Ann. Inst. Statist. Math.
Vol. 43, No. 3, 407–433 (1991)

ESTIMATION OF THE ARRIVAL TIMES OF SEISMIC WAVES BY MULTIVARIATE TIME SERIES MODEL*

TETSUO TAKANAMI[1] AND GENSHIRO KITAGAWA[2]**

[1] *Research Center for Earthquake Prediction, Faculty of Science, Hokkaido University, Kita-ku, Sapporo 060, Japan*
[2] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*

(Received July 13, 1990; revised March 25, 1991)

Abstract. A computationally efficient procedure was developed for the fitting of many multivariate locally stationary autoregressive models. The details of the Householder method for fitting multivariate autoregressive model and multivariate locally stationary autoregressive model (MLSAR model) are shown. The proposed procedure is quite efficient in both accuracy and computation. The amount of computation is bounded by a multiple of Nm^2 with N being the data length and m the highest model order, and does not depend on the number of models checked. This facilitates the precise estimation of the change point of the AR model. Based on the AICs' of the fitted MLSAR models and Akaike's definition of the likelihood of the models, a method of evaluating the posterior distribution of the change point of the AR model is also presented. The proposed procedure is, in particular, useful for the estimation of the arrival time of the S wave of a microearthquake. To illustrate the usefulness of the proposed procedure, the seismograms of the foreshocks of the 1982 Urakawa-Oki Earthquake were analyzed. These data sets have been registered to AISM Data Library and the readers of this Journal can access to them by the method described in this issue.

Key words and phrases: Locally stationary AR model, AIC, Householder transformation, P wave, S wave, arrival time, seismology.

1. Introduction

In recent years, the automatic processing of seismic signals for the detection of seismic activity has become realistic due to the establishment of a well-equipped nation-wide seismological network system (Hamaguchi and Suzuki (1979)).

* A part of this research was carried out under the ISM Cooperative Research Program (89-ISM.CRP-57).

** Also with the Faculty of Economics, the University of Tokyo. The author was supported in part by the Japanese Ministry of Education, Science and Culture under Grant-in-Aid for Developmental Scientific Research 63830002.

In earthquake prediction, it is expected that by analyzing many foreshocks, important information about the forthcoming mainshock can be obtained. According to the Gutenberg-Richiter's law, $\log n(M) = a - bM$, the incremental or cumulative number of earthquakes as a function of magnitude M, $n(M)$, increases exponentially with the decrease of the magnitudes, M. Therefore, by developing a system which can analyze microearthquakes with smaller magnitudes, we can get increasingly much information about the seismicity of the region and, hopefully, about the forthcoming mainshock.

In practice, however, the analysis of such microearthquakes causes two problems. Firstly, the seismic signals observed by seismometers are contaminated by various kinds of noises, such as microtremors, microseisms, and artificial vibration. Since the noise level is almost a constant independent of the signal, the effect of the background noise becomes more severe, for earthquakes with smaller magnitudes. Therefore, if we want to analyze earthquakes with smaller magnitudes, it is required to develop a more sophisticated procedure which can handle very noisy data. Secondly, the number of earthquakes increases exponentially with the decrease of the magnitude. Seismograms have been conventionally handled by empirical methods based on the expertise of the human operator to single out real seismic signal from the various noises. However, for the processing of so many micro-earthquakes, it thus becomes necessary to develop a computationally efficient method that can automatically detect seismic wave from noisy data.

Some attempts have been made based on the autoregressive modeling of the seismic signals (Tjøstheim (1975), Hamaguchi and Suzuki (1979), Hamaguchi and Morita (1980), Yokota *et al.* (1981), Maeda (1985) and Hasegawa *et al.* (1986)). The AR model is very useful for the analysis of a stationary time series. However, from the statistical point of view, the main feature of the seismic signal is the nonstationarity. Although seismic waves are nonstationary, it might be reasonable to approximate it by an AR model on each properly divided time interval. The use of the locally stationary AR model (Ozaki and Tong (1975), Kitagawa and Akaike (1978)) was thus motivated and it was shown that it is actually useful for the detection of arrival time of P waves in noisy data (Yokota *et al.* (1981)). A significant merit of the time series method is that we can automatically determine the arrival time of the P waves by just looking for the time point that attains the minimum value of the AIC of the locally stationary AR model. The CPU time of this method used to depend on the number of data, the order of the AR model and on the number of models checked, i.e., the number of candidates of the arrival time. In our previous paper (Takanami and Kitagawa (1988)), a computationally efficient procedure for the detection of the arrival time of the P wave has been developed based on the univariate locally stationary AR model. However, it is well known that the additional information from S wave improves the quality of location (Buland (1976)). Further, to get elastic parameters such as Poisson's ratios (σ), the velocity of the S wave as well as that of P wave is required. Therefore, in this paper we consider the extension of the method developed in the previous paper, so that it can be applied to the automatic processing of the S wave.

The objective of this paper is three-fold. Firstly, we develop a computationally

efficient algorithm for the fitting of the multi-variate locally stationary AR model so that it can be applied to the on-line processing of seismic wave. The procedure is, in particular, useful for the automatic determination of the arrival time of S waves of microearthquakes ($1 \leq M < 3$) . Secondly, we will present a method of evaluating the posterior probability of the arrival time. The posterior probability will be useful for the estimation of the hypocenters of the earthquakes. The third objective of the paper is to demonstrate the usefulness of the proposed procedure by applying it to the foreshocks of the 1982 Urakawa-Oki Earthquakes. These data sets have been registered to the AISM Data Library and are released to the readers.

The plan of the paper is as follows. In Section 2, a procedure for the estimation of arrival time of a seismic wave is developed based on a multi-variate locally stationary autoregressive (MLSAR) model fitting. In Section 3, a computationally efficient procedure for MLSAR model fitting is developed. In Section 4, the posterior probability of the arrival time of seismic wave is derived by using the likelihood of the MLSAR models. Section 5 is devoted to empirical study where the proposed procedure is applied to the estimation of the arrival times of the seismic waves. Especially, the main focus is put on the estimation of the arrival time of the S wave.

2. Estimation of the arrival time and multivariate locally stationary AR modeling

Let $y_n = (y_{nE}, y_{nN}, y_{nU})^t$, $(n = 1, \ldots, N)$ be a three variate time series where y_{nE}, y_{nN} and y_{nU} express the east–west (E–W), north–south (N–S) and up–down (U–D, vertical) components of the seismograms, respectively. The characteristics of the series, e.g., the variances and the spectra, change over time due to the arrival of seismic waves such as the P wave or the S wave. However, it might be reasonable to assume that each of the seismogram before and after the arrival of the seismic wave are stationary and can be expressed by a single time series model. This will be verified by the time-varying spectrum analysis shown in the discussion. For a stationary time series, an autoregressive model usually fits well and allows computationally efficient procedure for the identification. Therefore we will use an autoregressive model for the modeling of each stationary subseries. In this modeling, the arrival time of the seismic wave, n_A, corresponds to the change point of the autoregressive model. In the estimation of the arrival time of P wave, the use of univariate time series has been considered reasonable, since the P wave is a compression wave and a dominant part of the movement appears in the vertical component. However, since the S wave is a shear type wave, for the estimation of the arrival time of S wave the analysis of the movement in the horizontal plane, namely of the east–west and north–south components, seems to be necessary. In view of the fact that even after the arrival of S wave, the coda of P wave remains and that S wave also induces the vertical motion, the use of two or three components seems to be desirable. We also have an anticipation that even for the detection of P wave, the analysis of three-variate time series will give more precise information about the arrival time. We are thus motivated to use a MLSAR model which consisted of the following two local models.

Background noise model

$$y_n = \sum_{i=1}^{m_1} A_{i1} y_{n-i} + w_{n1}, \qquad (n = 1, \ldots, n_B). \tag{2.1}$$

Here m_1, is the autoregressive order, A_{i1} is the $k \times k$ autoregressive coefficient matrix for i-lag component, and w_{n1} is the innovation sequence with mean 0 and the covariance matrix Σ_1. In our applications, k is typically 2 or 3 and $y_n = (y_{nE}, y_{nN})^t$ or $y_n = (y_{nE}, y_{nN}, y_{nU})^t$. This model expresses the dynamics of the background motion. It should be noted that in the detection of S wave, the coda of P wave together with the background motion are expressed by this "background noise" model.

Signal model

$$y_n = \sum_{i=1}^{m_2} A_{i2} y_{n-i} + w_{n2}, \qquad (n = n_B + 1, \ldots, N). \tag{2.2}$$

Here m_2, A_{i2} and w_{n2} are autoregressive order, autoregressive coefficient matrix and the innovation of the signal model, respectively. The covariance matrix of the innovation w_{n2} is denoted by Σ_2. This model expresses the dynamics of the seismic wave.

Assuming the arrival time $n_A \equiv n_B + 1$ and the orders of autoregressions, m_1 and m_2 to be known, the distribution of the time series is given by

$$y_n \sim N\left(\sum_{i=1}^{m_1} A_{i1} y_{n-i}, \Sigma_1\right) \qquad (n = 1, \ldots, n_B),$$

$$y_n \sim N\left(\sum_{i=1}^{m_2} A_{i2} y_{n-i}, \Sigma_2\right) \qquad (n = n_B + 1, \ldots, N). \tag{2.3}$$

Therefore, given the observations $y_1, \ldots, y_N$, the log-likelihood of the MLSAR model is given approximately as follows:

$$\begin{aligned} \ell(A_1, A_2, \Sigma_1, \Sigma_2) = -\frac{1}{2}\Bigg\{ & k(N - m_1)\log 2\pi + (n_B - m_1)\log|\Sigma_1| \\ & + (N - n_B)\log|\Sigma_2| \\ & + \sum_{n=m_1+1}^{n_B} w_{n1}^t \Sigma_1^{-1} w_{n1} + \sum_{n=n_B+1}^{N} w_{n2}^t \Sigma_2^{-1} w_{n2} \Bigg\} \end{aligned} \tag{2.4}$$

where, $A_1 = (A_{11}, \ldots, A_{m_1,1})$, $A_2 = (A_{12}, \ldots, A_{m_2,2})$ and from (2.1) and (2.2), w_{nj} are obtained by

$$w_{nj} = y_n - \sum_{i=1}^{m_j} A_{ij} y_{n-i} \qquad (j = 1, 2). \tag{2.5}$$

The maximum likelihood estimates of A_{ij} and Σ_j $(i = 1, \ldots, m_j;\ j = 1, 2)$ are approximately given by maximizing (2.4).

