Reprinted from Applied Optics, Vol. 27, page 4514, November 1, 1988

Bayesian statistical inference to remove periodic noise in the optical observations aboard a spacecraft

Tomoyuki Higuchi, K. Kita, and Toshihiro Ogawa

Optical data taken aboard a spacecraft sometimes suffer from an unexpected modulation synchronized with the rotation or wobble of the spacecraft. This modulation may cause a serious error in estimating the volume emission rate from the observed column emission rate when an inversion method is used, since the differential operation in the inversion procedure is sensitive to noise. A new statistical technique that discriminates between random and periodic noise has been developed in this study. This technique, based on the concept of entropy maximization, gives the best-fit solution to the data. The criterion of fit used in this procedure is the logarithmic likelihood of a solution calculated in terms of the probability characteristics of the data. This method is very effective on data containing a periodic modulation due to the spin motion of a rocket as well as a random fluctuation. We use this technique in the analysis of 5577-Å airglow data obtained aboard a rocket.

I. Introduction

Inversion procedures are sensitive to noise in observed data, because they involve differential operations. Thus far, attempts to eliminate the noise have been made by smoothing and filtering the observed data. In particular, in observations obtained with an optical instrument on board a spacecraft, a spurious modulation of the observed data sometimes occurs in conjunction with a periodic change in the spacecraft attitude. Several procedures have been developed to recover the volume emission rate from a measurement of the column emission rate as a function of height.[1–3] However, a direct inversion by differentiating the column emission rate with height leads to an erroneous result, because the periodic noise due to the spacecraft wobble is amplified by the differential operation. This introduces a spurious oscillation in the recovered height profile of the volume emission rate.[1] Therefore it is necessary to remove this noise prior to applying the inversion procedure.

Smoothing the data or fitting a smooth curve to the data is useful for eliminating noise. However, the function form in curve fitting is arbitrarily chosen, because no *a priori* criterion is available. To overcome this problem, Akaike proposed an information criterion on goodness-of-fit based on the principle of maximization of likelihood.[4,5] Since the Bayesian approach to statistical inference is useful in many fields of data analysis, Akaike extended his information criterion to a choice of parameters of a Bayesian model.[6]

In this paper we introduce a Bayesian inference model for observed data. The proposed model has many parameters that are adjustable to random and periodic components of the data. The optimal model is determined using the information criterion of Akaike.[7,8] Our procedure is applied not only to the simulated data but also to observed 5577-Å airglow data in terms of the column emission rate.

II. Method

We consider a series of observed values, I_j, taken at constant time intervals, represented by $\mathbf{I} = [I_1, I_2, \ldots I_N]^T$, where N is the total number of data and T denotes matrix transposition. It is assumed that the I_j can be decomposed as

$$I_j = T_j + S_j + \epsilon_j, \tag{1}$$

where T_j, S_j, and ϵ_j are trend, periodic, and random components, respectively. In this model T_j and S_j represent the systematic part of the observed data. Therefore the number of parameters is $2N$ plus the variance of the random component which is assumed to have zero mean. Hereafter we use the following expression for these components:

$$\mathbf{a} = [T_1, T_2, \ldots T_N, S_1, S_2, \ldots S_N]^T. \tag{2}$$

Let the periodic component be $S_j = \sin(2\pi j f_c \Delta t)$, where f_c is a characteristic frequency and Δt is the time

The authors are with University of Tokyo, Geophysics Research Laboratory, Bunkyo-ku, Tokyo 113, Japan.
Received 12 February 1988.
0003-6935/88/214514-06$02.00/0.

interval of sampling. The constraint on the periodic component is given by the second-order difference equation

$$S_{j+2} - 2cS_{j+1} + S_j = 0, \tag{3}$$

where c is a constant defined by $c = \cos(2\pi f_c \Delta t)$. Since this equation describes the local characteristics of the sinusoidal variation, it weakly constrains the modulation of the amplitude. If c is not known beforehand, it may be determined by visual inspection of the data.

