Robust Functional Classification for Time Series

Andrés M. Alonso[1], David Casado[1], Sara López-Pintado[2] and Juan Romo[1]

[1] Universidad Carlos III de Madrid — 28903 Getafe (Madrid), Spain

[2] Columbia University — New York, NY 10032, USA

ERCIM, United Kingdom, 2011

Outline

- Time Series Classification
  - Introduction
  - The Method
  - Robustness
  - Results
- Further Work
- Conclusions

Time Series Classification

Introduction

- Time series can be studied from both **time and frequency domains**.

Introduction

- Time series can be studied from both **time and frequency domains**.
  - Short stationary series $\longrightarrow$ the usual multivariate techniques could be applied.

Introduction

- Time series can be studied from both **time and frequency domains**.
  - Short stationary series ⟶ the usual multivariate techniques could be applied.
  - Long stationary series ⟶ a frequency domain approach is more appropriate.

Introduction

- Time series can be studied from both **time and frequency domains**.
    - Short stationary series $\longrightarrow$ the usual multivariate techniques could be applied.
    - Long stationary series $\longrightarrow$ a frequency domain approach is more appropriate.
    - Nonstationary series $\longrightarrow$ the frequency domain is essential.

Introduction

- Time series can be studied from both **time and frequency domains**.
  - Short stationary series ⟶ the usual multivariate techniques could be applied.
  - Long stationary series ⟶ a frequency domain approach is more appropriate.
  - Nonstationary series ⟶ the frequency domain is essential.
- There are several works on classification methods from both domains.

Introduction

- Time series can be studied from both **time and frequency domains**.
  - Short stationary series $\longrightarrow$ the usual multivariate techniques could be applied.
  - Long stationary series $\longrightarrow$ a frequency domain approach is more appropriate.
  - Nonstationary series $\longrightarrow$ the frequency domain is essential.
- There are several works on classification methods from both domains.
- Most authors have studied the classification of stationary time series.

Introduction

Classification of Stationary Series

- Pulli (1996) considers the ratio of spectra.
- Kakizawa, Shumway and Taniguchi (1998) discriminate multivariate time series with the Kullback-Leibler's and the Chernoff's information measures.

Classification of Nonstationary Series

- Ombao et al. (2001) introduce the SLEX spectrum for a nonstationary random process.
- Caiado et al. (2006): define a measure, based on the periodogram, for both clustering and classifying stationary and nonstationary time series.

Introduction

Models for Nonstationary Series

- Priestley (1965) introduces the concept of a Cramér representation with time-varying transfer function.

$$X_t = \int_{-\pi}^{+\pi} e^{i\lambda t} A_t(\lambda) d\xi(\lambda)$$

- Dahlhaus (1996) establishes an asymptotic framework for locally stationary processes.

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{+\pi} e^{i\lambda t} A^0_{t,T}(\lambda) d\xi(\lambda)$$

Context

- The Problem. **Classification of time series**: (x_t) is a new series we want to classify in one of K different populations.

Context

- The Problem. **Classification of time series**: (x_t) is a new series we want to classify in one of K different populations.
- The Way. We transform the time series problem into a functional data question.

Context

- The Problem. **Classification of time series**: (x_t) is a new series we want to classify in one of K different populations.
- The Way. We transform the time series problem into a functional data question.
- The Tools

  Functional Data: Each element is a real function $\chi(t)$, $t \in I \subset \mathbb{R}$.

  Depth: The "centrality" or "outlyingness" of an observation within a set of data. It provides a criterion to order data from center-outward.

Functional Data

Let $(x_t) = (x_1, \ldots, x_T)$ be a time series, the **periodogram** and its cumulative version, the **integrated periodogram**, are:

$$I_T(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=+1}^{T} x_t e^{-it\lambda_j} \right|^2, \quad \lambda_j \in \mathcal{S}$$

$$F_T(\lambda_j) = \frac{1}{c_T} \sum_{i=1}^{j} I_T(\lambda_i), \qquad \lambda_i \in \mathcal{S}, \quad \lambda_j \in \mathcal{S}$$

where

$$\mathcal{S} = \left\{ \lambda_j = \frac{2\pi j}{T}, \ j = -\left[\frac{T-1}{2}\right], \ldots, -1, 0, +1, \ldots, +\left[\frac{T}{2}\right] \right\}$$

is the *Fourier set of frequencies*.

Classification Criterion

A new function χ is assigned to the group minimizing its distance to a reference function $\mathcal{R}$ of the group.

- The reference function: The **mean**

$$\mathcal{R}^{(k)}(t) = \bar{\chi}^{(k)}(t) = \frac{1}{N}\sum_{e=1}^{N}\chi_e^{(k)}(t)$$

- The distance: The distance

$$d(\chi_1,\chi_2) = \int_I |\chi_1(t) - \chi_2(t)|dt, \quad \chi_k \in \mathcal{L}^1(I), \; k = 1,2$$