However, from the form of the log-likelihood function given in (2.4), it can be easily seen that the parameters of the background model and the signal model can be obtained by minimizing

$$(2.6)\qquad \begin{aligned} &(n_B - m_1)\log|\Sigma_1| + \sum_{n=m_1+1}^{n_B} w_{n1}^t \Sigma_1^{-1} w_{n1}, \\ &(N - n_B)\log|\Sigma_2| + \sum_{n=n_B+1}^{N} w_{n2}^t \Sigma_2^{-1} w_{n2}, \end{aligned}$$

respectively. A computationally efficient procedure for the fitting of these models will be shown in the next section. The fitted model can be evaluated by the AIC criterion (Akaike (1973)) defined by

$$(2.7)\qquad \mathrm{AIC} = -2\ell(\hat{A}_1, \hat{A}_2, \hat{\Sigma}_1, \hat{\Sigma}_2) + 2(\text{the number of estimated parameters}).$$

In the estimation of the arrival time, the crucial problem is the estimation of the dividing point, n_B. This point can be determined by finding the minimum of the AIC.

3. Computationally efficient procedure for multi-variate locally stationary AR model fitting

3.1 *Householder method for multivariate AR model fitting*

We will first briefly review the procedure for the fitting of multivariate AR model developed for the program MULMAR in TIMSAC-78 (Akaike *et al.* (1979)). The program has been widely used since then. However, the details of the algorithm has never been described elsewhere. Assume that three variate time series $\{y_1, \ldots, y_N\}$ is given and we are going to fit multivariate AR (MAR) model

$$(3.1)\qquad y_n = \sum_{n=1}^{m} A_i y_{n-i} + w_n, \qquad w_n \sim N(0, \Sigma).$$

It should be noted that although an algorithm for fitting three-variate AR model is shown here, it can be readily extended to a general k-variate time series. The main idea of the MULMAR is to use an autoregressive model with instantaneous response

$$(3.2)\qquad y_n = B_0 y_n + \sum_{n=1}^{m} B_i y_{n-i} + v_n, \qquad v_n \sim N(0, V).$$

Here it is assumed that the coefficient of the instantaneous response is of the form

$$(3.3)\qquad B_0 = \begin{bmatrix} 0 & 0 & 0 \\ b_{21} & 0 & 0 \\ b_{31} & b_{32} & 0 \end{bmatrix}$$

and that the covariance matrix V is of diagonal form:

$$V = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}. \tag{3.4}$$

Since

$$y_n = (I - B_0)^{-1} \sum_{i=1}^{m} B_i y_{n-i} + (I - B_0)^{-1} v_n, \tag{3.5}$$

this AR model with instantaneous response has a one to one correspondence to the ordinary AR model with the relation

$$\begin{aligned} A_i &= (I - B_0)^{-1} B_i, \\ \Sigma &= (I - B_0)^{-1} V (I - B_0)^{-t}. \end{aligned} \tag{3.6}$$

It should be noted that these two models, (3.1) and (3.2), have the same number of parameters.

The significant merit of the use of the AR model with the instantaneous response is that it can be estimated by independently fitting the univariate models for each of the three components. This can be justified as follows. Since the covariance matrix is of diagonal form,

$$N \log |V| + \sum_{n=1}^{N} v_n^t V^{-1} v_n = \sum_{i=1}^{3} \left\{ N \log \sigma_i^2 + \frac{1}{\sigma_i^2} \sum_{n=1}^{N} v_{ni}^2 \right\}, \tag{3.7}$$

where v_{ni} denotes the i-th element of v_n. Therefore, if the 3×3 matrix B_i is divided as

$$B_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix}, \tag{3.8}$$

the parameter set $\{b_{ij}, (i = 1, \ldots, m_1), \sigma_j^2\}$, for $j = 1, 2, 3$ can be estimated by minimizing

$$N \log \sigma_j^2 + \frac{1}{\sigma_j^2} \sum_{n=1}^{N} v_{nj}^2. \tag{3.9}$$

For any given b_{ij}, (3.9) is minimized when

$$\sigma_j^2 = \frac{1}{N} \sum_{n=1}^{N} v_{nj}^2 \tag{3.10}$$

and by substituting this estimate into (3.9), it can be seen that b_{ij} are obtained by minimizing

$$N \log \sigma_j^2 + N, \tag{3.11}$$

or equivalently by minimizing σ_j^2. This means that by using the special expression for the multivariate AR (MAR) model given in (3.2), the maximum likelihood estimates of the MAR model are obtained by solving the least squares problem for each of the three component. Further, the log-likelihood and AIC of the MAR model are obtained as the sum of AIC's of three component models.

We will next show an algorithm which can solve these three least squares problems quite efficiently. The least squares estimation of the MAR model can be realized by first making $(N-m) \times (3m+3)$ matrix

$$X = \begin{bmatrix} y_m^t & \cdots & y_1^t & y_{m+1}^t \\ y_{m+1}^t & \cdots & y_2^t & y_{m+2}^t \\ \vdots & \ddots & \vdots & \vdots \\ y_{N-1}^t & \cdots & y_{N-m}^t & y_N^t \end{bmatrix} \tag{3.12}$$

and reduce this matrix to an upper triangular form by an orthonormal transformation (i.e., Householder transformation, Golub (1965), Sakamoto *et al.* (1986))

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1,3m+3} \\ & \ddots & \vdots \\ & & s_{3m+3,3m+3} \\ & \mathrm{O} & \end{bmatrix}. \tag{3.13}$$

The $(3m+1) \times (3m+1)$ upper left triangular matrix of S contains sufficient information for the fitting of the model for the first component (e.g., E–W component in this case). In particular, the innovation variance $\sigma_1^2(j)$ and AIC(j) of the j-th order model

$$y_{nE} = \sum_{i=1}^{j} b_{i1} y_n + w_{nE}, \tag{3.14}$$

where $b_{i1} = (b_i(1,1), b_i(1,2), b_i(1,3))$ and $y_n = (y_{nE}, y_{nN}, y_{nU})^t$, are obtained by (Kitagawa and Akaike (1980), Sakamoto *et al.* (1986))

$$\begin{aligned} \sigma_1^2(j) &= \frac{1}{N-m} \sum_{i=3j+1}^{3m+1} s_{i,3m+1}^2 \qquad (j = 0, 1, ..., m), \\ \mathrm{AIC}_1(j) &= (N-m) \log \sigma_1^2(j) + 2(3j+1). \end{aligned} \tag{3.15}$$

Incidentally, the regression coefficients of the E–W component model with order j are obtained by solving the linear equation

$$\begin{bmatrix} s_{11} & \cdots & s_{1,3j} \\ & \ddots & \vdots \\ \mathrm{O} & & s_{3j,3j} \end{bmatrix} \begin{bmatrix} b_{11}(1,1) \\ \vdots \\ b_{j1}(1,3) \end{bmatrix} = \begin{bmatrix} s_{1,3m+1} \\ \vdots \\ s_{3j,3m+1} \end{bmatrix}. \tag{3.16}$$

However, it should be emphasized that for the present purpose of the estimation of the arrival time, only the AIC value of the best model is necessary and we do not actually solve this linear equation.

For the computation of the AIC of the second (N–S) component model,

$$y_{nN} = b_{02}(2,1)y_{nE} + \sum_{i=1}^{j} b_{i2}y_{n-i} + w_{nN}, \tag{3.17}$$

we first transform the matrix (3.13) to the following form

$$S' = \begin{bmatrix} s'_{11} & \cdots & s'_{1,3m} & s'_{1,3m+1} & s'_{1,3m+2} & s'_{1,3m+3} \\ s'_{21} & \cdots & s'_{2,3m} & & s'_{2,3m+2} & s'_{2,3m+3} \\ & \ddots & \vdots & & \vdots & \vdots \\ & & s'_{3m+1,3m} & & s'_{3m+1,3m+2} & s'_{3m+1,3m+3} \\ & & & & s'_{3m+2,3m+2} & s'_{3m+2,3m+3} \\ & & & & & s'_{3m+3,3m+3} \\ & & \mathrm{O} & & & \end{bmatrix}. \tag{3.18}$$

This can be done by using an appropriate Householder transformation with only a little additional computations. Then the upper left $(3m+2)\times(3m+2)$ submatrix of S' contains sufficient information for the fitting of the regression model for the second component which has an instant response from the first (E–W) component. The residual variance and the AIC of the j-th order model is given by

$$\sigma_2^2(j) = \frac{1}{N-m}\sum_{i=3j+2}^{3m+2}(s'_{i,3m+2})^2, \tag{3.19}$$

$$\mathrm{AIC}_2(j) = (N-m)\log\sigma_2^2(j) + 2(3j+2).$$

It should be noted that the $(3m+1)$-st column of the matrix S which was used as the vector of objective variable in fitting the model for the first component, is now used as the vector of a regressor corresponding to the instantaneous response from the first variable.

