FREQUENCY DOMAIN CHARACTERISTICS OF LINEAR OPERATOR TO DECOMPOSE A TIME SERIES INTO THE MULTI-COMPONENTS

T. HIGUCHI

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

Abstract. Frequency domain properties of the operators to decompose a time series into the multi-components along the Akaike's Bayesian model (Akaike (1980, *Bayesian Statistics*, 143–165, University Press, Valencia, Spain)) are shown. In that analysis a normal disturbance-linear-stochastic regression prior model is applied to the time series. A prior distribution, characterized by a small number of hyperparameters, is specified for model parameters. The posterior distribution is a linear function (filter) of observations. Here we use frequency domain analysis or filter characteristics of several prior models parametrically as a function of the hyperparameters.

Key words and phrases: Time series, Bayesian approach, signal decomposition, linear filter, variable kernel, curve smoothing, smoothness prior, seasonal component model, quasi-sinusoidal wave extraction.

1. Introduction

A Bayesian approach for a time series analysis has been introduced in Akaike (1980) and additionally developed by others: see, for example, Gersch and Kitagawa (1988). There is mainly due to the development of a general Bayesian approach which facilitates fitting a large amount of the time series model parameters. Since the number of parameters in a Bayesian time series model generally exceeds that of observation, such parametric time series model is satisfactorily flexible to the observation data. The Bayesian approach is characterized by specifying a prior distribution of the time series model parameters (Akaike (1980), Silverman (1985), Titterington (1985)). Moreover the Bayesian model contains tradeoff parameters which balance the tradeoff between the infidelity of the model to data and the infidelity of the model to the constraints expressed by the prior models. In Bayesian terminology, this tradeoff parameter is referred to as the hyperparameter (Good and Gaskin (1980), Akaike (1980)). We use this terminology, "hyperparameter", hereafter (Lindley and Smith (1972)).

By a proper choice of hyperparameters, our estimate of the time series model parameter satisfies both the goodness of fit to the data and the conformability to

the prior constraints. The proper choice of hyperparameters was opened to the investigators (Titterington (1985)). One is to calculate the marginal density (Good (1965), Leonard (1978)). Akaike (1980) has developed a Bayesian smoothness idea to yield a likelihood computation for determining the smoothness hyperparameter. The critical idea in smoothness prior approach is the likelihood of a Bayesian model and the use of the likelihood as a measure of the goodness of fit of the model. Other criteria for choosing a hyperparameter were presented and can be referred in Titterington (1985). Among them, the method of "cross-validation" is popular and widely used, particularly for the spline smoothing approach to curve fitting (Wahba and Wold (1975), Silverman (1984b, 1985), Wahba (1990)).

Akaike (1980) also gave an explicit solution to the smoothing problem when both the prior and the conditional data distributions are normally distributed, and then extensively exploited applications to several important statistical data analysis problems. This procedure can be interpreted as one of the maximum penalized likelihood (MPL) method (Good and Gaskin (1971, 1980), Leonard (1978), Kitagawa and Gersch (1984), Titterington (1985), Silverman (1985, 1986)). It should be noticed that Akaike's approach is reduced to be the constrained least squares problem because the Bayesian model is linear with Gaussian system and observation noise is also assumed to be Gaussian (Lindley and Smith (1972)). His colleagues, primarily at the Institute of Statistical Mathematics, Tokyo, extensively applied this Bayesian smoothness approach to a variety of statistical problems (Ishiguro and Arahata (1982), Kashiwagi (1982), Ishiguro and Sakamoto (1983, 1984), Tanabe and Tanaka (1983), Nakamura (1986), Tamura (1987), Sakamoto and Ishiguro (1988), Gersch and Kitagawa (1988)).

An effort to efficiently obtain the best parameters in describing a time series model has been extensively implemented. The smoothness prior approach to a nonstationary time series with trend and seasonality by Akaike and Ishiguro (1983), which originates from Akaike (1980), has computational complexity $O(N^3)$. This complexity was also implemented to be $O(N)$ by the efficient scheme for the computation (Ishiguro (1984)), and by the computationally efficient recursive Kalman filter (Kitagawa (1981)). In the Kitagawa's procedure, the flexibility to the observed data is, of course, achieved by a state space modeling (SSM). In that model, the minimum AIC procedure (Akaike (1973), Sakamoto et al. (1986)) is used to determine the alternative best model. Although there is somewhat difference between Akaike's and Kitagawa's works in formulation, they are essentially the same modeling and give approximately the same estimates of parameters. Hence, from the computative efficiency, the procedures induced by Kitagawa probably favor the sequent data type such as the time series. His procedure using recursive Kalman filter and AIC criterion is widely applied to various data analysis such as the estimations of power spectrum density (Kitagawa (1983)) and time varying AR coefficients (Kitagawa and Gersch (1985a)), and transfer function estimation (Gersch and Kitagawa (1989)). According to an unprecedented development of computational facilities, the Bayesian approach is applied to various field data, and now becomes a handy and useful method for the analysis of the time series only if the prior is a linear model.

The Bayesian smoothing approach to a trend estimation in the discrete setting

is typically used to remove a small fluctuation from the observational data (Akaike
(1980), Kitagawa (1981), Kashiwagi (1982), Akaike and Ishiguro (1983), Tanabe
and Tanaka (1983), Green (1987), Higuchi *et al.* (1988)). Namely, we apply this
procedure to reject the higher frequency components, which contribute to the
irregularities and jitter in the original time series. The Bayesian smoothness model,
therefore, works as a lowpass filter. A spline smoothing approach is also used
to smooth a curve (Wahba (1975, 1990), Wahba and Wold (1975), Ishiguro and
Arahata (1982), Silverman (1984*a*, 1984*b*, 1985, 1986)) and then can be interpreted
as a lowpass filtering procedure (Silverman (1984*a*, 1985)). Silverman (1984*b*,
1985) has shown analytically that this spline smoothing corresponds approximately
to smoothing by a kernel method and that the local bandwidth can be given as a
function of the hyperparameter.

There is also a Bayesian model which is interpreted as a bandpass filter
(Higuchi *et al.* (1988)). This prior was proposed so as to extract a sinusoidal
wave with a characteristic frequency and time varying amplitude. The observed
data obtained by the rotational instruments, such as a rocket, balloon, and satel-
lite, always suffer from such quasi-sinusoidal noise. The time domain filtering by
using the Bayesian approach was successfully applied to the observation data by
the rockets, and its usefulness and flexibility in data analysis were demonstrated
in comparison with the conventional filtering procedures (Kita *et al.* (1989)).

Major purpose in this paper is to study the frequency domain or filter char-
acteristics of the Bayesian time series model, and to show the performance of this
model in frequency domain parametrically as a function of the hyperparameters.
In this study, we consider the decomposition of the observed time series into multi-
components with the observation noise. In Section 2, the model for decomposition
is treated. The prior is imposed on each of the multi-components. Here we assume
only that the Bayesian model is linear and Gaussian system. The definite form of
prior is unnecessary for discussions in Section 2. We will show a frequency domain
characteristics of the linear operator for a decomposition. The several models for
priors are specifically presented by the definite forms in Section 3. In addition
to the smoothness priors, the other priors, which work as filters, are explained.
Numerical examples that illustrate the methodology presented in both Section 2
and 3 are demonstrated in Section 4. The relationships between the performance
of the linear filters by the priors and the hyperparameters are closely examined in
this section.

2. Method

2.1 *Background: The Bayesian model and the least squares analysis*

Given the observations $\{y(i), i = 1, \ldots, N\}$, we consider the multi-component
model,

$$(2.1) \qquad y(i) = s_1(i) + s_2(i) + \cdots + s_M(i) + e(i) \qquad (i = 1, \ldots, N).$$

In (2.1) $e(i)$ is an i.i.d. white noise sequence with $e(i) \sim N(0, \sigma^2)$. The reader
should note that in (2.1) $N \times M$ parameters are to be estimated from the N

observations. Then, the result of the least squares computation is poor. To miti-
gate this difficulty, Akaike (1980) has presented a clear and complete least squares
computational-Bayesian analysis. He has introduced the smoothness prior con-
straint on $N \times M$ parameters in (2.1). As a general form for the smoothness
priors, the "signals", $s_m(i)$ $(m = 1, \ldots, M)$ should each follow in this study an
autoregressive (AR) model

$$(2.2) \qquad \sum_{j=0}^{J_m} a_m(j) s_m(i - j) = u_m(i),$$

where each $u_m(i)$ is a white noise sequence with $u_m(i) \sim N(0, \sigma^2/\tau_m^2)$, and the
$\{a_m(j)\}$ are the AR coefficients for the m-th component model and $a_m(0) = 1$ for
all m. J_m is the order of AR model for m component.

