Bulletin of the Seismological Society of America, Vol. 97, No. 6, pp. 1890–1910, December 2007, doi: 10.1785/0120060243

Real-Time Estimation of Fault Rupture Extent Using Near-Source versus Far-Source Classification

by Masumi Yamada, Thomas Heaton, and James Beck

Abstract To estimate the fault dimension of an earthquake in real time, we present a methodology to classify seismic records into near-source or far-source records. Characteristics of ground motion, such as peak ground acceleration, have a strong correlation with the distance from a fault rupture for large earthquakes. This study analyzes peak ground motions and finds the function that best classifies near-source and far-source records based on these parameters. We perform (1) Fisher's linear discriminant analysis and two different Bayesian methods to find the coefficients of the linear discriminant function and (2) Bayesian model class selection to find the best combination of the peak ground-motion parameters. Bayesian model class selection shows that the combination of vertical acceleration and horizontal velocity produces the best performance for the classification. The linear discriminant function produced by the three methods classifies near-source and far-source data, and in addition, the Bayesian methods give the probability for a station to be near-source, based on the ground-motion measurements. This discriminant function is useful to estimate the fault rupture dimension in real time, especially for large earthquakes.

Introduction

Recent studies show that earthquake early warning systems, such as the virtual seismologist (VS) method (Cua, 2005), can accurately estimate the location of the epicenter a few seconds after the first arrival station records the ground motion of the mainshock (Nakamura, 1988; Allen and Kanamori, 2003; Odaka *et al.*, 2003; Wu and Kanamori, 2005). The VS method assumes a point source model for the rupture, and it works well for small to moderate earthquakes ($M_w < 6.5$) (Cua, 2005). However, for large earthquakes, the fault rupture length can be on the order of tens to hundreds of kilometers, and the prediction of ground motion at a site requires approximated knowledge of the rupture geometry. Early warning information based on a point source model may underestimate the ground otion at a site, if a station is close to the fault and distant from the epicenter. This occurs because, for large earthquakes, the peak characteristics of ground motion, such as peak ground acceleration, have stronger correlation with the fault rupture distance rather than with the epicentral or hypocentral distance (Campbell, 1981). (The definition of the fault rupture distance in this article is the shortest distance between the station and the surface projection of the fault rupture surface.)

In order to construct an early warning system that is more reliable for large earthquakes, it is necessary to estimate the fault rupture extent and slip on the fault in real time. The objective of this article is to develop a methodology to classify stations into near source and far source because this can be used for identifying the fault geometry if there is a sufficiently dense seismic network. Peak ground motions recorded in past earthquakes are analyzed to predict whether a station recording ground motion is close to the earthquake fault area. This classification problem can be stated as follows: given ground-motion data from past earthquake records, what is the probability that a station is near source when a new observation is obtained?

To approach this problem, we take the following steps:

1. We collect strong-motion data from earthquake strong-motion archives and classify these samples into two predefined groups: records from near-source stations and from far-source stations. This particular set of data is called the training set.
2. We discover a discriminant function of the samples features (e.g., peak ground acceleration [PGA], peak ground velocity [PGV], peak ground displacement [PGD]) that provides the best performance in terms of near-source/far-source classification.
3. We allocate new observations, when they are obtained, to one of the two groups based on the discriminant function.

The first step is quite straightforward: strong-motion data from past earthquakes are collected based on certain selection criteria. The second step is the main topic of this article, and we investigate linear discriminant functions by using the traditional Fisher method and two Bayesian methods. The third step can then be accomplished in a real-time analysis. Given a new ground-motion observation from ongoing rupture, the discriminant function gives the probability that the observation is located in the near source.

Strong-Motion Data

We used strong-motion datasets from nine earthquakes with magnitudes greater than 6.0 and containing records of near-source stations. The selected earthquake dataset is shown in Table 1. Here, we define a near-source station as a station whose fault rupture distance is less than 10 km. 695 three-component strong-motion data are used for the classification analysis, and 14% (100 stations) are from near-source stations.

Data Sources

We obtained the strong-motion dataset for the Imperial Valley (15 October 1979), Loma Prieta (18 October 1989), Landers (28 June 1992), Northridge (17 January 1994), and Denali (3 November 2002) earthquakes from the COSMOS Virtual Data Center (http://db.cosmos-eq.org), which includes data from the California Strong Motion Instrumentation Program seismic network and the U.S. Geological Survey seismic network. The Northridge earthquake dataset in the COSMOS Virtual Data Center also includes records from seismic networks of the California Institute of Technology, Los Angeles Department of Water and Power, Metropolitan Water District, Southern California Earthquake Center, and University of Southern California. All these data were recorded by accelerometers and processed appropriately before distribution to users. The correction process may apply baseline corrections, band-pass filters to remove noise contamination, and instrument correction to remove the effects of frequency-dependent instrument response (http://nsmp.wr.usgs.gov/processing.html).

Strong-motion data from the Hyogoken-nanbu earthquake (16 January 1995) are provided by the Japan Meteorological Agency, the Committee of Earthquake Observation and Research in the Kansai Area (CEORKA) in Japan (Toki *et al.*, 1995), and the Japan Railway Institute, whose records were scanned and digitized by Wald (1996). Seismometers installed in the CEORKA network record velocity, and those records are differentiated once to obtain accelerograms.

The national strong-motion accelerograph network in Turkey recorded the strong motions during the Izmit earthquake (17 August 1999) (Akkar and Gülkan, 2002). They can be downloaded from the ftp site of the Earthquake Research Department of General Directorate of Disaster Affairs, Ministry of Public Works and Settlement, Ankara, Turkey (ftp://angora.deprem.gov.tr/). The COSMOS Virtual Data Center archived the dataset of another network operated by the Kandilli Observatory and Earthquake Research Institute, Earthquake Engineering Department, Bogaziçi University, Istanbul, Turkey. Stations with fault distance greater than 200 km are excluded because ground-motion amplitudes of those stations are quite small, which results in a low signal-to-noise ratio. We use four digital and six analog acceleration records from the national network and eight digital acceleration records from the Bogaziçi University network.

The Chi-Chi earthquake (20 September 1999) is one of the best recorded earthquakes with a large number of stations and a dense station distribution both in the near source and far source. Strong-motion records for the Chi-Chi earthquake are available on the attached CD in the Special Issue of the Bulletin of the Seismological Society of America, volume 93, number 5 (Lee *et al.*, 2001). These records were produced by the Central Weather Bureau Seismic Network, and they are the largest set of strong-motion data recorded from a major earthquake (Shin and Teng, 2001). Lee *et al.* (2001) classified the recorded accelerograms into four quality groups based on the existence of absolute timing, preevents, and defects. For this analysis, QA-class data (best for any studies) and QB-class data (next best but no absolute timing) are used.

Strong-motion data from the Niigataken-chuetsu earthquake (23 October 2004) were recorded by the K-NET and

Table 1
The Earthquake Dataset Used for the Classification Analysis

Earthquake	M_w	Near Source	Far Source	Total	Fault Model
Imperial Valley (1979)	6.5	14	20	34	Hartzell and Heaton, 1983
Loma Prieta (1989)	6.9	8	39	47	Wald *et al.*, 1991
Landers (1992)	7.3	1	112	113	Wald and Heaton, 1994
Northridge (1994)	6.6	17	138	155	Wald *et al.*, 1996
Hyogoken-Nanbu (1995)	6.9	4	14	18	Wald, 1996
Izmit (1999)	7.6	4	13	17	Sekiguchi and Iwata, 2002
Chi-Chi (1999)	7.6	42	172	214	Ji *et al.*, 2003
Denali (2002)	7.8	1	29	30	Tsuboi *et al.*, 2003
Niigataken-Chuetsu (2004)	6.6	9	58	67	Honda *et al.*, 2005
Total		147	623	770	

Moment magnitude (M_w) is cited from the Harvard Centroid Moment Tensor solution. The numbers of near-source and far-source data for each earthquake are also shown. The fault models are used as selection criteria to classify near-source and far-source stations.

KiK-net seismic networks operated by the National Research Institute for Earth Science and Disaster Prevention in Japan. Those data are available at their websites (http://www.knet.bosai.go.jp/ and http://www.kik.bosai.go.jp/). The stations with epicentral distance less than 100 km are used for this analysis.

Data Processing

We processed the accelerograms obtained from the nine earthquakes according to the following method. A bias is removed from the accelerograms by subtracting the preevent mean. Because a small bias has a large effect when the record is integrated, this process is applied to all accelerograms.

The peak amplitudes of the horizontal components are calculated by the square root of the sum of the squares of the peaks of north–south and east–west components. If one of the horizontal components (north–south or east–west) of a station has been clipped or is not well recorded, the square root of twice the other well-recorded horizontal component is used for the peak amplitude of the horizontal component.

The peak amplitude of the up-down component is used directly for the peak vertical component. The station records that have defects in the vertical component are excluded.

The following processes are completed for all the data.

- Jerk: The three-component accelerograms are differentiated in the time domain, using a simple finite-difference approximation. The peak value of each component is selected.
- Acceleration: Original accelerograms are used to select the peak value.
- Velocity: Some velocity records have a linear trend due either to tilting, the response of the transducer to strong shaking, or to problems in the analog-to-digital converter. The baseline correction scheme applied to obtain appropriate velocity records is as follows (Iwan *et al.*, 1985; Boore, 2001): (1) Determine the straight line to be subtracted from the velocity trace. The line is given by the equation $v_f(t) = a_1 t + a_2$, where coefficients a_1 and a_2 are determined by least-squares fitting to the velocity trace after the strong shaking. The segment of the record used for least-squares fitting is from t_1 to t_2 (see Fig. 1). t_1 is the time when the strong shaking has subsided. The results of baseline correction are not very sensitive to the choice of t_1 (Boore, 2001). The second cutoff time, t_2, is generally chosen as the end of the record. (2) Remove this linear trend from the velocity record. The initial time to subtract the linear trend is determined as the intersection between the linear trend and the x axis.

This baseline correction scheme assumes the baseline shift of the acceleration occurs only once. There may be records that have more than one baseline shift during strong shaking. However, our purpose is to get the peak value of each velocity record, and this does not require accurate integration of the entire record. After time-domain integration, the distortion is not very large in the first portion of the record where the peak value is generally recorded.

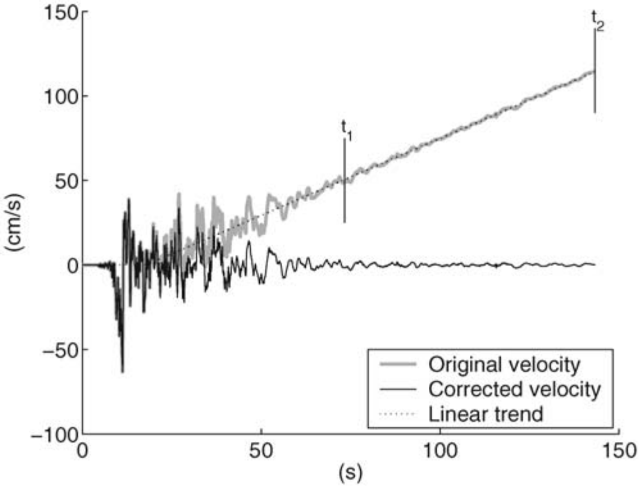


Figure 1. An example of baseline correction for a velocity record from the Chi-Chi earthquake. The corrected velocity trend is obtained by subtracting the linear trend from the original velocity record. The portion of the record from t_1 to t_2 is used for least-squares fitting to obtain the linear trend.

- Displacement: The corrected velocity records are integrated once in the time domain and are high-pass filtered using a fourth-order butterworth filter with a corner frequency of 0.075 Hz.

The peak features used for the classification analysis are shown in Table 2. Several combinations of these eight features are tried to find the best performance of the classification.