Similarly, the model for the third variable (U–D component) can be obtained from

$$S'' = \begin{bmatrix} s''_{11} & \cdots & s''_{1,3m} & s''_{1,3m+1} & s''_{1,3m+2} & s''_{1,3m+3} \\ s''_{21} & \cdots & s''_{2,3m} & & s''_{2,3m+2} & s''_{2,3m+3} \\ s''_{31} & \cdots & s''_{3,3m} & & & s''_{3,3m+3} \\ & \ddots & \vdots & & & \vdots \\ & & s''_{3m+2,3m} & & & s''_{3m+2,3m+3} \\ & & & & & s''_{3m+3,3m+3} \\ & & \mathrm{O} & & & \end{bmatrix}, \tag{3.20}$$

$$\sigma_3^2(j) = \frac{1}{N-m}\sum_{i=3j+3}^{3m+3}(s''_{i,3m+3})^2,$$

$$\mathrm{AIC}_3(j) = (N-m)\log\sigma_3^2(j) + 2(3j+3).$$

The AIC of the original MAR model is then given by

$$\text{AIC} = \min_j \text{AIC}_1(j) + \min_j \text{AIC}_2(j) + \min_j \text{AIC}_3(j). \tag{3.21}$$

By using the Householder transformation, we can further fit a more sophisticated model which, for example, allows that some part of the coefficients are zeros. The program for such models is given in the subroutine MARFIT of TIMSAC-78 (Akaike *et al.* (1979)). However, this will not be necessary for the present purpose.

3.2 *Augmentation of the data*

In the previous section, we showed an algorithm for the fitting of MAR model. We will now present a method of modifying the AR model when the augmented data set $\{y_1, \ldots, y_N, y_{N+1}, \ldots, y_{N+p}\}$ was obtained. Here $p \geq 1$ is the number of the new data. This can be performed by first organizing the following $(3m + 3 + p) \times (3m + 3)$ matrix R

$$R = \begin{bmatrix} & & S & \\ & & & \\ y_N^t & \cdots & y_{N+1-m}^t & y_{N+1}^t \\ \vdots & \ddots & \vdots & \vdots \\ y_{N+p-1}^t & \cdots & y_{N+p-m}^t & y_{N+p}^t \end{bmatrix} \tag{3.22}$$

with S being the upper triangular matrix given in (3.13) and then reducing to an upper triangular form. It should be noted that due to the orthogonality of the Householder transformation, non-zero elements of the Householder reduced form of R is one and the same as the upper triangular form obtained by the Householder reduction of the following $(N + p - m) \times (3m + 3)$ matrix:

$$X = \begin{bmatrix} y_m^t & \cdots & y_1^t & y_{m+1}^t \\ \vdots & \ddots & \vdots & \vdots \\ y_{N-1}^t & \cdots & y_{N-m}^t & y_N^t \\ y_N^t & \cdots & y_{N+1-m}^t & y_{N+1}^t \\ \vdots & \ddots & \vdots & \vdots \\ y_{N+p-1}^t & \cdots & y_{N+p-m}^t & y_{N+p}^t \end{bmatrix}. \tag{3.23}$$

This means that the upper triangular matrix which is necessary to fit AR models to the augmented data set can be obtained with only a few additional computations. By applying the same method as presented in the previous section to this matrix, we can get the AIC values of the best AR model fitted to this augmented data set.

3.3 *Fitting locally stationary AR model*

In order to determine the arrival time of a seismic wave based on the MLSAR model, we have to compare the goodness of the fit of many MLSAR models obtained by assuming various arrival time.

We assume that we have N observations and that n_B is $n_0 \leq n_B \leq n_1$. Note that the arrival time n_A is given by $n_A = n_B + 1$. It is also assumed that the required resolution is p points, thus we have to fit models for each dividing points $n_0, n_0 + p, \ldots, n_0 + \ell p \equiv n_1$. Therefore, we have to fit $\ell + 1$ different MLSAR models. In this subsection we shall present a computationally efficient procedure for the fitting of many MLSAR models based on the Householder method for MAR model fitting and the augmentation of the data. The procedure is constructed as follows:

1. Fit an MAR model to the data $\{y_1, \ldots, y_{n_0}\}$ by the method presented in Subsection 3.1. AIC_0^N denotes the AIC of the best MAR model fitted to the data.

2. For $i = 1, \ldots, \ell$, successively augment the upper triangular matrix obtained in step 1 by the additional data $\{y_{n_0+(i-1)p+1}, \cdots, y_{n_0+ip}\}$ and find the minimum value of AICs'. This value is denoted by AIC_i^N.

3. Similarly to step 1, fit an MAR model to the data $\{y_{n_1+1}, \ldots, y_N\}$. The minimum value of the AIC for this data set is denoted by AIC_ℓ^S.

4. For $i = \ell - 1, \ldots, 0$, successively augment the upper triangular matrix obtained in step 3, by the additional data $\{y_{n_0+ip+1}, \cdots, y_{n_0+(i+1)p}\}$. The minimum value of the AIC of the MAR models fitted to the data set is denoted by AIC_i^S.

5. Obtain the AIC of MLSAR model which assume the dividing point to be $n_B = n_0 + ip$ by

$$\mathrm{AIC}_i = \mathrm{AIC}_i^N + \mathrm{AIC}_i^S \qquad (i = 0, \ldots, \ell). \tag{3.24}$$

6. Find the minimum of $\mathrm{AIC}_0, \ldots, \mathrm{AIC}_\ell$. If AIC_i is the minimum, then $n_A = n_0 + ip + 1$ is our estimate of the arrival time of the seismic signal.

3.4 *The number of necessary operations*

For the Householder transformation of a $n \times k$ matrix to an upper triangular form, the amount of necessary multiplications (and additions) is approximately evaluated as (Golub (1965))

$$\sum_{i=1}^{k} (n + 1 - i)(k + 1 - i) \approx \frac{1}{2} nk^2. \tag{3.25}$$

Therefore, the number of necessary operations for fitting an ordinary 3-V MAR model to entire data set is approximately $9Nm^2/2$, and fitting 3-V MLSAR model without recursive formula shown in Subsection 3.3 requires

$$\sum_{i=0}^{\ell} \frac{1}{2}(n_0 + ip)(3m+3)^2 + \sum_{i=0}^{\ell} \frac{1}{2}(N - n_0 - ip)(3m+3)^2 \approx \frac{9}{2p} Nm^2(n_1 - n_0). \tag{3.26}$$

On the other hand, the necessary operations for the Householder transformation of the matrix (3.22) is

$$\sum_{i=1}^{k}(p+1)(k+1-i) \approx \frac{1}{2}pk^2. \tag{3.27}$$

Therefore, the total amount of multiplications (and additions) for the fitting of all possible 3-V MLSAR models by the present method is

$$\begin{aligned}\frac{1}{2}n_0(3m+3)^2 + \sum_{i=1}^{l}\frac{1}{2}p(3m+3)^2 + \frac{1}{2}(N-n_1)(3m+3)^2 \\ + \sum_{i=1}^{l}\frac{1}{2}p(3m+3)^2 \cong \frac{9}{2}Nm^2 + \frac{9}{2}m^2(n_1-n_0) \le 9Nm^2.\end{aligned} \tag{3.28}$$

This means that the total number of computation for the fitting of 3-V MLSAR model by the proposed method is less than $2p/N$ of the conventional method, and is less than twice of that for the fitting single 3-V MAR model.

Incidentally, fitting a 1-V LSAR model by the same recursion requires $(1/2)Nm^2 + (1/2)m^3(n_1 - n_0)$. Summarizing, the necessary computing time by the present method is only twice of that for ordinary 3-V MAR model, and is 9 times of that for 1-V LSAR model.

4. Posterior probabilities of the arrival time

So far, it has been shown that we can determine the arrival time of the seismic wave by using MLSAR model and AIC and that we can develop a computationally efficient algorithm for the computation of the AIC values.

In this section, we will present a method that allows to use more fully the information contained in the AIC values. In Akaike (1979), it was shown that $\exp\{-(1/2)\mathrm{AIC}\}$ can be considered as an appropriate definition of the likelihood of the model whose parameters are estimated by the maximum likelihood method. In our case

$$p(y \mid j) = \exp\left\{-\frac{1}{2}\mathrm{AIC}_j\right\} \tag{4.1}$$

is the likelihood of the MLSAR model which assumes that $n_0 + pj + 1$ is the arrival time.

Therefore, if the prior distribution of the arrival time is given, then the posterior distribution of the arrival time can be obtained by

$$p(j \mid y) = \frac{p(y \mid j)p(j)}{\sum_j p(y \mid j)p(j)}. \tag{4.2}$$

In the actual analysis shown in the next section, the uniform prior over the interval is used. It seems more reasonable to put more weight on the center of the

interval. However, since the likelihood, $p(y \mid j)$, usually takes significant values only at limited interval, only the local behavior of the prior is influential to the posterior probability. Therefore as long as very smooth functions are used, the choice of the prior is not so important in the present problem.

One of the most important use of the estimated arrival time is the determination of the hypocenter of the earthquake. Conventionally, this has been done by the weighted least squares method. However, the use of the likelihoods of various MLSAR models or the posterior distribution of the arrival time may yield more precise inference on the hypocenter by using the maximum likelihood method or by a Bayesian modeling. This is a subject of future study.