We assume that the trend component, T_j, varies as a smooth function locally. The constraints on both the periodic and trend components are then given in such a way that $M(\mathrm{a})$ has a minimal, where

$$M(\mathrm{a}) = (\alpha T_1)^2 + (\beta[T_2 - T_1])^2 + (s\alpha S_1)^2 + (s\beta[S_2 - S_1])^2 + \sum_{j=1}^{N-2} [T_{j+2} - 2T_{j+1} + T_j]^2 + (s[S_{j+2} - 2cS_{j+1} + S_j])^2. \tag{4}$$

α and β are constants for special treatment of T_1, T_2, S_1, and S_2 and should be properly chosen, because the second-order difference equation cannot be defined for T_1, T_2, S_1, and S_2. The particular choice of α and β has little influence on selecting a model. s is a hyperparameter whose function is to balance the constraints for S_j and T_j. The periodic component becomes more evident in the result obtained with an increase of s. We also assume that the systematic components, $T_j + S_j$, do not deviate very much from the observed data I_j. This assumption leads to the minimization of $K(\mathrm{a})$, where $K(\mathrm{a})$ is defined by

$$K(\mathrm{a}) = \sum_{j=1}^{N} [I_j - T_j - S_j]^2 = \sum_{j=1}^{N} \epsilon_j^2. \tag{5}$$

$K(\mathrm{a})$ can be rewritten as

$$\|\mathrm{I} - \bar{X}\mathrm{a}\|^2, \tag{6}$$

where $\bar{X}$ is an $(N \times 2N)$ matrix represented by

$$\bar{X} = \begin{pmatrix} 1 & & & 1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & 1 \end{pmatrix}, \tag{7}$$

and $\|\cdot\|$ denotes the standard Euclidean norm. In similar matrix–vector notation, $M(\mathrm{a})$ can be expressed as $M(\mathrm{a}) = \|\bar{D}\mathrm{a}\|^2$, where $\bar{D}$ is a $(2N \times 2N)$ matrix defined by

$$\bar{D} = \begin{pmatrix} \bar{D}_2 & \bar{O} \\ \bar{O} & s\bar{D}_f \end{pmatrix}, \tag{8}$$

with $\bar{D}_2$ and $\bar{D}_f$ being $(N \times N)$ matrices represented by

$$\bar{D}_2 = \begin{pmatrix} \alpha & & & & O \\ -\beta & \beta & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ O & & 1 & -2 & 1 \end{pmatrix}, \tag{9}$$

and

$$\bar{D}_f = \begin{pmatrix} \alpha & & & & O \\ -\beta & \beta & & & \\ 1 & -2c & 1 & & \\ & \ddots & \ddots & \ddots & \\ O & & 1 & -2c & 1 \end{pmatrix}. \tag{10}$$

In Eq. (8) $\bar{O}$ is a zero matrix. Note that c in the matrix $\bar{D}_f$ is the same as in Eq. (3). We solve the vector a by minimizing $K(\mathrm{a}) + d^2 M(\mathrm{a})$, where d is a hyperparameter used to balance the constraints minimizing $K(\mathrm{a})$ and $M(\mathrm{a})$. If the values of s and d are given, the solution for the vector a is reduced to that of a constrained least-square problem.

The hyper parameters s and d are determined using the method of Akaike.[6] We assume that the following prior distribution of a defined by

$$g(\mathrm{a}|\sigma^2,s,d) = \frac{1}{(2\pi\sigma^2)^N} (\det|d^2\bar{D}^T\bar{D}|)^{1/2} \exp\left(-\frac{d^2\|\bar{D}\mathrm{a}\|^2}{2\sigma^2}\right), \tag{11}$$

where $(\det|d^2\bar{D}^T\bar{D}|)^{1/2}$ is the normalization factor of $g(\cdot)$, and σ is a positive constant. When the data I_j are given, the posterior distribution of a determined by the data distribution is given by

$$h(\mathrm{a}) = f(\mathrm{I}|\sigma^2,\mathrm{a})g(\mathrm{a}|\sigma^2,s,d), \tag{12}$$

where $f(\cdot)$ is defined by

$$f(\mathrm{I}|\sigma^2,\mathrm{a}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\|\mathrm{I} - \bar{X}\mathrm{a}\|^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_{j=1}^{N} \epsilon_j^2}{2\sigma^2}\right). \tag{13}$$

