# Time-Frequency Clustering and Discriminant Analysis

Robert H. Shumway[1]

**Abstract:** We consider the use of time-varying spectra for classification and clustering of non-stationary time series. In particular, recent developments using local stationarity and Kullback-Leibler discrimination measures of distance are exploited for classifying earthquakes and mining explosions at regional distances

**Key words:** Spectral analysis, Kullback-Leibler, Seismology, Nuclear testing, Earthquakes and explosions.

## 1. Monitoring a CTBT

A fundamental problem faced in monitoring a potential comprehensive nuclear test ban treaty (CTBT) is that of discriminating between seismic records originating from nuclear explosions and those generated by other seismic events such as earthquakes and mining explosions. In areas where no nuclear testing has occurred, it is also of importance to be able to identify new events of suspicious origin that are substantially different from previously encountered events from that area. Hence, classification and clustering become important tools for analyzing potential violations of a CTBT.

Most current discriminants depend on measurements of the power spectrum read over specific frequency bands. The inherent non-stationarity of the data is accounted for by extracting spectral components corresponding to primary and secondary arrival phases. For example, Figure 1 shows a typical earthquake and mining explosion from a suite of eight earthquakes and eight explosions originating in the Scandinavian peninsula recorded by arrays in Scandinavia. Note that both series contain two phases or arrivals, generally denoted by P, the initial body wave and S, the later shear wave. The initial P-wave is very small in the typical earthquake, relative to the S-wave, as can be seen in Figure 1. Ratios and amplitudes of the two components as well as spectral ratios in different frequency bands (see, for example, McQuarrie et al, 1993, Kakizawa et al, 1998) are used as features in ordinary linear discriminant analysis, or as fused linear and quadratic discriminants (Anderson and Taylor, 2001).

Kakizawa et al (1998) also apply discriminants based on the assumption that the bivariate P and S waves differ only in their two-dimensional spectral matrices (see also Shumway and Stoffer, 1990, Chapter 5). However, it is clear that even the primary P and S-waves in Figure 1 are still not stationary. This observation leads us to consider the time-varying spectrum as an approach to classification and clustering for locally stationary processes. We consider applying locally stationary versions of Kullback-Leibler (K-L) discrimination information measures that give optimal time-frequency statistics for measuring the discrepancy between two non-stationary time series. We show that time-frequency profiles for earthquakes and nuclear explosions differ in important ways and that

[1]Robert H. Shumway is Professor of Statistics, University of California, Davis CA

