



TÉCNICAS MULTIVARIADAS DE REDUÇÃO DE DADOS:
COMPARAÇÃO, APLICABILIDADE E CONVERGÊNCIA DE RESULTADOS

FERNANDO JOSÉ DO NASCIMENTO SEBASTIÃO

ORIENTADORA:

Professora Doutora Irene Cristina Salgueiro de Oliveira

COORIENTADOR:

Professor Doutor Jorge Filipe Campinos Landerset Cadima

Professor Associado

Instituto Superior de Agronomia

Universidade de Lisboa

TESE ELABORADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM
MATEMÁTICA E ESTATÍSTICA

JÚRI:

Presidente: Doutor António José Guerreiro de Brito

Professor Associado com agregação

Instituto Superior de Agronomia

Universidade de Lisboa.

Vogais: Doutor António Gabriel da Silva St.Aubyn

Professor Catedrático Emérito
Instituto Superior de Agronomia
Universidade de Lisboa;

Doutora Fernanda Maria Campos de Sousa
Professora Auxiliar
Faculdade de Engenharia
Universidade do Porto;

Doutor Manuel Joaquim Piteira Minhoto
Professor Auxiliar
Escola de Ciências e Tecnologia
Universidade de Évora;

Doutora Irene Cristina Salgueiro de Oliveira
Professora Auxiliar
Escola de Ciências e Tecnologia
Universidade de Trás-os-Montes e Alto Douro;

Doutora Marta Guerreiro Duarte Mesquita de Oliveira
Professora Auxiliar
Instituto Superior de Agronomia
Universidade de Lisboa.

2016

Resumo

É muito comum a análise de dados recorrendo a técnicas multivariadas, nomeadamente para séries temporais. As principais metodologias consagradas, descritas ao longo desta dissertação, são a Análise em Componentes Principais (ACP), a Análise em Componentes Independentes (ACI), a Análise Espectral Singular (SSA) e a Análise Espectral Singular Multicanal (MSSA). A ACI tem sido considerada mais adequada que a ACP na análise de séries temporais, principalmente quando se admite a não normalidade e se exige que as componentes a estimar sejam independentes. Neste sentido, como alternativa à MSSA, é proposta uma nova abordagem baseada na ACI aplicada à matriz dos desfasamentos, que é designada *Lag*-Análise em Componentes Independentes (*Lag*ACI). Tal abordagem é desenvolvida e suportada teoricamente pelos conceitos algébricos e algoritmos existentes das duas técnicas envolvidas. São dados exemplos de aplicação, com especial relevância para dados climáticos.

O principal objetivo da tese é interligar o novo método com os restantes métodos estatísticos multivariados referidos, para permitir avaliar as semelhanças e diferenças entre estes. Analisam-se diferentes abordagens em cada método e estudam-se as implicações da escolha da matriz informativa das estruturas subjacentes aos dados, tendo em conta se estes estão ou não sujeitos a algum tipo de pré-processamento, como por exemplo a centragem. São ainda considerados alguns indicadores comparativos de avaliação do desempenho de modelos para averiguar o grau de semelhança nos resultados transversais às técnicas em estudo.

Considera-se que a nova abordagem representa uma contribuição para novos desenvolvimentos teóricos e aplicações no campo da análise de séries temporais.

Palavras-chave: Análise em Componentes Independentes, Análise em Componentes Principais, Análise Espectral Singular Multicanal, *Lag*-Análise em Componentes Independentes, Séries Temporais.

Abstract

The use of multivariate techniques in data analysis is very common, in particular for time series. The main established methods described in this thesis are Principal Component Analysis (PCA), Independent Component Analysis (ICA), Singular Spectrum Analysis (SSA) and Multichannel Singular Spectrum Analysis (MSSA). ICA has been considered more appropriate than PCA in time series analysis, especially when non-normality is assumed and it is required that the components to be estimated are independent. In this sense, as an alternative to MSSA, a new approach is proposed based on ICA applied to the trajectory matrix, which has been called *Lag-Independent Component Analysis* (*LagICA*). This approach is theoretically developed and supported by algebraic concepts and existing algorithms in the two relevant techniques. Application examples are given, with special relevance for climate data.

The main goal is to connect the new method with the established multivariate statistical methods, and to assess their similarities and differences. Different approaches to each method are considered and the implications of the choice of matrix with information on the underlying data structures are explored, taking into account whether or not these are subject to some type of pre-processing, such as centring. In addition, some comparative measures assessing the models' performance are used, to compare results that are transversal to the techniques under study.

It is considered that the new approach can be of assistance in new theoretical developments and applications in the field of time series analysis.

Keywords: Independent Component Analysis, Principal Component Analysis, Multichannel Singular Spectrum Analysis, *Lag-Independent Component Analysis*, Time Series.

Agradecimentos

Uma vez que seria praticamente impossível construir esta dissertação sem o apoio de várias pessoas, expresso aqui os meus mais sinceros agradecimentos.

À minha Orientadora, a Prof. Irene Oliveira, pela transmissão de grande parte do seu conhecimento, pelo incansável acompanhamento do desenvolvimento da minha investigação ainda que por vezes por "terrenos pantanosos", e pelo enorme encorajamento que demonstrou para comigo. Sem estes fatores teria sido difícil chegar a esta etapa da minha formação académica e a este nível de evolução pessoal.

Ao meu Coorientador, o Prof. Jorge Cadima, por toda a atenção e paciência desde os tempos de professor do mestrado, pelo rigor que sempre incutiu aos seus alunos e pelas suas sugestões construtivas que permitiram introduzir melhorias significativas no decorrer do trabalho.

Ao Prof. João Santos da Universidade de Trás-os-Montes e Alto Douro, pelos aconselhamentos relativamente à escolha de dados para analisar no contexto da climatologia.

À Prof. Manuela Neves, como Coordenadora do Doutoramento e como professora, pelo especial entusiasmo, incentivo e carinho que sempre manifestou ao longo destes anos que estudei no Instituto Superior de Agronomia.

A todos os amigos, familiares e colegas que sempre me apoiaram e sempre acreditaram nas minhas capacidades.

Agradeço à Universidade de Trás-os-Montes e Alto Douro que me acolheu nos meus primeiros anos de Doutoramento assim como ao Centro de Matemática da UTAD, CM-UTAD.

Ao Instituto Superior de Agronomia da Universidade de Lisboa agradeço por ter reunido condições para que eu pudesse realizar o meu Doutoramento, ao aceitar o processo da minha transferência da UTAD.

Agradeço ainda ao Instituto Politécnico de Leiria pelo apoio logístico e financeiro no início do Doutoramento e em particular à Escola Superior de Tecnologia e Gestão pelas flexibilidades e reduções de distribuição do serviço docente em alguns semestres.

Agradeço a todos os investigadores que me permitiram elaborar esta tese, e por isso dedico-a a todos os jovens investigadores, em particular aos deste país que me viu crescer.

Índice

1	Introdução	1
1.1	Visão Geral	1
1.2	Enquadramento do Trabalho	2
1.3	Organização da Dissertação	3
1.4	Preliminares	6
1.4.1	Tópicos de Séries Temporais	6
1.4.2	Decomposição Espectral e Decomposição em Valores Singulares	8
1.4.3	Indicadores Comparativos do Desempenho de um Modelo	10
2	Análise em Componentes Principais	17
2.1	Introdução e Descrição da ACP	17
2.2	Obtenção das Componentes Principais	21
2.3	Algumas Propriedades na ACP	22
2.4	Escolha do Número de Componentes Principais	26
2.5	Representação Gráfica dos Indivíduos e das Variáveis	27
2.6	Breve História da ACP	28
2.7	Para Além da ACP Clássica - Outras Abordagens	30
2.7.1	Rotação das Componentes	31
2.7.2	ACP Ponderada	32
2.7.3	ACP Centrada <i>versus</i> ACP Não Centrada	34
2.7.4	ACP para Dados Categóricos (CATPCA)	36
2.7.5	ACP Log-contraste	38
2.7.6	ACP em Três Modos	39

2.7.7	ACP para Séries Temporais	39
3	Análise em Componentes Independentes	43
3.1	Introdução	43
3.2	Breve História da ACI	44
3.3	Motivação	48
3.4	Descrição do Modelo	52
3.4.1	Propriedades Básicas do Modelo	54
3.4.2	Centragem e Branqueamento	55
3.5	Métodos de Estimação	59
3.5.1	Curtose como Medida de Não Normalidade	60
3.5.2	Neguentropia como Medida de Não Normalidade	65
3.5.3	Estimação de Máxima Verossimilhança	68
3.5.4	Princípio do <i>Infomax</i>	69
3.5.5	Minimização da Informação Mútua	70
3.5.6	Métodos Tensoriais	73
3.6	Métodos que Usam uma Estrutura Temporal	75
3.6.1	Separação por Autocovariâncias	76
3.6.2	Separação por Não Estacionariedade de Variâncias	79
3.6.3	Princípios de Separação Unificados	79
3.7	Algoritmos para Implementar a ACI	80
3.7.1	AMUSE	81
3.7.2	FastICA	81
3.7.3	FOBI	83
3.7.4	JADE	84
3.7.5	A Escolha do Algoritmo	85
3.8	Métodos de Ordenação das CIs	85
3.9	Número de CIs a Reter	89
3.10	Outros Modelos de ACI	90
3.10.1	ACI com Ruído	90
3.10.2	ACI com Número de Misturas Diferente do Número de CIs	92

3.10.3	ACI Não Linear	94
3.10.4	ACI com Restrição	95
3.10.5	ACI nos Complexos	96
3.10.6	ACI para Dados Binários	97
3.11	Áreas de Aplicações da ACI	98
3.12	Exemplo de Aplicação: Pressão ao Nível do Mar	104
3.12.1	Estudo no Domínio do Tempo	109
3.12.2	Estudo no Domínio da Frequência	114
3.12.3	Reconstrução	121
3.13	Resumo e Conclusões	122
4	Análise Espectral Singular	129
4.1	Introdução	129
4.2	Descrição da SSA	133
4.2.1	Decomposição	133
4.2.2	Reconstrução	135
4.3	Comprimento da Janela	137
4.4	Previsão na SSA	140
4.4.1	Algoritmo de Previsão Recorrente em SSA	141
4.4.2	Intervalos de Predição	143
4.5	Matriz de Toeplitz na Decomposição	144
4.6	Pares de Elementos na Decomposição	146
4.7	Séries Reconstruídas/Filtradas	152
4.8	Centragem em SSA	154
4.8.1	Centragem Simples por Colunas	154
4.8.2	Centragem Simples por Linhas	157
4.8.3	Centragem Dupla	159
4.8.4	Centragem na Série Original	160
4.9	<i>Softwares e Packages</i>	162
4.10	Exemplo: Vítimas Mortais em Acidentes de Viação	163
4.10.1	SSA Básica (Sem Centragem)	164

4.10.2	Comparações entre SSA Sem Centragem e SSA Com Centragem	178
4.11	Resumo e Conclusões	188
5	Análise Espectral Singular Multicanal	191
5.1	Introdução	191
5.2	Descrição da MSSA	192
5.2.1	Decomposição	192
5.2.2	Reconstrução	196
5.2.3	Comprimento da Janela	198
5.2.4	Algumas Propriedades	198
5.3	Estado da Arte	199
6	LagACI: Lag-Análise em Componentes Independentes	205
6.1	Introdução	205
6.2	Fundamentação Teórica	206
6.3	Exemplo de Aplicação: Pressão ao Nível do Mar	209
6.3.1	Dados e Metodologia	209
6.3.2	Análise dos Resultados	212
6.4	Resumo e Conclusões	220
7	Discussão e Conclusões	223
7.1	Introdução	223
7.2	Exemplo de Aplicação: Altura Geopotencial	226
7.2.1	Dados e Metodologia	226
7.2.2	Análise dos Resultados	229
7.2.3	Resumo	243
7.3	Conclusões Finais	244
7.4	Trabalho Futuro	249
A	Notação	251
A.1	Glossário dos Principais Símbolos para ACP	251
A.2	Glossário dos Principais Símbolos para ACI	252

A.3	Glossário dos Principais Símbolos para SSA e MSSA	255
B	Lista de Acrónimos	261
C	Apontamentos Teóricos Complementares	265
C.1	Tópicos Elementares de Vetores e Matrizes	265
C.1.1	Vetores	265
C.1.2	Matrizes	268
C.1.3	Operações com Matrizes	270
C.1.4	Valores e Vetores Próprios	272
C.2	Análise de Fourier e Transformada de Fourier	273
C.3	Análise Espectral	274
C.4	Momentos, Cumulantes e suas Propriedades	277
C.4.1	Caso Univariado	277
C.4.2	Caso Multivariado	278
C.4.3	Propriedades	281
D	Rotinas e Conjuntos de Dados	283
D.1	Rotinas e <i>Packages</i> Usados na Análise dos Dados	283
D.2	Dados da Pressão ao Nível do Mar	301
D.2.1	Resultados	302
D.3	Dados do Número de Vítimas Mortais em Acidentes de Viação	302
D.3.1	Resultados	302
D.4	Dados da Altura Geopotencial	308
D.4.1	Resultados	308
	Bibliografia	310

Lista de Figuras

3.1	Exemplo do problema <i>cocktail party</i> : três pessoas a conversar numa sala com cinco microfones a captar as conversas em diferentes localizações	49
3.2	Exemplo de três sinais (CIs) originais	51
3.3	Exemplo de três misturas correspondentes aos sinais da Figura 3.2	51
3.4	Exemplo das estimativas das três componentes independentes (ou seja, dos sinais originais) com base nas misturas dos sinais ilustradas na Figura 3.3	52
3.5	Esquema do processo de estimação do modelo linear de ACI básico	53
3.6	Esquema de blocos ilustrando o sistema de mistura das CIs e de separação das misturas ao longo do processo da ACI	55
3.7	Funções densidade de probabilidade da distribuição de Laplace, da distribuição Normal (a ponteados) e da distribuição Uniforme (a negrito) com médias nulas e variâncias unitárias	61
3.8	Cenário de otimização para a curtose: as linhas finas representam algumas curvas de nível da função descrita em (3.10) e a linha a grosso representa o círculo unitário	63
3.9	Séries temporais dos valores mensais médios de pressão ao nível do mar em 8 estações meteorológicas no Norte do Oceano Pacífico entre janeiro de 1979 e dezembro de 1996	105
3.10	Série temporal dos valores mensais de temperatura ao nível do mar (em °C) do Índice de <i>El Niño</i> entre janeiro de 1979 e dezembro de 1996	105
3.11	Coordenadas das três primeiras componentes principais da ACP sobre a matriz de correlações	106
3.12	Coordenadas das duas primeiras componentes independentes obtidas pelo algoritmo FastICA	106

3.13 Coordenadas das duas primeiras componentes independentes obtidas pelo algoritmo AMUSE	107
3.14 Representação das coordenadas das 216 observações nas duas primeiras componentes principais (CP1+CP2) e nas duas componentes independentes (CI1+CI2) obtidas através dos algoritmos FastICA e AMUSE	108
3.15 Coordenadas da CP1, da CI1-FastICA e da CI1-AMUSE normalizadas	110
3.16 Coordenadas da CP2, da CI2-FastICA e da CI2-AMUSE normalizadas	110
3.17 Coordenadas da CP3, da CI2-FastICA e da CI2-AMUSE normalizadas	111
3.18 Vetores próprios da ACP e vetores da matriz do branqueamento em ACI	112
3.19 Valores da função de autocorrelação para as CP1, CP2 e CP3. As linhas horizontais a tracejado representam os limites do intervalo a 95% de confiança para a autocorrelação	113
3.20 Valores da função de autocorrelação para as CI1 e CI2 obtidos através dos algoritmos FastICA e AMUSE. As linhas horizontais a tracejado representam os limites do intervalo a 95% de confiança para a autocorrelação	114
3.21 Espectros das coordenadas de CI1-FastICA, CI2-FastICA, CI1-AMUSE e CI2-AMUSE	115
3.22 Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA e CI2-FastICA	116
3.23 Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e da série temporal do Índice de <i>El Niño</i>	117
3.24 Espectros de fase para CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e série temporal do Índice de <i>El Niño</i>	118
3.25 Espectros de coerência para CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e série temporal do Índice de <i>El Niño</i>	118
3.26 Série temporal dos valores mensais standardizados de pressão ao nível do mar do Índice de Oscilação Sul entre janeiro de 1979 e dezembro de 1996	119
3.27 Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e da série temporal do Índice de Oscilação Sul	120
3.28 Esquema das etapas de aplicação do modelo ACI: fases do pré-processamento e da implementação de um algoritmo	125
3.29 Esquema ilustrativo de alguns métodos de ordenação das CIs	126

4.1	Nove pares de seqüências de senos e cossenos no círculo unitário	152
4.2	Número de vítimas mortais em acidentes de viação em Portugal Continental . . .	165
4.3	Logaritmos dos 48 valores próprios	166
4.4	Logaritmos dos 24 primeiros valores próprios exceto o primeiro	167
4.5	Componentes principais CP1 a CP8	167
4.6	Componentes principais CP9 a CP16	168
4.7	Pares de CPs harmónicas	169
4.8	Espectros dos sete pares de CPs	170
4.9	Cronogramas do número de vítimas mortais mensais entre os anos de 1998 e 2009	171
4.10	Evolução anual do número de vítimas mortais e medidas de segurança rodoviária	172
4.11	Reconstrução da série original com a CP1	173
4.12	Reconstruções da série original com as CPs 1-5, CPs 1-10 e CPs 1-16	174
4.13	Resíduos para a série original ao utilizar as reconstruções com as CPs 1-5, CPs 1-10 e CPs 1-16	175
4.14	Séries original e série prevista construída com base nas CPs 1-16; previsões dos 12 valores mensais para o ano de 2010	176
4.15	Séries original e série prevista construída com base nas CPs 1-16; previsões dos 12 valores mensais para o ano de 2010 e intervalos de predição a 95% obtidos através do método empírico	176
4.16	Comparação do número de vítimas mortais previstas <i>versus</i> observadas para os 12 meses do ano de 2010	177
4.17	Logaritmos dos 24 valores próprios para os 5 tipos de SSA em análise	179
4.18	16 CPs dominantes para os 5 tipos de SSA em análise	180
6.1	Metodologia para técnicas I, II e III com um comprimento da janela $m = 50$. \mathbf{Z} é a matriz das coordenadas das CPs, \mathbf{P} é a matriz dos vetores próprios, \mathbf{S} é a matriz das CIs e \mathbf{A} é a matriz dos coeficientes das misturas	210
6.2	Valores próprios para a ACP na matriz dos desfasamentos com $m = 50$	211
6.3	Coordenadas para $q = 5$ componentes obtidas pelas técnicas I, II e III, para $m = 50$	213
6.4	Vetores associados à ACP e à ACI a cada uma das três técnicas	216
6.5	Espectros das coordenadas das cinco CPs e CIs retidas para as técnicas I e II . .	217

6.6	Espectros das coordenadas das cinco CPs e CIs retidas para as técnicas I e III . . .	218
6.7	Espectros das coordenadas das cinco CIs retidas para cada uma das técnicas II e III	218
6.8	Coordenadas para $q = 5$ componentes obtidas pelas técnicas I, II e III, para $m = 100219$	
6.9	Valores das médias (das 8 estações meteorológicas) dos cinco indicadores para as técnicas I, II e III	219
7.1	Comparações das estruturas matriciais das cinco técnicas em estudo ao longo da dissertação: ACP, ACI, SSA e MSSA e <i>LagACI</i>	224
7.2	Setor do Hemisfério Norte entre 20°N - 80°N e 0°E - 357.5°E , onde foram recolhidos os dados da altura geopotencial	227
7.3	Primeiros 50 valores próprios (de um total de 600) obtidos através da ACP aplicada nos dados iniciais	228
7.4	Metodologia para MSSA e <i>LagACI</i> aplicada na matriz aumentada dos desfaseamentos, com um comprimento da janela $m = 180$, $n' = 600$ observações e $p' = 19$ CPs retidas. \mathbf{Z} é a matriz das coordenadas das CPs, \mathbf{P} é a matriz dos vetores próprios, \mathbf{S} é a matriz das CIs e \mathbf{A} é a matriz dos coeficientes das misturas . . .	229
7.5	Logaritmos dos primeiros 50 valores próprios (de um total de 421) obtidos através da ACP aplicada na matriz dos desfaseamentos dos dados	230
7.6	Coordenadas para $q = 20$ CPEs, com $m = 180$ e $p' = 19$	231
7.7	Coordenadas para $q = 20$ CIEs obtidas através do algoritmo FastICA, com $m = 180$ e $p' = 19$	232
7.8	Coordenadas para $q = 20$ CIEs obtidas através do algoritmo AMUSE, com $m = 180$ e $p' = 19$	232
7.9	Pares de CPEs harmónicas	233
7.10	Pares de CIEs harmónicas obtidas através do algoritmo AMUSE	234
7.11	Vetores associados a $q = 20$ CPEs, com $m = 180$ e $p' = 19$	235
7.12	Vetores associados a $q = 20$ CIEs obtidas através do algoritmo FastICA, com $m = 180$ e $p' = 19$	235

7.13	Comparações entre os espectros das coordenadas das CPEs e das CIEs de acordo com os valores dos períodos dominantes 144, 43, 33, 12, 8.5 e 4.5 meses, com $m = 180$ e $p' = 19$	236
7.14	Série temporal das anomalias mensais standardizadas da pressão do Índice de Oscilação do Atlântico Norte entre janeiro de 1958 e dezembro de 2007	238
7.15	Espectro da série temporal mensal do Índice de Oscilação do Atlântico Norte . .	240
7.16	Altura geopotencial a 500 milibar em junho de 1958, observada no setor 20°N - 80°N e 0°E - 357.5°E	241
7.17	Altura geopotencial a 500 milibar em dezembro de 1958, observada no setor 20°N - 80°N e 0°E - 357.5°E	241
7.18	1º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E	242
7.19	2º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E	242
7.20	3º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E	243
D.1	Séries temporais originais para cada uma das oito estações meteorológicas e respectivas reconstruções às custa das duas CPs (CP1+CP2) em simultâneo	303
D.2	Séries temporais originais para cada uma das oito estações meteorológicas e respectivas reconstruções às custa das duas CIs (CI1+CI2) em simultâneo	304
D.3	Número de vítimas mortais em acidentes de viação: componentes principais CP17 a CP24	305
D.4	Número de vítimas mortais em acidentes de viação: componentes principais CP25 a CP32	305
D.5	Número de vítimas mortais em acidentes de viação: componentes principais CP33 a CP40	306
D.6	Número de vítimas mortais em acidentes de viação: componentes principais CP41 a CP48	306
D.7	Altura geopotencial (a 500 milibar) em Portugal na região de Castelo Branco (40°N, 7.5°W) ao longo dos 50 anos em estudo	309

D.8	Altura geopotencial (a 500 milibar) em Filadélfia, no Estado da Pensilvânia (40°N , 75°W) ao longo dos 50 anos em estudo	309
-----	--	-----

Lista de Tabelas

3.1	Correlações entre as coordenadas das componentes principais em ACP e das componentes independentes em ACI através do FastICA e do AMUSE	109
3.2	Soma de quadrados dos resíduos para as reconstruções com a ACP e a ACI (através do FastICA e do AMUSE) à custa da primeira CP (CP1), da primeira CI (CI1), das duas primeiras CPs (CP12), das duas primeiras CIs (CI12) e das três primeiras CPs (CP123)	121
3.3	Resumo dos métodos de estimação do modelo ACI relacionados com a maximização da não normalidade	123
3.4	Continuação do resumo dos métodos de estimação do modelo ACI	124
4.1	Percentagens de variância de cada uma das primeiras 16 CPs e de sete pares dessas CPs, de um total de 48 CPs. A percentagem de variância de cada par de CPs foi calculada através da soma das percentagens individuais das respetivas CPs.	168
4.2	Percentagem de variância explicada pelas primeiras 5 CPs ao longo de diferentes valores do comprimento da janela - lag m	177
4.3	Valores próprios da SVD para a SSA sem centragem e para a SVD de cada um dos quatro tipos de SSA com centragem em análise	181
4.4	Percentagens de variâncias das CPs da SVD para a SSA sem centragem e para a SVD de cada um dos quatro tipos de SSA com centragem em análise	183
4.5	Cossenos entre os vetores próprios da SVD da SSA sem centragem e os vetores próprios da SVD de cada um dos quatro tipos de SSA com centragem em análise	184
4.6	Correlações entre as CPs da SVD da SSA sem centragem (SC) e as CPs da SVD de cada um dos quatro tipos de SSA com centragem em análise	185

6.1	Correlações entre as CPs obtidas pela técnica I e as CIs obtidas por ordem arbitrária pela técnica II, para um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas	214
6.2	Correlações entre as CPs obtidas pela técnica I e as CIs obtidas por ordem arbitrária pela técnica III, para um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas	214
6.3	Soma de quadrados dos resíduos para as reconstruções com as técnicas I, II e III, usando um número de componentes distintas; as técnicas II e III apresentam cada uma das ordens nos diferentes métodos de ordenação das CIs; as técnicas usaram séries desfasadas com um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas	215
7.1	Períodos dominantes dos pares de CPEs, para um comprimento da janela $m = 180$, $p' = 19$ e $q = 20$ componentes estendidas retidas	237
D.1	Matriz das correlações das variáveis originais dos dados Pacific2	302
D.2	Previsões do número de vítimas mortais para os meses do ano de 2010 e respetivos intervalos de predição obtidos através do método empírico	307

Capítulo 1

Introdução

*"O trabalho do pensamento é semelhante à perfuração de um poço,
a água é turva no início, mas logo se clarifica."*

Provérbio Chinês

1.1 Visão Geral

Em muitas áreas de investigação, os dados são recolhidos ao longo do tempo ou segundo uma outra ordem específica, o que geralmente acarreta uma dependência entre as observações. Tais dados abundam em áreas tão distintas como por exemplo na demografia, medicina, economia, processamento de sinais, climatologia, geologia, epidemiologia, ciências ambientais, entre outras.

Muitas vezes a matriz que representa um conjunto de dados é constituída por um elevado número de observações registadas para inúmeras variáveis. Várias técnicas da análise estatística multivariada são usadas para compreensão e sintetização de conjuntos de dados de grande dimensão, nas quais para uma matriz de dados se obtêm as principais fontes de variabilidade usando processos algébricos de condensação da informação. Estes processos envolvem usualmente o estudo das estruturas matriciais quer da matriz de dados, quer da matriz de variâncias-covariâncias, da matriz de correlações, ou de outra matriz auxiliar e informativa das inter-relações das variáveis e inter-relações das observações.

De entre as metodologias conhecidas a utilizar ao longo desta dissertação, destacam-se a Análise em Componentes Principais (ACP), a Análise em Componentes Independentes (ACI), a Análise Espectral Singular (SSA) e a Análise Espectral Singular Multicanal (MSSA). Nas

análises mais recentes e alternativas à ACP, opta-se por analisar outras matrizes informativas das estruturas dos dados, em vez da usual decomposição espectral da matriz de covariâncias ou de correlações. Uma nova concepção na compressão de dados com base na ACI aplicada na matriz dos dados desfasados é proposta como alternativa à MSSA, com o intuito de extrair mais informação que a aplicação da ACP usual, a qual é designada por *Lag*-Análise em Componentes Independentes (*Lag*ACI).

Os objetivos principais desta tese consistem em confrontar os métodos estatísticos multivariados da ACP, da ACI, da SSA e da MSSA, de forma a permitir encontrar semelhanças e diferenças relevantes entre as diferentes abordagens dos mesmos, consoante a existência ou não de pré-processamentos, e mediante os algoritmos que sejam implementados.

Nas aplicações a utilizar na área da climatologia, as observações possuem naturalmente uma estrutura temporal intrínseca e logo as componentes principais e as componentes independentes são encaradas sob a forma de séries temporais, com vista ao estudo nos domínios do tempo e da frequência.

As comparações entre diferentes métodos e suas abordagens são efetuadas através de vários indicadores tais como cossenos entre vetores, correlações entre componentes, somas de quadrados dos resíduos nas reconstruções dos dados originais, espectros de frequências, assim como outros índices estatísticos que permitem avaliar o desempenho de diferentes modelos, tendo sempre em mente a precisão e avaliação da convergência de resultados.

A implementação dos diferentes métodos baseados nos conceitos teóricos para cada técnica, assim como a comparação de resultados em cada técnica e entre técnicas, com base na aplicação a conjuntos de dados reais, são levadas a cabo essencialmente através de várias rotinas desenvolvidas no *software* estatístico de livre acesso *R*.

1.2 Enquadramento do Trabalho

Esta dissertação insere-se na área da Estatística e aborda essencialmente técnicas de Estatística Multivariada, na qual se enquadram a Análise em Componentes Principais, a Análise Espectral Singular e a Análise Espectral Singular Multicanal. Como técnica computacional e estatística adicional incorpora-se a Análise em Componentes Independentes, e propõe-se uma nova abordagem, designada por *Lag*ACI, como alternativa à MSSA para o estudo de séries temporais, e

que consiste na aplicação da ACI em vez da ACP usual à matriz dos desfasamentos das séries iniciais estendidas. Serão apresentados alguns exemplos de aplicação para séries temporais com vista à comparação de resultados entre as técnicas referidas.

1.3 Organização da Dissertação

Para além da Introdução, onde são apresentados, entre outros, alguns tópicos sobre séries temporais, e uma série de determinados indicadores de avaliação do desempenho de um modelo, esta dissertação está organizada em mais seis Capítulos, cujos assuntos estão divididos por Secções e respectivas Subsecções. No final da dissertação são ainda apresentados Apêndices independentes, com conceitos teóricos para complementar determinados assuntos abordados ao longo da mesma, e com alguns resultados adicionais provenientes da aplicação das metodologias sobre os conjuntos de dados utilizados. Para além disso, como usual, a tese encerra com a bibliografia de suporte indispensável a todo o trabalho de pesquisa.

Nos Capítulos 2, 3, 4 e 5, efetuam-se descrições dos aspetos teóricos de cada uma das técnicas multivariadas, respetivo estado da arte e alguns exemplos de aplicação a conjuntos de dados envolvendo séries temporais. Posteriormente é descrita a proposta de uma nova abordagem - a *LagACI*, como alternativa à *MSSA* para o estudo de séries temporais no Capítulo 6.

O Capítulo 7 descreve algumas comparações das técnicas da *MSSA* e *LagACI* e um exemplo ilustrativo de conjuntos de dados de dimensão complexa é utilizado com esse propósito. Por fim tecem-se algumas conclusões relevantes acerca do trabalho e fornecem-se sugestões para trabalho futuro de investigação.

No Capítulo 2 da ACP começam-se por descrever os conceitos teóricos, a forma de obtenção das Componentes Principais (CPs) e os objetivos da técnica. Algumas das principais propriedades das CPs são mencionadas assim como uma breve descrição da representação gráfica dos indivíduos e das variáveis. Para além dos aspetos históricos, são ainda mencionadas algumas abordagens relacionadas com a ACP. Este Capítulo é constituído pelas bases teóricas fulcrais para a descrição das técnicas introduzidas nos Capítulos seguintes.

A ACI é apresentada no Capítulo 3, no qual se efetua uma breve introdução e um resumo da história da técnica, e onde se descrevem exemplos de motivação da mesma, seguidos da descrição do modelo e acompanhada das propriedades básicas e das técnicas de pré-processamento dos

dados. Como parte fundamental da aplicação da técnica, são abordados vários métodos de estimação do modelo da ACI e referidos alguns dos algoritmos que podem ser utilizados para cada método. Dado que ao longo da tese são introduzidos exemplos de conjuntos de dados constituídos por séries temporais, apresentam-se métodos que captam a informação temporal, permitindo assim melhorar a estimação do modelo. Após a estimação das Componentes Independentes (CIs) através de determinado algoritmo, estas podem ser ordenadas consoante certos critérios de acordo com alguns dos métodos de pós-processamento de ordenação das CIs que são descritos. É ainda efetuada uma abordagem à questão do número adequado de CIs que devem ser retidas, mediante o propósito de obter a informação essencial para a interpretação do modelo. De uma forma sucinta mencionam-se algumas das principais áreas de aplicação da técnica assim como outros modelos de ACI que podem ser usados para além do modelo básico. Por fim, é explorado um exemplo de aplicação da técnica a um conjunto de dados temporais, de valores de pressão ao nível do mar, com o intuito de efetuar comparações de resultados entre as técnicas de ACP e ACI.

Para a técnica da Análise Espectral Singular (SSA), descrita no Capítulo 4, procede-se a uma introdução onde se mencionam algumas áreas de aplicação, com especial ênfase às aplicações de dados climáticos. Posteriormente efetua-se a descrição da técnica que, a partir de uma série temporal unidimensional, consiste na construção da matriz aumentada dos dados desfasados, seguida da respetiva decomposição em valores singulares, assim como da reconstrução da série original. É abordada a questão do comprimento da janela, ou seja, a escolha adequada do parâmetro de desfasamento a aplicar aquando da decomposição da série original, o qual vai influenciar a extração das componentes oscilatórias, da tendência e da estrutura subjacente de ruído. Para a previsão de valores futuros de uma série, sugere-se fundamentalmente o algoritmo de previsão recorrente em SSA, entre outros possíveis, e apresentam-se os intervalos de predição. São introduzidos alguns conceitos relacionados com a Análise de Fourier, os quais possibilitam a descrição de determinados pares de elementos provenientes da decomposição em valores singulares, que apresentam comportamentos similares ao nível da frequência e do tempo. É ainda discutida a utilização da centragem como contrapartida da SSA básica sem centragem. Novos desenvolvimentos teóricos são sugeridos para a centragem na série original, por analogia com os métodos de centragem mais usuais. Por último, recorre-se a uma série temporal de dados

relacionados com a sinistralidade rodoviária, registados mensalmente ao longo de 12 anos, para aplicação da SSA básica sem e com centragem com vista à comparação e discussão de resultados.

No Capítulo 5 descreve-se a técnica da Análise Espectral Singular Multicanal (MSSA), como uma extensão natural do caso univariado da SSA, onde se analisam em simultâneo várias séries temporais. Uma das principais áreas de aplicação é a Climatologia, na qual há interesse em identificar padrões climáticos relacionados não só com o tempo mas também com a localização espacial onde se registam as observações para determinadas variáveis meteorológicas. Serão descritas as etapas para implementação da técnica, de forma similar à SSA, com as respetivas adaptações. De forma sintética menciona-se como é importante a escolha adequada para o comprimento da janela (desfasamento temporal) na decomposição das séries e identificam-se algumas propriedades relativas à MSSA. Uma breve descrição do estado da arte relativamente à técnica e a alguns dos seus mais recentes desenvolvimentos serão apresentados.

No Capítulo da 6, para além de uma breve introdução incluindo a motivação para a nova proposta, descreve-se também a fundamentação teórica. No exemplo de aplicação, consideram-se três técnicas envolvendo a ACP e a ACI, duas das quais utilizam esta nova abordagem recorrendo ao uso da ACI aplicada na matriz dos dados desfasados (*LagACI*) como alternativa à técnica da MSSA, para as quais se usou o conjunto de dados da pressão ao nível do mar já explorados no Capítulo da ACI.

No Capítulo 7 efetua-se uma breve comparação envolvendo as técnicas da ACP, ACI, MSSA e *LagACI*. Como exemplo de aplicação considera-se um conjunto de dados da altura geopotencial. Os dados, de enorme complexidade, correspondem a um período de 50 anos observados num setor do Hemisfério Norte, segundo uma grelha espacial de 3600 localizações. Pretende-se analisar a eventual existência de convergência de resultados entre as técnicas e ainda entre os valores da série temporal do Índice de Oscilação do Atlântico Norte, um dos principais padrões de variabilidade da pressão atmosférica ao nível do mar no Hemisfério Norte. Serão construídos alguns mapas para ilustrar os modos principais de variabilidade da altura geopotencial ao nível do mar, patente nos dados. Finalmente apresentam-se algumas conclusões da investigação levada a cabo ao longo da tese, assim como uma visão geral de um possível trabalho a desenvolver no futuro.

1.4 Preliminares

1.4.1 Tópicos de Séries Temporais

Dada a extrema importância que as séries temporais (também conhecidas por séries cronológicas) desempenham ao longo de toda a tese, descrevemos nesta Subsecção alguns dos principais conceitos básicos de séries temporais, nomeadamente no domínio do tempo. Alguns destes e outros tópicos podem ser consultados em Murteira *et al.* (1993) ou em Oliveira (2003), assim como em von Storch e Zwiers (1999).

A parte referente a conceitos relacionados com o estudo de séries temporais no domínio da frequência pode ser encontrada no Apêndice C.3.

Definição 1.4.1 *Um **processo estocástico** é uma sequência ou coleção ordenada de variáveis aleatórias geralmente denotada por $\{X_t\}_{t \in T}$ ou simplesmente por $\{X_t\}$, em que T é um conjunto de índices representando o tempo.*

Definição 1.4.2 *Uma **série temporal** é uma sequência de n observações de um processo estocástico denotada por $X_t = (x_1, x_2, \dots, x_n)$, realizadas em pontos ou períodos de tempo durante um determinado intervalo.*

*Quando as observações são registadas em pontos isolados no tempo estamos na presença de uma **série temporal discreta**; para o caso de serem medidas ao longo de um intervalo de tempo, a série designa-se por **série temporal contínua**.*

As séries temporais discretas são muito comuns, de entre as quais se destacam aquelas cujas observações são efetuadas em pontos igualmente espaçados no tempo, como é o caso de dias, meses, anos, décadas, Para os exemplos desta natureza que iremos abordar, o conjunto de índices T pertence a \mathbb{N} .

Embora em determinados casos as séries temporais sejam originalmente contínuas, como nas medições de variáveis meteorológicas, no entanto passamos a trabalhar com séries discretizadas cujas observações são efetuadas em intervalos de tempo espaçados de forma regular ou irregular.

Definições 1.4.3 *Sejam X_t e X_s variáveis aleatórias do processo estocástico $\{X_t\}$.*

- A **média** de X_t é definida por $E(X_t) = \mu_t$.

- A **função de autocovariância** é dada por

$$\gamma_{t,s} = \text{cov}(X_t, X_s) = E((X_t - \mu_t)(X_s - \mu_s)).$$

Esta função fornece uma descrição, ao nível do domínio do tempo, dos momentos de segunda ordem do processo.

- A **função de autocorrelação**, é dada por

$$\rho_{t,s} = \text{corr}(X_t, X_s) = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}.$$

Definição 1.4.4 Sejam $\{X_t\}$ e $\{Y_s\}$ dois processos estocásticos. A **função de covariância cruzada** é definida por

$$\gamma_{xy}(t, s) = E((X_t - \mu_{X_t})(Y_s - \mu_{Y_s})).$$

Embora os momentos de primeira e de segunda ordem possam depender, por exemplo, de parâmetros como t e s , para muitas séries temporais é possível admitir que à medida que se desloca o tempo, as suas propriedades mantêm-se inalteráveis. Perante esta situação dizemos que a **série temporal** é **estacionária** ou **estritamente estacionária**.

Definição 1.4.5 Um processo estocástico $\{X_t\}$ diz-se **estacionário** (ou **estritamente estacionário**) se todas as propriedades estocásticas são invariantes a qualquer translação do tempo t . Se um processo $\{X_t\}$ é estacionário então

- 1) X_t tem a mesma função de distribuição para todo o t , e
- 2) Para todo o t e s , a função de distribuição conjunta de X_t e X_s depende apenas de $|t - s|$.

Dado que a estacionariedade no sentido estrito é demasiado exigente e de difícil verificação, é possível definir uma outra forma de estacionariedade mais fraca, geralmente conhecida como estacionariedade de segunda ordem (ou estacionariedade em variância).

Definição 1.4.6 Um processo estocástico $\{X_t\}$ diz-se **fracamente estacionário** (ou **estacionário de segunda ordem**) se

- 1) $E(X_t) = \mu_t, \forall t \in T$ (média constante), e
- 2) $E((X_t - \mu_t)(X_s - \mu_s)) = \gamma_{t,s} = \gamma_{|t-s|}$ (a autocovariância é função apenas da diferença temporal $|t - s|$).

Nota 1.4.7 *Se um processo estritamente estacionário possuir variância finita então também é um processo fracamente estacionário (ou de segunda ordem).*

Ao pressupor que existem momentos até à segunda ordem e que a média e variância são constantes, para além da função de autocovariância, também a função de autocorrelação passa a depender apenas da diferença temporal $|t - s|$. Neste caso podemos reescrever as funções de autocovariância e de autocorrelação como funções do tempo desfasado, ou seja, do desfasamento $k = |t - s|$, as quais ficam denotadas por

$$\gamma_{t,s} = \gamma(k) \text{ e } \rho_{t,s} = \rho(k).$$

Ao considerar uma realização do processo, dizemos que a **série temporal é fracamente estacionária** ou **estacionária de segunda ordem**. Neste âmbito, a função de autocorrelação $\rho(k)$ passa a medir a proximidade entre a série em causa e a série desfasada no tempo por k unidades.

Quando uma série não apresenta média ou covariância constantes, ou seja, quando uma destas quantidades muda consoante o tempo, dizemos que estamos perante uma **série temporal não estacionária**.

Definição 1.4.8 *Um processo estocástico $\{N_t\}$ diz-se **processo de ruído branco** quando é formado por uma sequência de variáveis aleatórias independentes e com a mesma distribuição, $N_t \sim N(0, \sigma_\varepsilon^2)$, donde vem que*

$$\gamma(k) = \begin{cases} \sigma_\varepsilon^2 & \text{se } k = 0 \\ 0 & \text{se } k \neq 0 \end{cases}.$$

Portanto um processo de ruído branco é um processo estacionário de segunda ordem.

Para um processo $\{N_t\}$ obtém-se um espectro $\Gamma(\omega)$ (Apêndice C.3) igual a σ_ε^2 para $0 \leq \omega \leq 1/2$, ou seja, uma função constante que indica que todas as oscilações contribuem de igual modo para explicar o processo de ruído branco tal como o espectro de luz branca. Contudo é preciso ter atenção à sua interpretação, uma vez que é expectável o espectro oscilar em torno do valor σ_ε^2 , numa amostra em concreto.

1.4.2 Decomposição Espectral e Decomposição em Valores Singulares

Nesta parte referimos dois conceitos importantes em teoria de matrizes que consideramos de extrema importância na abordagem de Capítulos seguintes: a Decomposição Espectral e a Decomposição em Valores Singulares (SVD). Enquanto que a Decomposição Espectral é válida para

matrizes simétricas (necessariamente quadradas), na obtenção de elementos conhecidos por valores e vetores próprios, a SVD permite fatorizar qualquer matriz, quadrada ou retangular, de uma forma simples através de elementos denominados de valores e vetores singulares. A SVD é uma ferramenta fundamental para a aplicação da ACP, assim como da SSA e da MSSA, entre outras técnicas estatísticas.

Estas decomposições não dispensam a consulta de conceitos elementares sobre matrizes e valores e vetores próprios (Apêndice C.1) assim como outros pormenores que podem ser encontrados, por exemplo, em Horn e Johnson (1985a, 1985b).

Teorema 1.4.9 (Decomposição Espectral) *Seja \mathbf{A} uma matriz quadrada de ordem n . Então \mathbf{A} é simétrica se e só se existir uma matriz ortogonal \mathbf{P} da mesma ordem (com colunas representadas por \mathbf{p}_i) e uma matriz diagonal \mathbf{L} da mesma ordem (com elementos da diagonal principal representados por λ_i) tal que*

$$\mathbf{A} = \mathbf{P}\mathbf{L}\mathbf{P}^T \Leftrightarrow \mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{p}_i \mathbf{p}_i^T.$$

Do teorema anterior podemos tirar as seguintes ilações:

Observações 1.4.10

- i) λ_i são os valores próprios de \mathbf{A} e \mathbf{p}_i são os vetores próprios de \mathbf{A} , os quais formam um conjunto ortonormado.
- ii) No caso de todos os valores próprios serem diferentes, os vetores próprios \mathbf{p}_i são únicos (a menos do sentido), pelo que uma ordenação dos valores próprios faz com que a decomposição seja única.
- iii) Se existirem valores próprios iguais, a decomposição não é única, uma vez que qualquer combinação linear de dois vetores próprios associados ao mesmo valor próprio λ também é um vetor próprio associado ao mesmo valor próprio λ de \mathbf{A} . Portanto nesse caso, as colunas da matriz \mathbf{P} associadas a esse mesmo valor próprio podem formar qualquer base ortonormada de um espaço vetorial de dimensão igual ao número de vezes que esse valor próprio se repete na matriz diagonal \mathbf{L} .

Teorema 1.4.11 *Seja \mathbf{A} uma matriz simétrica de ordem p . Então \mathbf{A} é definida positiva (Apêndice C.1.2) se e só se $\mathbf{A} = \mathbf{X}^T \mathbf{X}$ para alguma matriz \mathbf{X} do tipo $n \times p$ de característica $p \leq n$.*

Deste teorema podemos constatar que qualquer matriz de variâncias-covariâncias e de correlações é sempre uma matriz definida positiva desde que não haja multicolinearidade.

Teorema 1.4.12 (Decomposição em Valores Singulares) *Seja $\mathbf{X}_{n \times p}$ uma matriz qualquer de característica r . Então:*

$$\mathbf{X} = \mathbf{V}\mathbf{\Delta}\mathbf{U}^T \Leftrightarrow \mathbf{X} = \sum_{i=1}^r \delta_i \mathbf{v}_i \mathbf{u}_i^T,$$

onde:

$\mathbf{\Delta}_{r \times r}$ é uma matriz diagonal;

$\mathbf{V}_{n \times r}$ é uma matriz com colunas ortonormadas, ou seja, $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$;

$\mathbf{U}_{p \times r}$ é uma matriz com colunas ortonormadas, ou seja, $\mathbf{U}^T \mathbf{U} = \mathbf{I}_r$;

δ_i são os elementos da diagonal principal de $\mathbf{\Delta}$ designados por **valores singulares** de \mathbf{X} ;

\mathbf{v}_i são as colunas de \mathbf{V} designados por **vetores singulares esquerdos** de \mathbf{X} ;

\mathbf{u}_i são as colunas de \mathbf{U} designados por **vetores singulares direitos** de \mathbf{X} .

Observações 1.4.13 *De acordo com a Decomposição em Valores Singulares temos:*

- i) A matriz \mathbf{U} é uma matriz cujas colunas correspondem aos vetores próprios ortonormados da matriz $\mathbf{X}^T \mathbf{X}$ associados a valores próprios não nulos. Por analogia, a matriz \mathbf{V} tem correspondência com $\mathbf{X} \mathbf{X}^T$.
- ii) $\mathbf{\Delta}$ é a matriz das raízes quadradas dos valores próprios não nulos de $\mathbf{X}^T \mathbf{X}$ (iguais aos de $\mathbf{X} \mathbf{X}^T$).
- iii) É costume ordenar os valores singulares δ_i por ordem decrescente dos seus valores.

1.4.3 Indicadores Comparativos do Desempenho de um Modelo

Quando se pretende analisar o desempenho de diferentes modelos estatísticos de acordo com as suas abordagens, a precisão e avaliação de resultados é geralmente efetuada através de indicadores comparativos entre as técnicas utilizadas no estudo. Os principais indicadores que irão ser utilizados ao longo da tese são:

- Cossenos entre vetores;
- Correlações entre componentes;

- Espectros de frequências (Apêndice C.3);
- Soma de quadrados dos resíduos (SQR) nas reconstruções dos dados originais.

De seguida apresentamos outras formas de efetuar a avaliação do desempenho de um modelo, com vista à comparação transversal de várias técnicas que serão abordadas nos Capítulos seguintes, recorrendo às reconstruções das séries.

Indicadores Adicionais

Num conjunto de n observações, os resultados observados O_i ($i = 1, 2, \dots, n$), de uma variável aleatória O , correspondem aos valores originais das séries em estudo, e os valores estimados ou previstos P_i ($i = 1, 2, \dots, n$), de uma variável aleatória P , correspondem às reconstruções dos dados originais à custa de um determinado número de componentes.

Geralmente, as medidas estatísticas de avaliação do desempenho de um modelo comparam as estimativas ou previsões P_i com as respetivas observações O_i . Por uma questão de comparação, as unidades de medida das previsões e das observações devem ser as mesmas.

Definem-se os seguintes conceitos:

Definições 1.4.14 *Seja O_i cada um dos resultados observados e P_i cada um dos valores estimados para cada observação i ($i = 1, 2, \dots, n$). Seja \bar{P} o valor médio das estimativas e \bar{O} o valor médio das observações.*

- O **resíduo** (ou erro de previsão) R_i é dado por $P_i - O_i$.
- A **Soma de Quadrados dos Resíduos** denota-se por $SQR = \sum_{i=1}^n (P_i - O_i)^2$.
- O **Erro Absoluto Médio** é dado por $EAM = n^{-1} \sum_{i=1}^n |P_i - O_i|$.
- O **Erro Quadrático Médio** denota-se por $EQM = n^{-1} \sum_{i=1}^n (P_i - O_i)^2$.
- A **Raiz do Erro Quadrático Médio** é dada por

$$REQM = \sqrt{EQM} = \sqrt{n^{-1} \sum_{i=1}^n (P_i - O_i)^2}.$$

As definições anteriores servem de suporte para as definições de coeficientes e índices que se seguem.

Definição 1.4.15 *Coefficiente de determinação* - O quadrado do coeficiente de correlação de Pearson, entre valores observados e previstos ou reconstruídos é dado por

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2.$$

No contexto da regressão linear, o coeficiente de determinação representa a proporção da variância total nos dados observados que pode ser explicada pelo modelo e varia entre 0 e 1. Contudo, este coeficiente é limitado à avaliação linear das relações entre as variáveis e é também insensível a diferenças aditivas e proporcionais entre previsões e observações do modelo (Willmott, 1981, 1984). Por outro lado, a magnitude do R^2 não está relacionada de forma coerente com a precisão da previsão, segundo os exemplos mencionados em Willmott (1982).

Para além das consideradas medidas clássicas de erro absoluto e relativo em relação à precisão das previsões, também são usados outros indicadores de erro relativo na literatura em áreas como a climatologia ou a hidrologia. Passemos a definir alguns deles.

Definição 1.4.16 *Índice de concordância de Willmott* (Willmott, 1981) - O índice de medida relativo é dado por

$$d = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}. \quad (1.1)$$

Willmott argumenta que o índice, o qual varia entre 0 e 1, deveria descrever a covariância relativa de P e O em torno de uma estimativa da média \bar{O} . O índice assume o valor $d = 0$ quando todos os P_i são iguais a \bar{O} , ou quando, $P_i - \bar{O} = -(O_i - \bar{O})$, para todo o i . Perante tal situação, a variabilidade das previsões em torno de \bar{O} apresenta um comportamento inverso ao da variabilidade das observações.

O índice d foi pensado com o intuito de superar a falta de sensibilidade de medidas baseadas no coeficiente de correlação, no que se refere às diferenças de médias e variâncias das previsões e observações. Segundo Legates e McCabe (1999), o índice d revela-se mais eficaz relativamente ao clássico coeficiente de determinação R^2 na avaliação do modelo, mas no entanto é bastante sensível à presença de *outliers* devido aos quadrados das diferenças.

Definição 1.4.17 *Índice de concordância de Willmott revisto* (Willmott et al., 1985) - A nova versão do índice de concordância de Willmott de 1981 (1.1), baseado nas somas dos valores absolutos dos erros, é dado por

$$d_1 = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)}. \quad (1.2)$$

Embora d_1 também varie entre 0 e 1, uma vantagem comparativamente a d é que a aproximação da concordância é mais lenta à medida que P se aproxima de O , e portanto fornece uma maior separação quando se comparam dois modelos que apresentam desempenhos relativamente bons.

Definição 1.4.18 *Índice de concordância de Willmott refinado* (Willmott et al., 2012) - A versão reformulada do índice de concordância de Willmott revisto de 1985 (1.2), para uma constante $c > 0$, é descrito por

$$d_r = \begin{cases} 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{c \sum_{i=1}^n |O_i - \bar{O}|}, & \text{se } \sum_{i=1}^n |P_i - O_i| \leq c \sum_{i=1}^n |O_i - \bar{O}| \\ \frac{c \sum_{i=1}^n |O_i - \bar{O}|}{\sum_{i=1}^n |P_i - O_i|} - 1, & \text{se } \sum_{i=1}^n |P_i - O_i| > c \sum_{i=1}^n |O_i - \bar{O}| \end{cases}.$$

Ao substituir P_i por O_i no denominador em (1.2), removendo assim a influência de P no denominador, o índice revisto d_1 , é transformado em

$$1 - \frac{\sum_{i=1}^n |P_i - O_i|}{2 \sum_{i=1}^n |O_i - \bar{O}|},$$

o que corresponde ao caso particular de d_r em que $c = 2$, e desta forma pondera-se o número de desvios avaliados pelo numerador e pelo denominador da fração. Devido à sua flexibilidade pode ser aplicado a uma vasto leque de problemas de desempenho do modelo.

O índice d_r varia entre -1 e 1 , e está diretamente relacionado com a precisão do modelo, ao nível dos aumentos e diminuições dos valores do erro absoluto médio (EAM). Podemos considerar as seguintes interpretações (Willmott, 2012):

- $d_r = 0.5$ indica que a soma das magnitudes dos erros $P_i - O_i$ é metade da soma dos desvios ao modelo perfeito (quando $P_i = O_i$ para todo o i) e das magnitudes dos desvios das observações $O_i - \bar{O}$;
- $d_r = 0$ indica que a soma das magnitudes dos erros e a soma dos desvios ao modelo perfeito e as magnitudes dos desvios das observações são equivalentes;
- $d_r = -0.5$ significa que a soma das magnitudes dos erros é o dobro da soma dos desvios ao modelo perfeito e das magnitudes dos desvios das observações;
- Para valores de d_r próximos de -1 pode significar que os desvios estimados $P_i - \bar{O}$ do modelo em torno de \bar{O} sejam estimativas fracas dos desvios observados ou, por outro lado, que apenas exista uma pequena variabilidade nas observações.

Outros dois índices relacionados de uma forma muito próxima com o índice de concordância de Willmott refinado são os índices de Nash e Sutcliffe e de Legates e McCabe, uma vez que também nestes os valores de P não aparecem nos denominadores das frações.

Definição 1.4.19 *Coefficiente de eficiência de Nash e Sutcliffe (Nash e Sutcliffe, 1970)*

- O índice baseado em quadrados de diferenças é dado por

$$E = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}.$$

Este índice é correlacionado positivamente com d_r e assume valores inferiores ou iguais a 1, com valores próximos da unidade a indicarem um bom grau de concordância entre os valores observados e previstos.

Harmel e Smith (2007) argumentam que, para analisar a qualidade do ajustamento de um modelo é preferível utilizar o coeficiente de eficiência E relativamente ao coeficiente de determinação R^2 , porque embora ambos sejam sensíveis a valores extremos, o R^2 não é sensível a diferenças proporcionais ou aditivas entre as previsões e as observações. Contudo, em E as diferenças entre as previsões e as observações são calculadas ao quadrado, e como tal os valores elevados na série são sobrestimados enquanto que tal não acontece com os valores mais baixos (Krause *et al.*, 2005, Legates e McCabe, 1999). Com vista a reduzir o efeito dos termos quadráticos surge uma versão modificada do coeficiente de eficiência.

Definição 1.4.20 *Coeficiente de eficiência (ou índice) de Legates e McCabe (Legates e McCabe, 1999) - O coeficiente de eficiência modificado é dado por*

$$E_1 = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n |O_i - \bar{O}|}.$$

De entre os índices aqui apresentados, E_1 é o que apresenta mais semelhanças com o índice de concordância de Willmott refinado, d_r , sendo mesmo equivalente para $c = 1$ quando é positivo. Para valores positivos de d_r , $d_r = 0.5(E_1 + 1)$, enquanto que para valores negativos de d_r , tem-se $d_r = -[2(E_1 - 1)^{-1} + 1]$.

Krause *et al.* (2005) investigaram nove medidas para avaliar o desempenho do modelo em três exemplos com dados hidrológicos. Aplicaram os indicadores R^2 , d , d_1 , E , E_1 , assim como uma versão ponderada de R^2 , o coeficiente de eficiência de Nash e Sutcliffe com logaritmos dos valores observados e estimados e ainda dois critérios de eficiência relativos para E e d . De um modo geral, não foi possível de estabelecer quais os indicadores com melhor desempenho, uma vez que cada indicador apresenta especificidades que devem ser tidas em conta durante a calibração e avaliação do modelo.

Willmott *et al.* (2012) compararam o desempenho relativo do índice de concordância de Willmott refinado, d_r , com os desempenhos de outros seis indicadores (d , d_1 , E , E_1 , índice de Watterson (1996) e índice de Mielke e Berry (2001)) e efetuaram simulações de valores de P e de O para descrever o comportamento da variação destes índices relativamente ao d_r . O índice refinado d_r apresentou evidências de melhores resultados, não triviais segundo os autores, sobre as primeiras versões do índice assim como em relação aos restantes índices que foram comparados. O estudo sugere que d_r é flexível e aplicável a uma ampla variedade de situações em que se pretende avaliar o desempenho de modelos.

Capítulo 2

Análise em Componentes Principais

*"Minhas palavras sobem voando, meus pensamentos permanecem abaixo;
palavras sem pensamentos nunca chegam ao céu."*

William Shakespeare (1564-1616) - Escritor Britânico

2.1 Introdução e Descrição da ACP

A Análise em Componentes Principais (ACP) é uma das técnicas estatísticas multivariadas que mais se utiliza em quase todas as áreas do domínio científico.

O objetivo da ACP é reduzir a dimensão de um vasto conjunto de variáveis, obtendo um pequeno conjunto de novas variáveis designadas de Componentes Principais (CPs), através de combinações lineares das variáveis originais, de forma a que as CPs sejam não correlacionadas entre si e maximizem sucessivamente a variância. Desta forma as CPs sintetizam a informação essencial subjacente ao conjunto dos dados originais, mantendo tanto quanto possível nesse novo subconjunto a variabilidade presente no conjunto inicial.

Há a ter em conta que, se as variáveis iniciais forem não correlacionadas, então o estudo em causa preservará as variáveis, pelo que este não fará qualquer sentido lógico, uma vez que impossibilita a redução da dimensão dos dados. Quando as variáveis iniciais estão correlacionadas (positiva ou negativamente), a ACP possibilita reduções de dimensão sem grande perda de informação.

Consideremos o espaço \mathbb{R}^n , de dimensão n , constituído por vetores n -dimensionais de indivíduos, observações ou pontos e o subespaço p -dimensional de \mathbb{R}^n , gerado pelas p variáveis

aleatórias originais (admitindo $p < n$). Pretendemos encontrar combinações lineares das variáveis iniciais $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ para criar novas variáveis $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p$, não correlacionadas entre si e de variância máxima (por ordem decrescente) denominadas por Componentes Principais.

Estas componentes estão ordenadas de modo a que \mathbf{z}_1 tenha a maior variância, \mathbf{z}_2 tenha a segunda maior variância, e assim por diante. Logo $\text{var}(\mathbf{z}_1) \geq \text{var}(\mathbf{z}_2) \geq \dots \geq \text{var}(\mathbf{z}_p)$, onde $\text{var}(\mathbf{z}_k)$ ($k = 1, 2, \dots, p$) representa a variância de cada CP \mathbf{z}_k . Uma vez que a soma das variâncias das p CPs é igual à soma das variâncias das p variáveis originais de acordo com a relação (2.1) a estabelecer na Secção 2.3, e como tal, a redução da dimensão dos dados em causa consiste em reter um número q de CPs \mathbf{z}_i ($i = 1, 2, \dots, q$) menor do que p , desprezando assim as componentes \mathbf{z}_i , com $i > q$, as quais apresentam muito pouca informação, ou seja, as que contêm uma variação que se considera desprezável em relação à variabilidade total das variáveis originais. Deste modo, podemos salientar que a finalidade da ACP prende-se essencialmente com a procura de um subespaço q -dimensional do subespaço p -dimensional de \mathbb{R}^n , onde a projecção dos dados seja maximizada.

Considere-se inicialmente uma matriz de dados $\mathbf{X}_{n \times p}$ de n observações nas p variáveis. É usual considerar um número de indivíduos mais elevado em relação ao número de variáveis ($n > p$). No entanto, nos dias de hoje e com o recurso a novas tecnologias informáticas, cresce o número de aplicações em que $p > n$, como por exemplo nas ciências atmosféricas em que determinadas variáveis meteorológicas são registadas em p localizações geográficas ao longo de n momentos temporais (Wilks, 2006). Também na genética é usual analisarem-se muitas variáveis em simultâneo com um reduzido número de observações, como por exemplo quando se trabalha com expressões de genes (Johansson *et al.*, 2003), proteínas ou mesmo com dados de sequenciação de DNA. Trendafilov e Unkel (2011) propuseram uma nova abordagem na Análise Fatorial para explorar conjuntos de dados com mais variáveis que observações. Foi sugerido um novo modelo de análise descritiva de dados e redução da sua dimensionalidade, considerado como alternativo à ACP, assim como um novo algoritmo para estimar em simultâneo todos os parâmetros do modelo, baseado na decomposição em valores singulares (SVD).

Consideramos que as variáveis são quantitativas e que \mathbf{X} tem característica $p < n$, isto é, nenhuma coluna (variável) de \mathbf{X} é combinação linear das restantes.

Para cada variável \mathbf{x}_j da matriz \mathbf{X} é definida a sua média \bar{x}_j , obtendo-se assim o vetor das

médias dado por $\bar{\mathbf{x}}^T = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$, onde $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ com $j = 1, 2, \dots, p$ e onde T representa a transposta de um vetor ou de uma matriz.

A variância de cada variável \mathbf{x}_j , com $j = 1, 2, \dots, p$, da matriz \mathbf{X} é dada por

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2.$$

Na ACP é costume definir inicialmente a matriz dos dados \mathbf{X} como sendo a matriz dos dados centrados, onde a cada valor de cada variável foi subtraída a média da respetiva variável.

A matriz de variâncias-covariâncias de p variáveis é a matriz simétrica \mathbf{S} dada por

$$\mathbf{S} = \begin{bmatrix} s_1^2 & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \dots & s_p^2 \end{bmatrix}.$$

Os elementos da diagonal principal de \mathbf{S} correspondem às variâncias das p variáveis, enquanto os restantes valores s_{jk} correspondem a covariâncias entre as variáveis \mathbf{x}_j e \mathbf{x}_k , dadas por

$$s_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k),$$

com $j, k = 1, 2, \dots, p$ e $j \neq k$.

A matriz de correlações de p variáveis é a matriz simétrica \mathbf{R} dada por

$$\mathbf{R} = \begin{bmatrix} 1 & \dots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \dots & 1 \end{bmatrix}.$$

Os valores da diagonal principal de \mathbf{R} são iguais à unidade, enquanto os restantes valores r_{jk} correspondem a correlações entre as variáveis \mathbf{x}_j e \mathbf{x}_k , dadas por

$$r_{jk} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_{ij} - \bar{x}_j}{s_j} \right) \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right),$$

com $j, k = 1, 2, \dots, p$ e $j \neq k$.

A partir de \mathbf{X} , o estudo da ACP é efetuado sobre a matriz de variâncias-covariâncias \mathbf{S} (ou sobre a matriz de correlações \mathbf{R} , nos casos em que as variáveis, além de centradas, são também reduzidas), das p variáveis da amostra inicial.

Daqui em diante, designa-se por \mathbf{X} a matriz centrada dos dados.

Usando estruturas matriciais, a matriz \mathbf{S} é dada por

$$\mathbf{S} = \frac{1}{n} \mathbf{X}^T \mathbf{X}.$$

Depois de se obter a matriz de variâncias-covariâncias \mathbf{S} , aplica-se a decomposição espectral, tendo em conta o Teorema 1.4.9 da Subsecção 1.4.2, dada por

$$\mathbf{S} = \mathbf{P} \mathbf{L} \mathbf{P}^T,$$

onde \mathbf{P} representa a matriz dos vetores próprios de \mathbf{S} , e \mathbf{L} a matriz diagonal cuja diagonal principal contém os valores próprios de \mathbf{S} .

Sempre que a análise for efetuada com base na matriz \mathbf{S} diz-se que estamos perante uma ACP sobre a matriz de variâncias-covariâncias.

No entanto existem casos em que nem todas as p variáveis iniciais têm as mesmas unidades de medida, pelo que a ACP deixa de fazer sentido sobre as variáveis originais, uma vez que as CPs são sensíveis a mudanças de escala, em particular não são invariantes a mudanças multiplicativas diferenciadas (Jolliffe, 2002). Para colmatar esta situação, deveremos utilizar uma ACP sobre as variáveis centradas e reduzidas e consequentemente, as CPs passam a ser combinações lineares destas variáveis previamente centradas e reduzidas. A decomposição espectral relevante passa a recair sobre a matriz de correlações \mathbf{R} , em vez de ser sobre a matriz \mathbf{S} , e nestes casos diz-se que a ACP é realizada sobre a matriz de correlações. Por isso, é de ressaltar que não existe qualquer relação direta entre as CPs de uma ACP sobre a matriz de variâncias-covariâncias e as de uma ACP sobre a matriz de correlações.

Após a decomposição espectral $\mathbf{S} = \mathbf{P} \mathbf{L} \mathbf{P}^T$, ao projetar os dados centrados no espaço gerado pelos vetores próprios de \mathbf{P} , obtemos a matriz $\mathbf{Z} = \mathbf{X} \mathbf{P}$ cujas colunas são as coordenadas dos n indivíduos nas CPs. Designa-se por p_{jk} o valor do coeficiente associado à variável \mathbf{x}_j no k -ésimo vetor próprio, com $j, k = 1, 2, \dots, p$. Sendo assim, a k -ésima CP é dada por $\mathbf{z}_k = \mathbf{X} \mathbf{p}_k$ e cada um dos seus coeficientes z_{hk} designa a coordenada (*score* em inglês) da h -ésima observação na k -ésima CP, com $h = 1, 2, \dots, n$ e $k = 1, 2, \dots, p$.

2.2 Obtenção das Componentes Principais

Apresentamos de seguida uma abordagem para obtenção das CPs numa ACP com base no caso amostral. Um estudo mais aprofundado em termos da teoria subjacente a uma população é dado em Jolliffe (2002).

Seja \mathbf{X} a matriz dos dados iniciais centrada. Tendo em conta que cada CP é definida por $\mathbf{z}_k = \mathbf{X}\mathbf{p}_k$ ($k = 1, 2, \dots, p$), há que estruturar o método que permita encontrá-las, não esquecendo que pretendemos que sejam não correlacionadas e de variância máxima.

Para obter a primeira CP, \mathbf{z}_1 , há que encontrar um vetor \mathbf{p}_1 que maximize a variância da combinação linear $\mathbf{X}\mathbf{p}_1$, ou seja, pretendemos maximizar $\text{var}(\mathbf{z}_1) = \text{var}(\mathbf{X}\mathbf{p}_1) = \mathbf{p}_1^T \mathbf{S} \mathbf{p}_1$. Para que o máximo seja atingido de forma única, a menos do sentido, há que impor a restrição de que o vetor tenha norma 1, ou seja, $\mathbf{p}_1^T \mathbf{p}_1 = 1$.

O Método dos Multiplicadores de Lagrange é introduzido para encontrar a solução do problema (Jolliffe, 2002). Portanto pretendemos maximizar a função Lagrangeana

$$\mathcal{L}(\mathbf{p}_1) = \mathbf{p}_1^T \mathbf{S} \mathbf{p}_1 - \lambda(\mathbf{p}_1^T \mathbf{p}_1 - 1),$$

onde λ é o multiplicador de Lagrange.

Efetuando a derivada parcial em ordem a \mathbf{p}_1 e igualando-a a zero, vem

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}_1} = 0 \Leftrightarrow 2\mathbf{S}\mathbf{p}_1 - 2\lambda\mathbf{p}_1 = 0 \Leftrightarrow (\mathbf{S} - \lambda\mathbf{I}_p)\mathbf{p}_1 = 0,$$

onde \mathbf{I}_p é a matriz identidade de ordem p .

De acordo com a equação obtida, λ é um valor próprio de \mathbf{S} enquanto que \mathbf{p}_1 é o respetivo vetor próprio. A variância de \mathbf{z}_1 corresponde ao maior valor de λ possível, uma vez que $\text{var}(\mathbf{z}_1) = \mathbf{p}_1^T (\mathbf{S}\mathbf{p}_1) = \mathbf{p}_1^T (\lambda\mathbf{p}_1) = \lambda\mathbf{p}_1^T \mathbf{p}_1 = \lambda$. Logo \mathbf{p}_1 é o vetor próprio associado ao maior valor próprio de \mathbf{S} , e a variância da combinação linear $\mathbf{X}\mathbf{p}_1$ será o maior valor próprio, designado por λ_1 . Obtemos assim a primeira solução do problema onde $\mathbf{z}_1 = \mathbf{X}\mathbf{p}_1$ é a primeira CP.

Para determinar a segunda solução do problema pretendemos maximizar a variância de $\mathbf{z}_2 = \mathbf{X}\mathbf{p}_2$ sujeita a duas condições: $\mathbf{p}_2^T \mathbf{p}_2 = 1$ e o vetor \mathbf{p}_2 ortogonal ao novo vetor \mathbf{p}_1 ($\mathbf{p}_2^T \mathbf{p}_1 = 0$).

Esta é a forma de assegurar que as CPs \mathbf{z}_1 e \mathbf{z}_2 sejam não correlacionadas entre si, uma vez que $\text{cov}(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{p}_2^T (\mathbf{S}\mathbf{p}_1) = \mathbf{p}_2^T (\lambda_1\mathbf{p}_1) = \lambda_1(\mathbf{p}_2^T \mathbf{p}_1) = 0$.

Sendo assim, a função Lagrangeana toma o seguinte aspeto

$$\mathcal{L}(\mathbf{p}_2) = \mathbf{p}_2^T \mathbf{S} \mathbf{p}_2 - \lambda(\mathbf{p}_2^T \mathbf{p}_2 - 1) - \mu(\mathbf{p}_2^T \mathbf{p}_1),$$

onde λ e μ são os multiplicadores de Lagrange. Salientando apenas a derivada parcial em ordem a \mathbf{p}_2 igualada a zero, vem

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}_2} = 0 \Leftrightarrow 2\mathbf{S}\mathbf{p}_2 - 2\lambda\mathbf{p}_2 - \mu\mathbf{p}_1 = 0.$$

Multiplicando à esquerda por \mathbf{p}_1^T , vem

$$2\mathbf{p}_1^T\mathbf{S}\mathbf{p}_2 - 2\lambda\mathbf{p}_1^T\mathbf{p}_2 - \mu\mathbf{p}_1^T\mathbf{p}_1 = 0 \Leftrightarrow 2\lambda_1\mathbf{p}_1^T\mathbf{p}_2 - 2\lambda\mathbf{p}_1^T\mathbf{p}_2 - \mu\mathbf{p}_1^T\mathbf{p}_1 = 0 \Leftrightarrow \mu = 0,$$

uma vez que os dois primeiros termos se anulam e $\mathbf{p}_1^T\mathbf{p}_1 = 1$.

Portanto obtemos novamente

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}_2} = 0 \Leftrightarrow 2\mathbf{S}\mathbf{p}_2 - 2\lambda\mathbf{p}_2 = 0 \Leftrightarrow (\mathbf{S} - \lambda\mathbf{I}_p)\mathbf{p}_2 = 0,$$

o que significa que \mathbf{p}_2 também é um vetor próprio de \mathbf{S} .

Como este vai ser ortogonal em relação ao primeiro vetor próprio \mathbf{p}_1 e com a condição de maximizar o valor $\mathbf{p}_2^T\mathbf{S}\mathbf{p}_2$, logo \mathbf{p}_2 é o vetor próprio associado ao segundo maior valor próprio de \mathbf{S} , e a variância de \mathbf{z}_2 será esse segundo maior valor próprio, designado por λ_2 . Portanto determinamos assim a segunda solução do problema, obtendo \mathbf{z}_2 como sendo a segunda CP.

Através de um raciocínio análogo ao procedimento anterior, obtemos as restantes CPs.

Portanto, segundo o Método dos Multiplicadores de Lagrange, a variância da k -ésima CP \mathbf{z}_k , é igual ao k -ésimo maior valor próprio λ_k : $var(\mathbf{z}_k) = \lambda_k$, $k = 1, \dots, p$.

Caso a matriz \mathbf{X} fosse para além de centrada, também reduzida, passaríamos a usar a matriz de correlações \mathbf{R} em vez da matriz \mathbf{S} de variâncias-covariâncias para obter as CPs.

2.3 Algumas Propriedades na ACP

Passaremos a descrever algumas das propriedades mais relevantes na aplicação da ACP. Para consulta destas e outras propriedades, para um estudo mais alargado destacamos Jolliffe (2002) e Jackson (1991).

- Geralmente a característica da matriz inicial \mathbf{X} é igual ao número de variáveis originais, p , e portanto a matriz simétrica $\mathbf{S} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$ tem também característica p (Horn e Johnson, 1985a) e como tal possui todos os seus p valores próprios positivos ($\lambda_k > 0$, $k = 1, \dots, p$), ou seja, \mathbf{S} é definida positiva. Caso a característica da matriz inicial \mathbf{X} seja inferior ao

número p de variáveis originais, então os valores próprios de \mathbf{S} são não negativos e pelo menos um é zero, ou seja, \mathbf{S} passa a ser semidefinida positiva (Apêndice C.1.2, Definições C.1.13).

- Os p vetores próprios de \mathbf{S} formam uma base ortonormada do espaço vetorial \mathbb{R}^p , ou seja, são ortogonais entre si e de norma unitária, pelo que $\mathbf{P}^T \mathbf{P} = \mathbf{P} \mathbf{P}^T = \mathbf{I}$, onde \mathbf{I} é a matriz identidade. Portanto \mathbf{P} é uma matriz ortogonal e $\mathbf{P}^T = \mathbf{P}^{-1}$. Por outro lado, podemos estabelecer que $\mathbf{S} = \mathbf{P} \mathbf{L} \mathbf{P}^T = \mathbf{P} \mathbf{L} \mathbf{P}^{-1}$, e deduzir que as matrizes \mathbf{S} e \mathbf{L} são semelhantes (Apêndice C.1.4, Definição C.1.31).
- A soma das variâncias das p CPs é igual à soma das variâncias das p variáveis dos dados iniciais, que por sua vez é igual à soma dos p valores próprios provenientes da matriz de variâncias-covariâncias \mathbf{S} (ou da matriz de correlações \mathbf{R} , no caso de variáveis normalizadas)

$$\sum_{k=1}^p \text{var}(\mathbf{z}_k) = \sum_{k=1}^p \lambda_k = \sum_{j=1}^p \text{var}(\mathbf{x}_j). \quad (2.1)$$

De facto, e designando por $\text{tr}()$ o traço de uma matriz quadrada (ou seja, a soma dos elementos da sua diagonal principal) e tendo em conta a circularidade do traço (Apêndice C.1.3, Proposição C.1.22), tem-se

$$\text{tr}(\mathbf{S}) = \text{tr}(\mathbf{P} \mathbf{L} \mathbf{P}^T) = \text{tr}(\mathbf{L} \mathbf{P}^T \mathbf{P}) = \text{tr}(\mathbf{L}).$$

Desta forma concluímos que a variabilidade total das p variáveis iniciais mantém-se nas p CPs.

- Cada CP \mathbf{z}_k da matriz de variâncias-covariâncias explica uma proporção da variabilidade total dada por

$$\frac{\text{var}(\mathbf{z}_k)}{\sum_{k=1}^p \text{var}(\mathbf{z}_k)} = \frac{\lambda_k}{\text{tr}(\mathbf{S})}.$$

Para a ACP sobre a matriz de correlações \mathbf{R} vem

$$\frac{\text{var}(\mathbf{z}_k)}{\sum_{k=1}^p \text{var}(\mathbf{z}_k)} = \frac{\lambda_k}{\text{tr}(\mathbf{R})} = \frac{\lambda_k}{p}.$$

Se pretendermos reter apenas as primeiras q (com $q < p$) CPs, a percentagem da variância total explicada pelas q primeiras CPs é dada por

$$\frac{\sum_{k=1}^q \text{var}(\mathbf{z}_k)}{\sum_{k=1}^p \text{var}(\mathbf{z}_k)} \times 100\% = \frac{\sum_{j=1}^q \lambda_k}{\sum_{j=1}^p \lambda_k} \times 100\%.$$

- As CPs são não correlacionadas entre si, isto é, a covariância entre CPs é nula

$$\text{cov}(\mathbf{z}_k, \mathbf{z}_l) = 0, \forall k, l = 1, 2, \dots, p \text{ e } k \neq l.$$

- Tendo em conta que $\text{cov}(\mathbf{x}_j, \mathbf{z}_k)$ é o elemento na j -ésima linha e k -ésima coluna da matriz $\text{cov}(\mathbf{X}, \mathbf{Z}) = \mathbf{P}^T \mathbf{S} = \mathbf{L} \mathbf{P}^T$, então a correlação linear entre as variáveis iniciais e as CPs é expressa por

$$\text{corr}(\mathbf{x}_j, \mathbf{z}_k) = \frac{\text{cov}(\mathbf{x}_j, \mathbf{z}_k)}{\sqrt{\text{var}(\mathbf{x}_j)} \sqrt{\text{var}(\mathbf{z}_k)}} = \frac{\lambda_k p_{jk}}{\sqrt{s_j^2} \sqrt{\lambda_k}} = \sqrt{\lambda_k} \frac{p_{jk}}{s_j},$$

onde \mathbf{x}_j representa a j -ésima variável inicial, \mathbf{z}_k representa a k -ésima CP e p_{jk} o coeficiente da j -ésima variável na combinação linear que define a k -ésima CP.

Em termos matriciais, podemos considerar uma matriz $\tilde{\mathbf{P}}$ constituída por colunas iguais a $\sqrt{\lambda_k} \mathbf{p}_k$, tal que a matriz de correlações entre cada variável original e cada CP possa ser expressa por $\mathbf{C} = \mathbf{E}^{-1/2} \mathbf{P} \mathbf{L}^{1/2} = \mathbf{E}^{-1/2} \tilde{\mathbf{P}}$, onde \mathbf{E} é a matriz diagonal de variâncias de cada variável.

Numa ACP sobre a matriz de correlações viria apenas que $\text{corr}(\mathbf{x}_j, \mathbf{z}_k) = \sqrt{\lambda_k} p_{jk}$, pois $\text{var}(\mathbf{x}_j) = 1$, ou seja, $\mathbf{E}^{-1/2} = \mathbf{I}$.

Geralmente, na procura de uma correlação satisfatória entre variáveis e CPs consideram-se apenas os coeficientes dos vetores próprios (*loadings* na terminologia inglesa) mais elevados, em valor absoluto. Estes *loadings* podem constituir um indicador razoável na escolha das variáveis iniciais que mais contribuem para explicar as CPs. No entanto, Cadima e Jolliffe (1995) argumentam que não se devem utilizar apenas os *loadings* para escolher variáveis na explicação de CPs, uma vez que tal decisão pode induzir em erro.

Por vezes há tendência para se ignorar a contribuição de algumas variáveis apenas pelo facto dos seus *loadings* serem próximos de zero na formação da CP. Essa decisão, sem

ter em linha de conta pelo menos a análise das correlações em causa, pode significar uma ilusão. Por outro lado, há ainda que ter alguma precaução, pois podem existir variáveis iniciais cujos *loadings* sejam todos próximos de zero em todas as CPs, o que reflete uma situação em que essas variáveis contribuem muito pouco relativamente a outras para a formação de uma CP.

- A distância euclidiana, d , entre dois vetores $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ é dada por

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

e a distância de Mahalanobis, d_M , entre dois indivíduos (vetores correspondentes às linhas de uma matriz de dados $\mathbf{X}_{n \times p}$), \mathbf{x}^i e \mathbf{x}^j ($i, j = 1, 2, \dots, n$), de um conjunto multivariado com matriz de variâncias-covariâncias \mathbf{S} , é dada por

$$d_M(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}^i - \mathbf{x}^j\|_{\mathbf{S}^{-1}} = \sqrt{(\mathbf{x}^i - \mathbf{x}^j)^T \mathbf{S}^{-1} (\mathbf{x}^i - \mathbf{x}^j)}.$$

Enquanto a distância euclidiana considera contribuição idêntica para todas as variáveis, a distância de Mahalanobis atribui relativamente menos peso às variáveis com variâncias elevadas e para grupos de variáveis altamente correlacionadas (Jolliffe, 2002). A distância de Mahalanobis é igual à distância euclidiana das variáveis transformadas (PCs) em que é dada a mesma importância para todas as componentes (estandardizadas).

- Uma vez que $\mathbf{Z} = \mathbf{X}\mathbf{P}$, então a matriz dos dados iniciais, \mathbf{X} , pode representar-se por $\mathbf{X} = \mathbf{Z}\mathbf{P}^T = \sum_{k=1}^p \mathbf{z}_k \mathbf{p}_k^T$. Se considerarmos apenas as primeiras q CPs para obtermos valores aproximados da matriz dos dados iniciais \mathbf{X} , ou seja, para efetuarmos a reconstrução da matriz \mathbf{X} , utilizamos a matriz dos dados reconstruída $\tilde{\mathbf{X}}_q = \sum_{k=1}^q \mathbf{z}_k \mathbf{p}_k^T$, a qual irá explicar apenas a parte da percentagem da variância total de \mathbf{X} correspondente às q CPs. Quanto maior for o número q de CPs a utilizar na reconstrução, obviamente maior será a aproximação dos valores reconstruídos aos dados originais. Tal constatação pode ser levada a cabo através da minimização da soma dos quadrados dos erros entre cada valor de \mathbf{X} e de $\tilde{\mathbf{X}}_q$, como se poderá verificar no exemplo de aplicação na Secção 3.12.

2.4 Escolha do Número de Componentes Principais

Uma das finalidades da técnica da ACP é reduzir a dimensionalidade dos dados, mantendo a máxima variabilidade existente nos dados iniciais. Como tal, pretende-se considerar um espaço de dimensão inferior a p (dimensão do espaço das p variáveis iniciais), mediante alguns critérios, dos quais destacamos o critério de escolha do número de eixos ou de CPs que deve aparecer numa interpretação. Para substituir as p CPs por um número mais reduzido, q , de CPs, deve tomar-se q de forma a não se perder informação significativa em relação à variabilidade total dos dados, ignorando as restantes $p - q$ CPs.

De entre vários critérios existentes para a escolha de q , um dos mais usuais é conhecido por critério de Pearson e consiste em selecionar o número de eixos de forma que a percentagem total da variação atinja pelo menos os 80%. Nestas condições é costume dizer que estamos perante uma boa representação dos indivíduos e das variáveis. No entanto o valor desta percentagem é uma opção subjetiva que por vezes é tomada no intervalo entre os 70% e os 90%, conforme o número de CPs que se considere ser razoável na contribuição para essa variabilidade. Uma percentagem superior a 90% poderá ser apropriada em casos em que uma ou duas CPs são nitidamente predominantes em termos de variação relativamente às restantes. Em estudos de dados massivos que exigem a análise de milhares de variáveis ou observações, o número de componentes a reter, segundo este critério, poderá ser demasiado grande, o que tornará eventualmente impraticável a interpretação e aplicabilidade das diversas componentes retidas. Alguns autores apresentaram alternativas a essa escolha do valor de q (Jolliffe, 2002).

Determinadas tentativas têm sido efetuadas com vista a encontrar a distribuição da percentagem de variação explicada pelas primeiras q CPs, e consequentemente fornecer procedimentos formais na escolha de q . Alguns autores tais como Mandel (1972) descreveram algumas metodologias e resultados para a distribuição da percentagem em causa e Huang e Tseng (1992) apresentaram um procedimento de decisão acerca do valor de q para uma dada amostra, ao admitirem determinadas condições populacionais. Critérios baseados na soma dos valores próprios correspondentes às CPs que se pretendem desprezar, foram abordados em Jackson (1991).

Outro critério, conhecido por critério ou regra de Kaiser (1960), consiste em reter apenas as CPs cujas variâncias sejam superiores à unidade, ou seja, aquelas que apresentam variabilidade superior à média das variâncias das variáveis estandardizadas originais ($var(\mathbf{z}_k) = \lambda_k > 1$,

$k = 1, \dots, p$), uma vez que este critério foi criado essencialmente com o intuito de ser aplicado na ACP sobre a matriz de correlações, mas pode ser adaptado para matrizes de variâncias-covariâncias, bastando para tal considerar o corte no valor médio dos valores próprios. Jolliffe (1972) propõe, em alternativa, o corte em 70% desse valor médio.

Um outro critério que se baseia numa representação gráfica designada por *scree graph* ou *scree plot*, introduzido em 1966 por Cattell, consiste em observar o gráfico que representa nas abcissas a ordem decrescente dos p valores próprios e nas ordenadas os valores próprios λ_k ($k = 1, \dots, p$). O ponto de corte deve ocorrer no valor em que os declives, dos segmentos de reta que unem os pontos do gráfico, deixam de ser consideravelmente acentuados e passam a ser quase nulos. Os valores próprios que se situem mais à esquerda no gráfico, até imediatamente antes do ponto de quebra que venha a ser estabelecido, indicam quais as respetivas CPs que devem ser retidas. Este critério é ainda mais subjetivo do que os anteriores. A utilização do logaritmo dos valores próprios na representação gráfica constitui uma alternativa ao *scree plot* como por exemplo em Farmer (1971) ou em Hassani (2007).

A descrição de outras regras, para a escolha do número adequado de CPs a reter, podem ser encontradas em Jolliffe (2002), como por exemplo critérios baseados em teste de hipóteses, métodos de validação cruzada, métodos de computação intensiva ou critérios baseados em correlações parciais. Em Ferré (1995) também se apresenta um estudo comparativo que abrange um total de 18 regras.

2.5 Representação Gráfica dos Indivíduos e das Variáveis

Há vantagens em que um conjunto de dados constituído por muitas variáveis possa ser representado num subespaço a duas dimensões, o que permite obter uma representação visual aproximada da nuvem dos n pontos definida pelos dados em \mathbb{R}^p . Ainda é possível conseguir uma visualização dos dados caso estes fiquem bem representados num subespaço a três dimensões. A ACP permite medir a qualidade dessas aproximações. Para obtermos o melhor gráfico dos dados a duas dimensões, há que representar graficamente os valores para cada observação nas duas primeiras CPs.

Em muitas técnicas estatísticas, em certas representações gráficas, geralmente usa-se apenas a informação dos indivíduos, contudo a ACP permite acrescentar uma nova ferramenta de forma a

poder contemplar também as variáveis. Para efetuar uma representação gráfica em simultâneo dos indivíduos e das variáveis num espaço de dimensão reduzida, pode recorrer-se ao uso do *biplot*. Este tipo de gráfico fornece a informação acerca das relações entre os indivíduos e as componentes, e permite interpretar as posições relativas das variáveis e dos indivíduos.

A noção clássica de *biplot* surgiu com Gabriel (1971), o principal responsável pela sua popularização e por grandes desenvolvimentos em torno desse tipo de gráficos tendo como base a decomposição em valores singulares. Substanciais desenvolvimentos e extensões foram surgindo entretanto, nomeadamente a partir dos trabalhos de Gower e Hand (1996), onde se poderão encontrar algumas variantes à abordagem clássica de *biplot*.

De acordo com a representação gráfica correspondente ao número q de CPs a reter, os primeiros q elementos de cada vetor que representarão os correspondentes indivíduos e variáveis (no mesmo sistema de eixos) formam subvetores que se costumam designar por marcadores. Num *biplot*, o cosseno do ângulo entre cada par de marcadores das variáveis é uma aproximação do coeficiente de correlação entre essas variáveis, e os marcadores das variáveis são representados por vetores cujo comprimento é proporcional ao seu desvio padrão. Por outro lado, a distância euclidiana entre cada par de marcadores de indivíduos é aproximadamente a distância de Mahalanobis entre os correspondentes indivíduos, e os marcadores dos indivíduos são representados pelas coordenadas nas novas CPs.

Em termos de interpretação gráfica de um *biplot*, para a ACP sobre a matriz das correlações, quanto maior for o comprimento dos marcadores das variáveis (e numa ACP sobre a matriz de variâncias-covariâncias, quanto maior for o comprimento dos marcadores das variáveis, relativamente ao verdadeiro comprimento da totalidade das p variáveis) melhor será a sua representação, e quanto menor for a distância destas a um determinado eixo melhor é a sua explicação em relação a esse eixo. Por outro lado, quanto mais afastados da origem estiverem os indivíduos, maior será a sua contribuição para um dado eixo quanto mais se aproximarem deste, em detrimento do outro eixo.

2.6 Breve História da ACP

Em 1877 Francis Galton introduzira o coeficiente de correlação como medida de relação entre duas variáveis, época em que se começava a descrever numa forma muito simples, um método

prático para executar a ACP. O desenvolvimento empenhado dos métodos estatísticos envolvendo várias variáveis teve o seu início com o trabalho de Bumpus (1898).

A técnica de análise de dados multivariados ACP foi descrita pela primeira vez em 1901 por Pearson, mas sem essa designação. Durante 32 anos desde os artigos de Pearson até aos de Hotelling pouco material relevante fora publicado, apesar de Rao (1964) ter referido que Frisch (1929) seguiu uma abordagem semelhante à de Pearson. Também o artigo de Hotelling (1933) sugere que Thurstone (1931) tinha trabalhado em linhas semelhantes às de Hotelling (Jolliffe, 2002).

Hotelling introduziu o termo "componentes", distinto do conceito "fatores" usado na Psicologia, para descrever as variáveis de um conjunto de dados. Os "fatores" estão na origem da Análise Fatorial, na qual se considera um determinado número de variáveis observadas que podem ser expressas, a menos de um termo de erro, como funções lineares de um número menor de variáveis latentes, isto é, não observáveis, as quais também são designadas de fatores comuns. Tal como a ACP, esta técnica tem o objetivo de reduzir a dimensionalidade de um conjunto de dados, usando um número de fatores menor que o número de variáveis observadas (Jolliffe, 2002).

Por conseguinte, Hotelling descrevera as "componentes" tendo em vista a maximização das contribuições sucessivas para o total das variâncias das variáveis originais, e designou-as de "componentes principais" (CPs). Surge assim a análise que tem por base estas componentes, a qual foi batizada de "método das componentes principais". O método proposto por Hotelling tinha como principal objetivo analisar a correlação, olhando para as variáveis originais expressas como funções lineares das componentes e sem usar uma notação matricial. Um outro artigo descrito por Hotelling em 1936 apresentava uma nova versão onde se mencionava um maior poder do método para encontrar CPs, e nesse mesmo ano, Girshick (1936) forneceu algumas derivações alternativas das mesmas.

Nos 25 anos imediatamente a seguir à publicação do artigo de Hotelling de 1933, pouco foi o trabalho que surgiu referente ao desenvolvimento de diferentes aplicações da ACP. Desde então, emergiu uma grande quantidade de novas aplicações e outros desenvolvimentos teóricos. Esta expansão refletia o crescimento geral da estatística, acompanhado pela introdução de computadores.

A década de sessenta é a época em que a análise de dados evoluiu verdadeiramente, incluindo

a ACP que se tornou uma técnica utilizada em quase todas as áreas do conhecimento humano, desde as ciências da terra e da vida até às ciências económicas e sociais. Esta evolução é devida ao trabalho da escola francesa, em particular à equipa de Benzécri. Na escola anglo-saxónica temos como referência Anderson (1963) que desenvolveu tal como Girshick (1939) a amostragem assintótica das distribuições dos coeficientes e variâncias das CPs; Rao (1964) que surgiu com novas ideias, interpretações e extensões da ACP; Gower (1966) que estabeleceu algumas ligações entre a ACP e outras técnicas, e que apresentou critérios geométricos importantes e Jeffers (1967) que aprofundou o uso da ACP indo mais longe do que uma simples redução de dimensão do conjunto das variáveis.

Durante a década de setenta e oitenta, a investigação avançou essencialmente com o desenvolvimento de métodos para o tratamento de tabelas multidimensionais (Robert e Escoufier, 1976, Escoufier, 1986). São também de salientar ainda nomes tais como Morrison (1976) ou Harman (1976), que tiveram uma certa importância na publicação de obras onde algumas técnicas da análise de dados, entre as quais a ACP, são abordadas. É ainda relevante a coleção de artigos editada por Bryant e Atchley em 1975, nos quais aparece o trabalho de Pearson e parte da obra de Hotelling (Jolliffe, 2002).

A partir da década de 90, a utilização de métodos computacionais tem sido praticamente indispensável para tratar com maior eficácia e rapidez um grande volume de dados. Inúmeras publicações relacionadas com a ACP continuam ainda hoje a apresentar novos desenvolvimentos e aplicações.

Para se obter um conhecimento mais alargado sobre o desenvolvimento da técnica da ACP, podem ser consultados os livros de Jolliffe (2002) e Jackson (1991). Outras obras de significativa importância na área das ciências atmosféricas tais como Preisendorf e Mobley (1988), Wilks (2006) e von Storch e Zwiers (1999) apresentam exemplos da utilização da ACP em muitos campos, estabelecendo ligações e extensões a outras técnicas multivariadas.

2.7 Para Além da ACP Clássica - Outras Abordagens

Nesta Secção descrevem-se brevemente determinadas variações ou técnicas que derivam ou se relacionam com a ACP clássica, encaradas por vezes como novas abordagens no tratamento de dados multivariados.

2.7.1 Rotação das Componentes

Ao efetuar-se uma ACP usual ou outra técnica de projeção dos dados (por exemplo a Análise Fatorial) em determinadas componentes previamente fixadas, por vezes são utilizadas técnicas de rotação, nomeadamente nos casos em que essas componentes não são de fácil interpretação.

Quando uma dada CP apresenta todos os *loadings* de uma mesma ordem de grandeza, ou alguns muito grandes e os restantes muito pequenos, parece ser de fácil interpretação. Contudo, é necessário ter cuidado pois podem surgir situações enganadoras. Por exemplo, se existirem *loadings* com valores intermédios, a interpretação dessa CP torna-se geralmente mais complicada, e num caso como este a rotação das CPs envolvidas tem sido sugerida como forma de facilitar a interpretação através da alteração dos *loadings* nas combinações lineares das variáveis iniciais.

Muitos foram os autores que estudaram o problema relacionado com a aplicação da rotação das CPs. Entre eles destacamos Richman (1986, 1987), Jolliffe (1987) e Rencher (1995), os quais contribuíram com algumas vantagens e desvantagens da utilização da rotação das CPs essencialmente nas ciências atmosféricas onde esta técnica desempenha um papel mais relevante.