5. Empirical study

5.1 *The data*

The data sets we analyze in this paper are seismograms of four foreshocks of the 1982 Urakawa-Oki Earthquake ($M_{JMA} = 7.1$, 11:32 on March 21, 1982, Urakawa, Hokkaido, Japan) recorded at six stations of the Research Center for Earthquake Prediction (RCEP) of Hokkaido University (Suzuki *et al.* (1986)). These foreshocks occurred a few hours prior to the main shock. Table 1 summarizes the source parameters of these earthquakes. These parameters, the origin times and the locations of hypocenters, are estimated by the routine procedure of the RCEP. The locations of the epicenters of the four foreshocks and six stations are shown in Fig. 1. The epicenters are closely located compared with the spread of the stations.

Table 1. Source parameters of the four foreshocks and the main shock.

Date	Time	Longitude	Latitude	Depth	M	Code name	Start time
Mar. 21, 1982	07:45 53.0	142.557	42.158	31.0	1.9	1F	7:45:31.46
Mar. 21, 1982	08:42 52.0	142.555	42.131	26.0	2.0	2F	8:42:30.44
Mar. 21, 1982	08:49 20.7	142.561	42.158	33.7	2.1	3F	8:48:59.46
Mar. 21, 1982	09:33 15.0	142.574	42.133	31.1	2.3	4F	9:32:54.55
Mar. 21, 1982	11:32 05.7	142.600	42.150	40.0	7.1	—	—

At each station, the East–West (E–W), the North–South (N–S) and the Up–Down (U–D, vertical) components of the ground velocity signal were measured by seismometers with a natural frequency of 1 Hz. They are digitized by an 8 bit nonlinear AD converter at the rate of 92.3 samples per second (2,400 bps/26 bit), and transformed to PCM (pulse coded modulation) data. Each record has 7740 observations (83.85 seconds time span).

Figure 2 shows the seismograms of the 1F (the first foreshock). The station Erimo (ERM) is located 0.5 km from the shoreline of the Pacific and the epicentral distances are about 50 km. The seismograms are strongly affected by the attenuation in the crustal structure along the ray path, and the ratios of the P wave signal

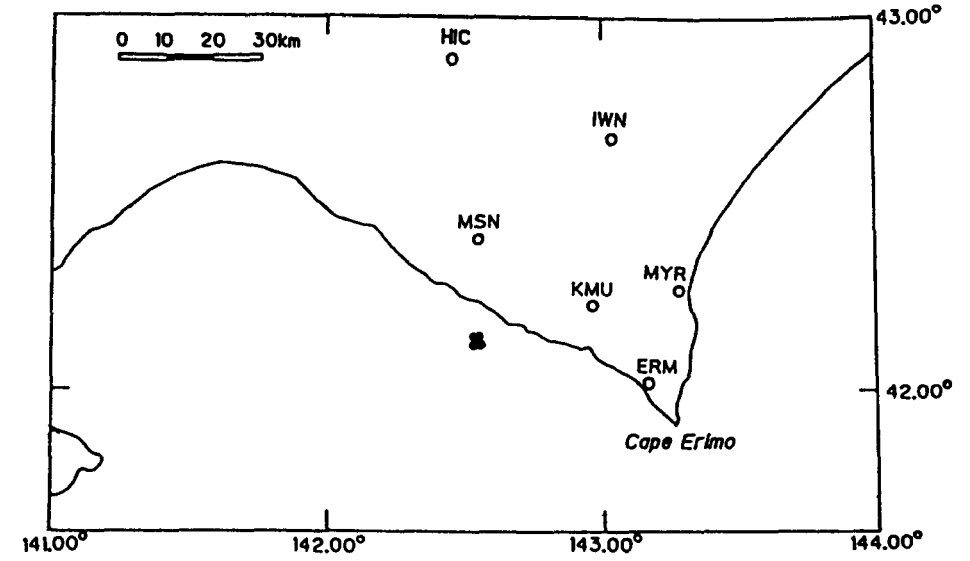

Fig. 1. Locations of the epicenters of four microearthquakes (•) and six stations (o).

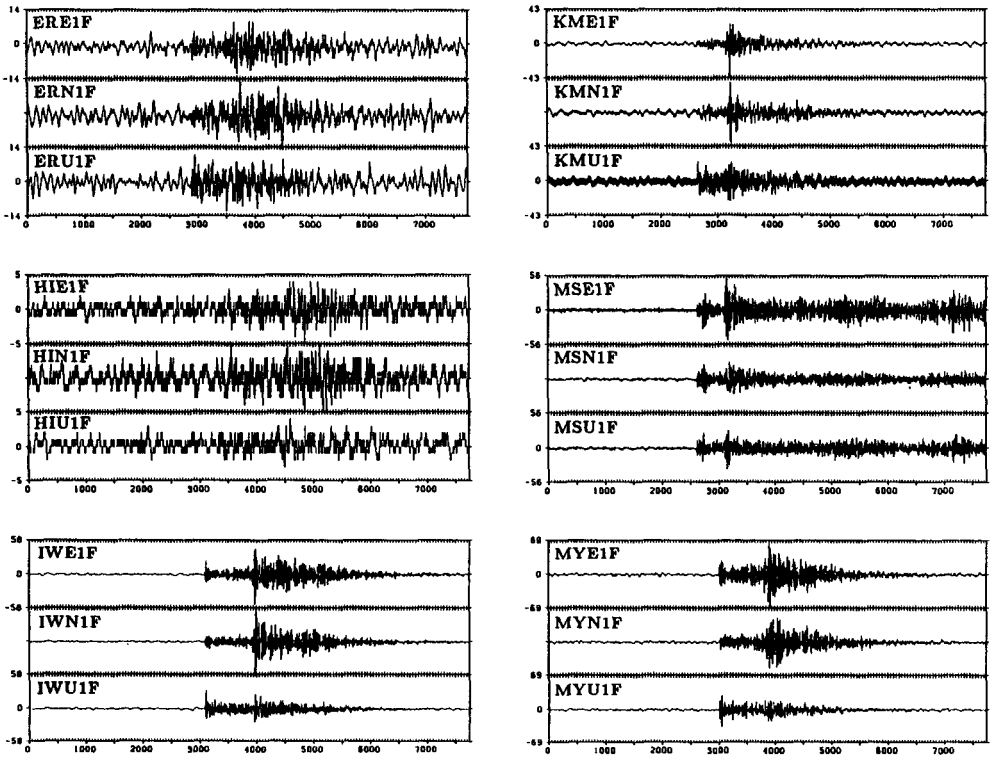

Fig. 2. Seismograms of the first foreshock (1F) observed at six stations, ERM, HIC, IWN, KMU, MSN and MYR.

to the background noise are reduced to about 0.5 (Takanami (1982)). Therefore, in the routine work, it is very hard to determine the arrival time of such weak P waves recorded on the paper chart.

The station Hidaka (HIC) is located 80 km from the epicenters of the foreshocks and is apart from towns and road and the amplitude of the background noise is usually less than two LSB (the least significant bit). The seismic signal observed at HIC was also very weak and almost equal to one LSB of the digital signal and are comparable to those of the background noise.

The station Iwanai (IWN) is located about 65 km from the epicenters. The observations at this station has higher signal to noise ratios.

The station Kamikineusu (KMU) is located about 27 km from the epicenters. On this occasion, the seismograms obtained at KMU were contaminated by a strong electronic hum noise with frequency of 50 Hz.

The station Misono (MSN) is located about 30 km from the epicenters and is near a load. Therefore, the seismograms obtained in this station occasionally suffer from the traffic noises, e.g., MSN-3F.

The station Moyori (MYR) is about 60 km from the epicenters. Good seismograms were obtained from this station.

5.2 *Detection of P wave*

Although it is not of our primary concern, we will first consider the estimation of the arrival time of P wave which is supposed to be much easier than that of S wave. LSAR model and MLSAR model presented in the previous sections have been applied to the microearthquake data presented in the previous subsection.

Figure 3.1 shows a part of MYR–2F data. Figure 3.2 shows the results of the LSAR model for this data set. Three figures from the top of this figure show the plot of AIC values versus assumed arrival time when LSAR models are fitted to E–W, N–S and U–D components. The minimum AIC estimates of the arrival time obtained by the LSAR model are $n_A = 3001$, 3011 and 2992 for the E–W, the N–S and the U–D components, respectively. AIC values of the estimated models are also shown in the figure. Judging from the original seismogram, the point 2992 seems to be the most reasonable estimate of the arrival time of the P wave. The estimates from E–W and N–S components are 9 points (0.098 seconds) and 19 points (0.206 seconds) later than that estimated from the vertical motion, respectively. This phenomenon can be typically seen in the estimation of the arrival time of P wave (see also Table 2). Further, the slope of the AIC value before and after the arrival time is the steepest for the U–D component. This means that the vertical component of the seismogram has more precise information about the arrival time of P wave than the other two components. This can be understand from the fact that the P wave is a compressional wave and the dominant part of the movement appears in the vertical motion.

