

Early Warning Systems for Large Earthquakes: Classification of Near-source and Far-source Stations by using the Bayesian Model Class Selection

M. Yamada, T.H. Heaton & J.L. Beck

California Institute of Technology, Pasadena, CA, USA

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. This study analyzes peak ground motions and finds the function which best classifies near-source and far-source records based on these parameters. We perform: (1) 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 classifies near-source and far-source data and it gives 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.

1 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 main shock (Nakamura, 1988; Allen & Kanamori, 2003). The VS method assumes a point source model for the rupture, and it works well for small to moderate earthquakes (magnitude < 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 ground motion at a site may be affected by the fault rupture direction or the fault rupture length.

The objective of this paper is to develop a methodology to classify stations into near-source and far-source since this can be used for identifying the fault geometry if there is a sufficiently dense seismic network. 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:

- 1) Collect strong motion data from earthquake strong motion archives and classify these samples into two predefined groups: records from near-source stations and far-source stations. This particular set of data is called the training set.

- 2) Discover a discriminant function of the samples

features (e.g. peak ground acceleration (PGA), velocity (PGV), displacement (PGD)) which provides the best performance in terms of near-source / far-source classification.

- 3) 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 paper; and we investigate linear discriminant functions by using the Bayesian method. The third step can then be accomplished in a real-time analysis. Given a new ground motion observation from on-going rupture, the discriminant function gives the probability that the observation is located in the near-source.

2 STRONG MOTION DATA

2.1 Data sources

We used strong motion datasets from nine earthquakes with magnitude 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 distance (the shortest distance between the station and the surface projection of the fault plane) is less than 10km. 695 three-component strong motion data are used for the classification analysis and 14% (100 sta-

Table 1: The earthquake dataset used for the classification analysis. Moment magnitude (M_w) is cited from Harvard CMT solution. The numbers of near-source (NS) and far-source (FS) data for each earthquake are also shown. The fault models are used as selection criteria to classify near-source and far-source stations.

Earthquake	M_w	NS	FS	Total	Fault Model	Reference
Imperial Valley (1979)	6.5	14	20	34	Hartzell & Heaton (1983)	BSSA, 1553-1583
Loma Prieta (1989)	6.9	8	39	47	Wald et al. (1991)	BSSA 1540-1572.
Landers (1992)	7.3	1	112	113	Wald & Heaton (1994)	BSSA, 668-691.
Northridge (1994)	6.6	17	138	155	Wald et al. (1996)	BSSA, S49-70.
Hyogoken-Nanbu (1995)	6.9	4	14	18	Wald (1996)	JPE 489-503.
Izmit (1999)	7.6	4	13	17	Sekiguchi & Iwata (2002)	BSSA, 300-311.
Chi-Chi (1999)	7.6	42	172	214	Ji et al. (2003)	JGR, 2412.
Denali (2002)	7.8	1	29	30	Tsuboi et al. (2003)	PEPI 305-312.
Niigataken-Chuetsu (2004)	6.6	9	58	67	Honda et al. (2004)	EPS (submitted).
Total		100	595	695		

tions) are near-source stations.

The classification as near-source or far-source in the training set is based on fault models used for waveform inversions. These fault models are typically determined from the aftershock distribution (Sekiguchi *et al.*, 1996), and the shape of the fault area is a rectangular box. Fault models used for classifying stations are also shown in Table 1. Figure 1 shows the surface projection of the fault rupture surface based on the fault models. Stations within 10 km of this fault projection (the white area in the figures) are classified as near-source, indicated by solid circles. Far-source stations are shown in open circles.

Table 2: Eight measurements of peak ground motions are calculated from three component accelerograms. Codes and units of the components used in this paper are shown.