Para se efetuar uma rotação das componentes, com vista a simplificar os seus *loadings*, existem alguns métodos chamados ortogonais, tais como o *Varimax* (Horel, 1981), o *Quartimax* e o *Equamax* (Gorsuch, 1983), ou métodos chamados oblíquos, como o *Direct Oblimin* e o *Oblimax* (Cohen, 1983) ou o *Promax* (Gorsuch, 1983). Listas mais completas de possíveis métodos encontram-se em Cattell (1978) e Richman (1986).

Para além da decisão sobre qual o método a escolher para a rotação das componentes, que possa permitir uma interpretação mais simples destas, outro inconveniente prende-se com o facto de que, embora a variabilidade total se mantenha no subespaço obtido após rotação, esta é redistribuída por entre as componentes rodadas de uma forma mais uniforme do que é geralmente antes da rotação, fazendo com que a informação acerca da natureza de qualquer componente dominante possa ser perdida. Também o número de CPs a reter pode ser uma preocupação uma vez que diferentes escolhas produzem certamente diferentes efeitos nos resultados depois da rotação. Para além disso, a escolha da restrição no que se refere ao tipo de normalização a aplicar aos vetores de *loadings* (vetores próprios) provoca uma mudança nas propriedades dos *loadings* após rotação (Jolliffe, 2002).

Kiers (1993) preocupou-se mais em realçar a simplicidade da solução do método de rotação

do que na própria maximização da variância das CPs, discutindo e comparando para tal quatro técnicas. Um outro método conhecido por SCoT (*Simplified Component Technique*) é sugerido por Jolliffe e Uddin (2000) onde sucessivas combinações lineares das variáveis são encontradas, de forma a maximizar um critério no qual a variância passa a ser combinada com uma função que força a combinação linear em direção à simplicidade.

Em Jolliffe (1989, 1995, 2002) discute-se os efeitos de diferentes formas de normalização dos vetores próprios e os problemas relacionados com a sua interpretação, assim como outros aspetos ligados à rotação.

Em Climatologia, onde a ACP é geralmente conhecida por Análise de Funções Ortogonais Empíricas (*Empirical Orthogonal Function (EOF) Analysis*), é frequente o uso da rotação de componentes (von Storch e Zwiers, 1999). Neste contexto os vetores próprios denominam-se por funções ortogonais empíricas (EOFs) e os *scores* por coeficientes EOF. Nesta área é comum encontrarmos observações relacionadas ao longo do tempo, uma vez que é usual recorrer-se ao estudo de observações por períodos temporais, como por exemplo nas recolhas em estações meteorológicas. Neste sentido a aplicação de EOFs ou a rotação de EOFs deve ser realizada com precaução pois a sua interpretação pode sofrer a influência das correlações existentes entre observações ao longo do tempo. A questão da aplicabilidade da rotação não é consensual, mas na área do clima uma grande parte dos investigadores defende o uso da mesma justificando que esta permite obter padrões espaciais estatisticamente interpretáveis (von Storch e Zwiers, 1999). Vários são os argumentos que apoiam a utilização da rotação das EOFs dos quais destacamos o facto da rotação produzir padrões compactos que podem ser utilizados para dividir uma área num número limitado de subáreas ou regiões homogêneas e o facto das EOFs após rotação serem menos sensíveis à distribuição das localizações observadas e geralmente estatisticamente mais estáveis do que as tradicionais EOFs.

2.7.2 ACP Ponderada

Um outro tipo de generalização muito utilizada é a ACP Ponderada (*Weighted Principal Component Analysis (WPCA)*), onde as diferentes variáveis e/ou indivíduos podem ter diferentes pesos. Aliás, a ACP sobre a matriz de correlações pode ser encarada como um caso particular da ACP Ponderada, em que a cada uma das variáveis é atribuído um peso correspondente ao inverso da variância da respetiva variável.

De seguida fazemos uma pequena referência a alguns autores que se debruçaram sobre a questão da ponderação.

- Rao (1964) e Gower (1966) foram dos primeiros a considerar a ideia de associar diferentes pesos a um conjunto de dados. Gower (1966) sugeriu a possibilidade de dividir as variáveis pelas suas amplitudes em vez dos desvios padrão, ou pelas suas médias ou mesmo pela raiz cúbica do terceiro momento amostral.
- Cochran e Horne (1977) aplicaram uma ACP Ponderada na área da Química no estudo de comprimentos de onda em experiências cinéticas, a fim de eliminar determinadas estimativas incorretas do número de espécies durante uma reação.
- Yue e Tomoyasu (2004) discutiram alguns aspetos práticos encontrados na manufatura de semicondutores ao aplicar a ACP à deteção e classificação de falsos alarmes. Propuseram uma ACP Ponderada em que os pesos foram utilizados de duas formas distintas: nas observações das amostras e nas variáveis. Os dados das amostras foram ponderados de acordo com a manutenção da técnica da ACP em termos temporais, a qual fora adaptada consoante as mudanças do processo permitindo melhoramentos da qualidade da sua aplicação ao longo do tempo. Por outro lado, para as variáveis formularam um método para incorporar ferramentas e conhecimento do processo na construção da técnica, de forma que a cada variável pudesse ser atribuído um fator de ponderação diferente. O estudo comparativo entre a ACP usual e a ACP Ponderada permitiu mostrar que esta última reduz a ocorrência de falsos alarmes, evidencia menos sensibilidade a variáveis com pouca variação na análise dos dados.
- Forbes e Fiume (2005) apresentaram um algoritmo de procura baseado na ACP Ponderada para dados do movimento, no qual introduziram uma métrica para a distância entre posturas. Foi demonstrada a utilização do algoritmo em dados reais e simulados e comparado o seu desempenho.
- Pinto da Costa *et al.* (2011) propõem um novo coeficiente de correlação como alternativa ao de Pearson o que leva como consequência à utilização da ACP Ponderada. Os autores consideram o problema que consiste em analisar conjuntos de dados de expressões genéticas, e para tal, é proposto um novo algoritmo iterativo baseado na ACP para selecionar

os genes mais importantes em *microarrays*, o qual produz geralmente melhores resultados quando a ACP Ponderada é usada em vez da ACP usual.

- Nas ciências atmosféricas, onde geralmente as variáveis (colunas da matriz) são as localizações espaciais, por vezes os pesos são fundamentais para colmatar as irregularidades de espaçamento existentes entre as localizações onde as observações são recolhidas. Os pesos refletem o tamanho da área onde se enquadra a observação. Quando as localizações estão espaçadas de uma forma regular segundo uma grelha ao longo da latitude e longitude, os pesos poderão também desempenhar um papel essencial. Nestes casos, as áreas das células da grelha decrescem à medida que nos aproximamos dos polos. Uma estratégia a adotar é atribuir às células da grelha pesos proporcionais às suas áreas. Por exemplo, Hannachi *et al.* (2009) utilizaram como pesos na análise dos seus dados climáticos as raízes dos cossenos da latitude.

2.7.3 ACP Centrada *versus* ACP Não Centrada

A ACP pode ser encarada como uma decomposição em valores singulares de uma determinada matriz. É prática corrente que essa matriz seja uma matriz de dados centrada (por colunas), ou seja, a cada observação subtrai-se a média das observações dessa mesma variável. Contudo, uma outra variante é a aplicação da decomposição em valores singulares a uma matriz de dados não centrada, e nestes casos dizemos que estamos perante uma ACP não centrada.

A ACP não centrada é uma técnica aplicada em Ecologia (ter Braak, 1983). Alguns dos exemplos em dados de contagens dos indivíduos para diferentes espécies, as quais representam as variáveis observadas em vários locais, indicam que a ACP não centrada é melhor que a ACP centrada para realçar a diversidade entre espécies dentro de cada local e entre locais. Também Reyment e Jöreskog (1993) aplicaram a ACP não centrada na abundância de vários microrganismos marinhos e referem que esta análise é relevante para o estudo de quantidades de constituintes químicos em amostras de solo ou rocha. A ACP não centrada é utilizada em várias outras áreas como por exemplo na Química (Cochran e Horne, 1977), na Climatologia (van den Dool *et al.*, 2000; van den Dool, 2007), no estudo de *microarrays* (Alter e Golub, 2006) entre outras e é aplicada nomeadamente em casos onde se pretende preservar a estrutura subjacente às colunas.

Cadima e Jolliffe (2009) analisaram extensivamente o paralelismo entre a ACP centrada por colunas e a ACP não centrada, apresentando os resultados referentes aos valores e vetores próprios das matrizes de variação conjunta e as componentes principais obtidas para cada um dos dois procedimentos. É destacado que os valores próprios desempenham um papel fundamental, por conterem muita da informação relevante no que diz respeito a uma comparação entre a ACP usual e a ACP não centrada, onde as variáveis iniciais são sujeitas a análises de diferentes formas. Em relação à escolha da combinação linear das variáveis, por um lado esta é efetuada através da maximização dos segundos momentos centrados (variâncias), por outro recorre-se à maximização dos segundos momentos não centrados. A ACP centrada por colunas obriga a que as componentes principais sejam não correlacionadas entre si, enquanto que a ACP não centrada exige que os segundos momentos não centrados cruzados sejam nulos. Os resultados apresentados quer a nível teórico quer através de exemplos, demonstraram existir mais semelhanças entre ambos os procedimentos do que seria de esperar.

Alguns autores argumentam que a centragem usual pode, em determinados casos particulares, levar à destruição dos principais padrões existentes nas linhas da matriz dos dados. De entre alguns casos relacionados com este assunto, tome-se como exemplo uma outra variante da ACP que é por vezes utilizada e que consiste na decomposição em valores singulares de matrizes centradas por linhas ou mesmo matrizes que são centradas em simultâneo por colunas e por linhas (Cangelosi e Goriely, 2007). A centragem por linhas é plausível em determinadas situações, nomeadamente quando as linhas da matriz também podem ser encaradas como variáveis ou quando as linhas correspondem a indivíduos e se deseja filtrar alguma medida relacionada com o tamanho global dos indivíduos (Cadima e Jolliffe, 2009).

Buckland e Anderson (1985) propuseram a dupla centragem no contexto ecológico, com dados de contagens de espécies em vários locais. Eles argumentavam que a centragem por linhas e colunas removia o efeito da abundância de cada espécie, e que aparecia na ACP clássica como CP dominante. Também Somers (1986) em estudos da relação entre o tamanho e a forma durante o processo de crescimento de organismos, sugeriu remover o efeito da componente relacionada com o tamanho, ficando apenas com aquelas que descreviam a forma. No entanto, esse procedimento acarreta alguns inconvenientes, nomeadamente a perda da ortogonalidade entre os vetores próprios e a correlação nula entre as componentes. Uma abordagem alternativa descrita por Somers (1989) para preservar a propriedade da ortogonalidade foi a dupla centragem

(por colunas e linhas), embora continuasse a verificar-se correlação entre as componentes. Para ultrapassar alguns inconvenientes dos métodos apresentados, assim como as desvantagens da dupla centragem, Cadima e Jolliffe (1996) propuseram uma nova modificação.

2.7.4 ACP para Dados Categóricos (CATPCA)

Embora a ACP usual apenas faça sentido para variáveis quantitativas, é comum encontrarem-se na literatura estudos da ACP para dados qualitativos de uma forma abusiva, uma vez que a ACP usual foi desenvolvida sobre a matriz de covariâncias ou de correlações, as quais são constituídas por quantidades que são características unicamente de variáveis quantitativas. Por exemplo, em numerosos estudos de ciências sociais e em estudos de mercado é frequente surgirem muitas variáveis numa escala nominal ou ordinal.

Enquadrada na designada ACP "não linear" (onde não se utilizam funções lineares para maximizar variâncias), uma das extensões da ACP mais apropriadas quando o objetivo consiste em reduzir a dimensionalidade de variáveis medidas em escalas diferentes num número reduzido, de componentes, que expliquem uma proporção razoável da variabilidade existente nos dados é a ACP para dados categóricos (variáveis nominais e ordinais), conhecida como *Categorical Principal Component Analysis* (CATPCA), a qual está preparada para estruturas não-lineares.

A abordagem descrita em Gifi (1990), desenvolvida inicialmente por investigadores holandeses da Universidade de Leiden, criada para variáveis categóricas, baseia-se num procedimento conhecido por escalonamento ótimo (*optimal scaling*), onde se atribuem quantificações numéricas às categorias de cada uma das variáveis qualitativas, possibilitando assim o recurso a métodos clássicos de análise numérica (Meulman, 1992). Após a quantificação das categorias das variáveis que constituem a matriz dos dados iniciais, obtém-se a matriz dos dados $\mathbf{X}_{n \times p}$, a partir da qual a técnica utiliza o critério de minimização de

$$\text{tr} \{(\mathbf{X} - \mathbf{YB}^T)^T(\mathbf{X} - \mathbf{YB}^T)\}$$

em relação à matriz $\mathbf{Y}_{n \times q}$ cujas colunas são funções lineares das colunas de \mathbf{X} , e em relação à matriz $\mathbf{B}_{q \times p}^T$ obtida a partir de \mathbf{B} cujas colunas são ortogonais e representam os vetores próprios da matriz de variâncias-covariâncias de \mathbf{Y} . Nestas condições a matriz ótima \mathbf{Y} é constituída pelas coordenadas (*scores*) das primeiras q CPs e a matriz ótima \mathbf{B} é constituída pelos coeficientes (*loadings*) das primeiras q CPs.

Os valores numéricos que se atribuem a cada uma das classes das variáveis originais, incluindo as quantitativas, são geralmente obtidos através de um algoritmo iterativo denominado por "método dos mínimos quadrados alternantes". Tal algoritmo foi implementado no programa computacional denominado PRINCALS (*Principal Component Analysis by means of Alternating Least Squares*) (Gifi, 1990) e incorporado no *software* SPSS, de forma que essas quantificações numéricas sejam dotadas de propriedades métricas. Os valores que resultam desta transformação de escala das variáveis são estandardizados para posterior aplicação da técnica CATPCA. Mais pormenores desta abordagem podem ser aprofundados em Lavado (2012).

Outra abordagem que coincide parcialmente com as ideias principais de Gifi relacionadas com a ACP não linear é sugerida por Meulman (1986), onde os dados categóricos são também transformados, com vista a fornecerem valores ótimos para cada categoria de cada variável, e um pequeno número de dimensões ótimas é encontrado de forma a representar esses mesmos valores. Para as variáveis quantitativas, a sua amplitude de valores é dividida num número finito de categorias (ou classes) e posteriormente atribuído um valor a cada categoria (o que vai envolver uma perda natural de informação mediante o agrupamento dos valores em classes). Esta proposta é conhecida como abordagem de distância para análise de dados multivariados não lineares e difere das ideias principais de Gifi na utilização de diferentes critérios envolvendo otimizações.

O escalonamento ótimo que permite produzir associações entre variáveis categóricas (análise de correspondência para duas ou mais variáveis) engloba vários procedimentos. Para além do referido PRINCALS, são geralmente mencionados outros procedimentos tais como ANACOR (*Correspondence Analysis*), HOMALS (*Homogeneity Analysis*) e OVERALS (*Nonlinear Canonical Correlation Analysis*). Todos estes procedimentos têm em comum a análise da relação de homogeneidade entre as variáveis, assim como a intenção de reduzir a dimensionalidade dos dados. Em particular o procedimento HOMALS, também conhecido por Análise de Correspondências Múltiplas ou Análise de Homogeneidade, permite analisar as correspondências subjacentes, na presença de mais de duas variáveis qualitativas onde a ordem implícita das categorias, mesmo que exista, não é tida em conta na análise. No caso particular em que todas as variáveis em estudo são escalonadas como numéricas, a CATPCA é equivalente à ACP usual (Maroco, 2007).

2.7.5 ACP Log-contraste

Por vezes podemos transformar os dados através de procedimentos de pré-processamento antes de aplicar a técnica da ACP. Para além da usual centragem por colunas ou da estandardização, é frequente encontrarmos outras aplicações como por exemplo a ACP Log-contraste (*Log-contrast PCA*) em que os dados iniciais são logaritmizados. No caso particular da análise de dados composicionais (Aitchison, 1986) é praticamente imprescindível o uso de logaritmos nos rácios. Considerem-se de seguida algumas referências em relação à ACP Log-contraste.

- Allen e Connor (2000) utilizaram a ACP Log-contraste num estudo ecológico sobre fatores estruturantes de grupos de aves que habitam as margens de uma série de lagos no nordeste dos EUA. Nesse estudo foram alcançados dois objetivos: identificar as relações dominantes entre os grupos funcionais nas composições de forragem e dieta, e isolar essas relações em eixos de variação independentes. Com esta técnica, os dados brutos organizados em termos de proporções foram inicialmente transformados utilizando logaritmos.
- Yu (2003) estudou a dinâmica, do final do Quaternário, da vegetação da floresta e da tundra no sul da Escarpa do Niágara no Canadá, tendo resumido as mudanças de vegetação através de uma ACP Log-contraste do registo de pólen. Para além disso foi utilizada também uma dupla centragem quer nas amostras quer nos diferentes tipos de pólen, a qual conjugada com a transformação através dos logaritmos fornece, segundo o autor, uma melhor representação da "distância ecológica" real entre amostras.
- Procedimentos muito semelhantes aos de Yu (2003) foram aplicados por Zhao *et al.* (2007) para estudar dados de vários tipos de pólen fóssil em amostras no Lago Hurleg no Planalto Tibetano, no noroeste da China, onde foram efetuadas novamente uma log-transformação e uma dupla centragem. São apresentadas interpretações climáticas a partir de um núcleo do sedimento e são analisadas as reconstruções da vegetação e do clima do Holoceno.
- Rodrigues e Lima (2009) apresentaram um estudo comparativo de três abordagens distintas, analisando os resultados das eleições do Parlamento Europeu em 2004, de onde se destacou a ACP Log-contraste.

2.7.6 ACP em Três Modos

Um outro exemplo de uma extensão da ACP a outros tipos de dados é a ACP em Três Modos (*Three-Mode Principal Component Analysis*), onde é adicionado um terceiro "modo" ou "camada" aos dados, para além dos indivíduos (linhas) e das variáveis (colunas). Um dos modos que geralmente mais se utiliza é o tempo, permitindo que os dados sejam observados em diferentes períodos temporais. Portanto, podemos considerar n indivíduos, p variáveis e t momentos temporais de forma a que indivíduos, variáveis e pontos definam os três modos. Neste caso temos n séries temporais de p variáveis ou, alternadamente, uma única série temporal de np variáveis.

Têm sido vários os investigadores a debruçar-se sobre a técnica, incluindo Kroonenberg (1983) que é considerado um dos principais. Várias extensões do modelo em três modos têm sido fruto de outras investigações, nomeadamente Timmerman e Kiers (2000) que analisaram as propriedades do modelo e a sua aplicabilidade a algumas situações especiais. Uma dessas extensões, que tem vindo a ser utilizada na psicometria e atualmente também na quimiometria, leva-nos a considerar mais do que três modos, cujos dados são geralmente denominados de "multiforme" em vez de "multimodal".

Também nas ciências atmosféricas o uso da ACP em três modos tem vindo a ganhar alguma relevância. Frequentemente um modo é a localização espacial, um segundo é o tempo e um terceiro é um conjunto de diferentes variáveis meteorológicas. Neste tipo de abordagem, em que os dados das p localizações e n variáveis meteorológicas se combinam em np variáveis medidas ao longo do tempo, estamos no contexto da Análise de Funções Ortogonais Empíricas Estendidas (*Extended EOF Analysis*).

2.7.7 ACP para Séries Temporais

Quando se está perante uma ACP clássica, com interesse nas estruturas de covariâncias dos dados, não existem preocupações em averiguar a possibilidade das observações estarem correlacionadas. Mesmo que se troquem linhas entre si na matriz dos dados, os resultados a obter não serão alterados. Embora, na maioria das aplicações, a ACP seja usada apenas em termos descritivos, onde a não independência não afeta a análise dos dados, para casos em que determinados tipos de dados possuem uma certa dependência entre as observações, alguns pro-

cedimentos usuais em termos de inferência são inválidos. Uma das abordagens da ACP com especial interesse é a extensão da ACP a dados de séries temporais (o tipo mais comum de dados não independentes), cujas observações não podem trocar entre si pela própria natureza dos dados.

Atualmente a ACP é utilizada em muitas áreas do conhecimento para estudar séries temporais multivariadas e dados que não apresentam independência. Noções gerais acerca de séries temporais podem ser encontradas, por exemplo, em Brillinger (1981) e Hamilton (1994).

Nas ciências atmosféricas, a ACP é sem dúvida uma das grandes ferramentas que permite compreender melhor o sistema climático recorrendo a dados de séries temporais (von Storch e Zwiers, 1999). Nesta área, é frequente obtermos dados de séries temporais em diferentes localizações espaciais que são analisadas como variáveis enquanto que as observações correspondem a diferentes períodos temporais.

Para além do estudo ao longo do tempo na procura de correlações entre sucessivas observações, as séries temporais também podem ser analisadas no domínio da frequência, e neste sentido surgiu a ACP com base na análise dos espectros da série temporal decomposta em oscilações de diferentes frequências.

A Análise de Funções Ortogonais Empíricas de Hilbert (*Hilbert EOF Analysis*), por vezes também designada por Transformada de Karhunen-Loève da Frequência, é uma extensão da ACP para dados com variáveis complexas, que utiliza a informação do espectro e do comportamento do processo ao longo do tempo, a qual é obtida segundo a matriz de covariâncias relacionada com o espectro cruzado (von Storch e Zwiers, 1999). A Análise de Funções Ortogonais Empíricas de Hilbert é portanto considerada um caso especial de ACP Complexa, onde a parte imaginária é um processo que visa explicar a evolução temporal do processo original.

Outras técnicas para analisar conjuntos de séries temporais, tais como a Análise Espectral Singular (*Singular Spectrum Analysis* (SSA)) e a Análise Espectral Singular Multicanal (*Multichannel Singular Spectrum Analysis* (MSSA)) (Golyandina *et al.*, 2001), serão abordadas em Capítulos seguintes. A SSA baseia-se na matriz de autocovariância e é encarada como um caso particular da ACP para uma única série temporal desfasada ao longo do tempo, enquanto a MSSA é a ACP da matriz estendida das séries temporais iniciais, ou seja com as séries originais desfasadas, motivo pela qual também é designada no contexto climatológico por Análise de Funções Ortogonais Empíricas Estendidas (EEOF).

Uma outra técnica semelhante às descritas anteriormente, também muito usada em dados climáticos, é a Análise dos Padrões de Oscilação Principal (*POP Analysis*). Na POP Análise é efetuada a decomposição em valores e vetores próprios da matriz que resulta do produto matricial entre a matriz de covariâncias das variáveis desfasadas a uma unidade no tempo e a inversa da matriz usual de covariâncias amostrais das variáveis iniciais (para mais detalhes veja-se, por exemplo, Hasselman, 1988 e von Storch e Zwiers, 1999, Oliveira, 2003). Mais conceitos relacionados com o estudo de séries temporais no domínio da frequência podem ser obtidos ainda em Murteira *et al.* (1993), assim como no Apêndice C.3.

Por vezes ocorrem problemas na estimação das variâncias-covariâncias porque a amostra não é suficientemente elevada para fornecer bons resultados, ou então porque existem valores em falta (não observados) ou porque os dados não foram recolhidos de forma igualmente espaçada no tempo. Uma técnica que pode resolver alguns destes problemas é a Análise em Componentes Principais Funcional (*Functional Principal Component Analysis*, FPCA), a qual é uma alternativa adequada a aplicações em que os dados são descritos por curvas (funções) (Ramsay e Silverman, 1997). Nestes casos, a matriz dos dados é constituída por séries temporais que são encaradas como curvas e consideradas como diferentes observações, e como tal a estrutura da variância-covariância (ou correlação) é analisada do ponto de vista temporal. Alguns exemplos são o registo contínuo de variáveis meteorológicas tais como a pressão ou a temperatura, dados obtidos por espectroscopia, a velocidade de um atleta durante uma prova, ou a trajetória de mergulho de animais marinhos (Oliveira, 2003).

Capítulo 3

Análise em Componentes Independentes

"Só existem dois dias no ano em que não podemos fazer nada.

Um chama-se ontem e o outro chama-se amanhã."

Dalai Lama, Tenzin Gyatso

3.1 Introdução

A Análise em Componentes Independentes é uma técnica estatística e computacional relativamente recente, que surgiu na década de 80 do século XX, tendo sido introduzida por Hérault e Ans (1984), Hérault *et al.* (1985) e Ans *et al.* (1985) para estudarem problemas relacionados com a decodificação de mensagens em circuitos neuronais, e mais tarde por uma formulação mais clara por Comon (1994).

O objetivo principal deste método é encontrar componentes ou fatores "escondidos" relacionados com conjuntos de variáveis aleatórias, medições ou sinais. No modelo básico da ACI admite-se que as variáveis são combinações (também designadas por misturas) lineares de variáveis latentes desconhecidas, isto é, que não podem ser diretamente observadas, as quais são denominadas de componentes independentes e abreviadas por CIs (por vezes também designadas de fatores, fontes ou sinais, conforme a área de aplicação dos dados).

Este método distingue-se de outros dentro do género (como a ACP, a Perseguição da Projecção

ou a Análise Fatorial) ao admitir que as componentes são estatisticamente independentes e que não seguem uma distribuição Normal. A técnica da ACI está intrinsecamente relacionada com a ACP, contudo é com a Análise Fatorial que apresenta mais afinidades, apesar desta exigir a normalidade dos dados. Para além do modelo básico da ACI, a partir dos meados da década de 90 ocorreu um vasto crescimento de métodos e aplicações da técnica. Segundo Hyvärinen e Oja (2000) e Hyvärinen *et al.* (2001b), podemos constatar que geralmente, a ACI é uma técnica que deteta mais facilmente os tais fatores "escondidos" quando estes dois métodos clássicos não permitem obter soluções satisfatórias ou quando a Análise Fatorial não verifica sequer as condições de aplicabilidade.

Ao longo deste Capítulo iremos apresentar a evolução histórica da técnica da ACI e a motivação do aparecimento da mesma, efetuar a descrição do modelo básico de ACI e mencionar alguns dos métodos de estimação do modelo mais usuais, entre os quais destacamos aqueles que utilizam uma estrutura temporal para o caso em que se admite que as CIs são séries temporais. Iremos salientar alguns algoritmos existentes para a implementação da técnica, identificar alguns dos métodos de ordenação das CIs mais utilizados, inclusive propor um novo, e abordar a questão da escolha do número de CIs a reter. Serão também descritas as principais áreas de aplicação da técnica e mencionados outros modelos sem ser o modelo básico da ACI, como por exemplo a ACI não linear e a ACI com restrição. Finalmente será ilustrado um exemplo de aplicação em séries temporais de dados de clima, cuja análise será efetuada com recurso à implementação de várias rotinas, umas criadas e outras já existentes, no *software R*.

3.2 Breve História da ACI

Foram os trabalhos de Hérault e Ans (1984), Hérault *et al.* (1985) e Ans *et al.* (1985) que estiveram na origem da designada separação de fontes (atualmente conhecida por separação cega de fontes, a qual pode ser implementada pela ACI), nos quais se modelou o problema biológico da codificação do movimento ao nível neurológico. Estes resultados foram recebidos com grande entusiasmo pela comunidade de investigadores destas áreas aplicadas, tendo-se verificado o crescimento de investigação e de artigos publicados a partir do final da década de 80 na área de processamento de sinal.

Lacoume e Ruiz (1988) propuseram o uso de cumulantes de quarta ordem para realçar a

suposição da independência e permitir a separação das fontes, através do desenvolvimento de uma abordagem da função de máxima verosimilhança. Deste modo, a discriminação entre os conceitos de separação cega de fontes e análise em componentes independentes apenas viria a surgir em 1987.

Cardoso (1989a) e Comon (1989) foram dos principais responsáveis pelo avanço da técnica da ACI, ao utilizarem a análise espectral de ordem superior. Posteriormente Comon (1994) formalizou e consolidou os aspetos essenciais da técnica.

Alguns algoritmos para aplicação da ACI começaram então a aparecer como o FOBI (Cardoso, 1989b) ou o JADE (Cardoso e Souloumiac, 1993). O trabalho de outros investigadores começa a ganhar realce com a obtenção de novos algoritmos, como por exemplo os de Cichocki e Moczczynski (1992) e Cichocki *et al.* (1994). A abordagem da ACI baseada no princípio da maximização da informação (princípio do infomax), introduzida por Bell e Sejnowski (1995a, 1995b), despertou grande interesse e o algoritmo do gradiente permitiu dar um novo impulso ao desenvolvimento da técnica, resultando no algoritmo do gradiente natural (Amari *et al.*, 1996). Os trabalhos de Perlmutter e Parra (1996) e de Cardoso (1997) permitiram ainda estabelecer que o princípio do infomax é equivalente à estimação de máxima verosimilhança.

Em 1997, Hyvärinen e Oja apresentaram o algoritmo do ponto-fixa FastICA, o qual contribuiu para a aplicação da ACI em larga escala devido à sua eficiência computacional e em Hyvärinen (1997, 1999a) descrevem-se as versões melhoradas. Em 2000, Jutten e Taleb abordam a evolução da técnica em termos históricos.

No início do novo milénio, o número de artigos relacionados com a ACI continuava a aumentar, tendo surgido novas abordagens da técnica. Os investigadores exploram novos modelos para além do modelo básico de ACI, como a ACI não linear (Jutten e Taleb, 2000, Hyvärinen *et al.*, 2001b), a ACI com restrição (Lu e Rajapakse, 2003, 2005) ou a ACI para números complexos (Ollila *et al.*, 2008) e para dados binários (Himberg e Hyvärinen, 2001) entre outras. O leque de áreas de aplicação começa a diversificar, e para além do típico processamento de sinais, e do processamento de imagens, outros campos ganham destaque, como o uso de indicadores biomédicos ou as telecomunicações e as aplicações financeiras. Outros estudos mais recentes ocorrem em áreas tão diversas, desde a análise de dados resultantes das expressões dos genes, da monitorização sísmica até aos processos industriais ou à interpretação de dados de clima.

Investigação Portuguesa em ACI

O número de trabalhos de Mestrado em Portugal envolvendo a técnica da ACI começa a ter alguma expressão, nomeadamente nos campos de processamento de imagens (Ferreira, 2003) e do tratamento de indicadores biomédicos (Pereira, 2003, Guerreiro, 2007, Penedo, 2012, Pereira, 2012). Os trabalhos de Doutoramento em Portugal que abordem a ACI são ainda em número muito reduzido, de onde se destaca Puga (2000).

Em relação a artigos científicos baseados na técnica da ACI elaborados por portugueses, mencionamos Marques (2002), Lima *et al.* (2003), Nascimento e Dias (2005), Ferreira e Figueiredo (2006) e um número bastante razoável de publicações realizadas por Luís Borges de Almeida (Professor da área da Engenharia Eletrotécnica do Instituto Superior Técnico da Universidade Técnica de Lisboa).

Conferências e Centros de Investigação em ACI

Só em 1999 é que surgiu a primeira conferência internacional em BSS e ACI (ICA'99), a qual decorreu em Aussois, França. A partir daí muitas outras conferências se têm sucedido, as quais têm sido organizadas a cada ano e meio. Em 2010, o nome da conferência foi alterado para *9th International Conference on Latent Variable Analysis and Signal Separation* (LVA/ICA 2010) com o intuito de corresponder aos avanços recentes na área. Para além das habituais publicações nos *Proceedings* destas conferências, muitas outras têm sido publicadas desde o início deste milénio em revistas internacionais como *Signal Processing*, *Neural Networks*, *Neural Computation*, *IEEE Transactions on Signal Processing*, etc..

Devido à crescente disseminação da técnica da ACI nas duas últimas décadas, atualmente existem muitos Centros de Investigação onde a ACI tem um papel fundamental em termos de produção científica na área: o *RIKEN* no Japão, o *Salk Institute for Biological Studies* sediado em La Jolla, San Diego na Califórnia, o *Swartz Center for Computational Neuroscience* que é um Laboratório do *Institute for Neural Computation*, da *University of California San Diego*, o *Laboratory of Computer and Information Science* da *Aalto University* em Helsínquia e o *Département Traitement du Signal et des Images* que pertence à Escola francesa de telecomunicações Télécom ParisTech.

Softwares de Aplicação da ACI

Uma vez que a ACI também é uma técnica que recorre a ferramentas computacionais, apenas faz sentido ser levada a cabo mediante a utilização de *softwares* suficientemente desenvolvidos que permitam a implementação eficaz dos algoritmos cada vez mais rápidos. De seguida enunciamos os *packages/toolboxes* mais utilizados para esta técnica entre a comunidade científica, cuja maioria está concebida para ser implementada em MATLAB, embora o *FastICA* também esteja implementado em R e noutros ambientes.

- *FastICA*¹: É um *package* de distribuição livre que implementa o algoritmo do ponto fixo, o FastICA, para a ACI e para a Perseguição da Projeção. Este algoritmo é um dos mais usados na ACI devido à sua rapidez de execução e aplicabilidade para problemas de grande escala.
- *ICALAB*²: ICALAB para o Processamento de Sinal e ICALAB para o Processamento de Imagem são duas *toolboxes* que integram vários algoritmos para a ACI empregando estatísticas de ordem superior, para a BSS utilizando estatísticas de segunda ordem e predição linear, e para a Extração Cega do Sinal usando vários métodos de estatísticas de segunda ordem e de ordem superior.
- *EEGLAB*³: É uma *toolbox* interativa que serve para processar dados de eletroencefalogramas, de magnetoencefalogramas e outros dados eletrofisiológicos utilizando a ACI, a análise do tempo e da frequência e outros métodos como por exemplo a rejeição de artefactos (sinais indesejados).
- *FMRLAB*⁴: É uma *toolbox* para análise de dados de imagens de ressonância magnética funcional através da ACI. O FMRLAB foi construído de forma a apresentar um ambiente integrado onde é possível gerir, processar e analisar dados de ressonância magnética numa única plataforma, de modo que os utilizadores possam completar a análise sem terem que

¹No software R: <http://cran.r-project.org/web/packages/fastICA/index.html>; no software MATLAB e outros ambientes: <http://research.ics.aalto.fi/ica/fastica/>

²A. Cichocki, S. Amari, K. Siwek, T. Tanaka, Anh Huy Phan et al., ICALAB Toolboxes, <http://www.bsp.brain.riken.jp/ICALAB>.

³<http://sccn.ucsd.edu/eeglab/>

⁴<http://sccn.ucsd.edu/fmrlab/>

alternar entre diferentes *softwares*.

- *Group ICA Toolbox*⁵: É constituída por duas aplicações que implementa uma série de algoritmos para ACI e BSS.
- *Fusion ICA Toolbox*⁶ (*FIT*): Utiliza a ACI para extrair a informação partilhada por várias fontes como imagens de ressonância magnética funcional, imagens de ressonância magnética estrutural, eletroencefalogramas e dados de polimorfismos de nucleótidos simples (dados SNP).
- *ICA:DTU Toolbox*⁷: É constituída por uma coleção de algoritmos para implementar a ACI.

3.3 Motivação

Passamos de seguida a ilustrar um dos exemplos elementares que serviu de motivação para a investigação da ACI, o qual é conhecido por problema *cocktail party* (Hyvärinen e Oja, 2000). Admita-se que existe um determinado número de fontes sonoras registadas por um certo número de microfones e que cada um dos sinais gravados é uma soma ponderada dos sinais emitidos pelas fontes. Pretendemos estimar os sinais originais (das fontes) usando apenas os sinais gravados. Muitas das vezes existe interesse em separar apenas um sinal, por exemplo o de uma pessoa que em determinado momento esteja a falar para o microfone enquanto que as outras fontes (pessoas) podem ser encaradas como emissoras de ruído.

O problema de separar ou estimar os sinais originais emitidos pelas fontes, a partir de dados amostrais recolhidos por sensores, sem o mínimo conhecimento das características dos canais de transmissão e das fontes envolvidas pode ser perfeitamente enquadrado no método designado por Separação Cega de Fontes (*Blind Source Separation (BSS)*). A ACI está diretamente relacionada com a Separação Cega de Fontes (Hyvärinen *et al.*, 2001b), uma vez que o termo "fontes" significa que existem sinais originais, ou seja, componentes independentes e a palavra "cega" significa que conhecemos muito pouco ou nada acerca das misturas. Neste contexto, a ACI talvez seja o método mais usado para implementar a Separação Cega de Fontes.

⁵<http://mialab.mrn.org/software/>

⁶<http://mialab.mrn.org/software/fit/>

⁷<http://cogsys.imm.dtu.dk/toolbox/ica/>

Como um exemplo particular do problema *cocktail party* consideremos a Figura 3.1 na qual estão três pessoas a conversar numa sala (embora pudesse ser qualquer número inteiro superior a um), na presença de cinco microfones que gravam as conversas em diferentes localizações. Cada orador (sinais originais) contribui com um determinado peso para a obtenção de cada um dos sinais gravados, os quais são misturas das falas originais.

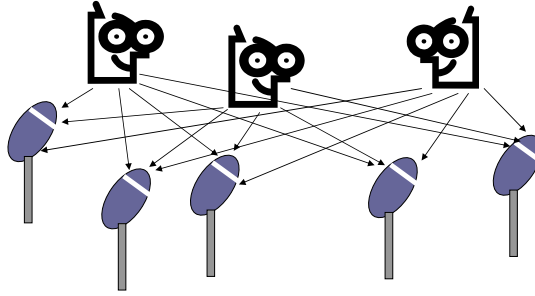


Figura 3.1: Exemplo do problema *cocktail party*: três pessoas a conversar numa sala com cinco microfones a captar as conversas em diferentes localizações

Se denotarmos por x_1, x_2, x_3, x_4 e x_5 os cinco sinais gravados a partir dos três sinais originais s_1, s_2 e s_3 ao longo do tempo t , obtemos o seguinte sistema de equações lineares

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\ x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \\ x_4(t) = a_{41}s_1(t) + a_{42}s_2(t) + a_{43}s_3(t) \\ x_5(t) = a_{51}s_1(t) + a_{52}s_2(t) + a_{53}s_3(t) \end{cases},$$

onde os coeficientes a_{ij} , com $i = 1, 2, 3, 4, 5$ e $j = 1, 2, 3$ são parâmetros desconhecidos que dependem das distâncias entre os oradores e os microfones. O sistema apenas seria invertível caso a matriz formada pelos coeficientes a_{ij} fosse quadrada e invertível. Nesse caso, conhecendo-se os parâmetros a_{ij} , a solução do sistema poderia facilmente ser determinada por inversão do mesmo. Contudo, neste exemplo concreto, o sistema não é invertível e para além de existirem os parâmetros desconhecidos a_{ij} , também não se conhecem os sinais originais s_1, s_2 e s_3 .

Uma forma razoável de abordar o problema seria usar alguma informação relacionada com propriedades estatísticas dos sinais para estimar, quer os parâmetros a_{ij} , quer os sinais originais. Neste contexto é aceitável admitir que é suficiente que os sinais sejam estatisticamente indepen-

dentes em cada instante de tempo t . Mesmo que na prática este pressuposto não seja verificado com precisão, é fundamental ser considerado para que se possam separar os sinais originais a partir das misturas x_i .

A necessidade crescente de investigar problemas semelhantes a este, levou a que a ACI suscitasse rapidamente grande interesse noutros campos de aplicação (veja-se a Secção 3.11).

Consideremos o exemplo na Figura 3.2 ilustrado por três sinais originais (de 1000 valores cada) obtidos computacionalmente, os quais constituem as CIs. Ao denotarmos por x_1 , x_2 e x_3 os três sinais misturados a partir dos três sinais originais s_1 , s_2 e s_3 ao longo do tempo t , obtemos o seguinte sistema de equações lineares

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\ x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \end{cases} .$$

Ao utilizarmos, por exemplo, a matriz quadrada invertível de coeficientes a_{ij} ($i, j = 1, 2, 3$) dada por

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & 2.5 & -0.2 \\ 0.7 & 0.3 & 0.1 \\ -1.3 & 1.0 & -1.8 \end{bmatrix} ,$$

as misturas correspondem às representações da Figura 3.3, e o sistema de equações lineares passa a ser possível e determinado, cuja solução é determinada por inversão do mesmo.

Partindo destas misturas, ao aplicar-se a técnica da ACI o que se pretende é separá-las com base na independência através de determinado algoritmo, com vista a obter as CIs estimadas, as quais irão permitir reconstruir as misturas observadas. Na Figura 3.4 estão representadas as estimativas das CIs com base num dos algoritmos mais utilizados, por exemplo o FastICA, o qual será apresentado na Secção 3.7.

Ao compararmos visualmente os sinais originais com as CIs estimadas, evidencia-se a potencialidade da técnica da ACI ao conseguir separar as diferentes componentes de uma forma eficaz (não obstante da ordem ou do sinal; ambiguidades a serem abordadas na Secção seguinte).

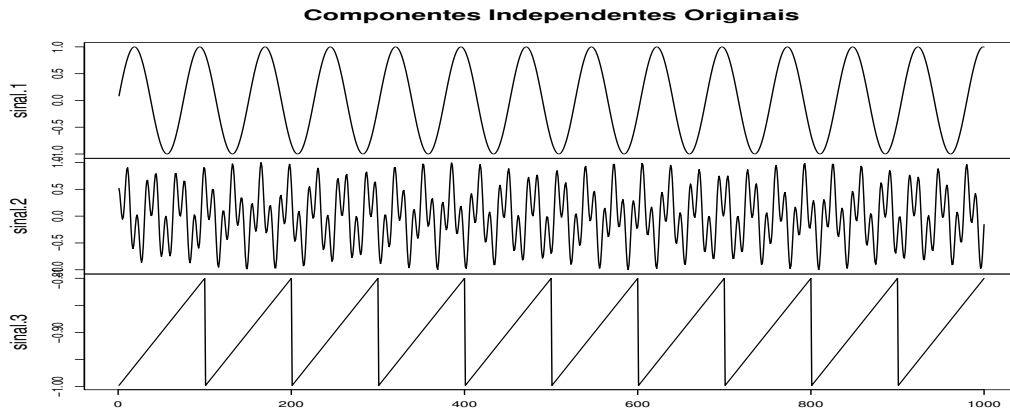


Figura 3.2: Exemplo de três sinais (CIs) originais

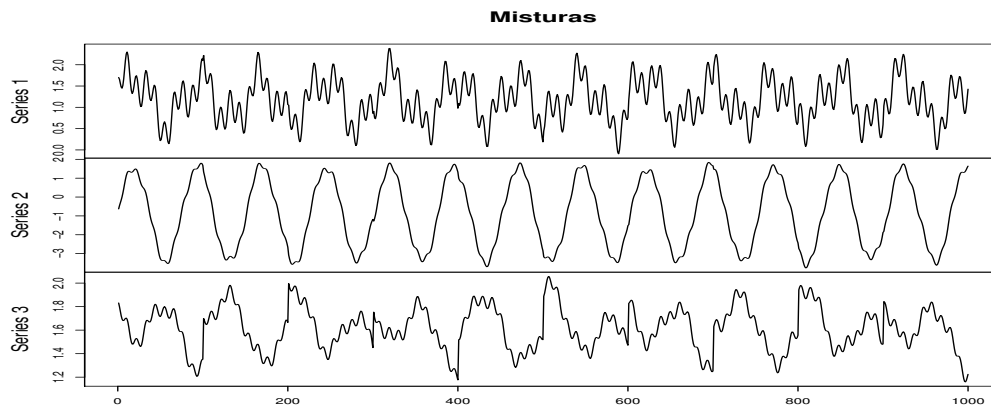


Figura 3.3: Exemplo de três misturas correspondentes aos sinais da Figura 3.2

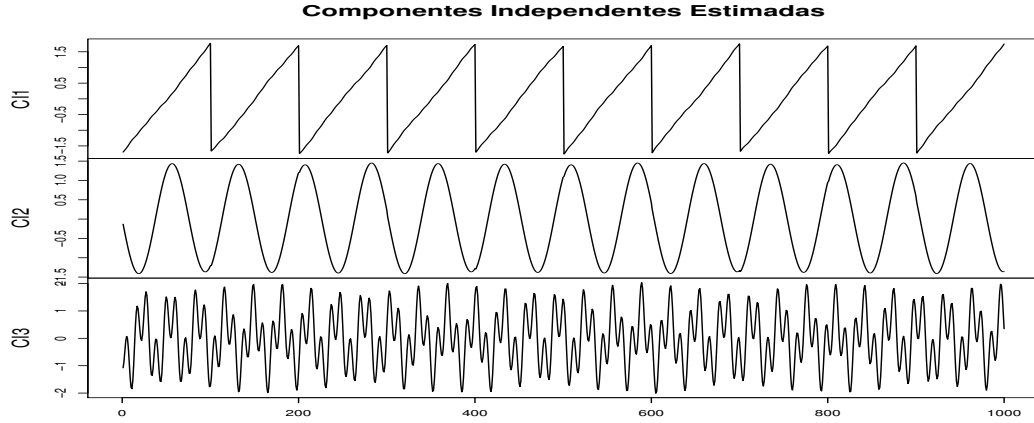


Figura 3.4: Exemplo das estimativas das três componentes independentes (ou seja, dos sinais originais) com base nas misturas dos sinais ilustradas na Figura 3.3

3.4 Descrição do Modelo

Definição 3.4.1 *Define-se o modelo linear clássico (sem ruído) constituído por n variáveis aleatórias x_1, x_2, \dots, x_n , descritas à custa de misturas lineares de n variáveis latentes s_1, s_2, \dots, s_n , na forma matricial*

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (3.1)$$

ou seja,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \quad (3.2)$$

onde

- \mathbf{x} é o vetor aleatório cujos elementos são as misturas x_1, x_2, \dots, x_n ,
- \mathbf{s} é o vetor aleatório cujos elementos são as componentes independentes s_1, s_2, \dots, s_n e
- \mathbf{A} é a matriz das misturas cujos elementos a_{ij} ($i, j = 1, 2, \dots, n$) são os parâmetros desconhecidos.

Neste modelo o que apenas se observa são valores das variáveis aleatórias x_i , enquanto que as componentes independentes não são diretamente observadas e, por outro lado, admite-se que

os coeficientes das misturas são desconhecidos. O objetivo é estimar quer os coeficientes a_{ij} das misturas quer as componentes independentes s_i somente à custa de \mathbf{x} (Hyvärinen e Oja, 2000), ilustrado no esquema da Figura 3.5, tendo em conta algumas suposições gerais.

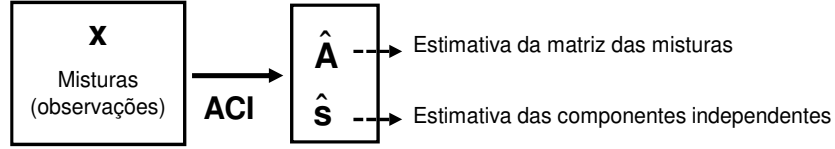


Figura 3.5: Esquema do processo de estimação do modelo linear de ACI básico

Embora esta definição de modelo de ACI seja a mais básica possível, não podemos excluir outras definições existentes e que serão mencionadas na Secção 3.10. Por exemplo, em muitas aplicações é necessário considerar que existe algum ruído nas medições, o que leva a acrescentar um termo de ruído ao modelo em causa. Para além disso, no modelo básico (3.1), admite-se também que o número de CIs é igual ao número de variáveis observadas, no entanto muitos problemas poderão apresentar um número de CIs diferente do número de misturas observadas. Apesar de se considerar este modelo como linear, pode ainda acontecer que as misturas não sejam lineares, dando portanto origem a um modelo não linear.

Se em vez de variáveis utilizarmos vetores aleatórios para definir o modelo da ACI dado por (3.1), e sem forçarmos a que o número de componentes independentes seja igual ao número de misturas, então passamos ao contexto matricial, o qual passa a ser reescrito matricialmente por

$$\mathbf{X} = \mathbf{S}\mathbf{A}, \quad (3.3)$$

onde:

- \mathbf{X} é a matriz $n \times p$ dos dados observados, ou seja, das p misturas,
- \mathbf{S} é a matriz $n \times k$ das k componentes independentes e
- \mathbf{A} é a matriz $k \times p$ dos coeficientes das misturas (parâmetros desconhecidos) e cujas colunas deverão ser linearmente independentes.

Note-se que, sem perda de generalidade, ao decidir apresentar o modelo definido desta forma matricial, a nova matriz \mathbf{A} corresponde à transposta da matriz \mathbf{A} definida em (3.1). Tal decisão

deve-se ao facto de termos em conta a forma como o modelo está implementado no *software R*, mas essencialmente para estarmos de acordo com a notação matricial usual para matriz de dados iniciais ($\mathbf{X}_{n \times p}$ - n observações e p variáveis) usada quer pela ACP quer pelas outras técnicas que iremos abordar, e uma vez que pretendemos efetuar comparações entre técnicas.

Como daqui em diante iremos basear-nos essencialmente no modelo clássico descrito a partir de (3.1) para apresentar os conceitos teóricos, então o modelo anterior dado por (3.3) não é passível de gerar confusão ao longo deste Capítulo.

3.4.1 Propriedades Básicas do Modelo

Para que se possa estimar o modelo básico de ACI, é preciso admitir que:

- as componentes s_i ($i = 1, 2, \dots, n$) sejam estatisticamente independentes, uma vez que este é o princípio no qual a ACI se baseia;
- pelo menos $n - 1$ das componentes s_i ($i = 1, 2, \dots, n$) não sigam uma distribuição Normal, pelo facto de ser impossível separar várias componentes que sigam distribuições Normais;
- por simplicidade, a matriz \mathbf{A} seja quadrada (o número de CIs igual ao número de misturas).

De acordo com os métodos de estimação que serão introduzidos na Secção 3.5, começa-se por estimar a matriz \mathbf{A} e o vetor \mathbf{s} , observando apenas \mathbf{x} , e de seguida admitindo que esta é invertível, determina-se a sua inversa, $\mathbf{W} = \mathbf{A}^{-1}$, para se obterem as CIs dadas por

$$\mathbf{s} = \mathbf{W}\mathbf{x}. \quad (3.4)$$

A Figura 3.6 representa o processo de misturar as CIs através de \mathbf{A} , seguido da transformação inversa \mathbf{W} de separação das misturas. As CIs são emitidas e combinadas linearmente originando as misturas de acordo com a equação (3.1), às quais é aplicada uma transformação linear inversa que permite obter as CIs estimadas através da equação (3.4).

Por analogia com o que habitualmente se faz na ACP, caso se pretendam comparar os valores observados de \mathbf{x} com os reconstruídos basta efetuar a reconstrução dos dados originais multiplicando a matriz das componentes estimadas pela matriz das misturas estimadas. Note-se que, no modelo apresentado em (3.1) não se podem determinar as variâncias das CIs (uma vez que a matriz \mathbf{A} e o vetor \mathbf{s} são desconhecidos), e portanto uma forma de contornar a questão

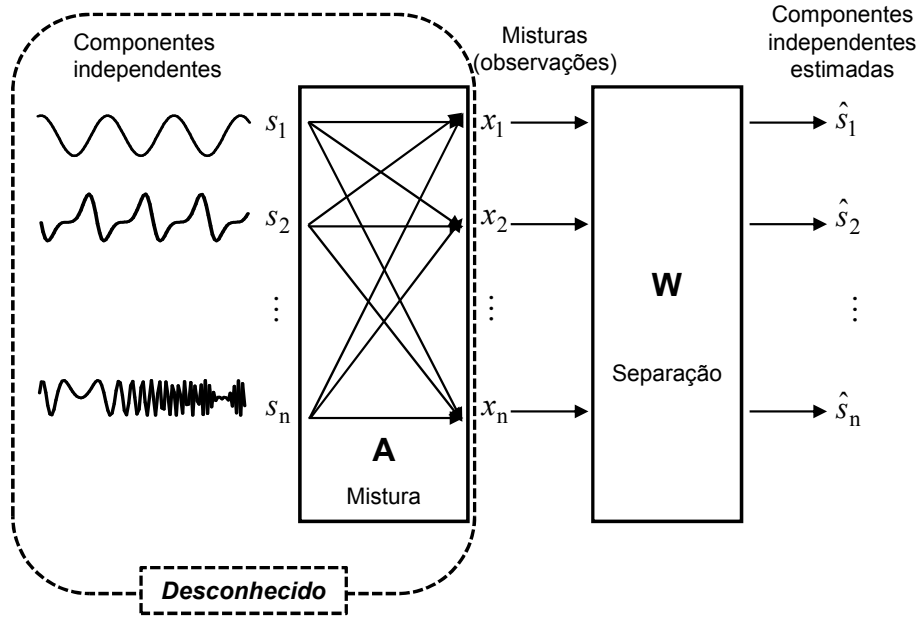


Figura 3.6: Esquema de blocos ilustrando o sistema de mistura das CIs e de separação das misturas ao longo do processo da ACI

é admitir que cada uma possui variância unitária. É de notar que a esta situação está inerente a ambiguidade do sinal de cada componente, ou seja, as CIs do modelo não serão afetadas a menos do sentido (Hyvärinen *et al.*, 2001b). Também não se podem ordenar as CIs à priori, uma vez que sendo ainda a matriz \mathbf{A} e o vetor \mathbf{s} desconhecidos, pode mudar-se arbitrariamente a ordem das equações resultantes do sistema de equações (3.2).

3.4.2 Centragem e Branqueamento

Tal como noutros modelos multivariados, para executar a ACI antes de aplicar um algoritmo a um determinado conjunto de dados, é comum proceder-se a técnicas de pré-processamento, as quais facilitam a estimação dos parâmetros do modelo.

A centragem, a mais básica das técnicas, corresponde a subtrair a média amostral de forma a que cada variável passe a ter média zero. Se designarmos as misturas originais por \mathbf{x}' , antes de implementar a ACI de acordo com a equação (3.1), aplicamos a centragem às variáveis observadas e obtemos

$$\mathbf{x} = \mathbf{x}' - E(\mathbf{x}'),$$

onde $E(\cdot)$ designa o operador valor esperado. Consequentemente, as CIs das novas variáveis centradas também passam a ter média nula, já que

$$E(\mathbf{x}) = E(\mathbf{A}\mathbf{s}) = \mathbf{A}E(\mathbf{s})$$

e portanto

$$E(\mathbf{s}) = \mathbf{A}^{-1}E(\mathbf{x}),$$

e logo centrar em \mathbf{x} implica centrar em \mathbf{s} . A matriz das misturas mantém-se inalterada depois da centragem, pelo que se pode efetuar sempre esta operação sem afetar a estimação da mesma. Designando as CIs originais (não centradas) por \mathbf{s}' tal que $\mathbf{s} = \mathbf{s}' - E(\mathbf{s}')$, então

$$E(\mathbf{s}') = \mathbf{A}^{-1}E(\mathbf{x}').$$

Depois de estimar a matriz das misturas e as CIs para os dados centrados, podemos efetuar a reconstrução através da adição do vetor das médias de \mathbf{s}' ,

$$\mathbf{A}^{-1}E(\mathbf{x}'),$$

às CIs centradas.

Outra técnica de pré-processamento dos dados, usada para estimar as componentes independentes, após a centragem e por forma a simplificar o processo de obtenção das CIs, é o branqueamento.

Definição 3.4.2 *Seja $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ um vetor de dados centrados. Diz-se que o vetor \mathbf{z} está branqueado se*

$$E(\mathbf{z}\mathbf{z}^T) = \mathbf{I}.$$

Deste modo, um vetor \mathbf{z} está branqueado se os elementos z_i ($i = 1, 2, \dots, n$) forem não correlacionados e tiverem variâncias unitárias, o que significa que as respetivas matrizes de covariâncias e de correlações são iguais à matriz identidade \mathbf{I} .

Propriedade 3.4.3 *Seja \mathbf{x} o vetor dos dados observados. O branqueamento transforma linearmente \mathbf{x} , multiplicando-o por uma determinada matriz \mathbf{V} , obtendo-se assim um vetor branqueado \mathbf{z} definido por*

$$\mathbf{z} = \mathbf{V}\mathbf{x}. \quad (3.5)$$

Um método frequente de branqueamento, embora não seja designado por esse nome no pré-processamento de dados, é o da obtenção prévia dos valores e vetores próprios da matriz de covariâncias

$$E(\mathbf{xx}^T) = \mathbf{E}\mathbf{D}\mathbf{E}^T,$$

onde $E(\mathbf{xx}^T)$ designa o valor esperado da matriz \mathbf{xx}^T , \mathbf{E} é a matriz ortogonal dos vetores próprios de $E(\mathbf{xx}^T)$ e \mathbf{D} é a matriz diagonal dos valores próprios positivos de $E(\mathbf{xx}^T)$.

Definição 3.4.4 *Considere-se a decomposição espectral da matriz $E(\mathbf{xx}^T) = \mathbf{E}\mathbf{D}\mathbf{E}^T$. Designa-se por matriz do branqueamento a*

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T,$$

onde $\mathbf{D}^{-1/2}$ é a matriz diagonal cujos elementos diagonais são os inversos das raízes quadradas dos valores próprios de \mathbf{D} .

Utilizando (3.1) e (3.5), tem-se uma nova matriz de misturas \mathbf{VA} , ou seja,

$$\mathbf{z} = \mathbf{VA}\mathbf{s}. \quad (3.6)$$

Como a não correlação é mais fraca que a independência, logo não é suficiente para a estimação do modelo ACI. No seguimento deste raciocínio, considere-se uma transformação ortogonal \mathbf{U} de \mathbf{z} dada por

$$\mathbf{y} = \mathbf{U}\mathbf{z} = \mathbf{UVA}\mathbf{s}.$$

Tem-se então

$$E(\mathbf{yy}^T) = E(\mathbf{Uzz}^T\mathbf{U}^T) = \mathbf{U}E(\mathbf{zz}^T)\mathbf{U}^T = \mathbf{U}\mathbf{U}^T = \mathbf{I},$$

uma vez que

$$\begin{aligned} E(\mathbf{zz}^T) &= E(\mathbf{Vxx}^T\mathbf{V}^T) = \mathbf{V}E(\mathbf{xx}^T)\mathbf{V}^T = \mathbf{VEDE}^T\mathbf{V}^T \\ &= \mathbf{ED}^{-1/2}\mathbf{E}^T\mathbf{EDE}^T\mathbf{ED}^{-1/2}\mathbf{E}^T = \mathbf{ED}^{-1/2}\mathbf{DD}^{-1/2}\mathbf{E}^T = \mathbf{EE}^T = \mathbf{I}. \end{aligned}$$

Portanto o vetor \mathbf{y} também é um vetor branqueado. No caso de \mathbf{y} ser uma transformação ortogonal de \mathbf{z} , o branqueamento fornece as CIs apenas para uma transformação ortogonal, o que nem sempre é suficiente para a maioria das aplicações.

A utilidade do branqueamento prende-se com o facto da nova matriz das misturas \mathbf{VA} ser ortogonal, o que pode ser visto a partir de

$$E(\mathbf{zz}^T) = E(\mathbf{VA}\mathbf{ss}^T(\mathbf{VA})^T) = \mathbf{VA}E(\mathbf{ss}^T)(\mathbf{VA})^T = \mathbf{VA}(\mathbf{VA})^T = \mathbf{I}.$$

Desta forma, pode restringir-se a procura de uma matriz de misturas ao espaço das matrizes ortogonais. Sendo assim, em vez de se estimarem n^2 elementos para a matriz das misturas apenas é necessário estimar aproximadamente metade dos parâmetros, uma vez que uma matriz ortogonal, de grandes dimensões, contém apenas $n(n-1)/2$ parâmetros de uma matriz arbitrária (Hyvärinen *et al.*, 2001b).

O facto de uma matriz de misturas ortogonal não poder ser estimada para variáveis com distribuição Normal, prende-se com a propriedade das variáveis não correlacionadas e com distribuição Normal conjunta serem obrigatoriamente independentes. Deste modo, a informação da independência das componentes não permite ir mais longe do que o branqueamento. Portanto, o modelo ACI não pode ser estimado para mais do que uma variável cuja distribuição seja Normal (uma vez que variáveis com distribuição Normal não se conseguem separar umas das outras).

Variáveis com Distribuição Normal

Veja-se como o branqueamento ajuda a perceber a observação anterior.

Admita-se que a distribuição conjunta de duas CIs s_1 e s_2 segue uma distribuição Normal. Nestas condições, a função densidade de probabilidade conjunta é dada por

$$f_{S_1, S_2}(s_1, s_2) = \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right).$$

Admita-se que a matriz das misturas \mathbf{A} é ortogonal, uma vez que os dados estão sujeitos ao branqueamento. Como $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$, e tendo em conta que $\mathbf{A}^{-1} = \mathbf{A}^T$ uma vez que \mathbf{A} é ortogonal, estabelecemos a definição seguinte.

Proposição 3.4.5 *Seja \mathbf{A} uma matriz ortogonal. A função densidade de probabilidade conjunta das misturas x_1 e x_2 é descrita por*

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{A}^T \mathbf{x}\|^2}{2}\right) |\det \mathbf{A}^T| = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right),$$

com $\|\mathbf{A}^T \mathbf{x}\|^2 = \|\mathbf{x}\|^2$ e $|\det \mathbf{A}^T| = |\det \mathbf{A}| = 1$.

Desta forma, verifica-se que a matriz de misturas ortogonal não altera a função densidade de probabilidade conjunta. Pelo facto das duas distribuições em causa serem idênticas, não existe forma de poder deduzir a matriz das misturas a partir das próprias misturas.

3.5 Métodos de Estimação

Nesta Secção debruçamo-nos sobre a descrição de alguns dos métodos de estimação do modelo ACI, nomeadamente aqueles que consideramos como mais importantes ou fundamentais para aprofundar o estudo da técnica da ACI ao longo deste trabalho.

Para estimar um modelo de ACI é preciso ter em conta certos princípios fundamentais para um bom desempenho do algoritmo a ser implementado. Um desses princípios simples de estimação da ACI é baseado na maximização de uma medida de não normalidade.

Considere-se como ponto de partida o teorema limite central, admitindo para tal que todas as componentes independentes s_i ($i = 1, 2, \dots, n$) têm distribuições idênticas.

Relembre-se que o modelo é descrito por $\mathbf{x} = \mathbf{A}\mathbf{s}$ e que $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$.

A combinação linear das misturas x_i ($i = 1, 2, \dots, n$) é dada por

$$y = \mathbf{b}^T \mathbf{x}, \quad (3.7)$$

onde \mathbf{b} é um vetor a ser determinado. Por consequência, y é uma combinação linear de s_i descrita por $y = \mathbf{b}^T \mathbf{A}\mathbf{s}$.

Consideremos o vetor y como combinação linear de s_i com coeficientes $\mathbf{b}^T \mathbf{A}$ e o vetor $\mathbf{q} = \mathbf{b}^T \mathbf{A}$ de modo que

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s} = \sum_i q_i s_i.$$

Para estimar uma das CIs, tem-se em conta (3.7).

Caso \mathbf{b} fosse uma das linhas da matriz inversa de \mathbf{A} , então esta combinação linear $\mathbf{b}^T \mathbf{x}$ deveria ser igual a uma das CIs, sendo \mathbf{q} constituído por um único coeficiente 1 e por 0 nos restantes coeficientes.

De seguida analisa-se como é que o teorema limite central pode ser utilizado de forma a permitir determinar \mathbf{b} . Ao variar os coeficientes de \mathbf{q} , ter-se-á a noção de como se comporta a distribuição de $y = \mathbf{q}^T \mathbf{s}$.

Partindo da ideia de que a soma de duas variáveis aleatórias idênticas e independentes apresenta geralmente "maior" normalidade que as variáveis originais, pode então deduzir-se que $y = \mathbf{q}^T \mathbf{s}$ possui geralmente "maior" normalidade que qualquer uma das CIs s_i , e tende a possuir "menor" normalidade quando é igual a uma das componentes s_i (e nesse caso só um dos elementos do vetor \mathbf{q} será não nulo).

Na prática, apenas se pode fazer variar os coeficientes de \mathbf{b} , e portanto averiguar o comportamento da distribuição de $y = \mathbf{b}^T \mathbf{x}$. Desta forma poderia considerar-se \mathbf{b} como um vetor que maximize a não normalidade de $\mathbf{b}^T \mathbf{x}$, o qual corresponderia a $\mathbf{q} = \mathbf{A}^T \mathbf{b}$, com apenas um elemento não nulo. Assim $y = \mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s}$ igualaria a uma das CIs, e portanto maximizar a não normalidade de $\mathbf{b}^T \mathbf{x}$ significava obter uma das CIs.

Podemos estipular assim, que a estimação da ACI está formulada para a procura das direções que sucessivamente maximizam a não normalidade, ou seja, para obter cada componente independente num máximo local. Nas Secções seguintes iremos analisar como é que a não normalidade de $y = \mathbf{b}^T \mathbf{x}$ pode ser medida e como podem ser obtidos os valores de \mathbf{b} que maximizam localmente tal medida.

Uma dessas medidas de não normalidade é a curtose, a qual é descrita na Subsecção 3.5.1. Com base na curtose podemos deduzir algoritmos através dos métodos do gradiente e do ponto fixo. Também a introdução na Subsecção 3.5.2 de uma medida alternativa de não normalidade, designada por neguentropia, permite resolver alguns problemas relacionados com a curtose, através da utilização de algoritmos correspondentes.

3.5.1 Curtose como Medida de Não Normalidade

Como medida de não normalidade para a estimação de um modelo ACI, utiliza-se frequentemente a informação do cumulante de quarta ordem, ou seja, a curtose (Apêndice C.4).

Definição 3.5.1 Para uma variável aleatória (v. a.) X com média zero, a curtose é dada por

$$\text{curt}(x) = E(x^4) - 3 [E(x^2)]^2,$$

onde, $E(x^k)$ são os designados momentos não centrados de ordem k .

Nota 3.5.2 Para uma v. a. X com $E(x) = 0$, no caso da variância ser unitária, vem apenas

$$\text{curt}(x) = E(x^4) - 3.$$

Para uma v. a. com distribuição Normal a curtose é zero, enquanto que em geral para as variáveis aleatórias com outras distribuições a curtose não é zero. As variáveis aleatórias que possuem valores negativos da curtose são designadas por sub-Gaussianas (ou platicúrticas) e as que possuem valores positivos da curtose são designadas por super-Gaussianas (ou leptocúrticas). A Figura 3.7 apresenta a comparação entre um exemplo de uma distribuição super-Gaussiana ilustrado pela distribuição de Laplace (ou Exponencial Dupla) com um pico e caudas pesadas, a distribuição Normal standard (ou Gaussiana) e uma distribuição sub-Gaussiana dada pela distribuição Uniforme. As três distribuições estão centradas e possuem variâncias unitárias.

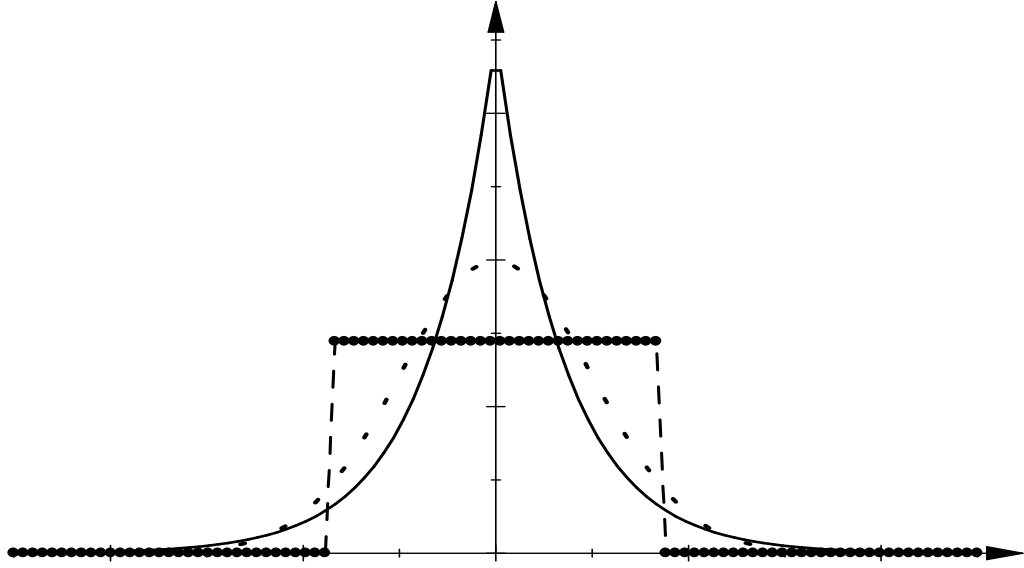


Figura 3.7: Funções densidade de probabilidade da distribuição de Laplace, da distribuição Normal (a ponteadado) e da distribuição Uniforme (a negrito) com médias nulas e variâncias unitárias

Propriedade 3.5.3 *Sejam x_1 e x_2 duas variáveis aleatórias independentes e α uma constante. A curtose, pelo facto de ser um cumulante, é dotada das seguintes propriedades (Hyvärinen et al., 2001b):*

$$\text{curt}(x_1 + x_2) = \text{curt}(x_1) + \text{curt}(x_2) \quad (3.8)$$

e,

$$\text{curt}(\alpha x_1) = \alpha^4 \text{curt}(x_1). \quad (3.9)$$

Considere-se o caso do modelo ACI (3.1) a duas dimensões, tendo como objetivo efetuar uma análise no que diz respeito à otimização para a curtose e à forma de obter as CIs.

Sejam s_1 e s_2 duas CIs de variâncias unitárias com valores não nulos $curt(s_1)$ e $curt(s_2)$ para a curtose e seja $y = \mathbf{b}^T \mathbf{x}$ considerada como uma das CIs a encontrar. Portanto, como visto anteriormente em (3.7), $y = \mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s}$. Neste caso concreto tem-se

$$y = q_1 s_1 + q_2 s_2,$$

pelo que de (3.8) e de (3.9) vem

$$curt(y) = curt(q_1 s_1) + curt(q_2 s_2) = q_1^4 curt(s_1) + q_2^4 curt(s_2).$$

Se a variância de y é igual à unidade, consequentemente restringe-se \mathbf{q} , ou seja

$$Var(y) = Var(q_1 s_1 + q_2 s_2) = q_1^2 Var(s_1) + q_2^2 Var(s_2) = q_1^2 + q_2^2 = 1$$

pois s_1 e s_2 são duas componentes independentes e centradas. Assim, o vetor \mathbf{q} fica restringido ao círculo unitário no plano bidimensional.

Para maximização da não normalidade procura-se o máximo do módulo da função $curt(y)$ no círculo unitário.

Admita-se por simplicidade que $curt(s_1) = curt(s_2) = 1$, e considere-se a função

$$F(\mathbf{q}) = q_1^4 + q_2^4. \quad (3.10)$$

Na Figura 3.8 ilustramos o que acontece em termos de otimização ao representarmos a projeção no plano de algumas curvas de nível da função, assim como a restrição do círculo unitário.

Aplicando o Método dos Multiplicadores de Lagrange, obtém-se a função

$$\mathcal{L}(q_1, q_2, \lambda) = q_1^4 + q_2^4 - \lambda(q_1^2 + q_2^2 - 1)$$

onde λ é o multiplicador de Lagrange.

Os pontos críticos desta função Lagrangeana determinam-se a partir do sistema de equações

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial q_1} = 0 \\ \frac{\partial \mathcal{L}}{\partial q_2} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \end{cases} \Leftrightarrow \begin{cases} 4q_1^3 - 2\lambda q_1 = 0 \\ 4q_2^3 - 2\lambda q_2 = 0 \\ q_1^2 + q_2^2 - 1 = 0 \end{cases}.$$

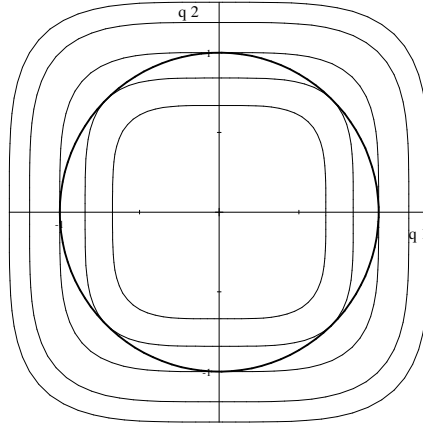


Figura 3.8: Cenário de otimização para a curtose: as linhas finas representam algumas curvas de nível da função descrita em (3.10) e a linha a grosso representa o círculo unitário

Uma vez que as suas soluções são da forma

$$\begin{cases} q_1 = 0 \vee q_1^2 = \frac{\lambda}{2} \\ q_2 = 0 \vee q_2^2 = \frac{\lambda}{2} \\ q_1^2 + q_2^2 = 1 \end{cases}$$

então os oito pontos críticos são os seguintes:

$$(1, 0), (0, 1), (-1, 0), (0, -1), \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \text{ e } \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

De entre os pontos críticos mencionados, aqueles em que a função F restringida ao círculo unitário atinge máximos são $(1, 0)$, $(0, 1)$, $(-1, 0)$ e $(0, -1)$. Estes são exatamente os quatro pontos obtidos em que $y = s_1$; $y = s_2$; $y = -s_1$ e $y = -s_2$ respetivamente.

Para o caso em que $\text{curt}(s_1) = \text{curt}(s_2) = -1$, a situação é análoga à anterior, uma vez que em módulo a função a maximizar é a mesma. No caso de $\text{curt}(s_1)$ e $\text{curt}(s_2)$ serem completamente arbitrárias e não nulas, o valor absoluto da curtose também é maximizado quando $y = \mathbf{b}^T \mathbf{x}$ é novamente igual a uma das CIs (Hyvärinen *et al.*, 2001b).

Para o vetor dos dados branqueados, designado anteriormente por \mathbf{z} , procura-se uma combinação linear $\mathbf{w}^T \mathbf{z}$ que maximize a não normalidade.

Dado que $\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s}$ e logo $\mathbf{w}^T \mathbf{z} = \mathbf{w}^T \mathbf{V}\mathbf{A}\mathbf{s}$, ocorre assim uma simplificação, desde que $\mathbf{q} = (\mathbf{V}\mathbf{A})^T \mathbf{w}$, e consequentemente

$$\|\mathbf{q}\|^2 = (\mathbf{w}^T \mathbf{V}\mathbf{A})(\mathbf{A}^T \mathbf{V}^T \mathbf{w}) = \|\mathbf{w}\|^2.$$

De acordo com a expressão obtida, restringir \mathbf{q} é o mesmo que restringir \mathbf{w} na esfera unitária, e desta forma maximiza-se o valor absoluto da curtose de $\mathbf{w}^T \mathbf{z}$ sob a condição $\|\mathbf{w}\| = 1$. Depois do branqueamento, as combinações lineares $\mathbf{w}^T \mathbf{z}$ podem ser encaradas como projeções na reta gerada pelo vetor \mathbf{w} , e cada ponto da esfera unitária corresponde a uma projeção.

Como exemplo, no caso de duas misturas branqueadas de componentes independentes uniformemente distribuídas, podem parametrizar-se os pontos da esfera unitária através do ângulo que o vetor \mathbf{w} faz com o eixo horizontal e pode efetuar-se o gráfico da curtose de $\mathbf{w}^T \mathbf{z}$ como uma função desse ângulo. As direções são aquelas em que o valor absoluto da curtose é maximizado, as quais correspondem aos lados do quadrado, que por sua vez fornecem as CIs (Hyvärinen *et al.*, 2001b).

Generalizando, os valores absolutos da curtose são maximizados pelas CIs (Delfosse e Loubaton, 1995) e as direções das CIs são ortogonais devido ao branqueamento.

Algoritmos para CIs com Base na Curtose

Um dos algoritmos para maximizar o valor absoluto da curtose é o algoritmo **do gradiente**. Este algoritmo inicia com um vetor \mathbf{w} arbitrariamente escolhido, e procura a direção na qual o valor absoluto da curtose de $y = \mathbf{w}^T \mathbf{z}$ cresce mais rapidamente, baseado numa amostra do vetor das misturas \mathbf{z} , e depois desloca o vetor \mathbf{w} nessa direção. Esta implementação é efetuada com base em métodos do gradiente assim como noutras suas extensões.

Outro algoritmo é o do **ponto fixo**, que pode ser aplicado como alternativa ao algoritmo do gradiente para casos em que este apresenta baixo nível de convergência ou a sua inexistência. Hyvärinen e Oja (1997) introduziram um algoritmo do ponto fixo denominado de FastICA que converge rapidamente, o qual apresenta certas propriedades que fazem dele um algoritmo mais eficiente em relação a outros baseados no gradiente. Este algoritmo iterativo permite encontrar os extremos locais da curtose de uma combinação linear das variáveis observadas, o que equivale a estimar as CIs que não sigam uma distribuição Normal, e não precisa de ajustar parâmetros como nos algoritmos do gradiente, tornando-o assim mais simples de usar.

Na Secção 3.7 serão descritos com detalhe alguns dos algoritmos mais utilizados em ACI, em particular o FastICA.

3.5.2 Neguentropia como Medida de Não Normalidade

Uma vez que nem sempre a curtose é a melhor medida de não normalidade para a estimação do modelo ACI, nomeadamente por esta ser por vezes sensível a *outliers* (Hyvärinen *et al.*, 2001b), surge como alternativa, entre outras, uma medida designada por neguentropia.

A noção de entropia de uma v. a. está relacionada com a quantidade de informação que a variável fornece. À partida, quanto maior for a "aleatoriedade" da variável, isto é, quanto mais imprevisível e menor estrutura apresentar, maior será a entropia. Consideremos H a medida de entropia a introduzir de seguida.

Definição 3.5.4 Para uma v. a. contínua X a entropia diferencial H é dada por

$$H(x) = - \int f_X(x) \log [f_X(x)] dx,$$

onde $f_X(\cdot)$ designa a função densidade de probabilidade da v. a. X , ou generalizada para um vetor aleatório \mathbf{x} como

$$H(\mathbf{x}) = - \int f_{\mathbf{x}}(\mathbf{x}) \log [f_{\mathbf{x}}(\mathbf{x})] d\mathbf{x}. \quad (3.11)$$

Uma vez que uma v. a. com distribuição Normal tem a maior entropia entre todas as variáveis aleatórias de igual variância (Hyvärinen *et al.*, 2001b), então a entropia pode ser usada como medida de não normalidade.

Muitas vezes, utiliza-se a versão normalizada da entropia diferencial, designada por neguentropia, para obter uma medida de não normalidade, que é zero para uma v. a. com distribuição Normal e sempre não negativa.

Definição 3.5.5 A neguentropia de um vetor aleatório \mathbf{x} , denotada por J , é dada por

$$J(\mathbf{x}) = H(\mathbf{x}_{Normal}) - H(\mathbf{x}), \quad (3.12)$$

onde \mathbf{x}_{Normal} é um vetor aleatório com distribuição Normal com a mesma matriz de covariâncias Σ que \mathbf{x} e cuja entropia é dada por

$$H(\mathbf{x}_{Normal}) = \frac{1}{2} \log |\det \Sigma| + \frac{n}{2} [1 + \log(2\pi)],$$

onde n é a dimensão de \mathbf{x} .

A neguentropia é ainda invariante para transformações lineares invertíveis. Para a transformação linear invertível $\mathbf{y} = \mathbf{M}\mathbf{x}$ obtém-se

$$H(\mathbf{y}) = H(\mathbf{x}) + \log |\det \mathbf{M}| \quad (3.13)$$

e, por Hyvärinen *et al.*, (2001b), vem

$$\begin{aligned} J(\mathbf{y}) &= H(\mathbf{y}_{Normal}) - H(\mathbf{y}) \\ &= \frac{1}{2} \log |\det \mathbf{M}\Sigma\mathbf{M}^T| + \frac{n}{2} [1 + \log(2\pi)] - (H(\mathbf{x}) + \log |\det \mathbf{M}|) \\ &= \frac{1}{2} \log |\det \Sigma| + 2\frac{1}{2} \log |\det \mathbf{M}| + \frac{n}{2} [1 + \log(2\pi)] - H(\mathbf{x}) - \log |\det \mathbf{M}| \\ &= \frac{1}{2} \log |\det \Sigma| + \frac{n}{2} [1 + \log(2\pi)] - H(\mathbf{x}) \\ &= H(\mathbf{x}_{Normal}) - H(\mathbf{x}) \\ &= J(\mathbf{x}). \end{aligned}$$

Como caso particular, é de notar que a neguentropia é invariante em escala.

Contudo, é mais difícil utilizar a neguentropia em termos computacionais, dado que a sua estimação, usando a própria definição, requereria uma estimativa da função densidade de probabilidade. No entanto, aproximações da neguentropia são muito úteis e podem ser utilizadas para obter um método eficiente para a ACI.

Para uma v. a. X com média zero e variância unitária, a medida clássica de aproximação da neguentropia, que utiliza cumulantes de ordem superior, é (Jones e Sibson, 1987)

$$J(x) \approx \frac{1}{12} [E(x^3)]^2 + \frac{1}{48} [curt(x)]^2. \quad (3.14)$$

De entre outras aproximações que foram desenvolvidas, para o caso particular em que G é uma função não quadrática, destaca-se a que é descrita por Hyvärinen (1998)

$$J(x) \approx (E[G(x)] - E[G(x_{Normal Std})])^2, \quad (3.15)$$

onde $X_{Normal Std}$ é uma v. a. com distribuição Normal standardizada.

Esta última aproximação representa uma generalização da aproximação descrita em (3.14) quando X tem uma distribuição simétrica, levando a que o primeiro termo em (3.14) desapareça. Para o caso em que $G(x) = x^4$ obtém-se uma aproximação baseada na curtose.

Segundo uma escolha cuidadosa da função G (também conhecida por função contraste), Hyvärinen (1997) obteve aproximações da neguentropia. Escolhendo uma função G que não cresça muito rapidamente, obteve estimadores mais robustos. As funções G_1 , G_2 e G_3 dadas por

$$G_1(x) = \frac{1}{\alpha} \log(\cosh(\alpha x)), \quad (3.16)$$

onde $1 \leq \alpha \leq 2$ é uma constante apropriada,

$$G_2(x) = -\exp(-x^2/2), \quad (3.17)$$

e

$$G_3(x) = \frac{1}{4}x^4, \quad (3.18)$$

têm sido das mais utilizadas na aproximação da neguentropia (Hyvärinen *et al.*, 2001b).

Algoritmos para CIs com Base na Neguentropia

De forma análoga ao que foi descrito para a curtose, podem obter-se algoritmos de forma a maximizar o valor da neguentropia. Um desses algoritmos é uma versão do FastICA (Hyvärinen, 1999a) que é baseado num esquema de iteração do ponto fixo para encontrar o máximo de não normalidade de $\mathbf{w}^T \mathbf{z}$, a qual é medida pela aproximação (3.15). Mais pormenores podem ser consultados na Secção 3.7.

ACI e a Perseguição da Projecção

Segundo Hyvärinen *et al.* (2001b), a Perseguição da Projecção está diretamente associada à ACI. Este método, encarado como um conjunto de técnicas exploratórias foi desenvolvida para encontrar determinadas projecções de dados multidimensionais que ostentem um "interesse" do ponto de vista espacial, ou seja, que permitam uma visualização ótima dos mesmos. O método é usado geralmente para calcular duas projecções unidimensionais, de acordo com algum critério que se estabeleça, e como tal, algumas estruturas existentes nos dados podem ser visualizadas no subespaço unidimensional ou nos planos bidimensionais gerados por duas direcções da Perseguição da Projecção. Portanto, este método é uma extensão do clássico método da ACP, no que diz respeito à visualização dos dados no plano gerado pelas duas primeiras CPs, obtidas de acordo com a maximização da variância.

Num contexto geral, como questão basilar, há que definir quais das projeções é que são consideradas interessantes numa Perseguição da Projeção. Argumentos de Huber (1985) e de Jones e Sibson (1987) indicam que a distribuição Normal é a menos interessante, e que as direções mais interessantes são aquelas que mostram uma distribuição menos parecida com a distribuição Normal.

Tendo em conta os pressupostos da ACI, a Perseguição da Projeção é portanto geralmente aplicada para encontrar as projeções dos dados que são menos parecidas com a distribuição Normal (direções que maximizam a não normalidade), ou seja, é praticamente o papel que a curtose e a neguentropia desempenham como medidas de não normalidade para a estimação de um modelo ACI. Por isso, as medidas de não normalidade podem também ser designadas por "índices" de Perseguição da Projeção e os algoritmos da ACI por algoritmos de Perseguição da Projeção.

3.5.3 Estimação de Máxima Verosimilhança

Para além da maximização da não normalidade, outra forma usual para estimar o modelo básico da ACI sem ruído é a estimação de máxima verosimilhança.

Para obter a verosimilhança num modelo de ACI utilizam-se as propriedades da densidade de uma transformação linear de um vetor aleatório. Para o modelo $\mathbf{x} = \mathbf{A}\mathbf{s}$ a função densidade de probabilidade do vetor de misturas \mathbf{x} pode ser descrita por

$$f_X(\mathbf{x}) = |\det \mathbf{W}| f_S(\mathbf{s}) = |\det \mathbf{W}| \prod_i f_i(s_i),$$

onde $\mathbf{W} = \mathbf{A}^{-1}$ e f_i representa a função densidade de cada uma das componentes independentes.

Admitindo que cada linha da matriz \mathbf{W} é um vetor da forma \mathbf{w}_i^T ($i = 1, 2, \dots, n$) e que existem K observações de \mathbf{x} , denotadas por $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(K)$, então a verosimilhança pode ser obtida através do produto das funções densidade nos K pontos como função de \mathbf{W}

$$L(\mathbf{W}) = \prod_{t=1}^K \prod_{i=1}^n f_i(\mathbf{w}_i^T \mathbf{x}(t)) |\det \mathbf{W}|,$$

e a log-verosimilhança por (Pham *et al.*, 1992)

$$\log L(\mathbf{W}) = \sum_{t=1}^K \sum_{i=1}^n \log f_i(\mathbf{w}_i^T \mathbf{x}(t)) + K \log |\det \mathbf{W}|. \quad (3.19)$$

Nota 3.5.6 Para simplificar a notação de (3.19), a soma ao longo da amostra de índice t passa à média da amostra observada, e ao dividir a log-verosimilhança por K obtém-se

$$\frac{1}{K} \log L(\mathbf{W}) = E \left(\sum_{i=1}^n \log f_i(\mathbf{w}_i^T \mathbf{x}) \right) + \log |\det \mathbf{W}|. \quad (3.20)$$

Alerta-se que a notação E representa a média da amostra e não o usual valor esperado teórico.

Algoritmos para a Estimação de Máxima Verosimilhança

Tal como na maximização da não normalidade, também aqui para se poder aplicar a estimação de máxima verosimilhança são necessários algoritmos que efetuem a maximização numérica da verosimilhança. Alguns métodos podem ser utilizados, nomeadamente algoritmos do gradiente, como o algoritmo de Bell-Sejnowski (Bell e Sejnowski, 1995b) e em especial o algoritmo do gradiente natural (Amari *et al.*, 1996). De entre os algoritmos do ponto fixo, salienta-se uma versão do FastICA que maximiza mais rapidamente a verosimilhança e com maior precisão (Hyvärinen *et al.*, 2001b). Para além da rapidez de convergência ser otimizada consoante a escolha de determinadas matrizes diagonais, outra das vantagens do FastICA é que consegue estimar componentes independentes (quer sejam sub-Gaussianas ou super-Gaussianas) sem ter que recorrer a quaisquer outros passos.

3.5.4 Princípio do *Infomax*

Um princípio de estimação para o modelo ACI que possui uma relação próxima com a máxima verosimilhança é o princípio do *Infomax*, o qual é baseado na maximização da entropia de saída, ou fluxo de informação, de uma rede neuronal⁸ com saídas não lineares (Bell e Sejnowski, 1995b).

Admita-se que \mathbf{x} é o vetor aleatório (das misturas) de entrada na rede neuronal e que as saídas são da forma

$$y_i = \phi_i(\mathbf{w}_i^T \mathbf{x}) + \mathbf{n},$$

onde ϕ_i são funções escalares não lineares, \mathbf{w}_i são os pesos dos vetores dos neurónios e \mathbf{n} é o vetor aditivo que segue uma distribuição Normal de ruído branco. Pretende-se maximizar a

⁸Rede neuronal é um processamento não linear distribuído dos dados de entrada e consiste na interligação de unidades computacionais simples chamadas neurónios.

entropia das saídas através de

$$H(\mathbf{y}) = H(\phi_1(\mathbf{w}_1^T \mathbf{x}), \dots, \phi_n(\mathbf{w}_n^T \mathbf{x})), \quad (3.21)$$

considerando para tal o fluxo de informação de uma rede neuronal. Para que a transmissão da informação seja eficiente é preciso que se maximize a informação mútua entre as entradas \mathbf{x} e as saídas \mathbf{y} . No entanto este problema só é significativo desde que exista alguma informação perdida na transmissão e portanto admite-se existir algum ruído na rede (Hyvärinen *et al.*, 2001b). Assim a maximização desta informação mútua é equivalente à maximização da entropia de saída de (3.21).

Nota 3.5.7 Ao admitir que o ruído tem variância nula e utilizando a fórmula clássica da entropia de uma transformação tem-se

$$H(\phi_1(\mathbf{w}_1^T \mathbf{x}), \dots, \phi_n(\mathbf{w}_n^T \mathbf{x})) = H(\mathbf{x}) + E \left(\log \left| \det \frac{\partial \mathbf{F}}{\partial \mathbf{W}}(\mathbf{x}) \right| \right), \quad (3.22)$$

onde $\mathbf{F}(\mathbf{x}) = (\phi_1(\mathbf{w}_1^T \mathbf{x}), \dots, \phi_n(\mathbf{w}_n^T \mathbf{x}))$ denota a função definida pela rede neuronal.

Ao calcular a derivada presente em (3.22), obtém-se (Bell e Sejnowski, 1995b)

$$E \left(\log \left| \det \frac{\partial \mathbf{F}}{\partial \mathbf{W}}(\mathbf{x}) \right| \right) = \sum_i E \left[\log \left(\phi'_i(\mathbf{w}_i^T \mathbf{x}) \right) \right] + \log |\det \mathbf{W}|,$$

que claramente é da mesma forma que (3.20), com as funções densidade das CIs substituídas pelas funções ϕ'_i . Se as funções ϕ_i usadas na rede neuronal forem escolhidas como sendo as funções de distribuição que correspondam às funções densidade f_i , então nestas condições a entropia de saída será igual à verosimilhança. Consequentemente o *Infomax* é equivalente à estimação de máxima verosimilhança (Cardoso, 1997).

3.5.5 Minimização da Informação Mútua

Com base ainda na teoria da informação, surge uma outra abordagem para a estimação do modelo ACI, a qual se prende com a minimização da informação mútua proposta por Comon (1994).

A motivação para esta abordagem prende-se com o facto de, em muitos casos, não ser possível admitir que os dados sejam descritos por um modelo ACI (Hyvärinen *et al.*, 2001b). Portanto, neste sentido, o que se pretende ter é uma medida genérica da dependência das componentes de

um vetor aleatório. Como tal, pode definir-se o modelo ACI como uma decomposição linear de forma a minimizar a medida dessa dependência, nomeadamente através da informação mútua, a qual é uma medida de informação teórica de dependência estatística. Uma das principais vantagens da informação mútua é que esta permite unificar muitos princípios de estimação, como é o caso da estimação de máxima verosimilhança e da maximização da não normalidade.

A informação mútua é uma medida da dependência entre variáveis aleatórias. É não negativa e em particular é zero se e só se as variáveis forem estatisticamente independentes. Na informação mútua tem-se em conta toda a estrutura de dependência das variáveis, e não apenas a covariância, como na ACP. Assim, a informação mútua pode usar-se como um critério para implementar a ACI. Define-se a ACI aplicada a um vetor aleatório \mathbf{x} como uma transformação invertível dada por $\mathbf{s} = \mathbf{W}\mathbf{x}$, onde a matriz \mathbf{W} é determinada de forma a que a informação mútua das componentes transformadas s_i seja minimizada.

Definição 3.5.8 *Tendo em conta (3.12), a informação mútua I , entre n variáveis aleatórias, y_i , $i = 1, 2, \dots, n$, de um vetor aleatório \mathbf{y} , define-se por*

$$I(y_1, y_2, \dots, y_n) = \sum_{i=1}^n H(y_i) - H(\mathbf{y}),$$

onde H representa a entropia diferencial (ver equação (3.11)).

Através da fórmula (3.13) da neguentropia, para uma transformação linear invertível da forma $\mathbf{y} = \mathbf{W}\mathbf{x}$, a informação mútua é dada pela expressão

$$I(y_1, y_2, \dots, y_n) = \sum_{i=1}^n H(y_i) - H(\mathbf{x}) - \log |\det \mathbf{W}|. \quad (3.23)$$

Nota 3.5.9 *Se as variáveis y_i ($i = 1, 2, \dots, n$) forem não correlacionadas com variâncias unitárias, a entropia e a neguentropia diferem apenas de uma constante e no sinal, obtendo-se assim a relação fundamental entre a neguentropia, J , e a informação mútua, I , (Hyvärinen et al., 2001b)*

$$I(y_1, y_2, \dots, y_n) = \text{const.} - \sum_{i=1}^n J(y_i). \quad (3.24)$$

Portanto, encontrar uma transformação linear invertível \mathbf{W} que minimize a informação mútua é aproximadamente equivalente a encontrar as direções nas quais a neguentropia é maximizada. Uma vez que a neguentropia é uma medida de não normalidade, com base em (3.24)

verifica-se que a estimação da ACI por minimização da informação mútua é equivalente a maximizar a soma das não normalidades das estimativas das CIs, desde que as estimativas sejam não correlacionadas.

Em termos práticos existem algumas diferenças entre minimização da informação mútua e a procura das direções nas quais a não normalidade é maximizada (Hyvärinen *et al.*, 2001b):

- A neguentropia, entre outras medidas de não normalidade, permite que as CIs sejam estimadas uma a uma, desde que se possa considerar o máximo de não normalidade de uma única projeção $\mathbf{b}^T \mathbf{x}$, enquanto que este processo não é possível com a informação mútua e com a verosimilhança (uma vez que a informação deve ser contemplada como uma medida global no conjunto das variáveis).
- Ao utilizar a não normalidade, força-se a que as estimativas das CIs sejam não correlacionadas, o que não é necessário na aplicação da informação mútua uma vez que pode usar-se a forma apresentada em (3.23).

A máxima verosimilhança e a informação mútua também estão estritamente relacionadas.