If d, s, and σ^2 are given along with the observational data I, then a is obtained by minimizing $\|Z(\mathrm{a}|d,s)\|^2$, where

$$\|Z(\mathrm{a}|d,s)\|^2 = \|\mathrm{I} - \bar{X}\mathrm{a}\|^2 + d^2\|\bar{D}\mathrm{a}\|^2. \tag{14}$$

The minimum value for given d and s is specified as $\|Z(\mathrm{a}^*|d,s)\|^2$, where a^* is given by

$$\mathrm{a}^* = [\bar{X}^T\bar{X} + d^2\bar{D}^T\bar{D}]^{-1}\bar{X}^T \cdot \mathrm{I}. \tag{15}$$

The essential concept behind determining the hyperparameters d, s, and σ^2 is to consider the mean of the posterior distribution $h(\cdot)$. We consider the marginal likelihood of (d,s,σ^2) defined by

$$L(d,s,\sigma^2) = \int_{-\infty}^{+\infty} h(\mathrm{a})d\mathrm{a} = \int_{-\infty}^{+\infty} f(\mathrm{I}|\sigma^2,\mathrm{a})g(\mathrm{a}|\sigma^2,s,d)d\mathrm{a}. \tag{16}$$

Integrating, we get

$$L(d,s,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\|Z(\mathrm{a}^*|d,s)\|^2}{2\sigma^2}\right) (\det|d^2\bar{D}^T\bar{D}|)^{1/2} \times (\det|d^2\bar{D}^T\bar{D} + \bar{X}^T\bar{X}|)^{-1/2}. \tag{17}$$

For given d and s, the maximum of the logarithm of $L(d,s,\sigma^2)$ with respect to σ^2 occurs at

$$\sigma^{2*} = \frac{\|Z(\mathrm{a}^*|d,s)\|^2}{N}. \tag{18}$$

In practical applications we consider finite sets of d and s values in choosing the model that maximizes

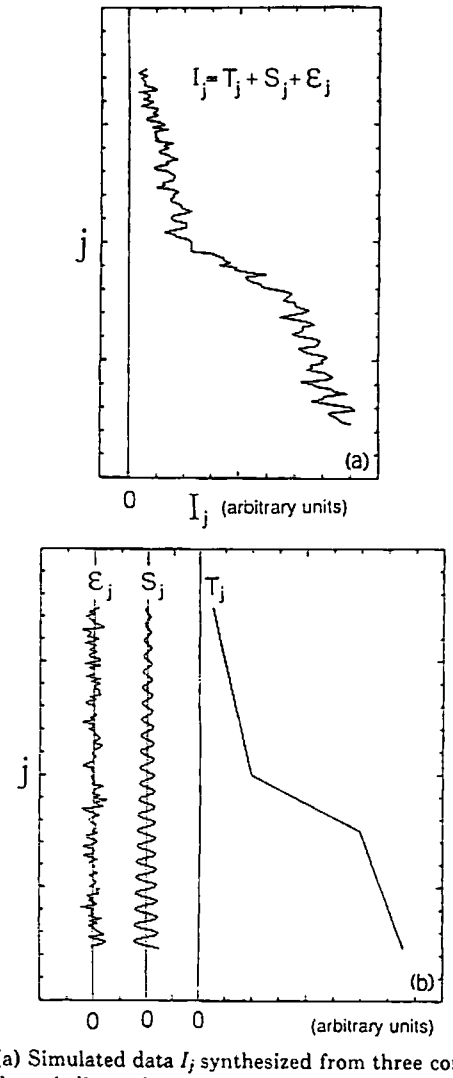

Fig. 1. (a) Simulated data I_j synthesized from three components: the trend, periodic, and random components shown in (b). (b) Three components of the data in (a). The random components ϵ_j were generated from a sequence of Gaussian white noise with an average of 0.