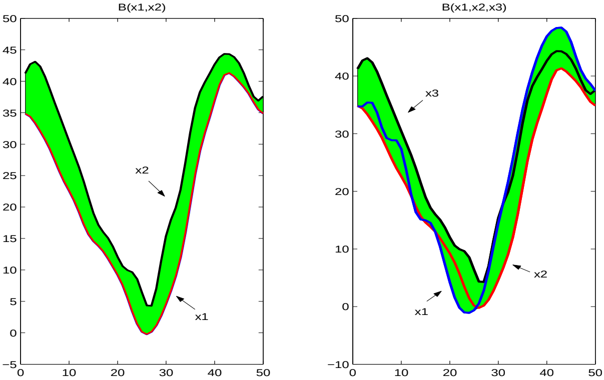
Remark: Our functional data belong to $\mathcal{L}^1(I)$

Depth

Let $G(\chi(t)) = \{(t, \chi(t)) : t \in [a, b]\}$ denote the graph of χ in $\mathbb{R}^2$, and let

$$B(\chi_{i_1},\ldots,\chi_{i_k})=\{(t,y) \mid t\in[a,b],\ \min_{r=1,\ldots,k} \chi_{i_r}(t)\leq y\leq\max_{r=1,\ldots,k} \chi_{i_r}(t)\}$$

be the band determined by k functions.



The **Band Depth**

- The proportion of bands containing the graph of χ is

$$BD_N^{(j)}(\chi(t)) = \binom{N}{j}^{-1} \sum_{1 \leq e_1 < e_2 < \ldots < e_j \leq N} \mathbb{I}\{G(\chi(t)) \subset B(\chi_{e_1}(t), \ldots, \chi_{e_j}(t))\}$$

The **Band Depth**

- The proportion of bands containing the graph of χ is

$$BD_N^{(j)}(\chi(t)) = \binom{N}{j}^{-1} \sum_{1 \le e_1 < e_2 < \ldots < e_j \le N} \mathbb{I}\{G(\chi(t)) \subset B(\chi_{e_1}(t), \ldots, \chi_{e_j}(t))\}$$

- This depth is defined, for $2 \le J \le N$, as

$$BD_{N,J}(\chi(t)) = \sum_{j=2}^{J} BD_N^{(j)}(\chi(t))$$

The **Band Depth**

- The proportion of bands containing the graph of χ is

$$BD_N^{(j)}(\chi(t))=\binom{N}{j}^{-1}\sum_{1\leq e_1<e_2<\ldots<e_j\leq N}\mathbb{I}\{G(\chi(t))\subset B(\chi_{e_1}(t),\ldots,\chi_{e_j}(t))\}$$

- This depth is defined, for $2\leq J\leq N$, as

$$BD_{N,J}(\chi(t))=\sum_{j=2}^{J}BD_N^{(j)}(\chi(t))$$

Population versions:

$$BD^{(j)}(\mathcal{X})=P\{G(\mathcal{X})\subset B(\mathcal{X}_{e_1},\ldots,\mathcal{X}_{e_j})\}$$

$$BD_J(\mathcal{X})=\sum_{j=2}^{J}BD^{(j)}=\sum_{j=2}^{J}P\{G(\mathcal{X})\subset B(\mathcal{X}_{e_1},\ldots,\mathcal{X}_{e_j})\}$$

The **Modified Band Depth**

- By taking the Lebesgue measure —instead of $\mathbb{I}$— of
  $A(\chi;\chi_{i_1},\ldots,\chi_{i_j}) \equiv \{t \in [a,b] \mid \min_{r=i_1,\ldots,i_j} \chi_r(t) \le \chi(t) \le \max_{r=i_1,\ldots,i_j} \chi_r(t)\}$,

$$MBD_N^{(j)}(\chi(t)) = \binom{N}{j}^{-1} \sum_{1 \le e_1 < e_2 < \ldots < e_j \le N} \nu_r(A(\chi(t);\chi_{e_1}(t),\ldots,\chi_{e_j}(t)))$$

The **Modified Band Depth**

- By taking the Lebesgue measure —instead of $\mathbb{I}$— of

$$A(\chi;\chi_{i_1},\ldots,\chi_{i_j})\equiv\{t\in[a,b] \mid \min_{r=i_1,\ldots,i_j}\chi_r(t)\leq\chi(t)\leq\max_{r=i_1,\ldots,i_j}\chi_r(t)\},$$

$$MBD_N^{(j)}(\chi(t))=\binom{N}{j}^{-1}\sum_{1\leq e_1<e_2<\ldots<e_j\leq N}\nu_r(A(\chi(t);\chi_{e_1}(t),\ldots,\chi_{e_j}(t)))$$

- The modified (generalized) band depth is defined as

$$MBD_{N,J}(\chi(t))=\sum_{j=2}^{J}MBD_N^{(j)}(\chi(t))$$

The **Modified Band Depth**

- By taking the Lebesgue measure —instead of $\mathbb{I}$— of
  $A(\chi;\chi_{i_1},\ldots,\chi_{i_j}) \equiv \{t \in [a,b] \mid \min_{r=i_1,\ldots,i_j} \chi_r(t) \leq \chi(t) \leq \max_{r=i_1,\ldots,i_j} \chi_r(t)\}$,

  $$MBD_N^{(j)}(\chi(t)) = \binom{N}{j}^{-1} \sum_{1 \leq e_1 < e_2 < \ldots < e_j \leq N} \nu_r(A(\chi(t);\chi_{e_1}(t),\ldots,\chi_{e_j}(t)))$$