$e(i)$ represents an observation noise and $u_m(i)$ describes a system (or pro-
cess) noise for $s_m(i)$ (Kitagawa (1981)). τ_m^2 is the hyperparameter which balances
the infidelity of the model (2.1) to the data and the infidelity of the model to
the constraints on $s_m(i)$ in (2.2) (Akaike (1980), Kitagawa and Gersch (1985b)).
This hyperparameter is sometimes called a *smoothing parameter* in smoothing
(Silverman (1984a, 1984b, 1985, 1986), Wahba (1990)), or a *roughness penalty*
within a framework of the penalized likelihood method (Good and Gaskin (1971,
1980), Green (1987)). Since the variances of the observation and the system noises,
respectively, are σ^2 and σ^2/τ_m^2, τ_m^2 is interpreted as observation-noise to system-
noise ratio. As τ_m^2 increases, $s_m(i)$ tends to follow a noise-free AR process. For
simplicity, we use hereafter the notation $\tau^2 = [\tau_1^2, \tau_2^2, \ldots, \tau_M^2]$.

By using vector-matrix notation, the model (2.1) can be expressed by

$$(2.3) \qquad y = Xs + e,$$

where

$$(2.4) \qquad y = [y(1), y(2), \ldots, y(N)]',$$
$$(2.5) \qquad e = [e(1), e(2), \ldots, e(N)]',$$
$$(2.6) \qquad s = [s_1', s_2', \ldots, s_M']',$$

and X is a $(N \times NM)$ matrix given by

$$(2.7) \qquad X = [I, I, \ldots, I],$$

where I is a $(N \times N)$ identity matrix. Here the apostrophe symbol $'$ denotes the
matrix transpose. In (2.6) we used the notation

$$(2.8) \qquad s_m = [s_m(1), s_m(2), \ldots, s_m(N)]',$$

for each m. Accordingly, y, e and s_m $(m = 1, \ldots, M)$ are a $(N \times 1)$ vector, therefore
s is a $(NM \times 1)$ vector. Similarly, the model (2.2) is also given in vector-matrix
representation as follows:

$$(2.9) \qquad D_m s_m = u_m,$$

where

$$(2.10) \qquad u_m = [u_m(1), u_m(2), \ldots, u_m(N)]',$$

and D_m is a $(N \times N)$ known design matrix defined by

$(2.11) \quad D_m$

$$= \begin{pmatrix}
\beta_m(1,1) & & & & & & \\
\beta_m(2,1) & \beta_m(2,2) & & & & & \\
\vdots & \vdots & & & & & \\
\beta_m(J_m,1) & \beta_m(J_m,2) & \cdots & \cdots & \beta_m(J_m,J_m) & & \\
a_m(J_m) & a_m(J_m-1) & \cdots & \cdots & a_m(1) & a_m(0) & \\
 & a_m(J_m) & a_m(J_m-1) & \cdots & \cdots & a_m(1) & a_m(0) \\
 & & \ddots & \ddots & & \ddots & \ddots \\
 & & & a_m(J_m) & a_m(J_m-1) & \cdots & \cdots & a_m(1) & a_m(0)
\end{pmatrix}$$

where $\beta_m(i,j)$ are significantly small numbers that are chosen to satisfy initial conditions: i.e., they are boundary conditions necessary for $s_m(i)$ $(i \le J_m)$.

Here we consider the constrained least squares problem which, for fixed value of τ^2, minimizes

$$(2.12) \qquad \sum_{i=1}^{N} \left(|e(i)|^2 + \sum_{m=1}^{M} |\tau_m u_m(i)|^2 \right).$$

In vector-matrix notation, the minimizor of this quantity is given by

$$(2.13) \qquad \|e\|^2 + \|\tau_1 u_1\|^2 + \cdots + \|\tau_M u_M\|^2.$$

$\|\cdot\|$ denotes the Euclidean norm. Introducing the following $(NM \times NM)$ matrix given by

$$(2.14) \qquad D = \begin{pmatrix} \tau_1 D_1 & & & \\ & \tau_2 D_2 & & \\ & & \ddots & \\ & & & \tau_M D_M \end{pmatrix},$$

(2.2) is equivalent to

$$(2.15) \qquad Ds = u,$$

where $u = [\tau_1 u'_1, \tau_2 u'_2, \ldots, \tau_M u'_M]'$. Using vector representations of (2.3) and (2.15), (2.13) can be rewritten as

$$(2.16) \qquad \|y - Xs\|^2 + \|Ds\|^2.$$

We denote s which minimizes (2.16) for fixed value of τ^2, by $\hat{s}$.

The solution for the constrained least squares problem presented by (2.16) is easily given by

$$(2.17) \qquad \hat{s} = (X'X + D'D)^{-1}X'y$$

(Kitagawa and Gersch (1985b)). This solution for the "signal", $\hat{s}$, is the expected value of the posterior distribution if we interpret this estimation procedure in a Bayesian way (Lindley and Smith (1972), Akaike (1980), Silverman (1985), Titterington (1985)).

2.2 Representation in a time domain

In (2.17), we denote an $(NM \times N)$ matrix, $(X'X + D'D)^{-1}X'$, as T and use the notation $T = [T_1, T_2, \dots, T_M]'$, where T_m is of an $(N \times N)$ matrix. Accordingly an optimum solution for m-th component, $\hat{s}_m$, can be expressed as

$$(2.18) \qquad \hat{s}_m = T_m y.$$

This means that the resulting estimate of the unknown model parameter, $\hat{s}_m$, is a linear function of the observations. T_m has the following convolution structure. From (2.18),

$$(2.19) \qquad \hat{s}_m(i) = \sum_{j=1}^{N} T_{m,i\cdot j}\, y(j) = \sum_{p=i-N}^{i-1} C_{m,i}(p) y(i-p),$$

where $T_{m,i\cdot j}$ is the (ij) component of T_m. In (2.19) we set $T_{m,i\cdot(i-p)} = C_{m,i}(p)$.

The form of (2.19) to estimate the signal $s_m(i)$ appears in the spline smoothing approach to nonparametric regression (Wahba (1975), Silverman (1984a, 1986)). In that case, $C_{m,i}(p)$ (or $T_{m,i\cdot(i-p)}$) is called a *weight function*. Silverman (1984a, 1985) has shown that this weight function is closely related to a *kernel* (or convolution or moving average) smoother and that a weight function for estimating the data points of i is independent of i under the following conditions:

(A1) large N,

(A2) small value of the hyperparameter (in this study τ_m^2),

(A3) not too close to the boundary (i.e., data point i is within certain range of $\lambda + 1 \le i \le N - \lambda$).

The condition of (A2) is intrinsically required for the smoothing problem. Under the conditions of (A1 and A3), $C_{m,i}(p)$ can take the similar form nearly independent of i for certain range of $\lambda + 1 \le i \le N - \lambda$, and shows $C_{m,i}(p) = 0$ for $p < -\lambda$ and $\lambda < p$. Obviously, the value of λ depends on the value of hyperparameters. Outside this range of i (within $i < \lambda$ and $N - \lambda < i$), $C_{m,i}(p)$ suffers from the boundary effect of data set. Within the range of $\lambda + 1 \le i \le N - \lambda$, (2.19) is approximately rewritten as

$$(2.20) \qquad \hat{s}_m(i) = \sum_{p=-\lambda}^{\lambda} C_m(p) y(i-p).$$

(2.20) exactly indicates that $\hat{s}_m(i)$ is obtained by convoluting $C_m(\cdot)$ with $y(\cdot)$. For simplicity, we rewrite (2.20) in vector notation as $\hat{s}_m = C_m \otimes y$, where C_m is a column vector defined by $[C_m(-\lambda), C_m(-\lambda+1), \ldots, C_m(0), \ldots, C_m(\lambda)]'$ and $\otimes$ denotes a convolution operator.

2.3   Representation in a frequency domain

We now consider the characteristics of the linear operator T_m in a frequency domain. Let a $(N \times N)$ matrix of a discrete Fourier transform be F. Since F is a unitary matrix, F^{-1} is given by F'. The operator T_m in the frequency domain, T_m^*, can be given by

$$(2.21) \qquad T_m^* = F T_m F^{-1} = F T_m F'.$$

T_m^* transforms y^* into $\hat{s}^*$, where y^* and $\hat{s}^*$ are the representations for y and $\hat{s}$ in a frequency domain, respectively: $y^* = Fy$ and $\hat{s}^* = F\hat{s}$. It should be noticed that since T_m is not unitary, T_m^* is also not unitary. Let F be as follows:

$$(2.22)$$
$$F_{1,j} = \frac{1}{\sqrt{N}},$$
$$F_{2i,j} = \sqrt{\frac{2}{N}} \sin\left(\frac{2\pi ij}{N}\right), \quad F_{2i+1,j} = \sqrt{\frac{2}{N}} \cos\left(\frac{2\pi ij}{N}\right) \quad \left(i = 1, 2, \ldots, \frac{N}{2} - 1\right),$$
$$F_{N,j} = \frac{1}{\sqrt{N}} \cos(\pi j),$$

for $j = 1, 2, \ldots, N$, where F_{ij} denotes the (ij) component of F.

