J. Geomag. Geoelectr., **43**, 957–978, 1991

Method to Subtract an Effect of the Geocorona EUV Radiation from the Low Energy Particle (LEP) Data by the Akebono (EXOS-D) Satellite

T. HIGUCHI*

The Institute of Statistical Mathematics, Tokyo 106, Japan

(Received March 16, 1991; Revised September 3, 1991)

It was found that the LEP data was sometimes contaminated with unexpectedly larger noise mainly due to the geocorona EUV radiation. Its noise level is so high that we are obliged to subtract the background noise in ground processing when we intend to analyze such LEP data. We therefore propose an efficient and useful procedure for eliminating this background noise, which is highly implemented so as to reduce the computing time. This method is based on a Bayesian smoothness prior approach with a state space modeling. The estimated background noise component is easily realized as a fixed-interval smoothed value by using the recursive Kalman filter and smoother algorithm. In addition, since we use an objective criterion ABIC to choose the best model for the LEP data with the background noise, then the background noise component is automatically and objectively (not *ad-hoc*) subtracted according to the characteristics of noise such as the signal-to-noise ratio, its intensity, and various nonstationarity. A detail description of the whole procedure based on a Bayesian approach can be shown.

1. Introduction

The Japanese scientific satellite designated as EXOS-D (renamed Akebono after launch) was successfully put into a semipolar, elliptical orbit to study the physical mechanism of auroral particle acceleration. A general description of this satellite and its onboard instruments can be found in the special issue for Akebono (EXOS-D) satellite observations (OYA, 1990). In this paper we discuss data from the Low Energy Particle (LEP) spectra analyzer. This instrument is designed to increase our understanding of physical acceleration process. We refer the reader to MUKAI *et al.* (1990) for details on the LEP instrumentation. Preliminary reports from the initial period of operation can also be found therein.

The measured raw LEP data, which is a count number of charged particle, is sometimes contaminated with increase in the background noise level (MUKAI *et al.*, 1990). This background noise can be attributed mainly to the geocorona EUV radiation (MUKAI *et al.*, 1990). The noise level is so high that we must subtract it in ground processing prior to analyzing the LEP data. Great effort was made to remove this background noise through conventional digital signal processing, however it was

* Present address: Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA, 90024-1567, U.S.A.

unsuccessful. The processed data still contained significant noise contamination. The failure of conventional digital signal processing to remove the noise contamination may stem from its inability to consider time-varying phenomena (personal communication with Mukai). Since the component attributed to the background noise strongly depends on time, it seems natural that we can not substantially subtract the background noise with a conventional data analysis method. Moreover, whenever we decompose the observed data into signal and noise components, the choice of an unknown signal-to-noise (S/N) ratio, which we in general provide *ad-hoc*, becomes a significant problem in conventional processing.

AKAIKE (1980) developed an effective and useful analysis method for nonstationary time series. His method was based on a Bayesian smoothness priors approach which is generally characterized by adding a smoothness constraint on the time series model parameters (AKAIKE, 1980; SILVERMAN, 1985; TITTERINGTON, 1985). The example shown below illustrates the Bayesian approach to a very simple problem to illustrate the basic principles of the approach. We consider in the time domain the model

$$y(n) = t(n) + e(n) \quad (n = 1, 2, \ldots, N), \tag{1.1}$$

where $t(n)$ is a trend (drifting mean value function) and $e(n)$ corresponds to observational errors. In this model, we treat the trend component $t(n)$ $(n = 1, \ldots, N)$ as a parameter vector and specify it θ. We assume the $t(n)$ is a *smooth curve*, and then we add one of the *smoothness constrains* to θ, expressed by minimizing the sum of the square of the distance between successive points in time

$$\sum_n \Big(t(n+1) - t(n)\Big)^2. \tag{1.2}$$

By this formulation, we explicitly incorporate the analyst's expectation (in this case, that the time series shows a smooth behavior). This assumption is therefore called *prior information.* Of course, any other prior information can be added to parameters.

We furthermore assume that the trend component $t(n)$ does not deviate far from the observations, $y(n)$. This assumption under smoothness constraint can be obtained by minimizing

$$E(\theta) = \sum_n \Big(y(n) - t(n)\Big)^2 + \frac{1}{\tau^2} \sum_n \Big(t(n+1) - t(n)\Big)^2 \tag{1.3}$$

where τ is a *tradeoff parameter* which controls the tradeoff between the goodness-of-fit of the estimated trend component $t(n)$ to the observations $y(n)$ (expressed by (1.1)) and the goodness-of-fit of $t(n)$ to a smoothness constraint (expressed by (1.2)) (AKAIKE, 1980; GERSCH and KITAGAWA, 1988). The estimation of $t(n)$ is reduced to the well-known *constrained least squares problem* and solution is uniquely determined, only if the tradeoff parameter τ is given (AKAIKE, 1980; TITTERINGTON, 1985).

This trade-off parameter is called in Bayesian terminology *hyperparameter* (LINDLEY and SMITH, 1972). The hyperparameter is sometimes called a *smoothing parameter*, because it controls the smoothness (or stability) of the estimated $t(n)$ (WAHBA, 1980; SILVERMAN, 1985; TITTERINGTON, 1985). By this reason, it also

called a *regularization parameter* in the context of image reconstruction (TIKHONOV and ARSENIN, 1977; MARROQUIN *et al.*, 1987). The choice of hyperparameter essentially corresponds to a selection of *cutoff frequency* when we apply the linear lowpass filter to remove the higher frequency components from the observations (SILVERMAN, 1985; HIGUCHI, 1991). In actual application of the lowpass filter, we often provide this cutoff frequency *subjectively* and *ad-hoc*, according to the analyst's tastes.

A key point in the Bayesian approach is to determine an optimal value of hyperparameter τ *objectively* and *automatically*. The constrained least squares problem has a clear Bayesian interpretation which facilitates the determination of the hyperparameters (LINDLEY and SMITH, 1972; KITAGAWA and GERSCH, 1985a, b; SILVERMAN, 1985; TITTERINGTON, 1985; GERSCH and KITAGAWA, 1988). We multiply (1.3) by $-1/2\sigma^2$ and exponentiate it. Then the minimization of (1.3) with respect to θ is equivalent to the maximization of

$$\begin{aligned} f(\theta) &= \exp(-\frac{E(\theta)}{2\sigma^2}) \\ &= \exp\Big\{-\frac{1}{2\sigma^2}\sum_n\Big(y(n)-t(n)\Big)^2\Big\}\cdot\exp\Big\{-\frac{1}{2\sigma^2\tau^2}\sum_n\Big(t(n+1)-t(n)\Big)^2\Big\} \end{aligned} \tag{1.4}$$

for the fixed values of σ^2 and τ^2. Here, for simplicity, we specify the observation vector $y(n)$ $(n=1,\ldots,N)$ by Y. The θ that maximizes (1.4) under the fixed values of σ^2 and τ^2 also maximizes the *posterior distribution* of θ, $p(\theta|Y,\sigma^2,\tau^2)$, which is proportional to the product of the *data distribution*

$$p(Y|\theta,\sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^N}\exp\Big\{-\frac{1}{2\sigma^2}\sum_n\Big(y(n)-t(n)\Big)^2\Big\} \tag{1.5}$$

and a *prior distribution*

$$p(\theta|\sigma^2,\tau^2) = \frac{1}{C}\exp\Big\{-\frac{1}{2\sigma^2\tau^2}\sum_n\Big(t(n+1)-t(n)\Big)^2\Big\}, \tag{1.6}$$

because the posterior distribution of θ can be rewritten as

$$p(\theta|Y,\sigma^2,\tau^2) = \frac{p(Y,\theta|\sigma^2,\tau^2)}{p(Y|\sigma^2,\tau^2)} = \frac{p(Y|\theta,\sigma^2)p(\theta|\sigma^2,\tau^2)}{p(Y|\sigma^2,\tau^2)} \propto p(Y|\theta,\sigma^2)p(\theta|\sigma^2,\tau^2) \propto f(\theta). \tag{1.7}$$

Of course, $p(A|B)$ means the probability of an occurrence of A on the condition that B is given. It should be noticed that C in (1.6) is a normalization factor required for $\int p(\theta|\sigma^2,\tau^2)d\theta = 1$ and that $p(Y|\sigma^2,\tau^2)$ in (1.7) does not depend on θ obviously. By using (1.3)–(1.7), it can be shown that the minimization of (1.3) over θ is in the Bayesian context equivalent to the maximization of the posterior distribution of θ. In addition, the constrained least squares approach implies that the observations Y and parameter vector θ are assumed to obey the multi-normal (*-Gaussian*) distributions.