$L(d,s,\sigma^{2*})$. The coefficient c may be treated as a hyperparameter, but in this study it is given beforehand. In place of $L(\cdot)$, we use

$$ABIC(d,s) = -2\log L(d,s,\sigma^{2*}) \tag{19}$$

after Akaike, where $ABIC(d,s)$ represents a Bayesian information criterion. Therefore, the hyperparameters s and d are optimal for minimizing $ABIC$. The minimum $ABIC$ value is denoted by $ABIC^*$, which occurs at $s = s^*$ and $d = d^*$. Although in this study the optimal value of a has been chosen at $s = s^*$, $d = d^*$, $\sigma = \sigma^*$, it is more appropriate to take an average of a, with a weight of $\exp[-ABIC(d,s)/2]$, for many hyperparameters s and d near $s = s^*$ and $d = d^*$.

III. Results

A. Application to Simulated Data

We apply our technique to the simulated data shown in Fig. 1(a). These data, $I_j (j = 1,2,\ldots,N)$, were synthesized from the trend, periodic, and random components shown in Fig. 1(b). The random component was generated from a sequence of Gaussian white noise which has an average of 0. The technique described in Sec. II is used to decompose the I_j [Fig. 1(a)] into their original three components [Fig. 1(b)]. $ABIC$ values are calculated for the data shown in Fig. 1(a) for many values of s and d. The results are shown in Table I. We see that $ABIC$ attains its minimum at $d = 16$ and $s = 1$. If we want to specify the values of d and s more precisely, we should search for the model with the smallest $ABIC$ in the vicinity of $(d,s) = (16,1)$.

Figure 2(a) shows the three resolved components, T_j, S_j, and ϵ_j, of the model with $ABIC^*$. Comparing Fig. 2(a) with Fig. 1(b), it is evident that our technique reproduces the periodic component as well as the trend component fairly well. There is, however, a small change in the amplitude of the periodic component, S_j, from the original one in Fig. 1(b). Also, because local smoothness has been assumed for the trend component, there is a small discrepancy near the corners of T_j. These differences are minor, however, and the overall agreement between Figs. 1(b) and 2(a) is excellent.

To demonstrate that the model with $ABIC^*$ is the best estimate of the original three components in Fig. 1(b), two models for $(d,s) = (8,0.25)$ and $(64,2)$ are shown in Figs. 2(b) and (c). The corresponding values of $ABIC$ are 799 and 829, respectively. Since the constraints on both the trend and periodic components becomes stricter as d increases, the trend component comes to a smooth and monotonically decreasing curve and the periodic component tends to a sinusoidal oscillation. The opposite is true for a small value of d where the constraint to the systematic part, $(S_j + T_j)$, is weak. Accordingly, the standard deviation of the random component becomes small; s determines the balance of constraints between the periodic and the trend components. The periodic component is approximately a sinusoidal oscillation for large s, while the trend component comes to a smooth curve for a small s. The above characteristics, arising from different values of the two hyperparameters d and s, are illustrated in Figs. 2(a)–(c).

Another test was done using the data of Fig. 3(a). These data were synthesized from the three simulated components in Fig. 3(b). Again $ABIC$ reaches its minimum at $(d,s) = (16,1)$ as shown in Table II. Fig-

Table I. $ABIC$ Values for Several Hyperparameters d and s

s	d = 64	32	16	8	4
4	827	786	770	769	782
2	829[a]	787	766	764	783
1	831	785	763[b]	766	793
0.5	831	784	766	777	814
0.25	832	788	778	799[c]	846

[a] $ABIC$ value for the model with $(d,s) = (64,2)$ in Fig. 2(c).
[b] $ABIC$ value for the model with $(d,s) = (16,1)$ shown in Fig. 2(a) ($ABIC^*$).
[c] $ABIC$ value for the model with $(d,s) = (8,0.25)$ in Fig. 2(b).