When the operator, T_m^*, selectively screens or filters certain frequency components of a signal, T_m^* can be thought of as filters. We now consider the characteristics of T_m^* in terms of the terminology of a digital filter: gain and phase. As previously mentioned, the commonly used linear filter takes the form of a *kernel* (or *convolution*) smoother. Convoluting in a time domain corresponds to multiplication in a frequency domain (according to the convolution theorem). Hence a convolution in a time domain is, in the frequency domain, equivalent to multiplying the discrete Fourier transformation (DFT) of a convolution function (Karl (1989)). In this study, we adopt (2.22) as one way for DFT, and thereby a multiplication in a frequency domain takes the following form:

$$(2.23) \quad Z = \begin{pmatrix} A_0 & & & & & & & \\ & A_1 & -B_1 & & & & & \\ & B_1 & A_1 & & & & & \\ & & & A_2 & -B_2 & & & \\ & & & B_2 & A_2 & & & \\ & & & & & \ddots & & \\ & & & & & & A_{N/2-1} & -B_{N/2-1} \\ & & & & & & B_{N/2-1} & A_{N/2-1} \\ & & & & & & & & A_{N/2} \end{pmatrix}.$$

In this case, a gain and phase for a frequency f ($f = k/N$), $G(f)$ and $\phi(f)$, are defined by

(2.24)
$$G(f) = G(k/N) = \sqrt{A_k^2 + B_k^2},$$
$$\phi(f) = \phi(k/N) = \tan^{-1}\left(\frac{B_k}{A_k}\right).$$

For $k = 0$ and $k = N/2$, $G(f)$ and $\phi(f)$ are defined by setting $B_k = 0$.

Although there exist somewhat effects from the finite data number, the actually calculated T_m^* through (2.21) with (2.22) in general takes the same form as (2.23) when the prior model is given by the linearly local constraints on $s_m(i)$ such as (2.2) and $2\lambda/N$ is sufficiently small. Namely, when the number of data points which suffer from the boundary effects, 2λ, is much smaller than N, the convolution function $C_m(p)$ exists for a wide range of i ($\lambda + 1 \leq i \leq N - \lambda$) and accordingly the values except for A_k and B_k in (2.23) are very negligible as compared with those of A_k and B_k. Hence, we hereafter assume that T_m^* follows the form expressed by (2.23). Comparing the elements of T_m^* with (2.23), here we define the following frequency domain quantities:

(2.25)
$$\hat{G}_m(0) = \left|T_{m,1\cdot1}^*\right|,$$
$$\hat{G}_m(f) = \hat{G}_m(k/N) = \sqrt{\left(T_{m,2k+1\cdot2k}^*\right)^2 + \left(T_{m,2k+1\cdot2k+1}^*\right)^2}$$
$$\left(k = 1, 2, \ldots, \frac{N}{2} - 1\right),$$
$$\hat{G}_m(1/2) = \left|T_{m,N\cdot N}^*\right|$$

and

(2.26)
$$\hat{\phi}_m(0) = \phi(1/2) = 0,$$
$$\hat{\phi}_m(f) = \hat{\phi}(k/N) = \tan^{-1}\left(\frac{T_{m,2k+1\cdot2k}^*}{T_{m,2k+1\cdot2k+1}^*}\right) \quad \left(k = 1, 2, \ldots, \frac{N}{2} - 1\right).$$

$\hat{G}(f)$ and $\hat{\phi}(f)$ are defined on $0 \leq f \leq 1/2$. $\hat{G}_m(f)$ and $\hat{\phi}_m(f)$, respectively, represent the gain and phase for the linear operator in frequency domain, T_m^*.

3. Models for signal constraints

Several Bayesian linear models such as (2.2) are considered here. First, the models which are widely used for smoothing are presented. We classify these models as a *smoothness model*. Secondly, we show the models which work in a frequency domain as a *bandpass filter*. These are sometimes used as one method of searching for a signal with known frequency buried in broadband noise. In addition, we demonstrate the *seasonal component models* which are naturally adopted to represent a signal clearly showing a seasonal pattern.

3.1 Smoothness model

Consider a model with smooth or lowpass filter behavior. In this case, a naturally adopted smoothness constraint is that the k-th order difference of $s_m(i)$, $\nabla^k s_m(i) = u_m(i)$, is assumed to be normally distributed zero-mean sequence with variance σ^2/τ_m^2. For example, for $k = 1$ and $k = 2$ those constraints are expressed in terms of the following $a_m(i)$:

(model d_1) $\qquad a_m(0) = 1, \qquad a_m(1) = -1,$

and

(model d_2) $\qquad a_m(0) = 1, \qquad a_m(1) = -2, \qquad a_m(2) = 1.$

The design matrix D_m in (2.11) is accordingly defined by $a_m(j)$. The resulting filtered data, hence, shows a smooth behavior. Then, the smoothness constraint is interpreted to work as a lowpass filter.

3.2 Bandpass filter

A bandpass filter is generally used to extract a sinusoidal signal with a characteristic frequency f_c from the original data. However, a general bandpass filter does not work well for a quasi-periodic wave which shows a gradual change in its amplitude and/or an abrupt shift in its phase. Such situations often occur in data taken aboard a rotational body such as a spacecraft, rocket, and balloon.

Assume a quasi-periodic signal is with the model

(model qp_1) $\quad a_m(0) = 1, \qquad a_m(1) = -2\cos(2\pi f_c \Delta t), \qquad a_m(2) = 1,$

where Δt is a sampling time (Higuchi *et al.* (1988)). The subscript in qp_1, 1, indicates that a signal consists of one quasi-periodic wave. Since the model qp_1 specifies a local relationship around $s_m(i)$ and the dependence of $s_m(i)$ on $s_m(i+h)$ for the larger value of h becomes smaller, this model satisfactorily represents a decaying (or growing) sinusoid. This model can describe a sinusoid whose amplitude gradually changes as a function of time (Higuchi *et al.* (1988)).

When the signal consists of multi-quasi-periodic signals, the regression coefficients $a_m(j)$ are obtained by comparing the coefficients of $s_m(\cdot)$ in the following equation:

(model qp_p) $\qquad \displaystyle\sum_{j=0}^{2p} a_m(j)s_m(i-j) = \prod_{j=1}^{p} \left(1 - 2\cos(2\pi f_{c,j}\Delta t)B + B^2\right)s_m(i),$

where $f_{c,j}$ is a characteristic frequency of j-th component and B is a backward operator: $Bs_m(i) = s_m(i-1)$.

In association with the model qp_p, we here present a model which describes a signal consisting of multi-decaying (and/or growing) sinusoids. The coefficients $a_m(j)$ of this model can be obtained by the following manner. When $s_m(i)$ is described as a set of decaying (and/or growing) sinusoids, $s_m(i)$ satisfies the following linear difference equation:

(3.1)
$$\prod_{j=1}^{p}(z_j - B)(z_j^\dagger - B)s_m(i) = 0,$$

$$z_j = \exp\left(2\pi(g_j + if_j)\Delta t\right),$$

where g_j and f_j are the growth rate and frequency of j-th component, respectively (Kay and Marple (1981)). Here, B is the backward operator as in model qp_p, and the dagger represents the complex conjugate operation. By setting $\gamma_j = \exp(2\pi g_j \Delta t)$, (3.1) can be rewritten as

$$(3.1') \qquad \prod_{j=1}^{p} \left(B^2 - 2\gamma_j \cos(2\pi f_j \Delta t)B + \gamma_j^2 \right) s_m(i) = 0.$$

Setting $\sum_{j=0}^{2p} a_m(j) s_m(i - j) = 0$, we compare the coefficients of $s_m(i - j)$ with those in (3.1), for each j, and then get the coefficients of $a_m(j)$ consequently.

3.3 Seasonal component model

For a signal with a "cyclic" behavior with a period of r, the following seasonable component model is satisfactorily useful. The seasonal component model is realized by the following $a_m(j)$

(model si_r) $\qquad a_m(0) = 1, a_m(1) = 1, \ldots, a_m(r - 1) = 1$

(Kitagawa and Gersch (1984)). For example, when $s_m(i)$ is monthly data and clearly shows a strong seasonal pattern, $s_m(i)$ satisfies the constraint which minimizes $\sum_{i=12}^{N} |\sum_{j=0}^{11} s_m(i - j)|^2$. This constraint has been applied to the economic data (Akaike and Ishiguro (1983), Gersch and Kitagawa (1983), Kitagawa and Gersch (1984)) and to observational data (Koike (1990)). For a seasonal component model with a period of r, the following simple constraint is naturally available:

(model sd_r) $a_m(0) = 1, a_m(1) = 0, \ldots, a_m(r - 1) = 0, a_m(r) = -1$

(Akaike and Ishiguro (1983), Ishiguro (1984)). It is impossible for the sd_r model to eliminate the DC level component, because the solution of $s_m(i) = s_m(i - 1)$ can also satisfy this sd_r condition (Kitagawa (1986)). Accordingly, the resulting decomposed signal by using this model for a seasonal component model intuitively contains the DC level component. When a signal is decomposed into a trend, seasonal and other components, the constraint of sd_r for a seasonal component should be used together with that of si_r because the trend component can not be uniquely defined, for a fixed value of the hyperparameters, without using the si_r model (Akaike and Ishiguro (1983)).

4. Numerical examples

4.1 Smoothness model

First we consider a simple model with $M = 1$ and one hyperparameter $\tau^2 = [\tau_1^2]$. That is, we decompose a time series into two components: $y(i) = s_1(i) + e(i)$. For $s_1(i)$, we adopt a model of d_2.