Code	Measurement	Unit
Hj	Horizontal Jerk	(cm/s^3)
Zj	Vertical Jerk	(cm/s^3)
Ha	Horizontal Acceleration	(cm/s^2)
Za	Vertical Acceleration	(cm/s^2)
Hv	Horizontal Velocity	(cm/s)
Zv	Vertical Velocity	(cm/s)
Hd	Horizontal Displacement	(cm)
Zd	Vertical Displacement	(cm)

2.2 Data processing

We processed the accelerograms obtained from the nine earthquakes according to the following method. The DC offset of the accelerograms is corrected by subtracting the mean of the pre-event portion.

The peak amplitude of the horizontal components is calculated by the square root of the sum of the squares of the peaks of NS and EW components. The peak amplitude of UD (up-down) component is used directly for the peak vertical component.

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 to either tilting, the response of the transducer to strong shaking, or 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 $v_f(t) = a_1t + a_2$ where coefficients a_1 and a_2 are determined by least-squares fitting to the velocity trace after the strong shaking;

2) Remove this linear trend from the velocity record.

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

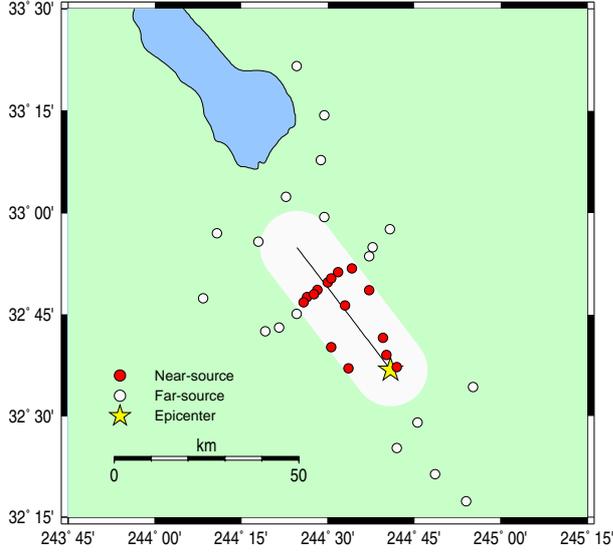
The peak features used for the classification analysis are shown in Table 2. Several combinations of these 8 features are tried to find the best performance of the classification.

3 BAYESIAN METHOD

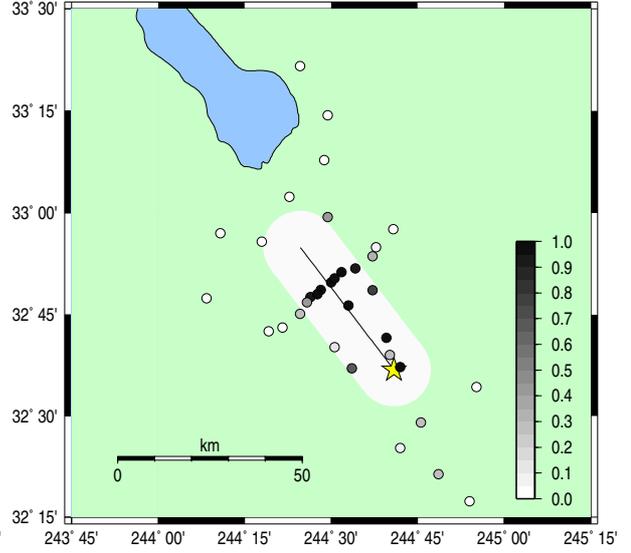
3.1 Near-source versus far-source discriminant function

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:

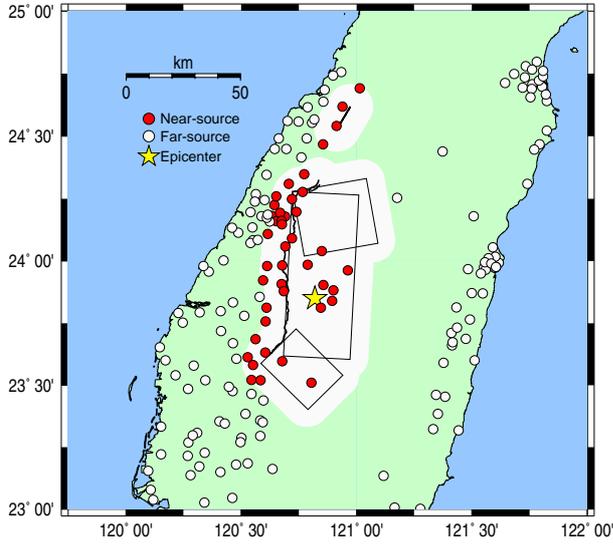
$$f(X_i|\theta) = c_1x_{i1} + c_2x_{i2} + \dots + c_mx_{im} - d \quad (1)$$