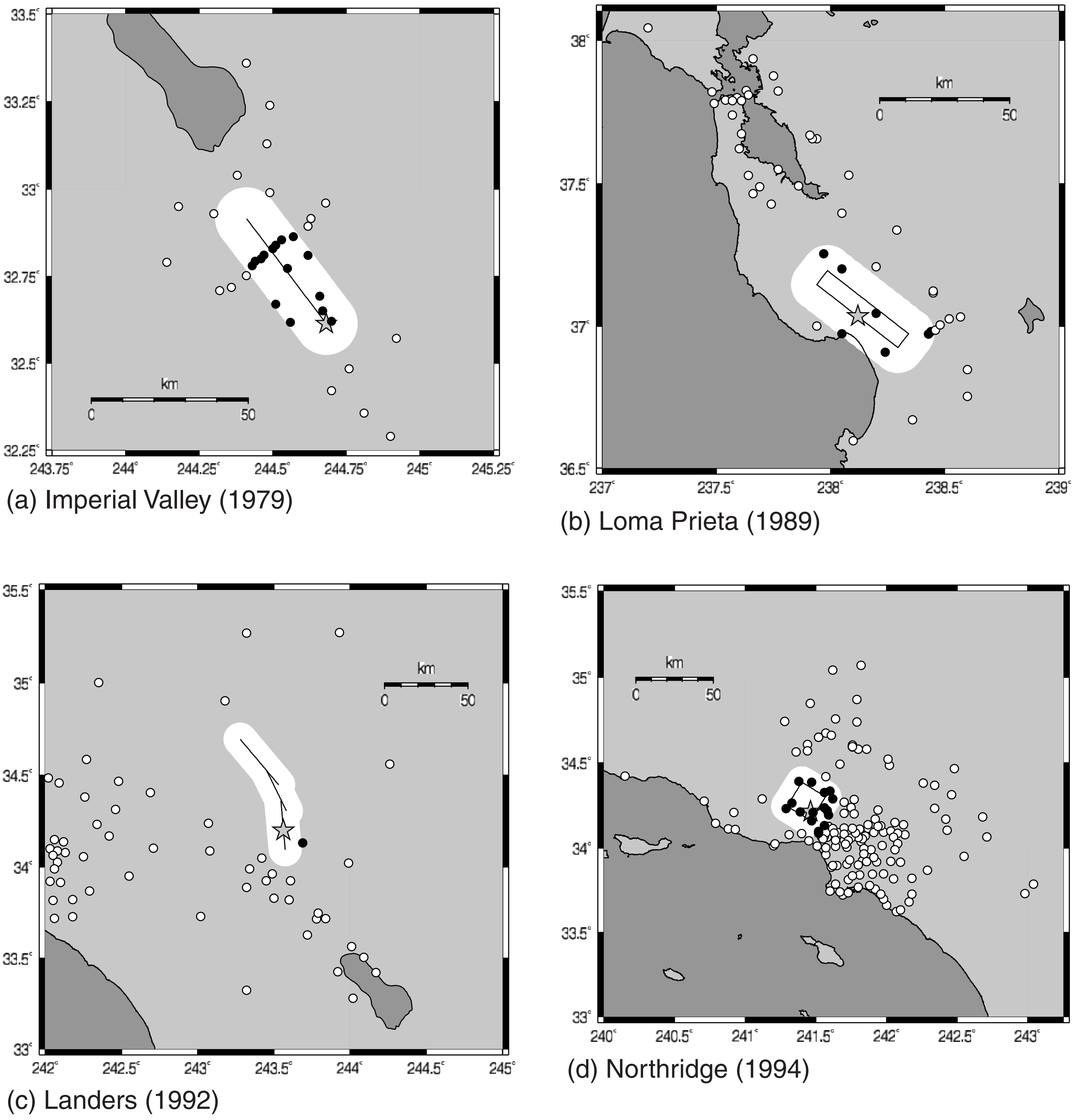
Data Classification

The classification as near source or far source in the training set is based on rupture area models used for waveform inversions. These rupture area models are typically determined from the aftershock distribution (Sekiguchi *et al.*, 1996), and the shape of the rupture area is approximated by a rectangular box. Fault models used for classifying stations

Table 2

Eight Measurements of Peak Ground Motions are Calculated from Three-Component Accelerograms; Codes and Units of the Components Used in This Article are Shown

Code	Measurement	Unit
Hj	Horizontal peak ground jerk	cm/sec^3
Zj	Vertical peak ground jerk	cm/sec^3
Ha	Horizontal peak ground acceleration	cm/sec^2
Za	Vertical peak ground acceleration	cm/sec^2
Hv	Horizontal peak ground velocity	cm/sec
Zv	Vertical peak ground velocity	cm/sec
Hd	Horizontal peak ground displacement	cm
Zd	Vertical peak ground displacement	cm



(a) Imperial Valley (1979)

(b) Loma Prieta (1989)

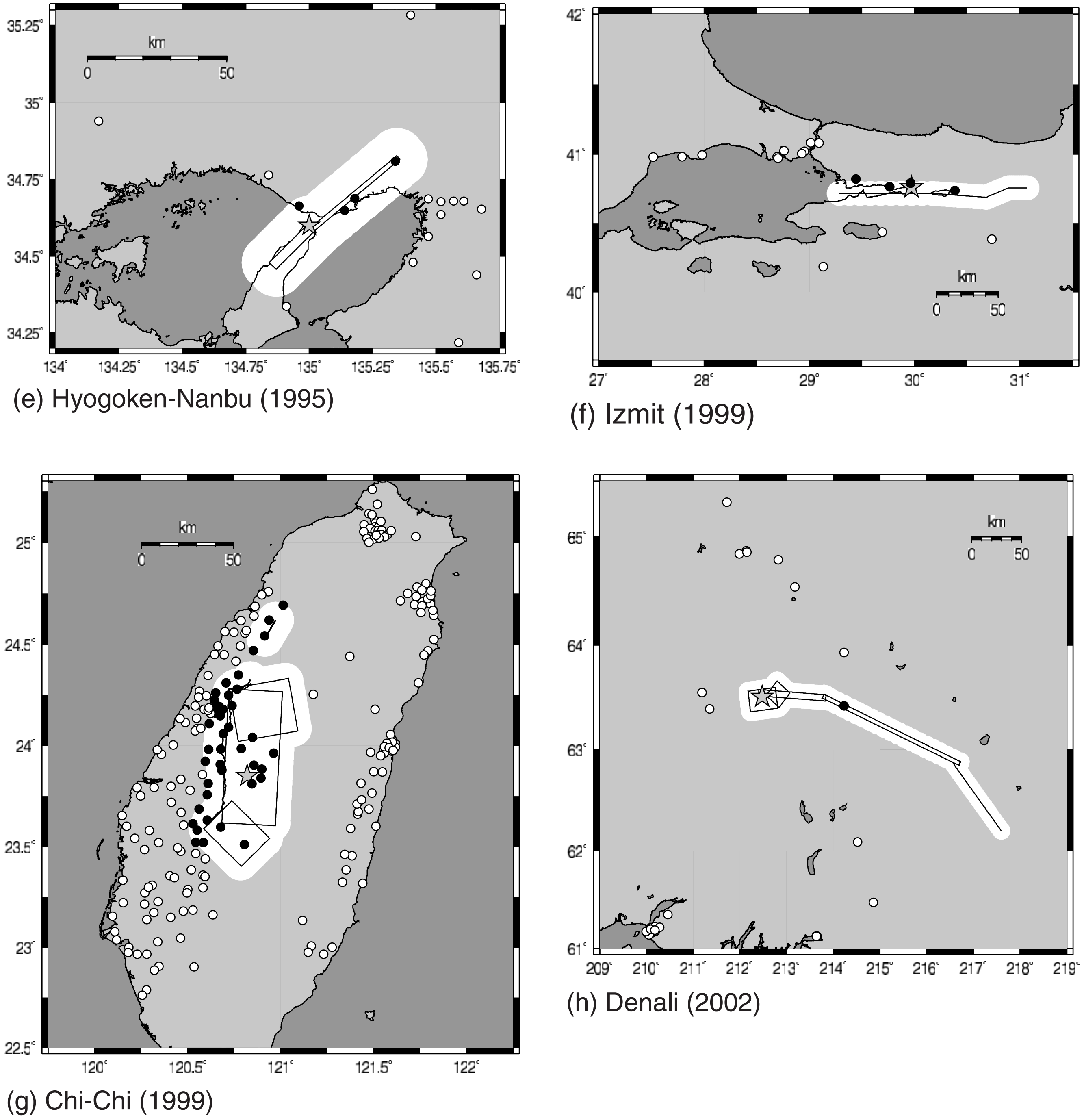
(c) Landers (1992)

(d) Northridge (1994)

Figure 2. Maps of the fault projections and station distributions. The fault projections are shown in the solid lines. The white areas around the fault lines indicate the area with distance less than 10 km from the fault projections. The stations in this area are classified as near source and are marked as solid circles. Far-source stations are shown as open circles. The star symbol denotes the epicenter of the earthquake. (*Continued*)

are shown in Table 1 and Figure 2. In Figure 2, black solid lines indicate the surface projection of the fault rupture surface based on the fault models. Stations within 10 km of this fault projection (the white areas in the figures) are classified as near source, indicated by solid circles. Far-source stations are shown in open circles.

High-frequency near-source ground motions have long been researched by engineers and seismologists. High-frequency ground motions depend weakly on magnitude in the near source (Hanks and Johnson, 1976; Hanks and McGuire, 1981; Joyner and Boore, 1981). This helps to analyze ground motions with a wide range of magnitudes. Fig-



(e) Hyogoken-Nanbu (1995)

(f) Izmit (1999)

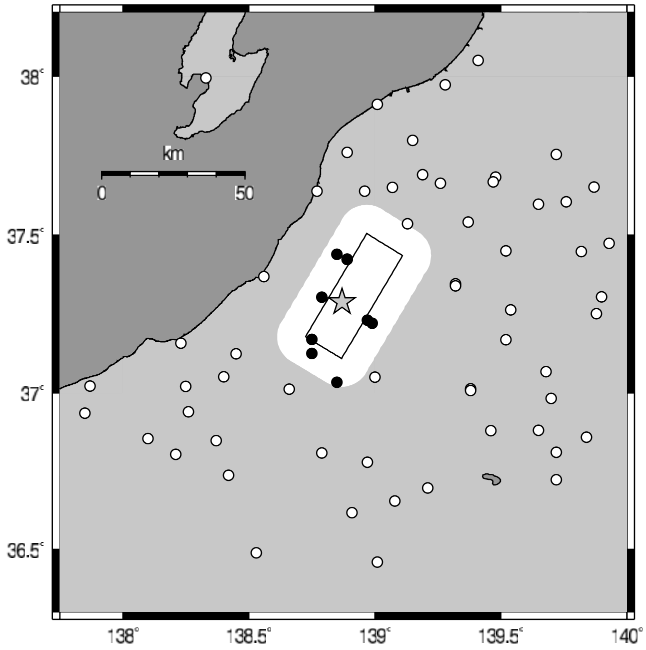
(g) Chi-Chi (1999)

(h) Denali (2002)

Figure 2. Continued.

ure 3 shows horizontal and vertical PGA of near-source records in our training set as a function of moment magnitude. The slope of a regression line would be almost equal to zero, which is consistent with past studies. On the other hand, low-frequency motion has a strong correlation with magnitude. Figure 4 shows horizontal and vertical PGD as a function of moment magnitude. The PGDs are log-proportional to the magnitude. Based on such observations, we assume that high-frequency motion does not depend on magnitude for large earthquake and that accelerations do not exceed 2g, whereas low-frequency motion is highly correlated with magnitude, and its amplitude increases as the magnitude becomes large.

High-frequency ground motion decays in amplitude more rapidly with distance than low-frequency motion (Hanks and McGuire, 1981). Therefore, high-frequency motions (e.g., acceleration and jerk) have high correlations with the fault distance. We compute the log of the ground-motion



(i) Niigataken-Chietsu (2004)

Figure 2. Continued.

amplitudes and find the means and standard deviations for the near-source and far-source records. Figure 5 shows the histograms and Gaussian densities given by the sample means and standard deviations for the near-source and far-source records. The Gaussian densities are good approximations of the histograms of the log of the ground-motion data. Figure 5 also shows that the distance between means for the near-source and far-source datasets is larger in high-frequency than low-frequency motions. Therefore, we expect that the high-frequency motions is a good measure to classify near-source and far-source records.

Near-Source Versus Far-Source Discriminant Function

We assume the discriminant function to classify records into near source and far source is expressed as a linear combination of the log of ground-motion amplitudes:

$$f(X_i|\theta) = c_1x_{i1} + c_2x_{i2} + \cdots + c_mx_{im} - d = \sum_{k=1}^{m} c_kx_{ik} - d = X_i \cdot c - d, \qquad (1)$$

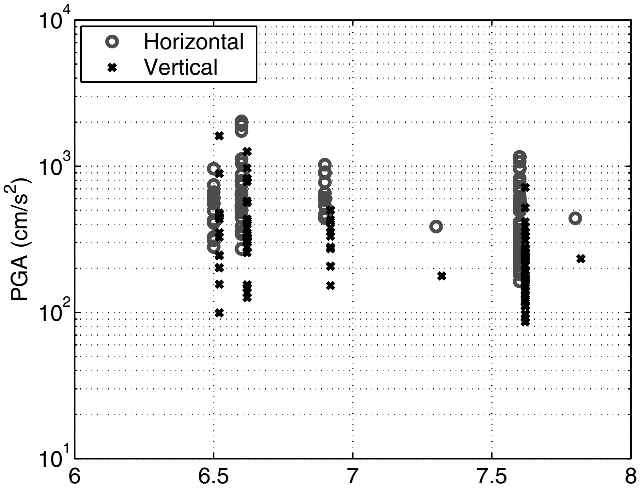


Figure 3. Distribution of horizontal and vertical PGA for near-source stations with respect to magnitude.