- The modified (generalized) band depth is defined as

  $$MBD_{N,J}(\chi(t)) = \sum_{j=2}^{J} MBD_N^{(j)}(\chi(t))$$

Population versions:

$$MBD^{(j)}(\mathcal{X}) = \mathbb{E}\left(\nu_r(A(\mathcal{X};\mathcal{X}_{e_1},\ldots,\mathcal{X}_{e_j}))\right)$$

$$MBD_J(\mathcal{X}) = \sum_{j=2}^{J} MBD^{(j)}(\mathcal{X}) = \sum_{j=2}^{J} \mathbb{E}\left(\nu_r(A(\mathcal{X};\mathcal{X}_{e_1},\ldots,\mathcal{X}_{e_j}))\right)$$

Adding Robustness

1. Our method depends on the group reference curve.
2. The mean function of a set of functions is not robust.
3. Robustness can be added to the process through the reference curve.

$$\Downarrow$$

We shall consider the α-**trimmed mean**, where only the deepest elements are averaged:

$$\mathcal{R}^{(k)}(t) = \overset{\alpha}{\bar{\chi}}(t) = \frac{1}{n-[n\alpha]} \sum_{e=1}^{n-[n\alpha]} \chi_{(e)}(t)$$

with $[\cdot]$ being the integer part function.

Algorithms 1 and 2

Consider the samples $(x_t)_e^{(k)}, \; e = 1, \ldots, n_k$ for $k = 1, 2$.

Algorithms 1 and 2

Consider the samples $(x_t)_e^{(k)},\ e = 1, \ldots, n_k$ for $k = 1, 2$.

1. For each series $(x_t)_e^{(k)}$, if $F_{g,e}^{(k)}(\lambda)$ is the integrated periodogram of the g-th block, we **construct the function** $\chi_e^{(k)}(\lambda) = (F_{1,e}^{(k)}(\lambda) \ldots F_{G,e}^{(k)}(\lambda))$ so that the functional data are $\{\chi_e^{(k)}(\lambda)\},\ e = 1, \ldots, n_k$ for $k = 1, 2$

Algorithms 1 and 2

Consider the samples $(x_t)_e^{(k)}$, $e = 1, \ldots, n_k$ for $k = 1, 2$.

1. For each series $(x_t)_e^{(k)}$, if $F_{g,e}^{(k)}(\lambda)$ is the integrated periodogram of the g-th block, we **construct the function** $\chi_e^{(k)}(\lambda) = (F_{1,e}^{(k)}(\lambda) \ldots F_{G,e}^{(k)}(\lambda))$ so that the functional data are $\{\chi_e^{(k)}(\lambda)\}$, $e = 1, \ldots, n_k$ for $k = 1, 2$
2. For both populations **the group reference function** is calculated: $\mathcal{R}(\lambda)^{(k)} = \bar{\chi}^{(k)}(\lambda)$, in algorithm 1, or $\mathcal{R}(\lambda)^{(k)} = \overset{\alpha}{\bar{\chi}}{}^{(k)}(\lambda)$, in algorithm 2, $k = 1, 2$

Algorithms 1 and 2

Consider the samples $(x_t)_e^{(k)},\ e = 1, \ldots, n_k$ for $k = 1, 2$.

1. For each series $(x_t)_e^{(k)}$, if $F_{g,e}^{(k)}(\lambda)$ is the integrated periodogram of the g-th block, we **construct the function** $\chi_e^{(k)}(\lambda) = (F_{1,e}^{(k)}(\lambda) \ldots F_{G,e}^{(k)}(\lambda))$ so that the functional data are $\{\chi_e^{(k)}(\lambda)\},\ e = 1, \ldots, n_k$ for $k = 1, 2$
2. For both populations **the group reference function** is calculated: $\mathcal{R}(\lambda)^{(k)} = \bar{\chi}^{(k)}(\lambda)$, in algorithm 1, or $\mathcal{R}(\lambda)^{(k)} = \overset{\alpha}{\bar{\chi}}{}^{(k)}(\lambda)$, in algorithm 2, $k = 1, 2$
3. A new series (x_t) **is classified** in

$$\begin{cases} k = 1 & \text{if} \quad d(\chi(\lambda), \mathcal{R}(\lambda)^{(1)}) < d(\chi(\lambda), \mathcal{R}(\lambda)^{(2)}) \\ k = 2 & \text{otherwise} \end{cases}$$

The SLEXbC Method

Using the SLEX (**s**mooth **l**ocalized complex **ex**ponential) model of a nonstationary random process, Huang et al. (2004) propose a classification method (SLEXbC).

We compare our algorithms with this method.

How SLEXbC works

1. It finds a basis from the SLEX library that can best detect the differences.
2. It assigns to the class minimizing the Kullback-Leibler divergence between the SLEX spectra.

Simulations

Let $(x_t)_e^{(k)}$ be the e-th series of the k-th population; let $\epsilon_t \sim N(0,1)$ be Gaussian noise.

Simulation 1. Series are stationary

$$X_t^{(1)} = \phi X_{t-1}^{(1)} + \epsilon_t^{(1)} \qquad t = 1, \ldots, T$$

$$X_t^{(2)} = \epsilon_t^{(2)} \qquad t = 1, \ldots, T$$

Training data sets sizes: $n = 8$ series of length $T = 1024$.
Testing data sets sizes: $n = 10$ series of length $T = 1024$.
Six comparisons: Values $\phi = -0.5, -0.3, -0.1, +0.1, +0.3$ and $+0.5$.