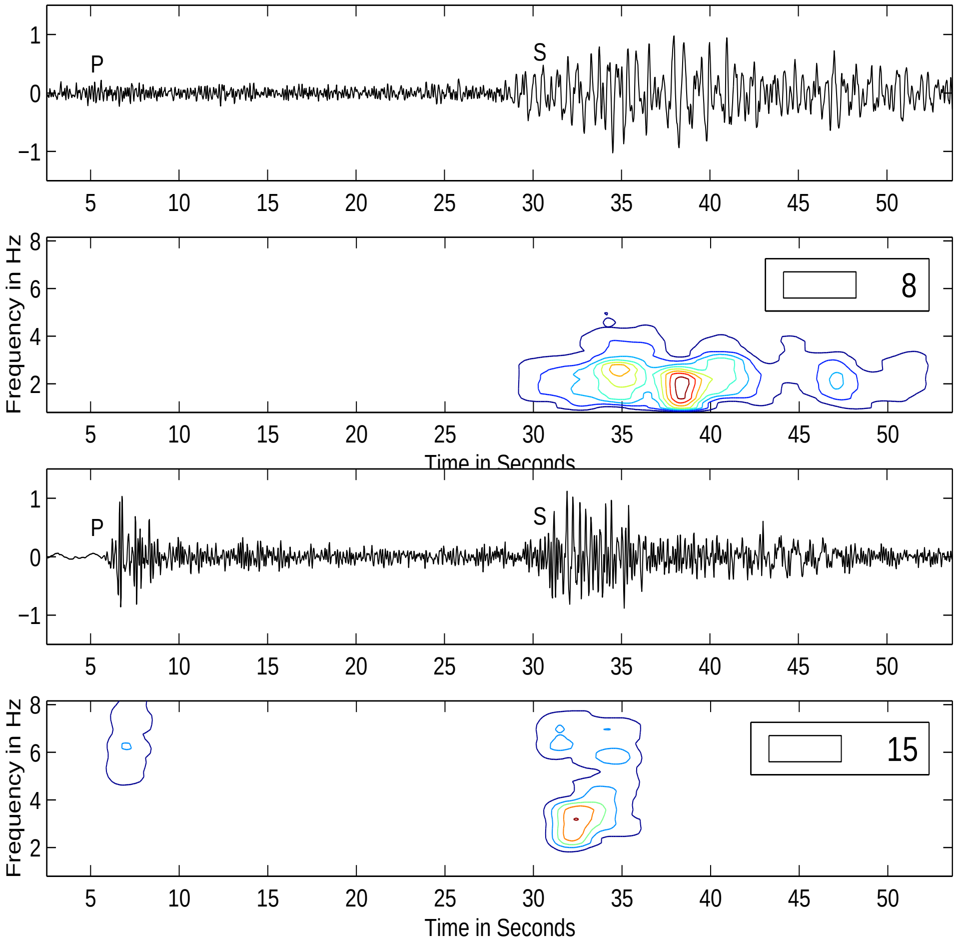


Figure 1: A typical earthquake (top two panels) and a typical explosion (bottom two panels) with time and frequency profiles (Events 8 and 15 in the file). The sampling rate is 40 points per second.

the K-L discrepancy measures, integrated over frequency and time, discriminate as well or better than many of the standard measures.

## 2. Locally Stationary Processes

Non-stationary processes with a time varying spectral representation, first considered in detail by Priestley (1965), have been recently be formulated in a rigorous asymptotic framework by Dahlhaus (1996, 1997). In particular, the existence of a spectral density $f(\nu, y)$ varying over both frequency $\nu$ and time $t$, is shown, where $-1/2 \leq \nu \leq 1/2$, with $\nu$ measured in cycles per unit time.

If we suppose that a given time series $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ has probability densities $p_1(\boldsymbol{x})$ and

$p_2(\boldsymbol{x})$ under hypotheses $H_1$ and $H_2$, the Kullback-Leibler discrimination information rate

$$I_n(1:2) = \frac{1}{n} \int p_1(\boldsymbol{x}) \log \frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})} \, d\boldsymbol{x}$$

measures the average discrepancy between two hypotheses, say $H_1$ : Earthquake and $H_2$ : Explosion. If the processes have zero means and differ only in the time-varying spectra, then results in Dahlhaus (1996, 1997) and Sakiyama and Taniguchi (2001) for the multivariate case, allow writing

$$I(f_1 : f_2) = \lim_{n\to\infty} I_n(1:2) = \int_{\nu,t} \Big[ \frac{f_1(\nu,t)}{f_2(\nu,t)} - \log \frac{f_1(\nu,t)}{f_2(\nu,t)} - 1 \Big] \, d\nu \, dt. \tag{1}$$

The discrimination information $I(f_1 : f_2)$ is a measure of the discrepancy between the hypotheses specifying either $f_1(\nu,t)$ or $f_2(\nu,t)$ and can be considered as a measure of the distance between the two densities $p_1(\boldsymbol{x})$ and $p_2(\boldsymbol{x})$. It is easy to show that $I(f_1 : f_2) \geq 0$, with equality if and only if $p_1(\boldsymbol{x}) = p_2(\boldsymbol{x})$ almost everywhere. It is not a real distance because it is not symmetric and doesn't satisfy the triangle inequality. For clustering, it is more convenient to use the symmetric information divergence

$$J(f_1 : f_2) = I(f_1 : f_2) + I(f_2 : f_1). \tag{2}$$

As approximations to the spectra in the above limiting expression, we use a tapered local spectral estimator for $f(\nu,t)$, smoothed over L contiguous frequencies and and over a local time interval of width $M$. We then add overlapping frequency-time smoothed estimators to form approximate versions of (1) and (2).

There is considerable latitude in selecting the parameters $L, M$, the degree of overlap in time and frequency, for obtaining reasonable estimates for the time-varying spectral densities $f_1(\nu,t), f_2(\nu,t)$. There will also be a frequencies where both spectra are essentially zero and these can be omitted from the approximating sum. For the seismic data in Figure 1, data are available at 40 points per second, so that the folding frequency is 20 cycles per second. A relatively large overlap in time was used with spectral estimators computed at increments of $\nabla t = 5$ points, with $M = 201$ time points or 5 seconds used for the spectral estimator at time $t$ Smoothing over $L = 9$ frequencies retained enough smoothness to produce sensible time-varying spectra such as the two given in Figure 1.