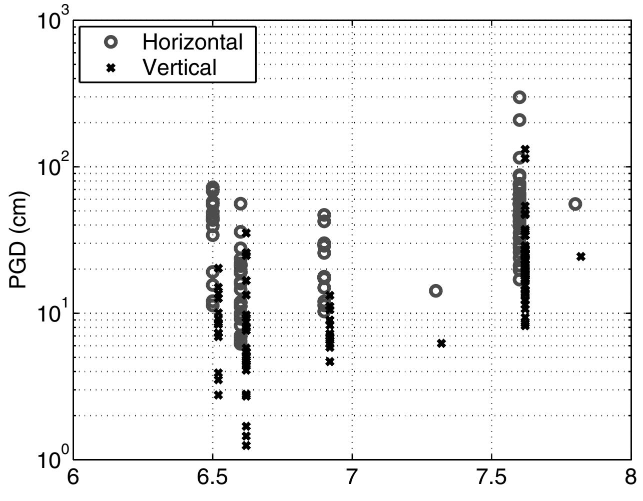


Figure 4. Distribution of horizontal and vertical PGD for near-source stations with respect to magnitude.

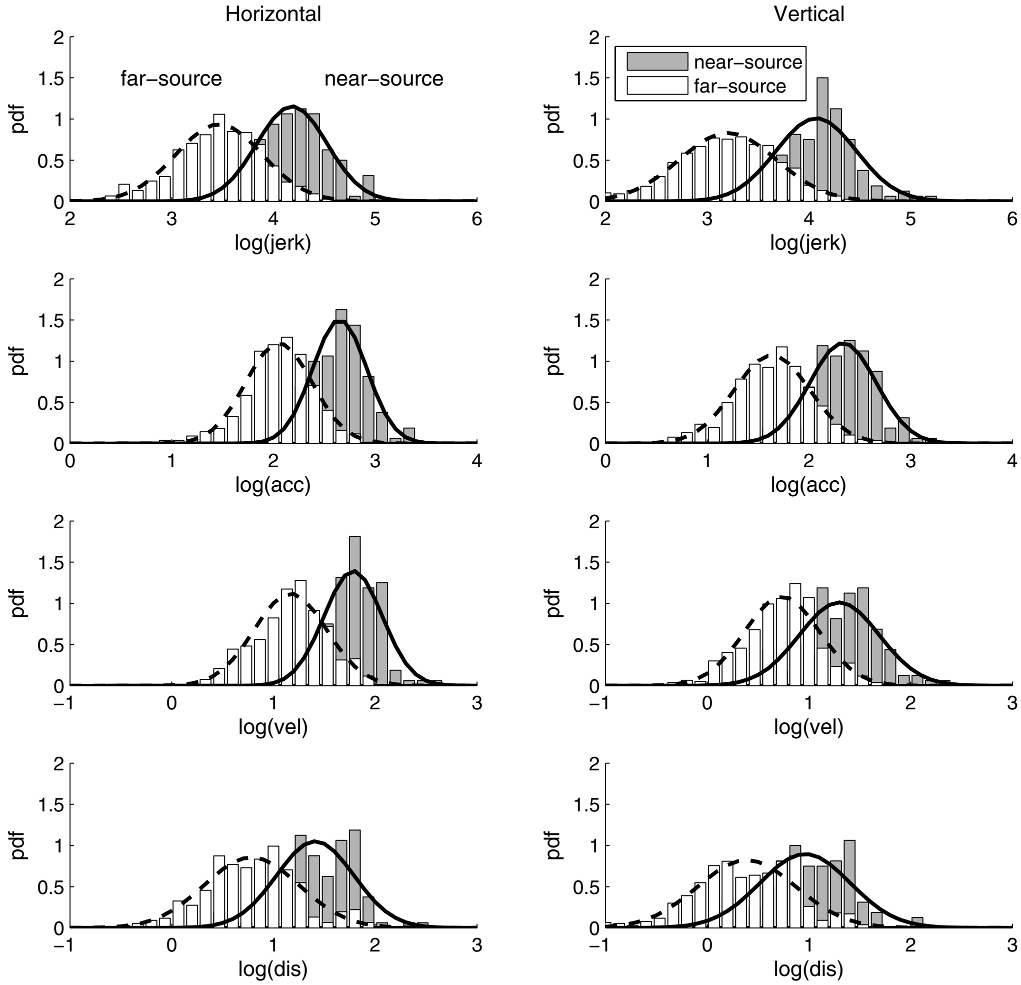


Figure 5. Histograms and Gaussian densities based on the sample means and standard deviations of the log of ground motions for the near-source and far-source records. These are distributions for jerk, acceleration, velocity, and displacement from the top.

where x_{ik} is the kth feature parameter of the ground motion at the ith station, m is the number of feature parameters, $X_i = [x_{i1}, x_{i2}, \ldots, x_{im}] = [\log_{10}(\text{component 1}), \log_{10}(\text{component 2}), \ldots, \log_{10}(\text{component } m)]$, $c_1, \ldots, c_m$ is the regression coefficients, d is the decision boundary constant, and $\theta = [c_1, c_2, \ldots, c_m, d]^T$. We may use m components out of the eight ground-motion components shown in Table 2. The coefficients $c_1, \ldots, c_m$, and d are determined from the training dataset by two different approaches: Fisher's linear discriminant analysis and Bayesian analysis.

This discriminant function is used to allocate new observations to one of the near-source or far-source groups, where $f(X_i|\theta) = 0$ is the boundary between the two groups in the feature parameter space. The station with observation X_i is classified as near source if $f(X_i|\theta)$ is positive. If $f(X_i|\theta)$ is negative, the station is classified as a far-source station. Note that the decision boundary may also be expressed using equation (1) as $X_i \cdot c = d$.

Fisher's Linear Discriminant Analysis

Fisher's linear discriminant analysis (LDA) is a method to classify data by using a linear function (1) that best discriminates two or more naturally occurring groups. LDA was first described by Fisher (1936) to separate two groups optimally. In general, LDA requires placing objects (e.g., humans) in predefined groups (e.g., Caucasoid, Mongoloid, and Negroid) based on certain feature parameters (e.g., related to physical characteristics) and finding a function to distinguish the groups. The parameters c_k in the linear function (1) are selected to minimize the within-group variance (variance of the samples centered on the group mean) and to maximize the between-group variance (variance of the between-group means). The following is a brief discussion about the procedure of linear discriminant analysis (Venables and Ripley, 2002).

Consider $n \times m$ data matrix X, where n is the number of samples and m is the number of different features of samples.

Each sample is assigned to one of g groups N_j, $j = 1, \ldots, g$, with n_j observations in each group. Let G denote the group indicator matrix, which indicates the group each sample is assigned to, and let M denote the group mean matrix. Then the within-group covariance matrix W and between-group covariance matrix B are

$$W = \frac{(X - GM)^T(X - GM)}{n - g}, \tag{2}$$

$$B = \frac{(GM - \mathbf{1}\mu)^T(GM - \mathbf{1}\mu)}{g - 1}, \tag{3}$$

where $X = [x_{ik}]$ is the $n \times m$ data matrix, $G = [g_{ij}]$ is the $n \times g$ group indicator matrix, $M = [m_{jk}]$ is the $g \times m$ group mean matrix, $\mu = [\mu_1, \mu_2, \ldots, \mu_m]$ is the $1 \times m$ mean vector, $\mathbf{1}$ is the $n \times 1$ column vector of 1s, x_{ik} is the kth feature of the ith sample, $g_{ij} = 1$ if and only if $X_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$ is assigned to group j, $m_{jk} = (1/n_j)\sum_{i \in N_j} x_{ik}$, and $\mu_k = (1/n)\sum_{i=1}^{n} x_{ik}$. We would like to find a linear combination $X \cdot c$ of the data such that the different groups are maximally separated, that is, maximizing the following separation ratio λ:

$$\lambda = \frac{c^T B c}{c^T W c} = \frac{\text{between-group variance}}{\text{within-group variance}}. \tag{4}$$

A necessary condition to maximize λ is $\partial\lambda/\partial c = 0$. By substituting equation (4) into this condition, we get

$$W^{-1}Bc = \lambda c, \tag{5}$$

assuming W is invertible. This is an eigenvalue problem, and the weight vector c and the separation ratio λ are eigenvectors and eigenvalues of $W^{-1}B$, respectively. $X \cdot c$ is called a canonical variate, and the canonical variate of the eigenvector c that corresponds to the largest eigenvalue is called the first canonical variate.

For the near-source versus far-source classification problem, the data matrix X is the dataset of peak seismic ground motions, where n is the number of stations and m is the number of the object features (PGA, PGV, PGD, etc.). We have two groups: the near-source group and the far-source group ($g = 2$). LDA finds the optimal set of coefficients of ground-motion amplitudes to classify near-source or far-source records.

Because the traditional LDA does not treat which choice of the ground-motion parameters is the best, Bayesian model class selection is performed (the results are shown later). According to this analysis, the best selection is the combination of Za (vertical acceleration) and Hv (horizontal velocity), and their coefficients obtained from LDA are shown in Table 3.

Table 3

Estimated Model Parameters by Fisher's LDA, Bayesian Approach with Asymptotic Approximation, and Bayesian Approach with Metropolis Algorithm; the Standard Deviations for Each Parameter Are Shown in Parentheses

Method	c_1 (Za)	c_2 (Hv)	d
LDA	7.233	6.813	25.903
Bayesian–asymptotic	6.046	7.886	27.090
(σ)	(±0.903)	(±1.206)	(±3.163)
Bayesian–Metropolis algorithm	6.194	8.150	27.872
(σ)	(±0.946)	(±1.224)	(±3.330)

We choose the decision boundary constant d to maximize the classification performance for the set of coefficients obtained by the LDA. The classification performance is evaluated by the following function:

$$\begin{aligned} P_c(d) = [&P(f(X_i|\theta) \geq 0 | Y_i = 1) \\ &+ P(f(X_i|\theta) \\ &< 0 | Y_i = -1)]/2, \end{aligned} \tag{6}$$

where

$$f(X_i|\theta) = X_i \cdot c - d, \qquad Y_i = \begin{cases} 1 & \text{if near source,} \\ -1 & \text{if far source.} \end{cases}$$

This is the average probability between the probability that a near-source station is classified correctly and the probability that a far-source is classified correctly. The parameter d, which maximizes this function for the given coefficients (Table 3), is 25.903, and the performance defined by the preceding function is 93.4%. Another way to compute d is to take the midpoint of the two group means of the first canonical variate. This method makes it easier to compute the value of d, and it gives $d = 25.045$, a good approximation to $d = 25.903$ that shows maximum performance.

As a conclusion, the discriminant function computed from the LDA is

$$f(X_i|\theta) = 7.233 \log_{10} \mathrm{Za} + 6.813 \log_{10} \mathrm{Hv} - 15.903$$
$$\text{if} \begin{cases} f(X_i|\theta) \geq 0 & \text{near source,} \\ f(X_i|\theta) < 0 & \text{far source.} \end{cases} \tag{7}$$

This discriminant function is applied to all the datasets, and the values of $f(X_i|\theta)$ are shown in Figure 6. The figure shows that most of the near-source data lie on the right-hand side of the decision boundary, which means the classification performance is very good. Although a fraction of the far-source records are misclassified, the misclassification of far-source data is less critical than that of near-source data.

Bayesian Approach

In this section, a Bayesian approach is applied to determine the coefficients of the discriminant function that classifies near-source and far-source data (Sivia, 1996; Jaynes, 2003). The probability density function (PDF) of parameter

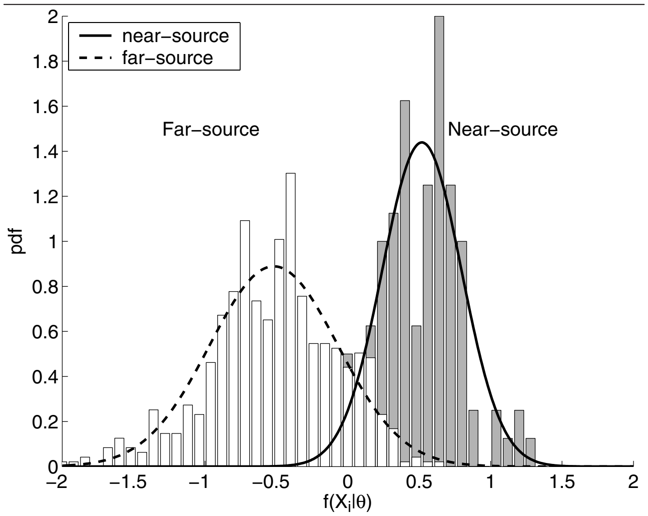


Figure 6. Histogram of the near-source and far-source data to which the discriminant function obtained from traditional LDA is applied. The column heights are normalized by the number of the data in each group. $f(X_i|\theta) = 0$ is the decision boundary between the two groups. The curves are the Gaussian distribution with the same mean and standard deviation as the values of $f(X_i|\theta)$ for each group.