Runs: 1000.

Simulations

Contamination 1

MA(ϕ) instead of AR(ϕ) (with the same parameter value).

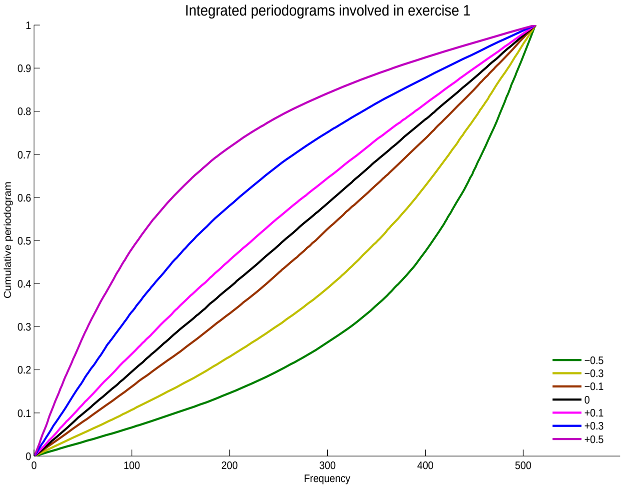
Contamination 2

$\phi = -0.9$ instead of the correct value ϕ (with the correct model).

Contamination 3

$\phi = +0.9$ instead of the correct value ϕ (with the correct model).

We always contaminate one series of the population $P^{(1)}$.


Integrated periodograms involved in exercise 1
Cumulative periodogram
Frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
−0.5
−0.3
−0.1
0
+0.1
+0.3
+0.5

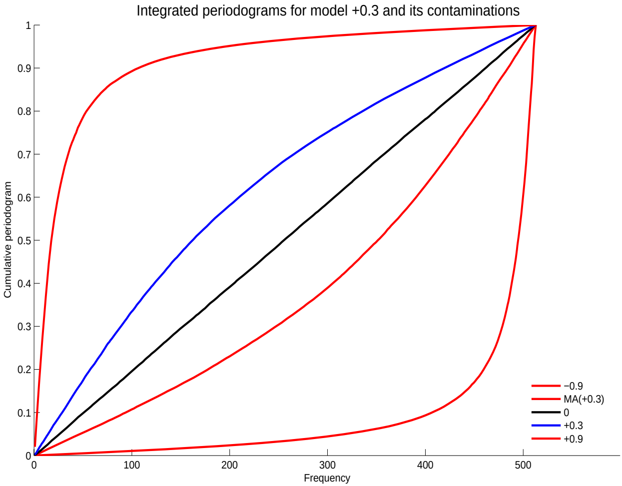
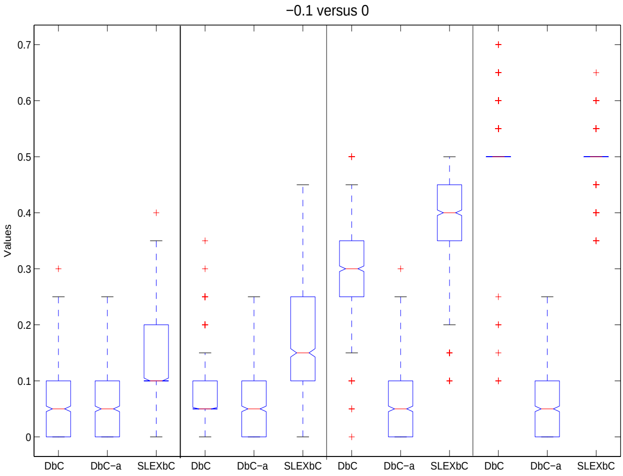

Integrated periodograms for model +0.3 and its contaminations
Cumulative periodogram
Frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
−0.9
MA(+0.3)
0
+0.3
+0.9

Table: Simulation 1 with and without contamination

	ϕ = **-0.3**	ϕ = **-0.1**	ϕ = **+0.1**	ϕ = **+0.3**
DbC	0.000	**0.063**	**0.060**	0.000
DbC-α	0.000	0.065	0.062	0.000
SLEXbC	0.000	0.131	0.127	0.000
DbC	0.000	0.077	0.074	0.000
DbC-α	0.000	**0.064**	**0.062**	0.000
SLEXbC	0.000	0.175	0.172	0.000
DbC	0.000	0.300	0.513	0.001
DbC-α	**0.000**	**0.065**	**0.062**	**0.000**
SLEXbC	0.001	0.377	0.491	0.002
DbC	0.001	0.512	0.300	0.000
DbC-α	**0.000**	**0.064**	**0.062**	**0.000**
SLEXbC	0.002	0.490	0.377	0.001


−0.1 versus 0
Values
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
DbC
DbC−a
SLEXbC
DbC
DbC−a
SLEXbC
DbC
DbC−a
SLEXbC
DbC
DbC−a
SLEXbC

Simulations

Let $(x_t)_e^{(k)}$ be the e-th series of the k-th population; let $\epsilon_t \sim N(0,1)$ be Gaussian noise.