The bottom one in Fig. 3.2 shows the trace of AIC of the 3-V MLSAR model. The minimum of the AIC, 1487, occurs at $n_A = 2992$ which is exactly the same as the one obtained by LSAR model from the U–D component. However, the slope of the AIC value is steeper than those of LSAR models indicating that the estimate by the MLSAR model is more reliable than the ones by LSAR models. As can be

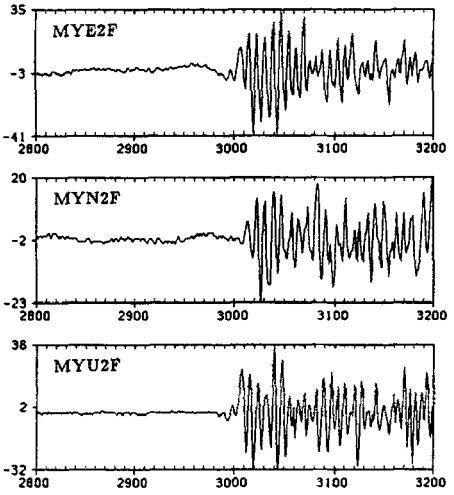

Fig. 3.1

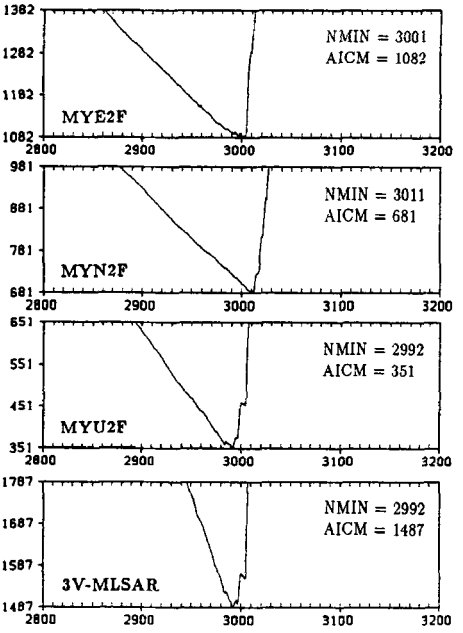

Fig. 3.2

Fig. 3.1. A part of seismogram of 2F observed at MYR. From top to bottom, E–W, N–S and U–D components.
Fig. 3.2. AIC values of LSAR and MLSAR models for the estimation of P waves of the MYR-2F data. From top to bottom, LSAR models for E–W, N–S and U–D components and the 3-V MLSAR model.

seen in Table 2, for about a half cases LSAR model for U–D component and the 3-V MLSAR model yielded the same estimates.

We will next show cases when the LSAR and the MLSAR models yield different estimates. In all cases except for HIC-3F and MSN-3F, the estimates by the MLSAR model are later than the ones by the LSAR models. Figure 4.1 shows the ERM-4F data. Figure 4.2 shows the traces of AIC of three LSAR models and 3-V MLSAR model for ERM-4F data. The estimate by the MLSAR model is 4 points (0.043 seconds) later than the one by LSAR model for U–D component. By a precise examination of the original record, it can be seen that the LSAR model for U–D component has an ability to detect the slight change of the slope of the data which is probably caused by the frequency characteristic of the seismometers. On the other hand, the MLSAR model yields more conservative estimates.

However, even when the signal to noise ratios are very low, such as the case of HIC, the MLSAR model yields a reasonable estimate. Later, by the P–S plot of the estimated arrival time, it becomes clear that in this case the estimates by the MLSAR model are more reasonable.

As a conclusion, for the detection of P wave, LSAR model is very sensitive to the slight change of the characteristics of the series caused by the P wave and that the estimates by the MLSAR model can be used to check these estimates.

Table 2. Estimated arrival time of P waves. The first column shows the code-names of the data. The second to fourth columns show the estimates by the LSAR models. The fifth column shows the results by the 3-V MLSAR models. The last column shows the estimates obtained from the sum of AICs' of three LSAR models.

Code name	E-W	N-S	U-D	3-D	E+N+U
ERM-1F	2881	2869	2867	2867	2866
ERM-2F	2856	2865	2857	2864	2857
ERM-3F	2870	2873	2846	2854	2854
ERM-4F	2791	2792	2782	2786	2787
HIC-1F	3393	3399	3320	3399	3399
HIC-2F	3553	3364	3396	3396	3396
HIC-3F	3479	3406	3318	3317	3317
HIC-4F	3340	3310	3278	3300	3300
IWN-1F	3079	3079	3074	3077	3077
IWN-2F	3076	3083	3076	3076	3076
IWN-3F	3076	3076	3075	3076	3076
IWN-4F	2980	2991	2986	2986	2986
KMU-1F	2636	2637	2622	2626	2624
KMU-2F	2638	2638	2622	2622	2635
KMU-3F	2630	2637	2614	2615	2615
KMU-4F	2540	2553	2518	2531	2531
MSN-1F	2619	2616	2609	2612	2613
MSN-2F	2623	2617	2599	2599	2599
MSN-3F	2679	2636	2616	2615	2615
MSN-4F	2558	2534	2514	2533	2554
MYR-1F	2999	3014	2986	2995	2995
MYR-2F	3001	3011	2992	2992	2992
MYR-3F	2988	2996	2986	2988	2987
MYR-4F	2909	2916	2902	2902	2902

5.3 *Detection of S wave*

We will examine the advantage of the use of the MLSAR model for determining the arrival time of S wave. The LSAR model and the MLSAR model used in the previous subsection are applied to a part of the three components seismograms where S waves presumably exist.

Figure 5.1 shows IWN-1F data. Figure 5.2 shows the results of LSAR model analyses. It can be seen that the minimum AIC estimates of the arrival time from the E–W, the N–S and the U–D component are 3929, 3906 and 3958, respectively. The discrepancy between the estimated arrival time by the three components indicates that it is difficult to determine the arrival time of S-wave from only one component of the seismogram. Moreover the shapes of AIC traces are much gentler than the case of P-waves suggesting the difficulty of the determination of the arrival time of S wave.

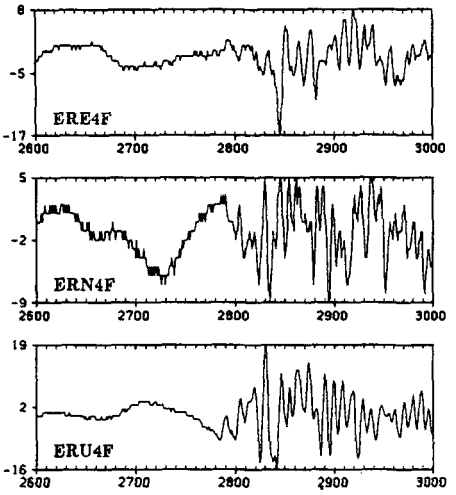

Fig. 4.1

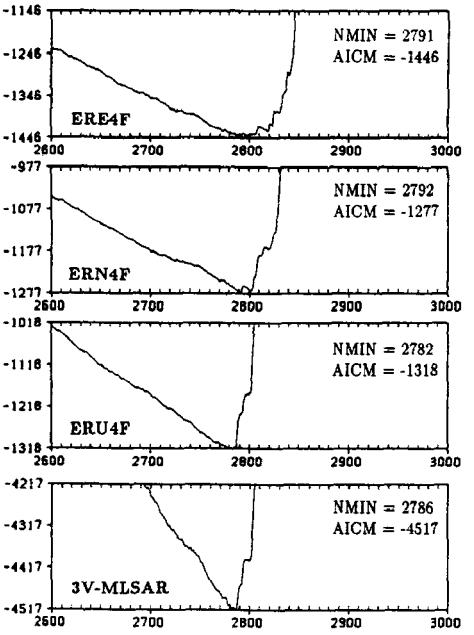

Fig. 4.2

Fig. 4.1. A part of seismogram of 4F observed at ERM.
Fig. 4.2. AIC values of LSAR and MLSAR models for the estimation of P wave of the ERM-4F data.

The last two panels of Fig. 5.2 shows the results by MLSAR models. In this case, the estimates by the 2-V MLSAR model and the 3-V MLSAR model coincide and are identical to the one by the N–S component.

Figures 6.1 and 6.2 show the results for MSN-1F. The estimated arrival time by 2-V and 3-V MLSAR models coincide and are identical to the estimate of the LSAR model obtained from the E–W component. However, the slope of AIC values in the neighborhood of its minimum is steeper than the one by LSAR model. Many local minima are found on each AIC curves of LSAR models as has been inferred from the behavior of the particle motion of S waves (Takanami (1990)). Therefore, it is usually hard to precisely determine the arrival time by using a single trace, especially only from the vertical component of the seismogram as seen in Fig. 6.2.

Figures 7.1 and 7.2 show the data and the results for MYR-2F. In this case, the left half of the trace of AIC of the LSAR models are almost flat indicating that this estimate is not reliable. The estimate by the LSAR model for E–W component is the earliest one. However, in the later analysis, it can be seen that this is not a good estimate. Even in this case, the AIC of the 2-V MLSAR model has clear minimum. The estimate only from U–D component is more than 100 points (about 1.33 seconds) earlier than these estimates. This clearly indicates that the U–D component is inadequate for the detection of S wave. This can be understood from Fig. 2, where in some cases such as ERU, KMU and MYU, the presence of the S wave is invisible. Figure 7.3 shows the posterior distributions

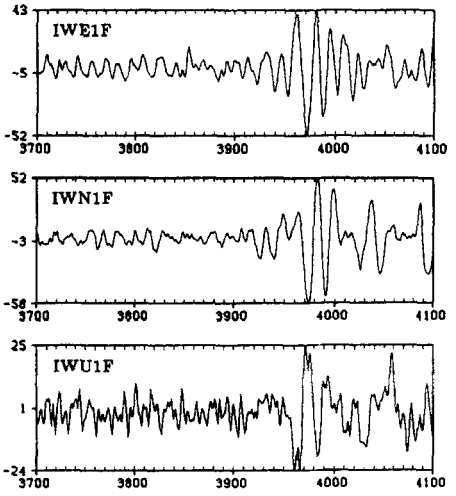

Fig. 5.1

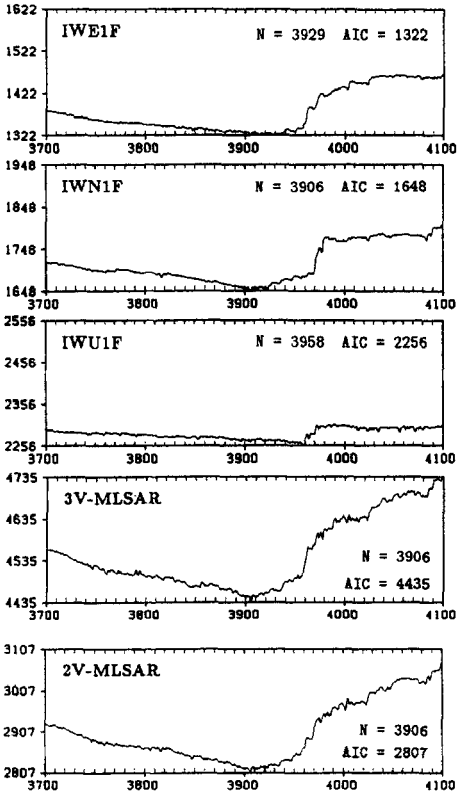

Fig. 5.2

Fig. 5.1. A part of seismogram of 1F observed at IWN. From top to bottom, E–W, N–S and U–D components.
Fig. 5.2. AIC values of LSAR and MLSAR models for the estimation of S waves of the IWN-1F data. From top to bottom, LSAR models for E–W, N–S and U–D components, 3-V MLSAR model and 2-V MLSAR model.

of the arrival time obtained by these models. The posterior distribution obtained by the LSAR model is distributed over a wide region. Whereas the one by the MLSAR model is concentrated on $n_A = 3812$.