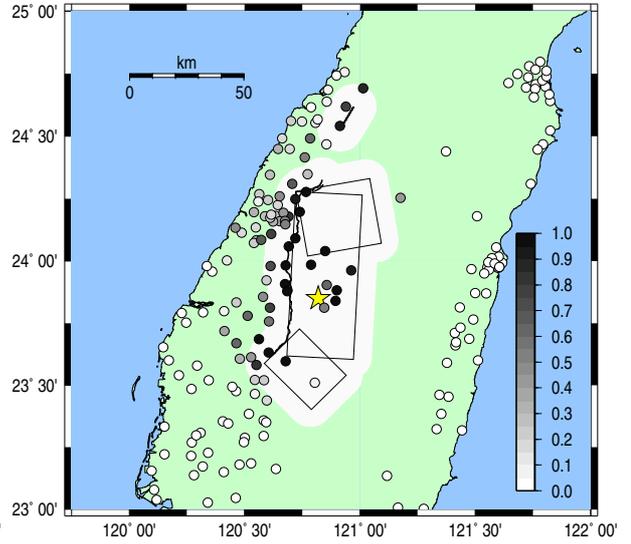
(a) Observations for the Imperial Valley (1979)



(b) Predictions for the Imperial Valley (1979)



(c) Observations for the Chi-Chi (1999)



(d) Predictions for the Chi-Chi (1999)

Figure 1: Observations and predictions of near-source and far-source stations. The fault projections are shown in the solid lines. The white area around the fault lines indicates the area with distance less than 10 km from the fault projections. The star symbol denotes the epicenter of the earthquake. Left: observations of near-source and far-source stations based on the fault models. Near-source and far-source stations are shown in solid and open circles, respectively. Right: 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.

where

x_{ik} = k th feature parameter at the station i

$X_i = [x_{i1} \ x_{i2} \ \dots \ x_{im}]$

c_1, \dots, c_m = the regression coefficients

d = decision boundary constant

$\theta = [c_1 \ c_2 \ \dots \ c_m \ d]^T$

We may use m components out of the eight ground motion components shown in Table 2.

We apply the logistic sigmoid function $\phi(x) = 1/(1 + e^{-x})$ to the linear function $f(X_i|\theta)$ to define the predictive probability that a station is near-source (Li *et al.*, 2002). The logistic sigmoid function is a smooth, positive, and monotonically increasing function, as shown in Figure 2. The predictive probabilities that a station is near-source or far-source are therefore defined here by:

$$P(Y_i|X_i, \theta) = \phi(Y_i f(X_i|\theta)) = \frac{1}{1 + e^{-Y_i f(X_i|\theta)}} \quad (2)$$

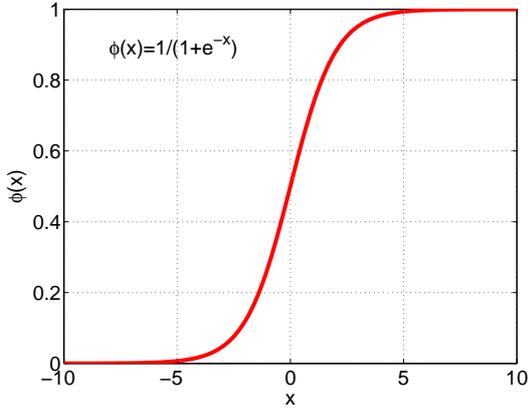


Figure 2: Logistic sigmoid function $\phi(x) = 1/(1 + e^{-x})$ is used to express the predictive probability for classification. The function approaches zero as $x \rightarrow -\infty$, and one as $x \rightarrow \infty$. The function is 0.5 when x is zero.