In a Bayesian framework, the θ obtained by the maximization of the posterior distribution is called the maximum *a posteriori* (MAP) solution, being the mode of

its posterior distribution (GEMAN and GEMAN, 1984; TITTERINGTON, 1985; BESAG, 1986; MARROQUIN *et al.*, 1987). We call in general the data distribution $p(Y|\cdot)$, *likelihood.* Thus it is easily understand that the constrained least squares approach to maximize (1.3) can be interpreted as one of the *maximum penalized likelihood* (MPL) method (GOOD and GASKIN, 1971; LEONARD, 1978), because (1.5) and (1.6) are a likelihood of Y and a penalty function, respectively, and thereby the posterior distribution also corresponds to the penalized likelihood function. In the MPL method, the hyperparameter is called a *roughness penalty.*

The integration of $p(Y, \theta|\sigma^2, \tau^2)$ over θ yields the likelihood of Y (called the *marginal likelihood*) for the unknown σ^2 and τ^2,

$$L(Y|\sigma^2, \tau^2) = \int p(\theta, Y|\sigma^2, \tau^2)d\theta = \int p(Y|\theta, \sigma^2)p(\theta|\sigma^2, \tau^2)d\theta. \tag{1.8}$$

We select the values σ^2 and τ^2 which maximize the logarithm of (1.8), $\log L(Y|\sigma^2, \tau^2)$. This concept was first introduced by GOOD (1965) and called *Type II likelihood method.* AKAIKE (1980) has explicitly shown the closed form of this marginal likelihood when both the data and prior distributions, $p(Y|\theta, \sigma^2)$ and $p(\theta|\sigma^2, \tau^2)$, are normally distributed. In place of $\log L(\cdot)$, we generally use

$$ABIC = -2\log L(Y|\sigma^2, \tau^2) \tag{1.9}$$

after Akaike, where $ABIC$ represents the *Akaike Bayesian Information Criterion.* In our smoothing problem, we choose the values of hyperparameter which minimize $ABIC$ and determine the trend component $t(n)$ for fixed values of hyperparameters.

AKAIKE (1980) has shown several practical results of its application to nonstationary time series modeling. Since the number of parameter in a Bayesian approach to nonstationary time series generally exceeds that of observations, such time series model is satisfactorily flexible to the observations. Along the lines suggested by Akaike, Akaike's 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; HIGUCHI *et al.*, 1988; KITA *et al.*, 1989). In particular, Kitagawa and Gersch have developed the Bayesian modeling to a variety of problems in nonstationary time series analysis (KITAGAWA, 1981, 1983, 1987, 1988, 1989a, b; KITAGAWA and GERSCH, 1984, 1985a, 1985b; KITAGAWA and TAKANAMI, 1985; GERSCH and KITAGAWA, 1988). It should be noticed that the modeling must be tailored to each application. An efficient algorithm for calculating ABIC is extensively explored so as to reduce a computing time and memory (KITAGAWA, 1981, 1983; ISHIGURO, 1984; GERSCH and KITAGAWA, 1988).

The noise reduction technique based on the Bayesian approach has been also applied to optical data taken aboard a spacecraft, which suffers from an unexpected modulations synchronized with the rotation or wobble of the spacecraft (HIGUCHI *et al.*, 1988; KITA *et al.*, 1989). In observations on board a spacecraft, it is very important and inevitable to subtract an periodic noise associated with the spin and/or precession of the spacecraft, because the objective quantity through observations is usually estimated by using inversion technique which is quite sensitive to noise (TANABE, 1976;

HIGUCHI *et al.*, 1988). By using ABIC, we can objectively determine the unknown signal and noise (S/N) ratio and eliminate the periodic noise from the observed data satisfactorily (HIGUCHI *et al.*, 1988; KITA *et al.*, 1989). In that model, the periodic noise was assumed to behave locally as a sinusoidal wave, and thus its amplitude was able to be estimated as a function of time. Obviously, except for the sinusoid, there occurs the periodic noise which shows in time domain a square, triangle with definite period. The Bayesian model for a seasonal adjustment is applicable and suitable for this case (ISHIGURO, 1984; KITAGAWA and GERSCH, 1984; GERSCH and KITAGAWA, 1988). The method for reduction of the complicated background noise has been demonstrated on a basis of the Bayesian approach (KITAGAWA and TAKANAMI, 1985). These Bayesian approach is characterized by involving the nonstationarity of the noise.

In this paper, we will propose a useful and efficient method to subtract the background noise observed in the LEP data, along the procedure proposed by KITAGAWA (1981, 1983). A detail explanation of our method will be shown together with its relation to the Bayesian approach along the Akaike's formula. A special modeling for the LEP data highly contaminated with the background noise will be presented together with its application to actual data. Some comments on practical application will be also summarized to help the reader understand our procedure of removing the background noise.

2. Data

For seeking of a spatial distribution of the particles as a function of time, Akebono carries ten channeltorons with their configuration as seen in Fig. 2 of MUKAI *et al.* (1990). Since we can not identify the channeltron only by its channel number, we hereafter specify Channel J of the LEP−S2 by Channel $J+5$; i.e., Channel 5 of the LEP−S2 is indicated by Channel 10. Hence, the channeltron with the channel number less than 5 (larger than 6) corresponds to one of the LEP−S1 (LEP−S2) hereafter. All channeltron is installed in the same plane which contains the spin axis. Accordingly all view direction (angle) of channeltron in a plane perpendicular to the spin axis is coherent at any time. Both the count number for the ion and that for the electron are simultaneously measured by each channeltron. Then the LEP data is a multivariate time series which consists of twenty (2×10) components.

An *E-t* (Energy-time) diagram enables us to easily examine the energy spectrum of the charged particle as a function of time. In order to make the energy spectrum, the energy range covered by the channeltron is divided into 64 equally spaced steps on a logarithmic scale and the count number is measured at each step. The energy step is labeled the integral number between 0 and 63. Of course, many scanning patterns can be considered, and the standard types among them are used in actual operation (MUKAI *et al.*, 1990). This energy scanning pattern is common to all channeltron. Since there may be the energy dispersion, we collect the data with the same energy step number from the original single time series, for each channeltron and each species. So that, the number of newly constructed time series is sixty four for each channeltron and each species. We specify the time series obtained like this by ${y_i}^{j,k}(n)$ $(n = 1, 2, \ldots, N)$, where i and j denote the energy step and channeltron numbers, respectively, and k represents the species of the charged particles (for the electron and ion, $k = 1$ and

$k = 2$, respectively). Then, obviously, $y_0^{10,2}(n)$ is composed of the ion data with energy step 0 and channeltron 10.

The sampling time of this newly constructed data $y_i^{j,k}(n)$, i.e., $\Delta t(n)$, defined by an interval of an observation time between $y_i^{j,k}(n)$ and that of $y_i^{j,k}(n-1)$, seems to be basically identical to the interval time required for fully energy scanning (indicated by T_E hereafter), but in fact depends on the the energy scanning pattern. In particular, for the scanning pattern as shown in Fig. 1 (on April 14, 89), the interval time also depends on n; $\Delta t(n)$ for n of the even number is different from that for the odd number. This is very inconvenient for the subsequent procedures, then in this case we distinguish between $y_i^{j,k}(n)$ with the even n (denoted by $\times$ in figure) and that with the odd n (denoted by $\circ$). Namely we furthermore generate two time series ($\times$ and $\circ$) from a single time series $y_i^{j,k}(n)$. In this study, these newly constructed time series are distinguished by a subscript: $y_{i(1)}^{j,k}(n)$ and $y_{i(2)}^{j,k}(n)$, respectively. Then the sampling time of the newly constructed time series comes to be constant with $\Delta t = 2T_E$ accordingly. In a case shown in Fig. 1, the sampling time Δt is given by $2T_E = 7.75$ second. As seen in this figure, the number of energy steps used on April 14, 1989 is thirty two. In short, no count number with an odd energy step number is measured during this interval shown in Fig. 1. It should be noticed in Fig. 1 that this special treatment for the sampling time is not necessary for the time series with the energy step number of 0 and 62. Then, in this case, sixty two $y_{i(l)}^{j,k}$ $(62 = 2 + (32 - 2) \times 2)$ are newly generated for each channeltron and each species. Consequently, since the count numbers both for electrons and for ions are simultaneously obtained by each channeltron, the number of the time series $y_{i(l)}^{j,k}$ is $2 \times 10 \times 62 = 1240$.

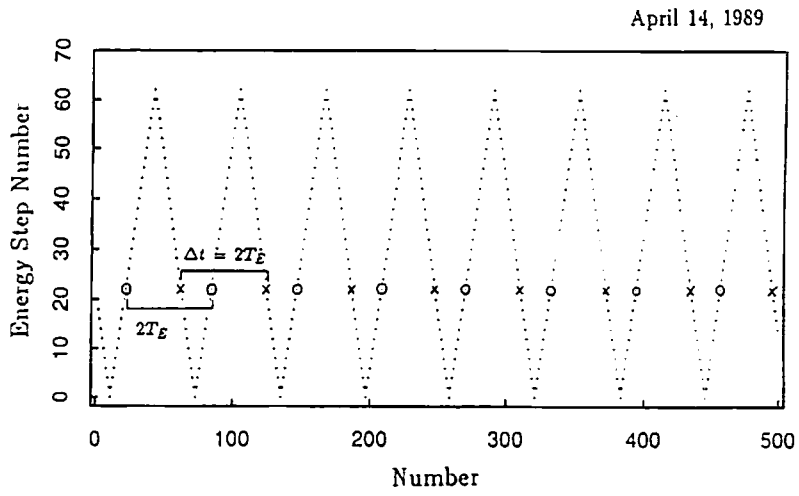

Fig. 1. The energy scanning pattern on April 14, 1989. The vertical axis denotes the energy step number.

