Pei-Jung Chung and Johann F. Böhme

The Methodology of the Maximum Likelihood Approach

Estimation, detection, and exploration of seismic events

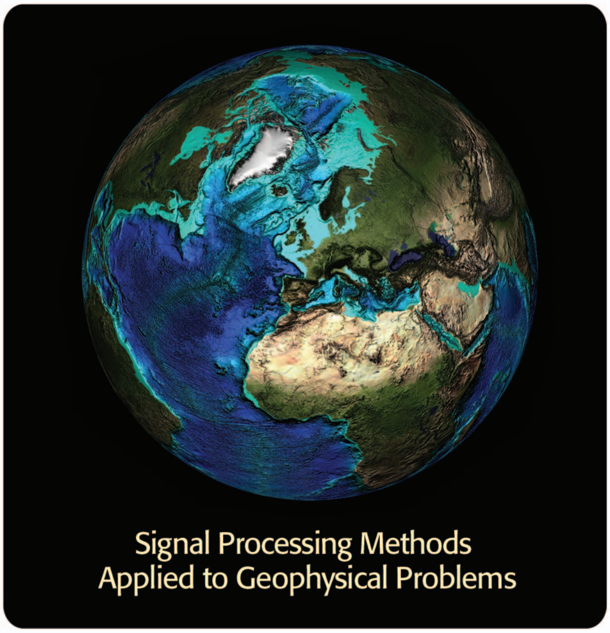


IMAGE COURTESY OF U.S. DEPARTMENT OF COMMERCE/NOAA/NESDIS/NATIONAL GEOPHYSICAL DATA CENTER

The retrieval of information conveyed in data recorded by seismic arrays plays a key role in seismology and geophysical exploration. Accurate localization and reliable detection of seismic events are major tasks in seismic monitoring systems. The nonstationarity, low signal-to-noise ratio (SNR), and weak signal coherence of seismic data remain challenging issues for signal processing algorithms. The maximum likelihood (ML) approach that performs well in such critical conditions is one of the best solutions for simultaneous detection and localization of seismic events. This article will discuss the methodology of ML for estimation and detection of seismic data and its extension to geoacoustic model selection.

INTRODUCTION

Since 1996, the overwhelming majority of countries have signed the Comprehensive Nuclear Test Ban Treaty, which bans nuclear testing underground, in oceans, and in the atmosphere. One important aspect of this treaty is the seismic monitoring regime. The ability to accurately detect and locate worldwide all seismic events above a local magnitude of four within a circle of about 18 km is a formidable scientific and technical challenge [1]. Recently, the 2011 magnitude nine earthquake in Tohoku (also known as the Great East Japan Earthquake) triggered extremely devastating tsunami waves that struck Japan's east coast. In addition to tremendous loss of life and destruction, nuclear accidents caused by the tsunami at the Fukushima power plant was a serious health and safety concern to countries around the globe. Hence, the task of reliable, efficient, and accurate detection of seismic events is of great importance to early warning systems and disaster mitigation.

To extract information contained in propagating waves, one has to automatically process signals measured by a network of seismic arrays and single stations with robust algorithms [2]–[4]. In this article, we address the aspect of signal detection and parameter extraction from a single regional seismic array [5]. In particular, we will investigate a statistically motivated ML

Digital Object Identifier 10.1109/MSP.2012.2182949
Date of publication: 9 April 2012

approach for seismic data analysis. Among existing array processing methods, the ML approach is characterized by optimal statistical properties and robustness against critical scenarios involving low SNRs, closely located sources, and signals coherence. In addition, the ML approach is applicable to both narrow band and broadband data. Unlike subspace methods, which require an additional focusing step in the broadband case, the ML approach combines broadband data in a statistically justified manner by exploiting asymptotic normality and independence of Fourier transformed data [5]. Based on the likelihood principle, a multiple test procedure can be formulated to detect seismic events with high accuracy and good time resolution [6]. The detection capability is of particular importance when multiple events happen within short time intervals. In [7], the experimental comparison with subspace methods [8] and the beamforming-based f-k analysis [9] shows that the ML approach provides the best estimation accuracy and resolution in space and time.

In the following, we first describe the array data model, their statistical properties in frequency domain, and the underlying parametric model. Next, the ML estimates for wave parameters are derived and the multiple hypothesis testing for signal detection is addressed. The application of the ML approach to geoacoustic model selection is briefly discussed. Finally we show experimental results obtained by processing data recorded by the German Experimental Seismic System (GERESS) array and make a comparison between the proposed ML method and f–k analysis in terms of their performance.

DATA MODEL

Seismic events generate waves propagating through the interior of the earth and along its surface layer. For propagation distance much larger than the array aperture, the wave fronts lie approximately on a flat plane perpendicular to the propagation direction. The wave field is observed by sensors located at N distinct positions $\boldsymbol{p}_i$, $i = 1, \ldots, N$. Suppose M wave types are present. The array outputs are sampled temporally by a properly chosen sampling frequency and short-time Fourier-transformed

$$\boldsymbol{X}^l(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} w_l(t)\boldsymbol{x}(t)\mathrm{e}^{-j\omega t}, \quad l = 1, \ldots, L, \tag{1}$$

where $w_l(t)$s are orthonormal window functions [10], [11]. Assuming the plane wave model, the steering vector associated with the ith wave is given by

$$\boldsymbol{d}_i(\omega) = [\mathrm{e}^{-j\omega\boldsymbol{\xi}_i^T\boldsymbol{p}_1}, \ldots, \mathrm{e}^{-j\omega\boldsymbol{\xi}_i^T\boldsymbol{p}_N}]^T, \tag{2}$$

where the superscript T defines matrix transpose. The expression (2) considers only sensitivity in the vertical direction. More details on polarization sensitive arrays can be found in [12]. The slowness vector $\boldsymbol{\xi}_i$ is related to the propagation velocity V_i, azimuth α_i, and elevation ϕ_i through the following equation:

$$\boldsymbol{\xi}_i = \frac{1}{V_i}[\cos\phi_i \sin\alpha_i, \quad \cos\phi_i \cos\alpha_i, \quad \sin\phi_i]^T. \tag{3}$$

