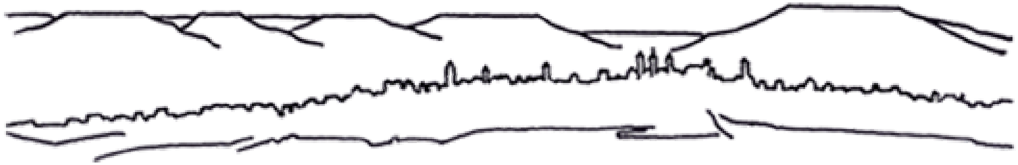
XXX CONGRESO NACIONAL ESTADISTICA Investigación OPERATIVA
y IV Jornadas de Estadística Pública



septiembre 25-28 <VALLADOLID’07>

Discrimination of Locally Stationary Time Series

Andres M. Alonso
Universidad Carlos III de Madrid

and

Elizabeth A. Maharaj
Monash University

Discrimination of Locally Stationary Time Series

- Time Series Patterns
- Some Existing Methods
- Wavelets and Wavelet Variances
- Wavelet Variances as Discriminating Variables
- Some Simulation Results
- Applications
- Further Research

Time Series Patterns

Examples

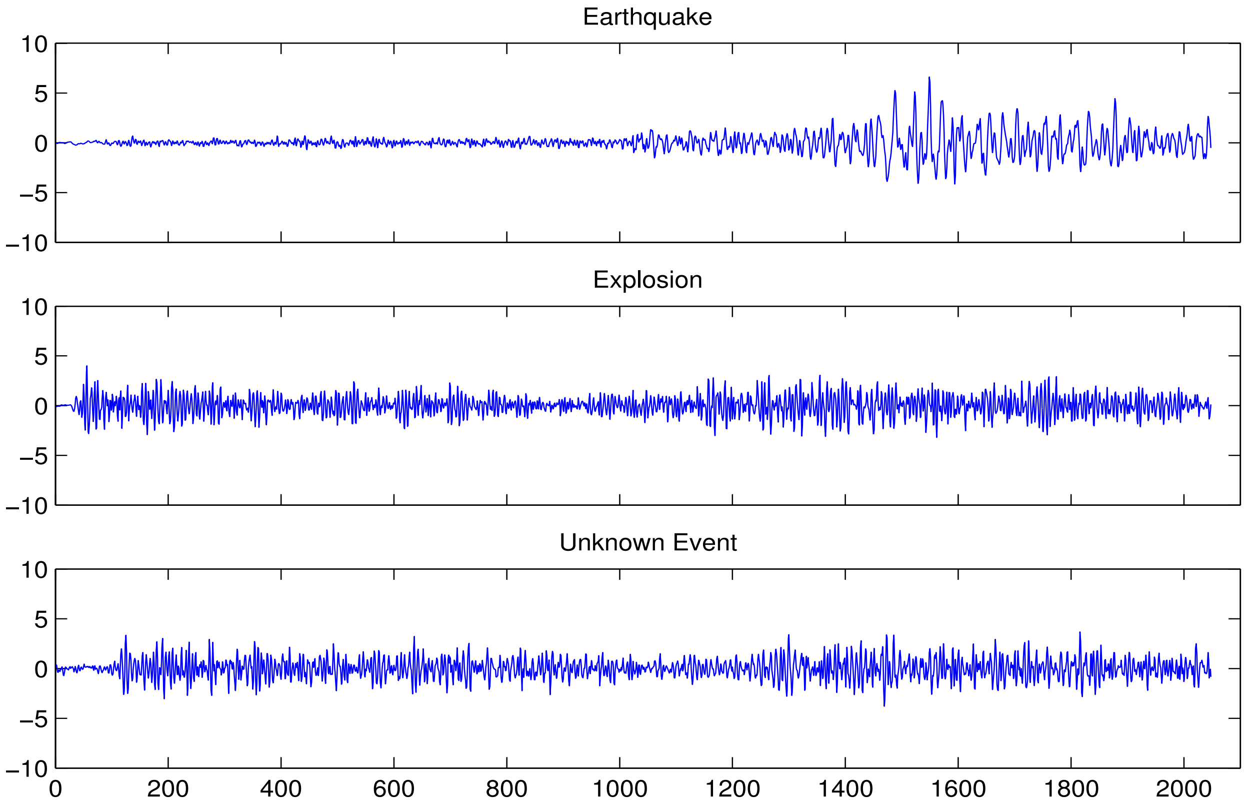
- Seismology
  - Earthquakes and Explosions
- Statistical Process Control
  - Control charts
- Medicine
  - Electroencephalogram