We show in Fig. 2(a) the newly generated time series, $y_0^{10,2}(n)$, observed on April 14, 1989. Since the energy scanning pattern at this period is the same one as shown in Fig. 1, it is evident that there exists only one time series with the energy step number

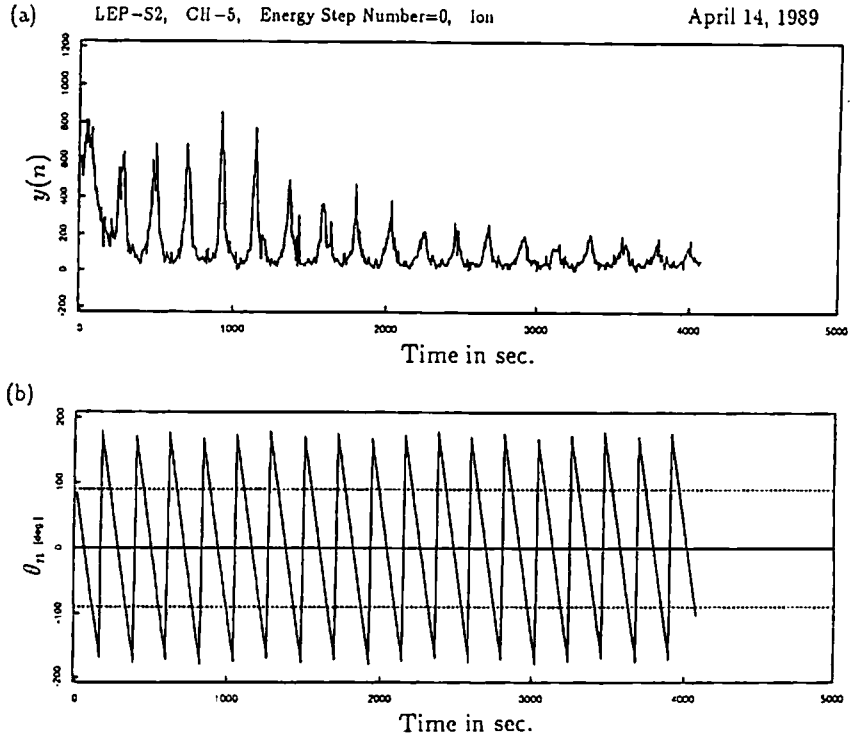

Fig. 2. (a) The ion count number obtained by the channeltron 10 (CH-5 of LEP-S2) for the energy step 0, on April 14, 1989. (b) The view direction (phase) in a plane perpendicular to the spin axis, $\theta_0^{10,2}(n)$, for given values of $T_{0,obs}^{10,2} = -220.4$ [sec.] and $\theta_0^{10,2}(0) = 87.43$ [deg.].

0 for each species. As previously defined, this time series is the ion count number for the channeltron number 10 and the energy step number 0. The background noise, which is attributed mainly to an effect of the geocorona, can be seen in this figure as a cyclically repeating pattern with period of approximately 220 second. The spin period T_s at this interval shown in figure is set to be approximately $T_s = 8.014$ second. Because of a constant difference between $2T_E = \Delta t$ and T_s, a view direction in a plane perpendicular to the spin axis changes cyclically as a function of time. We denote hereafter this view direction, i.e., phase in a plane, by $\theta_0^{10,2}(n)$ and set $\theta_0^{10,2}(n) = 0$ at the time when the channeltron points just toward the background noise source. We here take the satellite's spin direction as positive orientation in a phase $\theta_0^{10,2}(n)$. Of course, the superscripts 10 and 2 indicate the channeltron number and ion, respectively. The subscript 0 is required to specifically represent the phase for the energy step number 0, because, for example, an observation time of $y_0^{10,2}(1)$ is obviously different from that of $y_1^{10,2}(1)$, as seen in Fig. 1. The period of the changing view direction, T_d, is given by

$$T_d = \frac{2T_E T_s}{2T_E - T_s}, \tag{2.1}$$

where a sign of T_d means the phase shift of the view direction relative to a direction of satellite's rotation. Since in our case shown in Fig. 2(a) $2T_E$ is slightly smaller than T_s and a sign of T_d is negative, $|T_d|$ is much larger than T_s ($|T_d| \sim 235$ second). So that $\theta_0^{10,2}(n)$ gradually changes in a negative direction; in other words, a view direction rotates in a direction opposite to the satellite's spin.

The period estimated by a visual inspection is approximately 220 second and definitely shorter than $|T_d|$. This discrepancy stems from a gradual change in the satellite position relative to the background noise source during the interval shown in Fig. 2(a). Namely, even if the satellite orients its view direction just toward the background noise source at $t = t_0$, the view direction at $t = |T_d| + t_0$ is different from that at $t = t_0$; i.e., $\theta_0^{10,2}(t = t_0) \neq \theta_0^{10,2}(t = |T_d| + t_0)$.

The position of the background noise source is previously unknown, and then an effect of the change in the satellite position on the phase $\theta_0^{10,2}(n)$ can not be *a-priori* given. Hence, we describe $\theta_0^{10,2}(n)$ simply as a function of $T_{0,obs}^{10,2}$ and $\theta_0^{10,2}(0)$ as follows:

$$\theta_0^{10,2}(n) = Mod\ (\theta_0^{10,2}(0) + \frac{2\pi n \Delta t}{T_{0,obs}^{10,2}}, 2\pi), \tag{2.2}$$

where $\theta_0^{10,2}(0)$ is, of course, an initial phase for a time series $y_0^{10,2}$. $|T_{0,obs}^{10,2}|$, of course, corresponds to the period previously estimated by the visual inspection into Fig. 2(a). A sign of $T_{0,obs}^{10,2}$ has also the same meaning as that of T_d. Although $T_{0,obs}^{10,2}$ and $\theta_0^{10,2}(0)$ can be roughly given by checking the intervals between sharp peaks seen in Fig. 2(a), we treat them as parameters in the subsequent procedures and search an optimal value for them. In order to show a good performance of the model for the phase $\theta_0^{10,2}(n)$, which is defined by (2.2), we show in Fig. 2(b) $\theta_0^{10,2}(n)$ for $T_{0,obs}^{10,2} = -220.4$[sec.] and $\theta_0^{10,2}(0) = 87.43$[deg.].

3. Model for Decomposition

Hereafter we consider a time series for the fixed channeltron, energy step number, and species (electron or ion) and subtract the background noise component from it. Then the sub- and super-scripts to specify the channeltron, energy step number, and the species are unnecessary for the following descriptions. In short, a single time series data $y(n)$ represents a certain $y_{i(l)}^{j,k}$. Accordingly, $\theta_{i(l)}^{j,k}(n)$ and parameters for it, $T_{i(l),obs}^{j,k}$ and $\theta_{i(l)}^{j,k}(0)$ are simply indicated by $\theta(n)$, T_{obs}, and θ_0, respectively.

3.1 The basic model for decomposition

3.1.1. Observation model

To subtract the background noise from an observed time series $y(n)$ $(n = 1, \ldots, N)$, we will here consider the model

$$y(n) = f(n)I(n) + t(n) + e(n), \tag{3.1}$$

where the first term in (3.1), $f(n)I(n)$, is the background noise component, $t(n)$ is a trend, and $e(n)$ is an observational noise which is a white noise sequence with $e(n) \sim N(0, \sigma^2)$. Of course, $N(0, \sigma^2)$ is a Gaussian distribution with a mean of 0 and unknown variance of σ^2. The assumption of Gaussian distribution implies that we intend to apply the familiar least squares fit which minimizes $\sum_{n=1}^{N} e(n)^2$. In other words, a minimization of

$$\sum_{n=1}^{N} \Big(y(n) - f(n)I(n) - t(n) \Big)^2 \tag{3.2}$$

can be justified by the maximum likelihood method under the assumption that a deviation from the observed data, $e(n) = y(n) - f(n)I(n) - t(n)$, obeys the Gaussian distribution. It should be noticed that we need not give a value of σ^2 and that it is naturally defined by the variance of the residual sequence $e(n)$ which is derived through estimating $f(n)$, $I(n)$, and $t(n)$ (AKAIKE, 1980; KITAGAWA, 1983). It is in detail explained hereafter.

As seen in (3.1), the background noise component is expressed by a product of $f(n)$ and $I(n)$, where $f(n)$ is an arbitrary periodic function which satisfies $f(n) = 0$ for $90° < |\theta(n)|$, and $I(n)$ is a time varying intensity of the noise source. The representation like this for the background noise component implies that $f(n)t(n)$ shows a cyclically repeating pattern with period T_{obs} and that the form of cyclic pattern changes gradually as well as the power of noise. The form of cyclic pattern moderately depends on $f(n)$, so that we select it without special care. In our case, any function can be a good candidate for $f(n)$ if it roughly approximates a cyclic pattern of the component which is probably attributed to the background noise as seen in Fig. 2(a). For simplicity, we therefore select the following function as $f(n)$:

$$f(n) = \begin{cases} \cos\theta(n) & |\theta(n)| < 90° \\ 0 & 90° \leq |\theta(n)| \leq 180°, \end{cases} \tag{3.3}$$

where the phase $\theta(n)$ is given by (2.2). Although we can take an alternative function as $f(n)$, such as $\cos^2\theta(n)$, $\exp(-\sin^2\theta(n))$, or quadratic curve, a choice is well justified by using the minimum AIC procedure that was developed as a natural extension to the maximum likelihood method (AKAIKE, 1973; SAKAMOTO *et al.*, 1986). Here we remark that a computational time for estimating the background noise slightly depends on the function form for $f(n)$. We can therefore search the best function form which minimize AIC, without wasting more computational time. For the first step, we apply for simplicity $\cos\theta(n)$ and $\cos^2\theta(n)$, and the model with the smaller AIC is chosen as $f(n)$. We can consider a piecewise linear expression for an alternative approach to give a form of $f(n)$. This approach might be flexible for describing the form of $f(n)$, but unfortunately requires lots of computational time.

