

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Exatas  
Programa de Pós-Graduação em Estatística

Jonathan de Souza Matias

**Discriminant Analysis in stationary time series based on robust cepstral  
coefficients**

Belo Horizonte  
2024

Jonathan de Souza Matias

**Discriminant Analysis in stationary time series based on robust cepstral coefficients**

**Versão Final**

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estatística.

Orientador: Valdério Anselmo Reisen

Belo Horizonte  
2024

Matias, Jonathan de Souza.

M433d Discriminant analysis in stationary time series based on robust cepstral coefficients [recurso eletrônico] / Jonathan de Souza Matias – 2024.  
1 recurso online (165 f. il, color.) : pdf.

Orientador: Valdério Anselmo Reisen.

Tese (Doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f.104-106

1. Estatística – Teses. 2. Análise de séries temporais – Teses. 3.– Análise discriminatória – Teses. 4. Fourier, Transformações de - Teses. 5. Seqüências espectrais (Matemática) - Teses I. Reisen, Valdério Anselmo. II.Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III.Título.

CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA



## FOLHA DE APROVAÇÃO

**Discriminant Analysis in stationary time series based on robust cepstral coefficients**

**JONATHAN DE SOUZA MATIAS**

Tese submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Doutor em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 26 de julho de 2024, pela banca constituída pelos membros:

Prof. Valdério Anselmo Reisen - Orientador

DEST/UFMG

Prof. Glauro da Conceição Franco  
DEST/UFMG

Prof. Pascal Thierry Bondon

Centrale Supélec

Prof. Marton Spanyol

University of Debrecen

Assinado por: Maria Helena Mouriño Silva Nunes  
Num. de Identificação: 08446895  
Data: 2024.07.29 23:26:51+01'00'

Profa. Maria Helena Mouriño Silva Nunes

Universidade de Lisboa

**Ta-Hsin Li**  
Digitally signed by Ta-Hsin Li  
DN: cn=Ta-Hsin Li, o=ou,  
email=tahsinli@outlook.com, c=US  
Date: 2024.07.29 09:16:40 -0400'

Prof. Ta-Hsin Li

IBM T. J. Watson Research Center

Belo Horizonte, 26 de julho de 2024.

*Primeiramente a Deus e a Jesus Cristo, que sempre me carregaram em todos os momentos e cuja minha fé é inabalável. Sem a força divina e a orientação espiritual, não teria conseguido superar os desafios desta jornada.*

*Ao meu pai, José Eustáquio Matias (in memoriam), e à minha mãe, Maria Fátima de Souza, que foram não só os que criaram meu genótipo, mas também moldaram minha personalidade, equilibrando razão e amor. Suas lições, carinho e apoio constantes foram fundamentais para que eu alcançasse este marco em minha vida.*

*À minha irmã Andressa, que se tornou uma grande amiga em um contexto tão complicado. Sua presença, apoio e amizade foram essenciais para me manter firme e motivado ao longo deste percurso.*

*Ao meu melhor amigo Alan, por sua amizade incondicional e apoio constante tanto na esfera pessoal quanto acadêmica. Sem ele, esta tese não teria atingido o nível que alcançou.*

*A vocês, minha eterna gratidão e amor.*

# Agradecimentos

Primeiramente, agradeço à CAPES, CNPq e FAPEMIG pelo suporte, bolsa, financiamento e todos os materiais fornecidos ao Departamento de Estatística no ICEX/UFMG, sem os quais esta pesquisa não seria possível.

Agradeço a todos os professores do Departamento de Estatística, cujo conhecimento e dedicação foram fundamentais para minha formação acadêmica.

Minha gratidão especial à coordenadora Rosângela, que não somente me deu suporte e compreensão como coordenadora, mas também como minha professora de Inferência Bayesiana, contribuindo significativamente para minha formação.

Ao meu orientador, Professor Valdério Anselmo Reisen, expressei meu profundo agradecimento pelo direcionamento, incentivo e suporte. Desde o início de sua orientação, o Professor Valdério sempre se dispôs a discutir ideias, me propor novos desafios e me orientar nos momentos mais desafiadores. Sua dedicação e sabedoria foram fundamentais para o desenvolvimento desta tese. Sua expertise e visão acadêmica foram essenciais para que eu pudesse alcançar este resultado. A sua orientação não só enriqueceu meu conhecimento científico, mas também moldou meu caráter como pesquisador. Sem sua orientação, esta tese não teria sido possível. Agradeço profundamente por acreditar em mim e por seus questionamentos, os quais me faziam sempre avançar um passo mais, proporcionando um ambiente de contínuo aprendizado e crescimento.

À professora Glauro, minha gratidão por ter sido a primeira pessoa com quem tive contato oficial com o Departamento de Estatística em 2016, enquanto coordenadora e professora da disciplina de Séries Temporais, fazendo com que eu tivesse meu primeiro contato com o domínio da frequência, objeto central desta tese.

Ao professor Dani Gamerman, meu profundo agradecimento por ser um orientador central no projeto covidLP. Como meu primeiro professor de Séries Temporais no departamento, sua orientação durante a pandemia de Covid-19 foi crucial. Sua sugestão de um desafio em sua disciplina resultou no projeto covidLP, que culminou em aprendizado significativo e em um livro do qual tive a honra de ser co-autor de três capítulos. Sua dedicação e visão foram essenciais para o sucesso desse projeto e para o avanço do meu conhecimento acadêmico.

À professora Thais Paiva, sou grato por ter sido orientadora em uma parte do projeto covidLP e por me incentivar no uso do aplicativo Shiny, resultando em um dos produtos desta tese.

Ao professor Marcus O. Prates, expressei minha sincera gratidão por seu papel

fundamental no projeto covidLP e por ter sido meu professor em Machine Learning. Seu suporte foi crucial para que eu pudesse utilizar os computadores do Leste para implementar as simulações necessárias para esta tese. Agradeço especialmente pelo auxílio e suporte para meu aprendizado, uso e acesso remoto aos computadores do Leste, o que foi vital para o processo de simulações. Sem essa ajuda, minhas simulações não teriam sido possíveis.

Ao professor Vinícius, expresso minha gratidão por sua orientação em partes importantes do projeto covidLP e pelo suporte no final do meu doutorado como coordenador da pós-graduação.

À Rogéria, agradeço pelo extremo suporte e auxílio desde o início do processo de seleção ao doutorado até à marcação da minha defesa.

Minha mãe, Fátima, merece um agradecimento especial. Mesmo sem entender completamente meu trabalho, sempre foi minha maior apoiadora e referência.

À minha irmã Andressa, agradeço por ser meu ouvido durante meus lamentos incessantes ao longo do doutorado e por ser minha companhia em vários momentos solitários.

Ao meu melhor amigo Alan, cuja amizade e suporte foram além das tentativas de explicação, agradeço por sua excepcional amizade e por seu suporte na revisão dos meus textos, artigos, qualificação e desta tese de doutorado. Sem ele, esta tese não teria atingido o nível que alcançou. Um amigo incondicional tanto na esfera pessoal quanto acadêmica, Alan esteve ao meu lado em cada passo desta jornada, oferecendo conselhos, apoio emocional e encorajamento constante.

Finalmente, agradeço aos meus colegas discentes, que sempre me deram suporte desde o início do doutorado até o fim, tornando esta jornada mais leve e enriquecedora.

A todos, meu mais sincero obrigado.

*“Logic will get you from A to B. Imagination will take you everywhere.”*  
(Albert Einstein)

# Resumo

A análise de séries temporais é crucial em áreas como finanças, economia, meio ambiente e engenharia biomédica para entender mecanismos internos, fazer previsões e identificar padrões. Métodos tradicionais no domínio do tempo, focados em tendências, sazonalidade e ruído, muitas vezes não consideram periodicidades e estruturas harmônicas, que são melhor capturadas no domínio da frequência. Analisar séries temporais nesse domínio permite identificar propriedades espectrais, oferecendo insights mais profundos sobre os comportamentos das séries temporais. Tais insights ajudam a diferenciar processos geradores de dados de diferentes populações, auxiliando na discriminação e classificação de séries temporais. A literatura geralmente utiliza estimadores suavizados, como o periodograma suavizado, para minimizar o viés, obtendo um espectro médio a partir de réplicas dentro de uma população. No entanto, variabilidade espectral dentro de cada população e valores abruptos podem comprometer esses métodos, dificultando a classificação prática. Há uma lacuna na literatura sobre métodos que considerem essa variabilidade, separem o efeito do ruído branco das autocorrelações e usem estimadores robustos na presença de valores extremos. Este artigo aborda essa lacuna, apresentando uma estrutura robusta para classificar grupos de réplicas de séries temporais, transformando-as no domínio da frequência usando a Transformada de Fourier para calcular o *power spectra*, o *log spectra* e o *cepstra*. Para calcular o *cepstra*, toma-se o logaritmo do espectro para separar a contribuição do ruído branco das autocorrelações, aplicando a Transformada de Fourier inversa ao espectro logarítmico. Para mitigar os efeitos de outliers e melhorar a consistência das estimativas espectrais, utilizamos o periodograma multitaper junto com o M-periodograma. Essas características espectrais são então usadas na Análise Discriminante Linear (LDA) para aumentar a precisão e a interpretabilidade da classificação. Esta abordagem integrada é promissora para aplicações que requerem distinção temporal precisa e resiliência a anomalias nos dados.

**Palavras-chave:** análise discriminante linear; séries temporais; domínio da frequência; robustez; cepstral.

# Abstract

Time series analysis is essential in fields such as finance, economics, environmental science, and biomedical engineering for understanding underlying mechanisms, forecasting, and identifying patterns. Traditional time-domain methods, which focus on trends, seasonality, and noise, often overlook periodicities and harmonic structures that are better captured in the frequency domain. Analyzing time series in the frequency domain enables the identification of these spectral properties, providing deeper insights into the underlying processes. These insights can help differentiate data-generating processes of different populations and assist in the discrimination and classification of time series. The literature commonly uses smoothed estimators like the smoothed periodogram to minimize bias, obtaining an average spectrum from individual replicates within a population to classify new time series. However, if there is spectral variability among replicates within each population, such methods become unfeasible. Moreover, abrupt values can significantly impact spectrum estimators, complicating practical discrimination and classification. There is a gap in the literature for methods that consider within-population spectral variability, separate white noise effects from autocorrelations, and use robust estimators in the presence of outliers. This paper addresses this gap by presenting a robust framework for classifying replicate groups of time series by transforming them into the frequency domain using the Fourier Transform to compute the power spectrum. Then, after taking the logarithm of the spectra, the inverse Fourier Transform is used to achieve the cepstrum. To mitigate the effects of outliers and improve the consistency of spectral estimates, we employ the multitaper periodogram alongside the M-periodogram. These spectral features are then utilized in Linear Discriminant Analysis (LDA) to enhance classification accuracy and interpretability. This integrated approach offers significant potential for applications requiring precise temporal pattern distinction and resilience to data anomalies.

**Keywords:** linear discriminant analysis; time series; frequency domain; robustness; outlier; cepstral.

# Lista de Figuras

2.1	Processo estocástico . . . . .	21
2.2	Domínio do tempo <i>versus</i> Domínio da Frequência . . . . .	25
2.3	Espectro Ruído Branco . . . . .	27
2.4	Espectro MA(1) com $\theta = 0.5$ . . . . .	28
2.5	Espectro AR(1) com $\phi = 0.5$ . . . . .	30
2.6	Espectro ARMA(1,1) com $\theta = 0.5$ $\phi = 0.5$ . . . . .	31
2.7	8 funis Seno de tamanho 120 observações . . . . .	37
2.8	Dados de ciclos de caminhada do paciente 4 com doença de Huntington e estes dados afunilados até o sétimo funil seno . . . . .	38
2.9	Redução do vazamento espectral usando os dados do periodograma multifunilado . . . . .	40
2.10	Decomposition of White Noise and autocorrelations of total variability of Spectral Density of a AR(1) process with $\phi = 0.5$ . . . . .	42
2.11	$\phi = 0.5$ e $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra . . . . .	44
2.12	$\phi = -0.5$ e $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra . . . . .	45
2.13	$\theta = 0.5$ e $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra . . . . .	46
2.14	$\theta = -0.5$ e $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra . . . . .	46
3.1	LDA projection of Cepstral. . . . .	54
A.1	séries temporais de tamanho $N = 100$ , $k = 3$ réplicas para $j = 1, 2, 3$ populações. Painéis (a) população 1; (b) população 2; (c) população 3 . . . . .	108
A.2	LDA . . . . .	111
A.3	Convergência modelo AR(1) com $\phi = 0.1$ e $N = 250$ . . . . .	113
A.4	Convergência modelo AR(1) com $\phi = 0.5$ e $N = 250$ . . . . .	113
A.5	Convergência modelo AR(1) com $\phi = 0.9$ e $N = 250$ . . . . .	114
A.6	Box-plot dos coeficientes cepstrais até $L = 7$ periodograma clássico versus robusto com $N = 250$ e 15 amostras . . . . .	114
A.7	Convergência modelo AR(1) com $\phi = 0.1$ e $N = 1000$ . . . . .	115
A.8	Convergência modelo AR(1) com $\phi = 0.5$ e $N = 1000$ . . . . .	116
A.9	Convergência modelo AR(1) com $\phi = 0.9$ e $N = 1000$ . . . . .	116
A.10	Box-plot dos coeficientes cepstrais até $L = 7$ periodograma clássico versus robusto com $N = 1000$ e 15 amostras . . . . .	117

A.11 $\phi \in [0.1, 0.5, 0.9]$ and $\sigma_{\epsilon}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra. . . . .	119
A.12 $\phi \in [-0.1, -0.5, -0.9]$ and $\sigma_{\epsilon}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra. . . . .	119
A.13 $\theta \in [0.1, 0.5, 0.9]$ and $\sigma_{\epsilon}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra. . . . .	121
A.14 $\theta \in [-0.1, -0.5, -0.9]$ and $\sigma_{\epsilon}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra. . . . .	121
D.1 Séries Temporais modificada por <a href="#">Krafty (2016)</a> dos ciclos de passada por segundo de 16 indivíduos saudáveis de controle . . . . .	140
D.2 Boxplot das Séries Temporais modificadas por <a href="#">Krafty (2016)</a> dos ciclos de passada por segundo de 16 indivíduos saudáveis de controle . . . . .	141
D.3 Série Temporal modificada por <a href="#">Krafty (2016)</a> dos ciclos de passada por segundo de 12 indivíduos com doença de Esclerose . . . . .	142
D.4 Boxplot das Séries Temporais modificadas por <a href="#">Krafty (2016)</a> dos ciclos de passada por segundo de 12 indivíduos com doença de Esclerose . . . . .	143
D.5 Gráfico e Boxplot da Série Temporal modificada por <a href="#">Krafty (2016)</a> de um indivíduo com doença de Esclerose . . . . .	144
D.6 Séries Temporais modificadas por <a href="#">Krafty (2016)</a> de ciclos de passada por segundo a cada segundo de 20 indivíduos de controle . . . . .	145
D.7 Boxplot das Séries Temporais modificadas por <a href="#">Krafty (2016)</a> de ciclos de passada por segundo de 20 indivíduos de controle . . . . .	146
D.8 Séries Temporais de ciclos de passada por segundo de 16 indivíduos de controle	147
D.9 Boxplot das Séries Temporais de ciclos de passada por segundo de 16 indivíduos de controle . . . . .	148
D.10 Séries Temporais de ciclos de passada por segundo de 12 indivíduos com doença de Esclerose . . . . .	149
D.11 Boxplot das Séries Temporais de ciclos de passada por segundo de 12 indivíduos com doença de Esclerose . . . . .	150
D.12 Série Temporal e Boxplot de ciclos de passada por segundo de 1 indivíduo com doença de Esclerose . . . . .	151
D.13 Séries Temporais de ciclos de passada por segundo de 20 indivíduos com doença de Huntington . . . . .	152
D.14 Boxplot das Séries Temporais de ciclos de passada por segundo de 20 indivíduos com doença de Huntington . . . . .	153
D.15 Cepstral dados modificados . . . . .	154
D.16 Cepstral dados não modificados . . . . .	155
D.17 Cepstral dados não modificados exceto cep_0 . . . . .	156

D.18 Funções dos 7 primeiros funis (tapers) . . . . .	157
D.19 Dados do Paciente Huntingons 4 e dados afunilados pelos 7 funis (tapers) . . .	158

# Lista de Tabelas

3.1	Matriz de Confusão Normalizada para $J$ Populações . . . . .	54
A.1	Distribuição dos Parâmetros Condicionais . . . . .	107
A.2	Cepstral Conditional Estimates for Each Population . . . . .	108
A.3	Matriz de médias por cepstra para cada população $\hat{\mu}_{j\ell}$ . . . . .	109
A.4	Eigenvalues and Eigenvectors of $(\hat{\Omega}_b^M)^{-1}\hat{\Omega}_b^M$ . . . . .	110
A.5	Discriminant coordinates for $q = 1, 2$ . . . . .	110
A.6	Comparison of classical and M-cepstral with $N=250$ . . . . .	111
A.7	Comparison of classical and M-cepstral with $N=1000$ . . . . .	114
A.8	Erros Quadráticos para $N=250$ . . . . .	117
A.9	Erros Quadráticos para $N=1000$ . . . . .	118
A.10	Distribuição dos Parâmetros Condicionais . . . . .	122
A.11	Simulações de Monte Carlo: taxas médias de classificação corretas . . . . .	124
E.1	Séries Temporais Simuladas $AR(2)$ . . . . .	159
E.2	Réplica, População e População Predita - doenças neurodegenerativas com dados alterados por <b>Krafty (2016)</b> . . . . .	160
E.3	Réplica, População e População Predita <b>robusta</b> - doenças neurodegenerativas	161
E.4	Matriz Confusion para as 49 séries de indivíduos com doenças neurodegenerativas	162
E.5	Matriz Confusion <b>robusta</b> para as 49 séries de indivíduos com doenças neurodegenerativas . . . . .	162

# Lista de Algoritmos

1	Simulation of Time Series Data . . . . .	126
2	Simulation with Additive Outlier . . . . .	127
3	AR(2) Simulation . . . . .	128
4	Sin-Cos Calculation for Time Series . . . . .	128
5	AR(2) Simulation with Additive Outlier . . . . .	129
6	Calculation of Classical Periodogram . . . . .	130
7	Calculation of M-Periodogram based on Huber . . . . .	132
8	Cepstral Coefficients Calculation with Multitaper Periodogram (Part 1) . . . . .	134
9	Cepstral Coefficients Calculation with Multitaper Periodogram (Part 2) . . . . .	135
10	M-Cepstral Calculation Based on M-Periodogram (Part 1) . . . . .	136
11	M-Cepstral Calculation Based on M-Periodogram (Part 2) . . . . .	137
12	Classification Using Classical FFT by Group . . . . .	138
13	Classification Using M-Periodogram . . . . .	138
14	Classification Mean Classic using FFT by Group . . . . .	139
15	Classification Mean Huber . . . . .	139

# Sumário

<b>1</b>	<b>Introdução</b>	<b>17</b>
1.1	Introdução . . . . .	17
<b>2</b>	<b>Séries Temporais</b>	<b>20</b>
2.1	Domínio do Tempo e Séries Temporais Estacionárias . . . . .	20
2.2	Domínio da Frequência e Análise Espectral . . . . .	23
2.3	Exemplos de Processos ARMA . . . . .	26
2.4	Estimação do Espectro . . . . .	32
2.4.1	O Periodograma . . . . .	32
2.4.2	O M-Periodograma . . . . .	35
2.4.3	Multitaper . . . . .	36
2.5	Análise Cepstral . . . . .	41
2.6	O Estimador M-cepstral . . . . .	47
<b>3</b>	<b>Modelos Lineares de Discriminação e Classificação</b>	<b>50</b>
3.1	Análise Discriminante Linear (LDA) baseado no subespaço M-Cepstral estimado . . . . .	50
3.2	Taxas de Classificação e a Matriz de Confusão . . . . .	54
<b>4</b>	<b>Paper: “Discriminant analysis based on robust cepstral coefficients”</b>	<b>56</b>
<b>5</b>	<b>Conclusões</b>	<b>101</b>
	<b>Referências</b>	<b>104</b>
	<b>Apêndice A Simulações Adicionais</b>	<b>107</b>
A.1	Da simulação da série até a Análise Discriminante . . . . .	107
A.2	Análise de Convergência empírica dos Cepstrals . . . . .	111
A.3	Example of $AR(1)$ and $MA(1)$ . . . . .	118
A.4	Simulações de Monte carlo . . . . .	122
<b>B</b>	<b>Protótipo do Aplicativo Shiny para a Aplicação</b>	<b>125</b>
<b>C</b>	<b>Documentação de Algoritmos e funções Genéricas no Python</b>	<b>126</b>
C.1	Função de simulação para modelos $ARMA(p,q)$ sem outlier . . . . .	126
C.2	Função de simulação para modelos $ARMA(p,q)$ com outlier . . . . .	127

C.3	Função que simula o modelo AR(2) sem outlier . . . . .	128
C.4	Função que monta uma base de dados de senos e cossenos pra estimar o M-periodograma . . . . .	128
C.5	Função que simula o modelo AR(2) com outlier . . . . .	129
C.6	Função que retorna o periodograma clássico calculado usando o Fast Fourier Transform - FFT . . . . .	130
C.7	Função que retorna o M-periodograma baseado na função Huber . . . . .	132
C.8	Função que retorna o classical cepstra Multitaper periodogram . . . . .	134
C.9	Algoritmo que retorna a taxa de classificação e a Matriz de Confusão para o caso clássico . . . . .	138
C.10	Algoritmo que retorna a taxa de classificação e a Matriz de Confusão para o caso M-periodograma . . . . .	138
C.11	Retorna a Taxa de Classificação Média e o erro padrão baseado no número de amostras - Samples para o periodograma clássico . . . . .	139
C.12	Retorna a Taxa de Classificação Média e o erro padrão baseado no número de amostras - Samples para o M-periodograma . . . . .	139
<b>D</b>	<b>Figuras Adicionais</b>	<b>140</b>
<b>E</b>	<b>Tabelas Adicionais</b>	<b>159</b>
<b>F</b>	<b>Apêndice</b>	<b>163</b>
F.1	Demonstração da fórmula do Periodograma . . . . .	163
F.2	Formato Matricial do cálculo do Periodograma . . . . .	164
F.3	Cepstra Matrix for $J$ populations and $L$ cepstra . . . . .	164
F.4	Matriz de Variância e Covariância entre as populações . . . . .	165

# Capítulo 1

## Introdução

### 1.1 Introdução

A análise de séries temporais é uma ferramenta essencial em diversos campos, incluindo finanças, economia, ciência ambiental e engenharia biomédica. O objetivo principal da análise de séries temporais é entender os mecanismos subjacentes que governam os dados observados, prever valores futuros e identificar padrões significativos. As abordagens tradicionais de análise de séries temporais operam predominantemente no domínio do tempo, focando na modelagem de tendências, sazonalidade e ruído por meio de técnicas como média móvel autorregressiva (ARMA), vetores autoregressivos (VAR) e modelos de correção de erro vetorial (VECM).

No entanto, esses métodos muitas vezes são insuficientes para descobrir as propriedades espectrais subjacentes dos dados. Analisar séries temporais no domínio da frequência proporciona uma perspectiva complementar, enfatizando periodicidades e estruturas harmônicas que não são facilmente discerníveis no domínio do tempo. Essa transformação é tipicamente alcançada através da Transformada de Fourier, que decompõe uma série temporal em seus componentes sinusoidais constituintes, revelando a distribuição de potência entre diferentes frequências.

A fundamentação da análise de séries temporais foi significativamente avançada por [Box e Jenkins \(1970\)](#); [Box et al. \(1994\)](#), que introduziram uma abordagem sistemática para a construção de modelos estacionários *ARMA*. Seu trabalho lançou as bases para desenvolvimentos subsequentes no campo, que se concentraram em melhorar a precisão dos modelos e a eficiência computacional [Brockwell e Davis \(1991\)](#); [Anderson \(1976\)](#); [Shumway e Stoffer \(2011\)](#). Avanços recentes também exploraram a integração de técnicas de aprendizado de máquina para aprimorar o desempenho preditivo e a adaptabilidade a estruturas de dados complexas [Alpaydin \(2014\)](#); [Lazzeri \(2020\)](#).

Para aprimorar ainda mais a compreensão das séries temporais, [Brillinger \(1981\)](#) e [Priestley \(1981\)](#) forneceram tratamentos abrangentes dos métodos do domínio da frequência, destacando sua aplicabilidade a uma ampla gama de dados de séries temporais. O espectro

de potência, obtido através da forma quadrática da Transformada de Fourier de um sinal, é central para essas análises, oferecendo insights sobre as frequências e periodicidades dominantes dentro dos dados. Técnicas do domínio da frequência têm sido particularmente eficazes em campos como processamento de sinal e ondas luminosas, onde a identificação de características espectrais é crucial. [Bloomfield \(2000\)](#) elaborou ainda mais sobre a utilidade da análise de Fourier para dados de séries temporais.

Construindo sobre essas técnicas espectrais, o log espectro — uma transformação logarítmica do espectro de potência — melhora a visualização das características espectrais decompostas em forma de soma, em vez de uma forma de produto resultante da forma quadrática do espectro. Essa abordagem utilizando o logaritmo tem sido amplamente usada no processamento de fala e reconhecimento de imagens [A. V. Oppenheim e Schafer \(2004\)](#) e análise de sinais geofísicos [Bogert \(1963\)](#); [Shumway \(1982\)](#); [Alagón \(1986\)](#); [Kakizawa et al. \(1998\)](#); [Shumway e Stoffer \(2011\)](#). O cepstrum, derivado da aplicação da Transformada de Fourier inversa no log espectro, decompõe ainda mais a variação da série temporal em contribuições de ruído branco e autocorrelação. Essa técnica tem se mostrado valiosa para identificar padrões de eco e deconvoluir sinais complexos [Schafer \(1969\)](#); [A. v. Oppenheim et al. \(1968\)](#).

Apesar dos avanços, observações outliers dentro de dados de séries temporais podem distorcer significativamente as estimativas espectrais. Periodogramas tradicionais podem ser sensíveis a esses outliers, levando a estimativas espectrais enviesadas ou inconsistentes, mesmo utilizando periodogramas suavizados ([Kakizawa et al., 1998](#); [G. Zhang e Taniguchi, 1992](#); [Z. Zhang e Chan, 2005](#)). Para mitigar esses efeitos, técnicas robustas de estimativa espectral, como o M-periodograma e o periodograma multitaper, foram desenvolvidas. [Katkovnik \(1998\)](#) introduziu o M-periodograma como um método robusto para reduzir a influência de outliers. [Reisen, Lévy-Leduc, et al. \(2020\)](#) forneceu uma visão geral dos estimadores espectrais robustos e séries temporais no contexto da abordagem de memória longa, destacando os benefícios dessas abordagens no tratamento de outliers e na melhoria da estimativa espectral. O periodograma multitaper, conforme descrito por [Thomson \(1982\)](#), utiliza múltiplos tapers ortogonais para produzir um espectro médio, reduzindo a variância ao diminuir o vazamento causado pelos lobos laterais e melhorando a precisão da estimativa espectral.

Para refinar ainda mais a classificação e análise de dados de séries temporais, a Análise Discriminante Linear (LDA), introduzida por [Fisher \(1936\)](#), é um método estatístico usado para classificação e redução de dimensionalidade. A LDA busca encontrar uma combinação linear de características que melhor separe várias classes. Quando aplicada a dados de séries temporais, a LDA pode aproveitar as características espectrais para melhorar o poder discriminativo. Estudos recentes [Shumway e Stoffer \(2011\)](#) demonstraram a eficácia da incorporação de características do domínio da frequência na LDA para melhorar a classificação de dados de séries temporais. [Krafty \(2016\)](#) abordou

especificamente o uso da LDA na presença de variabilidade espectral dentro dos grupos.

Além disso, métodos estatísticos robustos foram desenvolvidos para melhorar o desempenho da LDA, mas há uma lacuna na literatura em relação a métodos robustos na análise de séries temporais, discriminação e classificação. [Huber e Ronchetti \(2009\)](#) introduziu estimadores robustos que podem ser aplicados no contexto da análise de séries temporais para melhorar a confiabilidade das funções discriminantes. Trabalhos mais recentes de [Kutz e Brunton \(2019\)](#) refinaram ainda mais essas técnicas, tornando-as mais aplicáveis a dados de séries temporais de alta dimensão.

Neste contexto, este trabalho visa desenvolver uma estrutura robusta para a classificação de grupos de réplicas de séries temporais com base em suas características espectrais, utilizando o M-periodograma como um estimador robusto e levando em conta a variabilidade espectral dentro dos grupos. Integrando técnicas avançadas de análise espectral com análise discriminante, esta pesquisa oferece uma abordagem inovadora para a classificação de séries temporais. O uso de métodos robustos de estimativa espectral garante que a análise seja resiliente a outliers, melhorando assim a confiabilidade dos resultados. A metodologia proposta não só melhora a precisão da classificação, mas também proporciona uma compreensão mais profunda da variabilidade espectral inerente aos dados de séries temporais. Esta abordagem tem um potencial significativo para aplicações em vários domínios, onde entender e distinguir padrões temporais complexos é crucial.

#### Estrutura dos Capítulos

Seguindo esta introdução, o capítulo 2 apresenta uma introdução às séries temporais, discutindo a relação entre o domínio do tempo e da frequência. Discute a estimação do espectro usando o periodograma, periodograma multitaper, periodograma robusto e estimador cepstra. O capítulo 3 discute sobre classificadores lineares e foca na análise discriminante linear. O capítulo 4 é o artigo resultante de todo o trabalho da tese. Ao final, existem 6 apêndices, contendo simulações adicionais, protótipo do aplicativo shiny para aplicação, documentação dos algoritmos e funções elaboradas em Python, figuras adicionais, tabelas adicionais e demais demonstrações.

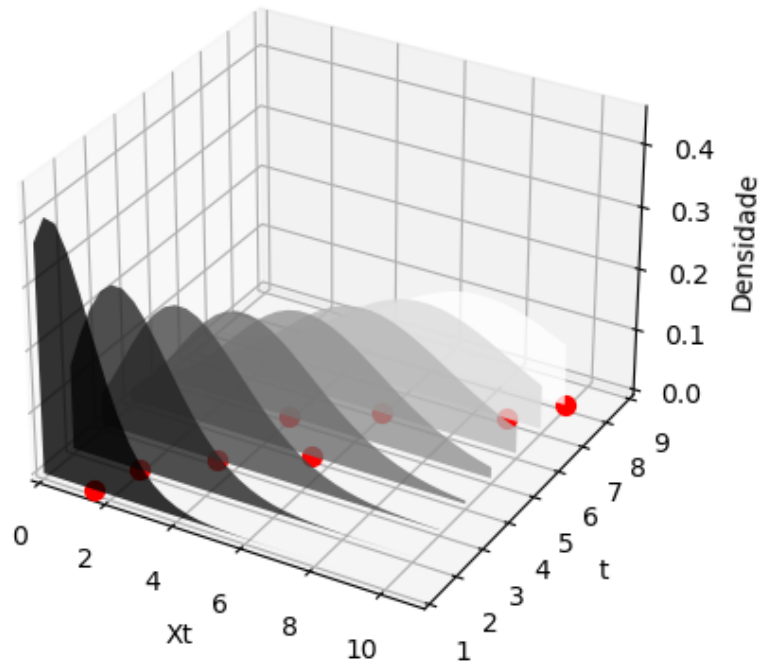
# Capítulo 2

## Séries Temporais

### 2.1 Domínio do Tempo e Séries Temporais Estacionárias

Em análise de séries temporais, a modelagem tem como foco tentar imitar o processo gerador de dados, tentando descobrir a dinâmica do sistema no qual a série foi gerada, permitindo assim simulação de novos valores de realização, bem como elaborar previsões de valores futuros. Para construir este modelo matemático, considere que  $X_t$  é uma variável aleatória indexada no tempo  $t \in Z$ , e tal variável aleatória tem suas realizações  $x_t$ . Então,  $X_t$  é chamado de processo estocástico que tem seus valores possíveis dependendo da distribuição de probabilidade, enquanto  $x_t$  é a "série temporal" ou realização do processo  $X_t$ . Este trabalho lida com séries temporais e este capítulo tem como objetivo fazer uma revisão básica dos principais conceitos e hipóteses em torno de séries temporais e processos estocásticos. Na figura 2.1 está apresentado um exemplo de processo estocástico, supondo que as densidades de cada variável aleatória  $X_t$  segue uma distribuição normal para cada  $t = 1, 2, \dots$ . Neste exemplo, a medida que  $t$  aumenta, tanto a média quanto as variâncias das distribuições aumentam. E para cada  $t$ , há uma realização do processo  $x_t$  marcados no gráfico em vermelho. Ao ligar todos esses pontos, tem-se o caminho de realização deste processo e é o que se chama de "série temporal" (Brockwell e Davis, 1991; Priestley, 1981).

Figura 2.1: Processo estocástico



Fonte: Elaborado pelo autor.

Nestes termos, para cada  $t$  existe uma variável aleatória  $X_t$  que tem sua distribuição marginal. E obviamente, esta variável pode ser entendida como um vetor multidimensional, representado por  $X = (X_1, \dots, X_N)' \in R^N$ . E nestes termos, terá uma densidade conjunta, sua matriz de variância e covariâncias, bem como o vetor do valor esperado, definido por  $E(X) = [E(X_1), \dots, E(X_N)]' \in R^N$ . Formalmente, pode-se definir uma série temporal de um processo estocástico como:

**(Processo Estocástico)** Um processo estocástico é uma família de variáveis aleatórias  $\{X_t, t \in T \subset Z\}$  definidas num espaço de probabilidade  $(\Omega, F, P)$ . (Brockwell e Davis, 1991, p.08)

Em que  $\Omega$  é o espaço amostral,  $F$  é a sigma-álgebra associada ao espaço amostral e  $P$  é a medida de probabilidade. Dado um espaço de probabilidade, a realização de uma série temporal,  $X_t$ , é conhecida como caminho amostral, no sentido que é um caminho de realização da variável aleatória assumindo um valor específico (Brockwell e Davis, 1991).

Uma vez realizada a série temporal, é esperado que esta tenha uma estrutura de correlação temporal que pode ser identificada pela função de autocovariância.

Para continuidade deste trabalho, algumas definições e observações se fazem necessárias, como será visto a seguir.

**(Função de Autocovariância)** Se um processo estocástico  $\{X_t, t \in T\}$  tem  $Var(X_t) < \infty$  para cada  $t \in N$ , a função de autocovariância  $\gamma_k$  pode ser definida como:

$$\gamma_k = E[X_t - E(X_t)][X_{t-k} - E(X_{t-k})] \quad (2.2)$$

for some  $k \in Z$ .

Neste trabalho, por hipótese os processos são assumidos ter média zero. Consequentemente, a função de autocovariância se torna apenas  $\gamma_k = E[X_t X_{t-k}]$ . Em particular, é o valor esperado do produto escalar entre dois vetores temporais <sup>1</sup>.

Finalmente, a estrutura de correlação da série temporal pode ser obtida pela divisão da equação (2.2) pela variância do processo  $\gamma_0$ , conhecida como Função de Autocorrelação <sup>2</sup>.

Outro conceito importante que será hipoteticamente utilizado neste trabalho é o conceito de processo fracamente estacionário, definido abaixo:

**(Processo Fracamente Estacionário)** Um processo estocástico é dito fracamente estacionário se:

- i.  $E(X_t) = \mu \forall t \in Z$
- ii.  $E(X_t - \mu)(X_t - \mu) = \gamma_j < \infty$
- iii.  $\gamma_k = E(X_t - E(X_t))(X_{t-k} - E(X_{t-k})) = E(X_{t+h} - E(X_{t+h}))(X_{t-k+h} - E(X_{t-k+h})) = \gamma_{k+h}$

for some  $k \in Z$ .

E se o processo é apenas um ruído com distribuição normal e média zero, é chamado de ruído branco:

**(Ruído Branco)** U processo  $X_t$  é dito ruído branco com média zero e variância  $\sigma^2$  quando pode ser escrito como:

$$X_t \sim RB(0, \sigma^2) \quad (2.3)$$

O processo estacionário mais comum é o processo *ARMA*, o qual também será utilizado por hipótese neste trabalho e definido como:

**(Processo ARMA estacionário)** Suponha que um processo  $X_t$  é fracamente estacionário. Então, ele é chamado de processo *ARMA*( $p, q$ ) se pode ser escrito como:

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \epsilon_t + \sum_{s=1}^q \theta_s \epsilon_{t-s} \quad (2.4)$$

em que  $\epsilon_t \sim RB(0, \sigma^2)$ ,  $p$  é a  $p$ -ésima ordem da parte autorregressiva e  $q$  é a  $q$ -ésima ordem da parte média móvel.

<sup>1</sup>Para mais sobre produto escalar ver [Hardle e Simar \(2007\)](#).

<sup>2</sup>Para mais, ver [Brockwell e Davis, 1991](#); [Hamilton, 1994](#).

Para linguagem, interpretação e análise deste trabalho, conceitos adicionais serão feitos neste contexto. Considere o corolário da proposição 3.2.1 de [Brockwell e Davis \(1991\)](#). O autor mostra que qualquer processo *ARMA* estacionário pode ser escrito em função do operador de defasagem usando uma variável  $z$  no lugar de  $L$ <sup>3</sup>, gerando uma razão de dois polinômios:

$$X_t = \frac{\mathcal{B}(z)}{\mathcal{A}(z)}\epsilon_t = \mathcal{T}(z)\epsilon_t \quad (2.5)$$

em que  $\mathcal{B}(z) = 1 - \phi_1 z - \dots - \phi_p z^p$  é um polinômio de grau  $p$  ( $p$ -ésima defasagem autorregressiva),  $\mathcal{A}(z) = 1 + \theta_1 z + \dots + \theta_q z^q$  é um polinômio de grau  $q$  ( $q$ -ésima defasagem média móvel),  $\mathcal{A}(z) \neq 0$ ,  $\epsilon_t \sim RB(0, \sigma^2)$ .

**(z-transform)** O uso de  $z$  no lugar de  $L$  é chamado na literatura de *z-transform* ([Hamilton, 1994](#); [Brockwell e Davis, 1991](#); ?). Trata-se de uma notação matemática, cuja importância se tornará evidente na análise do domínio da frequência, explicado na próxima seção.

**(Função de Transferência)** A função  $\mathcal{T}(z)$  na equação (2.5) é conhecida como função de transferência, estabelecendo a relação entre o ruído branco e o processo estocástico.

A função de transferência  $\mathcal{T}(z)$  tem sua importância pautada na relação de causalidade em processos estacionários *ARMA*, seguindo as diretrizes de da definição 3.1.3 de ([Brockwell e Davis, 1991](#), p.83). Por meio da função de transferência é possível avaliar, por exemplo, por quanto tempo um choque no distúrbio permanece ecoando no processo ao longo do tempo, fenômeno este conhecido na literatura como "Impulso-resposta". Ou seja, dada uma função de transferência definindo a dinâmica do sistema, pode-se encontrar qual é a resposta no processo estocástico dado um impulso (choque) no distúrbio presente<sup>4</sup>.

Na próxima subseção, serão apresentados os principais conceitos de domínio da frequência, bem como sua utilização no contexto deste trabalho.

## 2.2 Domínio da Frequência e Análise Espectral

Na última seção foram apresentados os conceitos básicos de séries temporais do domínio temporal com a variável  $X_t$  indexada pelo tempo e devido a este fato é conhecido na literatura como domínio temporal pois a série é uma função de  $t$ , cujo domínio é o

<sup>3</sup>O operador  $L$  vem do termo em inglês *lag operator*. Sua operação é simplesmente para gerar  $L^a X_t = X_{t-a}$ .

<sup>4</sup>Para mais ver [Hamilton \(1994\)](#).

tempo (Box e Jenkins, 1970; Anderson, 1976; Alagón, 1986; Brockwell e Davis, 1991; Hamilton, 1994). Neste contexto, a caracterização das séries temporais está pautada nas correlações da série com suas defasagens. Ou seja, investigar as representações das séries temporais em sua forma mais elementar (? , p.143).

Por outro lado, a mesma série temporal pode ser aproximada por séries de Fourier como uma soma ponderada de ondas periódicas para diferentes frequências. E neste caso, a série se torna função da frequência, de onde vem o termo *Domínio da Frequência* (Persival e Walden, 1993). Basicamente, as aproximações de Fourier são combinações lineares de funções senoidais e cossenoidais (harmônicas). Tais funções são periódicas no intervalo de  $0$  à  $2\pi$  (ou de  $-\pi$  à  $\pi$ ). Neste caso, supondo que  $\lambda_m = 2\pi m/z$  é a  $m$ -ésima frequência de Fourier <sup>5</sup> a série temporal  $X(\lambda_m)$  pode ser representada como uma soma de componentes harmônicos da seguinte forma:

$$X(\lambda_m) = \sum_{j=0}^{\infty} [A(\lambda_m)\cos(k\lambda_m) + B(\lambda_m)\sen(k\lambda_m)] \quad (2.6)$$

em que  $k = 1, 2, \dots$  e  $A(\lambda_m)$  e  $B(\lambda_m)$  são constantes representando a amplitude da onda para cada frequência.

Levando em consideração que  $\cos(\lambda_m) = \cos(-\lambda_m)$  e  $\sen(\lambda_m) = \sen(-\lambda_m) \forall m$ , pode-se mostrar que a equação (2.6) pode ser reescrita como <sup>6</sup>:

$$X(\lambda_m) = \sum_{m=-\infty}^{\infty} R(\lambda_m)e^{-i\lambda_m k} \quad (2.7)$$

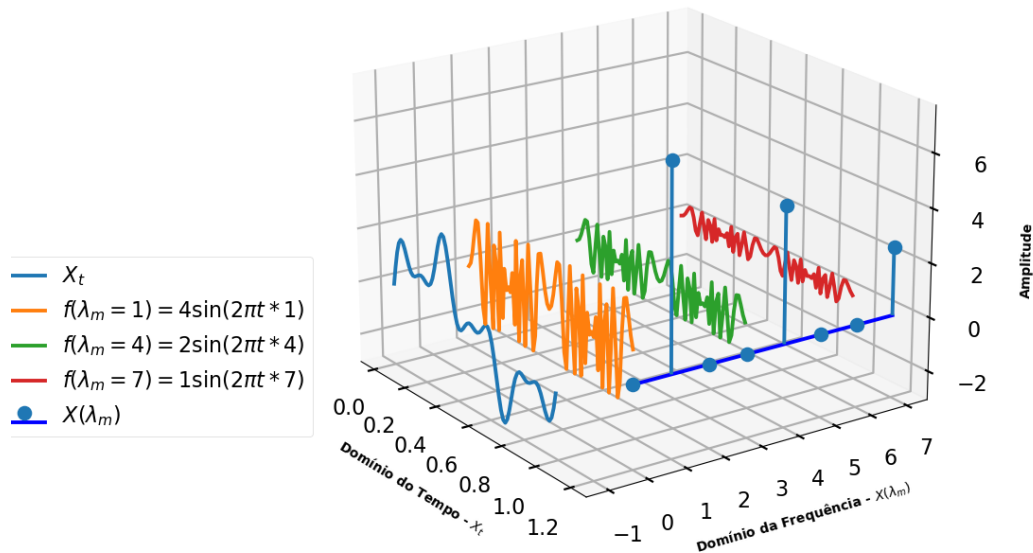
em que  $i = \sqrt{-1}$  e:

$$R(\lambda_m) = \begin{cases} \frac{A(\lambda_m)}{2} - i\frac{B(\lambda_m)}{2} & , se \quad m > 0 \\ \frac{A(\lambda_m)}{2} + i\frac{B(\lambda_m)}{2} & , se \quad m < 0 \end{cases} \quad (2.8)$$

Desta forma, as amplitudes e os termos harmônicos que geraram a série temporal fossem conhecido, bastaria escrevê-la como uma decomposição ortogonal de ondas harmônicas. Contudo, na prática o que se tem é a série temporal e o objetivo é descobrir quais foram as frequências que mais contribuíram para a geração da série. Para exemplificar, considere uma série temporal cujas frequências que contribuíram para sua geração foram primeira, quarta e sétima, com respectivas amplitudes de 4, 2 e 1, apresentadas na figura 2.2:

<sup>5</sup>Para mais ver (Persival e Walden, 1993, p.196).

<sup>6</sup>Basta usar Teorema de Euler Priestley (1981).

Figura 2.2: Domínio do tempo *versus* Domínio da Frequência

Fonte: Elaborado pelo autor.

No eixo temporal, a curva da série temporal é o resultado da soma de todas as ondas harmônicas. a frequência vai influenciar no grau de oscilação de cada onda, ou seja, quanto maior a frequência, maior a oscilação da onda. E quanto maior a amplitude, maior sua contribuição na soma para geração da série temporal. Esta amplitude é representada no eixo do domínio da frequência. Por meio desta figura fica clara a vantagem do domínio da frequência para este tipo de série temporal devida sua simplicidade em demonstrar a estrutura de composição de uma série temporal. Olhando para o domínio do tempo não é possível visualizar que a série temporal tem uma estrutura de combinação linear de harmônicos, enquanto tal estrutura é evidente no domínio da frequência.

A simplificação da série temporal no domínio da frequência se deve às características bem comportadas das funções senos e cossenos, tais como ortogonalidade em diferentes frequências. Esta ortogonalidade, por exemplo, irá levar às amplitudes não correlacionadas para cada frequência, tornando-se, assim, uma ferramenta alternativa extremamente importante na análise de processos estacionários.

Esta mudança de domínio do tempo para domínio da frequência é, basicamente, uma mudança de coordenadas de um espaço do temporal para o espaço das frequências usando as funções ortogonais de senos e cossenos como base do novo espaço vetorial para diferentes frequências. Para implementar esta mudança, considere a definição abaixo:

**(Função Geradora de Auto-covariâncias):** Seja  $z \in C$ , e suponha que as auto-covariâncias são absolutamente somáveis  $\sum_{\nu=-\infty}^{\infty} |\gamma_{\nu}| < \infty$ , então, a função geradora de auto-covariâncias pode ser definida como:

$$g_X(z) = \sum_{\nu=-\infty}^{\infty} \gamma_{\nu} z^{\nu} \quad (2.9)$$

Desta forma, considere um escalar  $\lambda$ ,  $z = e^{-i\lambda}$ <sup>7</sup>,  $\nu = 0, \pm 1, \pm 2, \dots, \pm(N-1)$   $i = \sqrt{-1}$  e dividindo a equação (2.9) por  $2\pi$ , tem-se a definição do espectro populacional de um processo:

**(Espectro Populacional)** Dadas as mesmas condições da definição 2.2 O espectro populacional é definido como<sup>8</sup> :

$$\begin{aligned} s_X(\lambda) &= \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} \gamma_{\nu} e^{-i\lambda\nu} \\ &= \frac{1}{2\pi} \left\{ \gamma_0 + 2 \sum_{\nu=0}^{\infty} \gamma_{\nu} \cos(\lambda\nu) \right\} \end{aligned} \quad (2.10)$$

A última parte da equação na definição 2.2 se deve à simetria das auto-covariâncias de processos estacionários, ou seja,  $\gamma_k = \gamma_{-k}$  (simetria em torno de zero) (Hamilton, 1994, p.153).

E assim, todo processo estocástico tem uma dinâmica do sistema que vai depender da função de ligação. E quando valorada em  $z = e^{-i\omega}$ , esta função é chamada de função de transferência.

Na próxima subseção serão apresentados vários processos estacionários *ARMA*, bem como suas funções polinomiais da z-transform, suas funções de ligação, suas funções de transferência e suas densidades espectrais populacionais.

## 2.3 Exemplos de Processos ARMA

Para todos os processos abaixo, considere que  $\epsilon_t \in RB(0, \sigma_{\epsilon}^2)$ . O exemplo 2.3 é o caso e mais simples, quando o próprio processo trata-se de um ruído branco puro. Note que as funções polinomiais e de ligação são iguais à 1. A função de transferência é igual à variância do distúrbio. Consequentemente, a densidade espectral é a variância do

<sup>7</sup>Note que o número  $z = e^{-i\lambda}$  é um número complexo particularmente importante devido o fato de estar no círculo unitário no plano complexo *RC*. Além disto, o sinal negativo indica uma rotação do vetor no plano complexo no sentido anti-horário.

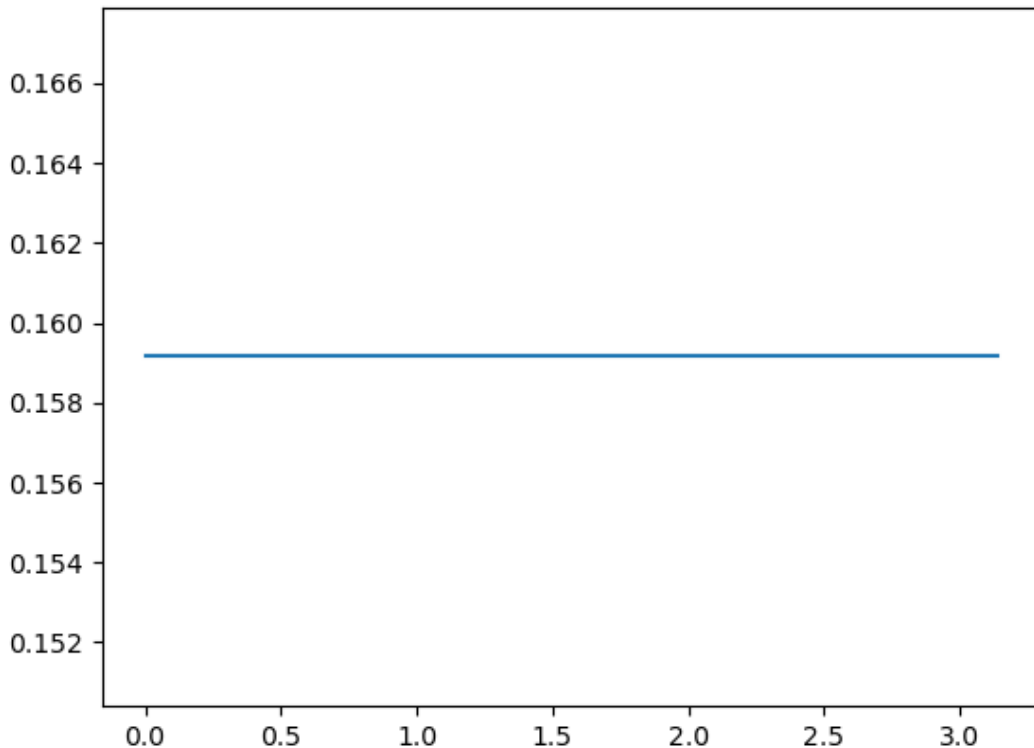
<sup>8</sup>Novamente, pelo Teorema de Euler,  $e^{-i\lambda\nu} = \cos(\lambda\nu) - i\text{sen}(\lambda\nu)$ . Para mais ver Priestley (1981).

distúrbio sobre  $2\pi$ , isto é, uma densidade espectral que não depende da frequência. Uma vez que a densidade espectral mede a contribuição de cada frequência para a variância do processo, percebe-se que todas as frequências contribuem igualmente para o processo. Não há uma frequência que contribua mais para o processo do que outra, é uma contribuição constante como pode ser observado na figura 2.7 com  $\sigma_\epsilon^2 = 1$ .

**(Ruído Branco)**  $X_t = \epsilon \sim RB(0, \sigma_\epsilon^2)$ , then:

- i. **MA(1):**  $X_t = \epsilon_t$
- ii. **Funções polinomiais:**  $A(L) = 1$  and  $B(L) = 1$
- iii. **Função de ligação:**  $\psi(z) = 1$
- iv. **Função de transferência:**  $g_X(z) = \sigma_\epsilon^2$
- v. **Densidade espectral:**  $s_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}$

Figura 2.3: Espectro Ruído Branco



Fonte: Elaborado pelo autor.

No exemplo 2.3, o processo trata-se de uma média móvel de primeira ordem representada por  $i$  com parâmetro  $\theta$ . Em particular, no caso do processo  $MA(1)$ , suas

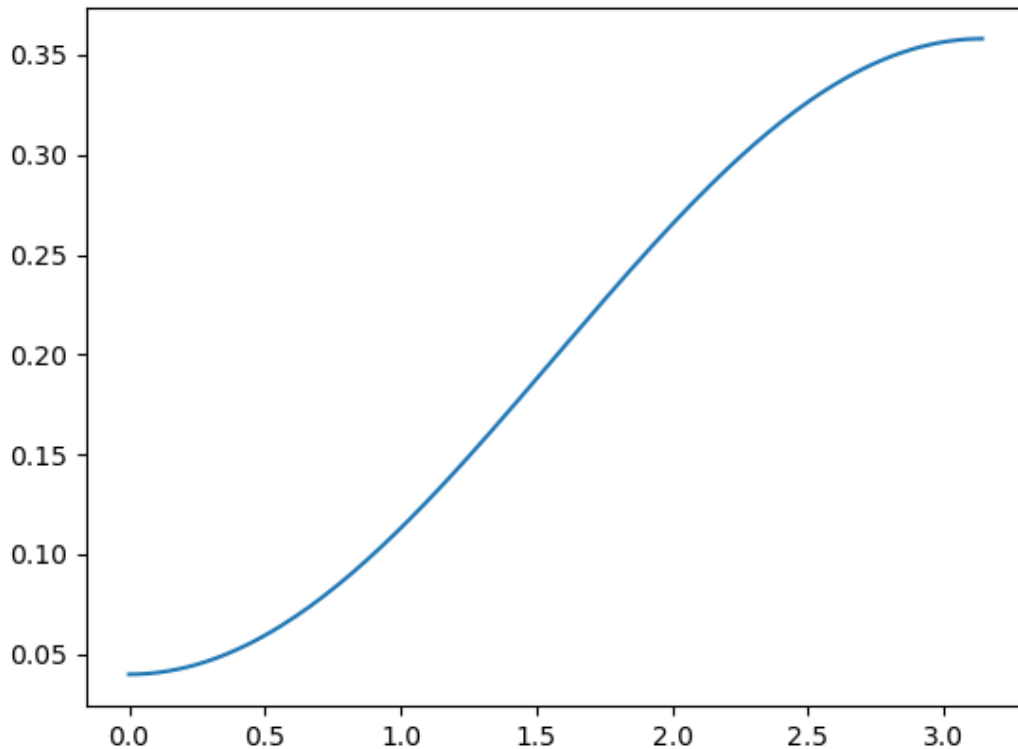
covariâncias são diferentes de zero apenas para defasagens de primeira ordem, o que faz da densidade espectral igual à  $s_X(\lambda) = (\gamma_{-1})e^{i\lambda} + \gamma_0 + (\gamma_1)e^{-i\lambda}$ , usando a equação 2.10, gerando o resultado apresentado em  $v$  neste exemplo.

Na figura 2.4 está apresentado o gráfico da densidade espectral para um  $MA(1)$  com  $\theta = 0.5$ . Note que neste caso, este processo é chamado "alta frequência" pois as frequências mais perto de  $\pi$  tem mais contribuição para a variância do processo do que as frequências próximas de zero.

Se  $X_t$  é um processo  $MA(1)$  e  $\epsilon \sim RB(0, \sigma_\epsilon^2)$ , então:

- i. **MA(1):**  $X_t = \epsilon_t + \theta\epsilon_{t-1}$
- ii. **Funções polinomiais:**  $A(L) = (1 + \theta L)$  and  $B(L) = 1$
- iii. **Função de ligação:**  $\psi(z) = 1 + \theta z$
- iv. **Função de transferências:**  $g_X(z) = \sigma^2\psi(z)\psi(z^{-1}) = \sigma^2(1 + \theta z)(1 + \theta z^{-1})$
- v. **Densidade espectral:**  $s_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}\psi(z)\psi(z^{-1}) = \frac{\sigma_\epsilon^2}{2\pi}[1 + \theta^2 + 2\theta\cos(\lambda)]$ , em que  $\lambda \in [0, \pi]$ .

Figura 2.4: Espectro  $MA(1)$  com  $\theta = 0.5$



Fonte: Elaborado pelo autor.

Utilizando  $\sigma_\epsilon^2 = 1$ ,  $\theta = 0.5$  e  $\lambda = 0$ , o espectro do processo  $MA(1)$  para a primeira frequência é igual à  $s_X(0) = \frac{1}{2\pi}(1 + 0.5^2 + 2*0.5) = 0.04$  que pode ser observado na figura 2.5 para a frequência zero. Isto quer dizer que a frequência zero contribui em 0.04 para a variância de todo o processo.

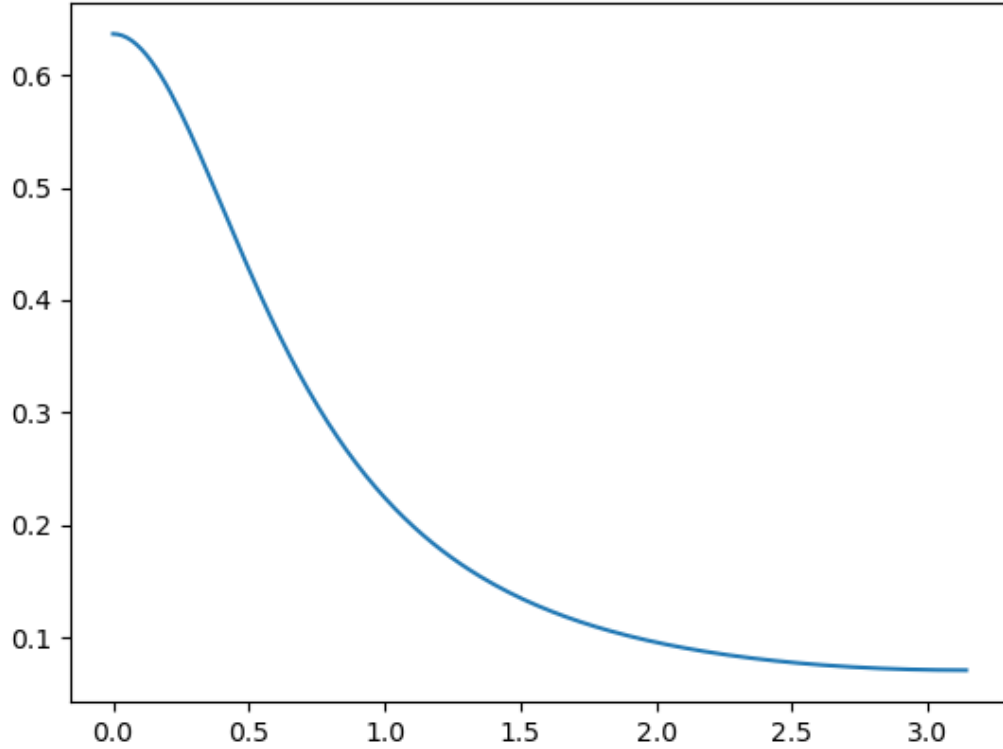
O exemplo 2.3 é um processo autorregressivo de primeira ordem e de maneira similar ao exemplo anterior, usando a equação 2.10 é possível mostrar que o resultado da densidade espectral de um processo  $AR(1)$  é igual à apresentada em  $v$ .

Na figura 2.5 é possível notar que a maior parte da variabilidade do processo é explicada pela variância do ruído branco reescalada por  $2\pi$  na frequência  $\lambda = 0$ .

Utilizando  $\sigma_\epsilon^2 = 1$ ,  $\phi = 0.5$  e  $\lambda = 0$ , o espectro do processo  $AR(1)$  para a primeira frequência é igual à  $s_X(0) = \frac{1}{2\pi} \frac{1}{1-2(0.5)+(0.5)^2} = 0.69$  que pode ser observado na figura 2.5 para a frequência zero. Isto quer dizer que a frequência zero contribui em 0.69 para a variância de todo o processo.