Since we use the discrete Fourier transformation expressed by (2.22), a frequency resolution, Δf, is given by $1/N$. A function in frequency domain, therefore, can be expressed in detail with increasing N. In the following results, we use $N = 200$.

We present the time domain form of linear operator, $T_{m,i\cdot j}$. Since $M = 1$, we omit $m = 1$ in $T_{m,i\cdot j}$, and simply denote $T[i, j]$. Figure 1 shows five curves of $T[i, j]$ for $i = 1, 25, 100, 175$ and 200 as a function of j. The vertical value is normalized by the maximum value of $T[i, j]$. Here we set $\tau^2 = [2^8]$. $T[i, j]$ for each i reaches its maximum value at $j = i$. There is little difference among three curves of $T[25, j]$, $T[100, j]$ and $T[175, j]$. $T[1, j]$ and $T[200, j]$ reflect the boundary condition and they are approximately a mirror image each other.
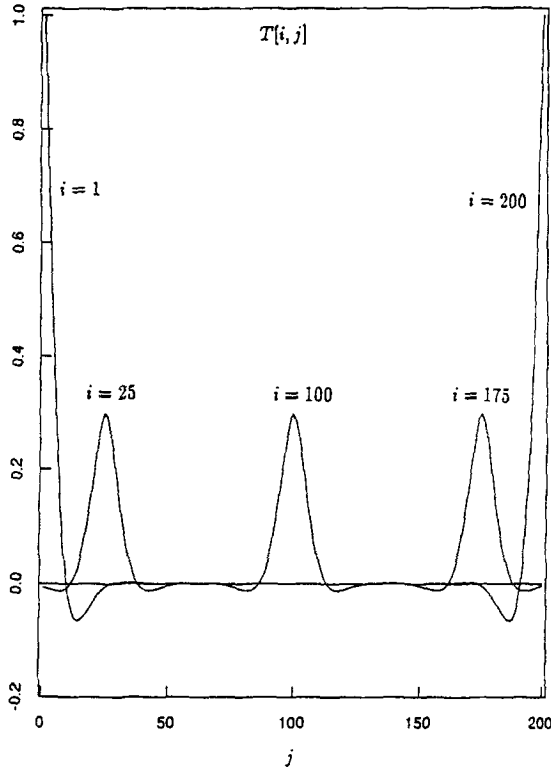
Fig. 1. $T[i, j]$ for $i = 1$, 25, 100, 175 and 200 as a function of j.

Since for $\tau_1^2 = 2^8$, $T[i, j]$ is almost independent of i for $30 \leq i \leq 170$, it is reasonable to consider that $T[100, j]$ represents the convolution function $C_1(p)$ in (2.20). Figure 2 illustrates $C_1(p)$ normalized by $C_1(0)$. The essential support of $C_1(p)$ is roughly $-30 \leq p \leq 30$ in this case, that is, λ in (2.20) is approximately 30. Accordingly $C_{1,i}(p)$ in (2.19) is independent of i only within $30 \leq i \leq 170$.

We obtain the frequency domain characteristics of $T[i, j]$, $T^*[i, j]$, by utilizing the aforementioned procedures, and thus calculate the gain $\hat{G}_1(f)$ and phase $\hat{\phi}_1(f)$. Here we demonstrate a typical example of $T^*_{m,i\cdot j}$ to easily understand an interpretation of $T^*_{m,i\cdot j}$ as a commonly used digital filter. Figure 3 shows a frequency domain behavior of the linear operator of the model d_2, $T^*[i, j]$. For the purpose of easier visual understanding, we exhibit in Fig. 3 only a part of $T^*[i, j]$: $T^*[i, j]$ ($i = 1, \ldots, 81$ and $j = 1, \ldots, 81$). The frequency range shown in this figure
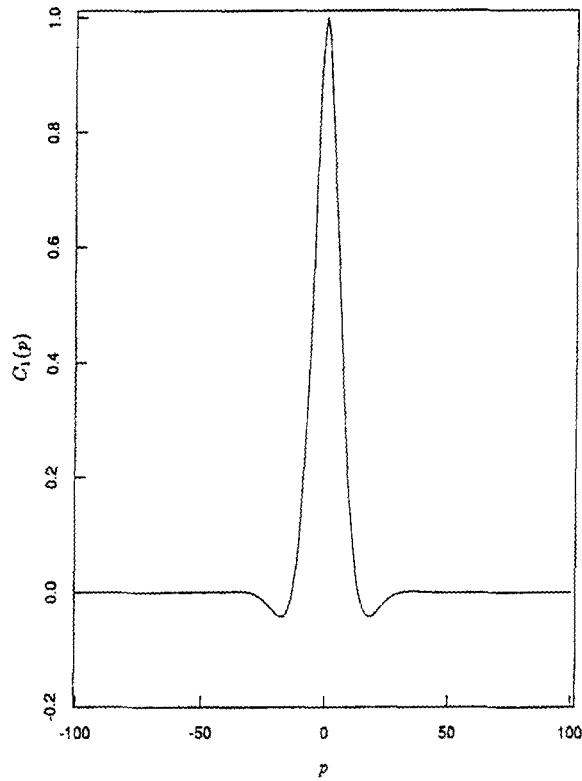
Fig. 2. $C_1(p)$ normalized by $C_1(0)$.
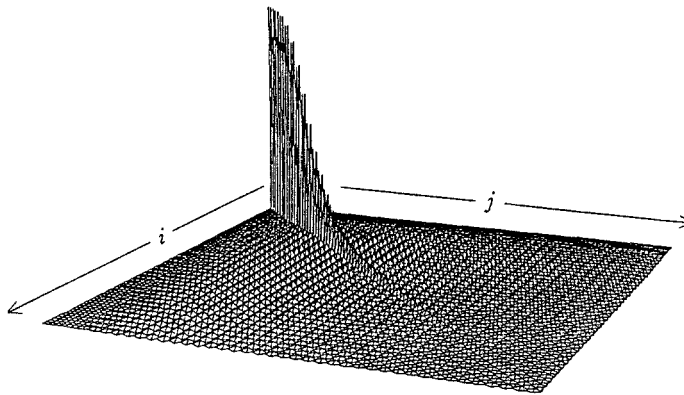
Fig. 3. $T^*[i,j]$.

is, hence, limited to $f \leq 0.2$. It is clear that $T^*[i,j]$ is nearly zero except for diagonal elements, and that the diagonal element, $T^*[i,i]$, is tapering toward zero with increasing i (f). These results sufficiently support the anticipation that $T^*[i,j]$ can be interpreted as a lowpass filter. From $T^*[i,j] = 0$ for $i \neq j$, it follows that the phase in (2.26), $\hat{\phi}_1(f) = 0$.

We display in Fig. 4(a) the gain of $T^*[i,j]$, $\hat{G}_1(f)$, approximately correspond-ing to an envelope of $T^*[i,i]$ in Fig. 3. The value of $\hat{G}_1(f)$ is calculated by (2.25), and normalized by $\hat{G}_1(0)$. From (2.25), the definition range of frequency is, as previously mentioned, limited within $0 \leq f \leq 1/2$. Figure 4(a) clearly shows that this linear operator, $T[i,j]$, works as a lowpass filter in frequency domain, and that the components with a frequency larger than 0.1 are rejected by $T^*[i,j]$. We note that $\hat{G}_1(f)$ monotonously decreases; i.e. $T[i,j]$ rejects higher frequencies. In Fig. 4(b), the phase $\hat{\phi}_1(f)$ on the definition of (2.26) is illustrated in degree. The phase is almost confined within the small range from -30 to 0, particularly, in the frequency range of $f \leq 0.1$, it approximately takes the value of $0°$. This means that $T^*[i,j]$ has almost zero phase shift at frequencies less than 0.1. In addition, since $\hat{\phi}_1(0) = \hat{\phi}_1(1/2)$, $T^*[i,j]$ is thought to have a *minimum phase* (Karl (1989)).

(a)

(b)
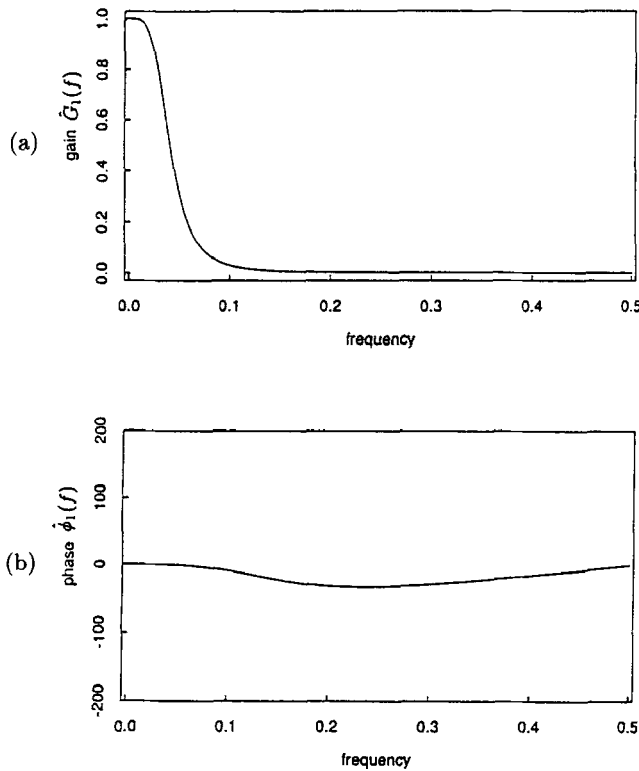
Fig. 4.   (a) The gain $\hat{G}_1(f)$ and (b) phase $\hat{\phi}_1(f)$.