3.1.2. System model

We assume that the noise intensity and trend components, $I(n)$ and $t(n)$ change gradually with time. Namely we consider that $I(n)$ ($t(n)$) and $I(n-1)$ ($t(n-1)$) are approximately equal at each n. This assumption can be represented by stochastically perturbed difference equation constraints

$$\begin{aligned} \nabla^{k_I} I(n) &= u_I(n), \\ \nabla^{k_t} t(n) &= u_t(n), \end{aligned} \tag{3.4}$$

where k_I and k_t denote an order of difference for $I(n)$ and $t(n)$, respectively. In (3.4), $u_I(n)$ and $u_t(n)$ are white noise sequences such that $u_I(n) \sim N(0, \tau_I^2(n)\sigma^2)$ and $u_t(n) \sim N(0, \tau_t^2\sigma^2)$. σ^2 is, as hereinbefore defined, an unknown variance of observation noise $e(n)$, and $\tau_I^2(n)$ and τ_t^2 give a ratio of each unknown variance to σ^2 accordingly. For a simple case of $k_I = k_t = 1$, these constraints of (3.4) imply that we intend to minimize

$$\sum_n \frac{\Big(t(n) - t(n-1)\Big)^2}{\tau_t^2} + \sum_n \frac{\Big(I(n) - I(n-1)\Big)^2}{\tau_I^2(n)}. \tag{3.5}$$

It should be here remarked that a variance of $u_I(n)$ is a time varying one and thereby $(I(n) - I(n-1))^2$ is weighted and summed up at each n. We here remind the reader that σ^2 is automatically defined through the residual sequence $e(n)$. In short, a value of σ^2 is unnecessary for estimating $I(n)$ and $t(n)$. On the other hand, the values of $\tau_I^2(n)$ $(n = 1, \ldots, N)$ and τ_t^2 are intrinsically required.

To definitely specify a basic model for decomposition, it remains to determine an order of difference, k_I and k_t. In this study, we take $k_I = 1$ and $k_t = 1$ because a first order difference is convenient for subsequent calculations. This simplification never reduce a goodness-of-fit our of model to data, and can be also justified by the aforementioned minimum AIC procedure among various selections for (k_I, k_t).

3.1.3. Constrained least squares approach

The constraints for $I(n)$ and $t(n)$ are given as expressed in (3.2) and (3.5), and thereby a solution for estimating $I(n)$ and $t(n)$ is reduced to be the constrained least squares problem in such a way that we minimize the quantity

$$\begin{aligned}\chi^2(a_{2N}) &= \sum_{i=1}^{N}\Big(y(n) - f(n)I(n) - t(n)\Big)^2 + \alpha^2\Big(\frac{I^2(1)}{\tau_I^2(1)} + \frac{t^2(1)}{\tau_t^2}\Big) \\ &\quad + \sum_{i=2}^{N}\frac{\Big(I(n) - I(n-1)\Big)^2}{\tau_I^2(n)} + \sum_{i=2}^{N}\frac{\Big(t(n) - t(n-1)\Big)^2}{\tau_t^2} \\ &= \sum_{i=1}^{N} e(n)^2 + \alpha^2\Big(\frac{I^2(1)}{\tau_I^2(1)} + \frac{t^2(1)}{\tau_t^2}\Big) + \sum_{i=2}^{N}\frac{u_I^2(n)}{\tau_I^2(n)} + \sum_{i=2}^{N}\frac{u_t^2(n)}{\tau_t^2},\end{aligned} \tag{3.6}$$

over a_{2N}, where a_{2N} represents the parameter vector defined by $a_{2N} = [I(1), I(2), \ldots, I(N), t(1), t(2), \ldots, t(N)]$, and α is constant for special treatments of $I(1)$ and $t(1)$. It should be noticed that this problem is also a weighted least squares problem, because each $u_I^2(n) = (I(n) - I(n-1))^2$ has its own weight, $1/\tau_I^2(n)$. This quantity $\chi^2(a_{2N})$ is also called "chi-square" in terms of the Chi-Square fitting (PRESS *et al.*, 1988). If τ_t^2, $\tau_I^2(n)$, $f(n)$ $(n = 1, \ldots, N)$, and α are given along with the observation $y(n)$ $(n = 1, \ldots, N)$, the solution of minimizing (3.6) is uniquely determined and takes a simple form referred in AKAIKE (1980), KITAGAWA and GERSCH (1985b), HIGUCHI *et al.* (1988), and HIGUCHI (1991). In other words, the solution to minimize (3.6) is determined by giving values of parameters

$$\Lambda = [\tau_t^2, \tau_I^2(1), \tau_I^2(2), \ldots, \tau_I^2(N), T_{obs}, \theta_0] \tag{3.7}$$

and supplemental parameter α.

The tradeoff parameters such as $\tau_t^2, \tau_I^2(n)$ $(n = 1, \ldots, N)$ balance the tradeoff between infidelity to data, which is expressed by the minimization of (3.2), and infidelity to the constraints given by minimizing (3.5), and thus it is obvious that the critical idea in the constrained least squares approach is a proper selection of the tradeoff parameters τ_t^2 and $\tau_I^2(n)$ $(n = 1, \ldots, N)$.

The constrained least squares problem has a clear Bayesian interpretation which facilitates the determination of the tradeoff parameters (AKAIKE, 1980; KITAGAWA and GERSCH, 1984, 1985b; TITTERINGTON, 1985; GERSCH and KITAGAWA, 1988). In detail, the reader should be referred to KITAGAWA and GERSCH (1985b) and

TITTERINGTON (1985). These tradeoff parameters are called as a hyperparameter in a Bayesian framework (LINDLEY and SMITH, 1972; TITTERINGTON, 1985; GERSCH and KITAGAWA, 1988). A performance of the hyperparameters on the estimated parameters can be visually shown in AKAIKE (1980), INOUE (1985), TANABE (1985), and HIGUCHI *et al.* (1988) for a simple case. HIGUCHI (1991) has examined the frequency domain characteristics of the various Bayesian approach parametrically as a function of the hyperparameters. In particular, the relationship between the hyperparameter and the halfband width in the lowpass filter has been shown when we adopt the smoothness prior model expressed by (3.4) (SILVERMAN, 1985; HIGUCHI, 1991).

3.1.4. Representation by the state space model

The Bayesian approach along the lines of AKAIKE (1980) has computational complexity $O(N^3)$. Since the constraints between the estimated parameters are usually local, as expressed in (3.4), in the Bayesian approach to model a time series, we can reduce this complexity to be $O(N)$ by the efficient scheme (ISHIGURO, 1984) and also by the computationally efficient recursive algorithm based on the Kalman filtering (KITAGAWA, 1981). The formulation by AKAIKE (1980), i.e., the constrained least squares computational approach by using smoothness prior models (LINDLEY and SMITH, 1972; O'HAGAN, 1976; AKAIKE, 1980; SILVERMAN, 1985; KITAGAWA and GERSCH, 1985b), can be rewritten by the state space model (SSM) approach for the linear model with Gaussian system and observation noises (GERSCH and KITAGAWA, 1988; KITAGAWA, 1989b). The SSM representation is very flexible enough to rewrite any constrained least squares approach only if the relationships between the estimated parameters are linear and local. Several examples for representations by the SSM enable us to understand the equivalence between the SSM and constrained least squares approaches (KITAGAWA, 1981, 1989b; KITAGAWA and GERSCH, 1984). The parameters determined by the constrained least squares approach can be in the SSM approach achieved by the Kalman smoother solution. The reader who is familiar with the constrained least squares approach along Akaike's formulation should refer to KITAGAWA (1981) and KITAGAWA and GERSCH (1985b) in order to comprehend the SSM approach, which is therein related to the constrained least squares approach.

We also rewrite the constrained least squares approach by the SSM approach. For fixed values of T_{obs} and θ_0, $f(n)$ is definitely determined, and then our model of (3.1) and (3.4) can be rewritten in the state space form

$$\begin{aligned} z(n) &= Fz(n-1) + Gu(n), \\ y(n) &= H(n)z(n) + e(n) \end{aligned} \tag{3.8}$$

with the state vector $z(n)$, and F, G, $H(n)$ matrices given by

$$z(n) = \begin{pmatrix} I(n) \\ t(n) \end{pmatrix}, \quad F = G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad H(n) = \begin{pmatrix} f(n) & 1 \end{pmatrix}. \tag{3.9}$$

The vector $u(n)$ is an uncorrelated sequence with

$$u(n) = \begin{pmatrix} u_I(n) \\ u_t(n) \end{pmatrix}, \tag{3.10}$$

where $u_I(n)$ and $u_t(n)$ are, of course, the identical in (3.4).

3.1.5. Kalman filter

Given the state space representation, the estimates of $I(n)$ and $t(n)$ are easily realized as the fixed-interval smoothed values by using the recursive Kalman filter and smoother algorithms (ANDERSON and MOORE, 1979; KATAYAMA, 1983; KITAGAWA, 1989b) for the fixed Λ (see, for detail descriptions of the procedure presented hereunder, KITAGAWA (1981, 1983) and KITAGAWA and GERSCH (1984, 1985a)). Since the system and observation noises, $u(n)$ and $e(n)$, are Gaussian in this study, any conditional probability density of $z(n)$, given observations $Y_{n'} = [y(1), y(2), \ldots, y(n')]$, $p(z(n)|Y_{n'})$, can be characterized by the mean vector and the covariance matrix. Here we denote the mean vector and covariance matrix, given $Y_{n'}$, by $z(n|n')$ and $V(n|n')$, respectively. If we give an initial mean vector and covariance matrix, $z(0|0)$ and $V(0|0)$, which represent the distribution of an initial state vector $z(0)$, we can define $z(1|0)$ and $V(1|0)$ through a one-ahead prediction formula. Moreover, the filtering formula gives us $z(1|1)$ and $V(1|1)$, from $z(1|0)$ and $V(1|0)$. This recursive algorithm is as follows.