In the frequency domain, array outputs can be approximately described through a nonlinear regression model

$$\boldsymbol{X}^l(\omega) = \boldsymbol{H}(\omega,\boldsymbol{\vartheta})\boldsymbol{S}^l(\omega) + \boldsymbol{U}^l(\omega), \tag{4}$$

where the transfer function $\boldsymbol{H}(\omega, \boldsymbol{\vartheta}) = [\boldsymbol{d}_1(\omega) \ldots \boldsymbol{d}_M(\omega)]$ consists of M steering vectors. The nonlinear wave parameters are summarized in $\boldsymbol{\vartheta} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M]$. $\boldsymbol{S}^l(\omega)$ and $\boldsymbol{U}^l(\omega)$ denote the Fourier-transformed signal and noise vectors, respectively.

The advantages of using orthonormal windows, for example, prolate spheroidal sequences suggested in [10] and [11] in (1) include the following:

1) The energy in the observation interval can be concentrated in a prespecified frequency band.
2) The orthogonality of window functions decouples the dependence of $\boldsymbol{X}^l(\omega)$, $(l = 1, \ldots, L)$ and leads to a reduced variance in estimation of the array spectral matrix.

The ML approach relies on proper probabilistic modeling of the data. According to asymptotic theory of Fourier transform [13][14], the array output $\boldsymbol{X}^l(\omega)$ is characterized by the following statistical features:

- For large T, $\boldsymbol{X}^l(\omega)$ is complex normally distributed.
- When the signal $\boldsymbol{S}^l(\omega)$ is considered as deterministic, the distribution of $\boldsymbol{X}^l(\omega)$ is completely specified by the mean $\boldsymbol{H}(\omega,\boldsymbol{\vartheta})\boldsymbol{S}^l(\omega)$ and covariance matrix $\boldsymbol{C}_{U^l}(\omega)$, denoted by $\boldsymbol{X}^l(\omega) \sim \mathcal{N}^c(\boldsymbol{H}(\omega, \boldsymbol{\vartheta})\boldsymbol{S}^l(\omega), \boldsymbol{C}_{U^l}(\omega))$.
- For different frequency bins, $\omega_i \neq \omega_j$, $\boldsymbol{X}^l(\omega_i)$ and $\boldsymbol{X}^l(\omega_j)$ are mutually independent.

In contrary to analysis based on time domain observation $\boldsymbol{x}(t)$, these properties do not assume normal distribution in time domain. In fact, they only require regularity conditions on the moments of $\boldsymbol{x}(t)$. The noise covariance matrix is defined as $\boldsymbol{C}_{U^l}(\omega) = \mathsf{E}[\boldsymbol{U}^l(\omega)\boldsymbol{U}^l(\omega)^H]$, where the superscript H denotes Hermitian transpose. For uncorrelated and identical sensor noise, $\boldsymbol{C}_{U^l}(\omega) = \nu(\omega)\boldsymbol{I}$ where the noise level $\nu(\omega)$ is constant for $l = 1, \ldots, L$.

The problem of central interest is to detect seismic events and estimate the wave parameters using the array observations. The frequencies $\omega_1, \ldots, \omega_J$ chosen for processing should cover most seismic energy. The unknown signal parameters $\boldsymbol{S}^l(\omega_j)$ and noise spectral parameters $\nu(\omega_j)$ are summarized in the vectors S and $\boldsymbol{\nu}$, respectively.

ML APPROACH

The ML approach for parameter estimation for sensor array processing is well known to have excellent statistical performance and robustness. More importantly, the statistical properties of Fourier transformed data allow a natural combination of different frequency bins. These features make the ML method an attractive solution in processing broadband data. For geophysical applications, determination of arrival times of seismic waves plays a crucial role. In the ML framework, localization and detection of seismic activities can be carried out simultaneously. A generalized likelihood ratio test can be easily constructed if an estimate for wave parameters is available.

PARAMETER ESTIMATION

Following the properties of normal distribution and independency discussed in the previous section, the broadband log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{\vartheta}, \boldsymbol{S}, \boldsymbol{\nu}) = -L\sum_{l=1}^{L}\sum_{j=1}^{J}\Big[N\log\nu(\omega_j) + \frac{1}{\nu(\omega_j)}(\boldsymbol{X}^l(\omega_j) - \boldsymbol{H}(\omega_j, \boldsymbol{\vartheta})\boldsymbol{S}^l(\omega_j))^H(\boldsymbol{X}^l(\omega_j) - \boldsymbol{H}(\omega_j, \boldsymbol{\vartheta})\boldsymbol{S}^l(\omega_j)\Big]. \quad (5)$$