Tendo em conta a log-verosimilhança dada por (3.19) e aplicando o valor esperado a (3.20) vem (Yang e Amari, 1997)

$$\frac{1}{K} E(\log L(\mathbf{W})) = \sum_{i=1}^n E(\log f_i(\mathbf{w}_i^T \mathbf{x})) + \log |\det \mathbf{W}|.$$

Se considerarmos o logaritmo de cada uma das funções densidade f_i denotado por $G_i(y_i) = \log f_i(y_i)$, então (3.23) aproxima-se por

$$I(y_1, y_2, \dots, y_n) = - \sum_{i=1}^n E(G_i(y_i)) - \log |\det \mathbf{W}| - H(\mathbf{x}).$$

Portanto, esta aproximação é praticamente igual à aproximação da máxima verosimilhança (3.20) exceto no que se refere aos sinais e à constante aditiva $H(\mathbf{x})$, e constitui uma alternativa para a aproximação da informação mútua em relação aos métodos de aproximação que utilizam a neguentropia (Hyvärinen *et al.*, 2001b).

Os algoritmos utilizados para a minimização da informação mútua são essencialmente os que são utilizados para a maximização da não normalidade ou para a estimação de máxima verosimilhança. Para uma maior abordagem da relação entre informação mútua e máxima verosimilhança, assim como alguns dos algoritmos relacionados podem ser consultados em Yang e Amari (1997).

3.5.6 Métodos Tensoriais

A utilização dos chamados métodos tensoriais, nomeadamente através de tensores cumulantes de ordem superior, constitui uma variante diferente de estimação do modelo ACI, quando comparada com os métodos abordados nas Subsecções anteriores.

Enquanto que os tensores podem ser generalizações de matrizes ou simplesmente operadores lineares, os tensores cumulantes são generalizações da matriz de covariâncias. A matriz de covariâncias é considerada o tensor cumulante de segunda ordem. As propriedades principais acerca de cumulantes podem consultar-se no Apêndice C.4.

Na ACP, a decomposição espectral da matriz de covariâncias é utilizada para branquear os dados, ou seja, para obter as CPs não correlacionadas entre si forçando a que as correlações de segunda ordem sejam nulas (obtendo v. a. não correlacionadas). Na sequência deste procedimento e de uma forma mais genérica, o tensor cumulante de quarta ordem pode ser usado para obrigar a que os cumulantes de quarta ordem sejam nulos ou quanto muito o mais pequenos possível. Esta forma de descorrelação de ordem superior fornece uma série de métodos para a estimação da ACI.

O tensor cumulante de quarta ordem é um operador linear a quatro dimensões cujas entradas são dadas pelos cumulantes cruzados de quarta ordem dos dados (de tamanho n) e é denotado por $cum(x_i, x_j, x_k, x_l)$ com $i, j, k, l = 1, 2, \dots, n$. Este operador é análogo ao caso a duas dimensões da matriz de covariâncias cujos elementos são $cov(x_i, x_j)$.

Definição 3.5.10 *Para o caso do tensor define-se uma transformação linear no espaço linear das matrizes de dimensão $n \times n$. O elemento na posição (i, j) da matriz dada pela transformação \mathbf{F} é definido por*

$$\mathbf{F}_{ij}(\mathbf{M}) = \sum_{kl}^n m_{kl} \, cum(x_i, x_j, x_k, x_l),$$

onde m_{kl} são os elementos da matriz \mathbf{M} que é transformada.

Os cumulantes de quarta ordem das misturas contêm praticamente toda a informação proveniente das estatísticas de quarta ordem subjacente aos dados, e podem ser usados para definir um tensor como uma generalização da matriz de covariâncias. Portanto, neste contexto, pode aplicar-se a decomposição espectral ao tensor cumulante e como qualquer operador linear simétrico, o tensor cumulante também possui uma decomposição espectral específica.

Apela-se aqui à atenção para evitar confusão entre a decomposição espectral sobre o tensor abordada de seguida e a decomposição espectral usual, sobre a matriz de variâncias-covariâncias, utilizada no Capítulo da ACP.

Definição 3.5.11 *Uma matriz própria do tensor é uma matriz \mathbf{M} tal que $\mathbf{F}(\mathbf{M}) = \lambda\mathbf{M}$, ou seja, para cada elemento de \mathbf{M} tem-se $\mathbf{F}_{ij}(\mathbf{M}) = \lambda\mathbf{M}_{ij}$. A constante λ designa-se valor próprio.*

No modelo ACI, tendo em conta os dados branqueados em (3.6), vem

$$\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s} = \mathbf{W}^T\mathbf{s},$$

onde \mathbf{W}^T é a matriz das misturas branqueadas. A estrutura do tensor cumulante de \mathbf{z} pode ser extraída da decomposição espectral, uma vez que cada matriz da forma $\mathbf{M} = \mathbf{w}_m\mathbf{w}_m^T$ é uma matriz própria do tensor para $m = 1, 2, \dots, n$. Por outro lado, os respetivos valores próprios são dados pela curtose das CIs enquanto que os restantes valores próprios do tensor são nulos (Hyvärinen *et al.*, 2001b).

No que se refere a algoritmos (alguns dos quais serão discutidos adiante na Secção 3.7) que utilizam o tensor, o mais simples é o método da potência o qual acaba por ser similar ao algoritmo FastICA com a não linearidade cúbica. No caso da decomposição espectral apresentar valores próprios iguais do tensor, uma das formas de resolver o problema é recorrendo à diagonalização conjunta aproximada das matrizes próprias, através do algoritmo JADE (*Joint Approximate Diagonalization of Eigenmatrices*) proposto por Cardoso e Souloumiac (1993). Este algoritmo assenta na escolha das matrizes próprias do tensor cumulante, uma vez que para se obter a informação essencial dos cumulantes basta seleccionar apenas n matrizes cujo subespaço gerado é o mesmo que o do tensor cumulante. Este é o princípio básico do algoritmo. Tal como outros métodos que apresentam problemas semelhantes ao usarem tensores no contexto da decomposição espectral, este algoritmo também não pode ser utilizado para espaços com elevado número de variáveis, segundo Hyvärinen *et al.*, (2001b).

Um método próximo do JADE que se designa por algoritmo FOBI (*Fourth-Order Blind Identification*), foi introduzido por Cardoso (1989b) e consiste na decomposição espectral efetuada a partir da matriz de correlações ponderada. Contudo o FOBI só pode ser aplicado para os casos em que as curtoses das CIs são todas diferentes, ou quanto muito, apenas quando algumas CIs possuem curtoses idênticas, de forma a estimar as CIs cujas curtoses sejam diferentes. Mesmo

apesar desta restrição, presume-se que este algoritmo seja o mais simples e mais eficiente método para efetuar a ACI (Hyvärinen *et al.*, 2001b).

3.6 Métodos que Usam uma Estrutura Temporal

Nas Secções consideradas anteriormente, o modelo linear ACI assenta em misturas de variáveis aleatórias independentes. No entanto, em muitas aplicações, o que se mistura não são as variáveis aleatórias, mas os sinais ao longo do tempo, cujas observações constituem as designadas séries temporais. Estes casos não estão de acordo com o modelo básico da ACI, no qual as amostras (misturas) de \mathbf{x} não têm uma ordem específica, uma vez que podemos trocá-las de qualquer forma que não afetam a validade do modelo nem os métodos de estimação apresentados. Porém, ao trabalhar com misturas de CIs que representam séries temporais, a situação é diferente pois convém que a estrutura temporal seja preservada na implementação do modelo.

Se as CIs forem sinais ao longo do tempo devem conter muito mais estrutura do que simples variáveis aleatórias, como é o caso das autocovariâncias, isto é, das covariâncias entre os valores da série em diferentes pontos desfasados no tempo. Tais estatísticas adicionais podem ser usadas para melhorar a estimação do modelo, e nesse sentido, esta informação pode tornar possível a estimação do modelo em casos onde os métodos básicos da ACI não podem estimá-lo, como por exemplo, se as CIs forem Normais mas correlacionadas ao longo do tempo.

De acordo com a Subsecção 1.4.1, e tendo em conta a noção de processo estocástico, nesta Secção considera-se a estimação do modelo ACI quando as CIs são séries temporais, $s_i(t)$, $t = 1, 2, \dots$, onde t é o índice temporal. Portanto t passa a ter um significado mais específico ao definir uma ordem em termos de estrutura de cada CI e o modelo (3.1) passa a ser representado por

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),$$

onde se admite que \mathbf{A} é a matriz quadrada das misturas e que as CIs s_i continuam a ser independentes. Por outro lado, as CIs não são obrigadas a seguir a não normalidade.

Passamos a descrever de seguida alguns pressupostos em relação à estrutura temporal subjacente das CIs, os quais permitem a estimação do modelo sendo considerados como alternativas ao pressuposto da não normalidade (Hyvärinen *et al.*, 2001b):

1. Admite-se que as CIs têm diferentes valores de autocovariâncias e todos diferentes de zero.

2. Considera-se o caso em que as variâncias das CIs são não estacionárias.
3. Analisa-se a complexidade de Kolmogoroff como uma estrutura para a ACI com misturas correlacionadas no tempo.

3.6.1 Separação por Autocovariâncias

Autocovariâncias Como uma Alternativa à Não Normalidade

Uma maneira relativamente simples de abordar a estrutura temporal é através de autocovariâncias, ou seja, pelas covariâncias entre os valores de uma série temporal em diferentes pontos no tempo, que são dadas por $cov(x_i(t), x_i(t - \tau))$, para uma certa constante de desfasamento $\tau \neq 0$, e com $i = 1, 2, \dots, n$. Para o caso em que os dados apresentam dependências no tempo, as autocovariâncias são diferentes de zero. Para além das autocovariâncias, consideram-se as covariâncias entre duas séries temporais, $cov(x_i(t), x_j(t - \tau))$, para $i \neq j$.

Definição 3.6.1 *Para um dado desfasamento τ e para o vetor das misturas \mathbf{x} , as estatísticas anteriores podem ser agrupadas na matriz de covariâncias dos desfasamentos no tempo dada por*

$$C_\tau^{\mathbf{x}} = E(\mathbf{x}(t)\mathbf{x}(t - \tau)^T). \quad (3.25)$$

Como se sabe, no branqueamento, encontrar uma matriz \mathbf{V} tal que $\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t)$ não é suficiente para estimar as CIs uma vez que existe uma infinidade de diferentes matrizes \mathbf{V} que devolvem componentes não correlacionadas. Esta é a razão pela qual no modelo básico da ACI tem que ser usada a não normalidade nas CIs, como acontece por exemplo, na variante de estimação do modelo para minimização das dependências de ordem superior através da informação mútua.

De acordo com Tong *et al.* (1991), a informação contida em $C_\tau^{\mathbf{x}}$ poderá ser usada em vez da informação de ordem superior para obter as CIs. Para isso, deverá encontrar-se uma matriz \mathbf{B} , de transformação dos dados, que para além de forçar as covariâncias instantâneas de $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ a tenderem para zero, também as covariâncias dos desfasamentos deverão tender para zero de tal forma que $E(y_i(t)y_j(t - \tau)) = 0$ para todos os valores de i, j e τ , uma vez que se pretende que as covariâncias dos desfasamentos das CIs sejam todas nulas devido à sua independência. Através destas covariâncias conseguir-se-á bastante informação adicional para estimar o modelo, sem necessidade de recorrer a informação de ordem superior.

Usando um Desfasamento no Tempo

No caso mais simples em que se usa apenas um desfasamento no tempo (geralmente denotado por $\tau = 1$), pode ser formulado um algoritmo para encontrar uma matriz que anule as covariâncias instantâneas e as que correspondem ao desfasamento τ em causa.

Considere-se o modelo $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ e os dados branqueados dados por

$$\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t) = \mathbf{V}\mathbf{A}\mathbf{s}(t).$$

De acordo com a Subsecção 3.4.2, em que $\mathbf{V}\mathbf{A} = \mathbf{W}^T$ é uma matriz ortogonal de separação, vem $\mathbf{z}(t) = \mathbf{W}^T\mathbf{s}(t)$, e logo $\mathbf{W}\mathbf{z}(t) = \mathbf{s}(t)$. Para um dado desfasamento vem $\mathbf{W}\mathbf{z}(t - \tau) = \mathbf{s}(t - \tau)$.

Tendo em conta (3.25), a matriz de covariâncias dos desfasamentos para o vetor $\mathbf{z}(t)$ é dada por

$$C_\tau^{\mathbf{z}} = E(\mathbf{z}(t)\mathbf{z}(t - \tau)^T), \quad (3.26)$$

e a matriz de covariâncias dos desfasamentos para o vetor $\mathbf{s}(t)$ é dada por

$$C_\tau^{\mathbf{s}} = E(\mathbf{s}(t)\mathbf{s}(t - \tau)^T). \quad (3.27)$$

Uma versão da matriz de covariâncias dos desfasamentos de (3.26) pode ser dada pela matriz simétrica (ver Apêndice C.1.2)

$$\bar{C}_\tau^{\mathbf{z}} = \frac{1}{2} \left(C_\tau^{\mathbf{z}} + (C_\tau^{\mathbf{z}})^T \right).$$

A partir de (3.26) e de (3.27), vem

$$\begin{aligned} \bar{C}_\tau^{\mathbf{z}} &= \frac{1}{2} \left[E(\mathbf{z}(t)\mathbf{z}(t - \tau)^T) + (E(\mathbf{z}(t)\mathbf{z}(t - \tau)^T))^T \right] \\ &= \frac{1}{2} \left[E(\mathbf{W}^T\mathbf{s}(t)(\mathbf{W}^T\mathbf{s}(t - \tau))^T) + \left(E(\mathbf{W}^T\mathbf{s}(t)(\mathbf{W}^T\mathbf{s}(t - \tau))^T) \right)^T \right] \\ &= \frac{1}{2} \mathbf{W}^T \left[E(\mathbf{s}(t)\mathbf{s}(t - \tau)^T) + E(\mathbf{s}(t - \tau)\mathbf{s}(t)^T) \right] \mathbf{W} \\ &= \frac{1}{2} \mathbf{W}^T \left[C_\tau^{\mathbf{s}} + (C_\tau^{\mathbf{s}})^T \right] \mathbf{W}. \end{aligned}$$

Devido à independência de $s_i(t)$, a matriz $C_\tau^{\mathbf{s}}$ é diagonal (e consequentemente simétrica), e portanto

$$\frac{1}{2} \left(C_\tau^{\mathbf{s}} + (C_\tau^{\mathbf{s}})^T \right) = C_\tau^{\mathbf{s}}.$$

Como tal estabelece-se a relação

$$\bar{C}_\tau^{\mathbf{z}} = \mathbf{W}^T C_\tau^{\mathbf{s}} \mathbf{W},$$

ou seja, as matrizes $C_\tau^{\mathbf{s}}$ e \mathbf{W} constituem a decomposição em valores e vetores próprios de $\bar{C}_\tau^{\mathbf{z}}$.

Desta forma a decomposição em valores e vetores próprios fica bem definida e relativamente simples de implementar em qualquer algoritmo quando se utiliza a matriz simétrica $\bar{C}_\tau^{\mathbf{z}}$ em vez da usual matriz de covariâncias dos desfasamentos.

Um algoritmo muito simples e rápido de aplicar, que será descrito na Secção 3.7, denomina-se por AMUSE (*Algorithm for Multiple Unknown Signals Extraction*) e poderá ser consultado em Tong *et al.* (1991). Este algoritmo possui uma desvantagem na aplicabilidade: no caso em que alguns valores próprios da matriz de covariâncias dos desfasamentos no tempo são iguais, os correspondentes vetores próprios não são únicos, e logo as correspondentes CIs não são estimadas. Os valores próprios são dados por $cov(s_i(t), s_i(t - \tau))$, e como tal estes são distintos se e só se as covariâncias desfasadas forem diferentes para todas as CIs.

Um outro algoritmo similar foi proposto por Molgedey e Schuster (1994), o qual também inclui informação acerca da estrutura temporal dos dados, requerendo que as correlações para certos instantes e as correlações desfasadas no tempo entre as diferentes misturas desapareçam. A estimação das CIs prende-se novamente com a decomposição em valores e vetores próprios que requiere a diagonalização simultânea de duas matrizes simétricas.

Extensão para Vários Desfasamentos no Tempo

Quando se consideram vários desfasamentos no tempo é possível utilizar uma certa extensão do algoritmo AMUSE por forma a melhorar o seu desempenho. Como tal é suficiente que as covariâncias sejam diferentes para um desses desfasamentos.

Pretendendo-se diagonalizar simultaneamente todas as matrizes de covariâncias dos desfasamentos, serão formuladas funções que indicam o grau de diagonalização obtido. Para avaliar o quanto uma matriz é ou não similar a uma matriz diagonal pode minimizar-se a soma de quadrados dos elementos que se encontram fora da diagonal. Essa minimização pode ser obtida com base no gradiente descendente ou através de métodos adaptados para a decomposição em valores e vetores próprios, à diagonalização aproximada de várias matrizes em simultâneo. Um dos algoritmos mais aplicados neste contexto, proposto por Belouchrani *et al.* (1997), denomina-se por SOBI (*Second-Order Blind Identification*). De entre outros trabalhos envolvendo assuntos relacionados com as autocovariâncias salientam-se Amari (2000) e Yeredor (2000).

3.6.2 Separação por Não Estacionariedade de Variâncias

A estrutura temporal dos sinais permite que a ACI seja executada utilizando uma aproximação alternativa à não estacionariedade das variâncias das CIs, admitindo que estas mudam suavemente ao longo do tempo. Matsuoka *et al.* (1995) foram os primeiros autores a abordar a não estacionariedade para separar as CIs, propondo para tal um algoritmo específico.

A separação de séries temporais não estacionárias pode ser levada a cabo através de uma variante de autocorrelações. Segundo Matsuoka *et al.* (1995), caso se encontre uma matriz \mathbf{B} de tal modo que as CIs de $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ sejam não correlacionadas em cada valor de tempo t então as CIs serão estimadas, pois devido à não estacionariedade, a covariância de $\mathbf{y}(t)$ depende de t e ao forçar as CIs a serem não correlacionadas para cada t , obtém-se uma condição muito mais forte que o branqueamento.

A não correlação local de $\mathbf{y}(t)$ pode ser avaliada utilizando as mesmas medidas que são usadas nos algoritmos da *Extensão para Vários Desfasamentos no Tempo*, assim como através de um algoritmo similar ao de Matsuoka *et al.* (1995).

A não estacionariedade da variância de uma série temporal $\mathbf{y}(t)$ pode também ser medida com base na correlação ao longo do tempo das energias $E(y(t)^2 y(t - \tau)^2)$, onde τ é um desfasamento que geralmente é igual à unidade. O cumulante correspondente à correlação de energias é dado pelo cumulante cruzado de 4ª ordem. Um algoritmo do ponto fixo semelhante ao FastICA pode ser usado para maximizar a não estacionariedade da variância medida pelo cumulante cruzado.

3.6.3 Princípios de Separação Unificados

Os dois princípios de separação apresentados (autocovariâncias e variâncias não estacionárias) complementam o princípio da não normalidade, o qual representa a base da estimação do modelo básico de ACI.

A escolha do princípio a utilizar deverá depender dos dados que se pretendem analisar e do objetivo do estudo, uma vez que critérios diferentes levam a considerar diferentes pressupostos nos próprios dados. Numa grande parte dos casos os dados não possuem uma estrutura temporal associada, pelo que a ordem das amostras é arbitrária e não tem qualquer significado. Neste caso só o modelo básico de ACI baseado na não normalidade pode ser usado. Por outro lado, as alternativas só são significativas se os dados vierem de uma fonte que tenha uma estrutura

temporal intrínseca.

Segundo Hyvärinen *et al.* (2001b), quando os dados possuem uma estrutura temporal, o modelo básico de ACI funciona muitas vezes bem e apesar de não usar tal informação não é afetado por esta. No entanto a estimação do modelo poderá ficar longe de ser a ideal por não usar toda essa estrutura dos dados quando existe.

Complexidade de Kolmogoroff como uma Estrutura de Unificação

É possível combinar diferentes tipos de informação, como por exemplo não normalidade com autocorrelações.

A complexidade de Kolmogoroff proporciona uma forma de medir a complexidade de um algoritmo.

De acordo com Pajunen (1998), a complexidade de Kolmogoroff para uma sequência de caracteres $x^n = x_1x_2...x_n$ é dada por

$$K_U(x^n) = \min_{p_U(x^n)} |p_U(x^n)|,$$

onde $p_U(.)$ é um algoritmo ou programa computacional, que permite gerar a sequência x^n , no universo U de todos os programas, e onde o módulo representa o comprimento da sequência usada por cada algoritmo.

A complexidade de Kolmogoroff mede a quantidade de estrutura de um sinal através da quantidade de compressão que é possível na codificação do sinal, de acordo com o comprimento do código do mesmo. Em termos computacionais é uma medida que pretende encontrar o melhor esquema de codificação para o respetivo sinal (Pajunen, 1998). A complexidade de Kolmogoroff permite englobar um tipo de estrutura mais geral em que praticamente todos os pressupostos podem ser considerados, desde as estruturas de tempo como as autocorrelações e a não estacionariedade até estruturas que usam a teoria da informação.

3.7 Algoritmos para Implementar a ACI

De entre os algoritmos existentes para a aplicação da ACI, e de acordo com as variantes de estimação do modelo, destacamos alguns que consideramos relevantes e que foram citados ao

longo do Capítulo. No *site* ICA CENTRAL⁹ podem ser consultados alguns algoritmos e *packages* para implementar a ACI ou a BSS em *softwares*.

Como todos os algoritmos são constituídos pelos passos 1 e 2 de centragem e de branqueamento, passamos a descrever apenas os passos seguintes.

3.7.1 AMUSE

O algoritmo AMUSE é utilizado nos casos em que se admite que as CIs possuem alguma dependência temporal, procurando encontrar uma matriz de transformação dos dados que anule as covariâncias instantâneas e também as covariâncias desfasadas.

O algoritmo descreve-se sumariamente através dos passos seguintes (Tong *et al.*, 1991), cujo código de implementação no *software R* pode ser consultado no Apêndice D:

3. Escolher o número m ($m \leq n$) de componentes independentes a estimar.
4. Selecionar um desfasamento no tempo, τ , para aplicar a decomposição em valores e vetores próprios da matriz $\bar{C}_\tau^{\mathbf{z}} = [C_\tau^{\mathbf{z}} + (C_\tau^{\mathbf{z}})^T] / 2$, onde $C_\tau^{\mathbf{z}} = E \{ \mathbf{z}(t) \mathbf{z}(t - \tau)^T \}$ é a matriz de covariâncias dos desfasamentos no tempo para um dado desfasamento τ .
5. Fazer a decomposição espectral da matriz $\bar{C}_\tau^{\mathbf{z}}$, em que \mathbf{W} é a respetiva matriz de vetores próprios.
6. A partir da matriz \mathbf{W} , obtida no passo 5, cujas linhas são os vetores próprios, e através de operações matriciais adequadas, estimar a matriz \mathbf{A} dos coeficientes das misturas e o vetor \mathbf{s} das componentes independentes.

3.7.2 FastICA

O algoritmo FastICA é baseado na maximização da não normalidade e permite ainda muitas variantes e extensões.

Para estimar várias CIs, consideremos duas versões do algoritmo FastICA (Hyvärinen, 1999a); uma que utiliza a ortogonalização deflacionária e outra que utiliza a ortogonalização simétrica.

⁹<http://perso.telecom-paristech.fr/~cardoso/icacentral/index.html>

Ortogonalização Deflacionária

A versão deste algoritmo, usando o método da orthogonalização deflacionária (segundo o método da orthogonalização de Gram-Schmidt) onde as componentes são estimadas uma a uma, descreve-se sumariamente por:

3. Escolher o número m ($m \leq n$) de componentes independentes a estimar. Aplicar a primeira iteração, tomando $k \leftarrow 1$.
4. Escolher aleatoriamente um vetor inicial com norma unitária para \mathbf{w}_k .
5. Implementar a iteração básica do ponto fixo $\mathbf{w}_k \leftarrow E \{ \mathbf{z}g(\mathbf{w}_k^T \mathbf{z}) \} - E \{ g'(\mathbf{w}_k^T \mathbf{z}) \} \mathbf{w}_k$, onde g é uma função contraste que serve para obter uma aproximação da neguentropia, como por exemplo as derivadas das funções referidas em (3.16), (3.17) e (3.18).
6. Efetuar a orthogonalização de Gram-Schmidt através da iteração $\mathbf{w}_k \leftarrow \mathbf{w}_k - \sum_{j=1}^{k-1} (\mathbf{w}_k^T \mathbf{w}_j) \mathbf{w}_j$.
7. Fazer a iteração $\mathbf{w}_k \leftarrow \mathbf{w}_k / \|\mathbf{w}_k\|$.
8. Se \mathbf{w}_k não convergir para os máximos da função contraste, sob a hipótese do modelo da ACI definido em (3.1), voltar ao passo 5.
9. Tomar a iteração $k \leftarrow k + 1$. Se $k \leq m$, voltar ao passo 4.
10. Estimar a matriz \mathbf{A} dos coeficientes das misturas (a partir da inversa ou pseudo-inversa da matriz \mathbf{W}) e o vetor \mathbf{s} das componentes independentes.

Ortogonalização Simétrica

Caso se pretenda estimar as componentes em simultâneo (em paralelo), utiliza-se o método da orthogonalização simétrica e a versão do algoritmo é descrita sumariamente por:

3. Escolher o número m ($m \leq n$) de componentes independentes a estimar.
4. Escolher os valores iniciais para cada vetor unitário \mathbf{w}_i ($i = 1, \dots, m$). Efetuar a orthogonalização de Gram-Schmidt através da iteração $\mathbf{w}_k \leftarrow \mathbf{w}_k - \sum_{j=1}^{k-1} (\mathbf{w}_k^T \mathbf{w}_j) \mathbf{w}_j$.

5. Para cada $i = 1, \dots, m$, implementar a iteração básica do ponto fixo $\mathbf{w}_i \leftarrow E \{ \mathbf{z} g(\mathbf{w}_i^T \mathbf{z}) \} - E \{ g'(\mathbf{w}_i^T \mathbf{z}) \} \mathbf{w}_i$, onde g é uma função contraste que serve para obter uma aproximação da neguentropia, como por exemplo as derivadas das funções referidas em (3.16), (3.17) e (3.18).
6. Após cada iteração ortogonalizar simetricamente a matriz $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)^T$ através de $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2} \mathbf{W}$.
7. Se não convergir, ou seja, se a matriz $\mathbf{W}\mathbf{W}^T$ não for praticamente idêntica à matriz identidade, voltar ao passo 5.
8. Estimar a matriz \mathbf{A} dos coeficientes das misturas (a partir da inversa ou pseudo-inversa da matriz \mathbf{W}) e o vetor \mathbf{s} das componentes independentes.

Uma alternativa iterativa mais simples ao passo 6 foi proposta por Hyvärinen (1999a) em que

- $\mathbf{W} \leftarrow \mathbf{W} / \|\mathbf{W}\|$.
- $\mathbf{W} \leftarrow \frac{3}{2} \mathbf{W} - \frac{1}{2} \mathbf{W}\mathbf{W}^T \mathbf{W}$.
- Se a matriz $\mathbf{W}\mathbf{W}^T$ não for praticamente idêntica à matriz identidade, voltar ao passo 4.

Relativamente à convergência do algoritmo FastICA, se as distribuições de probabilidade das CIs forem simétricas então este apresenta uma convergência cúbica, e nos outros casos a convergência é de natureza quadrática (Hyvärinen *et al.*, 2001b). Para o caso particular em que a curtose é usada como função contraste, a convergência também é cúbica. Assim, o FastICA apresenta uma convergência mais rápida comparativamente aos tradicionais métodos do gradiente descendente, nos quais a convergência é apenas linear (Hyvärinen, 1999a).

3.7.3 FOBI

O algoritmo não iterativo FOBI (Cardoso, 1989b), considerado como uma extensão do algoritmo AMUSE, foi criado para a identificação cega através de momentos de quarta ordem e é constituído pelas etapas da ortonormalização e da ponderação quadrática. O FOBI é baseado na decomposição espectral efetuada a partir da matriz de correlações ponderada, de tal modo

que as CIs são facilmente identificadas como os seus vetores próprios. Este algoritmo aplica-se apenas nos casos em que as curtoses das CIs são todas distintas, ou seja, apresenta a limitação de não ser capaz de discriminar componentes com distribuições de probabilidade semelhantes.

Tendo em conta os dados do modelo de ACI branqueado, os passos para implementar o algoritmo são:

3. Formar a matriz de covariâncias ponderada $R = E(\mathbf{z}\mathbf{z}^T \|\mathbf{z}\|^2) = \mathbf{W}^T E(\mathbf{s}\mathbf{s}^T \|\mathbf{s}\|^2) \mathbf{W}$.
4. Usar a independência de \mathbf{s}_i , para obter a decomposição em valores e vetores próprios de $R = \mathbf{W}^T \text{diag}(E(\mathbf{s}_i^4) + n - 1) \mathbf{W}$.
5. Estimar a matriz \mathbf{A} dos coeficientes das misturas e o vetor \mathbf{s} das componentes independentes.

3.7.4 JADE

O algoritmo mais utilizado no que diz respeito a métodos tensoriais (Subsecção 3.5.6) é o JADE (Cardoso e Souloumiac, 1993). A estimação do modelo da ACI consiste em captar a informação obtida a partir da diagonalização conjunta aproximada das matrizes próprias quando os valores próprios do tensor forem iguais. Este algoritmo surge com o propósito de conseguir separar as componentes com curtoses idênticas, e colmatar assim uma das situações que o FOBI não permitia.

Segundo Hyvärinen *et al.* (2001b), para dados branqueados, a definição de cumulante pode ser escrita por $\mathbf{F}(\mathbf{M}) = E((\mathbf{z}^T \mathbf{M} \mathbf{z}) \mathbf{z} \mathbf{z}^T) - 2\mathbf{M} - \text{tr}(\mathbf{M})\mathbf{I}$. Pode-se definir em alternativa a matriz ponderada à custa do tensor como $R = \mathbf{F}(\mathbf{I})$, uma vez que $\mathbf{F}(\mathbf{I}) = E(\|\mathbf{z}\|^2 \mathbf{z} \mathbf{z}^T) - (n + 2)\mathbf{I}$, e a matriz identidade não altera a decomposição espectral de forma significativa.

Deste modo, a implementação do algoritmo JADE, encarado como uma generalização do algoritmo FOBI, é descrita abreviadamente por:

3. Estimar o tensor cumulante e efetuar a sua decomposição espectral. Diagonalizar o conjunto de matrizes próprias.
4. Estimar a matriz \mathbf{A} dos coeficientes das misturas e o vetor \mathbf{s} das componentes independentes.

3.7.5 A Escolha do Algoritmo

Na escolha do algoritmo mais adequado para implementar a ACI devem-se ponderar algumas regras como as descritas a seguir.

- Regra geral, pretende estimar-se mais do que uma componente independente e para tal é necessário aplicar várias vezes, qualquer um dos algoritmos que encontram uma componente independente, considerando diferentes pontos iniciais.
- Dos métodos, descritos na Subsecção 3.7.2, que permitem que se consiga a decorrelação, destacam-se os que usam a ortogonalização deflacionária, na qual as componentes são estimadas uma a uma e a ortogonalização simétrica onde todas as componentes são estimadas em simultâneo (em paralelo). Segundo Hyvärinen *et al.* (2001b), a ortogonalização simétrica é preferível e mais vantajosa, comparativamente com a ortogonalização deflacionária num grande número de aplicações, uma vez que os métodos deflacionários apresentam o inconveniente das estimativas dos erros aumentarem à medida que as componentes vão sendo estimadas.
- Para além do caso usual em que as misturas são consideradas lineares, podemos escolher a não linearidade utilizada em certos algoritmos, nos quais a inclusão de determinadas restrições acarreta dificuldades adicionais.
- Pode optar-se por um algoritmo que funciona em bloco, como o FastICA, caso os dados estejam todos disponíveis antes da estimação, ou por um algoritmo adaptativo que deve ser utilizado em aplicações onde a matriz das misturas pode mudar durante o processo, como por exemplo no processamento de sinal. Nesta situação é aconselhável recorrer a algoritmos que usam métodos do gradiente estocásticos.

3.8 Métodos de Ordenação das CIs

Após a implementação de um algoritmo, é necessário ordenar as CIs, com o intuito de as comparar ou de identificar as componentes que extraem a informação mais significativa do estudo. Ao aplicar a técnica de ACI através de determinados algoritmos tais como o FastICA, as CIs não saem ordenadas de acordo com qualquer critério, ao contrário do que acontece na ACP em

que as CPs são ordenadas de acordo com a ordem decrescente da variância explicada. Têm sido propostos alguns métodos de pós-processamento para a ordenação das CIs, podendo diferentes métodos fornecer diferentes ordens das componentes.

Consideremos uma breve descrição de alguns de entre os vários métodos existentes na literatura. Consideremos ainda a introdução de um novo critério de ordenação, passível de ser aplicado quando utilizamos a ACI e a ACP no mesmo conjunto de dados.

1. *(M1) - Maximização da curtose das CIs:*

O método ordena as CIs de acordo com a diminuição do módulo da curtose, quando esta é calculada em termos de afastamento em relação à curtose de uma variável aleatória com distribuição Normal.

Como exemplo, considere-se Cichocki *et al.* (1997) em que foi desenvolvido um algoritmo capaz de extrair as CIs numa ordem decrescente de acordo com os valores absolutos das suas curtoses. Numa outra perspetiva, a informação disponível à priori em relação à forma de ordenação pode ser incorporada como uma restrição na função objetivo ao utilizar-se a técnica da ACI com restrição (cICA), como pode ser visto no algoritmo proposto por Lu e Rajapakse (2003).

2. *(M2) - Maximização da norma dos vetores das colunas da matriz estimada das misturas:*

As CIs são ordenadas de acordo com a diminuição do valor da norma de cada um dos vetores das colunas da matriz estimada das misturas, A , que tem similaridade com a ordenação da ACP, uma vez que as normas dos vetores das colunas da matriz A fornecem as contribuições das CIs para as variâncias das variáveis observadas (Hyvärinen, 1999c).

3. *(M3) - Minimização da soma dos quadrados dos erros na reconstrução da matriz dos dados originais:*

Este método consiste em ordenar as CIs de acordo com um algoritmo que se baseia no aumento da soma dos quadrados dos erros das reconstruções da matriz dos dados originais. Os resíduos são obtidos através das diferenças entre os dados originais e as reconstruções efetuadas a partir de subconjuntos de componentes.

Uma vez que alguns dos métodos existentes na literatura abordam a ordenação baseada nas componentes de forma individual sem considerarem as interações entre elas, foi proposto um método alternativo por Cheung e Xu (2001) para ordenar as CIs de acordo com

as contribuições conjuntas na reconstrução dos dados, que consiste na aplicação de um algoritmo para encontrar a ordem ótima.

4. (*M4*) - *Análise das Correlações Canônicas (ACC)*:

As correlações canônicas existentes entre as reconstruções dos dados originais à custa das CIs e os próprios dados originais pode ser usada para ordenar as CIs. A ordenação é feita consoante a diminuição do valor das correlações canônicas entre todo o conjunto dos dados originais e cada um dos subconjuntos formados à custa das CIs. O processo inicia-se com a CI que estiver mais correlacionada com os dados originais. Após fixar a primeira CI e utilizando correlações entre conjuntos de variáveis, escolhemos o subconjunto de duas CIs que seja o mais correlacionado com os dados originais e que inclua a primeira CI que foi previamente fixada. Usando novamente o critério anterior da maior correlação, o processo é repetido até se obter a ordenação final.

Um método similar de aplicação da ACC pode ser encontrado em Youssef *et al.* (2003), onde se apresenta um algoritmo para analisar Imagens de Ressonância Magnética funcional, com vista a ordenar as CIs depois destas estarem estimadas. Nesse artigo foram considerados dois grupos importantes de sinais, uns relevantes para a ativação neuronal e outros relevantes para o ruído fisiológico.

5. (*M5*) - *Correlações entre as CPs (da ACP) e as CIs*:

Nos casos em que se pode aplicar a ACP em paralelismo com a técnica de ACI, o método de comparação das correlações existentes entre as CPs e as CIs pode ser apresentado como um novo critério alternativo para ordenar as CIs. Uma vez que, ao aplicar a técnica da ACP, as CPs saem ordenadas naturalmente segundo a ordem decrescente da sua variância, ao manter fixa a ordem das CPs, ordenam-se as CIs tendo em conta a maximização do valor de correlação com cada uma das CPs. Esta é uma forma natural de ordenar as CIs, pois geralmente cada CI está altamente correlacionada com uma única e distinta CP.

Num exemplo de aplicação de séries temporais para dados meteorológicos, Sebastião e Oliveira (2013) compararam determinadas técnicas multivariadas, entre as quais a ACI. Diferentes métodos de ordenação das CIs são utilizados, nomeadamente este novo método que sugere resultados satisfatórios na ordenação das CIs ao avaliar-se a forma como influencia a qualidade da reconstrução dos dados originais. Este trabalho está descrito em

pormenor na Secção 6.3.

Para além dos métodos mencionados anteriormente é de referir o trabalho de Üzümcü *et al.* (2003), onde se utilizaram três algoritmos diferentes (FastICA, Infomax e JADE) para aplicar quatro métodos de ordenação das CIs, que produziram resultados muito semelhantes entre si, e cuja descrição se apresenta de seguida.

6. (M6) - *Variância do histograma:*

Método para ordenar as CIs, onde a variância pode ser encarada como medida de energia das CIs, que é similar ao esquema da ACP, na qual as componentes estão ordenadas por ordem decrescente de grandeza, uma vez que a variância destas é igual ao respetivo valor próprio. Neste método todas as formas (associadas ao reconhecimento de imagens) são projetadas em cada uma das CIs. É construído um histograma dessas projeções, ao qual é ajustado uma curva Normal. A variância dessa distribuição Normal é estimada para cada CI, as quais são ordenadas da maior variância para a menor.

7. (M7) - *Não normalidade:*

As CIs são calculadas através da maximização da não normalidade das projeções nas componentes. A ordenação das CIs está de acordo com a ordem crescente dos *p-values* associados a testes de hipóteses de Shapiro-Wilk.

8. (M8) - *Alinhamento das formas com CIs:*

Neste método de ordenação as CIs são consideradas vetores que geram um espaço designado espaço forma. Este método é baseado no conceito do alinhamento das CIs e na posição de cada CI em relação aos restantes vetores que apontam numa direção específica (as formas). Um ângulo médio é calculado entre cada CI e o conjunto das formas, de modo que as CIs sejam ordenadas consoante o aumento do ângulo médio.

9. (M9) - *Variação da localização da forma:*

A ordenação das CIs é feita de acordo com a localização e amplitude da variação da forma.

Üzümcü *et al.* referem que, relativamente à ordenação das CIs para três conjuntos de dados do sistema circulatório, foi o método M9 que apresentou melhores resultados, quando comparado com M6, M7 e M8.

Zhang *et al.* (2008) apresentam um resumo dos métodos de ordenação das CIs existentes até então e introduzem um novo algoritmo do ponto fixo com restrição para implementar a ACI, o qual é uma ferramenta simples e prática para recuperar e ordenar simultaneamente as CIs resultantes.

3.9 Número de CIs a Reter

Na ACI, um tópico relevante na implementação de qualquer algoritmo prende-se com a escolha do número adequado de CIs que se devem reter no estudo.

Numa grande parte dos estudos admite-se que se conhece o número de fontes (CIs) que originam as n misturas, embora este na prática seja muitas vezes desconhecido. Segundo Karhunen (1996), se a relação sinal/ruído¹⁰ é satisfatória, podemos estimar o número q de CIs a partir da ACP clássica, o que na prática é efetuado através da estimação da matriz de covariâncias dos dados observados, seguida do cálculo dos seus valores e vetores próprios. Dos n valores próprios, os q maiores valores próprios que representam os sinais deverão ser claramente maiores que os restantes $n - q$ valores próprios associados ao ruído. Neste caso, retêm-se as q primeiras componentes correspondentes aos q maiores valores próprios. Contudo, Karhunen considera que se a potência do ruído for suficientemente grande ou se algumas das CIs forem relativamente fracas, então torna-se mais difícil estimar o valor correto de q apenas através da inspeção dos valores próprios. Para o caso ideal onde não haja ruído, apenas os q valores próprios mais elevados que descrevem os sinais são não nulos, e os restantes $n - q$ valores próprios associados ao ruído são nulos (Cichocki *et al.*, 1999).

Em 2000, Nadal *et al.* analisaram uma forma de processar a informação contida numa mistura linear de fontes independentes, nas quais algumas podem ser consideradas fortes comparativamente com as restantes designadas de fracas. Mostraram que, nessa situação, é vantajoso projetar os dados no espaço gerado pelas fontes fortes, de modo a extrair a informação principal,

¹⁰A relação sinal/ruído (Signal-to-Noise Ratio: SNR) é um termo usual em telecomunicações. O seu valor é dado pela razão da potência de um sinal que contém algum tipo de informação e a potência do ruído que se sobrepõe ao sinal.

É costume utilizar-se a relação em decibéis que é dada por $SNR_{dB} = 10 \cdot \log_{10}(SNR)$.

Quanto maior a relação SNR, melhor a qualidade do sinal recebido e menor é o efeito do ruído sobre a medição ou deteção do sinal.

e atendendo a que as fontes fracas que contribuem para o sinal possam ser consideradas como termos de ruído, tal como na ACP que permite projetar os dados nas CPs dominantes. Contudo, caso o número de fontes originais seja maior do que o número de misturas, os autores sugerem que é possível introduzir a ACI com o intuito de procurar um número de CIs dominantes maior ainda que o número de CPs dominantes.

Du *et al.*, (2006) desenvolveram trabalho de investigação na aplicação dos algoritmos Fast-ICA e JADE para a classificação de imagens por detecção remota. Embora o uso da ACI permita classificar objetos desconhecidos numa imagem também ela desconhecida, a sua elevada complexidade computacional impede, geralmente, a sua aplicação à análise de dados de elevada dimensão. Para resolver o problema da elevada dimensionalidade dos dados, aplicaram um procedimento de pré-processamento seguido da implementação da classificação através da ACI. Em vez da ACP usual, e por se tratar de dados de imagens, propuseram o uso de uma transformação de componentes principais de ruído ajustado, por forma a que os dados fossem reorganizados de acordo com a relação sinal/ruído, aliás um critério que consideram mais apropriado para ordenar imagens comparativamente à variância na ACP. Por conseguinte, a informação de uma imagem pode ser comprimida nas CPs dominantes de um modo mais eficaz, e a posterior aplicação dos algoritmos da ACI nessas CPs pode proporcionar uma melhor classificação das imagens.

3.10 Outros Modelos de ACI

Muitos são os modelos da ACI que têm vindo a ser desenvolvidos desde o aparecimento da técnica para além do modelo básico descrito na Secção 3.4. Nesta Secção, abordam-se de uma forma sintética outros modelos ou extensões da ACI.

3.10.1 ACI com Ruído

Na recolha de observações é natural que estas apresentem algum tipo de ruído. Durante a análise dos dados, e na presença de ruído, o uso de um certo modelo que não previa a influência do mesmo, pode ser inadequado ou pouco preciso.

A matriz das misturas do modelo é mais difícil de estimar na presença de ruído. No entanto, algumas abordagens podem reduzir o ruído dos dados antes mesmo de ser implementada a ACI. A filtragem simples de sinais temporais ou a redução da dimensionalidade dos dados através da

ACP são alguns exemplos úteis para contornar essa questão. No modelo da ACI com ruído, um outro problema prende-se com a estimação das CIs livres de ruído, a qual consiste na aplicação de novos métodos.

O modelo básico de ACI apresentado em (3.1) passa a incluir uma estrutura de ruído.

Definições 3.10.1 *Admitindo que o ruído é um termo aditivo, o modelo com ruído aditivo é descrito por*

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (3.28)$$

onde \mathbf{n} é o vetor n -dimensional de ruído.

O ruído presente na expressão (3.28) também pode ser designado por ruído do sensor, pelo facto das variáveis que representam o ruído serem adicionadas em separado em cada variável observada de \mathbf{x} , e nesse caso dizemos que estamos perante um modelo com ruído do sensor.

Outro tipo de ruído, relacionado com a fonte, que é adicionado às CIs pode ser incorporado no modelo que passa a ser designado por modelo com ruído da fonte e expresso por

$$\mathbf{x} = \mathbf{A}(\mathbf{s} + \mathbf{n}).$$

Nota 3.10.2 *Geralmente admite-se que o ruído é independente das CIs e que segue uma distribuição Normal.*

Se considerarmos que cada CI com ruído é dada por $\tilde{s}_i = s_i + n_i$, ($i = 1, 2, \dots, n$), temos

$$\mathbf{x} = \mathbf{A}\tilde{\mathbf{s}},$$

o que efetivamente corresponde ao modelo básico da ACI, mas para CIs modificadas. Em casos como este, no modelo ACI com ruído, a matriz das misturas pode ser estimada através dos métodos básicos de ACI (admitindo as restrições de independência e não normalidade para cada \tilde{s}_i), e nos casos em que isso não é possível devem aplicar-se outras técnicas como a remoção do enviesamento ou o uso de cumulantes de ordem superior ou ainda a estimação de máxima verosimilhança para estimar a matriz. No caso do algoritmo FastICA, em Hyvärinen (1999b) também se encontra uma versão alternativa que usa medidas não enviesadas de não normalidade, que deve ser usada em dados que apresentem ruído.

Como o modelo com ruído (3.28) não apresenta uma estrutura matricial apropriada de forma a ser invertível, então a estimação de CIs sem ruído exige a introdução de novos procedimentos

que permitam remover esse ruído como, por exemplo, um que se designa por encolhimento do código esperso (*sparse code shrinkage*) (Hyvärinen *et al.*, 2001a).

3.10.2 ACI com Número de Misturas Diferente do Número de CIs

No modelo básico de ACI (3.1) admite-se que o número n de CIs é igual ao número de variáveis observadas, ou seja, de misturas. No entanto, podemos confrontar-nos com casos em que o modelo apenas faz sentido se considerarmos um número de CIs diferente do número de misturas.

Na maioria dos trabalhos iniciais de desenvolvimento da ACI, admitia-se que o número de CIs era conhecido com antecedência, embora tal situação não aconteça frequentemente em aplicações práticas. Para confrontar esta falaciosa suposição, mencionamos o artigo de Cichocki *et al.* (1999) em que, mediante o estudo do comportamento de várias estruturas de rede neuronais, foram propostas algumas soluções para um problema de BSS, cujo número de fontes (CIs) é diferente do número de *outputs* (misturas) e ainda onde o número de fontes é geralmente desconhecido.

Nesta Secção estabelecemos que o número de misturas é dado por p e o número de CIs continua a ser indicado por n , enquanto a matriz de misturas passa a ter dimensão $p \times n$.

Número de Misturas Superior ao Número de CIs ($p > n$)

Para o caso em que o número de CIs é inferior ao número de misturas, o modelo básico de ACI deixa de se poder aplicar, uma vez que a matriz das misturas não se mantém quadrada. O uso da ACP, como pré-processamento dos dados, permite reduzir a dimensionalidade destes de forma a ficarmos com um número de misturas igual ao número de CIs, permitindo assim que o modelo básico de ACI possa ser aplicado.

Há no entanto a ressaltar que a aplicação da ACP, para reduzir o número de misturas, funciona geralmente de uma forma correta ao reter as CIs que possuem uma "forte" informação, e que implica uma redução substancial do ruído, o que traduz uma grande vantagem no tratamento dos dados. Consequentemente, se a força das CIs for muito maior que a potência do ruído, então os maiores valores próprios do sinal, da matriz de covariâncias dos dados, serão ainda claramente maiores que os valores próprios do ruído.

De entre muitos trabalhos existentes, onde se admite que o número de misturas p é superior ou igual ao número n de CIs, refira-se por exemplo Karhunen (1996) que apresenta uma revisão

de algumas abordagens neuronais.

Número de Misturas Inferior ao Número de CIs ($p < n$)

Nestes casos a matriz das misturas \mathbf{A} do modelo também não é invertível, e como tal não é possível estimar as CIs como no modelo básico. Mesmo que se conheçam os coeficientes da matriz das misturas nem sempre é possível conseguir obter exatamente os valores das CIs, uma vez que a informação é perdida durante o processo de mistura. O modelo cujo número de misturas é inferior ao número de CIs é conhecido por "ACI com bases sobrecompletas" (*ICA with Overcomplete Bases*), uma vez que o número de vetores da base da matriz \mathbf{A} excede o número de misturas p . Um contexto onde ocorrem frequentemente situações como esta é a extração de características de imagens.

À semelhança do que acontece no modelo ACI com ruído, também aqui nos deparamos com a necessidade de efetuar duas estimações em separado, uma para a matriz das misturas e outra para as CIs.

- Por vezes, a estimação das CIs pode conseguir-se através do uso da pseudo-inversa \mathbf{A}^+ da matriz das misturas \mathbf{A} . Isto leva a que no modelo $\mathbf{x} = \mathbf{A}\mathbf{s}$, desde que $\mathbf{A}\mathbf{A}^T$ seja invertível, o vetor \mathbf{s} das CIs seja estimado por

$$\mathbf{s} = \mathbf{A}^+ \mathbf{x} = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{x}.$$

Em muitas situações mais complexas, em que a pseudo-inversa não forneça uma solução satisfatória, serão necessários outros métodos mais desenvolvidos para a estimação das CIs, como por exemplo a máxima verosimilhança.

- Na estimação da matriz das misturas, podem ser utilizados métodos baseados na máxima verosimilhança, embora por vezes considerados pouco eficientes em termos computacionais. Para melhorar os desempenhos dos algoritmos a utilizar, há que considerar outras aproximações para além desta (Hyvärinen *et al.*, 2001b), como por exemplo uma modificação do algoritmo FastICA baseada em conceitos de decorrelação aproximada.

3.10.3 ACI Não Linear

Podem admitir-se situações em que as misturas não sejam lineares, originando um modelo de ACI não linear. Uma das dificuldades adicionais do modelo não linear prende-se geralmente com a inclusão de determinadas restrições.

Definição 3.10.3 *Em vez do usual modelo básico de ACI descrito por $\mathbf{x} = \mathbf{A}\mathbf{s}$, o modelo de misturas não linear é expresso por*

$$\mathbf{x} = \mathbf{f}(\mathbf{s}), \quad (3.29)$$

em que \mathbf{f} é uma transformação não linear das misturas.

Admitindo que o número de misturas é igual ao número de CIs, n , o problema do modelo de ACI não linear consiste em encontrar uma transformação $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que as componentes obtidas através de

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \quad (3.30)$$

sejam estatisticamente independentes. Uma característica essencial do problema deste modelo é que embora geralmente existam soluções, estas são não únicas (Hyvärinen *et al.*, 2001b).

Alguns dos métodos existentes para resolver o caso na ACI não linear permitem também dar resposta à BSS não linear, nomeadamente uma abordagem Bayesiana ou métodos baseados em transformações de misturas não lineares. De acordo com Lappalainen *et al.* (2000), os métodos não lineares de ACI e BSS descritos na literatura podem dividir-se em dois grupos: abordagens generativas e abordagens de transformação do sinal. Nas abordagens generativas, pretende-se encontrar um modelo específico que permita explicar como é que as observações foram geradas, o que para o modelo descrito em (3.29) corresponde a estimar quer os sinais da fonte \mathbf{s} quer a transformação desconhecida das misturas \mathbf{f} , que originaram os dados observados \mathbf{x} . Nas abordagens de transformação do sinal, as fontes são estimadas diretamente a partir da transformação inversa (3.30).

Alguns exemplos de aplicações da ACI não linear incluem canais de satélites digitais, de micro-ondas e de registos magnéticos. Para mais detalhes sobre a técnica, incluindo a separação das CIs e um caso especial do modelo geral de misturas não lineares, podem ser vistos em Jutten e Taleb (2000). Num outro trabalho de Jutten e Karhunen (2003) são abordados assuntos como a existência e unicidade de soluções para os problemas dos modelos de ACI e BSS não lineares,

e são propostas algumas técnicas de separação para misturas não lineares. Singer e Coifman (2008) apresentaram aplicações recentes da técnica da ACI não linear a mapas de difusão.

3.10.4 ACI com Restrição

Uma outra abordagem da ACI, designada por ACI com restrição ou cICA (*constrained Independent Component Analysis*), cujo desenvolvimento incide essencialmente nesta última década, abrange a implementação da técnica da ACI sujeita a restrições prévias adicionais, visando melhorar o desempenho do modelo no sentido de extrair CIs que sejam o mais próximo possível das fontes originais.

A ACI com restrição surge como uma generalização que permite incorporar certos requisitos adicionais assim como alguma informação prévia, através de restrições na função contraste da neguentropia, introduzida no modelo básico de ACI na Subsecção 3.5.2, de modo a obter-se uma função Lagrangeana aumentada.

Lu e Rajapakse (2003) apresentaram uma abordagem em que usaram a ACI com restrição para eliminar a indeterminação existente nas soluções do modelo clássico da ACI. A introdução, à priori de certa informação pretendida, na função contraste do algoritmo permitiu melhorar o desempenho do modelo na identificação e separação das CIs. Foram efetuadas experiências com sinais aleatórios simulados, com sinais de áudio reais e com imagens de ressonância magnética funcional para salientar a polivalência da técnica em eliminar a indeterminação na ACI.

Em 2005, Lu e Rajapakse voltaram a propor soluções adaptativas para resolver o problema de otimização com restrições. As aplicações a imagens faciais e também a imagens de ressonância magnética funcional ilustraram a eficácia da técnica na separação de subespaços de CIs.

James e Hesse (2005) introduziram a ACI com restrição para permitir a extração de uma das CIs entre muitas outras que estão subjacentes ao sinal de medição, através do fornecimento de um sinal de referência de base. Com tal, incorporaram restrições espectrais de registos de um único canal em dados de eletroencefalogramas associados à epilepsia.

Como resultado da investigação de restrições de desigualdade, Huang e Mi (2007) apresentaram um algoritmo melhorado para a ACI com restrição, no qual compararam diferentes medições com exatidão. Os resultados experimentais permitiram demonstrar que esta nova versão do algoritmo exibe maior precisão, comparativamente com outros existentes para o efeito, nomeadamente na remoção de artefactos em sinais de eletroencefalogramas.

Em Zhang *et al.* (2008), apresenta-se um novo modelo de ACI com restrição composto por três partes: um critério de máxima verosimilhança como uma função objetivo, medidas estatísticas como restrições de desigualdade e a normalização da matriz que permite recuperar as CIs como restrições de igualdade. É incorporado um novo algoritmo do ponto fixo, mas com restrição. Foram efetuadas simulações computacionais em sinais sintetizados e de voz, os quais demonstraram que, até certo ponto, essa combinação pode eliminar a indeterminação das CIs e fornecer melhor desempenho na recuperação dos referidos sinais. Em relação à comparação dos resultados entre o algoritmo apresentado por Lu e Rajapakse (2003) e o novo algoritmo introduzido, verificou-se a eficiência deste evidenciando que é mais simples de implementar. O novo algoritmo foi ainda aplicado a dados reais de eletrocardiogramas fetais.

Em De Vos *et al.* (2011), foi proposto um algoritmo para a ACI com restrição espacial, o qual exhibe, de uma forma geral, melhores resultados numéricos quando comparados com os de outros algoritmos. Foram também propostas novas versões semicegas dos algoritmos JADE e SOBI, os quais funcionam quer para dados reais quer para complexos assim como para um número arbitrário de restrições espaciais. Foi ainda ilustrada uma aplicação à remoção do artefacto relacionado com o piscar de olhos em registos de eletroencefalogramas.

3.10.5 ACI nos Complexos

Como regra geral, é admitido que os modelos apresentados nas Secções anteriores estão descritos para dados reais. No entanto é possível trabalhar em ACI com dados que representam números complexos. Por exemplo, é relativamente comum na área de processamento de sinal, para estudos no domínio da frequência, ser feita uma abordagem à ACI com números complexos.

Definição 3.10.4 *O modelo de ACI para números complexos pode ser descrito por (Ollila et al., 2008)*

$$\mathbf{z} = \mathbf{A}\mathbf{s},$$

onde $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ é o vetor aleatório em que os elementos complexos são as misturas, $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ é o vetor aleatório cujos elementos complexos são as CIs e \mathbf{A} representa a matriz complexa de parâmetros a_{ij} ($i, j = 1, 2, \dots, n$) desconhecidos.

Por simplicidade admite-se que o número de misturas é igual ao número de CIs e como tal admite-se que a matriz \mathbf{A} é constituída por n colunas linearmente independentes.

Para além da habitual condição de independência das CIs, admite-se ainda que cada CI é branqueada, o que corresponde a dizer que as partes real e imaginária são não correlacionadas e as suas variâncias são iguais, o que é bastante realista em problemas práticos.

O algoritmo FastICA também pode ser estendido para sinais de valores complexos, onde as CIs e as misturas observadas contemplam números complexos (Bingham e Hyvärinen, 2000). Num trabalho, na área da Neurologia, foi aplicada a ACI espacial com números complexos a dados no domínio da frequência em imagens de ressonância magnética funcional, com vista a modelar padrões dinâmicos no espaço e no tempo do fluxo de sangue oxigenado (Anemüller *et al.*, 2006).

Ultimamente têm surgido algoritmos adaptativos em relação a parâmetros, como por exemplo em situações de canais que variam ao longo do tempo (Ranganathan e Mikhael, 2007) ou outros que conjugam o sistema adaptativo com o princípio do gradiente para a maximização da curtose com vista à separação dos sinais de valores complexos (Mikhael *et al.*, 2010). Estes algoritmos mais recentes apresentam bons resultados (por simulação) em relação ao desempenho e à rapidez assim como à precisão de convergência individuais para as partes real e imaginária.

Em Ollila *et al.* (2008) é efetuada uma abordagem teórica sobre a ACI para dados com valores complexos, apresentando-se um novo e eficiente algoritmo designado por DOGMA (*Diagonalization Of Generalized covariance MATrices*), que evita a descorrelação dos dados, o qual é uma generalização do algoritmo FOBI.

3.10.6 ACI para Dados Binários

Uma outra variante do modelo de ACI é a ACI para dados binários, em que as CIs e as misturas observadas assumem apenas os valores 0 ou 1.

Em Himberg e Hyvärinen (2001) são apresentados resultados experimentais em que as CIs, as misturas, as observações e o ruído são constituídos por valores binários. O objetivo principal consistiu em encontrar a matriz das misturas em vez das fontes binárias, tendo sido utilizado o algoritmo FastICA recorrendo ou à assimetria ou à curtose das CIs estimadas. Os autores salientaram que na sua utilização surgiram melhores resultados no caso de dados muito esparsos e ruidosos.

Também Nguyen e Zheng (2011) utilizaram a ACI para dados binários, considerando as observações como misturas disjuntivas de fontes binárias independentes, e propuseram um al-

goritmo iterativo determinístico com vista a determinar a distribuição das CIs e da matriz das misturas. Levaram a cabo um estudo de simulação para verificar a eficácia do algoritmo e apresentaram exemplos de possíveis aplicações a casos reais, como a otimização da monitorização em redes sem fio multicanal. Para além desta, outras aplicações podem ser consideradas, como por exemplo, diagnósticos médicos, tomografias e gestão de recursos da internet.

3.11 Áreas de Aplicações da ACI

De entre várias aplicações da ACI nas diferentes áreas destacam-se as seguintes que consideramos mais importantes.

- *Processamento de sinais áudio*: Para além do clássico problema *cocktail party* descrito na Secção 3.3, outras aplicações podem ser enumeradas como situações em que se pretende proceder à separação de sinais como por exemplo, quando se deseja separar o ruído ou a interferência de um sinal de voz de uma pessoa que esteja a falar ao telemóvel num carro com muito barulho ou separar os sinais de interferência de rádio num telemóvel.

A separação de sinais de áudio não é simples, uma vez que por aplicação do modelo básico da ACI nem sempre se obtém uma aproximação perfeita em relação ao original processo de misturas de sinais. Uma das questões prende-se com o facto da mistura de sinais áudio não ser instantânea pois estes propagam-se muito devagar e como tal são captados pelos recetores em diferentes momentos. Para além disso ainda podem ocorrer ecos, nomeadamente em recintos fechados. Uma melhor abordagem do problema em que as misturas são convolutivas, isto é são misturadas por meio de atrasos de propagação no tempo, pode ser implementada por uma versão convolutiva do modelo ACI (Hyvärinen *et al.*, 2001b). Outra questão importante a ter em conta é a não estacionariedade da mistura devido às possíveis mudanças rápidas de posições de emissores e/ou recetores.

Uma outra situação prende-se com o problema da ACI subdeterminada ou sobrecompleta (*underdetermined* ou *overcomplete ICA*), em que se admite que o número de componentes independentes a estimar é superior ao número de misturas.

E ainda pode ocorrer o caso em que o ruído seja considerável de tal forma que dificulte a estimação do modelo básico da ACI, mesmo com ruído normal. Nesses casos, aconselha-se a utilizar antes o modelo da ACI com ruído (Hyvärinen *et al.*, 2001b).

Geralmente os sinais acústicos registados num ambiente real não são instantâneos mas sim misturas convolutivas, quer seja devido a atrasos no tempo quer seja devido a reflexões, e por isso em Ikeda e Murata (1999) é proposto um método da ACI para separar misturas convolutivas de sinais que contêm informação de autocorrelações e de não estacionariedade para intervalos temporais longos. Tendo ainda em conta as estruturas temporais dos sinais, Brandstein (1998) defende o desenvolvimento de técnicas que incorporam de uma forma explícita a natureza do sinal da fala dentro de um contexto multicanal. Por outro lado, Torkkola (1999) apresentou uma revisão acerca das complicações dos sinais de áudio. Makino *et al.* (2004) apresentaram um outro estudo acerca destes conteúdos da separação de sinais acústicos de misturas convolutivas através da ACI.

Mencionamos ainda Lombard *et al.* (2011), onde se mostra que a minimização da dependência estatística utilizando a ACI para sinais da banda larga pode ser explorada com sucesso por localização da fonte acústica. Três métodos foram investigados para extrair a diferença de tempo de chegada de informação contida em filtros de banda larga de um esquema de ACI de dois canais.

- *Processamento de imagens:* Uma das aplicações mais importantes da ACI relaciona-se com o uso de sinais para extrair características de imagens naturais (paisagens, retratos humanos, animais, habitações, materiais, ...). Neste contexto, a ACI pode ser utilizada para comprimir, reduzir o ruído ou efetuar reconhecimento de padrões.

A ACI está relativamente próxima de um método que permite estimar características subjacentes a imagens (veja-se por exemplo em Hurri *et al.*, 1997), designado por método do código esparsa, uma vez que este fornece procedimentos similares à ACI para a estimação. O método do código esparsa permite extrair características cuja propriedade seja estarem ativas em simultâneo, mas apenas num pequeno número. Esta extração de características de imagens em que ocorre uma decomposição está relacionada também com a que é efetuada através de onduletas (Mallat, 1989) ou Análise de Gabor (Feichtinger e Strohmer, 1997).

Em Manduchi e Portilla (1999) é efetuada a implementação de uma técnica baseada na ACI para a escolha de um conjunto de filtros a aplicar em imagens com texturas. Em experiências que envolvem tanto a classificação como a síntese da textura, a ACI evidencia

um desempenho superior comparativamente à ACP.

Outros resultados como por exemplo a aplicação da contração do código esparsa para redução do ruído encontra-se em Hyvärinen *et al.* (2001a) ou a aplicação da ACI para extrair características de imagens ao nível das informações cromáticas e de estéreo (Hoyer e Hyvärinen, 2000). Um outro campo de aplicação importante em imagens está relacionado com o reconhecimento facial, no qual a ACI desempenha um papel crucial, por vezes com maior desempenho que a ACP usual, como ocorre em Bartlett *et al.* (2002).

Na tese de Doutoramento de Puga (2000), sobre codificação e compressão de imagem, foram desenvolvidas técnicas mais genéricas, com base na estacionariedade e na não normalidade, para maximizar a qualidade subjetiva das imagens tratadas através de processos como a ACI, a ACP e a Análise em Componentes Simétricas.

Mais recentemente, no campo da deteção remota de imagens podemos mencionar o trabalho de Du *et al.* (2006) em que a ACI foi aplicada na classificação de imagens com o auxílio da ACP. Também Benlin *et al.* (2008) tiraram partido da ACI ao usarem a técnica na classificação e deteção de alterações de imagens remotas multiespectrais.

Um dos campos de pesquisa recente incide na deteção de falsificações de imagens digitais, onde destacamos Kumar *et al.* (2013) que explora a ACI como uma ferramenta para obter pistas sobre a adulteração da imagem original.

- *Telecomunicações*: Um ramo específico das telecomunicações que possibilita várias aplicações da ACI enquadra-se nas denominadas técnicas de Acesso Múltiplo por Divisão em Código (em inglês: *Code Division Multiple Access* - CDMA), as quais permitem que todos os usuários possam transmitir simultaneamente nas mesmas frequências e utilizar toda a banda disponível. Uma característica específica de todas as aplicações da ACI em telecomunicações é que elas constituem quase sempre problemas não totalmente cegos. Com efeito, à priori, o recetor possui alguma informação no sistema de comunicação, mais que não seja pelo menos o código de difusão do utilizador pretendido é conhecido.

Alguns dos exemplos em que se aplica a ACI a dados CDMA podem ser consultados em Joutsensalo e Ristaniemi (1998), Ristaniemi e Joutsensalo (1999) e Cristescu *et al.* (2000). Numa outra referência mais recente (Parmar e Unhelkar, 2010) comparam-se os desempenhos de alguns dos algoritmos da ACI utilizados para resolver um problema de estimação

num sistema de comunicação de sequência direta em CDMA.

- *Indicadores biomédicos*: Um dos propósitos da Engenharia Biomédica é investigar as possíveis alterações ao nível fisiológico que se manifestam nos mais variados órgãos internos do corpo humano. Tais ocorrências podem ser quantificadas e modeladas através de equipamentos de saúde que fornecem indicadores biomédicos do funcionamento de certos sistemas fisiológicos tais como o eletroencefalograma, o eletrocardiograma, o magnetoencefalograma, a ressonância magnética funcional, a tomografia por emissão de positrões, entre outros. Geralmente estes indicadores são fenómenos que apresentam uma certa fragilidade, são não estacionários e distorcidos devido a ruídos e interferências. Os equipamentos permitem captar um certo número de observações sobrepostas no espectro que correspondem à mistura de múltiplas fontes independentes. Portanto, o objetivo consiste no tratamento dessas observações de forma a que os resultados pretendidos correspondam às fontes originais separadas. Existem várias técnicas de processamento de sinal (incluindo a ACI) que podem ser usadas na redução do ruído e na análise de tais indicadores.

São exemplos neste campo a análise dos sinais captados pela atividade bioelétrica cerebral através de elétrodos colocados no couro cabeludo (escalpe), em casos de pacientes com determinado tipo de epilepsia ou outras doenças relacionadas, bem como a deteção da apneia do sono, que é um outro fenómeno caracterizado por oscilações específicas, possíveis de serem medidas através de um eletrocardiograma.

De entre muitas das aplicações neste domínio em que a ACI se estabelece como um bom método, realçamos a identificação e remoção de artefactos (sinais que não são gerados pela atividade cerebral, mas por algum distúrbio externo como a atividade muscular do olho, a respiração ou o batimento cardíaco) que se podem encontrar, por exemplo, em Makeig *et al.* (1996), Vigário (1997), Vigário *et al.* (1998), Jung *et al.* (1998) e Delorme *et al.* (2007). Uma outra importante aplicação da ACI a sinais de eletroencefalogramas e magnetoencefalogramas, que também consiste na separação de artefactos, pode ser consultada em Vigário *et al.* (2000). Também em Faria *et al.* (2012), procedeu-se à investigação da viabilidade de utilização de um sistema experimental de estimulação elétrica transcraniana, através de correntes contínuas de baixa intensidade no tratamento de epilepsias refratárias, em crianças durante o sono. Para garantir uma maior segurança na aplicação

das correntes contínuas, recorreu-se, em simultâneo, ao registo do eletroencefalograma, no qual apareceram artefactos que inviabilizavam a sua leitura. A ACI, com base no algoritmo Infomax, revelou ser o método mais eficaz na eliminação desses artefactos e permitiu aferir que esta técnica de estimulação elétrica transcraniana pode ser considerada como ferramenta útil no tratamento deste tipo de epilepsias.

Num outro exemplo, Anemüller *et al.* (2006) utilizam a ACI espacial com números complexos aplicada ao domínio da frequência, para que seja possível modelar padrões dinâmicos no espaço e no tempo em dados de imagens de ressonância magnética funcional. Mais recentemente, Delorme *et al.* (2012) compararam os resultados da decomposição de treze conjuntos de dados de eletroencefalogramas através de 22 algoritmos de ACI e BSS.

- *Aplicações financeiras:* A ACI pode ajudar a encontrar fatores que se julgam desconhecidos em determinadas situações financeiras como por exemplo nas taxas de câmbio ou nos retornos das ações da bolsa.

Um dos estudos financeiros pioneiros na aplicação da ACI como complemento à ACP clássica foi levada a cabo por Back e Weigend (1997), no qual foi analisada uma carteira de ações. O estudo permitiu observar de uma forma mais rápida algumas estruturas que estavam subjacentes aos dados, ajudando assim a minimizar o risco a que se está sujeito quando se efetuam certos investimentos. Posteriormente, Kiviluoto e Oja (1998) aplicaram a ACI a séries temporais financeiras, cujos dados representavam o fluxo de caixa simultâneo em várias lojas pertencentes à mesma cadeia de retalho.

De entre outros trabalhos desenvolvidos nesta área, salientamos um artigo de Oja *et al.* (2000) em foi utilizada a ACI para análise e previsão em séries temporais financeiras e, um outro em que Lu *et al.* (2009), com vista a reduzir a influência de ruído, propuseram uma abordagem de um modelo bietápico usando para tal a ACI e um modelo de regressão baseado em algoritmos de redes neuronais, designado de regressão vetorial de suporte, na previsão de séries temporais financeiras.

Outras extensões podem-se encontrar para tratar dados de séries temporais distorcidos da realidade por desfasamentos, por ruído e por falta de linearidades.

- *Climatologia:* A noção de que o sistema climático é bastante complexo sugere que, para além das variáveis meteorológicas usualmente estudadas, possam existir outras compo-

nentes físicas que sejam independentes e não necessariamente observáveis. Como exemplo em Hannachi *et al.* (2009), a ACI é encarada como um método de rotação das CPs, admitindo-se que os sinais climáticos subjacentes possuem uma ação de força independente, e como tal é expetável encontrar *loadings* com padrões interpretáveis cujos coeficientes ao longo do tempo tenham propriedades que vão para além da não correlação observada nas CPs. A ACI têm vindo a ser utilizada com mais frequência na área do clima e alguns autores tais como Aires *et al.* (2000) ou Basak *et al.* (2004) afirmam que em determinados casos a ACI evidencia melhores resultados do que a ACP tradicional. Em Sebastião e Oliveira (2009) é abordado um conjunto de dados mensais médios de pressão ao nível do mar, em oito estações meteorológicas no Norte do Oceano Pacífico, com o intuito de obter resultados comparativos entre a ACP e a ACI, os quais irão ser analisados na Secção 3.12. O mesmo conjunto de dados foi utilizado em Sebastião e Oliveira (2013), onde se aplicam a ACP e a ACI e ainda uma nova abordagem da ACI como alternativa à Análise Espectral Singular Multicanal, a partir da matriz dos desfasamentos, e que serão discutidas na Secção 6.3.

Para além das aplicações descritas anteriormente outras aplicações e referências podem ser consideradas, tais como:

- Análise de dados das expressões de genes (Engreitz *et al.*, 2010);
- Análise de documentos de texto (Kolenda e Hansen, 1999);
- Análise de perfil proteómico e biomarcadores (Zhang *et al.*, 2011);
- Astrofísica (Allen *et al.*, 2013);
- Extração de informação de bases de dados - *data mining* (Draghici *et al.*, 2003; Wu e Law, 2010);
- Monitorização sísmica (Ham e Faour, 1999; Orihara *et al.*, 2009);
- Processos industriais (Tseng e Tsai, 2006);
- Sistemas mecânicos (Ypma e Pajunen, 1999).

3.12 Exemplo de Aplicação: Pressão ao Nível do Mar

A aplicação de conceitos teóricos abrangidos pelas técnicas da ACP e da ACI serão analisados nesta Secção através de um exemplo de dados climáticos (de pressão ao nível do mar), os quais já tinham sido usados na obtenção de resultados comparativos entre a ACP e outras técnicas, nomeadamente em Oliveira (2003). Para além da ACP usual, iremos implementar a ACI através dos algoritmos FastICA e AMUSE, e efetuar comparações dos dados em estudo com duas séries específicas de dados de temperatura e de pressão atmosférica ao nível do mar a oeste do Pacífico, devido à sua importância como fonte de variabilidade interanual no sistema climático global.

Consideremos um conjunto de 216 valores mensais médios de pressão ao nível do mar observados em cada uma das 8 estações meteorológicas (desde janeiro de 1979 a dezembro de 1996) no norte do Oceano Pacífico, nos estados do Alasca, Califórnia, Havai e Washington (Figura 3.9). As estações meteorológicas analisadas foram: 1 - Crescent City; 2 - San Diego; 3 - San Francisco; 4 - Hilo; 5 - Honolulu; 6 - Neah Bay; 7 - Seldovia e 8 - Sitka (Apêndice D). A amplitude temporal (num total de 216 meses) é suficiente para examinar as relações dos padrões obtidos pelas técnicas e pela série temporal do Índice de *El Niño*¹¹ (Figura 3.10).

A este conjunto de dados aplicaram-se as técnicas da ACP (sobre a matriz de correlações) e da ACI. Selecionaram-se as 3 primeiras CPs (Figura 3.11) e as 2 primeiras CIs (Figuras 3.12 e 3.13) por se verificar serem suficientes na obtenção de resultados satisfatórios para uma análise elementar com vista a realçar as diferenças e semelhanças entre as técnicas em análise quer ao nível do domínio do tempo quer ao nível da frequência. A quantidade de variância explicada nas 2 e nas 3 primeiras CPs é, respetivamente, de 72.5% e 87%. Foi utilizado o algoritmo FastICA para obtenção das 2 CIs cuja extração ocorreu em paralelo. A escolha recaiu sobre este algoritmo por ser um dos mais usados na estimação do modelo ACI. No entanto muitos dos resultados foram também comparados com os resultados obtidos segundo o algoritmo AMUSE, uma vez que este é uma boa alternativa (ver a Secção 3.6) para analisar os dados originais (de pressão ao nível do mar) ao admitirmos que as CIs possuem uma dependência temporal.

¹¹ A série do Índice de *El Niño* é constituída por 216 valores mensais (de janeiro de 1979 a dezembro de 1996) de temperatura (em °C) ao nível do mar (*Sea Surface Temperature* - SST). Os dados foram retirados de www.cpc.ncep.noaa.gov/data/indices/ e são provenientes da região designada por Niño 3 (150°W-90°W e 5°N-5°S).

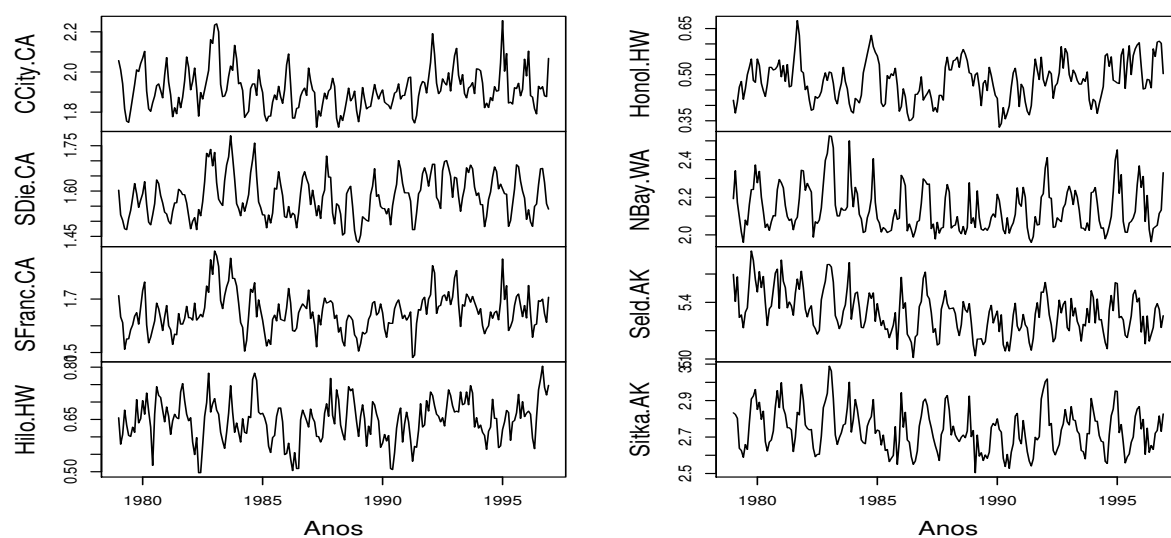


Figura 3.9: Séries temporais dos valores mensais médios de pressão ao nível do mar em 8 estações meteorológicas no Norte do Oceano Pacífico entre janeiro de 1979 e dezembro de 1996

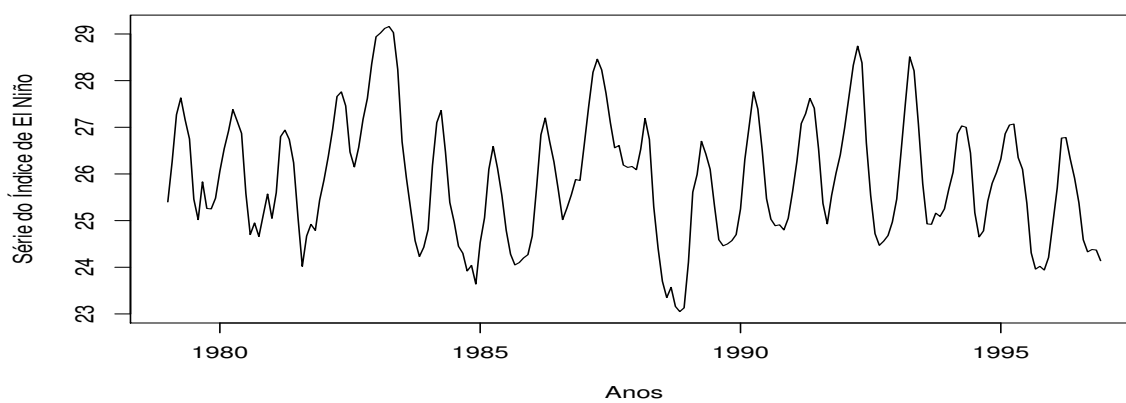


Figura 3.10: Série temporal dos valores mensais de temperatura ao nível do mar (em $^{\circ}\text{C}$) do Índice de *El Niño* entre janeiro de 1979 e dezembro de 1996

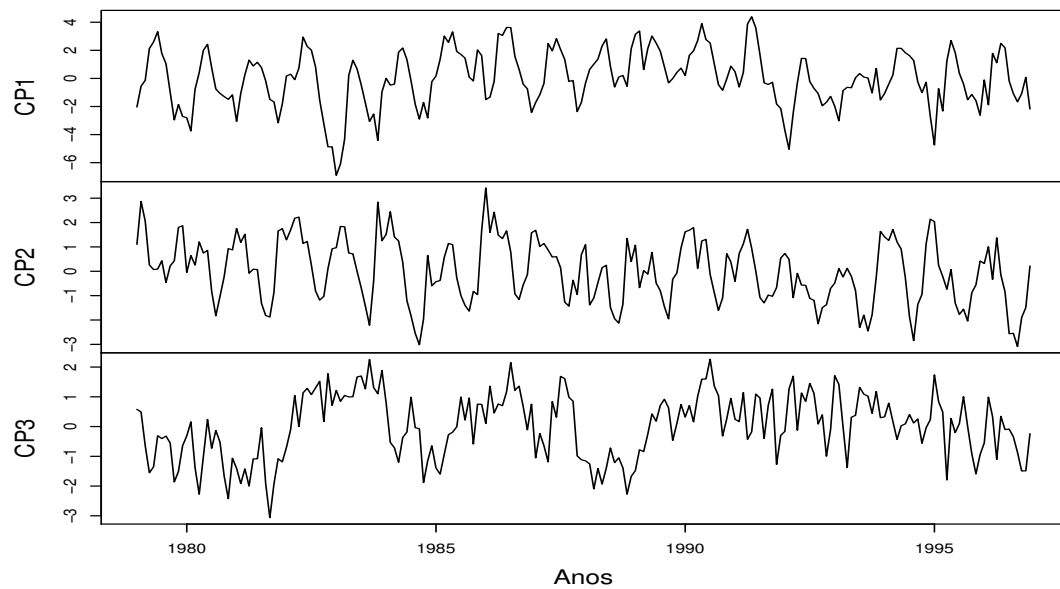


Figura 3.11: Coordenadas das três primeiras componentes principais da ACP sobre a matriz de correlações

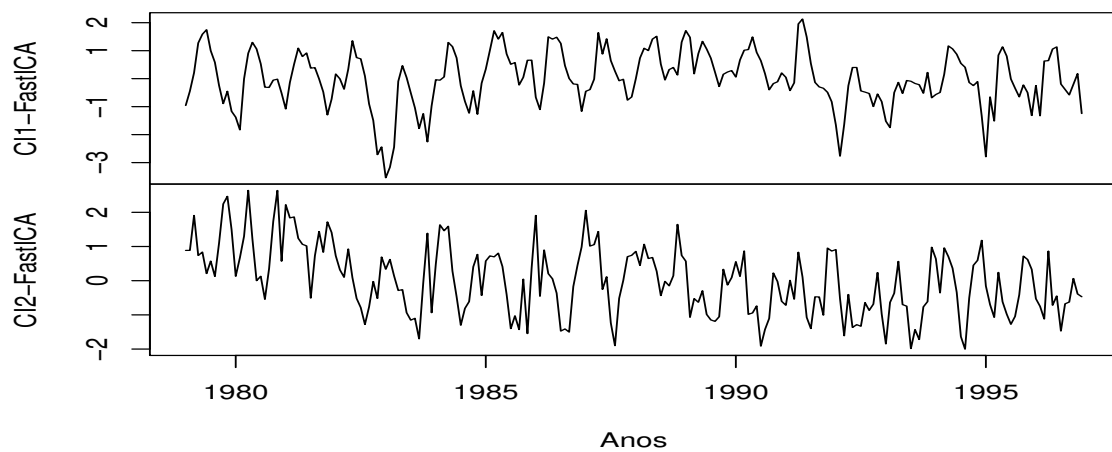


Figura 3.12: Coordenadas das duas primeiras componentes independentes obtidas pelo algoritmo FastICA

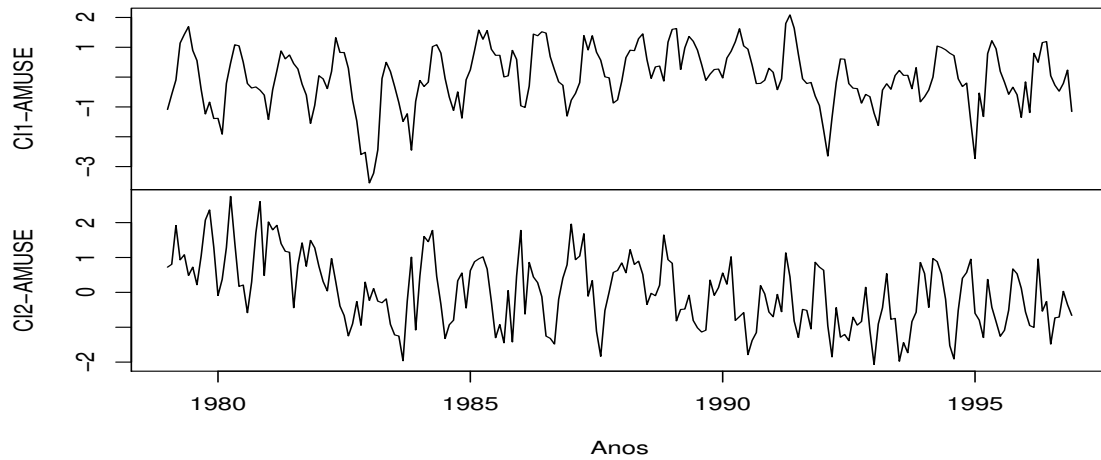


Figura 3.13: Coordenadas das duas primeiras componentes independentes obtidas pelo algoritmo AMUSE

A Figura 3.14 apresenta os gráficos que permitem comparar a representação das observações nas duas primeiras CPs (CP1+CP2) e nas duas únicas CIs extraídas (CI1+CI2), e como tal obtidas sem qualquer ordem específica, através dos algoritmos FastICA e AMUSE. Para quantificar as semelhanças existentes na representação gráfica, aplicamos um critério de comparação dos subespaços gerados pelas CPs e CIs mencionadas.

Coeficiente de Determinação Generalizado

O Coeficiente de Determinação Generalizado (*Generalized Coefficient of Determination* (GCD)) de Yanai (Ramsay *et al.*, 1984) é um indicador que mede o grau de semelhança entre dois subespaços, o qual é definido como o cosseno do ângulo entre as matrizes de projeção ortogonal (Apêndice C.1) nesses subespaços. Quanto mais próximos da unidade forem os valores do índice, maior será a semelhança entre dois subespaços. Como referido em Cadima *et al.* (2004), este coeficiente possui uma relação estreita com o conceito de distância entre subespaços abordada em Golub e Van Loan (1996).

No contexto deste exemplo de aplicação, consideramos que as matrizes de projeção ortogonal

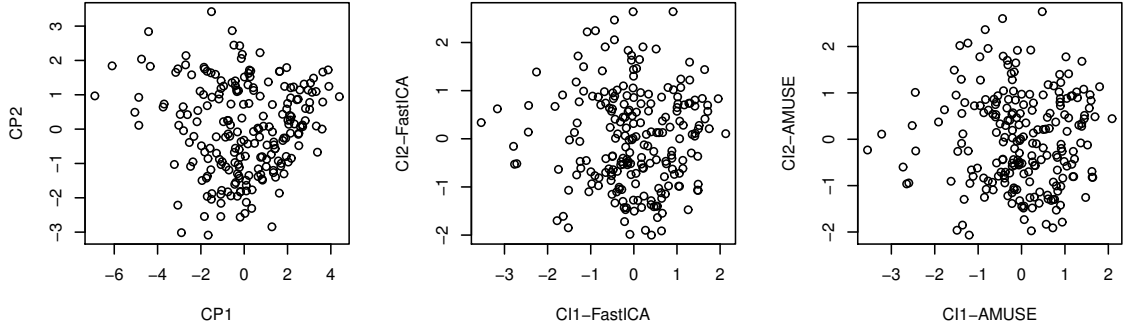


Figura 3.14: Representação das coordenadas das 216 observações nas duas primeiras componentes principais (CP1+CP2) e nas duas componentes independentes (CI1+CI2) obtidas através dos algoritmos FastICA e AMUSE

são obtidas a partir das matrizes cujas colunas são formadas pelas CPs e pelas CIs, respetivamente.

Definição 3.12.1 *Seja \mathbb{G} o subespaço gerado pelas q CPs retidas e \mathbb{H} o subespaço gerado pelas m CIs extraídas. Sejam $\mathbf{P}_{\mathbb{G}}$ e $\mathbf{P}_{\mathbb{H}}$ as matrizes de projeção ortogonal nos subespaços \mathbb{G} e \mathbb{H} , respetivamente.*

De acordo com Minhoto (2009), o indicador GCD é dado por

$$GCD(\mathbb{G}, \mathbb{H}) = \cos(\mathbf{P}_{\mathbb{G}}, \mathbf{P}_{\mathbb{H}}) = \frac{\text{tr}(\mathbf{P}_{\mathbb{G}}\mathbf{P}_{\mathbb{H}})}{\sqrt{q * m}}.$$

Os valores do coeficiente GCD para os subespaços gerados pelas duas primeiras CPs e pelas duas CIs obtidas através do algoritmo FastICA, assim como pelas duas CIs obtidas através do algoritmo AMUSE são ambos iguais a 0.663. Uma vez que o coeficiente GCD varia entre 0 e 1, podemos admitir que este valor representa uma semelhança razoável entre os subespaços gerados em ambos os casos (ou seja, entre o gráfico da esquerda e o gráfico central e entre o gráfico da esquerda e o gráfico da direita da Figura 3.14). Ao seleccionar as três primeiras CPs em vez das duas CPs e mantendo apenas duas CIs, a semelhança entre os subespaços gerados aumenta uma vez que o valor do coeficiente GCD é igual a 0.759.

Como não é de estranhar, o coeficiente GCD para os subespaços gerados pelas duas CIs obtidas através do algoritmo FastICA e pelas duas CIs obtidas através do algoritmo AMUSE

Tabela 3.1: Correlações entre as coordenadas das componentes principais em ACP e das componentes independentes em ACI através do FastICA e do AMUSE

	CP1	CP2	CP3
CI1-FastICA	0.965	0.001	-0.235
CI2-FastICA	-0.209	0.592	-0.692
CI1-AMUSE	0.986	-0.093	-0.122
CI2-AMUSE	-0.053	0.584	-0.721

é unitário, o que traduz uma semelhança "perfeita" entre os respectivos subespaços gerados, uma vez que ambos servem o mesmo propósito de representar as duas componentes que sejam estatisticamente independentes. Note-se que o padrão associado à mancha de pontos é similar no gráfico central e no gráfico da direita da Figura 3.14, ambos envolvendo CIs, os quais diferem ligeiramente em termos de localização a menos de uma rotação e translação.

3.12.1 Estudo no Domínio do Tempo

A análise das coordenadas das componentes sob a forma de séries temporais deverá consistir nas comparações das correlações (Tabela 3.1) entre pares de componentes das duas técnicas. A CP1 está altamente correlacionada com a CI1, enquanto que a CP2 apresenta um grau de correlação razoável com a CI2 embora seja um pouco mais relevante a correlação da CP3 com a CI2. De entre estas correlações, destacam-se valores (em módulo) ligeiramente mais altos para a utilização do AMUSE em relação ao FastICA, em particular nos pares (CP1;CI1-AMUSE) e (CP3;CI2-AMUSE). Como tal, para efetuarmos uma interpretação gráfica do comportamento comparativo entre as componentes mais correlacionadas, apresentamos as Figuras 3.15 e 3.16 cujas componentes foram previamente normalizadas. Estas permitem estabelecer comparações visuais conjuntas entre as CPs e as CIs e constatar que a primeira componente apresenta praticamente o mesmo comportamento ao longo do tempo independentemente da técnica utilizada, enquanto que a segunda componente revela um comportamento visual relativamente distinto para ambas as técnicas.

Uma vez que a terceira CP (CP3) apresenta uma correlação com a segunda CI relativamente significativa, decidimos efetuar ainda o gráfico da Figura 3.17 para comparar estas componentes,

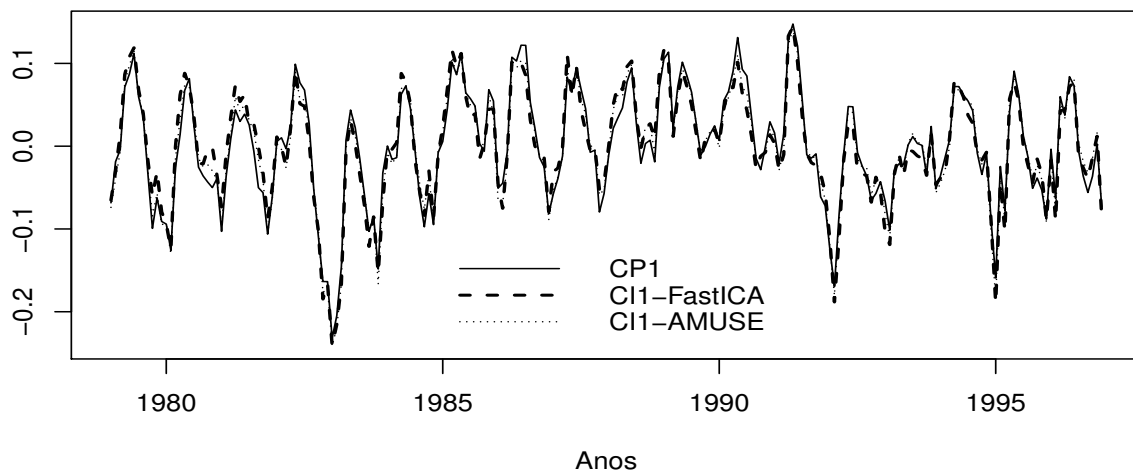


Figura 3.15: Coordenadas da CP1, da CI1-FastICA e da CI1-AMUSE normalizadas

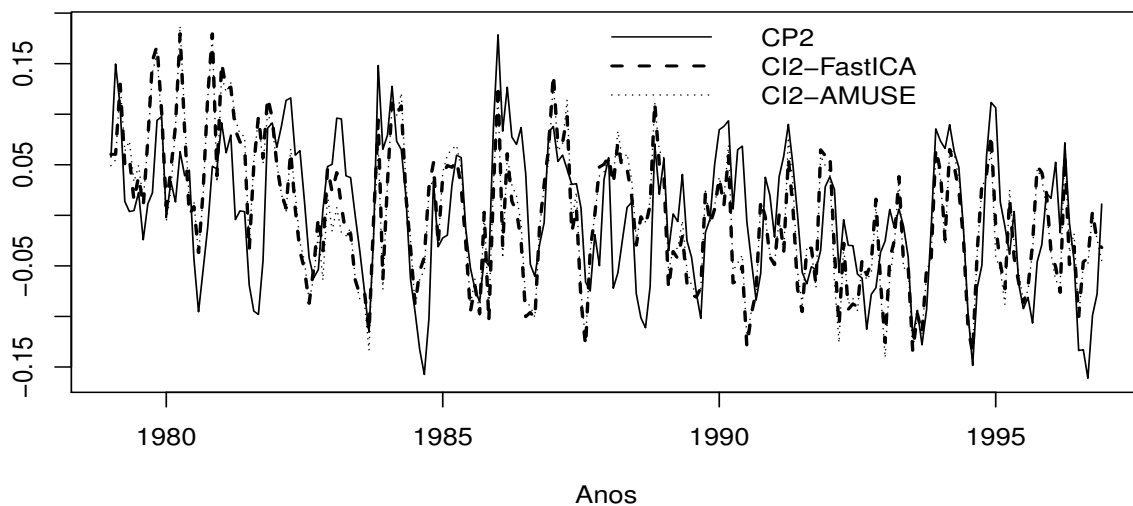


Figura 3.16: Coordenadas da CP2, da CI2-FastICA e da CI2-AMUSE normalizadas

as quais foram previamente normalizadas. As componentes apresentam um padrão geral de comportamento que é por vezes distinto entre técnicas tal como na Figura 3.16, mas ainda assim parece existir uma concordância razoável.

