Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®

Statistics & Probability Letters 63 (2003) 307–314

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Time-frequency clustering and discriminant analysis

Robert H. Shumway*

Department of Statistics, University of California, One Shields Ave., Davis, CA 95616-8705, USA

Received November 2002; received in revised form February 2003

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.

Keywords: Spectral analysis; Kullback–Leibler; Seismology; Nuclear testing; Earthquakes and explosions

1. Introduction

The analysis of relatively long stretches of time series data that may contain either slow or rapid changes in the spectrum is of interest in a number of areas. The use of the time-varying spectra has been endemic to underwater acoustics, where it is called the sonogram, and in recorded speech where the time-varying spectra are called voiceprints. Further applications are to non-stationary encephalographic (EEG) recordings that may contain spontaneous bursts of energy caused by epileptic seizures or longer term changes that are characteristic of diseases, for example, in Alzheimer patients. However, the sheer volume of seismic data as well as a rich empirical history suggesting that fundamental properties of seismic events are expressed through the time-varying spectrum make this kind of data ideal for illustrating the possibilities for time series discrimination and classification.

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.

* Tel.: +1-530-752-7099.
E-mail address: rhshumway@ucdavis.edu (R.H. Shumway).

0167-7152/03/$ - see front matter
doi:10.1016/S0167-7152(03)00095-6

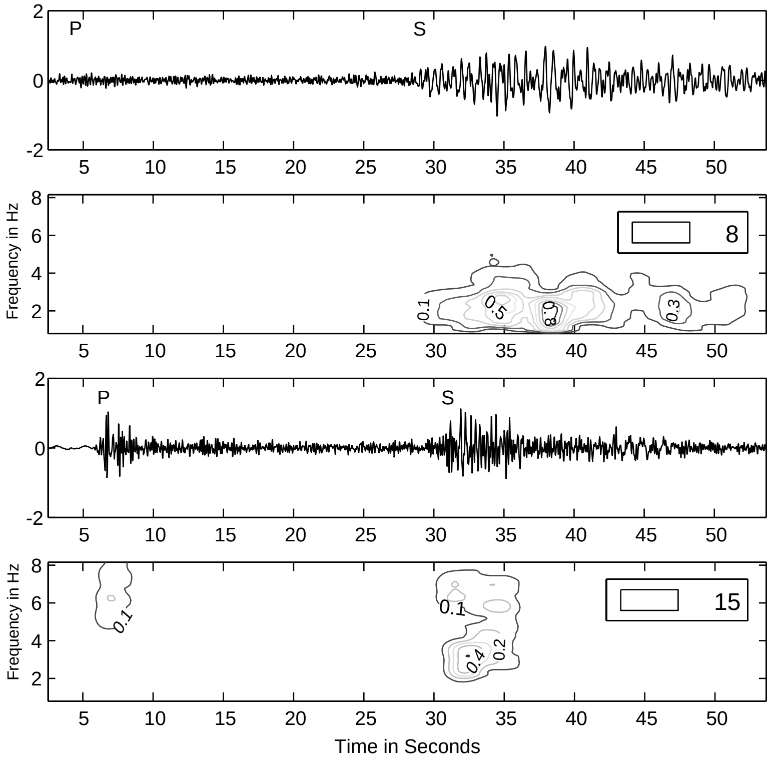

Fig. 1. A typical earthquake (top two panels) and a typical explosion (bottom two panels) with time and frequency profiles (Events 8 and 15 in the file). The sampling rate is 40 points per second.

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 seismic data is accounted for by extracting spectral components corresponding to primary and secondary arrival phases. This is particularly useful for seismic data for events observed at regional distances of less than 1000 km. For example, Fig. 1 shows a typical earthquake and mining explosion from a suite of eight earthquakes and eight explosions originating in the Scandinavian peninsula and 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 Fig. 1. Ratios involving amplitudes of the two components as well as spectral ratios in different frequency bands (see, for example, Kakizawa et al., 1998) are used as features in ordinary linear discriminant analysis or as fused linear and quadratic discriminants (Anderson and Taylor, 2002).