θ conditioned on data D_n and model class M can be expressed using Bayes's theorem:

$$p(\theta|D_n, M)_{\text{posterior}} \propto p(D_n|\theta, M)_{\text{likelihood}} \times p(\theta|M)_{\text{prior}} \propto \prod_{i=1}^{n} P(Y_i|X_i, \theta) \times p(\theta|M), \quad (8)$$

where $\theta = [c_1, c_2, \dots, c_m, d]^T$ is the parameter vector, $D_n = \{(X_i, Y_i): i = 1, \dots, n\}$ is the available set of data, $X_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ is the ground motion at the station i and is equal to [$\log_{10}$(component 1), $\log_{10}$(component 2), ..., $\log_{10}$(component m)], $Y_i = 1$ if the classification is near source at the station i, $Y_i = -1$ if the classification is far source at the station i, m is the number of object features, and n is the number of data. Note that the model class M defines the likelihood for each value of θ in some set of values and also the prior PDF $p(\theta)$.

We determine the parameters $c_1, \dots, c_m$, and d based on a Bayesian approach using the same notation as the LDA. The goal of the Bayesian approach is to obtain the posterior PDF of the model parameters (θ) and to determine the most plausible value of θ by maximizing this PDF.

Choice of Prior Distribution

Assume that the model class M is globally identifiable based on D_n (Beck and Katafygiotis, 1998), that is, there is a unique θ maximizing the likelihood $p(D_n|\theta, M)$. In this case, given a sufficiently large dataset D_n, the choice of prior PDF does not affect the resulting posterior PDF, and all posteriors with different priors will converge to the same answer (Sivia, 1996). Here, the prior is chosen to cover a wide range of the parameter space by selecting the prior of each model parameter to be a Gaussian PDF with zero mean and standard deviation $\sigma = 100$, so

$$p(\theta|M) = \frac{1}{(\sqrt{2\pi}\sigma)^{m+1}} \exp\left(-\frac{1}{2\sigma^2}\theta^T\theta\right) = \frac{1}{(\sqrt{2\pi}\sigma)^{m+1}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{k=1}^{m} c_k^2 + d^2\right)\right]. \quad (9)$$

Choice of Likelihood Function

Let the predictive probability that station i is near source be $P(Y_i = 1|X_i, \theta)$. The predictive probability that a station is far source is then $P(Y_i = -1|X_i, \theta) = 1 - P(Y_i = 1|X_i, \theta)$. A standard approach in Bayesian classification is to define the predictive probability by applying the logistic sigmoid function $\phi(x) = 1/(1 + e^{-x})$ to the linear function $f(X_i|\theta)$ that is also used in the traditional LDA (Li *et al.*, 2002). The logistic sigmoid function is a smooth, positive, and monotonically increasing function, as shown in Figure 7. Although there are other sigmoid functions that have these properties, the logistic sigmoid function is mathematically convenient, and the class probability (shown in the Bayesian model class selection) is robust to the choice of the sigmoid function for the following reason. Notice from equation (1) that the location of the separating boundary $f(X_i|\theta) = 0$ is independent of a uniform scaling of the parameters. The Bayesian updating automatically produces a scaling appropriate to the separation of the classes in the feature parameter

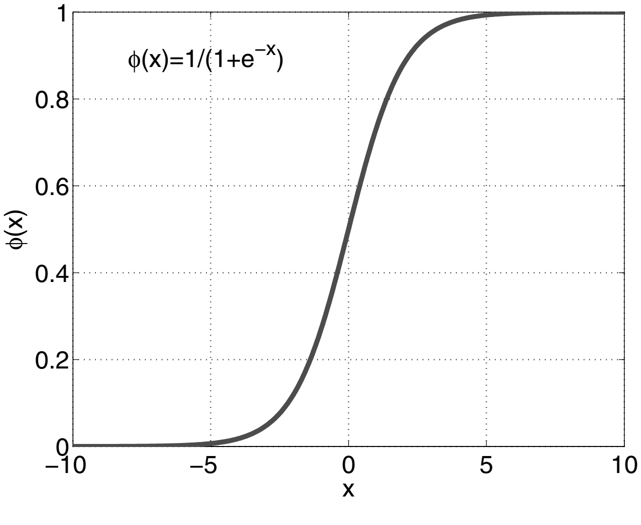


Figure 7. A logistic sigmoid function $\phi(x) = 1/(1 + e^{-x})$ is used to express the predictive probability for classification. The function approaches zero as $x \to -\infty$ and approaches one as $x \to \infty$. The function is 0.5 when x is zero.

space that is implied by the data. If the data are well separated, a large scale will be chosen so that there is a steep transition in the class probability as the separating boundary is crossed; on the other hand, if the data for the classes have significant overlap in the feature parameter space, then a smaller scale will be chosen to give a more gradual transition.

The predictive probability that the *i*th station is near source is therefore defined here by

$$P(Y_i = 1|X_i, \theta) = \phi(f(X_i|\theta)) = \frac{1}{1 + e^{-f(X_i|\theta)}}. \qquad (10)$$

As $f(X_i|\theta)$ becomes larger, the station is more likely to be near source, and the probability that the station is near source becomes closer to one. Note that the predictive probability that the station is far source is then

$$P(Y_i = 1|X_i, \theta) = 1 - \phi(f(X_i|\theta)) = \phi(-f(X_i|\theta)) = \frac{1}{1 + e^{f(X_i|\theta)}}, \qquad (11)$$

where from equation (1),

$$f(X_i|\theta) = \sum_{k=1}^{m} c_k x_{ik} - d = X_i \cdot c - d.$$

From equations (10) and (11), the likelihood function can be expressed as

$$p(D_n|\theta, M) = \prod_{i=1}^{n} P(Y_i|X_i, \theta) = \prod_{i=1}^{n} \phi(Y_i f(X_i|\theta)) = \prod_{i=1}^{n} \frac{1}{1 + e^{-Y_i f(X_i|\theta)}}. \qquad (12)$$

Posterior Distribution

By substituting equations (9) and (12) into equation (8), the posterior can be expressed as

$$p(\theta|D_n, M) \propto \frac{1}{(\sqrt{2\pi}\sigma)^{m+1}} \exp\left(-\frac{1}{2\sigma^2}\theta^T\theta\right) \times \prod_{i=1}^{n} \frac{1}{1 + e^{-Y_i f(X_i|\theta)}}. \qquad (13)$$

Both an asymptotic approximation and stochastic simulation are performed to characterize the PDF defined by equation (13). In the asymptotic approach, the posterior is represented by a Gaussian distribution for θ with mean $\hat{\theta}$, the most probable value of θ, and a covariance matrix $\hat{\Sigma}$ defined later. Stochastic simulation uses the Metropolis algorithm to generate random samples of the parameter vector θ from the posterior PDF. It is noted that it is computationally challenging to evaluate the proportionality constant in equation (13) that normalizes the posterior PDF because it requires numerical integration over a high-dimensional parameter space. However, this evaluation can be avoided in both the asymptotic approximation and stochastic simulation methods.

Asymptotic Approximation

We first find the optimal value $\hat{\theta}$ of θ that maximizes the posterior PDF. This multidimensional optimization problem is solved by a numerical optimization algorithm provided by Matlab.

Using Laplace's method of asymptotic approximation, Beck and Katafygiotis (1998) show that the posterior PDF for a set of model parameters θ for a globally identifiable model class M (which has a unique most probable value) may be approximated accurately by a Gaussian distribution with mean $\hat{\theta}$ and covariance matrix $\hat{\Sigma}$, given a large amount of data. Define $H(\theta)$ by

$$H(\theta) = -\nabla\nabla \log[p(D_n|\theta, M)p(\theta|M)] = -\nabla\nabla \log\left[\prod_{i=1}^{n} P(Y_i|X_i, \theta)p(\theta|M)\right], \qquad (14)$$

then $\hat{\Sigma} = H(\hat{\theta})^{-1}$. By substituting equations (9) and (12) into equation (14),

$$\begin{aligned}[H(\theta)]_{(\alpha,\beta)} &= \left[-\nabla\nabla \log \prod_{i=1}^{n} P(Y_i|X_i, \theta) - \nabla\nabla \log p(\theta|M)\right]_{(\alpha,\beta)} \\ &= -\frac{\partial^2}{\partial c_\alpha \partial c_\beta}\left(\log \prod_{i=1}^{n} \phi_i\right) + \frac{1}{\sigma^2}\delta_{\alpha\beta} \\ &= -\sum_{i=1}^{n} \frac{\partial^2}{\partial c_\alpha \partial c_\beta}(\log \phi_i) + \frac{1}{\sigma^2}\delta_{\alpha\beta} \\ &= -\sum_{i=1}^{n} \frac{\partial}{\partial c_\beta}\left[\frac{1}{\phi_i}\phi_i(1 - \phi_i)\frac{\partial(Y_i f(X_i|\theta))}{\partial c_\alpha}\right] \\ &\quad + \frac{1}{\sigma^2}\delta_{\alpha\beta} = \sum_{i=1}^{n} \phi_i(1 - \phi_i)x_{i\alpha}x_{i\beta} \\ &\quad + \frac{1}{\sigma^2}\delta_{\alpha\beta},\end{aligned} \qquad (15)$$

where $\phi_i = \phi(Y_i f(X_i|\theta))$ and equation (1), along with $Y_i^2 = 1$, has been used. The optimal parameter values and their standard deviations for the selection of features Za and Hv are shown in Table 3. Note that for large σ, the effect of the prior in equation (15) is negligible.

In order to examine the sensitivity of the Bayesian approach to the training dataset, we perform a cross-validation analysis. First, the training dataset is randomly divided into two datasets, and the discriminant function is constructed from one dataset (training set). This discriminant function is applied to the other dataset (validation set) to check its classification performance. We then switch the testing set and validation set and repeat this cross-validation analysis. We set the near-source/far-source boundary so that the probability is $1/2$ that the station is near source, that is, the station is classified as near source if the probability that it is near source is more than $1/2$. The confusion matrices of these two analyses and the previous analysis that uses all of the dataset are shown in Table 4. The classification error with half of the dataset is as small as that of the analysis that uses all of the dataset. Therefore, we confirm that the sensitivity to the training dataset is small, giving more confidence that the discriminant function from Bayesian analysis will perform well for future earthquake data.

Stochastic Simulation using Metropolis Algorithm

The asymptotic approximation is valid only if the posterior PDF for the model parameters can be approximated well with a Gaussian distribution. This requires a large sample size and that the class of models M is globally identifiable based on data D_n (Beck and Katafygiotis, 1998). On the other hand, a stochastic simulation algorithm can be applied to the problem that generates samples from a Markov chain, whose stationary PDF is the posterior PDF, that is, the samples are asymptotically distributed according to the posterior PDF for the parameters. The Metropolis algorithm is used to solve this high-dimensional problem, because it does not require evaluation of the normalizing constant for sampling the posterior PDF in equation (13).

The Metropolis algorithm is a Markov chain Monte Carlo (MCMC) method proposed by Metropolis *et al.* (1953). It is a simulation technique for generating random samples from any given probability distribution. The algorithm uses a proposal PDF Q that depends on the current sample of parameters, $\theta^{(t)}$ at the tth iteration (MacKay, 1998). Here, we choose as the proposal density a Gaussian PDF centered on the current parameters $\theta^{(t)}$ with the covariance matrix Σ of the parameters in the asymptotic approximation. The optimal parameters estimated from direct optimization of the posterior PDF are used as an initial $\theta^{(t)}$. The expression for Q is

$$Q(\theta'|\theta^{(t)}) = \frac{1}{(2\pi)^{m'/2}|\Sigma|^{1/2}} \times \exp\left[-\frac{1}{2}(\theta' - \theta^{(t)})^T\Sigma^{-1}(\theta' - \theta^{(t)})\right], \quad (16)$$

where $|\Sigma|$ is the determinant of the covariance matrix and m' is the dimension of the parameter θ, which is $m + 1$. A candidate sample is drawn from $Q(\theta'|\theta^{(t)})$. The ratio of the posterior PDF in equation (8) at the current sample $\theta^{(t)}$ and the candidate sample θ' determines whether to accept the candidate sample, according to

$$r = \frac{p(\theta'|D_n, M)}{p(\theta^{(t)}|D_n, M)}, \quad (17)$$

$$\theta^{(t+1)} = \begin{cases} \theta' & \text{with probability } \min(1, r), \\ \theta^{(t)} & \text{with probability } 1 - \min(1, r). \end{cases} \quad (18)$$

If $r \geq 1$, then the candidate is accepted as the next sample in the Markov chain. Otherwise, the candidate is accepted with probability r as follows: We generate a random number uniformly distributed between zero and one, and if it is less than r, the candidate is accepted, that is, $\theta^{(t+1)} = \theta'$. If it is not accepted, the current sample is repeated ($\theta^{(t+1)} = \theta^{(t)}$). This procedure is repeated until the desired number of samples are generated. There is a burn-in period at the beginning of the MCMC method until the probability distribution of the current sample $\theta^{(t)}$ is sufficiently close to the posterior PDF, which is the stationary PDF of the Markov chain, so judgment is used to discard initial samples.