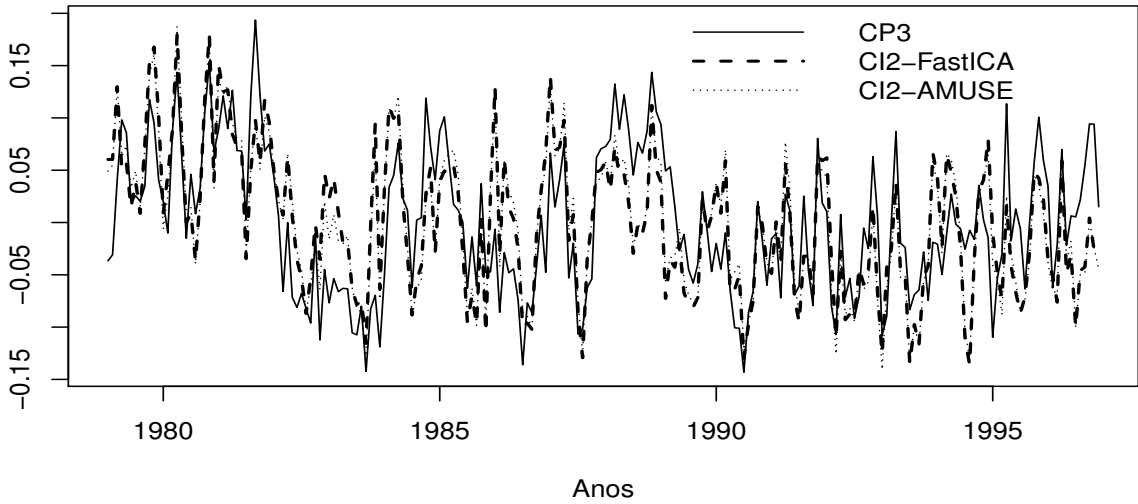


Figura 3.17: Coordenadas da CP3, da CI2-FastICA e da CI2-AMUSE normalizadas

Também é possível comparar os coeficientes dos vetores (associados às componentes) através dos ângulos entre cada par de vetores de componentes. Tendo em conta os passos da implementação dos dois algoritmos envolvidos na técnica da ACI, os vetores associados às CIs são exatamente iguais em termos de coeficientes, e como tal destacam-se os seguintes cossenos mais elevados (em módulo) válidos para ambos os algoritmos: $\cos(\text{CP1}, \text{CI1}) = 0.906$, $\cos(\text{CP2}, \text{CI2}) = 0.483$ e $\cos(\text{CP3}, \text{CI2}) = 0.687$. Estes permitem identificar relações existentes entre os coeficientes dos vetores de ambas as técnicas, e estão em concordância com os resultados das correlações entre as referidas componentes.

Por outro lado, os vetores próprios associados às CPs da ACP e os vetores da matriz do branqueamento da ACI também podem ser usados como meio de comparação entre técnicas, para avaliar até que ponto é que os seus coeficientes são ou não similares. Uma inspeção visual da Figura 3.18 permite constatar um elevado grau de semelhança entre os coeficientes dos vetores associados a CP1 e CI1-FastICA e algumas semelhanças entre o CP3 e CI2-FastICA, resultados

esses que são consistentes com os valores de cossenos existentes entre os vetores (associados às componentes) descritos anteriormente.

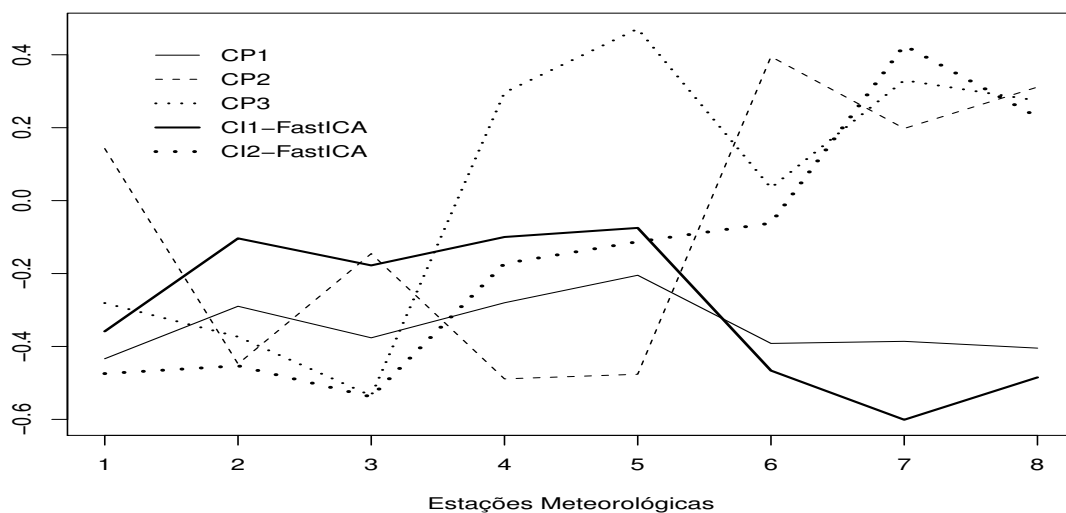


Figura 3.18: Vetores próprios da ACP e vetores da matriz do branqueamento em ACI

No que se refere às autocorrelações das componentes sob o ponto de vista de séries temporais, apresentamos na Figura 3.19 os valores da função de autocorrelação obtidos a partir das coordenadas de CP1, CP2 e CP3.

Os dois primeiros correlogramas evidenciam uma sazonalidade anual para as CP1 e CP2. A CP2 possui valores mais elevados (em módulo) de autocorrelação nomeadamente para um desfasamento até cinco anos, ou seja, a correlação entre pares de valores da série temporal separados por um intervalo até cinco anos é geralmente mais elevada quando comparada com a CP1. Por outro lado, o terceiro correlograma sugere a existência de um ciclo de periodicidade de cerca de 3 a 4 anos (o que será evidenciado no espectro) para a CP3.

Consideremos também as componentes independentes sob o ponto de vista de séries temporais, cujos valores da função de autocorrelação obtidos a partir das coordenadas de CI1-FastICA, CI2-FastICA, CI1-AMUSE e CI2-AMUSE se apresentam na Figura 3.20.

Comparativamente em termos gráficos, estes são muito semelhantes entre algoritmos FastICA e AMUSE para as mesmas CIs. Os correlogramas referentes à CI1 apresentam uma sazonalidade

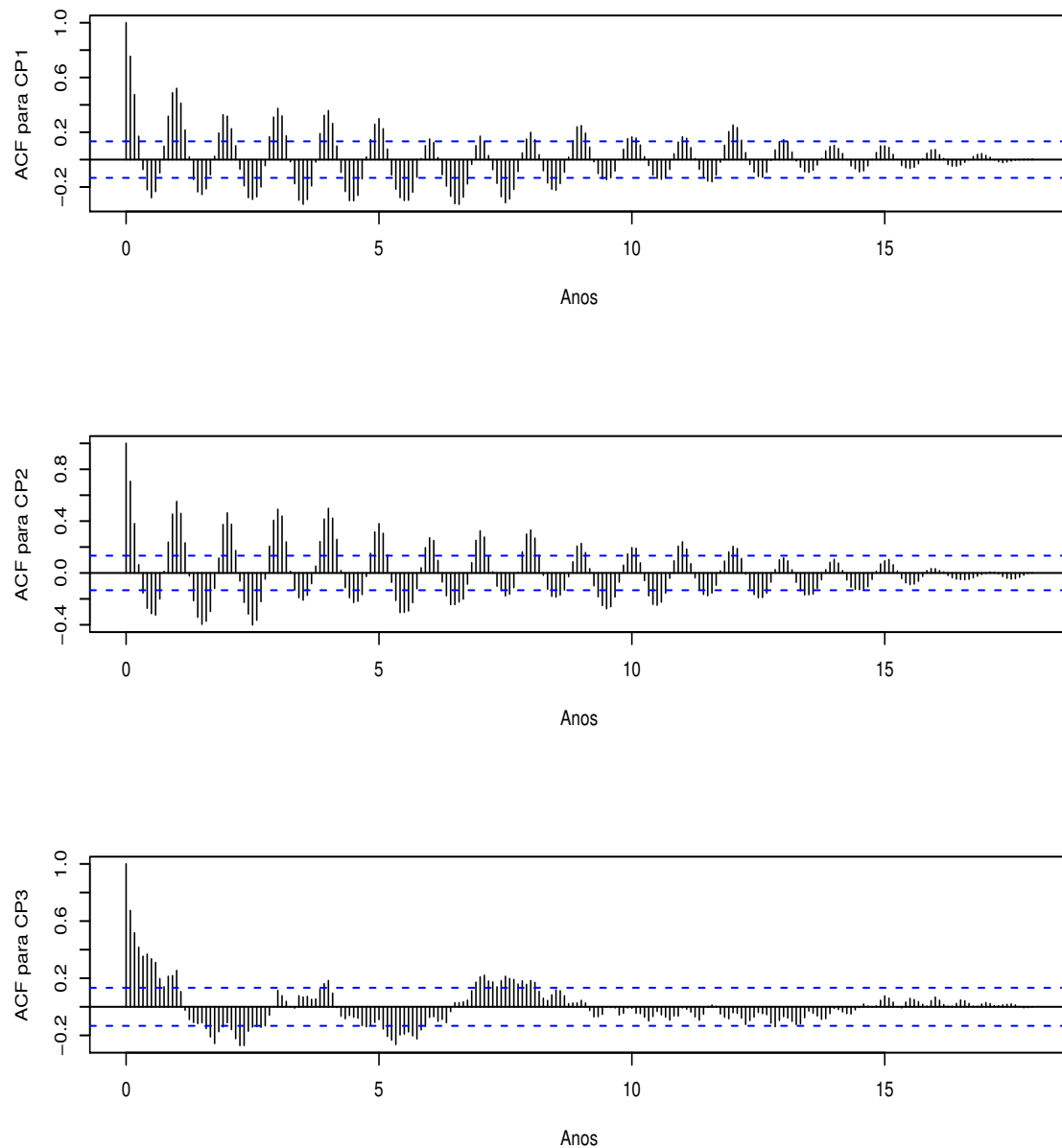


Figura 3.19: Valores da função de autocorrelação para as CP1, CP2 e CP3. As linhas horizontais a tracejado representam os limites do intervalo a 95% de confiança para a autocorrelação

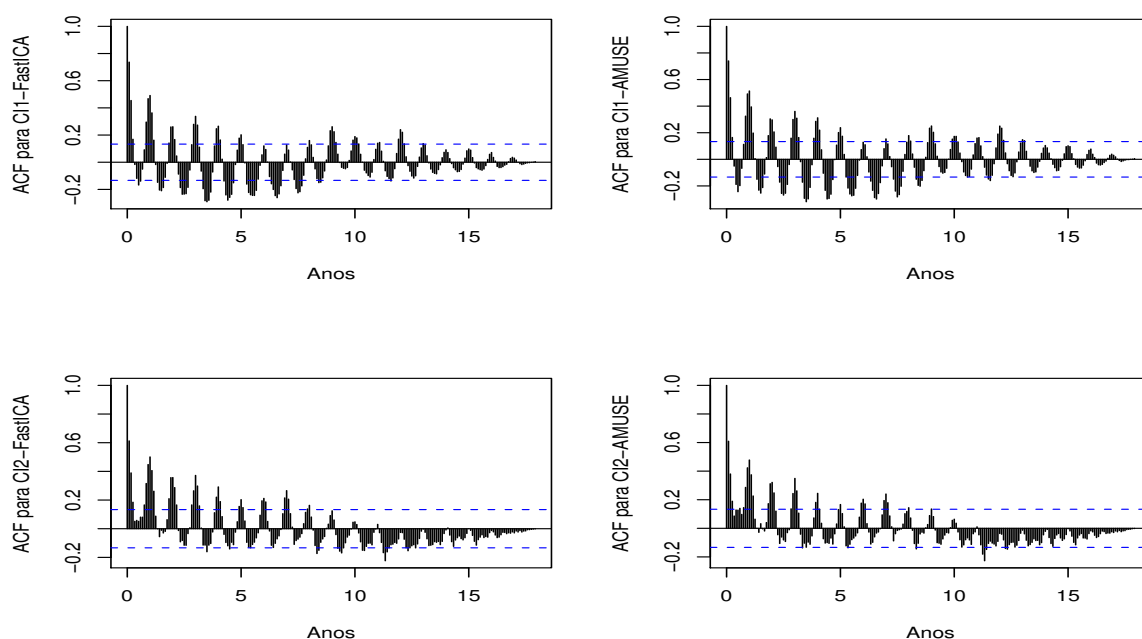


Figura 3.20: Valores da função de autocorrelação para as CI1 e CI2 obtidos através dos algoritmos FastICA e AMUSE. As linhas horizontais a tracejado representam os limites do intervalo a 95% de confiança para a autocorrelação

anual, enquanto que nos correlogramas da CI2 a sazonalidade não é evidente. Para além disso, constatamos que as autocorrelações tornam-se praticamente todas negativas em CI2 a partir dos 10 anos de desfasamento.

3.12.2 Estudo no Domínio da Frequência

Após o estudo detalhado ao nível do domínio do tempo, podemos efetuar o estudo de uma série estacionária no domínio da frequência. Para tal devemos analisar a estrutura resultante de diferentes oscilações em diferentes frequências. Alguns dos indicadores estatísticos apropriados são o espectro, o espectro cruzado, o espectro de fase e o espectro de coerência (Apêndice C.3).

O espectro das coordenadas de cada componente é uma função que permite geralmente identificar de uma forma simplista os principais picos de frequência. Os gráficos dos espectros podem ser utilizados entre diferentes coordenadas para averiguar se os padrões fornecem

resultados satisfatórios para diferentes valores de desfasamentos no tempo.

Na Figura 3.21 podemos averiguar que a CI1 apresenta espectros semelhantes para ambos os algoritmos utilizados, assim como a CI2. Em todos os espectros se destacam dois principais picos de frequência em 0.083 e 0.167 que correspondem aos períodos anual ($1/0.083 \simeq 12$) e semestral ($1/0.167 \simeq 6$) respetivamente.

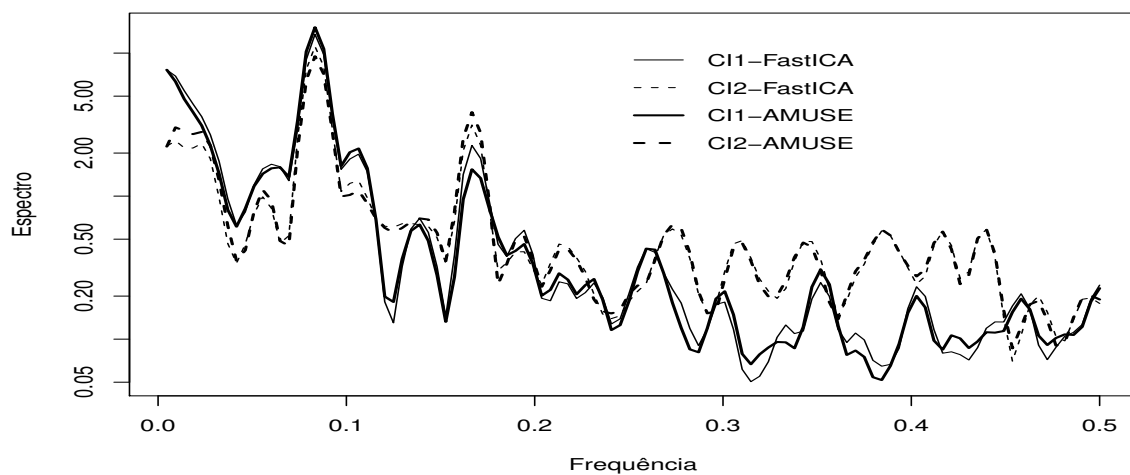


Figura 3.21: Espectros das coordenadas de CI1-FastICA, CI2-FastICA, CI1-AMUSE e CI2-AMUSE

Dada a semelhança existente entre os espectros das séries na Figura 3.21, seleccionámos apenas os espectros obtidos através do FastICA para representar conjuntamente com os espectros das CPs na Figura 3.22. Ao compararmos estes espectros, observa-se que as diferentes componentes assumem comportamentos idênticos e em particular os espectros CP1 e CI1-FastICA são muito semelhantes assim como o CP3 e CI2-FastICA. Para além dos picos referidos anteriormente, aquando do uso da ACI e também identificados pela ACP, ainda é notório no espectro da CP3 e da CP2 um pico inicial para uma frequência em 0.023 o que corresponde a um período de cerca de três anos e meio.

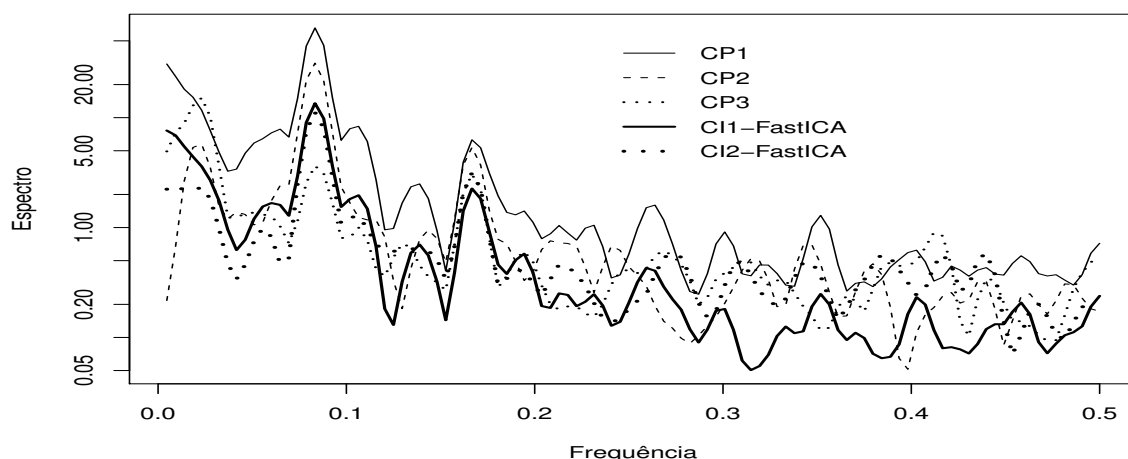


Figura 3.22: Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA e CI2-FastICA

Comparações com a Série Temporal do Índice de *El Niño*

Para além das comparações entre as análises dos espectros do conjunto de dados mensais de pressão ao nível do mar, decidimos ainda avaliar a concordância dos resultados com os da série temporal do Índice de *El Niño* (dados de temperaturas ao nível do mar, em graus Celsius). Vários estudos científicos indicam que o fenómeno *El Niño* afeta a temperatura da água do mar e como tal também influencia a pressão ao nível do mar (Ilin *et al.*, 2005).

A Oscilação Sul do *El Niño* (ENSO) é um fenómeno que ocorre no Oceano Pacífico e na atmosfera, mas também conhecido por ser uma das fontes mais relevantes da variabilidade interanual no clima (em geral) e no tempo (em particular) à escala global. A sua componente oceânica pode ser definida como um aumento (*El Niño*) ou diminuição (*La Niña*) da temperatura ao nível do mar no Oceano Pacífico central e oriental. As ocorrências de aquecimento do *El Niño* são conhecidas por serem acompanhadas de um decréscimo na pressão atmosférica ao nível do mar a oeste no Pacífico, o qual traduz a componente atmosférica do ENSO, conhecida por Oscilação Sul (SO).

Analisemos a relação entre o Índice de *El Niño* e as componentes em estudo.

A Figura 3.23 fornece comparações entre as coordenadas das 3 CPs da ACP, das 2 CIs da ACI implementada com o FastICA e a série temporal do Índice de *El Niño*. Os resultados são

semelhantes entre o espectro da série do Índice de *El Niño* e os espectros das coordenadas das técnicas, uma vez que estes possuem estruturas idênticas evidenciando mais uma vez os dois principais picos de frequência referentes aos períodos anual e semestral. Portanto tudo indica existirem ligações relevantes entre o Índice de *El Niño* e as outras coordenadas das componentes (séries temporais) obtidas por ACP e ACI. No espectro do Índice de *El Niño* ainda se evidencia o pico de frequência de 0.0185, o que corresponde a um período de cerca de quatro anos e meio. É de notar que o fenómeno de *El Niño* ocorre em média em cada 4 a 5 anos, mas com intervalos que podem ir desde os 2 aos 7 anos. A duração do fenómeno pode variar entre 6 a 18 meses.

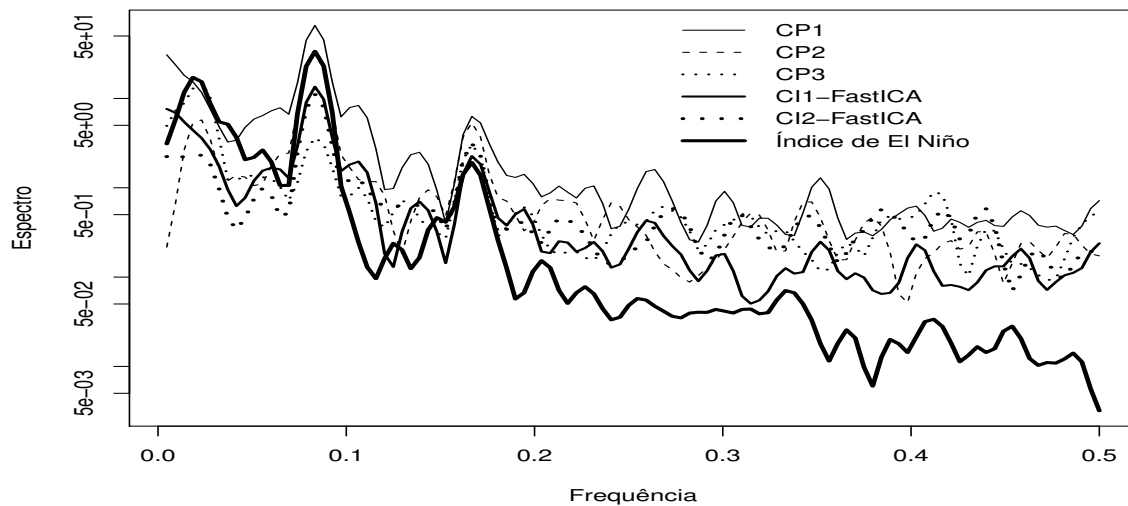


Figura 3.23: Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e da série temporal do Índice de *El Niño*

Ao analisar os espectros de fase na Figura 3.24 e os espectro de coerência na Figura 3.25, podemos constatar que as únicas séries temporais que estão em fase e que variam com coerência são CP1 e CI1-FastICA. Em relação ao Índice de *El Niño*, notamos que é difícil averiguar em que frequências é que este varia de uma forma coerente com as restantes coordenadas das componentes.

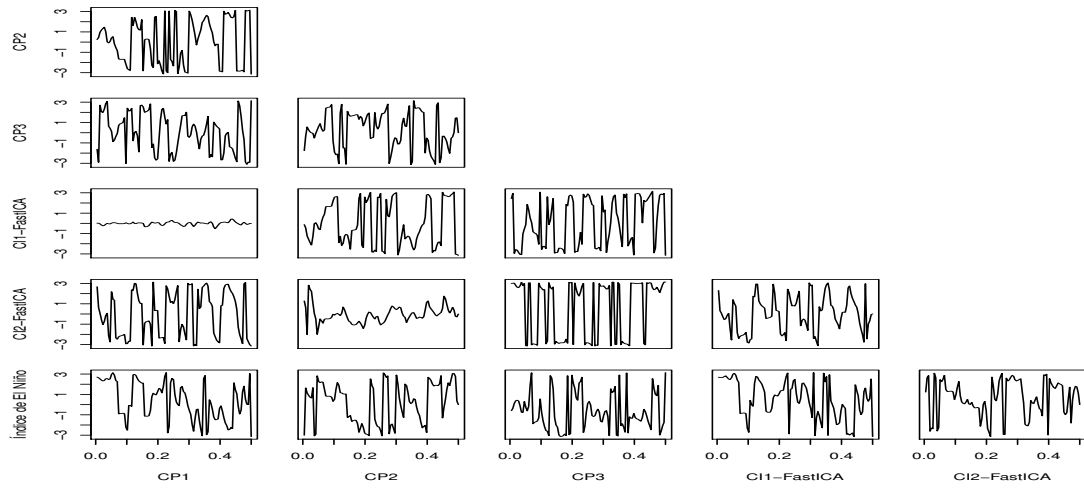


Figura 3.24: Espectros de fase para CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e série temporal do Índice de *El Niño*

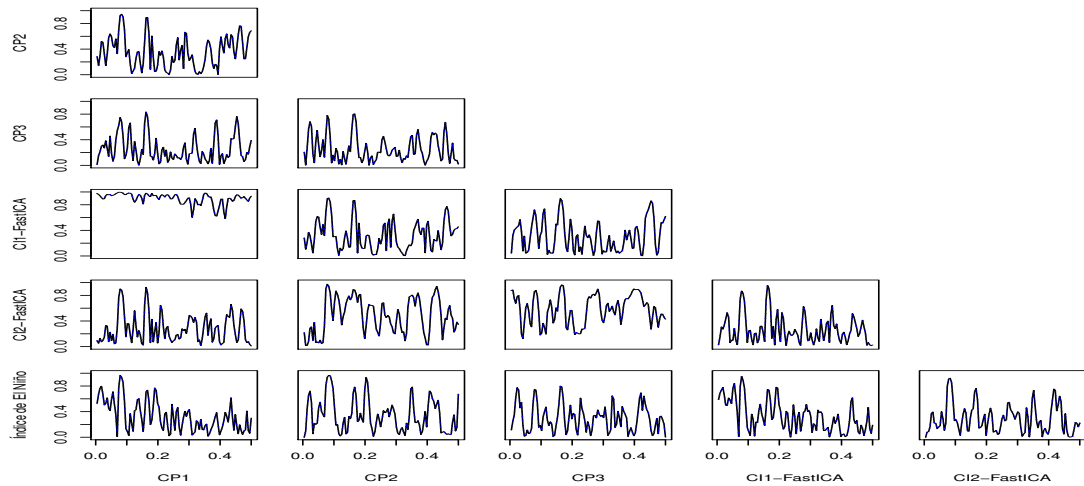


Figura 3.25: Espectros de coerência para CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e série temporal do Índice de *El Niño*

Comparações com a Série Temporal do Índice de Oscilação Sul (SOI)

Para além da inclusão no estudo do Índice de *El Niño* para a temperatura ao nível do mar, procedemos ainda a uma análise comparativa ao nível espectral entre o conjunto de dados em causa (valores mensais médios de pressão ao nível do mar no norte do Oceano Pacífico) e a pressão atmosférica ao nível do mar a oeste do Pacífico de acordo com o Índice de Oscilação Sul¹² (SOI) (Figura 3.26). Este Índice é uma medida das flutuações em grande escala na pressão do ar, que ocorrem no Pacífico tropical, durante os episódios dos fenómenos *El Niño* e *La Niña* (Trenberth, 1984). Sabe-se que períodos prolongados de valores negativos (positivos) do SOI coincidem com o aquecimento (arrefecimento) das águas em todo o leste do Pacífico tropical, o que é típico de *El Niño* (*La Niña*).

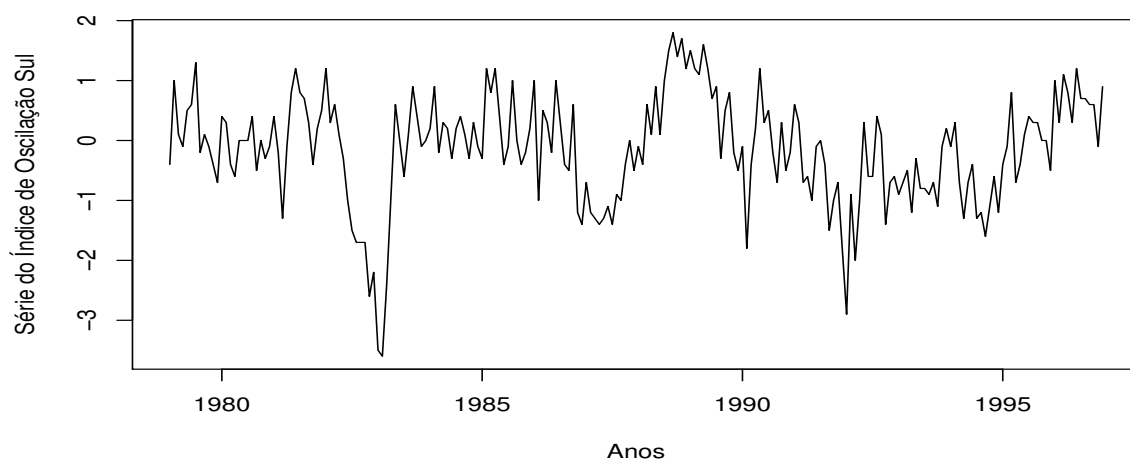


Figura 3.26: Série temporal dos valores mensais standardizados de pressão ao nível do mar do Índice de Oscilação Sul entre janeiro de 1979 e dezembro de 1996

Para analisar a relação entre o Índice de Oscilação Sul e as componentes envolvidas neste exemplo de aplicação, consideremos a Figura 3.27, que ilustra a comparação entre este índice e

¹²O Índice de Oscilação Sul (SOI) é um índice standardizado baseado nas diferenças de pressão ao nível do mar observadas entre Taiti e Darwin, na Austrália. Para este caso em estudo, a série do Índice de Oscilação Sul é constituída por 216 valores mensais (de janeiro de 1979 a dezembro de 1996) de pressão ao nível do mar. Os dados foram retirados de www.cpc.ncep.noaa.gov/data/indices/soi.

as coordenadas das 3 CPs da ACP e das 2 CIs da ACI obtida através do algoritmo FastICA.

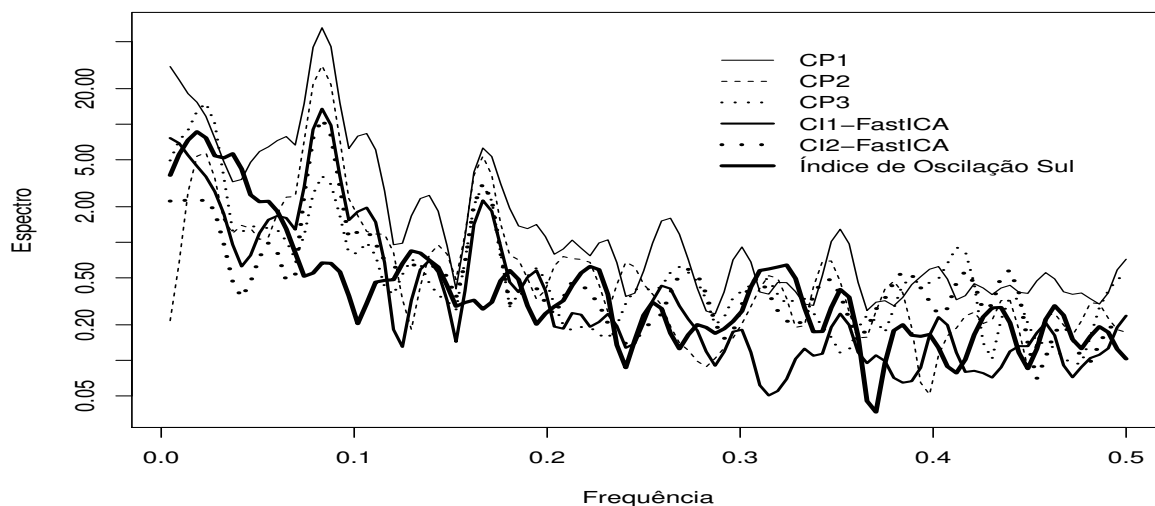


Figura 3.27: Espectros das coordenadas de CP1, CP2, CP3, CI1-FastICA, CI2-FastICA e da série temporal do Índice de Oscilação Sul

Os resultados obtidos não apresentam, de um modo geral, semelhanças entre o espectro do SOI e os espectros das coordenadas das restantes técnicas. Contudo, detetamos um pico de frequência de 0.0185 no espectro do SOI, que corresponde a um período de cerca de quatro anos e meio, o qual surge também no Índice de *El Niño*, permitindo assim estabelecer a relação que é conhecida entre o Índice de *El Niño* e o Índice de Oscilação Sul de acordo com o fenómeno ENSO. Ainda segundo a Figura 3.27 apercebemo-nos de que o SOI não apresenta picos de frequência destacáveis em torno de 0.083 nem em torno de 0.167 (períodos anual e semestral respetivamente), contrariamente à relevância destes em todas as outras componentes envolvidas na análise e também no Índice de *El Niño*. Uma vez que o SOI pretende captar essencialmente as variações que ocorrem associadas aos fenómenos *El Niño* e *La Niña*, os valores do índice são calculados através de expressões¹³ baseadas em anomalias seguidas de standardizações que levam à remoção do ciclo anual (Keppen e Ghil, 1992; Trenberth, 1984; Ropelewski e Jones, 1987).

¹³As fórmulas de cálculo do SOI utilizadas neste caso de estudo podem ser consultadas em <http://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/>

Tabela 3.2: Soma de quadrados dos resíduos para as reconstruções com a ACP e a ACI (através do FastICA e do AMUSE) à custa da primeira CP (CP1), da primeira CI (CI1), das duas primeiras CPs (CP12), das duas primeiras CIs (CI12) e das três primeiras CPs (CP123)

	CP1	CP12	CP123	CI1-FastICA	CI12-FastICA	CI1-AMUSE	CI12-AMUSE
CCity.CA	0.494	0.419	0.222	0.220	0.220	0.267	0.220
SDie.CA	0.684	0.329	0.160	0.629	0.465	0.717	0.465
SFranc.CA	0.466	0.426	0.057	0.262	0.111	0,392	0.111
Hilo.HW	0.527	0.211	0.132	0.602	0.599	0.613	0.599
Honol.HW	0.771	0.412	0.172	0.840	0.840	0.844	0.840
NBay.WA	1.159	0.331	0.326	1.203	0.688	0.926	0.688
Seld.AK	1.924	1.595	0.971	2.982	0.500	2.275	0.500
Sitka.AK	0.981	0.489	0.229	1.502	0.238	1.077	0.238
Sum	7.007	4.213	2.270	8.238	3.660	7.111	3.660

3.12.3 Reconstrução

Tendo em conta o conjunto de dados em análise, para cada uma das estações meteorológicas, é possível apresentar informação referente às reconstruções, através da ACP e da ACI, para os dados originais dos níveis de pressão ao nível do mar. Para avaliar a eficácia das reconstruções, recorre-se às somas de quadrados dos resíduos (SQR).

Em termos gerais, a reconstrução para a ACP realiza-se de acordo com a descrição da Secção 2.3 enquanto que para a ACI consiste na reconstrução dos dados originais multiplicando o vetor das componentes independentes estimadas pela matriz das misturas estimadas.

As reconstruções podem ser efetuadas com um determinado número de componentes dominantes para ambas as técnicas. Na Tabela 3.2 apresentam-se as somas de quadrados dos resíduos, para cada uma das séries temporais relacionadas com as estações meteorológicas, onde se utiliza um número distinto de componentes da ACP e da ACI (através dos algoritmos FastICA e AMUSE) nas reconstruções dos dados originais.

Quando se usa apenas a primeira componente, os valores das somas de SQR são praticamente idênticos entre as reconstruções de CP1 e CI1-AMUSE enquanto que a soma de SQR

de CII-FastICA é ligeiramente mais elevada. As reconstruções com as duas primeiras componentes apresentam valores de somas de SQR relativamente próximos para ambas as técnicas. Relembramos que, de acordo com o exposto no início da Secção 3.12, a comparação entre grupos de componentes analisada através do Coeficiente de Determinação Generalizado sugeria uma semelhança razoável entre os subespaços gerados pelas duas primeiras CPs e pelas duas CIs.

Há ainda a salientar a questão relacionada com o facto de cada técnica possuir o seu próprio critério de reconstrução. Com tais critérios distintos entre a ACP e a ACI era de esperar naturalmente resultados diferentes nas reconstruções. Contudo, quanto mais parecidos forem os valores das somas de SQR entre técnicas, significa que maior é a concordância das suas reconstruções. Por outro lado, quanto menores forem os valores das somas de SQR, maiores serão as semelhanças entre os dados originais e as suas reconstruções. Tal situação, como era de esperar, acontece à medida que o número de componentes intervenientes na reconstrução dos padrões aumenta.

3.13 Resumo e Conclusões

Na parte inicial deste Capítulo apresentámos a técnica da ACI, em que os dados observados são considerados como transformações lineares de fatores desconhecidos, designados de componentes independentes. O clássico problema *cocktail party* foi introduzido como exemplo que esteve na génese do desenvolvimento da técnica.

Na estimação do modelo básico da ACI, admite-se que as CIs são independentes entre si em termos estatísticos, e que todas as CIs ou todas exceto uma não seguem uma distribuição Normal. A estimação pressupõe a implementação de um algoritmo ao conjunto dos dados iniciais, aos quais é costume aplicarem-se técnicas de pré-processamento como a usual centragem das variáveis e o branqueamento, desde que seja efetuado através de uma transformação ortogonal, para que as CIs se tornem não correlacionadas e com variâncias unitárias. Este procedimento geral de aplicação do modelo ACI pode ser ilustrado no esquema da Figura 3.28.

De entre os vários métodos existentes para a estimação do modelo de ACI, muitos deles foram descritos ao longo do Capítulo e estão resumidos nas Tabelas 3.3 e 3.4. Também foram mencionados os algoritmos mais adequados para cada tipo de método.

Em relação a dados cujas CIs são constituídas por observações em que se admite à partida

Tabela 3.3: Resumo dos métodos de estimação do modelo ACI relacionados com a maximização da não normalidade

Curtose como Medida de Não Normalidade

- Método que consiste em encontrar as direções ortogonais das CIs cujos valores absolutos da curtose sejam maximizados.
- Utiliza algoritmos do gradiente e também do ponto fixo como por exemplo o FastICA.

Neguentropia como Medida de Não Normalidade

- Baseada na versão normalizada da entropia diferencial, alternativa ao uso da curtose.
- Problemas computacionais podem ser evitados, recorrendo a aproximações simples da neguentropia que utilizam cumulantes de ordem superior ou certas funções mais robustas.
- Usa algoritmos do gradiente e do ponto fixo (por exemplo uma versão do FastICA).

Tabela 3.4: Continuação do resumo dos métodos de estimação do modelo ACI

<p>Estimação de Máxima Verosimilhança</p> <ul style="list-style-type: none"> • É um dos princípios estatísticos mais usados de estimação. • Pode ser usada para estimar o modelo desde que sejam conhecidas as funções densidades das CIs à priori. • Recorre a algoritmos, desde o mais simples do gradiente ao rápido FastICA.
<p>Princípio do <i>Infomax</i></p> <ul style="list-style-type: none"> • Conceito com uma relação próxima da máxima verosimilhança usado em redes neuronais. • Baseia-se na maximização da entropia de saída ou fluxo de informação.
<p>Minimização da Informação Mútua</p> <ul style="list-style-type: none"> • Baseia-se na teoria da informação. • Tem a vantagem de permitir unificar muitos outros princípios de estimação. • Utiliza essencialmente os mesmos algoritmos que a maximização da não normalidade ou a estimação de máxima verosimilhança.
<p>Métodos Tensoriais</p> <ul style="list-style-type: none"> • Estimação do modelo através de tensores cumulantes de ordem superior, que constituem uma generalização da matriz de covariâncias, a partir da qual se pode aplicar a decomposição espectral cujos valores próprios são fornecidos pela curtose das CIs. • Podem usar-se o algoritmo da potência que é similar ao FastICA, o JADE para valores próprios iguais do tensor e o FOBI para casos de curtoses das CIs todas diferentes.

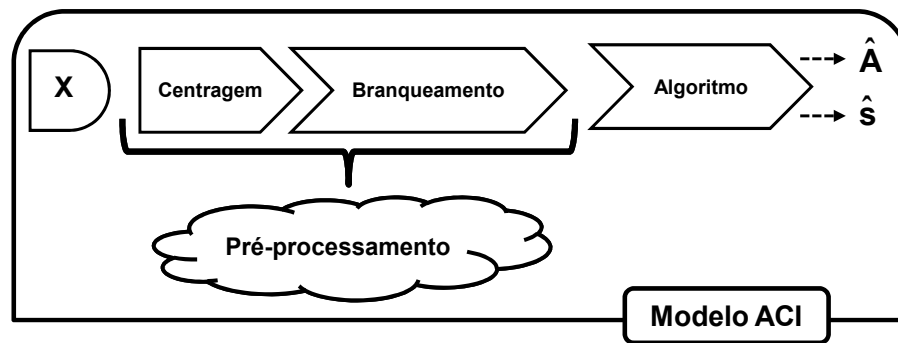


Figura 3.28: Esquema das etapas de aplicação do modelo ACI: fases do pré-processamento e da implementação de um algoritmo

possuírem uma estrutura de dependência temporal intrínseca, é preferível utilizar um modelo de ACI que retire o máximo de proveito dessa informação temporal como elemento essencial para melhorar a estimação do mesmo em detrimento, por exemplo, do modelo básico à custa da não normalidade. Foram apresentados dois casos em que a separação das CIs é possível (mesmo que estas sejam Normais) recorrendo a diferentes autocovariâncias e à não estacionariedade de variâncias. Para além disso, foi referido ainda o princípio da complexidade de Kolmogoroff que permite unificar todos os pressupostos anteriores.

Apesar de termos descrito apenas quatro algoritmos neste trabalho, o número de algoritmos existentes nem sempre proporciona ao investigador uma escolha rápida e fácil do algoritmo que se julgue ser o mais adequado possível ao seu problema. Para além disso, não existe um critério geral válido referente à escolha certa, embora esta deva depender da natureza das propriedades estatísticas das fontes (CIs) e da especificidade das aplicações (Cichocki e Amari, 2002). Podemos distinguir de grosso modo dois tipos de algoritmos, uns que funcionam em bloco para maximizar ou minimizar determinada função objetivo e outros que são adaptativos a cada instante de tempo consoante o processo, os quais são muitas vezes baseados em métodos do gradiente.

A implementação da maioria dos algoritmos acarreta a ambiguidade da permutação ao estimarem as CIs segundo uma ordem arbitrária. Geralmente, em muitas aplicações, é útil que as CIs estejam ordenadas, de acordo com determinadas propriedades estatísticas. Existem vários métodos de ordenação das CIs baseados em critérios específicos, no entanto neste trabalho são realçados apenas os cinco métodos ilustrados na Figura 3.29, os quais irão ter maior destaque

no exemplo de aplicação do Capítulo 5, no qual se retém um elevado número de CIs.

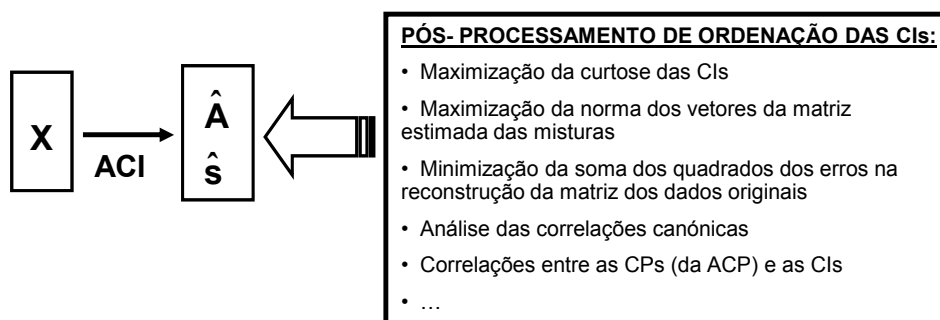


Figura 3.29: Esquema ilustrativo de alguns métodos de ordenação das CIs

Um outro aspeto importante, aquando do processo de aplicação de um dado algoritmo, reside na escolha do número adequado de CIs que devem ser retidas em detrimento de outras que contribuam com menor informação para a explicação do modelo. Esse número de CIs poderá ser o mesmo que é identificado pela ACP usual como o número adequado de componentes dominantes, embora tenham sido apontadas outras práticas.

Neste Capítulo foram ainda abordados superficialmente outros modelos, extensões ou derivações do modelo básico de ACI como é o caso do modelo com ruído, o modelo não linear, o modelo sujeito a determinado tipo de restrição, o modelo para números complexos ou o modelo para dados binários. Realçaram-se também as mais variadas áreas de atuação da ACI.

Por último foi efetuada uma aplicação prática que incidiu na comparação entre a utilização das técnicas da ACP e da ACI, em dados de pressão ao nível do mar, os quais foram ainda comparados com a série do Índice de *El Niño* para a temperatura ao nível do mar e com a série do Índice de Oscilação Sul para a pressão ao nível do mar no Pacífico. Para além do afamado algoritmo FastICA que pressupõe a não normalidade de todas as CIs (ou todas exceto uma), foi implementado o algoritmo AMUSE para confrontar com este último, uma vez que os dados em estudo são constituídos por observações ao longo do tempo, e como tal, admite-se à priori que possuem uma estrutura de dependência temporal, onde a questão da não normalidade das CIs não é imposta.

Verificou-se que na maior parte dos resultados do exemplo prático, o algoritmo AMUSE apresenta um desempenho ligeiramente superior ao do FastICA. A estrutura temporal contida

nos dados permitiu extrair a informação da autocorrelação nas primeiras componentes, a qual está diretamente relacionada com o Índice de *El Niño*. No que diz respeito ao estudo dos dados no domínio da frequência, também se constatou que a série do Índice de *El Niño* apresenta resultados semelhantes aos espectros das coordenadas das componentes dominantes obtidas pelas outras técnicas. Por outro lado, os índices de *El Niño* e de Oscilação Sul apresentam um mesmo período relevante correspondente a quatro anos e meio, característico da ocorrência média do fenómeno ENSO.

Neste Capítulo foi usado um único conjunto de dados com o intuito de ser o ponto de partida para explorar a técnica da ACI, comparando-a com a ACP, por forma a obter resultados que pudessem servir de suporte para eventuais extrapolações para conjuntos de dados mais complexos a abordar em Capítulos posteriores.

A análise comparativa entre grupos de componentes através do Coeficiente de Determinação Generalizado permitiu identificar semelhanças nos subespaços gerados por duas CPs e duas CIs. Tendo em conta as suas características algébricas, tal indicador pode ser usado como medida de escolha de pares de subespaços gerados por determinado número mais abrangente de componentes envolvendo a ACP e a ACI. Deste modo ao fixar as CPs dominantes podemos averiguar qual o melhor subconjunto de CIs que permite obter valores mais elevados deste indicador.

A comparação dos dados de pressão ao nível do mar com indicadores das mudanças temporais e espaciais em larga escala, que influenciam a variabilidade da circulação atmosférica, como o Índice de *El Niño* e o Índice de Oscilação Sul, mostram evidências de que a técnica de ACI apresenta convergência de resultados relativamente à ACP usual ao extrair a mesma informação no contexto de séries temporais.

Capítulo 4

Análise Espectral Singular

"Todos os rios correm para o mar, mas o mar não transborda."

Provérbio Chinês

4.1 Introdução

Neste Capítulo serão descritos conceitos relacionados com a Análise Espectral Singular (*Singular Spectrum Analysis* (SSA)), que é uma técnica multivariada de análise de séries temporais. A SSA que tem desempenhado um papel preponderante nas ciências atmosféricas, é conhecida em processamento de sinal (Pike *et al.*, 1984) onde está relacionada com a denominada expansão de Karhunen-Loève, assim como em processos biológicos (Basilevsky, 1979) e usada na teoria de sistemas dinâmicos (Broomhead e King, 1986a). Por vezes a SSA é também referida como Análise do Sistema Singular (von Storch e Zwiers, 1999).

Com base numa única variável registada ao longo do tempo considera-se que é possível aplicar uma análise estatística multivariada, uma vez que um único registo proveniente de um sistema dinâmico pode ser encarado como o resultado de todas as variáveis intervenientes no sistema, e como tal é sensato admitir que esse registo contenha informação razoável relacionada com a dinâmica de todos os fatores que contribuem para a evolução do próprio sistema.

O termo "espectral singular" está relacionado com a decomposição em valores singulares de uma matriz - que define o seu espectro singular. Por esta razão, a SSA poderia ser designada por análise de séries temporais usando o espectro singular.

A SSA é geralmente considerada como uma técnica multivariada para analisar uma série

temporal. A SSA envereda sobretudo pela parte exploratória onde o seu principal objetivo é decompor uma série original de dados num dado número de componentes para que se possam interpretar as componentes oscilatórias e a tendência, desprezando assim a estrutura de ruído. Após a decomposição é usual efetuar-se a reconstrução da série original dos dados através de um determinado número reduzido de componentes.

Em SSA, os três tipos de componentes interpretáveis numa série temporal são distinguidos da seguinte forma: a tendência é a componente de variação lenta de uma série temporal que não contenha componentes oscilatórias; as componentes oscilatórias descritas como aquelas que apresentam um comportamento harmónico com um determinado período associado, motivo pela qual algumas estão relacionadas com a sazonalidade; e o ruído descrito pelas componentes que aparentemente, do ponto de vista empírico, não contêm uma estrutura de tendência nem de oscilação. Há a acrescentar que a SSA é considerada uma ferramenta extremamente útil para a resolução de problemas tais como (Hassani, 2007):

- encontrar diferentes tendências;
- suavização;
- extração de componentes sazonais;
- extração simultânea de ciclos com pequenos e grandes períodos;
- extração de periodicidades com diferentes amplitudes;
- extração simultânea de tendências e periodicidades complexas;
- encontrar uma estrutura em séries temporais pequenas;
- detetar pontos de viragem.

De um modo geral, a série temporal original é transformada numa matriz de desfasamentos ao longo do tempo consoante o objetivo em causa, para posteriormente se aplicar a ACP usual a esta matriz aumentada. Além da análise exploratória de séries temporais, a técnica também pode ser usada para efetuar previsões. Veja-se, por exemplo, o artigo de Hassani *et al.* (2009) onde são analisadas oito séries da produção industrial na Alemanha, França e Reino Unido, e apresentados resultados acerca de previsões.

Nesse sentido, a SSA pode exibir resultados mais satisfatórios em relação às técnicas clássicas, nomeadamente quando os dados apresentam uma estrutura, como a do caso de séries que possuem componentes sazonais complexas. Na aplicação da SSA não há a necessidade de quaisquer suposições acerca da estacionariedade da série ou da normalidade dos resíduos. Uma

outra vantagem é a de que a técnica funciona geralmente bem, mesmo para amostras reduzidas.

Na decomposição da SSA, uma propriedade importante está relacionada com o facto de que a série original deve satisfazer uma fórmula recorrente linear, a qual permitirá a previsão de novos valores da série. No que se refere à obtenção dos intervalos de predição para essas previsões, para além do método empírico em que se pressupõe a mesma estrutura inerente aos novos dados, a SSA pode utilizar também a técnica do *bootstrapping* com base na continuação do sinal das componentes dominantes de toda a série.

Por outro lado, em estudos com séries de dados reais, surgem frequentemente dados omissos, o que pode reduzir a precisão dos resultados em diversas análises estatísticas caso o número de dados omissos seja demasiado grande. Atualmente existem alguns métodos baseados na SSA para preencher os dados omissos (por exemplo, Schoellhamer, 2001, Golyandina e Osipov, 2007).

Broomhead e King (1986a, 1986b) considerados os grandes impulsionadores da técnica, permitiram que esta fosse conhecida através das suas publicações na década de oitenta. De entre muitas outras referências existentes, destacamos os livros de Elsner e Tsonis (1996) para uma introdução à técnica e de Golyandina *et al.* (2001) para uma leitura mais aprofundada no que se refere a fundamentações teóricas.

Uma das aplicações mais importantes da SSA é a de análise de séries temporais de dados climáticos. Em 1989, Vautard e Ghil analisaram os núcleos de sedimentos oceânicos, onde os ciclos dominantes do paleoclima foram extraídos como pares de componentes oscilatórias na série dos rácios de isótopos de oxigénio. A SSA permitiu apresentar modulações fortes da amplitude desses ciclos ao longo da época quaternária. Uma outra aplicação foi levada a cabo por Ghil e Vautard (1991) em que a SSA foi usada nos registos globais da temperatura de superfície, a qual permitiu distinguir uma oscilação entre décadas, embora pouco significativa devido à falta de alguns registos. Em 2003, Baratta *et al.* apresentou um trabalho com novos resultados relativos à aplicação da SSA para a previsão individual de cada uma das séries temporais de precipitação diária, registadas por 135 estações meteorológicas distribuídas na bacia do Rio Tibre em Itália, assim como para a série de precipitação média diária, cujos valores foram calculados à custa das médias de todas as intensidades de precipitação das 135 estações. Também é comum encontrar algumas aplicações da SSA no estudo do Índice de Oscilação Sul, relacionado com o fenómeno do *El Niño* conhecido por ENSO (*El Niño/Southern Oscillation*), donde destacamos Rasmusson *et al.* (1990) e Keppenne e Ghil (1992). Outros exemplos podem ser analisados em Ghil e Mo

(1991), Vautard *et al.* (1992), Elsner e Tsonis (1996), Yiou *et al.* (1996), Yiou *et al.* (2000) e Ghil *et al.* (2002).

Embora se considere que a SSA seja ainda pouco utilizada na área da economia, das finanças ou das ciências sociais, outras áreas de aplicação têm vindo a ganhar destaque como é o caso da Geofísica (Varadi *et al.*, 1999), da Biomedicina (Maciel e Pereira, 2000), da Biomecânica (Alonso *et al.*, 2005) ou da Produção Industrial (Aldrich e Barkhuizen, 2003, Hassani *et al.*, 2009, Salgado e Alonso, 2006). Muitas são as referências e exemplos em que a SSA é aplicada com o objetivo principal de detetar as componentes oscilatórias, que muitas das vezes são extraídas nas primeiras componentes. Para analisar se as componentes oscilatórias extraídas com a SSA são ou não significativas, a simulação de Monte Carlo pode ser um bom algoritmo para efetuar os testes de hipóteses para discriminar entre potenciais oscilações e componentes de ruído (Allen e Smith, 1994 e 1996).

A SSA é encarada como um possível caso particular de aplicação da ACP, também designada Análise de Funções Ortogonais Empíricas (EOF) em contexto climático, onde os dados que constituem a série são observados na mesma localização mas em diferentes desfasamentos temporais. Nestes casos, os principais valores e vetores próprios conjuntamente com as respetivas CPs resultantes da decomposição espectral permitem representar os principais padrões temporais que ocorrem no sistema físico climático.

Pelo facto da SSA desempenhar desde cedo um papel fulcral na área da Climatologia, a técnica foi estendida para incorporar séries multivariadas ou multicanal, na designada Análise Espectral Singular Multicanal (MSSA) ou Análise de Funções Ortogonais Empíricas Estendidas (EEOF) que será abordada mais adiante. Para além da SSA existem outros métodos para decompor um sinal no domínio da frequência, os designados métodos de análise espectral tais como o Método de Máxima Entropia (MEM) e o Método *Multitaper* (MTM) que podem ser encontrados, por exemplo, em Vautard *et al.* (1992) ou em Ghil *et al.* (2002).

Neste Capítulo iremos descrever a SSA, salientar alguns tópicos relacionados como a escolha do desfasamento temporal, a previsão de novos valores da série temporal ou a centragem dos dados, assim como apresentar uma aplicação à série mensal do número de vítimas mortais em acidentes de viação em Portugal Continental entre 1998 e 2009.

4.2 Descrição da SSA

Na SSA parte-se de uma única variável registada ao longo do tempo - a série temporal - a qual é inicialmente decomposta em várias variáveis com o mesmo número de observações através de defasamentos no tempo. A partir desta matriz constituída pelas novas variáveis, aplica-se a ACP.

O método básico da SSA (Golyandina *et al.*, 2001) é constituído por duas etapas complementares: a decomposição e a reconstrução. A primeira etapa consiste em decompor uma única série temporal em várias componentes, enquanto a segunda etapa tem como finalidade reconstruir a série original para eventualmente usar essa nova série reconstruída, e sem ruído, na previsão de novos valores. Cada uma destas etapas contempla dois passos fundamentais que passaremos a descrever de seguida.

Consideremos um processo estocástico $\{X_t\}$, constituído por uma sequência de variáveis aleatórias no tempo. Sem perda de generalidade, denominemos uma sequência de n observações desse processo por série temporal $X_t = (x_1, x_2, \dots, x_n)$, realizadas em pontos regulares ou irregulares no tempo (valores finitos de t).

4.2.1 Decomposição

1º Passo - Incorporação (*Embedding*):

Dada a série original, consideremos um tamanho apropriado para o vetor desfasado, o qual se designa por comprimento da janela m , com $1 < m < n$. A série unidimensional é assim transformada numa série multidimensional constituída por uma sequência de $k = n - m + 1$ vetores desfasados, de dimensão m .

Consideremos a série temporal X_t , e os k vetores desfasados $\mathbf{x}_j = [x_j : x_{j+1} : \dots : x_{j+m-1}]^T$, com $j = 1, 2, \dots, k$.

Definição 4.2.1 A matriz de trajetória ou também designada por matriz aumentada dos dados desfasados \mathbf{X} é dada por

$$\mathbf{X} = [\mathbf{x}_1 : \mathbf{x}_2 : \dots : \mathbf{x}_k]^T = \begin{bmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m+1} & x_{n-m+2} & \dots & x_n \end{bmatrix}. \quad (4.1)$$

A matriz de trajetória \mathbf{X} é encarada como um conjunto multivariado de observações que contém o registo dos padrões que ocorrem dentro de uma janela de comprimento m . Esta matriz é constituída por um número de linhas igual a k vetores desfasados e um número de colunas igual ao comprimento da janela m . Designando por diagonal da matriz a qualquer conjunto de elementos cujos índices de linha i ($i = 1, 2, \dots, k$), e de coluna j ($j = 1, 2, \dots, m$), verificam $i + j = c$, para uma dada constante c , note-se que todos os elementos de uma diagonal são iguais. Deste modo, a matriz de trajetória (4.1) é uma matriz de Hankel (Focardi e Fabozzi, 2004).

2º Passo - Decomposição em Valores Singulares (SVD) de \mathbf{X} :

Os padrões repetidos que representam as oscilações na série original podem ser analisados com base na matriz de trajetória (4.1) a partir da decomposição em valores singulares (Teorema 1.4.12) da matriz \mathbf{X} .

Dada a matriz \mathbf{X} do tipo $k \times m$, constrói-se a matriz simétrica $\mathbf{X}^T \mathbf{X}$.

Denotemos por $\lambda_1, \lambda_2, \dots, \lambda_m$ os m valores próprios de $\mathbf{X}^T \mathbf{X}$ de modo que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ e por U_1, U_2, \dots, U_m os correspondentes m vetores próprios ortonormados de $\mathbf{X}^T \mathbf{X}$.

A SVD decompõe a matriz \mathbf{X} na expressão matricial

$$\mathbf{X} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T,$$

onde

- $\mathbf{V} = (V_1, V_2, \dots, V_m)$ é designada por matriz das componentes principais de \mathbf{X} ,
- $\mathbf{\Lambda} = (\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_m})$ representa a matriz diagonal dos valores singulares de \mathbf{X} e
- $\mathbf{U} = (U_1, U_2, \dots, U_m)$ é designada por matriz dos vetores singulares (direitos) de \mathbf{X} .

Sejam $d = \text{car}(\mathbf{X}) = \max\{i : \lambda_i > 0\} \leq m$ e V_i os vetores dados por $V_i = \mathbf{X}U_i/\sqrt{\lambda_i}$, para $i = 1, \dots, d$, então a SVD da matriz \mathbf{X} pode ser escrita como a soma de matrizes \mathbf{X}_i de

característica unitária de forma que

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d = \sum_{i=1}^d \sqrt{\lambda_i} V_i U_i^T, \quad (4.2)$$

onde $\mathbf{X}_i = \sqrt{\lambda_i} V_i U_i^T$, $i = 1, \dots, d$.

Definição 4.2.2 O terno $(\sqrt{\lambda_i}, U_i, V_i)$ é designado na terminologia inglesa por *eigentriple* da SVD de \mathbf{X} , onde $\sqrt{\lambda_i}$ ($i = 1, \dots, d$) são os valores singulares da matriz \mathbf{X} e o conjunto $\{\sqrt{\lambda_i}\}$ é chamado o espectro singular da matriz \mathbf{X} .

Os sucessivos valores singulares (e de forma análoga os respetivos valores próprios) podem ser ordenados de forma decrescente num gráfico onde o nível de ruído surge com o aspeto de uma cauda praticamente plana na parte final do espectro. Acima deste patamar de ruído situam-se os valores singulares associados à tendência e às componentes oscilatórias, cujo número é definido como dimensão estatística (Vautard e Ghil, 1989), a qual depende essencialmente da qualidade dos dados, do período temporal correspondente à totalidade dos mesmos e por vezes do valor do comprimento da janela considerado.

Na SVD apresentada, que na realidade corresponde a uma ACP da matriz \mathbf{X} mas sem a centragem prévia das variáveis da matriz e sem o escalamento por $1/\sqrt{k}$, há que considerar as propriedades relacionadas com as direções determinadas pelos vetores próprios de modo similar ao que acontece na ACP usual. Desta forma, o primeiro vetor próprio U_1 identifica a direção tal que a variabilidade das projeções dos vetores desfasados, segundo essa mesma direção, seja máxima. Cada um dos vetores próprios seguintes determina a direção que é ortogonal a todas as direções estabelecidas anteriormente, e de forma que a variabilidade das projeções dos vetores desfasados ao longo dessa nova direção continue a ser máxima, relativamente às restantes direções disponíveis.

4.2.2 Reconstrução

3º Passo - Agrupamento (*Grouping*):

O procedimento de agrupamento consiste em subdividir as matrizes \mathbf{X}_i ($i = 1, \dots, d$) obtidas em (4.2) em vários grupos e somá-las dentro de cada grupo, tendo em vista a reconstrução à custa das componentes de alguns desses grupos. Seja $I = \{i_1, \dots, i_p\}$, um grupo de índices i_1, \dots, i_p ;

para cada grupo I corresponde a matriz \mathbf{X}_I definida por $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. A partição do conjunto de índices $\{1, \dots, d\}$ em r subconjuntos disjuntos dois a dois, I_1, I_2, \dots, I_r , corresponde à representação

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_r}. \quad (4.3)$$

A contribuição da matriz \mathbf{X}_I , para cada grupo I de componentes principais, é medida por (Golyandina *et al.*, 2001)

$$\frac{\sum_{i \in I} \lambda_i}{\sum_{i=1}^d \lambda_i}.$$

4º Passo - Média ao Longo das Diagonais (*Diagonal Averaging*):

O último passo, que pode ser visto quase como que a operação inversa em relação ao primeiro passo, consiste em reconstruir a série de tamanho n a partir da matriz \mathbf{X}_I para cada grupo I , onde cada uma pode ser encarada como uma componente aditiva da série inicial.

Regra geral, em termos práticos, cada diagonal da forma $i + j = c$ (para uma constante c) de cada matriz \mathbf{X}_I não é constituída por elementos iguais. É necessário um processo que consista em transformar uma matriz arbitrária (que à partida não é uma matriz de Hankel) numa matriz de Hankel e conseqüentemente numa série. Os valores das novas séries reconstruídas são obtidos através da operação que corresponde a efetuar a média dos valores de cada matriz \mathbf{X}_I ao longo das diagonais $i + j = c$.

Definição 4.2.3 Consideremos uma matriz \mathbf{Y} com elementos y_{ij} , $1 \leq i \leq m$, $1 \leq j \leq k$. Sejam $m^* = \min\{m, k\}$, $k^* = \max\{m, k\}$ e $n = m + k - 1$ e seja $y_{ij}^* = y_{ij}$ se $m < k$ e $y_{ij}^* = y_{ji}$ se $m \geq k$. A média ao longo das diagonais transfere a matriz \mathbf{Y} para a série reconstruída \tilde{X}_t , com $t = 0, 1, \dots, n - 1$, definida através de

$$\tilde{x}_t = \begin{cases} \frac{1}{t+1} \sum_{r=1}^{t+1} y_{r,t-r+2}^* & , \text{ para } 0 \leq t < m^* - 1 \\ \frac{1}{m^*} \sum_{r=1}^{m^*} y_{r,t-r+2}^* & , \text{ para } m^* - 1 \leq t < k^* \\ \frac{1}{n-t} \sum_{r=t-k^*+2}^{n-k^*+1} y_{r,t-r+2}^* & , \text{ para } k^* \leq t < n \end{cases} \quad (4.4)$$

A expressão (4.4) corresponde a efetuar a média dos valores da matriz ao longo das diagonais $i + j = t + 2$, de tal forma que

$$\begin{aligned} t &= 0 \rightarrow \tilde{x}_0 = y_{11}, \\ t &= 1 \rightarrow \tilde{x}_1 = (y_{12} + y_{21})/2, \\ t &= 2 \rightarrow \tilde{x}_2 = (y_{13} + y_{22} + y_{31})/3, \end{aligned}$$

e assim por diante.

Ao aplicar o procedimento (4.4) a cada componente matricial \mathbf{X}_I de (4.3) obtemos a série $\tilde{X}_t = (\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_{n-1})$. Consequentemente, ao utilizar a contribuição dos r grupos, a série inicial X_t é decomposta na soma de r séries

$$x_t = \sum_{h=1}^r \tilde{x}_t^{(h)}, \quad (4.5)$$

onde a série reconstruída $\tilde{X}_t^{(h)} = (\tilde{x}_0^{(h)}, \tilde{x}_1^{(h)}, \dots, \tilde{x}_{n-1}^{(h)})$ corresponde à matriz \mathbf{X}_{I_h} . A igualdade na expressão (4.5) apenas ocorre se forem utilizadas as séries reconstruídas à custa de todas as componentes.

4.3 Comprimento da Janela

Na descrição da técnica da SSA na Secção anterior, no primeiro passo da decomposição, introduzimos o parâmetro m que é vulgarmente designado como comprimento da janela ou parâmetro de desfasamento (*delay parameter*). Este parâmetro é também conhecido por dimensão de incorporação (*embedding dimension*) na teoria de sistemas dinâmicos.

A questão inicial mais importante para a aplicação da SSA a uma série temporal é a decisão acerca do valor apropriado do comprimento da janela m . De uma forma geral considera-se que não existem cálculos numéricos estabelecidos com exatidão para a escolha do comprimento da janela. Contudo, este deve ter em conta o problema em estudo e a informação preliminar acerca da série temporal. Nos casos em que não se utiliza diretamente o conhecimento prévio acerca do contexto em que a série se insere, a única fonte de informação a utilizar na escolha do comprimento da janela para a série é a da estrutura da própria série. Um dos indicadores mais frequentes é a forma do gráfico da série, a qual permite identificar visualmente uma tendência ou oscilações.

Caso a série seja suficientemente grande, a análise do espectro pode ajudar a determinar as frequências e por consequência os períodos das componentes oscilatórias da série e a selecionar o comprimento da janela que seja recomendável para a sua separação. Para além disso, a presença de valores claramente distintos em termos de frequências no espectro permite relacionar essas frequências com as componentes que à partida irão existir na série.

Uma vez que existem vários princípios gerais para selecionar o comprimento da janela (Golyandina *et al.*, 2001), passaremos à descrição de alguns que consideramos mais relevantes.

1. Quanto maior for o comprimento da janela, maior será o detalhe da decomposição da série. Na análise da estrutura da série temporal através do modelo básico de SSA, o valor de m deverá ser suficientemente grande mas não ultrapassar metade do comprimento da série, ou seja deveremos considerar $m \leq n/2$.
2. Para evitar efeitos de fraca separação das componentes, deveremos evitar utilizar valores demasiado pequenos de m , os quais poderiam levar a que as componentes interpretáveis surgissem misturadas. Por outro lado, para valores de m relativamente grandes, os resultados da separação tendem a ser estáveis perante pequenas variações nos valores de m .
3. Por vezes poderão existir efeitos negativos provenientes da proximidade de valores singulares relacionados com diferentes componentes. Por exemplo, para uma série com uma estrutura mais complexa, um valor de m muito grande poderá produzir uma decomposição inadequada das componentes de interesse, uma vez que estas podem permanecer misturadas com outras componentes. Contudo, pode ainda acontecer que uma pequena variação no valor de m seja o suficiente para que a separação das componentes passe de fraca a forte ou o contrário.
4. Quando prevemos que a tendência seja uma curva relativamente suave, a sua extração e separação das restantes componentes oscilatórias e do ruído deverá requerer grandes valores de m .
5. Nos casos em que o comportamento da série depende de certas oscilações temporais como a sazonalidade, deveremos escolher um valor de m por forma a extrair as componentes periódicas.

No caso do último princípio enumerado, a referida escolha está intrinsecamente relacionada com o período dessas componentes. Ao medir o comprimento da série, n , é natural que este seja múltiplo de um determinado período T , e como tal, para separar a componente periódica do ruído, é costume considerar que a razão n/T seja um número inteiro não demasiado pequeno. Para séries curtas, na eventualidade de sabermos que a série possui componente oscilatória com período inteiro T , é recomendável usar um comprimento da janela proporcional a esse período para obtermos bons resultados na separação. Para séries longas, a proporcionalidade entre m , n e T não é uma questão deveras importante, apenas se aconselha a que o m seja próximo de $n/2$, e se possível mantendo a razão m/T como inteira (Golyandina *et al.*, 2001).

No caso em que a componente de tendência da série tem um aspeto similar a uma função conhecida (exponencial, função linear, função quadrática, ...) ou até casos mais complexos em que a componente residual apresente elevada variabilidade, e por isso, de difícil separação deveremos ser de alguma forma parcimoniosos, pois nem sempre os valores elevados serão os mais sensatos para a escolha de m .

Caso pretendamos extrair em simultâneo diferentes componentes da série, todas as recomendações anteriores deverão ser levadas em linha de conta. As periodicidades mais importantes deverão ser extraídas ao escolher um comprimento da janela múltiplo de todos esses períodos, e ao mesmo tempo m deverá permitir separar de forma razoável a tendência, não esquecendo que para isso o m deverá ser suficientemente grande.

Nos trabalhos de Vautard *et al.* (1992), um dos primeiros em que se extraíram tendências e componentes oscilatórias em meteorologia e geofísica, sugere-se que o comprimento da janela seja tal que permita analisar com sucesso os períodos que venham a pertencer ao intervalo $(m/5, m)$.

Por toda a panóplia de possibilidades existente para a escolha do comprimento da janela, de acordo com a natureza da série e os objetivos a atingir, nem sempre é fácil decidir pelo melhor valor, pelo que por vezes podem efetuar-se simulações para distintos valores de m até que o método pareça estar calibrado. Num artigo de Rodrigues e de Carvalho (2008) podemos encontrar um exemplo de uma análise comparativa entre várias medidas de precisão na previsão de futuros valores de uma série, mediante diferentes valores de m , onde se alerta para a precaução na sua escolha de forma a que essa decisão não prejudique a referida precisão.

4.4 Previsão na SSA

Um dos objetivos finais na SSA é a previsão de novos valores da série temporal. Segundo Golyandina *et al.* (2001), uma previsão razoável para uma determinada série só deverá ser levada a cabo se as seguintes condições forem conhecidas:

- a série possuir uma estrutura;
- existir um método ou algoritmo que identifique a estrutura;
- estar disponível um método para a continuação da série temporal, baseado na estrutura identificada;
- a série preservar a estrutura para um período futuro ao longo do qual se irá efetuar a previsão da série.

Geralmente um modelo de previsão é construído com base nos dados originais. Os modelos de previsão em SSA podem ser aplicados a uma série temporal que satisfaça aproximadamente a fórmula recorrente linear.

Definição 4.4.1 *Diz-se que uma série temporal $X_t = (x_1, x_2, \dots, x_n)$ satisfaz a fórmula recorrente linear de ordem ou dimensão não superior a d com $1 \leq d \leq n$ se existirem coeficientes a_1, a_2, \dots, a_d tais que*

$$x_{i+d} = \sum_{k=1}^d a_k x_{i+d-k}, \quad \text{com } 1 \leq i \leq n-d \quad \text{e} \quad a_d \neq 0.$$

A série que obedece a fórmulas recorrentes lineares admite uma continuação recorrente natural desde que cada termo da série seja igual a uma combinação linear de outros termos anteriores. Os coeficientes de tal combinação linear podem também ser utilizados para previsões futuras.

A classe de séries temporais que são regidas por fórmulas recorrentes lineares é muito vasta, uma vez que esta inclui séries harmónicas, polinomiais e exponenciais.

Admitamos que uma série X_t é decomposta por $X_t^{(1)} + X_t^{(2)}$ onde se pretende resolver o problema de previsão do sinal $X_t^{(1)}$ na presença do ruído $X_t^{(2)}$, segundo os pressupostos:

- a série $X_t^{(1)}$ admite previsões futuras através de uma fórmula recorrente linear de uma dimensão d relativamente pequena e;
- existe um comprimento da janela m em que se admite que $X_t^{(1)}$ e $X_t^{(2)}$ são aproximadamente separáveis.

Como exemplo de aplicação de métodos de previsão, considere-se Hassani (2007) em que a técnica da SSA foi utilizada e comparada com outros modelos e algoritmos de previsão. Os resultados mostraram que a técnica da SSA forneceu uma melhor precisão na previsão do que os outros métodos considerados.

4.4.1 Algoritmo de Previsão Recorrente em SSA

Uma vez que as primeiras componentes da SSA, V_i , (também aqui abreviadas por CPs tal como na técnica da ACP) são encaradas como versões filtradas do sinal de uma série, o seu comportamento é mais regular que o da série original e portanto mais previsível. Enquanto que as primeiras CPs contêm mais informação sobre a série original, e consideradas aquelas que são significativas em termos de tendência ou oscilação, as últimas representam o ruído. A escolha do m e naturalmente do subconjunto de CPs a reter terão implicações na decisão do que é o sinal e o que é ruído para se efetuar a previsão (Vautard *et al.*, 1992).

Apresentamos de seguida a descrição formal do algoritmo de Golyandina *et al.* (2001), designado por algoritmo de previsão recorrente em SSA (*SSA recurrent forecasting algorithm*), que permite selecionar apenas um subconjunto de CPs, para efetuar a previsão de futuros valores de uma série temporal. Como parâmetros de entrada do algoritmo temos:

- A série temporal $X_t = (x_1, x_2, \dots, x_n)$ com $n > 2$.
- O comprimento da janela m , com $1 < m < n$.
- O número r de CPs a reter.
- O espaço linear contido em \mathbb{R}^m de dimensão r inferior a m , definido à custa da base constituída pelos r vetores próprios ortonormados a reter.
- O número de pontos h a prever.

Passamos a descrever os passos necessários para a implementação do algoritmo, em que os primeiros 6 passos correspondem exatamente à anterior descrição do método SSA:

1. A partir da série temporal original $X_t = (x_1, x_2, \dots, x_n)$ constrói-se a matriz de trajetória $\mathbf{X} = [\mathbf{x}_1 : \mathbf{x}_2 : \dots : \mathbf{x}_k]^T$ para um determinado comprimento da janela m .
2. Aplica-se a decomposição em valores singulares da matriz \mathbf{X} e selecionam-se os r vetores próprios U_1, U_2, \dots, U_r de $\mathbf{X}^T \mathbf{X}$ que se pretendem reter.
3. Obtêm-se as componentes principais V_i e as componentes matriciais \mathbf{X}_i .
4. Identifica-se quantos e quais os grupos I a formar.
5. Criam-se as matrizes \mathbf{X}_I .
6. Procede-se à aplicação da média ao longo das diagonais de cada componente matricial \mathbf{X}_I e obtém-se a série reconstruída $\tilde{X}_t = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$.
7. Para cada vetor próprio $U = (u_1, u_2, \dots, u_{m-1}, u_m)^T \in \mathbb{R}^m$ denota-se o vetor dos primeiros $m-1$ elementos do vetor U por $U^\nabla \in \mathbb{R}^{m-1}$.
8. Considera-se o coeficiente $v^2 = \pi_1^2 + \pi_2^2 + \dots + \pi_r^2 < 1$, onde π_i é o último elemento do vetor próprio U_i ($i = 1, 2, \dots, r$).
9. Tendo em conta que o último elemento u_m de qualquer vetor próprio é uma combinação linear dos restantes elementos u_1, u_2, \dots, u_{m-1} :

$$u_m = a_1 u_{m-1} + a_2 u_{m-2} + \dots + a_{m-2} u_2 + a_{m-1} u_1,$$

considere-se o vetor dos coeficientes da combinação linear dado por $A = (a_{m-1}, a_{m-2}, \dots, a_1)^T$ que pode ser expresso por

$$A = \frac{1}{1-v^2} \sum_{i=1}^r \pi_i U_i^\nabla.$$

Finalmente as h previsões $\hat{x}_{n+1}, \hat{x}_{n+2}, \dots, \hat{x}_{n+h}$ são obtidas por

$$\hat{x}_i = \begin{cases} \tilde{x}_i & , \text{ para } i = 1, 2, \dots, n \\ \sum_{j=1}^{m-1} a_j \hat{x}_{i-j} & , \text{ para } i = n+1, n+2, \dots, n+h \end{cases}.$$

É de realçar que o algoritmo de previsão recorrente apresentado não deixa de ser um algoritmo básico que tem uma relação direta com a fórmula recorrente linear, o qual pode ser visto como o principal algoritmo de previsão em SSA. Golyandina *et al.* (2001) apresenta ainda modificações naturais deste algoritmo, as quais em determinadas situações particulares podem fornecer melhores resultados ao nível da previsão de futuros valores. São exemplos o *SSA vector forecasting algorithm*, o *Toeplitz SSA recurrent forecasting algorithm* ou o *Toeplitz SSA vector forecasting algorithm* assim como as suas variantes centradas.

4.4.2 Intervalos de Predição

Como o algoritmo de previsão recorrente depende do comprimento da janela e do número de CPs a reter na reconstrução, então ao se alterarem os parâmetros de entrada, os valores das previsões irão variar. Tal como num modelo clássico de predição, devem-se criar limites inferiores e superiores entre os quais seja expetável encontrar os valores previstos com uma certa confiança estipulada de acordo com a teoria dos intervalos de predição.

Admitamos as suposições de previsão, estabelecidas no início da Secção 4.4. Consideramos ainda que o sinal $X_t^{(1)}$ é uma subsérie finita de uma série infinita designada por $X^{(1)}$. Por analogia, $X_t^{(2)}$ é vista como uma subsérie finita de uma série infinita de ruído aleatório $X^{(2)}$ que perturba o sinal aleatório $X^{(1)}$.

Para construir intervalos de predição para a previsão de futuros valores da série, há que considerar dois problemas inerentes. Um deles é a construção de intervalos de predição para a série $X = X^{(1)} + X^{(2)}$ na sua totalidade, para os pontos futuros, admitindo que esta possui a mesma estrutura no futuro. O outro problema prende-se com a construção de intervalos de predição para o sinal $X^{(1)}$ para os mesmos pontos temporais futuros. Para a resolução destes problemas utiliza-se a informação acerca dos erros de previsão que são obtidos através do processamento da série, a qual se costuma designar por forma empírica, ou usa-se a informação adicional da série reconstruída $\tilde{X}_t^{(2)}$ para elaborar uma simulação não paramétrica (por métodos *bootstrap*) da série X_t .

De seguida descrevemos sumariamente apenas o caso empírico de construção de intervalos de predição para cada valor futuro x_{n+h} da série X para o algoritmo básico de previsão recorrente. Após a obtenção dos valores de previsão $\hat{x}_{n+h}^{(1)}$, ou seja, depois de se terem efetuado h passos do

algoritmo, cada valor $\hat{x}_{n+h}^{(1)}$ é definido como sendo a previsão do termo $x_{n+h}^{(1)}$ do sinal $X^{(1)}$.

No procedimento a adotar considera-se um número inteiro relativamente pequeno h e aplicam-se h passos de continuação recorrente. Em cada passo os dados iniciais serão atualizados. Os últimos pontos das continuações recorrentes são comparados com os valores da série inicial original X_t , estabelecendo assim uma série de resíduos.

Para um determinado nível de confiança $1 - \alpha$, e considerando $c_{\alpha/2}^-$ e $c_{\alpha/2}^+$ os quantis inferior e superior para $\alpha/2$ calculados a partir da função de distribuição empírica da série dos resíduos dos h passos, obtém-se o intervalo de predição empírico

$$(\hat{x}_{n+h}^{(1)} + c_{\alpha/2}^-, \hat{x}_{n+h}^{(1)} + c_{\alpha/2}^+),$$

o qual deverá conter o verdadeiro valor x_{n+h} da série X com um nível de confiança aproximado de $1 - \alpha$ (Golyandina *et al.*, 2001).

4.5 Matriz de Toeplitz na Decomposição

Consideremos um processo estocástico $\{X_t\}$, e sem perda de generalidade, a série temporal $X_t = (x_1, x_2, \dots, x_n)$ para a decomposição da SSA, onde a matriz dos desfaseamentos dos dados \mathbf{X} (definida em (4.1)) para um determinado comprimento da janela m , pode ser utilizada para obter uma estimativa da matriz $\mathbf{S} = \mathbf{X}^T \mathbf{X}$. Para evitar a confusão de notações matriciais, alerta-se que \mathbf{S} é definida no Capítulo da ACP como a matriz de variâncias-covariâncias (a qual envolve a centragem da matriz dos dados \mathbf{X}).

Na decomposição em valores singulares (SVD) da matriz \mathbf{X} , em vez de \mathbf{S} , passamos a considerar sem perda de generalidade,

$$\mathbf{S}^k = \frac{1}{n - m + 1} \mathbf{X}^T \mathbf{X} = \frac{1}{k} \mathbf{X}^T \mathbf{X} = \frac{1}{k} \mathbf{S}.$$

A diferença entre as decomposições SVD das matrizes \mathbf{S} e \mathbf{S}^k consiste apenas na diferença do tamanho dos correspondentes valores próprios, uma vez que estes são k vezes maiores para \mathbf{S} e os vetores próprios são idênticos. Portanto podemos utilizar indiferentemente qualquer das matrizes na SVD.

De entre as várias formas de estimar a matriz \mathbf{S}^k , que é usada para efetuar a decomposição espectral, passamos a descrever os dois estimadores mais importantes.

- *Estimador de Broomhead & King*, \mathbf{S}_{BK} : fornece uma estimação básica da matriz \mathbf{S}^k , onde cada elemento ij da matriz \mathbf{S}^k é dado por (Broomhead e King, 1986a)

$$S_{BK,ij} = \frac{1}{n-m+1} \sum_{t=1}^{n-m+1} x_{i+t-1}x_{j+t-1}, \text{ com } 1 \leq i, j \leq m.$$

- *Estimador de Vautard & Ghil*, \mathbf{S}_{VG} : uma alternativa foi sugerida em 1989 por Vautard e Ghil, em que cada elemento ij da matriz \mathbf{S}^k é definido por

$$S_{VG,ij} = \frac{1}{n-|i-j|} \sum_{t=1}^{n-|i-j|} x_{|i-j|+t}x_t, \text{ com } 1 \leq i, j \leq m.$$

É de realçar que para estimar a matriz de variâncias-covariâncias \mathbf{S} da ACP, pelos estimadores apresentados anteriormente, pode-se aplicar a ACP na matriz dos desfasamentos dos dados \mathbf{X} bastando subtrair a cada elemento a média da respetiva coluna.

Nos casos em que se admite a estacionariedade para a série temporal X_t , a covariância para um desfasamento de s unidades no tempo só depende do desfasamento s e pode ser designada por $cov(X_t, X_{t+s}) = \gamma(s)$, e por consequência a matriz de covariâncias teórica, $\Gamma_{\mathbf{X}}$, é dada por

$$\Gamma_{\mathbf{X}} = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \dots & \gamma(m-2) & \gamma(m-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \dots & \dots & \gamma(m-2) \\ \vdots & \gamma(1) & \gamma(0) & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ \gamma(m-2) & \vdots & \vdots & \vdots & \ddots & \gamma(1) \\ \gamma(m-1) & \gamma(m-2) & \dots & \dots & \gamma(1) & \gamma(0) \end{bmatrix}.$$

$\Gamma_{\mathbf{X}}$ tem uma estrutura de matriz de Toeplitz, isto é, todos os elementos ao longo de cada uma das diagonais (paralelas à diagonal principal, inclusive) com desfasamento $s = |i-j|$ são iguais.

De seguida analisemos quais as vantagens e desvantagens da utilização dos estimadores \mathbf{S}_{BK} e \mathbf{S}_{VG} da matriz de covariâncias de \mathbf{X} .

Caso se utilize a matriz de trajetória, com as m colunas centradas, o estimador \mathbf{S}_{BK} estima a matriz de covariâncias usual do processo como na ACP. Contudo em qualquer dos casos (com dados centrados ou não) a matriz de covariâncias resultante não apresenta uma estrutura de matriz de Toeplitz, exigida pela estacionariedade da série, pelo que o estimador tende a fornecer pesos maiores para a parte central da série o que irá produzir maiores desvios das estimativas

das covariâncias, motivos que representam desvantagens segundo alguns autores (Vautard *et al.*, 1992). No entanto, esta primeira desvantagem da aplicação de \mathbf{S}_{BK} para alguns autores pode ser encarada com uma vantagem para outros, pois não é necessário admitir uma estrutura subjacente aos dados e como tal este pode ser usado sem verificação prévia se a série é ou não estacionária.

Segundo Golyandina *et al.* (2001), a decomposição da matriz de Toeplitz pode levar a problemas de otimização uma vez que os principais efeitos da série são descritos pelos valores e vetores próprios dominantes, e mesmo assim os seguintes ainda são por vezes importantes, algo que não é garantido caso a solução não seja ótima.

Ainda de acordo com Golyandina *et al.* (2001), ambos os métodos apresentados fornecem praticamente os mesmos resultados para séries longas e estacionárias. Para séries relativamente pequenas, estacionárias e com ruído, a utilização de uma matriz de Toeplitz na SVD poderá ser vantajosa. Para mais detalhes acerca das semelhanças e diferenças entre os estimadores \mathbf{S}_{BK} e \mathbf{S}_{VG} na aplicação da SVD, veja-se por exemplo Allen e Smith (1996) ou Ghil e Taricco (1997).

Oliveira (2003) apresenta um exemplo sobre a série do número de manchas solares¹, na qual se analisa a influência de três valores distintos do comprimento da janela, na aplicação da SSA e mediante a utilização dos estimadores mencionados anteriormente. Apresentou ainda resultados das reconstruções da série para os seis diferentes casos em análise.

4.6 Pares de Elementos na Decomposição

Relativamente à decomposição espectral da série, também esta pode ser obtida pelo método tradicional de Análise de Fourier (Apêndice C.2) para caraterizar picos espectrais em registos temporais que correspondem a oscilações regulares das funções trigonométricas seno e cosseno.

Para determinados processos mais complexos que o usual, não são esperadas correspondências perfeitas entre o conjunto dos valores singulares (o espectro singular ou simplesmente espectro) e o espectro de potência (*power spectrum*), embora estejam patentes num grande número de casos. Segundo Vautard e Ghil (1989), alguns autores mostraram que a correspondência exata é

¹O número de manchas solares é uma das três séries que constitui o conhecido conjunto de dados de Quenouille de 1952, observados ao longo de 100 anos, entre os anos de 1770 e 1869. Esta série é bem conhecida em astronomia e a sua decomposição em componentes oscilatórias, assim como os períodos dominantes associados estão bem descritos na literatura.

alcançada quando o comprimento da janela m é suficientemente grande. Nesse caso, no limite, o conjunto dos elementos espectrais é constituído por pares de funções trigonométricas (pares e ímpares) que correspondem ao mesmo valor próprio. Cada valor próprio corresponde assim a um valor do espectro de potência multiplicado por um fator de escala cujas funções trigonométricas associadas são oscilações (harmónicas) puras da respetiva frequência, como iremos equacionar mais adiante.

Várias são as referências que se encontram para consulta da teoria da Análise de Fourier e da Análise Espectral tais como Murteira *et al.* (1993) ou von Storch e Zwiers (1999) e também alguns resultados que se resumem nos Apêndices C.2 e C.3.

Consideremos a constatação seguinte.

Se X_t é uma série temporal que tem uma oscilação dominante, na análise do seu espectro, então na decomposição espectral quando se destaca um par de valores próprios λ_1 e λ_2 consecutivos tais que $\lambda_1 \simeq \lambda_2$, os vetores próprios U^1 e U^2 e as CPs V^1 e V^2 associados também surgem relacionados entre si aos pares para um mesmo período T e diz-se que ambos estão em quadratura (Vautard e Ghil, 1989; Vautard *et al.*, 1992).

Para ilustrar esta característica consideremos algumas noções sobre a teoria de ondas, onde se admite a existência de sinais periódicos.

Definição 4.6.1 *Designa-se por onda a toda a perturbação que se repete em intervalos de tempo regulares e que ocorre num determinado meio ou no espaço.*

Definição 4.6.2 *Num sinal periódico, o comprimento de onda λ é definido como a distância entre valores repetidos consecutivos num padrão de onda, ou seja, é a distância descrita no espaço por um ciclo.*

Definição 4.6.3 *Uma onda progressiva² no intervalo finito $-a \leq y \leq a$, para um instante de tempo t , é definida por*

$$W(y, t) = A \cos(\mu y - \omega t), \quad (4.6)$$

²Em Meteorologia, uma onda progressiva é uma onda ou perturbação ondulatória que se move em relação à superfície da Terra. Em Física, é uma onda que transfere energia de um meio para outro.

Um exemplo simples, para observar uma onda progressiva, consiste em realizar a experiência de deixar cair à superfície de um líquido um objeto que vai originar uma série de cristas e vales no líquido, as quais se deslocam ao longo da superfície do mesmo.

onde A é a amplitude da onda, μ é o número de vezes que ocorre o comprimento de onda por unidade de distância (frequência espacial) e ω é a frequência angular (Plaut e Vautard, 1994).

Na equação de onda (4.6), para um comprimento de onda λ , vem $\mu = 2\pi/\lambda$ e para um intervalo de tempo T , a frequência angular é descrita por $\omega = 2\pi/T$.

Portanto a equação (4.6) pode ser reescrita por

$$W(y, t) = A \cos \left(2\pi \left(\frac{y}{\lambda} - \frac{t}{T} \right) \right).$$

Consideremos que o sinal é discretizado no tempo por $t_i = i\Delta t$ e no espaço por $y_s = s\Delta y$, ($s = 1, \dots, p$).

Com base nas propriedades trigonométricas, para uma translação h em relação ao tempo t , a equação de onda (4.6) vem

$$\begin{aligned} W(y, t + h) &= A \cos(\mu y - \omega(t + h)) \\ &= A \cos(\omega t) \cos(\mu y - \omega h) + A \sin(\omega t) \sin(\mu y - \omega h). \end{aligned} \quad (4.7)$$

Ao se adicionar uma fase inicial ϕ , a equação de onda (4.6) é escrita de um modo geral por

$$W(y, t) = A \cos(\mu y - \omega t + \phi),$$

e (4.7) passa a ser expressa por

$$W(y, t + h) = A \cos(\omega t + \phi) \cos(\mu y - \omega h + \phi) + A \sin(\omega t + \phi) \sin(\mu y - \omega h + \phi).$$

De acordo com as notações introduzidas na Subsecção 4.2.1 para a matriz dos desfasamentos dos dados, \mathbf{X} , cada elemento $X_{s,i}$ (para uma linha s e uma coluna i) pode ser descrito à custa das componentes principais V^l e dos respectivos vetores próprios U^l , de acordo com a expressão

$$X_{s,i} = \sum_{l=1}^m V_i^l \left(U_{s,j}^l \right)^T = W(y_s, t_i), \quad (4.8)$$

que é satisfeita quando se consideram as seguintes equações para duas CPs e para $j = 1, 2, \dots, m$

(Plaut e Vautard, 1994).

$$\begin{aligned}
 U_{s,j}^1 &= F_1 \cos(\mu s \Delta y - \omega j \Delta t + \phi), \\
 U_{s,j}^2 &= F_2 \sin(\mu s \Delta y - \omega j \Delta t + \phi), \\
 V_i^1 &= \frac{A}{F_1} \cos(\omega i \Delta t + \phi), \\
 V_i^2 &= \frac{A}{F_2} \sin(\omega i \Delta t + \phi), \\
 \phi &= \frac{2\mu \Delta y - \omega(m+1)\Delta t}{2}.
 \end{aligned}$$

Impondo que os vetores próprios U^1 e U^2 sejam unitários, calculam-se as constantes F_1 e F_2 que dependem dos parâmetros m , Δy e Δt . A fase ϕ é regulada pela condição de ortogonalidade. Como tal, estes dois vetores próprios são ortogonais em fase e em quadratura.

Assim para um desfasamento nulo, a covariância entre as duas CPs anula-se, fazendo com que estas fiquem também em fase e em quadratura (consultar Apêndices C.2 e C.3). O espectro fica essencialmente restringido aos dois primeiros valores próprios

$$\lambda_1 = \frac{A^2}{2F_1^2} \text{ e } \lambda_2 = \frac{A^2}{2F_2^2},$$

os quais apesar de não serem iguais são muito próximos para valores elevados do comprimento da janela m , e consequentemente

$$\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \xrightarrow{m \rightarrow +\infty} 0.$$

Ocorrem resultados semelhantes quando o sinal X_t é uma componente oscilatória mais genérica, da forma

$$W(y, t) = A(y) \cos(\omega t) + B(y) \sin(\omega t).$$

Ao generalizar para q ($\leq m$) componentes oscilatórias, de sinal quase periódico, cujo sinal X_t é definido pela sua soma, o número de valores próprios que não são aproximadamente nulos é $2q$, e os vetores próprios associados surgem aos pares, desde que o comprimento da janela seja suficientemente grande para permitir a separação dos pares.

Contudo para uma oscilação pura, os dois valores próprios associados não são exatamente iguais e por isso em termos práticos poderá ser difícil identificar o emparelhamento oscilatório. Vautard *et al.* (1992) apresentaram três critérios para reconhecimento de pares de elementos oscilatórios. Um é o critério de Ghil e Mo para a significância da quadratura entre as CPs de um

mesmo par com base na correlação desfasada, mas que é considerada uma abordagem com falta de fiabilidade na significância estatística dos estimadores da correlação desfasada. Como tal, foram propostos outros dois critérios baseados nas propriedades espectrais dos vetores próprios. Vautard *et al.* (1992) ainda apresentam pormenores e alguns exemplos de aplicação destes dois critérios.