Direct maximization of (5) is computationally prohibitive due to a high dimension of parameter space. Fortunately, the signal and noise parameters are separable and the dependence on $\boldsymbol{S}_l(\omega_j)$ and $\nu(\omega_j)$ can be removed through replacing the unknown signal and noise parameters by their ML estimates at fixed and unknown wave parameters $\boldsymbol{\vartheta}$ [13]. Omitting the constant term, the broadband concentrated likelihood function is given by

$$L(\boldsymbol{\vartheta}) = -\sum_{j=1}^{J}\log \operatorname{tr}\Big[(\boldsymbol{I} - \boldsymbol{P}(\omega_j, \boldsymbol{\vartheta}))\,\hat{\boldsymbol{C}}_x(\omega_j)\Big], \quad (6)$$

where $\boldsymbol{P}(\omega_j, \boldsymbol{\vartheta})$ is the projection matrix onto the column space of the transfer matrix $\boldsymbol{H}(\omega_j, \boldsymbol{\vartheta})$ and the sample spectral matrix is given by $\hat{\boldsymbol{C}}_x(\omega_j) = \frac{1}{L}\sum_{l=1}^{L}\boldsymbol{X}^l(\omega_j)\boldsymbol{X}^l(\omega_j)^H$. The ML estimate $\hat{\boldsymbol{\vartheta}}$ is obtained from maximizing (6)

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} L(\boldsymbol{\vartheta}). \quad (7)$$

It is worth noting that the noise level $\nu(\omega_i)$ is estimated by $\hat{\nu}(\omega_j) = 1/N \operatorname{tr}[(\boldsymbol{I} - \boldsymbol{P}(\omega_j, \boldsymbol{\vartheta})\hat{\boldsymbol{C}}_x(\omega_j)]$. The likelihood function (6) suggests that the optimizing parameter minimizes the geometric mean of estimated noise power over frequencies.

SIGNAL DETECTION

The detection of seismic events can be formulated as a multiple hypothesis test. Suppose the maximal number of signals to be $M_{\max}$. The detection procedure discovers the mth signal by testing the null hypothesis H_m against the alternative A_m.

For $m = 1$,

H_1 : Data contains only noise.

$$\boldsymbol{X}(\omega_j) = \boldsymbol{U}(\omega_j)$$

A_1 : Data contains at least one signal.

$$\boldsymbol{X}(\omega_j) = \boldsymbol{H}_1(\omega_j; \boldsymbol{\vartheta}_1)\boldsymbol{S}_1(\omega_j) + \boldsymbol{U}(\omega_j) \quad (8)$$

For $m = 2, \ldots, M_{\max}$

H_m : Data contains at most $(m-1)$ signals.

$$\boldsymbol{X}(\omega_j) = \boldsymbol{H}_{m-1}(\omega_j; \boldsymbol{\vartheta}_{m-1})\boldsymbol{S}_{m-1}(\omega_j) + \boldsymbol{U}(\omega_j)$$

A_m : Data contains at least m signals.

$$\boldsymbol{X}(\omega_j) = \boldsymbol{H}_m(\omega_j; \boldsymbol{\vartheta}_m)\boldsymbol{S}_m(\omega_j) + \boldsymbol{U}(\omega_j) \quad (9)$$

Starting from the noise only hypothesis H_1, the test decides if a signal is present. If no signal is present, the procedure stops. If a signal is detected, the procedure goes to the next step and decides if a second signal is present in the observation. This process continues until the maximum number of signals $M_{\max}$ is reached. The subscript $(\cdot)_m$ in (8) and (9) emphasizes the matrix dimension and the number of parameters associated with the assumed model.

The test statistic for testing H_m against A_m is constructed by the likelihood ratio principle [5]

$$T_m = \sup_{\boldsymbol{\vartheta}_m} L(\boldsymbol{\vartheta}_m) - \sup_{\boldsymbol{\vartheta}_{m-1}} L(\boldsymbol{\vartheta}_{m-1}). \quad (10)$$

Inserting the ML estimate $\hat{\boldsymbol{\vartheta}}_m$ and $\hat{\boldsymbol{\vartheta}}_{m-1}$ into the concentrated likelihood function, the resulting test statistic is given by

$$T_m = \sum_{j=1}^{J} T_m(\omega_j), \quad (11)$$

$$T_m(\omega_j) = \log\Big(1 + \frac{n_1}{n_2}F_m(\omega_j; \hat{\boldsymbol{\vartheta}}_m)\Big), \quad (12)$$

$$F_m(\omega_j; \hat{\boldsymbol{\vartheta}}_m) = \frac{n_2}{n_1}\frac{\operatorname{tr}[(\boldsymbol{P}_m(\omega_j; \hat{\boldsymbol{\vartheta}}_m) - \boldsymbol{P}_{m-1}(\omega_j; \hat{\boldsymbol{\vartheta}}_{m-1}))\hat{\boldsymbol{C}}_x(\omega_j)]}{\operatorname{tr}[(\boldsymbol{I} - \boldsymbol{P}_m(\omega_j; \hat{\boldsymbol{\vartheta}}_m))\hat{\boldsymbol{C}}_x(\omega_j)]}. \quad (13)$$

The statistic $F_m(\omega_j; \hat{\boldsymbol{\vartheta}}_m)$ is asymptotically F_{n_1, n_2}-distributed under null hypothesis with degrees of freedom n_1, n_2 [15]. Taking the estimated nonlinear parameters into account, the degrees of freedom are given by [16]

$$n_1 = L(2 + r_m), \quad n_2 = L(2n - 2m - r_m) \quad (14)$$

with $r_m = \dim(\boldsymbol{\vartheta}_m)$ denoting the dimension of the nonlinear parameter vector associated with the mth signal.