Earthquakes and Explosions

- In seismology, there is interest in differences and similarities between classes of events such as earthquakes, mining explosions and nuclear explosion.
- Monitoring nuclear proliferation critically depends on reliably being able to differentiate between small nuclear explosions and earthquakes.
- Because of the limited availability of past nuclear explosion data, researchers examine mining explosion data as surrogates that are expected to have similar patterns to low yield nuclear explosions.

Earthquakes and Explosions

- Kakizawa et al. (1998), Shumway (2003), Huang et al (2004), Chinipardaz and Cox (2004).
  - A suite of 8 earthquakes and 8 mining explosions originating from the Scandinavian Peninsula.
  - Unknown event originating from Novaya-Zemmlya, Russia.
  - Could this unknown event be an earthquake or an explosion?

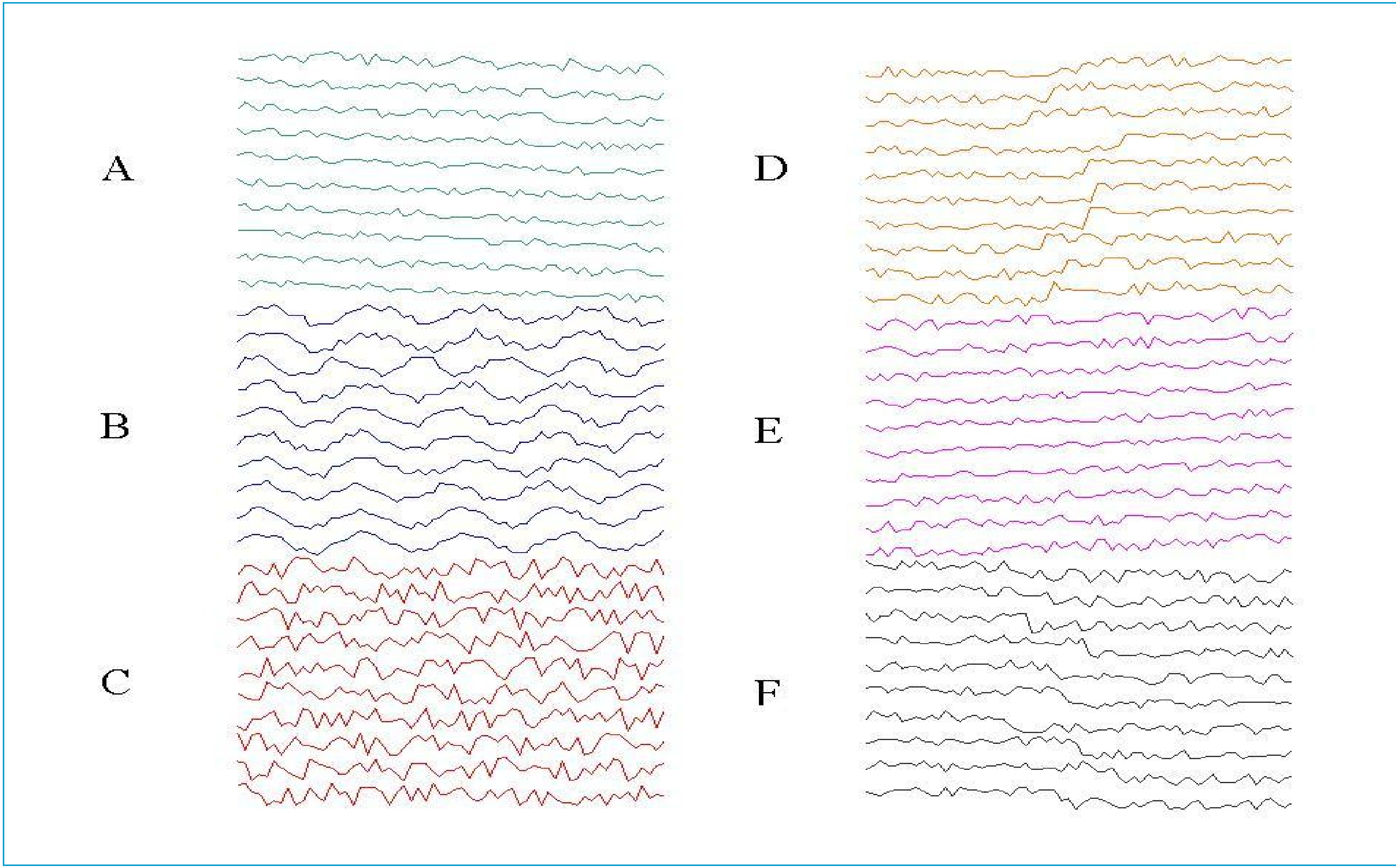


Both earthquakes and explosions contain two phases of arrival:

- the initial body wave, P-wave.
- the shear wave, S-wave.