It is typically seen for many of the seismograms (see Table 3) that one of the horizontal components (i.e. the minimum of estimates from E–W and N–S components) coincides with the ones by MLSAR models and that the estimate by U–D component is slightly earlier than this estimate or completely different.

Figures 8.1 and 8.2 show the analysis of ERM-2F data. All of three traces of AIC values obtained by LSAR models are flat and none of them yields reliable estimates. In this case, the AIC of the two MLSAR models have reasonable minima and yield the same estimate. They are quite different from all of the three LSAR models (see Table 3).

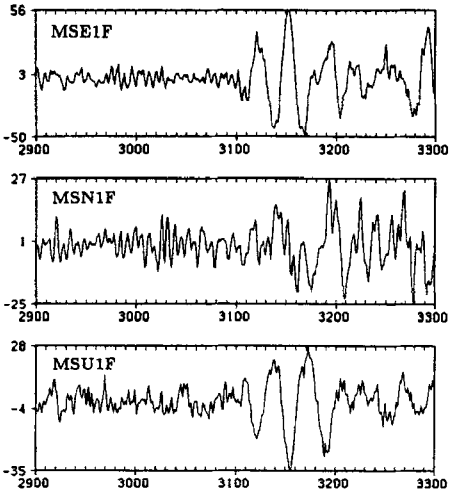

Fig. 6.1

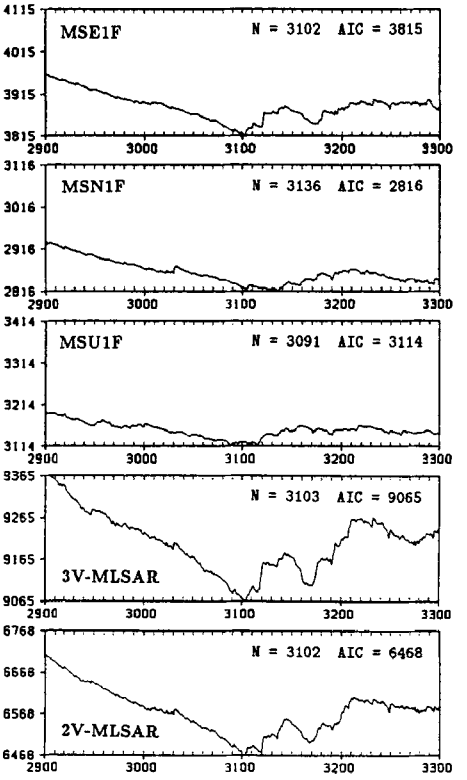

Fig. 6.2

Fig. 6.1. A part of seismogram of 1F observed at MSN. From top to bottom, E–W, N–S and U–D components.
Fig. 6.2. AIC values of LSAR and MLSAR models for the estimation of S waves of the MSN-1F data. From top to bottom, LSAR models for E–W, N–S and U–D components and 3-V and 2-V MLSAR models.

6. Discussion

Using the start time of the record, t_S given in Table 1 and the estimated arrival time point, n_A, given in Table 2 and Table 3, we can get the arrival time by

$$t_A = t_S + n_A \frac{26}{2400}. \tag{6.1}$$

Therefore, by denoting the arrival times of the P wave and S wave by t_A^P and t_A^S, the travel time of these waves are given by $t_A^P - t_0$ and $t_A^S - t_0$, respectively. Figure 9.1 shows the P wave travel time versus S wave travel time plot. Since the hypocenters of four foreshocks are very closely located, four points obtained from the same station should locate closely each other. Further, if the ratio of the velocities of P wave and S wave is a constant over the region, these points obtained for various stations lie on a straight line.

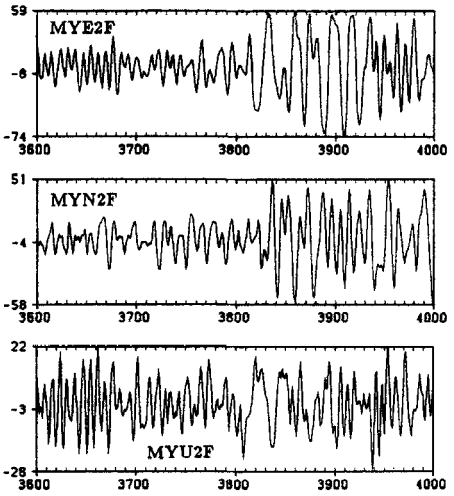

Fig. 7.1

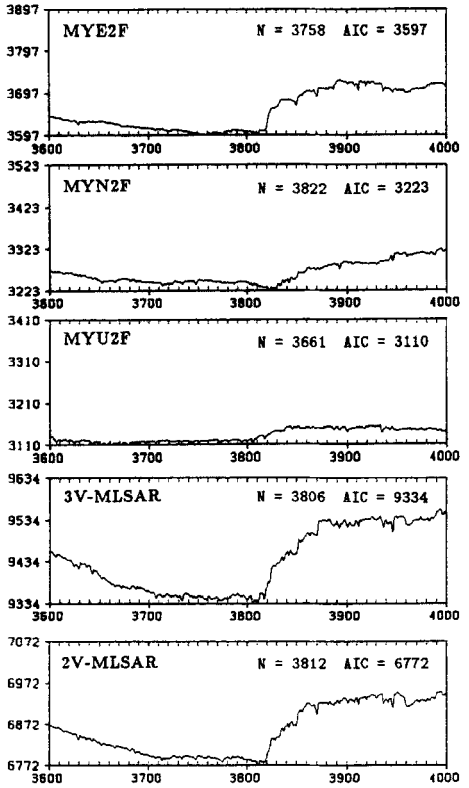

Fig. 7.2

Fig. 7.1. A part of seismogram of 2F observed at MYR. From top to bottom, E–W, N–S and U–D components.
Fig. 7.2. AIC values of LSAR and MLSAR models for the estimation of S waves of the MYR-2F data. From top to bottom, LSAR models for E–W, N–S and U–D components and 3-V and 2-V MLSAR models.

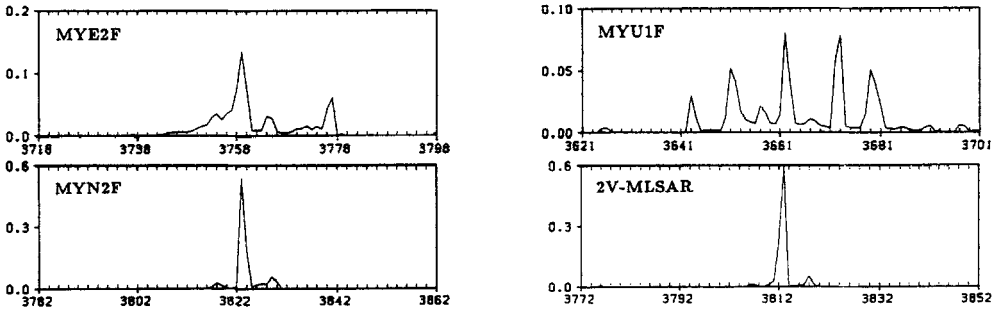

Fig. 7.3. Posterior probabilities of the S arrival time for MYR-2F data. LSAR models for E–W (top left), N–S (bottom left) and U–D (top right) components and the 2-V MLSAR model (bottom right).

Table 3. Estimated arrival time of S waves.