Se  $X_t$  é um processo  $AR(1)$  estacionário e  $\epsilon \sim RB(0, \sigma_\epsilon^2)$ , então:

- i. **AR(1):**  $X_t = \phi X_{t-1} + \epsilon_t$
- ii. **Funções polinomiais:**  $A(L) = 1$  and  $B(L) = (1 - \phi L)$
- iii. **Funções de ligação:**  $\psi(z) = \frac{1}{1+\phi z}$
- iv. **Função de transferência:**  $g_X(z) = \sigma^2 \psi(z) \psi(z^{-1}) = \sigma_\epsilon^2 \frac{1}{(1-\phi z)} \frac{1}{(1-\phi z^{-1})}$
- v. **Densidade espectral:**  $s_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \psi(z) \psi(z^{-1}) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1}{1+\phi^2-2\phi \cos(\lambda)}$

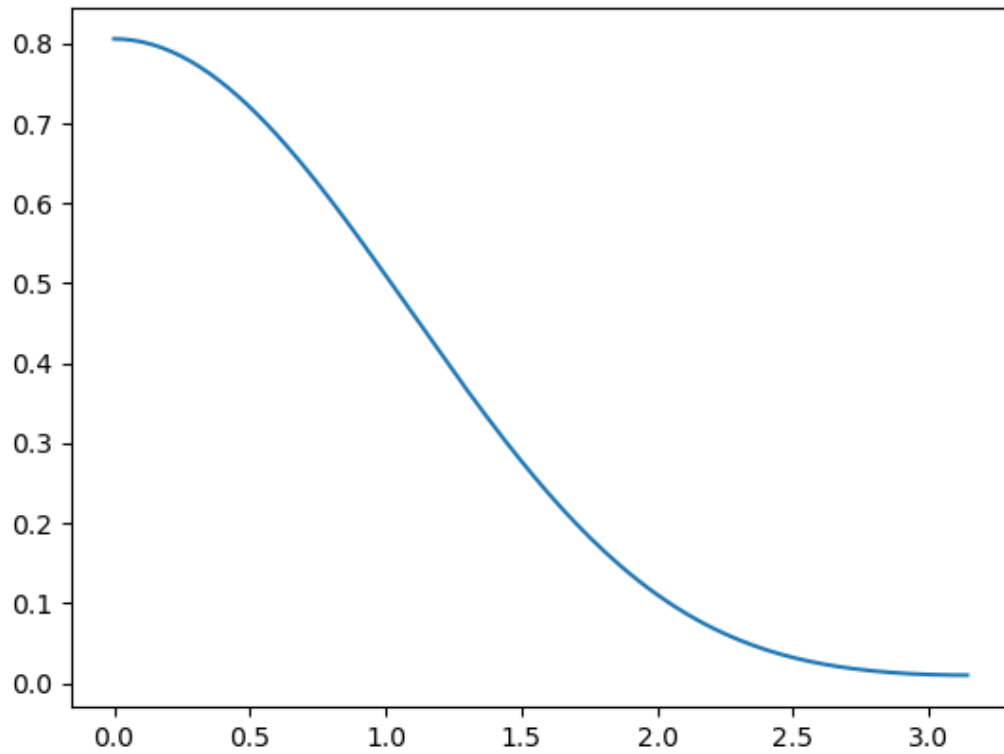
Figura 2.5: Espectro AR(1) com  $\phi = 0.5$ 

Fonte: Elaborado pelo autor.

O exemplo 2.3 é um processo  $ARMA(1, 1)$  que se trata de uma mistura do processo  $AR(1)$  com o processo  $MA(1)$ , o que gera uma densidade espectral misturada dos dois processos e cada frequência irá contribuirá mais para o processo dependendo dos valores de  $\phi$  e  $\theta$ . Por exemplo, na figura 2.6 está apresentado o gráfico da densidade espectral do processo  $ARMA(1, 1)$  para  $\phi = 0.5$  e  $\theta = 0.5$ .

Se  $X_t$  é um processo  $ARMA(1, 1)$  estacionário e  $\epsilon \sim RB(0, \sigma_\epsilon^2)$ , então:

- i. **ARMA(1,1):**  $X_t = \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$
- ii. **Funções Polinomiais:**  $A(L) = (1 + \theta L)$  and  $B(L) = (1 - \phi L)$
- iii. **Funções de Ligação:**  $\psi(z) = \frac{1+\theta z}{1-\phi z}$
- iv. **Função de transferência:**  $g_X(z) = \sigma_\epsilon^2 \psi(z) \psi(z^{-1}) = \sigma_\epsilon^2 \left( \frac{1+\theta z}{1-\phi z} \right) \left( \frac{1+\theta z^{-1}}{1-\phi z^{-1}} \right)$
- v. **Densidade espectral:**  $s_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \psi(z) \psi(z^{-1}) = \frac{\sigma_\epsilon^2}{2\pi} \frac{1+\theta^2+2\theta \cos(\lambda)}{1+\phi^2-2\phi \cos(\lambda)}$

Figura 2.6: Espectro ARMA(1,1) com  $\theta = 0.5$   $\phi = 0.5$ 

Fonte: Elaborado pelo autor.

O último exemplo se trata de uma generalização de processos  $ARMA(p, q)$  e sua densidade espectral é a mistura de todas as defasagens tanto do  $AR(p)$  quanto do  $MA(q)$ .

Se  $X_t$  é um processo  $ARMA(p, q)$  estacionário e  $\epsilon \sim RB(0, \sigma_\epsilon^2)$ , então:

- i. **ARMA(p,q):**  $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$
- ii. **Funções polinomiais:**  $A(L) = (1 + \theta_1 L + \dots + \theta_q L^q)$  and  $B(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$
- iii. **Função de ligação:**  $\psi(z) = \frac{1 + \theta_1 z + \dots + \theta_q z^q}{1 - \phi_1 z - \dots - \phi_p z^p}$
- iv. **Função de transferência:**  $g_X(z) = \sigma_\epsilon^2 \psi(z) \psi(z^{-1}) = \sigma_\epsilon^2 \left( \frac{1 + \theta_1 z + \dots + \theta_q z^q}{1 - \phi_1 z - \dots - \phi_p z^p} \right) \left( \frac{1 + \theta_1 z^{-1} + \dots + \theta_q (z^{-1})^q}{1 - \phi_1 z^{-1} - \dots - \phi_p (z^{-1})^p} \right)$
- v. **Densidade espectral:**  $s_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} \psi(z) \psi(z^{-1}) = \frac{\sigma_\epsilon^2}{2\pi} \frac{\prod_{i=1}^q [1 + \eta_i^2 + \eta_i \cos(\lambda)]}{\prod_{j=1}^p [1 + \omega_j^2 - 2\omega_j \cos(\lambda)]}$

Quanto mais defasagens e mais misturado o processo, mais complicada fica sua densidade espectral e mais difícil fica a identificação de qual frequência tem maior ou menor contribuição para toda a variabilidade do processo. Na próxima subseção serão discutidos a estimação de espectro de um processo quando não se sabe qual foi o processo gerador, bem como outros estimadores para redução de viés, variância e que acomode valores extremos.

## 2.4 Estimação do Espectro

### 2.4.1 O Periodograma

Conforme foi elucidado na seção anterior, a decomposição espectral de uma série temporal em harmônicos ortogonais tem uma série de vantagens na análise como replicar a série temporal, descobrir o processo gerador de dados e visualização. Contudo, a maior importância está centrada na decomposição da variabilidade de todo o processo em frequências, sendo esta decomposição feita pela densidade espectral usando as séries de Fourier ([Persival e Walden, 1993](#)). Mas o principal problema é o fato da não disponibilidade das ondas componentes do processo gerador de dados. Na prática, a série temporal final que, teoricamente, já é a soma de todas as componentes harmônicas.

Neste contexto, surge a necessidade de um estimador que consiga obter as estimativas das amplitudes de cada componente harmônico e esta subseção tem como objetivo explicar os principais estimadores.

Neste contexto, suponha que existam várias populações, cada uma com seu processo gerador de dados. Por meio da identificação do padrão de cada população, várias realizações podem ser obtidas e por meio de seu padrão é possível prever a população de pertencimento de uma nova série temporal desconhecida, sabendo por hipótese que

pertence a uma das populações existentes. Além disso, este tipo de análise faz sentido quando o processo tem características similares, pois se seu padrão é muito distinto fica mais fácil a identificação do processo gerador de dados de cada sistema <sup>9</sup>

Usando a mesma notação de [Krafty \(2016\)](#), considere que existem  $J$  populações, cada uma com seu respectivo processo gerador de dados e várias realizações do processo podem ser obtidas, chamadas aqui de  $k$ -ésima réplica, com  $k = 1, \dots, n_j$  e  $n_j$  sendo o número de realizações por população (cada população pode ter uma quantidade de realização distinta). Sem perda de generalidade, suponha que o tamanho de cada realização deve ser igual à  $N$  em todas as réplicas para todas as populações para que possam ser comparáveis. Definindo  $\Pi_j$  como sendo a  $j$ -ésima população com  $k$  realizações  $\pi_j$  a proporção de réplicas em comparação ao total de réplicas possíveis.

Com esta notação, suponha todas as séries temporais são números reais, fracamente estacionárias e com média zero, formando o conjunto  $\{X_{jkt}\}_{t \in Z}$ . Estas são hipóteses de regularidade que garantem séries bem comportadas, mas principalmente o fato de terem auto-covariâncias absolutamente somáveis, condição necessária para existência da densidade espectral. Logo, para  $|\nu| < N$ , o estimador imediado das auto-covariâncias como proposto na equação 2.2 é definido por ([Persival e Walden, 1993](#), p.196):

$$\hat{\gamma}_\nu^{(jk)} = \frac{1}{N} \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} \quad (2.11)$$

[Persival e Walden \(1993\)](#) sugere substituir  $\hat{\gamma}_\nu^{(jk)}$  na equação da densidade espectral definida na definição 2.2, resultando no estimador clássico do espectro, o Periodograma descrito na definição .

Dadas  $n_1, n_2, \dots, n_j$  realizações para as populações  $\Pi_j$ , supondo que todas tem média zero, fracamente estacionárias e compostas por números reais, o periodograma clássico para a  $k$ -ésima réplica e  $j$ -ésima população pode ser escrito como:

$$I_X^{(jk)}(\lambda_m) = \frac{1}{2\pi N} \left| \sum_{t=1}^N X_{jkt} e^{-i\lambda_m t} \right|^2 \quad (2.12)$$

em que  $m = 1, \dots, N/2$  e  $\lambda_m = 2\pi m/N$  são as frequências de Fourier <sup>10</sup>.

Com esta notação, o periodograma é uma medida de norma quadrática da transformada de Fourier de cada série temporal, estando, portanto, no espaço de Hilbert.

Embora existam tais vantagens na utilização da transformada de Fourier das séries temporais para estimação do espectro, seu cálculo é pouco intuitivo e não tem muita interpretação prática. Mas [Persival e Walden \(1993\)](#); [Priestley \(1981\)](#); [Reisen, Lévy-Leduc, et al. \(2020\)](#) sugerem outra formulação que gera os mesmos resultados mas com uma interpretação prática.

<sup>9</sup>Por exemplo, processos  $AR(1)$  que se diferem apenas pelo valor do coeficiente  $\phi$ .

<sup>10</sup>Para mais ver ([Persival e Walden, 1993](#), p.196).

uma vez que o espectro é uma soma de harmônicos ortogonais (senos e cossenos) para cada frequência, cada um com sua amplitude, é natural que possa tentar obter-se um estimador cujas variáveis explicativas são funções senoidais e cossenoidais para cada frequência, em que os parâmetros estimados lineares se tornam as amplitudes. E neste formato, as estimativas podem ser obtidas por modelo de regressão clássico e pelo método de mínimos quadrados caso o ruído branco tenha distribuição normal.

Usando esta abordagem, considere que a  $k$ -ésima réplica da  $j$ -ésima população pode ser escrita como:

$$X_{jkt} = S'_{jk}\beta_{jk} + \xi_{jkt} \quad (2.13)$$

em que  $k = 1, \dots, n_j$ ,  $j = 1, \dots, J$ ,  $X_{jkt}$  é um vetor ( $N \times 1$ ),  $S_{jk}$  é uma matriz ( $N \times N/2$ ),  $\beta_{jk}$  é um vetor ( $N/2 \times 1$ ) e  $\xi_{jkt}$  é um vetor ( $N \times 1$ )<sup>11</sup>.

A matriz  $S_{jk}$  é uma matriz contendo sub-vetores de tamanho ( $2 \times 1$ ), contendo termos ortogonais de senos e cossenos para cada frequência, enquanto  $\beta_{jk}$  é uma matriz contendo vetores também de tamanho ( $2 \times 1$ ) com os parâmetros a serem estimados, representando as amplitudes de cada harmônico (senos e cossenos) a serem estimados. O vetor de parâmetros é comumente estimado de dois em dois por meio de uma regressão com dois parâmetros de inclinação e sem intercepto, pois a séries tem média zero por hipótese e cada componente é ortogonal. Neste último caso, o produto escalar de qualquer vetor em um harmônico é igual à zero quando feito com dois vetores com frequências diferentes. Logo, cada vetor de betas pode ser escrito como:

$$\beta_{jk}(\lambda_m) = \begin{bmatrix} \beta_{jk}^c(\lambda) \\ \beta_{jk}^s(\lambda) \end{bmatrix}_{(2 \times 1)} \quad (2.14)$$

em que  $c$  e  $s$  representam os termos cosseno e seno respectivamente. E com esta notação, os parâmetros podem ser estimados por mínimos quadrados ordinários (MQO) resolvendo o problema de otimização abaixo:

$$\beta_{jk}^{MQO}(\lambda_m) = \underset{\beta \in R^2}{\operatorname{argmin}} \sum_{t=1}^N (X_{jkt} - S'_{jk}(\lambda_m)\beta_{jk}(\lambda_m))^2 \quad (2.15)$$

O resultado da otimização clássica mor MQO irá resultar na seguinte equação:

$$(S'_{jk}S_{jk})\beta_{jk}^{MQO} = S'_{jk}X_{jkt} \quad (2.1)$$

E obviamente, se a matriz  $(S'_{jk}S_{jk})$  é não singular, o estimador de MQO clássico neste caso pode ser definido por:

$$\beta_{jk}^{LS} = (S'_{jk}S_{jk})^{-1}S'_{jk}X_{jkt} \quad (2.16)$$

<sup>11</sup>Para maior clareza da equação 2.3, ver Anexo C.2

Finalmente, (Priestley, 1981, p.395) mostra que para cada  $j$ -ésima população e  $k$ -ésima réplica, o periodograma clássico pode ser escrito como a norma quadrática dos vetores de betas (amplitudes) como:

$$I_N(\lambda_m) = \frac{N}{8\pi} \left\| \hat{\beta}_{jk}(\lambda_m) \right\|^2 \quad (2.17)$$

em que  $\|\cdot\|$  denota a norma euclidiana. Para cada frequência de Fourier  $\lambda_m$  com  $m = 1, \dots, 1/N$ , as estimativas de MQO das amplitudes contidas nos vetores  $\hat{\beta}_{jk}(\lambda_m)$  são dadas por:

$$\beta_{jk}^{(c)}(\lambda_m) = \sqrt{\frac{2}{N}} \sum_{t=1}^N X_{jkt} \cos(\lambda_m t) \quad (2.18)$$

$$\beta_{jk}^{(s)}(\lambda_m) = \sqrt{\frac{2}{N}} \sum_{t=1}^N X_{jkt} \sin(\lambda_m t) \quad (2.19)$$

Além da formulação clássica do periodograma como a norma quadrática da transformada de Fourier da série temporal, existem outras formulações que não serão discutidas aqui por fugir do escopo do trabalho. Contudo, deve-se enfatizar a formulação do periodograma por meio da transformada de Fourier das auto-covariâncias.

#### (Periodograma como Transformada de Fourier das Auto-covariâncias)

O periodograma da  $j$ -ésima população e  $k$ -ésima réplica pode ser calculado usando  $\hat{\gamma}_\nu^{(jk)}$  definido na equação 2.11 como:

$$I_X^{(jk)}(\lambda_m) = \frac{1}{2\pi} \sum_{t=-(N-1)}^{N-1} \hat{\gamma}_\nu^{(jk)} \cos(\lambda_m t) \quad (2.20)$$

Na próxima seção será descrita outra formulação importante do periodograma que leva em consideração a estimação acomodando valores abruptos que se comportam como outliers aditivos.

### 2.4.2 O M-Periodograma

No contexto de séries temporais, vários são os casos em que valores extremos e abruptos sejam observados. E na última seção, foi feita uma detalhada formulação para obter as estimativas do espectro sem considerar tal possibilidade pois o estimador de MQO usado para obter as estimativas dos parâmetros da regressão descrita na equação 2.13 são já conhecidos na literatura por serem sensíveis a valores abruptos (Huber e Ronchetti, 2009). Se for levado em consideração que o periodograma é calculado como

uma medida quadrática dos parâmetros estimados, torna-se ainda mais sensível a tais valores aberrantes. E nesses caso, torna-se necessário o uso de um estimador que acomode tais valores.

Para acomodar valores abruptos que se comportem como outliers aditivos, defende-se neste trabalho o uso do M-Periodograma usando o estimador descrito por [Huber e Ronchetti \(2009\)](#) ao invés do MQO na equação 2.13. Trata-se de um estimador como uma função definida em partes, em que o estimador de MQO é utilizado caso o resíduo não ultrapasse um determinado valor, e uma medida de norma unitária é utilizada em seu lugar caso ultrapasse.

Considere novamente a equação 2.13, então a equação

$$\beta_{jk}^M(\lambda_m) = \underset{\beta \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{t=1}^N \psi \left[ X_{jkt} - S'_{jk}(\lambda_m) \beta_{jk}(\lambda_m) \right] \quad (2.21)$$

em que a função  $\psi(\cdot)$  é a função de influência Huber proposta por [Huber e Ronchetti \(2009\)](#) e definida como:

$$\psi_H(x) = \begin{cases} x, & |x| \leq c \\ c \operatorname{sign}(x), & \text{otherwise} \end{cases}, \quad (2.21)$$

em que  $\operatorname{sign}(x)$  é uma função que apenas troca o sinal de  $x$ .

Desta forma, o estimador obtido será o estimador robusto das amplitudes de cada harmônico para a  $j$ -ésima população e  $k$ -ésima réplica proposto por [Katkovnik \(1998\)](#) e explicado em [Li \(2008\)](#); [Reisen, Lévy-Leduc, et al. \(2020\)](#).

Na próxima seção, serão discutidas algumas propriedades dos estimadores do espectro, bem como as alternativas utilizadas na literatura para correção de viés e variância.

### 2.4.3 Multitaper

A técnica do periodograma multitaper é utilizada para melhorar tanto a precisão quanto a estabilidade das estimativas espectrais por meio da redução da variância e diminuição do vazamento espectral. No caso do vazamento espectral, ele ocorre quando a energia de uma frequência com muita amplitude é espalhada entre frequências próximas. Tal vazamento ocorre devido à natureza finita utilizada para o espectro, geralmente variando de  $0$  à  $\pi$ . Caso a série seja periódica mas não termina com a mesma altura mais próximas de  $\pi$ , comparada às alturas próximas de zero, há o vazamento do espectro, chamado de *sidelobe*. Para mitigar este sidelobe, uma janela é aplicada aos dados de forma

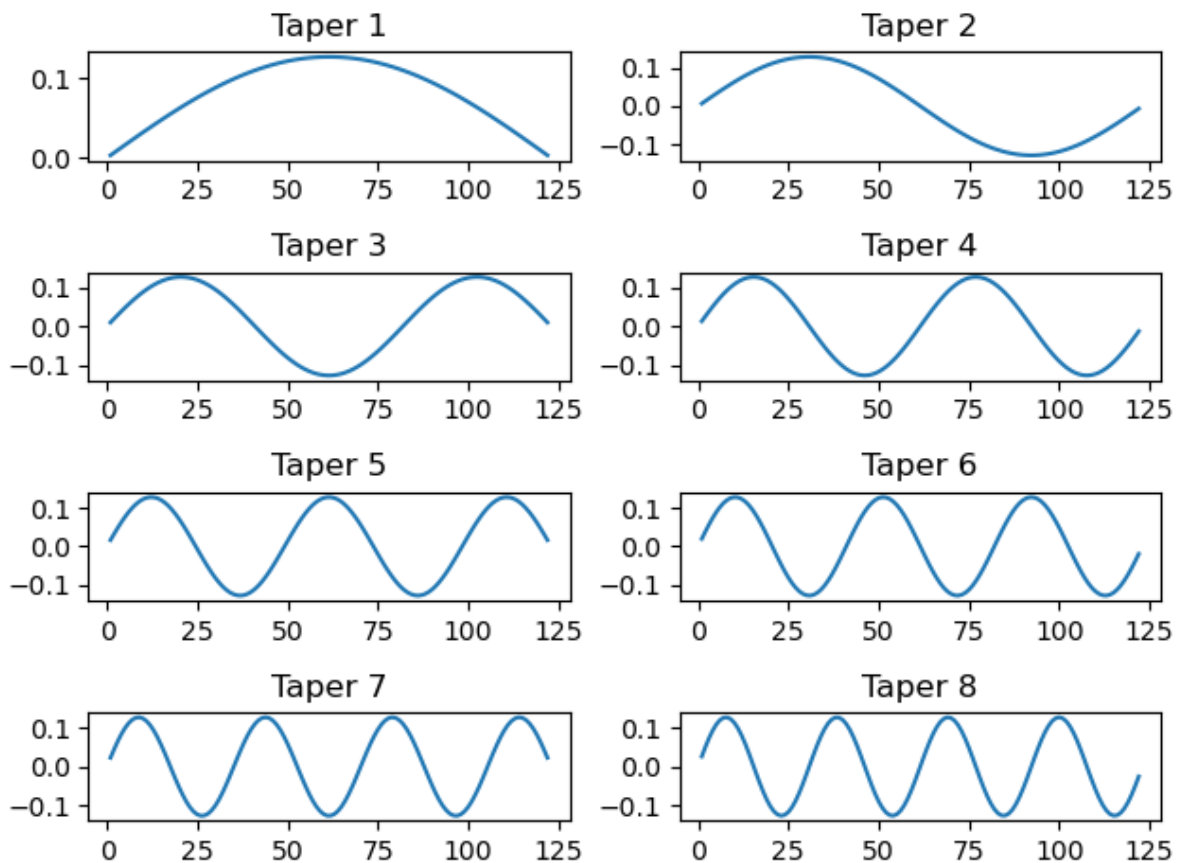
a ter valores próximos de zero tanto próximos à frequência 0 quanto à frequência  $\pi$ . Tal janela vai afunilar os dados para gerar tal estabilidade.

Existem vários funis (*tapers*) que podem ser utilizados neste contexto. Mas o mais comum utilizado na literatura é o funil seno devido suas características de ortogonalidade para diferentes funis. Considere a equação da equação do funil abaixo:

$$h_{rt} = \left( \frac{2}{N+1} \right)^{1/2} \sin \left( \pi t \frac{r}{N+1} \right). \quad (2.22)$$

eme que  $r = 1, \dots, R$  com  $R$  sendo o total de funis a serem utilizados e  $h_{rt}$  é a função de peso que deve ser utilizada na série temporal para afunilar e posteriormente calcular o periodograma afunilado.

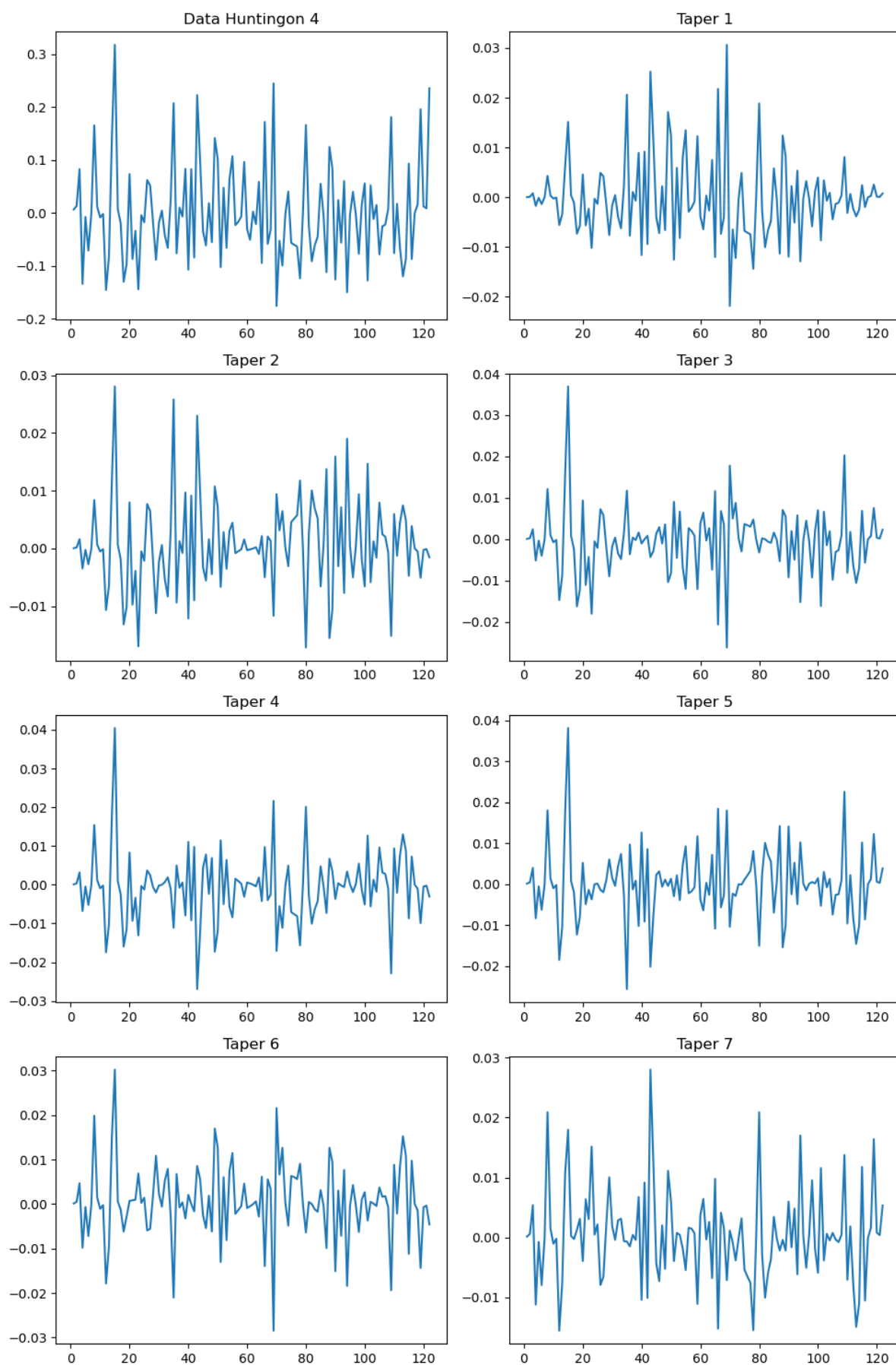
Figura 2.7: 8 funis Seno de tamanho 120 observações



Fonte: Elaborado pelo autor.

Na figura 2.7 estão apresentados os gráficos das funções funis para  $r = 1, \dots, 8$ .

Figura 2.8: Dados de ciclos de caminhada do paciente 4 com doença de Huntington e estes dados afunilados até o sétimo funil seno



Fonte: Elaborado pelo autor.

Já na figura 2.4, no primeiro painel estão apresentados os dados de uma série temporal <sup>12</sup>. Nos demais painéis, estão os dados resultantes da multiplicação da função funil seno apresentada na figura 2.7 pela série temporal. Observe que no primeiro funil (taper 1), os valores do centro estão livres com seus respectivos valores, enquanto os valores próximos a zero ou à 120 são afunilados para zero com a função seno. Já no segundo taper, há uma "dupla-onda" de liberdade para os dados. Isto permite que novas informações da série sejam coletadas, mas ortogonais ao funil anterior e ainda afunilando nas bordas. Assim sucessivamente, cada  $R$ -ésimo funil irá gerar " $R$ " ondas e novos dados ortogonais a todos os  $R - 1$  funis anteriores.

Para cada dado afunilado é calculado o periodograma. A estimativa final do espectro é obtida ao calcular a média das potências espectrais individuais, chamado de periodograma multitaper e obtido por:

$$I_{jk}^R(\lambda_m) = \frac{1}{2\pi NR} \sum_{r=1}^R \left\| \sum_{t=1}^N h_{rt} X_{jkt} e^{-i\lambda_m t} \right\|^2. \quad (2.23)$$

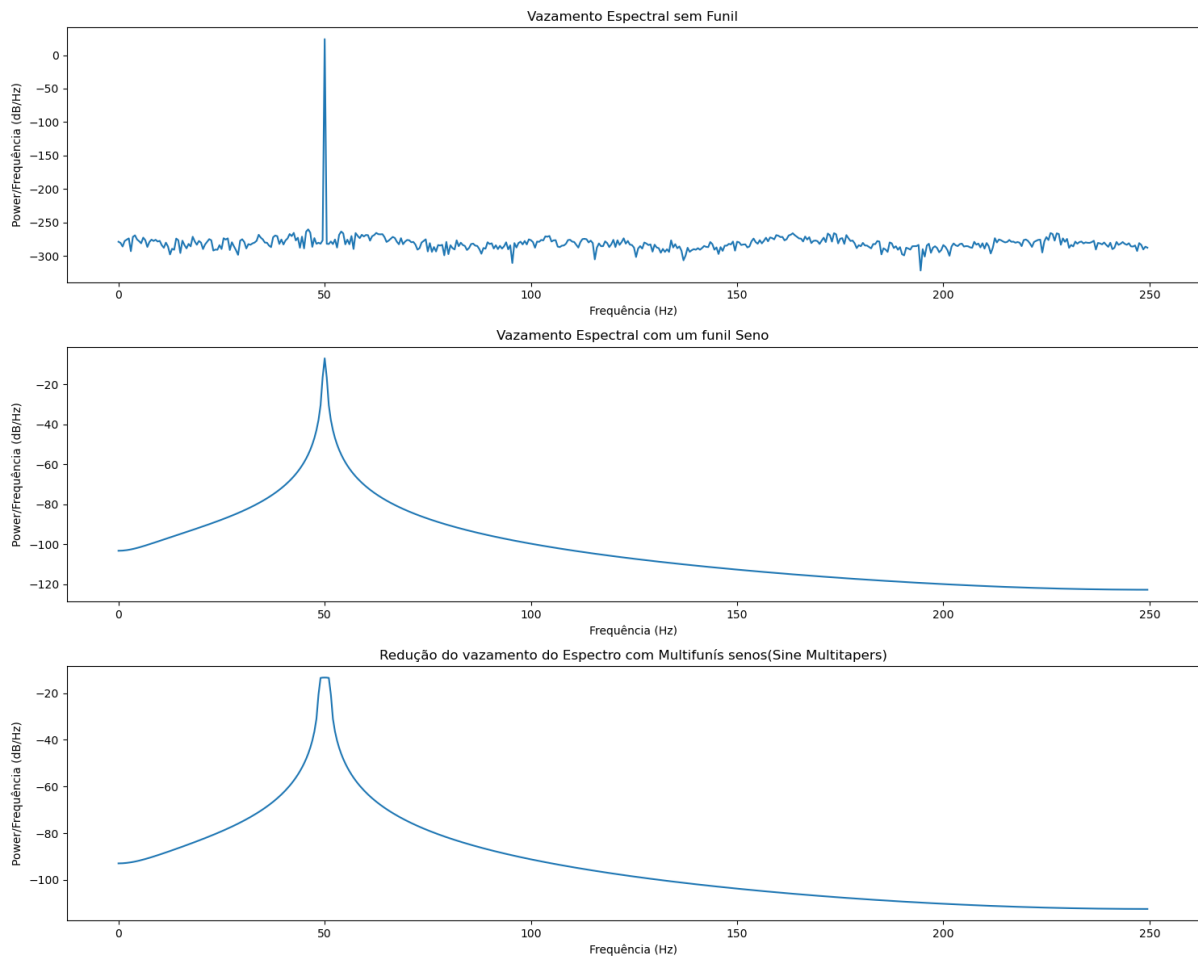
O método multitaper usa janelas senoidais (ou outras janelas otimizadas) que são projetadas para minimizar o vazamento espectral. As funções senoidais, por exemplo, têm propriedades de concentração espectral que limitam a dispersão da energia de uma frequência para frequências adjacentes. Adicionalmente, as janelas utilizadas no multitaper são ortogonais, o que significa que elas não têm sobreposição de energia entre si. Isso ajuda a assegurar que cada estimativa espectral seja obtida de uma forma que minimize o vazamento de frequências.

Ao combinar as estimativas espectrais obtidas de diferentes janelas, o método multitaper consegue suprimir ainda mais o vazamento, resultando em uma representação mais nítida e precisa das frequências presentes na série temporal.

---

<sup>12</sup>Série temporal de Huntingtons, apresentados no capítulo 5

Figura 2.9: Redução do vazamento espectral usando os dados do periodograma multifunilado



Fonte: Elaborado pelo autor.

Com relação à série temporal apresentado na figura 2.8, foi estimado o periodograma dos dados originais, dos com um funil e com 7 funis, apresentados na figura 2.9. Note que no primeiro painel, os dados apresentam distúrbios no periodograma, chamados na literatura de *sidelobes*. Estes *sidelobes* existem devido ao vazamento espectral e pode-se observar que com um funil, os *sidelobes* ficam mais suaves e menores. Note ainda que do primeiro para o segundo painel, a amplitude vertical dos dados é diminuída, o que significa que os valores estão mais concentrados relativamente na frequência máxima em 50 hz. O último painel tem como objetivo mostrar que há ainda mais concentração de *power* em torno da frequência 50 hz quando vários funis são aplicados e combinados em uma média. Neste caso, uma média de 7 funis foi utilizada para gerar o último gráfico da figura.

## 2.5 Análise Cepstral

Cepstral é uma ferramenta para investigar estruturas periódicas em espectros de frequência para extrair os componentes fundamentais de um sinal, no sentido de que separa sinais que foram combinados de uma maneira não aditiva. **A. V. Oppenheim e Schafer (2004)** deu o seguinte exemplo para esclarecer o significado de cepstral. Suponha que um sinal de um eco simples possa ser escrito como

$$X_t = s_t + \alpha s_{(t-\tau)}. \quad (2.23)$$

onde  $\alpha$  é uma constante que satisfaz condições para garantir a estabilidade do processo real  $x_t$ , e  $s_t$  é um processo de ruído estável. Note que  $x_t$  corresponde a um processo  $MA(1)$  com um processo de ruído branco  $s_t$  com defasagem  $\tau$ . A representação espectral da Eq. 2.23 pode ser escrita como

$$\|X_t(\lambda)\|^2 = \|S_s(\lambda)\|^2 [1 + \alpha^2 + 2\alpha \cos(\lambda\tau)]. \quad (2.24)$$

$$\log \|X_t(\lambda)\|^2 = \log \|S_t(\lambda)\|^2 + \log[1 + \alpha^2 + 2\alpha \cos(\lambda\tau)]. \quad (2.25)$$

A partir da Eq. 2.24, pode-se perceber que a densidade espectral do eco corresponde ao produto do espectro do ruído ( $S_s$ ) com a contribuição espectral do eco. Uma maneira de ver a contribuição individual da variabilidade temporal do eco é aplicar a transformação logarítmica, como mostrado na Eq. 2.25. A partir disso, é possível observar que o  $\log \|X_t\|^2$  tem uma forma de onda com o componente periódico com atraso  $\tau$ .

**Bogert (1963)**, em seu trabalho seminal, introduzem o cepstrum (um substantivo parafraseado da palavra spectrum), que é a transformada de Fourier inversa do  $\log \|X_t\|^2$ . Os coeficientes correspondentes são chamados de coeficientes cepstrais ( $c_\ell$ ),  $\ell = 0, 1, \dots$ . A análise cepstral está no domínio denominado quefrência. Essas terminologias foram introduzidas pelos autores onde cepstral e cepstrum são anagramas de espectral e espectro, respectivamente. Além da elegância matemática do cepstrum e dos coeficientes cepstrais, eles apresentam aplicações interessantes em várias áreas do conhecimento. O cepstrum é uma ferramenta para investigar estruturas periódicas em espectros de frequência, com principais aplicações em fala humana, música e sistemas de energia elétrica.

As estruturas periódicas estão relacionadas a ecos perceptíveis no sinal ou à ocorrência de frequências harmônicas. Matematicamente, lida com o problema da deconvolução de sinais no espaço de frequência. A definição típica para uma harmônica é “um componente sinusoidal de uma onda periódica ou quantidade com uma frequência que é um múltiplo inteiro da frequência fundamental.” Algumas referências se referem a “energia limpa” ou

“pura” como aquelas formas de onda sem harmônicas (“IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems” (2014)). Veja uma revisão em Cavicchioli (2020).

O cepstra também tem sido utilizado para discriminar e classificar densidades espectrais, ou seja, uma análise discriminante baseada nos domínios de frequência e quefrência. Fokianos e Savvides (2008) usou a metodologia para testar a similaridade de G funções densidade espectrais de processos estacionários G-independentes. Krafty (2016) introduz os coeficientes cepstrais na distância de Mahalanobis para construir uma função discriminante nos domínios de frequência e quefrência, considerando a variabilidade entre e dentro dos grupos.

Para uma análise visual, aa figura 2.10 apresenta a decomposição da variação total da série temporal em partes de ruído branco e autocorrelação, sem ser dividida por  $2\pi$ . Usando um processo  $AR(1)$  com  $\phi = 0.5$ ,  $\epsilon_t \sim \mathcal{N}(0, 1)$ , é possível observar que, para uma frequência fixa  $\lambda^*$ , a área verde representa a parte de ruído branco que é igual a 1. E a parte verde é devida à soma de várias autocorrelações de 0 a  $\lambda^*$ .

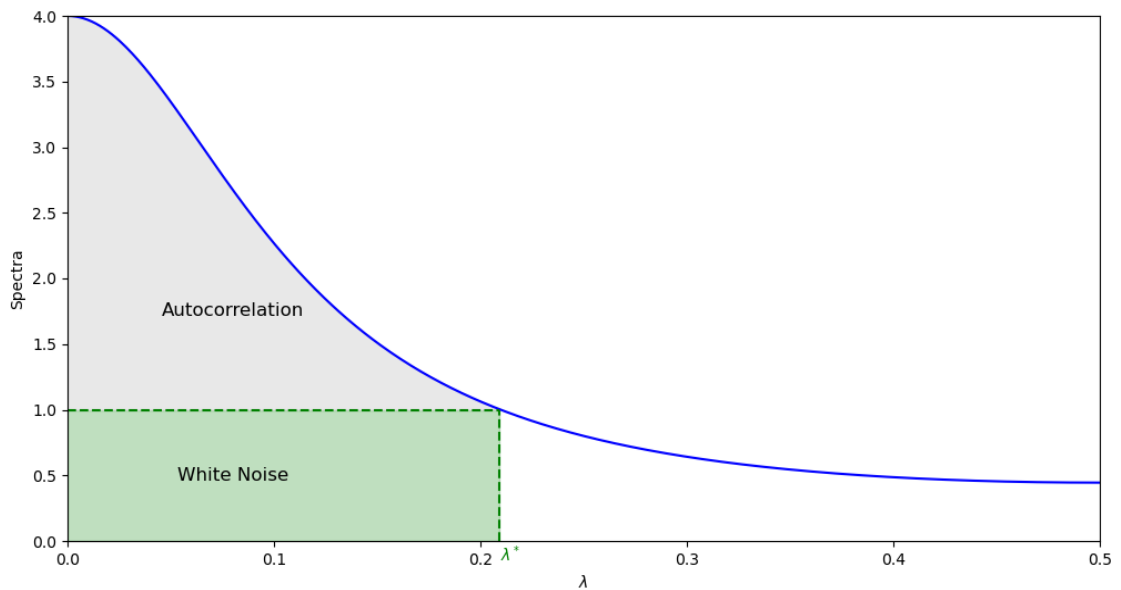


Figura 2.10: Decomposition of White Noise and autocorrelations of total variability of Spectral Density of a  $AR(1)$  process with  $\phi = 0.5$ .

Como este trabalho está considerando o problema discriminante com base em modelos  $ARMA$  estacionários de média zero, deriva-se os coeficientes cepstrais para alguns exemplos dessa classe de processos abaixo.

Seja  $X_t$  um processo estacionário  $AR(p, q)$  e  $\sigma^2$  a variância da inovação. Além disso, sejam  $\eta_i$  e  $\zeta_r$  a  $i$ -ésima e  $r$ -ésima raiz dos polinômios  $\Theta(z)$  e  $\Phi(z)$ , respectivamente (Hamilton, 1994, p.155). Então, “ “

i. O log do espectro de  $X_t$  pode ser escrito como :

$$\begin{aligned}
 \log S_X(\lambda) &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \left[ \frac{\prod_{i=1}^q (1 + \eta_i^2 + 2\eta_i \cos(\lambda))}{\prod_{r=1}^p (1 + \zeta_r^2 + 2\zeta_r \cos(\lambda))} \right] \right\} \quad (2.8) \\
 &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \sum_{i=1}^q \log [1 + \eta_i^2 + 2\eta_i \cos(\lambda)] - \sum_{r=1}^p \log [1 + \zeta_r^2 + 2\zeta_r \cos(\lambda)] \\
 &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \left\{ \sum_{i=1}^q \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} \cos(\lambda \ell) + \sum_{r=1}^p \sum_{\ell=1}^{\infty} \frac{\zeta_r^\ell}{\ell} \cos(\lambda \ell) \right\}. \\
 &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \left\{ \sum_{i=1}^q \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} \cos(\lambda \ell) + \sum_{r=1}^p \frac{\zeta_r^\ell}{\ell} \cos(\lambda \ell) \right\}.
 \end{aligned}$$

ii. Usando o resultado do log do espectro, os coeficientes de cepstra para o processo  $ARMA(p, q)$  pode ser escrito como:

$$c_\ell = \begin{cases} \log\left(\frac{\sigma^2}{2\pi}\right) & , \text{if } \ell = 0 \\ 2 \left( \sum_{i=1}^q \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} + \sum_{r=1}^p \frac{\zeta_r^\ell}{\ell} \right) & , \text{if } \ell \geq 1. \end{cases} \quad (2.9)$$

Em particular, os Corolários 2.5, 2.5 e 2.5 exibem os log espectros e os coeficientes cepstrais para  $AR(1)$ ,  $MA(1)$  e  $ARMA(1, 1)$ . As demonstrações completas são apresentadas nos Apêndices 7.1 e 7.2.

Seja  $X_t = \phi X_{(t-1)} + \epsilon_t$  um processo Gaussiano estacionário  $AR(1)$ . Então:

i. O log espectro pode ser escrito como:

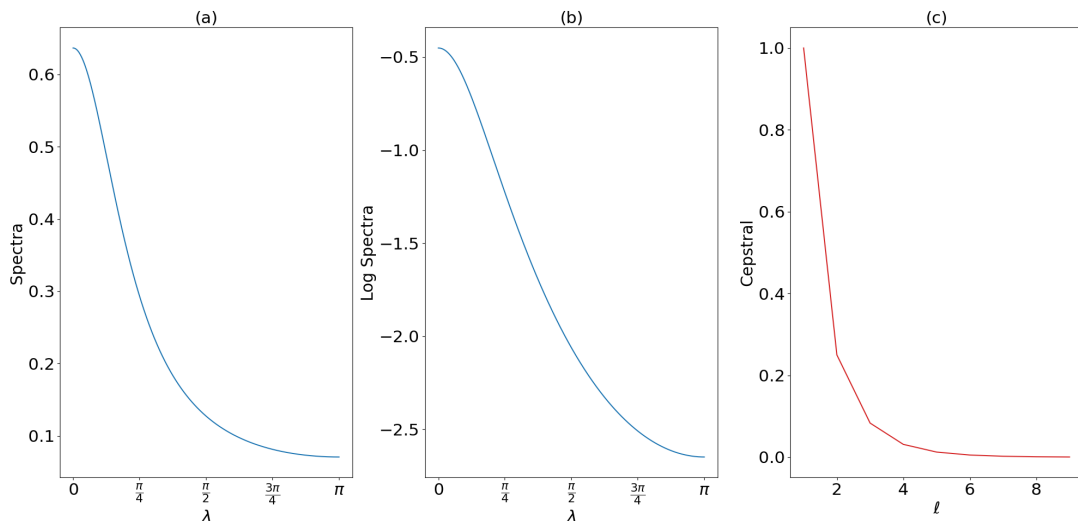
$$\begin{aligned}
 \log S_X(\lambda) &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} [1 + \phi^2 - 2\phi \cos(\lambda)]^{-1} \right\} \quad (2.10) \\
 &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} - \log [1 + \phi^2 - 2\phi \cos(\lambda)] \\
 &= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \frac{\phi^\ell}{\ell} \cos(\lambda \ell).
 \end{aligned}$$

ii. Usando o log espectro, os coeficientes cepstrais para  $AR(1)$  podem ser escritos como:

$$c_\ell = \begin{cases} \log\left(\frac{\sigma^2}{2\pi}\right) & , \text{ se } \ell = 0 \\ \frac{2\phi^\ell}{\ell} & , \text{ se } \ell \geq 1. \end{cases} \quad (2.11)$$

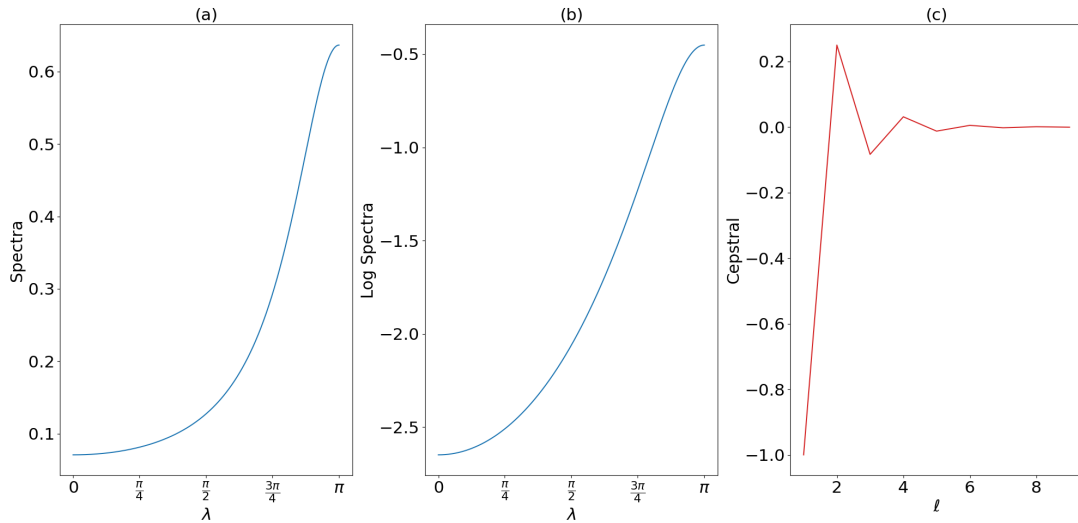
As Figuras 2.11 e 2.12 mostram espectros (a), log espectros (b) e cepstra (c) para o caso particular de  $AR(1)$  com  $\phi = 0.5$  e  $\phi = -0.5$ , respectivamente. Na Figura 2.11, os espectros mostram que a maior parte da variabilidade do processo é explicada por baixas frequências, próximas de zero. Isso indica uma série temporal com volatilidade relativamente baixa, o que é confirmado pelos cepstra. No painel (c), os cepstra diminuem mais rapidamente do que uma função exponencial em direção a zero. Isso implica que aproximadamente  $\ell = 8$  é suficiente para explicar a maior parte da variabilidade do processo.

Figura 2.11:  $\phi = 0.5$  e  $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra



Fonte: Elaborado pelo autor.

Por outro lado, é possível ver na Figura 2.12 que com  $\phi < 0$ , a maioria da variabilidade do processo é explicada por altas frequências, próximas a  $\pi$ . Como resultado, a série é mais volátil, e os cepstra exibem comportamento semelhante a uma função periódica oscilando em torno de zero. Além disso, essas características mostram que apenas cerca de  $\ell = 6$  é suficiente para explicar a maior parte da variabilidade do processo.

Figura 2.12:  $\phi = -0.5$  e  $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra

Fonte: Elaborado pelo autor.

Como pode ser observado no Corolário 2.5, o comportamento dos coeficientes cepstrais para um processo  $MA(1)$  difere daquele do processo  $AR(1)$  apresentado no Corolário 2.5 apenas por uma mudança de sinal para  $\ell \geq 1$  e pelo parâmetro  $\theta$ .

Seja  $X_t = \epsilon_t + \theta\epsilon_{t-1}$  um processo Gaussiano invertível  $MA(1)$ . Então

i. O log espectro pode ser escrito como:

$$\log[S_X(\lambda)] = \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} [1 + \theta^2 + 2\theta \cos(\lambda)] \right\} \quad (2.12)$$

$$= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \log [1 + \theta^2 + 2\theta \cos(\lambda)]$$

$$= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \theta^\ell}{\ell} \cos(\lambda\ell).$$

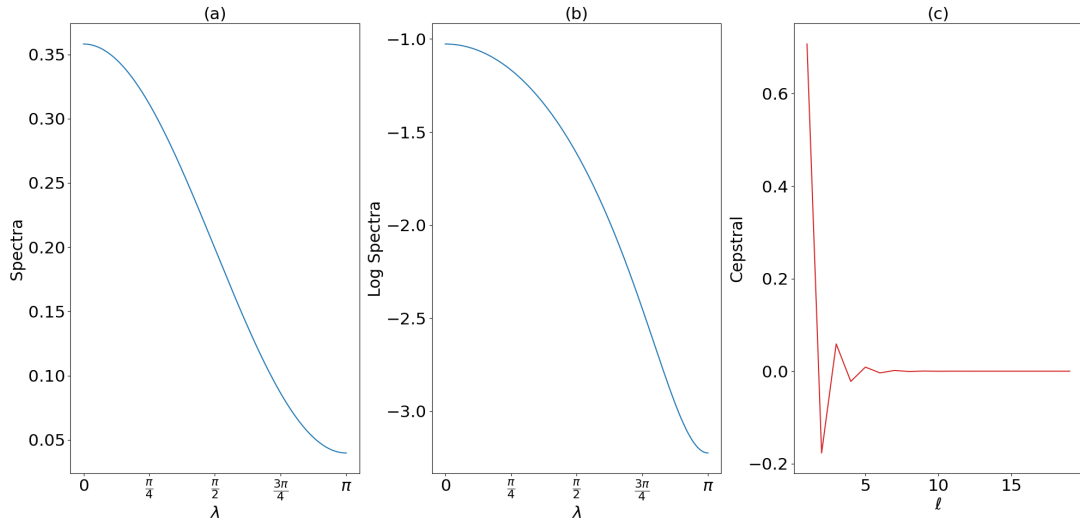
ii. Usando o log espectro, resultará em coeficientes cepstrais para  $MA(1)$ , dados por:

$$c_\ell = \begin{cases} \log\left(\frac{\sigma_\epsilon^2}{2\pi}\right) & , \text{ se } \ell = 0 \\ \frac{(-1)^{\ell+1} 2\theta^\ell}{\ell} & , \text{ se } \ell \geq 1. \end{cases} \quad (2.13)$$

Da mesma forma, nas Figuras 2.13 e 2.14, o painel (a) mostra baixas e altas frequências, respectivamente. Assim, quando  $\theta > 0$ , o processo exibe comportamento semelhante ao de  $\phi > 0$ , enquanto  $\theta < 0$  é semelhante a  $\phi < 0$ . No entanto, como pode ser observado no painel (c), os cepstra exibem comportamento oposto, indicando que a baixa frequência no processo  $MA(1)$  com  $\theta = 0.5$  está associada a uma função periódica

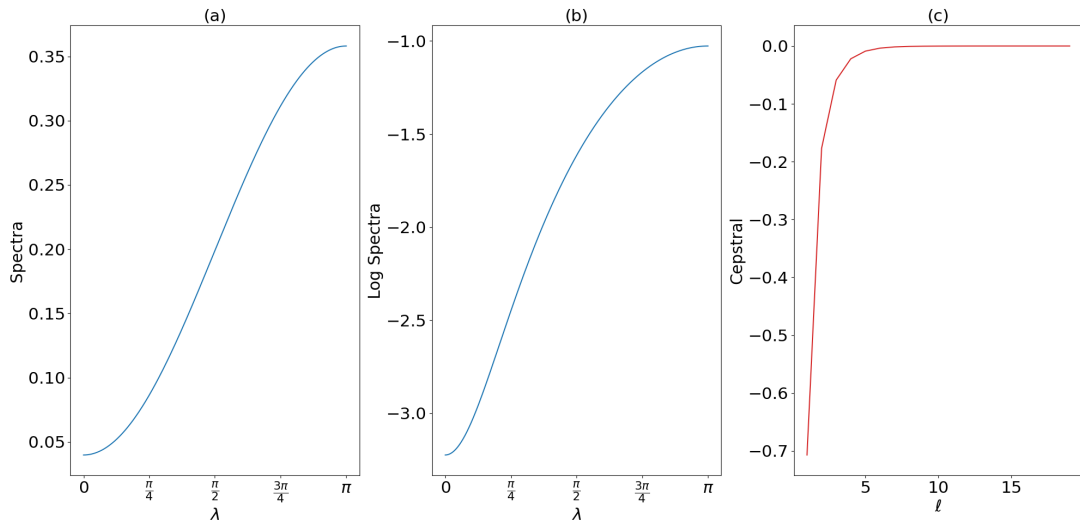
nos cepstra. Isso ocorre porque o espectro do  $MA(1)$  tem um sinal positivo no termo cosseno, enquanto o processo  $AR(1)$  tem um sinal negativo.

Figura 2.13:  $\theta = 0.5$  e  $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra



Fonte: Elaborado pelo autor.

Figura 2.14:  $\theta = -0.5$  e  $\sigma_\epsilon^2 = 1$ : (a) Espectros, (b) Log Espectros, (c) Cepstra



Fonte: Elaborado pelo autor.

Seja  $X_t = \phi X_{(t-1)} + \epsilon_t + \theta \epsilon_{t-1}$  um processo Gaussiano estacionário invertível  $ARMA(1, 1)$ . Então

i. O log espectro pode ser escrito como:

$$\log S_X(\lambda) = \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \left[ \frac{1 + \theta^2 + 2\theta \cos(\lambda)}{1 + \phi^2 - 2\phi \cos(\lambda)} \right] \right\} \quad (2.14)$$

$$\begin{aligned}
&= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \log [1 + \theta^2 + 2\theta \cos(\lambda)] - \log [1 + \phi^2 - 2\phi \cos(\lambda)] \\
&= \log \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \left\{ \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \theta^\ell}{\ell} \cos(\lambda\ell) + \sum_{\ell=1}^{\infty} \frac{\phi^\ell}{\ell} \cos(\lambda\ell) \right\}.
\end{aligned}$$

ii. Usando o log espectro, resultará em coeficientes cepstrais para  $ARMA(1, 1)$ , dados por:

$$c_\ell = \begin{cases} \log\left(\frac{\sigma^2}{2\pi}\right) & , \text{se } \ell = 0 \\ \frac{2\phi^\ell}{\ell} + \frac{(-1)^{\ell+1}2\theta^\ell}{\ell} & , \text{se } \ell \geq 1. \end{cases} \quad (2.15)$$

Sob a suposição de um modelo estacionário  $ARMA$ , os coeficientes cepstrais são de ordem  $c_\ell = \mathcal{O}(\ell^{-3/2}) = o(\ell^{-1/2})$  enquanto  $\ell \rightarrow \infty$ .

Note que o coeficiente cepstral decai à taxa  $c_\ell \sim \vartheta |\frac{\delta^\ell}{\ell}|$ ,  $-\infty < \ell < \infty$ ,  $\vartheta$  é uma constante e  $|\delta| < 1$ , enquanto a função de autocorrelação dos modelos padrão Box-Jenkins  $ARMA$  decai aproximadamente na forma exponencial. Por exemplo, a ACF do modelo  $AR(1)$  é  $\phi^\ell$ . A diferença matemática de comportamento entre os coeficientes do cepstrum e os valores da ACF mostram claramente que o primeiro exhibe a maior parte da variação temporal do processo com uma defasagem menor do que a função ACF, indicando que a maior parte da informação está contida nos coeficientes cepstrais de ordem inferior.

As Figuras 2.11, 2.12, 2.13 e 2.14 mostram o comportamento dos espectros, log espectros e cepstra dos modelos  $MA(1)$  e  $AR(1)$ , respectivamente. Em ambos os casos, vemos as frequências próximas de zero, que contribuem mais para a variância do processo. O cepstrum mostra que as defasagens iniciais contêm todas as informações necessárias para extrair as características dos processos. Para  $\phi$  positivo, a série é positivamente correlacionada, o espectro e o cepstrum são dominados por baixas frequências e pequenas defasagens, respectivamente, o que significa que a série é relativamente suave. Quando  $\phi < 0$ , a série é negativamente correlacionada, assim, o espectro e o cepstrum são dominados por altas frequências e defasagens, respectivamente. Isso significa que os dados são mais irregulares.

## 2.6 O Estimador M-cepstral

A maioria dos estimadores não paramétricos da densidade espectral são baseados no periodograma. Como é bem conhecido, o periodograma não é resistente a outliers ou distribuições com caudas pesadas como explicado, por exemplo, em Fox (1972) e Li

e Song (2008). Essa questão tem sido discutida recentemente por muitos autores que introduziram novos tipos de periodogramas que mostram superioridade em relação ao periodograma clássico quando os dados têm algumas observações atípicas, por exemplo, Reisen et al. (2017); Fajardo et al. (2018); Li (2008), entre outros.

Nesse contexto, consideramos aqui o  $M$ -periodograma proposto independentemente por Reisen, Lévy-Leduc, et al. (2020) para sugerir a ferramenta de cepstrum discriminante baseada nesse periodograma robusto, denominado aqui como estimador  $M$ -cepstral.

Seja  $\{X_{jk0}, \dots, X_{jk(N-1)}\}$  a  $k$ -ésima realização do processo previamente definido  $\{X_{jkt}\}$  na população  $j$ , um processo que tem autocovariância absolutamente somável. Seguindo a notação em Reisen, Lévy-Leduc, et al. (2020a),  $I_{jk}^M(\lambda_m)$  é definido como

$$I_{jk}^M(\lambda_m) = \frac{N}{8\pi} \left\| \hat{\beta}_{jk}^M(\lambda_m) \right\|^2 = \frac{N}{8\pi} \left( \hat{\beta}_{jk}^{(c)}(\lambda_m)^2 + \hat{\beta}_{jk}^{(s)}(\lambda_m)^2 \right) \quad (3.1)$$

onde  $\hat{\beta}_{jk}^M(\lambda_m) = [\beta_{jk}^{(c)}(\lambda_m), \beta_{jk}^{(s)}(\lambda_m)]^T$  é a solução de

$$\beta_{jk}^M(\lambda_m) = \operatorname{argmin}_{\beta \in R^2} \sum_{t=1}^N \psi [X_{jkt} - \mathbf{S}_{jk}^T(\lambda_m) \beta_{jk}(\lambda_m)] \quad (3.2)$$

onde a função de perda  $\psi(\cdot)$  satisfaz algumas condições e  $\mathbf{S}_{jk}(\lambda_m) = [\cos(t\lambda_m), \sin(t\lambda_m)]^T$ .

O multitaper  $M$ -periodograma ( $I_{jk}^{RM}(\lambda_m)$ ) é definido como

$$I_{jk}^{RM}(\lambda_m) = \frac{N}{8\pi R} \sum_{r=1}^R \left\| \hat{\beta}_{jk}^{rM}(\lambda_m) \right\|^2 = \frac{N}{8\pi} \sum_{r=1}^R \left( \hat{\beta}_{jk}^{r(c)}(\lambda_m)^2 + \hat{\beta}_{jk}^{r(s)}(\lambda_m)^2 \right) \quad (3.3)$$

onde  $\hat{\beta}_{jk}^{rM}(\lambda_m) = [\beta_{jk}^{r(c)}(\lambda_m), \beta_{jk}^{r(s)}(\lambda_m)]^T$  é a solução de

$$\beta_{jk}^M(\lambda_m) = \operatorname{argmin}_{\beta \in R^2} \sum_{t=1}^N \psi [h_{rt} X_{jkt} - \mathbf{S}_{jk}^T(\lambda_m) \beta_{jk}(\lambda_m)] \quad (3.4)$$

Para derivar propriedades assintóticas para séries temporais de curta e longa memória, os autores Reisen, Lévy-Leduc, et al. (2020a), entre outros, consideram a função de perda de Huber com a constante  $c_H = 1.345$ . Seguindo esses autores, também consideramos aqui a função de perda de Huber para calcular os coeficientes amostrais do  $M$ -cepstrum ( $\hat{c}_{jk\ell}^M$ ) e, conseqüentemente, a função discriminante amostral  $M$  ( $\hat{s}_{jkq}^M$ ), dadas como segue.

$$\hat{d}_{jks}^M = \left\langle \hat{\mathbf{p}}_\ell^M, \hat{\mathbf{c}}_{jk\ell}^{(k)M} \right\rangle = \sum_{\ell=0}^{L-1} \hat{p}_{s\ell}^M \hat{c}_{jk\ell}^M. \quad (3.5)$$

onde  $\hat{p}_{s\ell}^M$  são estimadores truncados do verdadeiro  $p_{s\ell}$ .

$$\hat{\Psi}^M(\hat{d}_{*1}^M, \dots, \hat{d}_{*q}^M; \hat{\mu}_j^M) = \min_j \left\{ \sum_{s=1}^q \left[ \left( \hat{d}_{*s}^M - \mu_{js}^M \right) \right]^2 - 2 \log(\hat{f}_j) \right\}. \quad (3.6)$$

---

onde  $\hat{\mathbf{d}}_{jks}^M$  e  $\hat{\Psi}^M$  são os estimadores de 2.20 e 2.21, respectivamente.