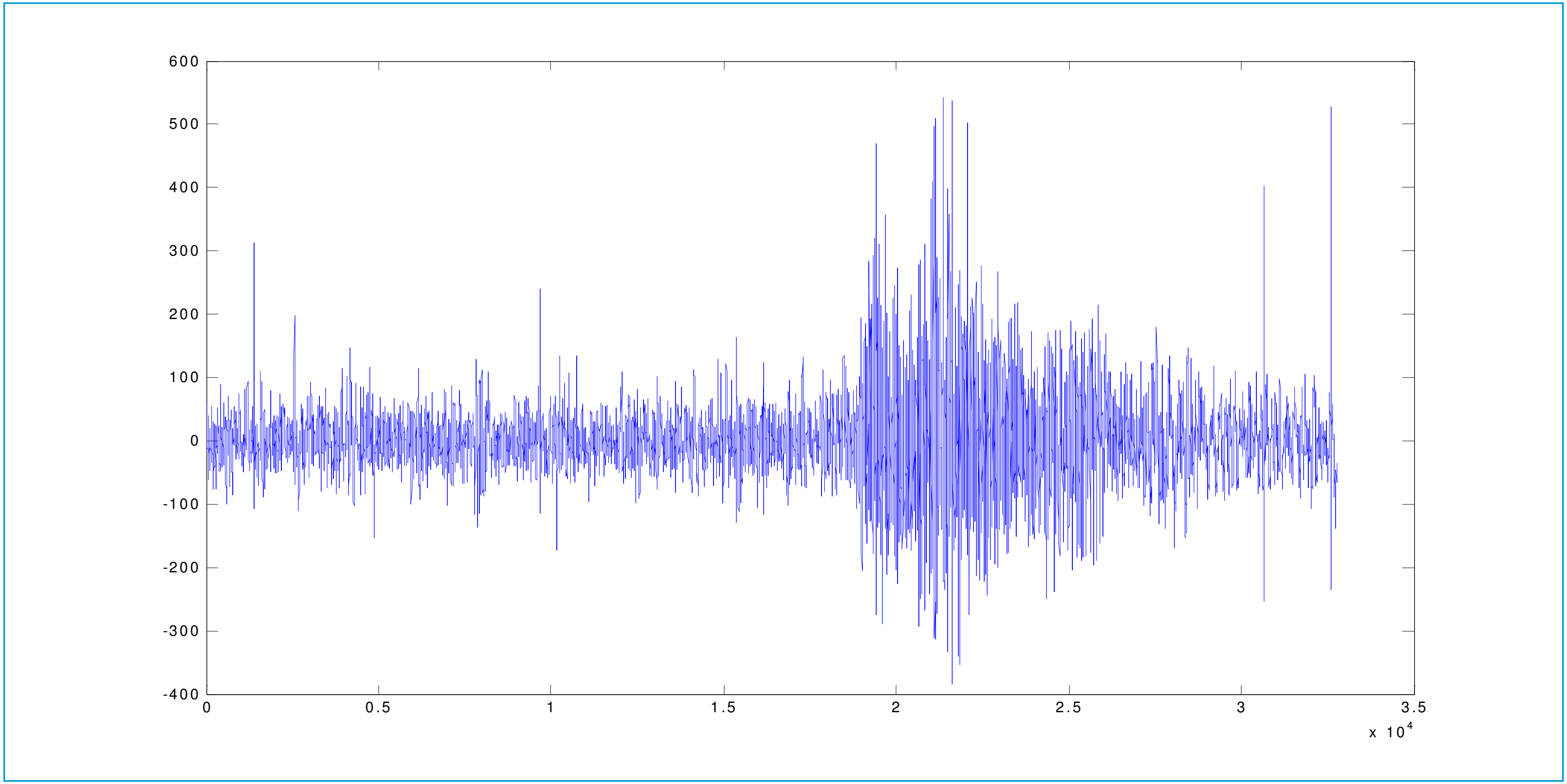
Control Charts

- Statistical process control uses statistical tools to observe the performance of the production line to predict significant deviations that may result in rejecting products.
- Control charts as used in statistical process control can exhibit six principal types of patterns: normal, cyclic, increasing trend, decreasing trend, upward shift and downward shift.
- Apart from normal patterns, all the other patterns indicate abnormalities in the process that must be corrected.
- Accurate and speedy detection of such patterns is important to achieving tight control of the process and ensuring good product quality.
- Example in Pham and Chan (1998).
  - Synthetically generated control charts.



(A) Downward Trend. (B) Cyclic. (C) Normal. (D) Upward Shift. (E) Upward Trend. (F) Downward Shift.

Electroencephalogram



Left temporal channel - T3

Discriminant Analysis

- When applying discriminant analysis to time series, use features associated with each time series.
- Approaches to the problem of discriminating among different classes of time series can be divided into two categories:
  - The optimality approach,
  - Feature extraction.

Time Series	Features				
	1	**2**	.	.	***p***
1	x_{11}	x_{12}	.	.	x_{1p}
2	x_{21}	x_{22}	.	.	x_{2p}
.	.	.	.	.	.
.	.	.	.	.	.
n	x_{n1}	x_{n2}	.	.	x_{np}

Discriminant Analysis

The Optimality Approach

- A time series is known to belong to one of *g* populations denoted by $\Pi_1, \Pi_2 \ldots \Pi_g$.
- General problem is to classify this time series into one of *g* groups in some optimal fashion.
- Makes specific Gaussian assumptions about the probability density function of the separate groups and then develops solutions that satisfy well-defined minimum error criteria.