For one-step-ahead prediction (time update),

$$z(n|n-1) = Fz(n-1|n-1), \qquad V(n|n-1) = FV(n-1|n-1)F^t + GQ(n)G^t, \tag{3.11}$$

where $Q(n)$ is a matrix defined by

$$Q(n) = \begin{pmatrix} \tau_I^2(n) & 0 \\ 0 & \tau_t^2 \end{pmatrix}. \tag{3.12}$$

Of course, t denotes the transposition. For filtering (observation update),

$$\begin{aligned} K(n) &= V(n|n-1)H^t(n)/[H(n)V(n|n-1)H^t(n) + 1], \\ z(n|n) &= z(n|n-1) + K(n)\Big(y(n) - H(n)z(n|n-1)\Big), \\ V(n|n) &= [I - K(n)H(n)]V(n|n-1), \end{aligned} \tag{3.13}$$

where I is an identical matrix. According to (3.11) and (3.13), we can estimate $z(n|n-1)$, $z(n|n)$, $V(n|n-1)$, and $V(n|n)$ at each n, if only $z(0|0)$ and $V(0|0)$ are given. These recursive formula have been schematically shown in KITAGAWA (1989b). Several ideas for giving $z(0|0)$ and $V(0|0)$ can be referred to KITAGAWA (1981) and KITAGAWA and GERSCH (1984). In our case, we set simply

$$z(0|0) = \begin{pmatrix} \beta \\ 0 \end{pmatrix} \tag{3.14}$$

and

$$V(0|0) = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}, \tag{3.15}$$

where β is a constant defined by an average of $y(n)/f(n)$ with a phase of $-80° \leq \theta(n) \leq 80°$ and c is a large value (for example, $c = 10^6$). The value of β approximately represents mean value of $I(n)$, because $y(n)$ within this phase range of $-80° \leq \theta(n) \leq 80°$ seems to be attributed mainly to the background noise and thereby $y(n)/f(n)$ gives us a rough estimate of $I(n)$ through (3.1). Taking a large value for c is equivalent to

setting a small value for α in (3.6). In short, a special treatment in the constrained least squares approach corresponds to setting $z(0|0)$ and $V(0|0)$ in a framework of the recursive Kalman filtering approach. This can be also easily interpreted in terms of the boundary condition in the 1 dimension space. In the time series analysis, an effect from the initial data decays rapidly and this problem is not so significant. Thus far, a choice of values of β and c slightly affects an estimation of $I(n)$ and $t(n)$.

In the approach along AKAIKE's formula (1980), values $I(n)$ and $t(n)$ are defined through the entire data set as the constrained least squares solution. However, $z(n|n)$ is based on the data $Y_n = [y(1), y(2), \ldots, y(n)]$ $(n \leq N)$, and thus values of the estimated parameters $I(n)$ and $t(n)$ should be defined by using all of the data. Namely we must calculate $z(n|N)$ for $I(n)$ and $t(n)$. Here it should be noticed that $1 \leq n \leq N$. To obtain $z(n|N)$, we can also use a recursive formula for the following backward smoothing. For smoothing,

$$\begin{aligned} A(n) &= V(n|n)F^t V(n+1|n)^{-1}, \\ z(n|N) &= z(n|n) + A(n)\Big(z(n+1|N) - z(n+1|n)\Big), \\ V(n|N) &= V(n|n) + A(n)\Big(V(n+1|N) - V(n+1|n)\Big)A^t(n), \end{aligned} \tag{3.16}$$

where $V(n+1|n)^{-1}$ represents the inverse matrix of $V(n+1|n)$. $z(n|n)$, $z(n|n+1)$, $V(n|n)$, and $V(n|n+1)$ have been already defined at each n by (3.11) and (3.13). We can therefore estimate $z(n|N)$ and $V(n|N)$ at each n when we start the above backward recursive calculation at $n = N-1$.

3.1.6. Minimum ABIC procedure

We use maximum likelihood estimation for choosing the unknown values of σ^2 and Λ. The likelihood $p(Y_N|\sigma^2, \Lambda)$ can be rewritten as follows.

$$\begin{aligned} p(Y_N|\sigma^2,\Lambda) &= p(y(N), y(N-1), \ldots, y(2), y(1)|\sigma^2, \Lambda) \\ &= p(y(N), y(N-1), \ldots, y(2)|y(1), \sigma^2, \Lambda) \cdot p(y(1)|\sigma^2, \Lambda) \\ &= p(y(N), y(N-1), \ldots, y(3)|y(2), y(1), \sigma^2, \Lambda) \cdot p(y(2)|y(1), \sigma^2, \Lambda) \\ &\quad \cdot p(y(1)|\sigma^2, \Lambda) \\ &= \cdots \\ &= p(y(1)|\sigma^2, \Lambda) \prod_{n=2}^{N} p(y(n)|Y_{n-1}, \sigma^2, \Lambda). \end{aligned} \tag{3.17}$$

$p(y(n)|Y_{n-1}, \sigma^2, \Lambda)$ is, of course, the conditional distribution of $y(n)$, given data $Y_{n-1} = [y(1), y(2), \ldots, y(n-1)]$, σ^2, and Λ. Along the aforementioned Kalman filtering procedure, an efficient algorithm for calculating the likelihood of a time series is available (KITAGAWA, 1981, 1983), because $p(y(n)|Y_{n-1}, \sigma^2, \Lambda)$ is approximately given by

$$\begin{aligned} p(y(n)|Y_{n-1}, \sigma^2, \Lambda) &= \int p(y(n)|z(n), \sigma^2) p(z(n)|Y_{n-1}, \sigma^2, \Lambda) dz(n) \\ &= \frac{1}{\sqrt{2\pi\sigma^2 d(n)}} \exp\left(-\frac{\varepsilon^2(n)}{2\sigma^2 d(n)}\right), \end{aligned} \tag{3.18}$$

where $\varepsilon(n) = y(n) - H(n)z(n|n-1)$ and $d(n) = [H(n)V(n|n-1)H^t(n) + 1]$, and they have been already appeared in filtering procedure, as seen in (3.13). The likelihood $p(Y_N|\sigma^2, \Lambda)$ is obviously equivalent to the integrated likelihood (marginal likelihood) which are defined by

$$p(Y_N|\sigma^2, \Lambda) = \int p(Y_N|a_{2N}, \sigma^2)p(a_{2N}|\sigma^2, \Lambda)da_{2N}, \tag{3.19}$$

where a_{2N} has been already defined in (3.6) (AKAIKE, 1980; KITAGAWA and GERSCH, 1985b; HIGUCHI *et al.*, 1988; GERSCH and KITAGAWA, 1988). The minus two times of the logarithm of this likelihood is called ABIC in the constrained least squares approach along AKAIKE's formula (1980). This integration is of computational complexity $O(N^3)$ (ISHIGURO, 1984; KITAGAWA and GERSCH, 1985a). In contrast, a computational complexity of $O(N)$ can be achieved along the algorithm of (3.17) and (3.18). We therefore adopt the state space modeling approach instead of the constrained least squares approach. Here we would like to comment on calculation of the marginal log-likelihood. When the system and observation noises are no longer Gaussian and/or the estimated parameters such as $I(n)$ and $t(n)$ are nonlinearly related, it is necessary to numerically realize an integration of the right hand side of (3.19). We usually treat the Bayesian model which contains many unknown parameters and requires a high dimensional integral accordingly. In fact, it is impossible to directly integrate the right hand side of (3.19). However the approach based on (3.17) enables us to avoid and to mitigate this problem (KITAGAWA, 1987, 1988, 1989a; GERSCH and KITAGAWA, 1988).

As previously mentioned, an effect from the initial data decays rapidly in the time series analysis. However we take account of this effect on the likelihood and adopt the following likelihood

$$p(Y_N|\sigma^2, \Lambda) \simeq \prod_{n=N_s}^{N} p(y(n)|Y_{n-1}, \sigma^2, \Lambda), \tag{3.17'}$$

instead of (3.17), where N_s is an integer of $1 < N_s << N$ (for example, $N_s = 2^4$ for $N = 2^9$). Since the likelihood function takes a form of (3.17′) and (3.18), the best estimate of σ^2 is obviously given by differentiating $\log p(Y_N|\sigma^2, \Lambda)$ with respect to σ^2 and thereby defined by

$$\widehat{\sigma}^2 = \frac{1}{N'} \sum_{n=N_s}^{N} \frac{\varepsilon^2(n)}{d(n)}, \tag{3.20}$$

where $N' = N - N_s + 1$. This means that the estimate of unknown variance σ^2 automatically results from the Kalman filtering algorithm (KITAGAWA, 1981) and that there is no necessity for giving it. The log-likelihood can be obtained from (3.17′), (3.18), and (3.20) as a function of Λ as follows.

$$\log p(Y_N|\Lambda) = -\frac{N'}{2} \log 2\pi\widehat{\sigma}^2 - \frac{N'}{2} - \frac{1}{2} \sum_{n=N_s}^{N} \log d(n). \tag{3.21}$$

The best model is determined so as to minimize the ABIC which is in our study defined by

$$ABIC = -2 \log p(Y_N|\Lambda) \tag{3.22}$$

with respect to Λ. Obviously the likelihood in the state space model corresponds to the marginal likelihood in the constrained least squares approach, so that the minimum ABIC procedure here is equivalent to the Type II maximum likelihood method (GOOD, 1965) and ABIC in (3.22) is essentially the same as ABIC defined by AKAIKE (1980).

