992. Smoothing serial count data through a state-space model. Biometrics 48, 1187–1194.

Kitagawa, G., 1987. Non-Gaussian state space modeling of nonstationary time series (with discussion). J. Am. Stat. Assoc. 79, 1032–1063.

Kitagawa, G., 1991. A nonlinear smoothing method for time series analysis. Statistica Sinica 1 (2), 371–388.

Kitagawa, G., 1993. A Monte Carlo filtering and smoothing method for non-Gaussian Nonlinear state space models. Proc. 2nd. Internat. US–Japan joint seminar on statistical time series analysis. 25–29 January, pp. 110–131, Honolulu, USA.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. J. Comput. Graph. Statist. 5 (1), 1–25.

Landsea, C.W., Nicholls, N., Gray, W.M., Avila. L.A., 1996. Downward trends in the frequency of intense Atlantic hurricanes during the past five decades. Geophys. Res. Lett. 23, No. 13, 1697–1700.

Lindley, D.V., Smith, A.F.M., 1972. Bayes estimates of the linear model (with discussion). J. Roy. Statist. Soc. B 34, 1–41.

Martin, R.D., Raftery, A.E., 1987. Robustness, computation, and non-euclidean models. J. Am. Statist. Assoc. 82 (400) 1044–1050.

Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

US Department of Commerce, National Oceanic and Atmospheric, Administration. Space environment Center, 1996. Preliminary report and forecast of solar geophysical data. SWO PRF 1083, 4 June.

West, M., 1995. Bayesian inference in cyclical component dynamic linear models. J. Am. Statist. Assoc. 90 (432), 1301–1312.

West, M., Harrison, P.J., 1989. Bayesian Forecasting and Dynamic Model. Springer, New York.

West, M., Harrison, P.J., Migon, H.S., 1985. Dynamic generalized linear models and Bayesian forecasting (with discussion). J. Am. Statist. Assoc. 80 (389), 73–97.

Zeger, S.L., 1988. A regression model for time series of counts. Biometrika 75, 621–629.