where

$$Y_i = \begin{cases} 1 & \text{if near-source} \\ -1 & \text{if far-source} \end{cases} \text{ at the station } i$$

3.2 Asymptotic approximation

The coefficients c_1, \dots, c_m , and d in the discriminant function are determined from the training data set by Bayesian approach with an asymptotic approximation. The probability density function (pdf) of parameter θ conditioned on data D_n and model class M can be expressed using Bayes' theorem:

$$p(\theta|D_n, M) \underset{\text{posterior}}{\propto} \underset{\text{likelihood}}{p(D_n|\theta, M)} \times \underset{\text{prior}}{p(\theta|M)} \\ \propto \prod_{i=1}^n P(Y_i|X_i, \theta) \times p(\theta|M) \quad (3)$$

where

$D_n = \{(X_i, Y_i) : i = 1, \dots, n\}$: available set of data

$X_i = [x_{i1} \ x_{i2} \ \dots \ x_{im}]$: ground motion at the station i

$$Y_i = \begin{cases} 1 & \text{if near-source} \\ -1 & \text{if far-source} \end{cases} \text{ at the station } i$$

We select a Gaussian prior with zero mean and standard deviation $\sigma=100$ to cover a wide range of the parameter space. The likelihood function is expressed by substituting equation (2) into (3):

$$p(D_n|\theta, M) = \prod_{i=1}^n \frac{1}{1 + e^{-Y_i f(X_i|\theta)}} \quad (4)$$

From equations (3) and (4), the posterior pdf is:

$$p(\theta|D_n, M) \propto \frac{1}{(\sqrt{2\pi}\sigma)^{m+1}} \exp\left(-\frac{1}{2\sigma^2}\theta^T\theta\right) \underset{\text{prior}}{\times} \prod_{i=1}^n \frac{1}{1 + e^{-Y_i f(X_i|\theta)}} \underset{\text{likelihood}}{\quad} \quad (5)$$

An asymptotic approximation is performed to characterize the posterior pdf defined by equation (5). 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 $\hat{\Sigma}$.

We first evaluate 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 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:

$$H(\theta) = -\nabla\nabla \log[p(D_n|\theta, M)p(\theta|M)] \quad (6)$$

then $\hat{\Sigma} = H(\hat{\theta})^{-1}$.

3.3 Optimal solution and sensitivity analysis

The optimal parameter values and their standard deviations for the selection of features Z_a and H_v are shown in Table 3. This combination of parameters are selected by Bayesian model class selection (the results are shown later). Note that for large standard deviations in the prior pdf, the effect of the prior on the posterior is negligible (Sivia, 1996).

Table 3: The optimal model parameters and standard deviations for parameters estimated by Bayesian method.

Coefficient	Optimal value	Std. deviation
c_1 (Z_a)	6.046	0.903
c_2 (H_v)	7.886	1.206
d	27.090	3.163

Figure 3 compares the Gaussian distribution with mean $\hat{\theta}$ and covariance matrix $\hat{\Sigma}$, and the distribution of samples for 3 parameters generated by Metropolis algorithm (Yamada *et al.*, 2006). The Gaussian approximation agrees well with the posterior pdf characterized by the stochastic simulation.