3.2 Estimation of time varying variance

The best model can be obtained by minimizing ABIC with respect to Λ, but in general, it is very difficult to select the optimal hyperparameters because ABIC is nonlinear function with many hyperparameters (the number of hyperparameters is $N+3$, where N is of course the number of time series data). Computational difficulties arise from $\tau_I^2(n)$ which is a function of time. Hence we adopt the following procedure feasible for optimizing ABIC.

In the first procedure, we roughly estimate the variance of $u_I(n)$, $\tau_I^2(n)\sigma^2$, by assuming that it appears to be modulated by a relatively slowly changing envelope function. This assumption has been justified through good results of its applications to actual data, in particular to seismic measurements during an earthquake (KITAGAWA and GERSCH, 1985a; KITAGAWA and TAKANAMI, 1985; GERSCH and KITAGAWA, 1988).

As seen in Fig. 2, the observation $y(n)$ for a phase within the limited range of $-\theta_{th} \leq \theta(n) \leq \theta_{th}$ around $\theta = 0°$ seems to be attributed mainly to the background noise component, where θ_{th} is a parameter which can range between 0° and 90°. Of course, although we can optimize ABIC with respect to θ_{th}, we here set $\theta_{th} = 80°$ simply. Accordingly, we assume that $y(n)$ is approximately given by $f(n)I(n)$ within $-80° \leq \theta(n) \leq 80°$. Based on this assumption, (3.4) yields a rough estimate of $u_I(n)$

$$\widehat{u}_I(n) = \nabla\widehat{I}(n) = \frac{y(n)}{f(n)} - \frac{y(n-1)}{f(n-1)}. \tag{3.23}$$

The next step is to estimate the time varying variance of $u_I(n)$. For this, we adopt a procedure to estimate a smoothed value of the instantaneous variance (KITAGAWA and GERSCH, 1985a; KITAGAWA and TAKANAMI, 1985), along the lines suggested by WAHBA (1980). This procedure is summarized as follows. First we assume that $\tau_I^2(2m) = \tau_I^2(2m-1)$ $(m = 1, \ldots, N/2)$ and calculate

$$s(m) = \frac{1}{2}(\widehat{u}_I{}^2(2m) + \widehat{u}_I{}^2(2m-1)). \tag{3.24}$$

If $\widehat{u}_I(n)$ is a realization of white noise from $N(0, \tau_I^2(n)\sigma^2)$, then the logarithm of $s(m)$ follows the almost normal distribution with the mean of $\log(\tau_I^2(2m)\sigma^2) - \gamma$ and variance of $\pi^2/6$, where γ is the Euler constant (WAHBA, 1980; KITAGAWA and GERSCH, 1985a). Secondly, based on an assumption that the variance of $u_I(n)$ changes slowly, we smooth $v(m)$ which is defined by

$$v(m) = \log s(m) + \gamma. \tag{3.25}$$

This transformation favors a estimation of $\tau_I(2m)$, because a least squares approach to smooth $v(m)$ can be justified from $v(m) \sim N(\log(\tau_I^2(2m)\sigma^2), \pi^2/6)$.

To obtain the smoothed $v(m)$, $r(m)$, we also apply the state space model approach of the linear model with Gaussian system and observation noises. In this case, we take the simple form for a system model

$$r(m) = r(m-1) + w_s(m), \qquad w_s(m) \sim N(0, \tau_s^2\tau_o^2), \tag{3.26}$$

and for an observation model

$$v(m) = r(m) + w_o(m), \qquad w_o(m) \sim N(0, \tau_o^2). \tag{3.27}$$

The value of a hyperparameter τ_s^2 is determined by the aforementioned minimum ABIC procedure. When we interpret these models in terms of the constrained least squares approach, the optimization problem for τ_s^2 is reduced to estimate the smoothing parameter, $1/\tau_s^2$, in an evaluation function of

$$\sum_{m=1}^{N/2}\Big(v(m) - r(m)\Big)^2 + (\alpha r(1))^2 + \frac{1}{\tau_s^2}\sum_{m=2}^{N/2}\Big(r(m) - r(m-1)\Big)^2. \tag{3.28}$$

Since this function takes a quadratic form with respect to the parameter $r(m)$, we can define the value of the hyperparameter τ_s^2, as previously mentioned in the section of (3.1), by introducing a Bayesian interpretation of this evaluation function (3.28). Namely the best value for τ_s^2 can be uniquely determined so as to minimize a Bayesian information criterion ABIC. We remind the reader that it is unnecessary for giving a value of τ_o^2. For a fixed τ_s^2, τ_o^2 is naturally defined by the variance of the residual sequence $w_o(m) = v(m) - r(m)$, which are derived from the estimated $r(m)$. In a case such smoothing, the hyperparameter τ_s^2 is closely related to the halfband width and its relationship has been shown (HIGUCHI, 1991). In short, a choice of the hyperparameter is completely equivalent to setting a halfband width when we use a digital lowpass filter for smoothing data. Hence, ABIC gives us an opportunity to objectively define the halfband width of the lowpass filter.

In searching the optimal τ_s^2, we use a coarse grid search method in the discrete point of τ_s^2. We feel that this simple approach can give a good estimate of $r(m)$, and this view is shared by many authors (AKAIKE, 1980; KITAGAWA and GERSCH, 1984, 1985a; TANABE, 1985; HIGUCHI *et al.*, 1988). However, we furthermore apply the Newton method to obtain the best τ_s^2 with the minimum ABIC, specified hereafter by $\tau_s^{2^*}$, around τ_s^2 which approximately minimizes ABIC in a coarse grid search. The obtained $\tau_s^{2^*}$ generates $r(m)$, and accordingly the unknown time varying variance can be determined by

$$\widehat{\tau}_I^2(2m)\sigma^2 = e^{r(m)}. \tag{3.29}$$

For the subsequent notation, we normalize $\widehat{\tau}_I^2(2m)\sigma^2$ by its maximum value;

$$\widetilde{\tau}_I^2(2m) = \frac{\widehat{\tau}_I^2(2m)\sigma^2}{max(\widehat{\tau}_I^2(2m)\sigma^2)} = \frac{\widehat{\tau}_I^2(2m)}{max(\widehat{\tau}_I^2(2m))}. \tag{3.30}$$

We denote this maximum value by τ_I^2.

For the given data $y(n)$ and hyperparameters of T_{obs} and θ_0, we fix this scaled $\widetilde{\tau}_I^2(2m)$ and take τ_I^2 as an unknown hyperparameter. Namely, instead of regarding $\tau_I(n)$ as hyperparameters, we hereafter consider only τ_I^2 as a hyperparameter. This means that a form of an envelope function of $u_I(n)$ holds in the successive procedure to search the best hyperparameters with the minimum ABIC in (3.22). By this simplification, the magnitude of the time varying variance of $u_I(n)$ depends only on the value of τ_I^2 at each n.

3.3 Choice of hyperparameters with the minimum ABIC

By the procedure presented in Subsection 3.2, we restrict the hyperparameter set from $\Lambda = [\tau_t^2, \tau_I^2(1), \ldots, \tau_I^2(N), T_{obs}, \theta_0]$ to $\lambda = [\tau_t^2, \tau_I^2, T_{obs}, \theta_0]$. ABIC should be therefore expressed as a function of λ and defined by

$$ABIC = -2 \log p(Y_N|\lambda). \tag{3.31}$$

We can numerically search the best hyperparameter set λ with the minimum ABIC of (3.31). Of course, to obtain the best hyperparameter set can be achieved by a coarse grid search over the λ space, but this approach is very redundant due to the relatively high dimension hypercube of hyperparameters and then requires lots of computing time. Thus far, in our case, we adopt a quasi-Newton method which is commonly used to minimize the nonlinear function with respect to parameters (NISHIKAWA *et al.*, 1982; PRESS *et al.*, 1988). To use the quasi-Newton method can definitely reduce the times of calculating the value of ABIC in (3.31) for the given hyperparameter set of λ. compared with that by the grid search.

4. Example

In this section, we demonstrate the procedure explained in the above section by using the data $y(n)$ shown in Fig. 2.

4.1 Estimation of $\tau_I^2(n)$

As mentioned in Subsection 3.2, we first estimate the time varying variance of $u_I(n)$ to decompose the data $y(n)$ into the multi-components expressed in (3.1). To do this. we calculate $\widehat{u}_I(n)$ according to (3.23). It should be again noticed that $\theta(n)$ is defined for the given (T_{obs}, θ_0) and then $\widehat{u}_I(n)$ also corresponds uniquely to the parameter set (T_{obs}, θ_0) through $f(n)$. The obtained $\widehat{u}_I(n)$, based on $\theta(n)$ shown in Fig. 2(b), is indicated by a solid line in Fig. 3(a). It is clearly seen in Fig. 3(a) that the relatively fast wiggles of $\widehat{u}_I(n)$ is modulated by a relatively slowly changing envelope function. Namely, the variance of $\widehat{u}_I(n)$, $\tau_I^2(n)\sigma^2$, appears to change slowly and thus this behavior justifies the assumption that the variance of $u_I(n)$ depends on time. Obviously, as seen in Fig. 3(a), there is no estimate of $u_I(n)$ for a phase within $|\theta(n)| > 80°$.

We calculate $v(m)$ defined in (3.25) and show its curve in Fig. 3(b). The smoothed $v(m)$, $r(m)$, at $\tau_s^{2*} = 3.753 \times 10^{-3}$ is in this figure denoted by a continuous solid curve. We emphasize the fact that the smoothing with the state space modeling automatically interpolates $r(m)$. The estimated $r(m)$ yields $\widehat{\tau}_I^2(2m)\sigma^2$ through (3.29). We superpose a curve of $\widehat{\tau}_I(n)\sigma$ on Fig. 2(a). It seems in Fig. 2(a) that the envelope of $\widehat{u}_I(n)$ can be satisfactorily represented by the estimated $\widehat{\tau}_I(n)\sigma$ curve.