## Capítulo 3

# Modelos Lineares de Discriminação e Classificação

Existem várias formas de discriminar e clusterizar um conjunto de dados, podendo estes serem não supervisionados (cujo qual não se sabe à população à qual pertence) tal como  $k$  vizinhos mais próximos e dendrograma ou supervisionado (sabe-se a população à qual pertence) como modelos Logit, SVM (*Support Vector Machine*) e Análise Discriminante linear e quadrática. Exceto o último que é usado apenas em alguns casos específicos, a maioria são modelos lineares cuja separação é feita por um plano de corte para separação dos dados.

A escolha do melhor classificador depende do contexto no qual os dados estão inseridos. E no contexto de séries temporais, é comum a literatura utilizar modelos lineares devido ao fato de gerar altas taxas de classificação e ter uma estrutura mais simples (Shumway e Unger, 1974; Kakizawa et al., 1998; Krafty, 2016)

A próxima seção irá discutir o modelo linear que será utilizado neste trabalho.

### 3.1 Análise Discriminante Linear (LDA) baseado no subespaço M-Cepstral estimado

Como mencionado anteriormente,  $\log S_X(\lambda)$ , para um  $\lambda$  fixo, é decomposto em uma soma linear infinita na qual  $\cos(\lambda\ell)$ ,  $\ell = 1, 2, \dots$  são ponderados por coeficientes,  $\hat{c}_\ell^M$  (cepstrum), que contêm a maior parte da informação do processo. Com base no Corolário 2.5 e na Observação 2.5, esses coeficientes decaem rapidamente. Portanto, o  $\log S_X(\lambda)$  pode ser aproximado usando um número finito de coeficientes, ou seja, daqui em diante  $\ell = 0, 1, 2, \dots, L - 1$ , onde  $L$  é um valor inteiro positivo tal que  $L < N, n$ .

Para a  $j$ -ésima população  $\Pi_j$ , seja  $\hat{\mathbf{c}}_{j\ell}^{(\ell)M} = [\hat{c}_{j1\ell}^M \quad \hat{c}_{j2\ell}^M \quad \dots \quad \hat{c}_{jn_j\ell}^M]^T$  o vetor cepstral para um  $\ell$  fixo da  $j$ -ésima população, centrado na média  $E(\hat{c}_{j\ell}^M \mid \Pi = j) = \mu_{j,\ell}$  e  $\mathbf{c}_{j\ell}^{(k)} =$

$[c_{jk0} \ c_{jk1} \ \cdots \ c_{jkL-1}]^T$  vetor cepstral para uma replicação fixa  $k$ . Os centróides (vetor média) são definidos como  $\boldsymbol{\mu}_{j\ell}^{(k)} = [\mu_{j0} \ \mu_{j1} \ \cdots \ \mu_{j(L-1)}]^T$ ,  $\boldsymbol{\mu}_{j\ell}^{(\ell)} = [\mu_{1\ell} \ \mu_{2\ell} \ \cdots \ \mu_{J\ell}]^T$  sendo, para uma dada população, o vetor médio nas replicações e o vetor médio na população, dado um cepstrum fixo  $\ell$ , respectivamente. Além disso, seja  $\mathbf{f} = [f_1 \ f_2 \ \cdots \ f_J]^T$  o vetor da proporção das séries temporais onde  $f_j = \langle \boldsymbol{t}_n, \boldsymbol{t}_n \rangle^{-1} \langle \boldsymbol{t}_{n_j}, \boldsymbol{t}_{n_j} \rangle$ ,  $\boldsymbol{t}_n$  e  $\boldsymbol{t}_{n_j}$  são vetores de uns com tamanho  $(n \times 1)$  e  $(n_j \times 1)$ , respectivamente,  $\langle \cdot, \cdot \rangle$  é o produto interno no espaço Euclidiano  $R^n$ .

Portanto, seja  $(\hat{\boldsymbol{\mu}}_{j\ell}^M)$  a média dentro do cepstrum,  $(\hat{\boldsymbol{\mu}}^M)$  o centróide da média geral,  $(\hat{\Omega}_W^M)$  a variância dentro do grupo e  $(\hat{\Omega}_B^M)$  a variância entre os grupos, definidos como segue.

**i. Média dentro da classe por cepstrum:**

$$\hat{\boldsymbol{\mu}}_{j\ell}^M = E(\hat{\boldsymbol{\mu}}_{j\ell}^M) = E(\hat{\mathbf{c}}_{jk\ell}^{(\ell)M} \mid \Pi = j) = \frac{1}{n_j} \sum_{k=1}^{n_j} \hat{\mathbf{c}}_{jk\ell}^{(\ell)M} = \langle \boldsymbol{t}_{n_j}, \hat{\mathbf{c}}_{jk\ell}^{(\ell)M} \rangle \langle \boldsymbol{t}_{n_j}, \boldsymbol{t}_{n_j} \rangle^{-1}$$

**ii. Centróide da média geral:**

$$\hat{\boldsymbol{\mu}}^M = E(\hat{\mathbf{c}}_{jk\ell}^{(\ell)M}) = E[E(\hat{\mathbf{c}}_{jk\ell}^{(\ell)M} \mid \Pi = j)] = \left[ \langle \hat{\boldsymbol{\mu}}_{j\ell}^{M(0)}, \hat{\mathbf{f}}^M \rangle \ \langle \hat{\boldsymbol{\mu}}_{j\ell}^{M(1)}, \hat{\mathbf{f}}^M \rangle \cdots \langle \hat{\boldsymbol{\mu}}_{j\ell}^{M(L-1)}, \hat{\mathbf{f}}^M \rangle \right]^T$$

**iii. Variância entre classes:**

$$\hat{\Omega}_B^M = \text{Var}[E(\hat{\mathbf{c}}_{jk\ell}^M \mid \Pi = j)] = \|\hat{\boldsymbol{\mu}}_j^M - \hat{\boldsymbol{\mu}}^M\|^2 \langle \boldsymbol{t}_n, \boldsymbol{t}_n \rangle^{-1}$$

**iv. Variância dentro das classes:**

$$\hat{\Omega}_W^M = E[\text{Var}(\hat{\mathbf{c}}_{jk}^{(k)M} \mid \Pi = j)] = E \left[ (\hat{\mathbf{c}}_{jk}^{(k)M} - \hat{\boldsymbol{\mu}}_j^M)(\hat{\mathbf{c}}_{jk}^{(k)M} - \hat{\boldsymbol{\mu}}_j^M)^T \right]$$

Como  $\hat{\Omega}_W^M$  e  $\hat{\Omega}_B^M$  são matrizes definidas positivas, os coeficientes discriminantes são derivados da seguinte generalização da maximização da forma quadrática na esfera unitária, geralmente chamada de Quociente de Rayleigh Generalizado (Shin, 2008, p.1192).

$$\hat{\mathbf{p}}_1^M = \underset{\hat{\mathbf{p}}^M \in R^L}{\text{argmax}} \left\{ \frac{\text{Var}[E((\hat{\mathbf{p}}^M)^T \hat{\mathbf{c}}_{jk}^M \mid \Pi_j = j)]}{E[\text{Var}((\hat{\mathbf{p}}^M)^T \hat{\mathbf{c}}_{jk}^M \mid \Pi_j = j)]} \right\} = \underset{\hat{\mathbf{p}}^M \in R^L}{\text{argmax}} \left\{ \frac{(\hat{\mathbf{p}}^M)^T \hat{\Omega}_B^M \hat{\mathbf{p}}^M}{(\hat{\mathbf{p}}^M)^T \hat{\Omega}_W^M \hat{\mathbf{p}}^M} \right\}. \quad (2.16)$$

ou equivalentemente

$$\begin{aligned} & \max_{\hat{\mathbf{p}}^M \in \mathbb{L}} \hat{\mathbf{p}}^{MT} \hat{\Omega}_B^M \hat{\mathbf{p}}^M \\ & \text{subject to } (\hat{\mathbf{p}}^M)^T \hat{\Omega}_W^M \hat{\mathbf{p}}^M = 1. \end{aligned} \quad (2.17)$$

o que fornece um dado  $\hat{\mathbf{p}}_1^M = [\hat{p}_{10}^M, \cdots, \hat{p}_{1(L-1)}^M]^T \neq 0$ .

O problema de otimização pode ser alcançado sequencialmente pelo multiplicador de Lagrange

$$\mathcal{L}(\hat{\mathbf{p}}^M) = \hat{\mathbf{p}}^{MT} \hat{\Omega}_B^M \hat{\mathbf{p}}^M - \lambda(\hat{\mathbf{p}}^{MT} \hat{\Omega}_w^M \hat{\mathbf{p}}^M - 1). \quad (2.18)$$

A condição de primeira ordem é dada pelo vetor gradiente igual a zero, ou seja,  $\nabla L(\mathbf{p}) = 0$ , então

$$2\hat{\Omega}_B^M \hat{\mathbf{p}}^M - 2\lambda \hat{\Omega}_w^M \hat{\mathbf{p}}^M = 0 \iff (\hat{\Omega}_B^M - \lambda \hat{\Omega}_w^M) \hat{\mathbf{p}}^M = 0 \iff (\hat{\Omega}_w^{M-1} \hat{\Omega}_B^M - \lambda I) \hat{\mathbf{p}}^M = 0. \quad (2.19)$$

onde  $I$  é a matriz identidade.

A solução da Eq. 2.16 (2.19) fornece o autovetor generalizado  $\mathbf{p}$  veja, por exemplo, (Golub e Loan, 2013, p.497). Com base na Desigualdade de Cauchy-Schwartz Estendida, é fácil mostrar que o primeiro vetor ortonormal, digamos  $\mathbf{p}_1$ , é dado pelo maior autovalor  $\lambda_1$  da matriz  $\Omega_w^{-1} \Omega_B$ . Em seguida, fornece a taxa das duas formas quadráticas onde o numerador considera a matriz de variabilidade entre os grupos ( $\Omega_B$ ) e o denominador corresponde à variabilidade dentro do grupo ( $\Omega_w$ ). Ou seja, as soluções acima também podem ser derivadas de

$$\mathbf{p}_1^* = \operatorname{argmax}_{\mathbf{p} \in R^L} \left\{ \frac{\mathbf{p}^T \Omega_w^{-1} \Omega_B \mathbf{p}}{\mathbf{p}^T \mathbf{p}} \right\}. \quad (2.20)$$

que leva à seguinte proposição.

Sejam  $A$  e  $B$  ambas simétricas definidas positivas ( $L \times L$ ) e  $\mathbf{x}$  e  $\mathbf{y}$  vetores ortonormais ( $L \times 1$ ), então  $\mathbf{x}_1 = \mathbf{y}_1$ , onde

$$\mathbf{x}_1 = \operatorname{argmax}_{\mathbf{x} \in R^L} \left\{ \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T B \mathbf{x}} \right\}.$$

e

$$\mathbf{y}_1 = \operatorname{argmax}_{\mathbf{y} \in R^L} \left\{ \frac{\mathbf{y}^T B^{-1} A \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}.$$

No contexto da equação 2.19, o processo de otimização é uma análise de componente principal de  $\hat{\Omega}_w^{M-1} \hat{\Omega}_B^M$ . Portanto, procuramos uma função discriminante, que é uma combinação linear do cepstrum tal que a variância entre-condicional seja o maior possível, em relação à variância condicional-dentro (Wichern e Johnson, 2007, p.431). Então, a primeira combinação linear (discriminantes) fornece a máxima variância condicional relativa.

Note que o denominador da equação 2.19  $\mathbf{p}^T \mathbf{p}$  é na verdade uma condição que  $\mathbf{p}^T \mathbf{p} = 1$ , ou seja, norma unitária. É uma condição necessária para o vetor  $p$  que será encontrado para a não crescente variância condicional relativa da função discriminante  $Var[E(\mathbf{d}_{jk} | \Pi_j = j)] = Var[E(\mathbf{p}_1^{*T} c_{jk} | \Pi_j = j)] = \mathbf{p}_1^{*T} \Omega_w^{-1} \Omega_B \mathbf{p}_1^*$ .

Note que  $\hat{\lambda}_1$  gerado pelos resultados acima corresponde à maior variabilidade dos hiperesferóides de  $\hat{\Omega}_w^{M-1} \hat{\Omega}_B^M$ , que está relacionado ao eixo gerado por  $\hat{\mathbf{p}}_1^M$ . Na sequência,  $\lambda_2 > \lambda_3 > \dots > \lambda_q$  são os eixos correspondentes gerados pela base ortonormal do espaço  $\hat{\mathbf{p}}_2^M, \dots, \hat{\mathbf{p}}_q^M$ , onde  $q = \min\{J-1, \text{rank}(A), \text{rank}(B)\}$  é a dimensão do subespaço. Assim, o conjunto de vetores  $\hat{\mathbf{p}}_1^M, \hat{\mathbf{p}}_2^M, \dots, \hat{\mathbf{p}}_q^M$  será uma base de um subespaço que melhor separa os centróides de cepstrum uns dos outros em relação à variância dentro do grupo. Então, todos os cepstra replicados devem ser projetados no novo subespaço por uma combinação linear, que fornece as funções discriminantes (coordenadas)

$$\hat{d}_{jkq}^M = \left\langle \hat{\mathbf{p}}_\ell^M, \hat{\mathbf{c}}_{j\mathbf{k}\ell}^{(\mathbf{k})M} \right\rangle = \sum_{\ell=0}^{\infty} \hat{p}_{q\ell}^M \hat{c}_{j\mathbf{k}\ell}^M. \quad (2.21)$$

Para estabelecer os limites discriminantes das elipsóides, ou seja, a área de decisão, é utilizada a função de perda com base na perda de erro quadrático dada por

$$\Psi(\hat{d}_{*1}, \dots, \hat{d}_{*q}; \text{hatmu}_j) = \left[ \left\| \hat{\mathbf{d}} - \hat{\boldsymbol{\mu}}_j \right\|^2 - 2 \log(f_j) \right] = \left[ \sum_{s=1}^q (\hat{d}_{*s} - \hat{\mu}_{js})^2 - 2 \log(\hat{f}_j) \right], \quad (2.22)$$

onde  $\hat{\mathbf{d}} = (\hat{d}_{*1}, \dots, \hat{d}_{*q})^T$  representa o vetor discriminante para a série temporal da população desconhecida

Por exemplo, a Figura 1 ilustra o subespaço formado pelos vetores ótimos, juntamente com a projeção ortogonal de um vetor cepstral. No contexto de duas populações,  $\hat{\mu}_1^M$  e  $\hat{\mu}_2^M$  representam seus respectivos centróides e o raio definido pela variância dentro dos grupos. O cepstra é  $L$  dimensional e será projetado em 2D. A base ótima que melhor separa os centróides, ao mesmo tempo que minimiza a variância dentro dos grupos, é representada por  $\hat{p}_1^M$  e  $\hat{p}_2^M$ . Para projetar o cepstra  $\hat{c}_{jk}^M$  neste subespaço, ele precisa ser uma combinação linear de  $\hat{p}_1^M$  e  $\hat{p}_2^M$ , definida como  $\hat{d}_{jk}^M = \hat{p}_1^M \hat{c}_{jk1}^M + \hat{p}_2^M \hat{c}_{jk2}^M = \hat{d}_{jk1}^M + \hat{d}_{jk2}^M$ .

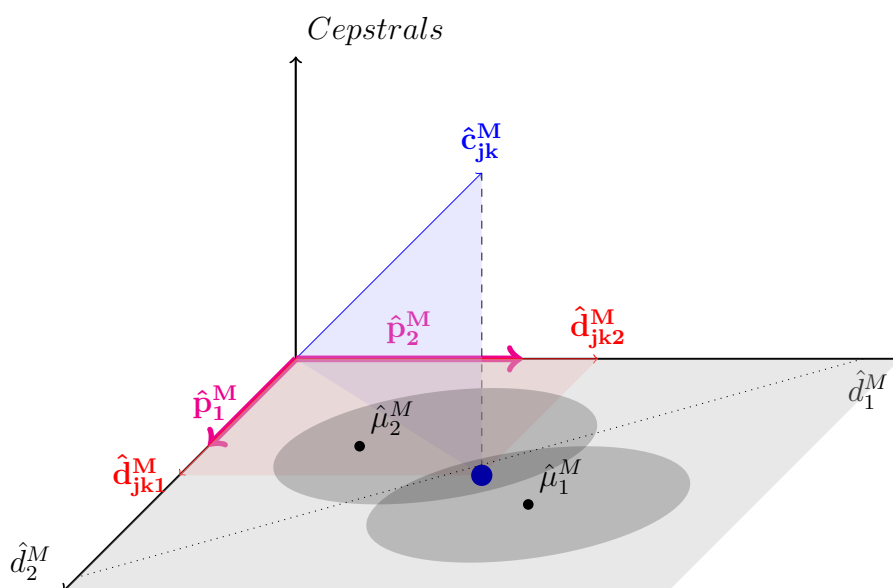


Figura 3.1: LDA projection of Cepstral.

## 3.2 Taxas de Classificação e a Matriz de Confusão

Uma matriz de confusão é uma ferramenta usada para avaliar o desempenho de um modelo de classificação. Ela apresenta um resumo dos resultados da previsão em um problema de classificação. A matriz em si é uma tabela com duas dimensões: "população real" e "população prevista", e é usada para visualizar o desempenho de um algoritmo. No contexto de classificar séries temporais em  $J$  populações usando coeficientes cepstrais de modelos ARMA estacionários, a matriz de confusão fornece *insights* sobre o quão bem o modelo de Análise Discriminante Linear (LDA) está, comparando os rótulos reais das populações das séries temporais com os rótulos previstos.

Tabela 3.1: Matriz de Confusão Normalizada para  $J$  Populações

População Prevista \ População Real	1	2	3	...	J
1	$\rho_{11}$	$\rho_{12}$	$\rho_{13}$	...	$\rho_{1J}$
2	$\rho_{21}$	$\rho_{22}$	$\rho_{23}$	...	$\rho_{2J}$
3	$\rho_{31}$	$\rho_{32}$	$\rho_{33}$	...	$\rho_{3J}$
...	...	...	...	...	...
J	$\rho_{J1}$	$\rho_{J2}$	$\rho_{J3}$	...	$\rho_{JJ}$
total	1	1	1	...	1

Fonte: Elaborado pelo autor.

No contexto desta matriz,  $\rho_{ij}$  representa a proporção das séries temporais da po-

pulação real  $i$  que foram classificadas como pertencentes à população prevista  $j$ . Cada célula  $\rho_{ij}$  é a razão entre o número de séries temporais corretamente ou incorretamente classificadas para a população  $j$  e o número total de séries temporais na população real  $i$ .

A soma de  $\rho_{ij}$  para cada coluna (ou seja, para cada população prevista) é igual a 1, pois todas as séries temporais das populações reais devem ser classificadas em uma das populações previstas, cobrindo todos os possíveis resultados de classificação. Isso reflete a distribuição total das classificações possíveis e garante que todas as previsões sejam contabilizadas, permitindo uma análise precisa do desempenho do modelo de classificação.

Ao usar LDA para classificar dados de séries temporais em  $J$  populações com base nos coeficientes cepstrais de modelos ARMA estacionários, cada série temporal é representada por seus coeficientes cepstrais. O LDA é treinado nos coeficientes cepstrais para aprender as funções discriminantes que separam as populações. O modelo LDA treinado é usado para prever a população de cada série temporal no conjunto de teste. A matriz de confusão é gerada comparando os rótulos reais das populações das séries temporais no conjunto de teste com os rótulos previstos.

As etapas para gerar e usar uma matriz de confusão são as seguintes: primeiro, treine o modelo LDA usando o conjunto de treinamento de dados de séries temporais com rótulos de populações conhecidos. Em seguida, aplique o modelo LDA treinado ao conjunto de teste para obter os rótulos de populações previstos. Compare os rótulos reais com os rótulos previstos para preencher a matriz de confusão. Por fim, identifique quais populações são frequentemente confundidas umas com as outras e ajuste o modelo ou as etapas de pré-processamento para melhorar o desempenho da classificação com base nos *insights* obtidos.

Os benefícios de usar uma matriz de confusão incluem análise detalhada de erros, que ajuda a identificar classes específicas que estão sendo classificadas incorretamente e a entender a natureza das classificações incorretas. Também auxilia na melhoria do modelo, fornecendo insights sobre como ajustar o modelo ou as etapas de pré-processamento para melhorar o desempenho da classificação. Ao usar a matriz de confusão em nossa análise de classificação com LDA, podemos obter uma compreensão profunda de quão bem nosso modelo distingue entre as populações e identificar áreas para melhorias.

## Capítulo 4

**Paper: “Discriminant analysis based on robust cepstral coefficients”**

# Discriminant analysis based on the M-cepstral coefficient estimates

August 2, 2024

## Abstract

Time series analysis is essential in fields such as finance, economics, environmental science, and biomedical engineering for understanding underlying mechanisms, forecasting, and identifying patterns. Traditional time-domain methods, which focus on trends, seasonality, and noise, often overlook periodicities and harmonic structures that are better captured in the frequency domain. Analyzing time series in the frequency domain enables the identification of these spectral properties, providing deeper insights into the underlying processes. These insights can help differentiate data-generating processes of different populations and assist in the discrimination and classification of time series. The literature commonly uses smoothed estimators like the smoothed periodogram to minimize bias, obtaining an average spectrum from individual replicates within a population to classify new time series. However, if there is spectral variability among replicates within each population, such methods become unfeasible. Moreover, abrupt values can significantly impact spectrum estimators, complicating practical discrimination and classification. There is a gap in the literature for methods that consider within-population spectral variability, separate white noise effects from autocorrelations, and use robust estimators in the presence of outliers. This paper addresses this gap by presenting a robust framework for classifying replicate groups of time series by transforming them into the frequency domain using the Fourier Transform to compute the power spectrum. Then, after taking the logarithm of the spectra, the inverse Fourier Transform is used to achieve the cepstrum. To mitigate the effects of outliers and improve the consistency of spectral estimates, we employ the multitaper periodogram alongside the M-periodogram. These spectral features are then utilized in Linear Discriminant Analysis (LDA) to enhance classification accuracy and interpretability. This integrated approach offers significant potential for applications requiring precise temporal pattern distinction and resilience to data anomalies.

## Keywords

Robust, discriminant, classification, cepstral, multitaper, time series, frequency domain.

## 1 Introduction

Time series analysis plays a crucial role in various disciplines such as finance, economics, environmental science, and biomedical engineering. The primary goal of time series analysis is to understand the underlying mechanisms governing the observed data, predict future values, and identify significant patterns. Traditional approaches to time series analysis predominantly

---

\* Federal University of Minas Gerais, ICEX, Brazil. • Federal University of Espirito Santo, PPGEA, Brazil.

operate in the time domain, focusing on modeling trends, seasonality, and noise through techniques such as autoregressive moving average (ARMA), vector autoregression (VAR), and vector error correction models (VECM).

However, these methods are often insufficient when it comes to uncovering the underlying spectral properties of the data. Analyzing time series in the frequency domain provides a complementary perspective, emphasizing periodicities and harmonic structures that are not easily observable in the time domain. To analyze the data in the frequency domain, it should be transformed by the Fourier transform, which decomposes a time series into the amplitude of harmonics for different frequencies, revealing the power distribution across those frequencies.

The foundation of time series analysis was significantly advanced by Box and Jenkins (1970); Box et al. (1994), who introduced a systematic approach for building stationary ARMA models. Their work laid the groundwork for subsequent developments in the field, which have focused on improving model accuracy and computational efficiency Brockwell and Davis (1991); Anderson (1976); Shumway and Stoffer (2011). Recent advancements have also explored the integration of machine learning techniques to enhance predictive performance and adaptability to complex data structures Alpydin (2014); Lazzeri (2020).

Further enhancing the understanding of time series, Brillinger (1981) and Priestley (1981) provided comprehensive treatments of frequency domain methods, highlighting their applicability to a wide range of time series data. The power spectrum, obtained through the quadratic form of the Fourier Transform of a signal, is central to these analyses, offering insights into the dominant frequencies and periodicities within the data. Frequency domain techniques have been particularly effective in fields such as signal processing and light wave, where the identification of spectral features is crucial. Bloomfield (2000) expanded on these foundational works by elaborating on the utility of Fourier analysis for time series data, demonstrating its effectiveness in isolating and analyzing the amplitude of harmonics across different frequencies.

Building on these spectral techniques, the ln spectrum—a lnarithmic transformation of the power spectrum—improves the visualization of spectral features by expressing them as a sum of terms rather than a product, resulted from the quadratic form of the time series in the frequency domain, giving the spectra. This lnarithmic approach has been extensively used in speech processing and image recognition Oppenheim and Schaffer (2004) and geophysical signal analysis Bogert et al. (1963); Shumway (1982); Alagón (1986); Kakizawa et al. (1998); Shumway and Stoffer (2011). The cepstrum, derived by taking the inverse Fourier Transform of the ln spectrum, further decomposes the time series variation into white noise and autocorrelation contributions. This technique has proven valuable for identifying echo patterns and deconvolving complex signals Schaffer (1969); Oppenheim et al. (1968).

Despite the advancements, outlier observations within time series data can significantly distort spectral estimates. Traditional periodograms can be sensitive to these outliers, leading to biased or inconsistent spectral estimates even using smooth periodograms (Kakizawa et al., 1998; Zhang and Taniguchi, 1992; Zhang and Chan, 2005). To mitigate these effects, robust spectral estimation techniques such as the M-periodogram and the multitaper periodogram have been developed. Katkovinik (1998) introduced the M-periodogram as a robust method to reduce the influence of outliers. Reisen et al. (2020b) provided an overview of robust spectral estimators and time series in the context of long memory approach, highlighting the benefits of these approaches in handling outliers and improving spectral estimation. The multitaper periodogram, as described by Thomson (1982), uses multiple orthogonal tapers to produce an averaged spectrum, reducing variance by diminishing the leakage caused by sidelobes and improving spectral estimation accuracy.

To further refine classification and analysis of time series data, Linear Discriminant Analysis (LDA), introduced by Fisher (1936), is a statistical method used for classification and dimensionality reduction. LDA seeks to find a linear combination of features that best sepa-

rates multiple classes. When applied to time series data, LDA can leverage spectral features to enhance discriminative power. Recent studies, such as the one of Shumway and Stoffer (2011) have demonstrated the effectiveness of incorporating frequency domain features into LDA for improved classification of time series data. Krafty (2016) specifically addressed the use of LDA in the presence of within-group spectral variability.

Additionally, robust statistical methods have been developed to improve the performance of LDA, but there is a gap in the literature regarding robust methods in time series analysis, discrimination, and classification. Huber and Ronchetti (2009) introduced robust estimators that can be applied in the context of time series analysis to enhance the reliability of discriminant functions. Building on these foundations, more recent works by Kutz and Brunton (2019) have further refined these techniques, making them more applicable to high-dimensional time series data.

In this context, this paper aims to develop a robust framework for the classification of replicate groups of time series based on their spectral characteristics, using the M-periodogram as a robust estimator and accounting for within-group spectral variability. By integrating advanced spectral analysis techniques with discriminant analysis, this research offers a novel approach to time series classification. The use of robust spectral estimation method ensures that the analysis is resilient to outliers, thereby improving the reliability of the results. The proposed methodology not only improves classification accuracy but also provides deeper insights into the spectral variability inherent in time series data.

Following this introduction, the paper is structured as follows. Section 2 presents the Cepstral Linear Discriminant Analysis (CLDA), detailing the theoretical foundation and methodology for applying cepstral features in discriminant analysis. Section 3 introduces the M-cepstral estimator, describing its formulation and the advantages it offers in robust spectral estimation, particularly in the presence of outliers. In Section 4, Monte Carlo simulations are conducted to evaluate the performance of the proposed methods, providing empirical evidence of their effectiveness and robustness. Section 5 applies the developed techniques to real-world data, specifically focusing on a study of neurodegenerative diseases and gait variability. This section demonstrates the practical utility of our approach in a complex biomedical context. Finally, Section 6 concludes the paper, summarizing the key findings and suggesting potential directions for future research.

## 2 Cepstral Linear Discriminat Analysis - CLDA

Discrimination and classification are multivariate tools used to distinguish objects based on their characteristics. According to (Wichern and Johnson, 2007, p. 573), discrimination is primarily exploratory, aiming to identify the main differences between populations. Meanwhile, (Anderson, 1976, p. 207) describes discrimination as applying algebraic or graphical rules to achieve maximal separation of time series data. On the other hand, classification involves assigning new observations to predefined populations, thereby facilitating their categorization into existing groups. The following section aims to demonstrate the extraction of time series properties for discrimination and classification, while subsequent sections will delve into the optimal procedures for effectively separating time series data.

Let  $\{X_{jkt}\}$ ,  $t \in \mathbb{Z}$ , be a family of process defined in probability space  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathcal{P})$ , such that  $X_{jkt} = \sum_{\tau=-\infty}^{\infty} v_{jk\tau} \epsilon_{jk(t-\tau)}$  where  $\{\epsilon_{jkt}\} \sim iid(0, \sigma^2)$  and ARMA coefficients with  $\sum_{\tau=-\infty}^{\infty} |v_{\tau}| < \infty$ . In the above,  $j = 1, \dots, J$ ,  $k = 1, \dots, n_j$ , where  $J$  and  $n_j$  are fixed values. The spectral density of  $\{X_{jkt}\}$  is given by

$$S_{jk}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_{jk}(\tau) \cos(\lambda\tau), \quad \text{for all } \lambda \in [-\pi, \pi]. \quad (2.1)$$

where  $\gamma_{jk}(\tau)$  is the covariance of the process.  $S_{jk}(\lambda)$  may be interpreted as a decomposition of the variance of the process.

As previously mentioned, this paper is an extension of Krafty (2016), that is, we address the linear discriminate analysis under time series contaminated by additive outliers or with the heavy-tailed distribution. These issues are discussed in the next sections.

## 2.1 Cepstra

Cepstral is a tool for investigating periodic structures in frequency spectra to extract the fundamental components of a signal, in the sense that it separates signals that have been combined in a non-additive way. Oppenheim and Schaffer (2004) gave the following example to clarify the meaning of cepstral. Suppose that a signal of a simple echo can be written as

$$X_t = s_t + \alpha s_{(t-\tau)}. \quad (2.2)$$

where  $\alpha$  is a constant which satisfies conditions to guarantee the stability of the real process  $x_t$ , and  $s_t$  is a white noise process. Note that  $x_t$  corresponds to MA(1) process with white noise process  $s_t$  with lag  $\tau$ . The spectral representation of Equation 2.5 can be written as

$$\|X_t(\lambda)\|^2 = \|S_s(\lambda)\|^2 [1 + \alpha^2 + 2\alpha \cos(\lambda\tau)]. \quad (2.3)$$

$$\ln \|X_t(\lambda)\|^2 = \ln \|S_t(\lambda)\|^2 + \ln[1 + \alpha^2 + 2\alpha \cos(\lambda\tau)]. \quad (2.4)$$

where  $\ln = \ln_e$ .

From Equation 2.3 we can see that the spectral density of the echo corresponds to the product of the spectral of the noise ( $S_s$ ) with the spectrum contribution of the echo. One way to see the individual contribution of echo time-variability is to apply the real ln transformation, as shown in 2.4. From this, we see that the  $\ln \|X_t\|^2$  has a waveform with the periodic component with delay  $\tau$ .

Bogert et al. (1963), in their seminal work, introduce the cepstrum (a noun paraphrased from the word spectrum), which is the inverse Fourier transform of the  $\ln \|X_t\|^2$ . The correspondent coefficients are called cepstral coefficients ( $c_\ell$ ),  $\ell = 0, 1, \dots$ . The cepstrum analysis is in the domain denoted as quefreny. These terminologies were introduced by the authors where cepstral and cepstrum are anagrams of spectral and spectrum, respectively. Apart from the mathematical elegance of the cepstrum and the cepstral coefficients, they display interesting applications in various areas of knowledge. The cepstrum is a tool for investigating periodic structures in frequency spectra, with main applications in human speech, music and electric power systems. The periodical structures are related to noticeable echos in the signal, or to the occurrence of harmonic frequencies. Mathematically it deals with the problem of deconvolution of signals in the frequency space. The typical definition for a harmonic is “a sinusoidal component of a periodic wave or quantity having a frequency that is an integral multiple of the fundamental frequency”. Some references refer to “clean” or “pure” power as those waveform without harmonics (IEE (2014)). See a review in Cavicchioli (2020).

The cepstrum power has also been to discriminate and classification spectral densities, that is, a discriminate analysis based on the frequency and quefreny domains. Fokianos and Savvides (2008) used the methodology for testing the similarity of G spectral density functions

from G-independent stationary processes. Krafky (2016) introduces the cepstral coefficients into the Mahalanobis distance to build a discriminant function in the frequency and quefrequency domains, considering variability between and within groups.

Since we are considering the discriminant problem based on stationary zero-mean ARMA models, we derive cepstral coefficients for some examples of this class process below.

**Proposition 1.** *Let  $X_t$  be a stationary ARMA( $p, q$ ) process and  $\sigma^2$  is the innovation variance. Additionally, let  $z = e^{-i\lambda}$ ,  $\eta_i$  and  $\zeta_r$  be the  $i$ th and  $r$ th roots of polynomials  $\Theta(z)$  and  $\Phi(z)$ , respectively (Hamilton, 1994, p. 155). Then:*

*i. The  $\ln$  spectra of  $X_t$  can be written as:*

$$\begin{aligned}
& \text{bb} \tag{2.5} \\
& = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \sum_{i=1}^q \ln [1 + \eta_i^2 + 2\eta_i \cos(\lambda)] - \sum_{r=1}^p \ln [1 + \zeta_r^2 + 2\zeta_r \cos(\lambda)] \\
& = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \left\{ \sum_{i=1}^q \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} \cos(\lambda\ell) + \sum_{r=1}^p \sum_{\ell=1}^{\infty} \frac{\zeta_r^\ell}{\ell} \cos(\lambda\ell) \right\}. \\
& = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \left\{ \sum_{i=1}^q \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} \cos(\lambda\ell) + \sum_{r=1}^p \frac{\zeta_r^\ell}{\ell} \cos(\lambda\ell) \right\}.
\end{aligned}$$

*ii. Using the  $\ln$  spectra, the cepstra coefficients for ARMA( $p, q$ ) can be written as:*

$$c_\ell = \begin{cases} \ln\left(\frac{\sigma_\epsilon^2}{2\pi}\right) & , \text{if } \ell = 0 \\ 2 \left( \sum_{i=1}^q \frac{(-1)^{\ell+1} \eta_i^\ell}{\ell} + \sum_{r=1}^p \frac{\zeta_r^\ell}{\ell} \right) & , \text{if } \ell \geq 1. \end{cases} \tag{2.6}$$

In particular, Corollaries 1, 2, and 3 display the  $\ln$  spectra and cepstral coefficients for AR(1), MA(1), and ARMA(1, 1). The complete demonstrations are presented in Appendices 7.1 and 7.2. Also, considering the  $\ln = \log_e$ , where  $e$  is the Neperian number.

**Corollary 1.** *Let  $X_t = \phi X_{(t-1)} + \epsilon_t$  be an stationary Gaussian AR(1) process. Then:*

*i. The  $\ln$  spectra can be written as:*

$$\begin{aligned}
\ln S_X(\lambda) & = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} [1 + \phi^2 - 2\phi \cos(\lambda)]^{-1} \right\} \tag{2.7} \\
& = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} - \ln [1 + \phi^2 - 2\phi \cos(\lambda)] \\
& = \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \frac{\phi^\ell}{\ell} \cos(\lambda\ell).
\end{aligned}$$

ii. Using the  $\ln$  spectra, the cepstra coefficients for AR(1) can be written as:

$$c_\ell = \begin{cases} \ln\left(\frac{\sigma^2}{2\pi}\right) & ,if \quad \ell = 0 \\ \frac{2\theta^\ell}{\ell} & ,if \quad \ell \geq 1. \end{cases} \quad (2.8)$$

Figures 1 and 2 display spectra (a),  $\ln$  spectra (b), and cepstra (c) for the particular case of AR(1) with  $\phi = 0.5$  and  $\phi = -0.5$ , respectively. In Figure 1, the spectra show that most of the variability of the process is explained by low frequencies, close to zero. This indicates a time series with relatively low volatility, which is confirmed by the cepstra. In panel (c), the cepstra decrease more rapidly than an exponential function toward to zero. This implies that approximately  $\ell = 8$  is sufficient to explain most of the variability of the process.

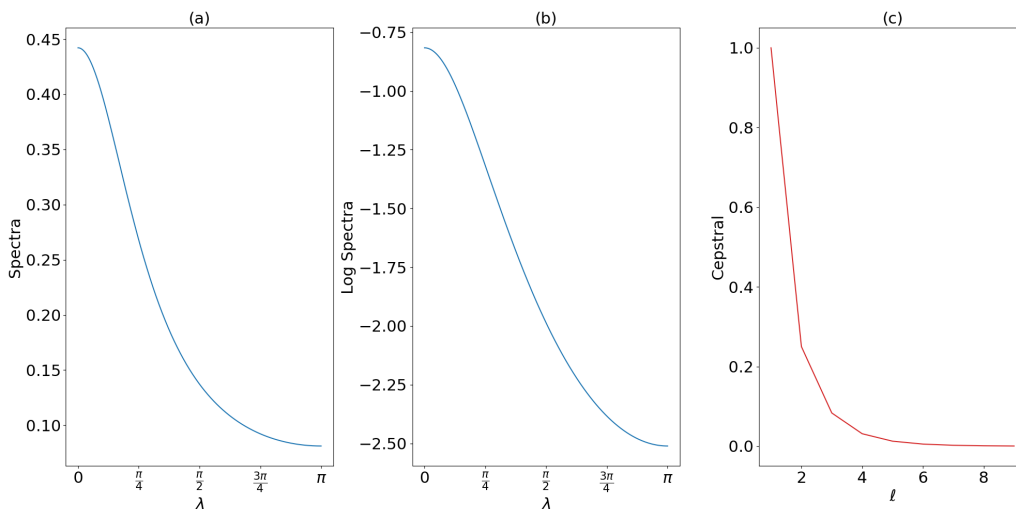


Figure 1:  $\phi = 0.5$  and  $\sigma_\epsilon^2 = 1$ : (a) Spectra, (b)  $\ln$  Spectra, (c) Cepstra

On the other hand, it is possible to see in Figure 2 that with  $\phi < 0$ , the majority of the variability of the process is explained by high frequencies, close to  $\pi$ . As a result, the series is more volatile, and the cepstra exhibit behavior akin to a periodic function oscillating around zero. Additionally, these characteristics show that only about  $\ell = 6$  is sufficient to explain the majority of the variability of the process.

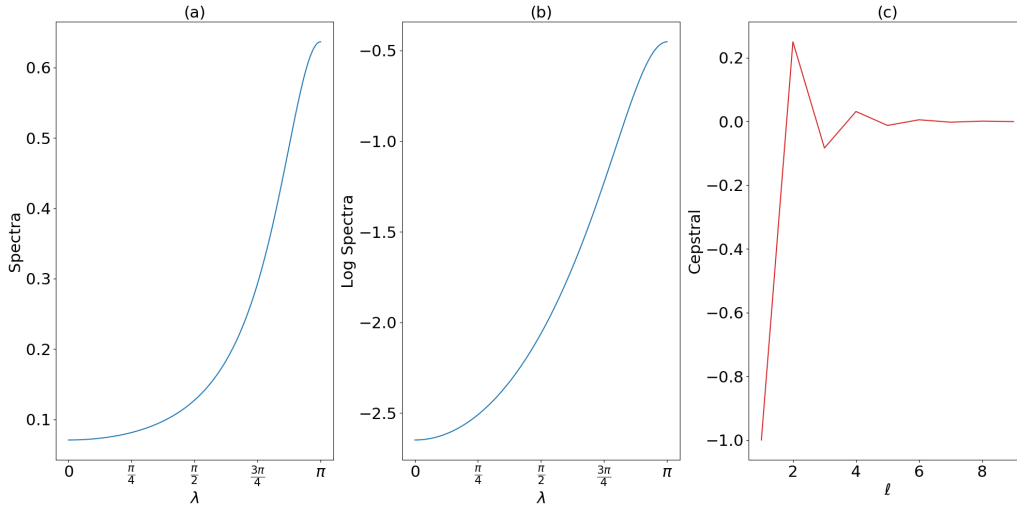


Figure 2:  $\phi = -0.5$  and  $\sigma_\epsilon^2 = 1$ : (a) Spectra, (b) ln Spectra, (c) Cepstra

As can be noted in Corollary 2, the behavior of cepstral coefficients for an MA(1) process differs from that of the AR(1) process presented in Corollary 1 only by a change in sign for  $\ell \geq 1$  and the parameter  $\theta$ .

**Corollary 2.** *Let  $X_t = \epsilon_t + \theta\epsilon_{t-1}$  be an invertible Gaussian MA(1) process. Then:*

*i. The ln spectra can be written as:*

$$\begin{aligned}
 \ln[S_X(\lambda)] &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} [1 + \theta^2 + 2\theta \cos(\lambda)] \right\} \\
 &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \ln [1 + \theta^2 + 2\theta \cos(\lambda)] \\
 &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \theta^\ell}{\ell} \cos(\lambda \ell).
 \end{aligned} \tag{2.9}$$

*ii. Using the ln spectra, it will result in cepstra coefficients for MA(1), given by:*

$$c_\ell = \begin{cases} \ln\left(\frac{\sigma_\epsilon^2}{2\pi}\right) & , \text{if } \ell = 0 \\ \frac{(-1)^{\ell+1} 2\theta^\ell}{\ell} & , \text{if } \ell \geq 1. \end{cases} \tag{2.10}$$

Similarly, in Figures 3 and 4, panel (a) shows low and high frequencies, respectively. Thus, when  $\theta > 0$ , the process exhibits behavior similar to that of  $\phi > 0$ , while  $\theta < 0$  is akin to  $\phi < 0$ . However, as can be noted in panel (c), the cepstra exhibit opposite behavior, indicating that low frequency in the MA(1) process with  $\theta = 0.5$  is associated with a periodic function in the cepstra. This is a result of the MA(1) spectra having a positive sign in the cosine term, whereas the AR(1) process has a negative one.

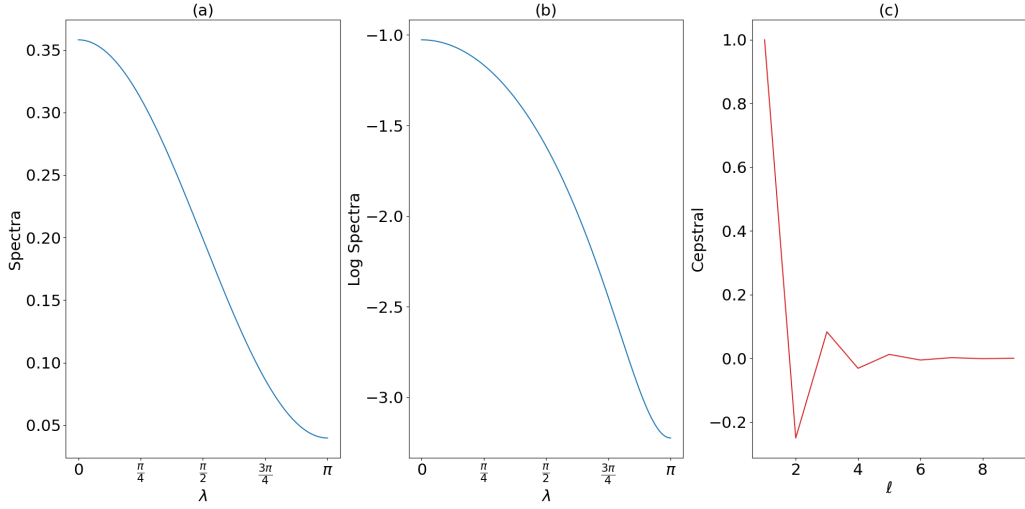


Figure 3:  $\theta = 0.5$  and  $\sigma_\epsilon^2 = 1$ : (a) Spectra, (b) ln Spectra, (c) Cepstra

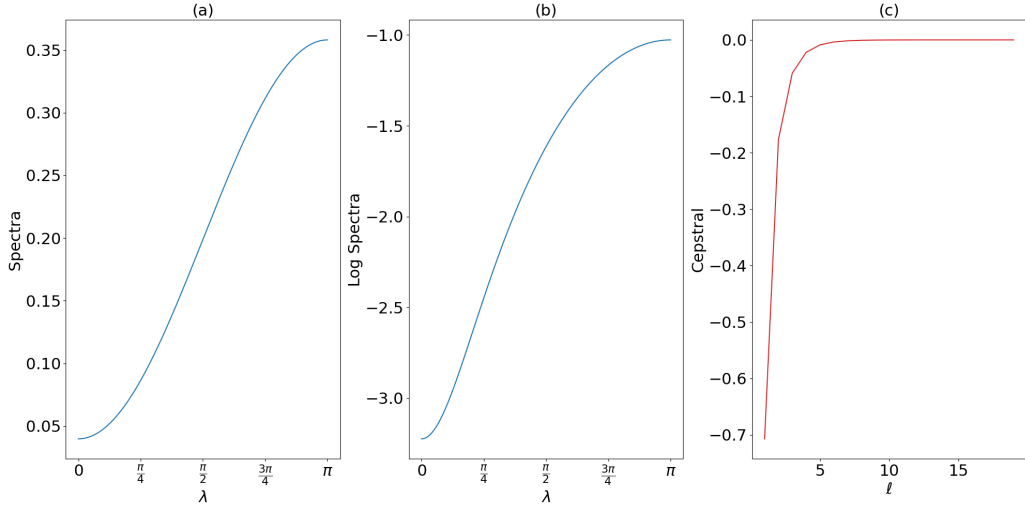


Figure 4:  $\theta = -0.5$  and  $\sigma_\epsilon^2 = 1$ : (a) Spectra, (b) ln Spectra, (c) Cepstra

**Corollary 3.** Let  $X_t = \phi X_{(t-1)} + \epsilon_t + \theta \epsilon_{t-1}$  be an invertible stationary Gaussian AR(1,1) process. Then:

i. The ln spectra can be written as:

$$\begin{aligned}
 \ln S_X(\lambda) &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \left[ \frac{1 + \theta^2 + 2\theta \cos(\lambda)}{1 + \phi^2 - 2\phi \cos(\lambda)} \right] \right\} \quad (2.11) \\
 &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + \ln [1 + \theta^2 + 2\theta \cos(\lambda)] - \ln [1 + \phi^2 - 2\phi \cos(\lambda)] \\
 &= \ln \left\{ \frac{\sigma_\epsilon^2}{2\pi} \right\} + 2 \left\{ \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \theta^\ell}{\ell} \cos(\lambda \ell) + \sum_{\ell=1}^{\infty} \frac{\phi^\ell}{\ell} \cos(\lambda \ell) \right\}.
 \end{aligned}$$

ii. Using the  $\ln$  spectra, it will result in cepstra coefficients for ARMA(1,1), given by:

$$c_\ell = \begin{cases} \ln\left(\frac{\sigma^2}{2\pi}\right) & , \text{if } \ell = 0 \\ \frac{2\theta^\ell}{\ell} + \frac{(-1)^{\ell+1}2\theta^\ell}{\ell} & , \text{if } \ell \geq 1. \end{cases} \quad (2.12)$$

**Corollary 4.** Under the Assumption of a stationary ARMA model, the cepstral coefficients are of order  $c_\ell = \mathcal{O}(\ell^{-3/2}) = o(\ell^{-1/2})$  as  $\ell \rightarrow \infty$ .

**Remark 1.** Note that the cepstral coefficient decays at the rate  $c_\ell \sim \vartheta|\frac{\delta^\ell}{\ell}|$ ,  $-\infty < \ell < \infty$ , where  $\vartheta$  is a constant and  $|\delta| < 1$ , while the autocorrelation function of the standard Box-Jenkins ARMA models decays approximately in the exponential form. For example, the ACF of the AR(1) model is  $\phi^l$ . The mathematical different behaviour between the cepstrum coefficients and the ACF values clearly show that the former displays most of the time-variation of the process with a smaller lag than the ACF function, indicating that most information is contained in lower-order cepstral coefficients.

Figures 1, 2, 3 and 4 display the behaviour of spectra,  $\ln$  spectra and cepstra of MA(1) and AR(1) models, respectively. In both cases, we see the frequencies near zero, which contribute most to the process variance. The cepstrum shows that the initial lags contain all the information necessary to extract the features of the processes. For positive  $\phi$ , the series is positively correlated the spectrum and cepstrum are dominated by low frequency and small lags, respectively, which means that the series is relatively smoothed. When  $\phi < 0$ , the series is negatively correlated, thus the spectrum and cepstrum are dominated by high-frequencies and lags, respectively. This means that the data is more ragged series.

## 2.2 Cepstral Linear Discriminant Analysis

Let  $\mathbf{c}_{j\mathbf{k}\ell}^{(\ell)} = [c_{j1\ell} \ c_{j2\ell} \ \cdots \ c_{jn_j\ell}]^T$  be the cepstral vector for a fixed  $\ell$  of the  $j$ th population  $\Pi_j$ , which is centered in the mean  $\mathbb{E}(c_{jk\ell} \mid \Pi = j) = \mu_{j,l}$  and  $\mathbf{c}_{j\mathbf{k}\ell}^{(\mathbf{k})} = [c_{jk0} \ c_{jk1} \ \cdots \ c_{jk\ell}]^T$  be the cepstral vector for a fixed replication  $k$ . The centroids (the mean vector) are defined as  $\boldsymbol{\mu}_{j\ell}^{(\mathbf{k})} = [\mu_{j0} \ \mu_{j1} \ \cdots \ \mu_{j\ell}]^T$ ,  $\boldsymbol{\mu}_{j\ell}^{(\ell)} = [\mu_{1\ell} \ \mu_{2\ell} \ \cdots \ \mu_{J1\ell}]^T$  being for a given population, the vector mean across replicates and the vector mean across population, given a fixed cepstra  $\ell$ , respectively.

Therefore, let  $(\mu_{j\ell})$  be the within-mean by cepstra,  $(\boldsymbol{\mu})$  be the overall mean centroid,  $(\Omega_W)$  be the within-variance and  $(\Omega_B)$  be the between-variance. Moreover, the operator  $\langle \cdot, \cdot \rangle$  is the inner product in the Euclidean space  $\mathbb{R}^n$ . Then, defined as follows:

i. **Within-class mean by cepstra:**

$$\mu_{j\ell} = \mathbb{E}(\mu_{jk\ell}) = \mathbb{E}(\mathbf{c}_{j\mathbf{k}\ell}^{(\ell)} \mid \Pi = j). \quad (2.13)$$

ii. **Overall mean centroid:**

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{c}_{j\mathbf{k}\ell}^{(\ell)}) = \mathbb{E}[\mathbb{E}(\mathbf{c}_{j\mathbf{k}\ell}^{(\ell)} \mid \Pi = j)] = \left[ \left\langle \boldsymbol{\mu}_{j\ell}^{(0)}, \mathbf{f} \right\rangle \ \left\langle \boldsymbol{\mu}_{j\ell}^{(1)}, \mathbf{f} \right\rangle \ \cdots \ \left\langle \boldsymbol{\mu}_{j\ell}^{(L-1)}, \mathbf{f} \right\rangle \right]^T. \quad (2.14)$$

iii. **Between-class variance:**

$$\Omega_B = \text{Var}[\mathbb{E}(\mathbf{c}_{j\ell} \mid \Pi = j)] = \|\mu_j - \boldsymbol{\mu}\|^2 \langle \boldsymbol{\iota}_{\mathbf{n}}, \boldsymbol{\iota}_{\mathbf{n}} \rangle^{-1} = d^2(\mu_j, \boldsymbol{\mu}) \langle \boldsymbol{\iota}_{\mathbf{n}}, \boldsymbol{\iota}_{\mathbf{n}} \rangle^{-1}. \quad (2.15)$$

iv. **Within-class variance:**

$$\Omega_W = \mathbb{E}[\text{Var}(\mathbf{c}_{jk}^{(k)} \mid \Pi = j)] = \mathbb{E} \left[ (\mathbf{c}_{jk}^{(k)} - \mu_j)(\mathbf{c}_{jk}^{(k)} - \mu_j)^T \mid \Pi = j \right]. \quad (2.16)$$

Since  $\Omega_W$  and  $\Omega_B$  are positive definite matrices, the discriminant coefficients are derived from the following generalization of the maximization of the quadratic form on the unit sphere, usually called Generalized Rayleigh Quotient (Shin, 2008, p. 1192)

$$\mathbf{p}_1 = \underset{\mathbf{p} \in \mathbb{R}^L}{\text{argmax}} \left\{ \frac{\text{Var}[\mathbb{E}(\mathbf{p}^T \mathbf{c}_{jk} \mid \Pi = j)]}{\mathbb{E}[\text{Var}(\mathbf{p}^T \mathbf{c}_{jk} \mid \Pi = j)]} \right\} = \underset{\mathbf{p} \in \mathbb{R}^L}{\text{argmax}} \left\{ \frac{\mathbf{p}^T \Omega_B \mathbf{p}}{\mathbf{p}^T \Omega_w \mathbf{p}} \right\}. \quad (2.17)$$

or equivalently

$$\begin{aligned} & \max_{\mathbf{p} \in \mathbb{R}^L} \mathbf{p}^T \Omega_B \mathbf{p} \\ & \text{subject to } \mathbf{p}^T \Omega_w \mathbf{p} = 1. \end{aligned} \quad (2.18)$$

which provide a given  $\mathbf{p}_1 = [p_{10}, \dots, p_{1(L-1)}]^T \neq 0$ .

The optimization problem can be achieved sequentially by Lagrange multiplier

$$\mathcal{L}(\mathbf{p}) = \mathbf{p}^T \Omega_B \mathbf{p} - \lambda(\mathbf{p}^T \Omega_w \mathbf{p} - 1). \quad (2.19)$$

The first order condition is given by the gradient vector equals to zero, i.e.,  $\nabla \mathcal{L}(\mathbf{p}) = 0$ , then

$$2\Omega_B \mathbf{p} - 2\lambda \Omega_w \mathbf{p} = 0 \iff (\Omega_B - \lambda \Omega_w) \mathbf{p} = 0 \iff (\Omega_w^{-1} \Omega_B - \lambda I) \mathbf{p} = 0. \quad (2.20)$$

where  $I$  is the identity matrix.

The solution of Equation 2.18 (2.20) gives the generalized eigenvector  $\mathbf{p}$  see, for example, (Golub and Loan, 2013, p. 497). Based on the Extended Cauchy-Schwartz Inequality, it is easy to show that the first orthonormal vector, say  $\mathbf{p}_1$ , is given from the largest eigenvalue  $\lambda_1$  of the matrix  $\Omega_w^{-1} \Omega_B$ . Then, it provides the rate of the two quadratic forms where the numerator takes in account the variability matrix between groups ( $\Omega_B$ ) and the denominator corresponds to variability within group ( $\Omega_w$ ). That is, the above solutions can also be derived from

$$\mathbf{p}_1^* = \underset{\mathbf{p} \in \mathbb{R}^L}{\text{argmax}} \left\{ \frac{\mathbf{p}^T \Omega_w^{-1} \Omega_B \mathbf{p}}{\mathbf{p}^T \mathbf{p}} \right\}. \quad (2.21)$$

which leads to the following proposition.

**Proposition 2.** *Let  $A$  and  $B$  both be symmetric positive definite ( $L \times L$ ) and  $\mathbf{x}$  and  $\mathbf{y}$  are orthonormal ( $L \times 1$ ) vectors, then  $\mathbf{x}_1 = \mathbf{y}_1$ , where*

$$\mathbf{x}_1 = \underset{\mathbf{x} \in \mathbb{R}^L}{\text{argmax}} \left\{ \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T B \mathbf{x}} \right\}.$$

and

$$\mathbf{y}_1 = \underset{\mathbf{y} \in \mathbb{R}^L}{\text{argmax}} \left\{ \frac{\mathbf{y}^T B^{-1} A \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}.$$

In the context of Equation 2.19, the optimization process is a component principal analysis of  $\Omega_w^{-1}\Omega_B$ . Therefore, we seek for the discriminant function, which is a linear combination of the cepstral such that the conditional variance-between is as large as possible, relative to the conditional variance-within (Wichern and Johnson, 2007, p. 431). Then, the first linear combination (discriminants) gives the maximum relative conditional variance.

**Remark 2.** Note that the denominator of the Equation 2.19  $\mathbf{p}^T\mathbf{p}$  is actually a condition that  $\mathbf{p}^T\mathbf{p} = 1$ , i.e., unitary norm. It is a necessary condition to the vector  $p$  that is going to be found for non increasing relative conditional variance of the discriminant function  $\text{Var}[\mathbb{E}(\mathbf{d}_{jk}|\Pi_j = j)] = \text{Var}[\mathbb{E}(\mathbf{p}_1^{*T}c_{jk}|\Pi_j = j)] = \mathbf{p}_1^{*T}\Omega_w^{-1}\Omega_B\mathbf{p}_1^*$ .

Note that  $\lambda_1$  generated by the above results corresponds to the largest variability of the hyperellipsoids of  $\Omega_w^{-1}\Omega_B$ , in which is related to the axis generated by  $\mathbf{p}_1$ . In the sequence,  $\lambda_2 > \lambda_3 > \dots > \lambda_q$  are the correspondent axis generated by the orthonormal basis space  $\mathbf{p}_2, \dots, \mathbf{p}_q$ , where  $q = \min\{J - 1, \text{rank}(A), \text{rank}(B)\}$  is the dimension of the subspace. Thus, the set of vectors  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_q$  will be a basis of a subspace that best separates cepstra centroids from each other related to within-variance. Then, all replicates cepstra should be projected into the new subspace by a linear combination, which gives the discriminant functions (coordinates)

$$d_{jkq} = \left\langle \mathbf{p}_q, \mathbf{c}_{j\mathbf{k}q}^{(\mathbf{k})} \right\rangle = \sum_{\ell=0}^{\infty} p_{q\ell} c_{j\mathbf{k}q\ell}. \quad (2.22)$$

### 2.2.1 Decision Rules

To establish the discriminant boundaries of the hyperellipsoids, that is, the decision area, this is computed using the loss (in making a decision) function based on square error loss given by

$$F(d_{*1}, \dots, d_{*q}; \mu_j) = \left[ \|\mathbf{d} - \boldsymbol{\mu}_j\|^2 - 2 \ln(f_j) \right] = \left[ \sum_{s=1}^q (d_{*s} - \mu_{js})^2 - 2 \ln(f_j) \right]. \quad (2.23)$$

where  $\mathbf{d} = (d_{*1}, \dots, d_{*q})^T$  stands as the discriminant vector for the unknown population time series, and  $2 \ln(f_j)$  is a correction factor for the number of time series in population  $j = 1, \dots, J$ .

Once each replicated time series is classified, the classification rate can be conducted by each population and can be represented in a confusion matrix, which is a tool used to evaluate the performance of a classification model. It presents a summary of the prediction results on a classification problem. The matrix itself is a Table with two dimensions: "population" and "predicted population," and it is used to visualize the performance of the algorithm.

In the context of classifying time series into  $J$  populations using cepstral coefficients from stationary ARMA models, the confusion matrix provides insights into how well your Linear Discriminant Analysis (LDA) model is performing by comparing the actual population labels of the time series against the predicted labels.