- Assume the difference between the classes is expressed through differences in the theoretical mean and covariance functions and use likelihood methods to develop an optimal classification function.

Discriminant Analysis

Feature extraction

- This is a heuristic approach which looks at quantities that tend to be good discriminators for well separated populations and have some basis in physical theory and intuition.

- For example, for the earthquake and explosion, features may be associated with the maximum amplitude of the P-waves and S-waves.

- Less attention is paid to finding functions that are approximations to some well-defined optimality criterion.

Some Existing Methods of Discriminating between Time Series Patterns

- Frequency Domain Methods
  - Stationary time series – spectral estimates over specific frequency bands, Karizawa (1998), Shumway and Stoffer (2000).
  - Non-stationary time series – time varying spectra which involve selecting specific bandwidths and window lengths Shumway (2003), Huang et al. (2004).

- Functional Data-Analytic Approach
  - Hall et al. (2001) Regard signals as curves – use a functional data analytic method for dimension reduction before applying discriminant analysis.

- Neural networks
  - Pham and Chan (1998) in discriminating between control chart patterns.
  - Nigam and Graupe (2004) in discriminating EEG patterns.

Wavelets

- Wavelets are mathematical tools for analyzing signals and images in one or more dimensions

- The discrete wavelet transform DWT re-expresses a time series in terms of coefficients that are associated with a particular time and a particular dyadic scale.