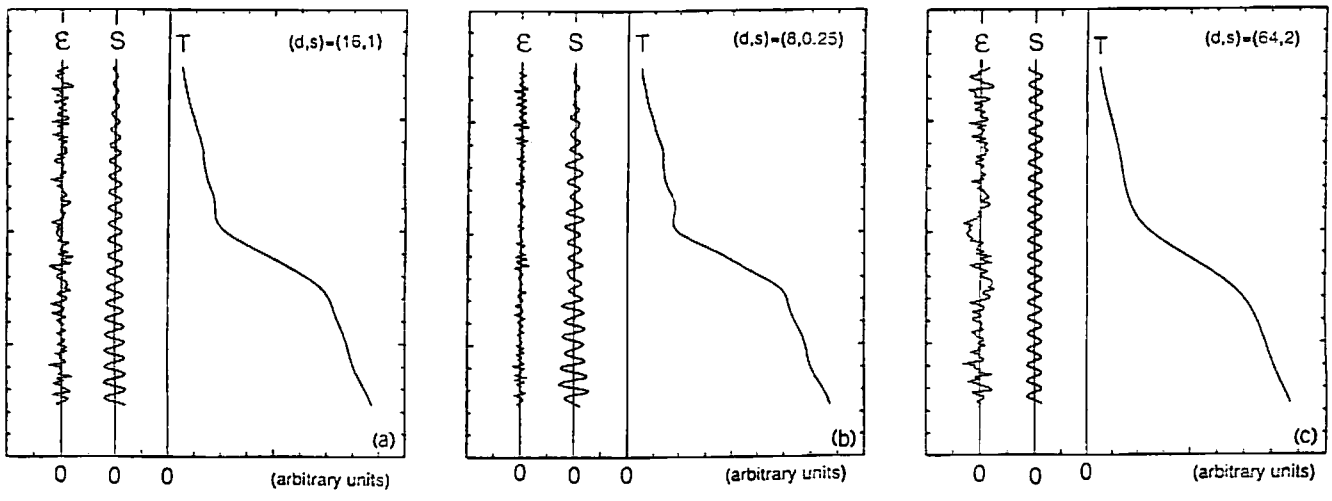

Fig. 2. Application to the simulated data in Fig. 1(a). I_j given in Fig. 1(a) is decomposed into three components, T_j, S_j, and ϵ_j, using the technique described in Sec. II. (a) The model at minimum *ABIC*. Three components T_j, S_j, and ϵ_j are resolved from I_j shown in Fig. 1(a). *ABIC* attains its minimum at $d = 16$ and $s = 1$. (b) The model with $(d,s) = (8,0.25)$. (c) The model with $(d,s) = (64,2)$.

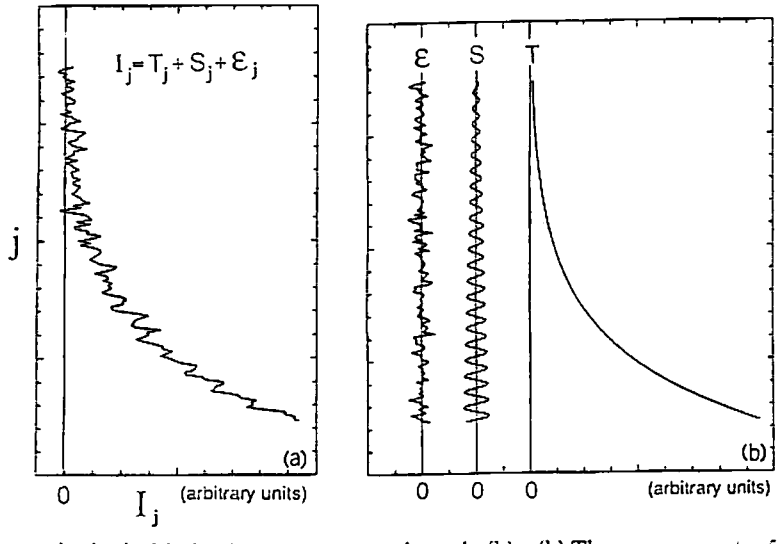

Fig. 3. (a) Simulated data synthesized with the three components shown in (b). (b) Three components of the data in (a).

ure 4 illustrates the model with *ABIC**. Both the trend and periodic components are in good agreement with their original counterparts [Fig. 3(b)] except for a small difference in amplitude in the periodic components.

Table II. *ABIC* Values for Several Hyperparameters d and s

S	d = 64	32	16	8	4
4	750	730	731	743	765
2	752	731	730	741	767
1	754	730	729*	744	775
0.5	754	730	732	753	793
0.25	756	734	742	772	823

* *ABIC* value of the model with $(d,s) = (16,1)$ shown in Fig. 4 (*ABIC**).