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

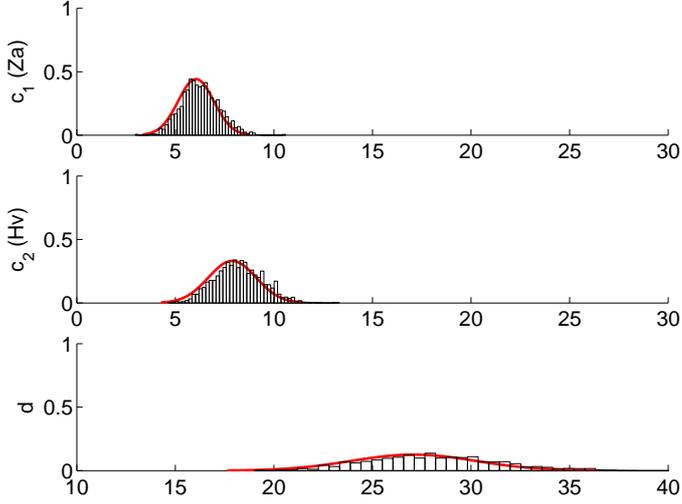


Figure 3: The Gaussian distributions for 3 parameters obtained from the asymptotic approximation. Distribution of samples generated by Metropolis algorithm (Yamada *et al.*, 2006) are added in the figure, and fit the Gaussian distribution well.

Table 4: The confusion matrix for the cross-validation analysis. ‘‘All dataset’’ is the analysis which uses the whole dataset as a training set and a validation set. ‘‘Half of dataset’’ is the analysis which uses half of dataset as a training set and the other half as a validation set. ‘‘Other half of dataset’’ is the analysis which switches the training and validation set. NS and FS stand for near-source and far-source, respectively.

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

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 corresponding to the probability that the station is near-source to be half, 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 analysis and the previous analysis which 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 which uses all of the dataset. Therefore, we confirm that the sensitivity to the training dataset is very small, giving more confidence that the discriminant function from Bayesian analysis will perform well for future earthquake data.

4 BAYESIAN MODEL CLASS SELECTION

4.1 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 M of candidate model classes M_j , $j = 1, \dots, J$ (Beck and Yuen, 2004). Applying Bayes’ theorem, the probability of model class M_j can be expressed as follows:

$$P(M_j|D_n, M) = \frac{\overset{\text{evidence}}{p(D_n|M_j)} \overset{\text{prior}}{P(M_j|M)}}{\underset{\text{normalizing constant}}{p(D_n|M)}} \quad (7)$$

We assume a uniform prior for the model class, i.e. each model class is equally plausible a priori. An asymptotic approximation for large sample sizes n can be used to compute the evidence for the model class (Papadimitriou *et al.*, 1997):

$$\begin{aligned} p(D_n|M_j) &\approx \frac{2\pi^{N_j/2} p(\hat{\theta}_j|M_j)}{\sqrt{|H_j(\hat{\theta}_j)|}} \times \underset{\text{likelihood}}{p(D_n|\hat{\theta}_j, M_j)} \quad (8) \\ &\quad \underset{\text{Ockham factor}}{} \\ &= \frac{\frac{1}{\sigma^{N_j}} \exp(-\frac{\|\hat{\theta}_j\|^2}{2\sigma^2})}{\sqrt{|H_j(\hat{\theta}_j)|}} \times p(D_n|\hat{\theta}_j, M_j) \end{aligned}$$

Here, $H_j(\theta_j)$ is given by equation (6) for the choice of parameters θ_j corresponding to model class M_j . $p(\hat{\theta}_j|M_j)$ is the Gaussian prior with mean 0 and standard deviation 100 and $p(D_n|\hat{\theta}_j, M_j)$ is the likelihood function defined in equation (4), evaluated at the optimal parameter vector for model class M_j . For the model class selection, the effect of the Gaussian prior is significant if the standard deviation, σ , is large. However, the most probable model class based on the data is robust to σ over a wide range of values (Yamada *et al.*, 2006).

4.2 Results of 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 fifteen possible combinations. We found out that the combination of acceleration and velocity is the preferred model with highest probability (see Table 5).

Table 5: Results for Bayesian model class selection when fifteen combinations of the ground motion parameters are examined under the condition that the horizontal and vertical components are used together. 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 (Ock.), likelihood (likeli.), and evidence (evi.) 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 fifteen model classes is 100 %.