Table 1: Normalized Confusion Matrix for  $J$  Populations

Predicted Population \ Actual Population	1	2	3	...	J
	1	$\rho_{11}$	$\rho_{12}$	$\rho_{13}$	...
2	$\rho_{21}$	$\rho_{22}$	$\rho_{23}$	...	$\rho_{2J}$
3	$\rho_{31}$	$\rho_{32}$	$\rho_{33}$	...	$\rho_{3J}$
...	...	...	...	...	...
J	$\rho_{J1}$	$\rho_{J2}$	$\rho_{J3}$	...	$\rho_{JJ}$
total	1	1	1	...	1

In this matrix,  $\rho_{ij}$  represents the proportion of time series from the actual population  $i$  that have been classified as belonging to the predicted population  $j$ . Each cell  $\rho_{ij}$  is the ratio of the number of time series correctly or incorrectly classified into population  $j$  to the total number of time series in the actual population  $i$ .

The sum of  $\rho_{ij}$  for each column (i.e., for each predicted population) equals 1 because all time series from the actual populations must be classified into one of the predicted populations, covering all possible classification outcomes. This reflects the total distribution of possible classifications and ensures that all predictions are accounted for, allowing for an accurate analysis of the classification model's performance.

When using LDA to classify time series data into  $J$  populations based on cepstral coefficients from stationary ARMA models, each time series is represented by its cepstral coefficients. The LDA is trained on the cepstral coefficients to learn the discriminant functions that separate the populations. The trained LDA model is used to predict the population for each time series in the test set. The confusion matrix is generated by comparing the actual population labels of the time series in the test set with the predicted labels.

The steps to generate and use a confusion matrix are as follows: First, train the LDA model using the training set of time series data with known population labels. Then, apply the trained LDA model to the test set to obtain predicted population labels. Compare the actual labels with the predicted labels to populate the confusion matrix. Finally, identify which populations are often confused with each other and adjust the model or preprocessing steps to improve classification performance based on the insights obtained.

The benefits of using a confusion matrix include detailed error analysis, which helps identify specific classes that are being misclassified and understand the nature of the misclassifications. It also aids in model improvement by providing insights into how to adjust the model or preprocessing steps to enhance classification performance. By using the confusion matrix in our classification analysis with LDA, we can gain a deep understanding of how well our model distinguishes between the populations and identify areas for improvement.

### 3 Estimation Methods

Let  $\Pi_j$  be the  $j$ th population under the study and  $\{X_{jk0}, \dots, X_{jk(N-1)}\}$  the  $k$ th realization of the process  $\{X_{jkt}\}$  in the population  $j = 1, \dots, J$ . The total number of realizations corresponds to  $n = \sum_{j=1}^J n_j$ . The proportion of time series in each population is  $f_j = n_j/n = \langle \mathbf{1}_n, \mathbf{1}_n \rangle^{-1} \langle \mathbf{1}_{n_j}, \mathbf{1}_{n_j} \rangle = \langle \mathbf{1}_n, \mathbf{1}_n \rangle^{-1} \langle \mathbf{1}_{n_j}, \mathbf{1}_{n_j} \rangle$ , whose vector is given by  $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_J]^T$ , the vectors  $\mathbf{1}_n$  and  $\mathbf{1}_{n_j}$  are vectors of ones with size  $(n \times 1)$  and  $(n_j \times 1)$ , respectively.

As stated previously, the cepstral estimated coefficients are the result of the inverse Fourier transform of the logarithm of the spectra, i.e., the inverse Fourier transform of  $\ln(S_X(\lambda))$  for a fixed  $\lambda$ . The logarithm is important to decompose the product caused by the convolution to

generate the spectra, separating the contribution of the white noise and the autocorrelations to the total variance of the process into sum terms. Since all processes are considered to be stationary, the Fourier transform needed is just using the cosine series  $\cos(\lambda\ell)$ ,  $\ell = 1, 2, \dots$ , generating the general cepstral coefficients given by

$$\hat{c}_{jk\ell} = \begin{cases} N^{-1} \sum_{m=0}^{N-1} \ln \left( \hat{S}_{jk}(\lambda_m) \right) & , if \quad \ell = 0 \\ N^{-1} \sum_{m=0}^{N-1} \ln \left( \hat{S}_{jk}(\lambda_m) \right) \cos(\lambda_m \ell) & , if \quad \ell \geq 1. \end{cases} \quad (3.1)$$

where  $\hat{S}_{jk}(\lambda_m)$  is an estimator of the spectra for the  $k$ th replicates of the population  $j$  and  $\lambda_m = 2\pi m/N$ ,  $m \in \mathbb{Z}$ . Any estimator of spectra is appropriate to estimate cepstral coefficients described in Equation 2.6, once for a fixed  $\ell$ ,  $\text{Var}(\hat{c}_{jk\ell} - c_{jk\ell}) \rightarrow 0$  when  $N \rightarrow \infty$ .

Based on Corollary 4 and Remark 1, these coefficients decay fast for ARMA( $p, q$ ) process. Therefore, the  $\ln(S_X(\lambda))$  can be approximated by using finite numbers fixed of the coefficients, that is, hereafter  $\ell = 0, 1, 2, \dots, L-1$ , where  $L$  is an integer positive value such that  $L < N, n$ .

Once the cepstral coefficients are estimated, for each  $j$ th population  $\Pi_j$ , the cepstral vectors can be estimated by  $\hat{\mathbf{c}}_{jk\ell}^{(\ell)} = [\hat{c}_{j1\ell} \quad \hat{c}_{j2\ell} \quad \dots \quad \hat{c}_{jn_j(L-1)\ell}]^T$  for each fixed  $\ell = 0, \dots, (L-1)$  vector centered in the mean  $\mathbb{E}(\hat{c}_{jk\ell} \mid \Pi = j) = \mu_{j,\ell}$  and  $\hat{\mathbf{c}}_{jk\ell}^{(k)} = [\hat{c}_{jk0} \quad \hat{c}_{jk1} \quad \dots \quad \hat{c}_{jk(L-1)}]^T$  cepstral vector for a fixed replication  $k$ . The centroids (mean vector) are defined as  $\hat{\boldsymbol{\mu}}_{j\ell}^{(k)} = [\hat{\mu}_{j0} \quad \hat{\mu}_{j1} \quad \dots \quad \hat{\mu}_{j(L-1)}]^T$ ,  $\hat{\boldsymbol{\mu}}_{j\ell}^{(\ell)} = [\hat{\mu}_{1\ell} \quad \hat{\mu}_{2\ell} \quad \dots \quad \hat{\mu}_{J1\ell}]^T$  being for a given population, the vector mean across replicates and the vector mean across population, given a fixed cepstra  $\ell$ , respectively. Then, all position and scale measures for estimated cepstral can be defined as:

i. **Within-class mean by cepstra:**

$$\hat{\mu}_{j\ell} = \frac{1}{n_j} \sum_{k=1}^{n_j} \hat{c}_{jk\ell}^{(\ell)} = \langle \mathbf{t}_{\mathbf{n}_j}, \hat{\mathbf{c}}_{jk\ell}^{(\ell)} \rangle \langle \mathbf{t}_{\mathbf{n}_j}, \mathbf{t}_{\mathbf{n}_j} \rangle^{-1}. \quad (3.2)$$

ii. **Overall mean centroid:**

$$\hat{\boldsymbol{\mu}} = \left[ \langle \boldsymbol{\mu}_{j\ell}^{(0)}, \mathbf{f} \rangle \quad \langle \boldsymbol{\mu}_{j\ell}^{(1)}, \mathbf{f} \rangle \dots \langle \boldsymbol{\mu}_{j\ell}^{(L-1)}, \mathbf{f} \rangle \right]^T. \quad (3.3)$$

iii. **Between-class variance:**

$$\hat{\Omega}_B = \|\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}\|^2 \langle \mathbf{t}_{\mathbf{n}}, \mathbf{t}_{\mathbf{n}} \rangle^{-1} = d^2(\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\mu}}) \langle \mathbf{t}_{\mathbf{n}}, \mathbf{t}_{\mathbf{n}} \rangle^{-1}. \quad (3.4)$$

iv. **Within-class variance:**

$$\hat{\Omega}_W = (n_j)^{-1} \sum_{k=1}^{n_j} \left[ (\hat{\mathbf{c}}_{jk}^{(k)} - \hat{\boldsymbol{\mu}}_j)(\hat{\mathbf{c}}_{jk}^{(k)} - \hat{\boldsymbol{\mu}}_j)^T \right]. \quad (3.5)$$

Once the variance of the cepstral coefficients given the population is calculated, it should have a minimum superior cote. Let it be defined as follows

$$\hat{\sigma}_L = \sup \left\{ E[(\hat{c}_{jk(L-1)} - \hat{\mu}_j)(\hat{c}_{jk(L-1)} - \hat{\mu}_j)^T] \right\} \quad (3.6)$$

Once the cepstral coefficients are calculated, the discriminant function can also be estimated as

$$\hat{d}_{jks} = \left\langle \hat{\mathbf{p}}_\ell, \hat{\mathbf{c}}_{jkl}^{(\mathbf{k})} \right\rangle = \sum_{\ell=0}^{L-1} \hat{p}_{s\ell} \hat{c}_{jkl}. \quad (3.7)$$

Then, the decision rule in Equation Equation 2.23 can be estimated  $\hat{F}(\hat{d}_{*1}, \dots, \hat{d}_{*q}; \hat{\mu}_j)$ , as well as the confusion matrix described in Table 1.

The estimation methods based on the periodograms will be discussed in next sections.

### 3.1 Cepstral Estimation based on Classical Periodogram

A natural estimator of the spectral density 2.1 is the periodogram function

$$\begin{aligned} I_{jk}(\lambda_m) &= \frac{1}{2\pi N} \left\| \sum_{t=0}^{N-1} X_{jkt} e^{-i\lambda_m t} \right\|^2 \\ &= \frac{1}{2\pi} \sum_{\nu=-(N-1)}^{N-1} \hat{\gamma}_{jk}(\tau) e^{-i\lambda_m \nu} \end{aligned} \quad (3.8)$$

where

$$\hat{\gamma}_{jk}(\tau) = \frac{1}{N} \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} \quad (3.9)$$

is the sample autocovariance function,  $\lambda_m = 2\pi m/N$ ,  $m = 1, \dots, \lfloor N/2 \rfloor$ , are the Fourier frequencies  $\|\cdot\|$  means Euclidean norm. As well known, the periodogram is an asymptotically unbiased estimator for the spectral density, however, its asymptotic variance is of order  $\mathcal{O}(1)$ , that is, the periodogram is not a consistent estimator of the spectral density.

As shown in (Persival and Waden, 1993, p. 206), the bias in periodogram is due to the sidelobes of Ferjér's kernel, which leads to leakage of the information of one frequency to another. To reduce the sidelobes, and consequent bias, Persival and Waden (1993) suggest the use of tapering the data as originally proposed by Thomson (1982). The idea is form another data tapered  $h_t X_{jkt}$ , where  $h_t$  is called a taper function. The main idea is construct another data base in order to  $X_{jkt} = 0$  for  $t$  near to 0 and  $N$ . It will reduce the leakage and the bias. Then, a new periodogram, called *modified periodogram* can be achieved for the new data. The taper function is, actually, a window function, but different from classical approach in the linterature, it is computationally more efficient and has high resolution if it is used the sine tapers, defined as

$$h_{rt} = \left( \frac{2}{N+1} \right)^{1/2} \sin \left( \pi t \frac{r}{N+1} \right). \quad (3.10)$$

where  $r = 1, \dots, R$  and  $h_{rt}$  is the  $r$ th weight taper of  $X_{jkt}$ . In this case, the  $r$ th taper will be multiplied by the time series realization and a periodogram is obtained. Then, it should be repeated  $R$  times and an average periodogram can be calculated by

$$I_{jk}^r(\lambda_m) = \frac{1}{2\pi N} \left\| \sum_{t=1}^N h_{rt} X_{jkt} e^{-i\lambda_m t} \right\|^2 \quad (3.11)$$

$$I_{jk}^R(\lambda_m) = \frac{1}{R} \sum_{r=1}^R I_{jk}^r(\lambda_m). \quad (3.12)$$

**Remark 3.** Note that Equation 3.10 is a type of smoothed periodogram, that is, it preserves the properties of being asymptotically unbiased and consistent estimator of the spectral density. In addition, for each fixed  $\lambda_m, j, k$  and  $r = 1, \dots, R$ , the variables  $I_{jk}^r(\lambda_m)$  are independent.

**Remark 4.** Under the Assumption that  $R$  is a function of  $N$  ( $R_N$ ) such  $\frac{R_N}{N} \rightarrow 0$  as  $R_N$  and  $N$  go to  $\infty$ ,  $I_{jk}^R(\lambda_m)$  also corresponds to a window spectral estimator, i.e. it is asymptotically unbiased and consistent estimator see, for example, (Brockwell and Davis, 1991, p. 358).

Now some additional Assumption made to guarantee the consistency property of the cepstral estimator.

*Assumptions*

**A 1.**  $\sum_{\tau=1}^{\infty} |v_{\tau}| |\tau^{1/2}| < \infty$  and  $E|\epsilon_{jkt}|^4 < \infty$ .

**A 2.** If  $\hat{\sigma}_L < \infty \Rightarrow n^{-1/2} \hat{\sigma}_L^{-1} \rightarrow 0$ ,  $N^{-1/2} \hat{\sigma}_L^{-1} \rightarrow 0$  and  $n^{-1/2} L \rightarrow 0$  when  $n, N, L \rightarrow \infty$ .

Assumption A1 is the standard Assumption made for a smoothed (window) periodogram estimator ((Brockwell and Davis, 1991, Theorem 10.4.1). It is also possible to prove that  $\hat{f}_j$  that Assumption A2 converge in probability of a binomial distribution with probability  $p$  estimated by  $n_j/n$ . Assumption 3 is the regularity condition of the  $\Omega_w$  to be non singular.

*Asymptotic Results*

**Proposition 3.** In 3.1, let  $\hat{S}_{jk}(\lambda_m) = I_{jk}^{R_N}(\lambda_m)$  (3.11), where  $R = R_N$  is defined in Remark 4. Thus,  $\hat{c}_{jkl} \xrightarrow{p} c_{jkl}$  when  $\frac{R_N}{N} \rightarrow 0$  and  $N, R_N \rightarrow \infty$ .

*Proof.* The proof is straightforward obtained using Theorem 9.4.2 in (Anderson, 1976, p. 541).  $\square$

**Lemma 1.** For  $n \rightarrow \infty$  and for a fixed  $j = 1, \dots, J$   $\hat{f}_j \xrightarrow{p} f_j$ .

Based on the above assumptions, Krafty (2016) establishes the following theorem.

**Theorem 1. ( $\Omega_w$ -norm consistent)** For every infinite vectors  $\mathbf{p}_1, \mathbf{p}_2, \dots$ , there exist a series of a finite eigenvectors  $\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \dots, \hat{\mathbf{p}}_q$  such that

$$\|(\hat{\mathbf{p}}_{q1} \ \hat{\mathbf{p}}_{q2} \ \dots \ \hat{\mathbf{p}}_q \ 0 \dots)^T - (\mathbf{p}_{q1} \ \mathbf{p}_{q2} \ \dots)^T\|_{\Omega_w} \xrightarrow{p} 0 \quad (3.13)$$

where  $\|\cdot\|_{\mathcal{K}}$  is the reproduced Hilbert space with kernel  $\mathcal{K}$ , where for a positive definite matrix  $\mathcal{K}$ . This notation follows the theorem 2.1 of Shin (2008).

Based on Theorem 1, we have

**Corollary 5.** As  $n, N, R_N, L \rightarrow \infty, \frac{R_N}{N} \rightarrow 0$ ,

i.  $\hat{d}_{jkq}^F \xrightarrow{p} d_{jkq}$ .

ii.  $\hat{F}(\hat{\mathbf{c}}_*) \xrightarrow{p} F(\mathbf{c}_*)$ .

### 3.2 Cepstral Estimator based on M-Periodogram

Most non-parametric estimators of the spectral density are based on the periodogram, which is not resistant against outliers or heavy-tailed distributions as explained for instance in Fox (1972) and Li and Song (2008). This issue has been recently discussed by many authors who have introduced new types of periodograms that display superiority over the classical one when the data has some atypical observations, for example, Reisen et al. (2017); Fajardo et al. (2018); Li (2008), among others. Even the averaged periodogram described in Equation 3.11, with small bias compared to the classical periodogram, is still affected by additive outliers, which are the most dangerous type of outlier in time series, drastically affecting the statistical quantities of the sample mean, variance, and periodogram Fajardo et al. (2018), and consequently, the estimators derived directly from these quantities, such as Equation 3.11.

Given this undesirable sampling property of these estimators, we propose here a multitaper based on the  $M$ -periodogram estimator that can cope well with the impact of additive outliers or heavy-tailed distributions on the discriminant analysis tools.

In this context, we consider here the  $M$ -periodogram by Katkovicnik (1998) and extended by Fajardo et al. (2018); Li and Song (2008) to suggest the discriminant cepstrum tool based on this robust periodogram, denoted here by  $M$ -cepstral estimator.

Following the notation in Fajardo et al. (2018)  $I_{jk}^M(\lambda_m)$  is defined as

$$I_{jk}^M(\lambda_m) = \frac{N}{8\pi} \left\| \hat{\beta}_{jk}^M(\lambda_m) \right\|^2 = \frac{N}{8\pi} \left( \hat{\beta}_{jk}^{(c)}(\lambda_m)^2 + \hat{\beta}_{jk}^{(s)}(\lambda_m)^2 \right). \quad (3.14)$$

$\hat{\beta}_{jk}^M(\lambda_m) = [\beta_{jk}^{(c)}(\lambda_m), \beta_{jk}^{(s)}(\lambda_m)]^T$  is the solution of

$$\beta_{jk}^M(\lambda_m) = \underset{\beta \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{t=1}^N \psi \left[ X_{jkt} - \mathbf{S}_{jk}^T(\lambda_m) \beta_{jk}(\lambda_m) \right] \quad (3.15)$$

where  $\mathbf{S}_{jk}(\lambda_m) = [\cos(t\lambda_m), \sin(t\lambda_m)]^T$  and  $\psi(\cdot)$  is the Huber (Huber and Ronchetti, 2009; Mukherjee, 2008) influence function defined as

$$\psi_H(x) = \begin{cases} x, & |x| \leq c \\ c \operatorname{sign}(x), & \text{otherwise.} \end{cases} \quad (3.16)$$

**A 3.** The sequence  $\{\epsilon_{jkt}\}$  is now assumed to be a Gaussian White Noise process, ie.  $\{\epsilon_{jkt}\} \sim WN(0, \sigma^2)$ .

**Remark 5.** Under Assumption 3, Lévy-Leduc et al. (2022) showed that: for each  $\tau$ ,  $\hat{\gamma}_N^M(\tau) \xrightarrow{P} \frac{\gamma(\tau)}{a^2}$ , where  $a = 2\Phi(c) - 1$ ,  $\Phi$  is the cumulative distribution function of a standard Gaussian random variable and  $c$  is from Equation 3.16. In addition, the authors show that  $\gamma_\psi(\tau) \leq c \left| \frac{\gamma(\tau)}{\gamma(0)} \right|$ , where  $\gamma_\psi(\tau)$  and  $\hat{\gamma}_N^M(\tau)$  are the theoretical ACF and its estimator derived from the use of Huber loss function, respectively. Also,  $\gamma(\tau)$  is the covariance of the process.

**Remark 6.** Lévy-Leduc et al. (2022) defined the spectral density of the process using  $\gamma_\psi(\tau)$  as

$$S_{\psi,jk}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_{\psi,jk}(\tau) \cos(\lambda\tau), \quad \text{for all } \lambda \in [-\pi, \pi]. \quad (3.17)$$

As previously, let  $R = R_N$  (see Remark 4), thus, multitaper  $M$ -periodogram ( $I_{jk}^{RM}(\lambda_m)$ ) is now defined as

$$I_{jk}^{RM}(\lambda_m) = \frac{1}{R_N} \sum_{r=1}^R I_{jk}^{rM}(\lambda_m) \quad (3.18)$$

where

$$I_{jk}^{rM}(\lambda_m) = \frac{N}{8\pi} \left\| \hat{\beta}_{\mathbf{jk}}^{rM}(\lambda_m) \right\|^2 = \frac{N}{8\pi} \left( \hat{\beta}_{jk}^{r(c)}(\lambda_m)^2 + \hat{\beta}_{jk}^{r(s)}(\lambda_m)^2 \right) \quad (3.19)$$

and  $\beta_{\mathbf{jk}}^{rM}(\lambda_m) = [\beta_{jk}^{r(c)}(\lambda_m), \beta_{jk}^{r(s)}(\lambda_m)]^T$  is the solution of

$$\beta_{\mathbf{jk}}^{rM}(\lambda_m) = \underset{\beta \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{t=1}^N \psi \left[ h_{rt} X_{jkt} - \mathbf{S}_{\mathbf{jk}}^T(\lambda_m) \beta_{\mathbf{jk}}(\lambda_m) \right] \quad (3.20)$$

To derive asymptotic properties for short and long-memory times series, the authors Fajardo et al. (2018), among others, consider the Huber loss-function with the constant  $c_H = 1.345$ . An additional Assumptions made in this paper as follows

As a result, the M-discriminant function can be estimated based on  $M$ -cepstrum coefficients ( $\hat{c}_{jkl}^M$ ), that can be defined as follows:

$$\hat{d}_{jks}^M = \left\langle \hat{\mathbf{p}}_\ell^M, \hat{\mathbf{c}}_{\mathbf{jk}}^{(\mathbf{k})M} \right\rangle = \sum_{\ell=0}^{L-1} \hat{p}_{s\ell}^M \hat{c}_{jkl}^M, \quad (3.21)$$

where  $\hat{p}_{s\ell}^M$  are truncated estimator of the true  $p_{s\ell}$ .

$$\hat{F}^M(\hat{d}_{*1}^M, \dots, \hat{d}_{*q}^M; \hat{\mu}_j^M) = \min_j \left\{ \sum_{s=1}^q \left[ \left( \hat{d}_{*s}^M - \mu_{js}^M \right) \right]^2 - 2 \ln(\hat{f}_j) \right\}, \quad (3.22)$$

where  $\hat{\mathbf{d}}_{jks}^M$  and  $\hat{F}^M$  are the estimators of 2.20 and 2.21, respectively.

**Proposition 4.** Under Assumptions 1 and 2,  $\hat{\gamma}_N^M(\tau) \xrightarrow{p} \hat{\gamma}_N(\tau)$ , as  $N \rightarrow \infty$ .

**Proposition 5.** In 3.1, let  $\hat{S}_{\psi,jk}(\lambda_m) = I_{jk}^{MRN}(\lambda_m)$  (3.11), where  $R = R_N$  is defined in Remark 4. Thus,  $\hat{c}_{jkl}^M \xrightarrow{p} c_{jkl}$  and  $\hat{d}_{jkl}^M \xrightarrow{p} d_{jkl}$  when  $\frac{R_N}{N} \rightarrow 0$  and  $N, R_N \rightarrow \infty$ .

*Proof.* Using the results of Proposition 4, the stationarity of the process allows us to use the cosine function for the Fourier transform, and the logarithm is a continuous function. Then, by applying the mapping theorem (Theorem 9.4.2 from (Anderson, 1976, p. 541)), the results follow directly.  $\square$

**Proposition 6.** As  $n, N, R_N, L \rightarrow \infty$ ,  $\frac{R_N}{N} \rightarrow 0$ ,  $\left\| \hat{d}_{jks}^M - \hat{d}_{jks} \right\| = o_p(1)$

**Theorem 2.** Under the Assumption 1 and using sine tapers as defined in Equation 2.3, each data taper are orthogonal and the periodogram are independent for each  $r$  taper. Then,  $I_{jk}^{RM}(\lambda_m) \xrightarrow{p} S_x(\lambda_m)$  for all  $\lambda_m$ , as  $R_N \rightarrow \infty$ .

**Lemma 2.**  $\operatorname{cov} \left( I_{jk}^{MR}(\lambda_m), I_{jk}^{MR}(\lambda_k) \right) \rightarrow 0$ , for all  $k \neq m$  and  $R_N \rightarrow \infty$ .

## 4 Simulations

### 4.1 AR(1) simulation with fixed parameters

Before show the results of Monte Carlo simulation, an example is conducted of the empirical cepstral function for the AR(1) model  $X_{jkt} = \phi_{jk}X_{jk(t-1)} + \epsilon_{jkt}$ , where  $\epsilon_{jkt} \sim \mathcal{N}(0, \sigma_{\epsilon_{jk}}^2)$ . We considered in the simulation  $J = 3$ ,  $n_j = 1$ ,  $N = 100$ , the AR parameters  $\phi \in \{0.25, 0.50, 0.75\}$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ . The classical periodogram was used in this example as the estimator of the spectral density and the results are presented in Figure 5. There are three panels for the time domain: (a) time series replicates; (b) ACF; and (c) PACF; and three panels for the frequency domain: (d) Periodogram; (e) ln-Periodogram; and (f) estimated Cepstra.

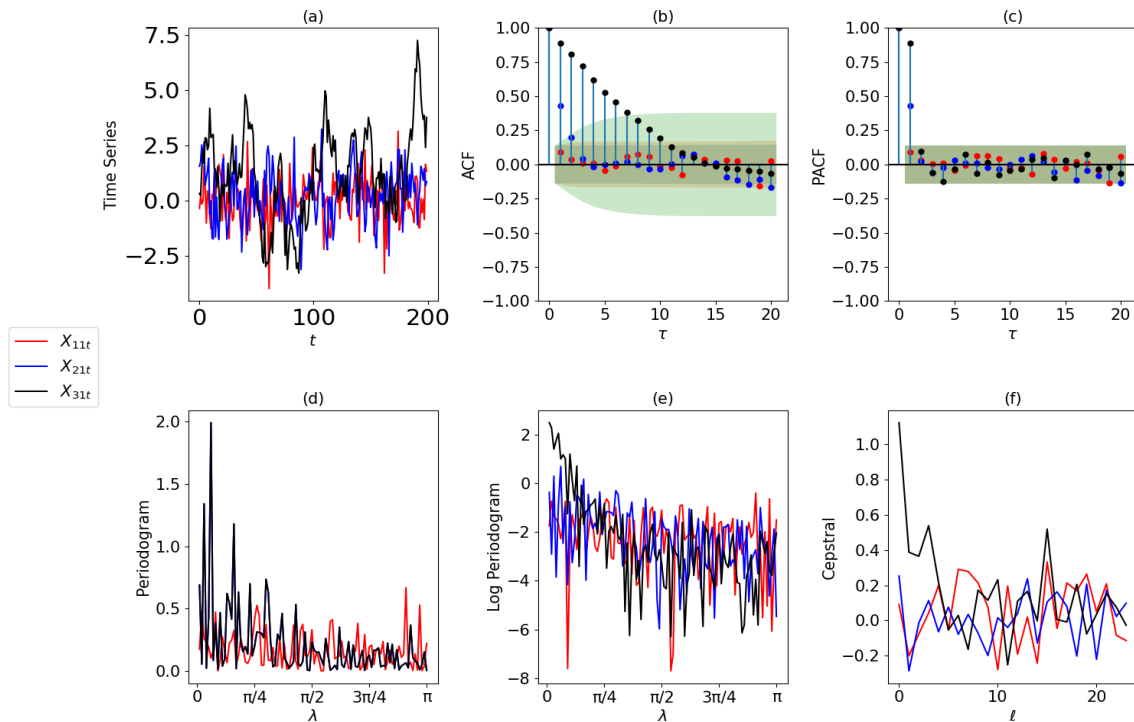


Figure 5:  $\phi \in [0.1, 0.5, 0.9]$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) ln-Periodogram, and (f) Cepstra

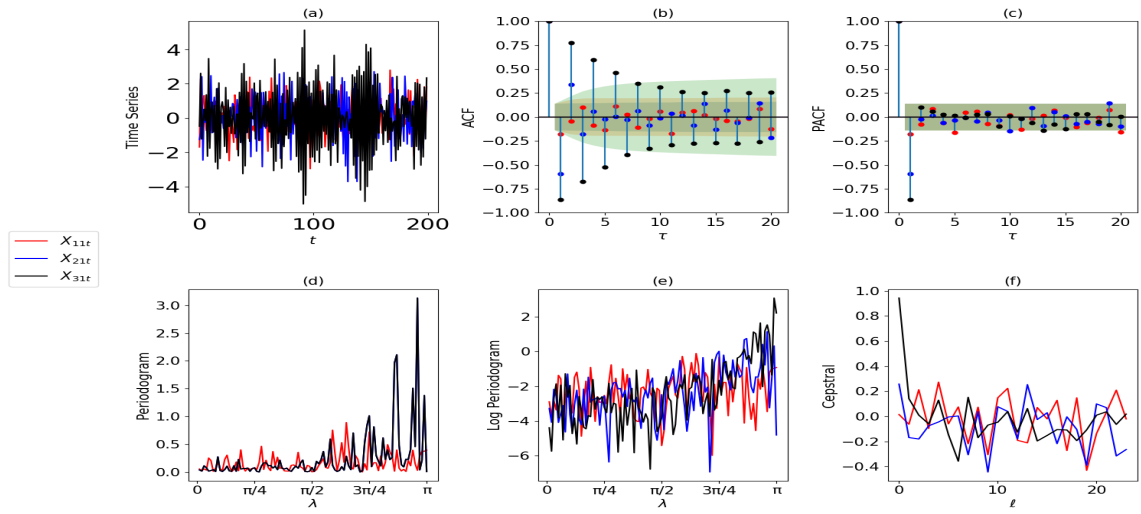


Figure 6:  $\phi \in [-0.1, -0.5, -0.9]$  and  $\sigma_{\epsilon}^2_{jk} = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) ln-Periodogram, and (f) Cepstra

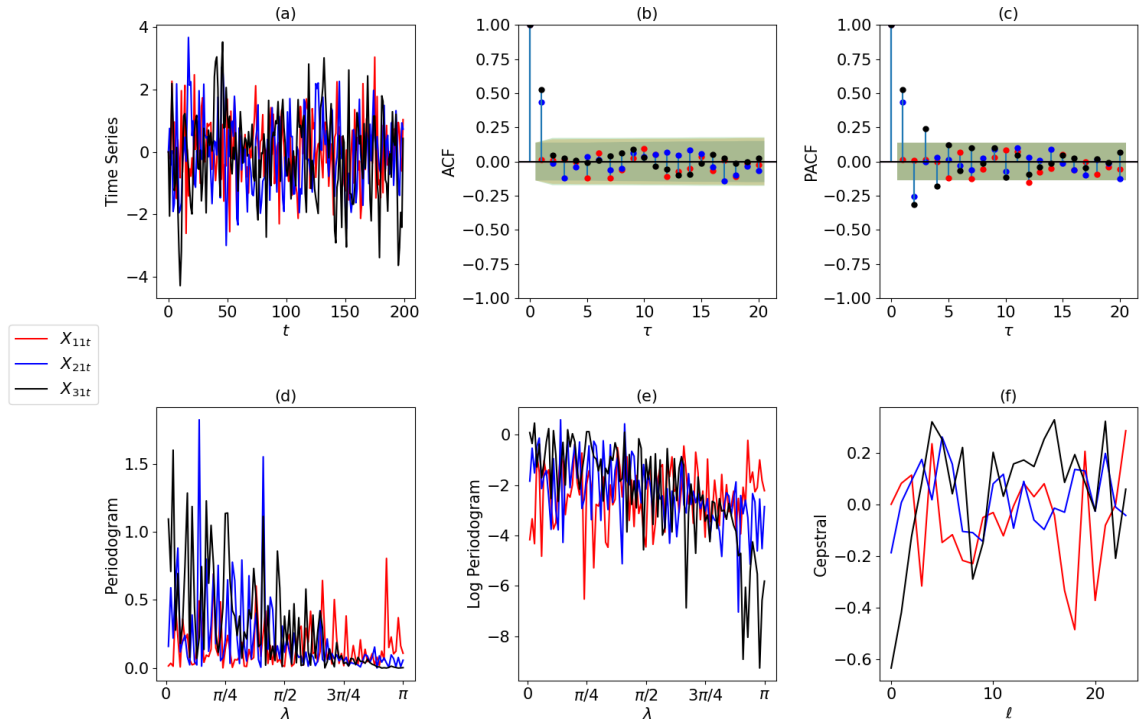


Figure 7:  $\theta \in [0.1, 0.5, 0.9]$  and  $\sigma_{\epsilon}^2_{jk} = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) ln-Periodogram, and (f) Cepstra

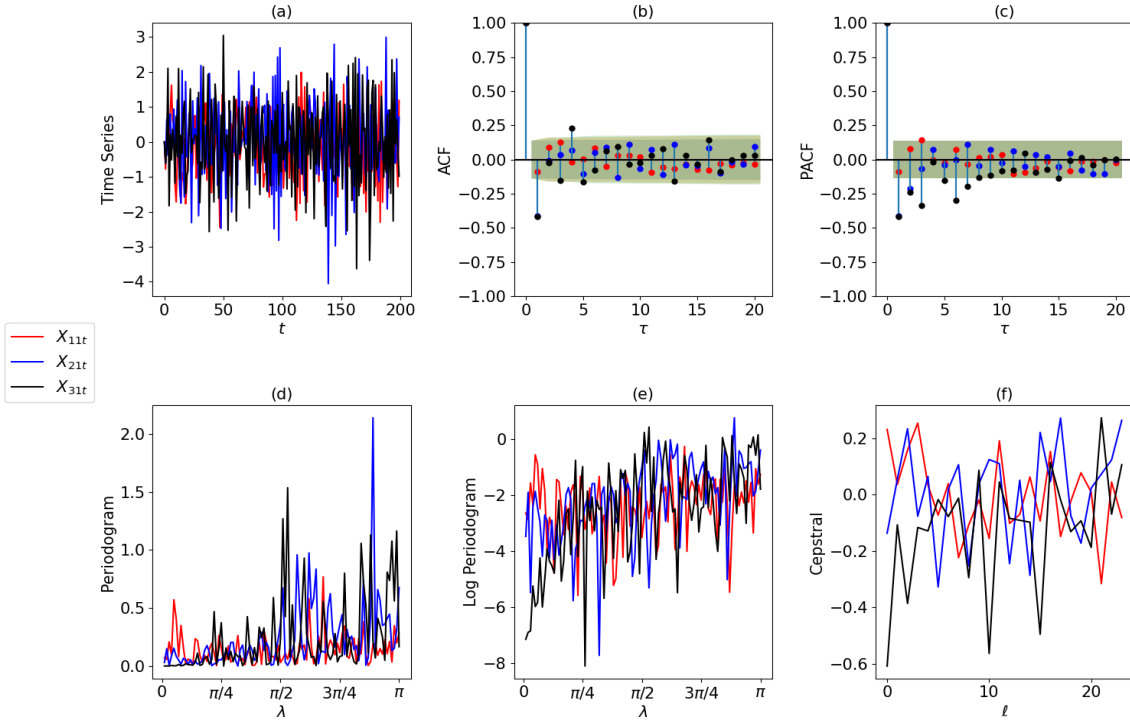


Figure 8:  $\theta \in [-0.1, -0.5, -0.9]$  and  $\sigma_{\epsilon}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) ln-Periodogram, and (f) Cepstra

On the one hand, the time domain is shown in panels (a), (b), and (c). It is possible to note that increasing the parameter will increase the variability for constant variance noise. This also influences the autocorrelations, which become more persistent over time as the coefficient increases. All coefficients can be seen in the partial ACF with just one significant peak.

On the other hand, the frequency domain is represented in panels (d) and (e). It can be noted that the estimated spectra of an AR(1) process exhibit lower frequency components. This means that the most variability of the process is due to the frequencies next to zero. Another important feature is the fact that for frequencies greater than 1, the time series cannot be differentiated. According to Krafty (2016), there is a contribution of this kind of frequencies that have amplitudes next to zero.

To observe more variability and to discriminate better, the  $\ln$ arithm is calculated on the periodogram, generating the graphs presented in panel (e). Clearly, they are now more easily distinguishable. More complicated processes are even better taken into account with this variability.

## 4.2 Monte Carlo Simulation using AR(2)

A Monte Carlo simulation was performed to analyze the sample properties of the M-cepstra and its impact on classification rates in the context of within spectral variability with and without contamination of abrupt observations that behave as additive outliers. For the simulations,  $J = 3$  populations are considered in 6 different situations:  $n_j = \{15, 50, 100\}$  and  $N = \{1000, 2000\}$ . In each scenario, the classification rate is calculated and the scenarios are repeated 100 times to obtain the mean and standard deviation of the classifications. In order to compare with Krafty (2016) and due to the characteristics of the data application presented in Section 5, the simulations are conducted using a conditional (on population) AR(2) model that can be represented by

$$X_{jkt} = \phi_{jk1}X_{jkt-1} + \phi_{jk2}X_{jkt-2} + \epsilon_{jkt} \quad \epsilon_{jkt} \sim N(0, \sigma_{jk}^2) \quad (4.1)$$

where the parameter for each population is a random variable presented in Table 2.

Table 2: Conditional Parameters Distribution

Population \ Parameters	$\phi_{jk1}$	$\phi_{jk2}$	White Noise	White Noise's Variance
$j = 1$	$(\phi_{1k1} j = 1) \sim UNI(0.05, 0.7)$	$(\phi_{1k2} j = 1) \sim UNI(-0.12, -0.06)$	$\epsilon_{1k} \sim N(0, \sigma_{1k}^2)$	$(\sigma_{1k}^2 j = 1) \sim UNI(0.1, 10)$
$j = 2$	$(\phi_{2k1} j = 2) \sim UNI(0.01, 1.2)$	$(\phi_{2k2} j = 2) \sim UNI(-0.36, -0.25)$	$\epsilon_{2k} \sim N(0, \sigma_{2k}^2)$	$(\sigma_{2k}^2 j = 2) \sim UNI(0.3, 3.0)$
$j = 3$	$(\phi_{3k1} j = 3) \sim UNI(0.12, 1.5)$	$(\phi_{3k2} j = 3) \sim UNI(-0.75, -0.56)$	$\epsilon_{3k} \sim N(0, \sigma_{3k}^2)$	$(\sigma_{3k}^2 j = 3) \sim UNI(0.9, 1.1)$

The conditional AR(2) model is chosen with parameters being random variables with uniform distribution in overlapping intervals. Additionally, the disturbances are modeled as normal distribution with zero mean and their variance as a uniform distribution with different increasing intervals. This characteristic is important to create time series that can be mixed and create difficulty for discrimination and classification, as would occur in a real application context.

Each scenario includes all three populations for each time series and replicas. All transformations are performed, including multitaper periodogram estimation for both classical and  $M$ -periodogram. The cepstral vector is then obtained for each replicate within its own population. An LDA parameter is estimated based on the cepstra and used for the classification test. For this test, a new group of time series for each population is simulated with the same parameters. For comparison with Krafty (2016),  $N = 50$  is chosen for the testing data. These new time series undergo the same transformations as the original replicas until achieving the cepstral vector, which is then used for testing.

Finally, the simulations are conducted in two situations: without outliers, where the model is fitted without extreme observations, and with outliers, where the replicas are contaminated with an additive outlier following Huber and Ronchetti (2009); Reisen et al. (2020b) with

$$Z_{jkt} = X_{jkt} + \omega I_{jkt}, \quad (4.2)$$

where  $\omega$  is the magnitude of the outlier,  $\{X_{jkt}\}$  and  $\{I_{jkt}\}$  are independent, and  $\{I_{jkt}\}$  is a sequence of IID random variables with  $P(I_{jkt} = -1) = P(I_{jkt} = 1) = p/2$  and  $P(I_{jkt} = 0) = 1 - p$ ,  $p \in (0, 1)$ . Here, we take  $p = 0.01$  and  $\omega = 7$ . The uncontaminated case corresponds to  $\omega = 0$ . The proposed robust cepstral discriminant analysis was compared to the cepstral multitaper discriminant analysis, both utilizing  $R = 7$  tapers. The tuning constant in the Huber model is set to 1.345 as the asymptotically optimal value defined by Reisen et al. (2020a). A test data set of 50 time series per group is simulated for each random sample to evaluate the classification rate of the training data. The correct classification rate is calculated in each scenario using the LDA model with cepstral coefficients. This scenario was repeated 100 times and the mean and standard deviation of the correct classification rates are calculated and presented in Table 3.

The performance of the two classification methods - Classical and Huber - are evaluated in different scenarios, with and without the presence of outliers. It is noted that for samples without outliers, the results show that both classification methods exhibit high correct classification rates. In  $\Pi_1$ , for instance, with  $n_j = 15$ , the average correct classification rates were 89.64% and 90.09% for the Classical method with  $N = \{1000, 2000\}$ , respectively. For the Huber method, the rates were slightly lower, with 85.37% and 89.85% for  $N = \{1000, 2000\}$ , respectively.

As the number of replicas per population increases, an improvement in correct classification rates is observed for both methods. With  $n_j = 100$ , the Classical and Huber methods achieved

91.02% and 93.00% correct classification for  $N = 1000$ , respectively. The same trend is observed for  $N = 2000$ , where the rates reach 93.00% and 92.19%.

In the presence of outliers, a decrease in correct classification rates is noted for both methods. For  $\Pi_1$  with  $n_j = 15$ , the average correct classification rates dropped to 67.06% and 68.52% for the Classical method with  $N = 1000$  and  $2000$ , respectively. However, the Huber method proved to be more robust, with rates of 79.13% and 83.64% for  $N = 1000$  and  $2000$ , respectively.

The robustness of the Huber method is more evident as the number of replicas per population increases. With  $n_j = 100$ , the correct classification rates for the Classical method were 76.18% and 82.78%, while the Huber method achieved 81.06% and 88.03% for  $N = 1000$  and  $2000$ , respectively. This indicates that the Huber method is more effective in the presence of outliers when compared to the Classical method.

Comparing the different variance scenarios ( $\sigma^2$ ), it is observed that for  $\Pi_1$  (0.01 - 10), the correct classification rates are generally lower than for  $\Pi_2$  (0.3 - 3) and  $\Pi_3$  (0.9 - 1.1). For example, in  $\Pi_3$ , for  $n_j = 15$ , the correct classification rates without outliers were 93.67% and 93.88% for the Classical method with  $N = \{1000, 2000\}$ , respectively, showing better performance compared to  $\Pi_1$  and  $\Pi_2$ .

With the inclusion of outliers,  $\Pi_3$  also showed overall better robustness. The correct classification rates for the Huber method in  $\Pi_3$  were consistently higher compared to  $\Pi_1$  and  $\Pi_2$ . This suggests that classifier performance can be strongly influenced by the variance of the data.

Table 3: Monte Carlo simulation: mean and standard deviation rate for correct classification

$\sigma_{jk}^2$	$n_j$	WITHOUT OUTLIER						WITH OUTLIER					
		Classic				M		Classic			M		
		$N$											
$\Pi_1$	(0.01 - 10)	15	250	83.94%	86.97%	89.64%	90.09%	85.37%	89.85%	67.06%	68.52%	79.13%	83.64%
				(6.72%)	(6.40%)	(5.80%)	(5.96%)	(6.10%)	(5.99%)	(9.10%)	(8.45%)	(7.45%)	(7.91%)
				85.55%	88.34%	90.85%	91.02%	86.40%	90.13%	75.08%	81.10%	80.34%	84.74%
		50		(5.58%)	(5.17%)	(4.47%)	(4.63%)	(5.20%)	(4.53%)	(4.97%)	(3.89%)	(4.89%)	(4.48%)
				86.00%	88.71%	91.02%	93.00%	89.02%	92.19%	76.18%	82.78%	81.06%	88.03%
				(5.40%)	(4.98%)	(4.19%)	(3.75%)	(6.28%)	(3.32%)	(3.77%)	(3.12%)	(3.39%)	(3.18%)
		100		79.25%	84.79%	88.80%	91.40%	83.10%	90.01%	72.28%	83.02%	85.26%	87.83%
				(8.09%)	(6.89%)	(6.07%)	(5.27%)	(4.20%)	(4.63%)	(6.44%)	(5.70%)	(5.01%)	(4.73%)
				82.92%	88.02%	91.98%	92.33%	89.60%	92.02%	84.78%	86.66%	87.12%	88.43%
$\Pi_2$	(0.3 - 3)	15		(6.01%)	(5.16%)	(4.17%)	(4.17%)	(4.60%)	(5.12%)	(4.77%)	(5.00%)	(4.23%)	(4.22%)
				83.76%	88.69%	92.53%	95.27%	91.14%	95.03%	71.04%	87.44%	89.02%	90.95%
				(5.44%)	(4.76%)	(3.91%)	(3.21%)	(1.74%)	(2.59%)	(2.97%)	(2.84%)	(1.93%)	(2.71%)
$\Pi_3$	(0.9 - 1.1)	15		(6.90%)	(6.47%)	(5.22%)	(6.15%)	(5.40%)	(4.33%)	(6.61%)	(6.63%)	(5.81%)	(4.91%)
				90.38%	93.11%	95.22%	95.83%	93.10%	95.16%	84.10%	88.52%	90.56%	93.26%
				(4.92%)	(4.18%)	(3.44%)	(4.57%)	(3.57%)	(4.23%)	(3.09%)	(4.57%)	(3.89%)	(2.81%)
50		90.48%	93.55%	95.30%	96.51%	94.98%	95.71%	84.22%	93.22%	91.90%	95.23%		
		(4.33%)	(3.61%)	(3.14%)	(2.74%)	(2.86%)	(2.93%)	(2.44%)	(2.91%)	(1.93%)	(1.95%)		
		87.73%	91.01%	93.67%	93.88%	91.15%	93.45%	81.94%	81.08%	89.43%	91.14%		

The results of the Monte Carlo simulations indicate that the Huber method is generally more robust in the presence of outliers compared to the Classical method. Furthermore, the performance of the classification methods is influenced by the sample size and the number of replicas per population. In high variance scenarios ( $\Pi_1$ ), both methods exhibit lower correct classification rates, while in low variance scenarios ( $\Pi_3$ ), the classification rates are higher. These findings highlight the importance of considering the robustness of classification methods in different contexts of variance and the presence of outliers.

## 5 The Real Data Application: Neurodegenerative Diseases

According to Hausdoff et al. (2000), gait patterns can provide tools to aid in the detection and monitoring of neurodegenerative diseases. For example, Amyotrophic Lateral Sclerosis (ALS) is a disease caused by the loss of motor neurons that directly affect balance. Another disease commonly known in the literature is Huntington’s disease, which degenerates parts of the brain responsible for controlling movement smoothness. As a consequence, individuals with these diseases tend to have more abrupt movements, affecting their gait pattern.

All these diseases affect gait and mobility, which can be identified by the cycle of walking patterns recorded by sensors in patients’ feet. Hausdoff et al. (2000) conducted a study for this purpose, involving 49 participants who were asked to walk. Among all the participants, 16 were healthy control individuals, 20 had Huntington’s disease, and 13 had ALS <sup>1</sup>. The recorded information includes the time spent for each individual to complete a stride cycle per second <sup>2</sup>.

The participants walked for exactly 6 minutes, resulting in a total of 360 observations in seconds. The objective is to capture how much time each person takes per second to complete a stride cycle, that is, how much time they take to place their foot on the ground and do so again during walking. Therefore, the data are unit of time per unit of time, in the sense that it evaluates how many milliseconds an individual took to complete a walking cycle (pushing one foot against the ground and doing so again) for each second. The first 20 observations (first 20 seconds) were excluded from the sample by Hausdoff et al. (2000) as the author considers them merely a start-up effect. The original data can be found at <https://www.physionet.org/content/gaitnnd/1.0.0/>.

According to Krafty (2016), the original data required some treatment. Firstly, the author decided to work with 3.5 minutes, which gives 210 observations out of the total 360 minutes, due to the slower pace of sick individuals compared to healthy ones, resulting in fewer samples. As this was still insufficient, two patients each from ALS and Huntington’s were eliminated due to their missing observations or abrupt ones among those who remained, leaving 11 and 18 patients, respectively. Additionally, in order to avoid further eliminations, the author opted to fill in the missing observations using cubic interpolation. These observations were also used to smooth spline sampled at 2Hz, effectively doubling the amount of information to 420 observations. Finally, the analyzed dataset contains 420 observations and 45 remaining participants (Krafty, 2016, p. 445).

Note that several transformations are needed to model could be used in frequency domain and to bring it back to the time domain with the main information in order to be used in linear discriminant analysis. In this work, the main supposition is that less transformation can be used for better classification, especially not using filter for extreme observations.

The main idea of this work centers on the fact that so many transformations can significantly influence the results, masking or generating false success rates. Furthermore, the fact of eliminating sick individuals seems to be a convenience choice once the healthy ones are more sTable and there are more observations.

Therefore, in this work, it is chosen not to filter the original data for abrupt observations, arguing that this information only behaves as additive outliers, but is in fact relevant information for identifying sick individuals. Additionally, all individuals are considered giving  $n = 49$  participants and  $N = 120$  observations per individual (representing 2.05 minutes). Subsequently, for comparison reasons, two cases are considered: i) **modified** data extracting

---

<sup>1</sup>There are still 16 individuals with Parkinson’s disease not considered in the sample by Krafty (2016). For this reason, they were not analyzed in this study either for comparison purposes.

<sup>2</sup>And because the data are given in seconds, the recorded data can be expressed in Hertz, representing here a measure of the frequency of strides per second.

abrupt observations by filtering using 3 standard deviation median exactly as done by Krafty (2016); ii) **non-modified** data without extracting the extreme observations. Next, in both cases, the resulting data are detrended.

In this context,  $J = 3$  populations (1: control; 2: ALS; 3: Huntington) are being considered, with  $n_1 = 16$ ,  $n_2 = 13$ , and  $n_3 = 20$  being their respective numbers of replicates. Figures 9, 10, and 11 present the time series, the boxplot, and ACF for 9 individuals, 3 for control healthy, 3 for ALS, and 3 for Huntington's with data **modified**. Similarly, figures 12, 13, and 14 are the data for **non-modified** data.

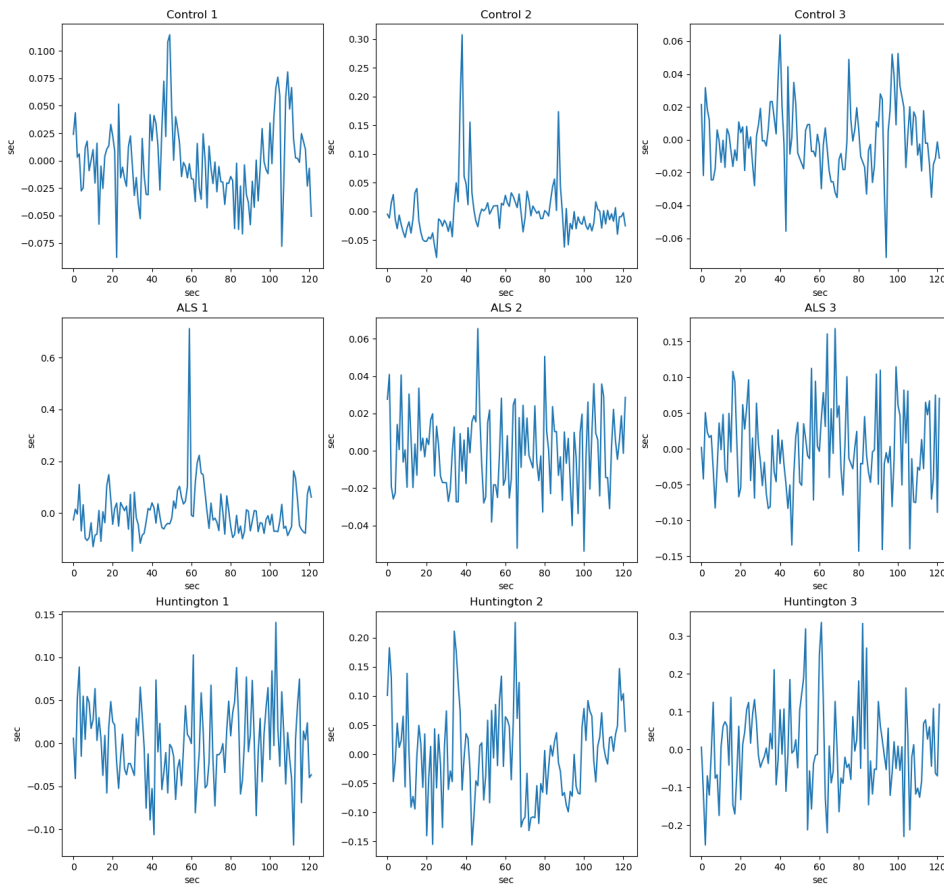


Figure 9: **Modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

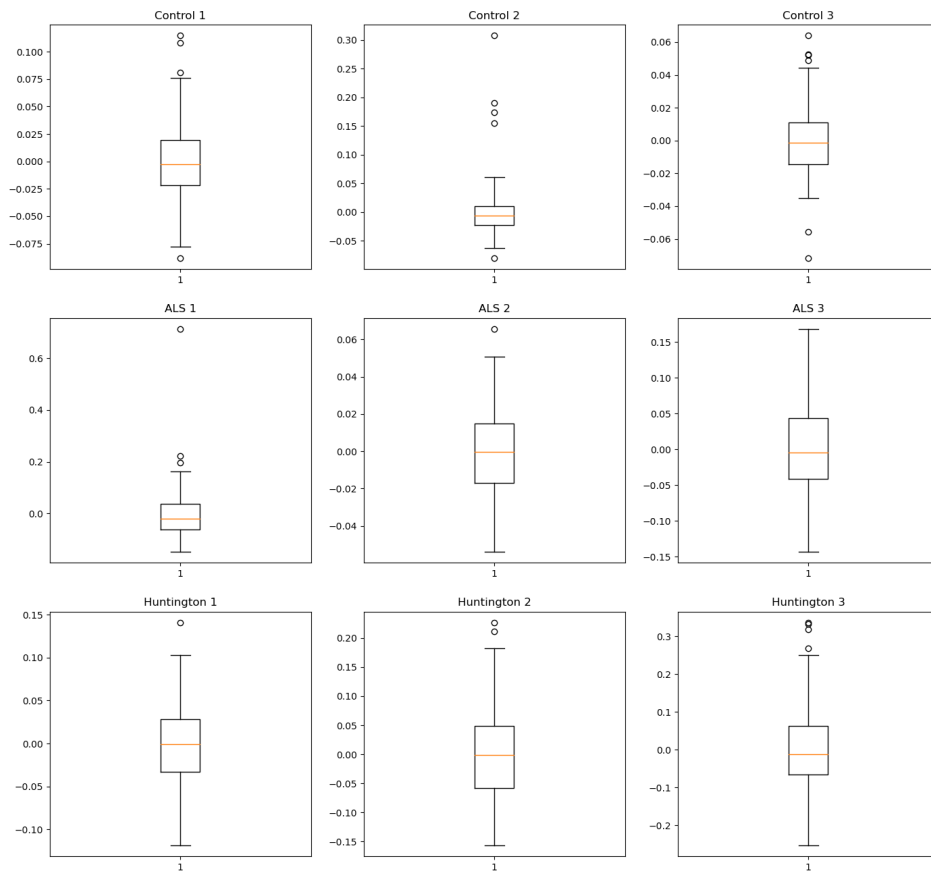


Figure 10: Box plot of **modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

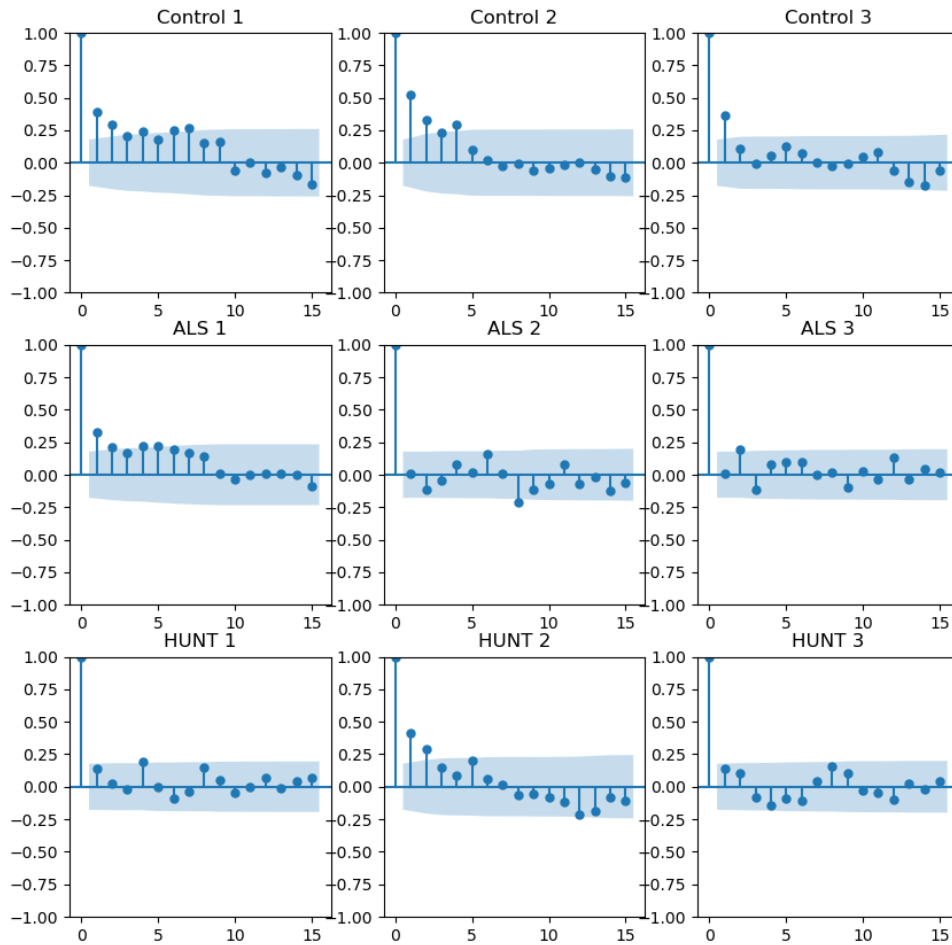


Figure 11: ACF of **modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

Comparing Figure 9 with 12, the modified data still have abrupt observations after filtering, as can be seen in Figure 10. Furthermore, healthy control individual 2, ALS 1, and ALS 3 exhibit values so extreme that they significantly affect the visualization of the data structure. Additionally, as observed in Figures 11, the autocorrelation structure of the same time series differs. In other words, the mechanical removal of extreme observations affects the autocorrelation structures, which can lead to spurious conclusions when using the periodogram based on such information.