At the mth stage, the test statistic is compared with a threshold t_m to decide whether to reject the hypothesis H_m or retain it. The rejection of H_m implies that a (further) signal is discovered; otherwise, no signal is detected. More specifically,

$$T_m \geq t_m \Rightarrow \text{reject } H_m, \quad (15)$$

$$T_m < t_m \Rightarrow \text{retain } H_m. \quad (16)$$

CALCULATION OF TEST THRESHOLD

The threshold t_m is chosen to keep a prespecified significance level (or false alarm rate) at α_m. The probability that a signal is wrongly detected should be kept less than α_m. Typical values for α_m are 1%, 5%, and 10%. Let $\mathcal{F}_m(\cdot)$ denote the null distribution of T_m. The threshold is determined by its inverse $\mathcal{F}_m^{-1}(\cdot)$ as follows:

$$t_m = \mathcal{F}_m^{-1}(\alpha_m). \quad (17)$$

In the narrow band case with $J = 1$, the test (9) is equivalent to the F-test [17], the threshold can be easily derived from the F-distribution. However, in the broadband case, the null distribution does not have a closed-form expression. To determine the threshold t_{α_m}, methods such as normal approximation was proposed by [18]. Note that the statistics $T_m(\omega_j)$, $(j = 1, \ldots, J)$ are independent, identically distributed with mean $\mu_m = ET_m(\omega_j)$ and variance $\sigma_m^2 = Var(T_m(\omega_j))$. For large samples J, the test statistic T_m is approximately normally distributed. The threshold derived from normal distribution is given by

$$t_m \approx \mu_m + \frac{\sigma_m}{\sqrt{J}} \Phi^{-1}(\alpha_m), \tag{18}$$

where $\Phi^{-1}(\alpha_m)$ is the inverse of the standard normal distribution function evaluated at α_m. The formulas for computing μ_m and σ_m^2 can be found in [19]. A more accurate approximation can be obtained from the Cornish-Fisher expansion that takes higher-order moments of T_m into account [20] or simulation-based bootstrap techniques [16], [21].

For simplicity, the test level α_m is kept constant in each detection stage. The problem of controlling of the global test level can be achieved by controlling the familywise error-rate (FWE) through the sequentially rejective Bonferroni-Holm procedure [22] or the false discovery rate (FDR) through the Benjamini-Hochberg procedure [23]. The former is discussed in [18] and the latter is addressed in [24].

Given the Fourier transformed data, the ML approach for estimation and detection of seismic waves is summarized in Table 1.

[TABLE 1] ML APPROACH FOR ESTIMATION AND DETECTION OF SEISMIC EVENTS.

INPUT: FOURIER TRANSFORMED DATA $\{\boldsymbol{X}^l(\omega_j), l = 1, \ldots, L, j = 0, \ldots, J-1\}$
MAXIMAL NUMBER OF SIGNALS $M_{\max}$
INITIAL VALUE FOR ESTIMATED NUMBER OF SIGNALS $\hat{M} = 0$.
FOR $m = 1, \ldots, M_{\max}$
1. WAVE PARAMETER ESTIMATION
FIND THE ML ESTIMATE: $\hat{\boldsymbol{\vartheta}}_m = \arg\max_{\boldsymbol{\vartheta}_m} L(\boldsymbol{\vartheta}_m)$.
2. COMPUTE THE TEST STATISTIC T_m (10).
3. SIGNAL DETECTION
IF $T_m \geq t_m$, REJECT H_m, A SIGNAL IS DETECTED
$\hat{M} = \hat{M} + 1$, UPDATE THE NUMBER OF SIGNALS
$m = m + 1$, THE PROCEDURE CONTINUES
ELSE BREAK THE LOOP
END
END
OUTPUT: ESTIMATED WAVE PARAMETERS $\hat{\boldsymbol{\vartheta}}_{\hat{M}}$ AND DETECTED NUMBER OF SIGNALS $\hat{M}$.

IMPLEMENTATION OF ML ESTIMATION

The maximization of the likelihood function (6) involves multidimensional search of nonlinear functions. The global maxima can be computed by stochastic optimization methods such as the genetic algorithm [25] or simulated annealing in a straightforward manner. A computationally more attractive implementation is to exploit the nature of sequential detection. Recall that the likelihood ratio (10) for testing H_m against A_m involves the estimates $\hat{\boldsymbol{\vartheta}}_{m-1}$ and $\hat{\boldsymbol{\vartheta}}_m$. Assuming that the components in $\hat{\boldsymbol{\vartheta}}_{m-1}$ for the first $(m-1)$ signals do not deviate from the corresponding components in the estimate $\hat{\boldsymbol{\vartheta}}_m$, we may employ the knowledge about the discovered signals and fix their values at $\hat{\boldsymbol{\vartheta}}_{m-1}$. Then the likelihood function is optimized globally over the slowness parameter $\boldsymbol{\xi}_m$ associated with the mth signal

$$\hat{\boldsymbol{\xi}}_m = \arg\max_{\boldsymbol{\xi}_m} L(\hat{\boldsymbol{\vartheta}}_{m-1}, \boldsymbol{\xi}_m). \tag{19}$$

The computational complexity for global optimization is significantly reduced. The so-obtained estimate can be refined by carrying out local optimization via local optimization methods such as Newton-type algorithms using the initial estimate

$$\hat{\boldsymbol{\vartheta}}_m^{[0]} = [\hat{\boldsymbol{\vartheta}}_{m-1}, \hat{\boldsymbol{\xi}}_m] \tag{20}$$

with a smaller search interval. This implementation shows excellent performance in both simulation and experimental results in seismic and sonar applications [18], [16], [19].

EXPLORATION OF GEOPHYSICAL STRUCTURES

The geophysical structure of Earth's surface has been of great interest to many areas including geology, environmental and civil engineering, and industrial applications. The geophysical inversion aims to construct a geophysical model that best fits experimental measurements. The majority of current inversion techniques assumes a constant model structure and computes the best model within this class of models [26], [27]. Contrary to these methods, the ML approach considers models of various complexity and finds the best model in a hierarchy of models. The best model within each class maximizes the likelihood function. The overall best model is selected by hypothesis testing.