Quanto ao problema de separação de valores próprios têm-se em conta várias regras. Uma é a verificação de quebras ao longo do espectro ao estabelecer que o ruído puro produz uma sequência de valores próprios que decresce lentamente. No caso de uma série temporal no sentido lato, com ruído e sinal, espera-se que alguns valores próprios sejam mais elevados, onde uma quebra no espectro pode distinguir nitidamente os valores próprios associados ao sinal e quais os associados a uma estrutura de ruído. Contudo convém ter a noção de que não existe nenhum procedimento formal que permita determinar tal quebra. Para além disso, em situações de séries complexas e de grande ruído, a parte do sinal poderá estar misturado com o ruído no que diz respeito à ordem do seus valores próprios, de tal forma que se torna impossível encontrar uma quebra que seja notória. Noutras situações alguns pares de valores próprios ainda podem ser extraídos da parte do ruído desde que se aumente o tamanho da janela, o que, em termos de representação gráfica dos valores próprios, faz com que nessa parte o espectro tenha tendência a ficar cada vez mais achatado.

Embora os valores próprios forneçam informação para a identificação e agrupamento das componentes oscilatórias, a estrutura dos vetores próprios também é fundamental para ajudar a complementar o estudo, e por isso passamos a analisar de seguida os pares de vetores próprios.

Quando os dois valores próprios de uma harmónica são geralmente muito próximos um do outro, torna-se fácil identificar as componentes harmónicas através de uma simples inspeção visual num gráfico de valores próprios (*scree plot*). Por outro lado, uma análise cuidada dos gráficos constituídos por pares de vetores próprios é considerada como uma ferramenta que permite distinguir visualmente os pares que correspondem às componentes harmónicas da série temporal.

Consideremos o caso particular de uma harmónica pura com fase e amplitude constantes e com um período inteiro $T = 1/\omega$ que é um divisor inteiro do comprimento da janela m e do número de vetores desfasados k . Nesta situação, os vetores próprios e as componentes principais têm a forma de sequências de senos e cossenos com o mesmo período T e a mesma

fase, de tal modo que a identificação das componentes correspondentes a harmónicas resume-se à determinação dos respetivos pares (exceto para o caso em que $\omega = 0.5$, que corresponde apenas a um valor próprio). Desta forma em Vautard e Ghil (1989), os vetores próprios são descritos pelas funções trigonométricas pares e ímpares que variam ao longo do tempo t ,

$$\rho_1(t) = \alpha_1 \cos(2\pi\omega t) \quad \text{e} \quad \rho_2(t) = \alpha_2 \sin(2\pi\omega t), \quad (4.9)$$

onde α_1 e α_2 são constantes apropriadas. Os correspondentes valores próprios são dados por

$$\lambda_1 = \frac{\sigma^2}{2} \left(1 + \frac{\sin(2\pi\omega m)}{2\pi\omega m} \right) \quad \text{e} \quad \lambda_2 = \frac{\sigma^2}{2} \left(1 - \frac{\sin(2\pi\omega m)}{2\pi\omega m} \right),$$

onde σ^2 é a variância total da série temporal, ou seja, é a soma de todos os valores próprios.

Sabe-se que a representação de um seno contra um cosseno, (\sin, \cos) , com iguais frequências, amplitudes e fases dão origem a um gráfico cujos pontos se situam sobre um círculo. No caso em que o período $T = 1/\omega$ é um número inteiro, os pontos são os vértices de um polígono regular com T vértices. Para o caso em que a frequência é dada por um número racional $\omega = p/q < 0.5$ com p e q primos inteiros, os pontos são os vértices de um polígono regular com q vértices. A Figura 4.1 ilustra os gráficos para os 9 pares de sequências de senos e cossenos (situados no círculo unitário) com fases zero, com a mesma amplitude e com períodos 12, 8, 6, 5, 4, 3, 2.5, 2.4 e 1.6 respetivamente.

A construção geral destes gráficos de acordo com (4.9), tendo por base o círculo unitário com $\alpha_1 = \alpha_2 = 1$, é efetuada a partir dos pontos do tipo

$$(\cos(2\pi\omega t), \sin(2\pi\omega t)) = \left(\cos \frac{2\pi t}{T}, \sin \frac{2\pi t}{T} \right).$$

Ao unir os pontos através de segmentos de reta, quando o período T é inteiro obtemos polígonos regulares para os instantes $t = 0, 1, 2, \dots, T - 1$. Quando o período não é um número inteiro obtemos estrelas poligonais regulares.

Por exemplo, a construção de um hexágono resulta da união através de segmentos de reta entre os pontos $(1, 0)$; $(\cos \frac{2\pi}{6}, \sin \frac{2\pi}{6})$; $(\cos \frac{4\pi}{6}, \sin \frac{4\pi}{6})$; $(\cos \frac{6\pi}{6}, \sin \frac{6\pi}{6})$; $(\cos \frac{8\pi}{6}, \sin \frac{8\pi}{6})$; $(\cos \frac{10\pi}{6}, \sin \frac{10\pi}{6})$ e $(1, 0)$ pela respetiva ordem.

Pequenos desvios em relação à situação teórica anterior leva a que os pontos nos gráficos não sejam exatamente os vértices dos polígonos regulares, embora venham a situar-se relativamente próximo destes. Na Secção 4.10 será apresentado um exemplo de aplicação onde poderemos constatar este efeito nos pares de CPs correspondentes a harmónicas de diversos períodos.

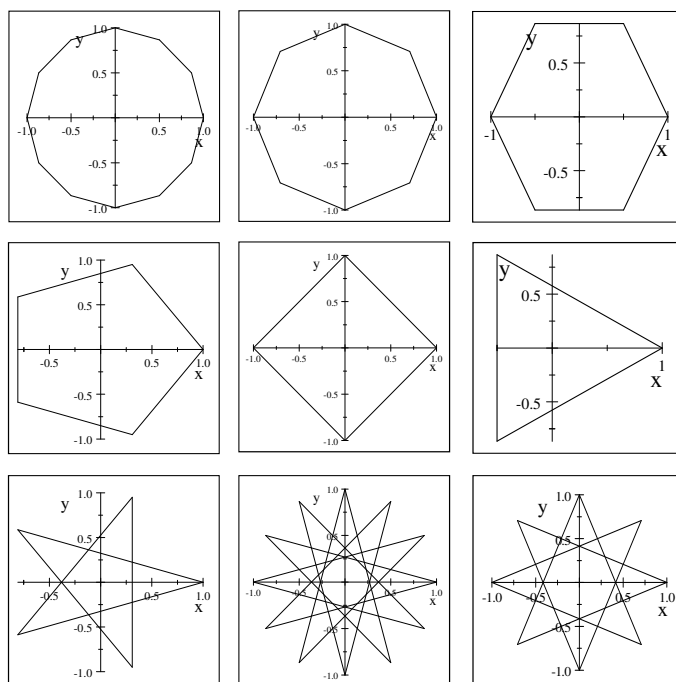


Figura 4.1: Nove pares de seqüências de senos e cossenos no círculo unitário

Como referências para estas construções gráficas consideramos Golyandina *et al.* (2001) e Hassani (2007), os quais descrevem exemplos de representações geométricas dos vetores próprios e das componentes principais como meio auxiliar na identificação dos períodos das oscilações mais relevantes presentes nos seus dados em análise.

4.7 Séries Reconstruídas/Filtradas

Após a etapa da decomposição segue-se frequentemente a reconstrução da série original.

De forma análoga ao que se passa na ACP, os vetores próprios obtidos na SSA são ortogonais entre si, pelo que as CPs representam a série temporal aumentada num novo sistema de coordenadas, em que cada uma das m CPs é definida como a projeção ortogonal da série original no respetivo vetor próprio.

Convém relembrar que, na SSA, as m CPs são de tamanho k e que numa primeira fase obtemos a matriz dos desfasamentos reconstruída, de dimensões $k \times m$, como descrita na Subsecção 4.2.2. Só posteriormente o processo de aplicação da média ao longo das diagonais nessa ma-

triz permite obter a série original reconstruída de comprimento n . Quando utilizarmos apenas um subconjunto próprio de q CPs ($q < m$) iremos obter reconstruções \tilde{X}_t consideradas versões filtradas da série original.

Tendo em conta a expressão (4.4), as reconstruções da série à custa de q CPs conforme descrito em Golyandina *et al.* (2001) pode ser comparado com o método de filtragem de Wiener que também fornece filtros ótimos para remover o ruído de um sinal. Segundo Vautard *et al.* (1992) as principais diferenças entre os métodos prendem-se com o facto do filtro de Wiener usar funções harmónicas e não ser totalmente adaptável por ter que se estabelecer a forma da densidade espectral filtrada. Para além disso, nesse método é impossível obter estimativas do espectro de potência para conjuntos de dados pequenos.

Relativamente à escolha do valor de q , este depende do comprimento da janela m previamente estabelecido, e como tal, por não ser consensual constitui um problema na decisão do número de componentes a reter, o qual foi abordado em Vautard e Ghil (1989) e desenvolvido em Vautard *et al.* (1992). Uma solução possível é reter todas as CPs associadas aos maiores valores próprios, de acordo com a descrição anterior na Secção 4.6 e acima de um determinado valor de quebra. Desta forma, admite-se que as q primeiras CPs passam a reter a informação dominante referente à estrutura subjacente dos dados, enquanto que as restantes $m - q$ CPs representam a parte que está relacionada com a estrutura de ruído.

Mesmo que surjam dificuldades em seleccionar m e q , a SSA é considerada uma técnica poderosa na reconstrução de dados mesmo que afetados por ruído significativo. Portanto, as primeiras CPs são usadas para remover o ruído intrínseco dos dados, aumentando assim a precisão na identificação dos picos espectrais.

No caso em que se utilizam todas as CPs na reconstrução, a soma dos espectros de potência de cada CP é igual ao espectro de potência da série original. Quando as CPs são usadas para definir o sinal e possuem diferentes picos espectrais de potência, então permitem isolar diferentes componentes oscilatórias. Para o caso particular dos pares de valores próprios quase iguais corresponderá um par de componentes oscilatórias com espectros idênticos mas que variam segundo uma forma desfasada no tempo, cuja periodicidade dominante irá estar relacionada com a oscilação que explica o sinal.

4.8 Centragem em SSA

Embora o modelo básico de SSA (Golyandina *et al.*, 2001) não utilize qualquer tipo de centragem aplicado à matriz de trajetória \mathbf{X} segundo os vetores desfasados, pode ser vantajoso efetuar, por exemplo, uma centragem simples (por colunas ou por linhas), caso a série original X_t possa ser representada pela soma de duas parcelas: $X_t = X_t^C + X_t^0$, onde X_t^C é uma série constante e X_t^0 é uma outra série que oscila em torno de zero. Se o tamanho da série for suficientemente grande, as suas componentes serão certamente extraídas através do modelo básico de SSA, mas para séries de pequena dimensão, a aplicação da centragem simples poderá funcionar melhor (Golyandina *et al.*, 2001).

Por outro lado, também é possível efetuar uma centragem dupla à matriz trajetória. Segundo Golyandina *et al.* (2001), geralmente a centragem dupla permite extrair aproximadamente a componente da tendência linear de uma série caso esta oscile relativamente perto da referida componente e, como tal, é preferível muitas das vezes a centragem dupla em relação à não centragem.

Neste contexto, a centragem da matriz de trajetória leva a que a nova matriz centrada deixe de ser uma matriz de Hankel. Em qualquer um dos casos mencionados anteriormente, passar-se-á a efetuar uma decomposição em valores singulares sobre a matriz de trajetória previamente centrada.

Para além da descrição das centragens simples e dupla, iremos ainda introduzir uma nova abordagem teórica onde se discute a centragem prévia da série original em detrimento do uso de qualquer um dos outros tipos de centragem aplicado na matriz trajetória.

4.8.1 Centragem Simples por Colunas

A centragem simples de uma matriz é usualmente definida por colunas, mas tal facto não invalida que a centragem não possa também ser efetuada ao longo das linhas. De seguida descrevemos a centragem simples por colunas e na Secção seguinte a centragem simples por linhas.

Tendo em conta o comprimento da janela m , e $k = n - m + 1$ vetores desfasados de dimensão m , consideremos a matriz (não centrada) de trajetória \mathbf{X} da série temporal X_t , constituída por k linhas e m colunas.

Definição 4.8.1 A transformação da matriz de trajetória \mathbf{X} numa nova matriz centrada por colunas é dada por $\mathbf{X}^{cc} = \mathbf{X} - \overline{\mathbf{X}}_c$, onde a matriz $\overline{\mathbf{X}}_c$ do tipo $k \times m$ é dada por

$$\overline{\mathbf{X}}_c = \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \quad (4.10)$$

e o vetor $\mathbf{1}_k = (1, \dots, 1)^T \in \mathbb{R}^k$ representa o vetor coluna com elementos todos iguais à unidade.

Portanto, $\overline{\mathbf{X}}_c$ é a matriz em que todos os elementos de cada coluna são iguais à média dos valores da coluna de \mathbf{X} correspondente.

Definição 4.8.2 O vetor das médias por coluna, $\overline{\mathbf{x}}_c \in \mathbb{R}^m$, é definido por

$$\overline{\mathbf{x}}_c = \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \quad (4.11)$$

onde cada elemento representa o resultado da média dos valores de cada coluna de \mathbf{X} .

Nota 4.8.3 A partir de (4.10) vem

$$\overline{\mathbf{X}}_c = \mathbf{1}_k \overline{\mathbf{x}}_c^T.$$

De forma análoga aos passos descritos nas Subsecções 4.2.1 e 4.2.2, constrói-se a matriz $\mathbf{S}^{cc} = (\mathbf{X}^{cc})^T \mathbf{X}^{cc}$ e efetua-se a SVD da matriz \mathbf{X}^{cc} para obter os respectivos m valores e vetores próprios de \mathbf{S}^{cc} .

Se representarmos por λ_i^{cc} os valores próprios, por U_i^{cc} os correspondentes vetores próprios ortonormados de \mathbf{S}^{cc} , e por d^{cc} o número de valores próprios não nulos de \mathbf{S}^{cc} , então os vetores V_i^{cc} , ou seja, as CPs são dadas por

$$V_i^{cc} = \mathbf{X}^{cc} U_i^{cc} / \sqrt{\lambda_i^{cc}}, \text{ com } i = 1, \dots, d^{cc} \text{ e } d^{cc} = \text{car}(\mathbf{X}^{cc}) \leq m, \quad (4.12)$$

e a matriz \mathbf{X} passa a ser reescrita por (Golyandina *et al.*, 2001)

$$\mathbf{X} = \overline{\mathbf{X}}_c + \sum_{i=1}^{d^{cc}} \mathbf{X}_i^{cc} = \overline{\mathbf{X}}_c + \sum_{i=1}^{d^{cc}} \sqrt{\lambda_i^{cc}} V_i^{cc} (U_i^{cc})^T.$$

De acordo com esta decomposição, vem ainda que

$$\mathbf{X} = \sqrt{\lambda_{0(c)}} V_{0(c)} U_{0(c)}^T + \sum_{i=1}^{d^{cc}} \sqrt{\lambda_i^{cc}} V_i^{cc} (U_i^{cc})^T = \overline{\mathbf{X}}_c + \sum_{i=1}^{d^{cc}} \mathbf{X}_i, \quad (4.13)$$

com

$$V_{0(c)} = \mathbf{1}_k / \sqrt{k}, \quad U_{0(c)} = \overline{\mathbf{x}}_c / \|\overline{\mathbf{x}}_c\| \quad \text{e} \quad \sqrt{\lambda_{0(c)}} = \|\overline{\mathbf{x}}_c\| \sqrt{k}.$$

A expressão (4.13) representa uma decomposição da matriz \mathbf{X} na soma de matrizes elementares e é designada por SVD com centragem simples, o que corresponde ao procedimento habitual de centragem da ACP.

Corolário 4.8.4 *Dada a decomposição (4.13) e tendo em conta a existência de ortogonalidade entre as matrizes intervenientes, verifica-se a seguinte igualdade*

$$\sum_{i=1}^d \lambda_i = \lambda_{0(c)} + \sum_{i=1}^{d^{cc}} \lambda_i^{cc}, \quad (4.14)$$

ou seja, a soma dos d valores próprios da SVD sem centragem é igual à soma do valor $\lambda_{0(c)}$ com os d^{cc} valores próprios da SVD com centragem simples por colunas.

Demonstração:

Para operações envolvendo matrizes, com base na circularidade e linearidade do traço, vem

$$\begin{aligned} \text{tr} \left((\mathbf{X}^{cc})^T \mathbf{X}^{cc} \right) &= \text{tr} \left((\mathbf{X} - \bar{\mathbf{X}}_c)^T (\mathbf{X} - \bar{\mathbf{X}}_c) \right) \\ &= \text{tr} \left(\left(\mathbf{X} - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right)^T \left(\mathbf{X} - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right) \right), \text{ por (4.10)} \\ &= \text{tr} \left(\left(\mathbf{X}^T - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \right) \left(\mathbf{X} - \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right) \right) \\ &= \text{tr} \left(\mathbf{X}^T \mathbf{X} - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} + \frac{1}{k^2} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right) \\ &= \text{tr} \left(\mathbf{X}^T \mathbf{X} - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} + \frac{1}{k^2} k \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right) \\ &= \text{tr} \left(\mathbf{X}^T \mathbf{X} - \frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right) \\ &= \text{tr}(\mathbf{X}^T \mathbf{X}) - \text{tr} \left(\frac{1}{k} \mathbf{X}^T \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} \right), \text{ pela linearidade do traço} \\ &= \text{tr}(\mathbf{X}^T \mathbf{X}) - \text{tr}(\bar{\mathbf{x}}_c k \bar{\mathbf{x}}_c^T), \text{ por (4.11).} \end{aligned}$$

Portanto, conclui-se que

$$\sum_{i=1}^{d^{cc}} \lambda_i^{cc} = \sum_{i=1}^d \lambda_i - k \text{tr}(\bar{\mathbf{x}}_c \bar{\mathbf{x}}_c^T) = \sum_{i=1}^d \lambda_i - k \|\bar{\mathbf{x}}_c\|^2,$$

ou seja,

$$\sum_{i=1}^d \lambda_i = \lambda_{0(c)} + \sum_{i=1}^{d^{cc}} \lambda_i^{cc}.$$

□

Algumas das relações anteriores foram analisadas e exploradas em Cadima e Jolliffe (2009), onde se apresenta um conjunto de resultados comparativos entre valores próprios, vetores próprios e componentes principais, entre a ACP sem centragem e a ACP com centragem por colunas.

Paralelismo com a ACP

A ACP utiliza habitualmente a centragem simples (nomeadamente por colunas) descrita anteriormente.

De acordo com o que foi referido na Secção 4.5, a diferença entre a SVD da matriz \mathbf{S}^{cc} na SSA com centragem simples e a SVD da matriz de variâncias-covariâncias da ACP prende-se apenas com a diferença do tamanho dos correspondentes valores próprios, onde estes são k vezes maiores para \mathbf{S}^{cc} . Por outro lado, os vetores próprios mantêm-se invariantes (a menos do sinal) entre os dois modelos.

Desta forma, podemos descrever a relação existente entre as CPs da SSA com centragem simples e as CPs da ACP.

A partir da expressão (4.12), vem que

$$\mathbf{X}^{cc}U_i^{cc} = \sqrt{\lambda_i^{cc}}V_i^{cc}.$$

No Capítulo 2 da ACP, cada CP \mathbf{z}_i associada ao vetor próprio \mathbf{p}_i é descrita por $\mathbf{z}_i = \mathbf{X}\mathbf{p}_i$, onde \mathbf{X} deverá ser encarada como \mathbf{X}^{cc} , uma vez que a ACP usa a centragem por colunas da matriz inicial, como medida de pré-processamento dos dados. Adaptando a notação, e denotando cada vetor próprio como U_i^* e cada valor próprio como λ_i^* para a ACP, vem

$$\mathbf{z}_i = \mathbf{X}U_i^* = \mathbf{X}^{cc}U_i^{cc} = \sqrt{\lambda_i^{cc}}V_i^{cc}.$$

Uma vez que $\lambda_i^{cc} = k\lambda_i^*$, vem então que

$$\mathbf{z}_i = \sqrt{\lambda_i^{cc}}V_i^{cc} = \sqrt{k\lambda_i^*}V_i^{cc}.$$

Portanto as CPs da ACP são descritas à custa das CPs da SSA com centragem simples (por colunas) através de $\mathbf{z}_i = \sqrt{k}\sqrt{\lambda_i^*}V_i^{cc}$.

4.8.2 Centragem Simples por Linhas

Por analogia ao procedimento descrito anteriormente para a centragem por colunas da matriz de trajetória \mathbf{X} , a centragem por linhas também pode ser implementada.

Definição 4.8.5 A matriz centrada por linhas passa a ser designada por $\mathbf{X}^{cl} = \mathbf{X} - \overline{\mathbf{X}}_l$, onde a matriz $\overline{\mathbf{X}}_l$ do tipo $k \times m$ é dada por

$$\overline{\mathbf{X}}_l = \frac{1}{m} \mathbf{X} \mathbf{1}_m \mathbf{1}_m^T \quad (4.15)$$

com $\mathbf{1}_m = (1, \dots, 1)^T \in \mathbb{R}^m$.

Desta forma $\overline{\mathbf{X}}_l$ é a matriz cujos elementos de cada linha são todos iguais à média dos valores da linha de \mathbf{X} correspondente.

Definição 4.8.6 O vetor das médias por linha, $\overline{\mathbf{x}}_l \in \mathbb{R}^k$, é dado por

$$\overline{\mathbf{x}}_l = \frac{1}{m} \mathbf{X} \mathbf{1}_m$$

onde cada elemento representa o resultado da média dos valores de cada linha de \mathbf{X} .

Nota 4.8.7 A partir de (4.15) vem

$$\overline{\mathbf{X}}_l = \overline{\mathbf{x}}_l \mathbf{1}_m^T.$$

A SVD é efetuada a partir da nova matriz $\mathbf{S}^{cl} = (\mathbf{X}^{cl})^T \mathbf{X}^{cl}$ com vista a obterem-se os respetivos m valores e vetores próprios de \mathbf{S}^{cl} .

Atribuindo uma notação análoga em relação à centragem por colunas para os valores próprios, vetores próprios e para as CPs, a matriz \mathbf{X} é dada por (Golyandina *et al.*, 2001)

$$\mathbf{X} = \overline{\mathbf{X}}_l + \sum_{i=1}^{d^{cl}} \mathbf{X}_i^{cl} = \overline{\mathbf{X}}_l + \sum_{i=1}^{d^{cl}} \sqrt{\lambda_i^{cl}} V_i^{cl} (U_i^{cl})^T$$

e a decomposição (4.13) é substituída por

$$\mathbf{X} = \sqrt{\lambda_{0(l)}} V_{0(l)} U_{0(l)}^T + \sum_{i=1}^{d^{cl}} \sqrt{\lambda_i^{cl}} V_i^{cl} (U_i^{cl})^T = \overline{\mathbf{X}}_l + \sum_{i=1}^{d^{cl}} \mathbf{X}_i,$$

onde

$$V_{0(l)} = \overline{\mathbf{x}}_l / \|\overline{\mathbf{x}}_l\|, \quad U_{0(l)} = \mathbf{1}_m / \sqrt{m} \quad \text{e} \quad \sqrt{\lambda_{0(l)}} = \|\overline{\mathbf{x}}_l\| \sqrt{m}.$$

Corolário 4.8.8 Por analogia a (4.14) do corolário anterior, verifica-se a igualdade

$$\sum_{i=1}^d \lambda_i = \lambda_{0(l)} + \sum_{i=1}^{d^{cl}} \lambda_i^{cl}, \quad (4.16)$$

cuja soma dos d valores próprios da SVD sem centragem é igual à soma do valor $\lambda_{0(l)}$ com os d^{cl} valores próprios da SVD com centragem por linhas.

A demonstração é similar à que foi efetuada para o corolário anterior.

4.8.3 Centragem Dupla

A centragem dupla é efetuada em simultâneo ao longo das linhas e das colunas da matriz de trajetória \mathbf{X} . Esta centragem pode ser encarada como uma centragem simples por colunas seguida de uma centragem simples por linhas ou vice-versa.

Para o comprimento da janela m , e para k vetores desfasados de dimensão m , consideremos a matriz de trajetória \mathbf{X} da série temporal X_t , constituída por k linhas e m colunas.

Definição 4.8.9 *Ao aplicarmos a centragem simples por linhas na matriz centrada por colunas $\mathbf{X}^{cc} = \mathbf{X} - \overline{\mathbf{X}}_c$, passamos a considerar a matriz centrada por colunas e por linhas definida por*

$$\mathbf{X}^{**} = \mathbf{X}^{cc} - \overline{\mathbf{X}^{cc}}_l, \text{ com } \overline{\mathbf{X}^{cc}}_l = \overline{\mathbf{x}^{cc}}_l \mathbf{1}_m^T,$$

e onde $\overline{\mathbf{x}^{cc}}_l$ representa o vetor das médias por linha da matriz previamente centrada por colunas \mathbf{X}^{cc} .

Para além desta definição da matriz \mathbf{X}^{**} , iremos apresentar uma outra como alternativa com base em Golyandina *et al.* (2001).

Definição 4.8.10 (Alternativa) *A matriz \mathbf{X} pode ser transformada numa nova matriz centrada por colunas e por linhas passando a ser denotada por $\mathbf{X}^{**} = \mathbf{X} - \overline{\mathbf{X}}^{(cl)}$, cuja matriz $\overline{\mathbf{X}}^{(cl)}$ do tipo $k \times m$ admite a seguinte expressão*

$$\overline{\mathbf{X}}^{(cl)} = \overline{\mathbf{X}}_c + \overline{\mathbf{X}}_l - \overline{\mathbf{X}}_{cl},$$

onde $\overline{\mathbf{X}}_{cl}$ é a matriz do tipo $k \times m$ em que cada elemento é igual ao valor médio de todos os elementos da matriz \mathbf{X} .

Designando a média de todos os elementos de \mathbf{X} por $\mathbf{a} = \frac{1}{km} \mathbf{1}_k^T \mathbf{X} \mathbf{1}_m$, a matriz $\overline{\mathbf{X}}_{cl}$ é igual a

$$\overline{\mathbf{X}}_{cl} = \mathbf{a} \mathbf{1}_k \mathbf{1}_m^T.$$

Na centragem dupla, a SVD é efetuada a partir da matriz $\mathbf{S}^{**} = (\mathbf{X}^{**})^T \mathbf{X}^{**}$ com o objetivo de se obterem os respetivos m valores e vetores próprios de \mathbf{S}^{**} .

Ao denotarmos por λ_i^{**} os valores próprios, por U_i^{**} os correspondentes vetores próprios ortonormados de \mathbf{S}^{**} , e por d^{**} o número de valores próprios não nulos de \mathbf{S}^{**} , então as CPs são dadas por

$$V_i^{**} = \mathbf{X}^{**} U_i^{**} / \sqrt{\lambda_i^{**}}, \text{ com } i = 1, \dots, d^{**} \text{ e } d^{**} = \text{car}(\mathbf{X}^{**}) \leq m,$$

e a matriz \mathbf{X} é decomposta por

$$\mathbf{X} = \overline{\mathbf{X}}^{(cl)} + \sum_{i=1}^{d^{**}} \mathbf{X}_i^{**} = \overline{\mathbf{X}}_c + \overline{\mathbf{X}}^{cc}_l + \sum_{i=1}^{d^{**}} \sqrt{\lambda_i^{**}} V_i^{**} (U_i^{**})^T,$$

a qual ainda pode ser descrita por

$$\mathbf{X} = \sqrt{\lambda_{0(c)}} V_{0(c)} U_{0(c)}^T + \sqrt{\lambda_{0(l)}^{(c)}} V_{0(l)}^{(c)} \left(U_{0(l)}^{(c)} \right)^T + \sum_{i=1}^{d^{**}} \sqrt{\lambda_i^{**}} V_i^{**} (U_i^{**})^T = \overline{\mathbf{X}}_c + \overline{\mathbf{X}}^{cc}_l + \sum_{i=1}^{d^{**}} \mathbf{X}_i, \quad (4.17)$$

onde

$$V_{0(c)} = \mathbf{1}_k / \sqrt{k}, \quad U_{0(c)} = \overline{\mathbf{x}}_c / \|\overline{\mathbf{x}}_c\| \quad \text{e} \quad \sqrt{\lambda_{0(c)}} = \|\overline{\mathbf{x}}_c\| \sqrt{k},$$

assim como

$$V_{0(l)}^{(c)} = \overline{\mathbf{x}}^{cc}_l / \|\overline{\mathbf{x}}^{cc}_l\|, \quad U_{0(l)}^{(c)} = \mathbf{1}_m / \sqrt{m} \quad \text{e} \quad \sqrt{\lambda_{0(l)}^{(c)}} = \|\overline{\mathbf{x}}^{cc}_l\| \sqrt{m}.$$

Dada a decomposição (4.17), e de uma forma semelhante aos Corolários (4.8.4) e (4.8.8) da centragem simples, verifica-se a seguinte igualdade

$$\sum_{i=1}^d \lambda_i = \lambda_{0(c)} + \lambda_{0(l)}^{(c)} + \sum_{i=1}^{d^{**}} \lambda_i^{**}, \quad (4.18)$$

ou seja, a soma dos d valores próprios da SVD sem centragem é igual à soma do valor $\lambda_{0(c)}$ e do valor $\lambda_{0(l)}^{(c)}$ com os d^{**} valores próprios da SVD com centragem dupla.

4.8.4 Centragem na Série Original

A centragem na SSA também pode ser estabelecida como um pré-processamento dos dados aplicado na série original, à qual se subtrai a média da série antes de se construir a matriz de trajetória. Passemos a estabelecer novos aspetos teóricos com base nas notações das centragens descritas anteriormente.

Consideremos a série temporal X_t de tamanho n , e a sua média $\bar{x} = \frac{1}{n} \mathbf{1}_n^T X_t$.

Definição 4.8.11 *A série original centrada é dada por $X_t - \bar{\mathbf{x}}$, onde*

$$\bar{\mathbf{x}} = \bar{x} \mathbf{1}_n$$

representa o vetor em que cada um dos n elementos é a média da série original e onde o vetor $\mathbf{1}_n$ é dado por $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$.

Posteriormente constrói-se a matriz de trajetória \mathbf{X}' , constituída por k linhas e m colunas. No entanto, trabalhar com esta matriz \mathbf{X}' é o mesmo que obter a transformação a partir da matriz de trajetória \mathbf{X} do modelo de SSA básico dada por

$$\mathbf{X}' = \mathbf{X} - \bar{\mathbf{X}},$$

onde $\bar{\mathbf{X}} = \bar{x} \mathbf{1}_k \mathbf{1}_m^T$ representa a matriz do tipo $k \times m$ cujos elementos são todos iguais à média da série original, \bar{x} .

Deste modo, podemos considerar esta transformação como um caso semelhante ao da centragem simples por colunas, com $\bar{\mathbf{X}}$ a desempenhar um papel similar ao de $\bar{\mathbf{X}}_c$ em (4.10) onde em vez de se subtrair a média de cada coluna de \mathbf{X} aos elementos da respetiva coluna, subtrai-se a todos os elementos de cada coluna o mesmo valor correspondente à média \bar{x} da série. Uma diferença é que esta centragem na série original leva a que \mathbf{X}' seja uma matriz de Hankel, o que não acontece nos outros tipos de centragem abordados anteriormente.

Recordemos a partir da SSA com centragem, por exemplo, consideremos a centragem simples por colunas em que $\mathbf{X}^{cc} = \mathbf{X} - \bar{\mathbf{X}}_c$, e que $\bar{\mathbf{X}}_c = \frac{1}{k} \mathbf{1}_k \mathbf{1}_k^T \mathbf{X} = \mathbf{1}_k \bar{\mathbf{x}}_c^T$, onde $\bar{\mathbf{x}}_c$ é o vetor cujos elementos são as médias dos valores de cada coluna de \mathbf{X} . De forma análoga podemos criar o vetor

$$\bar{\mathbf{x}}_{dif} = \bar{\mathbf{x}}_c - \bar{x} \mathbf{1}_m$$

como sendo o vetor em que cada elemento é dado pela diferença entre a média dos valores de cada coluna de \mathbf{X} e a média \bar{x} da série.

Portanto, denotando a matriz da diferença $\mathbf{X}_{dif} = \bar{\mathbf{X}}_c - \bar{\mathbf{X}}$, vem

$$\mathbf{X}_{dif} = \mathbf{1}_k \bar{\mathbf{x}}_c^T - \bar{x} \mathbf{1}_k \mathbf{1}_m^T = \mathbf{1}_k (\bar{\mathbf{x}}_c^T - \bar{x} \mathbf{1}_m^T) = \mathbf{1}_k (\bar{\mathbf{x}}_c - \bar{x} \mathbf{1}_m)^T = \mathbf{1}_k \bar{\mathbf{x}}_{dif}^T.$$

Por consequência, a matriz \mathbf{X}' pode reescrever-se como

$$\mathbf{X}' = \mathbf{X} - \bar{\mathbf{X}}_c + (\bar{\mathbf{X}}_c - \bar{\mathbf{X}}) = \mathbf{X}^{cc} + \mathbf{X}_{dif}.$$

Analogamente às duas Subsecções anteriores, efetua-se a SVD de \mathbf{X}' com base na matriz $\mathbf{S}' = (\mathbf{X}')^T \mathbf{X}'$ para obter os respetivos m valores e vetores próprios de \mathbf{S}' . Se denotarmos por λ'_i os valores próprios, por U'_i os correspondentes vetores próprios ortonormados de \mathbf{S}' , e por d' o número de valores próprios não nulos de \mathbf{S}' , então os vetores V'_i , isto é, as CPs são expressas por

$$V'_i = \mathbf{X}' U'_i / \sqrt{\lambda'_i}, \text{ com } i = 1, \dots, d' \text{ e } d' = \text{car}(\mathbf{X}') \leq m,$$

e a matriz \mathbf{X} passa a ser reescrita por

$$\mathbf{X} = \bar{\mathbf{X}}_c - \mathbf{X}_{dif} + \sum_{i=1}^{d'} \mathbf{X}'_i = \bar{\mathbf{X}}_c - \mathbf{X}_{dif} + \sum_{i=1}^{d'} \sqrt{\lambda'_i} V'_i (U'_i)^T.$$

A partir desta decomposição podemos estabelecer a seguinte relação

$$\mathbf{X} = \sqrt{\lambda_{0(c)}} V_{0(c)} U_{0(c)}^T - \sqrt{\lambda_{0(dif)}} V_{0(dif)} U_{0(dif)}^T + \sum_{i=1}^{d'} \sqrt{\lambda'_i} V'_i (U'_i)^T = \bar{\mathbf{X}}_c - \mathbf{X}_{dif} + \sum_{i=1}^{d'} \mathbf{X}'_i, \quad (4.19)$$

onde

$$V_{0(c)} = \mathbf{1}_k / \sqrt{k}, \quad U_{0(c)} = \bar{\mathbf{x}}_c / \|\bar{\mathbf{x}}_c\| \quad \text{e} \quad \sqrt{\lambda_{0(c)}} = \|\bar{\mathbf{x}}_c\| \sqrt{k},$$

assim como

$$V_{0(dif)} = \mathbf{1}_k / \sqrt{k}, \quad U_{0(dif)} = \bar{\mathbf{x}}_{dif} / \|\bar{\mathbf{x}}_{dif}\| \quad \text{e} \quad \sqrt{\lambda_{0(dif)}} = \|\bar{\mathbf{x}}_{dif}\| \sqrt{k}.$$

De acordo com a decomposição (4.19) e admitindo a existência de ortogonalidade entre as matrizes intervenientes, verifica-se a seguinte igualdade

$$\sum_{i=1}^d \lambda_i = \lambda_{0(c)} - \lambda_{0(dif)} + \sum_{i=1}^{d'} \lambda'_i. \quad (4.20)$$

4.9 Softwares e Packages

Existem vários *softwares/packages* para analisar séries temporais com a SSA como uma das suas funcionalidades.

*CaterpillarSSA*³

Destacamos aqui o *software* denominado de *CaterpillarSSA*, desenvolvido pela equipa de investigação *GistaT Group* da Universidade de São Petersburgo na Rússia. Este programa permite ao utilizador analisar passo a passo cada uma das etapas da SSA, realizar previsões e detetar pontos de viragem (mudanças de estrutura da série) para séries temporais univariadas assim como efetuar análise e previsão de séries temporais multivariadas.

No caso univariado, para além da SVD onde são determinados os valores e vetores próprios e as CPs, é possível obter gráficos para identificação da tendência, das componentes oscilatórias e

³<http://www.gistatgroup.com/cat/index.html>

do ruído, proceder-se à reconstrução das componentes da série temporal e à análise dos resíduos correspondentes e, relativamente à parte da previsão, o programa permite proceder à análise de métodos recorrentes e vetoriais de previsão, construir intervalos de predição empíricos e de *bootstrap*, contruir envelopes e testar os resultados previstos no período de validação.

Rssa⁴

Recentemente surgiu o *package* *Rssa* para a implementação da SSA, incorporado no *software* estatístico *R*, que fornece várias rotinas para analisar as etapas da técnica, visualizar resultados através de representações gráficas, efetuar previsões, entre outras (Golyandina e Korobeynikov, 2013).

SSA-MTM Toolkit⁵

Realçamos ainda o *software* *Singular Spectrum Analysis - MultiTaper Method* (SSA-MTM) *Toolkit*, que é um programa para analisar pequenas séries temporais com ruído, assim como séries multivariadas. Este *software* é desenvolvido pelo grupo de investigação *Theoretical Climate Dynamics* (TCD), liderado por Michael Ghil na Universidade da Califórnia em Los Angeles, que estuda a dinâmica do clima em todas as escalas de tempo (sazonais, anuais, entre décadas, milenares).

Este programa inclui 3 tipos de estimação do espectro de potência: o correlograma de Blackman-Tukey tradicional, o Método *Multitaper* e o Método de Máxima Entropia. Para além destas, o SSA-MTM *Toolkit* possui ainda ferramentas próprias para implementar a SSA e a MSSA.

4.10 Exemplo: Vítimas Mortais em Acidentes de Viação

As políticas de segurança rodoviária visam sobretudo contribuir para que a sinistralidade rodoviária em Portugal tenha cada vez menos impacto na sociedade. Analisar o comportamento do número de vítimas mortais que resultam dos acidentes de viação permite uma melhor perceção

⁴<http://cran.r-project.org/web/packages/Rssa/index.html>

⁵<http://web.atmos.ucla.edu/tcd/ssa/>

da influência das políticas adotadas nos últimos anos, e pode contribuir para um planeamento de novas políticas de sensibilização, de prevenção e de aplicação de legislação.

Para além do mais, este assunto continua a suscitar uma grande motivação de estudo, uma vez que no ano de 2010, a Organização das Nações Unidas (ONU) veio proclamar o período entre 2011 e 2020 como a Década de Ação para a Segurança Rodoviária com o objetivo de reduzir os acidentes de viação a nível mundial, permitindo implementar medidas que evitem a previsão de que os acidentes rodoviários se tornem na quinta causa de morte em 2030, segundo a Organização Mundial de Saúde (OMS).

Embora no Capítulo anterior, tenhamos abordado conjuntos de dados climatológicos, neste Capítulo optámos por um exemplo "mais familiar" que se enquadra no campo da sinistralidade rodoviária nacional, pelo facto da série do número de vítimas mortais constituir, como referido, um fator de preocupação dos sucessivos governos, mas também por existirem evidências desta vir a revelar-se uma "boa" fonte informativa para extrair componentes interpretáveis, do ponto de vista temporal, à luz do conhecimento geral no contexto rodoviário e até mesmo ao nível do comportamento social. Porém, nos Capítulos finais iremos novamente explorar dados climatológicos com o objetivo de efetuar comparação de métodos.

Sendo assim, nesta Secção apresentamos o estudo da série mensal do número de vítimas mortais em acidentes de viação que ocorreram em Portugal Continental entre os anos de 1998 e 2009, através da SSA. Em particular, pretende-se extrair e interpretar as componentes dominantes e períodos associados, efetuar reconstruções dos dados originais e previsões da mortalidade rodoviária mensal. Serão também efetuadas algumas comparações de resultados entre a SSA sem centragem e a SSA com centragem.

4.10.1 SSA Básica (Sem Centragem)

Tomemos como exemplo de aplicação da técnica básica da SSA, e abordado em Sebastião e Oliveira (2015), os valores referentes à série mensal do número de vítimas mortais em acidentes de viação em Portugal Continental entre 1998 e 2009.

Os dados, representados na Figura 4.2, recolhidos ao longo dos 12 anos em estudo ($n = 144$ meses) foram provenientes da Direção Geral de Viação (DGV) que exerceu funções até 2006, dando origem à atual Autoridade Nacional de Segurança Rodoviária (ANSR), onde estes se

encontram disponíveis⁶. No entanto, a entidade responsável pelo tratamento e divulgação dos dados é o Instituto Nacional de Estatística (INE). Segundo o INE, a definição de vítima de acidente, utilizada neste período em estudo, é aquela cujo óbito ocorra no local do evento ou no seu percurso até à unidade de saúde. Alguns resultados complementares deste estudo podem ser encontrados no Apêndice D.

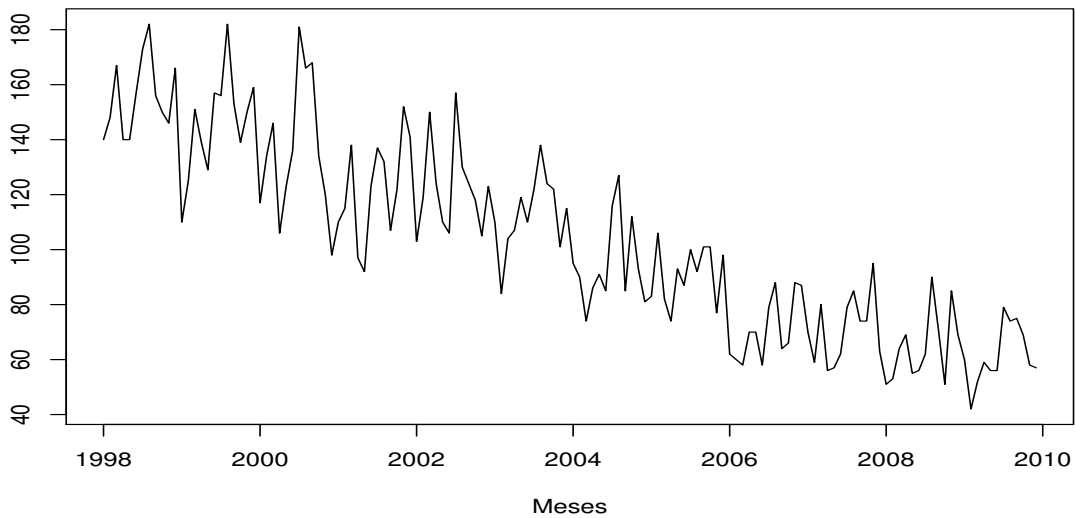


Figura 4.2: Número de vítimas mortais em acidentes de viação em Portugal Continental

Decomposição

Uma questão inicial importante na aplicação da SSA é a decisão da escolha do comprimento da janela, m . Tendo em conta a especificidade do problema em análise e alguma informação prévia da natureza da série temporal, utilizou-se o comprimento da janela $m = n/3 = 48$ meses, uma vez que era expetável que a série tivesse componentes sazonais de período inteiro, pelo que era aconselhável usar um comprimento da janela proporcional a esse período para obtermos uma melhor separação das componentes periódicas (Golyandina *et al.*, 2001; Hassani, 2007). Consequentemente, a matriz de trajetória é constituída por $k = n - m + 1 = 97$ vetores

⁶Os dados podem ser encontrados no site da ANSR, <http://www.ansr.pt/>, a partir dos relatórios anuais de sinistralidade rodoviária.

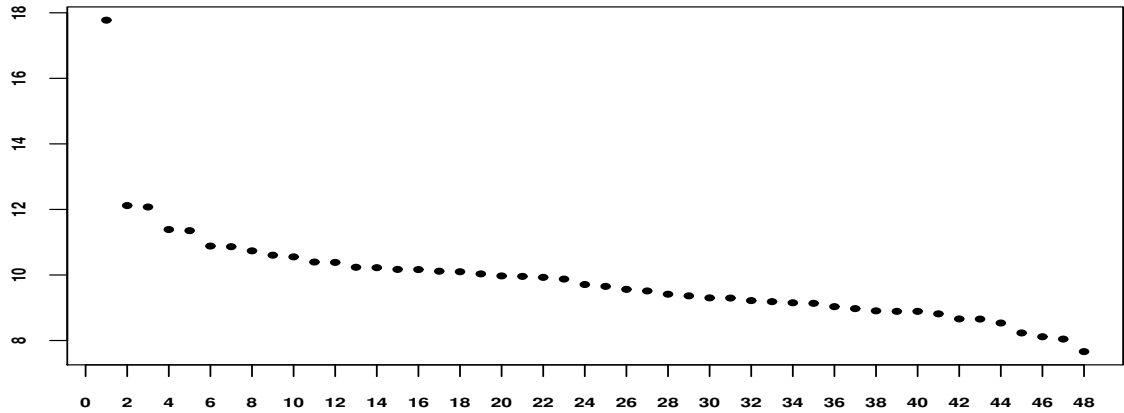


Figura 4.3: Logaritmos dos 48 valores próprios

desfasados correspondentes ao número de linhas e por $m = 48$ colunas.

Com base na matriz de trajetória \mathbf{X} e através da decomposição em valores singulares, selecionaram-se as primeiras 16, de 48 componentes principais, essencialmente com base nos respetivos valores próprios de $\mathbf{X}^T \mathbf{X}$ ilustrados nas Figuras 4.3 e 4.4, associados às CPs apresentadas nas Figuras 4.5 e 4.6. Desta forma, as restantes componentes foram desprezadas por se admitir que continham a parte correspondente ao ruído. O primeiro valor próprio é muito elevado em relação aos restantes, o que é natural uma vez que estamos a usar o modelo básico da SSA sem qualquer tipo de centragem de dados, e os valores próprios obtêm-se da matriz $\mathbf{X}^T \mathbf{X}$ em vez da matriz de variâncias covariâncias.

Às 16 CPs que ficaram retidas associaram-se sete pares de componentes oscilatórias (CP2-3, CP4-5, CP6-7, CP9-10, CP11-12, CP13-14, CP15-16), por ostentarem um comportamento oscilatório semelhante assim como valores próprios muito similares, enquanto a CP1 corresponde à tendência decrescente do número de vítimas mortais (de 1998 para 2009 ocorreu uma redução de 60% no valor total anual, o que corresponde, em média, a uma redução de 5 para 2 vítimas mortais por dia) e a CP8 exhibe uma mistura de tendência com componente oscilatória.

Os valores das percentagens de variância explicadas pelos pares de CPs descritos anteriormente (entre as primeiras 16 CPs de um total de 48 CPs) são apresentadas na Tabela 4.1.

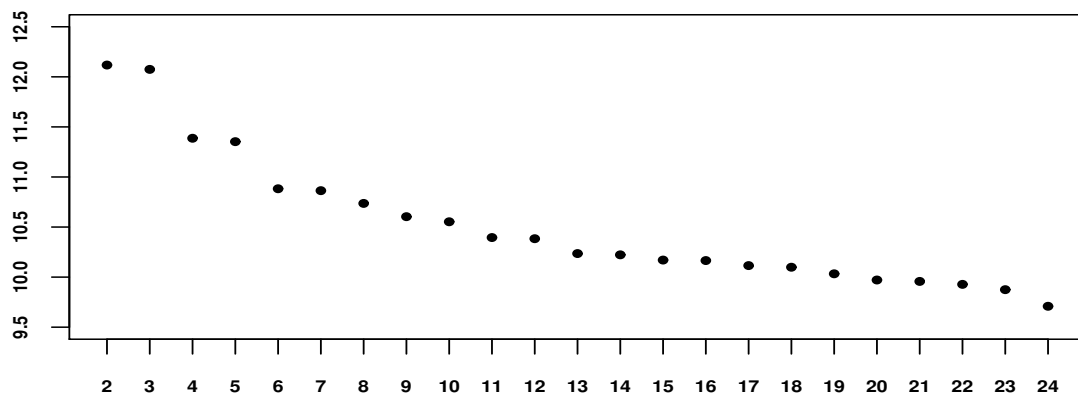


Figura 4.4: Logaritmos dos 24 primeiros valores próprios exceto o primeiro

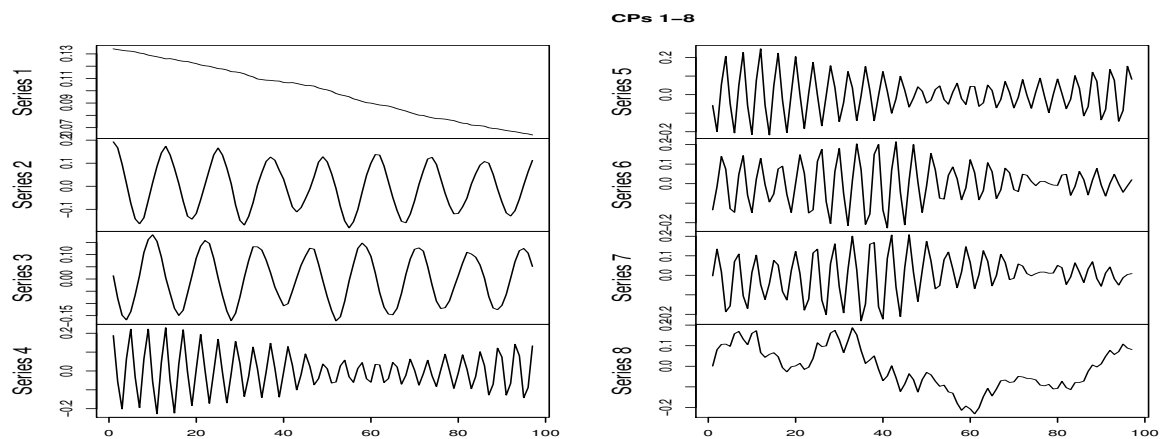


Figura 4.5: Componentes principais CP1 a CP8

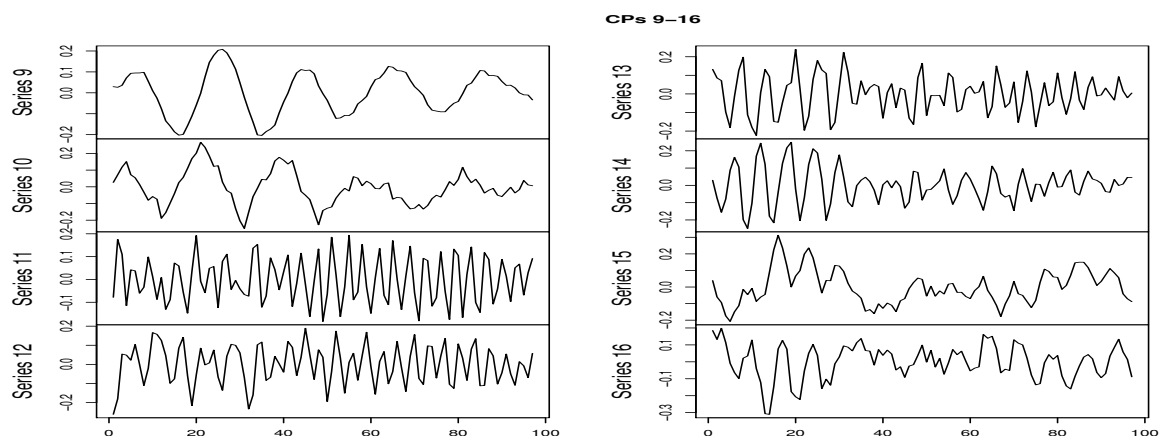


Figura 4.6: Componentes principais CP9 a CP16

Tabela 4.1: Percentagens de variância de cada uma das primeiras 16 CPs e de sete pares dessas CPs, de um total de 48 CPs. A percentagem de variância de cada par de CPs foi calculada através da soma das percentagens individuais das respectivas CPs.

CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12	CP13	CP14	CP15	CP16
97.57	0.34	0.33	0.164	0.158	0.10	0.10	0.086	0.075	0.071	0.061	0.060	0.052	0.051	0.049	0.048
CP1	CP2-3	CP4-5	CP6-7	CP8	CP9-10	CP11-12	CP13-14	CP15-16							
97.57	0.67	0.322	0.20	0.086	0.146	0.121	0.103	0.097							

O total da percentagem de variância explicada pelas 16 primeiras CPs é 99.307%. A CP1 associada à tendência dos dados apresenta uma percentagem de variância muitíssimo elevada (97.57%) comparativamente com as restantes CPs. Podemos argumentar que tal ocorrência deve-se a uma nítida e predominante tendência decrescente do número de vítimas mortais em acidentes de viação ao longo dos anos. Situações extremas e similares a esta em que a tendência se destaca fortemente ocorrem com alguma frequência em séries de valores relacionados com a produção industrial (Hassani *et al.*, 2009; Golyandina *et al.*, 2001). Embora muitas das vezes se retire a tendência aos dados (como nalguns exemplos de ACP), decidimos manter esta componente uma vez que a aplicação da SSA a este conjunto de dados consiste essencialmente em realçar a capacidade da técnica em extrair as componentes de tendência, de oscilação e de ruído, procedimento também utilizado em Hassani (2007).

Análise das Periodicidades das CPs Retidas

Para auxiliar na identificação dos períodos dos pares de componentes que são geradas por uma harmônica, para além da inspeção visual do comportamento individual das CPs como séries temporais e mantendo em linha de conta os pares de valores próprios similares associados, recorreremos também à visualização gráfica dos pares de CPs. A Figura 4.7 apresenta os gráficos para os sete pares de CPs assinalados anteriormente com as respetivas percentagens de variância associadas a cada CP, evidenciando-se principalmente que os dois primeiros pares de CPs em análise correspondem a harmónicas com períodos 12 e 4 meses.

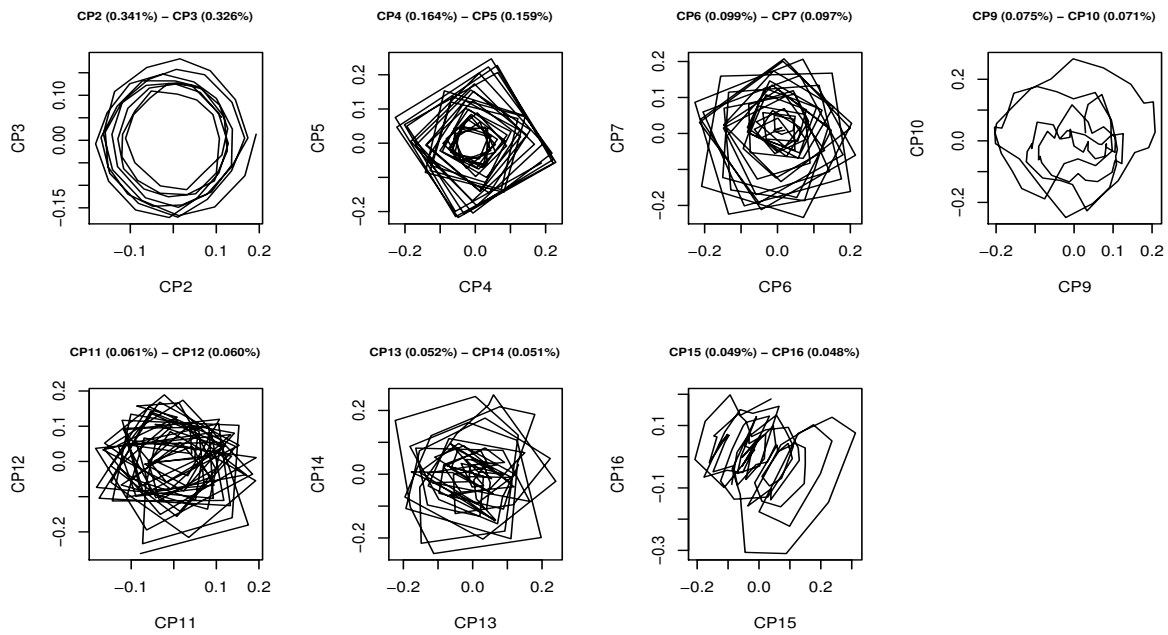


Figura 4.7: Pares de CPs harmónicas

Recorremos ainda à visualização gráfica dos espectros (Figura 4.8) dos sete pares de CPs oscilatórias retidas, donde se concluiu que os períodos dominantes dos respetivos pares indicados são aproximadamente iguais a 12, 4, 4, 20, 4, 6 e 33 meses respetivamente.

Note-se que embora em Vautard *et al.* (1992) se refira que o comprimento da janela deverá permitir analisar com sucesso os períodos pertencentes ao intervalo $(m/5, m)$, e que neste caso corresponde a $(9.6, 48)$, tal recomendação não invalida que se possam considerar também os períodos de 4 e 6 meses explícitos em alguns pares de CPs.

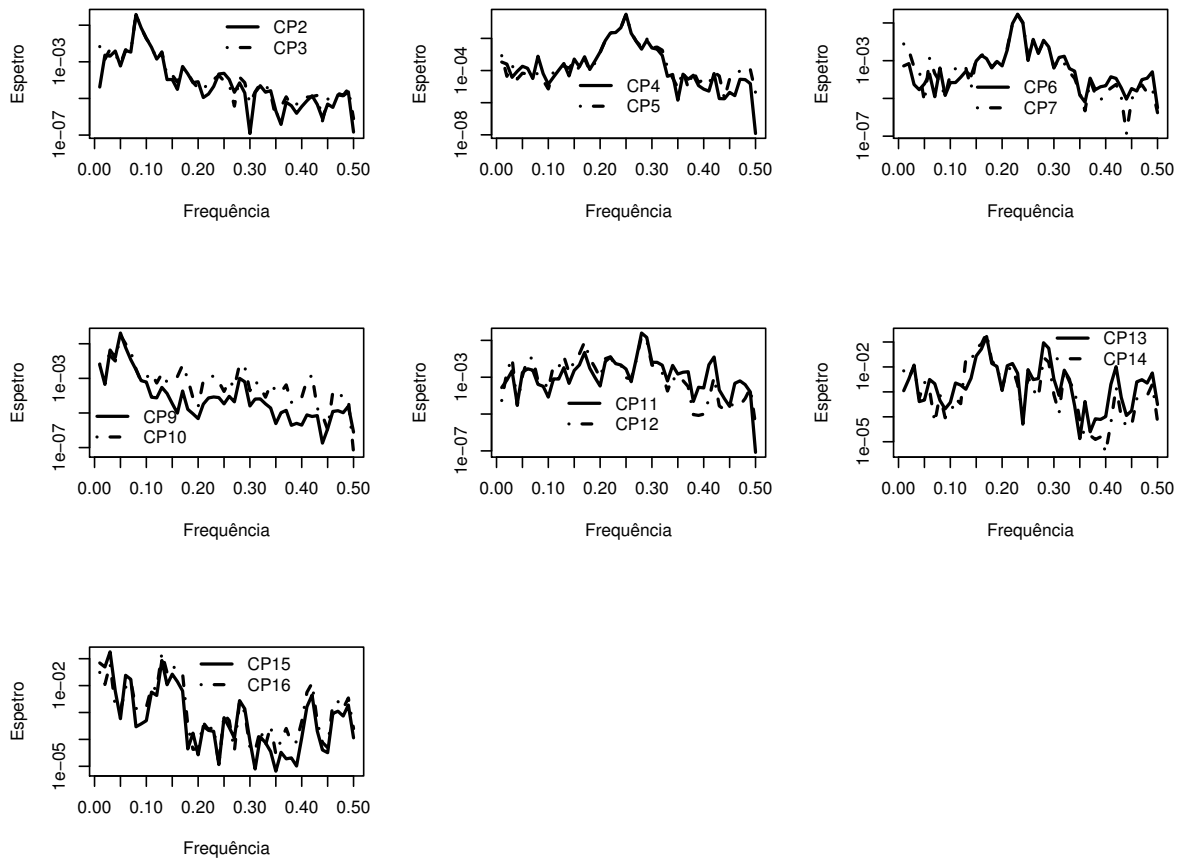


Figura 4.8: Espectros dos sete pares de CPs

Para complementar as análises anteriores e testar periodicidades escondidas de uma determinada frequência, consideremos a aplicação do teste de hipóteses de Fisher g (Brockwell e Davis, 1991; de Carvalho e Rua, 2014), sob a hipótese nula de que um pico espectral não é significativo contra a alternativa de que existe uma componente periódica. Todas as CPs dos sete pares retidos revelaram a existência de uma componente periódica estatisticamente significativa, para qualquer nível de significância usual.

Tendo em conta a natureza da série em estudo, é possível efetuar uma interpretação dos ciclos temporais mais relevantes no contexto da sinistralidade rodoviária. Para além da compreensível sazonalidade anual, destaca-se o ciclo de 4 meses, uma vez que encontramos 3 principais picos de frequência em que o número de vítimas mortais é elevado, como podemos constatar nos

cronogramas ao longo de cada ano na Figura 4.9.

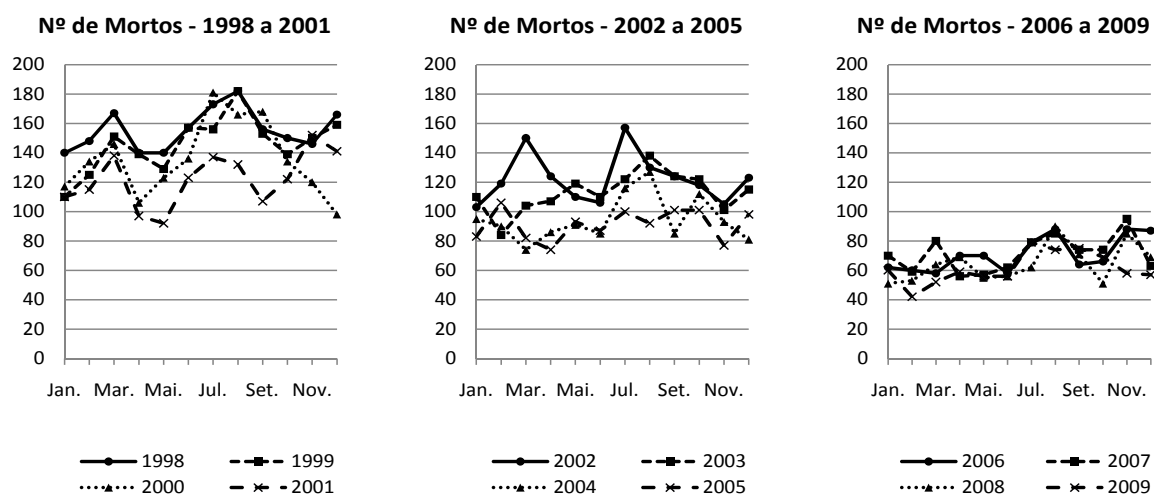


Figura 4.9: Cronogramas do número de vítimas mortais mensais entre os anos de 1998 e 2009

Estes 3 picos correspondem aos períodos de maior tráfego rodoviário em que muitas famílias gozam férias - férias escolares da Páscoa (março a maio), férias de verão (julho-agosto) e época natalícia. No período natalício, a falta de cuidados de alguns condutores na condução em condições climatéricas adversas provoca um aumento de acidentes e naturalmente o número de mortos tem tendência a aumentar. Contudo há que ter em consideração que os dados em causa correspondem à totalidade dos mortos registados em todo o território nacional continental e não desagregados ao nível dos distritos, nos quais se verificam geralmente discrepâncias significativas entre as condições meteorológicas.

Há que realçar ainda o período com cerca de 33 meses (2 anos e 9 meses aproximadamente) na análise dos dados, o qual leva a pressupor uma relação com as alterações à legislação em vigor e com as campanhas de sensibilização introduzidas essencialmente na última década ilustradas na Figura 4.10.

Simão (2010) descreve que entre os fatores mais significativos que têm contribuído para o decréscimo dos índices de sinistralidade em Portugal, onde se inclui a mortalidade rodoviária, destacam-se as sucessivas alterações à legislação em vigor e a eficiência na sua fiscalização.

Por outro lado, quando se compara Portugal com os países da União Europeia a 27 Estados Membros, este sobressai-se por ser um dos países com melhores resultados na diminuição dos valores de sinistralidade rodoviária, onde verificamos uma subida do 25º lugar em 1998 para

o 15º lugar em 2009 dos países com menor número total de vítimas mortais por milhão de habitantes.

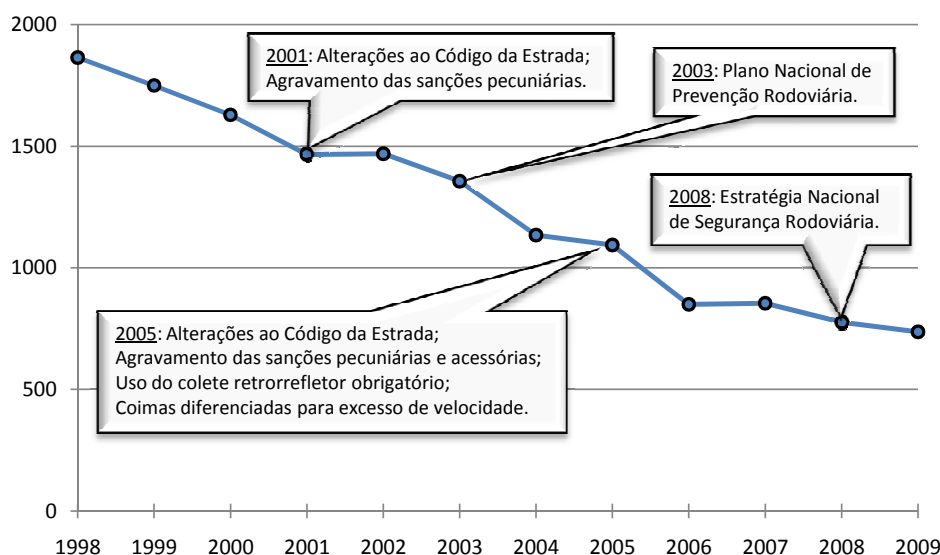


Figura 4.10: Evolução anual do número de vítimas mortais e medidas de segurança rodoviária

Reconstrução da Série Original do Número de Vítimas Mortais e Previsão

No que se refere à reconstrução da série original, começemos por analisar a reconstrução à custa apenas da CP1. Geralmente a primeira CP descreve a tendência geral dos dados, o que se pode comprovar ao analisar o seu comportamento na Figura 4.11.

A Figura 4.12 permite mostrar a evolução da reconstrução da série à medida que se aumenta o número de CPs consideradas, nomeadamente à custa das primeiras 5 CPs, das primeiras 10 CPs e das primeiras 16 CPs. Ao utilizar as primeiras 16 CPs retidas inicialmente, podemos constatar que a qualidade da reconstrução é bastante satisfatória, e na Figura 4.13 verificamos que existe uma grande parte dos resíduos com valores próximos de zero. Tal elevado nível de satisfação poderia eventualmente suscitar algumas dúvidas, indagando até que ponto é que poderia estar a ser incorporado algum ruído da reconstrução por mais residual que fosse, podendo mesmo pôr em causa a periodicidade de algumas das últimas CPs que foram retidas, senão tivéssemos utilizado a informação auxiliar presente nas Figuras 4.6, 4.7 e 4.8.

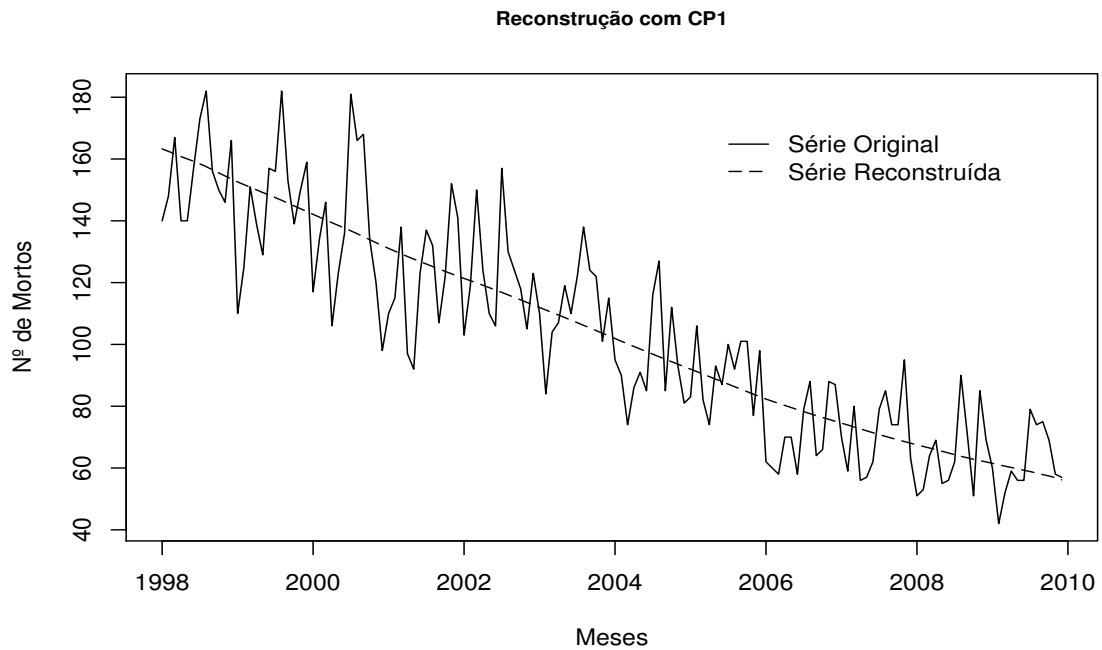


Figura 4.11: Reconstrução da série original com a CP1

As primeiras 16 CPs identificadas anteriormente foram utilizadas na reconstrução da série original sem ruído e permitiram aplicar o algoritmo de previsão recorrente em SSA com o intuito de prever novos valores da série. Utilizámos o *software R*, no qual implementámos todos os passos da técnica assim como uma rotina (Apêndice D) para o algoritmo de previsão recorrente.

Na Figura 4.14 observamos as séries original e prevista, a qual inclui a previsão dos 12 valores mensais para o ano de 2010. A linha vertical no gráfico corresponde ao chamado ponto de truncagem, a partir do qual deixamos de ter valores observados e passamos a ter apenas previsões.

Constatamos que os valores previstos do número de vítimas mortais para os meses do ano de 2010 variam aproximadamente entre os 48 e os 73, os quais acompanham a tendência de uma certa estabilização suave que já era notória nos últimos anos desde 2006. É de notar que embora a série temporal em estudo seja constituída apenas por valores inteiros, podemos efetuar previsões cujos resultados sejam valores não inteiros. Tal situação aparece com alguma frequência em estudos como por exemplo em Hassani (2007) ou em Golyandina *et al.* (2001).

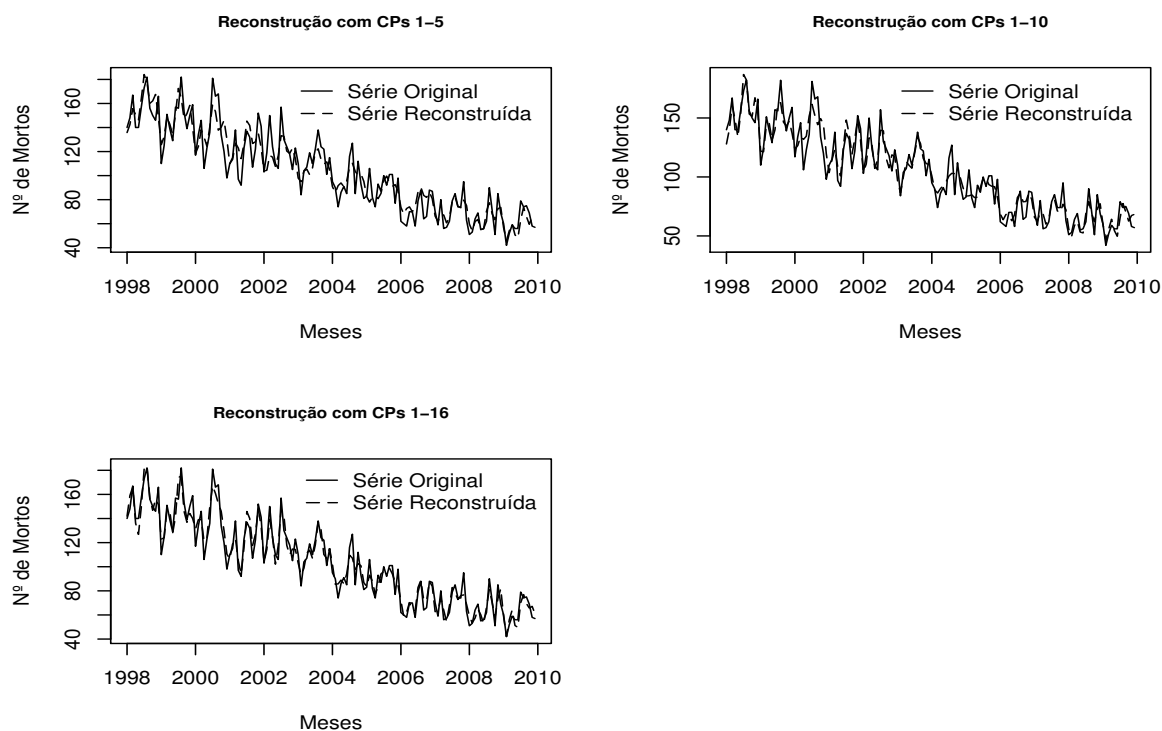


Figura 4.12: Reconstruções da série original com as CPs 1-5, CPs 1-10 e CPs 1-16

Tendo em conta que as previsões são consideradas um pouco voláteis consoante os parâmetros estipulados no algoritmo, ilustramos na Figura 4.15 para além das séries original e prevista, as previsões com os respetivos intervalos de predição empíricos a 95%, construídos através do método empírico descrito na Subsecção 4.4.2, para os 12 meses de 2010.

A amplitude dos intervalos de predição empíricos é praticamente constante dada a homogeneidade dos resíduos existente aquando da reconstrução (ver Figura 4.13). Os gráficos das Figuras 4.14 e 4.15 em particular foram obtidos através do *software CaterpillarSSA*.

Uma vez que atualmente estão divulgados os dados oficiais do número de vítimas mortais em acidentes de viação ocorridos no ano de 2010, podemos efetuar a comparação dos 12 valores mensais previstos para o ano de 2010 com os valores observados (Figura 4.16).

Verificamos que em julho de 2010, a diferença entre predição e observação é significativa e de aproximadamente 30 mortes, o que poderá estar a influenciar as previsões posteriores e por isso admite-se a necessidade de um eventual reajuste da série.

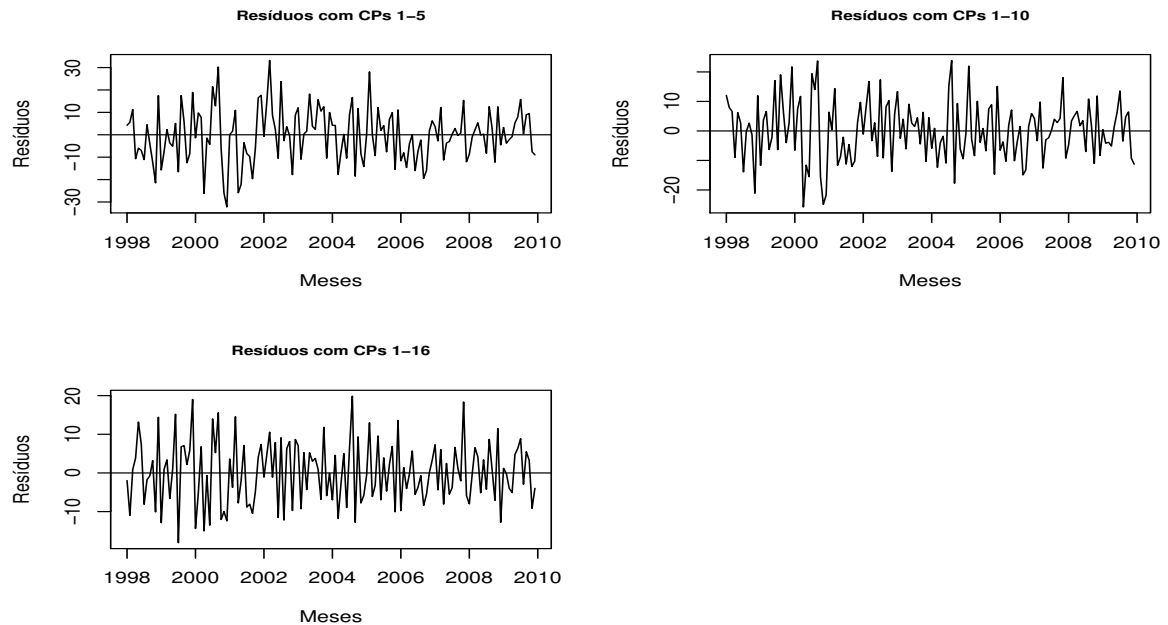


Figura 4.13: Resíduos para a série original ao utilizar as reconstruções com as CPs 1-5, CPs 1-10 e CPs 1-16

Como medidas de precisão das previsões, verificamos que a raiz do erro quadrático médio é igual a 13.05 e que a precisão direcional média (percentagem de vezes em que a previsão e a série temporal variam no mesmo sentido entre dois momentos temporais consecutivos) é apenas cerca de 27.27%, o que nos leva a considerar que o ajustamento é fraco.

Simulações para Valores do Comprimento da Janela

Como a escolha adequada do comprimento da janela m , é algo que apresenta uma certa liberdade dentro de certos limites, efetuámos algumas simulações para valores credíveis do comprimento da mesma, para analisar se existem discrepâncias na informação que se pode extrair através das componentes dominantes. Começámos por considerar $m = 6$ meses, seguindo-se os valores múltiplos de 12 meses até um limite de metade do número de observações da série em estudo que corresponde a $m = 72$ meses, ou seja 6 anos. Na Tabela 4.2 apresentamos a percentagem de variância explicada por cada uma das 5 primeiras CPs, assim como a percentagem acumulada por essas 5 CPs para os diferentes valores do comprimento da janela descritos.

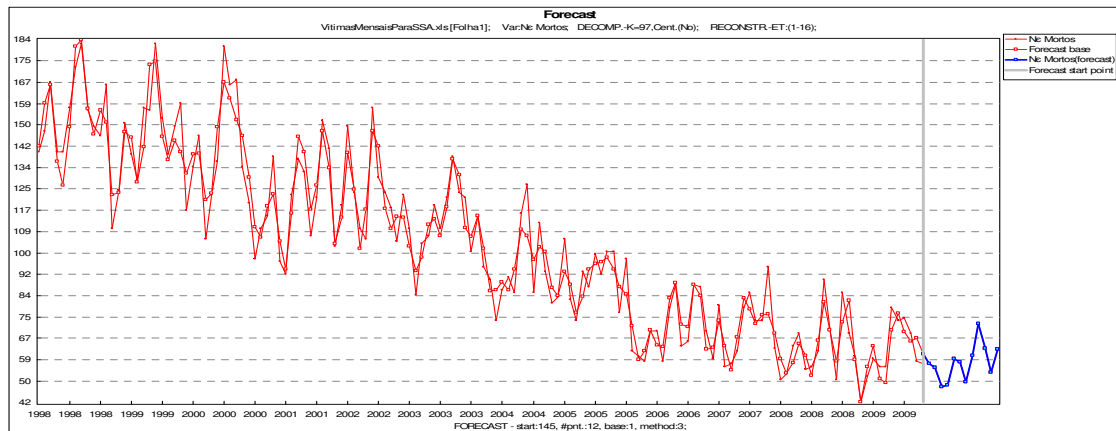


Figura 4.14: Séries original e série prevista construída com base nas CPs 1-16; previsões dos 12 valores mensais para o ano de 2010

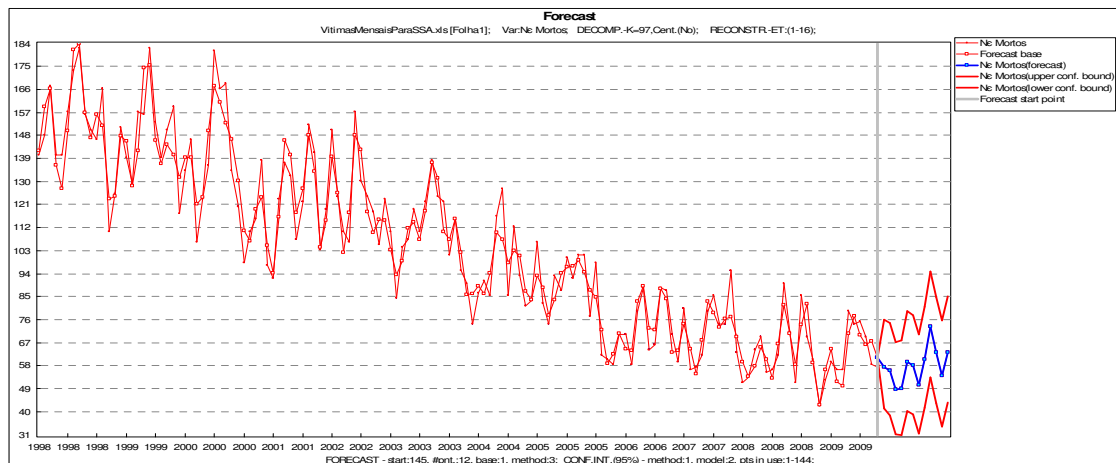


Figura 4.15: Séries original e série prevista construída com base nas CPs 1-16; previsões dos 12 valores mensais para o ano de 2010 e intervalos de predição a 95% obtidos através do método empírico

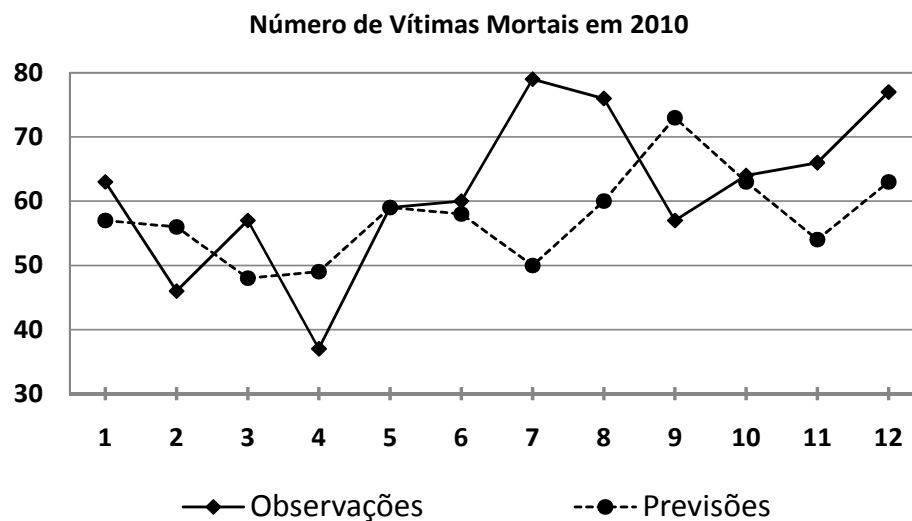


Figura 4.16: Comparação do número de vítimas mortais previstas *versus* observadas para os 12 meses do ano de 2010

Tabela 4.2: Percentagem de variância explicada pelas primeiras 5 CPs ao longo de diferentes valores do comprimento da janela - lag m

Lag m	CP1 (%)	CP2 (%)	CP3 (%)	CP4 (%)	CP5 (%)	CPs 1-5 (%)
6	98.26	0.55	0.43	0.39	0.19	99.84
12	97.79	0.46	0.43	0.31	0.29	99.26
24	97.65	0.38	0.37	0.23	0.22	98.86
36	97.58	0.36	0.34	0.19	0.18	98.65
48	97.57	0.34	0.33	0.16	0.16	98.56
60	97.56	0.34	0.33	0.14	0.13	98.51
72	97.55	0.35	0.34	0.14	0.13	98.51

Os resultados evidenciam uma dominância geral bastante elevada da primeira CP sobre as restantes, mesmo para os diferentes valores do comprimento da janela considerados. As percentagens acumuladas pelas primeiras 5 CPs indicam pequenas variações entre os diferentes valores de m , permitindo assim concluir que a troca de um valor de m por um outro qualquer destes pouco influencia na variabilidade que se pretende captar nos dados.

4.10.2 Comparações entre SSA Sem Centragem e SSA Com Centragem

Aspetos Gerais

Nesta Subsecção pretendemos efetuar algumas considerações sobre as comparações entre os diferentes tipos de abordagens da técnica de SSA: a SSA sem centragem e quatro tipos de SSA com centragem (a centragem simples por colunas, a centragem simples por linhas, a centragem dupla e a centragem na série original), seguindo uma mesma linha orientadora de raciocínio da Secção 3.12, usando o exemplo prático dos dados do número de vítimas mortais em acidentes de viação.

Tais comparações contemplam uma abordagem numérica dos valores próprios e dos vetores próprios, assim como uma análise das coordenadas das CPs retidas no estudo ao nível temporal e espectral. Na discussão iremos ainda dar um especial enfoque à questão da influência da componente que descreve a tendência, caso ela exista.

É expetável que na centragem simples por colunas da matriz de trajetória se obtenham resultados semelhantes aos da centragem na série original, embora não seja obrigatório uma vez que ao centrar todas as colunas trabalha-se com uma série de média zero, mas em que os valores correspondentes, com desfasamentos diferentes, já não são iguais (ou seja, \mathbf{X}^{cc} não é uma matriz de Hankel).

Ilustração Prática

Recordemos que ao aplicar a SSA básica (sem centragem) após a decomposição em valores singulares da matriz de trajetória, foram obtidas 48 CPs. Foram efetuados procedimentos análogos na aplicação da SSA com centragem.

Valores Próprios e CPs

Tendo como referência os gráficos da Figura 4.17 e da 4.18, foram selecionadas as primeiras 16 CPs obtidas segundo os métodos sem centragem, da centragem simples por colunas e da centragem na série original. No entanto, para os métodos da centragem simples por linhas e da centragem dupla optámos por seleccionar apenas as primeiras 15 CPs através de critérios análogos aos utilizados nos métodos anteriores (inspeção visual dos gráficos dos valores próprios e das CPs), permitindo assim desprezar as restantes componentes em que se admite possuírem a parte correspondente ao ruído.

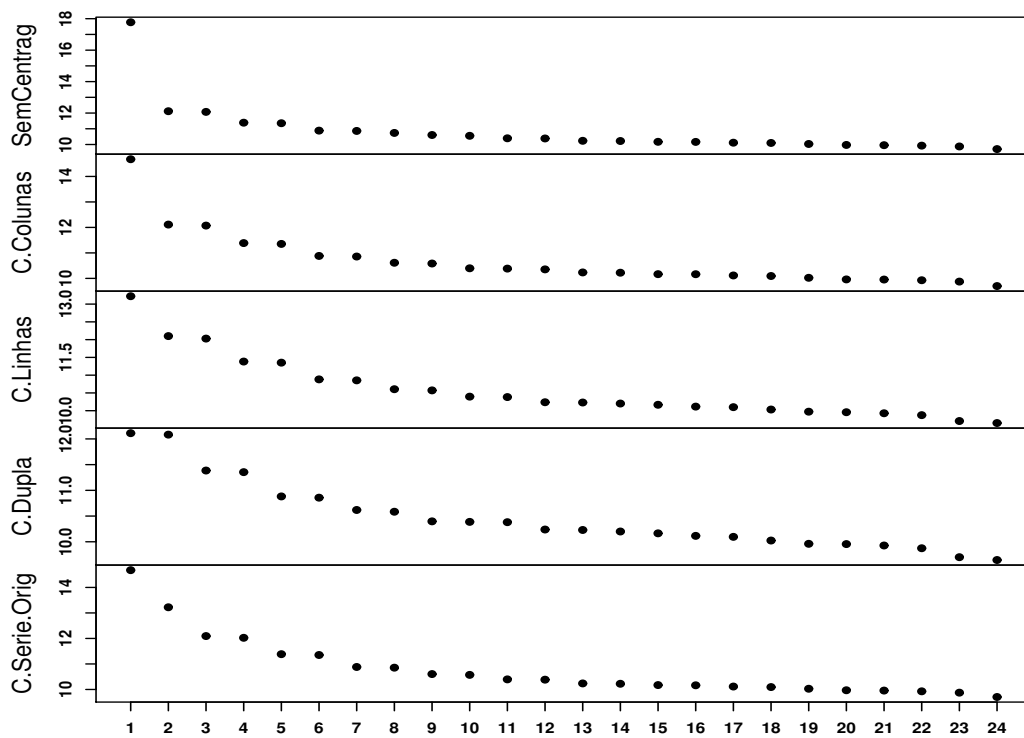


Figura 4.17: Logaritmos dos 24 valores próprios para os 5 tipos de SSA em análise

A tendência decrescente dos dados ao longo do tempo destaca-se na primeira CP, para o caso sem centragem, na centragem simples por colunas e na centragem na série original (não aparecendo nos outros dois casos, levando portanto à escolha de apenas 15 CPs em vez das 16).

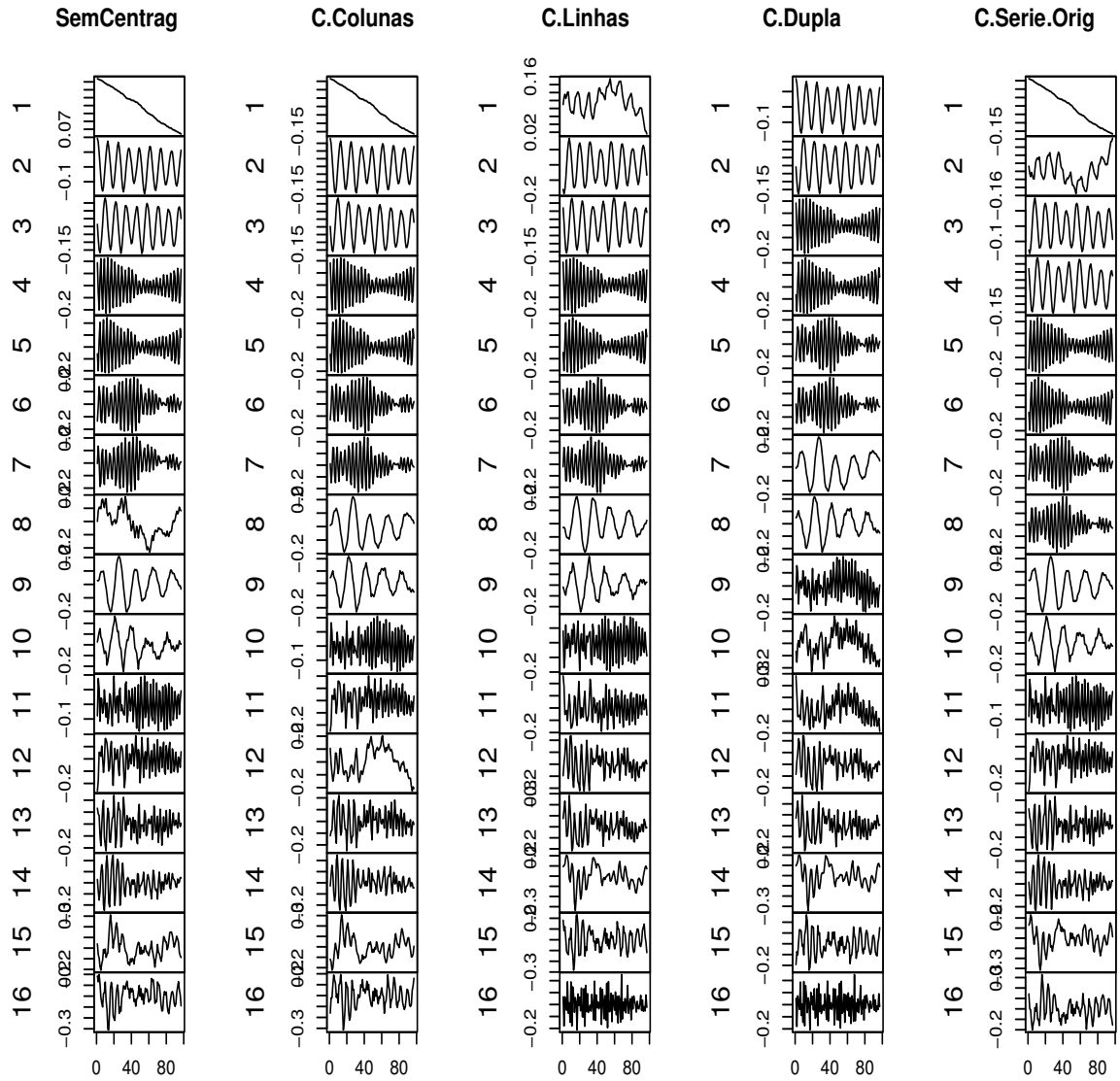


Figura 4.18: 16 CPs dominantes para os 5 tipos de SSA em análise

Uma outra CP que exibe uma mistura de tendência com componente oscilatória não se encontra na mesma posição em todos os modelos, uma vez que corresponde no modelo sem centragem à CP8, na centragem simples por colunas corresponde à CP12, na centragem simples por linhas à CP1, na centragem dupla (embora menos perceptível ao nível visual) à CP11 e no modelo com centragem na série original à CP2.

Há a realçar que no modelo de SSA com centragem dupla deixamos de ter presente a CP que representa a tendência dos dados, o que está de acordo com a conclusão de Golyandina *et al.* (2001), de que, ao efetuar-se a centragem dupla, geralmente a componente de tendência é eliminada.

Na Tabela 4.3 encontram-se os três maiores e os três menores valores próprios resultantes da SVD para cada um dos modelos. Para os modelos sem centragem, com centragem simples por colunas e centragem na série original, o primeiro valor próprio apresenta uma ordem de grandeza muitíssimo superior aos restantes valores próprios, a qual é ligeiramente menos acentuada na centragem simples por linhas, e ainda menos acentuada na centragem dupla.