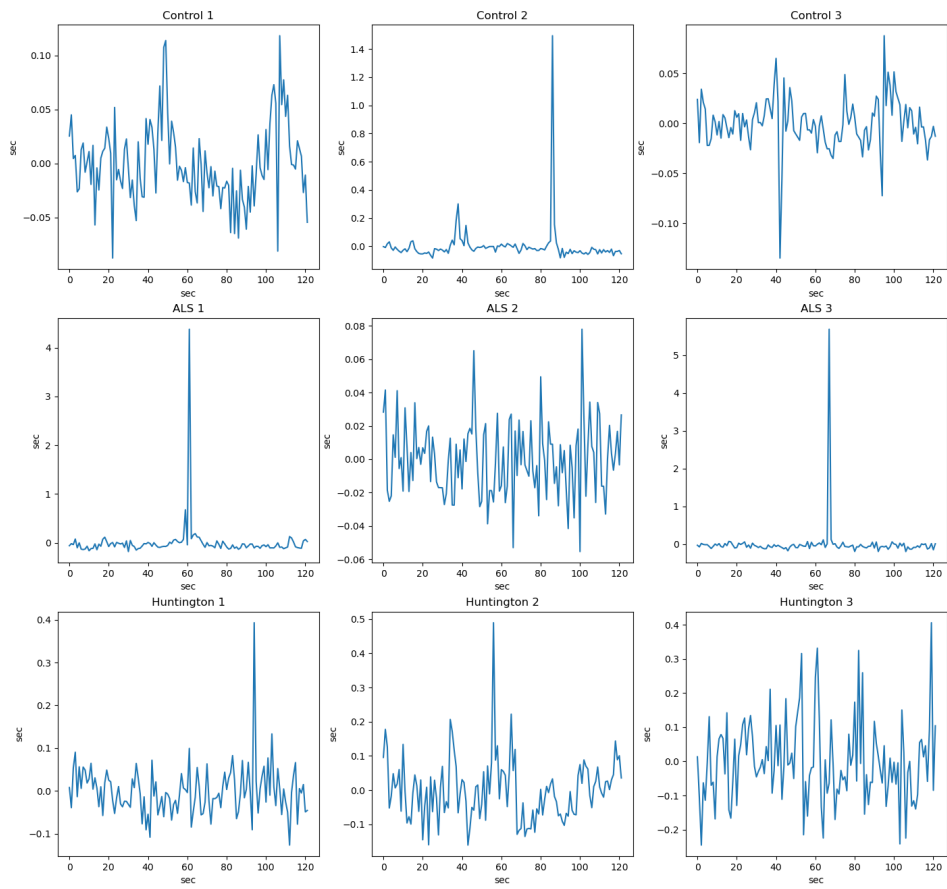


Figure 12: **Non modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

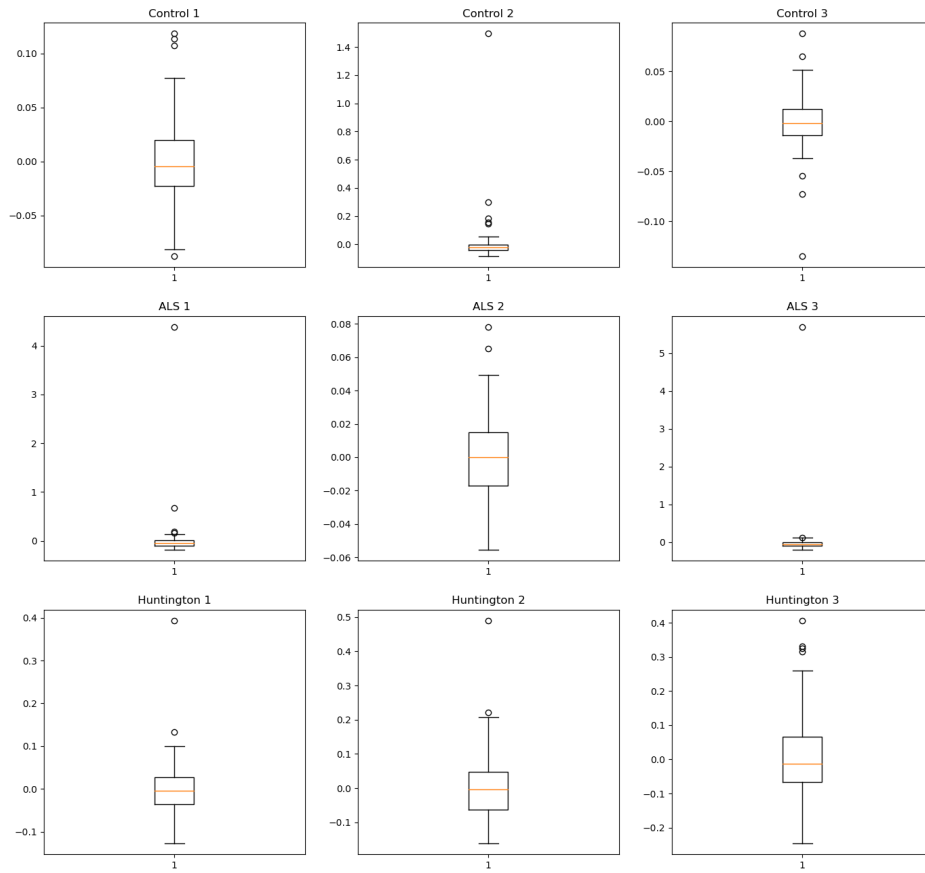


Figure 13: Box plot of **non modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

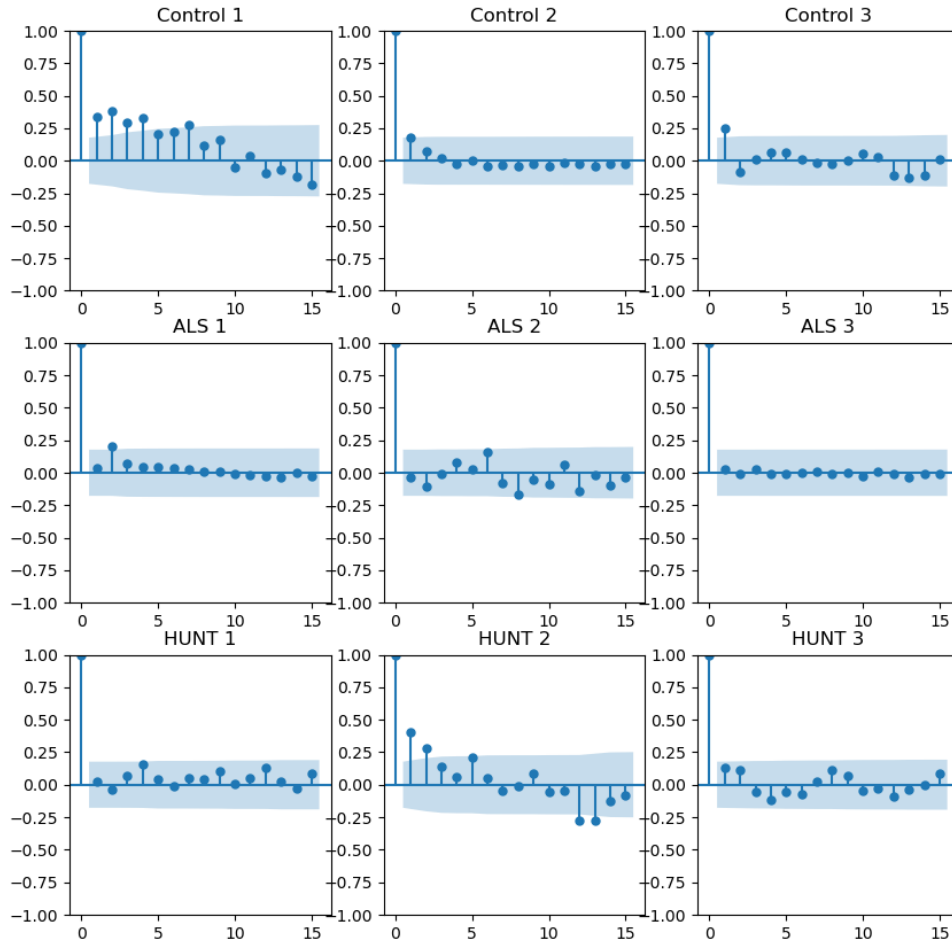


Figure 14: ACF of **non modified** detrended stride interval time series from 9 participants in neurodegenerative disease study with 3 healthy controls, 3 ALS and 3 Huntington's

This work focuses on the best classifier for linear discriminant analysis using multitaper periodogram estimation for classic analysis and modified data, and multitaper M-periodogram for robust analysis and non-modified data. As a result, the parameters of the number of multitapers ( $R$ ) and the number of cepstrals ( $L$ ) need to be chosen. Since the multitaper impacts only asymptotic properties, and for comparison, it will be chosen  $R = 7$ , the same used by Krafty (2016).

The number of cepstrals needs to be chosen in a parsimonious way, opting for the smallest possible number that still generates the highest classification rate. Table ?? contains the classification rate for the first 20 cepstrals. The maximum classification rate is achieved when  $L = 9$ .

Table 4: Number of Cepstrals by classification rate

<b>L</b>	<b>Classification rate (%)</b>
3	69.39
4	69.39
5	73.47
6	75.51
7	81.63
8	81.63
<b>9</b>	<b>83.67</b>
10	75.51
11	71.43
12	73.47
13	73.47
14	75.51
15	73.47
16	75.51
17	75.51
18	79.59
19	79.59
20	81.63

In Figure 15, the horizontal axis presents the number of cepstrals, and the vertical axis shows the classification rate presented in the Table 4. Clearly, the classification rate grows until it reaches its maximum value when  $L = 9$  with 83.67% of success.

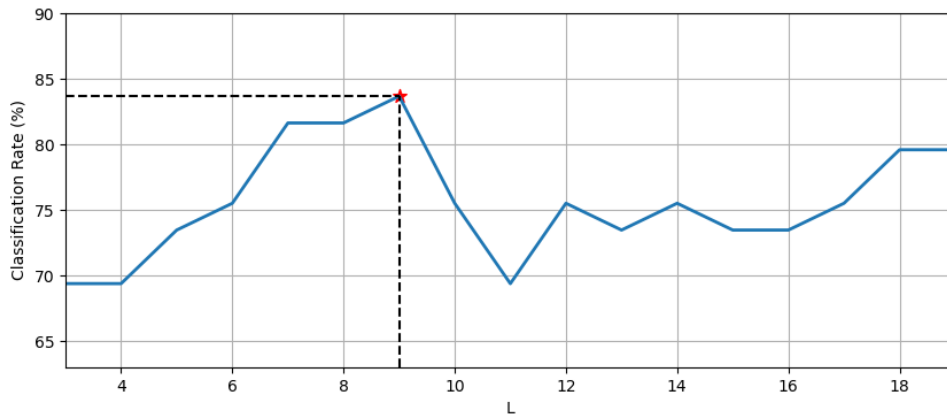


Figure 15: Number of Cepstrals by classification rate

Cepstrals are defined as the result of taking the inverse Fourier transform of the logarithm of the spectrum. . Consequently, they carry all the information regarding not only the variability of the time series but also the variability of the spectrum, enabling a more profound and parsimonious analysis of discrimination and facilitating their use for classifying new time series.

Furthermore, Krafty (2016) states that when using multitaper periodogram estimates, most of the information is contained in the first cepstrals. According to the author, no more than  $L = 4$  cepstrals are necessary for this purpose. However, to achieve the classification

rates shown by the author, many transformations were necessary in the original database, particularly the exclusion of 4 series from sick individuals, which contain important information for the analysis. Considering the database used in this work without excluding any sick individuals, the average classification rate across all groups is indeed equal to 69.39%. It is worth noting that Krafty (2016) did not report this general average rate.

The classification rate can be decomposed by population based on conditional probabilities and use Bayes' formula to compute the probabilities as follows:

$$\rho_{ij} = \mathbb{P}(\hat{c}_{i\ell} \in \Pi = i | \hat{c}_{j\ell} \in \Pi = j) = \frac{\mathbb{P}(\hat{c}_{j\ell} \in \Pi = i \cap \hat{c}_{j\ell} \in \Pi = j)}{\mathbb{P}(\hat{c}_{j\ell} \in \Pi = j)}, \quad (5.1)$$

where  $\sum_{i=1}^J \rho_{ij} = 1$ . Once in this application case  $J = 3$ , then the  $(3 \times 3)$  normalized confusion matrix can be summarized in Table 4.

Table 5: Normalized Confusion Matrix for  $J = 3$

Predicted population \ Population	Population		
	1	2	3
1	$\rho_{11}$	$\rho_{21}$	$\rho_{31}$
2	$\rho_{12}$	$\rho_{22}$	$\rho_{23}$
3	$\rho_{31}$	$\rho_{32}$	$\rho_{33}$
total	1	1	1

Note that in Table 5, the main diagonal represents the correct classification rates per population, while off-diagonal entries represent the misclassification rates. The sum of each column is equal to 1, and each entry can be interpreted as the probability of being classified into each of the populations (rows) given that they belong to the known population (columns).

Table 6: Confusion Matrix for 49 **changed** time series of neurodegenerative disease in %

Predicted population \ Population	Population		
	1	2	3
1	<b>100.00%</b>	23.08%	30.00%
2	0.00%	<b>38.46%</b>	5.00%
3	0.00%	38.46%	<b>65.00%</b>

\* 1: Healthy control; 2: ALS; 3: Huntington's

It is evident that the accuracy rate of healthy individuals reaches 100% accuracy. However, among individuals with Sclerosis, only 30.77% were correctly classified, with the majority being confused with individuals with Huntington's disease, resulting in a classification error rate of 38.46%. Regarding individuals with Huntington's disease, 65% of the 20 individuals were correctly classified, with 30% incorrectly classified as healthy and 5% wrongly classified as having Sclerosis.

Consider now the data non-modified in the sense that the extreme values are not removed. The Assumption of normality serves as a foundation in the context of linear discriminant analysis. Cepstrals are, by definition, the Fourier transform of the lnarithm of the spectrum. Additionally, ? demonstrates that the periodogram follows an asymptotic chi-square distribution with two degrees of freedom. The lnarithm serves as a monotone transformation that

linearizes the quadratic term, and the inverse Fourier transform simply shifts the coordinates from frequency domain back to the time domain. Consequently, such variables are expected to exhibit a normal asymptotic distribution. To assess the normality of the Cepstrals for application purposes, a Shapiro test was conducted in estimated cepstrals generating a p-value of 56.18%. This means that the normality distribution is not rejected.

The confusion matrix is constructed for this data using the M-periodogram as spectral estimation and the results are presented in Table 7. Exactly as the classical periodogram for modified data, the healthy control.

Table 7: Confusion Matrix for 49 **non-modified** time series of neurodegenerative disease

Population			
Predicted population	1	2	3
<b>1</b>	<b>100%</b>	15.38%	15.00%
<b>2</b>	0.00%	<b>76.92%</b>	10.00%
<b>3</b>	0.00%	7.69%	<b>75.00%</b>

\* 1: control; 2: ELA; 3: Huntington's

Figure 9 displays the estimated discriminant coefficients obtained through the robust proposed approach and the classical proposed by Krafty et al. (2016).

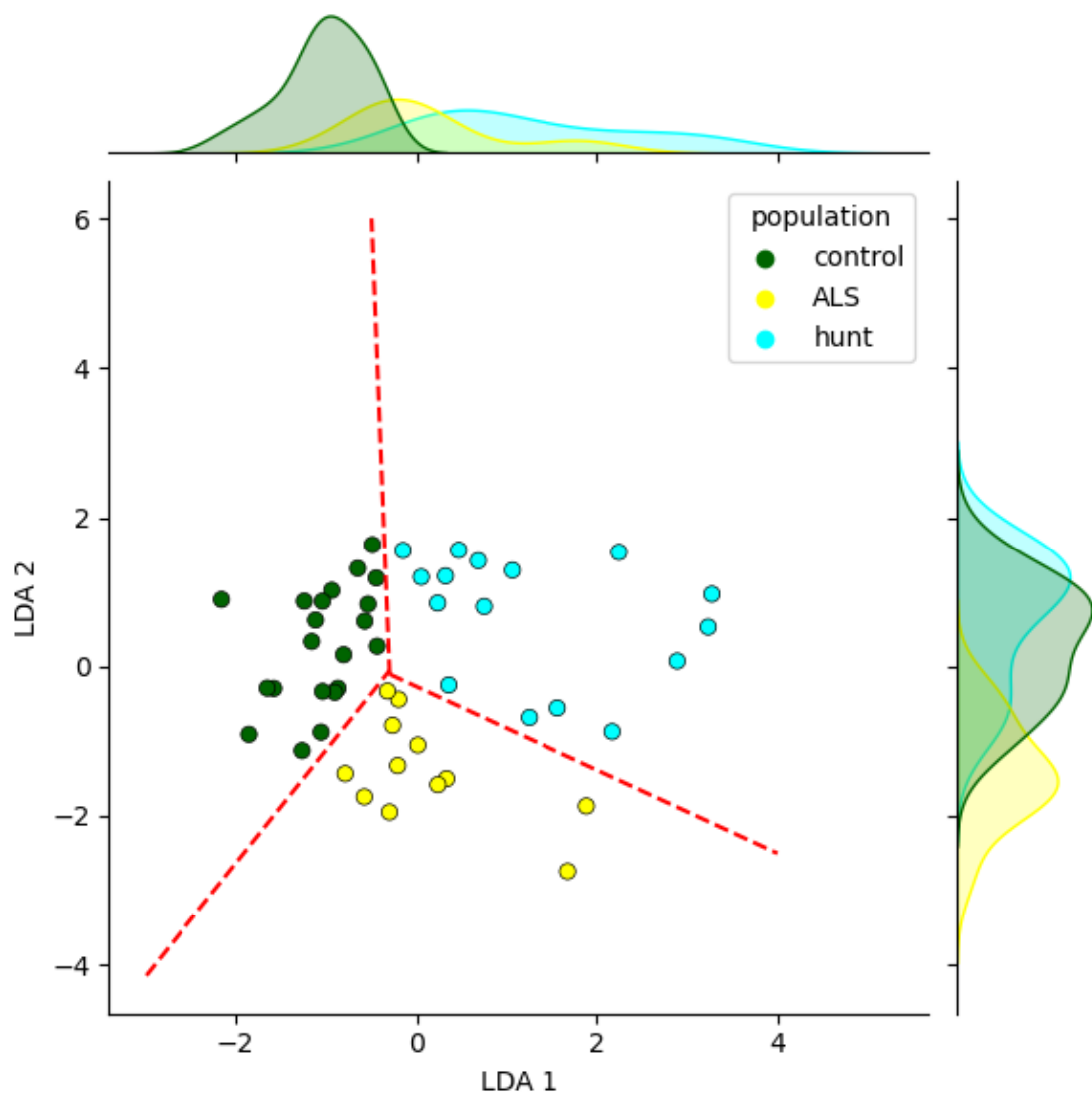


Figure 16: Linear discriminant analysis using two discriminants for **non-modified** data.

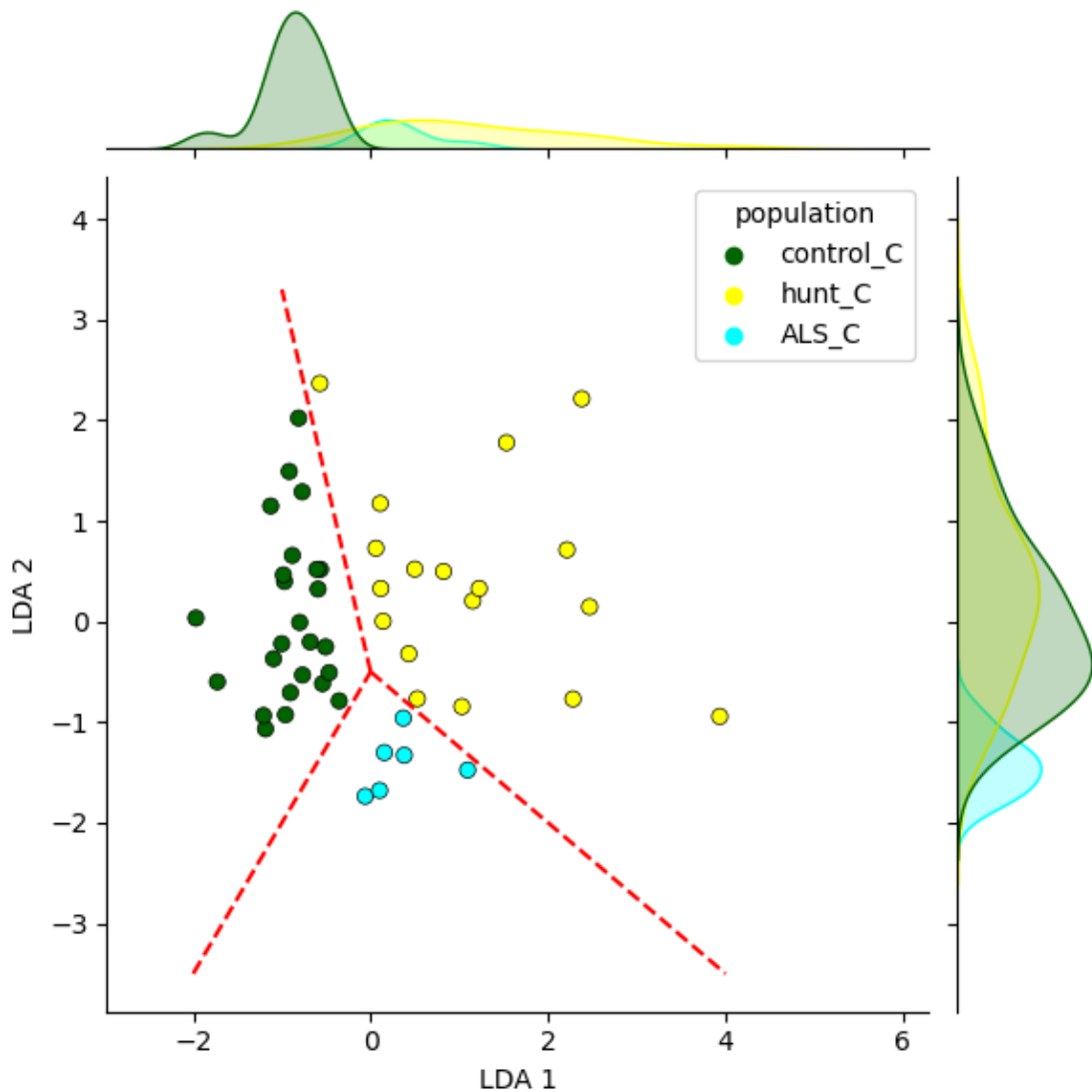


Figure 17: Linear discriminant analysis using two discriminants for **modified** data.

Specifically, coefficients estimated from the modified dataset are represented in black, while those derived from the original dataset are shown in red. Furthermore, the coefficients derived from our proposed robust method using the original dataset are presented in blue. Notably, there are distinct differences between the black and red coefficients. In the scenario of the adjusted dataset, the black discriminant primarily distinguishes individuals with Huntington’s disease from the other two groups. The second discriminant primarily separates healthy controls from ALS patients. Conversely, in the original dataset scenario, the first discriminant predominantly separates control participants from the other two groups. In contrast, the second discriminant primarily distinguishes individuals with Huntington’s disease from those with ALS. However, under the presence of atypical observations in the dataset, the discriminants represented in blue demonstrate better separability performance than the red coefficients. Our proposed robust method’s empirical mean classification rate for the original data is approximately 75%, which closely aligns with the classical approach calculated for the modified dataset (76%).

## 6 Conclusion

Time series in the time domain have proven to be ineffective for classifying data when these processes have similar systems or very close variability. Given this context, the frequency domain was chosen as the source of information to separate the series instead of the time domain. In the frequency domain, it is possible to discriminate and separate the series by frequency rather than the origin of the entire series. Additionally, noise information and the process autocorrelation structure are mixed in the spectrum due to the convolution used to transform the data from the time domain to the frequency domain. Considering the constant contribution of white noise to the total variability of the process, the most important factor for distinguishing series among populations is the autocorrelation structure. Furthermore, the spectrum of a time series always has an internal product characteristic between the autocorrelations and the noise caused by the transformation. Consequently, they can be separated using a logarithmic function, preserving all information as it is a monotonic transformation of the spectrum. Through the inverse Fourier transform, or deconvolution, this information can return to the time domain with the noise and autocorrelation information separated and summarized, referred to as cepstral coefficients. These coefficients can then be used to distinguish between processes from different populations.

Additionally, this study aims to implement a model considering the spectral variability of each replica (time series) within each population. Since the processes of each population can be very similar, time series from different populations can mix, making discrimination and classification even more difficult. If the series are contaminated by an extreme value that behaves as an additive outlier, the entire correlation structure of the process is lost, along with its predictive capacity for discrimination and classification among populations.

To address this challenge, this study proposed a new model that considers the spectral variability of the replicas within each population and obtains robust cepstral estimates using the  $M$ -periodogram estimator. This approach allows the model to statistically accommodate contamination by abrupt values. In this study, the robust cepstrum is referred to as  $M$ -cepstrum.

In the context of asymptotic analysis, the asymptotic results of the  $M$ -periodogram can be used to show that the cepstral coefficients also tend to converge as the number of replicas ( $n_j$ ), the length of the series ( $N$ ), and the number of cepstral coefficients ( $L$ ) increase. Additionally, it was shown that the cepstral coefficients can be truncated to a fixed value for the  $M$ -cepstrum, similar to what was shown by Krafty (2016) for the cepstrum based on the classical periodogram. Empirically, the mean squared errors of both estimators tend to converge to the true value of the cepstrum as the sample size increases  $N$ , confirming the results demonstrated conduct the same analysis increasing  $R$ ,  $n_j$ , and  $L$ .

Once the  $M$ -periodogram estimates are obtained, it was demonstrated that  $M$ -cepstra can be used for discrimination and classification using linear discriminant analysis based on Fisher (1936)'s method. It was observed that, through this method, a significantly smaller dimension can be used for this purpose, as the echoes containing most of the time series information are the closest in terms of autocorrelation. This information is contained in the first  $M$ -cepstral coefficients.

To validate the effectiveness of the model, various types of simulations were conducted, with a focus on Monte Carlo simulations to analyze the asymptotic properties of classification rates. These simulations demonstrated the effect of extreme information loss in the classifier using the classical periodogram when the series is contaminated by an abrupt value. The  $M$ -cepstrum proved to be more effective in discriminating and classifying the time series in this context, especially for larger series sizes  $N$ , more replicas  $n_j$ , and a wide range of white noise variance possibilities. Although the robust model proposed in this study has lower

classification rates in the absence of extreme values, it performs better with the presence of such values, which behave as additive outliers, providing important information for time series discrimination and classification.

A practical application was conducted to evaluate the proposed M-cepstrum using gait cycle data from healthy individuals, those with amyotrophic lateral sclerosis, and those with Huntington's disease. It was shown that, in the same context used by Krafty (2016), fewer transformations were necessary, resulting in more satisfactory outcomes for determining whether an individual has a disease and, if so, identifying the type of disease.

For future work, there are several perspectives for theoretical, applied, and computational advancements. In the theoretical and applied contexts, as proposed by Krafty (2016), data from the right foot sensor of individuals are used for analysis. Additionally, information from heart rate and electroencephalograms could also be used for the same time units as a source of information. However, the database contains several other sources of information, such as data from the left foot. In this context, more than one piece of information can be used for each time unit, resulting in a multivariate time series in the context of unit cuts. In other words, more than one measure per unit of time. Instead of using only one piece of information per individual at each time, multiple pieces of information can be used in the analysis. The model should include not only the spectra of the left and right foot series but also the cross-spectrum of these two series. Consequently, a multivariate model would emerge for the cepstra, with information not only from the cepstral coefficients but also a matrix of coefficients containing both these coefficients and a cross-cepstrum coefficient. In this context, the M-cepstrum can be developed into a multivariate vector. This could result in higher classification rates for the identification and monitoring of neurodegenerative diseases.

In terms of computational advancements, it is noteworthy that the M-periodogram is currently calculated using a robust regression method. In this case, a matrix of the type  $X'X$  must be calculated with an internal product of order  $N^2$ . However, it is already known that, in pairs, the vectors of the  $X$  matrix are orthogonal due to the orthogonal complement characteristics of Fourier series. Consequently, several of these products are being calculated unnecessarily, making the estimations and simulations extremely slow. This issue, besides hindering Monte Carlo simulations, limited the progress of the Shiny app developed in this study and, consequently, the practical use of the technique suggested. In the classical model estimation process, this problem is mitigated by using the *Fast Fourier Transform* (FFT) algorithm, which, under certain conditions regarding the series length, yields the same result. In this study, the FFT can be used if the regression residual does not exceed the Huber constant in absolute value. This change could make the model significantly faster and more efficient, and consequently more useful for the practical use of time series.

## References

- (2014). Ieee recommended practice and requirements for harmonic control in electric power systems. *IEEE*, (7):1–29.
- Alagón, J. (1986). Spectral discrimination for two groups of time series. *Journal of time Series Analysis*, 10.
- Alpydin, E. (2014). *Introduction to Machine Learning*. MIT press, 3 edition.
- Anderson, O. D. (1976). *Time Series Analysis and Forecasting: The Box and Jenkins Approach*. Butterworths.

- Bloomfield, P. (2000). *Fourier Analysis of time Series: an introduction*. John Wiley Sons, 2 edition.
- Bogert, B. P., Healy, M. J. R., and Turkey, J. W. (1963). The quefreny alanalysis of time series for echoes: Cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. In *Proc. Symposium Time Series Analysis, 1963*, pages 209–243.
- Box, G. E. P. and Jenkins, G. M. (1970). *Times Series Analysis: Forecasting and Control*. Holden-Day.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis - Forecasting and control*. Prentice Hall, <http://www.prenticehall.com/>.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, <https://link.springer.com/>.
- Cavicchioli, M. (2020). Generalised cepstral models for the spectrum of vector time series. *Electronic Journal of Statistics*, 14(14):1935–7524.
- Fajardo, F. A., Reisen, V. A., Lévy-Leduc, C., and Taqqu, M. (2018). M-periodogram for the analysis of long-range-dependent time series. *Statistics*, 52(3):665–683.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Fenetics*, 8(2):179–188.
- Fokianos, K. and Savvides, A. (2008). On comparing several spectral densities. *Technometrics*, 50:317–331.
- Fox, A. J. (1972). Outliers in time series. *J. R. Stat. Soc., Ser. B (Methodol.)*, 34(3):350–363.
- Golub, G. H. and Loan, C. F. V. (2013). *Matrix Computations*. The Johns Hopkins University Press, <https://www.press.jhu.edu/>.
- Hamilton, J. D. (1994). *Times Series Analysis*. Princeton.
- Hausdoff, J. M., Lertratanakul, A., Cudkowicz, M. E., Peterson, A. L., Kaliton, D., and Goldberger, A. L. (2000). Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of Applied Physiology*.
- Huber, P. J. and Ronchetti, E. (2009). *Robust Statistics*. Hohn Wiley Sons, Inc, <https://www.wiley.com/en-us>.
- Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441):328–340.
- Katkovinik, V. (1998). Robust M-periodogram. *IEEEk Transactions on Signal Processing*, 46(11):3104–3109.
- Krafty, R. T. (2016). Discriminant analysis of time series in the presence of within-group spectral variability. *Journal of Time Series*, 37:435–450.
- Kutz, J. N. and Brunton, S. L. (2019). *Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control*. Cambridge Univesity Press.
- Lazzeri, F. (2020). *Machine Learning for Time Series Forecasting with Python*. O'REILLI.

- Li, T.-H. (2008). Laplace periodogram for time series analysis. *J. Am. Stat. Assoc.*, 103(482):757–768.
- Li, T.-H. and Song, K.-S. (2008). Estimation of the parameters of sinusoidal signals in non-Gaussian noise. *IEEE Trans. Signal Process.*, 57(1):62–72.
- Lévy-Leduc, C., Bondon, P., and Reisen, V. (2022). A spectral approach to estimate the autocovariance function. *Journal of Statistical Planning and Inference*, 221:281–298.
- Mukherjee, K. (2008). M-estimation in garch models. *Econometric Theory*, (24):1530–1553.
- Oppenheim, A. v. and Schaffer, R. (2004). From frequency to quefrequency: A history of the cepstrum. *IEEE Signal Processing Magazine*, (4):95–106.
- Oppenheim, A. v., Schaffer, R., and Stockham, T. (1968). Nonlinear filtering of multiplied and convolved signals. *IEEE transactions on audio and electroacoustics*, 16(3):437–466.
- Persival, D. B. and Waden, A. T. (1993). *Spectral Analysis for Physical Applications - Multitaper and Conventional Univariate Techniques*. Cambridge University press.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, [https://openlibrary.org/publishers/Academic\\_press\\_inc\\_\(London\)\\_Ltd](https://openlibrary.org/publishers/Academic_press_inc_(London)_Ltd).
- Reisen, V. A., Lévy-Leduc, C., Cotta, H. H. A., Bondon, P., Ispany, M., and Filho, P. R. P. (2020a). An overview of robust spectral estimators. In Chaari, F., Leskow, J., Zimroz, R., Wyłomańska, A., and Dudek, A., editors, *Cyclostationarity: Theory and Methods – IV*, pages 204–224, Cham. Springer International Publishing.
- Reisen, V. A., Lévy-Leduc, C., and Taqqu, M. S. (2017). An M-estimator for the long-memory parameter. *J. Stat. Plan. Inference*, 187:44 – 55.
- Reisen, V. A., Lévy-Leduc, C., Cotta, H. H. A., Bondon, P., Ispany, M., and Filho, P. R. P. (2020b). An overview of robust spectral estimators. *Springer Nature Switzerland*, pages 01–23.
- Schafer, R. W. (1969). Echo removal by discrete generalized linear filtering.
- Shin, H. (2008). An extension of fisher’s discriminant analysis for stochastic processes. *Journal of Multivariate Time Series*, 99:1191–1216.
- Shumway, R. H. (1982). Discriminant analysis for time series. *Handbook of Statistics*, 2:1–46.
- Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Application*. Springer, 3 edition.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70:1055–1096.
- Wichern, D. W. and Johnson, R. A. (2007). *Applied Multivariate Statistical Analysis*. Pearson, 6 edition.
- Zhang, G. and Taniguchi, M. (1992). Discriminant analysis for time vector time series. *Journal of Time Series*, 15:117–126.
- Zhang, Z. and Chan, S.-C. (2005). Robust adaptive lomb periodogram for time-frequency analysis of signals with sinusoidal and transient components. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP’05). IEEE International Conference on*, volume 4, pages iv–493. IEEE.

## 7 Appendix: PROOFS

### 7.1 Proof of Proposition 1:

*Proof.* Let  $\mathbb{A} = \ln \left\{ 1 + \eta_{i,jk}^2 + 2\eta_{i,jk} \cos(\lambda) \right\}$  and  $\mathbb{B} = \ln \left\{ 1 + \zeta_{r,jk}^2 + 2\zeta_{r,jk} \cos(\lambda) \right\}$  for all  $i = 1, \dots, q$  and  $r = 1, \dots, p$ . Then:

Applying Euler's Formula to cosine term in  $\mathbb{A}$ :

$$\begin{aligned} \mathbb{A} &= \ln \left\{ 1 + \eta_{i,jk}^2 + 2\eta_{i,jk} \left[ \frac{e^{i\lambda} + e^{-i\lambda}}{2} \right] \right\} = \ln \left\{ 1 + \eta_{i,jk}^2 + \eta_{i,jk} e^{i\lambda} + \eta_{i,jk} e^{-i\lambda} \right\} \\ \mathbb{A} &= \ln \left\{ (1 + \eta_{i,jk} e^{i\lambda}) (1 + \eta_{i,jk} e^{-i\lambda}) \right\} = \left\{ \ln (1 + \eta_{i,jk} e^{i\lambda}) + \ln (1 + \eta_{i,jk} e^{-i\lambda}) \right\} \\ \mathbb{A} &= \mathbb{C} + \mathbb{D} \end{aligned}$$

Using Taylor expansion for  $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$

$$\begin{aligned} \mathbb{C} &= \ln (1 + \eta_{i,jk} e^{i\lambda}) = (\eta_{i,jk} e^{i\lambda}) - \frac{(\eta_{i,jk} e^{i\lambda})^2}{2} + \frac{(\eta_{i,jk} e^{i\lambda})^3}{3} - \frac{(\eta_{i,jk} e^{i\lambda})^4}{4} + \dots \\ &= \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} e^{i\lambda \ell}}{\ell} \\ \mathbb{D} &= \ln (1 + \eta_{i,jk} e^{-i\lambda}) = (\eta_{i,jk} e^{-i\lambda}) - \frac{(\eta_{i,jk} e^{-i\lambda})^2}{2} + \frac{(\eta_{i,jk} e^{-i\lambda})^3}{3} - \frac{(\eta_{i,jk} e^{-i\lambda})^4}{4} + \dots \\ &= \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} e^{-i\lambda \ell}}{\ell} \end{aligned}$$

Therefore,  $\mathbb{A}$  results in:

$$\begin{aligned} \mathbb{A} &= \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} e^{i\lambda \ell}}{\ell} + \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} e^{-i\lambda \ell}}{\ell} \\ \mathbb{A} &= \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} [e^{i\lambda \ell} + e^{-i\lambda \ell}]}{\ell} \\ \mathbb{A} &= 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} \cos(\lambda \ell)}{\ell} \end{aligned}$$

Following the same logic, it is possible to show that

$$\mathbb{B} = 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \zeta_{r,jk}^{\ell} \cos(\lambda \ell)}{\ell}$$

Summing  $\mathbb{A}$  and  $\mathbb{B}$  we have

$$\begin{aligned} \mathbb{A} + \mathbb{B} &= 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \eta_{i,jk}^{\ell} \cos(\lambda \ell)}{\ell} + 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \zeta_{r,jk}^{\ell} \cos(\lambda \ell)}{\ell} \\ &= 2 \sum_{\ell=1}^{\infty} \frac{1}{\ell} \left( (-1)^{\ell+1} \eta_{i,jk}^{\ell} + (-1)^{\ell+1} \zeta_{r,jk}^{\ell} \right) \cos(\lambda \ell) \end{aligned}$$

The first part *i)* is proven. And the terms of the series in the last sum are  $c_{\ell}$  for each fixed  $\ell$  as described in second part of the corollary *ii)*.

□

## 7.2 Proof of Corollary 1:

*Proof.*  $\textcircled{A} = \ln \left\{ 1 + \theta_{jk}^2 + 2\theta_{jk}\cos(\lambda) \right\}$

Applying Euler's Formula to cosine term:

$$\textcircled{A} = \ln \left\{ 1 + \theta_{jk}^2 + 2\theta_{jk} \left[ \frac{e^{i\lambda} + e^{-i\lambda}}{2} \right] \right\} = \ln \left\{ 1 + \theta_{jk}^2 + \theta_{jk}e^{i\lambda} + \theta_{jk}e^{-i\lambda} \right\}$$

$$\textcircled{A} = \ln \left\{ (1 + \theta_{jk}e^{i\lambda}) (1 + \theta_{jk}e^{-i\lambda}) \right\} = \left\{ \ln (1 + \theta_{jk}e^{i\lambda}) + \ln (1 + \theta_{jk}e^{-i\lambda}) \right\}$$

$$\textcircled{A} = \textcircled{B} + \textcircled{C}$$

Using Taylor expansion for  $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$

$$\textcircled{B} = \ln (1 + \theta_{jk}e^{i\lambda}) = (\theta_{jk}e^{i\lambda}) - \frac{(\theta_{jk}e^{i\lambda})^2}{2} + \frac{(\theta_{jk}e^{i\lambda})^3}{3} - \frac{(\theta_{jk}e^{i\lambda})^4}{4} + \dots = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell}e^{i\lambda\ell}}{\ell}$$

$$\textcircled{C} = \ln (1 + \theta_{jk}e^{-i\lambda}) = (\theta_{jk}e^{-i\lambda}) - \frac{(\theta_{jk}e^{-i\lambda})^2}{2} + \frac{(\theta_{jk}e^{-i\lambda})^3}{3} - \frac{(\theta_{jk}e^{-i\lambda})^4}{4} + \dots = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell}e^{-i\lambda\ell}}{\ell}$$

Therefore,  $\textcircled{A}$  results in:

$$\textcircled{A} = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell}e^{i\lambda\ell}}{\ell} + \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell}e^{-i\lambda\ell}}{\ell}$$

$$\textcircled{A} = \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell} [e^{i\lambda\ell} + e^{-i\lambda\ell}]}{\ell}$$

$$\textcircled{A} = 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}\theta_{jk}^{\ell} \cos(\lambda\ell)}{\ell}$$

$$\textcircled{A} = 2 \left( \sum_{\ell=1}^{\infty} \frac{1}{\ell} (-1)^{\ell+1} \theta_{jk}^{\ell} \cos(\lambda\ell) \right) d$$

The first part *i)* is proven. And the terms of the series in the last sum are  $c_{\ell}$  for each fixed  $\ell$  as described in second part of the corollary *ii)*.  $\square$

## 7.3 Proof of Corollary 2:

*Proof.*  $\textcircled{A} = \ln \left\{ \frac{1}{1 + \phi_{jk}^2 - 2\phi_{jk}\cos(\lambda)} \right\}$

$$\textcircled{A} = \ln \left\{ 1 + \phi_{jk}^2 - 2\phi_{jk}\cos(\lambda) \right\}^{-1} = -\ln \left\{ 1 + \phi_{jk}^2 - 2\phi_{jk}\cos(\lambda) \right\}$$

Applying Euler's Formula to cosine term:

$$\textcircled{A} = -\ln \left\{ 1 + \phi_{jk}^2 - 2\phi_{jk} \left[ \frac{e^{i\lambda} + e^{-i\lambda}}{2} \right] \right\} = -\ln \left\{ 1 + \phi_{jk}^2 - \phi_{jk}e^{i\lambda} - \phi_{jk}e^{-i\lambda} \right\}$$

$$\textcircled{A} = -\ln \left\{ (1 - \phi_{jk}e^{i\lambda}) (1 - \phi_{jk}e^{-i\lambda}) \right\} = - \left\{ \ln (1 - \phi_{jk}e^{i\lambda}) + \ln (1 - \phi_{jk}e^{-i\lambda}) \right\}$$

$$\textcircled{A} = - [\textcircled{B} + \textcircled{C}]$$

Using Taylor expansion for  $\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots$

$$\begin{aligned} \textcircled{B} &= \ln(1 - \phi_{jk}e^{i\lambda}) = -(\phi_{jk}e^{i\lambda}) - \frac{(\phi_{jk}e^{i\lambda})^2}{2} - \frac{(\phi_{jk}e^{i\lambda})^3}{3} - \frac{(\phi_{jk}e^{i\lambda})^4}{4} - \dots = -\sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{i\lambda\ell}}{\ell} \\ \textcircled{C} &= \ln(1 - \phi_{jk}e^{-i\lambda}) = -(\phi_{jk}e^{-i\lambda}) - \frac{(\phi_{jk}e^{-i\lambda})^2}{2} - \frac{(\phi_{jk}e^{-i\lambda})^3}{3} - \frac{(\phi_{jk}e^{-i\lambda})^4}{4} - \dots = \\ &= -\sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{-i\lambda\ell}}{\ell} \end{aligned}$$

Therefore,  $\textcircled{A}$  results in:

$$\begin{aligned} \textcircled{A} &= -\left[ -\sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{i\lambda\ell}}{\ell} - \sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{-i\lambda\ell}}{\ell} \right] \\ \textcircled{A} &= \sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{i\lambda\ell}}{\ell} + \sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} e^{-i\lambda\ell}}{\ell} \\ \textcircled{A} &= \sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell} [e^{i\lambda\ell} + e^{-i\lambda\ell}]}{\ell} \\ \textcircled{A} &= 2 \sum_{\ell=1}^{\infty} \frac{\phi_{jk}^{\ell}}{\ell} \cos(\lambda\ell) \\ \textcircled{A} &= 2 \left( \sum_{\ell=1}^{\infty} \frac{1}{\ell} \phi_{jk}^{\ell} \cos(\lambda\ell) \right) \end{aligned}$$

The first part *i)* is proven. And the terms of the series in the last sum are  $c_{\ell}$  for each fixed  $\ell$  as described in second part of the corollary *ii)*.  $\square$

## 7.4 Proof of Proposition 2:

*Proof.* Suppose that  $\mathbf{y} = B^{1/2}\mathbf{x}$  a changing coordinates, which implies  $\mathbf{x} = B^{-1/2}\mathbf{y}$ . Also, suppose that  $D_A$  and  $D_B$  are diagonal matrices with of  $A$  and  $B$ , with its respective eigenvalues in its principal diagonal. Additionally, suppose that  $P_A = P_B$ , where  $P_A$  and  $P_B$  are the respective eigenvector matrices, associated to respective eigenvalues. Then,

$$\begin{aligned} \mathbf{x}_1 &= \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^L} \left\{ \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T B \mathbf{x}} \right\} \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{(B^{-1/2}\mathbf{y})^T A (B^{-1/2}\mathbf{y})}{(B^{-1/2}\mathbf{y})^T B (B^{-1/2}\mathbf{y})} \right\} \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T B^{-1/2} A B^{-1/2} \mathbf{y}}{\mathbf{y}^T B^{-1/2} B B^{-1/2} \mathbf{y}} \right\} \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T B^{-1/2} A B^{-1/2} \mathbf{y}}{\mathbf{y}^T B^{-1/2} B^{1/2} B^{1/2} B^{-1/2} \mathbf{y}} \right\} \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T B^{-1/2} A B^{-1/2} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\} \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T (P_B D_B^{-1/2} P_B^T) (P_A D_A P_A^T) (P_B D_B^{-1/2} P_B^T) \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\} \end{aligned}$$

$$\begin{aligned}
&= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T P_B D_B^{-1/2} D_A D_B^{-1/2} P_B^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}. \\
&\equiv \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T P_B D_B^{-1} D_A P_B^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}. \\
&= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T (P_B D_B^{-1}) (D_A) P_B^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}. \\
&= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T (B^{-1} P_B) (P_A^T A P_A) P_B^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\}. \\
&= \operatorname{argmax}_{\mathbf{y} \in \mathbb{R}^L} \left\{ \frac{\mathbf{y}^T B^{-1} A \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \right\} = \mathbf{y}_1.
\end{aligned}$$

□

**Lemma 3.** As a direct consequence of proposition 4,  $E(I_{jk}^{MR}(\lambda_m)) \approx E(I_{jk}^R(\lambda_m)) = S_X(\lambda_m)$

## 7.5 Proof of Theorem 2:

*Proof.* After constructing the multitaper M-periodogram through an averaging process of periodograms derived from orthogonal data sine tapers, the proof can proceed directly by applying Chebyshev inequality. So, consider that the for each  $r$ , the periodograms are independent due to orthogonality of the tapers and under 1,  $E|X_{jkt}|^4 < \infty \Rightarrow \operatorname{Var}(I_{jk}^{rM}(\lambda_m)) < \infty$  when  $R = R_N \rightarrow \infty$  (as defined in 4. Also, from Equation 3.3,

$$I_{jk}^{MR}(\lambda_m) = \frac{1}{R} \sum_{r=1}^R I_{jk}^{rM}(\lambda_m)$$

Since  $I_{jk}^{rM}(\lambda_m)$  are independent with finite variance, the variance of the mean  $I_{jk}^{MR}(\lambda_m)$  is:

$$\begin{aligned}
\operatorname{Var}(I_{jk}^{MR}(\lambda_m)) &= \operatorname{Var}\left(\frac{1}{R} \sum_{r=1}^R I_{jk}^{rM}(\lambda_m)\right) \\
&= \frac{1}{R^2} \sum_{r=1}^R \operatorname{Var}(I_{jk}^{rM}(\lambda_m))
\end{aligned}$$

For each  $r$ , let the finite variance be  $\operatorname{Var}(I_{jk}^{rM}(\lambda_m)) = \sigma_r^2$ . Then, for any given  $\epsilon > 0$  and

$$\begin{aligned}
\mathbb{P}(|I_{jk}^{MR}(\lambda_m) - S_X(\lambda_m)| > \epsilon) &= \mathbb{P}\left(|I_{jk}^{MR}(\lambda_m) - S_X(\lambda_m)|^2 > \epsilon^2\right) \leq \frac{\operatorname{Var}[I_{jk}^{MR}(\lambda_m)]}{\epsilon^2} \\
&= \frac{\sigma_r^2}{R\epsilon^2} \xrightarrow{p} 0, \text{ when } R \rightarrow \infty
\end{aligned}$$

Note that under Assumption1, the fourth moment of  $X_{jkt}$  is bounded, then  $E|X_{jkt}|^4 < \infty \implies \operatorname{var}(I_{jk}^{Mr}(\lambda_m)) < \infty$  for fixed  $j, k, \lambda_m$ .

□

## 7.6 Proof of Lemma 2:

*Proof.* Under the theorem 1 and the continuous mapping theorem,  $E[I^M(\lambda_m)] \rightarrow S_X(\lambda_m) \implies E\{\ln[I^M(\lambda_m)]\} \rightarrow \ln[S_X(\lambda_m)]$ .  $\square$

## 7.7 Proof of Proposition 4:

*Proof.*

$$|\hat{\gamma}^M(\tau) - \hat{\gamma}(\tau)| = |\hat{\gamma}^M(\tau) - \hat{\gamma}(\tau) + \gamma(\tau) - \gamma(\tau)| \quad (1)$$

$$= |\hat{\gamma}^M(\tau) - \gamma(\tau) + \gamma(\tau) - \hat{\gamma}(\tau)| \quad (2)$$

$$\leq |\hat{\gamma}^M(\tau) - \gamma(\tau)| + |\gamma(\tau) - \hat{\gamma}(\tau)| = 0 \quad (3)$$

The last term in the right term of the inequality is zero as showed in Lévy-Leduc et al. (2022) and the first term because

$$\lim_{N \rightarrow \infty} |\hat{\gamma}_N^M(\tau) - \hat{\gamma}(\tau)| = |\gamma_\psi(\tau) - \gamma(\tau)| \quad (4)$$

$$\leq |a\gamma(\tau) - \gamma(\tau)| = |(1-a)\gamma(\tau)| = 0, \text{ when } a \rightarrow 1 \quad (5)$$

The final result for (3) shows that  $|\hat{\gamma}^M(\tau) - \hat{\gamma}(\tau)| \xrightarrow{P} 0$ . And using proposition 6.1.3 in Brockwell and Davis (1991), when  $|\hat{\gamma}_N^M(\tau) \xrightarrow{P} \gamma_\psi(\tau)|$ ,  $|\hat{\gamma}_N(\tau) \xrightarrow{P} \gamma(\tau)|$  and (3), then  $\hat{\gamma}^M(\tau) \xrightarrow{P} \gamma(\tau)$   $\square$

## 7.8 Proof of Theorem 3:

*Proof.* For all  $j$ th population and  $k$ th replicates, if  $\hat{c}_{jkL}$  is the cepstral classical periodogram based, Krafty (2016) showed that  $|\hat{c}_{jkL} - c_{jk\ell}| = \mathcal{O}_p(1)$  and it is easy to show that  $|\hat{c}_{jkL}^M - \hat{c}_{jkL}| = \mathcal{O}_p(1)$ . Then, by the triangular inequality we have:

$$|\hat{c}_{jkL}^M - c_{jk\ell}| = |\hat{c}_{jkL}^M - \hat{c}_{jkL} + \hat{c}_{jkL} - c_{jk\ell}| \leq |\hat{c}_{jkL}^M - \hat{c}_{jkL}| + |\hat{c}_{jkL} - c_{jk\ell}| = \mathcal{O}_p(1) \quad \square$$

## 7.9 Proof of Theorem 4:

*Proof.* Let  $\{\hat{c}_{jk\ell}^M\}_{\ell=1}^\infty$  is a sequence of estimated cepstral coefficients based on M-Periodogram and  $\{\hat{p}_{s\ell}^M\}_{\ell=1}^\infty$  the truncated weights functions. The distance between discriminant estimator in the Equation 3.5 of the true discriminant function defined in the Equation 2.21 is defined as

$$\left| \hat{d}_{jks}^M - d_{jks} \right| = \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M \hat{c}_{jk\ell}^M - \sum_{\ell=1}^{\infty} p_{jks\ell} c_{jk\ell} \right|$$

Adding and subtracting  $\sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jk\ell}$  inside of the modulo on the right side we get

$$= \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M \hat{c}_{jk\ell}^M - \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jk\ell} + \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jk\ell} - \sum_{\ell=1}^{\infty} p_{jks\ell} c_{jk\ell} \right|$$

$$= \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M (\hat{c}_{jkl} - c_{jkl}) + \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jkl} - \sum_{\ell=1}^{\infty} p_{jks\ell} c_{jkl} \right|$$

And using triangular inequality we get

$$\leq \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M (\hat{c}_{jkl} - c_{jkl}) \right| + \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jkl} - \sum_{\ell=1}^{\infty} p_{jks\ell} c_{jkl} \right|$$

Taking the limit of probability in both sides we get

$$\begin{aligned} \lim_{L \rightarrow \infty} \mathbb{P} \left| \hat{d}_{jks}^M - d_{jks} \right| &\leq \lim_{L \rightarrow \infty} \mathbb{P} \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M (\hat{c}_{jkl} - c_{jkl}) \right| + \lim_{L \rightarrow \infty} \mathbb{P} \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M c_{jkl} - \sum_{\ell=1}^{\infty} p_{jks\ell} c_{jkl} \right| \\ &= \lim_{L \rightarrow \infty} \mathbb{P} \left| \sum_{\ell=1}^{L-1} \hat{p}_{jks\ell}^M (\hat{c}_{jkl} - c_{jkl}) \right| + \lim_{L \rightarrow \infty} \mathbb{P} \left| \sum_{\ell=1}^{L-1} (\hat{p}_{jks\ell}^M - p_{jks\ell}) c_{jkl} \right| = \mathcal{O}_p(1) \end{aligned}$$

The first term is due to the 2 and the second term is due  $\hat{p}_{s\ell}$  is surprisingly to be bounded in probability.

□

## Capítulo 5

### Conclusões

Séries temporais no domínio do tempo mostram-se ineficazes quanto ao contexto de classificá-las no respectivo processo gerador de dados quando estes processos têm sistemas similares ou com variabilidade muito próxima. Devido a este contexto, escolheu-se utilizar o domínio da frequência como fonte de informação para separar as séries, ao invés do domínio do tempo. Contudo, no domínio da frequência, é possível discriminar e separar as séries por frequência e não de onde a série inteira é originada. Adicionalmente, as informações do ruído e da estrutura de autocorrelação do processo são misturadas no espectro por um produto, devido à convolução utilizada para transformar os dados do domínio do tempo no domínio da frequência. E devido à constância da contribuição do ruído branco para a variabilidade total do processo, o mais importante para distinguir as séries dentre as populações é a estrutura de autocorrelação. Além disso, o espectro de uma série temporal tem sempre uma característica de produto interno entre as autocorrelações e o ruído causado pela transformação dos dados originais no domínio do tempo no domínio da frequência. Conseqüentemente, elas podem ser separadas por meio da função logarítmica, preservando toda a informação por se tratar de uma transformação monótona do espectro. E por meio da transformada inversa de Fourier, uma deconvolução, estas informações podem retornar ao domínio temporal com as informações do ruído e das autocorrelações separadas e resumidas, chamadas de coeficientes cepstrais. E tais coeficientes podem ser utilizados para discernir entre os processos oriundos de populações distintas.

Adicionalmente, este trabalho se propôs a implementar um modelo no contexto de variabilidade espectral de cada réplica (série temporal) dentro de cada população. Como os processos de cada população podem ser muito similares, as séries temporais oriundas de populações distintas podem se misturar, dificultando ainda mais sua discriminação e classificação. E mais ainda, caso as séries sejam contaminadas por um valor extremo com comportamento de *outlier* aditivo, toda a estrutura de correlação do processo é perdida, assim como sua capacidade preditiva de discriminação e classificação entre as populações.

Para lidar com esse desafio, este trabalho propôs um novo modelo que não apenas leve em consideração a variabilidade espectral das réplicas dentro de cada população, mas também que a estimação dos coeficientes cepstrais possa ser obtida de forma robusta, originada no estimador M-periodograma. Neste caso, as contaminações por valores abruptos

podem ser acomodadas de forma estatística pelo processo de estimação do próprio modelo. Neste trabalho, o cepstra robusto foi chamado de M-cepstra.

No contexto de análise assintótica, os resultados assintóticos do M-periodograma podem ser utilizados para mostrar que os coeficientes de cepstrais também tendem a se aproximar à medida que o número de réplicas ( $n_j$ ), tamanho das séries ( $N$ ) e quantidade de coeficientes cepstrais ( $L$ ) aumentam. Além disso, mostrou-se também que os coeficientes cepstrais podem ser truncados em um valor fixo para o M-cepstra, de forma similar àquela mostrada por [Krafty \(2016\)](#) para o cepstra baseado no periodograma clássico. Empiricamente, os erros quadráticos médios de ambos os estimadores tendem a se aproximar do verdadeiro valor do cepstra à medida que o tamanho da amostra aumenta  $N$ , o que confirma os resultados feitos pelas demonstrações (*fazer a mesma análise aumentando  $R$ ,  $n_j$  e  $L$* ).

Uma vez obtidas as estimativas do M-periodograma, foi demonstrado que os M-cepstrais podem ser utilizados para discriminação e classificação utilizando a análise discriminante linear baseada no método de [Fisher \(1936\)](#). Observou-se que, por meio deste método, uma dimensão significativamente menor pode ser utilizada para tal finalidade, já que os ecos contendo a maior parte da informação das séries temporais são os mais próximos em termos de autocorrelação. Essas informações estão contidas nos primeiros coeficientes M-cepstrais.

Para validar a eficácia do modelo, diversos tipos de simulações foram conduzidas, destacando-se a simulação de Monte Carlo para análise das propriedades assintóticas das taxas de classificação. Por meio destas simulações, foi possível mostrar o efeito de perda de informação extrema no classificador utilizando o periodograma clássico quando a série é contaminada por um valor abrupto. O M-cepstra mostrou-se mais eficaz na discriminação e classificação da série temporal neste contexto, sobretudo para maiores tamanhos de séries  $N$ , para mais réplicas  $n_j$  e com alto intervalo de possibilidade da variância do ruído branco. Embora o modelo robusto proposto neste trabalho tenha taxas de classificação menores na ausência de valores extremos, ele performa melhor com a presença de tais valores com comportamento de outliers aditivos, sendo ainda uma fonte de informação importante para a discriminação e classificação de séries temporais.

Uma aplicação prática foi realizada para avaliar o desempenho do M-cepstra proposto utilizando dados de ciclos de passadas de indivíduos saudáveis, com doença de esclerose lateral amiotrófica e *Huntington*. Demonstrou-se que, no mesmo contexto utilizado por [Krafty \(2016\)](#), menos transformações foram necessárias, gerando resultados mais satisfatórios para discernir se um indivíduo tem ou não a doença e qual é o tipo de doença caso tenha.

Como extensão deste trabalho, existem vários pontos de vista para avanços no contexto teórico, aplicado e computacional. No contexto dos dois primeiros, assim como proposto por [Krafty \(2016\)](#), utilizam-se os dados do sensor do pé direito dos indivíduos para a

---

análise. Além disso, informações do batimento cardíaco e do eletroencefalograma também poderiam ser utilizados para mesmas unidades de tempo como fonte de informação. Contudo, na base de dados, existem várias outras fontes de informação, como os dados do pé esquerdo. Neste contexto, mais de uma informação pode ser utilizada para cada unidade de tempo. Assim, trata-se de uma série temporal multivariada no contexto de unidade de corte. Ou seja, mais de uma medida por unidade de tempo. E neste caso, ao invés de utilizar apenas uma informação por indivíduo em cada tempo, várias informações podem ser utilizadas na análise. O modelo deve incluir não somente os espectros da série do pé esquerdo e do pé direito, mas também o espectro cruzado destas duas séries, por exemplo. Conseqüentemente, um modelo multivariado surgiria para os cepstrais, com informações não somente dos coeficientes cepstrais mas uma matriz de coeficientes contendo não somente tais coeficientes, mas também um coeficiente cruzado. E neste contexto, o M-cepstra pode ser desenvolvido para um vetor multivariado. Neste caso, a identificação e o acompanhamento de doenças neurodegenerativas podem gerar taxas de classificação maiores.

No contexto do avanço computacional, destaca-se o fato de que o M-periodograma é atualmente calculado por meio de um método de regressão robusto. Neste caso, uma matriz do tipo  $X'X$  deve ser calculada com o produto interno de ordem  $N^2$ . Contudo, já se sabe que, de duas em duas, os vetores da matriz  $X$  são ortogonais devido à característica de complementos ortogonais das séries de Fourier. Conseqüentemente, vários desses produtos estão sendo calculados desnecessariamente, tornando as estimações e simulações extremamente lentas. Este fato, além de ter dificultado as simulações de Monte Carlo, limitou o avanço do aplicativo Shiny desenvolvido neste trabalho e, conseqüentemente, o uso da técnica sugerida na prática. No processo de estimação do modelo clássico, tal problema é subjulgado por meio do uso do algoritmo *Fast Fourier Transform* (FFT), que sob algumas condições sobre o tamanho das séries gera o mesmo resultado. Neste trabalho, poder-se-ia utilizar o FFT caso o resíduo da regressão não ultrapasse, em valor absoluto, a constante de Huber. Essa mudança poderia tornar o modelo significativamente mais rápido, eficiente e, conseqüentemente, mais útil no uso prático de séries temporais.

# Referências

- Alagón, J. (1986). Spectral discrimination for two groups of time series. *Journal of Time Series Analysis*, 10.
- Alpaydin, E. (2014). *Introduction to machine learning* (3rd ed.). MIT Press.
- Anderson, O. D. (1976). *Time series analysis and forecasting: The box and jenkins approach*. Butterworths.
- Bloomfield, P. (2000). *Fourier analysis of time series: An introduction* (2nd ed.). John Wiley Sons.
- Bogert, B. P. (1963). The quefreny alalysis of time series for echoes: Cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. In *Proc. symposium time series analysis, 1963* (pp. 209–243).
- Box, G. E. P., e Jenkins, G. M. (1970). *Times series analysis: Forecasting and control*. Holden-Day.
- Box, G. E. P., Jenkins, G. M., e Reinsel, G. C. (1994). *Time series analysis - forecasting and control*. Prentice Hall.
- Brillinger, D. R. (1981). *Time series: Data analysis and theory*. Holden-Day.
- Brockwell, P. J., e Davis, R. A. (1991). *Time series: Theory and methods*. Springer.
- Cavicchioli, M. (2020). Generalised cepstral models for the spectrum of vector time series. *Electronic Journal of Statistics*, 14(14), 1935-7524.
- Fajardo, F. A., Reisen, V. A., Lévy-Leduc, C., e Taqqu, M. (2018). M-periodogram for the analysis of long-range-dependent time series. *Statistics*, 52(3), 665–683.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics*, 8(2), 179-188.
- Fokianos, K., e Savvides, A. (2008, 08). On comparing several spectral densities. *Technometrics*, 50, 317-331. doi: 10.1198/004017008000000244
- Fox, A. J. (1972). Outliers in time series. *J. R. Stat. Soc., Ser. B (Methodol.)*, 34(3), 350-363.