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, 2000, Chapter 5). However, it is clear that even the primary P- and S-waves in Fig. 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 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 formulated in a rigorous asymptotic framework by Dahlhaus (1996, 1997). In particular, the existence of a spectral density $f(v,y)$ varying over both frequency v and time t, is shown, where $-\frac{1}{2} \leqslant v \leqslant \frac{1}{2}$, with v 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 K–L discrimination information rate

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

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) allow writing

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

The discrimination information $I(f_1 : f_2)$ is a measure of the discrepancy between the hypotheses specifying either $f_1(v,t)$ or $f_2(v,t)$ 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 $I(f_1 : f_2) \geqslant 0$, 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 $f(v,t)$, smoothed over L contiguous frequencies and over a local time interval of width M. We then add overlapping time-frequency 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 $f_1(v,t), f_2(v,t)$. We also note that there are less arbitrary methods for segmenting the series into locally stationary pieces (see, for example, Adak, 1998 or Ombao et al., 2001). For the purposes of the present example, it seemed sensible to require spectra that varied smoothly in frequency and time rather than the blocky spectra that are typically given by the dyadic partitioning methods given in the references mentioned above.

For the seismic data in Fig. 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 s used for the spectral estimator at time t. Smoothing over $L = 9$ frequencies and using a cosine bell for time tapering retained enough smoothness to produce sensible time-varying spectra such as the two given in Fig. 1. Some kind of time tapering seems essential and it may be that the tapers proposed by Ombao et al. (2001) will be more useful in applications where there are sudden changes in the spectral content or where orthogonality becomes important. It is interesting that Ombao et al. (2001) use the discrimination information measure (1) to choose the bandwidth, with the periodogram and smoothed spectrum corresponding to f_1 and f_2 in (1).

There will also be a frequencies where both in the approximations to (1) and (2) are essentially zero and these can be omitted from the approximating sum by the following device. Noting that the magnitudes of the typical spectra in Fig. 1 and later were in the range 0.1 to 0.8, it seemed reasonable to set $f_1 = f_2$ when either were less than 10^{-5} so that there is essentially no contribution to the approximate versions of (1) and (2) for negligible spectral content.

The two events in Fig. 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 s time window, whereas the spectrum of the explosion S-wave is only about 5 s 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 3 Hz for the earthquake and above 3 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 convenient to construct averages of the earthquake and explosion spectra. These mean spectral estimators, shown in Fig. 2, suggest that the average characteristics for the two populations tend to emulate those suggested by Fig. 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 extend 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

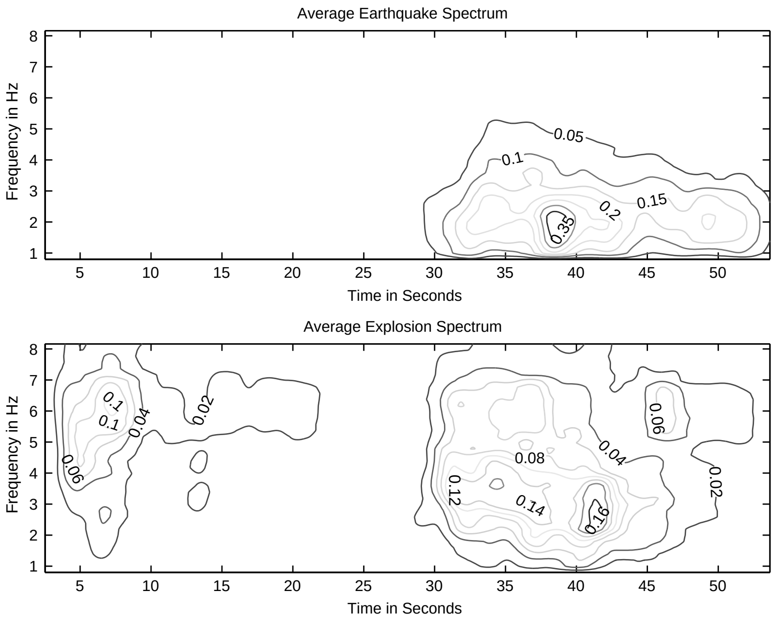

Fig. 2. Mean time varying spectra for earthquake and explosion population.

bivariate discriminants using the multivariate time-invariant forms of the discriminant functions (1) and (2); these turned out to be slightly better than any of the features (see Fig. 3).