Simulation 2. Series are made of stationary blocks

$$\begin{array}{ll} X_t^{(1)} = \epsilon_t^{(1)} & \text{if } t = 1, \dots, T/2 \\ X_t^{(1)} = -0.1 X_{t-1}^{(1)} + \epsilon_t^{(1)} & \text{if } t = T/2+1, \dots, T \end{array}$$

$$\begin{array}{ll} X_t^{(2)} = \epsilon_t^{(2)} & \text{if } t = 1, \dots, T/2 \\ X_t^{(2)} = +0.1 X_{t-1}^{(2)} + \epsilon_t^{(2)} & \text{if } t = T/2+1, \dots, T \end{array}$$

Training data sets sizes: $n = 8$ and 16; $T = 512$, 1024 and 2048.
Testing data sets sizes: $n = 10$; $T = 512$, 1024 and 2048.

Runs: 1000.

Simulations

Contamination 1

MA(ϕ) instead of AR(ϕ) (with the same parameter value).

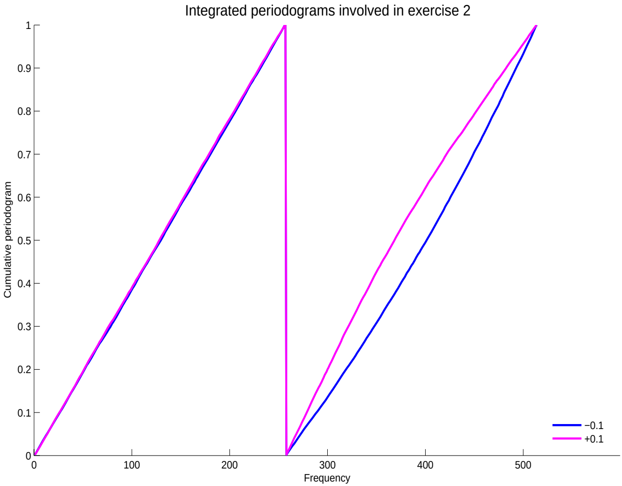
Contamination 2

$\phi = -0.9$ instead of the correct value ϕ (with the correct model).

Contamination 3

$\phi = +0.9$ instead of the correct value ϕ (with the correct model).

We always contaminate one series of the population $P^{(1)}$.


Integrated periodograms involved in exercise 2
Cumulative periodogram
Frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
−0.1
+0.1

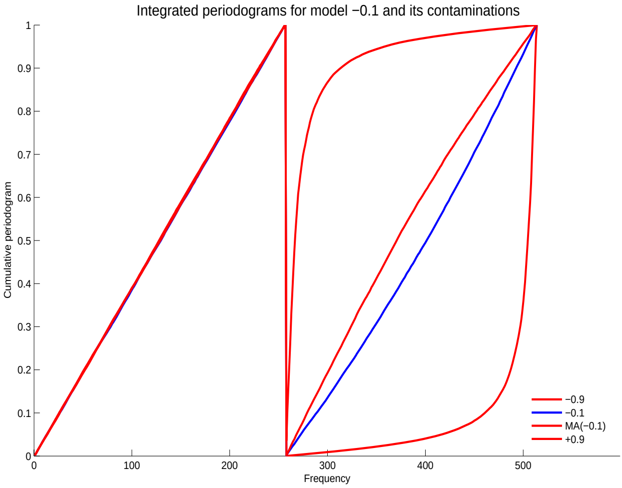
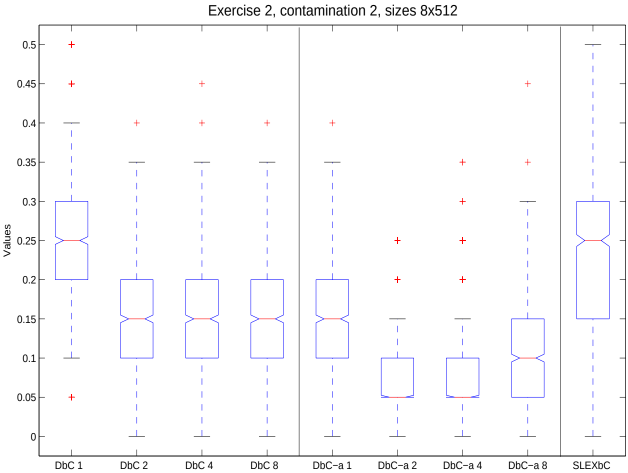

Integrated periodograms for model −0.1 and its contaminations
Cumulative periodogram
Frequency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
−0.9
−0.1
MA(−0.1)
+0.9

Table: Simulation 2 without contamination

		8x512	16x512	8x1024	16x1024	8x2048	16x2048
DbC	**1**	0.141	0.131	0.062	0.060	0.014	0.014
	2	**0.066**	**0.061**	**0.015**	**0.014**	**0.001**	**0.001**
	4	0.078	0.069	**0.015**	0.014	**0.001**	**0.001**
	8	0.090	0.080	0.020	0.018	0.002	**0.001**
DbC-α	**1**	0.143	0.132	0.063	0.061	0.015	0.014
	2	0.069	0.064	0.016	0.015	**0.001**	**0.001**
	4	0.083	0.073	0.017	0.016	0.002	**0.001**
	8	0.105	0.088	0.024	0.019	0.002	0.002
SLEXbC		0.114	0.086	0.038	0.025	0.007	0.003

Table: Simulation exercise 2, sizes 8x512

	Without contam.	Contam. 1	Contam. 2	Contam. 3
DbC 1	0.141	0.143	0.258	0.457
2	**0.066**	**0.070**	0.135	0.147
4	0.078	0.083	0.137	0.187
8	0.090	0.102	0.143	0.225
DbC-α 1	0.143	0.145	0.145	0.145
2	0.069	0.072	**0.070**	**0.073**
4	0.083	0.086	0.081	0.083
8	0.105	0.114	0.104	0.108
SLEXbC	0.114	0.128	0.239	0.376


Exercise 2, contamination 2, sizes 8x512
Values
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
DbC 1
DbC 2
DbC 4
DbC 8
DbC−a 1
DbC−a 2
DbC−a 4
DbC−a 8
SLEXbC

Simulations

Let $(x_t)_e^{(k)}$ be the e-th series of the k-th population; let $\epsilon_t \sim N(0,1)$ be Gaussian noise.