Consider the following classes of models of increasing orders

$$\mathcal{M}_1 \subset \cdots \subset \mathcal{M}_m \subset \cdots \mathcal{M}_{M_{\max}}. \tag{21}$$

The major difference between the previously discussed seismic location and detection is that the propagation waves are caused by man-made signals in a controlled environment. The array outputs collected during an experiment can be described by

$$\boldsymbol{X}^l(\omega_j) = \boldsymbol{d}(\omega_j; \boldsymbol{\vartheta}_m) S^l(\omega_j) + \boldsymbol{U}^l(\omega_j), \tag{22}$$

where $\boldsymbol{\vartheta}_m$ contains the geophysical parameters associated with the model class $\mathcal{M}_m$. Note that only one signal source is present; hence, $S^l(\omega_j)$ is a scalar and $\boldsymbol{d}(\omega_j; \boldsymbol{\vartheta}_m)$ is a column vector. The transfer function $\boldsymbol{d}(\omega_j; \boldsymbol{\vartheta}_m)$ is derived from underlying geophysical laws. In a multilayered half-space model discussed in [28], $\boldsymbol{\vartheta}_m$ may contain thickness parameters of layers, P-wave and S-wave propagation velocities. The model complexity increases with an increasing numbers of layers.

To select the best model, a sequential generalized likelihood ratio test is suggested in [19] and [28]. Unlike the detection problem, the dimension of signal subspace remains the same for all models. To obtain a test statistic with known null distribution, the two models are tested against a joint signal subspace spanned by both models. Starting from the smallest model in (21), $\mathcal{M}_1$, we compare two adjacent models $\mathcal{M}_m$ and $\mathcal{M}_{m+1}$ by carrying out a three-step test.

Step 1

$H_{1,m}$	$\boldsymbol{X} = \boldsymbol{U}$	No signal in the data.
$A_{1,m}$	$\boldsymbol{X} = [\boldsymbol{d}_m \; \boldsymbol{d}_{m+1}] \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} + \boldsymbol{U}$	$\mathcal{M}_m$ or $\mathcal{M}_{m+1}$ generates the data.

[TABLE 2] EVENT LIST FROM THE NATIONAL EARTHQUAKE INFORMATION CENTER (NEIC).

ORIGIN TIME HHMMSS.SS	LAT. DEG. N	LONG. DEG. E	DIST. DEG.	THEOR. BAZ DEG.	MAG. MB	LOCATION
17:42:50	34.88	32.81	19.86	127.6	3.1	CYPRUS

Step 2

$H_{2,m}$ $\quad \boldsymbol{X} = \boldsymbol{d}_m + \boldsymbol{U}$ $\quad \mathcal{M}_m$ generates the data.

$A_{2,m}$ $\quad \boldsymbol{X} = [\boldsymbol{d}_m \, \boldsymbol{d}_{m+1}] \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} + \boldsymbol{U}$ $\quad$ Some components cannot be modeled by $\mathcal{M}_m$.

Step 3

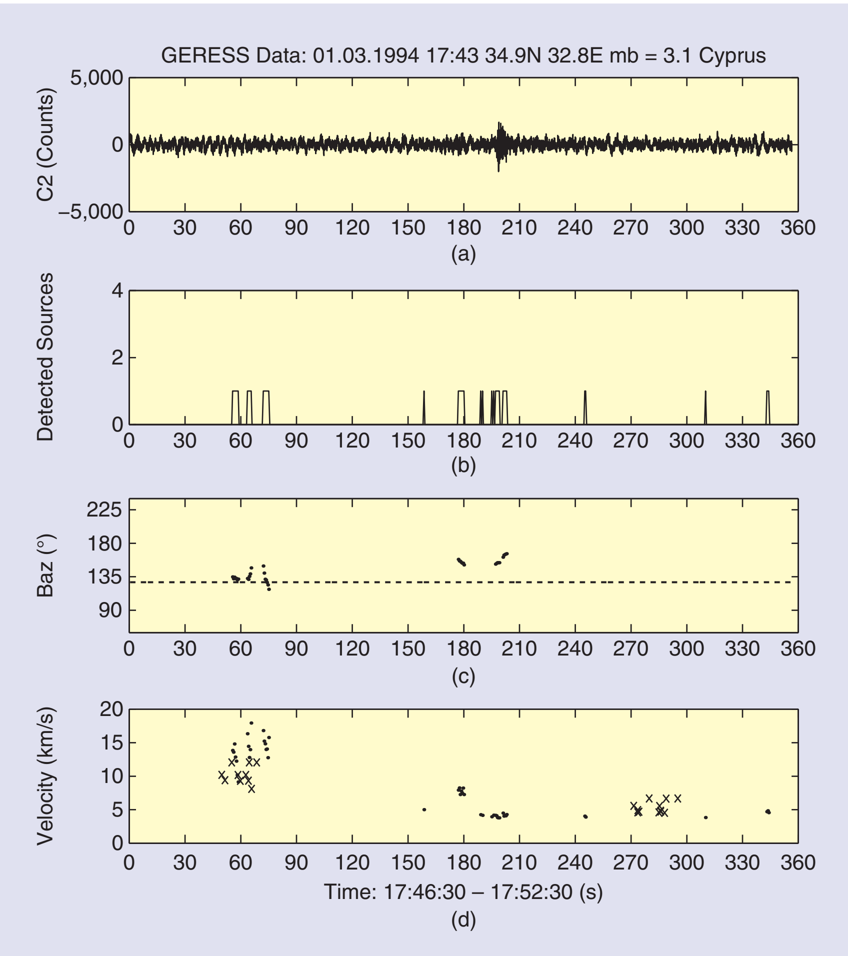
$H_{3,m}$ $\quad \boldsymbol{X} = \boldsymbol{d}_{m+1} + \boldsymbol{U}$ $\quad \mathcal{M}_{m+1}$ generates the data.