The two events in Figure 1 illustrate some of the essential differences between earthquakes and explosions. Note that the initial or P-wave is larger for the explosion (labeled 15) than for the earthquake (labeled 8), relative to the S-wave. A simple, but very effective discriminant is the logarithm of the ratio of the amplitude of the P-wave to the amplitude of the S-wave, denoted herein by $\log(P/S)$. Also, the spectrum of the S-wave for earthquakes is spread over a 20 second time window, whereas the spectrum of the explosion S-wave is only about 5 seconds long. The primary frequency difference seems to be in the lower frequency power present in the S-wave of the earthquake. Most of the power is below three Hz for the earthquake and above three Hz for the explosion. Unfortunately, these features may be intermingled in other earthquakes and explosions.

If the differences in the preceding discussion seem to be consistent throughout the sample, we could hope to be successful using a sample approximation for the discrimination information measure (1). To investigate whether there are consistent differences in the two population spectra, it is

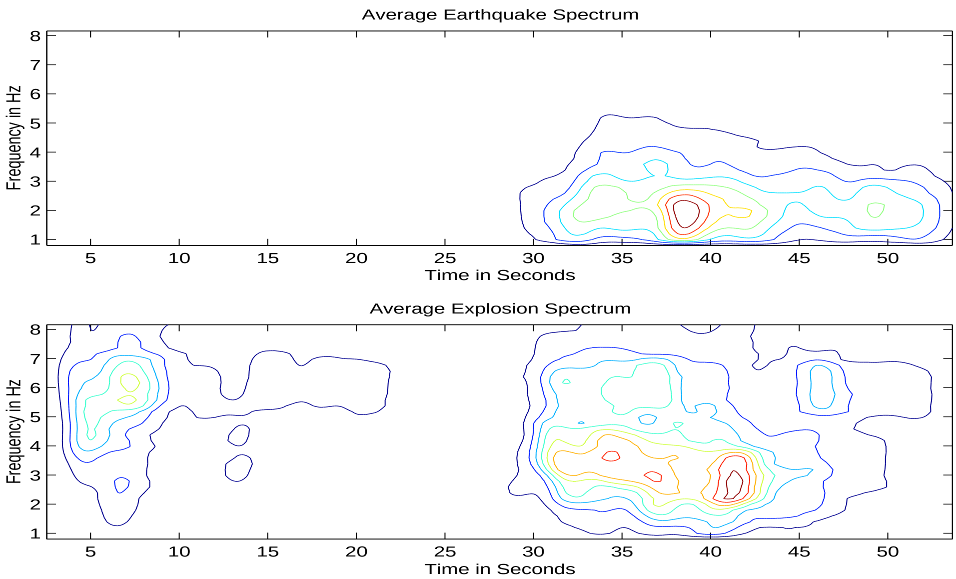


Figure 2: Mean time varying spectra for earthquake and explosion population

convenient to construct averages of the earthquake and explosion spectra. These mean spectral estimators, shown in Figure 2, suggest that the average characteristics for the two populations tend to emulate those suggested by Figure 1. It is certainly clear that the power in the explosion P-wave is larger, relative to the S-wave power, for the explosion than for the earthquakes. However, it is also clear that there is no completely consistent pattern for the spectra of the earthquakes and explosions. The average earthquake spectrum still has lower frequencies, concentrated mainly in the 0-4 Hz range, whereas the explosion spectra go out to about 8 Hz.

## 3. Discriminant Analysis and Clustering

In general, there are two approaches that can be taken to discriminant analysis, characterized usually as feature extraction and optimal signal discrimination. Feature extraction proceeds by defining characteristics of the process, namely, the spectrum, in this case, and then combining these features in a classical discriminant analysis. Cavanaugh et al (1993) investigated amplitude ratios and spectral ratios in the bands 0-3 Hz , 3-6 Hz, and 6-9 Hz for the P and S components. They concluded that simple amplitude ratios worked as well as spectral ratios. Kakizawa et al (1998) developed optimal bivariate discriminants using the multivariate time-invariant forms of the disciminant functions (1) and (2); these turned out to be slightly better than any of the features.

In the present paper, we have looked at the measure (1) as a possible discrimination procedure and (2) for cluster analysis. Suppose we define $\bar{f}_1(\nu, t), \bar{f}_2(\nu, t)$ as the group means of the estimated spectra in the earthquake and explosion group respectively. The sample means are plotted in Figure 2 and we use these for the theoretical spectra. Then, denoting $\hat{f}(\nu, t)$ for the estimated sample spectrum of a particular event, we classify the observation into the population with density $p_1(\boldsymbol{x})$