The gain presented in Fig. 4(a) is obtained with $\tau_1^2 = 2^8$ for the model d_2. Obviously, as mentioned in Section 2, the form of $\hat{G}_1(f)$ completely depends on the value of τ_1^2. Increasing τ_1^2 increases the smoothing action on y and reduces the cutoff frequency.

To examine a dependency of the function form of $\hat{G}_1(f)$ on τ_1^2, Fig. 5 shows $\hat{G}_1(f)$ vs. f for five values of τ_1^2: $\tau_1^2 = 2^{12}$, 2^8, 2^4, 2^0 and 2^{-4} with each curve normalized by its $\hat{G}_1(0)$. We also specify the normalized gain by $\hat{G}_1(f)$. Each curve peaks at $f = 0$, and displays a monotonically decreasing response. As shown
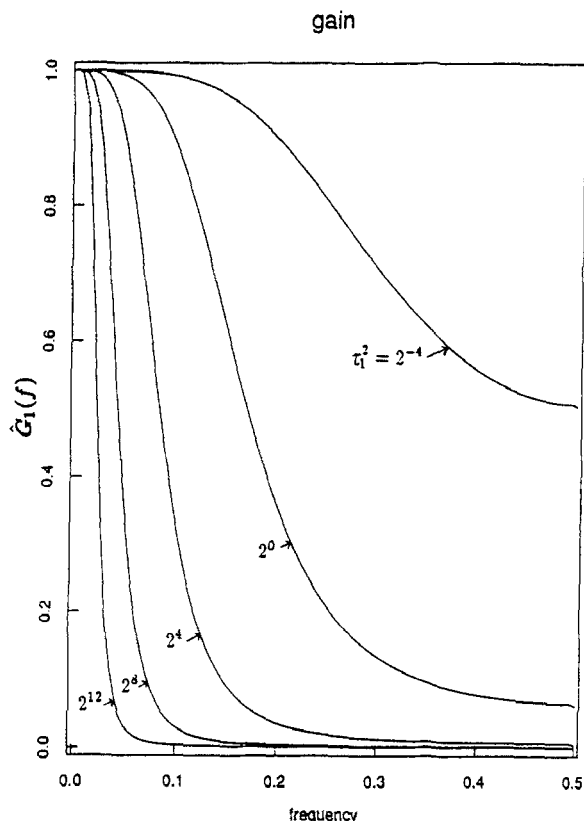
gain

Fig. 5.   The gains for several values of τ_1^2: $\tau_1^2 = 2^{12}$, 2^8, 2^4, 2^0 and 2^{-4}.

in the figure, the lowpass filter achieved by using d_2 model with small τ_1^2 has a broad transition zone, where the transition zone is a frequency range of $\hat{G}_1(f) = 1 \rightarrow \hat{G}_1(f) = 0$; $\hat{G}_1(f)$ for small τ_1^2 displays a gradual decrease.

For small τ_1^2 ($\tau_1^2 \le 2^4$), $\hat{G}_1(f)$ has no cutoff. Because $\hat{G}_1(f)$ is obtained numerically, it is inevitably contaminated with numerical errors. Then, instead of the cutoff, we now adopt

$$(4.1) \qquad\qquad f_{1/2} = \hat{G}_1^{-1}(1/2).$$

Namely, to examine the dependency on τ_1^2, we use the frequency at which $\hat{G}_1(f)$ takes a half value ($\hat{G}_1(f_{1/2}) = 1/2$).

Figure 6 is a curve of the half value frequency, $f_{1/2}$, against τ_1^2 on doubly logarithmic scale (the base of logarithm is 2). It is clearly seen that $\log_2 f_{1/2}$ linearly decreases with increasing $\log_2 \tau_1^2$. The correlation coefficient between them is -0.9995. Using a least square straight line fit to $\log_2 \tau_1^2$ vs. $\log_2 f_{1/2}$ plots yields

$$(4.2) \qquad\qquad \log_2 f_{1/2} = -2.565 - 0.2534 \log_2 \tau_1^2.$$

Accordingly, the relation between τ_1^2 and $f_{1/2}$ is approximately given by

$$(4.3) \qquad\qquad f_{1/2} \simeq \frac{1}{6\sqrt{\tau_1}}.$$
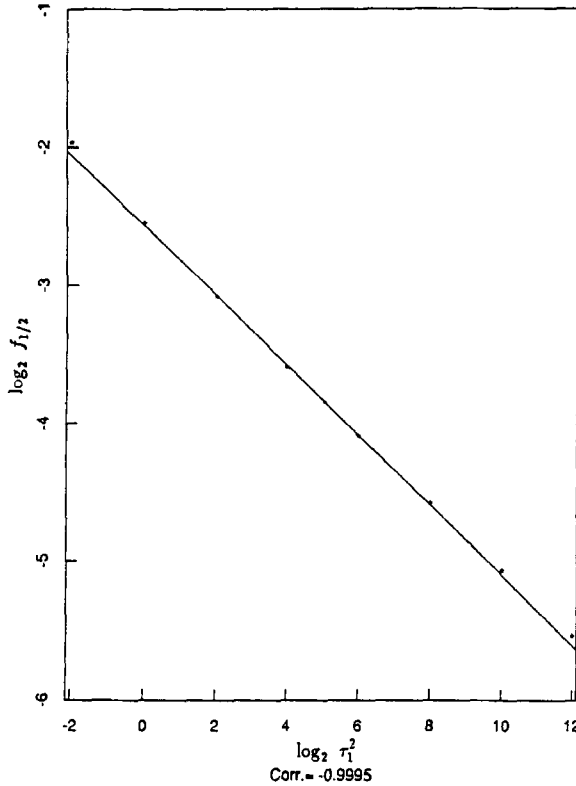
Fig. 6.   The logarithm of the half value width, $\log_2 f_{1/2}$, against $\log_2 \tau_1^2$.

Using this relation, the desired lowpass filter can be constructed in the following manner. If we give a value of $f_{1/2}$, the hyperparameter τ_1^2 is directly defined according to (4.3). The resulting τ_1^2 gives a form of $T[i, j]$.

4.2   Quasi-sinusoidal model

In the next example, we try to examine the performance of the qp_1 model as a bandpass filter. We here add one component $s_2(i)$, which follows a qp_1 model, to the model previously used as a lowpass filter; we consider the model such that $y(i) = s_1(i) + s_2(i) + e(i)$, where $s_1(i)$ and $s_2(i)$ follow the d_2 and qp_1 models, respectively. Two hyperparameters, τ_1^2 and τ_2^2, are required to represent this model. To examine a performance of the model corresponding to the bandpass filter, we fix the value of τ_1^2 which controls the property of the lowpass filter: $\tau_1^2 = 1$. In the qp_1 model, we set a characteristic frequency f_c to be $1/8$.

Figure 7 shows four curves of $\hat{G}_2(f)$ for $\tau_2^2 = 2^0, 2^4, 2^8$ and 2^{12}, where $\hat{G}_2(f)$ denotes the value normalized by $\hat{G}_2(f_c)$ for each curve. It was previously mentioned that as the hyperparameter τ_2^2 becomes larger, the constraint on $s_2(i)$ is intensified to emphasize the pure sinusoidal pattern. Accordingly, for $\tau_2^2 \to +\infty$, it is expected that $s_2(i)$ is completely a sinusoid. In this case, only the component with the frequency of f_c is extracted by T_2. This is a consequence of the fact that $\hat{G}_2(f)$ is given by $\delta(f - f_c)$, where $\delta(\cdot)$ is a delta function. Clearly, Fig. 7 supports
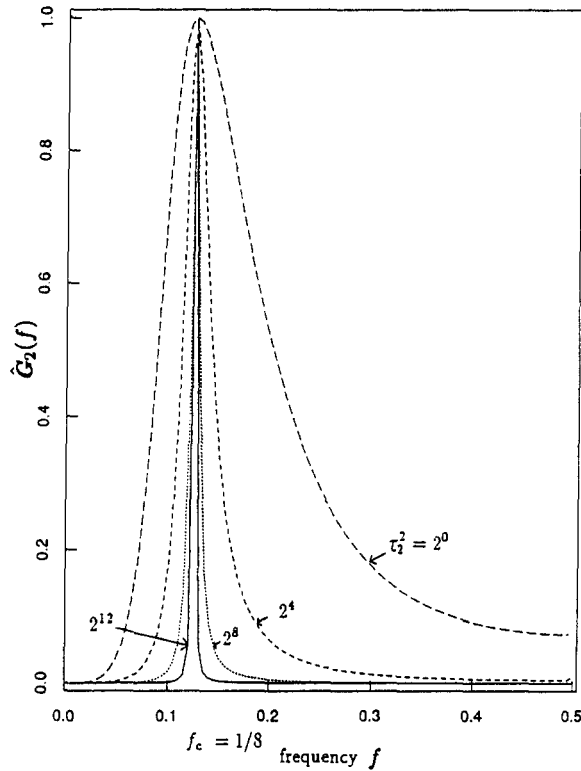
Fig. 7. The gains $\hat{G}_2(f)$ for $\tau_2^2 = 2^0$, 2^4, 2^8 and 2^{12}. For each τ_2^2, the value is normalized by $\hat{G}_2(0)$.

this expectation for large τ_2^2. For example, $\hat{G}_2(f)$ for $\tau_2^2 = 2^{12}$ is nearly to a delta function with a sharp peak. The smaller τ_2^2 shows that $s_2(i)$ is more weakly subject to restriction for the sinusoidal pattern, and the Periodgram of $\hat{s}_2(i)$ shows a broader band around f_c. This means that for smaller τ_2^2, $\hat{G}_2(f)$ exhibits a broader skirt. Figure 7 clearly supports, as expected, these anticipations on a dependence of $\hat{G}_2(f)$ on τ_2^2.