- Golub, G. H., e Loan, C. F. V. (2013). *Matrix computations*. <https://www.press.jhu.edu/>: The Johns Hopkins University Press.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton.
- Hardle, H., e Simar, L. (2007). *Applied multivariate statistical analysis*. Springer Press.
- Huber, P. J., e Ronchetti, E. (2009). *Robust statistics*. John Wiley Sons, Inc.
- Ieee recommended practice and requirements for harmonic control in electric power systems. (2014). *IEEE*(7), 1-29. doi: 10.1109/IEEESTD.2014.6826459
- Kakizawa, Y., Shumway, R. H., e Taniguchi, M. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association*, 93(441), 328-340.
- Katkovinik, V. (1998). Robust m-periodogram. *IEEE Transactions on Signal Processing*, 46(11), 3104-3109.
- Krafty, R. T. (2016). Discriminant analysis of time series in the presence of within-group spectral variability. *Journal of Time Series*, 37, 435-450.
- Kutz, J. N., e Brunton, S. L. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- Lazzeri, F. (2020). *Machine learning for time series forecasting with python*. O'Reilly.
- Li, T.-H. (2008). Laplace periodogram for time series analysis. *J. Am. Stat. Assoc.*, 103(482), 757-768.
- Li, T.-H., e Song, K.-S. (2008). Estimation of the parameters of sinusoidal signals in non-Gaussian noise. *IEEE Trans. Signal Process.*, 57(1), 62-72.
- Oppenheim, A. V., e Schafer, R. (2004). From frequency to quefreny: A history of the cepstrum. *IEEE Signal Processing Magazine*, 95-106.
- Oppenheim, A. v., Schafer, R., e Stockham, T. (1968). Nonlinear filtering of multiplied and convolved signals. *IEEE transactions on audio and electroacoustics*, 16(3), 437-466.
- Persival, D. B., e Walden, A. T. (1993). *Spectral analysis for physical applications - multitaper and conventional univariate techniques*. Cambridge University Press.
- Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.

- Reisen, V. A., Lévy-Leduc, C., Cotta, H. H. A., Bondon, P., Ispany, M., e Filho, P. R. P. (2020a). An overview of robust spectral estimators. In F. Chaari, J. Leskow, R. Zimroz, A. Wyłomańska, e A. Dudek (Eds.), *Cyclostationarity: Theory and methods – iv* (pp. 204–224). Cham: Springer International Publishing.
- Reisen, V. A., Lévy-Leduc, C., Cotta, H. H. A., Bondon, P., Ispany, M., e Filho, P. R. P. (2020b). An overview of robust spectral estimators. In F. Chaari, J. Leskow, R. Zimroz, A. Wyłomańska, e A. Dudek (Eds.), *Cyclostationarity: Theory and methods – iv* (pp. 204–224). Cham: Springer International Publishing.
- Reisen, V. A., Lévy-Leduc, C., e Taqqu, M. S. (2017). An M-estimator for the long-memory parameter. *J. Stat. Plan. Inference*, 187, 44 - 55.
- Reisen, V. A., Lévy-Leduc, C., Cotta, H. H. A. C., Bondon, P., Ispany, M., e Filho, P. R. P. (2020). An overview of robust spectral estimators. *Springer Nature Switzerland*, 01-23.
- Schafer, R. W. (1969). Echo removal by discrete generalized linear filtering.
- Shin, H. (2008). An extension of fisher’s discriminant analysis for stochastic processes. *Journal of Multivariate Time Series*, 99, 1191-1216.
- Shumway, R. H. (1982). Discriminant analysis for time series. *Handbook of Statistics*, 2, 1-46.
- Shumway, R. H., e Stoffer, D. S. (2011). *Time series analysis and its application* (3rd ed.). Springer.
- Shumway, R. H., e Unger, A. N. (1974). Linear discriminant functions for stationary time series. *Journal of the American Statistical Association*, 69(2), 948-956.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70, 1055-1096.
- Wichern, D. W., e Johnson, R. A. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson.
- Zhang, G., e Taniguchi, M. (1992). Discriminant analysis for time vector time series. *Journal of Time Series*, 15, 117-126.
- Zhang, Z., e Chan, S.-C. (2005). Robust adaptive lomb periodogram for time-frequency analysis of signals with sinusoidal and transient components. In *Acoustics, speech, and signal processing, 2005. proceedings.(icassp’05). ieee international conference on* (Vol. 4, pp. iv–493).

# Apêndice A

## Simulações Adicionais

### A.1 Da simulação da série até a Análise Discriminante

Uma simulação foi realizada para explicar o procedimento passo-a-passo até a discriminação e classificação das 3 réplicas de séries temporais condicionais a 3 grupos. Para tal análise, o M-periodograma foi utilizado como exemplo e os M-cepstra foram calculados e utilizados para discriminação e classificação. Neste contexto em particular, as séries não foram contaminadas com valores extremos por se tratar apenas de um exemplo. Portanto, foram consideradas  $J = 3$ ,  $k = 3$  e  $N = 100$ . E devido às características da aplicação dos dados apresentadas no capítulo 5, as simulações são conduzidas usando um modelo condicional (na população)  $AR(2)$  que pode ser representado por

$$X_{jkt} = \phi_{jk1}X_{jkt-1} + \phi_{jk2}X_{jkt-2} + \epsilon_{jkt} \quad \epsilon_{jkt} \sim N(0, \sigma_{jk}^2) \quad (4.1)$$

onde os parâmetros para cada população são variáveis aleatórias apresentadas na tabela

Tabela A.1: Distribuição dos Parâmetros Condicionais

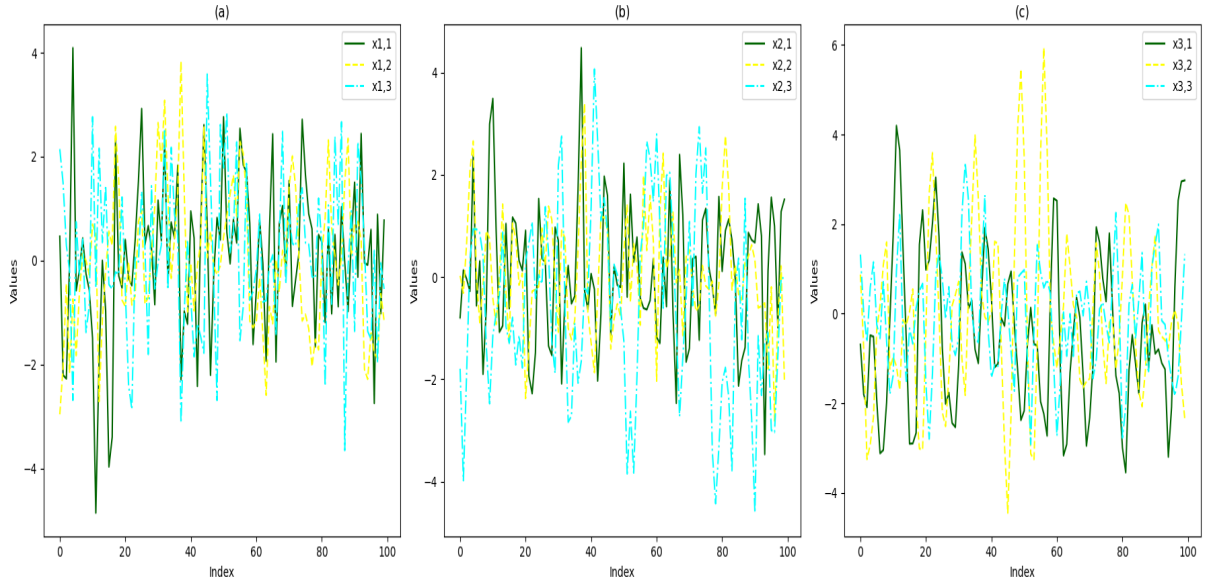
População	Parâmetros	$\phi_{jk1}$	$\phi_{jk2}$	Ruído Branco	Variância do Ruído Branco
$j = 1$		$(\phi_{1k1} j = 1) \sim UNI(0.05, 0.7)$	$(\phi_{1k2} j = 1) \sim UNI(-0.12, -0.06)$	$\epsilon_{1k} \sim N(0, \sigma_{1k}^2)$	$(\sigma_{1k}^2 j = 1) \sim UNI(0.1, 10)$
$j = 2$		$(\phi_{2k1} j = 2) \sim UNI(0.01, 1.2)$	$(\phi_{2k2} j = 2) \sim UNI(-0.36, -0.25)$	$\epsilon_{2k} \sim N(0, \sigma_{2k}^2)$	$(\sigma_{2k}^2 j = 2) \sim UNI(0.3, 3.0)$
$j = 3$		$(\phi_{3k1} j = 3) \sim UNI(0.12, 1.5)$	$(\phi_{3k2} j = 3) \sim UNI(-0.75, -0.56)$	$\epsilon_{3k} \sim N(0, \sigma_{3k}^2)$	$(\sigma_{3k}^2 j = 3) \sim UNI(0.9, 1.1)$

Fonte: Elaborado pelo autor.

Fonte: Elaborado pelo autor.

A Tabela ?? no anexo B estão apresentadas a tabela dos das séries (réplicas) simulados para cada população. Para apresentação visual, as séries foram separadas por população e apresentadas na figura A.2. Note que apenas com o gráfico das séries, caso elas fossem misturadas, não seria fácil discernir a qual população cada réplica pertence de forma clara no domínio do tempo.

Figura A.1: séries temporais de tamanho  $N = 100$ ,  $k = 3$  réplicas para  $j = 1, 2, 3$  populações. Painéis (a) população 1; (b) população 2; (c) população 3



Fonte: Elaborado pelo autor.

A Tabela A.2 mostra as estimativas condicionais cepstrais para cada população. Para cada replicação  $k$  dentro de cada população  $j$ , são fornecidos os valores dos coeficientes  $\hat{c}_{jk}^0$ ,  $\hat{c}_{jk}^1$ ,  $\hat{c}_{jk}^2$  e  $\hat{c}_{jk}^3$ . Estes coeficientes representam as características fundamentais da estrutura espectral das séries temporais simuladas e são essenciais para distinguir entre as diferentes populações na análise discriminante.

Tabela A.2: Cepstral Conditional Estimates for Each Population

$k$	$\hat{c}_{jk}^0$	$\hat{c}_{jk}^1$	$\hat{c}_{jk}^2$	$\hat{c}_{jk}^3$	$j$
1	-6.5487	0.3261	0.1986	0.0914	1
2	-5.8591	0.2123	0.1422	-0.2428	1
3	-7.1160	1.0702	0.4868	-0.0938	1
1	-6.0919	0.3522	-0.1860	0.1929	2
2	-6.5031	0.5871	-0.0018	-0.2458	2
3	-6.1342	1.6420	0.2162	-0.0043	2
1	-6.9406	0.8399	-0.2971	-0.3441	3
2	-6.9986	1.0223	-0.1747	0.2035	3
3	-6.8599	0.7979	-0.2902	-0.2275	3

Fonte: Elaborado pelo autor.

A Tabela A.2 exibe a matriz de médias por cepstra para cada população  $\hat{\mu}_{j\ell}$ . Nesta tabela, cada linha representa uma população ( $j=1, 2, 3$ ) e cada coluna representa um cepstrum ( $\ell = 0, 1, 2, 3$ ). As médias cepstrais fornecem uma visão agregada das características espectrais médias de cada população, que são cruciais para a análise discriminante. Estas

médias são utilizadas para calcular as matrizes de variância dentro e entre as populações, que são componentes essenciais na construção dos discriminantes lineares.

Tabela A.3: Matriz de médias por cepstra para cada população  $\hat{\mu}_{j\ell}$

$j\ell$	0	1	2	3
1	-6.507925	0.536182	0.275872	-0.081739
2	-6.243080	0.860460	0.009477	-0.019073
3	-6.933049	0.886705	-0.253989	-0.122691

Fonte: Elaborado pelo autor.

As matrizes  $\hat{\Omega}_w^M$  e  $\hat{\Omega}_b^M$  representam, respectivamente, a variabilidade dentro das populações e a variabilidade entre as populações. A matriz  $\hat{\Omega}_w^M$  reflete a dispersão dos cepstra dentro de cada população, enquanto a matriz  $\hat{\Omega}_b^M$  captura a dispersão entre as populações. Essas matrizes são fundamentais para a análise discriminante linear, pois ajudam a determinar as fronteiras de decisão que melhor separam as diferentes populações.

$$\hat{\Omega}_w^M = \begin{bmatrix} 2.304777 & -1.405175 & -0.656148 & -0.436119 \\ -1.405175 & 1.386275 & 0.788575 & 0.284454 \\ -0.656148 & 0.788575 & 0.689227 & 0.183678 \\ -0.436119 & 0.284454 & 0.183678 & 0.102275 \end{bmatrix}$$

$$\hat{\Omega}_b^M = \begin{bmatrix} 6.331926 & -0.668449 & 1.474091 & -0.068772 \\ -0.668449 & 0.190275 & -0.037819 & 0.141757 \\ 1.474091 & -0.037819 & 0.459091 & 0.116341 \\ -0.068772 & 0.141757 & 0.116341 & 0.151861 \end{bmatrix}$$

A Tabela A.4 apresenta os autovalores e os autovetores da matriz  $(\Omega_b)^{-1}\Omega_b$ . Os autovalores indicam a quantidade de variância explicada por cada componente discriminante, enquanto os autovetores fornecem os coeficientes das combinações lineares dos cepstra que formam os discriminantes lineares. Observa-se que o terceiro e o quarto autovalores são aproximadamente zero, indicando que a terceira e a quarta coordenada do subespaço não contribui significativamente para a variância entre as populações ponderada pela variância dentro. Portanto, não é necessário utilizar as coordenadas do terceiro e do quarto subespaço para a projeção dos dados, pois essas coordenadas não melhorariam a separação entre as populações, facilitando assim a classificação apenas com as duas primeiras coordenadas.

Tabela A.4: Eigenvalues and Eigenvectors of  $(\hat{\Omega}_b^M)^{-1}\hat{\Omega}_b^M$ 

Eigenvalues	29.3356	2.2853	0.0000	0.0000
	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
	-0.2278	-0.0023	-0.2405	-0.0699
	-0.1342	0.0390	-0.5820	-0.7427
	0.1716	-0.3272	0.7623	0.0057
	-0.9490	0.9442	-0.1496	0.6660

Fonte: Elaborado pelo autor.

A Tabela A.5 mostra as coordenadas discriminantes para  $q = 1, 2$ . Essas coordenadas são as projeções das séries temporais simuladas nos dois primeiros discriminantes lineares. As coordenadas discriminantes permitem visualizar como as diferentes populações se distribuem no espaço discriminante, ajudando a identificar quais séries temporais pertencem a quais populações com base nas características cepstrais.

Tabela A.5: Discriminant coordinates for  $q = 1, 2$ 

$\hat{d}_{jk1}$	$\hat{d}_{jk2}$
1.1961	-0.0350
1.3014	-0.1601
1.2200	-0.2510
1.2971	0.1254
1.3511	0.0233
1.3601	0.2573
1.6420	-0.1162
1.6964	-0.0285
1.6516	-0.0401

Fonte: Elaborado pelo autor.

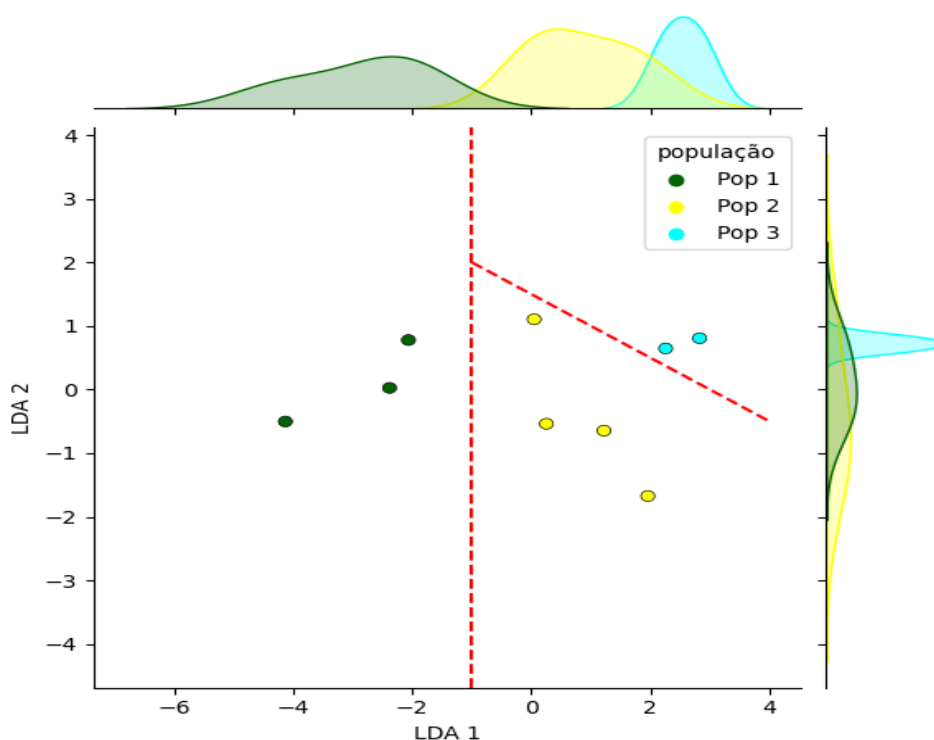
A Análise Discriminante Linear (LDA) foi utilizada para projetar os dados de séries temporais no espaço discriminante, resultando nos componentes discriminantes LDA 1 e LDA 2. No gráfico gerado (Figura 1), os pontos foram coloridos para representar três diferentes populações: População 1 (verde escuro), População 2 (amarelo) e População 3 (ciano). Foram adicionadas duas regras de decisão por hiperplano: uma linha vertical pontilhada vermelha e uma linha diagonal pontilhada vermelha com inclinação negativa.

A População 1 está bem separada das outras duas populações utilizando LDA 1. A maioria dos pontos verdes está à esquerda da linha vertical pontilhada vermelha, indicando que LDA 1 é eficaz para discriminar a População 1 das demais. Isso sugere que o primeiro componente discriminante captura a variação principal que distingue a População 1 das Populações 2 e 3.

Entretanto, apesar da eficácia do LDA 1 na separação da População 1, as Populações 2 e 3 apresentam uma sobreposição significativa. Mesmo utilizando o LDA 2,

não há uma separação clara entre essas duas populações. A linha diagonal foi desenhada para tentar minimizar a sobreposição, mas ainda existem classificações erradas, uma vez que os pontos amarelos e cianos estão misturados tanto acima quanto abaixo da linha diagonal pontilhada vermelha. Isso indica que, embora o segundo componente discriminante adicione alguma informação adicional, ele não é suficiente para separar completamente as Populações 2 e 3.

Figura A.2: LDA



Fonte: Elaborado pelo autor.

## A.2 Análise de Convergência empírica dos Cepstrals

A Tabela A.6 apresenta os resultados para  $N = 250$ , enquanto a Tabela A.7 mostra os resultados para  $N = 1000$ . Além disso, em cada caso foram calculados os valores reais, os cepstrais e os M-cepstrais estimados para  $\phi = \{0.1, 0.5, 0.9\}$ . Para a análise empírica, foram simuladas 15 amostras de cepstrais e a comparação do verdadeiro valor foi feita com o real.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

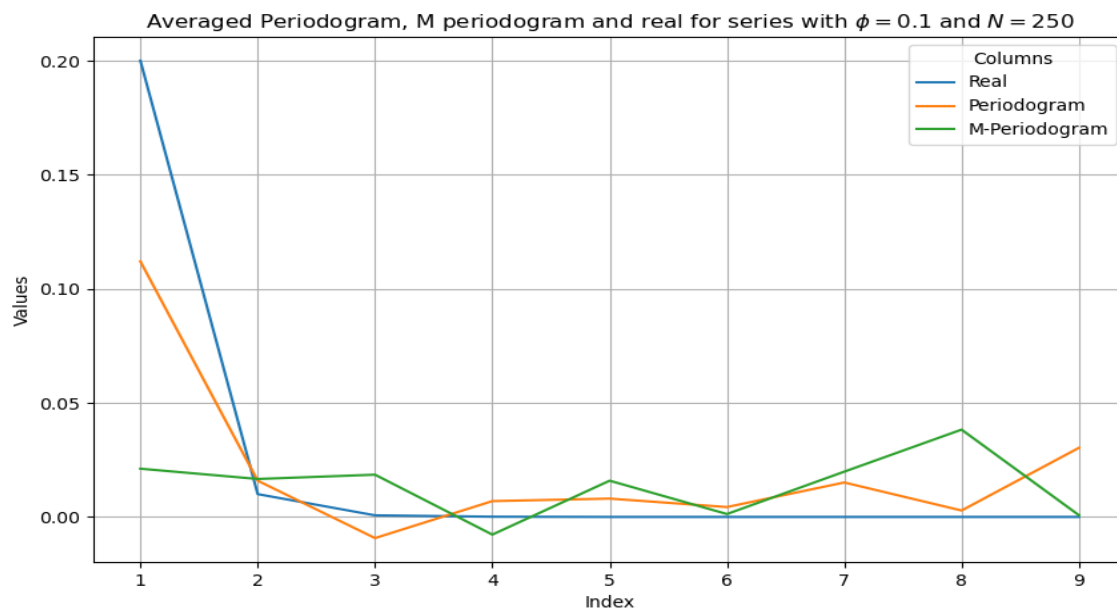
Analisando essas tabelas, pode-se observar a convergência dos valores empíricos dos coeficientes cepstrais em direção aos valores reais à medida que o tamanho da amostra aumenta ou quanto menor for o valor de  $\phi$ .

Para cada  $\phi$  foram calculados os valores médios e o valor real, apresentados nas Figuras 4.3, 4.4 e 4.5 para  $N = 250$  e 4.7, 4.8 e 4.9 para  $N = 1000$ .

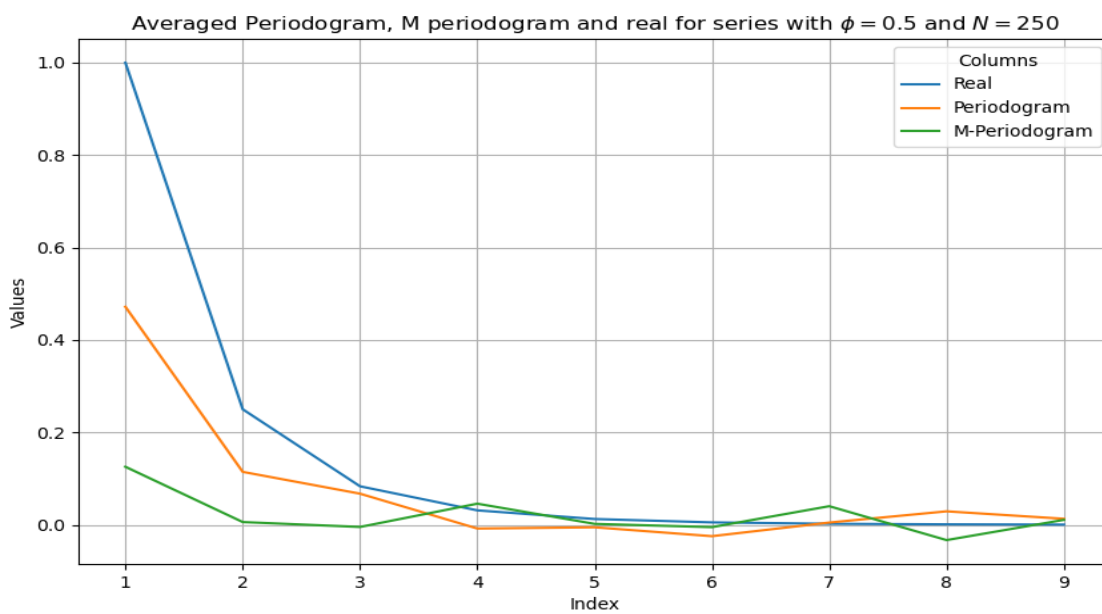
Já nas Figuras 4.6 e 4.10 estão apresentados os box-plots comparativos dos primeiros 7 cepstrais dois a dois, comparando o cepstra e o M-cepstra, respectivamente. Note que o valor mediano do M-cepstra sempre está abaixo do valor mediano dos cepstrais. Além disso, nota-se que, para os primeiros cepstrais, o cepstral está mais próximo de zero do que o M-cepstra, indicando uma convergência mais rápida do baseado no periodograma do que o clássico. Contudo, para  $N = 1000$ , os valores tendem a ficar mais próximos, indicando uma convergência empírica de um para o outro.

No caso do periodograma clássico, os valores das médias empíricas estão mais próximos das médias reais quando  $N = 1000$  comparado a  $N = 250$ , indicando uma melhor estimativa com um tamanho de amostra maior. Por exemplo, a média do coeficiente cepstral para  $x_1$  melhora de -2.424069 ( $N = 250$ ) para -2.395369 ( $N = 1000$ ), aproximando-se do valor real de -1.837900.

Semelhante ao método do cepstra baseado no periodograma, o M-cepstra mostra uma melhor convergência com tamanhos de amostra maiores. Os valores das médias dos coeficientes cepstrais para  $c_2$  e  $c_3$  também se aproximam dos valores reais à medida que o tamanho da amostra aumenta.

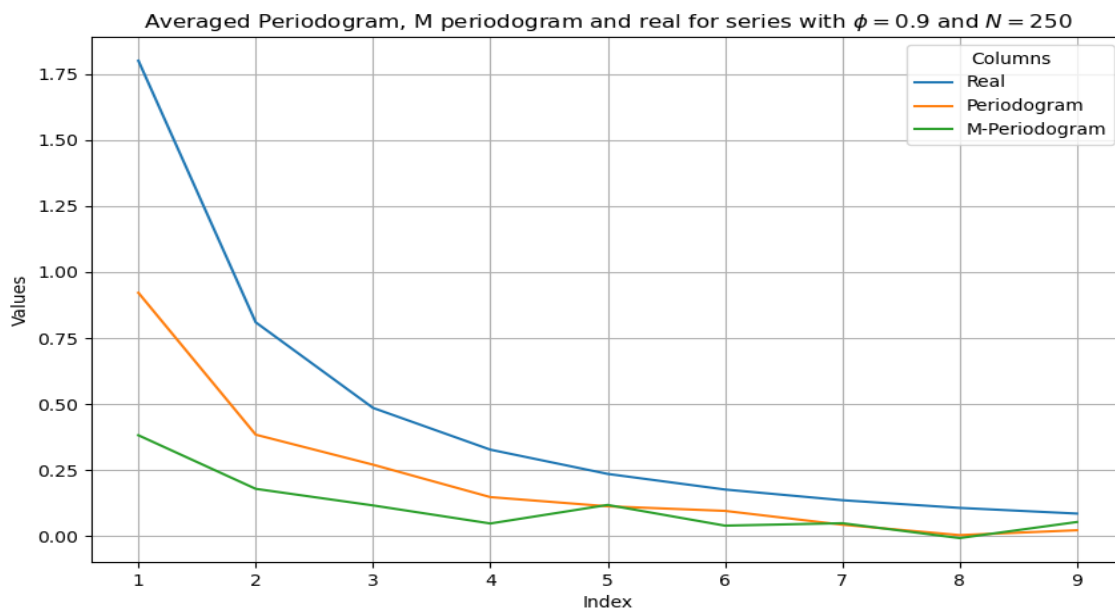
Figura A.3: Convergência modelo  $AR(1)$  com  $\phi = 0.1$  e  $N = 250$ 

Fonte: Elaborado pelo autor.

Figura A.4: Convergência modelo  $AR(1)$  com  $\phi = 0.5$  e  $N = 250$ 

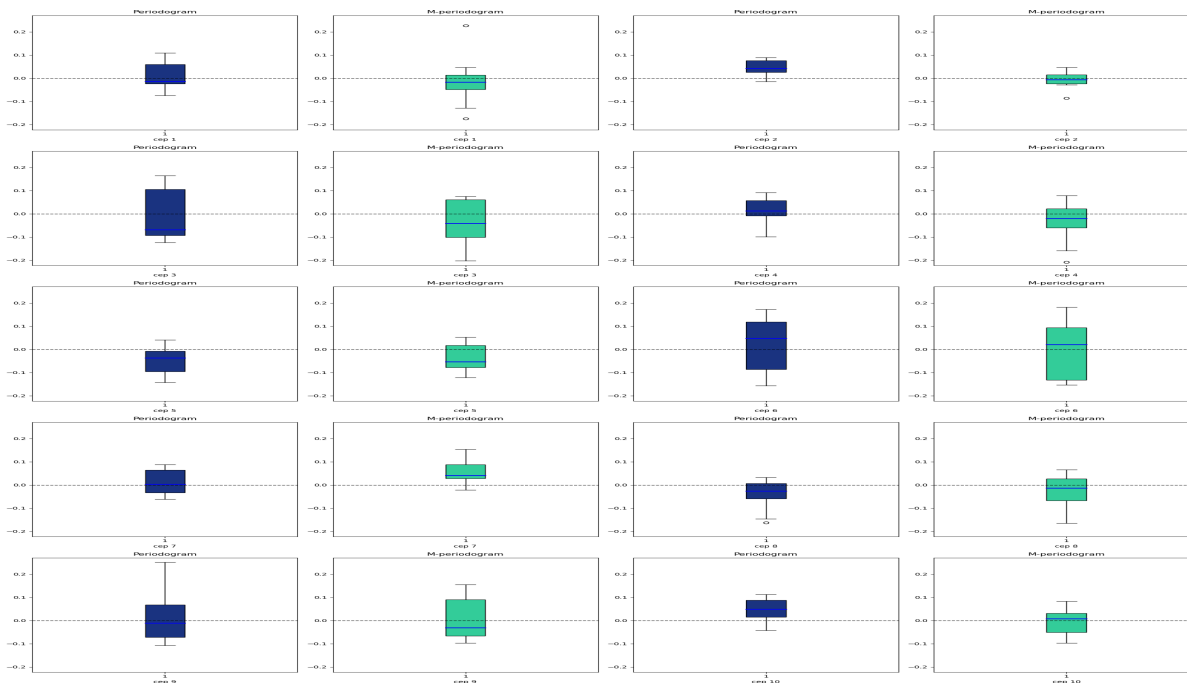
Fonte: Elaborado pelo autor.

Figura A.5: Convergência modelo  $AR(1)$  com  $\phi = 0.9$  e  $N = 250$



Fonte: Elaborado pelo autor.

Figura A.6: Box-plot dos coeficientes cepstrais até  $L = 7$  periodograma clássico versus robusto com  $N = 250$  e 15 amostras



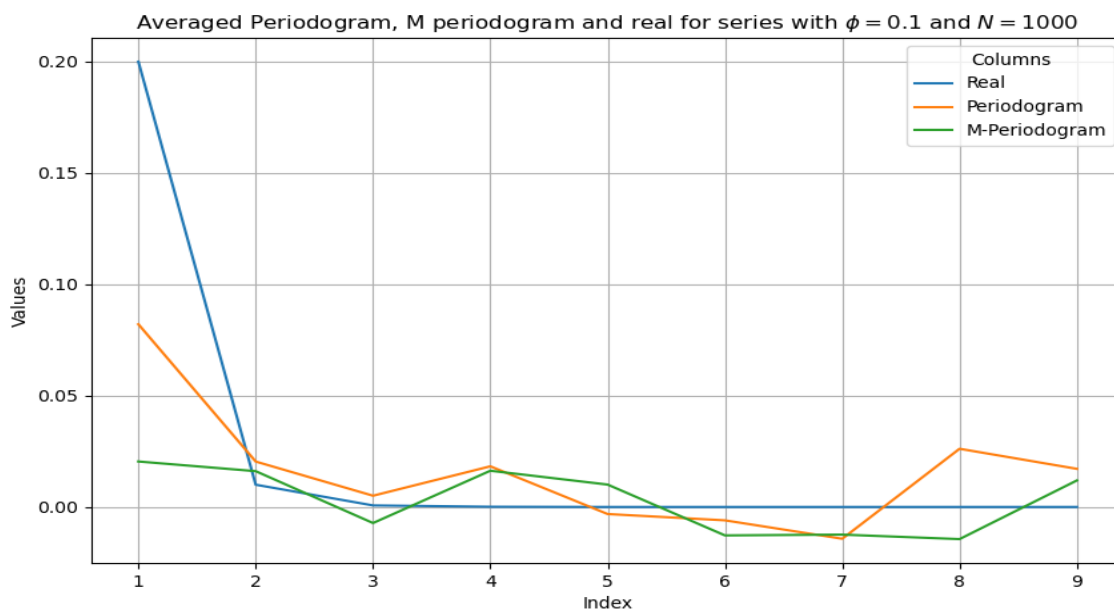
Fonte: Elaborado pelo autor.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque.

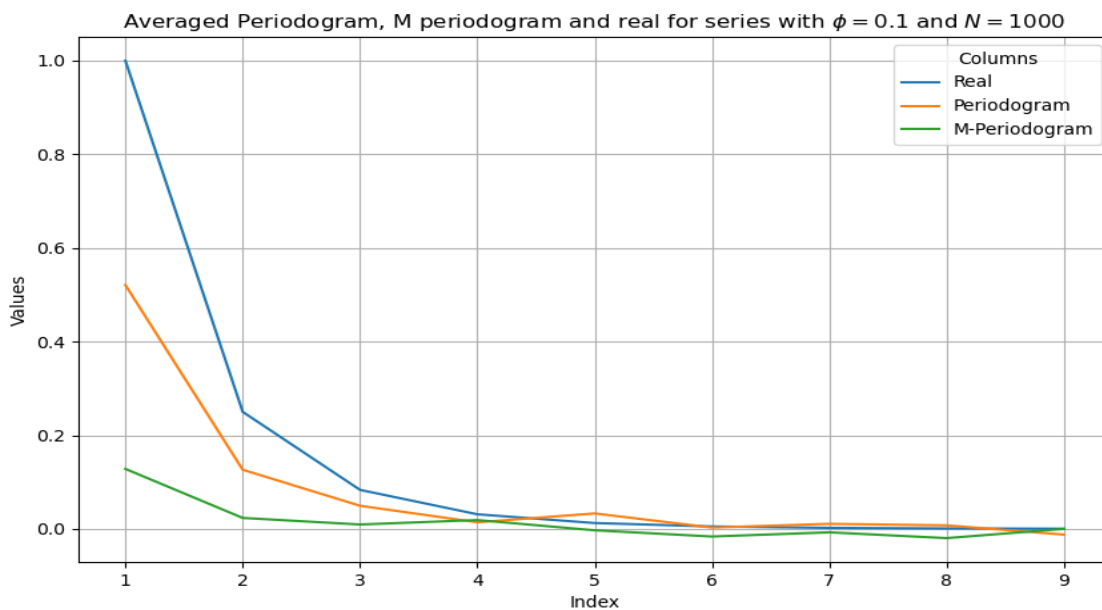
Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

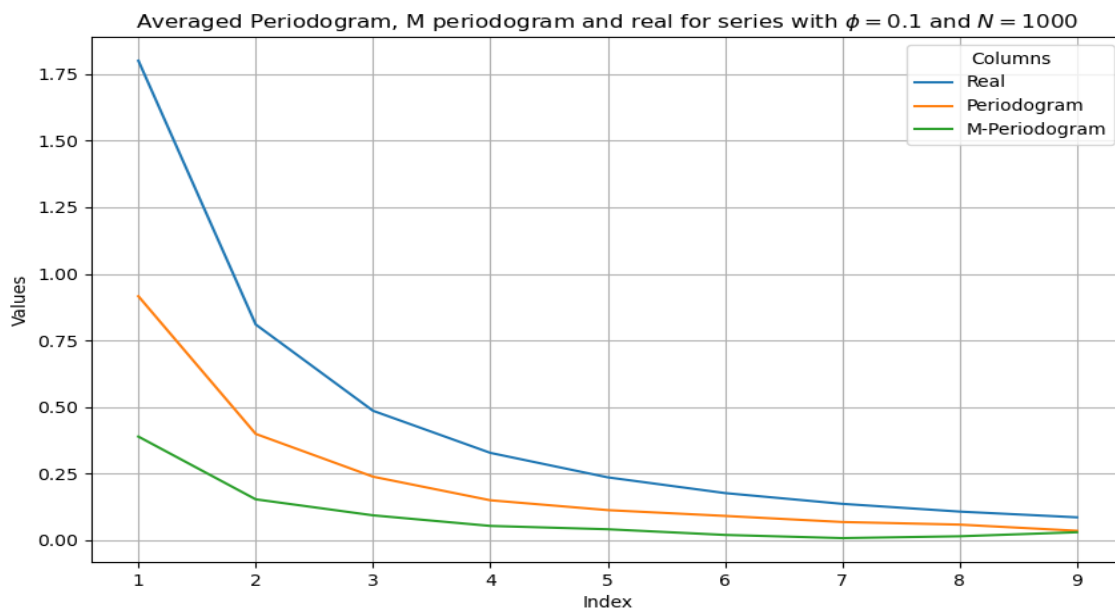
Figura A.7: Convergência modelo  $AR(1)$  com  $\phi = 0.1$  e  $N = 1000$



Fonte: Elaborado pelo autor.

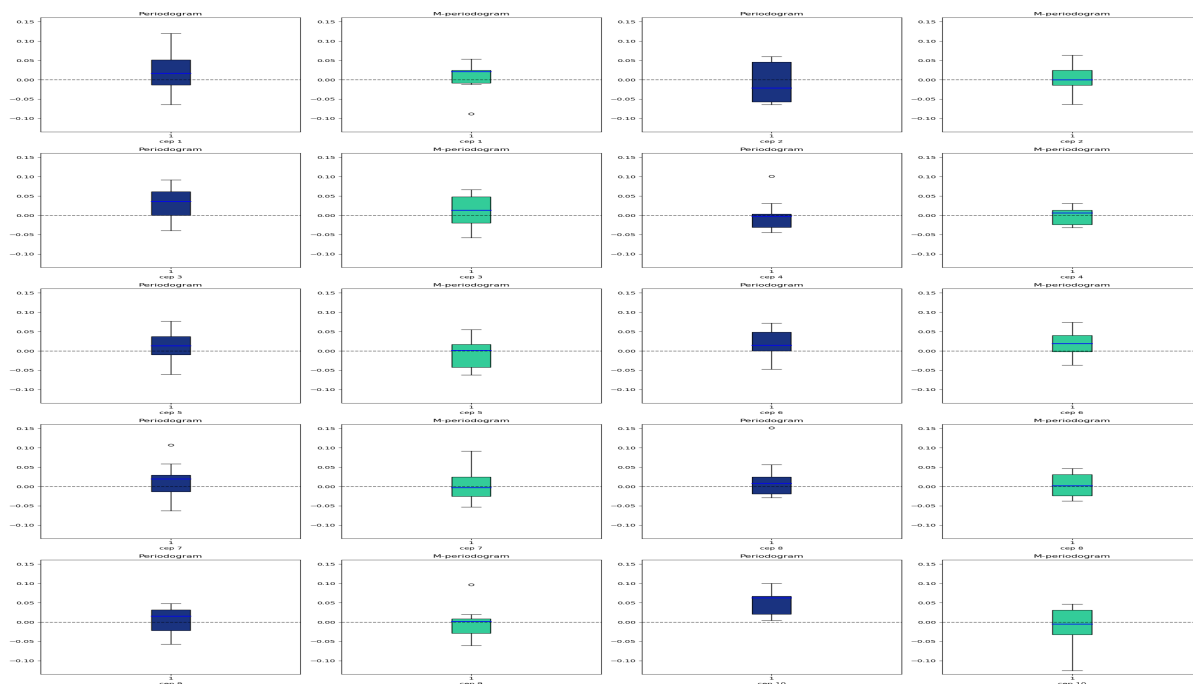
Figura A.8: Convergência modelo  $AR(1)$  com  $\phi = 0.5$  e  $N = 1000$ 

Fonte: Elaborado pelo autor.

Figura A.9: Convergência modelo  $AR(1)$  com  $\phi = 0.9$  e  $N = 1000$ 

Fonte: Elaborado pelo autor.

Figura A.10: Box-plot dos coeficientes cepstrais até  $L = 7$  periodograma clássico versus robusto com  $N = 1000$  e 15 amostras



Fonte: Elaborado pelo autor.

Tabela A.8: Erros Quadráticos para  $N=250$

$c_\ell$	$ c_\ell - \hat{c}_\ell ^2$	$ c_\ell - \hat{c}_\ell^M ^2$	$ \hat{c}_\ell - \hat{c}_\ell^M ^2$
0	0.3448	0.2677	0.0047
1	0.0065	0.0257	0.0690
2	0.0047	0.0335	0.0428
3	0.0127	0.0105	0.0307
4	0.0013	0.0163	0.0275
5	0.0034	0.0182	0.0278
6	0.0325	0.0245	0.0272
<b>Soma</b>	0.4060	0.3964	0.2297
<b>EQM</b>	0.0541	0.0529	0.0306

Fonte: Elaborado pelo autor.

Ambos os métodos (periodograma e Huber) melhoram suas estimativas com tamanhos de amostra maiores. O método Huber parece ter um viés ligeiramente menor em alguns casos, como visto nas médias dos coeficientes cepstrais para  $x_2$  e  $x_3$ .

A análise empírica dos coeficientes cepstrais mostra que tanto o periodograma quanto o método Huber fornecem melhores aproximações dos valores reais com tamanhos de amostra maiores. Isso indica a convergência das estimativas empíricas para os valores reais, validando a robustez desses métodos para análise espectral.

Tabela A.9: Erros Quadráticos para  $N=1000$ 

$c_\ell$	$ c_\ell - \hat{c}_\ell ^2$	$ c_\ell - \hat{c}_\ell^M ^2$	$ \hat{c}_\ell - \hat{c}_\ell^M ^2$
0	0.3097	0.2657	0.0025
1	0.0139	0.0214	0.0606
2	0.0210	0.0234	0.0599
3	0.0167	0.0246	0.0674
4	0.0101	0.0115	0.0718
5	0.0205	0.0117	0.0552
6	0.0323	0.0162	0.0586
<b>Soma</b>	0.4242	0.3745	0.3760
<b>EQM</b>	0.0566	0.0499	0.0501

Fonte: Elaborado pelo autor.

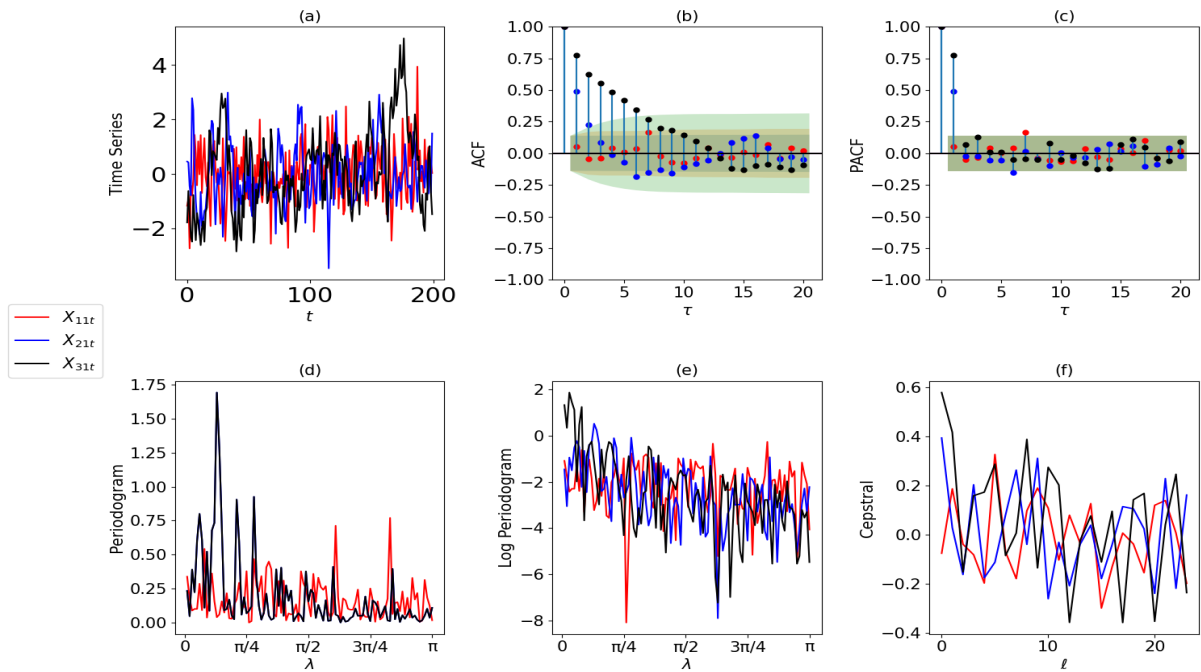
### A.3 Example of $AR(1)$ and $MA(1)$

Esta seção tem como objetivo apresentar a simulação de séries  $AR(1)$  e  $MA(1)$  para diferentes parâmetros, como se cada um pertencesse a uma população distinta. E assim, verificar seus efeitos nos diferentes domínios. As séries foram simuladas para os parâmetros  $\phi = \{0.1, 0.5, 0.9\}$ ,  $\phi = \{-0.1, -0.5, -0.9\}$ ,  $\theta = \{0.1, 0.5, 0.9\}$ ,  $\theta = \{-0.1, -0.5, -0.9\}$ . Além disso, o ruído branco utilizado foi  $\epsilon_t \sim N(0, \sigma^2)$  e o tamanho das séries  $N = 100$ .

Tais simulações estão apresentadas nas figuras [A.11](#), [A.12](#), [A.13](#) e [A.14](#). Estas figuras estão compostas por 6 painéis: (a) gráfico das séries temporais; (b) gráfico das Funções de Autocorrelação (FAC); (c) Gráfico das Funções de Autocorrelações Parciais (FACP); (d) Periodogramas; (e) Log dos periodogramas; e (f) coeficientes cepstrais.

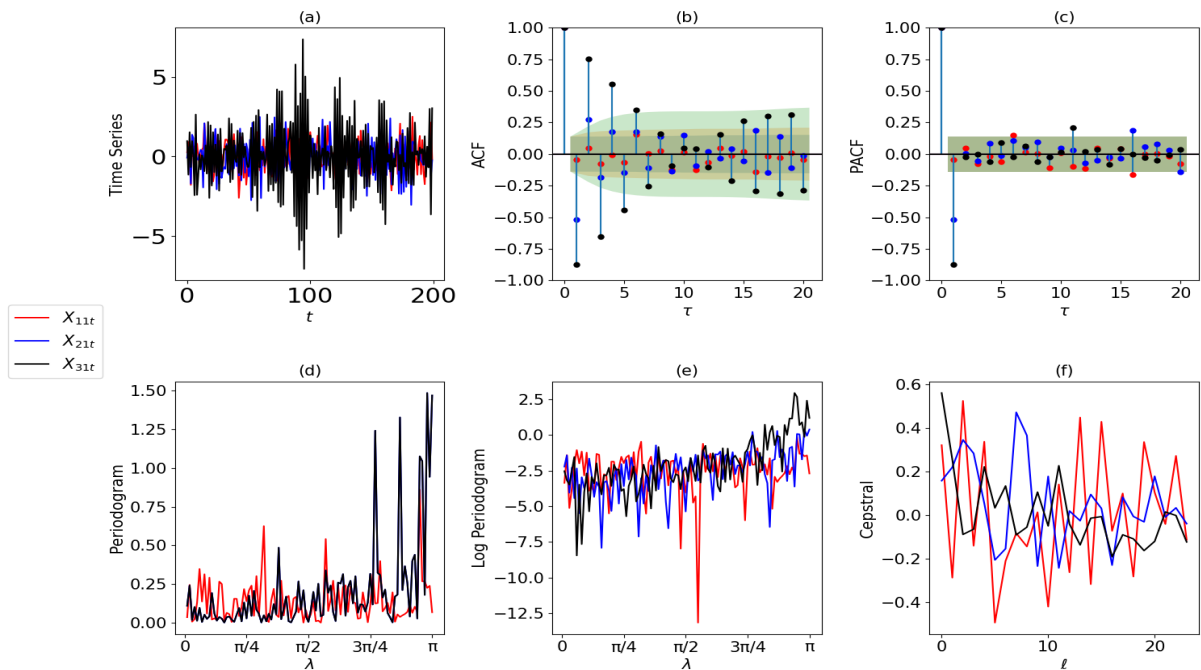
Em todos os gráficos, as figuras em vermelho significam os parâmetros iguais à 0.1 em valor absoluto, enquanto as figuras azuis e pretas representam, respectivamente, os parâmetros em valor absoluto 0.5 e 0.9.

Figura A.11:  $\phi \in [0.1, 0.5, 0.9]$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra.



Fonte: Elaborado pelo autor.

Figura A.12:  $\phi \in [-0.1, -0.5, -0.9]$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra.



Fonte: Elaborado pelo autor.

Note que no modelo  $AR(1)$ , suas características são apresentadas na primeira linha dos gráficos e como esperado, quanto maior o valor do parâmetro, mas suave são as FAC

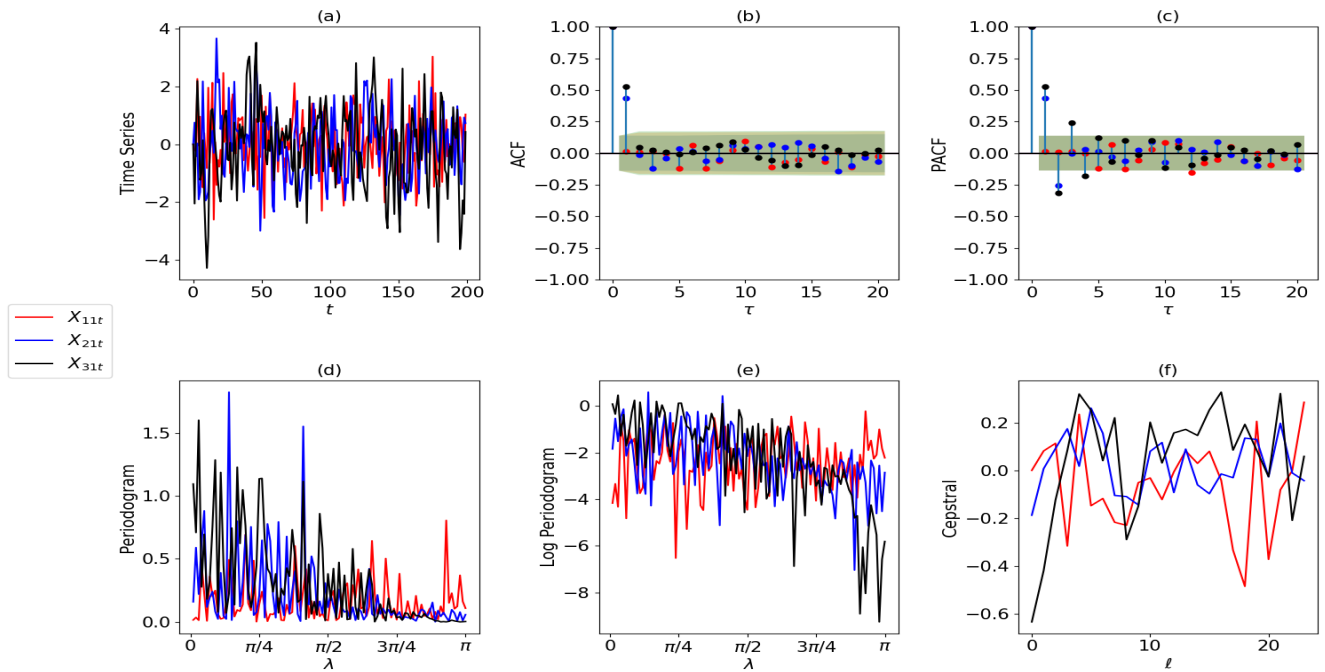
e apenas um valor significativo na FACP, exatamente no valor do parâmetro utilizado. No domínio da frequência, observa-se um decaimento claro no periodograma e no log do periodograma no gráfico A.11, sobretudo para a série com maior parâmetro (preta). Ou seja, o espectro estimado é como esperado com as frequências próximas de zero sendo as que mais contribuem para a variabilidade do processo  $AR(1)$  positivo, sendo tanto mais perto de zero quanto maior o valor do parâmetro.

Por outro lado, observa-se que a FAC e FACP também seguem o padrão esperado com decaimento exponencial alternando de sinal na FAC e apenas um valor negativo significativo na FACP, exatamente o valor do parâmetro. E no domínio da frequência, observa-se que o periodograma apresenta comportamento crescente, assim como o log do periodograma. Isto indica que a variabilidade deste processo é mais explicada por harmônicas de mais alta frequência com baixo período.

Note que independentemente do sinal e do tamanho dos coeficientes, os cepstral mais importantes são os próximos de zero, sobretudo o primeiro, que guarda a memória de apenas um eon do processo  $AR(1)$ . A medida que a quantidade de cepstra aumenta, sua importância tende para zero. Contudo, quando os coeficientes são positivos e, consequentemente, de baixa frequência, a importância do primeiro coeficiente cepstra em comparação aos demais torna-se mais evidente.

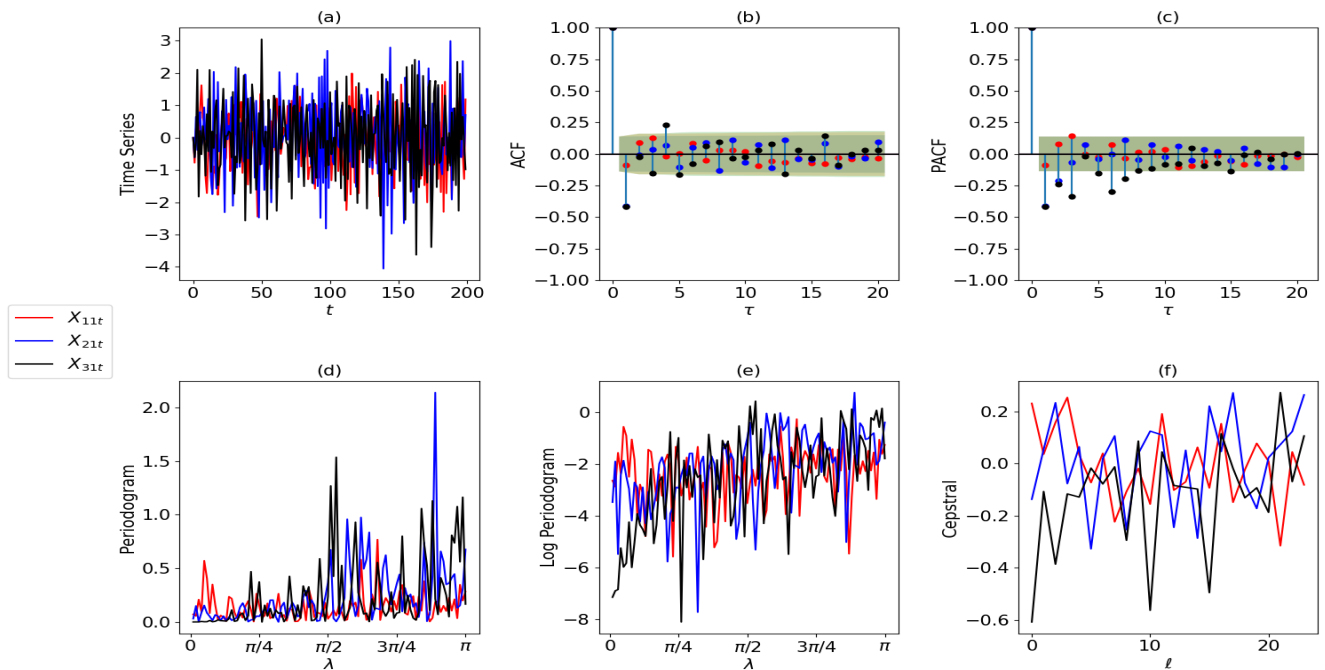
Nas figuras A.13 e A.14 estão apresentados os mesmos gráficos, contudo agora para processos  $MA(1)$  do tipo  $X_t = \epsilon_t + \theta\epsilon_{t-1}$  com sinais positivos e negativos, respectivamente. Note que os padrões esperados no domínio do tempo também são esperados com a FAC com apenas um valor significativo no valor do parâmetro e com a FACP com decaimento exponencial oscilatório ou negativo, dependendo do sinal de  $\theta$ . Nota-se que com sinais positivos, o periodograma tem baixa frequência similar ao  $AR(1)$  positivo, gerando resultados mais claros de que apenas o primeiro cepstra é importante. Por outro lado, no caso dos sinais negativos, o modelo se torna de alta frequência e curto período e embora é possível ver mais claramente que o primeiro cepstra é importante quando  $\theta = -0.9$ , para os outros dois não é tão clara tal importância.

Figura A.13:  $\theta \in [0.1, 0.5, 0.9]$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra.



Fonte: Elaborado pelo autor.

Figura A.14:  $\theta \in [-0.1, -0.5, -0.9]$  and  $\sigma_{\epsilon_{jk}}^2 = 1$ : (a) Time Series, (b) ACF, (c) PACF, (d) Periodogram, (e) Log-Periodogram, and (f) Cepstra.



Fonte: Elaborado pelo autor.

## A.4 Simulações de Monte carlo

Uma simulação de Monte Carlo foi realizada para analisar as propriedades da amostra dos M-cepstra e seu impacto nas taxas de classificação no contexto da variabilidade espectral dentro das populações, com e sem a contaminação de observações abruptas que se comportam como outliers aditivos. Para as simulações, foram consideradas  $J = 3$  populações em 6 situações diferentes:  $n_j = \{15, 50, 100\}$  e  $N = \{1000, 2000\}$ . Em cada cenário, a taxa de classificação é calculada e os cenários são repetidos 100 vezes para obter a média e o desvio padrão das classificações. E para comparar com [Krafty \(2016\)](#) e devido às características da aplicação dos dados apresentadas no capítulo 5, as simulações são conduzidas usando um modelo  $AR(2)$  condicional (à população) que pode ser representado por

$$X_{jkt} = \phi_{jk1}X_{jkt-1} + \phi_{jk2}X_{jkt-2} + \epsilon_{jkt} \quad \epsilon_{jkt} \sim N(0, \sigma_{jk}^2) \quad (4.1)$$

onde o parâmetro de cada população é uma variável aleatória apresentada na Tabela A.10.

Tabela A.10: Distribuição dos Parâmetros Condicionais

População	Parâmetros	$\phi_{jk1}$	$\phi_{jk2}$	Ruído Branco	Variância do Ruído Branco
$j = 1$		$(\phi_{1k1} j = 1) \sim UNI(0.05, 0.7)$	$(\phi_{1k2} j = 1) \sim UNI(-0.12, -0.06)$	$\epsilon_{1k} \sim N(0, \sigma_{1k}^2)$	$(\sigma_{1k}^2 j = 1) \sim UNI(0.1, 10)$
$j = 2$		$(\phi_{2k1} j = 2) \sim UNI(0.01, 1.2)$	$(\phi_{2k2} j = 2) \sim UNI(-0.36, -0.25)$	$\epsilon_{2k} \sim N(0, \sigma_{2k}^2)$	$(\sigma_{2k}^2 j = 2) \sim UNI(0.3, 3.0)$
$j = 3$		$(\phi_{3k1} j = 3) \sim UNI(0.12, 1.5)$	$(\phi_{3k2} j = 3) \sim UNI(-0.75, -0.56)$	$\epsilon_{3k} \sim N(0, \sigma_{3k}^2)$	$(\sigma_{3k}^2 j = 3) \sim UNI(0.9, 1.1)$

Fonte: Elaborado pelo autor.

O modelo condicional  $AR(2)$  é escolhido com parâmetros sendo variáveis aleatórias com distribuição uniforme em intervalos sobrepostos. Além disso, os distúrbios são modelados como distribuição normal com média zero e sua variância como uma distribuição uniforme com intervalos crescentes diferentes. Essa característica é importante para criar séries temporais que podem ser misturadas e criar dificuldade para discriminação e classificação, como ocorreria em um contexto de aplicação real.