- Long time scales give more low frequency information while short time scales give more high frequency information.

- The coefficients are fully equivalent to the original time series in that the time series can be perfectly reconstructed from its DWT coefficients.

- The DWT of a time series is an orthonormal transformation of the original series.

Wavelets and wavelets variance

- Given a time series with X_t with N data points
  - assumed to have decomposed into wavelet series, $W_{j,t}$, using the MODWT, the time dependent MODWT wavelet variance at the dyadic scale $\tau_j = 2^{j-1}$ is defined as

$$\upsilon^2_{X,t}(\tau_j) \equiv \mathrm{var}\{\tilde{W}_{j,t}\}$$

where $\tilde{W}_{j,t} = \sum_{\ell=0}^{L_j-1} \tilde{h}_{j,\ell} X_{t-\ell}, \quad t = 0, 1, \ldots, N-1$

$\tilde{h}_{j,\ell}$ is the wavelet filter, and L_j is the length of the j-th level filter.

Wavelet variance estimator

- Suppose that X can be divided into K blocks and each block is considered as a stationary time series. An estimator of the MODWT variance is

  $$\hat{\upsilon}^2_{X_i}\left(\tau_j\right) \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{W}^2_{j,t}$$

  where the wavelets coefficients, $W_{j,t}$, use X_i instead of X.

- The MODWT wavelet variance estimator is preferred because it has been shown to be asymptotically more efficient than an estimator based on the DWT.

Distribution of wavelet variance estimator

- Suppose that $\tilde{W}_{j,t}$ is a Gaussian stationary process with mean 0 and spectral density function S_j.

- If S_j is finitely squared integrable and strictly positive almost everywhere

- Then it has been shown that the estimator $\hat{\upsilon}_X^2(\tau_j)$ is asymptotically normal with mean $\upsilon_X^2(\tau_j)$ and large sample variance $2A_j/M_j$, where

$$A_j = \int_{-1/2}^{1/2} S_j^2(f)\,df$$

Wavelet Variances as Discriminating Variables

- Given a number of time series that belong to one of *g* groups
  - Obtain MODWT for each series
  - Determine the MODWT variance at each scale
- MODWT variances are features associated with each time series
  - Asymptotically normal
  - Leads to optimal discriminant solution
    - Linear or quadratic

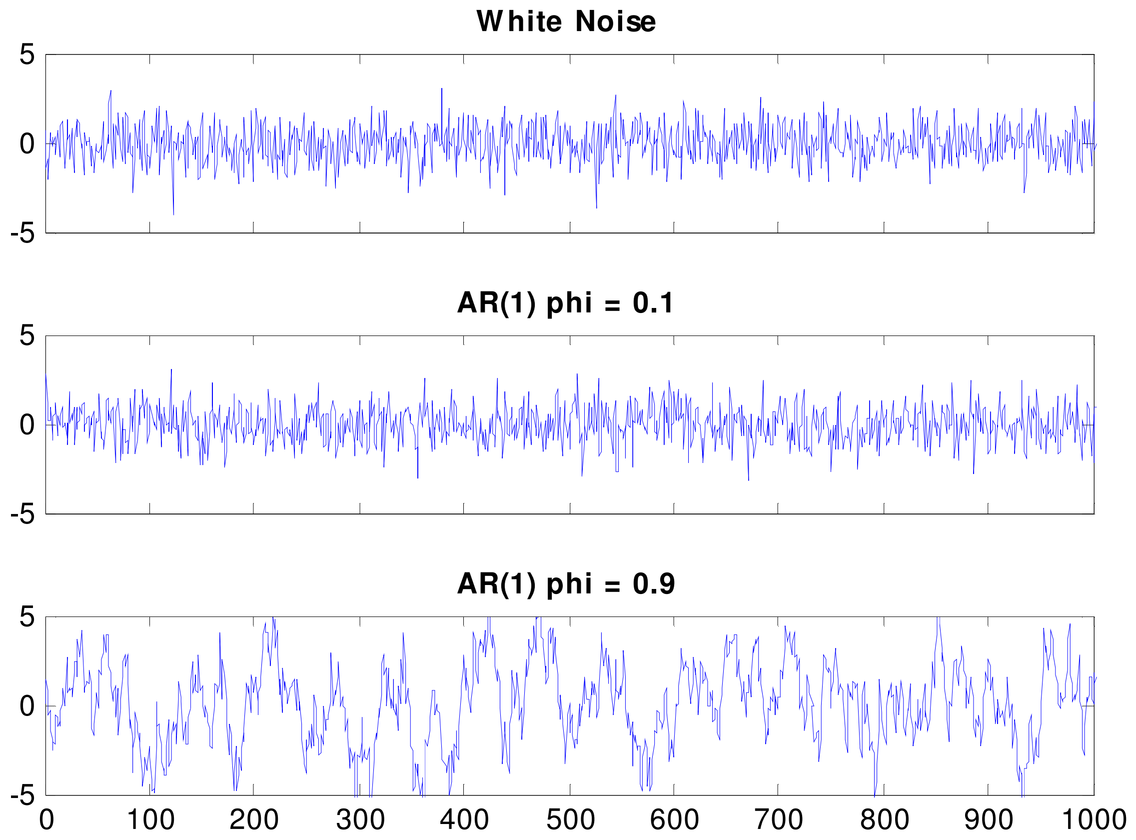
Wavelet Variances as Discriminating Variables