Code name	E-W	N-S	U-D	E+N+U	3-D	E+N	2-D
ERM-1F	3582	3565	3423	3500	3497	3582	3481
ERM-2F	3622	3515	3733	3577	3577	3577	3576
ERM-3F	3538	3635	3466	3537	3537	3538	3538
ERM-4F	3382	3553	3623	3543	3458	3358	3458
HIC-1F	4462	4484	4376	4484	4505	4484	4484
HIC-2F	4501	4600	4507	4501	4501	4501	4501
HIC-3F	4477	4423	4426	4423	4612	4602	4423
HIC-4F	4399	4532	4572	4540	4533	4540	4421
IWN-1F	3929	3906	3958	3906	3906	3906	3906
IWN-2F	3927	3911	3898	3919	3887	3919	3919
IWN-3F	3947	3925	3880	3947	3925	3925	3925
IWN-4F	3855	3864	3843	3843	3843	3850	3854
KMU-1F	3156	3174	3116	3156	3117	3156	3156
KMU-2F	3154	3164	3218	3154	3149	3154	3154
KMU-3F	3130	3168	3207	3146	3130	3146	3130
KMU-4F	3089	3100	3176	3093	3089	3093	3089
MSN-1F	3102	3136	3091	3102	3103	3102	3102
MSN-2F	3106	3074	3108	3105	3067	3105	3106
MSN-3F	3042	3002	3015	3066	3100	3066	3097
MSN-4F	3014	3002	3033	3032	3033	3002	3002
MYR-1F	3819	3843	3939	3812	3812	3816	3816
MYR-2F	3758	3822	3661	3805	3806	3812	3812
MYR-3F	3775	3763	3756	3778	3778	3763	3781
MYR-4F	3758	3757	3501	3758	3758	3758	3758

However, in Fig. 9.1, these points look rather scattered. To get finer relation between P and S travel time, we first refined the origin times of the foreshocks given in Table 1. To do that, we compute the mean of the four travel times and then compute the deviations from the means. Then the mean of these deviations at five stations, excluding HIC, gives the bias of the origin time of each foreshock. Figure 9.2 shows the same plot as Fig. 9.1 obtained by correcting for the bias of the origin time. Here the arrival time, n_A, is obtained by the MLSAR model. It can be seen that, four points corresponding to the same station, mostly locate closer than Fig. 9.1. This shows that by the above method, better estimates of origin times were obtained. In these figures, horizontal variation corresponds to the estimation error of P wave, whereas the vertical variation does that of the S-wave. Comparing these two figures, it can be seen that the points by MLSAR models with the correction of the origin time are much closely located.

One way to simplify the present procedure is to use the sum of the AICs' of three LSAR models obtained by fitting to each of E–W, N–S and U–D components.

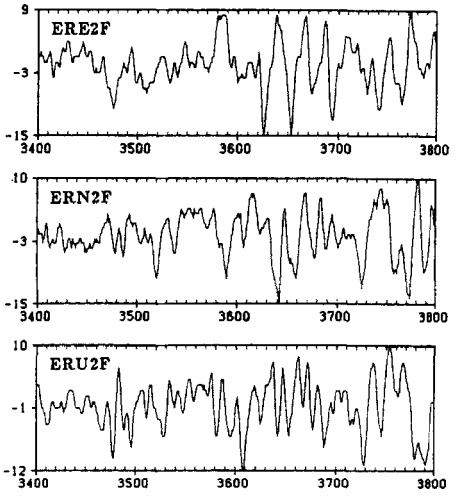

Fig. 8.1

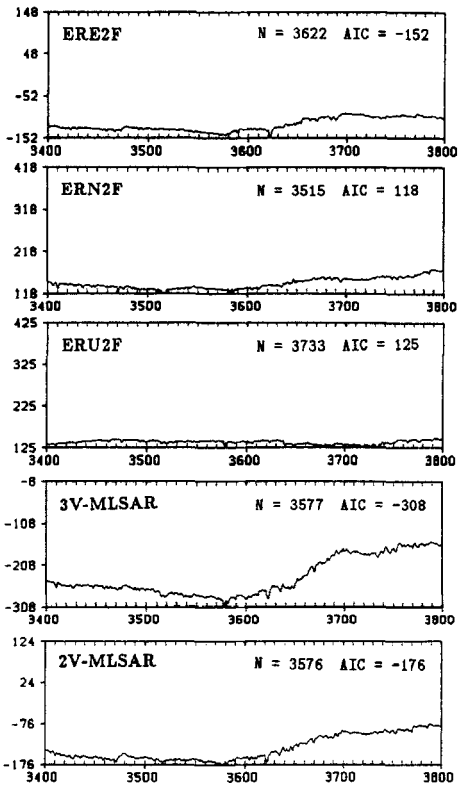

Fig. 8.2

Fig. 8.1. A part of seismogram of 2F observed at ERM. From top to bottom, E–W, N–S and U–D components.
Fig. 8.2. AIC values of LSAR and MLSAR models for the estimation of S waves of the ERM-2F data. From top to bottom, LSAR models for E–W, N–S and U–D components and 3-V and 2-V MLSAR models.

This is equivalent to assume that these three components are independent. The estimated arrival times by the method are shown in the extreme right column of the Tables 2 and 3. In all cases, the sum of three (or two) AICs' are significantly larger than the AIC of the MLSAR model. This indicates that the series are actually not independent. However, as can be seen in Tables 2 and 3, for many of the series, the minimum points obtained from these two models coincide. Therefore, at least for the seismograms with high signal to noise ratios, the use of this simplified procedure might be reasonable.

It might be interesting to compare the estimates shown in Tables 2 and 3 with the ones obtained by experienced persons. Figure 10.1 shows the histograms of the P arrival times read from the seismograms by 38 researchers of RCEP or students of Geophysical department of several universities. The mid points of the histograms show the P arrival times estimated by the LSAR models for U–D components. The number of unanswered person is also shown in the right margin of each histogram. It can be seen that the estimated P arrival times are distributed around the center

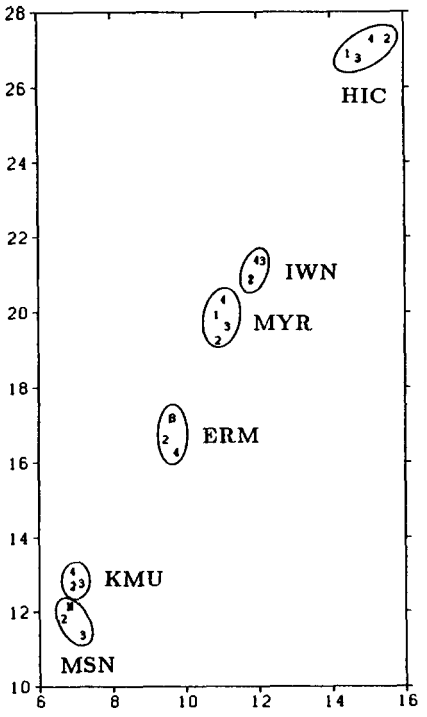

Fig. 9.1

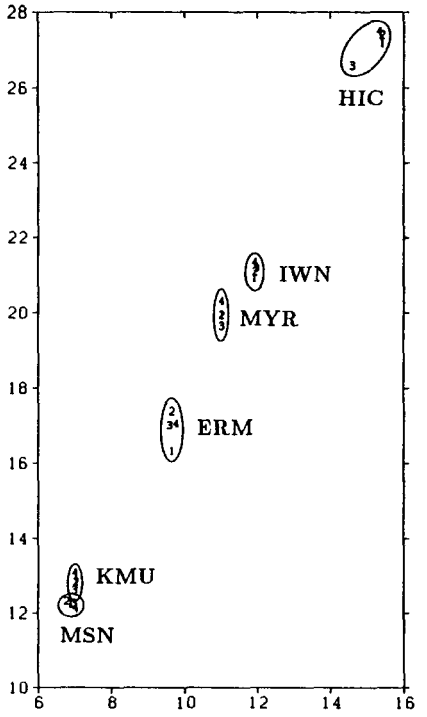

Fig. 9.2

Fig. 9.1. P–S Plot for LSAR models. Origin times of earthquakes are estimated by routine programs. Horizontal axis indicates the estimated P arrival times. Vertical axis indicates the estimated S arrival times.
Fig. 9.2. P–S Plot for MLSAR models. Origin times of earthquakes are modified by the method shown in discussion. Horizontal axis indicates the estimated P arrival times. Vertical axis indicates the estimated S arrival times.

and generally have good coincidence with the estimates by LSAR model. However, for KMU data, the estimates by LSAR model precede 10–40 points. This is due to the presence of hum noise and the human operators could not detect small signal barried in the hum noise.

Figure 10.2 shows the cases of S arrival times. The mid points of the histograms show the estimates by the 2-V MLSAR models. The histograms are scattered over wide region indicating the difficulty of the estimation of S arrival time. On the other hand, however, from the P–S plot shown in Figs. 9.1 and 9.2, it was shown that the MLSAR model can yield reasonable estimates for most of the data sets. Thus the merit of the time series method becomes clear.

Figure 11 shows the changing spectra (Kitagawa (1983)) of the three components of MYR-1F data obtained by the program TVCAR of TIMSAC-84 (Akaike *et al.* (1985)). In the computation, the estimated arrival times were used to specify the change points of the spectra. Based on this information, the change of spectra are clearly detected. It can be seen that after the arrival of S waves, the spectra of the series gradually change and go back to the original shape of background noise.