$A_{3,m}$ $\quad \boldsymbol{X} = [\boldsymbol{d}_m \, \boldsymbol{d}_{m+1}] \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} + \boldsymbol{U}$ $\quad$ Some components cannot be modeled by $\mathcal{M}_m$.

For simplicity, the frequency and parameter dependence has been omitted in the formulation. The vector $\boldsymbol{d}_i = \boldsymbol{d}(\omega_j; \hat{\boldsymbol{\vartheta}}_i)$, $i = m, m+1$, is the transfer vector computed at the ML estimate obtained under $\mathcal{M}_i$. In Step 1, if $H_{1,m}$ is rejected, we conclude that a signal is present in the data and proceed to Step 2. If the hypothesis $H_{2,m}$ is accepted, then model $\mathcal{M}_m$ is the true model and the test stops here. Otherwise, we proceed to Step 3. Step 3 is a cross-check to test if the larger model $\mathcal{M}_{m+1}$ is a better one. The test statistics based on likelihood ratio have the same form as (10)–(13). The null distribution and test threshold can be derived in a similar manner as those for signal detection. The three-step procedure for model selection has been successfully applied to near surface seismic model reconstruction. More details can be found in [28].

EXPERIMENTAL RESULTS

In this section, we demonstrate feasibility and performance of the ML approach by processing seismic data recorded by the 25 vertical seismometers of the GERESS array located in the Bavarian Forest, Germany. The output of each sensor is sampled with 40 Hz. Details about the GERESS arrays can be found in [29]. Because of the relatively small vertical aperture of GERESS array, the elevation ϕ_i is taken to be zero.



[FIG1] Application of wideband ML method to a weak teleseismic event that occurred on 1 March 1994: (a) seismic data recorded at station C2, GERESS array in Germany, (b) number of detected signals, (c) estimated values for back-azimuth (·) and theoretical values for back-azimuth (– –), and (d) estimated values for velocity (·) and theoretical values for velocity (×).

A teleseismic event, which occurred on 1 March 1994 in the Eastern Mediterranean, was selected for analysis. We also applied the conventional sliding f–k analysis [30] to the same batch of data for comparison. Theoretical slowness values for each event are derived from the AK135 Earth model [31] as reference. Relevant information about the selected events is collected in Table 2.

For the ML-based algorithm, we employed sliding windows of length 3.2 s with a shift of 0.5 s. In each window, data is processed by the method described before and the estimated parameters are plotted at the center of the data window (e.g., Figure 1). The spectral density matrix is estimated with $L = 3$ Thomson's orthonormal windows. $J = 7$ frequency bins between 0.9 and 3.1 Hz are selected for estimation and detection. The maximum number of signals is assumed to be three. The log–likelihood function is optimized over search space V_i (apparent velocity) $\in [0 \ldots 50]$ km/s, α_i (back–azimuth) $\in [0 \ldots 360]$ degree. The corresponding slowness vector is calculated according to (3). These constraints avoid nonreasonable estimates for v_i and offer better accuracy for α_i. A genetic algorithm is used in global optimization similar to those described in [25]. The sequential testing is carried out with test level $\alpha_m = 0.033$.

The conventional sliding f–k analysis also uses a window length of 3.2 s and a shift of 0.5 s. The data in each window were first filtered with a Butterworth bandpass filter of 0.7−2.0 Hz, order three. Next, a wideband frequency-wavenumber spectrum analysis following [9] was applied and the corresponding results (back-azimuth and apparent velocity) were plotted at the center of the data window (e.g., Figure 2). Furthermore, the quality factor of the f–k analysis (ratio of the incoherent noise power to the total power of the sensor output) is displayed in addition to the beam steered in the estimated direction and centered at the main station C2 of the GERESS array.

It is shown in Figure 1 that earthquakes can be detected with good time resolution by its P phases. No signals are detected for the S phases. Sometimes there are false alarms. The estimates for back azimuth are quite accurate while the estimation of apparent velocity is slightly higher than theoretical values (× , *). More accurate estimates can be obtained by including frequency bins between 0.6 and 0.9 Hz. It was not done in our analysis because the presence of coherent noise structure in this frequency interval severely affects the detection performance. The signals detected between 170 and 210 s after begin of analysis is another local seismic event that can be recognized by slightly different azimuths and different velocities at 180 s and 190 s. The application of f–k analysis to this event shows that under this critical condition it is very difficult to claim that a seismic signal is present with help of the beamformer output and quality of the estimates (see Figure 2). In the absence of a seismic signal quality lies in the same range as in the presence of a seismic signal. The estimates for back azimuth are similar to those given by ML approach and apparent velocities seem slightly better. There is no indication for the regional event detected by ML algorithm in f-k analysis.