Figure 8 shows 5000 parameter samples generated with the Metropolis algorithm for the optimal selection of features Za and Hv. This selection of the ground-motion features comes from Bayesian model class selection explained later. After discarding the samples in the burn-in period (taken as the first 100 samples), the mean and standard deviation of the samples are shown in Table 3. The average acceptance ratio of the candidate samples θ' is 44%, which indicates the method works well (Roberts *et al.*, 1997). The stability of the sample mean and standard deviation of each parameter is examined in Figure 9. The mean and standard deviation of the samples converge after the first 1000 samples are added. The most probable values of the parameters from maximization of the posterior PDF are also shown in Figure 9. Note that the means of the marginal PDFs and the most prob-

Table 4
The Confusion Matrix for the Cross-Validation Analysis with the Bayesian Method with Asymptotic Approximation

Dataset	Near Source/Far Source	Near Source	Far Source
All dataset	Near source	78 (78%)	22 (22%)
	Far source	12 (2%)	583 (98%)
Half of dataset	Near source	39 (74%)	14 (26%)
	Far source	4 (1%)	291 (99%)
Other half of dataset	Near source	37 (79%)	10 (21%)
	Far source	8 (3%)	292 (97%)

"All dataset" is the analysis that uses the whole dataset as a training set and a validation set. "Half of dataset" is the analysis that uses half of the dataset as a training set and the other half as a validation set. "Other half of dataset" is the analysis that switches the training and validation set.

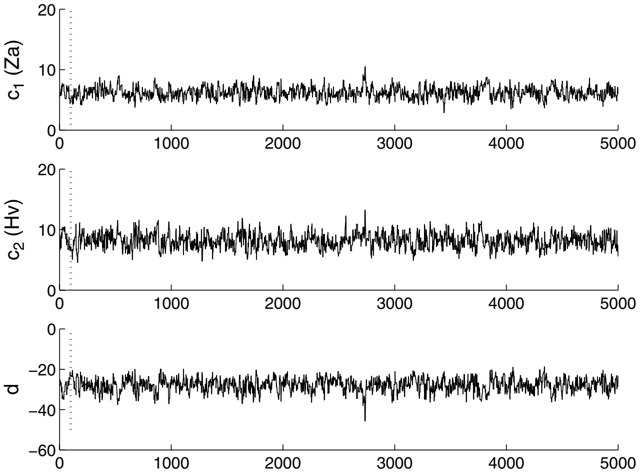


Figure 8. Samples generated by the Metropolis algorithm plotted in the parameter space. The x axis denotes the sample number. The vertical dotted lines indicate the end of the burn-in period (100 samples).

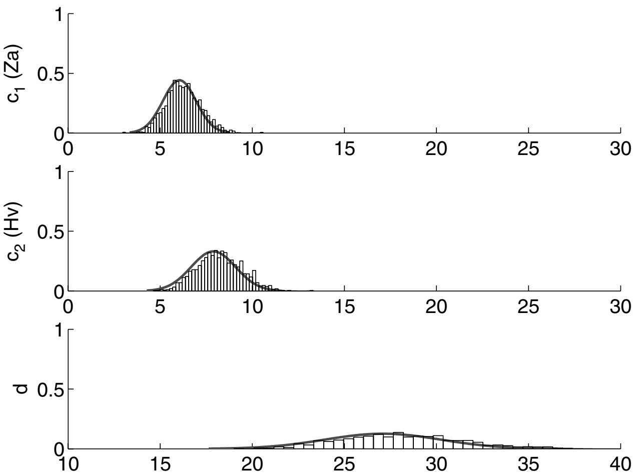


Figure 10. Distribution of samples for three parameters generated by the Metropolis algorithm. The Gaussian distributions obtained from the asymptotic approximation are added in the figure and fit the histogram well.

able values of the joint posterior PDF need not agree if these PDFs are skewed.

The distribution of sample values for parameters θ and the resulting histogram of probability that a station is near source calculated by the generated set of parameters are shown in Figure 10. The distribution of parameter samples agrees well with the Gaussian distribution defined by the optimal parameters and standard deviations given by the asymptotic approximation. The standard deviations of c_1 and c_2 are similar to each other, and the distribution is peaked close to the mean of the samples. The distribution of samples for the decision boundary constant d has a standard deviation almost three times as large as that of c_1 and c_2. However, in terms of th coefficient of variation, the uncertainty in d is smaller than that of other parameters (11.7% compared with 14.9% and 15.3% for c_1 and c_2, respectively).

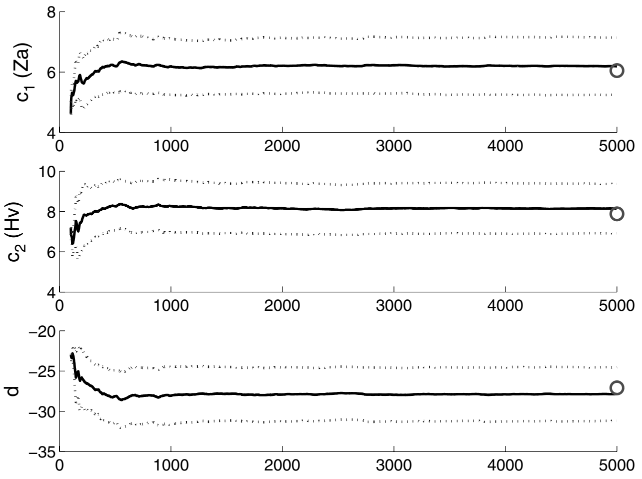


Figure 9. Mean and standard deviation of samples plotted against the number of samples included (excluding the first 100 samples). The solid lines are the sample mean, and the dashed lines represent the mean plus and minus one standard deviation. The small circles are the most probable values of the model parameters estimated from optimization.

Figure 11 shows the correlation of samples of model parameters generated from the simulation. This is the result of the model class with all parameters corresponding to the eight ground-motion parameters, not the result of the optimal model class. The figure shows that the parameter d is not correlated significantly with any other parameter. The combinations of parameters that have significant interaction are horizontal and vertical jerk (c_1 and c_2), horizontal and vertical acceleration (c_3 and c_4), and horizontal and vertical displacement (c_7 and c_8). Parameters with the same component and similar frequency range (e.g., jerk and acceleration [c_1 and c_3, and c_2 and c_4], acceleration and velocity [c_3 and c_5, and c_4 and c_6], and velocity and displacement [c_5 and c_7, and c_6 and c_8]) are also strongly correlated. This result agrees with our intuition: horizontal and vertical components of the same quantity are correlated, and records with similar frequency ranges have similar attenuation relationships and so are correlated.

Comparison between Traditional LDA and the Bayesian Approach

Parameters for the linear discriminant function $f(X_i|\theta) = \sum_{k=1}^{m} c_k x_{ik} - d$ are estimated by traditional LDA and by the Bayesian approach with two different techniques to characterize the posterior PDF. The results are shown in Table 3. The parameters for LDA are scaled such that the norm of the vector $c = [c_1, c_2]$ is equal to the norm of the vector from the asymptotic approximation. Note that the discriminant function $f(X_i|\theta)$ is a linear function, so for the

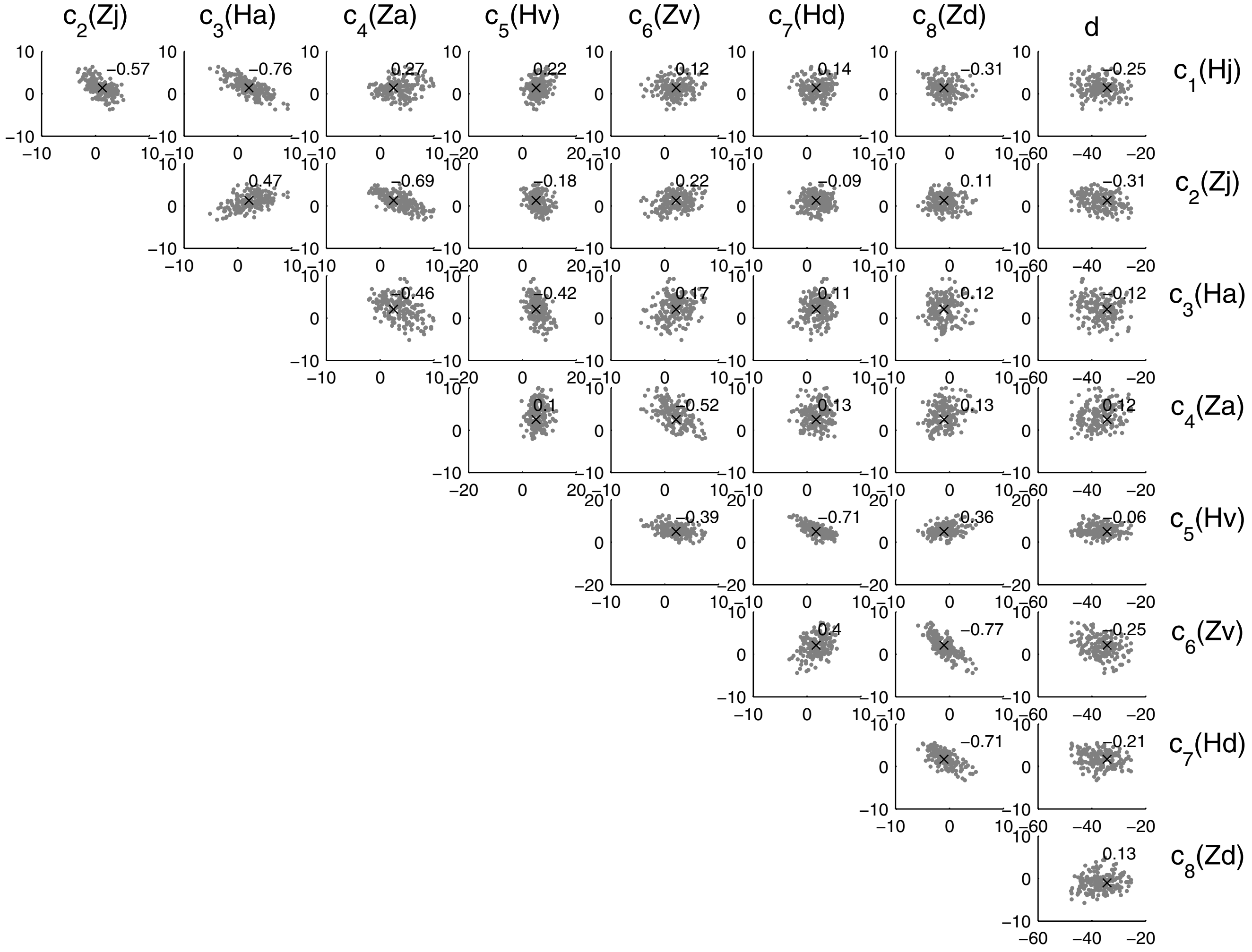


Figure 11. Correlation plot of posterior samples of the model parameters generated by the Metropolis algorithm. The most probable values of the parameters are shown as crosses (×). The numbers in the figure are the correlation coefficients of the parameters.

traditional LDA, multiplying all c_k and d by an arbitrary positive constant does not change the result of classification. However, this is not true for the Bayesian approach, where the modulus of $f(X_i|\theta)$ affects the probability that a station is near source.

The estimated parameters are close for the three methods. The coefficients from LDA are within one standard deviation of those from both Bayesian methods, except that c_1 from LDA is slightly over one standard deviation from the corresponding mean and most probable values from the Bayesian methods.

For the asymptotic approximation and Metropolis algorithm, the estimates and standard deviations for the posterior parameter distribution are very close. If the posterior is a skewed PDF, the mean is not necessarily equal to the most probable value (e.g., consider lognormal distribution), as mentioned before. However, Figure 10 suggests that the posterior PDF is almost symmetric, and the means of the samples and the most probable values should show very good agreement. In this case, the Gaussian distribution is a good approximation for the posterior PDF of the parameters.

By using the discriminant functions defined by the values of the parameters in Table 3, we performed a classification analysis using the whole dataset. The classification performance for the discriminant function from LDA and two Bayesian approaches is shown in Table 5. The results for LDA show 100% of near-source data and 86% of far-source data are classified correctly, and the result of Bayesian approach shows 78% of near-source data and 98% of far-source data are classified correctly. This discriminant function is the function that has the smallest prediction error. To obtain this function, the misclassification of near-source data and that of far-source data are considered to be of equal importance. Generally speaking, the misclassification of near-source data is more critical than that of far-source data, and we may want to decrease the misclassification rate of near-

Table 5
The Confusion Matrix for Near-Source versus Far-Source Classification by the Discriminant Function Obtained from LDA, from the Bayesian Approach with Asymptotic Approximation, and from the Bayesian Approach with Metropolis Algorithm

Dataset	Near Source/Far Source	Near Source	Far Source
LDA	Near source	100 (100%)	0 (0%)
	Far source	82 (14%)	513 (86%)
Bayesian–assymptotic	Near source	78 (78%)	22 (22%)
	Far source	12 (2%)	583 (98%)
Bayesian–Metropolis algorithm	Near source	78 (78%)	22 (22%)
	Far source	12 (2%)	583 (98%)

source data. This misclassification rate can be easily controlled by changing the decision boundary constant d. We also can control this by shifting the near-source/far-source boundary in the Bayesian approach to correspond to some other probability than the half used in this classification analysis.