Tabela 4.3: Valores próprios da SVD para a SSA sem centragem e para a SVD de cada um dos quatro tipos de SSA com centragem em análise

Val. Próprios	Sem Centragem	C.Colunas	C.Linhas	C.Dupla	C.Serie.Orig
1	52439123.58	2363970.75	552599.08	181967.23	2366139.24
2	183087.15	181934.56	179785.82	176886.76	552495.89
3	175322.60	174855.57	167595.46	88023.62	178740.87
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
46	3345.71	3345.19	3120.043	3084.221	3348.99
47	3113.39	3079.26	2132.825	2116.224	3111.47
48	2125.63	2116.15	0	0	2132.35

Nos modelos da centragem por linhas e da centragem dupla em que se eliminou a componente de tendência, verifica-se a existência de um último valor próprio nulo, uma vez que estes tipos de centragens introduzem uma componente com valor próprio nulo na SVD, devido à restrição adicional da soma dos elementos de cada linha ser nula (Jolliffe, 2002). Tal restrição leva a que o (i) -ésimo valor próprio da centragem por colunas seja relativamente semelhante ao $(i - 1)$ -ési-

mo (com $i = 2, 3, \dots, 48$) valor próprio na centragem dupla em termos de grandeza, e que os respectivos coeficientes dos vetores próprios e das componentes principais sejam também similares entre modelos. Como consequência, uma CP da SSA com centragem por colunas passa a ter variância nula na centragem dupla, obrigando portanto a que esta ocupe a última posição.

De acordo com resultados teóricos em Cadima e Jolliffe (2009) para o caso da ACP centrada por colunas e da ACP não centrada, e também deduzidos para a SSA na Subsecção 4.8.1, sabe-se que $tr(\mathbf{X}^T \mathbf{X}) - tr((\mathbf{X}^{cc})^T \mathbf{X}^{cc}) = k \|\bar{\mathbf{x}}_c\|^2$, onde $\|\bar{\mathbf{x}}_c\|^2$ representa o quadrado da distância da origem ao centro de gravidade da nuvem de pontos e k corresponde às 97 linhas da matriz não centrada \mathbf{X} . Ainda segundo os autores é pouco provável que qualquer uma das matrizes $\mathbf{X}^T \mathbf{X}$ ou $(\mathbf{X}^{cc})^T \mathbf{X}^{cc}$ tenha um vetor próprio que seja exatamente colinear com o vetor de médias das colunas da matriz \mathbf{X} , $\bar{\mathbf{x}}_c$, mas muitas das vezes o primeiro vetor próprio da matriz $\mathbf{X}^T \mathbf{X}$ está próximo dessa direção, que une a origem e o centro de gravidade da nuvem dos pontos.

No exemplo em estudo, para o caso da SSA sem centragem, uma vez que o traço de $\mathbf{X}^T \mathbf{X}$ é $tr(\mathbf{X}^T \mathbf{X}) = 53745457$, então $97 \times \|\bar{\mathbf{x}}_c\|^2 = 50092597$, o que representa 93.20% de variabilidade relativamente ao traço de $\mathbf{X}^T \mathbf{X}$. Esta percentagem relativamente elevada, reforça a direção principal de variação em torno da origem dos dados, que por sua vez é próxima da direção que liga a origem e o centro de gravidade da nuvem. Por outro lado, os valores próprios para a SSA sem centragem e para a SSA com centragem por colunas são relativamente semelhantes, exceto para o maior valor próprio de cada caso, uma vez que, nomeadamente na centragem por colunas, este está associado ao vetor próprio que se situa muito próximo da direção de $\bar{\mathbf{x}}_c$. Tal constatação fundamenta-se no facto de $|\cos(U_1, \bar{\mathbf{x}}_c)| = 0.9999923$, para o primeiro vetor próprio, U_1 , da SSA sem centragem e $|\cos(U_1^{cc}, \bar{\mathbf{x}}_c)| = 0.9960638$, para o primeiro vetor próprio, U_1^{cc} , da SSA com centragem por colunas. Para estes vetores próprios, por analogia, tendo em conta as deduções em torno da ACP em Cadima e Jolliffe⁷ (2009), a similaridade estabelece-se entre $\lambda_1^{cc} = 2363970.75$ e $\lambda_1 - 97 \times \|\bar{\mathbf{x}}_c\|^2 = 2346527$. Portanto pode-se argumentar que é mais relevante a direção dos vetores próprios (de acordo com os cossenos anteriormente apresentados) que a grandeza dos valores próprios.

Uma comparação (Tabela 4.4) em relação à percentagem de variância explicada por cada

⁷É de salientar que neste artigo, os autores utilizaram um exemplo relativo a medidas morfométricas de lagostins provenientes do Lago Opeongo, no Canadá, previamente estudados por Somers (1986) em que não existe estrutura temporal entre os dados.

umas das CPs para os cinco modelos considerados, permite destacar as diferenças na primeira CP em que na SSA sem centragem a percentagem é extremamente elevada de aproximadamente 97.57%, comparativamente com as primeiras CPs dos restantes modelos. No modelo da centragem dupla evidencia-se um maior equilíbrio na distribuição das percentagens por entre as CPs. No Apêndice D podem ser encontrados alguns cálculos, para este conjunto de dados, alusivos às relações teóricas entre os traços das matrizes envolvidas em cada SVD dos modelos em análise.

Tabela 4.4: Percentagens de variâncias das CPs da SVD para a SSA sem centragem e para a SVD de cada um dos quatro tipos de SSA com centragem em análise

CPs	Sem Centragem	C.Colunas	C.Linhas	C.Dupla	C.Serie.Orig
CP1	97.5694	64.7156	30.6455	14.0629	56.8048
CP2	0.3407	4.9806	9.9704	13.6703	13.2640
CP3	0.3262	4.7868	9.2943	6.8027	4.2911
CP4	0.1640	2.4097	4.8708	6.5927	4.0081
CP5	0.1585	2.3313	4.7221	4.1107	2.1084
CP6	0.0990	1.4558	2.9498	4.0127	2.0415
⋮	⋮	⋮	⋮	⋮	⋮
CP46	0.0062	0.0916	0.1730	0.2384	0.0804
CP47	0.0058	0.0843	0.1183	0.1635	0.0747
CP48	0.0040	0.0579	0	0	0.0512

Embora as CPs não estejam todas ordenadas da mesma forma em termos do seu tipo (tendência, oscilatórias e misturas de tendência com oscilatórias), contudo as CPs que captam as variações mais relevantes ocorridas nos dados são praticamente as mesmas em todos os modelos nas primeiras 15/16 CPs. Para os cinco modelos analisados, destacamos os seguintes sete pares mais relevantes de CPs oscilatórias, cujos comportamentos são semelhantes e cujos respetivos valores próprios são muito similares.

- Sem centragem: CP2-3, CP4-5, CP6-7, CP9-10, CP11-12, CP13-14, CP15-16;
- Centragem simples por colunas: CP2-3, CP4-5, CP6-7, CP8-9, CP10-11, CP13-14, CP15-16;

- Centragem simples por linhas: CP2-3, CP4-5, CP6-7, CP8-9, CP10-11, CP12-13, CP14-15;
- Centragem dupla: CP1-2, CP3-4, CP5-6, CP7-8, CP9-10, CP12-13, CP14-15;
- Centragem na série original: CP3-4, CP5-6, CP7-8, CP9-10, CP11-12, CP13-14, CP15-16.

Vetores Próprios

No que se refere às comparações dos vetores próprios da SVD entre os diferentes modelos em estudo, apresentamos na Tabela 4.5 o valor de alguns dos cossenos (em valor absoluto) dos ângulos entre os respectivos pares de vetores.

Tabela 4.5: Cossenos entre os vetores próprios da SVD da SSA sem centragem e os vetores próprios da SVD de cada um dos quatro tipos de SSA com centragem em análise

SSA sem centragem vs SSA com centragem simples por colunas						
$\cos(U_1, U_1^{cc})$	$\cos(U_2, U_2^{cc})$	$\cos(U_3, U_3^{cc})$	\dots	$\cos(U_{46}, U_{46}^{cc})$	$\cos(U_{47}, U_{47}^{cc})$	$\cos(U_{48}, U_{48}^{cc})$
0.9964	0.9944	0.9948	\dots	0.9998	0.9990	0.9997
SSA sem centragem vs SSA com centragem simples por linhas						
$\cos(U_1, U_1^{cl})$	$\cos(U_2, U_2^{cl})$	$\cos(U_3, U_3^{cl})$	\dots	$\cos(U_{46}, U_{45}^{cl})$	$\cos(U_{47}, U_{46}^{cl})$	$\cos(U_{48}, U_{47}^{cl})$
0.0970	0.8406	0.8205	\dots	0.9997	0.9996	0.9999
SSA sem centragem vs SSA com centragem dupla						
$\cos(U_2, U_1^{**})$	$\cos(U_3, U_2^{**})$	$\cos(U_4, U_3^{**})$	\dots	$\cos(U_{46}, U_{45}^{**})$	$\cos(U_{47}, U_{46}^{**})$	$\cos(U_{48}, U_{47}^{**})$
0.9978	0.9979	0.9995	\dots	0.9997	0.9989	0.9997
SSA sem centragem vs SSA com centragem na série original						
$\cos(U_1, U_1')$	$\cos(U_2, U_3')$	$\cos(U_3, U_4')$	\dots	$\cos(U_{46}, U_{46}')$	$\cos(U_{47}, U_{47}')$	$\cos(U_{48}, U_{48}')$
0.9951	0.8707	0.8551	\dots	0.9995	0.9994	0.9999

Entre o modelo da SSA sem centragem e o modelo da SSA com centragem por colunas, os valores dos cossenos dos ângulos dos vetores correspondentes são praticamente todos muito próximos da unidade (em módulo), o que reforça o grau de semelhança de ambos os modelos. É de notar que, como nos modelos com centragem simples por linhas e com centragem dupla não temos a componente da tendência, a maior parte dos pares de vetores foram constituídos à custa do (i) -ésimo vetor da SSA sem centragem com o $(i - 1)$ -ésimo vetor (com $i = 2, 3, \dots, 48$) da SSA com centragem (simples por linhas ou dupla).

Ao comparar o modelo com centragem dupla relativamente ao modelo sem centragem, grande parte dos valores dos cossenos dos ângulos dos vetores são muito próximos da unidade (em módulo), o que de alguma forma evidencia o efeito que a centragem dupla provoca no modelo, eliminando apenas o vetor correspondente à tendência dos dados e mantendo os restantes vetores a desempenharem praticamente o mesmo papel.

Correlações entre CPs

Passemos a efetuar de seguida as comparações entre as coordenadas das CPs do modelo básico sem centragem e as coordenadas das CPs de cada um dos modelos com centragem. Para tal consideremos a Tabela 4.6 constituída pelos valores mais elevados das correlações (em valor absoluto) entre os respetivos pares de CPs.

Tabela 4.6: Correlações entre as CPs da SVD da SSA sem centragem (SC) e as CPs da SVD de cada um dos quatro tipos de SSA com centragem em análise

SC vs C.Colunas		SC vs C.Linhas		SC vs C.Dupla		SC vs C.Serie.Orig	
(V_1, V_1^{cc})	0.9999	—	—	—	—	(V_1, V_1')	0.9999
(V_2, V_2^{cc})	0.9914	(V_2, V_2^{cl})	0.8495	(V_2, V_1^{**})	0.9959	(V_2, V_3')	0.8769
(V_3, V_3^{cc})	0.9935	(V_3, V_3^{cl})	0.8475	(V_3, V_2^{**})	0.9964	(V_3, V_4')	0.8809
(V_4, V_4^{cc})	0.9988	(V_4, V_4^{cl})	0.9970	(V_4, V_3^{**})	0.9990	(V_4, V_5')	0.9971
(V_5, V_5^{cc})	0.9994	(V_5, V_5^{cl})	0.9986	(V_5, V_4^{**})	0.9995	(V_5, V_6')	0.9988
(V_6, V_6^{cc})	0.9909	(V_6, V_6^{cl})	0.9906	(V_6, V_5^{**})	0.9906	(V_6, V_7')	0.9898
(V_7, V_7^{cc})	0.9744	(V_7, V_7^{cl})	0.9597	(V_7, V_6^{**})	0.9717	(V_7, V_8')	0.9599
(V_8, V_{12}^{cc})	0.6044	(V_8, V_1^{cl})	0.4667	—	—	(V_8, V_2')	0.4971
(V_9, V_8^{cc})	0.8527	(V_9, V_8^{cl})	0.9896	(V_9, V_7^{**})	0.8340	(V_9, V_9')	0.9862
(V_{10}, V_9^{cc})	0.7744	(V_{10}, V_9^{cl})	0.9352	(V_{10}, V_8^{**})	0.7622	(V_{10}, V_{10}')	0.9361
(V_{11}, V_{10}^{cc})	0.9899	(V_{11}, V_{10}^{cl})	0.9997	(V_{11}, V_9^{**})	0.9076	(V_{11}, V_{11}')	0.9998
(V_{12}, V_{11}^{cc})	0.9943	(V_{12}, V_{11}^{cl})	0.99997	(V_{12}, V_{11}^{**})	0.7645	(V_{12}, V_{12}')	0.99995
(V_{13}, V_{13}^{cc})	0.9926	(V_{13}, V_{12}^{cl})	0.8427	(V_{13}, V_{12}^{**})	0.7990	(V_{13}, V_{13}')	0.9881
(V_{14}, V_{14}^{cc})	0.99996	(V_{14}, V_{13}^{cl})	0.7808	(V_{14}, V_{13}^{**})	0.7623	(V_{14}, V_{14}')	0.9864
(V_{15}, V_{16}^{cc})	0.8016	(V_{15}, V_{14}^{cl})	0.7423	(V_{15}, V_{14}^{**})	0.6941	(V_{15}, V_{16}')	0.8726
(V_{16}, V_{15}^{cc})	0.8214	(V_{16}, V_{15}^{cl})	0.8483	(V_{16}, V_{15}^{**})	0.8365	(V_{16}, V_{15}')	0.8758

De um modo geral as CPs quando comparadas visualmente (ver Figura 4.18) aos pares na mesma posição, nem sempre evidenciam uma forte relação em termos das suas coordenadas entre o modelo sem centragem e cada um dos modelos com centragem, apesar de algumas CPs nas mesmas posições apresentarem correlações próximas da unidade (em módulo). No entanto é possível estabelecer, mesmo até para posições de ordens diferentes das CPs entre modelos, correlações elevadas praticamente em todas as CPs do modelo sem centragem relativamente a alguma CP consoante o outro modelo em comparação. Constata-se que a primeira CP do modelo sem centragem (que representa a tendência dos dados) não apresenta relações fortes com qualquer uma das CPs dos modelos onde ela não aparece (centragem simples por linhas e centragem dupla).

Períodos

No que se refere ao estudo no domínio da frequência, foram analisados os espectros das coordenadas dos sete pares de CPs oscilatórias retidas (anteriormente mencionados) e concluiu-se que os períodos dominantes dos pares em causa são aproximadamente iguais a 12, 4, 4, 20, 4, 6 e 33 meses, respetivamente, para todos os cinco modelos. Para o último par de CPs de cada modelo, para além dos 33 meses ainda é geralmente de igual modo dominante o período de aproximadamente 8 meses.

Considerações Finais

Para este conjunto de dados, em que se efetuou uma comparação transversal abrangendo os cinco tipos de modelos analisados (SSA sem centragem, SSA com centragem simples por colunas, SSA com centragem simples por linhas, SSA com centragem dupla e SSA com centragem na série original), podemos enumerar as seguintes constatações.

- Mantendo o mesmo comprimento da janela ($m = 48$) foram selecionadas 16 das 48 CPs, exceto nos modelos da centragem simples por linhas e na centragem dupla em que foram escolhidas 15 CPs, uma vez que a componente que descrevia nos outros modelos a tendência dos dados não surge nestes, o que comprova o senso comum, de que a centragem por linhas geralmente remove a tendência aos dados.
- Uma vez que a série temporal em estudo apresenta uma tendência, no geral semelhante a

uma tendência linear, é esperado que a SSA com dupla centragem, segundo Golyandina *et al.* (2001), produza melhores resultados na eliminação dessa componente linear. Efetivamente, ao comparar os cinco modelos, o modelo da centragem dupla evidencia ser o mais eficiente a eliminar o efeito de tendência dos dados, uma vez que exclui a componente de tendência, assim como componentes que apresentem uma mistura de oscilação com tendência como componentes dominantes. Contudo, de uma forma geral, a eliminação da tendência não revelou grande utilidade, uma vez que se conseguiu retirar a mesma informação relativamente às oscilatórias com base nos outros modelos.

- A presença de uma CP que representa uma mistura de tendência com componente oscilatória verificou-se em todos os modelos, embora a sua relevância varie consoante o modelo.
- Para a série temporal em estudo, o modelo da SSA com centragem na série original, capta de uma forma mais proeminente em relação aos restantes, as CPs envolventes com comportamentos de tendência ao exibir como CP dominante a evidente tendência geral dos dados, seguida de uma CP que descreve um comportamento oscilatório misturado com tendência.
- Em relação aos valores próprios constatámos que estes apresentam geralmente valores da mesma ordem de grandeza para as mesmas posições entre modelos, exceto para o caso do primeiro valor próprio.
- Embora seja pouco provável que qualquer uma das matrizes \mathbf{X} ou \mathbf{X}^{cc} tenha um vetor próprio que seja exatamente colinear com o vetor $\bar{\mathbf{x}}_c$ das médias das colunas da matriz não centrada \mathbf{X} , contudo em muitos casos como este, o primeiro vetor próprio é próximo da direção de $\bar{\mathbf{x}}_c$, a qual une a origem e o centro de gravidade da nuvem das k linhas de \mathbf{X} , uma vez que a quantidade $k \|\bar{\mathbf{x}}_c\|^2$ quando comparada com o $tr(\mathbf{X})$ em termos proporcionais é muito grande.
- Em todos os modelos foram captados os sete pares mais relevantes de CPs oscilatórias entre as primeiras 16 CPs, cujos sete períodos dominantes são aproximadamente iguais a 12, 4, 4, 20, 4, 6 e 33 meses respetivamente.

- Foi possível comparar os coeficientes dos vetores próprios através dos cossenos dos seus ângulos entre o modelo básico de SSA (sem centragem) e cada um dos restantes quatro modelos. Os cossenos exibiram valores geralmente elevados (em módulo) para pares de vetores associados a CPs que apresentavam um comportamento visual similar entre modelos.
- Quanto à forma de aferir o nível de comparação de CPs entre modelos, utilizaram-se as correlações entre os pares de CPs que apresentavam um comportamento visual similar entre modelos. A grande maioria dos pares de CPs identificadas com comportamentos similares (nomeadamente as oscilatórias) entre o modelo de SSA básica e cada um dos restantes quatro modelos, apresentaram correlações próximas da unidade (em módulo). Tal facto vem consolidar as evidências de que todos os modelos exibem dentre as componentes relevantes, componentes oscilatórias associadas aos mesmos períodos dominantes.

4.11 Resumo e Conclusões

A SSA surge como uma técnica que permite expandir uma série temporal com o intuito de analisar o seu comportamento através da informação retirada da matriz dos desfasamentos de forma análoga como na ACP usual. A grande finalidade da SSA consiste na decomposição da série original em componentes que possam ser interpretadas como tendências, oscilatórias ou ruído.

Após a descrição dos passos essenciais à implementação do modelo básico de SSA, foi abordado um dos aspetos mais importantes a ter em consideração que se prende com a escolha do tamanho adequado da janela a utilizar no primeiro passo. É imprescindível que se efetue uma escolha sensata do comprimento da janela mediante o tamanho da série original que se pretende analisar. Foram expostos alguns princípios orientadores e algumas questões relevantes acerca do assunto.

Para além da parte exploratória com o auxílio da decomposição espectral para analisar as componentes predominantes indicadoras de maior variabilidade nos dados, a SSA também pode ser usada para a previsão de novos valores da série temporal. Foram apresentadas as condições ideais de aplicação da previsão, assim como um algoritmo de previsão recorrente e intervalos de predição.

Foram introduzidas noções básicas sobre a teoria de ondas que permitem caraterizar alguns pares de elementos na SVD, que ostentem oscilações e picos espectrais idênticos.

Um assunto de especial interesse e que pode suscitar discussão prende-se com a centragem ou não dos dados na SSA. Como alternativas ao modelo básico de SSA sem centragem, foram descritas abordagens que contemplam centragens na matriz aumentada dos dados desfasados, assim como a centragem na série original, para a qual se introduziram novos aspetos teóricos tendo por base os resultados dos outros tipos de centragem.

Como forma de efetuar uma análise que englobasse os conceitos essenciais introduzidos ao longo do Capítulo, discutiu-se um caso prático em que utilizámos a série mensal do número de vítimas mortais em acidentes de viação em Portugal Continental entre 1998 e 2009. Como resultados principais, para além do usual ciclo anual presente nos dados, detetou-se o ciclo de 4 meses cujos picos evidenciam uma relação com as férias escolares dos portugueses, nas quais aumenta a circulação automóvel, registando-se tendencialmente um maior número de acidentes envolvendo vítimas mortais. Também o período presente de 33 meses pressupõe uma ligação às alterações à legislação em vigor e às campanhas de sensibilização introduzidas na última década. Os resultados obtidos mostram como a SSA pode desempenhar um papel fundamental no planeamento atempado para a introdução de novas políticas que se julguem mais eficazes no combate à mortalidade rodoviária.

De um modo geral, relativamente aos resultados apresentados, a SSA sem centragem possui mais semelhanças com a SSA com centragem simples por colunas e com a SSA com centragem na série original. Por um lado o tamanho da série (144 valores) pode ser considerado como suficiente para que componentes de comportamentos similares sejam extraídas através destes modelos, mas por outro lado, se admitirmos que a série possui uma dimensão relativamente pequena então é natural que a SSA com centragem dupla se destaque como o modelo que melhor elimina ou atenua a tendência que os dados apresentam.

Geralmente, nestes casos, em que a SVD ostenta comportamentos da tendência nas primeiras componentes, a reconstrução da série original apenas à custa dessas mesmas CPs leva à obtenção de uma suavização dessa tendência.

Não é de estranhar as semelhanças existentes em termos de informação dominante entre a SSA sem centragem, a SSA com centragem simples por colunas e a SSA com centragem na série

original, uma vez que de acordo com a Secção 4.8, o desfasamento por colunas é análogo aquando da construção da matriz de trajetória (sem centragem) \mathbf{X} , da matriz de trajetória centrada por colunas \mathbf{X}^{cc} e da matriz de trajetória correspondente à centragem na série original \mathbf{X}' , e como tal as colunas destas matrizes possuem exatamente as mesmas variâncias. Tal acontece dado que a variância de uma variável centrada é igual à variância da mesma variável não centrada, ou seja, para qualquer variável x_j de média \bar{x}_j , vem $var(x_j - \bar{x}_j) = var(x_j)$.

Capítulo 5

Análise Espectral Singular Multicanal

"Climate is what we expect but weather is what we get."

Mark Twain - Escritor Norte Americano

5.1 Introdução

A ACP é uma técnica útil para reduzir a dimensionalidade dos dados e para representar em larga escala padrões que expliquem a maior parte da variabilidade dos mesmos. Uma possível escolha da matriz inicial de dados é aquela cujas colunas (variáveis) correspondem a diferentes localizações espaciais e cujas linhas (indivíduos) estão associadas a diferentes momentos temporais. A ACP utiliza a informação espacial ao nível da correlação existente entre as variáveis, mas no entanto, a correlação que existe ao longo do tempo não é tida em conta. As autocorrelações e correlações cruzadas no tempo podem ser muito úteis em questões relacionadas com previsões e com a construção de modelos probabilísticos de séries temporais. Em dados que apresentem uma estrutura temporal, como sejam os de clima, pode-se considerar a técnica da Análise Espectral Singular Multicanal (MSSA) ou da Análise de Funções Ortogonais Empíricas Estendidas (EEOF) que é a extensão da Análise de Funções Ortogonais Empíricas (EOF). Portanto, a MSSA é uma extensão da Análise Espectral Singular (SSA), quando o estudo passa a envolver mais do que uma série temporal em simultâneo.

Para além dos principais objetivos já descritos na SSA, a MSSA pretende ainda identificar padrões no espaço e no tempo, a partir de uma sequência de vários mapas ao longo de uma escala temporal, assim como extrair a informação temporal conjunta das inter-relações das variáveis e das inter-relações destas desfasadas no tempo.

Ao longo deste Capítulo apresenta-se uma descrição da evolução histórica da técnica da MSSA. Serão expostos os principais conceitos teóricos, incluindo indicadores comparativos do desempenho do modelo, e explorar-se-á o conjunto de valores de pressão ao nível do mar, no norte do Oceano Pacífico em 8 estações meteorológicas, introduzido na Secção 3.12.

5.2 Descrição da MSSA

Analogamente à SSA, a técnica da MSSA subdivide-se em duas etapas fundamentais designadas por decomposição e reconstrução. Consideremos um determinado número p (> 1) de séries temporais (também designadas por canais) formadas por n pontos igualmente espaçados ao longo do tempo.

Consideremos a matriz das observações representada por

$$[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

onde $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tp})^T$, com $t = 1, 2, \dots, n$, representa cada vetor no instante de tempo t (Hannachi *et al.*, 2007).

5.2.1 Decomposição

Tal como na SSA, o principal objetivo da MSSA consiste em decompor séries originais num dado número de componentes que sejam interpretáveis, desprezando aquelas que não contenham informação relevante para o estudo.

Esta etapa é constituída por dois passos: a incorporação e a decomposição em valores singulares.

1º Passo - Incorporação (*Embedding*):

A matriz dos dados iniciais, do tipo $n \times p$, vai ser transformada numa matriz de novas dimensões $k \times p'$, onde $k = n - m + 1$ e $p' = mp$, para um determinado desfasamento ou comprimento da janela m ($1 < m < n$).

O vetor aleatório $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tp})$ com $t = 1, 2, \dots, n$, usado na ACP tradicional, é estendido de forma similar à SSA, para incluir a informação temporal como

$$\mathbf{x}_t = (x_{t1}, \dots, x_{t+m-1,1}, x_{t2}, \dots, x_{t+m-1,2}, \dots, x_{tp}, \dots, x_{t+m-1,p}) \quad (5.1)$$

em que o instante t sem perda de generalidade passa a ser $t = 1, 2, \dots, k$.

Se para cada série inicial s ($s = 1, 2, \dots, p$) descrita à custa de $n - m + 1$ vetores desfasados denotarmos por

$$\mathbf{x}_t^s = (x_{ts}, x_{t+1,s}, \dots, x_{t+m-1,s}),$$

então o vetor aleatório estendido (5.1) reescreve-se de forma similar por

$$\mathbf{x}_t = (\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^p).$$

Definição 5.2.1 *A matriz de trajetória \mathbf{X} é definida como a matriz aumentada dos dados desfasados dada por*

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{n-m+1} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{x}_1^1 & \mathbf{x}_1^2 & \dots & \mathbf{x}_1^p \\ \mathbf{x}_2^1 & \mathbf{x}_2^2 & \dots & \mathbf{x}_2^p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n-m+1}^1 & \mathbf{x}_{n-m+1}^2 & \dots & \mathbf{x}_{n-m+1}^p \end{bmatrix}. \end{aligned} \quad (5.2)$$

A matriz \mathbf{X} do tipo (5.2) pode definir-se de duas formas distintas para aplicar a MSSA, pela ordenação das variáveis (colunas). Uma delas é

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{m1} & x_{12} & \dots & x_{m2} & \dots & x_{1p} & \dots & x_{mp} \\ x_{21} & \dots & x_{m+1,1} & x_{22} & \dots & x_{m+1,2} & \dots & x_{2p} & \dots & x_{m+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n-m+1,1} & \dots & x_{n1} & x_{n-m+1,2} & \dots & x_{n2} & \dots & x_{n-m+1,p} & \dots & x_{np} \end{bmatrix},$$

onde os vetores próprios de dimensão mp podem ser interpretados como p subestruturas que permitem estudar a evolução temporal dos pontos no espaço desde a ausência de desfasamento até ao desfasamento $m - 1$. Deste modo, as primeiras m colunas correspondem aos valores de cada desfasamento da primeira variável (série), as m colunas seguintes correspondem aos valores de cada desfasamento da segunda variável e assim sucessivamente até que as últimas m colunas correspondem aos valores de cada desfasamento da p -ésima variável. A outra forma de reordenar as colunas da matriz \mathbf{X} , consiste em dispor as mp colunas, tal que as primeiras p colunas correspondem ao desfasamento nulo para todas as p variáveis, as p colunas seguintes correspondem ao primeiro desfasamento para todas as p variáveis e assim sucessivamente até que as últimas p colunas correspondem ao $(m - 1)$ -ésimo desfasamento. Neste caso dizemos que os vetores próprios constituem m mapas temporais que caracterizam padrões no espaço e no tempo.

Ambas as formas de definir a matriz \mathbf{X} irão produzir resultados equivalentes, uma vez que apenas muda a ordem das colunas, o que faz com que na SVD, os coeficientes dos vetores próprios sejam reordenados. Posteriormente aquando da reconstrução, embora se venha a reter a mesma estrutura principal das séries originais, há que ter em conta essas mudanças de ordem das colunas.

2º Passo - Decomposição em Valores Singulares (SVD):

Dada a matriz \mathbf{X} do tipo $k \times mp$, constrói-se a matriz simétrica $\mathbf{X}^T \mathbf{X}$.

Denotemos por $\lambda_1, \lambda_2, \dots, \lambda_{p'}$ os p' valores próprios de $\mathbf{X}^T \mathbf{X}$ de modo que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p'} \geq 0$ e por $U_1, U_2, \dots, U_{p'}$ os correspondentes p' vetores próprios ortonormados de $\mathbf{X}^T \mathbf{X}$.

A SVD decompõe a matriz \mathbf{X} na expressão matricial

$$\mathbf{X} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T,$$

onde

- $\mathbf{V} = (V_1, V_2, \dots, V_{p'})$ é designada por matriz das componentes principais de \mathbf{X} ,
- $\mathbf{\Lambda} = (\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_{p'}})$ representa a matriz diagonal dos valores singulares de \mathbf{X} e
- $\mathbf{U} = (U_1, U_2, \dots, U_{p'})$ é designada por matriz dos vetores singulares (direitos) de \mathbf{X} .

Sejam $d = \max \{i : \lambda_i > 0\} \leq p'$ e as componentes principais dadas pelos vetores $V_i = \mathbf{X} U_i / \sqrt{\lambda_i}$, com $\mathbf{X}_i = \sqrt{\lambda_i} V_i U_i^T$, $i = 1, \dots, d$, então a SVD da matriz \mathbf{X} pode ser escrita por

(Golyandina e Stepanov, 2005)

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d. \quad (5.3)$$

Paralelismo da MSSA com a ACP

Na SVD da matriz \mathbf{X} descrita anteriormente, é costume obterem-se os valores e vetores próprios a partir da matriz $\mathbf{X}^T \mathbf{X}$.

No entanto, podemos aplicar a ACP usual sobre a matriz dos dados desfasados, a partir da matriz correspondente $\frac{1}{k} \mathbf{X}^T \mathbf{X}$, cuja decomposição em valores e vetores próprios é dada por $\mathbf{P} \mathbf{L} \mathbf{P}^T$, onde \mathbf{P} representa a matriz cujas colunas são os p' vetores próprios ortogonais e \mathbf{L} a matriz diagonal cujos elementos da diagonal principal são os p' valores próprios.

Na MSSA, os p' valores próprios são k vezes mais elevados relativamente aos da ACP usual. Há ainda a ressaltar que na MSSA, não se efetua à priori a centragem por colunas na matriz $\mathbf{X}^T \mathbf{X}$, mas em contrapartida a centragem (por colunas) está implícita na ACP usual.

Definição 5.2.2 *Admitindo que a matriz \mathbf{X} de (5.2) foi previamente centrada (por colunas), então a matriz $p' \times p'$ de variâncias-covariâncias dos desfasamentos de \mathbf{X} define-se por*

$$\mathbf{S}' = \frac{1}{k} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \dots & \mathbf{S}_{1p} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \dots & \mathbf{S}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{p1} & \mathbf{S}_{p2} & \dots & \mathbf{S}_{pp} \end{bmatrix}, \quad (5.4)$$

onde cada bloco $\mathbf{S}_{ij} = \mathbf{S}_{ji}^T$, com $i, j = 1, 2, \dots, p$, é uma matriz $m \times m$ de variâncias-covariâncias desfasada entre a localização i e a localização j até ao desfasamento $m - 1$, dada por

$$\mathbf{S}_{ij} = \frac{1}{k} \sum_{t=1}^k \mathbf{x}_t^{iT} \mathbf{x}_t^j.$$

Os elementos de cada matriz \mathbf{S}_{ij} de variâncias-covariâncias dos desfasamentos entre duas variáveis i e j podem ser estimados através do estimador de Broomhead & King ou através do estimador de Vautard & Ghil, conforme o desenvolvimento na Secção 4.5.

Se as séries forem estacionárias então cada bloco \mathbf{S}_{ij} tem aproximadamente a estrutura de uma matriz simétrica de Toeplitz, para amostras de tamanho elevado. A matriz simétrica \mathbf{S}' não é em geral de Toeplitz, uma vez que diferentes blocos representam covariâncias entre diferentes pares de localizações dos pontos (Hannachi *et al.*, 2007).

5.2.2 Reconstrução

Após a decomposição, e a partir dos vetores próprios e das respectivas componentes principais, efetua-se a reconstrução das p séries originais, para que se possam comparar as séries reconstruídas, através de um determinado número adequado de componentes, com as séries originais.

3º Passo - Agrupamento (*Grouping*):

Por analogia com SSA, dada a expressão (5.3), procede-se à fase do agrupamento.

Seja $I = \{i_1, \dots, i_p\}$, um grupo de índices i_1, \dots, i_p ; para cada grupo I corresponde a matriz \mathbf{X}_I definida por $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. A partição do conjunto de índices $\{1, \dots, d\}$ em r subconjuntos disjuntos I_1, I_2, \dots, I_r corresponde a

$$\mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_r}. \quad (5.5)$$

4º Passo - Média ao Longo das Diagonais (*Diagonal Averaging*):

Neste último passo, transforma-se cada matriz \mathbf{X}_I da decomposição (5.5) num sistema de séries reconstruídas de comprimento n , através de um processo que, em vez de transformar a matriz arbitrária numa matriz de Hankel (como na SSA), transforma neste caso numa matriz constituída por p blocos de submatrizes de Hankel $k \times m$, e posteriormente em p séries.

Na reconstrução de novas séries, os valores obtêm-se através da operação que corresponde a efetuar a média dos valores de cada matriz \mathbf{X}_I ao longo das diagonais $i + j = c$ (para um valor fixo de c) de cada bloco. Estas médias correspondem à transferência dos valores da matriz por blocos para cada série reconstruída que é definida em cada bloco de forma análoga a (4.4) na SSA.

Consideremos a l -ésima Componente Principal da matriz \mathbf{X} dada por

$$V_l = (v_l(1), v_l(2), \dots, v_l(k))^T, \text{ para } l = 1, 2, \dots, p'.$$

Para remover a contribuição de componentes pouco significativas, com vista à reconstrução, podemos utilizar as CPs como forma de filtragem dos dados. A partir da decomposição em valores singulares de \mathbf{X} , cada linha \mathbf{x}_t da matriz \mathbf{X} de (5.2) representa-se por

$$\mathbf{x}_t^T = \sum_{l=1}^{p'} \sqrt{\lambda_l} v_l(t) U_l, \text{ para } t = 1, 2, \dots, k. \quad (5.6)$$

Uma alternativa para reescrever a matriz dos desfasamentos \mathbf{X} é denotar cada linha sob a forma do vetor (5.1) por

$$\mathbf{x}_t = (\mathbf{x}_t, \mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+m-1}),$$

onde $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tp})$, com $t = 1, 2, \dots, k$. Como tal, e de acordo com as variáveis originais \mathbf{x}_t , a partir da expressão (5.6) obtém-se

$$\mathbf{x}_{t+j-1}^T = \sum_{l=1}^{p'} \sqrt{\lambda_l} v_l(t) U_l^j, \quad (5.7)$$

onde $U_l^j = (u_{j,l}, u_{j+m,l}, \dots, u_{(p-1)m,l})^T$, para $j = 1, 2, \dots, m$.

Numa grande parte das aplicações, existe essencialmente o interesse em truncar a construção (5.7), retendo apenas um conjunto Cq com um número pequeno q de CPs com vista à filtragem, como por exemplo na reconstrução de um par de CPs que correspondam a uma determinada oscilação. Em tais situações, a expressão (5.7) é truncada para um conjunto Cq (com $q < p'$) por

$$\mathbf{y}_{t+j-1}^T = \sum_{l \in m, q} \sqrt{\lambda_l} v_l(t) U_l^j,$$

onde $\mathbf{y}_t = (y_{t1}, y_{t2}, \dots, y_{tp})$ é o vetor reconstruído ou filtrado.

As variáveis reconstruídas a partir de um subconjunto de q CPs são obtidas por

$$\mathbf{y}_t^T = \begin{cases} \frac{1}{t} \sum_{j=1}^t \sum_{i \in Cq} \sqrt{\lambda_l} v_l(t-j+1) U_l^j & , \text{ para } 1 \leq t \leq m-1 \\ \frac{1}{m} \sum_{j=1}^m \sum_{i \in Cq} \sqrt{\lambda_l} v_l(t-j+1) U_l^j & , \text{ para } m \leq t \leq k \\ \frac{1}{n-t+1} \sum_{j=t-k+1}^m \sum_{i \in Cq} \sqrt{\lambda_l} v_l(t-j+1) U_l^j & , \text{ para } k+1 \leq t \leq n \end{cases}.$$

Sendo $\mathbf{Z}_q = \mathbf{X} \mathbf{P}_q$ a matriz de q CPs, então a reconstrução de \mathbf{X} à custa de apenas q CPs é dada por $\tilde{\mathbf{X}} = \mathbf{Z}_q \mathbf{P}_q^T$. Cada série é agora também reconstruída com base no procedimento de cálculo da média ao longo das diagonais, efetuado de forma idêntica à SSA para uma série.

No contexto de dados com estrutura espacial, como seja em climatologia, por vezes o número de pontos p da grelha espacial é demasiado elevado para garantir uma adequada decomposição espectral da matriz de variâncias-covariâncias de dimensão pm . Perante tal situação Kimoto *et al.* (1991) e Plaut e Vautard (1994) recomendam proceder-se à compressão prévia da informação espacial, a qual pode ser efetuada através do cálculo de uma média espacial ou através de uma pré-filtragem recorrendo apenas a algumas CPs dominantes obtidas a partir de uma ACP, seguida

da MSSA nesse subconjunto de CPs. Estes e outros aspetos adicionais acerca da reconstrução das séries originais podem ser encontrados, por exemplo, em Hannachi *et al.* (2007).

5.2.3 Comprimento da Janela

No que diz respeito à escolha do comprimento da janela, face às recomendações introduzidas em SSA (Secção 4.3) e que continuam a ser admissíveis em MSSA, são aconselhados valores suficientemente grandes para que seja possível distinguir entre dois picos espectrais relativamente próximos, visto não ser possível distinguir oscilações de período maior que o comprimento da janela (Plaut e Vautard, 1994). Na MSSA, segundo Ghil *et al.* (2002), deve escolher-se um comprimento da janela m maior ou igual que o número p de séries.

5.2.4 Algumas Propriedades

Apresentamos de seguida algumas propriedades relacionadas com a técnica da MSSA.

Considerem-se:

- dois valores próprios λ_1 e λ_2 consecutivos tais que $\lambda_1 \simeq \lambda_2$;
- duas correspondentes sequências temporais descritas pelos vetores próprios U^1 e U^2 de dimensão mp , aproximadamente periódicas com o mesmo período T e em quadratura e
- duas CPs V^1 e V^2 associadas a λ_1 e λ_2 , também em quadratura.

Então, nestas condições, e de acordo com Plaut e Vautard (1994), existe nas p séries uma oscilação em que o período e o padrão espacial são os mesmos que os dos vetores próprios U^1 e U^2 .

Consideremos alguns conceitos sobre a teoria de ondas, introduzidos no Capítulo da SSA. Relembramos que, de acordo com a ACP e com as notações da Subsecção 5.2.1, para a matriz dos desfasamentos dos dados, \mathbf{X} , cada elemento $X_{s,i}$ pode ser descrito à custa das componentes principais V^l e dos respetivos vetores próprios U^l . Ao considerar o sinal discretizado no tempo por $t_i = i\Delta t$ e no espaço por $y_s = s\Delta y$, ($s = 1, \dots, p$), adaptando a notação para MSSA, com $p' = mp$, vem a expressão

$$X_{s,i} = \sum_{l=1}^{p'} V_i^l \left(U_{s,j}^l \right)^T = W(y_s, t_i),$$

que é satisfeita quando se consideram as seguintes equações para duas CPs e para $j = 1, 2, \dots, m$

(Plaut e Vautard, 1994).

$$\begin{aligned}
 U_{s,j}^1 &= F_1 \cos(\mu s \Delta y - \omega j \Delta t + \phi), \\
 U_{s,j}^2 &= F_2 \sin(\mu s \Delta y - \omega j \Delta t + \phi), \\
 V_i^1 &= \frac{A}{F_1} \cos(\omega i \Delta t + \phi), \\
 V_i^2 &= \frac{A}{F_2} \sin(\omega i \Delta t + \phi), \\
 \phi &= \frac{\mu(p+1) \Delta y - w(m+1) \Delta t}{2}.
 \end{aligned}$$

Por analogia com a Secção 4.6, os vetores próprios U^1 e U^2 são ortogonais em fase e em quadratura. Ainda de forma análoga, para um desfasamento nulo, a covariância entre duas CPs anula-se, permanecendo em fase e em quadratura (consultar Apêndices C.2 e C.3). Portanto, quando o sinal é uma soma finita de q ($\leq p'$) componentes oscilatórias de sinal próximo do sinal periódico, o número de valores próprios não nulos é $2q$ e como tal, os vetores próprios associados surgem aos pares.

Ainda segundo Plaut e Vautard (1994), na MSSA, a separação de duas componentes atua segundo o espaço e segundo o tempo, isto é, distinguem-se duas oscilações com o mesmo padrão espacial mas com diferentes períodos assim como oscilações com o mesmo período e padrões ortogonais ao nível espacial.

Um exemplo da vantagem da MSSA face a outros métodos espectrais é a sua capacidade de distinguir vários padrões espaciais na mesma frequência, pois se uma oscilação estiver sobreposta com algum ruído e possuir alguma força em torno da frequência dessa oscilação, então distingue-se a parte correspondente à oscilação e a parte correspondente ao ruído, mediante a análise do espectro.

Para o caso multivariado em que se podem encontrar várias séries detentoras de periodicidades e picos de frequências em comum, apenas uma parte da variância no respetivo pico de frequência pode ser explicada para cada par oscilatório.

5.3 Estado da Arte

Ao nível do clima, a MSSA é considerada vantajosa pelo facto dos campos geofísicos apresentarem geralmente correlações significativas tanto no espaço como ao nível do tempo através

das autocorrelações e correlações cruzadas como mencionado na primeira abordagem à técnica em Weare e Nasstrom (1982). Também de acordo com estes autores, uma outra vantagem prende-se com o facto das componentes dominantes parecerem muitas vezes interpretáveis não só ao nível dos modos¹ dominantes de variabilidade mas também em termos de identificação dos modos de sequências de acontecimentos no espaço e no tempo.

Em 1991, Kimoto *et al.* usaram a mesma metodologia, que viria a ser desenvolvida posteriormente por Plaut e Vautard, direccionada para a análise de dados associados com a oscilação extratropical de 30-60 dias e as suas ligações à circulação tropical.

Na MSSA, sucessivos padrões (ou subestruturas) de uma mesma variável numa CP são interpretados como a propagação ou evolução temporal do primeiro padrão presente nessa mesma CP (Chen e Harr, 1993), e tais subestruturas subjacentes a cada uma das CPs podem representar os desfasamentos no tempo. No entanto argumentam que são necessárias precauções na interpretação das inter-relações entre as CPs, uma vez que estas são obtidas com vista à otimização da variância de cada CP, mas não têm em conta as correlações entre as subestruturas em cada uma das CPs nem entre outras subestruturas noutras CPs.

Segundo Plaut e Vautard (1994), a principal propriedade a destacar da MSSA é permitir detetar comportamentos oscilatórios quer no tempo quer ainda ao nível de padrões espaciais oscilatórios. Com o objetivo de isolar componentes quase periódicas que fazem parte do sistema atmosférico, Plaut e Vautard (1994) desenvolveram e aplicaram a MSSA a um conjunto extenso de dados correspondentes a 32 anos de medições da altura geopotencial² a 700 milibar, cobrindo a região do Hemisfério Norte acima do trópico de Câncer. O estudo incidiu sobre oscilações de baixa frequência e as suas relações, nomeadamente as oscilações intermensais. Para além das oscilações anuais e sazonais usuais, a MSSA revelou a existência de uma oscilação forte, no Atlântico, com um período de 70 dias cujo padrão está essencialmente relacionado com o padrão da Oscilação do Atlântico Norte³ (NAO) descrito em Wallace e Gutzler (1981). Esta oscilação

¹No contexto geofísico, modo é um padrão espacial recorrente bastante acentuado. Geralmente os modos são ortogonais entre si.

²A altura geopotencial é uma variável meteorológica que representa a altitude acima do nível do mar em que se encontra um determinado nível de pressão. Os valores de pressão mais utilizados no geopotencial são 1000, 850, 500 e 200 *hPa*.

³A Oscilação do Atlântico Norte (NAO) é o modo dominante de variabilidade da pressão ao nível do mar no Norte do Oceano Atlântico. O seu padrão espacial consiste num dipolo com orientação aproximada

propaga-se em direção aos polos com uma componente persistente sobre a Sibéria. Foram ainda detetadas duas oscilações diferentes no intervalo de 30 a 60 dias no Pacífico e no Atlântico.

Em 2000, van den Dool *et al.* propuseram uma nova variante para calcular funções empíricas e ortogonais, em vez dos usuais vetores próprios da ACP, conhecidos na área do clima como as tradicionais funções ortogonais empíricas (EOFs), a partir de um conjunto de dados no espaço e no tempo. O método relacionado com a regressão linear múltipla permite soluções que sejam ortogonais apenas numa das duas direções: ou espaço ou tempo. Este consiste em procurar saber qual o ponto no espaço (ponto base encarado como preditor) que, através da regressão linear, explica o máximo da variância de todos os outros pontos combinados (os preditores). O primeiro padrão espacial é descrito pelos coeficientes da regressão entre o ponto base e os restantes pontos, e a primeira série temporal é considerada a série dos dados originais no ponto base. Os dados originais são de seguida reduzidos, ou seja, remove-se o que é explicado no primeiro passo (ou modo 1) e volta-se a procurar um novo ponto no espaço que maximize a variância para voltar a reduzir os dados, repetindo-se o processo um determinado número de vezes. Estas novas funções são designadas por teleconexões ortogonais empíricas (EOTs), as quais desempenham um papel semelhante às teleconexões de Wallace e Gutzler (1981). Algumas vantagens da utilização das EOTs são defendidas por van den Dool *et al.* (2000). Os mesmos autores propuseram ainda algumas possíveis aplicações das EOTs, tais como: modelos numéricos de previsão; outros níveis como por exemplo o espaço a ser percorrido na horizontal e na vertical, e adição de variáveis no estudo; um tipo especial de Análise das Correlações Canónicas; análise retroativa do clima; sistema de observação do clima.

Ghil *et al.* (2002) publicaram um trabalho sobre os avanços da teoria de sistemas dinâmicos, no qual efetuaram uma descrição das relações entre a análise de séries temporais e os sistemas dinâmicos não lineares. Como exemplo, a MSSA foi aplicada a um conjunto de dados de temperaturas mensais da superfície do mar, medidos ao longo de 45 anos, acerca dos quais foram Norte/Sul, em que um dos centros se localiza sobre a Islândia e o outro, de sinal oposto, estendendo-se pelas latitudes médias do Atlântico Norte, centra-se a cerca dos 35°N, na zona dos Açores.

A fase positiva da NAO regista-se quando uma depressão intensa na Islândia tende a ocorrer com um anticiclone forte nos Açores, o que leva à ocorrência de tempestades mais frequentes e mais intensas na região do Atlântico Norte. Em contrapartida, a fase negativa da NAO regista-se quando uma depressão fraca na Islândia tende a ocorrer com um anticiclone fraco nos Açores, o que por sua vez leva à ocorrência de tempestades menos frequentes e menos intensas ao nível do Atlântico Norte.

contemplados estudos em comparação a outros relacionados com o fenómeno ENSO.

Golyandina e Stepanov (2005) aplicaram a MSSA e a Análise Espectral Singular Complexa (CSSA) no estudo da reconstrução e previsão de sinais multivariados na presença de ruído. Algumas abordagens e características inerentes às técnicas foram comparadas através de métodos de simulação.

Uma revisão acerca da ACP e de outras técnicas relacionadas, com grande utilização nas ciências atmosféricas, foi efetuada por Hannachi *et al.* (2007) com vista a fornecer uma atualização relativamente a novos desenvolvimentos e a recentes abordagens ainda pouco exploradas, incluindo a MSSA. Para além da descrição de alguns conceitos teóricos, foram usadas várias aplicações, baseadas em diferentes conjuntos de dados, para ilustrar a implementação das técnicas apresentadas.

Groth e Ghil (2011) utilizaram a MSSA como contributo para o estudo da sincronização de fase num sistema de osciladores na presença de elevados níveis de ruído observados. Mostraram que a técnica pode identificar automaticamente múltiplos modos oscilatórios, assim como detectar se esses modos são partilhados por grupos de osciladores bloqueados por fase e frequência. Na MSSA foi introduzida uma modificação da rotação dos vetores próprios, com base na maximização da variância, para identificar o grupo ótimo de osciladores sincronizados.

Patterson *et al.* (2011) aplicaram a MSSA num estudo de dados económicos sobre o índice de produção industrial do Reino Unido, os quais foram sujeitos a vários processos de revisão com base em múltiplas recolhas. As previsões subsequentes evidenciaram uma maior precisão em relação às recolhas preliminares e a outros modelos paramétricos de previsão convencionais. Nos trabalhos de Hassani *et al.* (2012) também foram usadas oito séries de produção industrial do Reino Unido com o objetivo de melhorar as previsões dos índices de produção industrial mensais através da MSSA. Os resultados foram comparados com os obtidos através de modelos autorregressivos integrados de médias móveis e modelos autorregressivos vetoriais.

Outras análises mais genéricas foram sugeridas na literatura por forma a contemplarem diferentes variáveis e diferentes desfasamentos temporais (Barnett e Hasselmann, 1979, Preisendorfer, 1988). Uma dessas análises foi apresentada por Mote *et al.* (2000) e designada por EEOF multivariada (MEEOF).

Uma outra variação designada por *Principal Sequence Pattern Analysis* (PSPA), foi proposta por Compagnucci *et al.* (2001) em que se aplicou a ACP à transposta da matriz estendida

habitual da MSSA, ou seja, em vez das variáveis corresponderem ao usual modo espacial passam a corresponder ao modo temporal. Portanto, nesta abordagem, para a obtenção de padrões espaciais e sua evolução no tempo, as variáveis são sequências de campos espaciais de uma determinada variável meteorológica ao longo de sucessivos momentos. No artigo aplicou-se a técnica a uma amostra de valores diários da altura geopotencial registados em 12 momentos temporais numa grelha regular de 178 pontos, cobrindo uma região que abrange a parte sul da América do Sul, a Península Antártica e mares adjacentes. Os resultados da aplicação a sequências de 5-dias demonstraram as vantagens do processo na identificação de sequências de padrão de campos espaciais. Foi também possível determinar quais as sequências mais frequentes, e portanto, deduzir quais as características que permitem que um sistema evolua para a próxima sequência, assim como quantificar a frequência de tais sequências, utilizando a variação explicada para cada sequência (componente) principal. Segundo os autores, esta técnica, quando comparada com a ACP usual, apresenta a vantagem de poder ser usada para determinar os tipos de evolução de sistemas meteorológicos, cuja contribuição é uma mais valia para o estudo da climatologia dinâmica. Como tal, é possível estudar a variabilidade de evolução desses sistemas e tendências de mudanças climáticas, produzidos por modificações resultantes da circulação a partir de mudanças a médio e longo prazo no desenvolvimento dos próprios sistemas meteorológicos.

Capítulo 6

*Lag*ACI: *Lag*-Análise em Componentes Independentes

*"Se queremos progredir, não devemos repetir a história,
mas fazer uma história nova."*

Mahatma Gandhi - Pensador Indiano

6.1 Introdução

Uma nova abordagem com base na Análise em Componentes Independentes aplicada na matriz dos dados desfasados foi introduzida em Sebastião e Oliveira (2013), como alternativa à técnica da MSSA no estudo de séries temporais estendidas.

Nesta tese propomos que esta nova abordagem seja encarada como uma nova técnica, a qual denominamos por *Lag*-Análise em Componentes Independentes, ou apenas *Lag*ACI sob a forma de acrónimo. O prefixo *Lag* pretende dar a noção de que a técnica da ACI recorre a um desfasamento no tempo (ou comprimento da janela). Na terminologia inglesa virá *Lag*ICA: *Lag*-Independent Component Analysis.

Uma vez que a ACP utiliza apenas a informação estatística relativa aos momentos de segunda ordem, condiciona a que as componentes principais sejam não correlacionadas entre si, e logo estas não são necessariamente independentes, ao contrário do que se pretende que aconteça na ACI. Em certos contextos, julga-se que a ACI é mais apropriada do que a ACP para analisar

séries temporais, dado que a extração das componentes independentes envolve estatísticas de ordem superior a dois, o que faz com que as CIs revelem mais informação útil do que as CPs usuais (Cheung e Xu, 2001).

Desta forma encontrámos motivação suficiente para, com o intuito de extrair mais informação ou pelo menos semelhante, aquando da aplicação da ACP usual na matriz dos desfasamentos, decidir aplicar a ACI em vez da ACP, criando assim esta nova abordagem *LagACI* como alternativa à MSSA.

Enquanto que na MSSA o principal objetivo é decompor séries originais num dado número de componentes que revelem a informação predominante, através do uso da ACP para extrair as componentes principais que maximizem sucessivamente a variância dos dados, na *LagACI* pretende-se decompor as séries originais de igual modo num número de componentes relevantes, mas recorrendo à implementação da ACI através de um algoritmo apropriado para extrair as componentes independentes. Em qualquer um dos casos há que desprezar as componentes que não revelem informação útil para o estudo em causa.

Uma parte desta Secção foi publicada em Sebastião e Oliveira (2013). Começamos por apresentar os fundamentos teóricos da *LagACI*, seguidos de um exemplo de aplicação.

6.2 Fundamentação Teórica

Embora a *LagACI* seja aqui descrita como uma alternativa à MSSA, ou seja, admitindo a presença de mais do que uma série temporal original, também pode ser encarada de forma mais simples como alternativa à SSA, caso se use apenas uma série temporal para construir a matriz dos desfasamentos.

As bases teóricas da técnica assentam na junção de alguns passos da MSSA com os procedimentos de implementação de um modelo de ACI.

Consideremos um determinado número p (> 1) de séries temporais constituídas por n pontos igualmente espaçados ao longo do tempo, e a matriz das observações (dados iniciais) representada

por

$$[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix},$$

onde $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tp})^T$, com $t = 1, 2, \dots, n$, representa cada vetor no instante de tempo t .

Passemos de seguida a descrever os passos para aplicar a *LagACI*.

1º Passo - Incorporação (Idêntico a MSSA):

A matriz dos dados iniciais, do tipo $n \times p$, é transformada numa matriz de novas dimensões $k \times p'$, onde $k = n - m + 1$ e $p' = mp$, para um determinado desfasamento (*lag*) ou comprimento da janela m ($1 < m < n$), o qual deve ser escolhido tendo em conta as recomendações enunciadas nas Secções 4.3 e 5.2.3.

A matriz aumentada dos dados desfasados (matriz trajetória ou matriz dos desfasamentos) \mathbf{X} é definida de modo idêntico ao da MSSA na Subsecção 5.2.1 e é dada por

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{n-m+1} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{x}_1^1 & \mathbf{x}_1^2 & \dots & \mathbf{x}_1^p \\ \mathbf{x}_2^1 & \mathbf{x}_2^2 & \dots & \mathbf{x}_2^p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n-m+1}^1 & \mathbf{x}_{n-m+1}^2 & \dots & \mathbf{x}_{n-m+1}^p \end{bmatrix}, \end{aligned}$$

onde, para $t = 1, 2, \dots, n - m + 1$,

$$\mathbf{x}_t = (\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^p),$$

e para cada série inicial s ($s = 1, 2, \dots, p$) descrita à custa dos vetores desfasados,

$$\mathbf{x}_t^s = (x_{ts}, x_{t+1,s}, \dots, x_{t+m-1,s}).$$

Esta matriz dos desfasamentos \mathbf{X} , corresponde a uma matriz $k \times p'$ dos dados observados num modelo ACI, ou seja, das p' misturas.

2º Passo - Aplicação da ACI na Matriz dos Desfasamentos:

Neste passo há que ter em conta os procedimentos de aplicação da ACI descritos de seguida.

- Aplica-se o modelo da ACI à matriz \mathbf{X} (do tipo $k \times p'$) das p' misturas (ou séries desfasadas) como descrito na Secção 3.4. Tendo em conta o modelo da ACI escrito na forma matricial por $\mathbf{X} = \mathbf{SA}$, e os pressupostos de estimação do mesmo, o objetivo é estimar a matriz \mathbf{A} (do tipo $l \times p'$) dos coeficientes das misturas, cujas colunas deverão ser linearmente independentes e a matriz \mathbf{S} (do tipo $k \times l$) das l componentes independentes.
- Estipula-se o número l de componentes independentes a estimar, que de acordo com o modelo geral da ACI pode ser igual ou diferente do número de misturas, p' (embora uma parte significativa dos algoritmos mais usuais permita extrair apenas um número de componentes independentes $l \leq p'$).
- Utiliza-se o método de estimação (Secção 3.5) que se julgue mais adequado perante a natureza dos dados e de acordo com o algoritmo mais apropriado (Secção 3.7).
- Após a extração das componentes independentes aplicam-se os diferentes métodos de pós-processamento de ordenação das CIs definidos na Secção 3.8.

3º Passo - Agrupamento (Idêntico a MSSA)

A partir da matriz das componentes independentes previamente ordenadas, \mathbf{S} , e da matriz dos coeficientes das misturas, \mathbf{A} , procede-se ao agrupamento das componentes em vários grupos tendo em conta a sua natureza de acordo com a Subsecção 5.2.2. Neste contexto a reconstrução da matriz dos desfasamentos denota-se por $\tilde{\mathbf{X}} = \mathbf{SA}$, de modo que a cada grupo I vai corresponder a uma matriz reconstruída \mathbf{X}_I .

4º Passo - Média ao Longo das Diagonais (Idêntico a MSSA)

No último passo, utiliza-se o procedimento que permite transformar cada matriz \mathbf{X}_I numa matriz constituída por p blocos de submatrizes de Hankel $k \times m$, e posteriormente reconstruir a matriz dos dados originais constituída por p séries. Os detalhes da reconstrução através da média ao longo das diagonais de cada matriz \mathbf{X}_I estão descritos na Subsecção 5.2.2.

6.3 Exemplo de Aplicação: Pressão ao Nível do Mar

6.3.1 Dados e Metodologia

Consideremos novamente o conjunto de dados de médias mensais de pressão ao nível do mar observados em 8 estações meteorológicas, no norte do Oceano Pacífico (entre 1979 e 1996), introduzido na Secção 3.12 para uma aplicação e comparação das técnicas da ACP e da ACI aplicada na matriz dos desfasamentos.

Para complementar o estudo já iniciado com as análises prévias de ACP e ACI, neste exemplo de aplicação, pretendemos introduzir a MSSA e a nova abordagem baseada na ACI aplicada na matriz dos desfasamentos - a *LagACI* - como alternativa à MSSA. De acordo com o contexto das ciências atmosféricas, cada vetor desfasado vai ser obtido a partir de uma série temporal em cada instante de tempo t , para o valor da variável atmosférica correspondente a um ponto no espaço.

A matriz dos dados originais, designada por *Pacific2*, constituída por $p = 8$ séries temporais de $n = 216$ valores mensais cada, foi transformada na matriz estendida das séries temporais iniciais $\mathbf{X}_{167 \times 400}$ ($\mathbf{X}_{\text{desfasada}}$), ao ser considerado um desfasamento (comprimento da janela) $m = 50$, próximo de $n/4$ como sugerido por Vautard *et al.* (1992). Sebastião e Oliveira (2013) aplicaram, a partir da matriz dos desfasamentos, três diferentes técnicas que se descrevem de seguida e que são ilustradas na Figura 6.1. No caso da MSSA na obtenção da matriz das variâncias-covariâncias dos desfasamentos, recorreu-se à estimação de Broomhead & King. Na ACI a escolha do FastICA deve-se apenas a uma questão de simplicidade de aplicabilidade e por este já existir num *package* do *R*.

A maior parte dos procedimentos necessários à aplicação das técnicas em causa foram implementados no *software R*, no qual elaborámos inúmeras rotinas de apoio às mesmas, incluindo novas rotinas para executar os métodos de ordenação das CIs.

Técnica I - MSSA

- Aplica-se a ACP usual na matriz dos desfasamentos.
- Retêm-se apenas as primeiras 5 CPs por apresentarem valores próprios nitidamente mais elevados quando comparados com os restantes, o que permite separar as CPs que capturam

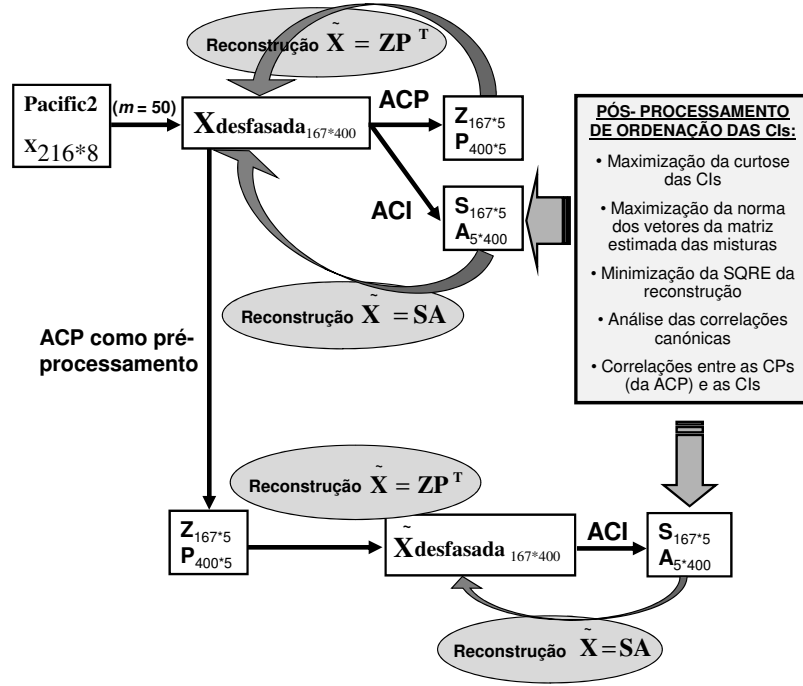


Figura 6.1: Metodologia para técnicas I, II e III com um comprimento da janela $m = 50$. \mathbf{Z} é a matriz das coordenadas das CPs, \mathbf{P} é a matriz dos vetores próprios, \mathbf{S} é a matriz das CIs e \mathbf{A} é a matriz dos coeficientes das misturas

a variabilidade dominante das restantes CPs que representam o ruído (Figura 6.2).

- A partir da matriz das coordenadas das CPs, $\mathbf{Z}_{167 \times 5}$, e da matriz dos vetores próprios, $\mathbf{P}_{400 \times 5}$, efetuam-se as reconstruções da matriz dos desfasamentos ($\tilde{\mathbf{X}} = \mathbf{ZP}^T$) e da matriz dos dados originais.
- Comparam-se as reconstruções com as séries originais.

Técnica II - LagACI

- Aplica-se a ACI através do algoritmo FastICA na matriz dos desfasamentos.
- Extraem-se apenas 5 CIs e aplicam-se os diferentes métodos de pós-processamento de ordenação das CIs definidos na Secção 3.8. O critério de extrair apenas 5 CIs prende-se com o facto de querermos comparar até que ponto é que um mesmo número de CIs

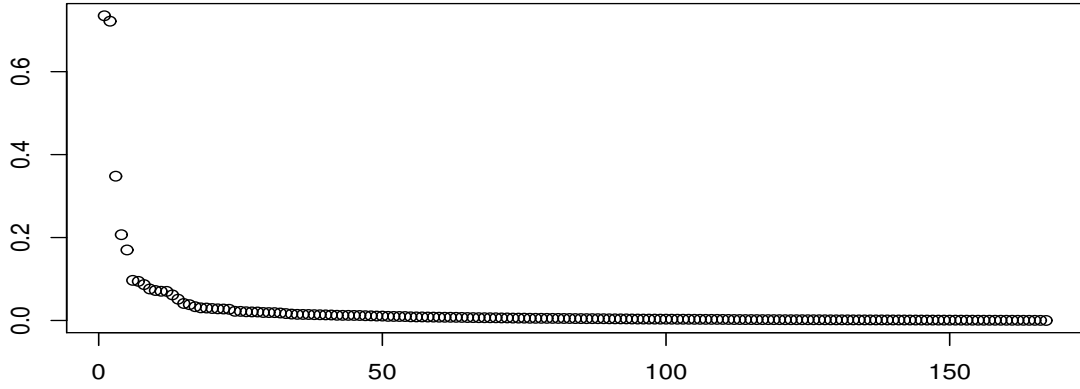


Figura 6.2: Valores próprios para a ACP na matriz dos desfasamentos com $m = 50$

extraem ou não a mesma informação quando comparada com a que é obtida pelas 5 CPs dominantes.

- A partir da matriz das coordenadas das CIs previamente ordenadas, $\mathbf{S}_{167 \times 5}$, e da matriz dos coeficientes das misturas, $\mathbf{A}_{5 \times 400}$, efetuam-se as reconstruções da matriz dos desfasamentos ($\tilde{\mathbf{X}} = \mathbf{S}\mathbf{A}$) e da matriz dos dados originais.
- Comparam-se as reconstruções com as séries originais.

Técnica III - *Lag*ACI após ACP na Matriz dos Desfasamentos

- Aplica-se a ACP usual na matriz dos desfasamentos como pré-processamento da aplicação da *Lag*ACI, de forma a reter-se a maior variabilidade possível dos dados num pequeno número de CPs, ignorando assim a maior parte do ruído.
- A partir da matriz das coordenadas das CPs, $\mathbf{Z}_{167 \times 5}$, e da matriz dos vetores próprios, $\mathbf{P}_{400 \times 5}$, das 5 primeiras CPs retidas, efetua-se a reconstrução da matriz dos desfasamentos ($\tilde{\mathbf{X}} = \mathbf{Z}\mathbf{P}^T$).
- Aplica-se a ACI através do algoritmo FastICA na matriz dos desfasamentos reconstruída

$$(\tilde{\mathbf{X}}_{167 \times 400}).$$

- Extraem-se apenas 5 CIs (pelas mesmas razões já mencionadas para serem comparadas com as 5 componentes extraídas nas outras duas técnicas) e aplicam-se os diferentes métodos de pós-processamento de ordenação das CIs definidos na Secção 3.8.
- A partir da matriz das coordenadas das CIs previamente ordenadas, $\mathbf{S}_{167 \times 5}$, e da matriz dos coeficientes das misturas, $\mathbf{A}_{5 \times 400}$, efetuam-se as reconstruções da matriz dos desfasamentos ($\tilde{\mathbf{X}} = \mathbf{S}\mathbf{A}$) e da matriz dos dados originais.
- Comparam-se as reconstruções com as séries originais.

6.3.2 Análise dos Resultados

Para as três técnicas obtivemos alguns resultados comparativos associados às componentes.

A percentagem da variância explicada pelas primeiras cinco (num total de 167) CPs é próxima de 55%.

Na Figura 6.3 comparam-se as coordenadas das primeiras cinco componentes para as 3 técnicas com $m = 50$. Enquanto as CPs estão ordenadas pela grandeza do valor próprio (a variância de cada componente), a indexação das CIs é arbitrária, não necessariamente como as CPs, e como tal há que considerar os vários métodos de ordenação mencionados.

Contudo, as CPs obtidas pela técnica I e as CIs obtidas de acordo com a ordem arbitrária através da técnica II, apresentam elevados valores de correlações entre certos pares de componentes como ilustra a Tabela 6.1. Por analogia são apresentadas também na Tabela 6.2 as correlações entre as CPs obtidas pela técnica I e as CIs obtidas de acordo com a ordem arbitrária através da técnica III.

Para decidir quais os métodos mais apropriados de ordenação das CIs, comparámos as reconstruções com os dados originais. Foram calculadas as somas dos quadrados dos resíduos (por séries temporais) e apresentadas na Tabela 6.3 as somas desses valores para cada uma das estações meteorológicas, utilizando um número distinto de componentes para as técnicas I, II e III na reconstrução dos dados originais para $m = 50$.

Na Tabela 6.3, nas técnicas II e III, a linha que representa a ordem arbitrária (1,2,3,4,5) das CIs é obtida por aplicação do algoritmo FastICA. A partir da ordem arbitrária, as CIs são

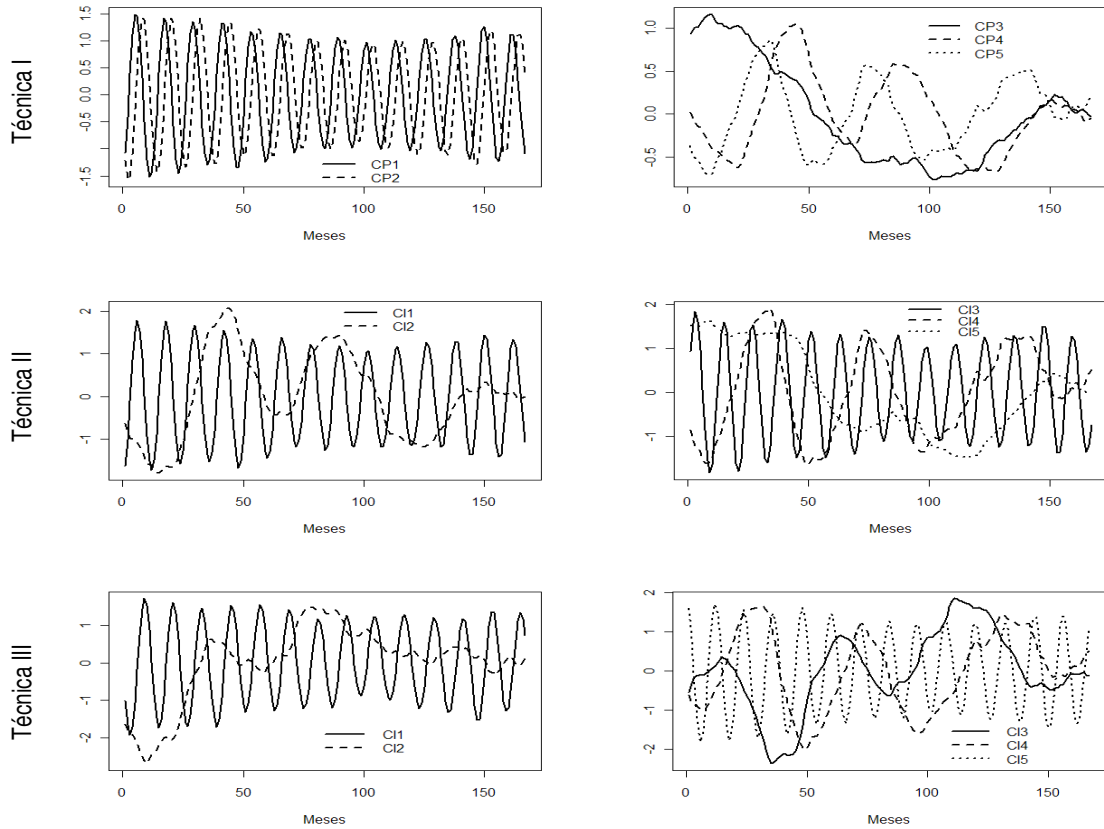


Figura 6.3: Coordenadas para $q = 5$ componentes obtidas pelas técnicas I, II e III, para $m = 50$

reordenadas de acordo com os correspondentes métodos para obtermos uma nova ordem. Por exemplo, na técnica II na linha do método M1 (maximização da curtose), a ordem (1,3,5,4,2) significa que a primeira CI é a mesma que na ordem arbitrária, a segunda nova CI é a terceira CI na ordem arbitrária, e assim sucessivamente até à quinta CI, a segunda CI na ordem arbitrária.

As últimas cinco colunas apresentam os valores de SQR para as reconstruções à custa de uma componente até cinco componentes, de acordo com a ordem do método correspondente. Por exemplo, na técnica II na linha do método M1, o valor 8.683 representa a SQR para a reconstrução à custa das três primeiras CIs (CI1, CI3 e CI5) na respetiva ordem (1,3,5,4,2).

Após comparação entre os cinco métodos de ordenação das CIs aplicados na técnica II, três (M2, M3 e M5) dos métodos consideram a mesma ordem (1,3,5,2,4) de modo a minimizar a SQR para todas as estações meteorológicas. Por outro lado, a técnica III sugere o algoritmo da minimização da SQR (M3) com a ordem (5,2,1,3,4) como o melhor método para minimizar

Tabela 6.1: Correlações entre as CPs obtidas pela técnica I e as CIs obtidas por ordem arbitrária pela técnica II, para um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas

	<i>CI1</i>	<i>CI2</i>	<i>CI3</i>	<i>CI4</i>	<i>CI5</i>
<i>CP1</i>	0.9487	0.0135	0.3149	-0.0228	-0.0134
<i>CP2</i>	0.3133	0.0710	-0.9470	-0.0026	-0.00015
<i>CP3</i>	0.0217	-0.2685	-0.0131	-0.0054	0.9629
<i>CP4</i>	-0.0335	0.9493	0.0605	-0.1536	0.2655
<i>CP5</i>	0.0176	0.1466	0.0141	0.9879	0.0462

Tabela 6.2: Correlações entre as CPs obtidas pela técnica I e as CIs obtidas por ordem arbitrária pela técnica III, para um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas

	<i>CI1</i>	<i>CI2</i>	<i>CI3</i>	<i>CI4</i>	<i>CI5</i>
<i>CP1</i>	-0.3170	0.0454	0.0449	-0.0400	-0.9454
<i>CP2</i>	0.9439	-0.0109	-0.0697	0.0262	-0.3215
<i>CP3</i>	-0.0685	-0.8298	-0.5410	0.1087	-0.0471
<i>CP4</i>	-0.0331	0.4396	-0.7656	-0.4682	0.0156
<i>CP5</i>	-0.0519	0.3405	-0.3381	0.8756	-0.0193

as somas das SQR das estações meteorológicas, no entanto uma alternativa à maximização das correlações canônicas do M4 (ACC) com a ordem (5,2,1,4,3) apresenta resultados similares. O método designado como "Alternativa ao M4" difere apenas na escolha de qual é a primeira CI a ser considerada para a ordenação, uma vez que cada uma das correlações canônicas da CI5 e da CI1 (obtidas a partir da matriz desfasada reconstruída) com os dados originais são muito próximas.

Por análise dos valores de SQR das respectivas reconstruções, salientamos os resultados muito semelhantes entre a técnica I e a ordem (1,3,5,2,4) que minimiza a SQR na técnica II, para os três métodos incluindo o novo método M5. Destacamos ainda os valores de SQR que são, de uma forma geral, relativamente mais baixos para a técnica III quando comparada com as técnicas I e II, essencialmente a partir de duas componentes na reconstrução, excluindo a ordem arbitrária

Tabela 6.3: Soma de quadrados dos resíduos para as reconstruções com as técnicas I, II e III, usando um número de componentes distintas; as técnicas II e III apresentam cada uma das ordens nos diferentes métodos de ordenação das CIs; as técnicas usaram séries desfasadas com um comprimento da janela $m = 50$ e $q = 5$ componentes extraídas

Técnica I						
	ORDEM	CP1	CP12	CP123	CP1234	CP12345
Ordem natural	(1,2,3,4,5)	12.103	10.433	8.666	7.751	7.344
Técnica II						
Métodos de ordenação das CIs	ORDEM	CI1	CI12	CI123	CI1234	CI12345
Ordem arbitrária	(1,2,3,4,5)	12.120	10.941	9.264	8.610	7.344
M1	(1,3,5,4,2)	12.120	10.430	8.683	7.887	7.344
M2; M3; M5	(1,3,5,2,4)	12.120	10.430	8.683	7.749	7.344
M4	(1,5,3,4,2)	12.120	10.460	8.683	7.887	7.344
Técnica III						
Métodos de ordenação das CIs	ORDEM	CI1	CI12	CI123	CI1234	CI12345
Ordem arbitrária	(1,2,3,4,5)	12.251	10.238	9.455	9.057	7.344
M1	(5,1,4,2,3)	12.118	10.415	9.387	7.722	7.344
M2; M5	(5,1,2,3,4)	12.118	10.415	8.467	7.718	7.344
M3	(5,2,1,3,4)	12.118	10.261	8.467	7.718	7.344
M4	(1,2,5,4,3)	12.251	10.238	8.467	7.722	7.344
Alternativa ao M4	(5,2,1,4,3)	12.118	10.261	8.467	7.722	7.344

e o M1.

Em relação aos vetores que se obtêm quando se implementa cada uma das três técnicas mencionadas, na Figura 6.4 constatamos visualmente que cada vetor associado a uma determinada componente apresenta a mesma estrutura entre técnicas e que cada um apresenta um comportamento repetitivo por 8 vezes a cada janela temporal de 50 meses, embora por vezes com magnitudes diferentes. Tais saltos de 50 em 50 meses devem-se à constituição dos vetores de 400 coordenadas, obtidos a partir da matriz desfasada do tipo 167×400 .

Em relação aos espectros entre as técnicas I e II (Figura 6.5), podemos destacar nas duas primeiras componentes o pico de frequência em 0.083 correspondente ao período de 12 meses, enquanto que nas componentes CP4, CP5, CI4 e CI5 surgem como relevantes os picos correspondentes às periodicidades de 5 anos, 45 meses, 9.5 meses, 6 meses e aproximadamente 3.8 e 2.8 meses.

Procedeu-se a análises dos espectros das componentes entre as técnicas I e III e entre as

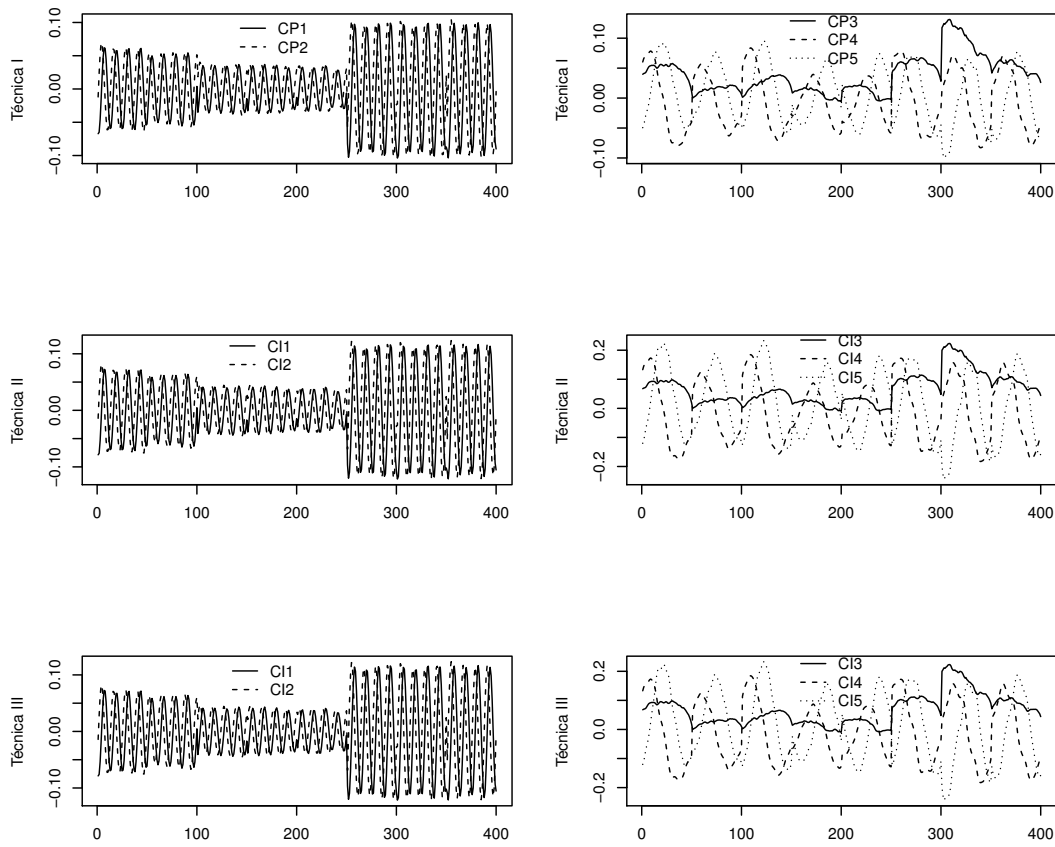


Figura 6.4: Vetores associados à ACP e à ACI a cada uma das três técnicas

técnicas II e III, e como podemos constatar nos gráficos das Figuras 6.6 e 6.7, as interpretações relativamente às periodicidades predominantes são similares às descritas anteriormente entre as técnicas I e II.

Ainda se implementaram as técnicas expostas anteriormente, alterando o comprimento da janela para $m = 100$, por ser um valor mais próximo do "limite" superior recomendado $n/2$, e porque, segundo Vautard *et al.* (1992), permite analisar com sucesso componentes com períodos mais elevados. Os resultados para $m = 100$, no que se refere ao comportamento das cinco componentes extraídas e ilustradas na Figura 6.8, foram similares entre as três técnicas (tendo em conta que as CIs não estão ordenadas) e também muito semelhantes em termos de interpretações

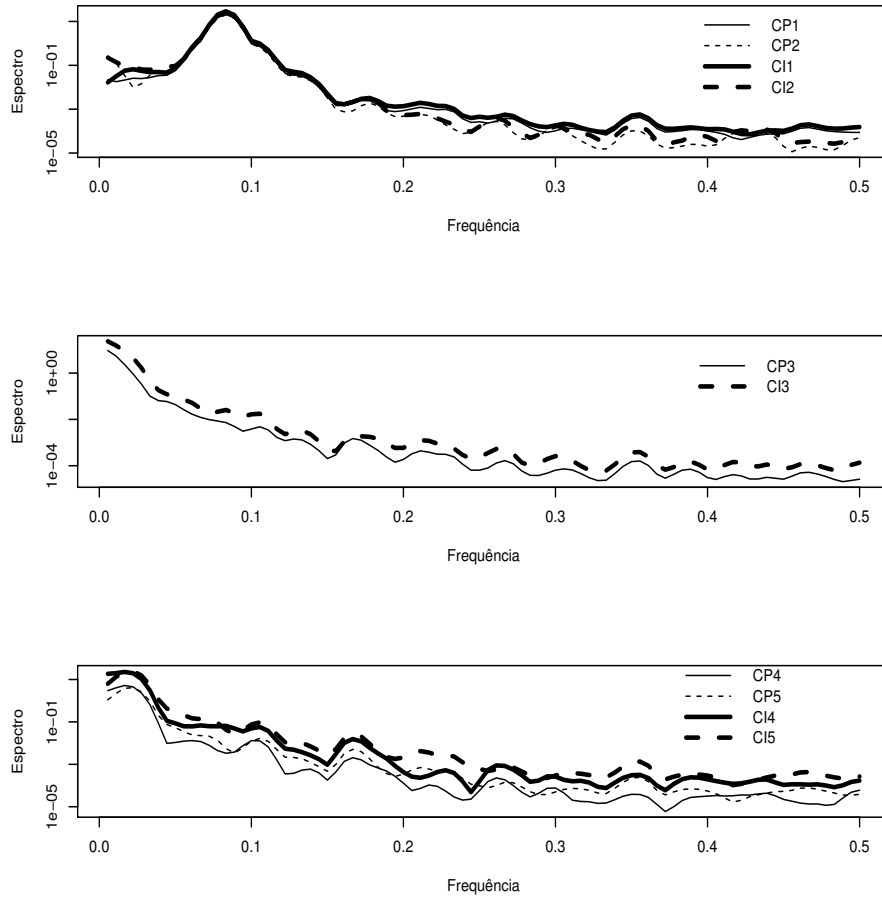


Figura 6.5: Espectros das coordenadas das cinco CPs e CIs retidas para as técnicas I e II

gráficas nas análises quando $m = 50$.

Indicadores Adicionais do Desempenho das Três Técnicas

Para as três técnicas em análise são apresentados alguns resultados comparativos adicionais com base nos cinco indicadores de avaliação do desempenho de um modelo (d, d_1, d_r, E, E_1) , definidos na Subsecção 1.4.3 (e rotinas implementadas em R no Apêndice D) com o intuito de analisar a existência ou não de convergência de resultados. Deste modo pretendemos analisar até que ponto é que podemos argumentar se uma técnica apresenta melhores resultados relativamente às restantes, quando se efetuam as reconstruções às custa das componentes envolvidas em cada

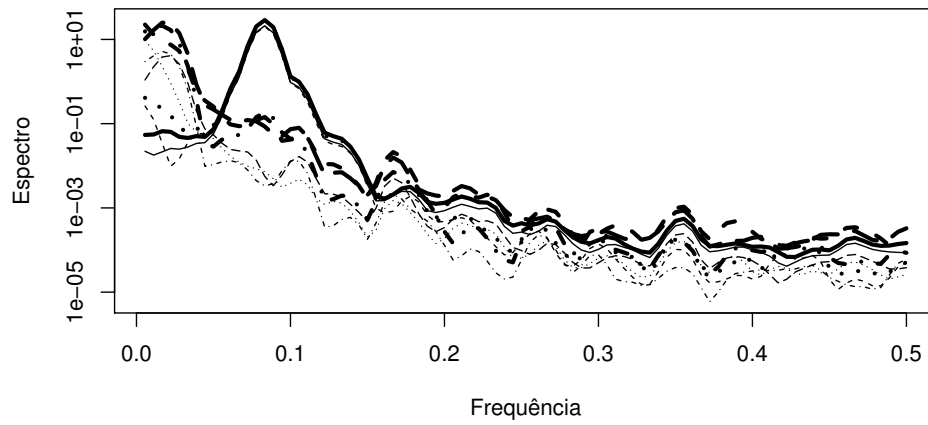


Figura 6.6: Espectros das coordenadas das cinco CPs e CIs retidas para as técnicas I e III

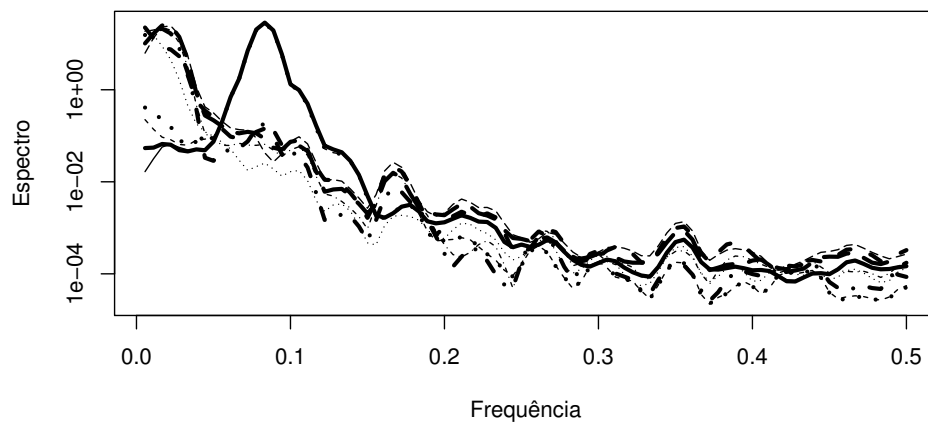


Figura 6.7: Espectros das coordenadas das cinco CIs retidas para cada uma das técnicas II e III

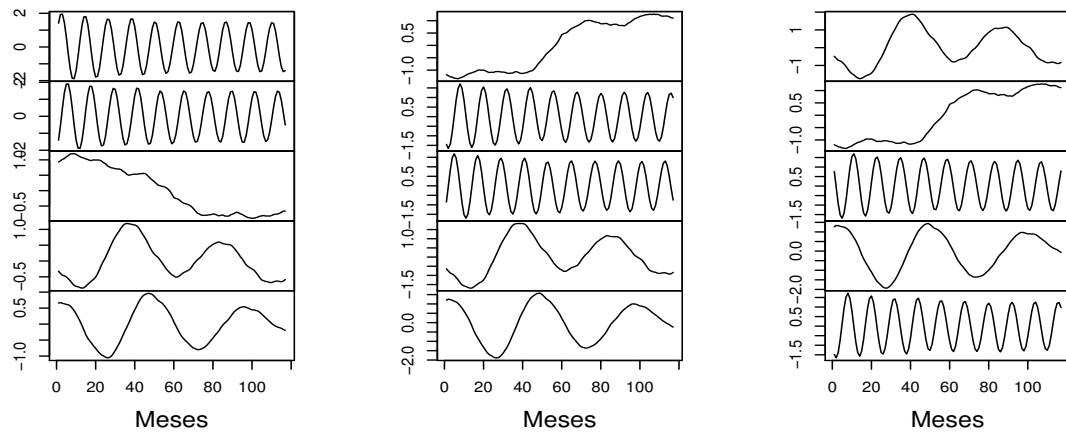


Figura 6.8: Coordenadas para $q = 5$ componentes obtidas pelas técnicas I, II e III, para $m = 100$

uma das técnicas.

Consideremos a Figura 6.9 que compara os indicadores entre as três técnicas para os modelos baseados nas reconstruções desde uma a cinco componentes.

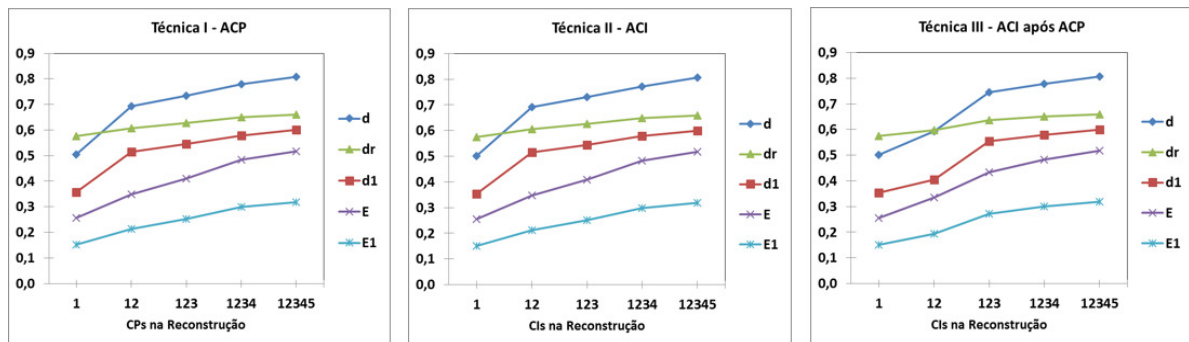


Figura 6.9: Valores das médias (das 8 estações meteorológicas) dos cinco indicadores para as técnicas I, II e III

Na análise comparativa dos vários indicadores, para cada uma das técnicas separadamente, constatamos que os respectivos valores são muito similares entre técnicas e que o comportamento de cada indicador consoante o número de componentes envolvidas na reconstrução também é muito semelhante entre técnicas, apesar da existência de uma ligeira flutuação na técnica III

para a reconstrução à custa das duas primeiras componentes, onde os valores dos índices de Willmott d e de Willmott revisto d_1 são um pouco mais baixos.

Relativamente a cada um dos indicadores analisados separadamente, o índice de Willmott d apresenta valores mais elevados de concordância relativamente ao índice de Willmott revisto d_1 . O índice de Willmott refinado d_r exhibe valores próximos de 0.6. Para os coeficientes de eficiência (não limitados inferiormente e limitados superiormente por 1), o coeficiente de Nash e Sutcliffe E apresenta valores relativamente mais elevados.

No geral podemos argumentar que parece não existirem vantagens relevantes, na utilização destes cinco indicadores adicionais de desempenho, para decidir qual das três técnicas apresenta as reconstruções mais próximas dos dados originais, uma vez que é notória uma grande estabilidade dos valores dos indicadores entre técnicas.

6.4 Resumo e Conclusões

Neste Capítulo é proposta uma nova técnica para o estudo de séries temporais estendidas designada por *LagACI*, cujo objetivo é identificar padrões que se julguem independentes e que possam ocorrer ao longo do tempo mas também ao nível espacial, à semelhança do que se pretende na MSSA. Apresentam-se os principais fundamentos teóricos descritos em quatro passos, recorrendo a uma fusão das bases da MSSA com o modelo da ACI mediante as necessárias adaptações.

No exemplo de aplicação acerca da pressão ao nível do mar, foram propostas três técnicas envolvendo a ACP e a ACI para analisar séries temporais estendidas, realçando a aplicação da nova abordagem *LagACI* na matriz aumentada dos dados desfasados em duas delas. As cinco componentes dominantes que são naturalmente ordenadas em ACP, não saem necessariamente pela mesma ordem na ACI. Como tal aplicaram-se cinco métodos distintos de ordenação das CIs. Nas três técnicas foram obtidos resultados semelhantes nas coordenadas das componentes, depois da ordenação das CIs, para um desfasamento $m = 50$.

A qualidade das reconstruções dos dados originais foi analisada através de tabelas de somas de quadrados dos resíduos, com resultados similares para as técnicas I e II, enquanto que a técnica III evidenciou resultados ligeiramente melhores. O novo método de ordenação das CIs baseado nas correlações entre as CPs e a CIs sugere para este exemplo ser um dos melhores, principalmente para a técnica III e como tal pode ser tido em consideração aquando da utilização

futura dos métodos de ordenação das CIs.

Relativamente ao estudo no domínio do tempo, foi possível identificar os pares formados por CPs e CIs entre técnicas com elevados valores de correlações, enquanto que a análise comparativa dos espectros das componentes (as principais e as independentes) entre quaisquer duas das três técnicas permitiu detetar as mesmas periodicidades, com relevância para o período de 12 meses nas duas primeiras componentes, ao usar uma janela de comprimento $m = 50$ meses. Ao efetuar-se uma pequena simulação para $m = 100$, as cinco componentes extraídas evidenciaram o mesmo comportamento visual entre as 3 técnicas e similar ao caso analisado em que $m = 50$.