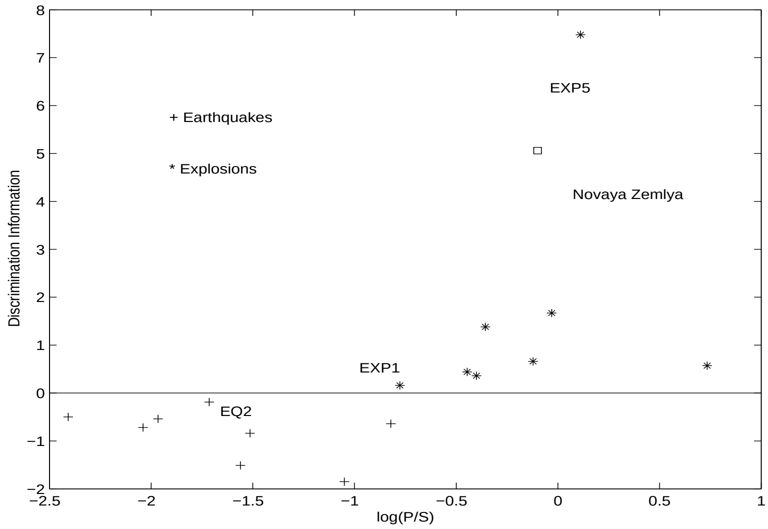


Figure 3: Amplitude ratio discriminant (abcissa) compared to the time-varying discrimination information (ordinate).

when $\hat{I}(\hat{f}, \bar{f}_1) \leq \hat{I}(\hat{f}, \bar{f}_2)$ and classify into $p_2(\boldsymbol{x})$ otherwise, where $\hat{I}$ denotes the discrimination information evaluated using the smoothed time-varying spectral estimators. This implies that the test statistic

$$T = \hat{I}(\hat{f}, \bar{f}_1) - \hat{I}(\hat{f}, \bar{f}_2) \tag{3}$$

can be compared with zero for each series, with $T \leq 0$ implying $H_1$ : Earthquake and $T > 0$ implying $H_2$ : Explosion

Figure 3 shows the statistic $T$ in (3) plotted on the vertical scale compared to one of the best feature discriminants, $\log P/S$ plotted on the abscissa. Note that setting the cut-off at zero results in no errors for the discimination information statistic. Parenthetically, setting the cut-off for the amplitude feature at about -.75 will also result in no misclassifications. The event of unknown origin from the Russian test-site Novayla Zemlya appears to be in the explosion population, but, like Explosion 5, it is somewhat of an outlier. The Russians have not admitted to a test ocurring at the time this event was recorded.

We also considered hierarchichal cluster analysis using the quasi-distance measure $J(\hat{f}_i, \hat{f}_j)$ in (2) for relating the $i^{th}$ and $j^{th}$ population members. Performing the hierarchichal analysis, i.e., always adding an element to the nearest member of a cluster leads to a final set of two clusters. The first cluster contained all earthquakes and the eighth explosion whereas the second cluster contained all of the other seven explosions plus the unknown Novaya-Zemlya event. The closest events were the sixth earthquake and the eighth explosion, shown in Figure 4. This is an unusual case, where the distance measure was small due to the fact that the S-phases both contain the strong low frequencies that are dispersed in time, characteristic of an earthquake, whereas the measure fails to detect the stronger P-Phase (relative to S) in the explosion, shown in the bottom two panels. This is a case