We performed the leave-one-out cross validation to compare the misclassification rate between LDA and the Bayesian method with asymptotic approximation. The idea of this method is to predict the probability of a station from the discriminant function constructed from the dataset from which that station is excluded. This process is repeated for all 695 data, and the accuracy of prediction is computed. The percentage of misclassified data is shown in Table 6. It shows the prediction error of the Bayesian approach is much smaller than that of LDA. In other words, the Bayesian approach is able to construct a more robust discriminant function. Therefore, we use the discriminant function obtained from the Bayesian method with asymptotic approximation for further analysis.

Table 6
Results of Leave-One-Out Cross-Validation for LDA and the Bayesian Approach

Method	Prediction	Error
LDA	82/695	(12%)
Bayesian approach	36/695	(5%)

Bayesian Model Class Selection

Method

Bayesian model class selection determines which combination of the eight ground-motion parameters gives the best classification for the near source and far source. The essential idea is to find the most probable model class based on data D_n within a set of candidate model classes M_j, $j = 1, \ldots, J$ (Gull, 1988; Beck and Yuen, 2004). Applying Bayes's theorem, the probability of model class M_j can be expressed as follows:

$$P(M_j|D_n, M) = \frac{p(D_n|M_j)_{\text{evidence}} P(M_j|M)_{\text{prior}}}{p(D_n|M)_{\text{normalizing constant}}}, \qquad (19)$$

where $M = \{M_1, M_2, \ldots, M_J\}$ is a set of candidate model classes and J is the number of model classes. The left-hand side of equation (19) is the probability of a particular model class M_j given the dataset and a set of candidate model classes. On the right-hand side, $p(D_n|M_j)$ is the evidence for each model class, $P(M_j|M)$ is the prior over the candidate model classes evaluated for M_j, and $p(D_n|M)$ is a normalizing constant given by

$$a(D_n|M) = \sum_{j=1}^{J} p(D_n|M_j)P(M_j|M). \qquad (20)$$

Assuming a uniform prior for the model class, $P(M_j|M)$ in the numerator and denominator of equation (19) cancel. By the total probability theorem, the evidence for M_j provided by the dataset D_n is given as

$$p(D_n|M_j) = \int_{\theta_j} p(D_n|\theta_j, M_j)p(\theta_j|M_j)d\theta_j. \qquad (21)$$

This is simply the integral of the likelihood of the data for a vector of parameters weighted by its prior probability integrated over the whole parameter set for θ_j for model class M_j.

An asymptotic approximation for large sample sizes n can be used to compute the evidence of the model (Papadimitriou *et al.*, 1997):

$$p(D_n|M_j) \approx \frac{2\pi^{N_j/2} p(\hat{\theta}_j|M_j)}{(\sqrt{|H_j(\hat{\theta}_j)|})_{\text{Ockham factor}}} \times p(D_n|\hat{\theta}_j, M_j)_{\text{likelihood}}, \qquad (22)$$

where $H_j(\theta_j) = -\nabla\nabla \log[p(D_n|\theta_j, M_j)P(\theta_j|M_j)]$, $\hat{\theta}_j$ is the optimal parameter vector (most probable value) for model class M_j, and N_j is the number of parameters for model class M_j. Here, $H_j(\theta_j)$ is given by equation (15) for the choice of parameters θ_j corresponding to model class M_j. $p(\hat{\theta}_j|M_j)$ is the prior defined in equation (9), and $p(D_n|\hat{\theta}_j, M_j)$ is the likelihood function defined in equation (12), evaluated at the optimal parameter vector for model class M_j. For the model class selection results, the effect of the standard deviation of the Gaussian prior on the choice of most probable model class is examined later.

Results of Bayesian Model Class Selection

We used Bayesian model class selection to find the best combination of the eight ground-motion parameters with the same dataset as the previous classification problem. First, we impose the condition that both horizontal and vertical components be included in the model for any selected ground-motion quantity. Under this condition, there are four groups of ground-motion parameters (peak jerk, acceleration, velocity, and filtered displacement) giving 15 possible combinations. These 15 candidate model classes are shown in Table 7.

The results in Table 7 indicate that the combination of acceleration and velocity is the model with highest probability, although the jerk and velocity combination also has significant probability. The log of prior ($p(\hat{\theta}_j|M_j)$) is simply a function of N_j and becomes smaller as the number of parameters increases. The factor $p(\hat{\theta}_j|M_j)(2\pi^{N_j/2})/\sqrt{|H_j(\hat{\theta}_j)|}$ in equation (22) is called the Ockham factor by Gull (Gull, 1988; Beck and Yuen, 2004). It penalizes a more complicated model and so makes a simpler model preferable. The Ockham factor is also shown in Table 7. Although the coefficient $2\pi^{N_j/2}$ and $\sqrt{|H_j(\hat{\theta}_j)|}$ are included in the Ockham factor, the effect of prior $p(\hat{\theta}_j|M_j)$ is dominant.

The log of the likelihood function $p(D_n|\hat{\theta}_j, M_j)$ becomes larger as the number of the parameters in the model class increases because a more complicated model class will fit the data better than a less complicated one. However, the Bayesian model class selection automatically accounts for the trade-off between the complexity of the model (e.g., the number of parameters) and the fit of the data to find a well-balanced model (Beck and Yuen, 2004). A useful information-theoretic interpretation of this trade-off is given in Muto and Beck (2007).

To examine the possible model classes further, the constraint that horizontal and vertical components be used together is removed. We test all 255 model classes created from the combinations of eight parameters. The results for the best five model classes are shown in Table 8. The sum of the posterior probability of the five model classes is 95% out of all 255 model classes.

Model class 1, which has the coefficients of the vertical acceleration and horizontal velocity, is the most probable model within the set of 255 model classes. The discriminant function for the most probable model in model class 1 is

$$f(X_i|\theta) = 6.046 \log_{10} \mathrm{Za} + 7.885 \log_{10} \mathrm{Hv} - 27.091, \tag{23}$$

where

$$P(Y_i = 1|X_i, \theta) = \frac{1}{1 + e^{-f(X_i|\theta)}} \tag{24}$$

is the probability that station i is near source. This result indicates that the amplitude of high-frequency components is effective in classifying near-source and far-source stations. Note that the probability that the station is near source is higher, if f is larger.

Table 7

Results for Bayesian Model Class Selection when 15 Combinations of the Ground-Motion Parameters Are Examined Under the Condition That the Horizontal and Vertical Components Are Used Together

Model	Hj	Zj	Ha	Za	Hv	Zv	Hd	Zd	*d*	Ockham Factor	Likelihood	Evidence	Probability
j	1.53	4.30	—	—	—	—	—	—	23.84	−17	−140	−156	0.00
a	—	—	4.38	4.37	—	—	—	—	21.43	−16	−117	−133	0.00
v	—	—	—	—	8.57	0.87	—	—	16.33	−16	−118	−134	0.00
d	—	—	—	—	—	—	2.49	1.44	5.76	−17	−192	−209	0.00
ja	−2.74	2.04	6.60	2.95	—	—	—	—	20.82	−25	−114	−139	0.00
jv	2.57	2.79	—	—	7.00	2.00	—	—	36.09	−25	−80	−105	0.32
jd	3.44	3.43	—	—	—	—	3.48	0.79	33.17	−26	−94	−120	0.00
av	—	—	2.54	4.38	7.01	0.91	—	—	29.47	−24	−80	−104	0.62
ad	—	—	4.93	5.02	—	—	3.89	0.22	29.40	−25	−82	−106	0.05
vd	—	—	—	—	12.55	2.30	−3.38	−0.25	19.99	−25	−106	−131	0.00
jav	1.36	1.47	1.36	2.28	6.93	1.50	—	—	33.75	−33	−78	−111	0.00
jad	0.55	0.43	4.35	4.49	—	—	3.89	0.27	30.72	−33	−81	−115	0.00
jvd	2.72	2.68	—	—	6.66	2.91	0.66	−1.12	36.66	−34	−80	−113	0.00
avd	—	—	3.47	4.50	4.58	1.06	1.80	−0.47	30.16	−33	−79	−112	0.00
javd	1.40	1.29	2.05	2.49	5.05	2.11	1.69	−1.02	34.31	−41	−78	−119	0.00

The most probable value of the decision boundary parameter corresponding to each ground-motion parameter is given first for each model class. The values for the Ockham factor, likelihood, and evidence of each model class are log-scaled. The last column is the posterior probability that measures how plausible the model class is. It is scaled such that the total probability of the 15 model classes is 1.0.

Table 8

The Best Five Model Classes in the Bayesian Model Class Selection when 255 Combinations of the Ground-Motion Parameters Are Examined

Model	Hj	Zj	Ha	Za	Hv	Zv	Hd	Zd	*d*	Ockham Factor	Likelihood	Evidence	Probability
1	—	—	—	6.05	7.89	—	—	—	27.09	−15	−81	−96	0.81
2	1.91	—	—	4.41	8.31	—	—	—	31.92	−20	−79	−99	0.07
3	—	—	1.86	4.88	7.86	—	—	—	29.17	−20	−80	−100	0.03
4	—	1.59	—	4.31	8.02	—	—	—	29.71	−20	−80	−100	0.03
5	—	4.43	—	—	8.52	—	—	—	32.22	−16	−84	−100	0.02

The columns are in the same format as in Table 7.

Effect of the Choice of Prior

In this section, we examine the choice of prior for the parameters in the model class selection. As we stated, for the Gaussian prior distribution, the effect of the number of parameters, N_j, is significant if the prior standard deviation, σ, is large. We demonstrate this feature by performing model class selection with a Gaussian prior with different values of σ and a uniform prior with different widths of boundary b. The posterior probabilities of the model classes are shown in Table 9.

In the table, we can see the effect of the prior standard deviation in the Gaussian prior. As we increase σ, it tends to bias the posterior probability towards simpler models (i.e., models with less parameters). For example, the probability of model jav slightly decreases as σ increases. The small probability of model jv with Gaussian prior ($\sigma = 10$) is caused by the narrow prior range. If σ is too small, it restricts the range of parameters as shown in Table 10. Also, for the uniform prior case the results are very similar to the Gaussian prior with $\sigma = 100$. Based on these results, we judge that the choice of $\sigma = 100$ for the Gaussian prior is a reasonable one for Bayesian model class selection in our classification application.

Table 9

The Posterior Probability of the Model Class Selection with Different Types of Prior Distribution for Parameters

Model	Gaussian Prior			Uniform Prior	
	$\sigma = 10$	$\sigma = 100$	$\sigma = 1000$	$\lvert b\rvert < 20$	$\lvert b\rvert < 100$
j	0.0	0.0	0.0	0.0	0.0
a	0.0	0.0	0.0	0.0	0.0
v	0.0	0.0	0.0	0.0	0.0
d	0.0	0.0	0.0	0.0	0.0
ja	0.0	0.0	0.0	0.0	0.0
jv	7.2	32.4	33.0	31.5	32.9
jd	0.0	0.0	0.0	0.0	0.0
av	78.9	62.1	61.7	59.0	61.6
ad	7.3	5.3	5.3	5.0	5.3
vd	0.0	0.0	0.0	0.0	0.0
jav	3.3	0.1	0.0	3.0	0.1
jad	0.1	0.0	0.0	0.0	0.0
jvd	0.1	0.0	0.0	0.3	0.0
avd	3.0	0.0	0.0	1.1	0.0
javd	0.1	0.0	0.0	0.0	0.0

σ is the standard deviation for the Gaussian distribution, and $|b|$ is the width of the boundary for the uniform distribution.

Results and Discussion

We apply the optimal discriminant function from the Bayesian approach (in equations 23 and 24) to all the stations in the dataset. Figure 12 shows the classification results. The distribution of stations with a high probability of being in the near source is consistent with the fault geometry. As mentioned before, the fault models that are used here are those from the source inversion, and they are not necessarily the best indicator of near-source and far-source stations.

To examine the application for real-time analysis, the optimal discriminant function in equations (23) and (24) is applied to the Chi-Chi earthquake strong-motion records. We generated snapshots of the probability that a station is near source from 10 to 40 sec after the beginning of rupture. Peak ground motions used for this classification analysis are computed from the observed data every 10 sec for each station and evaluated in the discriminant function. The results are shown in Figure 13. A darker mark at a station in Figure 13 indicates that the station is more likely to be near source, and a lighter mark indicates that the station is more likely to be far source.