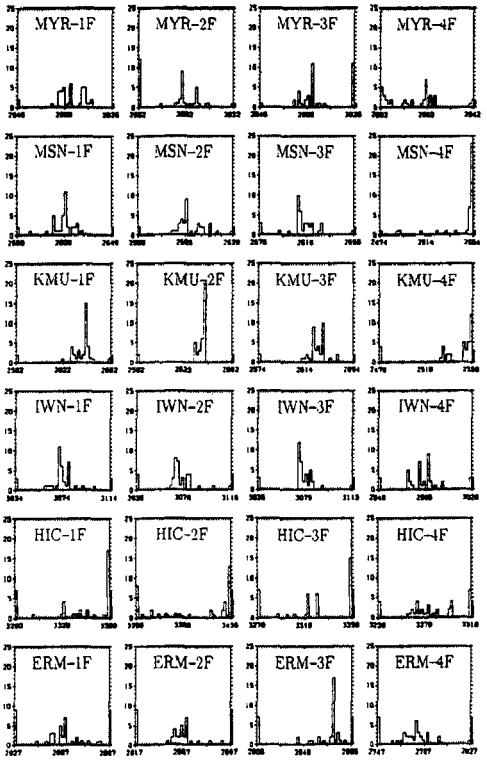

Fig. 10.1

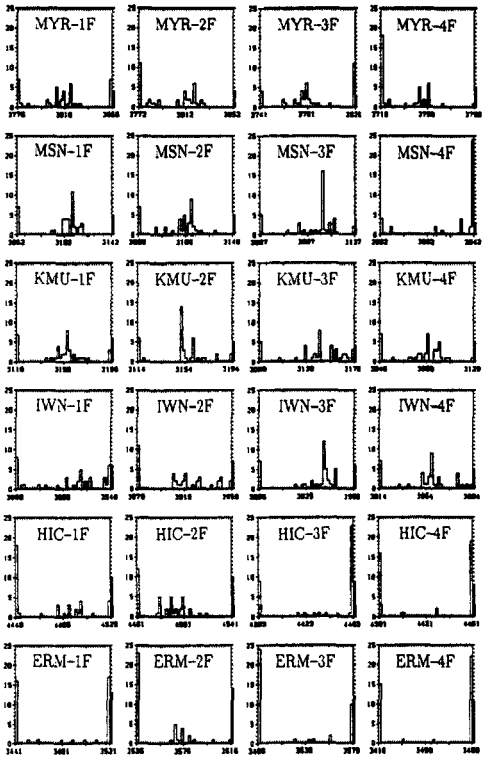

Fig. 10.2

Fig. 10.1. The histograms of the estimated P arrival times by human operators. The center of each histogram indicates the estimate by the MLSAR model.
Fig. 10.2. The histograms of the estimated S arrival times by human operators. The center of each histogram indicates the estimate by the MLSAR model.

However, the significant change occurred only when the P wave and the S wave arrived. This supports our assumption that the seismogram before and after the arrival of the new wave are reasonably expressed by a single AR model.

The reader may wonder the relation between the proposed numerical algorithm and the Kalman filter. For simplicity, we consider the univariate case. For the application of the Kalman filter, we will use the following state space representation of the AR model:

$$\begin{aligned} x_n &= x_{n-1}, \\ y_n &= H_n x_n + w_n, \end{aligned} \tag{6.2}$$

where $x_n = (a_1^{(n)}, \ldots, a_m^{(n)})^t$, $H_n = (y_{n-1}, \ldots, y_{n-m})$ and w_n is the Gaussian white noise with mean 0 and the variance σ^2. By using this expression, the Kalman filter can yield the LSAR model with the same order of computations. However, the Householder method has the following two merit compared with the Kalman filter. Firstly, by the use of Kalman filter, the necessary computation is approximately $4k^2N$, whereas by our method it is less than k^2N. This difference appears

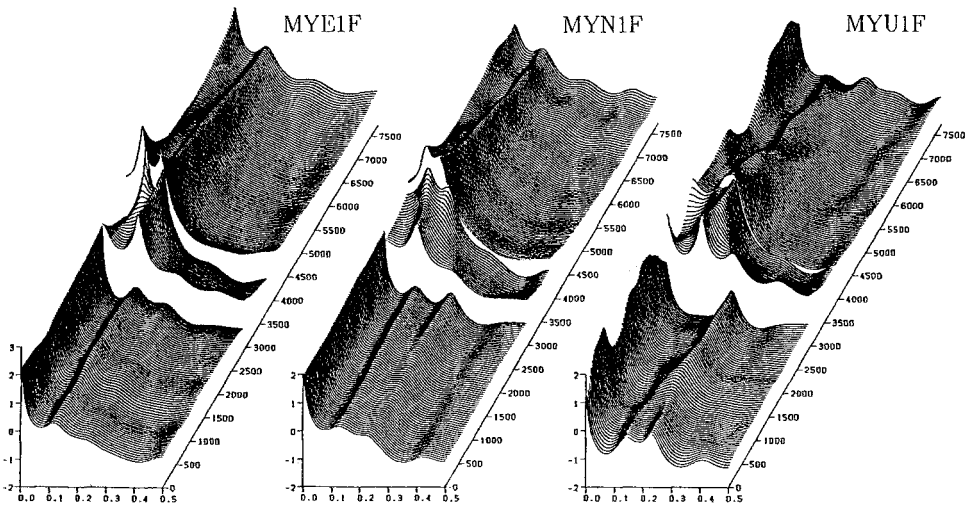

Fig. 11. Changing spectra of MYR-1F data.

since the Householder method can automatically use the merit of symmetry of the matrix and since in our method the variances of the regression coefficients are not explicitly evaluated. Secondly, by our method we can automatically get the minimum value of the AICs' of the AR models with orders less than or equal to m, whereas the Kalman filter can yield only the AIC of the AR models of order m.

7. Conclusion

The objectives of the paper were three-fold. Firstly, we have presented a computationally efficient procedure for the fitting of multivariate locally stationary autoregressive (MLSAR) models. By the proposed procedure, it becomes possible to fit all of the necessary MLSAR models for determining the arrival time of a signal with only a few times as much as the computation for fitting single multivariate AR model. This makes it practical to use the MLSAR model in an on-line system for automatic detection of earthquakes.

Secondly, we have shown a method of obtaining the posterior probability of the arrival time. It is expected that this posterior probability will be useful for the estimation of the hypocenter of the earthquakes. It was also used for the changing spectrum estimation.

Thirdly, the proposed method was applied to various seismograms of foreshocks of the 1982 Urakawa-Oki Earthquake. By this empirical study, it has been seen that

1. For the estimation of the arrival time of P wave, the information on the U–D component is usually sufficient. However, even when it is difficult to determine the arrival time only from the U–D component, the 3-V MLSAR model might give reasonable estimates.

2. For the estimation of the arrival time of S wave, 2-V MLSAR model for E–W and N–S components yields the most reliable estimate. In particular, even

when the LSAR model yielded very flat AIC values and was difficult to determine the minimum, the AIC of the 2-V MLSAR model had clear minimum.

In summary, for the estimation of the arrival times of S waves, the use of 2-V MLSAR model is adequate and the fast algorithm presented in this paper facilitates the application of this model in an on-line system.

Acknowledgements

The authors are grateful to professor Hiroshi Okada of Hokkaido University who motivated the research on the automatic detection of the S wave. The careful reading and the useful comments by the referees are gratefully acknowledged.

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. (1979). A bayesian extension of the minimum AIC procedure of autoregressive model fitting, *Biometrika*, **66**, 237–242.

Akaike, H., Kitagawa, G., Arahata, E. and Tada, F. (1979). TIMSAC-78, *Comput. Sci. Monographs*, No. 11.

Akaike, H., Ozaki, T., Ishliguro, M., Ogata, Y., Kitagawa, G., Tamura, Y-H., Arahata, E., Katsura, K. and Tamura, Y. (1985). TIMSAC-84 Part 1, *Comput. Sci. Monographs*, No. 22.

Buland, R. (1976). The mechanics of locating earthquakes, *Bull. Seismol. Soc. Amer.*, **66**, 173–187.

Golub, G. (1965). Numerical methods for solving linear least squares problems, *Numer. Math.*, **7**, 206–219.

Hamaguchi, H. and Morita, Y. (1980). Second order autoregressive representation of short-period seismic waves, *Zisin*, **33**, 131–140 (in Japanese with English abstract).

Hamaguchi, H. and Suzuku, Z. (1979). Automatic detection of onset time of microearthquake and its confidence, Report of grant-in-aid for scientific research project on natural disaster, No. 54-2, 62-83, Ministry of Education, Science and Culture, Japan.

Hasegawa, A., Umino, N., Yamamoto, A. and Takagi, A. (1986). Automatic event detection and location system of microearthquake observation network, *Zisin*, **39**, 381–395 (in Japanese with English abstract).

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

Kitagawa, G. and Akaike, H. (1978). A procedure for the modeling of non-stationary time series, *Ann. Inst. Statist. Math.*, **30**, 351–363.

Kitagawa, G. and Akaike, H. (1980). On TIMSAC-78, *Applied Time Series Analysis II*, 499–547, Academic Press, New York.

Maeda, N. (1985). A method for reading and checking phase time in auto-processing system of seismic wave data, *Zisin, Ser. 2*, **38**, 365–379 (in Japanese with English abstract).

Ozaki, T. and Tong, H. (1975). On the fitting of non-stationary autoregressive models in the time series analysis, *Proceedings of the 8th Hawaii International Conference on System Science*, Western Periodical, Hawaii.

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

Suzuki, S., Takanami, T., Motoya, Y., Kazahara, M. and Nakanishi, I. (1986). Automatic processing system for microearthquake network of Hokkaido University, *Proceedings of the Annual Meeting of Seismological Society of Japan*, April (in Japanese).

Takanami, T. (1982). Three-dimensional seismic structure of the crust and upper mantle beneath the orogenic belts in the southern Hokkaido, Japan, *Journal of Physics of the Earth*, **30**, 87–104.

Takanami, T. (1990). A study of detection and extraction methods for microearthquake waves by autoregressive models, Doctoral dissertation, Hokkaido University (unpublished).

Takanami, T. and Kitagawa, G. (1988). A new efficient procedure for the estimation of on-set times of seismic waves, *Journal of Physics of the Earth*, **36**, 267–290.

Tjøstheim, D. (1975). Autoregressive representation of seismic P wave signals with an application to the problem of short-period discriminants, *Geophysical Journal of Royal Astronomical Society*, **43**, 269–291.

Yokota, T., Zhou, S., Mizoue, M. and Nakamura, I. (1981). An automatic measurement of arrival time of seismic waves ant its application to an on-line processing system, *Bulletin of Earthquake Research Institute*, **55**, 449–484 (in Japanese with English abstract).