Simulation 3. Series are not stationary
If $a_{t;\tau} = 0.8 \cdot [1 - \tau \cos(\pi t/1024)]$, then

$$X_t^{(1)} = a_{t;0.5} X_{t-1}^{(1)} - 0.81 X_{t-2}^{(1)} + \epsilon_t^{(1)} \qquad t = 1, \dots, T$$

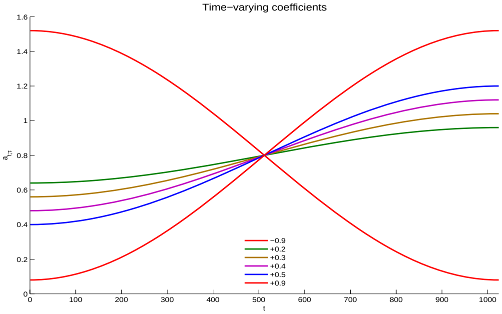
$$X_t^{(2)} = a_{t;\tau} X_{t-1}^{(2)} - 0.81 X_{t-2}^{(2)} + \epsilon_t^{(2)} \qquad t = 1, \dots, T$$

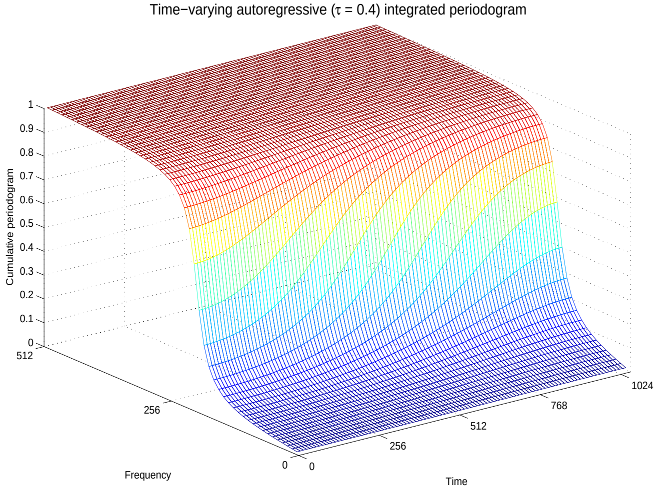
Training and testing data sets sizes: $n = 10$; $T = 1024$.
Three comparisons: τ values 0.4, 0.3 and 0.2.
Runs: 1000.

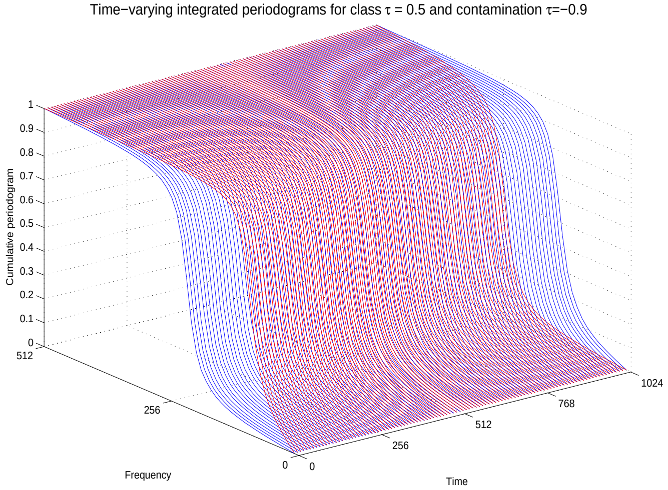
Simulations

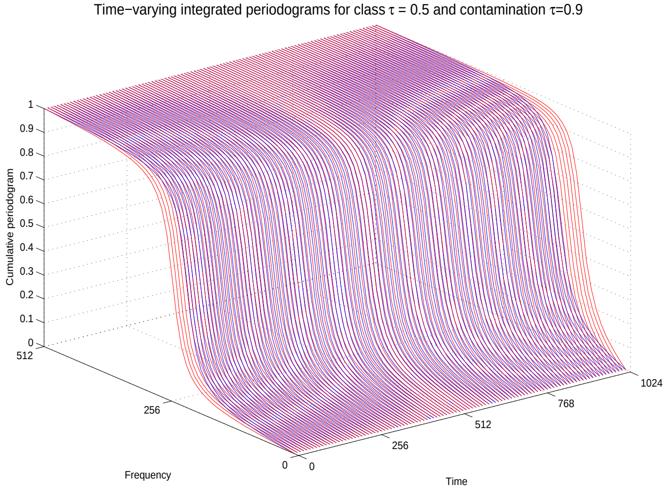
Contamination 1: $\tau = +0.2$ instead of $\tau = +0.5$.
Contamination 2: $\tau = -0.9$ instead of the correct value τ.
Contamination 3: $\tau = +0.9$ instead of the correct value τ.