In the further step, we present a model with four components: $y(i) = s_1(i) + s_2(i) + s_3(i) + e(i)$. Three hyperparameters, τ_1^2, τ_2^2 and τ_3^2, are therefore required to represent this model. $s_1(i)$ follows a d_2 model, and both $s_2(i)$ and $s_3(i)$ are required to obey a qp_1 model. In constructing qp_1 model, we specify the characteristic frequency, f_c, for $s_2(i)$ and $s_3(i)$ components by f_{c_2} and f_{c_3}, and here set to $1/4$ and $1/8$, respectively. Although many models can be considered according to a value of a set of the hyperparameters, $\tau^2 = [\tau_1^2, \tau_2^2, \tau_3^2]$, we now give two examples.

Figure 8(a) shows three curves for $\hat{G}_1(f)$, $\hat{G}_2(f)$ and $\hat{G}_3(f)$. The values of hyperparameters are set to be $\tau^2 = [2^{-2}, 2^6, 2^{10}]$. $\hat{G}_m(f)$ indicates the value normalized by a maximum value of $\hat{G}_m(f)$ for each m. As seen in Fig. 5, the small value of τ_1^2 leads both a broad transition zone and a high half value frequency $f_{1/2}$. However, on comparison with Fig. 5, we should note that the frequency components just with $f = 1/4$ and $f = 1/8$ are completely rejected through $T_1[i,j]$.
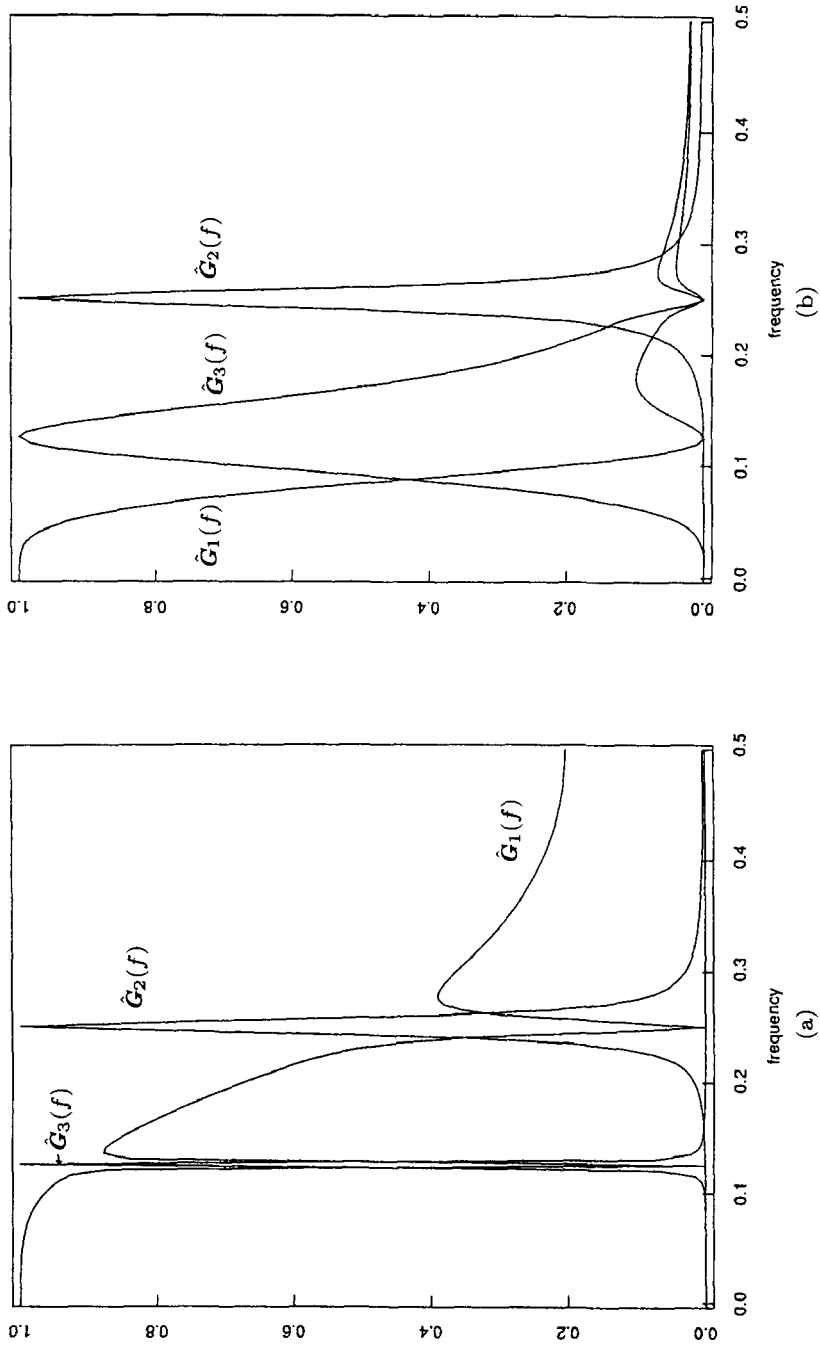
Fig. 8. (a) The three gains, $\hat{G}_1(f)$, $\hat{G}_2(f)$ and $\hat{G}_3(f)$, for $\tau^2 = [2^{-2}, 2^6, 2^{10}]$. (b) The three gains for $\tau^2 = [2^2, 2^6, 2^2]$.

Figure 8(a) distinctly shows that these frequency components with $f = 1/4$ and $f = 1/8$ are, respectively, extracted by $T_2[i,j]$ and $T_3[i,j]$. Because of the very large value of τ_3^2, $\hat{G}_3(f)$ comes near the delta function. The constraint of the smoothness prior is weak due to the small value of τ_1^2. $f_{1/2}$ seen in $\hat{G}_1(f)$ is high thereby.

In addition, we exhibit in Fig. 8(b) three gains for three components with $\tau^2 = [2^2, 2^6, 2^2]$. The model and characteristic frequencies are the same used in Fig. 8(a). Because of the same value of $\tau_2^2 = 2^6$, $\hat{G}_2(f)$ is identical to $\hat{G}_2(f)$ shown in Fig. 8(a). The support of $\hat{G}_3(f)$ becomes wider according to small value of τ_3^2. That is distinctly different appearance from that in Fig. 8(a), which is nearly to the delta function. Since τ_1^2 is relatively large, the smoothness constraint becomes stricter and the cutoff frequency is reduced to the lower frequency.

4.3 Decaying (or growing) sinusoidal wave model

Furthermore, we demonstrate a gain of the model closely related to the quasi-sinusoidal wave. This model, as previously presented in Section 3, describes a set of decaying and/or growing sinusoids. For simplicity, we use a simple model with two components: $y(i) = s_1(i) + e(i)$. In addition, we set $s_1(i)$ to represent a growing or decaying sinusoidal wave. Then, only values of two coefficients in (3.1), g_1 and f_1, are required. Of course, the value of a hyperparameter, τ_1^2, is also necessary for constructing a prior model. Since the value of Δt is not essential in this study, we set $\Delta t = 1$. Since, for the quasi-sinusoidal wave, we satisfactorily understood the relationship between the gain and hyperparameters, it is easy to assume the relationship for a growing or decaying model. We therefore examine the dependency of g_1 on a gain under the fixed values of τ_1^2 and f_1: $\tau_1^2 = 2^{10}$ and $f_1 = 0.25$.