Model	Hj	Zj	Ha	Za	Hv	Zv	Hd	Zd	d	Ock.	Likeli.	Evi.	Pr.(%)
j	1.53	4.30	-	-	-	-	-	-	-23.84	-17	-140	-156	0.0
a	-	-	4.38	4.37	-	-	-	-	-21.43	-16	-117	-133	0.0
v	-	-	-	-	8.57	0.87	-	-	-16.33	-16	-118	-134	0.0
d	-	-	-	-	-	-	2.49	1.44	-5.76	-17	-192	-209	0.0
ja	-2.74	2.04	6.60	2.95	-	-	-	-	-20.82	-25	-114	-139	0.0
jv	2.57	2.79	-	-	7.00	2.00	-	-	-36.09	-25	-80	-105	32.4
jd	3.44	3.43	-	-	-	-	3.48	0.79	-33.17	-26	-94	-120	0.0
av	-	-	2.54	4.38	7.01	0.91	-	-	-29.47	-24	-80	-104	62.1
ad	-	-	4.93	5.02	-	-	3.89	0.22	-29.40	-25	-82	-106	5.3
vd	-	-	-	-	12.55	2.30	-3.38	-0.25	-19.99	-25	-106	-131	0.0
jav	1.36	1.47	1.36	2.28	6.93	1.50	-	-	-33.75	-33	-78	-111	0.1
jad	0.55	0.43	4.35	4.49	-	-	3.89	0.27	-30.72	-33	-81	-115	0.0
jvd	2.72	2.68	-	-	6.66	2.91	0.66	-1.12	-36.66	-34	-80	-113	0.0
avd	-	-	3.47	4.50	4.58	1.06	1.80	-0.47	-30.16	-33	-79	-112	0.0
javd	1.40	1.29	2.05	2.49	5.05	2.11	1.69	-1.02	-34.31	-41	-78	-119	0.0

Table 6: The best five model classes in the Bayesian model class selection when 255 combinations of the ground motion parameters are examined. The columns are in the same format as in Table 5.

Model	Hj	Zj	Ha	Za	Hv	Zv	Hd	Zd	d	Ock.	Likeli.	Evi.	Pr.(%)
1	-	-	-	6.05	7.89	-	-	-	-27.09	-15	-81	-96	80.8
2	1.91	-	-	4.41	8.31	-	-	-	-31.92	-20	-79	-99	6.6
3	-	-	1.86	4.88	7.86	-	-	-	-29.17	-20	-80	-100	2.9
4	-	1.59	-	4.31	8.02	-	-	-	-29.71	-20	-80	-100	2.5
5	-	4.43	-	-	8.52	-	-	-	-32.22	-16	-84	-100	1.9

The factor $p(\hat{\theta}_j|M_j)(2\pi^{N_j/2})/\sqrt{|H_j(\hat{\theta}_j)|}$ in equation (8) 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 5.

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 (here it can be interpreted as the number of parameters) and the fit of the data to find a well-balanced model (Beck and Yuen, 2004).

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 8 parameters. The results for the best five model classes are shown in Table 6. 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:

$$f(X_i|\theta) = 6.046 \log_{10} Za + 7.885 \log_{10} Hv - 27.091 \quad (9)$$

where

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

is the probability that station i is near-source. Note that the probability that the station is near-source is higher, if f is larger.

This result indicates that the amplitude of high-frequency components is effective in classifying near-source and far-source stations. It is obvious that more complicated model classes have a better goodness-to-fit, as the same reason as a higher-degree polynomial