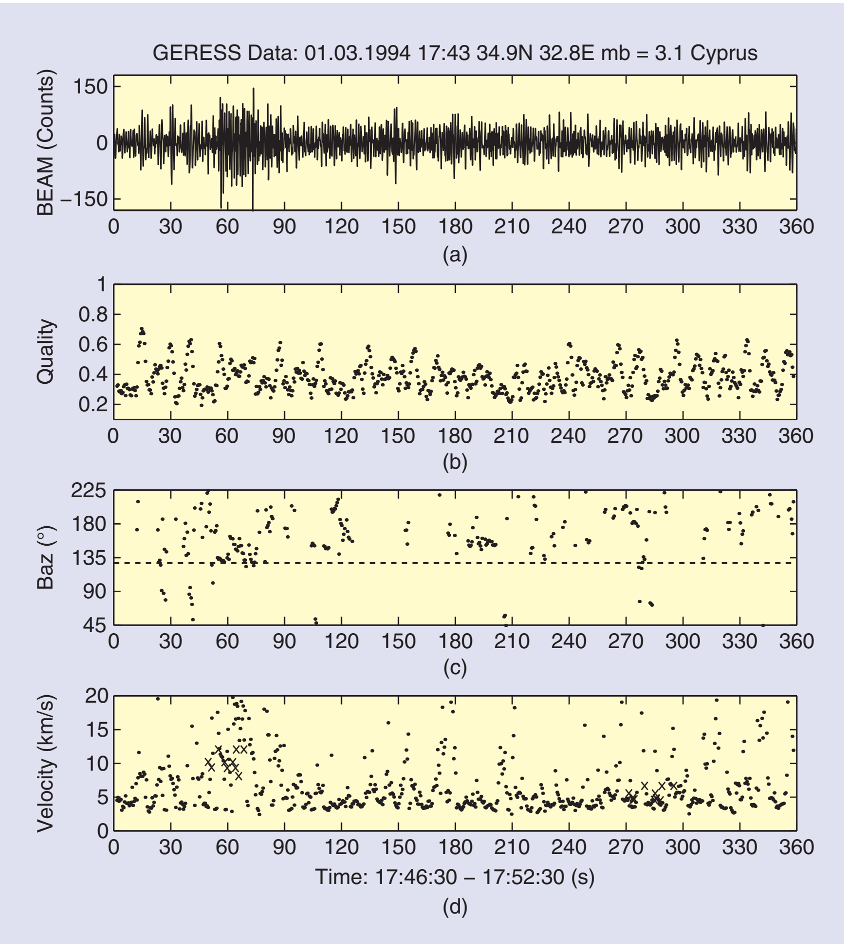
In another analysis, an earthquake originated from Gulf of Aqaba in the Middle East is contaminated by a one magnitude-unit smaller preshock, located about 37 km from the main event. For the first time, both events can be detected and localized accurately by applying the ML approach. The f-k analysis was only able to identify the strong event and could not discover the weak one. More details can be found in [6].

Experimental results show that the ML method provides not only reasonable estimates for wave parameters but also a reliable indication about the presence of weak signals or multiple sources for low SNRs and long propagation paths. The detection ability is a significant advantage over routinely used f–k method. Thus the ML method is a promising alternative to the conventional method in seismic application.

CONCLUSIONS AND FUTURE WORK

We discussed the ML approach for seismic parameter estimation and signal detection, and its extension to geophysical model reconstruction. The ML approach has the advantage of excellent performance, high-resolution capability, and robustness against low SNRs and small samples. Exploiting the asymptotic normality of Fourier transformed data, the ML method allows an optimal combination of various frequency components. More importantly, localization and detection of seismic events can be carried out simultaneously. The multiple hypothesis test provides a statistically justified framework for signal detection and model fitting.

Experimental results show that the ML approach provides both accurate estimates for velocity and location and reliable indication about the presence of seismic events in critical scenarios and long propagation paths. The detection ability will significantly enhance the power of modern seismic monitoring systems in minimizing the impact of natural disasters.



[FIG2] Application of sliding f-k analysis to a weak teleseismic event that occurred on 1 March 1994: (a) beamformer output, (b) quality of estimates, (c) estimated values for back-azimuth (·) and theoretical values for back-azimuth (−−), and (d) estimated values for velocity (·) and theoretical values for velocity (×).

In this article, we have considered ML methods under the plane wave model. Since the ML method is a model-based approach, it can be expected that further improvement in estimation accuracy can be achieved if a more realistic seismic model is incorporated into the algorithm. Another important issue is computational complexity. Although the processing speed of modern computers has grown rapidly, computational efficiency still plays an important role in practical systems. The recursive algorithms suggested in [32] and [33] facilitate computationally efficient implementation of ML methods. Their online processing capability can be very beneficial to seismic processing systems. As mentioned in the beginning of the article, the ML approach discussed here addresses seismic data processing from the view of a regional seismic array. How to integrate ML techniques into global seismic monitoring networks would be a challenging and very important topic in the future.

AUTHORS

Pei-Jung Chung (peijung.chung@gmail.com) received the Dr.-Ing. with distinction in 2002 from Ruhr-Universität Bochum, Germany. From 2002 to 2004, she held a postdoctoral position at Carnegie Mellon University and the University of Michigan, Ann Arbor, respectively. From 2004 to 2006, she was an assistant professor with National Chiao Tung University, Hsin Chu, Taiwan. In 2006, she joined the Institute for Digital Communications, School of Engineering, the University of Edinburgh, United Kingdom as a lecturer. Currently, she is an associate member of the IEEE Signal Processing Society Sensor Array Multichannel Technical Committee and serves on the IEEE Communications Society, Multimedia Communications Technical Committee as vice chair of the Interest Group on Acoustic and Speech Processing for Communications. Her research interests include array processing, statistical signal processing, wireless multiple-input and multiple-output communications, and distributed processing in wireless sensor networks.

Johann F. Böhme (Johann.Boehme@rub.de) received the Diplom in mathematics in 1966, the Dr.-Ing. in 1970, and the Habilitation in 1977, both in computer science, from the Technical University of Hannover, Germany, the University of Erlangen, Germany, and the University of Bonn, Germany, respectively. From 1967 to 1974, he was with the Sonar-Research Laboratory of Krupp Atlas Elektronik in Bremen, Germany. He was with the University of Bonn until 1978 and the FGAN in Wachtberg, Werthhoven. He has been a professor of signal theory in the Department of Electrical Engineering and Information Sciences, Ruhr-Universität Bochum, Germany, since 1980. His research interests are in the areas of statistical signal processing and its applications. He is an IEEE Life Fellow and recipient of the 2003 IEEE Signal Processing Society Technical Achievement Award.