Figure 9(a) demonstrates $\hat{G}_1(f)$ against several values of g_1: $g_1 = 0, 0.025, 0.05$ and 0.125. The negative (positive) value of g_1 represents that s_1 is a growing (decaying) sinusoid. It should be noticed that the model with $g_1 = 0$ is equivalent to qp_1 model with $f_{c,1} = 0.25$. All gain shows a peak at $f = 0.25$, indicating that the filter works as a bandpass filter. With increasing g_1, the skirt of $\hat{G}_1(f)$ becomes wider and finally $\hat{G}_1(f)$ with $g_1 = 0.125$ shows a w-shape behavior. This means that for a comparatively small value of the growth rate, the linear filter works as a bandpass filter like that of qp_1 model. However, for the larger g_1, the components in wider frequency range can be passed through this filter. We examine $\hat{G}_1(f)$ for the negative value of g_1; i.e. we demonstrate $\hat{G}_1(f)$ for a growing sinusoidal wave. The absolute values of the used g_1 are the same as those for a growing sinusoidal wave: $g_1 = 0, -0.025, -0.05$ and -0.125. Figure 9(b) shows $\hat{G}_1(f)$ against the several values of g_1. It is also seen that the skirt of $\hat{G}_1(f)$ becomes wider as an increase in $|g_1|$ (decrease in g_1). Nevertheless, it should be noticed that even if g_1 takes the same absolute value (except for $g_1 = 0$), $\hat{G}_1(f)$ in Fig. 9(a) is different from that in Fig. 9(b). In particular, $\hat{G}_1(f)$ in Fig. 9(b) with $g_1 = -0.125$ shows significantly different behavior from that in Fig. 9(a) with $g_1 = 0.125$.
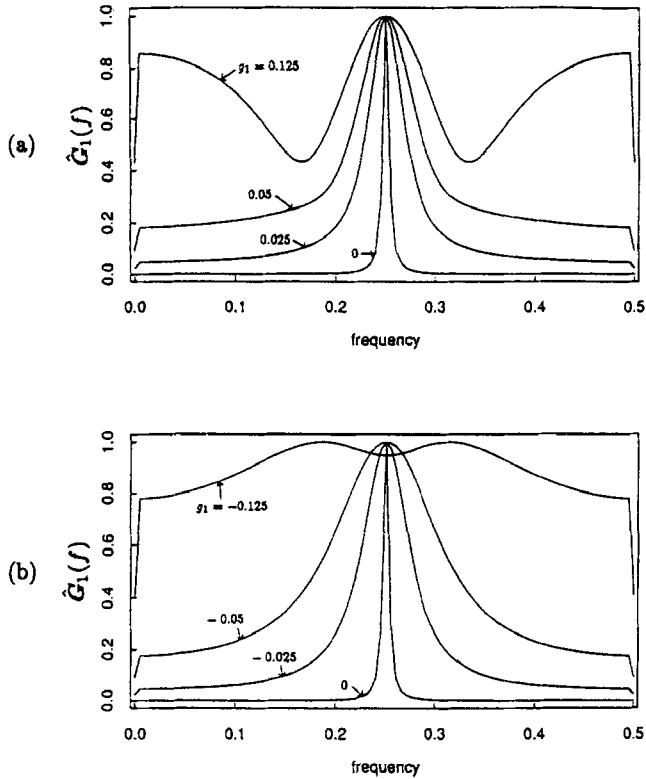
Fig. 9.   (a) The gains of the decaying sinusoidal wave model, $\hat{G}_1(f)$, against several values of g_1: $g_1 = 0, 0.025, 0.05$ and 0.125. (b) The gains of the growing sinusoidal wave model, $\hat{G}_1(f)$ against $g_1 = 0, -0.025, -0.05$ and -0.125.

4.4   Seasonal component model

As previously mentioned, when the signal shows a cyclic pattern with a certain period, a seasonal component model is satisfactorily applied. We examine the gain of this model by using a simple model of $y(i) = s_1(i) + e(i)$. To construct the prior model, we set $r = 10$; the period, r, is 10.   Then, since we use $\Delta t = 1$ and $N = 200$ in this study, the frequency of this cyclic signal is given by $1/r\Delta t = 0.1$. Figure 10(a) demonstrates $\hat{G}_1(f)$ for the si_{10} model against hyperparameters: $\tau_1^2 = 2^0, 2^2, 2^4$ and 2^6.   All curves show peaks at the frequency of an integral times of 0.1: $f = 0.1, 0.2, 0.3, 0.4$ and 0.5. It is shown that as τ_1^2 is smaller, the skirt around each peak becomes wider. Moreover, the tendency that the higher frequency is preferentially passed becomes enhanced with decreasing in τ_1^2.

We examine $\hat{G}_1(f)$ of the sd_{10} model under using the same parameters used above. $\hat{G}_1(f)$ is demonstrated in Fig. 10(b) against the hyperparameters.   Of course, all curves show peaks at the frequency $f = 0.1, 0.2, 0.3, 0.4$ and 0.5.   In addition, it has a peak at $f = 0$, because as previously mentioned, the sd_{10} model for a seasonal component obviously contains the DC level (i.e., $s_1(i) = s_1(i-1)$). It is also seen in Fig. 10(b) that increasing in τ_1^2 makes each skirt wider. This tendency of broadening is independent of the frequency of peaks. Therefore, it does not occur that the higher frequency is preferentially passed by this model.
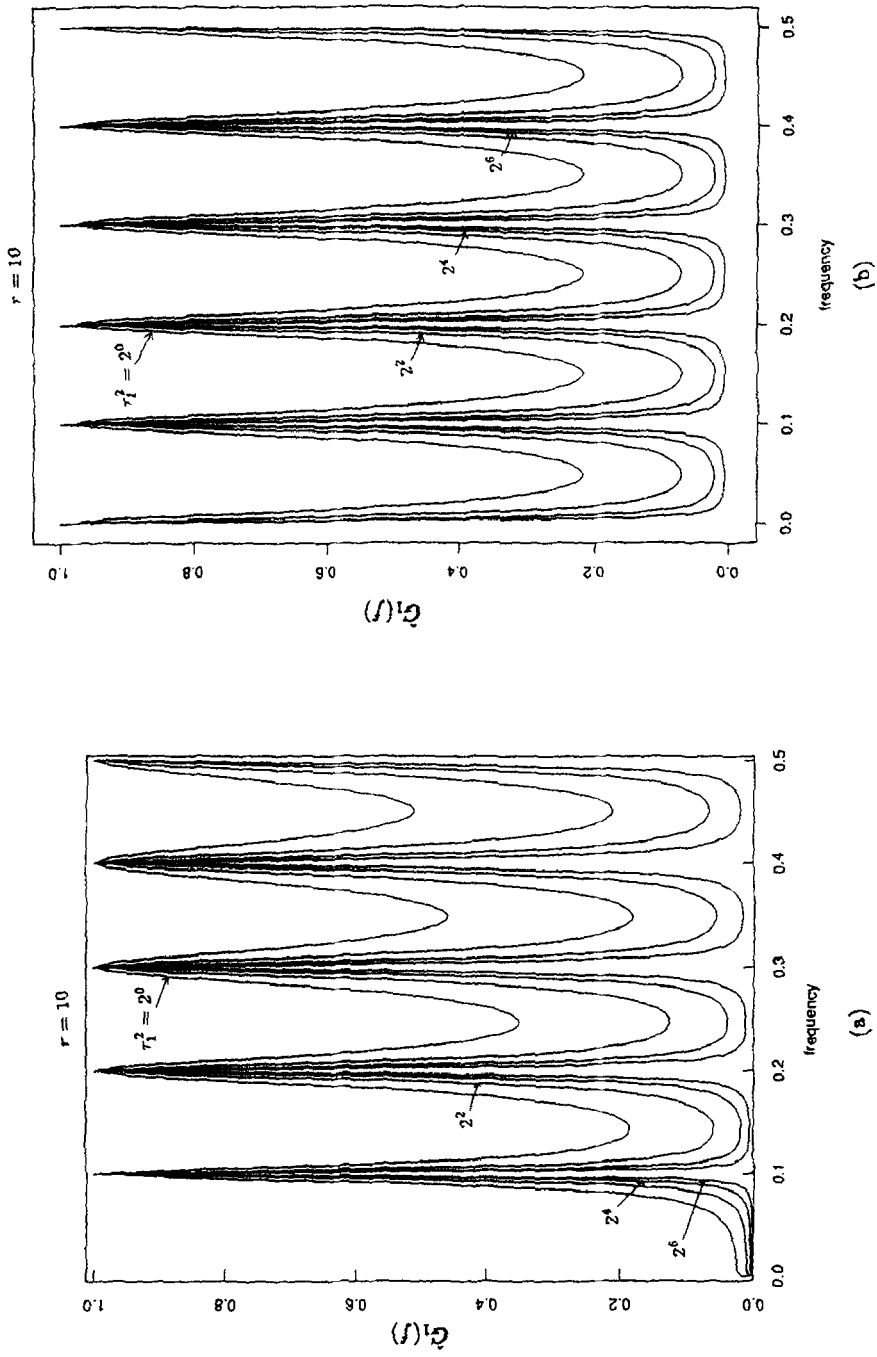
Fig. 10.   (a) The gains of the seasonal component model si_{10}, $\hat{G}_1(f)$, for several values of τ_1^2; $\tau_1^2 = 2^0, 2^2, 2^4$ and 2^6.   (b) $\hat{G}_1(f)$ of the sd_{10} model, for several values of τ_1^2: $\tau_1^2 = 2^0, 2^2, 2^4$ and 2^6.

This result is a performance distinguishable from that for the si_{10} model.