Cada cenário inclui todas as três populações para cada série temporal e réplicas. Todas as transformações são feitas, incluindo a estimativa do periodograma Multitaper para clássico e M-periodograma, e o vetor cepstral é obtido para cada réplica dentro de sua própria população. Em seguida, um parâmetro LDA é estimado com base nos cepstrais e usado para o teste de classificação. Para este teste, um novo grupo de séries temporais para cada população é simulado para os mesmos parâmetros e, novamente, para comparação com [Krafty \(2016\)](#),  $N = 50$  é escolhido para os dados de teste. Esta nova série temporal passa pela mesma transformação das réplicas originais até alcançar o vetor cepstral, que é então usado para teste.

Finalmente, as simulações são conduzidas em duas situações: sem outliers, onde o modelo é modelado sem observações extremas, e com outliers, onde as réplicas são

contaminadas com um outlier aditivo seguindo [Huber e Ronchetti \(2009\)](#); [Reisen, Lévy-Leduc, et al. \(2020\)](#) com

$$Z_{jkt} = X_{jkt} + \omega I_{jkt}, \quad (4.2)$$

onde  $\omega$  é a magnitude do outlier,  $\{X_{jkt}\}$  e  $\{I_{jkt}\}$  são independentes, e  $\{I_{jkt}\}$  é uma sequência de variáveis aleatórias IID com  $P(I_{jkt} = -1) = P(I_{jkt} = 1) = p/2$  e  $P(I_{jkt} = 0) = 1 - p$ ,  $p \in (0, 1)$ . Aqui, utiliza-se  $p = 0.01$  e  $\omega = 7$ . O caso não contaminado corresponde a  $\omega = 0$ . A análise discriminante cepstral robusta proposta foi comparada com a análise discriminante cepstral multitaper, ambas utilizando  $R = 7$  tapers. A constante de ajuste limítrofe no modelo Huber é definida como 1.345 conforme valor ótimo assintoticamente definido por [Reisen, Lévy-Leduc, et al. \(2020b\)](#). Um conjunto de dados de teste de 50 séries temporais por grupo é simulado para cada amostra aleatória para avaliar a taxa de classificação dos dados de treino. A taxa de classificação correta é calculada em cada cenário utilizando o modelo LDA com os coeficientes de cepstra. Este cenário foi repetido 100 vezes e a taxa média e o erro padrão das taxas de corretas classificações é calculado e apresentados na tabela [E.3](#)

O desempenho dos dois métodos de classificação - Clássico e robusto - são avaliados em diferentes cenários, com e sem a presença de outliers. Nota-se que para as amostras sem outliers, os resultados mostram que ambos os métodos de classificação apresentam taxas de classificação corretas elevadas. Em  $\Pi_1$ , por exemplo, com  $n_j = 15$ , as taxas médias de classificação corretas foram de 89.64% e 90.09% para o método Clássico com  $N = 1000$  e  $2000$ , respectivamente. Para o método Huber, as taxas foram ligeiramente menores, com 85.37% e 89.85% para  $N = 1000$  e  $2000$ , respectivamente.

À medida que o número de réplicas por população aumenta, observa-se uma melhora nas taxas de classificação corretas para ambos os métodos. Com  $n_j = 100$ , os métodos Clássico e Huber alcançaram 91.02% e 93.00% de classificação correta para  $N = 1000$ , respectivamente. A mesma tendência é observada para  $N = 2000$ , onde as taxas atingem 93.00% e 92.19%.

Na presença de outliers, nota-se uma queda nas taxas de classificação corretas para ambos os métodos. Para  $\Pi_1$  com  $n_j = 15$ , as taxas médias de classificação corretas caíram para 67.06% e 68.52% para o método Clássico com  $N = 1000$  e  $2000$ , respectivamente. No entanto, o método Huber mostrou-se mais robusto, com taxas de 79.13% e 83.64% para  $N = 1000$  e  $2000$ , respectivamente.

A robustez do método Huber é mais evidente à medida que o número de réplicas por população aumenta. Com  $n_j = 100$ , as taxas de classificação corretas para o método Clássico foram de 76.18% e 82.78%, enquanto o método Huber alcançou 81.06% e 88.03% para  $N = 1000$  e  $2000$ , respectivamente. Isso indica que o método Huber é mais eficaz na presença de outliers em comparação com o método Clássico.

Comparando os diferentes cenários de variância ( $\sigma^2$ ), observa-se que para  $\Pi_1$  (0.01 - 10), as taxas de classificação corretas são geralmente mais baixas do que para  $\Pi_2$  (0.3 - 3) e  $\Pi_3$  (0.9 - 1.1). Por exemplo, em  $\Pi_3$ , para  $n_j = 15$ , as taxas de classificação corretas sem outliers foram de 93.67% e 93.88% para o método Clássico com  $N = 1000$  e  $2000$ , respectivamente, mostrando uma melhor performance em comparação com  $\Pi_1$  e  $\Pi_2$ .

Com a inclusão de outliers,  $\Pi_3$  também apresentou uma melhor robustez geral. As taxas de classificação corretas para o método Huber em  $\Pi_3$  foram consistentemente mais altas em comparação com  $\Pi_1$  e  $\Pi_2$ . Isso sugere que o desempenho dos classificadores pode ser fortemente influenciado pela variância dos dados.

Tabela A.11: Simulações de Monte Carlo: taxas médias de classificação corretas

$I_N^{jk}(\lambda_m)$	$\sigma^2$	$n_j$	SEM OUTLIER				COM OUTLIER			
			Clássico		Huber		Clássico		Huber	
			1000	2000	1000	2000	1000	2000	1000	2000
$6^*\Pi_1$	$6^*(\mathbf{0.01 - 10})$	$2^*\mathbf{15}$	89.64%	90.09%	85.37%	89.85%	67.06%	68.52%	79.13%	83.64%
			5.80%	5.96%	6.10%	5.99%	9.10%	8.45%	7.45%	7.91%
		$2^*\mathbf{50}$	90.85%	91.02%	86.40%	90.13%	75.08%	81.10%	80.34%	84.74%
			4.47%	4.63%	5.20%	4.53%	4.97%	3.89%	4.89%	4.48%
		$2^*\mathbf{100}$	91.02%	93.00%	89.02%	92.19%	76.18%	82.78%	81.06%	88.03%
			4.19%	3.75%	6.28%	3.32%	3.77%	3.12%	3.39%	3.18%
$6^*\Pi_2$	$6^*(\mathbf{0.3 - 3})$	$2^*\mathbf{15}$	88.80%	91.40%	83.10%	90.01%	72.28%	83.02%	85.26%	87.83%
			6.07%	5.27%	4.20%	4.63%	6.44%	5.70%	5.01%	4.73%
		$2^*\mathbf{50}$	91.98%	92.33%	89.60%	92.02%	84.78%	86.66%	87.12%	88.43%
			4.17%	4.17%	4.60%	5.12%	4.77%	5.00%	4.23%	4.22%
		$2^*\mathbf{100}$	92.53%	95.27%	91.14%	95.03%	71.04%	87.44%	89.02%	90.95%
			3.91%	3.21%	1.74%	2.59%	2.97%	2.84%	1.93%	2.71%
$6^*\Pi_3$	$6^*(\mathbf{0.9 - 1.1})$	$2^*\mathbf{15}$	93.67%	93.88%	91.15%	93.45%	81.94%	81.08%	89.43%	91.14%
			5.22%	6.15%	5.40%	4.33%	6.61%	6.63%	5.81%	4.91%
		$2^*\mathbf{50}$	95.22%	95.83%	93.10%	95.16%	84.10%	88.52%	90.56%	93.26%
			3.44%	4.57%	3.57%	4.23%	3.09%	4.57%	3.89%	2.81%
		$2^*\mathbf{100}$	95.30%	96.51%	94.98%	95.71%	84.22%	93.22%	91.90%	95.23%
			3.14%	2.74%	2.86%	2.93%	2.44%	2.91%	1.93%	1.95%

Fonte: Elaborado pelo autor.

Os resultados das simulações de Monte Carlo indicam que o método Huber é geralmente mais robusto na presença de outliers em comparação com o método Clássico. Além disso, o desempenho dos métodos de classificação é influenciado pelo tamanho da amostra e pelo número de réplicas por população. Em cenários de alta variância ( $\Pi_1$ ), ambos os métodos apresentam taxas de classificação corretas mais baixas, enquanto em cenários de baixa variância ( $\Pi_3$ ), as taxas de classificação são mais elevadas. Esses achados destacam a importância de considerar a robustez dos métodos de classificação em diferentes contextos de variância e presença de outliers.

## Apêndice B

# Protótipo do Aplicativo Shiny para a Aplicação

Um aplicativo Shiny foi criado em codificação Python para análise comparativa entre os resultados visuais do modelo clássico versus o modelo robusto, que pode ser acessado pelo link: [https://jonathanestatistica.shinyapps.io/robust\\_lda\\_app/](https://jonathanestatistica.shinyapps.io/robust_lda_app/)

Seu desenvolvimento ainda é um protótipo devido a lentidão das estimativas do modelo robusto por meio das regressões. Portanto, os parâmetros que podem ser modificados apenas modificam as discriminações e classificações do modelo clássico, mas o mesmo será futuramente desenvolvido para que ambos os modelos sejam alterados com a alteração dos parâmetros. Consequentemente, o modelo robusto é estático e foi colocado apenas para análise comparativa. Os parâmetros que podem ser alterados neste protótipo são o número de multitapers  $R = 1, \dots, 10$  e o número de cepstra  $L = 1, \dots, 10$ .

### Avanços futuros:

1. Inclusão de modificação de outros parâmetros fixos como o limite do *Huber*;
2. Nos códigos, foi utilizado o modelo  $AR(2)$  como um caso particular. Contudo, qualquer modelo  $ARMA(p, q)$  pode ser utilizado. Portanto, estes parâmetros também serão passíveis de alteração;
3. Os dados da aplicação foram utilizados, mas no aplicativo é possível introduzir uma busca de uma base de dados dentro de algum diretório como do próprio computador. Isto permitirá que o modelo possa ser analisado e utilizado para qualquer base de dados como dados macroeconômicos, dados de poluição do ar, dados de nível do mar, etc.
4. Para inclusão do modelo M-cepstra no Shiny, é necessário avanço na estimação do M-periodograma por meio do FFT. Isto permitirá que a estimação seja mais rápida e possa ser implementada no aplicativo;
5. Com relação ao item anterior, tal avanço pode permitir o download do M-periodogram, periodograma clássico, bem como de suas respectivas taxas de classificação.

# Apêndice C

## Documentação de Algoritmos e funções Genéricas no Python

### C.1 Função de simulação para modelos ARMA(p,q) sem outlier

---

#### Algorithm 1 Simulation of Time Series Data

---

**Require:**  $i$ : Identifier for the current simulation

**Require:**  $n_j$ : Number of time series per group

**Require:**  $N$ : Size of time series

**Require:**  $\phi_{1i}$ : Lower bound for  $\phi_1$

**Require:**  $\phi_{1s}$ : Upper bound for  $\phi_1$

**Require:**  $\phi_{2i}$ : Lower bound for  $\phi_2$

**Require:**  $\phi_{2s}$ : Upper bound for  $\phi_2$

**Require:**  $\sigma_{li}$ : Lower bound for noise variance

**Require:**  $\sigma_{ls}$ : Upper bound for noise variance

**Require:**  $\theta_1$ : First moving average parameter

**Require:**  $\theta_2$ : Second moving average parameter

```

1: Import necessary libraries
2: Initialize  $X$  as an empty DataFrame
3: for  $j \leftarrow 1$  to  $n_j$  do
4:    $\phi_1 \leftarrow \text{uniform}(\phi_{1i}, \phi_{1s})$ 
5:    $\phi_2 \leftarrow \text{uniform}(\phi_{2i}, \phi_{2s})$ 
6:    $\sigma^2 \leftarrow \text{uniform}(\sigma_{li}, \sigma_{ls})$ 
7:    $ar \leftarrow \text{array}([1, -\phi_1, -\phi_2])$  ▷ AR parameters with zero-lag and negation
8:    $ma \leftarrow \text{array}([1, \theta_1, \theta_2])$  ▷ MA parameters with zero-lag
9:    $z \leftarrow \text{arma\_generate\_sample}(ar, ma, N, \text{scale} = \sqrt{\sigma^2})$ 
10:   $X[f'x\{i\}, \{j\}'] \leftarrow \text{DataFrame}(z)$ 
11: end for return  $X$ 

```

---

## C.2 Função de simulação para modelos ARMA(p,q) com outlier

---

### Algorithm 2 Simulation with Additive Outlier

---

**Require:**  $i$ : Identifier for the current simulation

**Require:**  $nj$ : Number of time series per group

**Require:**  $N$ : Size of time series

**Require:**  $\phi_{1l}$ : Lower bound for  $\phi_1$

**Require:**  $\phi_{1s}$ : Upper bound for  $\phi_1$

**Require:**  $\phi_{2l}$ : Lower bound for  $\phi_2$

**Require:**  $\phi_{2s}$ : Upper bound for  $\phi_2$

**Require:**  $\sigma_{1l}$ : Lower bound for noise variance

**Require:**  $\sigma_{1s}$ : Upper bound for noise variance

**Require:**  $\theta_1$ : First moving average parameter

**Require:**  $\theta_2$ : Second moving average parameter

**Require:**  $aop$ : Proportion parameter for additive outlier

**Require:**  $\mu$ : Mean of normal distribution for additive outlier

**Require:**  $aom$ : Standard deviation of normal distribution for additive outlier

1: **Import necessary libraries**

2: Initialize  $X$  as an empty DataFrame

3: **for**  $j \leftarrow 1$  to  $nj$  **do**

4:    $\phi_1 \leftarrow \text{uniform}(\phi_{1l}, \phi_{1s})$

5:    $\phi_2 \leftarrow \text{uniform}(\phi_{2l}, \phi_{2s})$

6:    $\sigma^2 \leftarrow \text{uniform}(\sigma_{1l}, \sigma_{1s})$

7:    $ar \leftarrow \text{array}([1, -\phi_1, -\phi_2])$  ▷ AR parameters with zero-lag and negation

8:    $ma \leftarrow \text{array}([1, \theta_1, \theta_2])$  ▷ MA parameters with zero-lag

9:    $z \leftarrow \text{arma.generate\_sample}(ar, ma, N, \text{scale} = \sqrt{\sigma^2}) + \text{binomial}(1, aop, N) * \text{normal}(\mu, aom, N)$

10:    $X[f'x\{i\}, \{j\}'] \leftarrow \text{DataFrame}(z)$

11: **end for** **return**  $X$

---

## C.3 Função que simula o modelo AR(2) sem outlier

---

### Algorithm 3 AR(2) Simulation

---

**Require:**  $nj$ : Number of time series per group

**Require:**  $N$ : Size of time series

```

1: Import necessary libraries
2:  $\theta_1 \leftarrow 0$ 
3:  $\theta_2 \leftarrow 0$ 
4: Define parameter bounds for 3 groups
5:  $\phi_{1i} \leftarrow [0.05, 0.01, 0.12]$ 
6:  $\phi_{1s} \leftarrow [0.7, 1.2, 1.5]$ 
7:  $\phi_{2i} \leftarrow [-0.12, -0.36, -0.75]$ 
8:  $\phi_{2s} \leftarrow [-0.06, -0.25, -0.56]$ 
9:  $\text{noise}_i \leftarrow [0.1, 0.3, 0.9]$ 
10:  $\text{noise}_s \leftarrow [10, 3, 1.1]$ 
11: Initialize  $X$  as an empty DataFrame
12: for  $i \leftarrow 1$  to 3 do
13:    $ts \leftarrow \text{simulation}(i, nj, N, \phi_{1i}[i-1], \phi_{1s}[i-1], \phi_{2i}[i-1], \phi_{2s}[i-1], \text{noise}_i[i-1], \text{noise}_s[i-1], \theta_1, \theta_2)$ 
14:    $X \leftarrow \text{concat}([X, ts], \text{axis} = 1)$ 
15: end for
16:  $y \leftarrow \text{DataFrame}(\{y' : [1] * nj + [2] * nj + [3] * nj\})$ 
   return  $y, X$ 

```

---

## C.4 Função que monta uma base de dados de senos e cossenos pra estimar o M-periodograma

---

### Algorithm 4 Sin-Cos Calculation for Time Series

---

**Require:**  $N$ : Length of the time series

```

1: Initialize  $df$  as an empty DataFrame
2:  $\text{freq} \leftarrow \frac{\text{np.arange}(N)}{N}$ 
3:  $df \leftarrow \text{pd.DataFrame}(\{l' : \text{np.arange}(N) + 1\})$ 
4: for  $i \leftarrow 1$  to  $N - 1$  do
5:    $df[f'c\{i\}] \leftarrow \cos(df[l'] * \text{freq}[i])$ 
6:    $df[f's\{i\}] \leftarrow \sin(df[l'] * \text{freq}[i])$ 
7: end for
8:  $df \leftarrow df.drop(df.columns[0], \text{axis} = 1)$  return  $df$ 

```

---

## C.5 Função que simula o modelo AR(2) com outlier

---

**Algorithm 5** AR(2) Simulation with Additive Outlier

---

**Require:**  $nj$ : Number of time series per group

**Require:**  $N$ : Size of time series

**Require:**  $aop$ : Proportion parameter for additive outlier

**Require:**  $mu$ : Mean of normal distribution for additive outlier

**Require:**  $aom$ : Standard deviation of normal distribution for additive outlier

```

1: Import necessary libraries
2:  $theta1 \leftarrow 0$ 
3:  $theta2 \leftarrow 0$ 
4: Define parameter bounds for 3 groups
5:  $phi1_{li} \leftarrow [0.05, 0.01, 0.12]$ 
6:  $phi1_{ls} \leftarrow [0.7, 1.2, 1.5]$ 
7:  $phi2_{li} \leftarrow [-0.12, -0.36, -0.75]$ 
8:  $phi2_{ls} \leftarrow [-0.06, -0.25, -0.56]$ 
9:  $noise_{li} \leftarrow [0.1, 0.3, 0.9]$ 
10:  $noise_{ls} \leftarrow [10, 3, 1.1]$ 
11: Initialize  $X$  as an empty DataFrame
12: for  $i \leftarrow 1$  to 3 do
13:    $ts \leftarrow simulation\_with\_outlier(i, nj, N, phi1_{li}[i - 1], phi1_{ls}[i - 1], phi2_{li}[i - 1], phi2_{ls}[i - 1], noise_{li}[i - 1], noise_{ls}[i - 1], theta1, theta2, aop, mu, aom)$ 
14:    $X \leftarrow concat([X, ts], axis = 1)$ 
15: end for
16:  $y \leftarrow DataFrame(\{ 'y' : [1] * nj + [2] * nj + [3] * nj \})$ 
   return  $y, X$ 

```

---

## C.6 Função que retorna o periodograma clássico calculado usando o Fast Fourier Transform - FFT

---

### Algorithm 6 Calculation of Classical Periodogram

---

**Require:**  $X$ : Group of time series

**Require:**  $J$ : Number of groups

```

1:  $N \leftarrow \text{len}(X)$ 
2:  $n_j \leftarrow \frac{\text{len}(X.\text{columns})}{J}$  ▷ Number of replicates for group  $j$ 
3: function  $\text{JFFT}(X, j, k)$ 
4:    $\text{per\_ols} \leftarrow \text{pd.DataFrame}()$ 
5:    $\text{fft} \leftarrow \text{np.fft.fft}(X[f'x\{j\}, \{k\}'])$ 
6:    $\text{per} \leftarrow \frac{(\text{np.abs}(\text{fft})^2)}{(N*2*\pi)}$ 
7:    $\text{per\_ols}[f'h\{j\}, \{k\}'] \leftarrow \text{per}$ 
8:    $\text{per\_ols} \leftarrow \text{per\_ols.iloc}[: N, ]$ 
9:   return  $\text{per\_ols}$ 
10: end function
11:  $Z \leftarrow \text{pd.DataFrame}()$ 
12: for  $j \leftarrow 1$  to  $J$  do
13:   for  $k \leftarrow 1$  to  $n_j$  do
14:      $z \leftarrow \text{jfft}(X, j, k)$ 
15:      $Z \leftarrow \text{pd.concat}([Z, z], \text{axis} = 1)$ 
16:   end for
17: end for
   return  $Z$ 

```

---



## C.7 Função que retorna o M-periodograma baseado na função Huber

---

**Algorithm 7** Calculation of M-Periodogram based on Huber

---

**Require:**  $X_{\text{train}}$ : Group of time series

**Require:**  $J$ : Number of groups

**Require:**  $\delta$ : Huber parameter

```

1:  $N \leftarrow \text{len}(X_{\text{train}})$ 
2:  $n_j \leftarrow \frac{\text{len}(X_{\text{train.columns}})}{J}$  ▷ Number of replicates for group  $j$ 
3:  $df \leftarrow \text{sin\_cos}(N)$ 
4: Initialize  $Z_h$  as an empty DataFrame
5: for  $j \leftarrow 1$  to  $J$  do
6:   for  $k \leftarrow 1$  to  $n_j$  do
7:     Initialize  $per\_hub$  as an empty list
8:     for  $a \leftarrow 1$  to  $\lfloor \frac{\text{len}(df)}{2} \rfloor$  do  $X \leftarrow df[[f'c\{a\}', f's\{a\}']]$ 
9:        $Y \leftarrow \text{pd.DataFrame}(X_{\text{train}}[f'x\{j\}, \{k\}'])$ 
10:       $huber\_t \leftarrow \text{sm.RLM}(Y, X, M=\text{sm.robust.norms.HuberT}(t = \delta))$ 
11:       $model2 \leftarrow huber\_t.fit()$ 
12:       $B2 \leftarrow model2.params[0]$ 
13:       $A2 \leftarrow model2.params[1]$ 
14:       $H2 \leftarrow \frac{(A2^2 + B2^2) * N}{8 * \pi}$ 
15:      Append  $H2$  to  $per\_hub$ 
16:
17:
18:       $c \leftarrow \text{np.ones}(\text{len}(X_{\text{train}}))$ 
19:      for  $m \leftarrow 0$  to  $\lfloor \frac{\text{len}(X_{\text{train}})}{2} \rfloor - 1$  do  $c[2*m] \leftarrow -1$ 
20:
21:
22:        Initialize  $df\_beta$  as an empty list
23:        for  $l \leftarrow 0$  to  $\text{len}(X_{\text{train}}) - 1$  do
24:           $d \leftarrow X_{\text{train}}[f'x\{j\}, \{k\}'].iloc[l] * c[l]$ 
25:          Append  $d$  to  $df\_beta$ 
26:        end for
27:         $df\_beta \leftarrow \text{pd.DataFrame}(df\_beta)$ 
28:         $H\_final \leftarrow df\_beta.sum() / \text{len}(X_{\text{train}})$ 
29:        Append  $\frac{(H\_final[0]^2) * N}{8 * \pi}$  to  $per\_hub$ 
30:        for  $a \leftarrow \lfloor \frac{\text{len}(df)}{2} \rfloor + 1$  to  $\text{len}(df) - 1$  do  $X \leftarrow$ 
31:           $df[[f'c\{a\}', f's\{a\}']]$ 
32:           $Y \leftarrow \text{pd.DataFrame}(X_{\text{train}}[f'x\{j\}, \{k\}'])$ 
33:           $huber\_t \leftarrow \text{sm.RLM}(Y, X, M=\text{sm.robust.norms.HuberT}(t = \delta))$ 
34:           $model2 \leftarrow huber\_t.fit()$ 
35:           $B2 \leftarrow model2.params[0]$ 
36:           $A2 \leftarrow model2.params[1]$ 
37:           $H2 \leftarrow \frac{(A2^2 + B2^2) * N}{8 * \pi}$ 
38:          Append  $H2$  to  $per\_hub$ 
39:        end for
40:         $Z_h[f'h\{j\}, \{k\}'] \leftarrow per\_hub$ 
41:
42:   return  $Z_h = 0$ 

```

---



## C.8 Função que retorna o classical cepstra Multitaper periodogram

---

### Algorithm 8 Cepstral Coefficients Calculation with Multitaper Periodogram (Part 1)

---

**Require:**  $y_{\text{train}}$ : Vector of labels

**Require:**  $X_{\text{train}}$ : Group of time series for  $J$  populations

**Require:**  $L$ : Number of cepstral coefficients

**Require:**  $R$ : Number of tapers for periodogram multitaper

**Require:**  $c\_periodogram\_fft$ : Function to calculate the periodogram using FFT

```

1:  $N \leftarrow \text{length}(X_{\text{train}})$ 
2:  $n_j \leftarrow \frac{\text{number of columns in } X_{\text{train}}}{J} \quad \triangleright$  Number of time series (replicates) for the  $j$ th population
3: Initialize taper as an empty DataFrame with column  $t \leftarrow \{1, \dots, N\}$ 
4: for  $r \leftarrow 1$  to  $R$  do
5:   Initialize  $h$  as an empty list
6:   for  $t \in \{1, \dots, N\}$  do
7:      $g \leftarrow \sqrt{\frac{2}{N+1}} \sin\left(\frac{\pi tr}{N+1}\right)$ 
8:     Append  $g$  to  $h$ 
9:   end for
10:   $\text{taper}[\text{taper } r] \leftarrow h$ 
11: end for
12: Initialize  $X_{\text{train\_tapered}}$  as an empty dictionary
13: for  $r \leftarrow 1$  to  $R$  do
14:   $X_{\text{train\_tapered}}[r] \leftarrow$  empty DataFrame
15:  for  $j \leftarrow 1$  to  $J$  do
16:    for  $k \leftarrow 1$  to  $n_j$  do
17:       $X_{\text{train\_tapered}}[r][x_{j,k}] \leftarrow X_{\text{train}}[x_{j,k}] \times \text{taper}[\text{taper } r]$ 
18:    end for
19:  end for
20: end for
21: Initialize  $Z_h$  as an empty dictionary
22: for  $r \leftarrow 1$  to  $R$  do
23:   $Z_h[r] \leftarrow$  empty DataFrame
24:   $z_h \leftarrow c\_periodogram\_fft(X_{\text{train\_tapered}}[r], J)$ 
25:   $Z_h[r] \leftarrow \text{concat}(Z_h[r], z_h, \text{axis} = 1)$ 
26: end for
27: if  $R == 7$  then
28:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{7} \sum_{r=1}^7 Z_h[r]\right)$ 
29: else if  $R == 6$  then
30:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{6} \sum_{r=1}^6 Z_h[r]\right)$ 
31: else if  $R == 5$  then
32:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{5} \sum_{r=1}^5 Z_h[r]\right)$ 
33: else if  $R == 4$  then
34:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{4} \sum_{r=1}^4 Z_h[r]\right)$ 
35: else if  $R == 3$  then
36:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{3} \sum_{r=1}^3 Z_h[r]\right)$ 
37: else if  $R == 2$  then
38:   $D_{\text{hat}} \leftarrow \log\left(\frac{1}{2} \sum_{r=1}^2 Z_h[r]\right)$ 
39: else
40:   $D_{\text{hat}} \leftarrow \log(Z_h[1])$ 
41: end if

```

**Algorithm 9** Cepstral Coefficients Calculation with Multitaper Periodogram (Part 2)

---

```

1: Initialize  $df3$  as an empty DataFrame
2: for  $j \leftarrow 1$  to  $J$  do
3:   for  $k \leftarrow 1$  to  $n_j$  do
4:     Initialize  $df$  and  $df2$  as empty DataFrames
5:     Initialize  $z2$  as an empty list
6:      $c_0 \leftarrow \frac{\sum D_{\text{hat}}[h_{j,k}]}{N}$ 
7:     for  $l \leftarrow 1$  to  $L - 1$  do
8:       Initialize  $z$  as an empty list
9:       for  $i \leftarrow 0$  to  $N - 1$  do
10:         $g \leftarrow D_{\text{hat}}[h_{j,k}][i] \cos\left(\frac{il}{N}\right) / N$ 
11:        Append  $g$  to  $z$ 
12:      end for
13:       $df[c_{j,k}] \leftarrow z$ 
14:      Append  $df.sum()$  to  $z2$ 
15:    end for
16:     $df2 \leftarrow \text{concat}(c_0, z2).reset\_index(drop = True)$ 
17:     $df2.iloc[0, 1] \leftarrow \frac{\sum D_{\text{hat}}[h_{j,k}]}{N}$ 
18:     $df2 \leftarrow df2.drop(df2.columns[0], axis = 1)$ 
19:     $df3 \leftarrow \text{concat}(df3, df2, axis = 1)$ 
20:  end for
21: end for
22:  $df3 \leftarrow df3.T.reset\_index(drop = True)$ 
23:  $df3 \leftarrow df3.add\_prefix('cep\_')$ 
24:  $df3['y'] \leftarrow y_{\text{train}}$  return  $df3$ 

```

---

**Algorithm 10** M-Cepstral Calculation Based on M-Periodogram (Part 1)**Require:**  $y$ : Vector of labels**Require:**  $X$ : Group of time series for  $J$  populations**Require:**  $L$ : Number of cepstral coefficients**Require:**  $R$ : Number of tapers for periodogram multitaper**Require:**  $\delta$ : Huber parameter used in robust estimation**Require:** *huber\_per*: Function to calculate the periodogram using the Huber estimator

```

1:  $N \leftarrow \text{length}(X)$ 
2:  $n_{j\_train} \leftarrow \frac{\text{number of columns in } X}{J} \triangleright$  Number of time series (replicates) for the  $j$ th population
3: Initialize taper as an empty DataFrame with column  $t \leftarrow \{1, \dots, N\}$ 
4: for  $r \leftarrow 1$  to  $R$  do
5:   Initialize  $h$  as an empty list
6:   for  $t \in \{1, \dots, N\}$  do
7:      $g \leftarrow \sqrt{\frac{2}{N+1}} \sin\left(\frac{\pi tr}{N+1}\right)$ 
8:     Append  $g$  to  $h$ 
9:   end for
10:   $\text{taper}[\text{taper } r] \leftarrow h$ 
11: end for
12: Initialize  $X\_train\_tapered$  as an empty dictionary
13: for  $r \leftarrow 1$  to  $R$  do
14:   $X\_train\_tapered[r] \leftarrow$  empty DataFrame
15:  for  $j \leftarrow 1$  to  $J$  do
16:    for  $k \leftarrow 1$  to  $n_{j\_train}$  do
17:       $X\_train\_tapered[r][x_{j,k}] \leftarrow X[x_{j,k}] \times \text{taper}[\text{taper } r]$ 
18:    end for
19:  end for
20: end for
21: Initialize  $Z_h$  as an empty dictionary
22: for  $r \leftarrow 1$  to  $R$  do
23:   $Z_h[r] \leftarrow \text{huber\_per}(X\_train\_tapered[r], J, \delta)$ 
24: end for

```

**Algorithm 11** M-Cepstral Calculation Based on M-Periodogram (Part 2)

---

```

1: Calculate  $D_{\text{hat}}$ :
2: if  $R == 7$  then
3:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{7} \sum_{r=1}^7 Z_h[r]\right)$ 
4: else if  $R == 6$  then
5:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{6} \sum_{r=1}^6 Z_h[r]\right)$ 
6: else if  $R == 5$  then
7:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{5} \sum_{r=1}^5 Z_h[r]\right)$ 
8: else if  $R == 4$  then
9:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{4} \sum_{r=1}^4 Z_h[r]\right)$ 
10: else if  $R == 3$  then
11:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{3} \sum_{r=1}^3 Z_h[r]\right)$ 
12: else if  $R == 2$  then
13:    $D_{\text{hat}} \leftarrow \log\left(\frac{1}{2} \sum_{r=1}^2 Z_h[r]\right)$ 
14: else
15:    $D_{\text{hat}} \leftarrow \log(Z_h[1])$ 
16: end if
17: Initialize  $df3$  as an empty DataFrame
18: for  $j \leftarrow 1$  to  $J$  do
19:   for  $k \leftarrow 1$  to  $nj\_train$  do
20:     Initialize  $df$  and  $df2$  as empty DataFrames
21:     Initialize  $z2$  as an empty list
22:      $c_0 \leftarrow \frac{\sum D_{\text{hat}}[h_{j,k}]}{N}$ 
23:     for  $l \leftarrow 1$  to  $L - 1$  do
24:       Initialize  $z$  as an empty list
25:       for  $i \leftarrow 0$  to  $N - 1$  do
26:          $g \leftarrow D_{\text{hat}}[h_{j,k}][i] \cos\left(\frac{il}{N}\right) / N$ 
27:         Append  $g$  to  $z$ 
28:       end for
29:        $df[c_{j,k}] \leftarrow z$ 
30:       Append  $df.sum()$  to  $z2$ 
31:     end for
32:      $df2 \leftarrow \text{concat}(c_0, z2).reset\_index(drop = True)$ 
33:      $df2.iloc[0, 1] \leftarrow \frac{\sum D_{\text{hat}}[h_{j,k}]}{N}$ 
34:      $df2 \leftarrow df2.drop(df2.columns[0], axis = 1)$ 
35:      $df3 \leftarrow \text{concat}(df3, df2, axis = 1)$ 
36:   end for
37: end for
38:  $df3 \leftarrow df3.T.reset\_index(drop = True)$ 
39:  $df3 \leftarrow df3.add\_prefix('cep\_')$ 
40:  $df3['y'] \leftarrow y$  return  $df3$ 

```

---

## C.9 Algoritmo que retorna a taxa de classificação e a Matriz de Confusão para o caso clássico

---

**Algorithm 12** Classification Using Classical FFT by Group

---

**Require:**  $y_{\text{meu}}$ : Vector of labels for training data

**Require:**  $X_{\text{meu}}$ : Training data matrix

**Require:**  $n_{j_{\text{test}}}$ : Number of test samples

**Require:**  $X_{\text{test2}}$ : Test data matrix

**Require:**  $y_{\text{test}}$ : Vector of labels for test data

- 1: **Import necessary libraries**
- 2:  $clf \leftarrow \text{LDA}(n\_components = 2)$
- 3:  $clf.fit(X_{\text{meu}}, y_{\text{meu}})$
- 4:  $y_{\text{pred}} \leftarrow clf.predict(X_{\text{test2}}.iloc[:])$
- 5:  $score \leftarrow \text{pd.DataFrame}()$
- 6:  $score[y'_{\text{pred}}] \leftarrow y_{\text{pred}}$
- 7:  $score[y'] \leftarrow y_{\text{test}}$
- 8:  $confussion \leftarrow \text{confusion\_matrix}(score[y'], score[y'_{\text{pred}}])$
- 9:  $c \leftarrow \text{pd.DataFrame}(data = confussion)/n_{j_{\text{test}}}$
- 10:  $d \leftarrow \text{np.diag}(c)$  **return**  $d$

---

## C.10 Algoritmo que retorna a taxa de classificação e a Matriz de Confusão para o caso M-periodograma

---

**Algorithm 13** Classification Using M-Periodogram

---

**Require:**  $y_{\text{meu}}$ : Vector of labels for training data

**Require:**  $X_{\text{meu}}$ : Training data matrix

**Require:**  $n_{j_{\text{test}}}$ : Number of test samples

**Require:**  $X_{\text{test2}}$ : Test data matrix

**Require:**  $y_{\text{test}}$ : Vector of labels for test data

- 1: **Import necessary libraries**
- 2:  $clf \leftarrow \text{LDA}(n\_components = 2)$
- 3:  $clf.fit(X_{\text{meu}}, y_{\text{meu}})$
- 4:  $y_{\text{pred}} \leftarrow clf.predict(X_{\text{test2}}.iloc[:])$
- 5:  $score \leftarrow \text{pd.DataFrame}()$
- 6:  $score[y'_{\text{pred}}] \leftarrow y_{\text{pred}}$
- 7:  $score[y'] \leftarrow y_{\text{test}}$
- 8:  $confussion \leftarrow \text{confusion\_matrix}(score[y'], score[y'_{\text{pred}}])$
- 9:  $c \leftarrow \text{pd.DataFrame}(data = confussion)/n_{j_{\text{test}}}$
- 10:  $d \leftarrow \text{np.diag}(c)$  **return**  $d$

---

## C.11 Retorna a Taxa de Classificação Média e o erro padrão baseado no número de amostras - Samples para o periodograma clássico

---

**Algorithm 14** Classification Mean Classic using FFT by Group

---

**Require:**  $J$ : Number of groups  
**Require:**  $n_j$ : Number of time series per group  
**Require:**  $N$ : Size of time series  
**Require:**  $L$ : Number of cepstral coefficients  
**Require:**  $R$ : Number of tapers for periodogram multitaper  
**Require:**  $n_{\text{test}}$ : Number of test samples

- 1: **Import necessary libraries**
- 2: Initialize  $df3$  as an empty DataFrame
- 3: **for**  $i \leftarrow 0$  to  $n_{\text{test}} - 1$  **do**
- 4:      $y_{\text{train}}, X_{\text{train}} \leftarrow \text{ar2}(n_j, N)$
- 5:      $a \leftarrow \text{tx\_class\_C\_fft\_by\_group}(y_{\text{train}}, X_{\text{train}}, J, L, R)$
- 6:      $df3[f'_{\text{test}}\{i + 1\}'] \leftarrow a.T$
- 7: **end for** **return**  $df3$

---

## C.12 Retorna a Taxa de Classificação Média e o erro padrão baseado no número de amostras - Samples para o M-periodograma

---

**Algorithm 15** Classification Mean Huber

---

**Require:**  $J$ : Number of groups  
**Require:**  $n_j$ : Number of time series per group  
**Require:**  $N$ : Size of time series  
**Require:**  $L$ : Number of cepstral coefficients  
**Require:**  $R$ : Number of tapers for periodogram multitaper  
**Require:**  $n_{\text{test}}$ : Number of test samples  
**Require:**  $t$ : Huber parameter

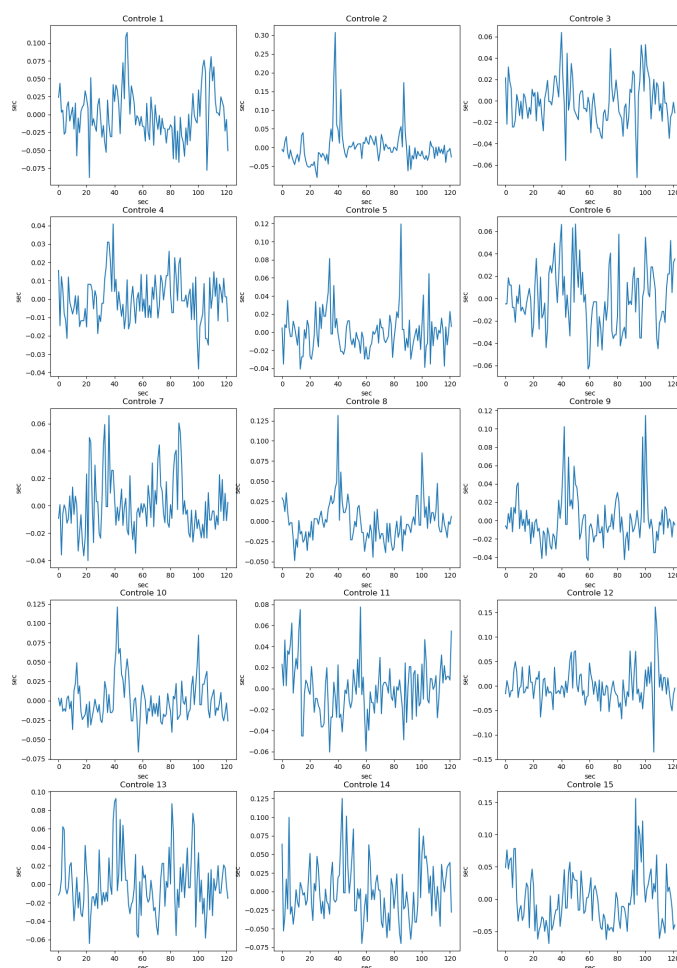
- 1: **Import necessary libraries**
- 2: Initialize  $df3$  as an empty DataFrame
- 3: **for**  $i \leftarrow 0$  to  $n_{\text{test}} - 1$  **do**
- 4:      $y_{\text{train}}, X_{\text{train}} \leftarrow \text{ar2}(n_j, N)$
- 5:      $D_{\text{hat}} \leftarrow \text{tx\_class\_M}(y_{\text{train}}, X_{\text{train}}, J, L, R, t)$
- 6:      $df3[f'_{\text{test}}\{i + 1\}'] \leftarrow D_{\text{hat}}.T$
- 7: **end for** **return**  $df3$

---

# Apêndice D

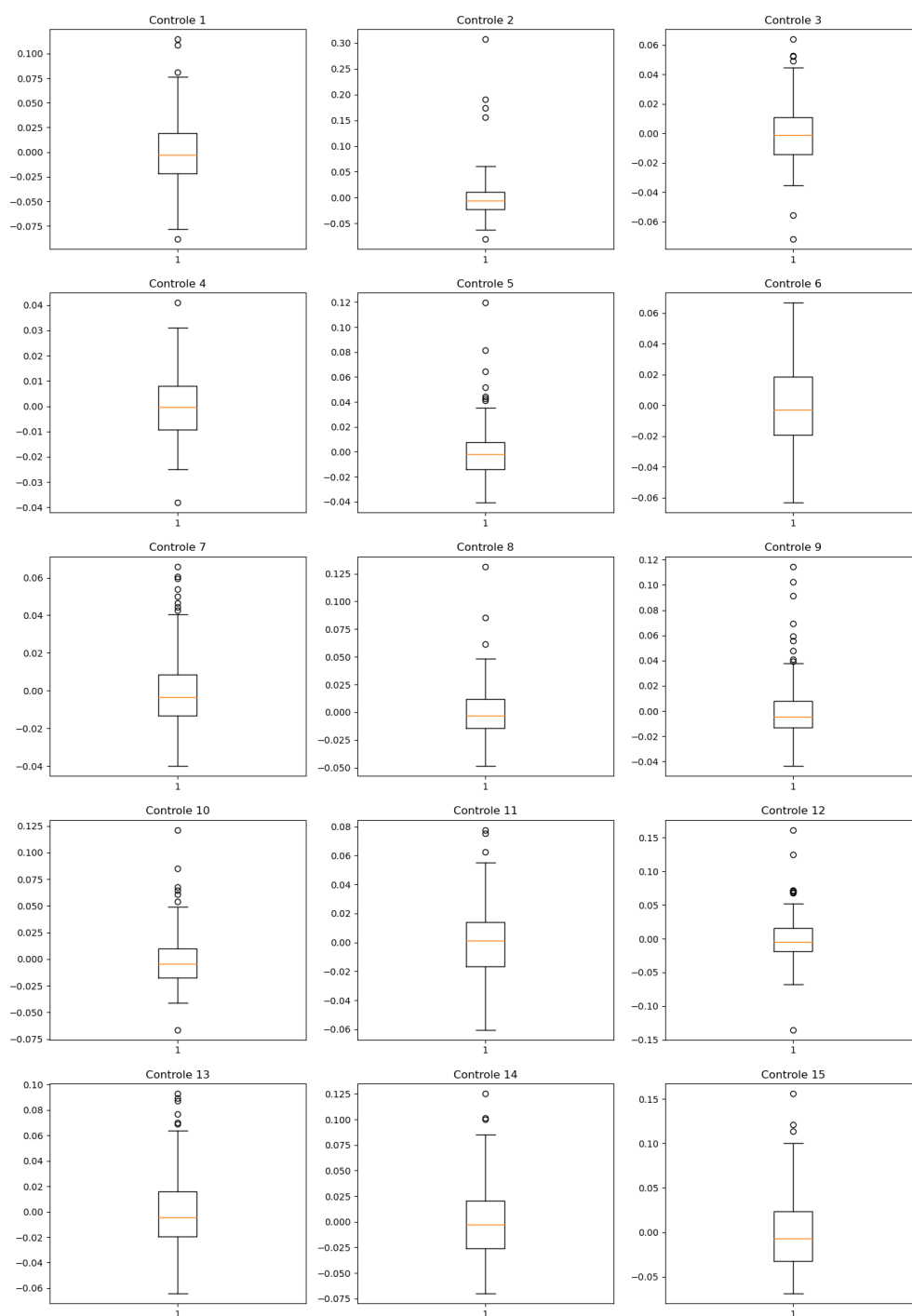
## Figuras Adicionais

Figura D.1: Séries Temporais modificada por [Krafty \(2016\)](#) dos ciclos de passada por segundo de 16 indivíduos saudáveis de controle



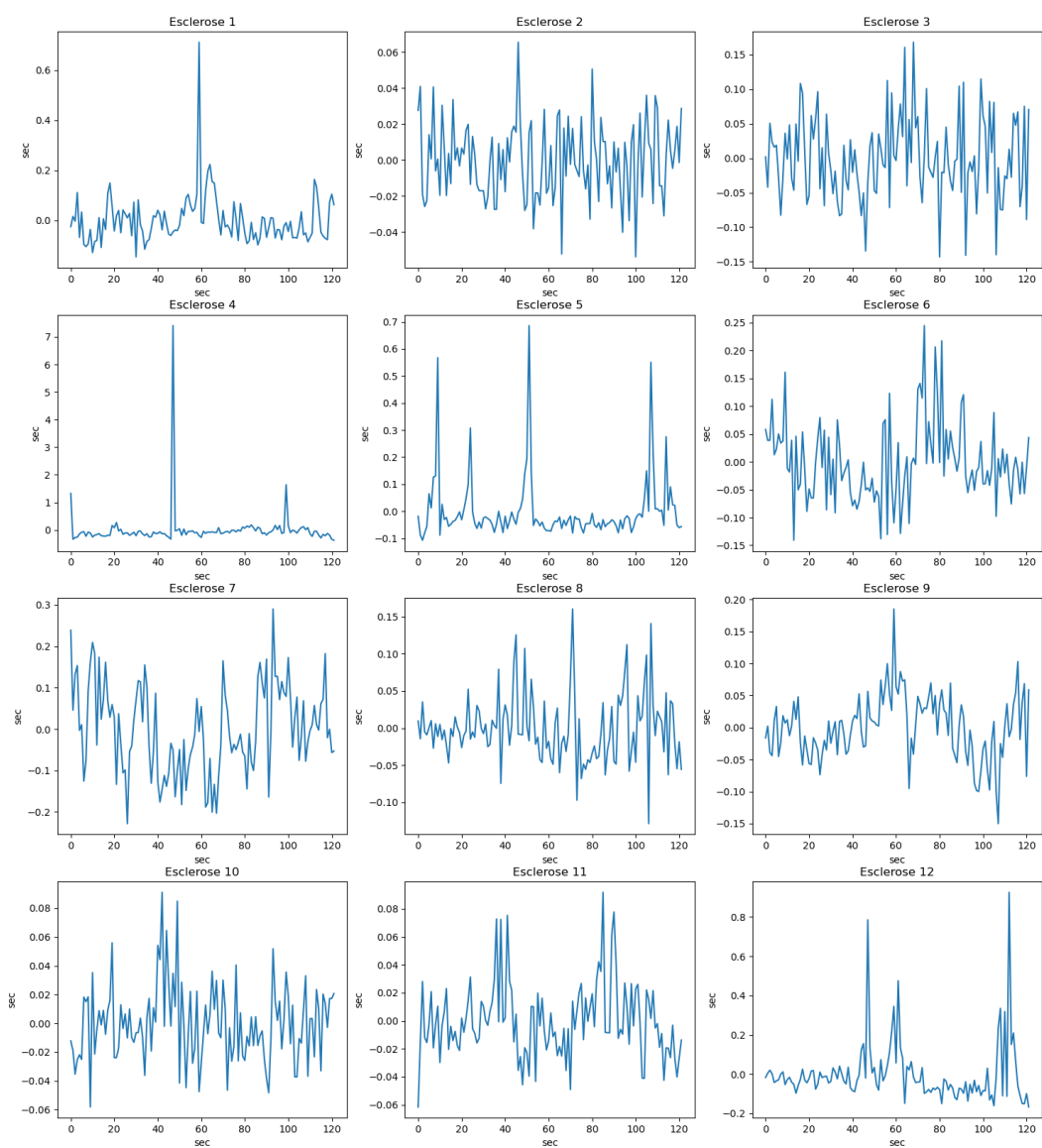
Fonte: Elaborado pelo autor.

Figura D.2: Boxplot das Séries Temporais modificadas por [Krafty \(2016\)](#) dos ciclos de passada por segundo de 16 indivíduos saudáveis de controle



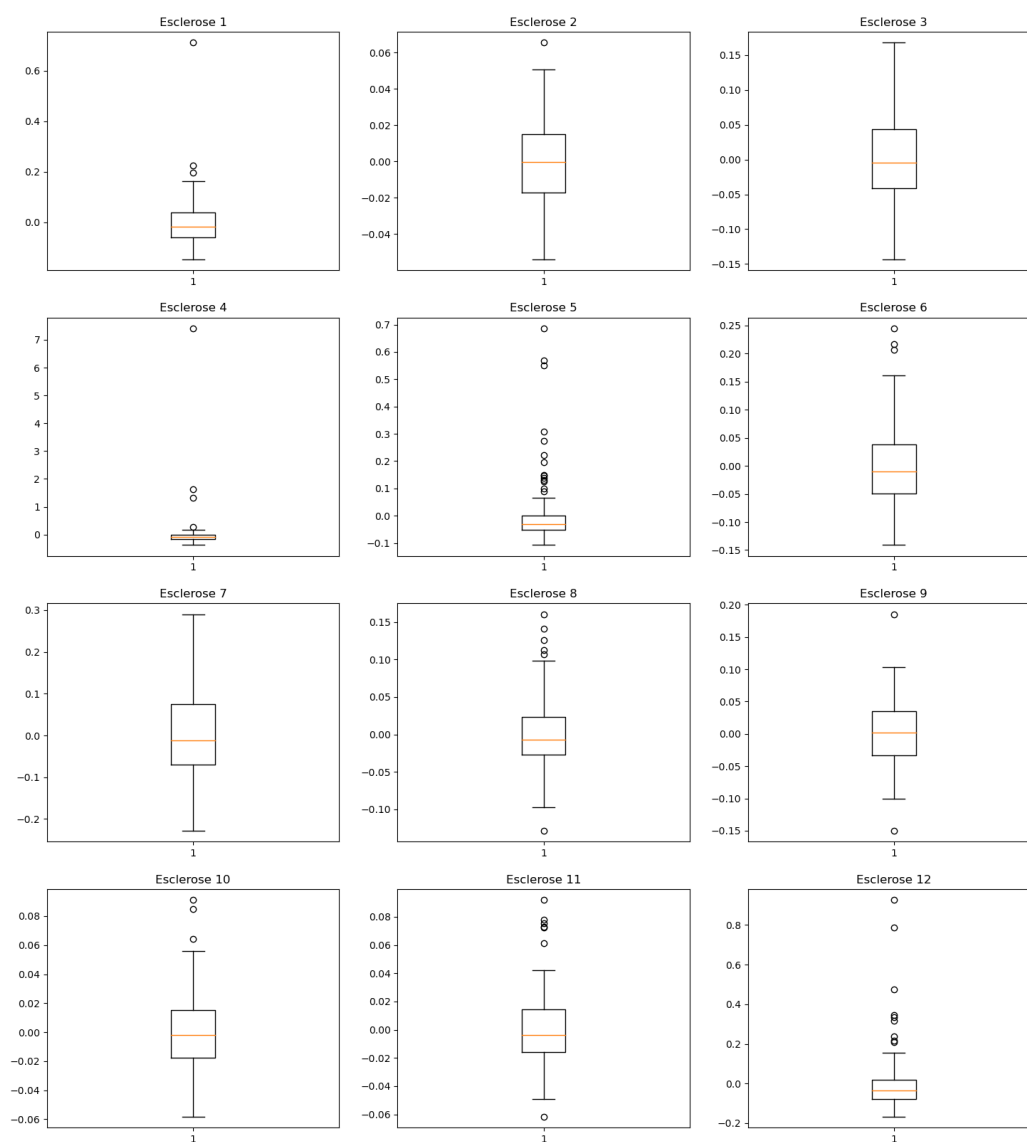
Fonte: Elaborado pelo autor.

Figura D.3: Série Temporal modificada por Krafty (2016) dos ciclos de passada por segundo de 12 indivíduos com doença de Esclerose



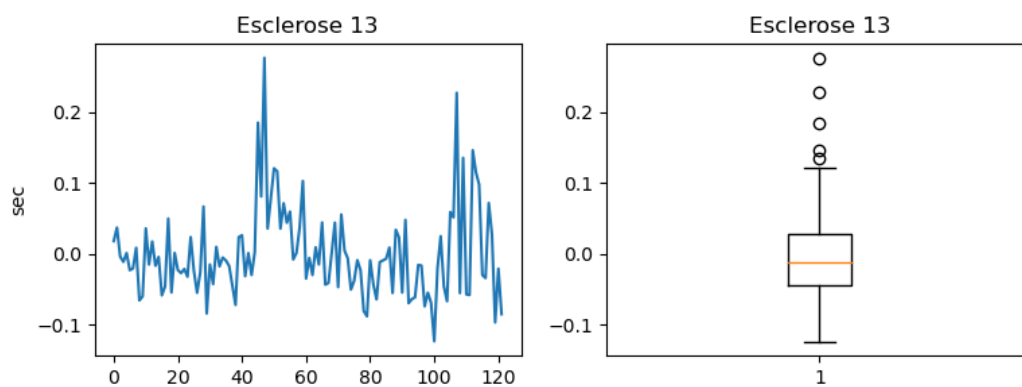
Fonte: Elaborado pelo autor.

Figura D.4: Boxplot das Séries Temporais modificadas por [Krafty \(2016\)](#) dos ciclos de passada por segundo de 12 indivíduos com doença de Esclerose



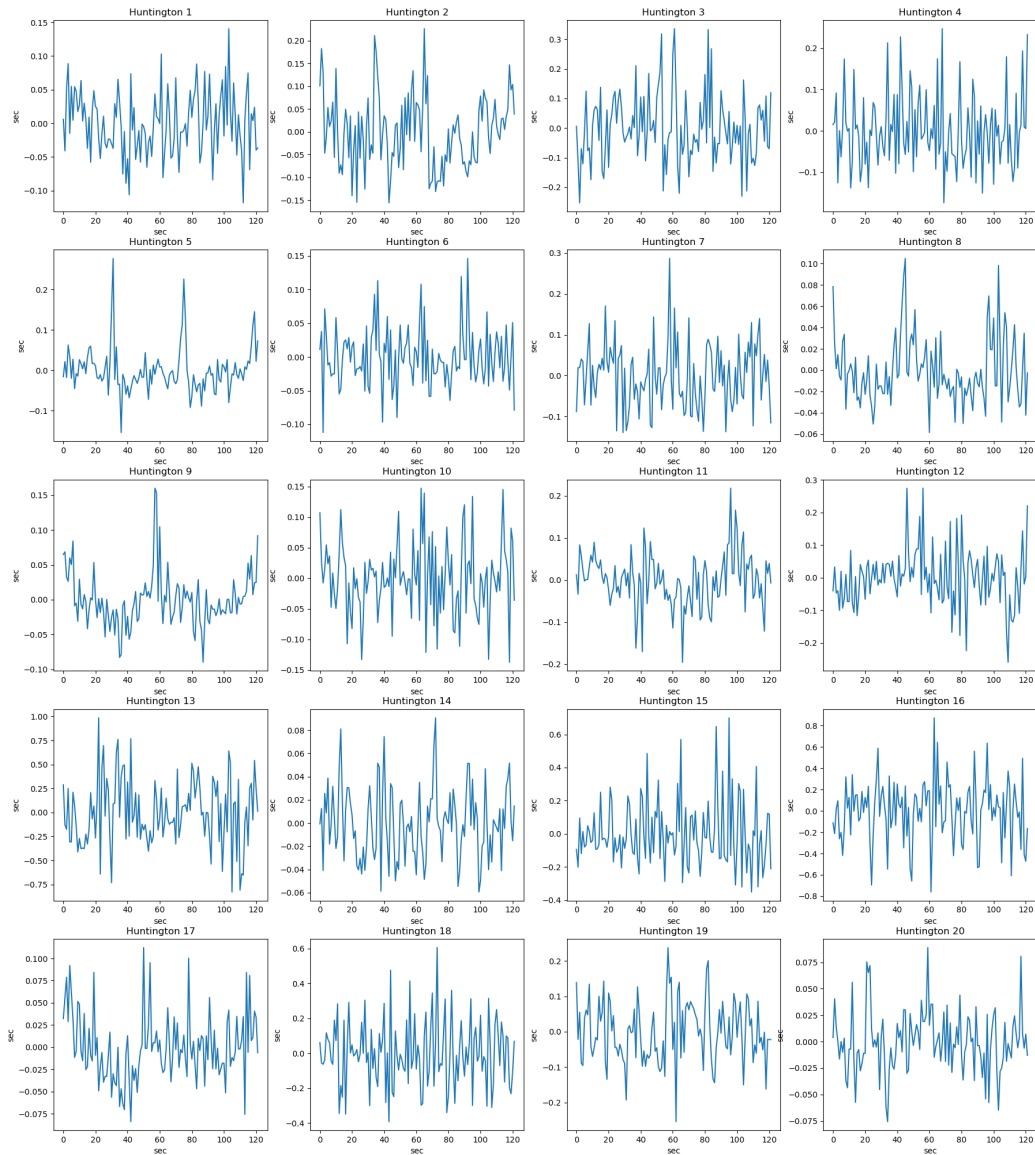
Fonte: Elaborado pelo autor.

Figura D.5: Gráfico e Boxplot da Série Temporal modificada por Krafty (2016) de um indivíduo com doença de Esclerose



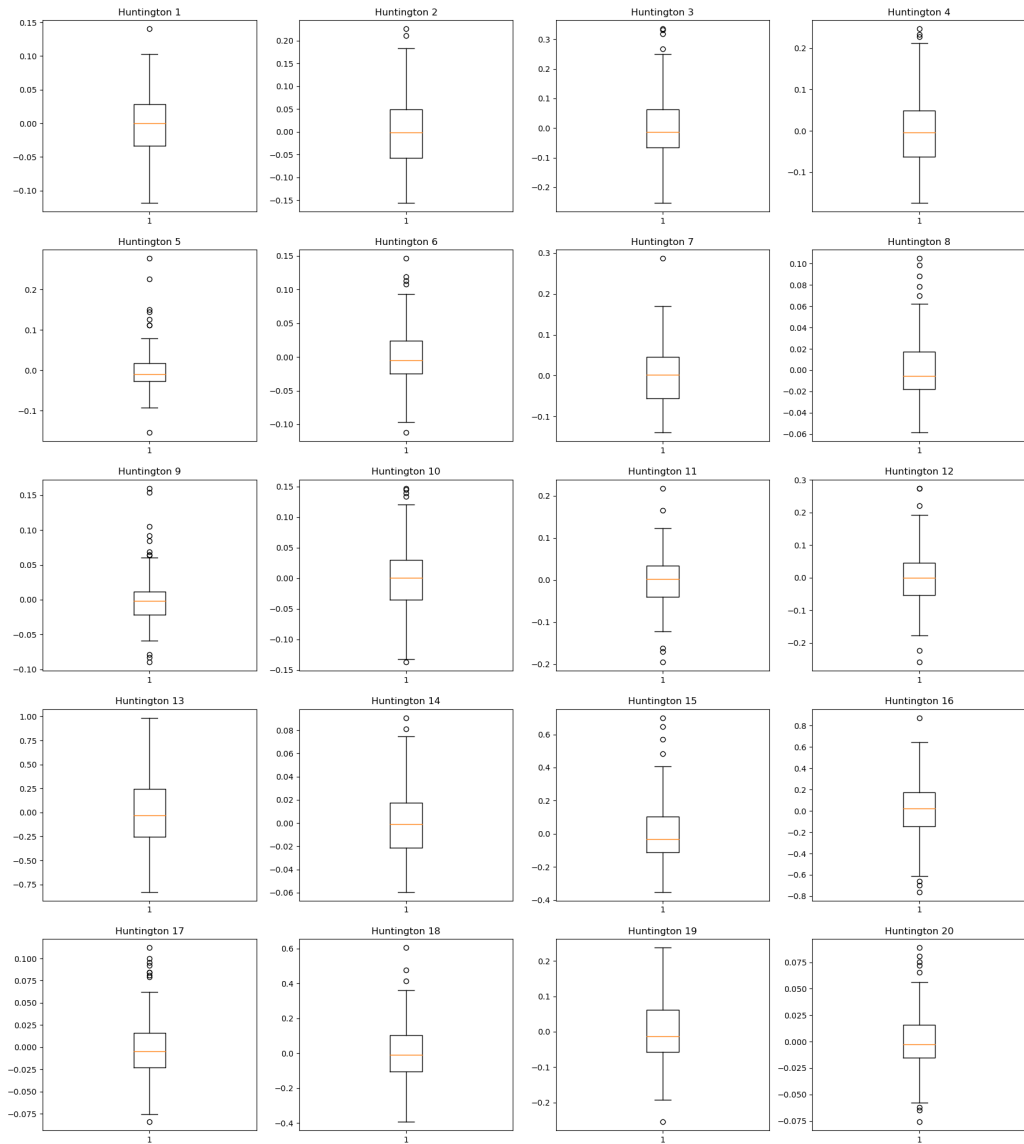
Fonte: Elaborado pelo autor.