4.2 Extraction of the background noise

To eliminate the background noise component from the observation data, we search the hyperparameter set λ with the minimum ABIC of (3.31). The estimated background noise component, $f(n)I(n)$, with the minimum ABIC is demonstrated in Fig. 4(a). We show in Fig. 4(b) the estimated $t(n) + e(n)$ which is free from the background noise component. At this estimation, the hyperparameter set, which has the minimum ABIC, is $\tau_t^2 = 0.1149$, $\tau_I^2 = 48.47$, $T_{obs} = -220.4$[sec.], and $\theta_0 = 87.43$[deg.].

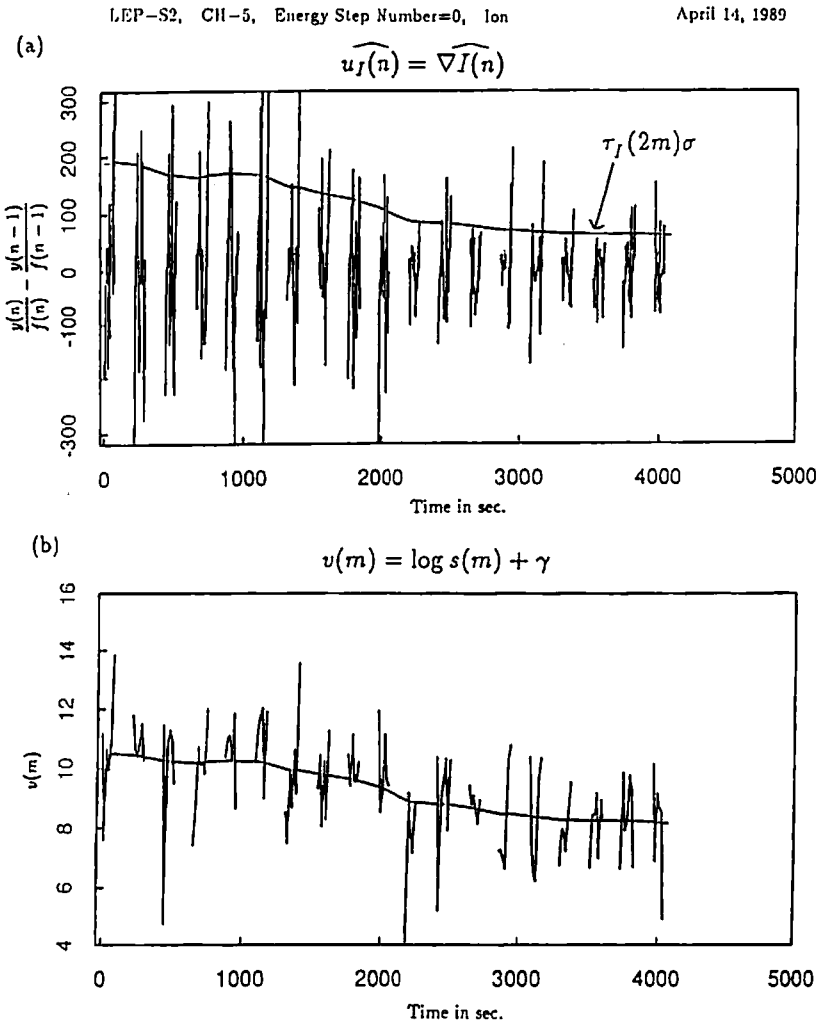

Fig. 3. (a) $\widehat{u}_I(n)$, based on $\theta(n)$ shown in Fig. 2(b). The estimated $\widehat{\tau}_I(n)\sigma$ is superposed on it. (b) $v(m)$ defined in (3.27). The smoothed $v(m)$, $r(m)$, is indicated by a solid curve.

We also exhibit the original data $y(n)$ in Fig. 4(c) in order to facilitate its comparison with Fig. 2(a). It is seen that the background noise which significantly contaminates the data can be satisfactorily subtracted from $y(n)$.

5. Some Comments on Actual Procedure

As demonstrated in Sections 3 and 4, we considered the time series for the fixed channeltron, energy step numbers, and species, $y_{i(l)}^{j,k}$, and removed the background noise component from it. In a case of energy scanning pattern shown in Fig. 1, the number of $y_{i(l)}^{j,k}$ is 1240, and so that we apply the aforementioned procedure for removing to each time series $y_{i(l)}^{j,k}$. Namely we repeat, for 1240 times over, an optimization to search values of hyperparameters λ for each $y_{i(l)}^{j,k}$. We obtain 1240 sets of optimal hyperparameters accordingly. However, the optimization like this is highly redundant and unsuited for an actual procedure. Several reasonable improvements are possibly

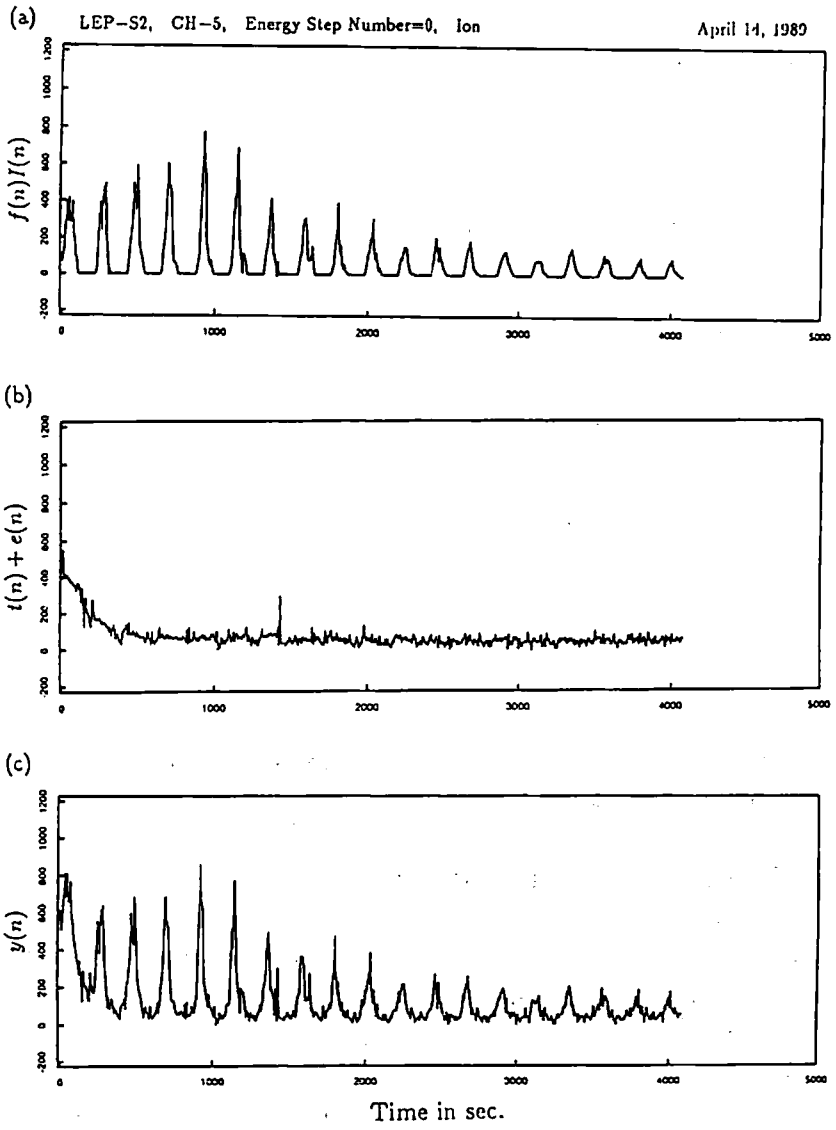

Fig. 4. (a) The estimated background noise component, $f(n)I(n)$. (b) The component free from the background noise, $t(n) + e(n)$. (c) The row data, $y(n)$, which is identical to Fig. 2(a).

considered to reduce a computing time required for filtering the background noise component out.

Here the reader should be reminded of that all channeltron is installed in a same plane which includes the spin axis (see Fig. 2 in MUKAI *et al.* (1990)). It is easily suggested from the configuration of the channeltorons that, for a time series $y_{i(l)}^{j,k}(n)$ with the same energy step number $i(l)$, the phase with the channeltron number j of $1 \sim 5$, i.e., $\theta_{i(l)}^{j,k}(n)$ with $1 \leq j \leq 5$, should be identical each other. Similarly, $\theta_{i(l)}^{j,k}(n)$ is common to the channeltron with its number j of $6 \sim 10$ and exactly anti-phase with that with $j = 1 \sim 5$.

The other redundancy in optimization is that the value of $T_{i(l),obs}^{j,k}$ should be essentially common to all $y_{i(l)}^{j,k}$. Even if we search the best hyperparameters for each $y_{i(l)}^{j,k}$, a

resulting $T^{j,k}_{i(l),obs}$ approximately gives the same value. There is little, if any, discrepancy of $T^{j,k}_{i(l),obs}$ from its mean. Consequently, once we estimate the value of $T^{j,k}_{i(l),obs}$, we need not optimize ABIC with respect to $T^{j,k}_{i(l),obs}$ for each $y^{j,k}_{i(l)}$.