Time Series	Wavelet Levels				
	1	2	.	.	*J*
1	var_{11}	var_{12}	.	.	var_{1J}
2	var_{21}	var_{22}	.	.	var_{2J}
.	.	.	.	.	.
.	.	.	.	.	.
n	var_{n1}	var_{n2}	.	.	var_{nJ}

Simulation Study

- 15 series of length 256, 1024, 2048 from each of
  - $X_1(t)$: White noise
  - $X_2(t)$: AR(1) : ϕ = -0.9 to 0.9 in increments of 0.2
- 20 series: training sample, 10 series: holdout sample
- Wavelet variance obtained on 8, 10, 11 levels for series lengths 256, 1024 and 2048 respectively
  - Number of discriminating variables: p = 8, 10, 11 respectively.
- Wavelet filters: Daubechies, Symmlets, Coiflets – different widths
- 1000 simulations

Simulation Study

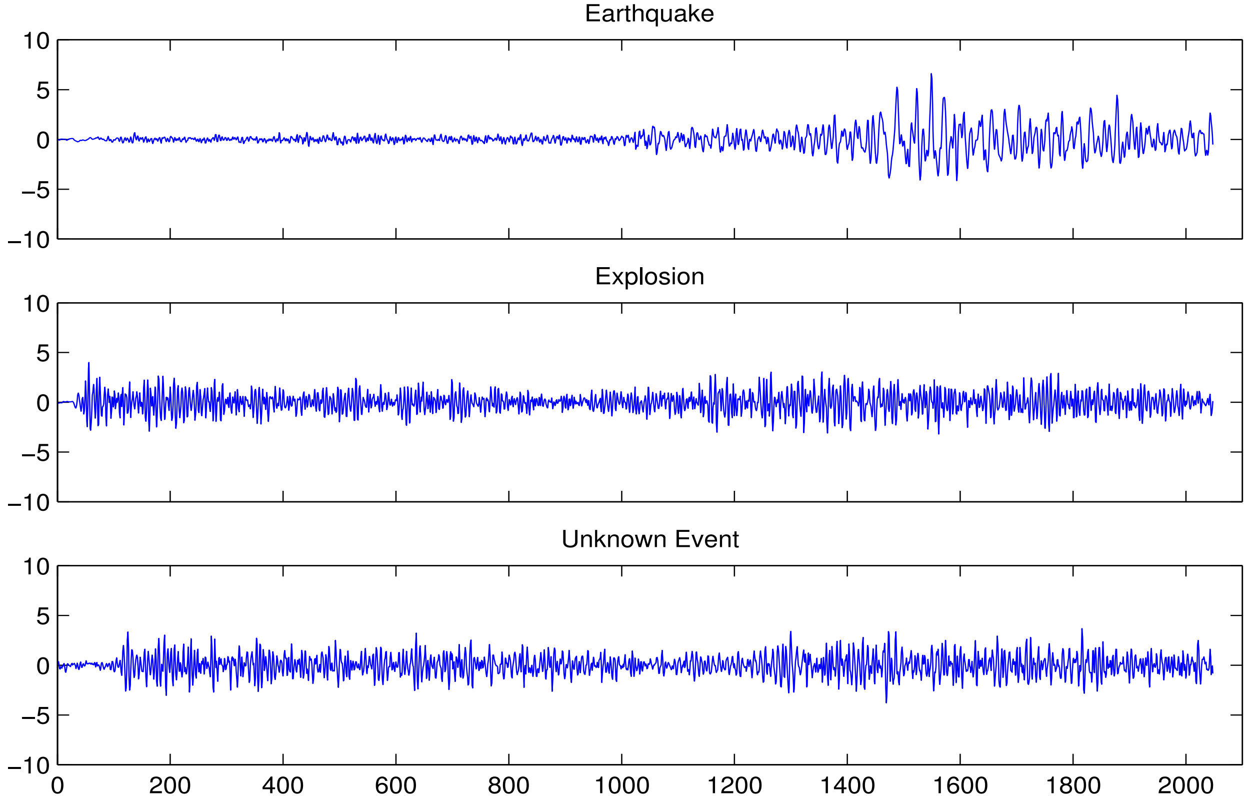


Simulation Study

Misclassification Rates:

		-0.5	-0.3	-0.1	0.1	0.3	0.5
T=256	Training	0.00	0.01	0.14	0.14	0.01	0.00
	Hold-out	0.00	0.04	0.33	0.32	0.04	0.00
T=1024	Training	0.00	0.00	0.04	0.03	0.00	0.00
	Hold-out	0.00	0.00	0.11	0.12	0.00	0.00
T=2048	Training	0.00	0.00	0.01	0.01	0.00	0.00
	Hold-out	0.00	0.00	0.04	0.04	0.01	0.00

Application - Earthquakes and Explosions



Application - Earthquakes and Explosions

Classification Results for Holdout one Procedure		
Predicted Patterns	**Real Patterns**	
	Earthquake	**Explosion**
Earthquake	4	4
Explosion	2	6

- First, we apply the procedure to the complete series assuming stationarity.
- Misclassification rate: 6/16. Very poor performance.

Application - Earthquakes and Explosions

Classification Results for Holdout one Procedure		
Predicted Patterns	Real Patterns	
	Earthquake	Explosion
Earthquake	8	1