We always contaminate one series of the population $P^{(1)}$.




Time−varying autoregressive (τ = 0.4) integrated periodogram
Cumulative periodogram
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Frequency
0
256
512
Time
0
256
512
768
1024


Time−varying integrated periodograms for class τ = 0.5 and contamination τ=−0.9
Cumulative periodogram
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
512
256
0
Frequency
0
256
512
768
1024
Time



Time−varying integrated periodograms for class τ = 0.5 and contamination τ=0.9
Cumulative periodogram
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
512
256
0
Frequency
0
256
512
768
1024
Time

Table: Simulation 3 without contamination

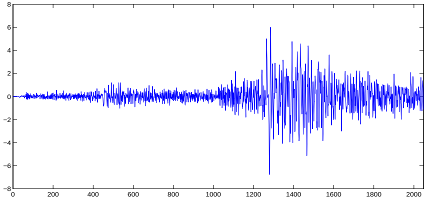
		$\tau = \mathbf{0.4}$	$\tau = \mathbf{0.3}$	$\tau = \mathbf{0.2}$
DbC	**1**	0.218	0.063	0.019
	2	0.119	0.006	**0.000**
	4	**0.101**	**0.002**	**0.000**
	8	0.123	0.003	**0.000**
DbC-α	**1**	0.226	0.065	0.021
	2	0.128	0.006	**0.000**
	4	0.112	**0.002**	**0.000**
	8	0.139	0.004	**0.000**
SLEXbC		0.181	0.011	**0.000**

Table: Simulation 3, $\tau = +0.4$

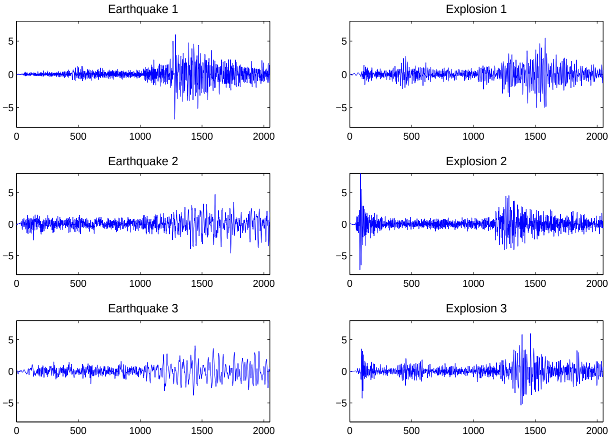
	Without contam.	Contam. 1	Contam. 2	Contam. 3
DbC 1	0.218	0.232	0.254	0.257
2	0.119	0.143	0.500	0.153
4	**0.101**	0.144	0.500	0.128
8	0.123	0.177	0.499	0.132
DbC-α 1	0.226	0.241	0.231	0.234
2	0.128	0.131	0.128	0.125
4	0.112	**0.121**	**0.113**	**0.114**
8	0.139	0.150	0.141	0.138
SLEXbC	0.181	0.234	0.492	0.173

Real Data

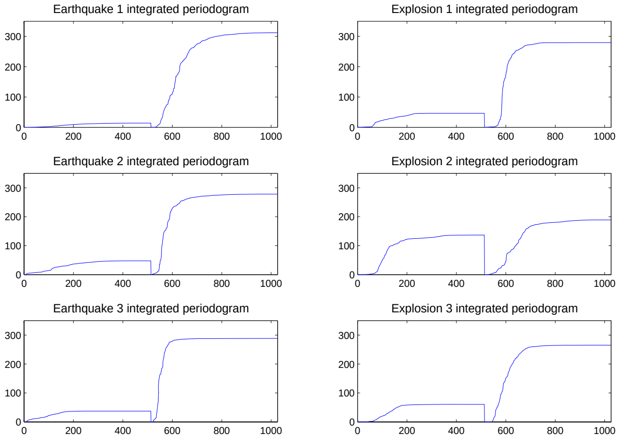
We have evaluated our proposal in a benchmark data set containing 8 explosions, 8 earthquakes and 1 extra series—known as NZ event—not classified (but being an earthquake or an explosion). Each series has two parts: the first half is the part P, and the second half is S.



Real data


Earthquake 1
Explosion 1
Earthquake 2
Explosion 2
Earthquake 3
Explosion 3
5
0
−5
0
500
1000
1500
2000

Real data


Earthquake 1 integrated periodogram
Explosion 1 integrated periodogram
Earthquake 2 integrated periodogram
Explosion 2 integrated periodogram
Earthquake 3 integrated periodogram
Explosion 3 integrated periodogram
0
100
200
300
0
200
400
600
800
1000

Real Data

Exercise 1

Group 1 = { 8 earthquakes }

Group 2 = { 8 explosions }

NZ event

Real Data

Exercise 1

Group 1 = { 8 earthquakes }

Group 2 = { 8 explosions }

NZ event

- Applying leave-one-out cross validation, **both of our algorithms misclassify only the first series of the group 2** (explosions).