In addition, a simple relationship between $\theta^{j,k}_{i(l)}(0)$ even for different energy step number is naturally suggested from a energy scanning pattern (see Fig. 1). From the obtained $T^{j,k}_{0,obs}$ and $\theta_0^{j,k}(0)$, $\theta^{j,k}_{i(l)}(0)$ can be given by

$$\theta^{j,k}_{i(l)}(0) \simeq Mod\Big\{\theta_0^{j,k}(0) + 2\pi\Big(g(j) + \frac{2T_E}{T_s}\frac{f(i(l))}{2+2\times(32-2)}\Big), 2\pi\Big\} \tag{5.1}$$

where $g(j)$ and $f(i(l))$ are defined by

$$g(j)\begin{cases} 0 & j = 1 \sim 5 \\ 1/2 & j = 6 \sim 10\ , \end{cases} \tag{5.2}$$

and

$$f(i(l))\begin{cases} i/2 & l = 1 \\ -i/2 & l = 2\ , \end{cases} \tag{5.3}$$

respectively. This relationship does not hold in a strict sense, because no effect of the change in a position of satellite relative to the noise source during the one spin period is taken into account. In addition, since the resulting $f(n)I(n)$ drastically changes according to $\theta^{j,k}_{i(l)}(0)$, thereby we had better definitely search a better $\theta^{j,k}_{i(l)}(0)$ with the minimum ABIC around the roughly estimated $\theta^{j,k}_{i(l)}(0)$ by (5.1). Namely, $f(n)I(n)$ is sensitive $\theta^{j,k}_{i(l)}(0)$ but insensitive to other hyperparameters.

Finally, we would like to comment on a limitation of our model expressed by (3.1). Since we consider the LEP data highly contaminated with the background noise, our model is unsuitable for the data which slightly contains the component probably attributed to the background noise. For example, we can see little effect from the background noise in the LEP data for the channeltorons of which the view direction is anti-solar (the spin axis is controlled to be directed toward the sun). Even if our model is applied to the LEP data like this, the nonlinear optimization does not converge during the finite computing time. Then, we should not apply our approach to such data (obviously, we need not use our procedure, because there is little background noise component in the raw data). In such a case, the Bayesian approach for removing the periodic noise (HIGUCHI *et al.*, 1988) can be sufficiently applied. By using it, we can reduce the computing time drastically.

I am grateful to Dr. Mukai for his providing the LEP data on board Akebono (EXOS-D) satellite and for his kind explanation of the LEP instruments. I thank Dr. Kitagawa for his useful comments on the nonlinear optimization and for supplying me with the FORTRAN program for the quasi-Newton method. The author is also grateful to G. K. Crawford (Institute of Geophysics and Planetary Physics, University of California, Los Angeles) for his helpful comments for improving the manuscript.

This work was supported by the Grant-in-Aid for Encouragement of Young Scientists project 02740218, Ministry of Education, Science and Culture, Japan, and partly carried out under by the ISM Cooperative Research Program (90-ISM.CRP-46).

REFERENCES

AKAIKE, H., Information theory and an extension of the maximum likelihood principle, in *Proc. 2nd International Symposium in Information Theory*, edited by B. N. Petrov and F. Csáki, Akademiai Kiado, Budapest, 1973.

AKAIKE, H., Likelihood and the Bayes procedure (with discussion), in *Bayesian Statistics*, edited by J. M. Bernardo, M. H. De Groot, D. V. Lindley, and A. F. M. Smith, pp. 143–166, University Press, Valencia, Spain, 1980.

ANDERSON, B. D. O. and J. B. MOORE, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, N.J., 1979.

BESAG, J., On the statistical analysis of dirty pictures (with discussion), *J. R. Statist. Soc. B*, **48**, 259–302, 1986.

GEMAN, S. and D. GEMAN, Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-6, 721–741, 1984.

GERSCH, W. and G. KITAGAWA, Smoothness priors in time series, in *Bayesian Analysis of Time Series and Dynamic Models*, edited by J. C. Spall, pp. 431–476, Marcel Dekker, INC., New York and Basel, 1988.

GOOD, I. J., *The Estimation of Probabilities*, Cambridge, Mass.:MIT Press, 1965.

GOOD, I. J. and R. A. GASKIN, Non-parametric roughness penalties for probability densities, *Biometrika*, **58**, 255–277, 1971.

HIGUCHI, T., Frequency domain characteristics of linear operator to decompose a time series into the multi-components, to be published in *Ann. Inst. Statist. Math.*, 1991.

HIGUCHI, T., K. KITA, and T. OGAWA, Bayesian statistical inference to remove periodic noise in the optical observations aboard a spacecraft, *Appl. Opt.*, **27**, 4514–4519, 1988.

INOUE, H., A smooth fitting based on a generalized inverse theory, *Jyouhou Chishitsu*, **10**, 105–134, 1985 (in Japanese).

ISHIGURO, M., Computationally Efficient Implementation of a Bayesian seasonal adjustment procedure, *J. Time Ser. Anal.*, **5**, 245–253, 1984.

ISHIGURO, M. and E. ARAHATA, A Bayesian spline regression, *Proc. Inst. Stat. Math.*, **30**, 29–36, 1982 (in Japanese).

ISHIGURO, M. and Y. SAKAMOTO, A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115–137, 1983.

ISHIGURO, M. and Y. SAKAMOTO, A bayesian approach to the probability density estimation, *Ann. Inst. Statist. Math.*, **36**, 523–538, 1984.

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

KATAYAMA, T., *Applied Kalman Filter*, Asakura Syoten, Tokyo, 1983 (in Japanese).

KITA, K., T. HIGUCHI, and T. OGAWA, Bayesian statistical inference of airglow profiles from rocket observational data: comparison with conventional methods, *Planet. Space Sci.*, **37**, 1327–1331, 1989.

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

KITAGAWA, G., Changing spectrum estimation, *J. Sound Vib.*, **89**(3), 433–445, 1983.

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

KITAGAWA, G., Numerical approach to non-Gaussian smoothing and its applications, *Proc. 20th Symp. Interface: Comput. Sci. Statist.*, edited by E. J. Wegmar, D. T. Gants, and J. J. Miller, pp. 379–388, Am. Statist. Assoc., VA, U.S.A., 1988.

KITAGAWA, G., Non-Gaussian time series modeling, *Oper. Res.*, **10**, 541–546, 1989a (in Japanese).

KITAGAWA, G., Kalman filter, in *Jikeiretsu-Ron*, edited by T. Ozaki, pp. 91–100, Housou-Daigaku Kyouiku Shinkou-Kai, Tokyo, 1989b (in Japanese).

KITAGAWA, G. and W. GERSCH, A smoothness priors-state space modeling of time series with trend and seasonality, *J. Am. Stat. Assoc.*, **79**, 378–389, 1984.

KITAGAWA, G. and W. GERSCH, A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series, *IEEE Trans. Autom. Control*, **AC-30**, 48–56, 1985a.

KITAGAWA, G. and W. GERSCH, A smoothness priors long AR model method for spectral estimation, *IEEE Trans. Autom. Control*, AC-30, 57–65, 1985b.

KITAGAWA, G. and T. TAKANAMI, Extraction of signal by a time series model and screening out micro earthquakes, *Signal Process.*, 8, 303–314, 1985.

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

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

MARROQUIN, J., S. MITTER, and T. POGGIO, Probabilistic solution of ill-posed problem in computational vision, *J. Am. Stat. Assoc.*, 82, 76–89, 1987.

MUKAI, T., N. KAYA, E. SAGAWA, M. HIRAHARA, W. MIYAKE, T. OBARA, H. MIYAOKA, S. MACHIDA, H. YAMAGISHI, M. EJIRI, H. MATSUMOTO, and T. ITOH, Low energy charged particle observations in the "Auroral" Magnetosphere first results from the Akebono (EXOS-D) satellite, *J. Geomag. Geoelectr.*, 42, 479–496, 1990.

NAKAMURA, T., Bayesian cohort models for general cohort table analysis, *Ann. Inst. Statist. Math. Part* B, 38, 353–370, 1986.

NISHIKAWA, Y., N. SANNOMIYA, and T. IBARAKI, *Saitekika*, Iwanami Syoten, Tokyo, 1982 (in Japanese).

O'HAGAN, A., On posterior joint and marginal modes, *Biometrika*, 63, 329–333, 1976.

OYA, H., Preface, *J. Geomag. Geoelectr.*, 42, 365–366, 1990.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY, and W. T. VETTERLING, *Numerical Recipes in C*, Cambridge University Press, New York, 1988.

SAKAMOTO, Y. and M. ISHIGURO, A bayesian approach to nonparametric test problems, *Ann. Inst. Statist. Math.*, 40, 587–602, 1988.

SAKAMOTO, Y., M. ISHIGURO, and G. KITAGAWA, *Akaike Information Criterion Statistics*, D. Reidel Publishing Company, Holland, 1986.

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

TAMURA, Y. H., An approach to the nonstationary process analysis, *Ann. Inst. Statist. Math.*, 39-B, 227–241, 1987.

TANABE, K., Statistical approach to incorrectly posed problem, *Suuri Kagaku*, 3, 1–5, 1976 (in Japanese).

TANABE, K., Bayesian model and ABIC, *Oper. Res.*, 3, 178–183, 1985 (in Japanese).

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

TIKHONOV, A. N. and V. Y. ARSENIN, *Solutions of Ill-posed Problems*, Winston & Suns, Washington, DC., 1977.

TITTERINGTON, D. M., Common structure of smoothing techniques in statistics, *Int. Stat. Rev.*, 53, 141–170, 1985.

WAHBA, G., Automatic smoothing of the log periodgram, *J. Am. Stat. Assoc.*, 75, 122–132, 1980.