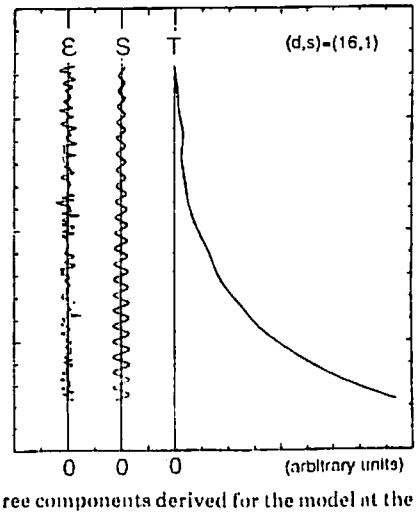

Fig. 4. Three components derived for the model at the minimum of *ABIC*.

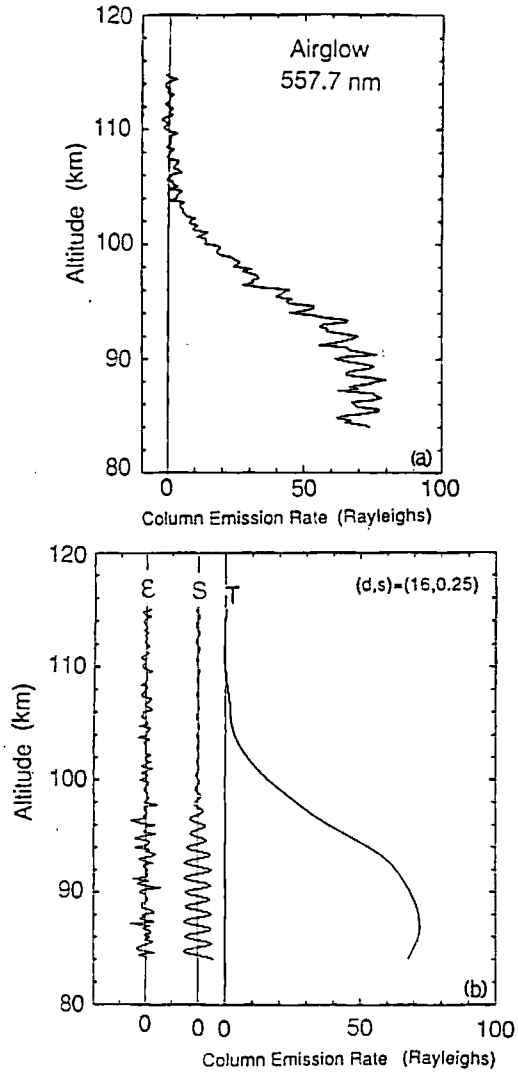

Fig. 5. (a) Height distribution of the vertical column emission rate of the 5577-Å airglow. (b) The model at the minimum of *ABIC* with $(d,s) = (16,0.25)$ is shown. T_j represents the estimate of the column emission rate.

B. Application to Observed Data

Our procedure is applied to the actual data observed aboard a rocket. The data set is the height variation of the vertical column emission rate of 5577-Å airglow and our objective is to infer the height profile of the volume emission rate. The data were obtained with a photometer on board the sounding rocket S310.17 flown from Uchinoura (31°15′N,131°05′E) at 1300 UT on 6 Sept. 1986, and are shown in Fig. 5(a). The side-looking photometer aboard a spin-stabilized rocket was used, and its look angle was determined by means of a geometric aspectmeter and a star sensor. Star background was subtracted from the data, and all the data are converted to the values at the zenith. Despite this shaping of the data, at low altitudes a modulation in synchronization with the rocket spin is superposed on the column emission data, in addition to noisy fluctuations. The characteristic period of the modulation

Table III. Values of *ABIC* for Several Hyperparameters *d* and *s*

	d		
S	32	16	8
0.5	879	841	854
0.25	852	840[a]	853
0.125	842	842	873

[a] Value for the model with $(d,s) = (16,0.25)$ shown in Fig. 5(b) (*ABIC*[*]).