In the present paper, we have looked at the measure (1) as a possible discriminant and at (2) for cluster analysis. Suppose we define $\bar{f}_1(v,t)$, $\bar{f}_2(v,t)$ as the group means of the estimated spectra in the earthquake and explosion group, respectively. The sample means are plotted in Fig. 2 and we use these for the theoretical spectra. Then, denoting $\hat{f}(v,t)$ as the estimated sample spectrum of a particular event, we classify the observation into the population with density $p_1(\boldsymbol{x})$ when $\hat{I}(\hat{f},\bar{f}_1) \leqslant \hat{I}(\hat{f},\bar{f}_2)$ and classify into $p_2(\boldsymbol{x})$ otherwise, where $\hat{I}$ denotes the discrimination information evaluated using the smoothed time-varying spectral estimators. 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 $T \leqslant 0$ implying H_1 : Earthquake and $T > 0$ implying H_2 : Explosion. Note that a hold-out one classification procedure was applied to provide a realistic estimator for the misclassification probabilities.

We also considered hierarchical cluster analysis using the quasi-distance measure $J(\hat{f}_i,\hat{f}_j)$ in (2) for relating the ith and jth population members. Performing the hierarchical 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 (X8) whereas the second cluster contained all of the other seven explosions plus the unknown Novaya–Zemlya event. The closest events were the fifth earthquake (Q5) and the eighth explosion (X8), shown in Fig. 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

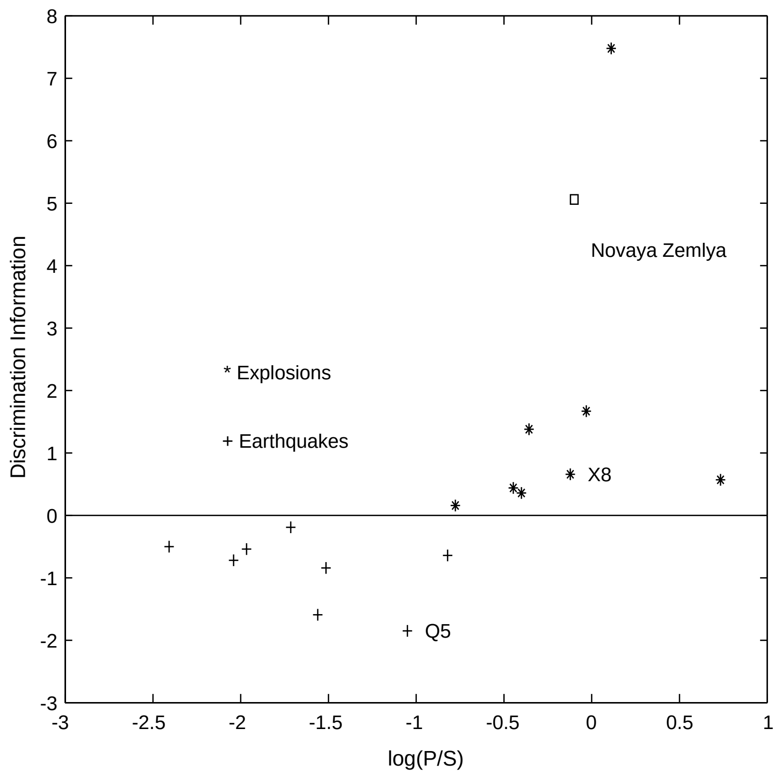

Fig. 3. Amplitude ratio discriminant (abcissa) compared to the time-varying discrimination information (ordinate).

bottom two panels. This is a case 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 an 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. The arrival times of the phases depend on such factors as the path taken from the source to the receiver and will differ for different event and receiver locations. Phases can be matched manually by analysts or automatically by applying one of the change-point detection algorithms. The cumulative sum of squares algorithm of Inclán and Tiao (1994) has worked well in preliminary experiments.