Real Data

Exercise 1

Group 1 = { 8 earthquakes }

Group 2 = { 8 explosions }

NZ event

- Applying leave-one-out cross validation, **both of our algorithms misclassify only the first series of the group 2** (explosions).
- Respecting the NZ event, **both algorithms agree on assigning it to the explosions group**, as, for example, Kakizawa et al. (1998) and Huang et al. (2004).

Real Data

Exercise 2

Group 1 = { 8 earthquakes + NZ event }

Group 2 = { 8 explosions }

We can consider that a atypical observation is presented in group 1.

Real Data

Exercise 2

Group 1 = { 8 earthquakes + NZ event }

Group 2 = { 8 explosions }

We can consider that a atypical observation is presented in group 1.

- In this situation, **algorithm 1 misclassifies the first and third elements of group 2** (explosions), not only the first.

Real Data

Exercise 2

Group 1 = { 8 earthquakes + NZ event }

Group 2 = { 8 explosions }

We can consider that a atypical observation is presented in group 1.

- In this situation, **algorithm 1 misclassifies the first and third elements of group 2** (explosions), not only the first.
- But **again algorithm 2 misclassifies only the first series of group 2**. This illustrates the robustness of our second algorithm.

Real Data

Results

Table: Misclassified series

	Exercise 1	**Exercise 2**
DbC	Explosion 1	Explosions 1 and 3
DbC-α	Explosion 1	Explosions 1

Further Work

Time Series Method

1. K-group classification can be dealt with

$$k = argmin_{\{1,\dots,K\}} \left\{ d(\chi(\lambda), \mathcal{R}^{(k)}(\lambda)) \right\}.$$

Time Series Method

1. K-group classification can be dealt with

$$k = argmin_{\{1,\ldots,K\}} \left\{ d(\chi(\lambda), \mathcal{R}^{(k)}(\lambda)) \right\}.$$

2. Clustering of time series, by tackling the associated functional data problem in the frequency domain.

Time Series Method

1. K-group classification can be dealt with

$$k = argmin_{\{1,\ldots,K\}} \left\{ d(\chi(\lambda), \mathcal{R}^{(k)}(\lambda)) \right\}.$$

2. Clustering of time series, by tackling the associated functional data problem in the frequency domain.

3. Other different definitions of depth can be considered, for example: Fraiman and Muniz (2001), Cuevas et al. (2007).

Conclusions

Time Series Method

- We define a new time series classification method based on the **integrated periodogram**.

Time Series Method

- We define a new time series classification method based on the **integrated periodogram**.
- The method can also work with **nonstationary series** by splitting them into blocks and computing the integrated periodogram of each block.

Time Series Method

- We define a new time series classification method based on the **integrated periodogram**.
- The method can also work with **nonstationary series** by splitting them into blocks and computing the integrated periodogram of each block.
- By substituting the mean by the α-trimmed mean the method becomes **robust**.

Time Series Method

- We define a new time series classification method based on the **integrated periodogram**.
- The method can also work with **nonstationary series** by splitting them into blocks and computing the integrated periodogram of each block.
- By substituting the mean by the α-trimmed mean the method becomes **robust**.
- The method has shown **good behavior** in a wide range of simulation exercises and with real data, improving on existing methods.

Time Series Method

- We define a new time series classification method based on the **integrated periodogram**.
- The method can also work with **nonstationary series** by splitting them into blocks and computing the integrated periodogram of each block.
- By substituting the mean by the α-trimmed mean the method becomes **robust**.
- The method has shown **good behavior** in a wide range of simulation exercises and with real data, improving on existing methods.
- It suggests that the **integrated periodogram contains useful information to classify** time series.

References

Caiado, J., N. Crato and D. Peña (2006). A Periodogram-Based Metric for Time Series Classification. *Computational Statistics & Data Analysis*. 50, 2668–2684.

Dahlhaus, R. (1996). Asymptotic Statistical Inference for Nonstationary Processes with Evolutionary Spectra. *Athens Conference on Applied Probability and Time Series Analysis, Vol. 2 (P.M. Robinson and M. Rosenblatt, eds.). Lecture Notes in Statist. 115 145–159. Springer- -Verlag*. 6, 171–191.

Huang, H., H. Ombao and D.S. Stoffer (2004). Discrimination and Classification of Nonstationary Time Series Using the SLEX Model. *Journal of the American Statistical Association*. 99 (467), 763–774.

Kakizawa, Y., R.H. Shumway and M. Taniguchi (1998). Discrimination and Clustering for Multivariate Time Series. *Journal of the American Statistical Association*. 93 (441), 328–340.

References

López-Pintado, S., and J. Romo (2009). On the Concept of Depth for Functional Data. *Journal of the American Statistical Association*. 93 (441), 328–340.

Ombao, H.C., J.A. Raz, R. von Sachs and B.A. Malow (2001). Automatic Statistical Analysis of Bivariate Nonstationary Time Series. *Journal of the American Statistical Association*. 104 (486), 704–717.

Pulli, J. (1996). Extracting and processing signal parameters for regional seismic event identification, in Monitoring a Comprehensive Test Ban Treaty.

Priestley, M. (1965). Evolutionary Spectra and Non-Stationary Processes. *Journal of the Royal Statistical Society*. Series B, 27 (2), 204–237. *NATO Advanced Study Institute Series*. Vol. 303, Kluwer Press, 743–754 (eds E. Husebye and A. Dainty).