Com base nos vetores associados às componentes detetaram-se padrões similares entre técnicas que se repetem ao longo do tempo a cada 50 meses, associados a cada uma das 8 estações meteorológicas (padrões no espaço). Esta aplicação indica que, de um modo geral, a introdução dos desfasamentos com a MSSA não sugere melhorias substanciais notórias face à ACP. Contudo, podemos constatar que a MSSA permite captar as oscilações, essencialmente aquelas que não são as dominantes, de uma forma mais eficaz.

Ainda para o exemplo, ao comparar os indicadores adicionais do desempenho das técnicas em estudo, as técnicas I e II apresentam resultados muito semelhantes e a técnica III evidencia umas pequenas diferenças para um número intermédio de componentes nas reconstruções, mas pouco relevantes. Deste modo não existem claras evidências para que se possa dizer que um modelo associado a uma técnica apresente melhor desempenho que o modelo de uma outra técnica. Apesar de se usarem estatísticas de ordem superior para extrair mais informação, a *LagACI* apresenta de um modo geral uma consistente convergência de resultados relativamente à clássica MSSA para esta aplicação.

Capítulo 7

Discussão e Conclusões

"Pensar contra a corrente do tempo é heroico, dizê-lo é uma loucura."

Eugene Ionesco (1912-1994) - Dramaturgo Francês



7.1 Introdução

Neste Capítulo pretendemos efetuar uma comparação transversal das técnicas apresentadas nos Capítulos anteriores, essencialmente a ACP, a ACI, a MSSA e a *LagACI*. Para ilustrar de uma forma generalizada quais as estruturas matriciais que estão subjacentes a cada uma das técnicas da ACP, ACI, SSA, MSSA e *LagACI*, apresentamos um breve resumo comparativo das mesmas na Figura 7.1.

Como exemplo de aplicação, iremos debruçar-nos sobre a discussão de resultados de um conjunto bastante complexo de dados, na área do clima, referentes a medições da altura geopotencial num setor do Hemisfério Norte. Uma vez que estudos da variabilidade ao longo de escalas temporais são extremamente importantes para perceber as alterações climáticas, um dos principais motivos deste estudo, ao longo de 50 anos, é ajudar a definir estados periódicos da atmosfera, identificando essencialmente as oscilações predominantes de períodos longos na altura

Técnica	Matriz de dados iniciais	Pré-processamento	Procedimento ou algoritmo a implementar	Estimação e Reconstrução
ACP	\mathbf{X} – matriz dos dados observados	Centragem + Redução (se necessário)	Decomposição espectral de $\mathbf{S} = (\mathbf{1}/n) \mathbf{X}'\mathbf{X}$: $\mathbf{S} = \mathbf{P}\mathbf{L}\mathbf{P}^T$, onde \mathbf{P} - matriz dos vetores próprios \mathbf{L} - matriz dos valores próprios	Coordenadas das CPs: $\mathbf{Z} = \mathbf{X}\mathbf{P}$ Reconstrução: $\mathbf{X} \sim \mathbf{Z}\mathbf{P}^T$
ACI (Modelo linear clássico)	\mathbf{X} - matriz dos dados observados (matriz das misturas)	Centragem + Branqueamento (descorrelação e variâncias unitárias)	Utilização do FastICA e do AMUSE entre dezenas de algoritmos existentes	Estimação do modelo $\mathbf{X} \sim \mathbf{S}\mathbf{A}$, obtida a partir de: \mathbf{S} - matriz das Cls \mathbf{A} - matriz dos coeficientes das misturas
SSA	X_t - série temporal observada de tamanho n	Desfasamento m ; $k = n-m+1$ vetores desfasados de dimensão m ; Matriz de desfasamentos (matriz trajetória) $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2 : \dots : \mathbf{X}_k]^T$ de dimensão $k \times m$	Decomposição em valores singulares de $\mathbf{X}'\mathbf{X}$, (Golyandina <i>et al.</i> , 2001): m valores próprios λ_i m vetores próprios \mathbf{U}_i <u>OU</u> Decomposição espectral de $\mathbf{S}' = (\mathbf{1}/k) \mathbf{X}'\mathbf{X}$, (Broomhead e King, 1986a, 1986b): $\mathbf{S}' = \mathbf{P}\mathbf{L}\mathbf{P}^T$	Coordenadas de cada CP: $\mathbf{V}_i = \mathbf{X}\mathbf{U}_i / \sqrt{\lambda_i}$, Reconstrução de \mathbf{X} a partir da soma de matrizes $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{V}_i \mathbf{U}_i^T$, seguida da média ao longo das diagonais para obter a série X_t reconstruída
MSSA	$\mathbf{X}' n \times p$ - matriz dos dados observados	Desfasamento m ; $k = n-m+1$ vetores desfasados de dimensão $p' = mp$; Matriz de desfasamentos (matriz trajetória) \mathbf{X} de dimensão $k \times p'$	Decomposição em valores singulares de $\mathbf{X}'\mathbf{X}$, (Golyandina <i>et al.</i> , 2001): p' valores próprios λ_i p' vetores próprios \mathbf{U}_i <u>OU</u> Decomposição espectral de $\mathbf{S}' = (\mathbf{1}/k) \mathbf{X}'\mathbf{X}$, (Broomhead e King, 1986a, 1986b): $\mathbf{S}' = \mathbf{P}\mathbf{L}\mathbf{P}^T$	Coordenadas de cada CP: $\mathbf{V}_i = \mathbf{X}\mathbf{U}_i / \sqrt{\lambda_i}$, Reconstrução de \mathbf{X} a partir da soma de matrizes $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{V}_i \mathbf{U}_i^T$, seguida da média ao longo das diagonais (de cada bloco de submatrizes) para obter a matriz \mathbf{X}' reconstruída
LagACI	$\mathbf{X}' n \times p$ - matriz dos dados observados	Desfasamento m ; $k = n-m+1$ vetores desfasados de dimensão $p' = mp$; Matriz de desfasamentos (matriz trajetória e das misturas) \mathbf{X} de dimensão $k \times p'$ Centragem + Branqueamento (descorrelação e variâncias unitárias)	Utilização do FastICA e do AMUSE entre dezenas de algoritmos existentes	Estimação do modelo $\mathbf{X} \sim \mathbf{S}\mathbf{A}$, obtida a partir de: \mathbf{S} - matriz das Cls \mathbf{A} - matriz dos coeficientes das misturas Reconstrução de \mathbf{X} a partir da soma de matrizes \mathbf{X}_i , seguida da média ao longo das diagonais (de cada bloco de submatrizes) para obter a matriz \mathbf{X}' reconstruída

Figura 7.1: Comparações das estruturas matriciais das cinco técnicas em estudo ao longo da dissertação: ACP, ACI, SSA e MSSA e *LagACI*

geopotencial.

Para entender o contexto amplo das alterações climáticas, há que distinguir as noções de tempo e clima. O tempo que ocorre num determinado local e num certo instante define-se por um conjunto de elementos atmosféricos tais como temperatura do ar, pressão atmosférica, humidade relativa do ar, vento à superfície, nebulosidade, entre outros. Por sua vez, o clima é definido como uma síntese dos estados de tempo característicos de um certo local ou região num determinado intervalo temporal estabelecido. Para auxiliar na interpretação do clima utilizam-se técnicas e modelos estatísticos baseados em variáveis meteorológicas através dos seus valores médios, valores extremos, probabilidades de ocorrência, etc.. Deste modo, podemos afirmar que o tempo e o clima são duas maneiras de descrever as mesmas variáveis meteorológicas mas em diferentes escalas temporais.

Algumas Referências a Estudos Usando a Altura Geopotencial

Algumas aplicações que envolvem o estudo da altura geopotencial têm vindo a ser efetuadas essencialmente nas últimas décadas, muito em parte devido ao crescente desempenho dos computadores no tratamento de conjuntos de dados extensos. Como tal, consideremos alguns exemplos que podem ser mencionados como percursos no estudo desta variável meteorológica a um nível mais complexo do sistema climático.

Ghil e Mo (1991) analisaram os modos oscilatórios que ocorrem no Hemisfério Norte e nos trópicos. Utilizaram dados da altura geopotencial a 700 milibar no estudo das oscilações extratropicais, entre os anos de 1949 e 1986, na região entre 15°N - 90°N. Os dados utilizados foram sujeitos a filtros para realçar a banda intrasazonal de 10 a 120 dias e os padrões espaciais dominantes de variabilidade foram obtidos através da aplicação da ACP. Posteriormente, as CPs dominantes foram submetidas à aplicação da SSA. No Hemisfério Norte, foram identificados dois modos principais de oscilação com períodos próximos de 48 dias e de 23 dias. O modo dominante correspondeu ao padrão do Pacífico-América do Norte (PNA) descrito anteriormente por Wallace e Gutzler (1981).

Também Plaut e Vautard (1994) aplicaram a MSSA num conjunto de dados da altura geopotencial a 700 milibar, observados durante 32 anos, sobre uma região do Hemisfério Norte. Com o intuito de analisarem as oscilações de baixa frequência, a MSSA permitiu detetar uma oscilação forte com um período de 70 dias, à qual está associado o padrão da Oscilação do Atlântico Norte

(NAO). Outras duas oscilações foram encontradas no intervalo de 30 a 60 dias no Pacífico e no Atlântico.

Para extrair o sinal do ciclo solar com base em dados da estratosfera, Coughlin e Tung (2004) aplicaram um novo método a partir do qual a altura geopotencial no Hemisfério Norte (entre janeiro de 1958 e julho de 2002, entre 20°N - 90°N, observada em vários níveis desde abaixo dos 10 milibar a 1000 milibar) pode ser decomposta em cinco modos temporais não estacionários, os quais incluem um ciclo anual, uma oscilação quase bianual, um modo de oscilação semelhante à ENSO, um sinal de ciclo solar e por outro lado uma tendência. Dada a existência de uma correlação significativamente elevada (cerca de 0.70) entre a altura geopotencial (a 30 milibar) e o ciclo de manchas solares, estabeleceu-se que o quarto modo descrevia um sinal do ciclo solar de 11 anos. Uma vez que a mudança climática natural, por vezes controversa relativamente às suas causas, continua a ser um assunto deveras importante, em Labitzke e Matthes (2005) é também discutida a influência dos ciclos solares como um fator de variabilidade natural na atmosfera. Estudos efetuados com dados (dos arquivos de reanálise NCEP/NCAR da *National Oceanic and Atmospheric Administration* dos E.U.A. (NOAA)) que abrangem cerca de quatro ciclos solares, mostraram correlações elevadas entre o ciclo solar de 11 anos e a altura geopotencial entre outros parâmetros meteorológicos, na troposfera¹ assim como na baixa estratosfera².

7.2 Exemplo de Aplicação: Altura Geopotencial

7.2.1 Dados e Metodologia

Consideremos o seguinte conjunto de dados da variável meteorológica altura geopotencial (gráficos ilustrativos da gama de valores da variável são apresentados no Apêndice D):

- *Dados*: valores mensais médios da altura geopotencial observada para um valor de pressão atmosférica a 500 milibar;
- *Período em estudo*: 50 anos, desde janeiro de 1958 a dezembro de 2007;

¹A troposfera é a região mais baixa da atmosfera, na qual ocorre a maior parte dos fenómenos meteorológicos como precipitação, nuvens e trovoadas. A sua altitude varia entre os 6 km (polos) e os 17 km (equador).

²A estratosfera é a região situada acima da troposfera entre os 6-17 km e até cerca de 50 km de altitude.

- *Fonte*: arquivos de reanálise NCEP/NCAR do *National Oceanic and Atmospheric Administration* dos E.U.A.³ (Kalnay *et al.*, 1996);
- *Setor*: Hemisfério Norte entre 20°N - 80°N e 0°E - 357.5°E (Figura 7.2);
- *Grelha espacial*: uniforme de 2.5° ao longo da latitude e da longitude, de tamanho $25 \times 144 = 3600$ pontos, o que corresponde a 3600 variáveis iniciais.

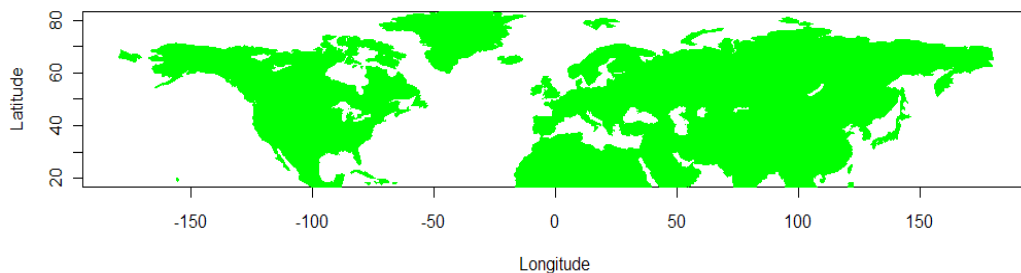


Figura 7.2: Setor do Hemisfério Norte entre 20°N - 80°N e 0°E - 357.5°E, onde foram recolhidos os dados da altura geopotencial

Uma vez que geralmente os dados climáticos não estão uniformemente distribuídos sobre a superfície da Terra, tal situação pode influenciar a estrutura das componentes que venham a ser extraídas de acordo com uma determinada técnica que envolva CPs. Um exemplo, muito comum, é aquele em que os dados são provenientes de uma grelha espacial com 5° de latitude por 5° de longitude, e com uma distribuição mais densa na direção dos polos. Como forma de evitar o efeito geométrico associado à superfície da Terra, habitualmente os dados são previamente ponderados. Uma maneira simples é ponderar cada ponto (com dados observados) através da área local referente à sua localização, como é o caso da ponderação através da raiz quadrada do cosseno da sua latitude como em Hannachi *et al.* (2006, 2007, 2009).

Desta forma, para os dados originais da altura geopotencial em estudo foram inicialmente obtidas as suas anomalias⁴ ponderadas para remover o efeito de alguma sazonalidade, como por exemplo o ciclo anual dominante. Tais anomalias foram calculadas através da subtração em cada

³<http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.pressure.html>

⁴Anomalia é o desvio em relação à designada climatologia, a qual é usualmente definida como o valor médio de uma variável meteorológica num dado intervalo temporal (ano, semana, mês, trimestre, ...).

mês das respectivas médias mensais ao longo dos 50 anos e para cada ponto da grelha espacial. Para reduzir o efeito dos dados recolhidos em latitudes com valores elevados, que correspondem a grelhas de tamanho mais pequeno à medida que nos aproximamos do polo norte, foi aplicada uma ponderação na área através da multiplicação das anomalias da altura geopotencial pela raiz quadrada do cosseno da correspondente latitude. Obtivemos assim a matriz denotada por $\mathbf{X}_{\text{Geopotencial}}$ ($\mathbf{X}_{600 \times 3600}$).

Em seguida, devido à elevada dimensionalidade dos dados, a qual iria acarretar grandes dificuldades computacionais na aplicação da MSSA, comprimimos a informação através da aplicação da ACP como método de pré-processamento na matriz $\mathbf{X}_{\text{Geopotencial}}$ para reter um menor número de componentes principais (espaciais). Com base na Figura 7.3, a partir de 600 CPs iniciais optou-se por reter as primeiras 19 CPs dominantes, as quais correspondem a 87.17% do total de variabilidade explicada nos dados.

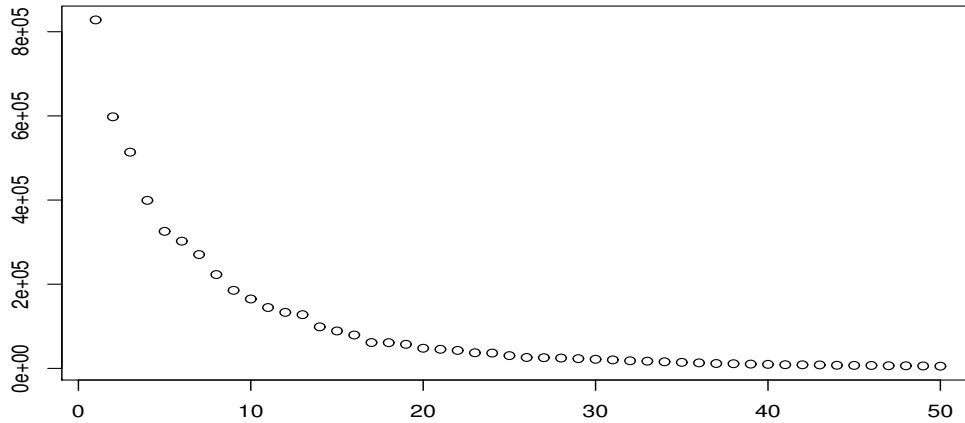


Figura 7.3: Primeiros 50 valores próprios (de um total de 600) obtidos através da ACP aplicada nos dados iniciais

A partir da matriz resultante com $p' = 19$ CPs retidas e $n' = 600$ observações, $\mathbf{X}'_{600 \times 19}$, construiu-se a matriz das séries estendidas, $\mathbf{X}_{\text{desfasada}}$ de dimensões $n \times p = 421 \times 3420$, obtida segundo a aplicação de um desfasamento de $m = 180$ meses (15 anos).

A abordagem a utilizar para este exemplo ilustra-se na Figura 7.4 em que se aplica novamente a *LagACI* como uma alternativa à MSSA para analisar séries temporais estendidas. Na *LagACI*, utilizámos os algoritmos FastICA e AMUSE introduzidos na Secção 3.7. A comparação de resultados entre as duas técnicas foi realizada nos domínios do tempo e da frequência a partir da matriz $\mathbf{X}_{\text{desfasada}}$. Novamente todas as rotinas necessárias à obtenção de resultados gráficos e numéricos foram implementadas no *software R*, as quais estão descritas no Apêndice D.

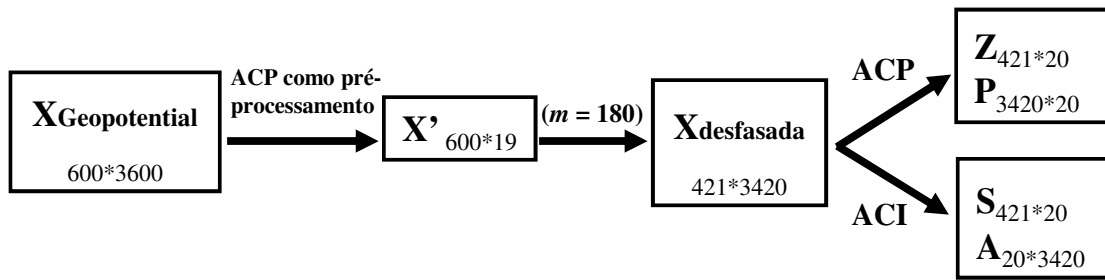


Figura 7.4: Metodologia para MSSA e *LagACI* aplicada na matriz aumentada dos desfasamentos, com um comprimento da janela $m = 180$, $n' = 600$ observações e $p' = 19$ CPs retidas. \mathbf{Z} é a matriz das coordenadas das CPs, \mathbf{P} é a matriz dos vetores próprios, \mathbf{S} é a matriz das CIs e \mathbf{A} é a matriz dos coeficientes das misturas

De acordo com o gráfico dos logaritmos dos valores próprios (Figura 7.5), foram retidas apenas $q = 20$ componentes que passam a ser designadas por Componentes Principais Estendidas (CPEs) no caso da ACP e por Componentes Independentes Estendidas (CIEs) no caso da ACI, como forma de as distinguir das componentes principais (CPs) e das componentes independentes (CIs) que se extraem diretamente da matriz inicial dos dados originais, $\mathbf{X}_{\text{Geopotencial}}$ do tipo 600×3600 . É de ter em conta que as CPEs são ordenadas por ordem decrescente, enquanto que as CIEs não saem ordenadas segundo uma ordem específica.

7.2.2 Análise dos Resultados

Nesta Secção foram utilizados os algoritmos FastICA e AMUSE, como recurso à implementação da técnica da *LagACI*, na obtenção de 20 CIEs cuja extração ocorreu em paralelo.

Para as duas técnicas em análise alguns resultados comparativos com base nas CPEs e nas CIEs (ao nível do tempo e da frequência) serão apresentados com o intuito de detetar as

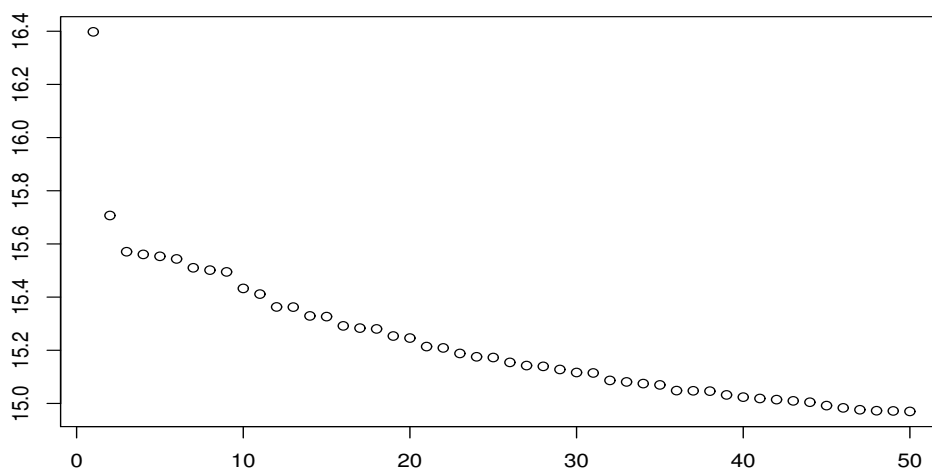


Figura 7.5: Logaritmos dos primeiros 50 valores próprios (de um total de 421) obtidos através da ACP aplicada na matriz dos desfasamentos dos dados

principais oscilações existentes na circulação da atmosfera, ao nível da altura geopotencial em consonância com a pressão atmosférica.

Nas Figuras 7.6, 7.7 e 7.8 apresentam-se respetivamente as coordenadas das 20 CPEs dominantes, das 20 CIEs obtidas pelo algoritmo FastICA e das 20 CIEs obtidas pelo algoritmo AMUSE que foram retidas. As CIEs não foram ordenadas segundo qualquer método de ordenação.

Uma inspeção visual comparativa das figuras permite concluir que as CIEs extraídas pelo algoritmo AMUSE (o qual entra em linha de conta com a dependência temporal das observações) ostentam uma ordem aparentemente crescente segundo a frequência das oscilações (o que não é obrigatório), e para além disso a partir da CIE5 as componentes apresentam-se com comportamentos de pares de oscilações, característica que é típica da MSSA. Todavia, é possível encontrar determinados pares de CPEs e CIEs fortemente correlacionados, e entre estes a grande maioria varia, em módulo, entre 0.4 e 0.6 (valores não apresentados).

Nos resultados da MSSA é de salientar que a CPE1 representa a tendência dos dados correspondente a um crescimento dos valores médios mensais da altura geopotencial ao longo do

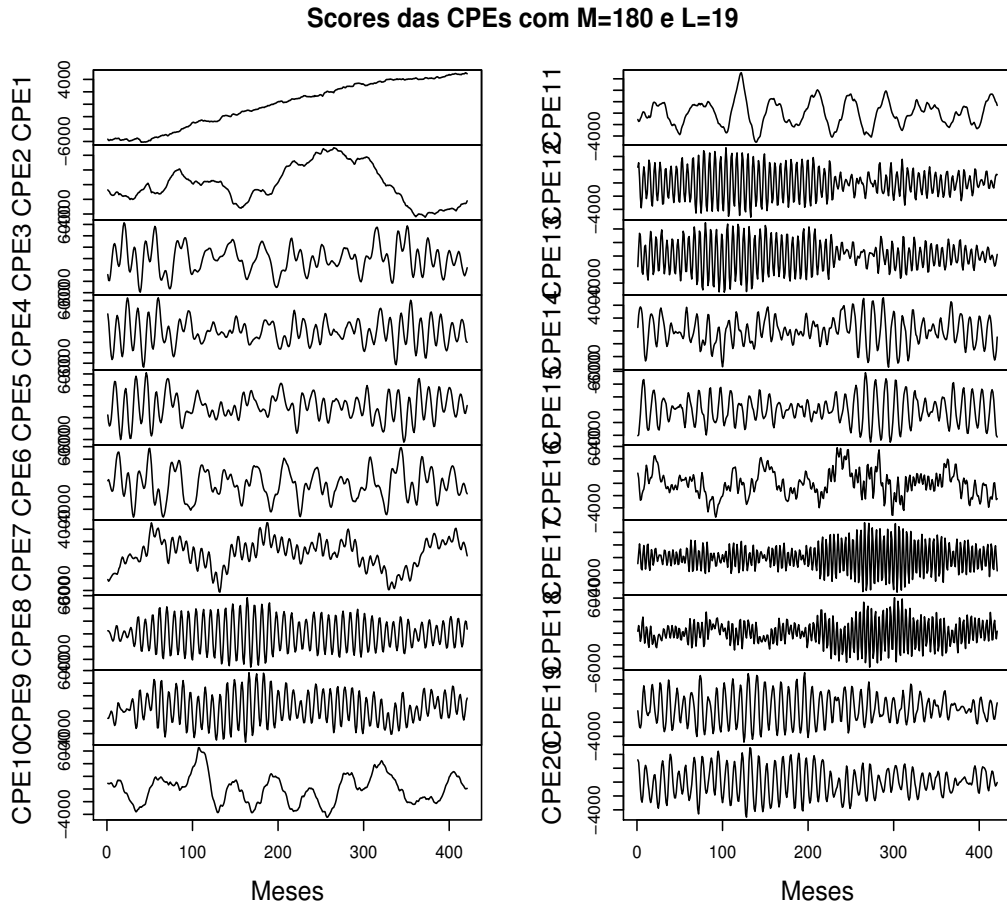


Figura 7.6: Coordenadas para $q = 20$ CPEs, com $m = 180$ e $p' = 19$

período em estudo, que a CPE2 representa uma mistura de tendência com oscilação, enquanto que a maioria das componentes seguintes exibem comportamentos oscilatórios que podem ser analisados ao nível da frequência. Tendo em conta alguns pares de valores próprios que são muito similares e a Figura 7.6, podem ser identificados os pares de componentes associadas que descrevem uma mesma harmónica, utilizando para tal a visualização gráfica de pares de CPEs. Para a Figura 7.9 foram seleccionados os sete pares de CPEs (entre as 20 CPEs retidas) correspondentes aos pares em que foi possível associar o mesmo período dominante.

Decidimos ainda apresentar, na Figura 7.10, oito aparentes pares de CIEs (que se destacam a partir da CIE5 inclusive com base na inspeção visual da Figura 7.8) referentes ao algoritmo AMUSE, onde as CIEs de cada par exibem um comportamento oscilatório semelhante.

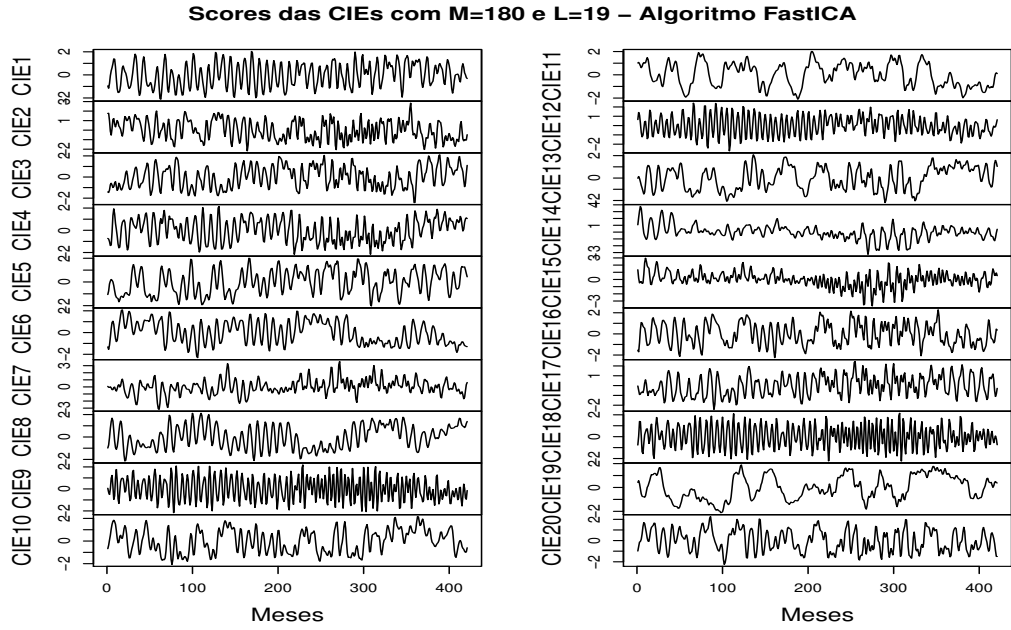


Figura 7.7: Coordenadas para $q = 20$ CIEs obtidas através do algoritmo FastICA, com $m = 180$ e $p_l = 19$

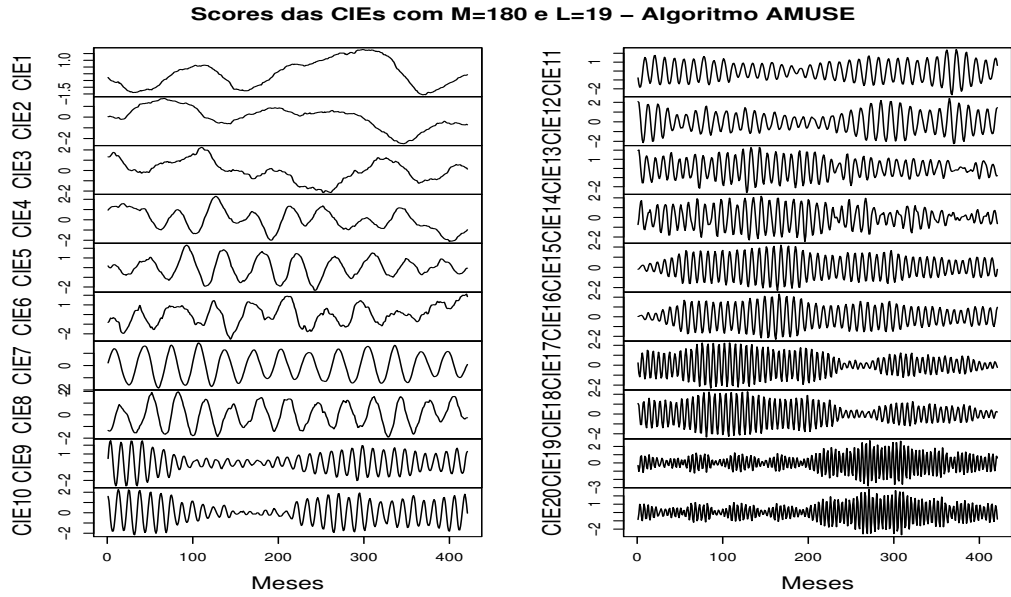


Figura 7.8: Coordenadas para $q = 20$ CIEs obtidas através do algoritmo AMUSE, com $m = 180$ e $p_l = 19$

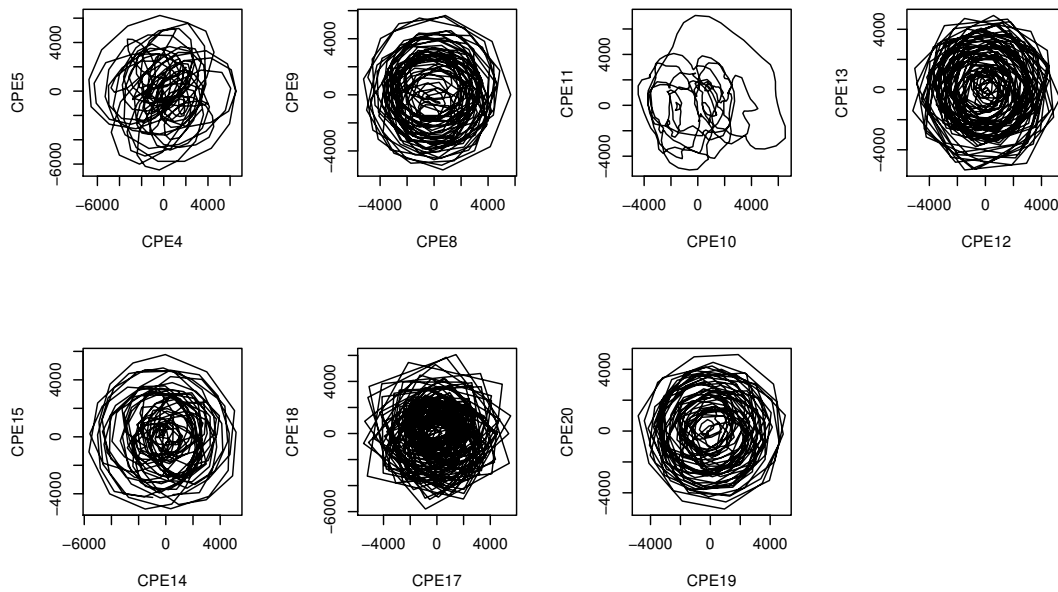


Figura 7.9: Pares de CPEs harmônicas

Os vetores associados às componentes para cada uma das duas técnicas em análise, ilustrados nas Figuras 7.11 e 7.12 apresentam estruturas de difícil interpretação à primeira vista. Por isso, utilizamos os cossenos dos ângulos entre cada par de vetores para as CPEs e CIEs para comparar os vetores. Os cossenos são mais elevados para determinados pares, nomeadamente para aqueles pares cujas respectivas componentes possuem correlações mais elevadas. Não se apresentam os vetores obtidos segundo o algoritmo AMUSE, pois estes exibem estruturas similares às da ACP.

Relativamente à análise da frequência, seguindo as sugestões de Plaut e Vautard (1994) para escolha do tamanho da janela $m = 180$ meses (15 anos), é esperada a distinção de oscilações com períodos no intervalo $(m/5, m) = (36, 180)$. Na Figura 7.13 estão representados, entre outros, os espectros das CPEs e das CIEs para o FastICA com períodos nesse intervalo, na qual podemos constatar que ambas as técnicas detetaram os períodos em torno dos 43 meses e dos 33 meses. A *LagACI* com base no algoritmo AMUSE também detetou os períodos de 43 e 33 meses.

O período mais elevado foi detetado na CPE7, a qual corresponde a uma oscilação com um período de 144 meses (12 anos). Para além destes períodos longos dominantes, outros considerados relevantes para as componentes ilustradas foram assinalados tais como 12, 8.5 e 4.5 meses,

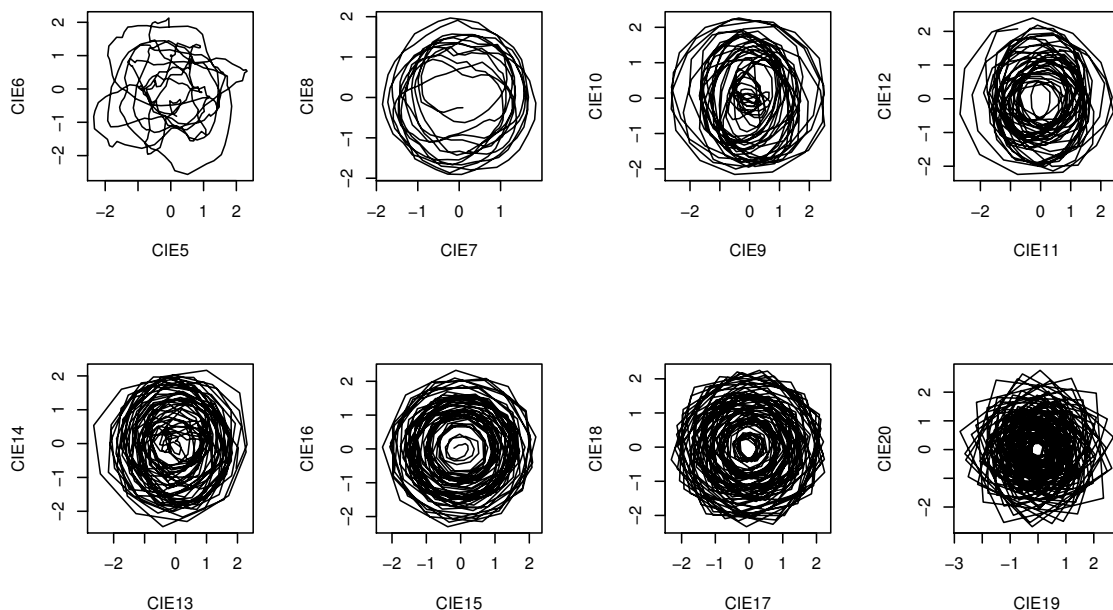


Figura 7.10: Pares de CIEs harmónicas obtidas através do algoritmo AMUSE

os quais surgem como dominantes noutras componentes não apresentadas. Em relação aos sete pares de CPEs apresentados na Figura 7.9, os espectros revelaram períodos dominantes mencionados na Tabela 7.1, na qual são também apresentadas as percentagens de variância explicada por cada par de CPEs. Por outro lado, para as 20 CIEs obtidas segundo o algoritmo FastICA, embora não sejam extraídas aos pares para descrever uma mesma oscilação, conseguimos encontrar os mesmos períodos dominantes que foram identificados na MSSA. Esta capacidade da *LagACI* em extrair a mesma informação dominante que é extraída pela MSSA, já era uma característica notória no exemplo da Secção 6.3, mas fica aqui reforçada ao ser contemplada para um conjunto de dados de elevada complexidade.

No entanto para o caso do algoritmo AMUSE e de acordo com a Figura 7.10, os espectros associados aos pares de CIEs em destaque evidenciaram períodos dominantes de aproximadamente 43, 33, 12, 11.5, 9.5, 8.5, 6.5 e 4.5 meses, respetivamente. É de notar que existem dois períodos dominantes muito próximos nos valores de 11.5 e 12 meses. Não havendo certezas sobre o significado de tal facto, eventualmente este poderá resultar de perturbações (flutuações climáticas) que criam falsas separações em torno do que na realidade corresponde ao ciclo anual,

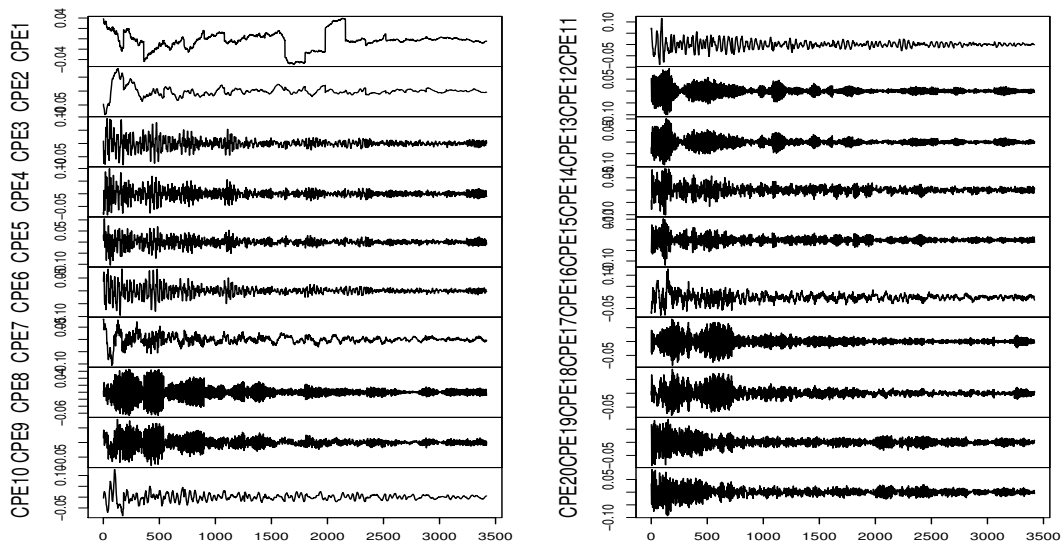


Figura 7.11: Vetores associados a $q = 20$ CPEs, com $m = 180$ e $p' = 19$

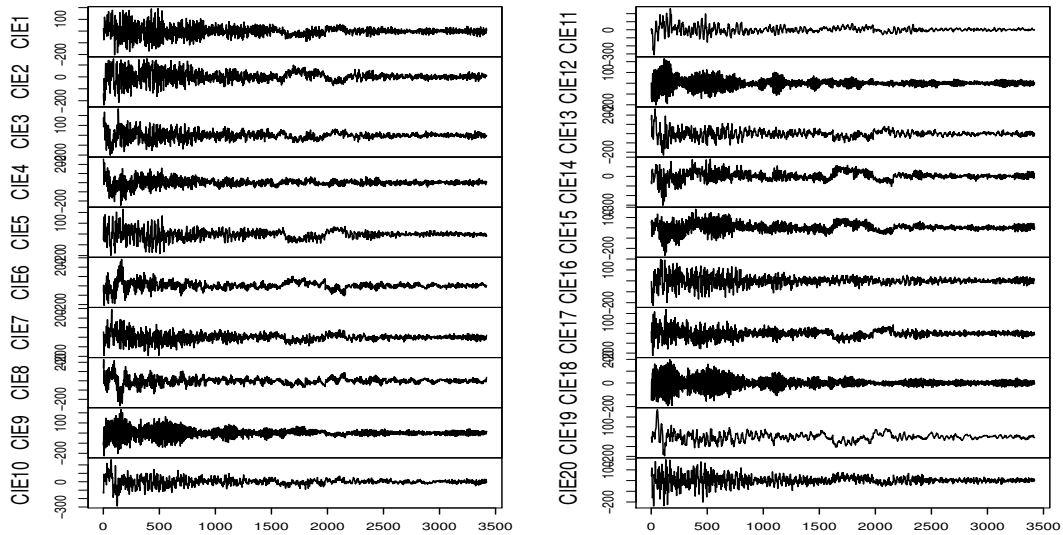


Figura 7.12: Vetores associados a $q = 20$ CIEs obtidas através do algoritmo FastICA, com $m = 180$ e $p' = 19$

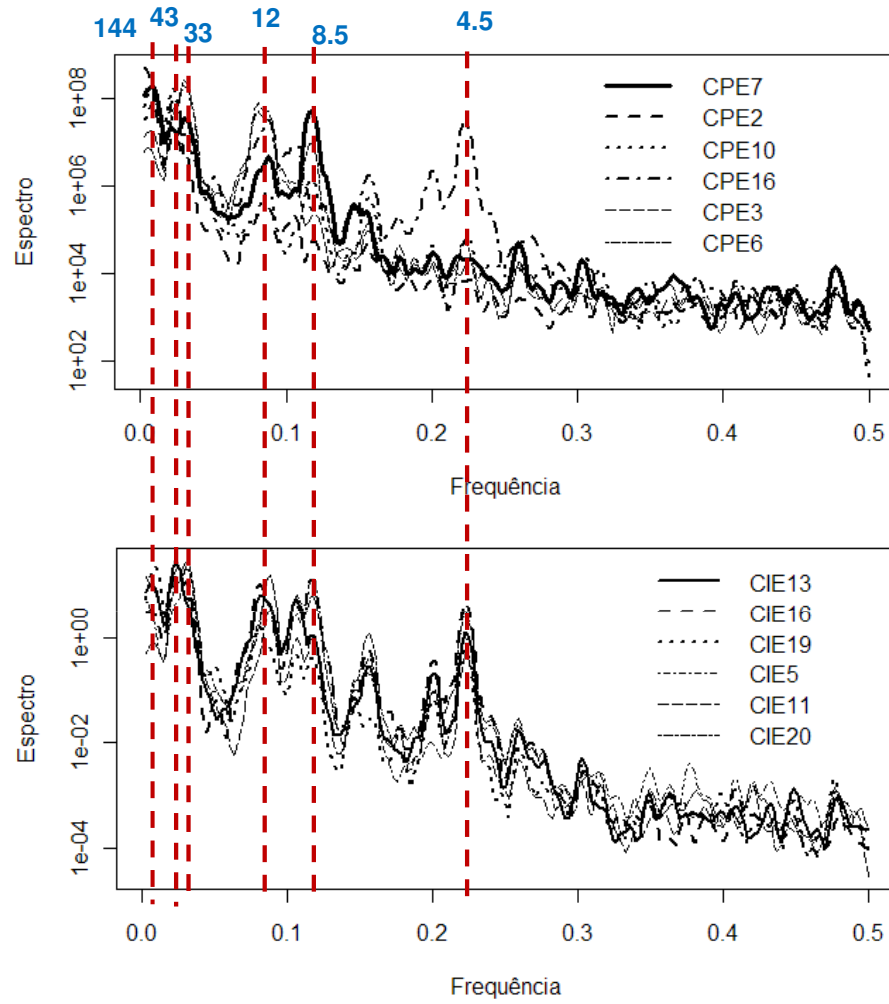


Figura 7.13: Comparações entre os espectros das coordenadas das CPEs e das CIEs de acordo com os valores dos períodos dominantes 144, 43, 33, 12, 8.5 e 4.5 meses, com $m = 180$ e $p = 19$

dando a ilusão de um ciclo de 11.5 meses. A *LagACI* também permitiu detetar o período longo de 144 meses como dominante na CIE2 através do AMUSE e como período secundário através do FastICA em algumas das suas CIEs. Portanto, a *LagACI* revela ser uma alternativa à MSSA no contexto da frequência para séries temporais estendidas ao detetar praticamente os mesmos períodos das principais oscilações nos 50 anos de altura geopotencial analisados.

Tabela 7.1: Períodos dominantes dos pares de CPEs, para um comprimento da janela $m = 180$, $p' = 19$ e $q = 20$ componentes estendidas retidas

Pares de CPEs	Período dominante (meses)	Variância (%)
4-5	12	1.36
8-9	8.5	1.28
10-11	43	1.19
12-13	6.5	1.12
14-15	11.5	1.08
17-18	4.5	1.03
19-20	9.5	1.00

Comparações com a Série Temporal do Índice de Oscilação do Atlântico Norte (NAOI)

Devido à grande importância que a Oscilação do Atlântico Norte (NAO) exerce sobre o sistema climático no Hemisfério Norte, como teleconexão⁵ dominante da variabilidade da pressão atmosférica ao nível do mar no Atlântico Norte, e tendo em conta a relação existente com a altura geopotencial, decidimos levar a cabo uma análise comparativa ao nível da frequência entre o conjunto de dados em destaque (valores mensais médios da altura geopotencial para 500 milibar no Hemisfério Norte) e os valores mensais da pressão no Norte do Oceano Atlântico de acordo com o Índice de Oscilação do Atlântico Norte (NAOI). Os dados do NAOI foram retirados da página

⁵Neste contexto, uma teleconexão é uma forte relação estatística existente no tempo em diferentes partes do globo, constituindo portanto um padrão espacial.

oficial de internet do NOAA, www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml, e estão representados na Figura 7.14.

Hurrell⁶ argumenta que existem várias formas de definir a estrutura espacial da NAO, e como tal não existe nenhum índice que seja aceite em termos universais que permita descrever a evolução temporal do fenómeno. Neste sentido, a nossa escolha recaiu sobre o índice extraído do NOAA uma vez que o considerámos mais adequado tendo em conta a forma como foi calculado.

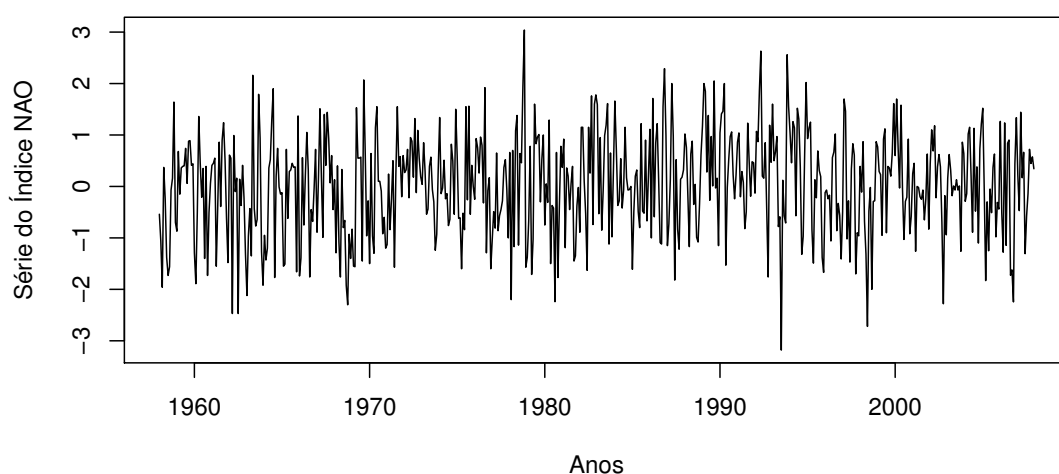


Figura 7.14: Série temporal das anomalias mensais estandardizadas da pressão do Índice de Oscilação do Atlântico Norte entre janeiro de 1958 e dezembro de 2007

Embora a NAO seja caracterizada por uma variação em oposição de fase entre as anomalias de pressão registadas na região da Islândia e as registadas na região dos Açores, as várias definições (mais modernas) do Índice NAO que têm sido usadas em muitos estudos (Gillett *et al.*, 2003), são obtidas a partir da diferença normalizada da pressão à superfície entre dois locais no Atlântico Norte; um mais setentrional (que geralmente corresponde à Islândia) e um outro mais meridional (que geralmente é situado nos Açores, embora também possa ser considerado em Lisboa ou em Gibraltar).

⁶Na página de internet <https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based>, acedida a 26-02-2014

Outros índices alternativos podem ser calculados a partir do vetor próprio associado à série temporal da componente principal dominante da pressão ao nível do mar. A NAO combina, segundo o NOAA, partes do padrão do Atlântico Este e do padrão do Atlântico Oeste, os quais foram originalmente identificados por Wallace e Gutzler (1981). O NOAA garante ainda que os padrões de teleconexões mensais são atualmente interpolações lineares dos respectivos valores diários e as anomalias estandardizadas são calculadas com base nos valores diários entre 1950 e 2000, enquanto que os índices mensais das teleconexões são calculados com base na ACP através de rotações das anomalias médias mensais estandardizadas.

Efetivamente, Wallace e Gutzler (1981) estudaram as teleconexões, relativamente à altura geopotencial para 500 milibar no Hemisfério Norte, com vista a identificar e descrever padrões espaciais recorrentes que sugeriam ser indicativos de oscilações permanentes durante o inverno, ao longo de um mês ou mais. Foram identificados 5 padrões principais (teleconexões) - o do Atlântico Este, o Americano do Pacífico Norte, o Euro-asiático, o do Pacífico Oeste e o do Atlântico Oeste.

De acordo com Hurrell (1995), a NAO apresenta consideráveis variabilidades intersazonal e interanual, assim como usuais períodos prolongados de vários meses de ambas as fases positiva e negativa do padrão. Consideremos a Figura 7.15 que representa o espectro do NAOI, no qual se destacam essencialmente três picos de frequência, os quais correspondem a períodos de aproximadamente 2 anos e meio, 1 ano e meio ano. O período de cerca de 2 anos e meio (31.6 meses) está muito próximo do período longo e dominante de 33 meses que fora encontrado nos dados da altura geopotencial em estudo. Podemos ainda deduzir que a influência dos ciclos anual e semestral na pressão ao nível do mar é notória no Atlântico Norte, o que denota alguma coerência com os resultados espectrais na altura geopotencial do Hemisfério Norte.

Portanto os resultados permitem estabelecer uma associação entre o espectro do NAOI e os espectros analisados na Figura 7.13, realçando assim a importância das técnicas de redução da dimensão de dados para extrair informação do sistema climático com base nas oscilações predominantes de períodos longos, quer seja através das técnicas clássicas, quer seja com os métodos de análise mais recentes, incluindo a nova metodologia da *LagACI* aplicada na matriz dos desfasamentos dos dados.

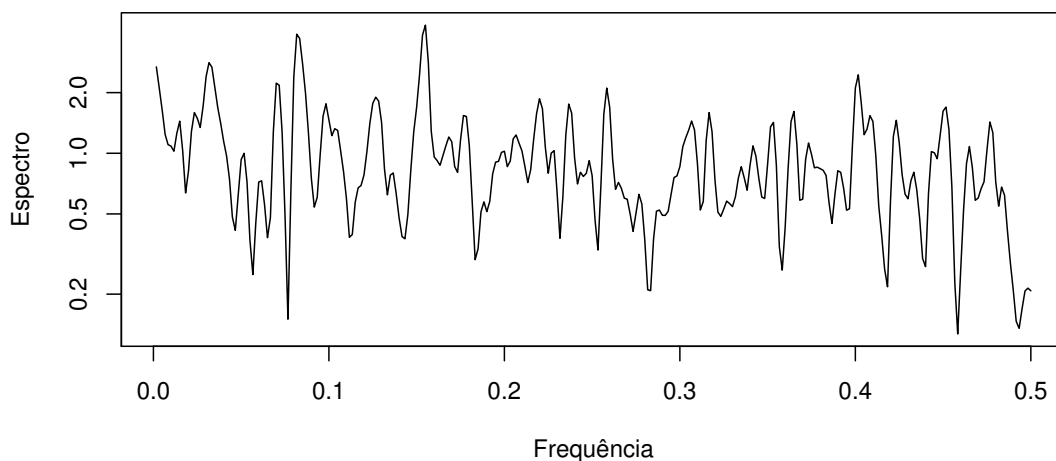


Figura 7.15: Espectro da série temporal mensal do Índice de Oscilação do Atlântico Norte

Mapas da Altura Geopotencial

Para complementar o estudo construímos alguns mapas, referentes aos dados em análise da altura geopotencial, recorrendo a comandos já existentes do *package graphics* do *software R*, nomeadamente a funções como *image.plot* e *filled.contour*, entre outras.

Nas Figuras 7.16 e 7.17 apresentamos os mapas de distribuição da altura geopotencial (dados em bruto) para junho e dezembro de 1958, como ilustrativos das sazonalidades verão e inverno que ocorrem no Hemisfério Norte, respetivamente. Podemos verificar que em junho os valores são mais elevados comparativamente aos de dezembro e, de uma forma generalizada, mais altos na América do Norte e mais baixos na Europa e Norte de África em ambos os casos.

Relativamente aos dados iniciais das anomalias, e no caso em que apenas aplicámos a ACP usual correspondente ao primeiro passo do esquema da Figura 7.4 (antes do desfasamento), obtivemos as Figuras 7.18, 7.19 e 7.20 para os modos (padrões espaciais) principais de variabilidade com base nos vetores próprios associados às três CPs dominantes.

A variabilidade explicada por cada um dos modos principais é cerca de 15.5%, 11.2% e 9.6% respetivamente. O primeiro e segundo modos principais incidem essencialmente ao nível do Atlântico Norte, o que leva novamente a reforçar a evidência da relação existente com o padrão

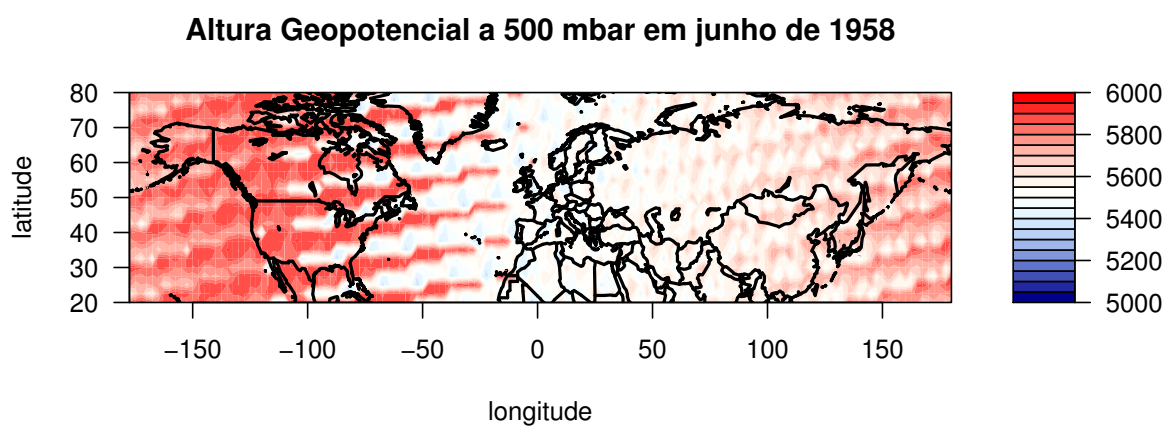


Figura 7.16: Altura geopotencial a 500 milibar em junho de 1958, observada no setor 20°N - 80°N e 0°E - 357.5°E

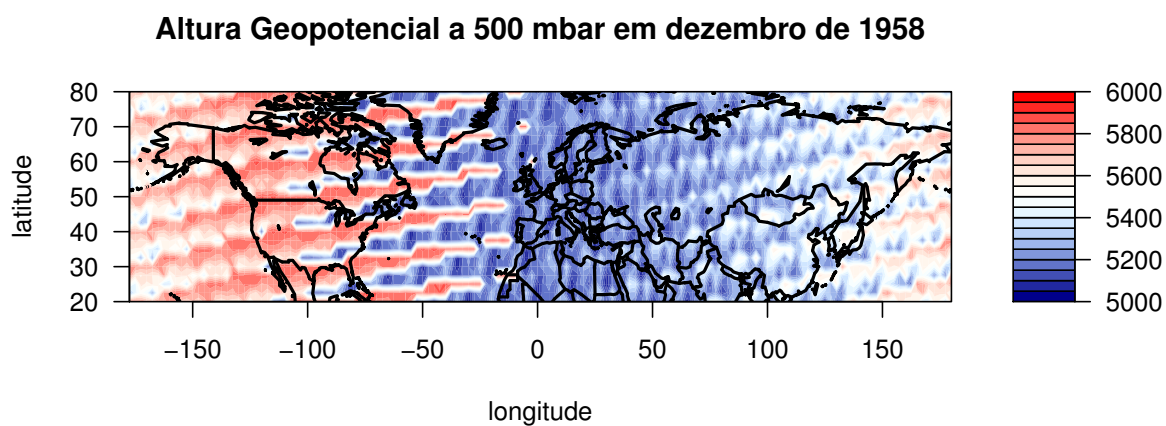


Figura 7.17: Altura geopotencial a 500 milibar em dezembro de 1958, observada no setor 20°N - 80°N e 0°E - 357.5°E

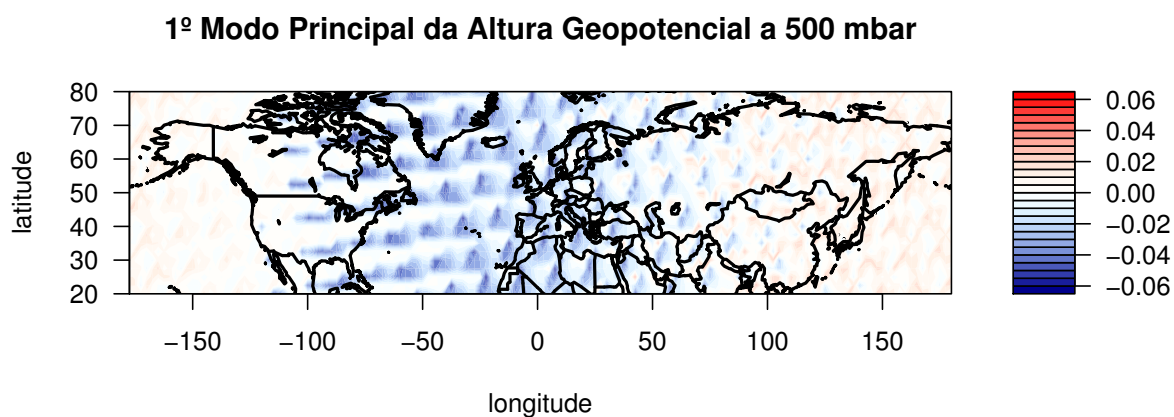


Figura 7.18: 1º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E

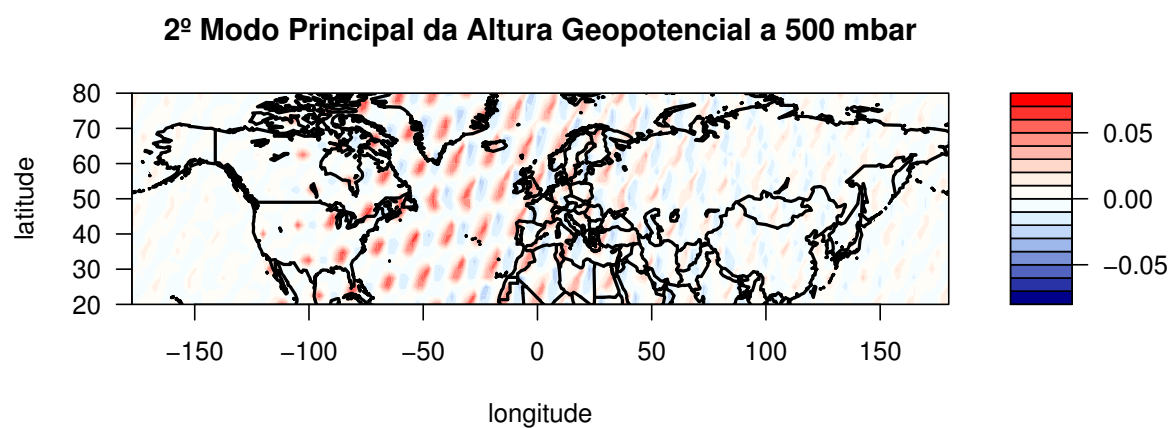


Figura 7.19: 2º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E

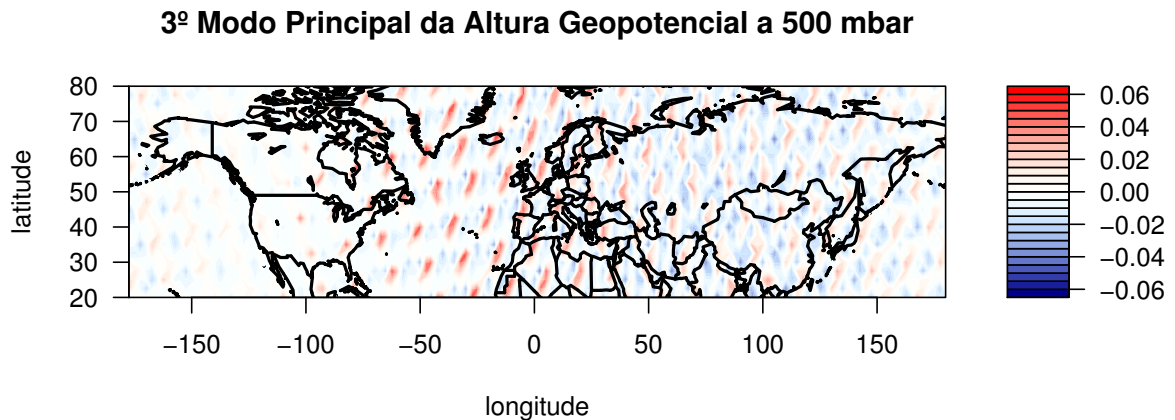


Figura 7.20: 3º modo principal de variação das anomalias da altura geopotencial a 500 milibar, observada no setor 20°N - 80°N e 0°E - 357.5°E

da Oscilação do Atlântico Norte. Para além disso, ainda é possível constatar que a correlação entre a CP dominante correspondente ao primeiro modo principal de variação e o Índice de Oscilação do Atlântico Norte é cerca de 0.62.

7.2.3 Resumo

Algumas conclusões relativamente às análises e baseadas em determinados indicadores da convergência de resultados, para os dados da altura geopotencial em estudo, são descritas de seguida.

Através das correlações entre CPEs e CIEs foi possível detetar correlações elevadas entre determinados pares de CPEs e CIEs que foram estabelecidos, cujos valores, em módulo, variaram de um modo geral entre 0.4 e 0.6.

Os cossenos entre vetores permitiram identificar convergência de resultados nas relações existentes entre os coeficientes dos vetores dos pares mais semelhantes entre ambas as técnicas, uma vez que estão de acordo com os resultados das correlações entre os pares das correspondentes componentes. Desta forma, os valores mais elevados dos cossenos surgem, na generalidade, associados aos pares cujas respetivas componentes apresentam correlações mais elevadas.

Na análise dos períodos dominantes, de acordo com os espectros das componentes extraídas, constatou-se que ambas as técnicas da MSSA e da *LagACI* (quer através do FastICA quer através do AMUSE) detetaram períodos em torno dos 43 meses e dos 33 meses. De um modo geral, as

20 CIEs obtidas através da *LagACI* detetaram os mesmos períodos dominantes identificados na CPEs através da MSSA, em particular nos sete pares de CPEs que foram selecionados. Ainda foi possível estabelecer algumas relações entre o espectro do NAOI e os espectros das componentes estendidas obtidas através das duas técnicas.

Relativamente à construção de alguns mapas referentes à altura geopotencial, aquando da aplicação da ACP usual na matriz das anomalias, e sem desfasamentos, ilustrámos os três principais modos de variação global que ocorrem ao nível do Atlântico Norte, onde sobressaem algumas evidências da relação destes com o padrão da NAO. Destacamos também a correlação entre a CP dominante e o Índice da NAO que é aproximadamente 0.62, o que reforça a presença da NAO no Hemisfério Norte.

Para o caso concreto da altura geopotencial, a *LagACI* pode então ser considerada como alternativa à MSSA para o estudo de séries temporais estendidas, uma vez que é capaz de detetar períodos idênticos para as oscilações mais relevantes que descrevem o comportamento dos dados. Deste modo, em termos de interpretação dos padrões de variabilidade dominantes que ocorrem no sistema climático, não existem evidências de que a *LagACI* permita obter melhores resultados relativamente à MSSA, mas podemos argumentar a existência da convergência de resultados entre técnicas. Contudo, a constatação de uma convergência de resultados para este conjunto de dados que apresentam uma elevada complexidade, não significa que o mesmo aconteça noutros conjuntos de dados.

Dado que a aplicação da ACI em estudos meteorológicos ainda é relativamente recente, novas formas de leitura poderão auxiliar a interpretar a complexa informação dominante no sistema climático, pelo que outras abordagens e aplicações futuras poderão vir a ser determinantes na extração dessa informação.

7.3 Conclusões Finais

Nas notas introdutórias da tese foram descritos os aspetos teóricos da ACP clássica e algumas das suas derivações no tratamento de dados multivariados. A ACI, técnica amplamente usada em processamento de sinal e de imagem, começa a ganhar expressão em aplicações onde se admite que as fontes de dados (não diretamente observáveis) são estatisticamente independentes. Neste sentido surgiu a ideia de efetuar um estudo aprofundado acerca das bases da ACI para perceber

as suas potencialidades com o intuito de a comparar e interligar com a ACP e outras técnicas afins.

A ACP tem como objetivo reduzir a dimensão de um conjunto de dados, obtendo as CPs através de combinações lineares das variáveis originais, de forma a que as CPs sejam não correlacionadas entre si e possuam sucessivas maximizações da variância. Por outro lado, e partindo do pressuposto que a não correlação não leva necessariamente à independência, a ACI tem como objetivo principal separar as componentes de modo que sejam independentes em termos estatísticos, o que traduz uma condição mais forte em relação à não correlação, fazendo com que as componentes independentes revelem entre si, em alguns casos, mais informação do que as CPs usuais.

Deste modo foram descritos os suportes teóricos da ACI, nomeadamente os métodos de estimação do modelo, alguns dos algoritmos existentes para implementar a estimação, as técnicas de pré-processamento dos dados como a centragem e branqueamento. Uma vez que as componentes independentes são extraídas sem uso de qualquer critério de ordenação aparente, foram apresentados alguns dos métodos mais usuais que permitem ordená-las. Para além da descrição do modelo básico de ACI e da nova abordagem da ACI proposta neste trabalho como alternativa à MSSA, denominada de *LagACI*, foram ainda mencionados outros tipos de modelos ACI. Foi também descrito o estado da arte para compreender as direções em que a ACI se desenvolve.

O estudo do clima é um tema que suscita o interesse da comunidade científica e das entidades governamentais. Nas últimas décadas, a investigação nessa área tem evoluído no sentido da aposta em novas metodologias alternativas, que permitam verificar alguma nova utilidade relativamente às técnicas usuais. Como tal desenvolveu-se esta tese, com particular relevância para aplicações na área da climatologia, visando encontrar uma técnica que pudesse através da ACI aperfeiçoar a interligação com algumas das já existentes, nomeadamente na comparação e interpretação de resultados.

Numa primeira abordagem foi explorado um conjunto de séries temporais multivariadas de dados meteorológicos (valores médios mensais de pressão ao nível do mar), onde se salientaram algumas diferenças e semelhanças entre as técnicas da ACP e da ACI, ao nível dos domínios do tempo e da frequência. Devido à estrutura temporal intrínseca dos dados, foi ainda possível estabelecer uma relação direta entre a informação das autocorrelações das componentes dominantes e a série temporal do Índice de *El Niño* (dados de temperaturas ao nível do mar).

Implementaram-se os algoritmos FastICA e AMUSE, tendo-se verificado que, na generalidade dos resultados práticos e em destaque nas reconstruções, o algoritmo AMUSE apresentou um desempenho ligeiramente superior ao do FastICA. Tais resultados permitem corroborar os argumentos teóricos de que, mediante a presença de dados que possuam uma dependência temporal, é preferível utilizar o AMUSE que contempla a estrutura temporal de autocovariâncias, em detrimento do FastICA.

Os dados foram ainda comparados com a série do Índice de Oscilação Sul para a pressão ao nível do mar no Oceano Pacífico. De um modo geral, não se estabeleceram semelhanças entre o espectro do Índice de Oscilação Sul e os espectros das coordenadas, do conjunto de dados da pressão, obtidos quer pela ACP quer pela ACI. Por outro lado, constatou-se que os índices de *El Niño* e de Oscilação Sul apresentam o mesmo período de quatro anos e meio como característica da ocorrência média do fenómeno ENSO.

Uma vez que, ao usarmos séries temporais, a ACP não contempla a estrutura temporal de correlação entre as observações, alguns recentes desenvolvimentos a partir da ACP têm surgido com o objetivo de aproveitar a estrutura interna das observações ao longo do tempo, tais como a SSA e a MSSA.

A SSA foi introduzida como método particular de aplicação da ACP, cujas variáveis iniciais são versões desfasadas da série temporal unidimensional, cujo objetivo é decompor a série original nas componentes oscilatórias, na tendência e na estrutura de ruído. Foram abordados alguns dos vários princípios gerais para escolher o desfasamento (comprimento da janela) mais adequado. Para uma série estacionária ou quase estacionária, as componentes principais relacionam-se com os padrões principais obtidos pela decomposição espectral da série. Neste caso os valores e vetores próprios dominantes vão identificar os picos de frequência mais elevados que traduzem os períodos mais relevantes da série. Mesmo que a série não seja estacionária, pode aplicar-se a SSA usando métodos para remover a tendência como a centragem por linhas ou a dupla centragem, captando na mesma as restantes componentes oscilatórias mais relevantes.

O modelo básico da SSA não contempla a centragem dos dados, contudo podemos considerar vários tipos de centragem. Em relação à discussão em torno da centragem ou não centragem dos dados na técnica SSA, foi apresentado um novo tipo de centragem aplicado na série original (antes dos dados desfasados) e descritos os respetivos conceitos teóricos.

Neste contexto, como exemplo de aplicação foi explorada a série mensal do número de vítimas mortais em acidentes de viação em Portugal Continental ao longo de 12 anos, na qual foram aplicados vários tipos de centragem. Constatou-se que as centragens por linhas e dupla removeram a tendência aos dados. Por outro lado, o modelo da SSA com centragem na série original destaca os comportamentos associados à tendência nas duas primeiras componentes dominantes. Todos os modelos de SSA, desde o básico sem centragem até aos que envolveram centragem, incluindo o novo tipo de centragem na série original, permitiram captar os mesmos pares de componentes oscilatórias (entre as primeiras 16) que ostentavam a maior fonte de variabilidade dos dados, e por sua vez associadas aos mesmos períodos dominantes. Também os indicadores utilizados na comparação dos vetores próprios e das componentes principais, entre o modelo básico de SSA e cada um dos quatro modelos com centragem abordados, permitiram evidenciar convergências de resultados.

Apesar das CPs obtidas pelos vários tipos de centragem em SSA serem distintas, a informação dominante que é captada pela SSA sem centragem, pela SSA com centragem simples por colunas e pela SSA com centragem na série original, é semelhante pois as colunas das respetivas matrizes dos desfasamentos dos dados possuem exatamente as mesmas variâncias, devido ao facto da variância de uma variável centrada ser igual à variância da mesma variável não centrada.

A MSSA, versão multivariada de SSA, é usada para o estudo simultâneo de várias séries temporais, com informação extraída das versões desfasadas de cada variável (série temporal). Os principais objetivos da SSA continuam a ser contemplados, mas a MSSA pretende ainda obter a informação conjunta das inter-relações entre observações das variáveis e das inter-relações entre as variáveis originais desfasadas no tempo. Por exemplo ao nível da climatologia, a MSSA pretende também identificar padrões no espaço e de evolução no tempo com base em mapas de clima. Os conceitos teóricos subjacentes à técnica da MSSA foram apresentados, como uma generalização da SSA, ou seja como uma ACP aplicada na matriz dos desfasamentos das séries originais.

Foi proposta uma nova abordagem na síntese de dados, recorrendo à ACI aplicada na matriz dos desfasamentos como alternativa ao método de MSSA no estudo de séries temporais estendidas: a *LagACI*. Apresentaram-se os principais fundamentos da técnica recorrendo aos conceitos teóricos da MSSA e da ACI.

Para o mesmo conjunto de dados da pressão ao nível do mar usados no Capítulo da ACI, aplicaram-se as técnicas MSSA, *LagACI* e *LagACI* após a utilização da ACP na matriz dos desfasamentos, para comparação de resultados. Analisaram-se cinco métodos diferentes de ordenação das CIs, incluindo o novo método que foi proposto baseado nas correlações entre as CPs e as CIs, que se destacou na minimização das somas de quadrados dos resíduos das reconstruções dos dados originais como sendo um dos melhores. As coordenadas das componentes forneceram resultados similares entre as três técnicas para o mesmo desfasamento temporal. As análises comparativas das reconstruções dos dados originais, através das somas de quadrados dos resíduos, exibiram resultados semelhantes para a MSSA e *LagACI* e um pouco mais satisfatórios para a terceira técnica. Também as análises dos espectros das componentes permitiram concluir que as interpretações das periodicidades predominantes são similares entre as três técnicas.

Ao analisar os indicadores adicionais do desempenho das técnicas em estudo (que comparam as reconstruções dos dados com as respectivas observações), verificamos a existência de semelhanças nos respectivos valores entre técnicas. Também existem evidências de semelhanças relativamente ao comportamento de cada indicador consoante o número de componentes (de uma até cinco) utilizadas na reconstrução. Constatou-se que o índice de Willmott d forneceu valores mais elevados de concordância em comparação ao índice de Willmott revisto d_1 , enquanto que o coeficiente de Nash e Sutcliffe E foi o que evidenciou maior concordância entre os coeficientes de eficiência analisados. Portanto podemos argumentar que, de um modo geral, os vários conjuntos de dados usados nas reconstruções apresentam resultados consistentes entre técnicas, para os vários indicadores, uma vez que não existem evidências suficientes para argumentar que um modelo associado a uma técnica exiba melhor desempenho em relação a um modelo de uma outra técnica.

Finalmente, foi apresentada uma síntese comparativa das técnicas discutidas ao longo da tese, relativamente aos vários passos e estruturas matriciais subjacentes à aplicação das mesmas, aos métodos de estimação e reconstrução dos dados originais.

A elevada complexidade como critério de escolha de um conjunto de dados, constituído por valores médios mensais de altura geopotencial a 500 milibar registados no Hemisfério Norte, ao longo de 50 anos, foi usada para o estudo de séries temporais estendidas, através da comparação transversal das técnicas da MSSA e *LagACI*. Após a aplicação da ACP como pré-processamento

dos dados originais, ao contemplar 19 canais (componentes retidas) e ao aplicar a MSSA e a *LagACI* na matriz dos dados desfasados com um desfasamento de 15 anos, foram obtidos resultados similares em relação aos períodos temporais dominantes, considerando as 20 componentes estendidas extraídas para ambas as técnicas. Foram também estabelecidas semelhanças ao nível dos espectros entre o Índice de Oscilação do Atlântico Norte, como padrão dominante no sistema climático da variabilidade da pressão atmosférica ao nível do mar, e as componentes estendidas obtidas pelas duas técnicas para a altura geopotencial. Para além disso, a influência dos ciclos anual e semestral na pressão ao nível do mar que ocorre no Atlântico Norte apresenta resultados convergentes no domínio da frequência na altura geopotencial do Hemisfério Norte.

Podemos concluir que a ACI, neste contexto, revela ser uma alternativa à ACP no estudo de séries temporais estendidas, extraindo a mesma informação existente nos dados, ao detetar períodos idênticos para as principais oscilações que permitem descrever o comportamento temporal dos mesmos. Portanto, para dados complexos, podemos argumentar que até para períodos temporais longos de várias décadas, algumas das relações dos padrões obtidos pelas técnicas podem ser validadas.

7.4 Trabalho Futuro

Relativamente a trabalho de investigação futuro, um dos caminhos a enveredar consiste no interesse em construir mapas de clima, para certas variáveis meteorológicas, consoante as diversas abordagens exploradas nesta tese e eventualmente outras novas que possam ser consideradas complementares, recorrendo às técnicas multivariadas da ACP, da ACI, da MSSA e da *LagACI*.

Um tipo particular de gráfico de contornos usado para representar dados meteorológicos através da propagação sob a forma de ondas é o Diagrama de Hovmöller, no qual os eixos representam geralmente a variável espacial longitude ou latitude (no eixo dos x) e a variável de tempo (no eixo dos y). Estes diagramas são muito úteis para determinar a velocidade de várias perturbações atmosféricas, no tempo e no espaço, tais como temperatura, pressão, radiação, velocidade do vento, concentração de constituintes na atmosfera ou no mar, A construção de Diagramas de Hovmöller pode ser utilizada para ilustrar as componentes principais estendidas obtidas segundo a MSSA para desfasamentos no tempo num determinado intervalo espacial, como por exemplo em Hannachi *et al.* (2007). A sua aplicação no contexto da ACI para

matrizes de dados dos desfasamentos seria também um assunto a explorar.

Um outro assunto que ainda pode ser alvo de investigação futura recai sobre a escolha do comprimento da janela utilizado na SSA e MSSA, uma vez que atualmente não existe nenhum indicador ou princípio concreto que permita deduzir qual o valor "ótimo" que deva ser usado em função da dimensão do conjunto de dados. Talvez a simulação, por ser uma ferramenta que, geralmente, ajuda a consolidar a implementação de novas abordagens baseadas noutras existentes, nomeadamente na comparação de desempenhos entre modelos e por vezes recorrendo a indicadores estatísticos, seja uma possibilidade.

Uma vez que as técnicas SSA e MSSA permitem a reconstrução "com sucesso" de séries com valores omissos, outras possíveis abordagens, eventualmente interligadas à ACI, poderão passar pelo uso de dados com valores omissos ou não equitativamente distanciados.

A técnica designada por Análise em Componentes Principais Independentes (ACPI) proposta por Yao *et al.* (2012), a qual combina as vantagens da ACP e da ACI, foi criada com o intuito de tratar conjuntos de dados biológicos de grande dimensão que ostentem um ruído considerável (em que muitas vezes o número de observações é muito menor que o número de variáveis biológicas em estudo). Neste contexto, o seu objetivo é obter componentes independentes que reflitam melhor os aspetos biológicos subjacentes ao estudo assim como conseguir uma redução da dimensão de forma mais eficaz do que apenas aplicando a ACP ou a ACI. Transpor a ACPI para a realidade do sistema climático complexo, onde se considera que atuam diversas componentes físicas independentes (não necessariamente observáveis), poderá ser útil no tratamento de várias variáveis meteorológicas em simultâneo.

Também o campo de ampla aplicação da Biomedicina pode ser um caminho a explorar, introduzindo a estrutura temporal da MSSA conjugada com a ACI, na análise de dados resultantes de exames médicos associados ao acompanhamento de pacientes em várias consultas (dados em painel).

Apêndice A

Notação

A.1 Glossário dos Principais Símbolos para ACP

- \mathbf{B} - matriz dos vetores próprios da matriz de variâncias-covariâncias de \mathbf{Y}
- \mathbf{I}_p - matriz identidade de ordem p
- \mathcal{L} - função Lagrangeana
- \mathbf{L} - matriz diagonal cujos elementos da diagonal principal são os valores próprios
- n - n.º de indivíduos ou observações
- p - n.º de variáveis aleatórias originais e n.º de componentes principais
- \mathbf{p}_k - vetor próprio associado ao k -ésimo maior valor próprio
- p_{jk} - coeficiente associado à variável \mathbf{x}_j no k -ésimo vetor próprio
- \mathbf{P} - matriz cujas colunas são os vetores próprios ortonormados
- q - n.º de componentes principais retidas
- \mathbb{R}^n - espaço vetorial de dimensão n
- \mathbf{R} - matriz de correlações da matriz \mathbf{X}
- r_{jk} - correlação entre as variáveis \mathbf{x}_j e \mathbf{x}_k

- \mathbf{S} - matriz de variâncias-covariâncias da matriz \mathbf{X}
- s_j^2 - variância da variável \mathbf{x}_j
- s_{jk} - covariância entre as variáveis \mathbf{x}_j e \mathbf{x}_k
- $\bar{\mathbf{x}}^T$ - vetor das médias da matriz \mathbf{X}
- \mathbf{x}_i - i -ésima variável inicial
- \bar{x}_j - média da variável \mathbf{x}_j
- \mathbf{X} - matriz de n observações nas p variáveis
- $\tilde{\mathbf{X}}_q$ - matriz dos dados reconstruída à custa das primeiras q componentes principais
- \mathbf{Y} - matriz cujas colunas são funções lineares das colunas de \mathbf{X}
- \mathbf{z}_k - k -ésima componente principal
- z_{hk} - coordenada da h -ésima observação na k -ésima componente principal
- \mathbf{Z} - matriz cujas colunas são as componentes principais
- λ e μ - multiplicadores de Lagrange
- λ_k - k -ésimo valor próprio

A.2 Glossário dos Principais Símbolos para ACI

- \mathbf{A} - matriz dos coeficientes das misturas, a_{ij} ($i, j = 1, 2, \dots, n$)
- \mathbf{A}^+ - matriz pseudo-inversa da matriz das misturas \mathbf{A}
- \mathbf{b} - vetor dos coeficientes da combinação linear de \mathbf{x}
- \mathbf{B} - matriz auxiliar para forçar as covariâncias a tenderem para zero
- cov - covariância
- cum - cumulante

- $curt$ - curtose de uma variável aleatória
- $C_{\tau}^{\mathbf{s}}$ - matriz de covariâncias dos desfasamentos no tempo para o vetor $\mathbf{s}(t)$
- $C_{\tau}^{\mathbf{x}}$ - matriz de covariâncias dos desfasamentos no tempo para o vetor $\mathbf{x}(t)$
- $C_{\tau}^{\mathbf{z}}$ - matriz de covariâncias dos desfasamentos no tempo para o vetor $\mathbf{z}(t)$
- $\bar{C}_{\tau}^{\mathbf{z}}$ - matriz simétrica com base na matriz de covariâncias $C_{\tau}^{\mathbf{z}}$
- \det - determinante de uma matriz
- $diag$ - matriz diagonal
- \mathbf{D} - matriz diagonal dos valores próprios positivos de $E(\mathbf{xx}^T)$
- \exp - função exponencial
- E - valor esperado
- \mathbf{E} - matriz ortogonal dos vetores próprios de $E(\mathbf{xx}^T)$
- f - função densidade de probabilidade
- \mathbf{f} - transformação não linear desconhecida das misturas
- F - função de otimização da curtose
- \mathbf{F} - função definida numa rede neuronal
- G - função não quadrática
- \mathbb{G} - subespaço gerado pelas q componentes principais
- \mathbf{h} - transformação
- H - entropia diferencial de uma variável aleatória
- \mathbb{H} - subespaço gerado pelas m componentes independentes
- I - informação mútua de n variáveis aleatórias
- \mathbf{I} - matriz identidade

- J - neguentropia de uma variável aleatória
- k - índice de contagem do n^o de componentes independentes a estimar
- K - n^o de observações de \mathbf{x}
- $K(\cdot)$ - complexidade de Kolmogoroff
- \log - função logaritmo
- \mathcal{L} - função Lagrangeana
- L - função verosimilhança
- m - n^o de componentes independentes a estimar
- \mathbf{M} - matriz de transformação de \mathbf{x}
- n - n^o de variáveis aleatórias / n^o de componentes independentes
- \mathbf{n} - vetor aditivo de ruído branco
- p - n^o de misturas
- $p_U(\cdot)$ - algoritmo ou programa computacional que gera uma sequência de caracteres
- $\mathbf{P}_{\mathbb{G}}$ - matriz de projeção ortogonal no subespaço \mathbb{G}
- $\mathbf{P}_{\mathbb{H}}$ - matriz de projeção ortogonal no subespaço \mathbb{H}
- q - n^o de componentes independentes retidas
- \mathbf{q} - vetor tal que $\mathbf{q}^T = \mathbf{b}^T \mathbf{A}$
- \mathbf{s} - vetor aleatório das componentes independentes s_1, s_2, \dots, s_n
- $\tilde{\mathbf{s}}$ - vetor aleatório das componentes independentes com ruído $\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_n$
- \mathbf{s}' - vetor aleatório das componentes independentes originais (não centradas)
- \mathbf{S} - matriz das componentes independentes
- U - universo de todos os programas computacionais

- \mathbf{U} - matriz ortogonal
- \mathbf{V} - matriz do branqueamento
- x_i - i -ésima variável
- \mathbf{x} - vetor aleatório das misturas x_1, x_2, \dots, x_n
- \mathbf{x}' - vetor aleatório das misturas originais (não centradas)
- \mathbf{X} - matriz dos dados observados (misturas)
- y - combinação linear das variáveis x_1, x_2, \dots, x_n
- \mathbf{y} - vetor de componentes não correlacionadas e de variâncias unitárias
- \mathbf{z} - vetor branqueado
- w - vetor associado à matriz \mathbf{W}
- \mathbf{W} - inversa da matriz \mathbf{A}
- Σ - matriz de covariâncias de um vetor aleatório com distribuição Normal
- λ - multiplicador de Lagrange / valor próprio de uma matriz
- ϕ_i - i -ésima função escalar não linear
- τ - desfasamento no tempo

A.3 Glossário dos Principais Símbolos para SSA e MSSA

- \mathbf{a} - média de todos os valores de uma matriz
- A - amplitude de onda
- Cq - conjunto de q CPEs a reter
- $c_{\alpha/2}^-$ - quantil inferior para $\alpha/2$
- $c_{\alpha/2}^+$ - quantil superior para $\alpha/2$

- car - característica de uma matriz
- cov - covariância
- d - característica da matriz \mathbf{X}
- d - índice de concordância de Willmott
- d_1 - índice de concordância de Willmott revisto
- d_r - índice de concordância de Willmott refinado
- d^{cc} - característica da matriz \mathbf{X}^{cc}
- d^{cl} - característica da matriz \mathbf{X}^{cl}
- d^{**} - característica da matriz \mathbf{X}^{**}
- d' - característica da matriz \mathbf{X}'
- E - coeficiente de eficiência de Nash e Sutcliffe
- E_1 - coeficiente de eficiência de Legates e McCabe
- F_1 e F_2 - constantes
- h - translação no tempo
- i - índice de tempo
- k - n^o de vetores desfasados de dimensão m
- m - comprimento da janela (lag)
- n - comprimento da série temporal
- O_i - observação
- p - n^o de séries temporais
- p' - n^o de colunas da matriz \mathbf{X} (em MSSA)
- P_i - estimativa ou previsão

- q - n^o de CPs a reter
- s - unidades de desfasamento no tempo
- \mathbf{S} - matriz de variâncias-covariâncias da matriz \mathbf{X}
- \mathbf{S}_{ij} - matriz de variâncias-covariâncias dos desfasamentos entre duas séries temporais
- \mathbf{S}^k - matriz dada por $\frac{1}{k}\mathbf{X}^T\mathbf{X}$
- \mathbf{S}^{cc} - matriz para efetuar a SVD para a centragem por colunas a partir da matriz \mathbf{X}^{cc}
- \mathbf{S}^{cl} - matriz para efetuar a SVD para a centragem por linhas a partir da matriz \mathbf{X}^{cl}
- \mathbf{S}^{**} - matriz para efetuar a SVD para a centragem dupla a partir da matriz \mathbf{X}^{**}
- \mathbf{S}' - matriz para efetuar a SVD para a centragem na série original a partir da matriz \mathbf{X}'
- \mathbf{S}_{BK} - estimador de *Broomhead & King*
- \mathbf{S}_{VG} - estimador de *Vautard & Ghil*
- t - instante temporal
- tr - traço de uma matriz
- T - período de uma componente oscilatória
- U_i - i -ésimo vetor próprio da matriz $\mathbf{X}^T\mathbf{X}$
- U_i^∇ - i -ésimo vetor dos primeiros $m - 1$ elementos do vetor próprio U_i
- U_i^{cc} - i -ésimo vetor próprio da matriz \mathbf{S}^{cc}
- U_i^{cl} - i -ésimo vetor próprio da matriz \mathbf{S}^{cl}
- U_i^{**} - i -ésimo vetor próprio da matriz \mathbf{S}^{**}
- U_i' - i -ésimo vetor próprio da matriz \mathbf{S}'
- V_i - i -ésima CP da matriz \mathbf{X}
- V_i^{cc} - i -ésima CP da matriz \mathbf{X}^{cc}

- V_i^{cl} - i -ésima CP da matriz \mathbf{X}^{cl}
- V_i^{**} - i -ésima CP da matriz \mathbf{X}^{**}
- V_i' - i -ésima CP da matriz \mathbf{X}'
- \bar{x} - média da série temporal X_t
- \hat{x}_i - previsão para um valor da série no instante temporal i
- x_j - j -ésima variável
- \bar{x}_j - média da variável x_j
- \mathbf{x}_t - vetor de séries temporais
- $\bar{\mathbf{x}}$ - vetor em que cada um dos elementos é a média da série
- $\bar{\mathbf{x}}_c$ - vetor das médias por coluna da matriz \mathbf{X}
- $\bar{\mathbf{x}}_l$ - vetor das médias por linha da matriz \mathbf{X}
- $\overline{\mathbf{x}}^{cc}_l$ - vetor das médias por linha da matriz previamente centrada por colunas \mathbf{X}^{cc}
- $\bar{\mathbf{x}}_{dif}$ - vetor da diferença entre a média dos valores de cada coluna de \mathbf{X} e a média da série
- \mathbf{X} - matriz de trajetória (não centrada)
- \mathbf{X}^{cc} - matriz de trajetória centrada por colunas
- \mathbf{X}^{cl} - matriz de trajetória centrada por linhas
- \mathbf{X}^{**} - matriz de trajetória centrada por colunas e por linhas (centragem dupla)
- \mathbf{X}' - matriz de trajetória centrada na série original
- $\bar{\mathbf{X}}$ - matriz cujos elementos são todos iguais à média da série
- $\bar{\mathbf{X}}_c$ - matriz cujos elementos de cada coluna são iguais à média dos valores da coluna da matriz \mathbf{X}
- $\bar{\mathbf{X}}_l$ - matriz cujos elementos de cada linha são iguais à média dos valores da linha da matriz \mathbf{X}

- $\overline{\mathbf{X}}_{cl}$ - matriz em que cada elemento é igual ao valor médio de todos os elementos da matriz \mathbf{X}
- $\overline{\mathbf{X}}^{cc}_l$ - matriz cujos elementos de cada linha são iguais à média dos valores da linha da matriz \mathbf{X}^{cc}
- $\overline{\mathbf{X}}_{dif}$ - matriz da diferença entre as matrizes $\overline{\mathbf{X}}_c$ e $\overline{\mathbf{X}}$
- X_t - série temporal
- X_t^C - série constante
- X_t^0 - série que oscila em torno de zero
- $X_t^{(1)}$ - sinal da série temporal X_t
- $X_t^{(2)}$ - ruído da série temporal X_t
- \tilde{X}_t - série temporal reconstruída
- $\{X_t\}$ - processo estocástico
- \mathbf{Y} - matriz genérica
- W - onda progressiva
- $\mathbf{1}_k$ - vetor com todos os k elementos iguais à unidade
- α - nível de significância
- α_1 e α_2 - constantes apropriadas
- $\Gamma_{\mathbf{X}}$ - matriz de covariâncias de \mathbf{X}
- Δ - variação
- λ - comprimento de onda
- λ_i - i -ésimo valor próprio da matriz $\mathbf{X}^T \mathbf{X}$
- λ_i^{cc} - i -ésimo valor próprio da matriz \mathbf{S}^{cc}

- λ_i^{cl} - i -ésimo valor próprio da matriz \mathbf{S}^{cl}
- λ_i^{**} - i -ésimo valor próprio da matriz \mathbf{S}^{**}
- λ_i' - i -ésimo valor próprio da matriz \mathbf{S}'
- $\sqrt{\lambda_i}$ - i -ésimo valor singular da matriz \mathbf{X}
- $\{\sqrt{\lambda_i}\}$ - espectro da matriz \mathbf{X}
- μ - n^o de vezes que ocorre o λ por unidade de distância (frequência espacial)
- π_i - último elemento do vetor próprio U_i
- ρ_1 e ρ_2 - funções trigonométricas
- ϕ - fase inicial da onda
- ω - frequência angular

Apêndice B

Lista de Acrónimos

- ACC: Análise das Correlações Canônicas
- ACI: Análise em Componentes Independentes
- ACP: Análise em Componentes Principais
- ACPI: Análise em Componentes Principais Independentes
- AMUSE: *Algorithm for Multiple Unknown Signals Extraction*
- ANACOR: *Correspondence Analysis* (Análise de Correspondências)
- ANSR: Autoridade Nacional de Segurança Rodoviária
- BSS: *Blind Source Separation*
- cICA: *constrained Independent Component Analysis* (Análise em Componentes Independentes com restrição)
- CATPCA: *Categorical Principal Component Analysis* (ACP para dados Categóricos)
- CDMA: *Code Division Multiple Access*
- CIEs: Componentes Independentes Estendidas
- CIs: Componentes Independentes
- CPEs: Componentes Principais Estendidas

- CPs: Componentes Principais
- CSSA: *Complex Singular Spectrum Analysis* (Análise Espectral Singular Complexa)
- DGV: Direção Geral de Viação
- DOGMA: *Diagonalization Of Generalized covariance MAtrices*
- EAM: Erro Absoluto Médio
- EEOF: *Extended Empirical Orthogonal Functions Analysis* (Análise de Funções Ortogonais Empíricas Estendidas)
- ENSO: *El Niño/Southern Oscillation* (Oscilação Sul do *El Niño*)
- EOF: *Empirical Orthogonal Functions Analysis* (Análise de Funções Ortogonais Empíricas)
- EOFs: Funções Ortogonais Empíricas
- EOTs: Teleconexões Ortogonais Empíricas
- EQM: Erro Quadrático Médio
- FOBI: *Fourth-Order Blind Identification*
- FPCA: *Functional Principal Component Analysis* (Análise em Componentes Principais Funcional)
- GCD - *Generalized Coefficient of Determination* (Coeficiente de Determinação Generalizado)
- HOMALS: *Homogeneity Analysis* (Análise de Homogeneidade)
- INE: Instituto Nacional de Estatística
- JADE: *Joint Approximate Diagonalization of Eigenmatrices*
- LagACI: *Lag*-Análise em Componentes Independentes
- MEM: Método de Máxima Entropia

- MTM: Método *Multitaper*
- MSSA: *Multichannel Singular Spectrum Analysis* (Análise Espectral Singular Multicanal)
- NAO: *North Atlantic Oscillation* (Oscilação do Atlântico Norte)
- NAOI: *North Atlantic Oscillation Index* (Índice de Oscilação do Atlântico Norte)
- NOAA: *National Oceanic and Atmospheric Administration*
- OMS: Organização Mundial de Saúde
- ONU: Organização das Nações Unidas
- OVERALS: *Nonlinear Canonical Correlation Analysis*
- PNA: *Pacific/North American* (Pacífico-América do Norte)
- POP: *Principal Oscillation Patterns Analysis* (Análise dos Padrões de Oscilação Principal)
- PRINCALS: *Principal Component Analysis by means of Alternating Least Squares*
- PSPA: *Principal Sequence Pattern Analysis*
- REQM: Raiz do Erro Quadrático Médio
- SCoT: *Simplified Component Technique*
- SNR: *Signal-to-Noise Ratio* (Relação sinal/ruído)
- SOBI: *Second-Order Blind Identification*
- SOI: *Southern Oscillation Index* (Índice de Oscilação Sul)
- SPSS: *Statistical Package for the Social Sciences*
- SQR: Somas de Quadrados dos Resíduos
- SSA: *Singular Spectrum Analysis* (Análise Espectral Singular)
- SSA-MTM: *Singular Spectrum Analysis - MultiTaper Method*
- SVD: Decomposição em Valores Singulares

- v. a.: variável aleatória
- WPCA: *Weighted Principal Component Analysis* (Análise em Componentes Principais Ponderada)

Apêndice C

Apontamentos Teóricos

Complementares

C.1 Tópicos Elementares de Vetores e Matrizes

Pretendemos introduzir alguns conceitos elementares ao nível algébrico, nomeadamente sobre vetores e matrizes, os quais podem ser encontrados em inúmeros livros de texto, como por exemplo, em Anton e Rorres (2010) ou em Horn e Johnson (1985a, 1985b).