5. Summary and concluding remarks

The problem of interest is to exhibit the frequency domain characteristics of the linear operator to decompose an observed time series into the multi-components. We used the Bayesian model for each of the decomposed multi-components. Assuming that both the prior and conditional data distributions are normally distributed, the solution to the best decomposition is then reduced to be the constrained least squared problem for a fixed value of the hyperparameters. Modifying this solved equation, we obtained the convenient form that each of decomposed components can be given by operating the matrix to the column vector composed of the observed data. By the discrete Fourier transformation, we got the representation of this matrix (linear operator) in a frequency domain. After such representation, we examined the characteristics of the Bayesian model in a frequency domain and discussed its performance in terms of terminology of the linear filters, as a function of the hyperparameters.

To illustrate our methodology, we have presented several Bayesian models which are considered to work as filters, and made their frequency domain representation. First, we investigated the smoothness prior model which is commonly used for removing the high fluctuations from the observed data. In this study we have qualitatively shown its relationship between the hyperparameters and the cutoff frequency. This cutoff frequency is closely related to the bandwidth defined in kernel-function representation. The relationship between the bandwidth and hyperparameters for a spline smoothing has been already established by Silverman (1984a, 1985). Although our approach to smooth a curve is different from that made by a spline smoothing method, the cutoff frequency is substantially the same concept as the bandwidth.

The prior model to extract a quasi-sinusoidal wave with a characteristic frequency can be widely applied to quasi-sinusoidal waves even with the time varying amplitude. In relation to this model, we presented the prior model to get the waves consisting of the multi-sinusoidal waves from an observational data. For the present there is no actual application of this model, but it may be useful for reaching a signal which consists solely of sinusoids in additive white noise. Additionally we demonstrated the decaying (or growing) sinusoidal prior model, which is a generalized model of a quasi-sinusoidal wave prior. This prior model implies that the signal can be expressed by a linear combination of exponentials. Using such specific, rather complicated model for an examination of the free oscillations (e.g., Earth's free oscillations (Fukao and Suda (1989))) is reasonably justified, because it precisely specifies the relationship necessary for the free oscillations.

The seasonal component model enables us to extract a signal showing a seasonal pattern. This model has been successfully applied to the economic data (Akaike and Ishiguro (1983), Gersch and Kitagawa (1983), Kitagawa and Gersch (1984)) and to observational data (Koike (1990)). Naturally it is suggested to be useful and efficient for the experimental data which suffer from artificial instrumental noise showing a strong cyclic pattern, due to the periodicity inherent in a experimental method such as a rotation of the rocket, or radiation of radio waves

with a characteristic frequency, or the instruments using an alternating current (AC).

In this study we use the Akaike's formulation to decompose a time series into multi-components. In an actual data analysis for a signal decomposition, a simple approach to convolute $C_1(\cdot)$ with $y(\cdot)$ ($C_1 \otimes y$) can give us the satisfactory results, instead of one to multiply $T[i,j]$ by y. Moreover we note that if τ_1^2 is given, the algorithm by using a state-space form signal model (Kitagawa (1981)) is useful and effective rather than computing the operator $T[i,j]$. This state-space model is accompanied by the smoothness priors, and equivalent to the d_1 or d_2 model presented in this study.

Finally, we would mention the treatment of non-linear Bayesian model. In this study, we used the linear model for the priors. The linear model is, of course, applicable in various fields and its usefulness has been already established. In addition to its flexibility to the observed data, its computational efficiency is very favorable for actual analysis. However, more general prior, which is no longer a linear model, can be implemented by recursive formula for the one-ahead prediction, filtering and smoothing (Kitagawa (1987)). In his model, the assumption of Gaussian system is also no longer required. Namely, the non-Gaussian distributions are induced for more general modeling. Because of its non-linearity, an estimate of the decomposed signal is no longer a linear function of the observations and hence it is impossible to make its frequency domain representation as demonstrated in this study. A more comprehensive treatment of this approach is referred in Kitagawa (1987).

Acknowledgements

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiadó, Budapest.

Akaike, H. (1980). Likelihood and the Bayes procedure (with discussion), *Bayesian Statistics*, 143–165, University Press, Valencia, Spain.

Akaike, H. and Ishiguro, M. (1983). Comparative study of the X-11 and BAYSEA procedures of seasonal adjustment, *Applied Time Series Analysis of Economic Data* (ed. A. Zellner), 17–53, U.S. Department of Commerce, Bureau of the Census, Washington, D.C.

Fukao, Y. and Suda, N. (1989). Core modes of the earth's free oscillations and structure of the inner core, *Geophysical Research Letter*, **16**, 401–404.

Gersch, W. and Kitagawa, G. (1983). The prediction of time series with trends and seasonalities, *J. Business & Economic Statistics*, **1**, 253–264.

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

Gersch, W. and Kitagawa, G. (1989). Smoothness priors transfer function estimation, *Automatica—J. IFAC*, **25**, 603–608.

Good, I. J. (1965). *The Estimation of Probabilities*, MIT Press, Cambridge, Massachusetts.

Good, I. J. and Gaskin, R. A. (1971). Non-parametric roughness penalties for probability densities, *Biometrika*, **58**, 255–277.

Good, I. J. and Gaskin, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data (with discussion), *J. Amer. Statist. Assoc.*, **75**, 42–73.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models, *Internat. Statist. Rev.*, **55**, 245–259.

Higuchi, T., Kita, K. and Ogawa, T. (1988). Bayesian statistical inference to remove periodic noise in the optical observations aboard a spacecraft, *Applied Optics*, **27**, 4514–4519.

Ishiguro, M. (1984). Computationally efficient implementation of a Bayesian seasonal adjustment procedure, *J. Time Ser. Anal.*, **5**, 245–253.

Ishiguro, M. and Arahata, E. (1982). A Bayesian spline regression, *Proc. Inst. Statist. Math.*, **30**, 29–36 (in Japanese).

Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115–137.

Ishiguro, M. and Sakamoto, Y. (1984). A Bayesian approach to the probability density estimation, *Ann. Inst. Statist. Math.*, **36**, 523–538.

Karl, J. H. (1989). *An Introduction of Digital Signal Processing*, Academic Press, San Diego.

Kashiwagi, N. (1982). Estimation of fertilities in field experiments, *Proc. Inst. Statist. Math.*, **30**, 1–10 (in Japanese).

Kay, S. M. and Marple, S. L., Jr. (1981). Spectrum analysis—A modern perspective, *Proc. IEEE*, **69**, 1380–1419.

Kita, K., Higuchi, T. and Ogawa, T. (1989). Bayesian statistical inference of airglow profiles from rocket observational data: comparison with conventional methods, *Planet Space Science*, **37**, 1327–1331.

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

Kitagawa, G. (1983). Changing spectrum estimation, *J. Sound Vibration*, **89**, 433–445.

Kitagawa, G. (1986). Decomposition of a nonstationary time series—An introduction of the program DECOMP—, *Proc. Inst. Statist. Math.*, **34**, 255–271 (in Japanese).

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

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. (1985a). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series, *IEEE Trans. Automat. Control*, **AC-30**, 48–56.

Kitagawa, G. and Gersch, W. (1985b). A smoothness priors long AR model method for spectral estimation, *IEEE Trans. Automat. Control*, **AC-30**, 57–65.

Koike, M. (1990). A study on stratospheric ozone distribution with the backscattered ultraviolet spectrometer aboard the satellite OZORA, Doctor Thesis, University of Tokyo.

Leonard, T. (1978). Density estimation, stochastic process and prior information (with discussion), *J. Roy. Statist. Soc. Ser. B*, **40**, 113–146.

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

Nakamura, T. (1986). Bayesian cohort models for general cohort table analysis, *Ann. Inst. Statist. Math.*, **38**, 353–370.

Sakamoto, Y. and Ishiguro, M. (1988). A Bayesian approach to nonparametric test problems, *Ann. Inst. Statist. Math.*, **40**, 587–602.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*, Reidel, Dordrecht.

Silverman, B. W. (1984a). Spline smoothing: the equivalent variable kernel method, *Ann. Statist.*, **12**, 898–916.

Silverman, B. W. (1984b). A fast and efficient cross-validation method for smoothing parameter choice in spline regression, *J. Amer. Statist. Assoc.*, **79**, 584–589.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **47**, 1–52.

Silverman, B. W. (1986). Density estimation for statistics and data analysis, *Monographs on Statist. Appl. Probab.*, Chapman and Hall, London.

Tamura, Y. H. (1987). An approach to the nonstationary process analysis, *Ann. Inst. Statist. Math.*, **39**, 227–241.

Tanabe, K. and Tanaka, T. (1983). Fitting curves and surface by Bayesian modeling, *Gekkan Chikyu*, **5**, 179–186 (in Japanese).

Titterington, D. M. (1985). Common structure of smoothing techniques in statistics, *Internat. Statist. Rev.*, **53**, 141–170.

Wahba, G. (1975). Smoothing noisy data with spline functions, *Numer. Math.*, **24**, 383–393.

Wahba, G. (1990). *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia.

Wahba, G. and Wold, S. (1975). A completely automatic French curve—Fitting spline functions by cross validation—, *Comm. Statist.*, **4**, 1–17.