A further innovation to the approach here that focused on univariate spectra 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

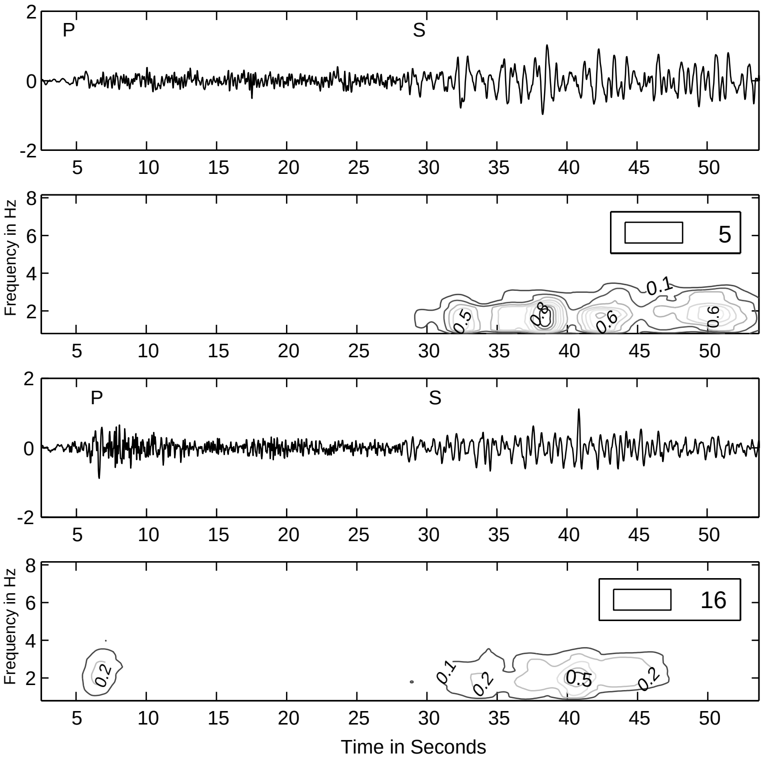

Fig. 4. Closest events, Earthquake 5 (Q5) and Explosion 8 (X8) (Events 5 and 16 in the file), as determined by the J-Divergence (3). The sampling rate is 40 points per second.

$\boldsymbol{x_t} = (x_{t1}, \ldots, x_{tp})'$, with $p \times p$ spectral matrices $S_1(v,t)$ and $S_2(v,t)$ under hypotheses H_1 and H_2 becomes

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

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 again that there are other 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.

Acknowledgements

The support of the nuclear monitoring office of the Air Force Technical Applications Center, Patrick Air Force Base, Florida is gratefully acknowledged. An anonymous reviewer made a number of thoughtful comments that improved the exposition.

References

Adak, S., 1998. Time dependent spectral analysis of non-stationary time series. J. Amer. Statist. Assoc. 93, 1488–1501.

Anderson, D.N., Taylor, S.R., 2002. Application of regularized discrimination analysis to regional seismic event identification. Bull. Seismolog. Soc. Amer. 92, 2391–2399.

Cavanaugh, J.E., McQuarrie, A.D.R., Shumway, R.H., 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 Process. Appl. 62, 139–168.

Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Ann. Statist. 25, 1–37.

Inclán, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. J. Amer. Statist. Assoc. 92, 739–747.

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

Ombao, H.C., Raz, J.A., von Sachs, R., Malow, B.A., 2001. Automatic statistical analysis of bivariate nonstationary time series. J. Amer. Statist. Assoc. 96, 543–560.

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

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

Shumway, R.H., Stoffer, D.S., 2000. Time Series Analysis and its Applications. Springer, New York.