REFERENCES

[1] E. S. Husebye and A. Dainty, Eds., *Monitoring a Comprehensive Test Ban Treaty*. Dordrecht, The Netherlands: Kluwer, 1996.

[2] T. C. Bache, S. R. Bratt, H. J. Swanger, G. W. Beall, and F. K. Dashiell, "Knowledge-based interpretation of seismic data in the intelligent monitoring system," *Bull. Seism. Soc. Amer.*, vol. 83, no. 5, pp. 1507–1526, 1993.

[3] L. M. Haikin, A. F. Kushnir, and A. M. Dainty, "Combined automated and off-line computer processing system for seismic monitoring with small aperture arrays," *Seism. Res. Lett.*, vol. 69, no. 3, pp. 235–247, 1998.

[4] G. Ekstrom, "Global detection and location of seismic sources by using surface waves," *Bull. Seism. Soc. Am. A*, vol. 96, no. 4, pp. 1201–1212, 2006.

[5] J. F. Böhme, "Retrieving signals from array data," in *Monitoring a Comprehensive Test Ban Treaty*, E. S. Husebye and A. M. Dainty, Eds. Alvor, Portugal: NATO ASI, Jan. 23–Feb. 2, 1995, pp. 587–610.

[6] P.-J. Chung, M. L. Jost, and J. F. Böhme, "Seismic wave parameter estimation and signal detection using broadband maximum likelihood methods," *Comput. Geosci.*, vol. 27, no. 2, pp. 147–156, Mar. 2001.

[7] P.-J. Chung, A. B. Gershman, and J. F. Böhme, "Comparative study of two-dimensional maximum likelihood and interpolated root-MUSIC with application to teleseismic source localization," in *Proc. IEEE Signal Processing Workshop Statistical Signal and Array Processing*, Pocono Manor, PA, Aug. 14–16, 2000, pp. 68–72.

[8] D. V. Sidorovich and A. B. Gershman, "Two-dimensional wideband interpolated root-MUSIC applied to measured seismic data," *IEEE Trans. Signal Processing*, vol. 46, no. 8, pp. 2263–2267, Aug. 1998.

[9] J. Capon, "High-resolution frequency–wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[10] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept. 1982.

[11] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[12] D. Maiwald, D. V. Sidorovitch, and J. F. Bohme, "Broadband maximum likelihood wave parameter estimation using polarization sensitive arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, 1993, pp. 356–359.

[13] J. F. Böhme, "Array processing," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 1–63.

[14] D. R. Brillinger, *Time Series: Data Analysis and Theory*, expanded ed. San Francisco, CA: Holden-Day, 1981.

[15] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions-2*. New York: Wiley, 1970.

[16] D. Maiwald, "Breitbandverfahren zur Signalentdeckung und -ortung mit Sensorgruppen in Seismik- und Sonaranwendungen," Dissertation, Faculty Elect. Eng., Ruhr Univ. Bochum, Germany, May 1995.

[17] R. H. Shumway, "Replicated time-series regression: An approach to signal estimation and detection," in *Handbook of Statistics*, vol. 3, D. R. Brillinger and P. R. Krishnaiah, Eds. New York: Elsevier Science Publishers B.V., 1983, pp. 383–408.

[18] D. Maiwald and J. F. Böhme, "Multiple testing for seismic data using bootstrap," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Adelaide, South Australia, 1994, vol. 6, pp. 89–92.

[19] C. F. Mecklenbräuker, P. Gerstoft, J. F. Böhme, and P.-J. Chung, "Hypothesis testing for geoacoustic environmental models using likelihood ratio," *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1738–1748, Mar. 1999.

[20] P.-J. Chung, M. Viberg, and C. F. Mecklenbräuker, "Broadband ML estimation under model order uncertainty," *Signal Process.*, vol. 90, no. 5, pp. 1350–1356, May 2010.

[21] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.

[23] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995.

[24] P.-J. Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. O. Hero, "Detection of the number of signals using the Benjamini–Hochberg procedure," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2497–2508, June 2007.

[25] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[26] M. Sambridge, "Geophysical inversion with a neighbourhood algorithm. I. Searching a parameter space," *Geophys. J. Int.*, vol. 138, pp. 479–494, 1999.

[27] N. Qiu, Q.-S. Liu, Q.-Y. Gao, and Q.-L. Zeng, "Combining genetic algorithm and generalized least squares for geophysical potential field data optimized inversion," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 4, pp. 660–664, Oct. 2010.

[28] M. Westebbe, J. F. Böhme, and H. Krummel, "Model fitting and testing in near surface seismics using maximum likelihood in frequency domain," in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1998, pp. 1311–1315.

[29] H.-P. Harjes, "Design and sitting of a new regional array in Central Europe," *Bull. Seism. Soc. Amer. B*, vol. 80, no. 6, pp. 1801–1817, 1990.

[30] S. Mykkeltveit and H. Bungum, "Processing of regional seismic events using data from small-aperture arrays," *Bull. Seism. Soc. Am.*, vol. 74, no. 6, pp. 2313–2333, 1984.

[31] B. L. N. Kennett, E. R. Engdahl, and R. Buland, "Constraints on seismic velocities in the earth from traveltimes," *Geophys. J. Int.*, vol. 122, pp. 108–124, 1995.

[32] P.-J. Chung and J. F. Böhme, "EM and SAGE algorithms for towed array data," in *The Applications of Space–Time Adaptive Processing*, R. Klemm, Ed. London: IEE Publishers, 2004, pp. 733–753.

[33] P.-J. Chung and J. F. Böhme, "Recursive EM and SAGE-inspired algorithms with application to DOA estimation," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2664–2677, Aug. 2005.

[SP]