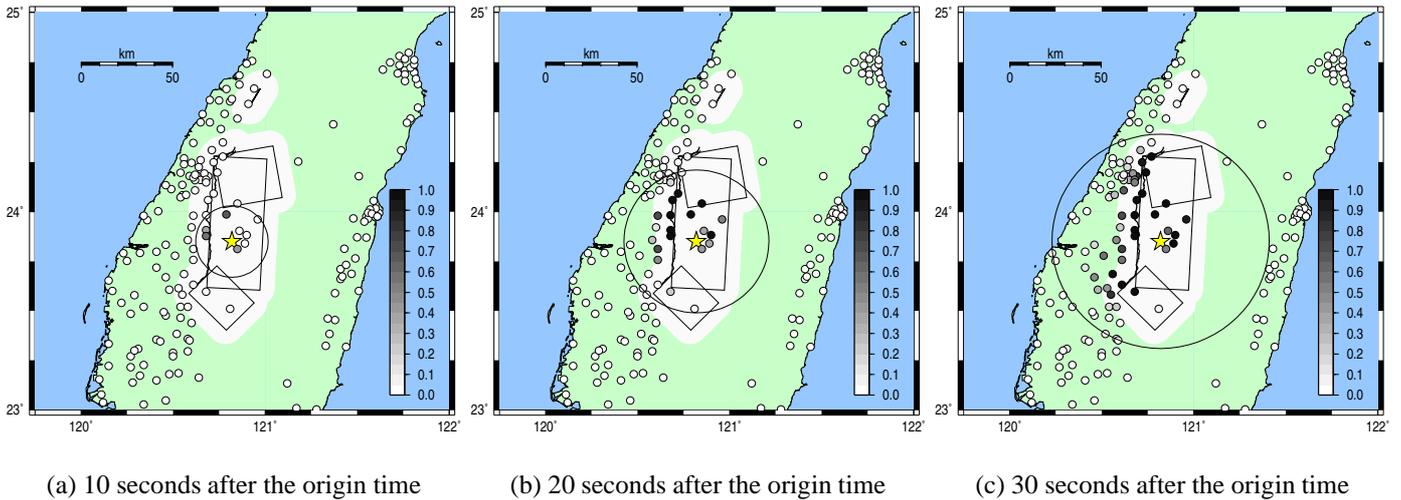


Figure 4: Snapshots of the probabilities of near-source for the Chi-Chi earthquake, based on the optimal discriminant function from the Bayesian approach. The large circle is the theoretical rupture front assuming the rupture velocity 2km/s.

function fits better to a dataset. However, the Ockham factor penalizes a more complicated model class. Model class selection solves the trade-off between the simplicity and reliability of the model class, and model class 1 is selected with the highest evidence.

We found that the horizontal and vertical components of same frequency range (e.g., horizontal and vertical accelerations) have significant interaction (Yamada *et al.*, 2006). Parameters with the same component and similar frequency range (e.g., horizontal acceleration and horizontal velocity) are also strongly correlated (Yamada *et al.*, 2006). These correlated parameters do not have significant effect on the performance of the classification. Therefore, models with these correlated parameters are rejected by the Ockham factor and the simpler model class is selected.

5 RESULTS AND DISCUSSION

We apply the optimal discriminant function from the Bayesian approach (in equations (9) and (10)) to all the stations in the dataset. Figure 1 (right) shows the classification results for the Imperial Valley and Chi-Chi earthquake dataset. The distribution of stations with a high probability of being in the near-source is consistent with the fault geometry.

To examine the application for real-time analysis, the optimal discriminant function in equations (9) and (10) is applied to the Chi-Chi earthquake strong motion records. We generated snapshots of the probability that a station is near-source from 10 seconds to 40 seconds after the beginning of rupture. Peak ground motions used for this classification analysis are computed from observed data every 10 seconds for each station and evaluated in the discriminant function. The results are shown in Figure 4. A darker mark at a station in Figure 4 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 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 seconds, the probability of being in the near-source at thirteen 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/s (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 seconds of rupture. At 30 seconds, 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.