Figura D.6: Séries Temporais modificadas por Krafty (2016) de ciclos de passada por segundo a cada segundo de 20 indivíduos de controle



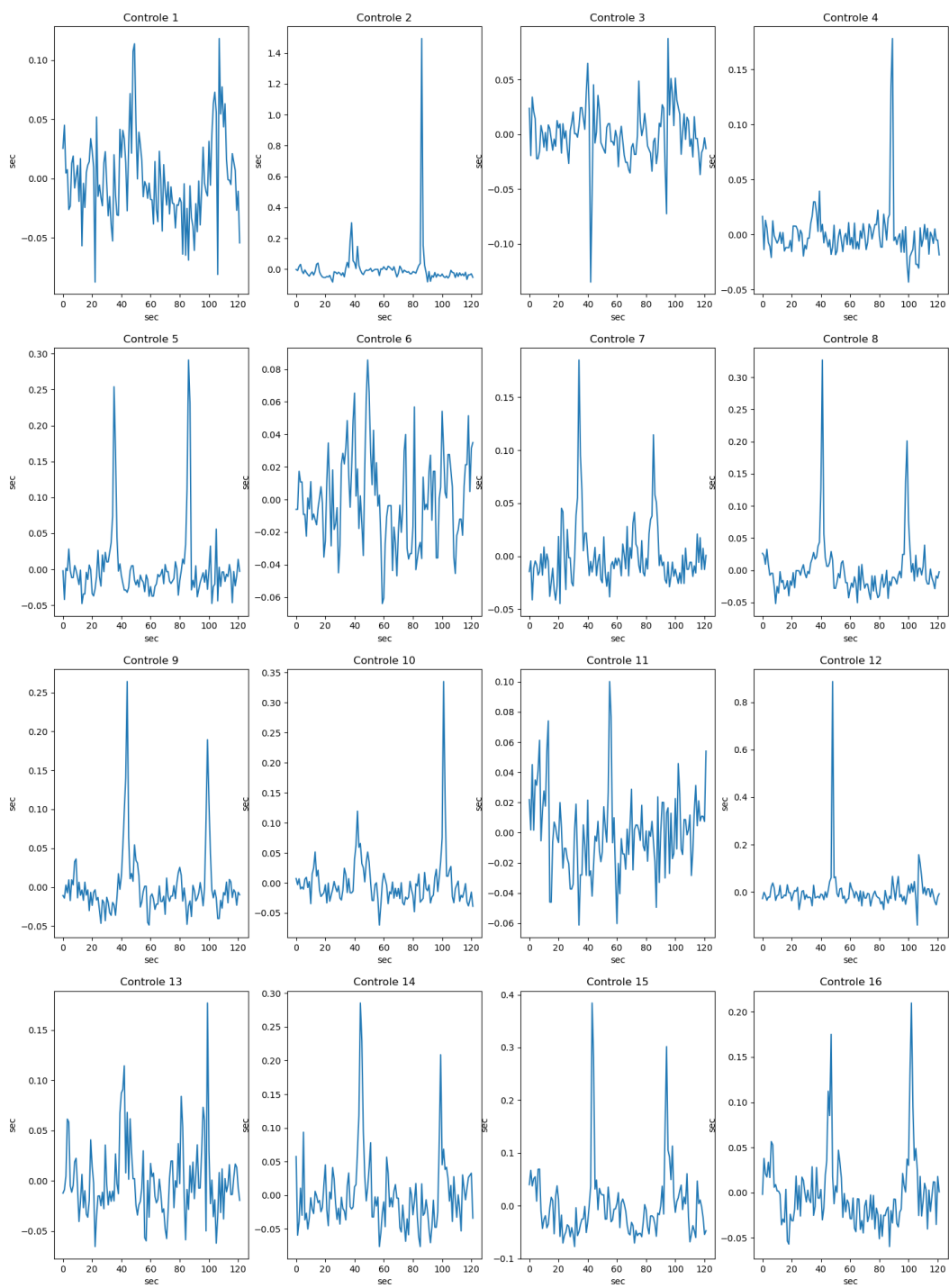
Fonte: Elaborado pelo autor.

Figura D.7: Boxplot das Séries Temporais modificadas por **Krafty (2016)** de ciclos de passada por segundo de 20 indivíduos de controle



Fonte: Elaborado pelo autor.

Figura D.8: Séries Temporais de ciclos de passada por segundo de 16 indivíduos de controle



Fonte: Elaborado pelo autor.

Figura D.9: Boxplot das Séries Temporais de ciclos de passada por segundo de 16 indivíduos de controle

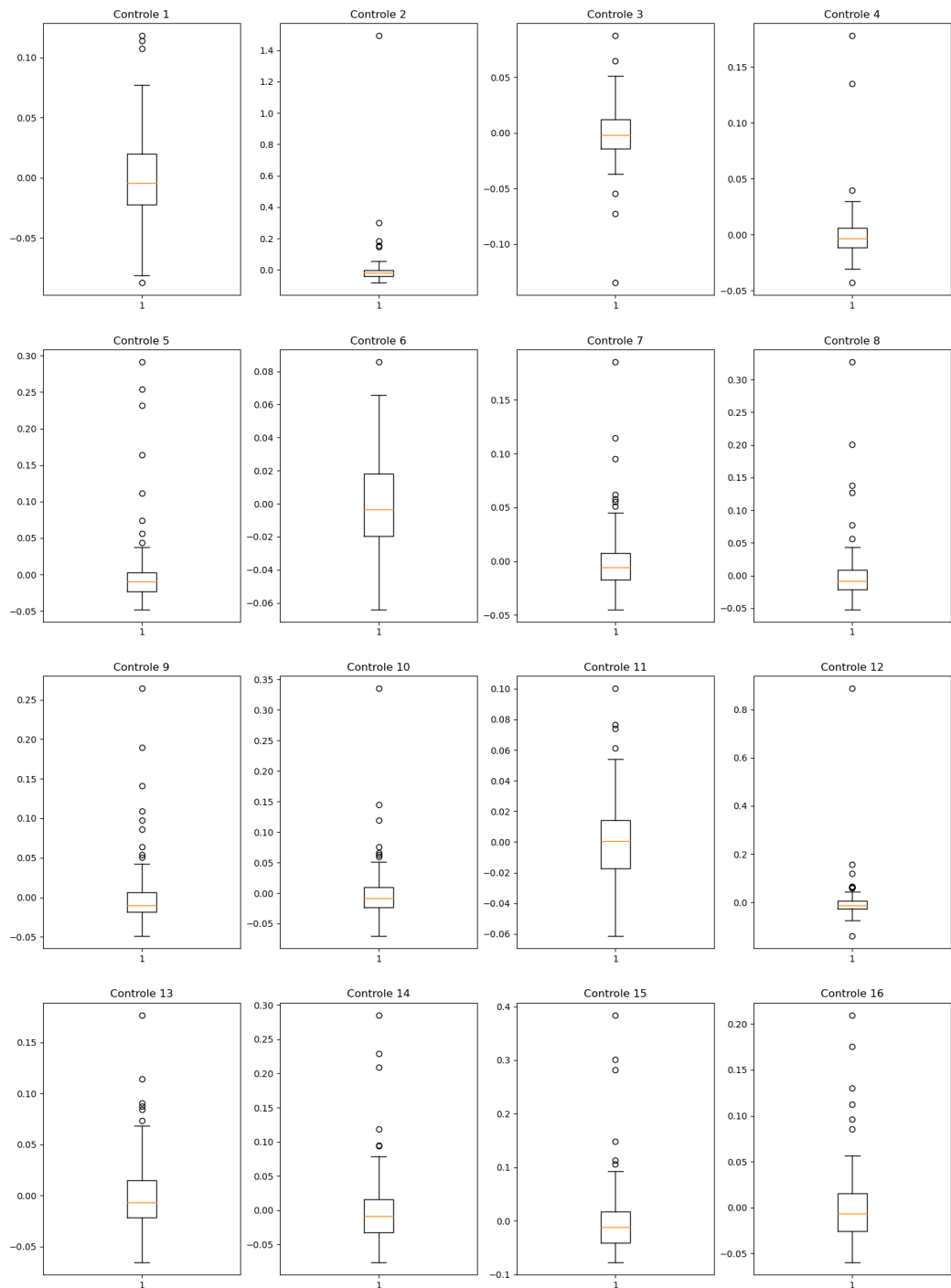
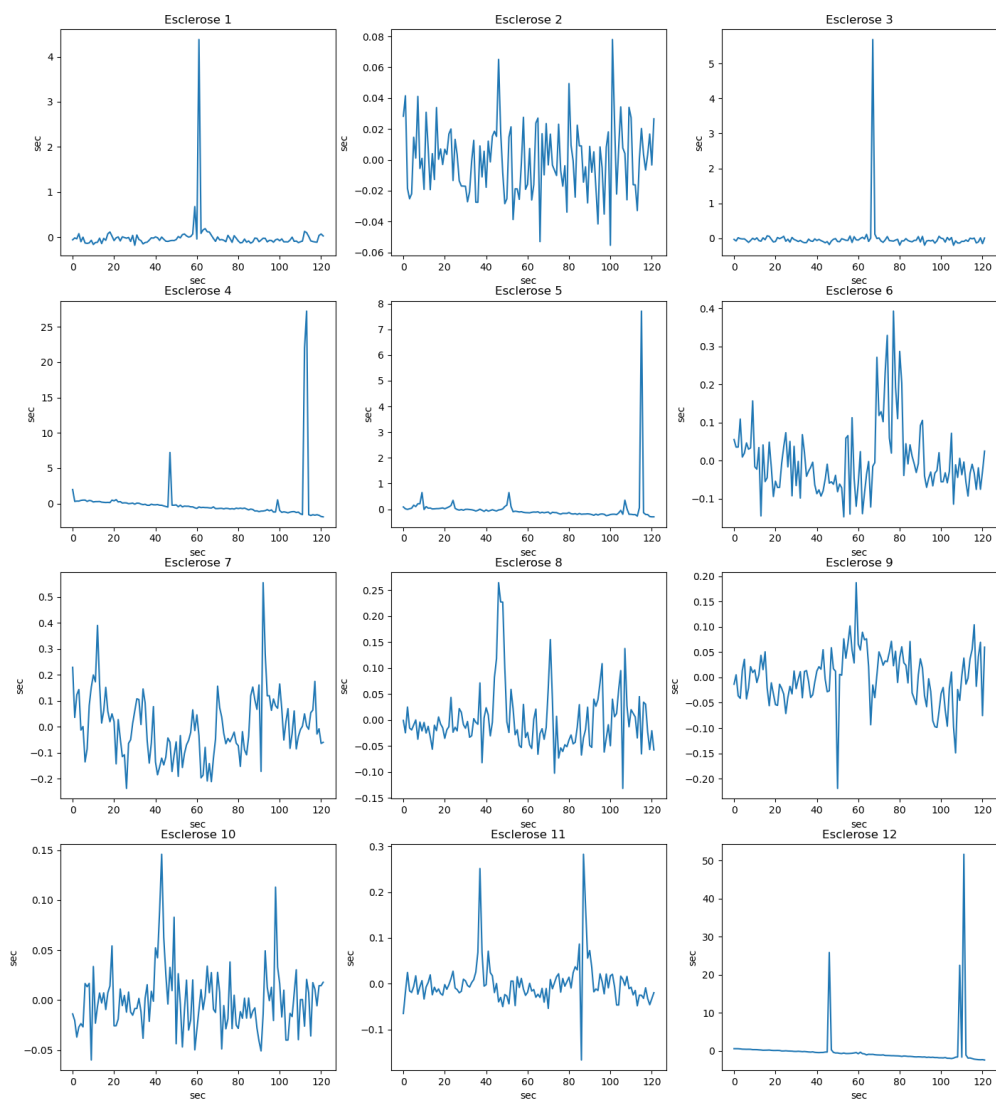
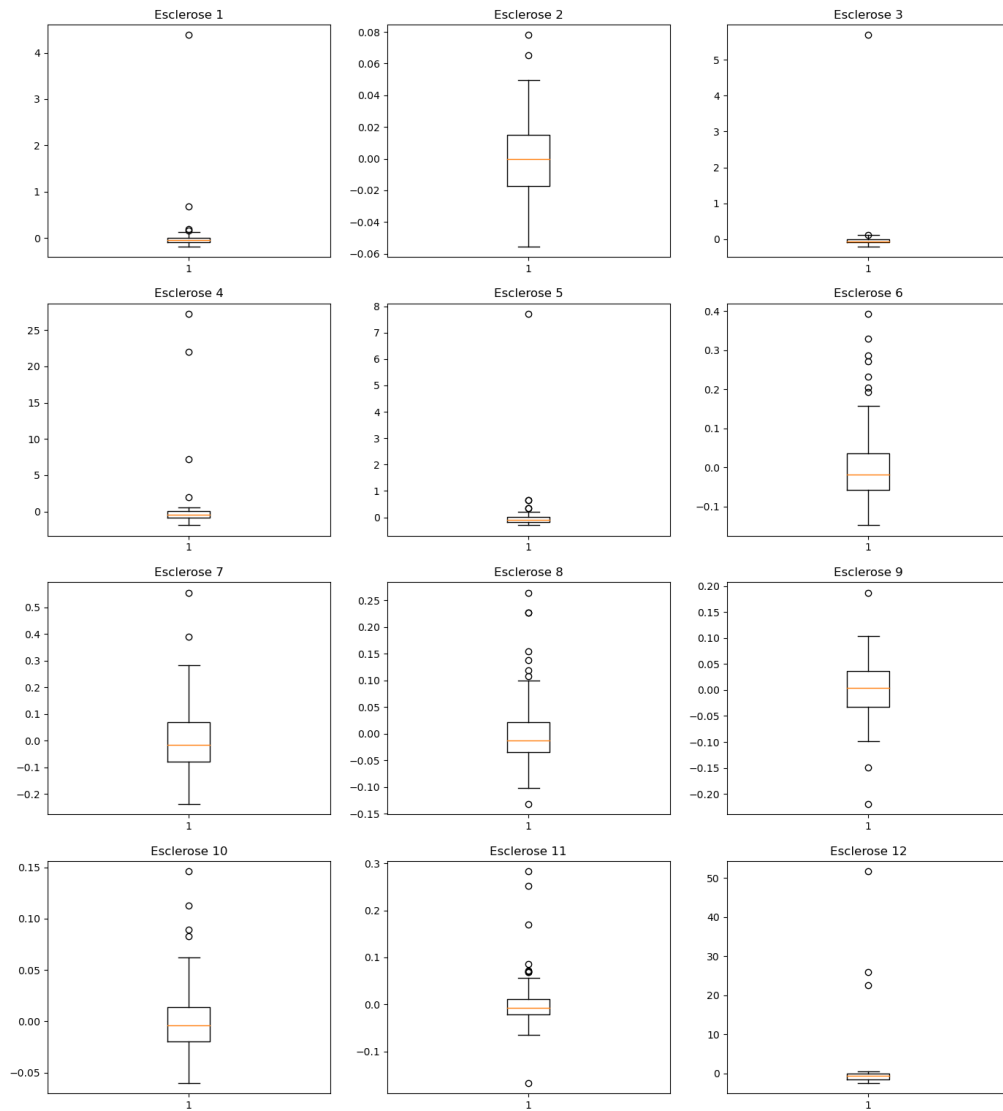


Figura D.10: Séries Temporais de ciclos de passada por segundo de 12 indivíduos com doença de Esclerose



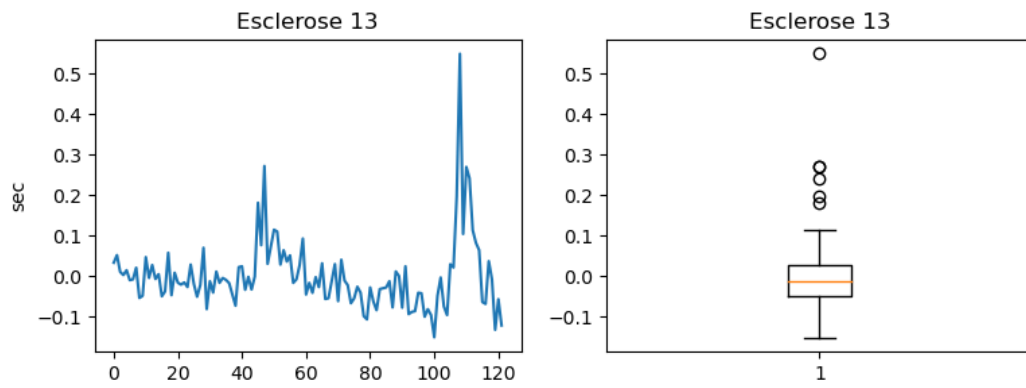
Fonte: Elaborado pelo autor.

Figura D.11: Boxplot das Séries Temporais de ciclos de passada por segundo de 12 indivíduos com doença de Esclerose



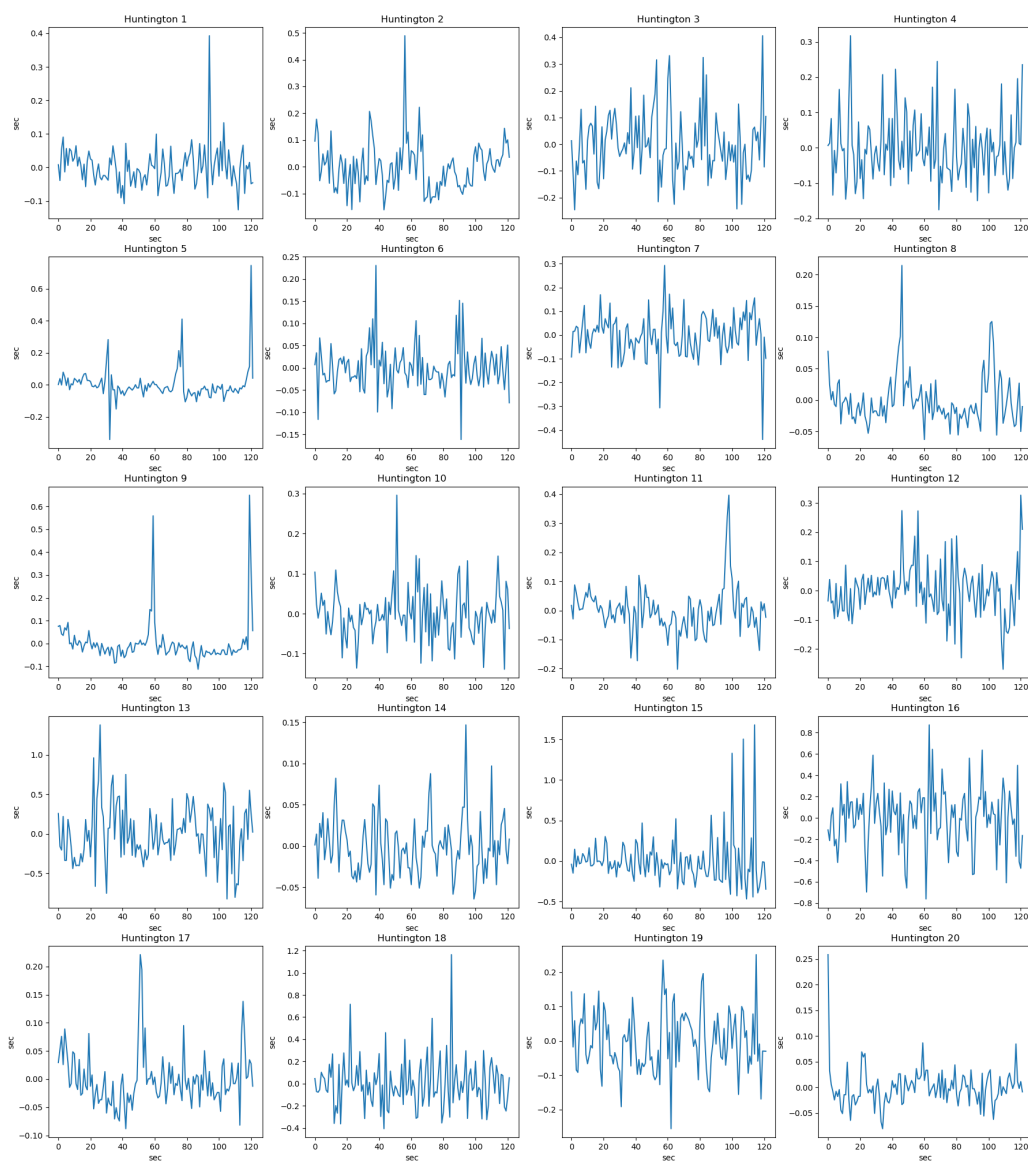
Fonte: Elaborado pelo autor.

Figura D.12: Série Temporal e Boxplot de ciclos de passada por segundo de 1 indivíduo com doença de Esclerose



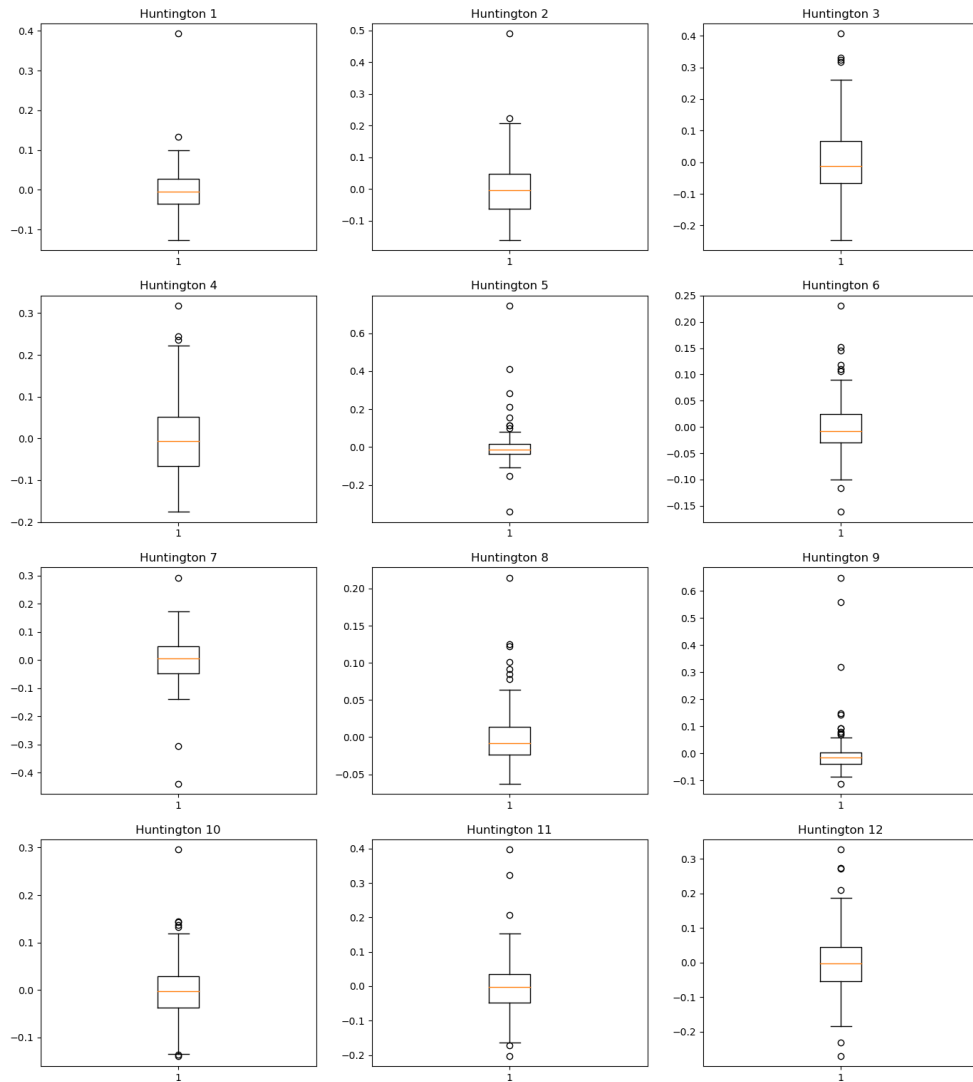
Fonte: Elaborado pelo autor.

Figura D.13: Séries Temporais de ciclos de passada por segundo de 20 indivíduos com doença de Huntington



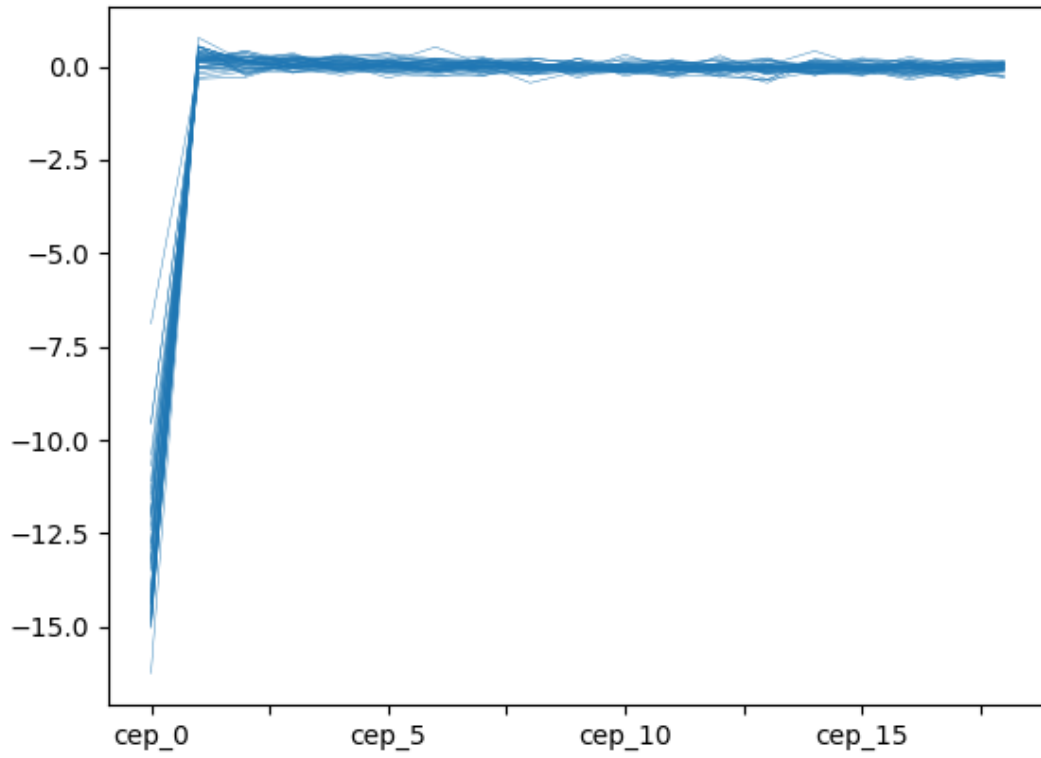
Fonte: Elaborado pelo autor.

Figura D.14: Boxplot das Séries Temporais de ciclos de passada por segundo de 20 indivíduos com doença de Huntington



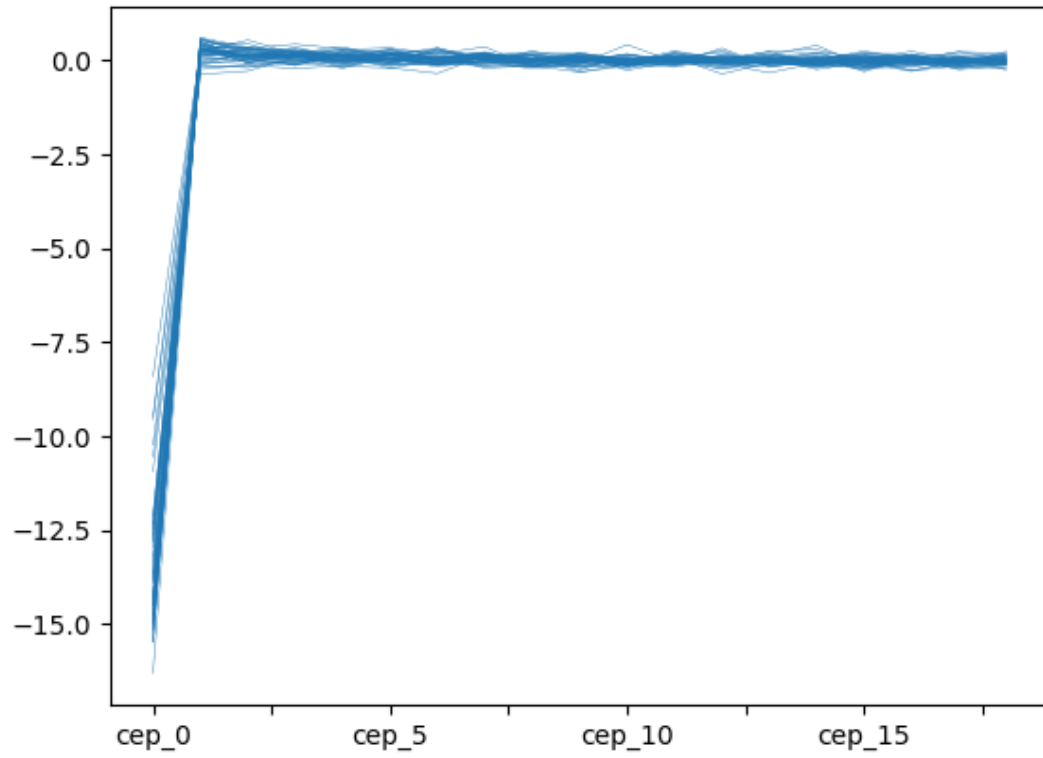
Fonte: Elaborado pelo autor.

Figura D.15: Cepstral dados modificados



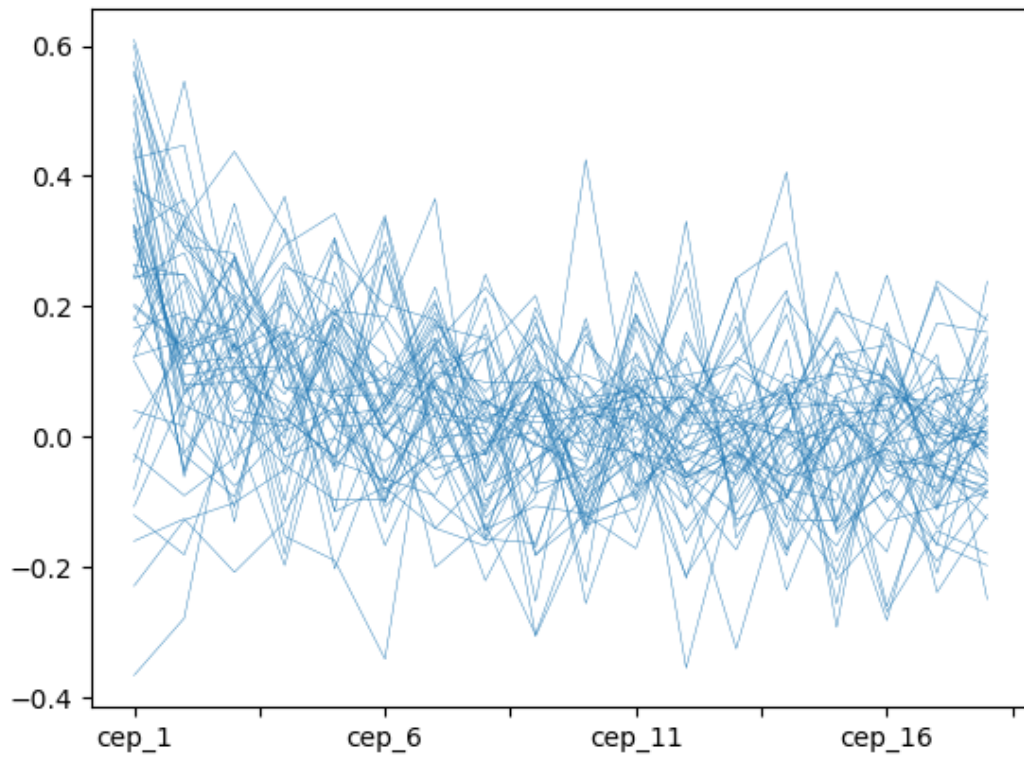
Fonte: Elaborado pelo autor.

Figura D.16: Cepstral dados não modificados



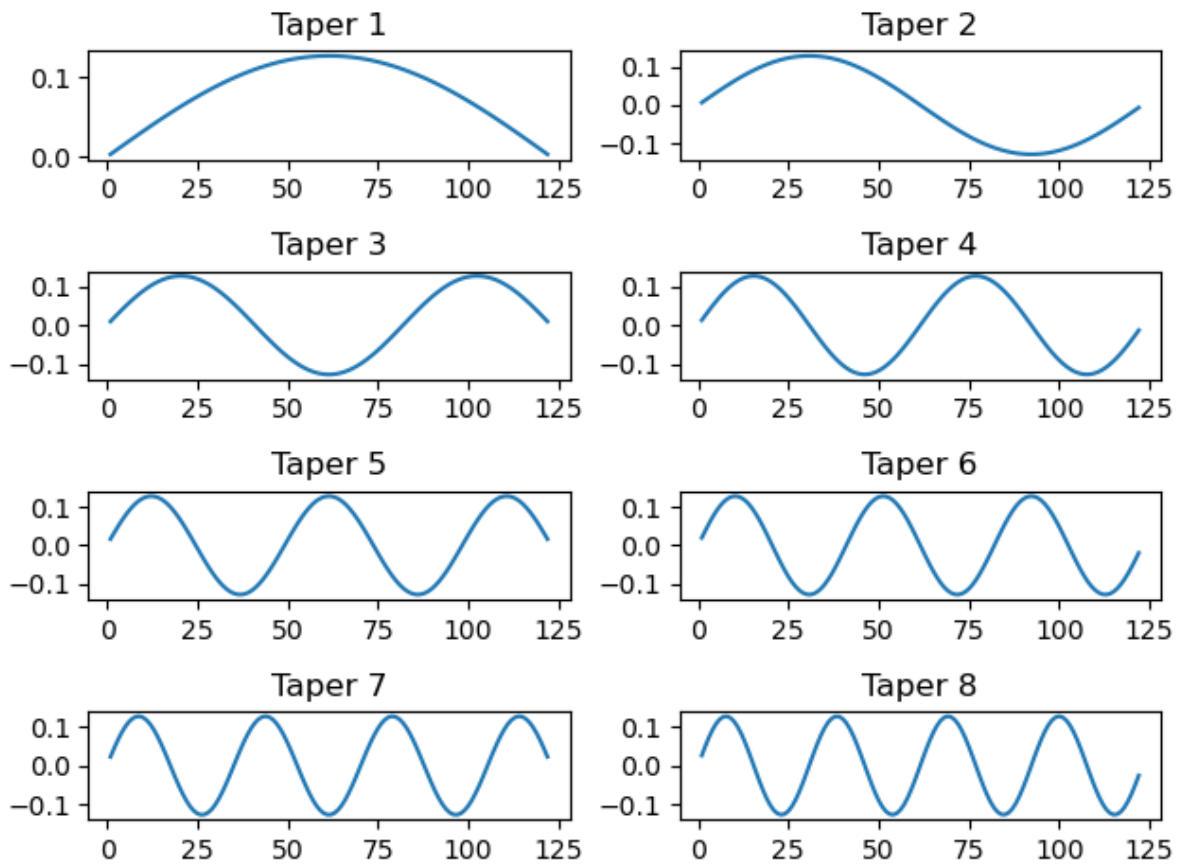
Fonte: Elaborado pelo autor.

Figura D.17: Cepstral dados não modificados exceto cep\_0



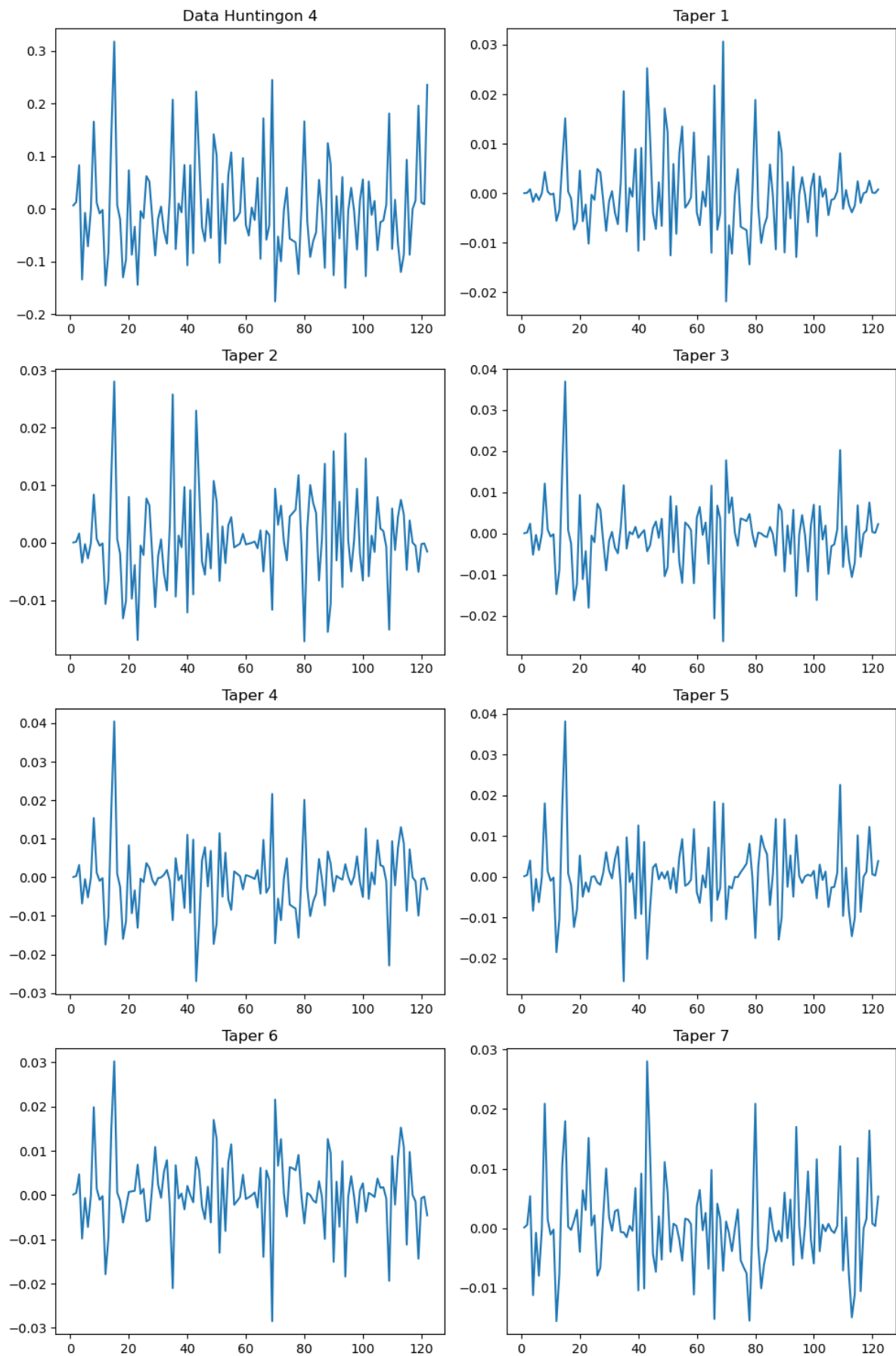
Fonte: Elaborado pelo autor.

Figura D.18: Funções dos 7 primeiros funis (tapers)



Fonte: Elaborado pelo autor.

Figura D.19: Dados do Paciente Huntingtons 4 e dados afunilados pelos 7 funis (tapers)



Fonte: Elaborado pelo autor.

# Apêndice E

## Tabelas Adicionais

Tabela E.1: Séries Temporais Simuladas  $AR(2)$ 

x1.1	x1.2	x1.3	x2.1	x2.2	x2.3	x3.1	x3.2	x3.3
1.284535	-0.489888	0.061092	0.878726	0.682875	-0.992572	-0.094215	0.406791	-0.303113
2.890179	2.040215	-2.262732	0.682748	1.618575	-0.327177	0.147001	1.766487	-0.428533
-0.709959	1.823572	0.297549	-0.810833	-0.706065	-1.198293	-1.069093	0.939732	-0.734861
-1.020753	1.721583	2.020002	-0.247929	-1.797392	-3.436155	-1.091637	-1.633075	1.183729
0.123715	2.203600	2.024877	-0.493071	-3.659337	-6.233442	-1.974667	-0.073450	1.464495
-0.964705	-1.551678	1.669049	-2.776125	-0.755178	-5.196307	-1.071355	1.248294	1.008848
0.650511	-1.487502	0.553862	0.673100	-0.784682	-2.733651	0.802048	-0.184022	-1.455063
-0.276814	-0.481883	-1.205400	1.498145	2.032537	-3.456000	0.028342	-0.598117	-3.268442
-1.197225	3.570000	-0.751665	0.592991	0.514059	-5.285018	0.087341	-0.039959	-1.010450
1.002555	0.508653	0.458078	-1.689306	-0.921678	-5.823183	-1.086243	1.162741	0.768166
2.763250	1.044478	-0.766479	1.689998	-1.343159	-4.879042	-0.322273	1.589891	-0.135100
-0.533302	1.385001	-0.332033	-1.529355	-0.602857	-1.754773	0.008354	-0.169192	-0.290268
2.811227	-0.855700	0.781337	-2.006596	-0.140465	-3.698724	0.788601	-2.483932	-0.097657
2.323982	-1.704478	-0.527899	-1.044143	0.520522	-1.538905	-0.692278	-2.736908	2.026427
-0.002349	3.048189	-0.061656	0.603081	-1.343082	0.683729	-2.643863	0.885359	1.903463
0.516615	1.237739	-0.827607	3.521896	1.422519	3.403306	-1.532381	1.401748	0.144157
1.613718	-3.190071	-1.404107	-0.021410	0.169974	4.896215	-0.579136	0.702743	-1.901372
-0.924165	-0.898937	0.959265	0.519827	-0.962453	4.984422	-1.388027	0.057399	-0.937047
-0.451179	-1.880320	1.109013	2.875234	0.298458	1.542777	-0.628886	1.598621	1.025749
0.193016	-2.109563	0.675310	2.776192	-0.503263	3.087564	1.299146	1.782251	1.407946
-0.270929	0.97426	2.470502	-0.658176	0.989892	1.871162	0.403409	1.328306	0.350234
0.322266	-3.712282	-1.73031	2.262624	-2.100782	1.923111	-0.375351	-2.178487	1.059102
1.931453	-1.860992	-2.04832	2.925534	0.789180	2.160729	-1.177869	1.254437	1.281648
2.412552	1.193858	-1.642976	0.087161	3.035311	1.905044	-0.217767	3.093250	-0.308900
0.643478	-0.220835	-0.281307	-0.792874	1.828364	0.330306	0.617904	0.549668	-0.040409
-0.706388	-0.767645	-1.060288	0.831577	-0.703866	0.432296	0.747409	-0.070700	1.628908
-0.229901	-3.727406	-1.546813	-1.060593	-2.506255	3.133138	-0.670001	1.586659	2.265335
0.194875	-0.170886	-0.695507	0.208428	-0.096791	4.539955	0.958957	1.111321	-1.318211
-2.950995	3.043341	1.016412	-2.462872	-0.523058	2.740494	1.009255	-0.134793	-2.999032
1.702959	0.853280	1.748255	2.910104	0.880905	1.595226	1.898229	0.524143	-0.316486
2.248967	-1.014717	0.588968	-1.011311	-1.509594	2.807992	1.350052	0.335671	2.696059
2.398054	2.880473	-0.549511	-0.957918	-0.736242	0.790529	-1.250514	-0.584180	0.633379
2.089417	-2.534844	0.448780	-3.093429	1.092916	2.274721	-3.110388	0.261701	-1.351457
0.653898	1.783244	1.253108	-3.279920	1.622837	1.070943	-1.427890	0.808948	-0.572129
0.880325	-0.455614	2.977114	-0.602828	-0.684308	-0.544552	1.382172	-0.570720	0.862652
0.270124	2.260949	3.205491	0.815639	-1.353532	-0.173106	2.566084	-1.669472	1.257132
0.728868	5.355701	1.414283	-0.610127	-1.044569	0.098962	0.449077	-2.648130	1.904144
-0.761521	-0.352086	2.153786	2.282626	2.292910	-0.305965	-3.595353	-0.241336	-0.137044
1.005675	-1.824642	0.435238	4.014972	2.939065	-0.724517	-3.472035	1.332056	-1.197377
3.100997	-1.292042	2.407230	0.441129	0.369050	0.432550	-0.402021	1.244804	0.776476
-0.071239	3.067484	0.853885	-3.729977	-0.088718	-0.891441	2.331778	0.833198	1.451362
0.348488	2.437268	0.091329	1.399467	-0.038488	-2.181409	-0.333238	1.784050	3.933654
0.172833	-0.132329	-0.835546	1.593558	2.618004	-1.804576	-0.041130	-1.211377	-0.997930
0.531595	1.285549	-0.965631	1.556370	-1.158410	-3.144047	-2.161323	-0.098333	-0.825333
2.464099	0.204549	-0.754310	-0.860318	-0.847069	-3.258229	-2.067965	1.747525	-0.403817
-0.363285	0.861574	0.603409	-0.719754	2.509429	-2.717097	0.957355	0.688660	-0.095618
0.639327	-2.480759	0.434655	0.467152	-2.551173	0.746655	3.741811	0.809266	0.847240
-0.654231	2.424461	-0.683395	2.802890	0.385338	3.012097	3.789096	0.794617	1.699763
-2.048442	-1.275455	1.653851	-0.540848	-0.863644	0.765774	0.786051	0.822892	-0.258439
-1.241555	1.464352	0.483865	-1.248509	-0.114870	-1.919349	-2.611811	-1.442184	-2.172303
-1.625736	-0.303012	2.367879	-0.099609	-1.184334	-3.388202	-1.670411	-2.423825	-0.383085
2.206010	-1.750872	1.305300	-3.241338	-0.604451	-2.441819	1.060699	-0.720370	1.568289
1.983168	-0.950921	0.294960	0.464027	-0.021288	-2.037700	2.425205	0.174830	2.781288
-0.031316	2.246569	-0.680336	-0.541129	1.290407	-1.363680	1.325539	1.146496	1.792024
1.048210	1.397192	0.315354	-0.271852	0.308645	0.639577	0.434705	-1.047181	-2.006823
-0.037704	1.973770	0.629858	1.502725	3.153744	2.280611	0.165831	-1.196362	-1.342646
1.257833	0.857288	1.208821	1.198238	3.050495	0.765529	1.126049	-0.837523	1.335706
-0.089759	2.274186	-1.317651	-3.338671	2.343801	0.521093	0.627741	-0.627740	3.560862
-0.710162	4.828491	-0.345763	0.965615	1.759724	-2.858107	-0.167677	-0.037555	2.300370
-0.155865	1.038462	2.200731	2.358511	-0.665350	-2.623654	-1.349983	0.299321	-1.096367
-0.939728	2.448396	0.762859	0.232512	-1.624617	-2.080558	-0.967699	0.105152	-3.404381
0.365883	-2.911472	1.865498	-2.165099	-2.365319	-0.446226	0.956695	0.174545	-2.145744
0.085022	-1.201069	0.917170	1.056330	-1.223811	1.191543	0.897565	0.429797	-0.083763
1.276391	-0.020541	1.326578	2.284644	1.449305	1.411882	0.015104	0.184122	0.881274
-0.110223	0.447875	2.085602	0.513287	0.108663	0.702795	-1.164040	-0.373124	-0.318734
1.910515	-0.953909	1.795776	0.368246	-0.642008	-0.131593	0.363827	-1.376421	-1.533281
1.279281	0.818567	0.598257	-1.002412	0.570973	-0.656853	0.525081	-0.854916	-2.042950
2.259836	4.817877	1.253123	0.526208	3.286851	-0.29458	1.252091	-0.036880	-0.098676
0.850984	0.927655	-1.217337	1.792813	2.344648	-1.804413	1.238539	-1.333434	-1.193978
-1.229149	-0.179490	-0.006512	0.863259	-0.568025	-2.816809	-0.491497	-0.681778	1.554898
-1.097207	1.120425	0.390992	-1.184072	-2.058006	-2.85247	-2.851016	-1.316286	0.402385
0.990916	-1.479227	0.831311	-3.400846	0.998548	-4.536295	-1.113017	-1.508168	-1.088548
0.537906	-2.473259	0.760352	0.069599	0.483346	-2.534652	-1.348295	-0.698643	-1.053212
2.563151	0.836314	-1.157898	-2.469871	0.597077	-1.069476	0.989943	0.181623	-0.620454
-0.929008	0.706269	-1.054712	-0.380291	0.079573	-1.440840	3.673723	0.523540	-0.010754
-1.967640	0.454156	0.214295	0.682812	0.027195	-1.667383	1.578958	-0.732722	0.665778
-0.293418	0.314314	0.144712	1.219161	0.085913	-2.455332	-0.785927	-2.991679	1.764777
1.615589	-1.104922	-0.975620	0.943144	-1.592240	-1.879800	-2.453418	-3.037382	0.613896
1.003894	0.851106	-1.631878	-1.277248	-1.431212	-2.060700	-1.478879	-0.774449	0.104150
0.778714	-2.774109	-3.063860	-0.543504	1.892768	-4.739986	0.760711	1.953574	-0.381966
-0.810653	-2.772291	-3.760046	-0.085751	-0.733134	-0.686550	0.970675	2.166296	0.055071
-0.522303	1.847904	-2.397823	-0.103623	0.342243	-3.919704	1.814882	-0.261780	0.726284
1.055884	0.563083	-1.977381	1.092617	-0.250018	-1.438670	0.894535	-1.768418	0.773491
-2.275332	-1.039457	-0.913469	1.265540	0.452643	0.494995	-1.492762	0.766396	1.585889
-2.096399	-1.699665	0.129702	0.233456	2.012897	1.032245	-3.344690	-0.010509	-0.288333
-2.406400	-2.023926	1.190425	-1.869066	2.301685	1.166209	-2.456968	0.053643	-0.142040
1.353205	0.403200	-0.629308	1.398630	-1.039771	2.349852	0.717709	1.214639	0.507010
3.138925	-1.176252	-0.894511	2.477344	-0.017027	-0.272763	2.145344	0.388329	0.461308
-0.542642	-0.541262	-2.114859	1.067690	-0.562443	-1.893818	1.208957	-0.828456	0.607233
0.406510	-2.252640	-0.882441	-1.405422	-1.197800	-1.373860	-1.213555	-0.878660	1.568593
-0.655326	-0.191070	-0.543717	0.007071	-1.515169	0.900655	-0.655485	0.296574	0.286803
0.934987	-3.366979	0.121112	-2.430695	-0.314053	-0.072127	1.746704	2.155145	-1.471768
0.759178	0.861998	0.253806	-0.641777	-1.167921	-1.941274	1.820495	1.868378	-1.109865
0.913892	1.243631	0.379176	1.774217	0.532989	-2.642195	-1.342730	1.069152	0.203409
2.083466	1.531179	2.904543	-1.222959	1.946541	-0.537148	-1.460348	-0.477229	0.464075
0.967028	2.024515	-0.312465	0.331754	0.862384	2.119855	-3.192690	-2.430390	-0.407940
-0.883705	-1.795311	-0.697936	-0.918823	0.811925	1.173888	-0.347110	-1.370140	-1.350995

Fonte: Elaborado pelo autor.

Tabela E.2: Réplica, População e População Predita - doenças neurodegenerativas com dados alterados por **Krafty (2016)**

Réplica	População*	População Predita*
0	1	1
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	<b>2</b>	<b>3</b>
17	<b>2</b>	<b>3</b>
18	<b>2</b>	3
19	<b>2</b>	3
20	2	2
21	2	2
22	2	2
23	2	3
24	<b>2</b>	<b>1</b>
25	<b>2</b>	<b>1</b>
26	<b>2</b>	<b>1</b>
27	2	2
28	2	2
29	3	3
30	3	3
31	3	3
32	3	3
33	3	3
34	3	3
35	3	3
36	<b>3</b>	<b>1</b>
37	<b>3</b>	<b>1</b>
38	3	3
39	<b>3</b>	<b>2</b>
40	3	3
41	<b>3</b>	<b>2</b>
42	<b>3</b>	<b>1</b>
43	3	3
44	3	3
45	<b>3</b>	<b>1</b>
46	3	3
47	3	3
48	<b>3</b>	<b>1</b>

\* 1: controle; 2: Esclerose; 3: Huntingtons

Fonte: Elaborado pelo autor.

Tabela E.3: Réplica, População e População Predita **robusta** - doenças neurodegenerativas

Réplica	População*	População Predita*
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	2	2
<b>18</b>	<b>2</b>	<b>3</b>
19	2	2
20	2	2
21	2	2
22	2	2
23	2	2
24	2	2
25	2	2
<b>26</b>	<b>2</b>	<b>1</b>
<b>27</b>	<b>2</b>	<b>1</b>
28	2	2
29	2	2
30	3	3
31	3	3
32	3	3
33	3	3
34	3	2
35	3	3
36	3	3
<b>37</b>	<b>3</b>	<b>1</b>
<b>38</b>	<b>3</b>	<b>1</b>
39	3	3
<b>40</b>	<b>3</b>	<b>2</b>
41	3	3
42	3	3
43	3	3
44	3	3
45	3	3
<b>46</b>	<b>3</b>	<b>1</b>
47	3	3
48	3	3
49	3	3

\* 1: controle; 2: Esclerose; 3: Huntingtons

Fonte: Elaborado pelo autor.

Tabela E.4: Matriz Confusion para as 49 séries de indivíduos com doenças neurodegenerativas

<b>População Predita</b>	<b>População</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>1</b>		<b>16</b>	3	5
<b>2</b>		0	<b>4</b>	2
<b>3</b>		0	6	<b>13</b>

Fonte: Elaborado pelo autor.

Tabela E.5: Matriz Confusion **robusta** para as 49 séries de indivíduos com doenças neurodegenerativas

<b>População Predita</b>	<b>População</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>1</b>		<b>16</b>	2	3
<b>2</b>		0	<b>10</b>	2
<b>3</b>		0	1	<b>15</b>

Fonte: Elaborado pelo autor.

# Apêndice F

## Apêndice

### F.1 Demonstração da fórmula do Periodograma

Conforme sugerido por (Persival e Walden, 1993, p.196) e (Priestley, 1981, 398), substituindo o estimador das auto-covariâncias da equação (??) na equação (2.10) com a alteração que é um estimador do espectro da k-ésima réplica para a j-ésima população, tem-se:

$$\begin{aligned}
 \hat{S}_X^{(jk)}(\lambda_m) &= \frac{1}{2\pi} \sum_{\nu=-(N-1)}^{N-1} \hat{\gamma}_\nu^{(jk)} e^{-i\lambda_m \nu} \\
 &= \frac{1}{2\pi} \left\{ \hat{\gamma}_0^{(jk)} + 2 \sum_{\nu=1}^N \hat{\gamma}_\nu^{(jk)} e^{-i\lambda_m \nu} \right\} \\
 &= \frac{1}{2\pi} \left\{ \hat{\gamma}_0^{(jk)} + 2 \sum_{\nu=1}^N \left[ \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} \right] e^{-i\lambda_m \nu} \right\} \\
 &= \frac{1}{2\pi} \left\{ \hat{\gamma}_0^{(jk)} + 2 \sum_{\nu=1}^N \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} e^{-i\lambda_m \nu} \right\} \\
 &= \frac{1}{2\pi N} \sum_{\nu=-(N-1)}^{N-1} \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} e^{-i\lambda_m \nu} \\
 &= \frac{1}{2\pi N} \sum_{\nu=-(N-1)}^{N-1} \sum_{t=1}^{N-|\nu|} X_{jkt} X_{jk(t-|\nu|)} e^{-i\lambda_m (t-\tau)}
 \end{aligned}$$

## F.2 Formato Matricial do cálculo do Periodograma

A equação (2.13) pode ser reescrita como:

$$X_{jkt} = \begin{bmatrix} s_{jk}(\lambda_1) & s_j(\lambda_2) & \dots & s_{jk}(\lambda_{N/2}) \end{bmatrix} \begin{bmatrix} \beta_{jk}(\lambda_1) \\ \beta_{jk}(\lambda_2) \\ \vdots \\ \beta_{jk}(\lambda_{N/2}) \end{bmatrix} + \xi_{jkt} \quad (\text{C.2.2})$$

em que

$$s_{jk}(\lambda_m) = \begin{bmatrix} \cos_{jk}(\lambda_m) & \sin_{jk}(\lambda_m) \\ \cos_{jk}(2\lambda_m) & \sin_{jk}(2\lambda_m) \\ \vdots & \vdots \\ \cos_{jk}(N\lambda_m) & \sin_{jk}(N\lambda_m) \end{bmatrix}_{(N \times 2)}$$

## F.3 Cepstra Matrix for $J$ populations and $L$ cepstra

$$(C, \Pi) = \begin{bmatrix} \begin{bmatrix} c_{110} & c_{111} & \dots & c_{11(L-1)} & 1 \\ c_{120} & c_{121} & \dots & c_{12(L-1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{1n_1 0} & c_{1n_1 1} & \dots & c_{1n_1(L-1)} & 1 \end{bmatrix} \\ \begin{bmatrix} c_{210} & c_{211} & \dots & c_{21(L-1)} & 2 \\ c_{220} & c_{221} & \dots & c_{22(L-1)} & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{2n_2 0} & c_{2n_2 1} & \dots & c_{2n_2(L-1)} & 2 \end{bmatrix} \\ \vdots \\ \begin{bmatrix} c_{J10} & c_{J11} & \dots & c_{J1(L-1)} & J \\ c_{J20} & c_{J21} & \dots & c_{J2(L-1)} & J \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{Jn_J 0} & c_{Jn_J 1} & \dots & c_{Jn_J(L-1)} & J \end{bmatrix} \end{bmatrix}$$

## F.4 Matriz de Variância e Covariância entre as populações

$$E \begin{bmatrix} \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)^2 & \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) & \cdots & \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k(L-1)} - \mu_{(L-1)}) \\ \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) & \sum_{k=1}^{n_1} (c_{1k1} - \mu_1)^2 & \cdots & \sum_{k=1}^{n_1} (c_{1k1} - \mu_1)(c_{1k(L-1)} - \mu_{(L-1)}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k(L-1)} - \mu_{(L-1)}) & \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) & \cdots & \sum_{k=1}^{n_1} (c_{1k(L-1)} - \mu_{(L-1)})^2 \end{bmatrix} \quad (\text{F.1})$$

$$\begin{bmatrix} \sum_{j=1}^J \sum_{k=1}^{n_j} (c_{1k0} - \mu_0)^2 f_j & \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) f_j & \cdots & \sum_{j=1}^J \sum_{k=1}^{n_j} (c_{1k0} - \mu_0)(c_{1k(L-1)} - \mu_{(L-1)}) f_j \\ \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) f_j & \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k1} - \mu_1)^2 f_j & \cdots & \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k1} - \mu_1)(c_{1k(L-1)} - \mu_{(L-1)}) f_j \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^J \sum_{k=1}^{n_j} (c_{1k0} - \mu_0)(c_{1k(L-1)} - \mu_{(L-1)}) f_j & \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k0} - \mu_0)(c_{1k1} - \mu_1) f_j & \cdots & \sum_{j=1}^J \sum_{k=1}^{n_1} (c_{1k(L-1)} - \mu_{(L-1)})^2 f_j \end{bmatrix} \quad (\text{F.2})$$

$$E \begin{bmatrix} \sigma_{00}^2 & \sigma_{01}^2 & \cdots & \sigma_{0(L-1)}^2 \\ \sigma_{10}^2 & \sigma_{11}^2 & \cdots & \sigma_{1(L-1)}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{(L-1)0}^2 & \sigma_{(L-1)1}^2 & \cdots & \sigma_{(L-1)(L-1)}^2 \end{bmatrix} \quad (\text{F.3})$$

where  $\sigma_{km}^2 = \langle c_{jkl} - \mu_j, c_{jkm} - \mu_j \rangle \langle l_n, l_n \rangle^{-1} \langle l_{n_j}, l_{n_j} \rangle$  and  $\ell, m = 1, \dots, L-1$ .