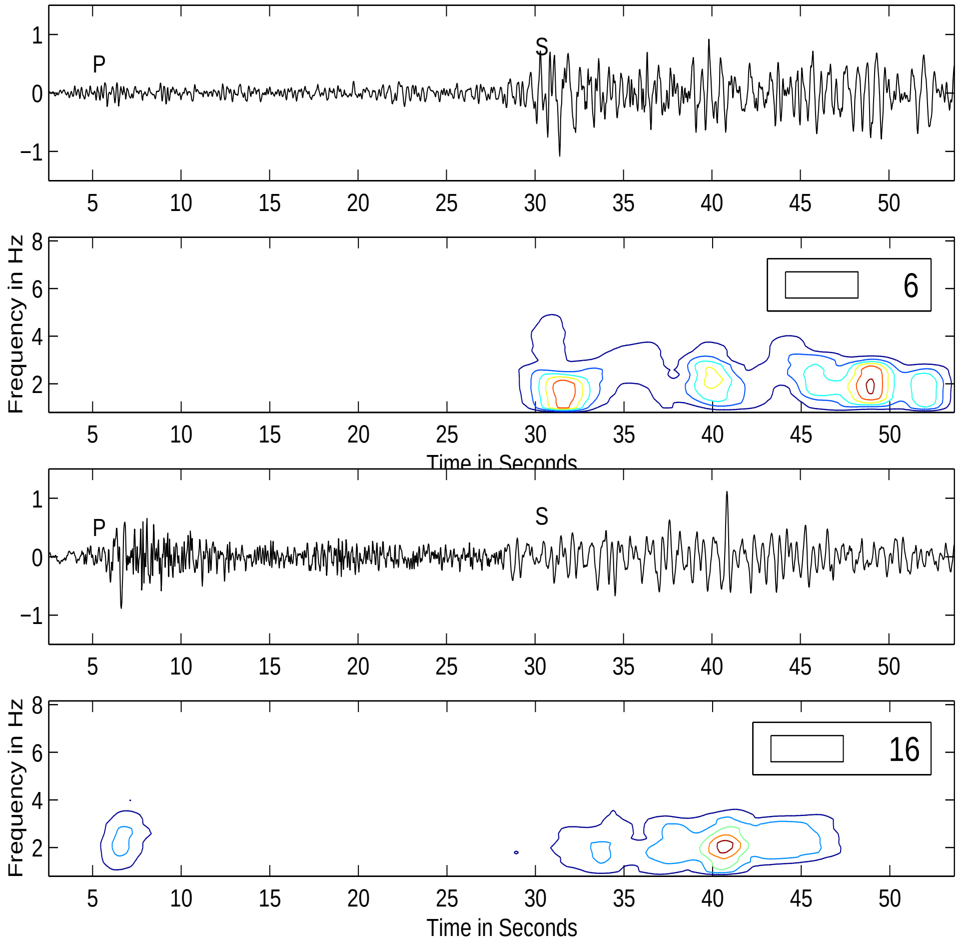


Figure 4: Closest events, EQ6 and EXP8 (Events 6 and 16), as determined by the J-Divergence (3). The sampling rate is 40 points per second.

that would not have been confused using the amplitude ratio discriminant.

## 4. Discussion

We have focused here on development of a discrimination procedure for non-stationary time series that could be implemented with and online processor. Since no features are extracted, this procedure would eliminate arbitrary decisions about which frequencies to apply and how to combine the features. The only potential difficulty of automating would be the matching the arrival times of the P and S phases on the record. This could be done directly from the underlying time varying spectra, i.e., we simply automatically match the two P and S spectral profiles by applying one of the change-point detection algorithms.

A further innovation could be provided if it is thought that differences between populations could be described through differences in the coherence structure. In this case, the generalization of (1) to the case of a p-variate vector series $\boldsymbol{x}_t = x_{t1}, \ldots, x_{tp})'$, with $p \times p$ spectral matrices $S_1(\nu, t)$ and $S_2(\nu, t)$ under hypotheses $H_1$ and $H_2$ becomes

$$I(S_1 : S_2) = \int_{\nu,t} \left[ \text{tr}\{S_1(\nu,t) S_2^{-1}(\nu,t)\} - \log \frac{|S_1(\nu,t)|}{|S_2(\nu,t)|} - p \right] d\nu \; dt. \tag{4}$$

Kakizawa et al (1998) have considered the stationary version of this problem and have given an example using the data of this paper, regarding the P and S phases as components of a bivariate time series. It should also be noted that large-sample expressions for misclassification rates have been worked out for both the stationary (see Kakizawa et al, 1998) and non-stationary cases (Sakiyama and Taniguchi, 2001).

Finally, we should emphasize that there are similar data situations, such as in the analysis of functional magnetic resonance imaging (fMRI) or electro-encephalographic (EEG) series where the above procedures might apply in whole or in part. EEG analysis, in particular is dependent on analyzing simultaneously many channels of brain wave data where non-stationary bursts at different frequencies (alpha, beta, gamma) may be associated with transient phenomena such as dreaming or epileptic seizures.

## 5. References


Anderson, D.N. and S.R. Taylor (2001). Application of regularized discrimination analysis to regional seismic event identification. To Appear: *Bull. Seismolog. Soc. Amer.*.

Cavanaugh, J.E., A.D.R. McQuarrie and R.H. Shumway (1993). Parametric and Nonparametric Discriminants for Regional Earthquakes and Explosions. Scientific Report PL-TR-93-2164, Phillips Lab, Hanscom Air Force Base, MA 01731-3010.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes. Appl.*, 62, 139-168.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.*, 25, 1-37.

Kakizawa, Y., R.H. Shumway and M. Tanaguchi (1998). Discrimination and clustering for multivariate time series. *J. Amer. Statist. Assoc.*, 441, 328-340.

Priestley, M.B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc., B*, 28, 228-240.

Sakiyama, K. and M. Taniguchi (2001). Discriminant analysis for locally stationary processes. Report S-58, Research Reports in Statistics, Osaka University, Toyonaka, Osaka, Japan.

Shumway, R.H. and D.S. Stoffer (2000). *Time Series Analysis and Its Applications.* New York: Springer-Verlag.