6 CONCLUSION

We presented a methodology to classify seismic records into near-source or far-source records in order to estimate fault dimension in an earthquake early warning system. Ground motion records for past earthquakes are analyzed to find a function that best discriminates near-source and far-source records. Peak values of jerk, acceleration, velocity, and displacement are used to find the linear combination of peak values which provides the best performance to classify near-source and far-source records. We also analyzed which combination of ground motion features had the best performance for classification using Bayesian model class selection, and the best discrim-

inant function among the model classes examined is:

$$f(X_i|\theta) = 6.046 \log_{10} Za + 7.885 \log_{10} Hv - 27.091 \quad (11)$$

$$P(Y_i = 1|X_i, \theta) = \frac{1}{1 + e^{-f(X_i|\theta)}} \quad (12)$$

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 magnitude greater than 6.0, so it only works for large earthquakes.

7 REFERENCES

- Allen, R. M., & Kanamori, H. 2003. The potential for earthquake early warning in Southern California, *Science* 300, 786-789
- Beck, J. L., & Yuen, K. 2004. Model selection using response measurements: Bayesian probabilistic approach, *J. Eng. Mech.*, 130, 192-203.
- Beck, J. L., & Katsfygiotis, L. S. 1998. Updating models and their uncertainties. I: Bayesian statistical framework, *J. Eng. Mech.*, 124(4), 455-461.
- Boore, D. M. 2001. Effect of baseline correction on displacement and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, earthquake, *Bull. Seism. Soc. Am.* 91, 1199-1211.
- 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, 2005.
- 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 eds., 53-74.
- Hartzell, S., & Heaton, T. 1983. Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. Seism. Soc. Am.* 73, 1553-1583.
- Honda, R., Aoi, S., Morikawa, N., Sekiguchi, H., Kunugi, T., & Fujiwara, H. 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 Planets Space*, (submitted).
- Iwan, W. D., Moser, M. A., & Peng, C.-Y. 1985. Some observations on strong-motion earthquake measurement using a digital accelerograph, *Bull. Seism. Soc. Am.* 75, 1225-1246.
- Ji, C., Helmberger, D. V., Wald, D. J., & Ma, K. F. 2003. Slip history and dynamic implication of 1999 Chi-Chi earthquake. *J. Geophys. Res.*, 108(B9), 2412, doi:10.1029/2002JB001764.
- Li, Y., C. Campbell, & Tipping, M. 2002. Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics*, 18, 1332-1339.
- Nakamura, Y. 1988. On the urgent earthquake detection and alarm system (UrEDAS), *Proc. of the 9th World Conference on Earthquake Engineering, Tokyo-Kyoto, Japan. VII.* 673-678.
- Papadimitriou, C., Beck, J. L., & Katsfygiotis, L. S. 1997. Asymptotic expansions for reliabilities and moments of uncertain dynamic systems, *J. Eng. Mech.*, 123(12), 1219-1229.
- Sekiguchi, H., Irikura, K., Iwata, T., Kakehi, Y., & Hoshihara, M. 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.
- Sekiguchi, H., & Iwata, T. 2002. Rupture process of the 1999 Kocaeli, Turkey, earthquake estimated from strong-motion waveforms, *Bull. Seism. Soc. Am.*, 92, 300-311.
- Sivia, D. S. 1996. *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Oxford.
- Tsuboi, S., D. Komatitsch, Ji, C., & Tromp, J. 2003. Broadband modeling of the 2002 Denali fault earthquake on the Earth Simulator, *Phys. Earth Planet. Interiors* 139, 305-312.
- 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., Heaton, T. H. & D. V. Helmberger, 1991. Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong motion and broadband teleseismic data, *Bull. Seism. Soc. Am.*, 81, 1540-1572.
- Wald, D. J., & Heaton, T. H. 1994. Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake, *Bull. Seism. Soc. Am.* 84, 668-691.
- Wald, D. J., Heaton, T. H., & Hudnut, K. W. 1996. A dislocation model of the 1994 Northridge, California, earthquake determined from strong-motion, GPS, and leveling-line data, *Bull. Seism. Soc. Am.* 86, S49-70.
- Yamada, M., Heaton, T. H. & Beck, J. L. 2006. Early warning systems for large earthquakes: near-source versus far-source classification, *Bull. Seism. Soc. Am.* (submitted)