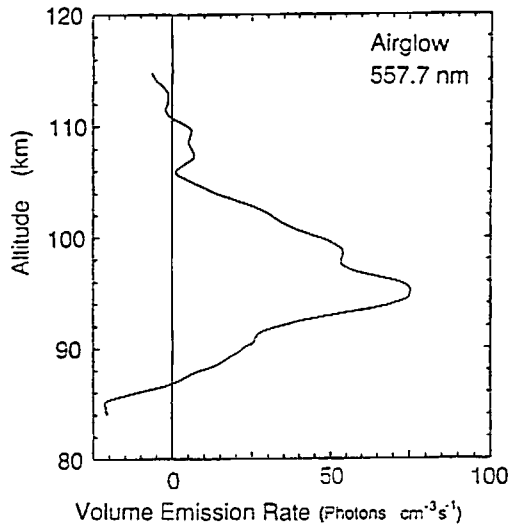

Fig. 6. Height distribution of the volume emission rate estimated by our inversion technique.

is exactly the same as that of the rocket spin. Before applying an inversion to reconstruct the volume emission rate, we removed the periodic component as well as the random fluctuation from the data as described in the previous subsection. In this case the model gives the minimum value of *ABIC* at $(d,s) = (16,0.25)$ (see Table III). The three components of the model are shown in Fig. 5(b). It is interesting to note that the periodic component is damped at high altitudes as is the trend component.

The vertical column emission rate I is related to the volume emission rate J by

$$I(z) = \int_z^{+\infty} J(z')dz', \tag{20}$$

or in discrete form,

$$I_k = \sum_{l=k}^{+\infty} J_l \Delta z, \tag{21}$$

where z denotes altitude. We calculate the volume emission rate, J_l, using

$$J_l = \frac{I_l - I_{l+1}}{\Delta z}, \tag{22}$$

where Δz in the present study is 200 m. The volume emission rate obtained is shown in Fig. 6. A comparison with other inversion techniques and a discussion from a physical viewpoint will be presented in a forthcoming paper.

IV. Summary

The major advantage of our inversion technique is that our procedure uses an objective criterion. In discussing a small-scale structure obtained in any inversion procedure, one should make sure that it is intrinsic to the inversion procedure. When analyzing observed data containing noisy fluctuations, one should, if possible, adopt a procedure that is independent of subjective assumptions.

In this study, we took into account only three components: trend, periodic, and random components and constructed the prior distributions for the trend and periodic components. We may specify other constraints in the model according to prior information on the trend and periodic components or add one or more components in the model. Our present procedure is useful for analyzing 5577-Å airglow data obtained by a rocket observation.

References

1. D. P. Murtagh, R. G. H. Greer, I. C. Mcdade, E. J. Llewellyn, and M. Bantle, "Representative Volume Emission Profiles from Rocket Photometer Data." Ann. Geophysicae 2, 467 (1984).
2. S. C. Solomon, P. B. Hayes, and V. J. Abreu, "Tomographic Inversion of Satellite Photometory," Appl. Opt. 23, 3409 (1984).
3. E. Yee, K. V. Paulson, and G. G. Shepherd, "Minimum Cross-Entropy Inversion of Satellite Photometer Data," Appl. Opt. 26, 2106 (1987).
4. H. Akaike, "A New Look at the Statistical Model Identification," IEEE Trans. Autom. Control AC-19, 716 (1974).
5. Y. Sakamoto, M. Ishiguro, and G. Kitagawa, *Akaike Information Criterion Statistics* (Reidel, Holland, 1986).
6. H. Akaike, "Likelihood and the Bayes Procedure," in *Bayesian Statistics* (Valencia U. P., Spain, 1980), p. 143.
7. H. Akaike, "Seasonable Adjustment by a Bayesian Modeling," J. Time Ser. Anal. 1, 1 (1980).
8. M. Ishiguro, "Computationally Efficient Implementation of Bayesian Seasonal Adjustment Procedure," J. Time Ser. Anal. 5, 245 (1984).

The authors thank K. Kohketsu and K. Takano for providing the terminal processing program ETERM. One of the authors (T.H.) is grateful to M. Ishiguro for his helpful discussions.