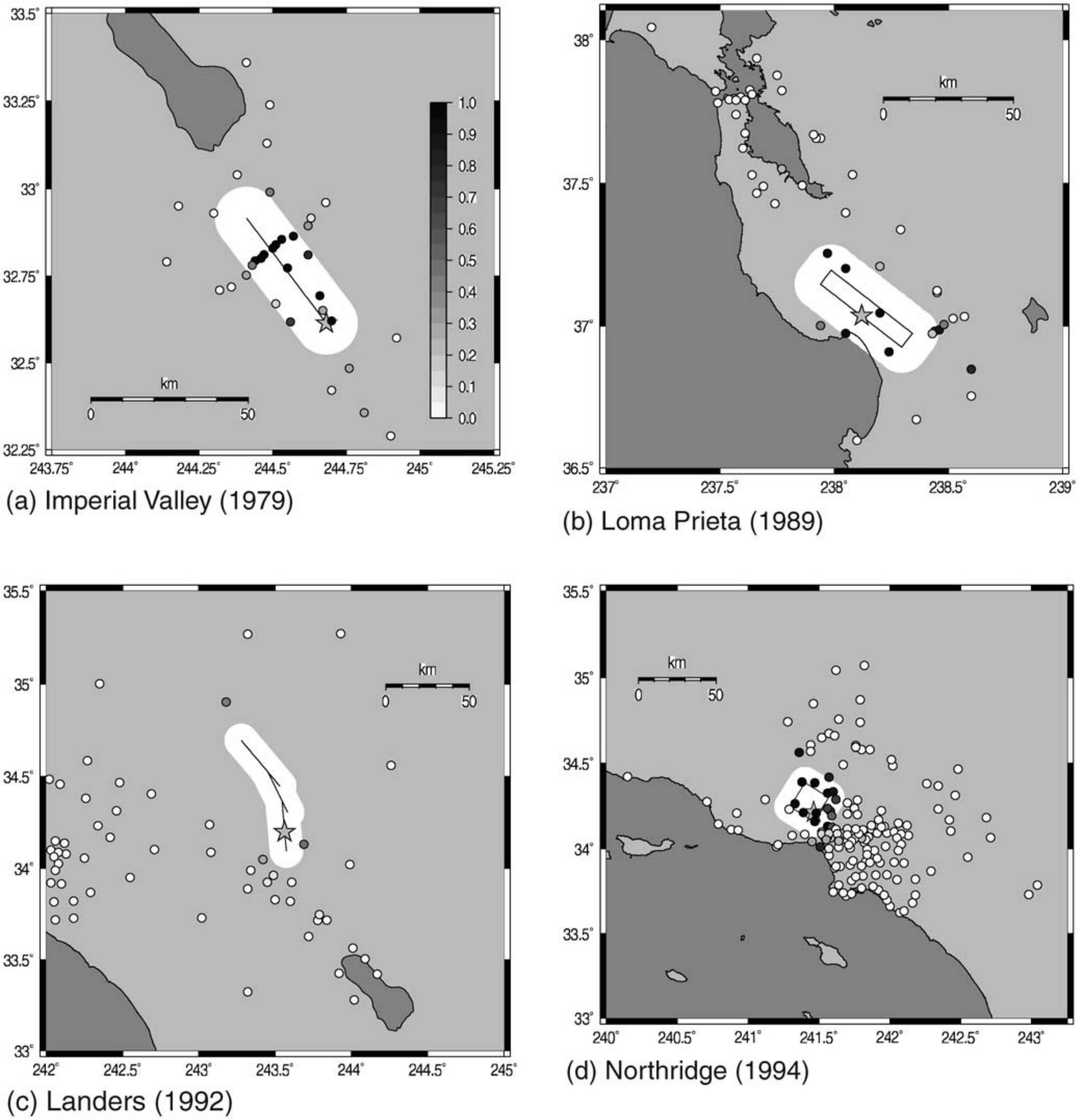
Ten seconds after the rupture initiation, the map shows that stations with a high probability of being in the near source are located near the epicenter, and it indicates that the rupture area is propagating concentrically. At 20 sec, the probability of being in the near source at 13 stations is computed to be greater than 50%, but the concentric station distribution makes it difficult to identify any directivity of rupture propagation. The average slip velocity is 2 km/sec

Table 10

The Estimated Parameters from Bayesian Approach with Different Types of Prior Distribution for Parameters

Prior	c_1 (Za)	c_2 (Hv)	*d*
Gaussian ($\sigma = 10$)	5.522	7.147	24.686
Gaussian ($\sigma = 100$)	6.046	7.885	27.091
Gaussian ($\sigma = 1000$)	6.053	7.895	27.122
Uniform cases	6.053	7.895	27.122

(Ji *et al.*, 2003), and the rupture front propagates 40 km from the hypocenter at this point. We can see the north–south character of the rupture direction clearly after 30 sec of rupture. At 40 sec, the distribution of stations with high near-source probability agrees with the fault surface projection, and stations at the near source and far source boundary have around 50% probability. Even though the fault geometries used for the wave inversion are not necessarily the actual extent of the fault, to a first-order approximation, the classification results are in good agreement with them. The near-source region at the north of the main rupture is a secondary rupture at the Shihtan fault, which is suggested by Shin and Teng (2001). This event may not be clear in the low-frequency ground motions, so it is not considered in the waveform in-



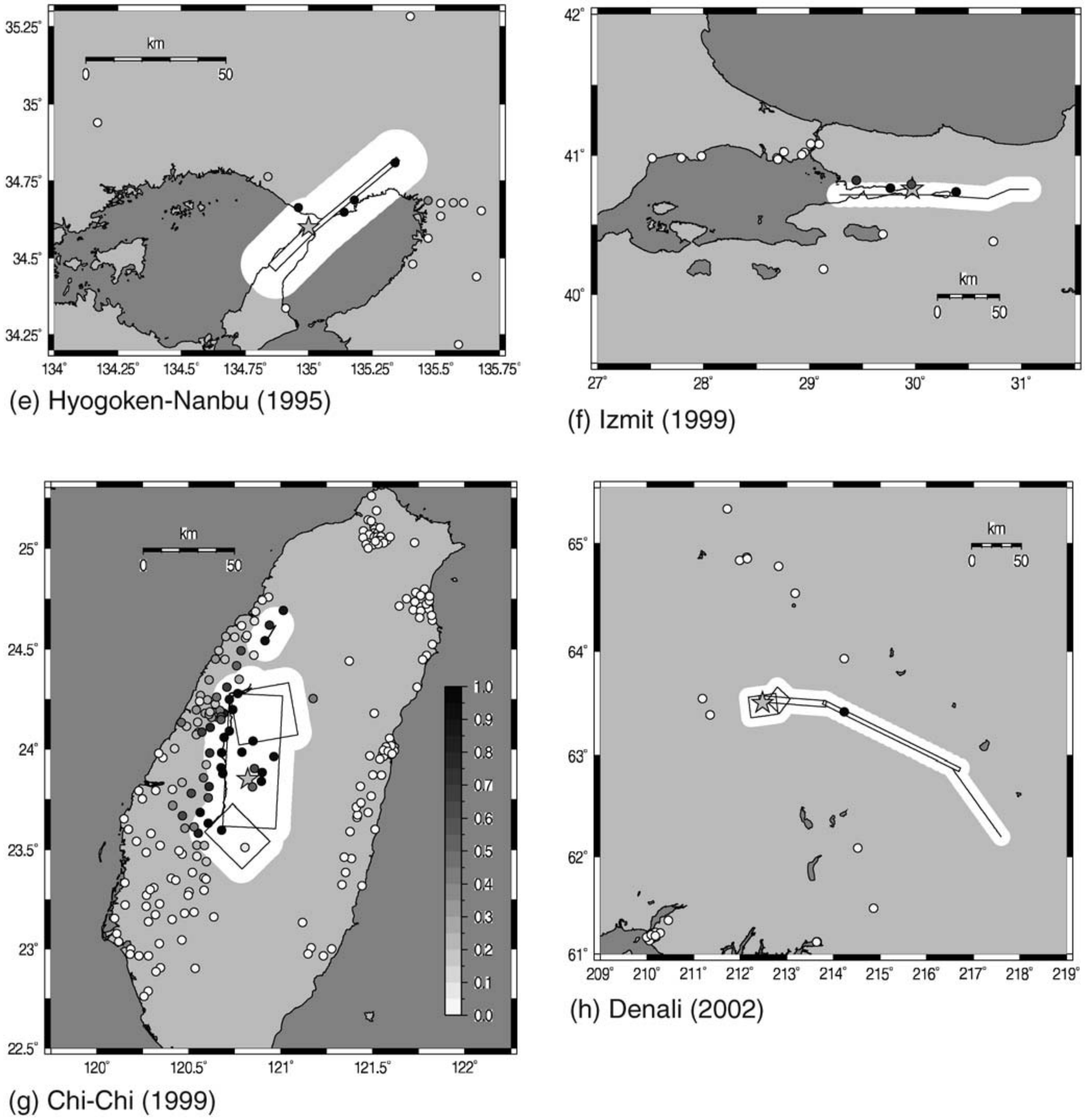
(a) Imperial Valley (1979)

(b) Loma Prieta (1989)

(c) Landers (1992)

(d) Northridge (1994)

Figure 12. Probabilities of near source based on the optimal discriminant function obtained by the Bayesian approach. Darker marks have higher probability that the station is located at near source. All stations in the figures use the same color code for scale. The symbols for the fault and epicenter are the same as in Figure 2. (*Continued*)



(e) Hyogoken-Nanbu (1995)

(f) Izmit (1999)

(g) Chi-Chi (1999)

(h) Denali (2002)

Figure 12. Continued.

version. However, the accelerograms at that region are clearly larger than those of neighboring region, and the classification results detected this secondary rupture.

Conclusions

We presented a methodology to classify seismic records into near-source or far-source records as a prelude to estimating fault dimensions in an earthquake early warning system. Ground-motion records from some past earthquakes are analyzed to find a linear function that best discriminates near-source and far-source records. Peak values of jerk, acceleration, velocity, and displacement are used in a traditional LDA and in a Bayesian approach to find the linear combination of peak values that provides the best performance to classify near-source and far-source records. All methods gave similar discriminant functions. We also analyzed which combination of ground-motion features had the best performance for clas-

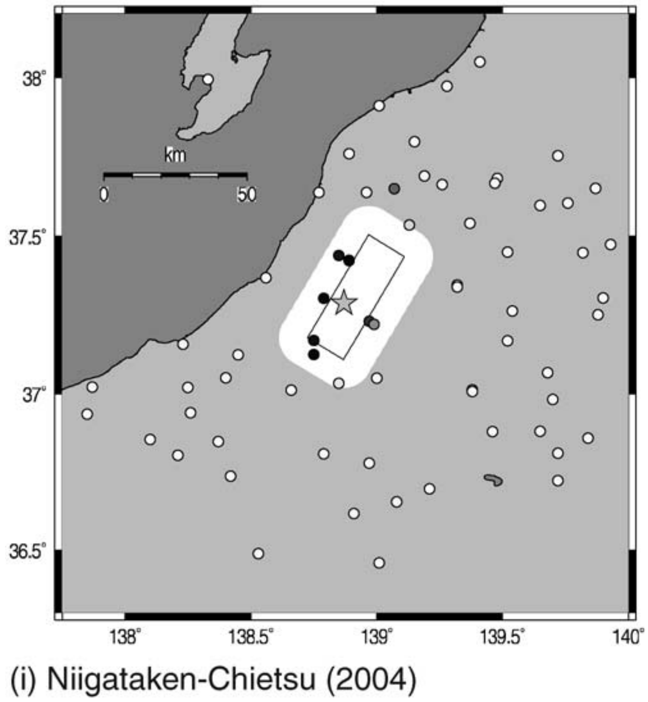


Figure 12. Continued.