Explosion	0	7

- Misclassification rate: 1/16 – one explosion classified as an earthquake
- Unknown event classified as an explosion
  - Consistent with observation from graphs
- Consistent with results obtained by Kakizawa et al. (1998), Shumway (2003) and Huang et al. (2004).

Application - Control Charts

- Pham and Chan (1998) use self organizing neural networks to discriminate among the different patterns
  - They presented ten different networks
  - Their misclassification rates were
    - Training sample: between 4.6% and 37.4%
    - Holdout sample: between 4.9% and 37.9%

- Using the wavelet variances in discriminant analysis our misclassification rates are
  - Training sample: 2.6%.
  - Holdout sample: 3%.

Application - Control Charts

Classification Results for Holdout one Procedure						
Predicted Patterns	**Real Patterns**					
	N	**C**	**IT**	**DT**	**US**	**DS**
N	100	0	0	0	0	1
C	0	100	0	0	0	0
IT	0	0	95	0	5	0
DT	0	0	0	96	0	4
US	0	0	5	0	95	0
DS	0	0	0	4	0	95

- The misclassifications can be easily explained
  - 5 increasing trend patterns classified as upward shift
  - 4 decreasing trend patterns classified as downward shift
  - 5 upward shift patterns classified as increasing trend
  - 4 downward shift patterns classified as decreasing trend

Further Research

- To develop a procedure to select the number and the length of stationary blocks.
- Extension to multivariate time series using the wavelets variance-covariance matrices.
- Application in medical data: EEG (20 channels) and ECG (12-15 channels).