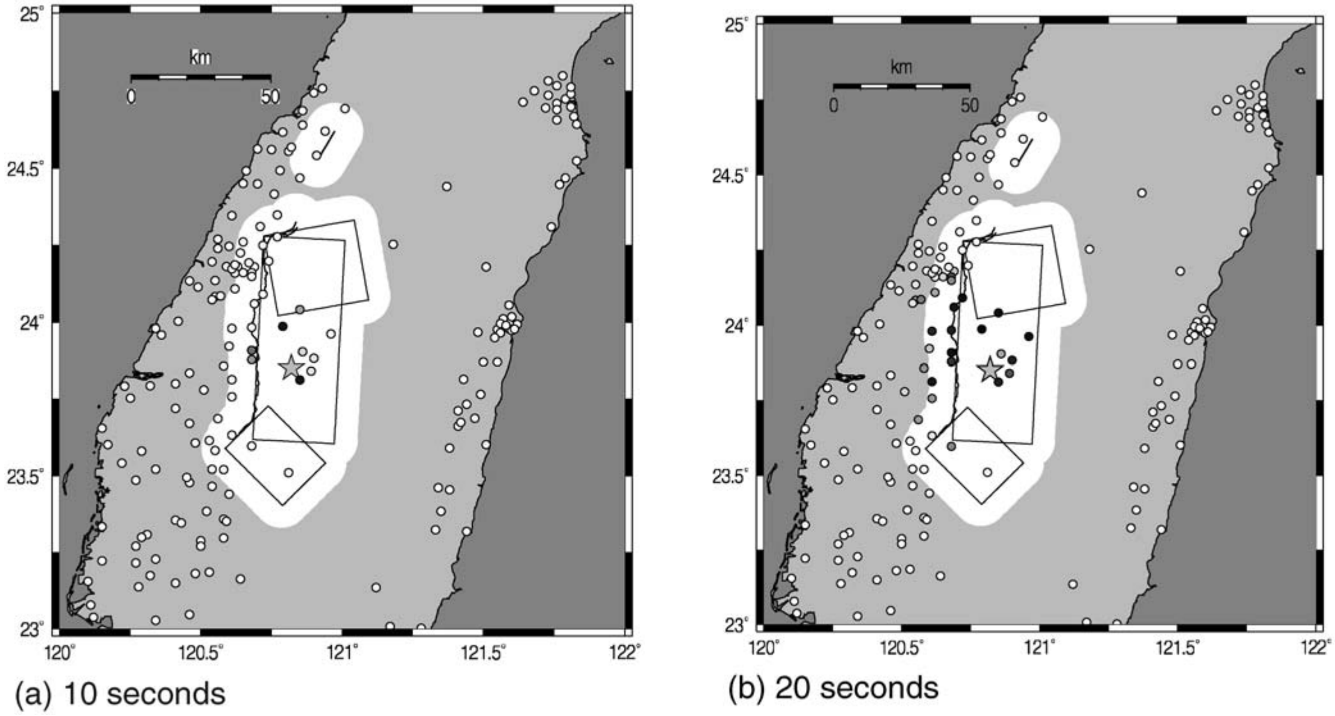


Figure 13. Snapshots of the probabilities of near source for the Chi-Chi earthquake, based on the optimal discriminant function from the Bayesian approach. The large circle is the theoretical rupture front assuming the rupture velocity 2 km/sec. (*Continued*)

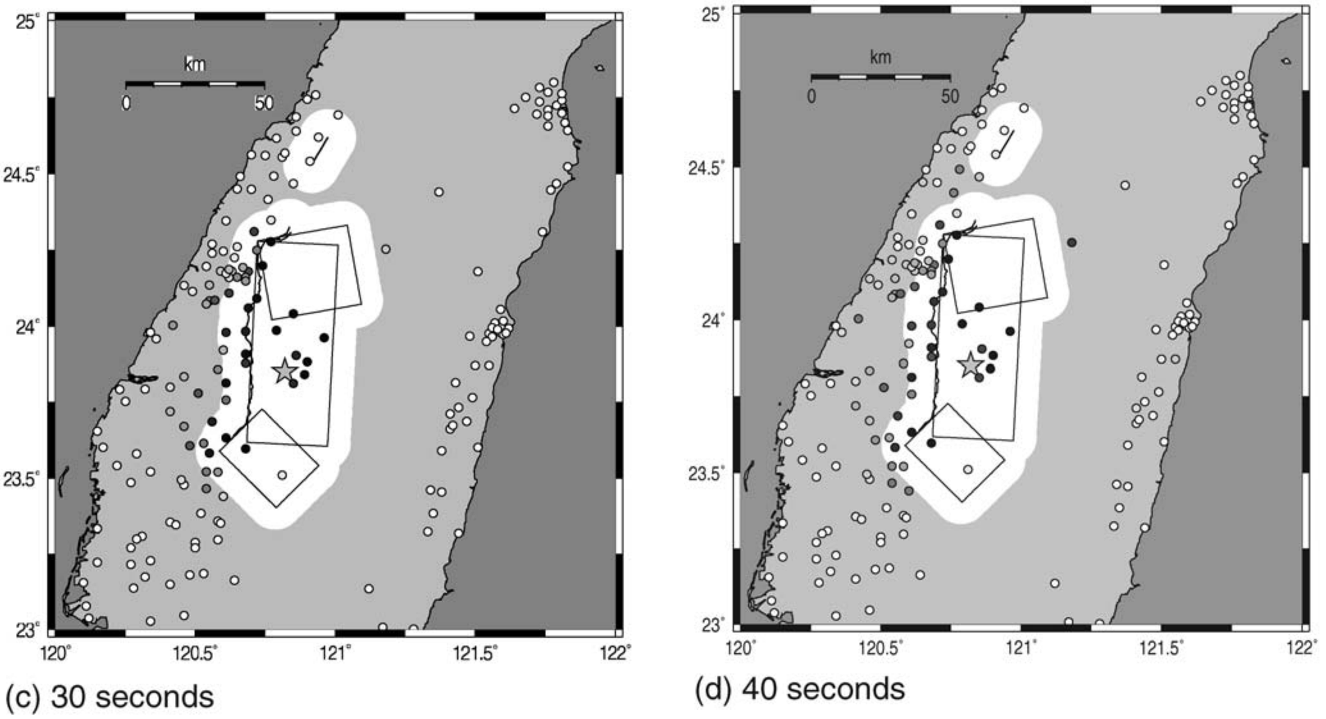


Figure 13. Continued.

sification using Bayesian model class selection, and the best discriminant function is

$$f(X_i|\theta) = 6.046 \log_{10} \mathrm{Za} + 7.885 \log_{10} \mathrm{Hv} - 27.091, \tag{25}$$

$$P(Y_i = 1|X_i, \theta) \frac{1}{1 + e^{-f(X_i|\theta)}}, \tag{26}$$

where Za and Hv denote the peak values of the vertical acceleration and horizontal velocity, respectively, and $P(Y_i = 1|X_i, \theta)$ is the probability that a station is near source. This function indicates that the amplitude of high-frequency components is effective in classifying near-source and far-source stations.

The probability that a station is near source obtained using this optimal discriminant function for all the earthquakes shows the extent of the near source area quite well, suggesting that the approach provides a good indicator of near-source and far-source stations for real-time analyses. Note that this function is constructed by the training dataset with a magnitude greater than 6.5, so it works only for large earthquakes.

Acknowledgments

We thank David Wald of the U.S. Geological Survey and Hiroo Kanamori of the California Institute of Technology for their help with obtaining data for the Hyogoken-nanbu earthquake. We also appreciate Chen Ji for providing the fault model for the Chi-Chi earthquake. Members of the Caltech Earthquake Engineering Group, especially Sai Hung Cheung, gave us useful comments for the Bayesian classification problem. Some of the figures are generated by Generic Mapping Tools (Wessel and Smith, 1991). This research was supported by the U.S. Geological Survey.

References

Akkar, S., and P. Gülkan (2002). A critical examination of near-field accelerograms from the sea of Marmara region earthquakes, *Bull. Seismol. Soc. Am.* **92,** 428–447.

Allen, R. M., and H. Kanamori (2003). The potential for earthquake early warning in Southern California, *Science* **300,** 786–789.

Beck, J. L., and L. S. Katafygiotis (1998). Updating models and their uncertainties. I: Bayesian statistical framework, *J. Eng. Mech.* **124,** no. 4, 455–461.

Beck, J. L., and K. Yuen (2004). Model selection using response measurements: Bayesian probabilistic approach, *J. Eng. Mech* **130,** 192–203.

Boore, D. M. (2001). Effect of baseline correction on displacement and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, earthquake, *Bull. Seismol. Soc. Am.* **91,** 1199–1211.

Campbell, K. W. (1981). Near-source attenuation of peak horizontal acceleration, *Bull. Seismol. Soc. Am.* **71,** 2039–2070.

Cua, G. (2005). Creating the virtual seismologist: developments in earthquake early warning and ground motion characterization, *Ph.D. Thesis*, Department of Civil Engineering, California Institute of Technology.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7,** 179–188.

Gull, S. (1988) Bayesian inductive reference and maximum entropy, in *Maximum Entropy and Bayesian Methods in Science and Engineering*, C. J. Erickson and C. R. Smith (Editors), Kluwer Academic, Dordrecht, 53–74.

Hanks, T. C., and D. A. Johnson (1976). Geophysical assessment of peak accelerations, *Bull. Seismol. Soc. Am.* **66,** 959–968.

Hanks, T. C., and R. K. McGuire (1981). The character of high-frequency strong ground motion, *Bull. Seismol. Soc. Am.* **71,** no. 6, 2071–2095.

Hartzell, S., and T. Heaton (1983). Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am.* **73,** 1553–1583.

Honda, R., S. Aoi, N. Morikawa, H. Sekiguchi, T. Kunugi, and H. Fujiwara (2005). Ground motion and rupture process of the 2004 mid Niigata prefecture earthquake obtained from strong motion data of K-NET and KiK-net, *Earth Planet Space* **57,** 527–532.

Iwan, W. D., M. A. Moser, and C.-Y. Peng (1985). Some observations on strong-motion earthquake measurement using a digital accelerograph, *Bull. Seismol. Soc. Am.* **75,** 1225–1246.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*, Cambridge U. Press, New York.

Ji, C., D. V. Helmberger, D. J. Wald, and K. F. Ma (2003). Slip history and dynamic implication of 1999 Chi-Chi earthquake, *J. Geophys. Res.* **108,** no. B9, 2412, doi 10.1029/2002JB001764.

Joyner, W. B., and D. M. Boore (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am.* **71,** 2011–2038.

Lee, W. H. K., T. C. Shin, K. W. Kuo, K. C. Chen, and C. F. Wu (2001). CWB Free-Field Strong-Motion Data from the 21 September Chi-Chi, Taiwan, Earthquake, *Bull. Seismol. Soc. Am.* **91,** no. 5, 1370–1376.

Li, Y., C. Campbell, and M. Tipping (2002). Bayesian automatic relevance determination algorithms for classifying gene expression data, *Bioinformatics* **18,** 1332–1339.

MacKay, D. J. C. (1998). Introduction to Monte Carlo methods, in *Learning in Graphical Models*, MIT Press, Cambridge, Massachusetts, 175–204.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller (1953). Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21,** no. 6, 1087–1092.

Muto, M., and J. L. Beck (2007). Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, *J. Vib. Control* (in press).

Nakamura, Y. (1988). On the urgent earthquake detection and alarm system (UrEDAS), *Proc. of the 9th World Conference on Earthquake Engineering*, Tokyo-Kyoto, Japan Vol. VII, 673–678.

Odaka, T., K. Ashiya, S. Tsukada, S. Sato, K. Otake, and D. Nozaka (2003). A new method of quickly estimating epicentral distance and magnitude from a single seismic record, *Bull. Seismol. Soc. Am.* **93,** no. 1, 526–532.

Papadimitriou, C., J. L. Beck, and L. S. Katafygiotis (1997). Asymptotic expansions for reliabilities and moments of uncertain dynamic systems, *J. Eng. Mech.* **123,** no. 12, 1219–1229.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Probab.* **7,** 110–120.

Sekiguchi, H., and T. Iwata (2002). Rupture process of the 1999 Kocaeli, Turkey, earthquake estimated from strong-motion waveforms, *Bull. Seismol. Soc. Am.* **92,** 300–311.

Sekiguchi, H., K. Irikura, T. Iwata, Y. Kakehi, and M. Hoshiba (1996). Minute locating of faulting beneath Kobe and the waveform inversion of the source process during the 1995 Hyogo-ken Nanbu, Japan, earthquake using strong ground motion records, *J. Phys. Earth* **44,** 473–488.

Shin, T.-C., and T.-L. Teng (2001). An overview of the 1999 Chi-Chi, Taiwan, earthquake, *Bull. Seismol. Soc. Am.* **91,** 895–913.

Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*, Oxford U. Press, Oxford.

Toki, K., K. Irikura, and T. Kagawa (1995). Strong motion data recorded in the source area of the Hyogoken-nanbu earthquake, January 17, 1995, Japan, *J. Nat. Disaster Sci.* **16,** 23–30.

Tsuboi, S., D. Komatitsch, C. Ji, and J. Tromp (2003). Broadband modeling of the 2002 Denali fault earthquake on the Earth Simulator, *Phys. Earth Planet. Interiors* **139,** 305–312.

Venables, W. N., and B. D. Ripley (2002). *Modern Applied Statistics with S*, Fourth Ed., Springer, New York.

Wald, D. J. (1996). Slip history of the 1995 Kobe, Japan, earthquake determined from strong motion, teleseismic, and geodetic data, *J. Phys. Earth* **44,** 489–503.

Wald, D. J., and T. H. Heaton (1994). Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake, *Bull. Seismol. Soc. Am.* **84,** 668–691.

Wald, D. J., T. H. Heaton, and D. V. Helmberger (1991). Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong motion and broadband teleseismic data, *Bull. Seismol. Soc. Am* **81,** 1540–1572.

Wald, D. J., T. H. Heaton, and K. W. Hudnut (1996). A dislocation model of the 1994 Northridge, California, earthquake determined from strong-motion, GPS, and leveling-line data, *Bull. Seismol. Soc. Am.* **86,** S49–70.

Wessel, P., and W. H. F. Smith (1991). Free software helps map and display data, *EOS. Trans. AGU* **72,** 445–446.

Wu, Y. M., and H. Kanamori (2005). Experiment on an onsite early warning method for the Taiwan early warning system, *Bull. Seismol. Soc. Am.* **95,** 347–353.

Kyoto University
Gokasyo, Uji, 611-0011 Japan
masumi@eqh.dpri.kyoto-u.ac.jp
(M.Y.)

California Institute of Technology
MC104-44, 1200 E. California Boulevard
Pasadena, California 91125
heaton_t@caltech.edu
jimbeck@caltech.edu
(T.H., J.B.)

Manuscript received 21 November 2006