C.1.1 Vetores

Nesta parte apresentamos alguns conceitos básicos relacionados com vetores.

Definições C.1.1 *Sejam $\mathbf{u} = (u_1, u_2, \dots, u_n)$ e $\mathbf{v} = (v_1, v_2, \dots, v_n)$ dois vetores não nulos de \mathbb{R}^n .*

- *Designa-se por **produto escalar** ou **produto interno** de \mathbf{u} e \mathbf{v} , ao valor*

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n.$$

- *Designa-se por **norma** ou **comprimento** de um vetor \mathbf{u} , ao valor definido por*

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2} = \sqrt{\mathbf{u} \cdot \mathbf{u}}.$$

- *Diz-se que um vetor \mathbf{u} é **unitário** quando $\|\mathbf{u}\| = 1$.*

- O **ângulo** entre dois vetores \mathbf{u} e \mathbf{v} , $\angle(\mathbf{u}, \mathbf{v})$, define-se por

$$\angle(\mathbf{u}, \mathbf{v}) = \arccos \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \right).$$

Logo $0 \leq \angle(\mathbf{u}, \mathbf{v}) \leq \pi$.

- Os vetores \mathbf{u} e \mathbf{v} designam-se por **ortogonais** ou **perpendiculares**, $\mathbf{u} \perp \mathbf{v}$, quando $\angle(\mathbf{u}, \mathbf{v}) = \frac{\pi}{2}$, ou seja, quando $\mathbf{u} \cdot \mathbf{v} = 0$.

Definição C.1.2 Dois ou mais vetores dizem-se **ortonormados** quando todos os vetores são unitários e todos os pares de vetores são ortogonais.

Descrevemos de seguida alguns conceitos relacionados com espaços vetoriais.

Definição C.1.3 Seja E um conjunto não vazio. Diz-se que $(E, +, \cdot)$ é um **espaço vetorial real** sobre o conjunto dos números (também designados por **escalares**) reais quando estão definidas duas operações: a adição $(+)$ e a multiplicação por um escalar (\cdot) .

A adição tem as seguintes propriedades de grupo comutativo, para $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in E$:

- Comutatividade: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$;
- Associatividade: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$;
- Existência de elemento neutro ($\mathbf{0}_E$): $\mathbf{u} + \mathbf{0}_E = \mathbf{0}_E + \mathbf{u} = \mathbf{u}$;
- Existência de elemento simétrico: $\forall \mathbf{u} \in E, \exists \mathbf{u}' = (-\mathbf{u}) \in E : \mathbf{u} + \mathbf{u}' = \mathbf{u}' + \mathbf{u} = \mathbf{0}_E$.

A multiplicação por um escalar tem as seguintes propriedades, para $\forall \mathbf{u}, \mathbf{v} \in E, \forall \alpha, \beta \in \mathbb{R}$:

- Distributividade em relação à adição de E : $\alpha \cdot (\mathbf{u} + \mathbf{v}) = \alpha \cdot \mathbf{u} + \alpha \cdot \mathbf{v}$;
- Distributividade em relação à adição de escalares: $(\alpha + \beta) \cdot \mathbf{u} = \alpha \cdot \mathbf{u} + \beta \cdot \mathbf{u}$;
- Associatividade mista: $\alpha \cdot (\beta \cdot \mathbf{u}) = (\alpha\beta) \cdot \mathbf{u}$;
- Unidade: $1 \cdot \mathbf{u} = \mathbf{u}$.

A **adição** associa a cada par de elementos (**vetores**) \mathbf{u} e \mathbf{v} no conjunto E , um elemento $\mathbf{u} + \mathbf{v} \in E$.

A **multiplicação por um escalar** associa a cada escalar $\alpha \in \mathbb{R}$ e cada elemento $\mathbf{u} \in E$, um elemento $\alpha \cdot \mathbf{u} \in E$.

Definição C.1.4 Seja S um subconjunto não vazio de um espaço vetorial E . Diz-se que S é um **subespaço vetorial** de E quando S , com as operações definidas para E , ainda é um espaço vetorial.

Definição C.1.5 Diz-se que um vetor \mathbf{u} é **combinação linear** dos vetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ quando existem escalares reais ou complexos $\alpha_1, \alpha_2, \dots, \alpha_n$ tais que

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_n \mathbf{u}_n.$$

Definições C.1.6 A equação linear

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_n \mathbf{u}_n = \mathbf{0}, \quad (\text{C.1})$$

admite sempre a solução nula $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$.

Quando a solução nula é a única solução da equação (C.1), os vetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ dizem-se **linearmente independentes**.

Quando existem outras soluções da equação (C.1) para além da nula, os vetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ dizem-se **linearmente dependentes**.

Definição C.1.7 O conjunto S de todas as combinações lineares dos vetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$, para os escalares $\alpha_1, \alpha_2, \dots, \alpha_n$, dado por

$$S = \{\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \dots + \alpha_n \mathbf{u}_n\}$$

é designado por **subespaço vetorial gerado** por $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ e os vetores são designados por **vetores geradores** de S .

Definições C.1.8 Diz-se que o conjunto $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ é uma **base** do subespaço vetorial S quando os vetores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ são linearmente independentes e geradores de S .

Diz-se que um subespaço vetorial S tem **dimensão** n ($\dim(S) = n$) quando admite uma base B com n vetores.

Proposição C.1.9 Num subespaço vetorial S que admite uma base com n vetores, quaisquer $n + 1$ vetores de S são linearmente dependentes.

Proposição C.1.10 Todas as bases de um subespaço vetorial S têm o mesmo número de elementos.

C.1.2 Matrizes

Consideremos algumas definições de matrizes.

Definição C.1.11 Um **matriz** é um conjunto de números dispostos sob a forma de uma tabela retangular com n filas horizontais denominadas por **linhas** e com p filas verticais denominadas por **colunas**, denotada por:

$$\mathbf{A}_{n \times p} = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}.$$

A matriz \mathbf{A} designa-se por **matriz do tipo** $n \times p$ ou **matriz** $n \times p$, onde a_{ij} representa o elemento da linha i ($i = 1, 2, \dots, n$) e da coluna j ($j = 1, 2, \dots, p$).

Definições C.1.12 Consideremos uma matriz do tipo $n \times p$.

- **Matriz real**, se constituída por números reais.
- **Matriz complexa**, se constituída por números complexos.
- **Matriz nula** quando todos os elementos são iguais a zero, e denota-se por $\mathbf{O}_{n \times p}$.
- **Matriz quadrada de ordem** n ou **matriz de ordem** n , se $n = p$.
- **Matriz retangular** $n \times p$, se $n \neq p$.
- **Matriz (ou vetor) linha** $1 \times p$, se é constituída por apenas uma linha.
- **Matriz (ou vetor) coluna** $n \times 1$, se é constituída por apenas uma coluna.
- **Matriz transposta** de $\mathbf{A}_{n \times p} = (a_{ij})$ denota-se por $\mathbf{A}_{p \times n}^T = (a_{ji})$, $\forall i, j$, ou seja, é a matriz que resulta da operação que consiste em trocar as linhas pelas colunas ou vice-versa.

Definições C.1.13 Para o caso de matrizes quadradas $\mathbf{A} = (a_{ij})$ de ordem n , consideremos as seguintes noções:

- **Diagonal principal** é constituída pelos elementos em que $i = j$.
- **Diagonal secundária** é constituída pelos elementos em que $(i + j = n + 1)$.
- **Traço** é a soma dos elementos da diagonal principal, ou seja, $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.
- **Matriz diagonal** quando todos os elementos fora da diagonal principal são nulos, isto é,

$i \neq j \Rightarrow a_{ij} = 0$.

- **Matriz identidade de ordem n** quando a matriz é diagonal com elementos da diagonal principal todos iguais a 1, e denota-se por \mathbf{I}_n (ou \mathbf{I}_p).
- **Matriz simétrica** quando $\mathbf{A}^T = \mathbf{A}$, isto é, se $a_{ij} = a_{ji}$, $\forall i, j$.
- **Matriz antissimétrica** quando $\mathbf{A}^T = -\mathbf{A}$, isto é, se $a_{ii} = 0$ e $a_{ij} = -a_{ji}$, $\forall i \neq j$.
- **Matriz ortogonal** quando $\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}_n$.
- **Matriz inversa** de \mathbf{A} denota-se por \mathbf{A}^{-1} se $\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}_n$.
- **Matriz definida positiva** se para todo o vetor $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_{\mathbb{R}^n}\}$, $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$.
- **Matriz semidefinida positiva** se para todo o vetor $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_{\mathbb{R}^n}\}$, $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$.

Nota C.1.14 Qualquer matriz quadrada \mathbf{A} pode ser expressa como a soma de uma matriz simétrica e de uma matriz antissimétrica da seguinte forma

$$\mathbf{A} = \frac{1}{2} (\mathbf{A} + \mathbf{A}^T) + \frac{1}{2} (\mathbf{A} - \mathbf{A}^T).$$

Observações C.1.15 Quando uma matriz \mathbf{A} de ordem n admite inversa diz-se **invertível**, e quando uma matriz não admite inversa designa-se por **matriz não invertível** ou **singular**. Só as matrizes quadradas admitem inversa, mas nem todas as matrizes quadradas são invertíveis.

Definição C.1.16 Seja \mathbf{A} uma matriz retangular do tipo $n \times p$, cujas p colunas formam uma base do subespaço S do espaço vetorial E . Designa-se por **matriz de projeção ortogonal** sobre S e denota-se por \mathbf{P}_S à matriz dada por

$$\mathbf{P}_S = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T.$$

Definição C.1.17 Seja \mathbf{A} uma matriz retangular do tipo $n \times p$. Designa-se por **matriz pseudo-inversa** da matriz \mathbf{A} e denota-se por \mathbf{A}^+ à matriz do tipo $p \times n$ que satisfaça os seguintes 4 critérios.

- i) $\mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A}$
- ii) $\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+$
- iii) $(\mathbf{A} \mathbf{A}^+)^* = \mathbf{A} \mathbf{A}^+$
- iv) $(\mathbf{A}^+ \mathbf{A})^* = \mathbf{A}^+ \mathbf{A}$

Quando $(\mathbf{A}^* \mathbf{A})$ é invertível, a única matriz \mathbf{A}^+ que satisfaz os 4 critérios pode ser calculada por

$$\mathbf{A}^+ = (\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^*.$$

Quando $(\mathbf{A}\mathbf{A}^*)$ é invertível, a única matriz \mathbf{A}^+ que satisfaz os 4 critérios pode ser calculada por

$$\mathbf{A}^+ = \mathbf{A}^* (\mathbf{A}\mathbf{A}^*)^{-1}.$$

Definição C.1.18 Seja $\mathbf{A} = (a_{ij})$ uma matriz retangular do tipo $n \times p$. Uma **matriz de Hankel** é uma matriz em que todos os elementos ao longo de cada diagonal $i+j = k$, são iguais para um valor fixo de k .

C.1.3 Operações com Matrizes

Sejam $\mathbf{A}_{n \times p} = (a_{ij})$ e $\mathbf{B}_{n \times p} = (b_{ij})$ duas matrizes. Definem-se as seguintes operações:

- **Adição:** $\mathbf{A}_{n \times p} + \mathbf{B}_{n \times p} = \mathbf{C}_{n \times p} = (c_{ij})$ onde $c_{ij} = a_{ij} + b_{ij}$.
- **Subtração:** $\mathbf{A}_{n \times p} - \mathbf{B}_{n \times p} = \mathbf{C}_{n \times p} = (c_{ij})$ onde $c_{ij} = a_{ij} - b_{ij}$.

Observações C.1.19 Sejam \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{O} matrizes do mesmo tipo, onde \mathbf{O} é a matriz nula.

A adição de matrizes:

- i) é comutativa: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$;
- ii) é associativa: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$;
- iii) admite elemento neutro: $\mathbf{A} + \mathbf{O} = \mathbf{A}$;
- iv) admite, para cada matriz \mathbf{A} , uma matriz $\mathbf{A}' = -\mathbf{A}$ tal que $\mathbf{A} + \mathbf{A}' = \mathbf{O}$.

Sejam $\mathbf{A}_{n \times p} = (a_{ij})$, $\mathbf{B}_{p \times q} = (b_{jk})$ duas matrizes e λ um escalar.

- **Multiplicação de λ por \mathbf{A} :** $\lambda \mathbf{A}_{n \times p} = \mathbf{C}_{n \times p} = (c_{ij})$ onde $c_{ij} = \lambda a_{ij}$.
- **Produto ou multiplicação de \mathbf{A} por \mathbf{B} :** $\mathbf{A}_{n \times p} \mathbf{B}_{p \times q} = \mathbf{C}_{n \times q} = (c_{ik})$ onde $c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \dots + a_{ip}b_{pk}$.

Nota C.1.20 Ao escrever dois vetores \mathbf{u} e \mathbf{v} como matrizes colunas, o produto interno entre \mathbf{u} e \mathbf{v} passa a ser reescrito como o produto matricial entre duas matrizes dado por $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$.

Observações C.1.21 Sejam \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{O} quaisquer matrizes de dimensões adequadas e λ um escalar. A matriz \mathbf{O} representa a matriz nula e a matriz \mathbf{I} representa a matriz identidade.

- i) O produto de matrizes \mathbf{AB} só é possível quando o número de colunas de \mathbf{A} é igual ao número de linhas de \mathbf{B} .

ii) A multiplicação de matrizes:

- não é comutativa, pois em geral $\mathbf{AB} \neq \mathbf{BA}$;
- é associativa: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$;
- é associativa em relação à multiplicação escalar: $(\lambda\mathbf{A})\mathbf{B} = \mathbf{A}(\lambda\mathbf{B}) = \lambda(\mathbf{AB})$;
- é distributiva para a adição de matrizes: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$; $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$.

iii) $\mathbf{A}_{n \times p} \mathbf{I}_p = \mathbf{A}_{n \times p}$ e $\mathbf{I}_n \mathbf{A}_{n \times p} = \mathbf{A}_{n \times p}$.

iv) $\mathbf{A}_{n \times p} \mathbf{O}_{p \times q} = \mathbf{O}_{n \times q}$ e $\mathbf{O}_{k \times n} \mathbf{A}_{n \times p} = \mathbf{O}_{k \times p}$.

v) Duas matrizes \mathbf{A} e \mathbf{B} de ordem n , dizem-se **comutáveis** ou que **comutam entre si** quando $\mathbf{AB} = \mathbf{BA}$.

Proposição C.1.22 (Circularidade do Traço) *Seja \mathbf{A} uma matriz do tipo $n \times p$ e \mathbf{B} uma matriz do tipo $p \times n$. Então $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$.*

Proposição C.1.23 (Linearidade do Traço) *Sejam \mathbf{A} e \mathbf{B} matrizes do tipo $n \times p$ e α e β escalares. Então $\text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\text{tr}(\mathbf{A}) + \beta\text{tr}(\mathbf{B})$.*

Consideremos ainda umas breves noções de característica e determinante de uma matriz.

Definição C.1.24 *Designa-se por **característica** da matriz \mathbf{A} , e denota-se por $\text{car}(\mathbf{A})$ ou $r(\mathbf{A})$, ao número máximo de linhas (ou colunas) linearmente independentes.*

Teorema C.1.25 *Uma matriz quadrada \mathbf{A} é invertível se e só se $\text{car}(\mathbf{A})$ é igual à sua ordem, ou seja, \mathbf{A} é invertível se e só se as linhas e as colunas de \mathbf{A} são linearmente independentes.*

Proposição C.1.26 *Seja \mathbf{A} uma matriz do tipo $n \times p$. Então $\text{car}(\mathbf{A}) = \text{car}(\mathbf{A}^T \mathbf{A}) = \text{car}(\mathbf{A} \mathbf{A}^T)$.*

Definição C.1.27 *O **determinante** de uma matriz quadrada \mathbf{A} de ordem n é a soma de $n!$ produtos de elementos da matriz da forma $a_{1,j_1} a_{2,j_2} a_{3,j_3} \dots a_{n,j_n}$ (denotados por **produtos elementares**), cujos índices (j_1, j_2, \dots, j_n) correspondem a todas as permutações dos inteiros de 1 a n , e em que os produtos são multiplicados por $+1$ ou -1 de acordo com o sinal da permutação que lhe está associado seja par ou ímpar. O determinante de \mathbf{A} representa-se por $\det(\mathbf{A})$ ou $|\mathbf{A}|$.*

C.1.4 Valores e Vetores Próprios

Considere-se uma matriz quadrada \mathbf{A} de ordem n .

Definição C.1.28 Diz-se que $\mathbf{v} \in \mathbb{C}^n \setminus \{0_{\mathbb{C}^n}\}$ é um **vetor próprio** de \mathbf{A} associado ao **valor próprio** $\lambda \in \mathbb{C}$ quando $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

Definição C.1.29 O polinómio de grau n , $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}_n)$, designa-se por **polinómio característico** de \mathbf{A} . As suas raízes são os valores próprios de \mathbf{A} , e portanto uma matriz \mathbf{A} , de ordem n , tem n valores próprios, embora alguns deles possam ser iguais, ou seja, o polinómio pode apresentar raízes com multiplicidades algébricas superiores à unidade.

Proposição C.1.30 Sejam $\lambda_1, \lambda_2, \dots, \lambda_n$ os valores próprios de uma matriz \mathbf{A} de ordem n .

Então:

- $\det(\mathbf{A}) = \lambda_1 \times \lambda_2 \times \dots \times \lambda_n = \prod_{i=1}^n \lambda_i$;
- $\text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + \dots + \lambda_n = \sum_{i=1}^n \lambda_i$.

Para o caso particular de uma matriz simétrica \mathbf{A} de ordem n , os valores e vetores próprios obedecem aos seguintes critérios:

1. Os valores e os respetivos vetores próprios são sempre reais.
2. Os vetores próprios associados a valores próprios diferentes são sempre ortogonais.
3. Mesmo no caso da existência de valores próprios repetidos, é sempre possível determinar um conjunto de n vetores próprios ortogonais e de norma unitária (ortonormados).

Definição C.1.31 Duas matrizes \mathbf{A} e \mathbf{B} de ordem n dizem-se **semelhantes** (ou diz-se que \mathbf{B} é **semelhante** a \mathbf{A}) se existir uma matriz invertível \mathbf{P} tal que $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

Teorema C.1.32 Sejam \mathbf{A} e \mathbf{B} duas matrizes semelhantes de ordem n . Então \mathbf{A} e \mathbf{B} têm:

- o mesmo polinómio característico.
- os mesmos valores próprios com as mesmas multiplicidades algébricas.
- o mesmo determinante.
- o mesmo traço.

Teorema C.1.33 *Seja \mathbf{A} uma matriz simétrica de ordem n . Então:*

- \mathbf{A} é definida positiva quando todos os valores próprios de \mathbf{A} são positivos.
- \mathbf{A} é semidefinida positiva quando todos os valores próprios de \mathbf{A} são positivos ou nulos, com pelo menos um nulo.

C.2 Análise de Fourier e Transformada de Fourier

A Análise de Fourier (a qual não deve ser confundida com a Transformada de Fourier), é um conceito geométrico que se aplica a funções periódicas discretas ou contínuas. No contexto da análise de séries temporais, o seu objetivo é o de aproximar uma função do tempo por uma combinação linear de funções trigonométricas ortogonais formada à custa de senos e cossenos da chamada série de Fourier.

A decomposição de uma função periódica X_t de período T pode ser descrita com base num conjunto de funções periódicas simples.

- No caso de funções discretas da forma $(x_0, x_1, \dots, x_{T-1})$ com $x_T = x_0$, vem

$$\begin{aligned} X_t &= \sum_{k=0}^{T-1} a_k e^{i2\pi kt/T} \\ &= \sum_{k=0}^{T-1} a_k [\cos(2\pi kt/T) + i \sin(2\pi kt/T)], \text{ com } t = 0, 1, \dots, T-1 \end{aligned}$$

cujos coeficientes são dados por

$$a_k = \frac{1}{T} \sum_{t=0}^{T-1} X_t e^{-i2\pi kt/T}.$$

- No caso de funções contínuas da forma X_t , $t \in \mathbb{Z}$, vem

$$\begin{aligned} X_t &= \sum_{k=-\infty}^{+\infty} a_k e^{i2\pi kt/T} \\ &= \sum_{k=-\infty}^{+\infty} a_k [\cos(2\pi kt/T) + i \sin(2\pi kt/T)], \text{ com } t \in [0, T] \text{ e } k \in \mathbb{Z} \end{aligned}$$

cujos coeficientes são descritos por

$$a_k = \frac{1}{T} \int_0^T X_t e^{-i2\pi kt/T} dt.$$

Uma vez que a Análise de Fourier só pode ser aplicada a funções periódicas, utiliza-se a Transformada de Fourier para trabalhar com funções não periódicas.

Para o caso discreto, a Transformada de Fourier \mathcal{F} de uma série discreta ou sinal X_t , $t \in \mathbb{Z}$, é uma função de valores complexos de frequência contínua ω (em ciclos por unidade de tempo discreto) dada por

$$\mathcal{F}_{X_t}(\omega) = \sum_{k=-\infty}^{+\infty} X_t e^{-i2\pi\omega k}, \text{ para } -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

Para o caso de funções não periódicas contínuas, a Transformada de Fourier admite fórmulas similares às do caso discreto.

Uma vez que as representações de uma série temporal nos domínios do tempo e da frequência contêm praticamente a mesma informação então uma pode ser obtida a partir da outra através da Transformada de Fourier ou da sua inversa. A inversa da Transformada de Fourier é definida por

$$X_t = \int_{-1/2}^{1/2} \mathcal{F}_{X_t}(\omega) e^{i2\pi\omega k} d\omega.$$

Mais pormenores relacionados com a Análise de Fourier podem ser consultados em (Murteira *et al.*, 1993; Robison, 1982; von Storch e Zwiers, 1999; Oliveira, 2003).

C.3 Análise Espectral

Em séries estacionárias, a análise no domínio da frequência permite estudar a forma como diferentes frequências ou componentes periódicas contribuem para a análise da variabilidade numa série temporal.

Passaremos a descrever alguns conceitos relacionados com o estudo de séries temporais no domínio da frequência.

Considere-se o processo estocástico para uma série temporal dado por $\{X_t\}_{t \in \mathbb{N}}$ e uma série temporal $X_t = (x_1, x_2, \dots, x_n)$. Consideremos algumas definições e observações seguintes.

- **Função periódica de período T :**

Uma função f , definida em \mathbb{R} , é periódica com período T , $T > 0$, se

$$f(x + T) = f(x), \forall x \in \mathbb{R}.$$

- **Harmónica:**

É uma função periódica da forma

$$\alpha_1 \cos(2\pi\omega t) + \alpha_2 \sin(2\pi\omega t)$$

de período $T = 1/\omega$, onde $2\pi\omega$ é a frequência angular em radianos e ω é a frequência em ciclos por unidade de tempo.

- **Relações de ortogonalidade para funções trigonométricas:**

A representação de funções periódicas pode efetuar-se em termos das funções trigonométricas simples

$$1, \cos \omega x, \sin \omega x, \cos 2\omega x, \sin 2\omega x, \dots, \cos n\omega x, \sin n\omega x.$$

Estas funções constituem um sistema ortogonal num intervalo $a \leq x < a + p$, para uma constante a e período $p = 2\pi/\omega$, e obedecem a um conjunto de relações, denominadas relações de ortogonalidade.

- **Espectro ou função densidade espectral** (ou espectro de potência - *power spectrum*):

É a Transformada de Fourier da função de autocovariância, descrito como alternativa aos momentos de segunda ordem do processo. É dado por

$$\Gamma(\omega) = \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-i2\pi\omega k}, \text{ para } -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

O espectro descreve a distribuição da energia do sinal ao nível do domínio da frequência, fornecendo a proporção de variação nos dados originais em diferentes frequências. Permite conhecer quais são as oscilações que mais contribuem para o processo, embora não se consiga conhecer o ponto inicial para cada ciclo associado a cada harmónica (não transmite informação acerca da fase das oscilações).

- **Espectro cruzado** (*cross-spectrum*):

É a Transformada de Fourier da função de covariância cruzada, descrito para dois processos estocásticos da seguinte forma

$$\Gamma_{xy}(\omega) = \sum_{k=-\infty}^{+\infty} \gamma_{xy}(k) e^{-i2\pi\omega k}, \text{ para } -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

- **Matriz de espectro** (*spectrum matrix*):

É uma generalização para um processo sob a forma de vetor $\mathbf{X} = (X_1, X_2, \dots, X_p)$. É a Transformada de Fourier da matriz dos desfasamentos de autocovariâncias, $\Sigma_{\mathbf{X}}(k)$, onde cada elemento da matriz na posição (ij) representa a covariância cruzada $\gamma_{xy}(k)$ entre o i -ésimo e o j -ésimo processos, e é dada por

$$\Gamma_{\mathbf{X}}(\omega) = \sum_{k=-\infty}^{+\infty} \Sigma_{\mathbf{X}}(k) e^{-i2\pi\omega k}, \text{ para } -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

- A função de covariância cruzada nem sempre é uma função par, pelo que o espectro cruzado é uma função complexa descrita por

$$\Gamma_{xy}(\omega) = \Lambda_{xy}(\omega) + i\Psi_{xy}(\omega),$$

ou em coordenadas polares

$$\Gamma_{xy}(\omega) = A_{xy}(\omega) e^{i\Phi_{xy}(\omega)}.$$

$\Lambda_{xy}(\omega)$ é conhecido por **co-espectro** e $\Psi_{xy}(\omega)$ por **espectro de quadratura**.

$A_{xy}(\omega)$ é designado por **espectro de amplitude** e $\Phi_{xy}(\omega)$ é conhecido por **espectro de fase**.

São válidas as seguintes relações

$$A_{xy}^2(\omega) = \Lambda_{xy}^2(\omega) + \Psi_{xy}^2(\omega) \text{ e}$$

$$\Phi_{xy}(\omega) = \arctg(\Psi_{xy}(\omega)/\Lambda_{xy}(\omega)), \text{ para } \Psi_{xy}(\omega) \neq 0 \text{ e } \Lambda_{xy}(\omega) \neq 0.$$

Para o caso em que $\Psi_{xy}(\omega) = 0$, vem

$$\Phi_{xy}(\omega) = \begin{cases} 0 & \text{se } \Lambda_{xy}(\omega) > 0 \\ \pm\pi & \text{se } \Lambda_{xy}(\omega) < 0 \end{cases}.$$

Para o caso em que $\Lambda_{xy}(\omega) = 0$, vem

$$\Phi_{xy}(\omega) = \begin{cases} \pi/2 & \text{se } \Psi_{xy}(\omega) > 0 \\ -\pi/2 & \text{se } \Psi_{xy}(\omega) < 0 \end{cases}.$$

- **Espectro de coerência ou quadrado da coerência:**

É uma função que varia no intervalo $[0; 1]$, a qual é utilizada como alternativa ao espectro de amplitude e é dada por

$$K_{xy}^2(\omega) = \frac{A_{xy}^2(\omega)}{\sqrt{\Gamma_x(\omega)\Gamma_y(\omega)}}.$$

Outros conceitos relacionados com a Análise de Fourier e Análise Espectral podem ser consultados em Murteira *et al.* (1993) ou em von Storch e Zwiers (1999).

C.4 Momentos, Cumulantes e suas Propriedades

Esta Secção pretende esclarecer alguns conceitos relacionados com momentos e cumulantes. Considera-se inicialmente o caso univariado para uma variável aleatória e o caso multivariado para um vetor aleatório constituído por várias variáveis aleatórias, e na última parte são destacadas algumas propriedades de momentos e cumulantes.

C.4.1 Caso Univariado

Admita-se que X é uma variável aleatória contínua, de média zero, com função densidade de probabilidade $f_X(\cdot)$.

Uma primeira função característica $\varphi(\omega)$ de X , também denominada por função geradora de momentos, é definida como a transformada de Fourier contínua de $f_X(\cdot)$ dada por

$$\varphi(\omega) = E(\exp(i\omega x)) = \int_{-\infty}^{+\infty} \exp(i\omega x) f_X(x) dx,$$

onde $i = \sqrt{-1}$ e ω é a variável transformada correspondente a X .

Ao aplicar o desenvolvimento em série de Taylor à função característica, vem

$$\varphi(\omega) = \int_{-\infty}^{+\infty} \left(\sum_{k=0}^{\infty} \frac{x^k (i\omega)^k}{k!} \right) f_X(x) dx = \sum_{k=0}^{\infty} E(x^k) \frac{(i\omega)^k}{k!}.$$

Os coeficientes dos termos desta expressão são os momentos (não centrados) de X de ordem k , $E(x^k)$.

Uma segunda função característica $\phi(\omega)$ de X , ou função geradora de cumulantes, é definida pelo logaritmo natural da primeira função característica

$$\phi(\omega) = \ln(\varphi(\omega)) = \ln(E(\exp(i\omega x))).$$

De forma análoga, ao aplicar o desenvolvimento em série de Taylor à segunda função característica, vem

$$\phi(\omega) = \sum_{k=0}^n k_k \frac{(i\omega)^k}{k!}.$$

Os coeficientes k_k desta série são os cumulantes de X de ordem k , e são descritos por

$$k_k = (-i)^k \left. \frac{d^k \phi(\omega)}{d\omega^k} \right|_{\omega=0}.$$

Para o caso em que a média da variável aleatória X é não nula, os primeiros quatro cumulantes surgem expressos em função dos momentos e são descritos por

$$\begin{aligned} k_1 &= E(x) \\ k_2 &= E(x^2) - [E(x)]^2 \\ k_3 &= E(x^3) - 3E(x^2)E(x) + 2[E(x)]^3 \\ k_4 &= E(x^4) - 3[E(x^2)]^2 - 4E(x^3)E(x) + 12E(x^2)[E(x)]^2 - 6[E(x)]^4. \end{aligned} \tag{C.2}$$

O primeiro cumulante coincide com a média, enquanto que o segundo cumulante coincide com a variância da variável.

Para o caso particular em que temos uma variável aleatória X de média zero, a partir de (C.2) os primeiros quatro cumulantes resumem-se a

$$\begin{aligned} k_1 &= 0 \\ k_2 &= E(x^2) \\ k_3 &= E(x^3) \\ k_4 &= E(x^4) - 3[E(x^2)]^2. \end{aligned}$$

Os três primeiros cumulantes são iguais aos respetivos momentos, enquanto que o cumulante de quarta ordem é conhecido como sendo a curtose de uma variável aleatória.

C.4.2 Caso Multivariado

Consideremos agora o caso multivariado onde se admite que \mathbf{x} é o vetor aleatório com função densidade de probabilidade $f_{\mathbf{x}}(\cdot)$.

A função característica $\varphi(\omega)$ de \mathbf{x} , definida pela transformada de Fourier de $f_{\mathbf{x}}(\cdot)$ é dada por

$$\varphi(\omega) = E(\exp(i\omega\mathbf{x})) = \int_{-\infty}^{+\infty} \exp(i\omega\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}.$$

Os momentos e cumulantes (também designados por cumulantes cruzados para o caso multivariado) de \mathbf{x} obtêm-se de forma análoga ao caso univariado, ou seja, os momentos (não centrados) são provenientes de $\varphi(\boldsymbol{\omega})$ enquanto que os cumulantes extraem-se da função característica $\phi(\boldsymbol{\omega}) = \ln(\varphi(\boldsymbol{\omega}))$ de \mathbf{x} .

De seguida vamos utilizar uma notação que foi inspirada em Puga (2000), o qual por sua vez se baseou em McCullagh (1995).

Um determinado momento que diga respeito às variáveis denotadas por $X_{x_1}, X_{x_2}, \dots, X_{x_r}$ é geralmente designado por $E(x_1, x_2, \dots, x_r)$ e o cumulante por $cum(x_1, x_2, \dots, x_r)$.

Para um vetor aleatório multidimensional $\mathbf{x} = (X_{x_1}, X_{x_2}, \dots, X_{x_r})$, os elementos que contêm todos os momentos de ordem r deveriam ser designados, por exemplo, por $E_{\mathbf{x}}(x_1, x_2, \dots, x_r)$ assim como os elementos que contêm todos os cumulantes de ordem r deveriam ser designados por $cum_{\mathbf{x}}(x_1, x_2, \dots, x_r)$. No entanto, por uma questão de simplificação de notação, e uma vez que a informação está toda na quantidade (e não na forma) de índices que determinam a ordem do momento ou do cumulante, apenas utilizaremos E e cum .

Consideremos um vetor aleatório multidimensional $\mathbf{x} = (X_{x_1}, X_{x_2}, \dots, X_{x_n})$ e o conjunto definido por $I_n = \{x_1, x_2, \dots, x_n\}$ correspondente a um conjunto de índices das componentes de \mathbf{x} . A partição do conjunto I_n é definida como a coleção não ordenada de conjuntos disjuntos não vazios I_p , de modo que $\bigcup_p I_p = I_n$.

As partições possíveis de obter a partir de I_2 são

$$\{(x_1), (x_2)\}; \{(x_1, x_2)\}$$

e a partir de I_3 são

$$\begin{aligned} &\{(x_1), (x_2), (x_3)\}; \\ &\{(x_1, x_2), (x_3)\}; \{(x_1, x_3), (x_2)\}; \{(x_2, x_3), (x_1)\}; \\ &\{(x_1, x_2, x_3)\}. \end{aligned}$$

Para I_4 , temos as seguintes partições

$$\begin{aligned} & \{(x_1), (x_2), (x_3), (x_4)\}; \\ & \{(x_1), (x_2), (x_3, x_4)\}; \{(x_1), (x_3), (x_2, x_4)\}; \{(x_1), (x_4), (x_2, x_3)\}; \\ & \{(x_2), (x_3), (x_1, x_4)\}; \{(x_2), (x_4), (x_1, x_3)\}; \{(x_3), (x_4), (x_1, x_2)\}; \\ & \{(x_1, x_2), (x_3, x_4)\}; \{(x_1, x_3), (x_2, x_4)\}; \{(x_1, x_4), (x_2, x_3)\}; \\ & \{(x_1, x_2, x_3), (x_4)\}; \{(x_1, x_2, x_4), (x_3)\}; \{(x_1, x_3, x_4), (x_2)\}; \{(x_2, x_3, x_4), (x_1)\}; \\ & \{(x_1, x_2, x_3, x_4)\}. \end{aligned}$$

A fórmula que permite descrever os cumulantes em função dos momentos é então descrita por

$$cum(I_n) = \sum_{\bigcup_{p=1}^q I_p = I_n} (-1)^{q-1} (q-1)! \prod_{p=1}^q E(I_p),$$

onde $\bigcup_{p=1}^q I_p = I_n$ indica que o somatório contempla todas as partições do conjunto I_n .

Tendo em conta os casos particulares de partições apresentados anteriormente, as relações para os cumulantes cruzados em função dos momentos até ao quarto cumulante são

$$\begin{aligned} cum(x_1) &= E(x_1) \\ cum(x_1, x_2) &= E(x_1, x_2) - E(x_1)E(x_2) \\ cum(x_1, x_2, x_3) &= E(x_1, x_2, x_3) \\ &\quad - E(x_1)E(x_2, x_3) - E(x_2)E(x_1, x_3) - E(x_3)E(x_1, x_2) \\ &\quad + 2E(x_1)E(x_2)E(x_3) \\ cum(x_1, x_2, x_3, x_4) &= E(x_1, x_2, x_3, x_4) \\ &\quad - E(x_1)E(x_2, x_3, x_4) - E(x_2)E(x_1, x_3, x_4) \\ &\quad - E(x_3)E(x_1, x_2, x_4) - E(x_4)E(x_1, x_2, x_3) \\ &\quad - E(x_1, x_2)E(x_3, x_4) - E(x_1, x_3)E(x_2, x_4) - E(x_1, x_4)E(x_2, x_3) \\ &\quad + 2E(x_1)E(x_2)E(x_3, x_4) + 2E(x_1)E(x_3)E(x_2, x_4) \\ &\quad + 2E(x_1)E(x_4)E(x_2, x_3) + 2E(x_2)E(x_3)E(x_1, x_4) \\ &\quad + 2E(x_2)E(x_4)E(x_1, x_3) + 2E(x_3)E(x_4)E(x_1, x_2) \\ &\quad - 6E(x_1)E(x_2)E(x_3)E(x_4). \end{aligned}$$

Para o caso de momentos em torno da média (em que a média de cada uma das componentes do vetor aleatório \mathbf{x} é zero, $E(x_1) = E(x_2) = E(x_3) = E(x_4) = 0$), os cumulantes até à 4ª ordem são expressos apenas por

$$\begin{aligned} cum(x_1) &= 0 \\ cum(x_1, x_2) &= E(x_1, x_2) \\ cum(x_1, x_2, x_3) &= E(x_1, x_2, x_3) \\ cum(x_1, x_2, x_3, x_4) &= E(x_1, x_2, x_3, x_4) \\ &\quad - E(x_1, x_2)E(x_3, x_4) - E(x_1, x_3)E(x_2, x_4) - E(x_1, x_4)E(x_2, x_3). \end{aligned}$$

Os três primeiros cumulantes coincidem com os momentos de ordem 1, 2 e 3 respetivamente, e em particular o primeiro cumulante coincide com a média, e o segundo cumulante coincide com a covariância entre duas variáveis. Por outro lado, o quarto cumulante é diferente do momento de ordem 4.

C.4.3 Propriedades

De uma forma geral, os momentos de ordem superior correspondem a correlações que são utilizadas em estatísticas de segunda ordem, enquanto que os cumulantes de ordem superior são usados nas covariâncias. Os momentos e os cumulantes contêm a mesma informação estatística, uma vez que os cumulantes podem ser expressos em função de somas de produtos dos momentos. Desta forma é muitas vezes preferível trabalhar com cumulantes em vez de momentos, dado que apresentam de uma forma mais clara a informação adicional fornecida por estatísticas de ordem superior.

Os cumulantes possuem as seguintes propriedades, as quais não se verificam para os momentos (Hyvärinen *et al.*, 2001b).

- Para dois vetores aleatórios estatisticamente independentes de iguais dimensões, o cumulante da sua soma é igual à soma dos cumulantes de cada um deles. Esta propriedade generaliza-se para a soma de mais de dois vetores aleatórios independentes.
- Se a distribuição do vetor aleatório é Normal multivariada, todos os seus cumulantes de ordem superior a dois são identicamente nulos.

Os cumulantes de ordem superior medem o afastamento de um dado vetor aleatório em relação a um vetor aleatório com distribuição Normal, e com vetor médio e matriz de covariâncias idênticos. Para além de outras propriedades, salienta-se ainda o facto de momentos e cumulantes possuírem propriedades de simetria, as quais podem ser utilizadas com vista à redução em termos de carga computacional no que diz respeito à sua estimação.

Apêndice D

Rotinas e Conjuntos de Dados

D.1 Rotinas e *Packages* Usados na Análise dos Dados

Ao longo da tese, no tratamento dos dados, foi essencialmente utilizado o *software* estatístico de acesso livre *R*, o qual é constituído por milhares de *packages*.

No Capítulo 3 da ACI foi usado o *package* FastICA para implementar o algoritmo com o mesmo nome. Foram ainda criadas rotinas para implementar o algoritmo AMUSE (descrito de seguida), outras para ordenar as componentes independentes e outras para aplicar alguns indicadores comparativos entre as técnicas ACP e ACI.

```
-----  
AMUSElag1<-function(X,lag,ncomp)  
{  
  # Algoritmo AMUSE para um dado desfasamento no tempo #  
  # X: matriz dos dados iniciais (não centrada) #  
  # lag: desfasamento no tempo #  
  # ncomp: n° de componentes a extrair #  
  # VAR: matriz de covariâncias #  
  # K: matriz do branqueamento #  
  # Z: matrix dos dados branqueados #  
  # C: matriz de covariâncias dos desfasamentos no tempo para um dado lag #  
  # CM: matriz simétrica obtida a partir da matriz C #
```

```

# V: matriz dos vectores próprios #
# A: matriz estimada dos coeficientes da mistura #
# S: matriz estimada dos sinais (componentes) apenas para ncomp sinais #

n<-nrow(X)
p<-ncol(X)
X<-scale(X,scale=FALSE)
VAR<-t(X)%*%X/n
s1<-svd(VAR)
D<-diag(c(1/sqrt(s1$d)))
K<-D%*%t(s1$u)
if(ncomp>p)
  stop("o ncomp deverá ser menor ou igual a p")
K<-matrix(K[1:ncomp,],ncomp,p)
Z<-K%*%t(X)
Z<-t(Z)
if(n-lag<1)
  stop("o lag deverá ser menor que n")
C<-t(matrix(Z[1:(n-lag),],n-lag,ncomp))%*%matrix(Z[(1+lag):n,],n-lag,ncomp)/n
CM<-(C+t(C))/2
s2<-eigen(CM)
V<-s2$vectors
T<-D[1:ncomp,1:ncomp]%*%t(s1$u[1:ncomp])
s3<-svd(T)
PseudoinversaT<-s3$v%*%diag(c(1/s3$d))%*%t(s3$u)
A<-PseudoinversaT%*%V
Z<-t(Z)
S<-t(V)%*%Z
return(list(K=K,Z=t(Z),C=C,CM=CM,V=V,A=t(A),S=t(S)))
}

```

Rotinas de SSA

Para o exemplo de aplicação da SSA no Capítulo 4, para além das próprias funcionalidades do *R*, foram criadas inúmeras rotinas para implementar a técnica, incluindo aquelas que permitem tratar os vários tipos de centragem associados à SSA. Na parte das previsões foi também usado o *software CaterpillarSSA*.

Para implementar a rotina para o algoritmo de previsão corrente (*ssaRFA*) é necessário implementar antes duas outras funções, uma para criar a matriz dos desfasamentos (*MssaMLmatlagdados*) e outra para a reconstrução de uma série temporal (*RCT.1serie*), de acordo com os códigos que se apresentam de seguida.

```
-----
MssaMLmatlagdados<-function (matriz, M)
{
  # Inserir uma matriz com L séries temporais e um comprimento da janela, M (M<=N/2).#
  # Devolve uma lista com 2 matrizes com L*M colunas #
  # que representam M desfasamentos para L séries.#
  # 1ª série nas primeiras M colunas#
  # 2ª série nas M colunas seguintes,...#
  # Matriz com colunas não centradas e centradas.#

  matriz<-as.matrix(matriz)

  N <- nrow(matriz)
  L <- ncol(matriz)

  xmat <- matrix(0, nrow = (N - M + 1), ncol = M * L)
  for (j in 1:L) (for (i in 1:M) xmat[, (j - 1) * M + i] <- c(matriz[, j])[i:(N - M + i)])

  xmatcent <- matrix(0, nrow = (N - M + 1), ncol = M*L)
  for (i in 1:M) (for (j in 1:L)
    xmatcent[, (j-1)*M+i] <- xmat[, (j-1)*M+i] -mean(xmat[, (j-1)*M+i] ))
  colnames <- rep("", M)
  rownames <- rep("", N - M + 1)
  for (j in 1:L) (for (i in 1:M) colnames[(j - 1) * M + i] <- paste("lag", i, "series", j, sep = ""))
  for (i in 1:N - M + 1) rownames[i] <- paste("row", i, sep = "")
}
```

```

list("Matrix of lagged data \n "=xmat, "Matrix of lagged data centered \n "=xmatcent, colnames,
      rownames)
}

-----
RCT.1serie<- function (serie, M, k, p)
{
  # serie - inserimos 1 serie original (por coluna) #
  # M - comprimento da janela #
  # k - n° mínimo da componente principal a reter na reconstrução #
  # p - n° máximo da componente principal a reter na reconstrução #
  # Esta função irá devolver as CPs e a reconstrução da L série original.#

  MatrizDesfasada<-MssaMLmatlagdados(serie,M)[[1]]
    # Matriz dos desfasamentos calculada a partir da função MssaMLmatlagdados #
  MatrizS<-t(MatrizDesfasada)%*%MatrizDesfasada
  D<-svd(MatrizS)$d
  U<-svd(MatrizS)$u
  MatrizV<-MatrizDesfasada%*%U%*%diag(1/sqrt(D))
    # Vi=X*Ui/sqrt(lambda i) matriz das componentes principais #
  ListaXi.kAte.p<-list()
  for(i in k:p)
  {
    MatrizXi<-sqrt(D[i])*MatrizV[,i]%*%t(U[,i])
    # Xi=sqrt(lambda i)*Vi*t(Ui) #
    ListaXi.kAte.p[[i]]<-MatrizXi
  }
  if(k==1 & p==1)
    SomaXi.kAte.p<-ListaXi.kAte.p[[1]]
  else {
    SomaXi.kAte.p<-ListaXi.kAte.p[[k]]
    for(i in k:(p-1))

```

```

      SomaXi.kAte.p<-SomaXi.kAte.p+ListaXi.kAte.p[[i+1]]
    }
  N <- nrow(as.matrix(serie))
  L <- ncol(as.matrix(serie))
  k1 <- nrow(MatrizDesfasada)
  M <- ncol(MatrizDesfasada)
  N1 <- k1 + M - 1
  if (N != N1)
    stop("Error on number of elements in our data")
  matrec2 <- matrix(0, nrow = N, ncol = M*L)
  for (i in 1:M)(for (j in 1:L)
    matrec2[i:(N - M + i), (j-1)*M+i] <- SomaXi.kAte.p[, (j-1)*M+i])
  med2 <- vector(mode = "numeric", length = N)
  for (i in 1:M) med2[i] <- i
  for (i in (M + 1):(N - M + 1)) med2[i] <- M
  for (i in (N - M + 2):N) med2[i] <- (N - i + 1)
  matrizrec <- as.matrix(matrix(0,nrow = N, ncol = L))
  for (i in 1:N) (for (j in 1:L)
    matrizrec[i,j] <- sum(matrec2[i,((j-1)*M+1):(j*M)])/med2[i])
  list("Matriz V das CPs \n" = MatrizV, "Matriz Reconstruida desfasada \n" = SomaXi.kAte.p,
    "Serie Reconstruida \n" = matrizrec, "Matriz U dos vectores próprios \n" = U)
}

-----
ssaRFA<- function (serie, M, r, h)
{
  # SSA - Algoritmo de Previsão Recorrente #
  # serie - inserimos 1 serie original (por coluna) #
  # M - comprimento da janela #
  # r - n° de componentes principais a reter na reconstrução #
  # h - n° de valores previstos a obter #
  # Esta rotina irá devolver valores previstos. #

```



```

MatrizU<-RCT.lserie(serie, M, 1, r)[[4]]

# Matriz dos vetores próprios; admitimos sempre que k=1 e p=r #
v<-MatrizU[M,1:r]
v2<-sum(v^2)
A<-matrix(0, nrow = M-1, ncol = r)
for(i in 1:r)
  A[,i]<-(1/(1-v2))*v[i]*MatrizU[1:(M-1),i]
vectorA<-rowSums(A)
N<-length(serie)
SerieReconstruida<-RCT.lserie(serie, M, 1, r)[[3]]
VectorValPrev<-c()
ValorPrevisto<-vectorA%*%SerieReconstruida[(N-M+2):N]
VectorValPrev[1]<-ValorPrevisto
if (h==1)
  VectorValPrev<-ValorPrevisto
else {
  for (i in 2:h)
    {
      SerieReconstruida[N+(i-1)]<-VectorValPrev[i-1]
      NovoValorPrevisto<-vectorA%*%SerieReconstruida[(N-M+2+i-1):(N+i-1)]
      VectorValPrev[i]<-NovoValorPrevisto
    }
  }
  return(VectorValPrev)
}

```

Rotinas de *LagACI*

Para o Capítulo 6 da *LagACI*, foram produzidas inúmeras rotinas para implementar a MSSA e para sustentar a nova abordagem *LagACI*, tendo em vista a análise do conjunto de dados da

pressão ao nível do mar. As linhas de código não são aqui apresentadas devido à sua enorme quantidade e extensão de muitas delas.

Rotinas de Indicadores do Desempenho de um Modelo

Para o exemplo de aplicação da *LagACI*, na parte da comparação entre as técnicas apresentadas no Capítulo 6, foram criadas funções para implementar cada um dos indicadores do desempenho de um modelo, de acordo com os códigos apresentados de seguida.

```
-----
INDICEdeWILLMOTT81<-function (matrizorig, matrizreconst, nlinhas, ncolunas)
{
  # matrizorig - matriz dos dados originais #
  # matrizreconst - matriz reconstruída #
  # nlinhas - n° de linhas das matrizes matrizorig e matrizreconst #
  # ncolunas - n° de colunas das matrizes matrizorig e matrizreconst #
  # devolve o valor do índice #

  MATRIZdeMEDIAS <- as.matrix(matrix(0, nrow = nlinhas, ncol = ncolunas))
  for(i in 1:nlinhas) {for(j in 1:ncolunas){
    MATRIZdeMEDIAS[i,j] <- mean(matrizorig[,j])}
  }

  SQRE <- c()
  SQMod <- c()
  for(j in 1:ncolunas){
    SQRE[j] <- sum((matrizreconst-matrizorig)[,j]^2)
    SQMod[j] <- sum((abs(matrizreconst[,j]-MATRIZdeMEDIAS[,j])
                     +abs(matrizorig[,j]-MATRIZdeMEDIAS[,j]))^2)}

  somaSQRE <- sum(SQRE)
  somaSQMod <- sum(SQMod)
  INDICE <- 1-(SQRE/SQMod)
  MEDIAdosVALORESdoINDICE <- mean(INDICE)

  list("SQRE de cada variável\n" = SQRE, "Soma das SQRE\n" = somaSQRE, "SQMod de cada
variável\n" = SQMod, "Soma das SQMod\n" = somaSQMod, "Indice de Willmott 1981\n" = INDICE,
"Média dos Valores do Indice de Willmott 1981" = MEDIAdosVALORESdoINDICE)
```

```

}

-----
INDICEdeWILLMOTTrevisto85<-function (matrizorig, matrizreconst, nlinhas, ncolunas)
{
  # matrizorig - matriz dos dados originais #
  # matrizreconst - matriz reconstruída #
  # nlinhas - n° de linhas das matrizes matrizorig e matrizreconst #
  # ncolunas - n° de colunas das matrizes matrizorig e matrizreconst #
  # devolve o valor do índice #
  # EAM - Erro Absoluto Médio vezes n de observações (soma de |Pi-Oi|) #
  # SMod - Soma de Módulos [|Pi-Obarra| + |Oi-Obarra|] #
  MATRIZdeMEDIAS <- as.matrix(matrix(0, nrow = nlinhas, ncol = ncolunas))
  for(i in 1:nlinhas) {for(j in 1:ncolunas){
    MATRIZdeMEDIAS[i,j] <- mean(matrizorig[,j])}
  }
  EAM <- c()
  SMod <- c()
  for(j in 1:ncolunas){
    EAM[j] <- sum(abs(matrizreconst[,j]-matrizorig[,j]))
    SMod[j] <- sum(abs(matrizreconst[,j]-MATRIZdeMEDIAS[,j])
      +abs(matrizorig[,j]-MATRIZdeMEDIAS[,j]))}
  INDICE <- 1-(EAM/SMod)
  MEDIAdosVALORESdoINDICE <- mean(INDICE)
  list("EAM de cada variável\n" = EAM, "SMod de cada variável\n" = SMod, "Índice de Willmott
revisto 1985\n" = INDICE, "Média dos Valores do Índice de Willmott revisto 1985" = MEDIAdosVAL-
ORESdoINDICE)
}

-----
INDICEdeWILLMOTTrefinado2012<-function (matrizorig, matrizreconst, nlinhas, ncolunas)
{
  # matrizorig - matriz dos dados originais #
  # matrizreconst - matriz reconstruída #

```

```

# nlinhas - n° de linhas das matrizes matrizorig e matrizreconst #
# ncolunas - n° de colunas das matrizes matrizorig e matrizreconst #
# devolve o valor do índice #
# EAM - Erro Absoluto Médio vezes n de observações (|Pi-Oi|) #
# DAM - Desvio Absoluto Médio vezes n de observações (|Oi-Obarra|)#
MATRIZdeMEDIAS <- as.matrix(matrix(0, nrow = nlinhas, ncol = ncolunas))
for(i in 1:nlinhas) {for(j in 1:ncolunas){
  MATRIZdeMEDIAS[i,j] <- mean(matrizorig[,j])}
EAM <- c()
DAM <- c()
for(j in 1:ncolunas){
  EAM[j] <- sum(abs(matrizreconst[,j]-matrizorig[,j]))
  DAM[j] <- sum(abs(matrizorig[,j]-MATRIZdeMEDIAS[,j]))}
NDICE <- c()
for(i in 1:ncolunas){
  if (EAM[i]<=(2*DAM[i])) INDICE[i] <- 1-(EAM[i]/(2*DAM[i])) else
  INDICE[i] <- ((2*DAM[i])/EAM[i])-1}
MEDIAdosVALORESdoINDICE <- mean(INDICE)

list("EAM de cada variável\n" = EAM, "DAM de cada variável\n" = DAM, "Índice de Willmott
refinado 2012\n" = INDICE, "Média dos Valores do Índice de Willmott refinado 2012" = MEDIAdos-
VALORESdoINDICE)

}

-----
COEFdeNASHeSUTCLIFFE70<-function (matrizorig, matrizreconst, nlinhas, ncolunas)
{
# matrizorig - matriz dos dados originais #
# matrizreconst - matriz reconstruída #
# nlinhas - n° de linhas das matrizes matrizorig e matrizreconst #
# ncolunas - n° de colunas das matrizes matrizorig e matrizreconst #
# devolve o valor do índice #
# SQRE - Soma de Quadrados dos Resíduos (Pi-Oi)^2 #

```

```

# SQDO - Soma de Quadrados dos Desvios das Observações (Oi-Obarra)^2 #
MATRIZdeMEDIAS <- as.matrix(matrix(0, nrow = nlinhas, ncol = ncolunas))
for(i in 1:nlinhas) (for(j in 1:ncolunas){
    MATRIZdeMEDIAS[i,j] <- mean(matrizorig[,j]))
SQRE <- c()
SQDO <- c()
for(j in 1:ncolunas){
    SQRE[j] <- sum((matrizreconst-matrizorig)[,j]^2)
    SQDO[j] <- sum((matrizorig[,j]-MATRIZdeMEDIAS[,j])^2)}
COEF <- 1-(SQRE/SQDO)
MEDIAdosVALORESdoCOEF <- mean(COEF)

list("SQRE de cada variável\n" = SQRE, "SQDO de cada variável\n" = SQDO, "Coeficiente de
Nash e Sutcliffe 1970\n" = COEF, "Média dos Valores do Coeficiente de Nash e Sutcliffe 1970" =
MEDIAdosVALORESdoCOEF)
}

-----
COEFdeLEGATESeMCCABE99<-function (matrizorig, matrizreconst, nlinhas, ncolunas)
{
# matrizorig - matriz dos dados originais #
# matrizreconst - matriz reconstruída #
# nlinhas - n° de linhas das matrizes matrizorig e matrizreconst #
# ncolunas - n° de colunas das matrizes matrizorig e matrizreconst #
# devolve o valor do índice #
# EAM - Erro Absoluto Médio vezes n de observações (|Pi-Oi|) #
# DAM - Desvio Absoluto Médio vezes n de observações (|Oi-Obarra|)#
MATRIZdeMEDIAS <- as.matrix(matrix(0, nrow = nlinhas, ncol = ncolunas))
for(i in 1:nlinhas) (for(j in 1:ncolunas){
    MATRIZdeMEDIAS[i,j] <- mean(matrizorig[,j]))
EAM <- c()
DAM <- c()
for(j in 1:ncolunas){

```

```

EAM[j] <- sum(abs(matrizreconst[,j]-matrizorig[,j]))
DAM[j] <- sum(abs(matrizorig[,j]-MATRIZdeMEDIAS[,j]))}
COEF <- 1-(EAM/DAM)
MEDIAdosVALORESdoCOEF <- mean(COEF)

list("EAM de cada variável\n" = EAM, "DAM de cada variável\n" = DAM, "Coeficiente de Legates
e McCabe 1999\n" = COEF, "Média dos Valores do Coeficiente de Legates e McCabe 1999" = MEDI-
AdosVALORESdoCOEF)

}

```

Rotinas do Capítulo 7

No Capítulo que engloba a comparação entre várias técnicas, para análise do conjunto de dados da altura geopotencial, ao nível de um setor do Hemisfério Norte, e tendo em conta o seu elevado grau de complexidade, recorreu-se à construção de novas rotinas que permitissem analisar uma possível convergência de resultados entre as técnicas. De seguida apresentam-se as linhas de código usadas no *software* R.

```

### DADOS DA ALTURA DO GEOPOTENCIAL - 500 (Dados Médios Mensais)

### De Janeiro de 1958 a Dezembro de 2007 (para uma Latitude 20N-80N)

### UTILIZAR O PACKAGE CLIMPACT ATRAVÉS DA FUNÇÃO retrieve.nc

### Ir buscar o ficheiro netCDF com os dados guardados numa pasta do disco:

AltGeop.500.MM.1958.2007.Lat20N80N<-retrieve.nc("C:/SEBAST/Doutoramento/DADOS_CLIMA_netCDF/

DADOS_MENSAIS_Geopotencial_20_80_MAI0_2010/Jan1958_De2007_AGeop500_Lat20N80N.nc")

MAT2D.AltGeop<-function(X)

{

# Função para construir a matriz 2-D dos dados da Altura do Geopotencial #

# X: matriz 4-D dos dados com os valores por tempo, nível, latitude e longitude #

tim<-dim(X)[1]

lat<-dim(X)[3]

lon<-dim(X)[4]

```

```

matriz <- matrix(0, nrow = tim, ncol = lat*lon )

for (i in 1:tim)

{

vector<- c()

for (j in 1:lon)

vector <- append(vector, X[i,,j], after = length(vector))

matriz[i,] <- vector

}

return(matriz)

}

*****

MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N<-MAT2D.AltGeop(AltGeop.500.MM.1958.2007.Lat20N80N$dat)

dim(MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N)

MAT2D.AltGeop.MediasMensais.50anos<-function(X)

{

# Função para construir a matriz das médias (por cada ponto da grelha e por cada mês) #

# para cada um dos meses ao longo do intervalo temporal de 50 anos dos dados da Altura do Geopotencial #

# X: matriz 2-D dos dados com os valores por tempo, latitude*longitude #

# linhas de X: valores do tempo #

# colunas de X: cada um dos pontos da grelha #

nr<-nrow(X)

nc<-ncol(X)

anos<-nr/12

matmediasBloco12 <- matrix(0, nrow = 12, ncol = nc)

vector <- c()

for (j in 1:nc)

{

for (mes in 1:12)

{

for (k in 1:anos)

```

```

    {

        vector[k] <- X[12*(k-1)+mes,j]

    }

    matmediasBloco12[mes,j] <- mean(vector)

}

}

matmedias <- matrix(0, nrow = nr, ncol = nc)

for (i in 1:anos)

{

    matmedias[(12*(i-1)+1):(12*i),] <- matmediasBloco12

}

return(matmedias)

}

MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N.MatMedias.AR<-

MAT2D.AltGeop.MediasMensais.50anos(MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N)

dim(MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N.MatMedias.AR)

### ANOMALIAS E ANOMALIAS PONDERADAS:

Anomalias.AR<-MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N-MAT2D.AltGeop.500.MM.1958.2007.Lat20N80N.MatMedias.AR

MatLatBloco1<-matrix(AltGeop.500.MM.1958.2007.Lat20N80N$lat,600,25,byrow=TRUE)

MatLat<-matrix(rep(MatLatBloco1),600,3600)

SqrtCosMatLat<-sqrt(cos(MatLat*(pi/180)))

AnomaliasPonderadas.AR<-Anomalias.AR*SqrtCosMatLat

dim(AnomaliasPonderadas.AR)

-----

### GRÁFICOS DE VALORES PRÓPRIOS:

plot(AltGeop.500.MM.ACP.Lat20N80N.AR[[1]][1:50]^2, xlab="", ylab="")

plot(log(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[1]][1:50]), xlab="", ylab="")

##### Para MSSA (ACP Estendida):

### SÉRIES TEMPORAIS DOS SCORES:

```



```

CPEstendidasParaPlot<-AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,1:20]

colnames(CPEstendidasParaPlot)<-c("CPE1","CPE2","CPE3","CPE4","CPE5",

"CPE6","CPE7","CPE8","CPE9","CPE10","CPE11","CPE12","CPE13",

"CPE14","CPE15","CPE16","CPE17","CPE18","CPE19","CPE20")

CPEstendidasParaPlot

library(zoo)

plot(as.zoo(ts(CPEstendidasParaPlot)), nc = 2, type="l", cex.main = 1.5,

main="Scores das CPEs com M=180 e L=19", xlab="Meses")

### GRÁFICOS DE PARES DE CPEs:

par(mfrow=c(2,4))

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,4:5], type="l", xlab="CPE4", ylab="CPE5")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,8:9], type="l", xlab="CPE8", ylab="CPE9")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,10:11], type="l", xlab="CPE10", ylab="CPE11")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,12:13], type="l", xlab="CPE12", ylab="CPE13")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,14:15], type="l", xlab="CPE14", ylab="CPE15")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,17:18], type="l", xlab="CPE17", ylab="CPE18")

plot(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,19:20], type="l", xlab="CPE19", ylab="CPE20")

### GRÁFICOS DOS LOADINGS (dos VETORES PRÓPRIOS):

VETORESdasCPEstendidasParaPlot<-AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[2]][,1:20]

colnames(VETORESdasCPEstendidasParaPlot)<-c("CPE1","CPE2","CPE3","CPE4","CPE5","CPE6","CPE7","CPE8",

"CPE9","CPE10","CPE11","CPE12","CPE13","CPE14","CPE15","CPE16","CPE17","CPE18","CPE19","CPE20")

VETORESdasCPEstendidasParaPlot

plot(as.zoo(ts(VETORESdasCPEstendidasParaPlot)), nc = 2, type="l", main="", xlab="")

##### Para LagACI (ACI Estendida)

### SÉRIES TEMPORAIS DOS SCORES:

CIEstendidasParaPlot<-AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20$$

colnames(CIEstendidasParaPlot)<-c("CIE1","CIE2","CIE3","CIE4","CIE5",

"CIE6","CIE7","CIE8","CIE9","CIE10","CIE11","CIE12","CIE13","CIE14",

"CIE15","CIE16","CIE17","CIE18","CIE19","CIE20")

```

```

CIEstendidasParaPlot

plot(as.zoo(ts(CIEstendidasParaPlot)), nc = 2, type="l", cex.main = 1.5,

main="Scores das CIEs com M=180 e L=19 - Algoritmo FastICA", xlab="Meses")

### GRÁFICOS dos VETORES PRÓPRIOS:

VETORESdasCIEstendidasParaPlot<-t(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20$A)

colnames(VETORESdasCIEstendidasParaPlot)<-c("CIE1","CIE2","CIE3","CIE4","CIE5",

"CIE6","CIE7","CIE8","CIE9","CIE10","CIE11","CIE12","CIE13","CIE14","CIE15",

"CIE16","CIE17","CIE18","CIE19","CIE20")

VETORESdasCIEstendidasParaPlot

plot(as.zoo(ts(VETORESdasCIEstendidasParaPlot)), nc = 2, type="l", main="", xlab="")

-----

### CORRELAÇÕES:

cor(CPEstendidasParaPlot,CIEstendidasParaPlot)

### COSSENOS DE VETORES:

> CosAngulo

function(vector1,vector2)

{

# Cosseno do ângulo entre dois vectores #

norma<-c(0,0)

norma[1]<-sqrt(t(vector1)%*%vector1)

norma[2]<-sqrt(t(vector2)%*%vector2)

cosseno<-(t(vector1)%*%vector2)/(norma[1]*norma[2])

return(cosseno)

}

### ALGUNS EXEMPLOS:

CosAngulo(VETORESdasCPEstendidasParaPlot[,2],VETORESdasCIEstendidasParaPlot[,11])

CosAngulo(VETORESdasCPEstendidasParaPlot[,3],VETORESdasCIEstendidasParaPlot[,5])

#### ESPECTROS (todas as componentes relevantes):

spectrum(cbind(AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,7],

AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,2],AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,10],

AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,16],AG.500.MM.Lat20N80N.M180L19.p421.ACP.AR[[3]][,3],

```



```

plot(as.zoo(ts(CIEstendidasParaPlot.AMUSE)), nc = 2, type="l", cex.main = 1.5,

main="Scores das CIEs com M=180 e L=19 - Algoritmo AMUSE", xlab="Meses")

### GRÁFICOS DE PARES DE CIEs:

par(mfrow=c(2,4))

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[5:6], type="l", xlab="CIE5", ylab="CIE6")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[7:8], type="l", xlab="CIE7", ylab="CIE8")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[9:10], type="l", xlab="CIE9", ylab="CIE10")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[11:12], type="l", xlab="CIE11", ylab="CIE12")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[13:14], type="l", xlab="CIE13", ylab="CIE14")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[15:16], type="l", xlab="CIE15", ylab="CIE16")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[17:18], type="l", xlab="CIE17", ylab="CIE18")

plot(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[19:20], type="l", xlab="CIE19", ylab="CIE20")

### GRÁFICOS dos VETORES PRÓPRIOS:

VETORESdasCIEstendidasParaPlot.AMUSE<-t(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$A)

colnames(VETORESdasCIEstendidasParaPlot.AMUSE)<-c("CIE1","CIE2","CIE3","CIE4","CIE5","CIE6","CIE7","CIE8",

,"CIE9","CIE10","CIE11","CIE12","CIE13","CIE14","CIE15","CIE16","CIE17","CIE18","CIE19","CIE20")

VETORESdasCIEstendidasParaPlot.AMUSE

plot(as.zoo(ts(VETORESdasCIEstendidasParaPlot.AMUSE)), nc = 2, type="l", cex.main = 1.5,

main="Loadings dos vetores com M=180 e L=19 - Algoritmo AMUSE", xlab="")

### ESPECTROS da LagACI AMUSE (todas as componentes relevantes):

### Para ver as FREQUÊNCIAS que correspondem às 20 CIs:

FREQUENCIAS.M180L19.ACI.20.AMUSE<-spectrum(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[1:20],

spans=c(3,3),col=c(1,1), cex.main = 1.0, main="Espectro para CIEs - Algoritmo AMUSE",sub=""),

xlab="Frequência",ylab="Espectro",ci=0)

FREQUENCIAS.M180L19.ACI.20.AMUSE$freq

FREQUENCIAS.M180L19.ACI.20.AMUSE$spec

spectrum(cbind(AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[2],

AG.500.MM.Lat20N80N.M180L19.lagdados.ACI.20.AMUSE$$[4:8]),lty=c(1,2,3,4,5,6,7),

lwd=c(2,2,2,1,1,1,1),spans=c(3,3),col=c(1,1,1,1,1,1,1), cex.main = 1.0,

main="Espectro para CIEs - Algoritmo AMUSE",sub="",xlab="Frequência",ylab="Espectro",ci=0)

legend(locator(1),c("CIE2","CIE4","CIE5","CIE6","CIE7","CIE8"),bty="n",lty=c(1,2,3,4,5,6,7),

```


D.2 Dados da Pressão ao Nível do Mar

O conjunto de dados designado de "Pacific2" refere-se a um total de 8 variáveis, as quais representam 8 estações meteorológicas situadas no Pacífico Norte. Em cada uma das estações foi medida a Pressão Média ao Nível do Mar (MSLP - *Mean Sea Level Pressure*). Estes dados foram abordados nos Capítulos 3 e 6.

As designações e as respetivas localizações das estações em estudo são:

1. Crescent City - "CCity.CA", $41^{\circ} 45'$ N, $124^{\circ} 11'$ W;
2. San Diego - "SDie.CA", $32^{\circ} 43'$ N, $117^{\circ} 10'$ W;
3. San Francisco - "SFranc.CA", $37^{\circ} 48'$ N, $122^{\circ} 28'$ W;
4. Hilo - "Hilo.HW", $19^{\circ} 44'$ N, $155^{\circ} 03'$ W;
5. Honolulu - "Honol.HW", $21^{\circ} 18'$ N, $157^{\circ} 52'$ W;
6. Neah Bay - "NBay.WA", $48^{\circ} 22'$ N, $124^{\circ} 37'$ W;
7. Seldovia - "Seld.AK", $59^{\circ} 26'$ N, $151^{\circ} 43'$ W;
8. Sitka - "Sitka.AK", $57^{\circ} 03'$ N, $135^{\circ} 21'$ W.

As medições da Pressão foram efetuadas durante o período de janeiro de 1979 a dezembro de 1996, e correspondem a 216 observações mensais. As abreviaturas de referência nas estações são:

- AK - para o Estado do Alasca;
- CA - para o Estado da Califórnia;
- HW - para o Estado do Havai;
- WA - para o Estado de Washington.

Tabela D.1: Matriz das correlações das variáveis originais dos dados Pacific2

	CCity.CA	SDie.CA	SFranc.CA	Hilo.HW	Honol.HW	NBay.WA	Seld.AK	Sitka.AK
CCity.CA	1.000	0.435	0.815	0.293	0.1605	0.8122	0.565	0.645
SDie.CA	0.435	1.000	0.733	0.532	0.3349	0.1074	0.261	0.182
SFranc.CA	0.815	0.733	1.000	0.355	0.1804	0.4768	0.329	0.373
Hilo.HW	0.293	0.532	0.355	1.000	0.6664	0.1547	0.359	0.292
Honol.HW	0.161	0.335	0.180	0.666	1.0000	0.0791	0.260	0.203
NBay.WA	0.812	0.107	0.477	0.155	0.0791	1.0000	0.651	0.843
Seld.AK	0.565	0.261	0.329	0.359	0.2598	0.6506	1.000	0.850
Sitka.AK	0.645	0.182	0.373	0.292	0.2032	0.8429	0.850	1.000

D.2.1 Resultados

Na tabela D.1 e gráficos seguintes apresentamos alguns resultados referentes aos dados Pacific2.

Nas Figuras D.1 e D.2 apresentamos as séries originais e as correspondentes reconstruções à custa das três primeiras CPs e das duas CIs respetivamente, que permitem uma comparação visual ao longo do tempo para cada uma das oito estações meteorológicas.

D.3 Dados do Número de Vítimas Mortais em Acidentes de Viação

Este conjunto de dados constitui a série temporal do número mensal de vítimas mortais em acidentes de viação em Portugal Continental entre 1998 e 2009 e foi usado no Capítulo 4. Os dados correspondem a 144 observações mensais e foram obtidos por intermédio do INE, embora as entidades responsáveis pela recolha tenham sido a DGV e a ANSR.

D.3.1 Resultados

Nos gráficos das Figuras D.3, D.4, D.5 e D.6 apresentamos alguns resultados referentes a este conjunto de dados, nomeadamente as componentes principais que não ficaram retidas aquando da aplicação da técnica da SSA, por se considerar que apresentam uma estrutura de ruído,

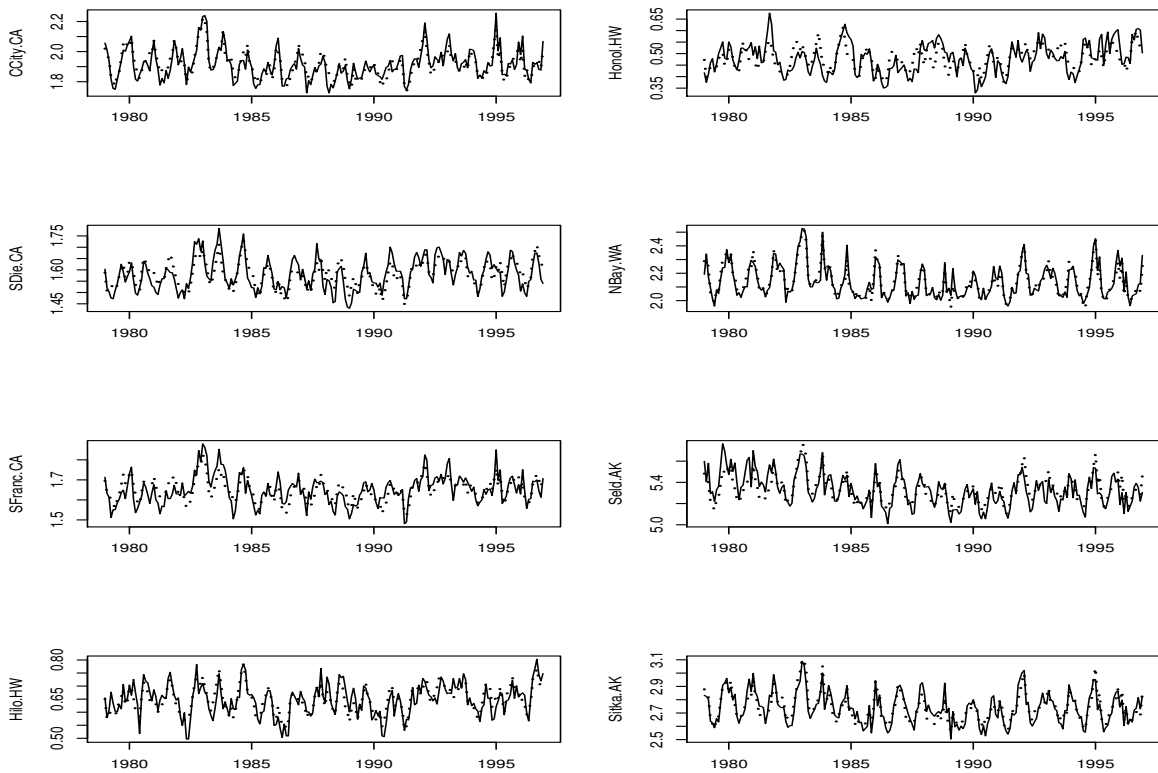


Figura D.1: Séries temporais originais para cada uma das oito estações meteorológicas e respectivas reconstruções às custa das duas CPs (CP1+CP2) em simultâneo

conforme se poderá verificar visualmente.

A Tabela D.2 apresenta os valores das 12 previsões mensais do número de vítimas mortais para o ano de 2010 e os limites inferiores e superiores dos intervalos de predição a 95%, para as correspondentes previsões, obtidas através do método empírico com base nos quantis calculados a partir da função de distribuição empírica da série de resíduos.

Tendo em conta a centragem na SSA, descrita na Secção 4.8, apresentamos alguns resultados, com base neste conjunto de dados, que comprovam alguns dos aspetos teóricos de cada tipo de centragem abordada.

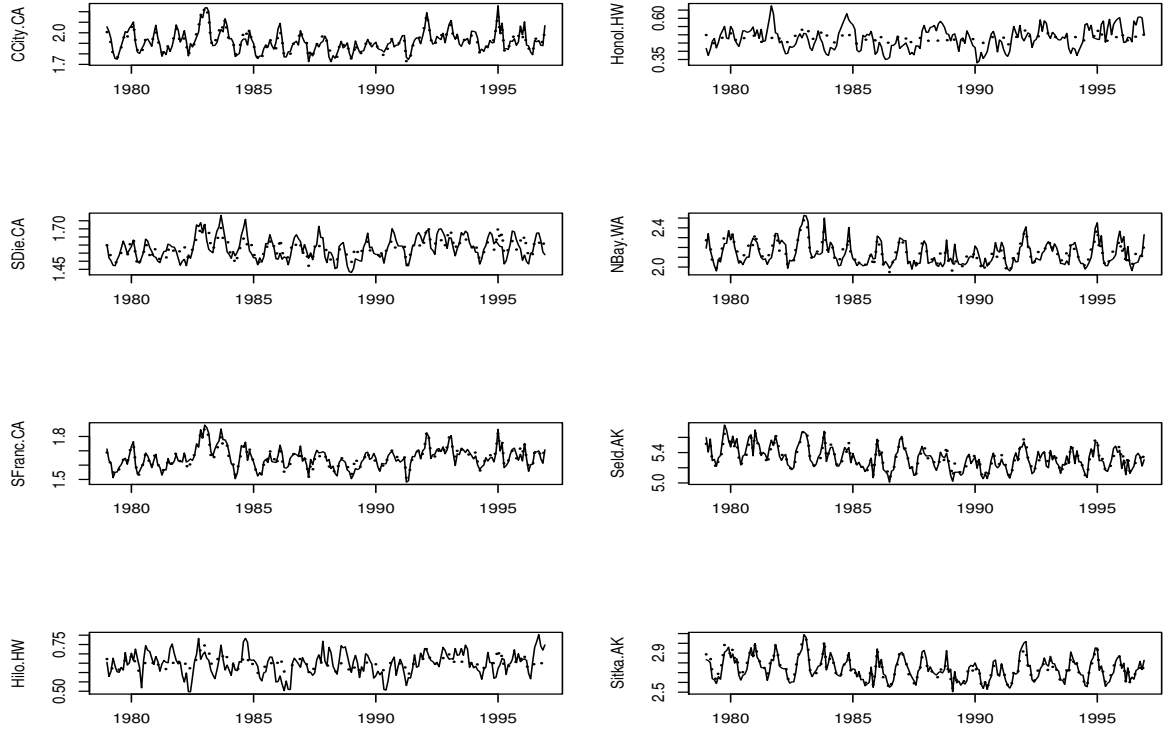


Figura D.2: Séries temporais originais para cada uma das oito estações meteorológicas e respectivas reconstruções às custa das duas CIs (CI1+CI2) em simultâneo

Centragem por colunas

A igualdade (4.14) está verificada, uma vez que para $k = 97$, $m = 48$ e $d = d^{cc} = 48$, obtém-se

$$\sum_{i=1}^{48} \lambda_i = 53745457$$

e

$$\lambda_{0(c)} + \sum_{i=1}^{48} \lambda_i^{cc} = 97 \times \|\bar{\mathbf{x}}_c\|^2 + \sum_{i=1}^{48} \lambda_i^{cc} = 50092596.69 + 3652860.31 = 53745457.$$

Centragem por linhas

A igualdade (4.16) está verificada, dado que para $k = 97$, $m = 48$, $d = 48$ e $d^{cl} = 47$ (pois o

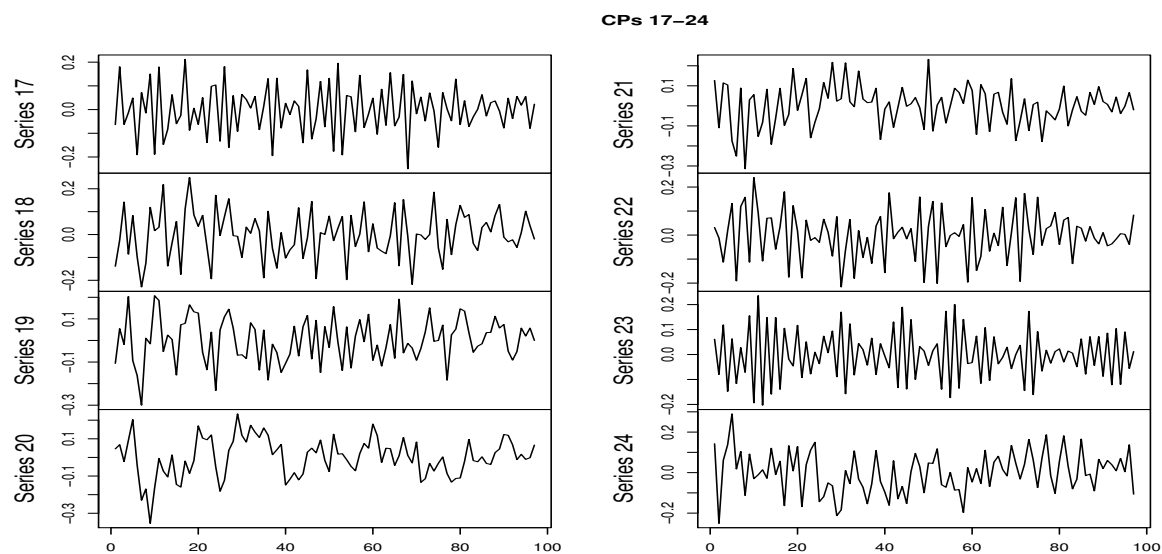


Figura D.3: Número de vítimas mortais em acidentes de viação: componentes principais CP17 a CP24

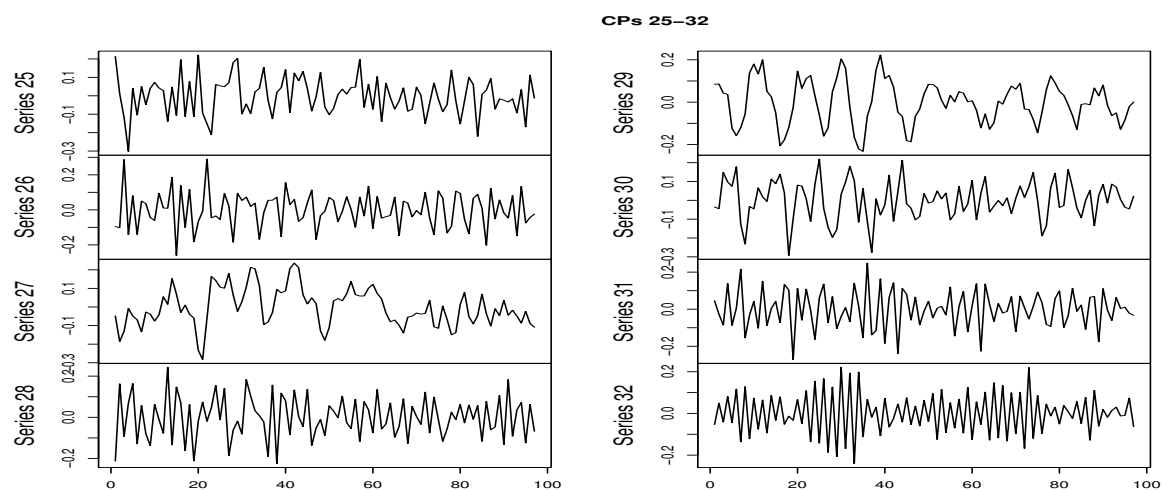


Figura D.4: Número de vítimas mortais em acidentes de viação: componentes principais CP25 a CP32

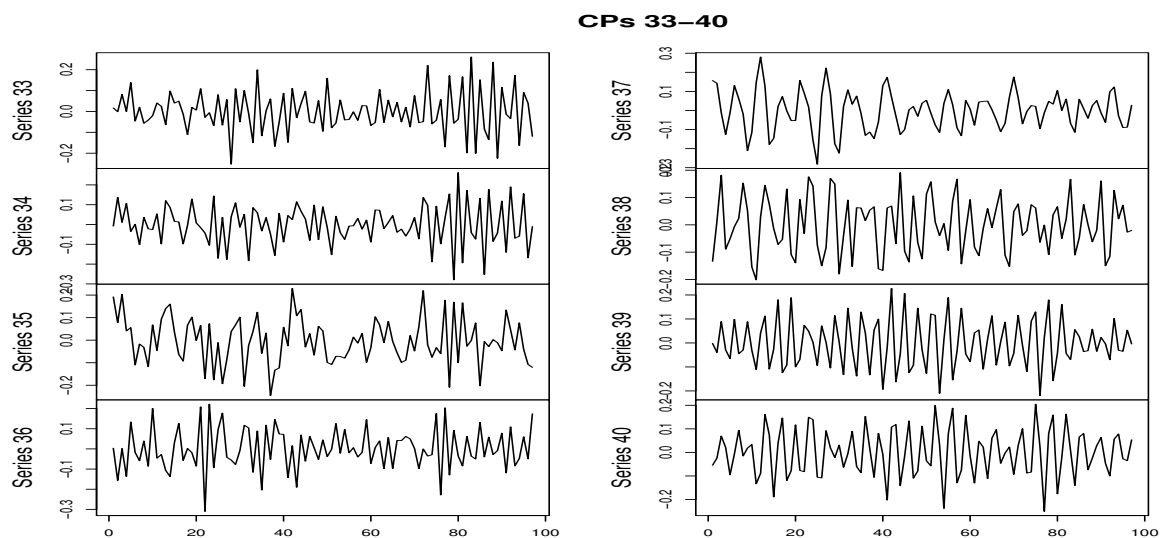


Figura D.5: Número de vítimas mortais em acidentes de viação: componentes principais CP33 a CP40

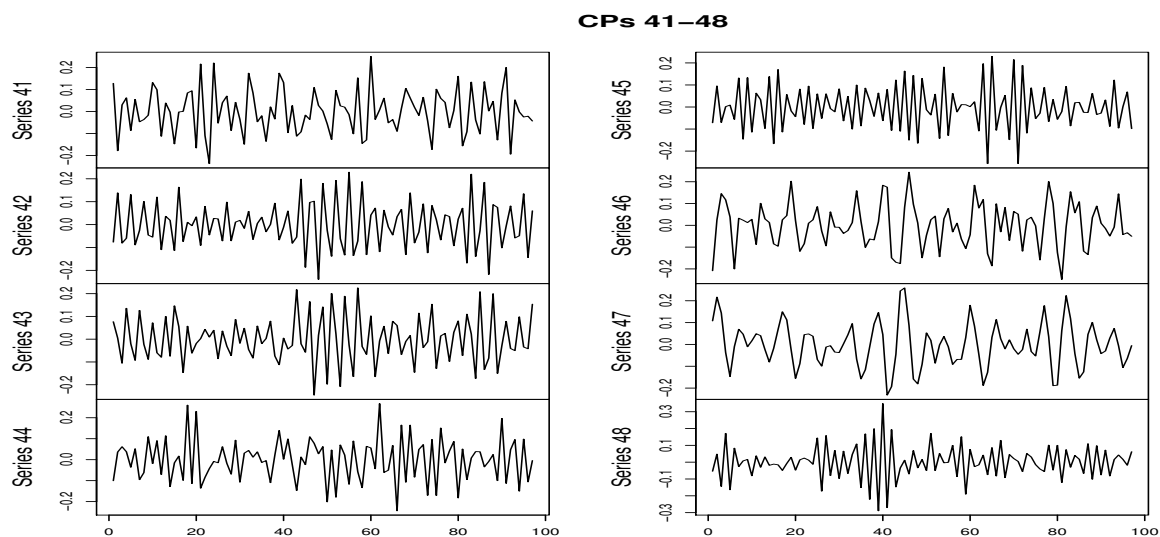


Figura D.6: Número de vítimas mortais em acidentes de viação: componentes principais CP41 a CP48

Tabela D.2: Previsões do número de vítimas mortais para os meses do ano de 2010 e respetivos intervalos de predição obtidos através do método empírico

Pontos	Previsões	Limite Inferior	Limite Superior
145	57.275	40.874	75.705
146	55.678	38.367	74.220
147	48.233	30.972	66.788
148	48.621	30.654	67.738
149	59.089	40.096	78.897
150	57.709	38.559	77.453
151	49.948	31.129	69.945
152	60.240	41.127	80.707
153	72.897	53.113	94.571
154	62.933	43.072	84.608
155	53.620	33.812	75.221
156	62.835	43.158	84.769

último valor próprio de \mathbf{S}^{cl} é praticamente nulo),

$$\sum_{i=1}^{48} \lambda_i = 53745457$$

e

$$\lambda_{0(l)} + \sum_{i=1}^{47} \lambda_i^{cl} = 48 \times \|\bar{\mathbf{x}}_l\|^2 + \sum_{i=1}^{47} \lambda_i^{cl} = 51942257.6 + 1803200.4 = 53745457.$$

Centragem dupla

A igualdade (4.18) está verificada para $k = 97$, $m = 48$, $d = 48$ e $d^{**} = 47$ (pois o último valor próprio de \mathbf{S}^{**} é praticamente nulo) uma vez que

$$\sum_{i=1}^{48} \lambda_i = 53745457$$

e

$$\begin{aligned} \lambda_{0(c)} + \lambda_{0(l)}^{(c)} + \sum_{i=1}^{47} \lambda_i^{**} &= 97 \times \|\bar{\mathbf{x}}_c\|^2 + 48 \times \|\bar{\mathbf{x}}^{cc}_l\|^2 + \sum_{i=1}^{47} \lambda_i^{**} \\ &= 50092596.69 + 2358908.36 + 1293951.95 = 53745457. \end{aligned}$$

Centragem na série original

Mais uma vez, efetuando a aplicação da teoria introduzida, a igualdade (4.20) fica verificada, para $k = 97$ e $m = 48$, uma vez que

$$\sum_{i=1}^{48} \lambda_i = 53745457$$

e

$$\begin{aligned} \lambda_{0(c)} - \lambda_{0(dif)} + \sum_{i=1}^{48} \lambda'_i &= 97 \times \|\bar{\mathbf{x}}_c\|^2 - 97 \times \|\bar{\mathbf{x}}_{dif}\|^2 + \sum_{i=1}^{48} \lambda'_i \\ &= 50092596.69 - 512526.37 + 4165386.68 = 53745457. \end{aligned}$$

D.4 Dados da Altura Geopotencial

Este conjunto de dados corresponde a valores mensais médios da altura geopotencial observada para um valor de pressão atmosférica a 500 milibar, desde janeiro de 1958 a dezembro de 2007 (50 anos). Os dados foram obtidos a partir dos arquivos de reanálise NCEP/NCAR do NOAA. O setor abrangido pelo estudo insere-se no Hemisfério Norte, entre 20°N - 80°N e 0°E - 357.5°E. Foi utilizada uma grelha espacial uniforme de 2.5° ao longo da latitude e da longitude, de tamanho $25 \times 144 = 3600$ pontos. A discussão dos dados encontra-se no Capítulo 7.

D.4.1 Resultados

Dada a grande complexidade dos dados, apresentamos como exemplos ilustrativos nas Figuras D.7 e D.8 dois gráficos referentes à altura geopotencial em duas localizações distintas para a mesma latitude, correspondentes a Portugal na região de Castelo Branco (40°N, 7.5°W) e a Filadélfia, no Estado da Pensilvânia (40°N, 75°W) ao longo dos 50 anos em estudo.

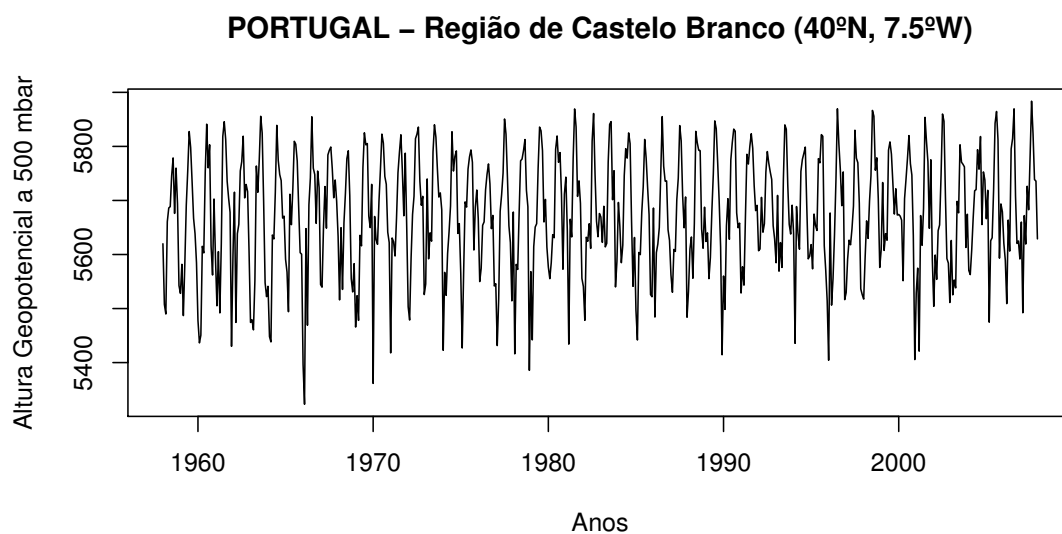


Figura D.7: Altura geopotencial (a 500 milibar) em Portugal na região de Castelo Branco (40°N, 7.5°W) ao longo dos 50 anos em estudo

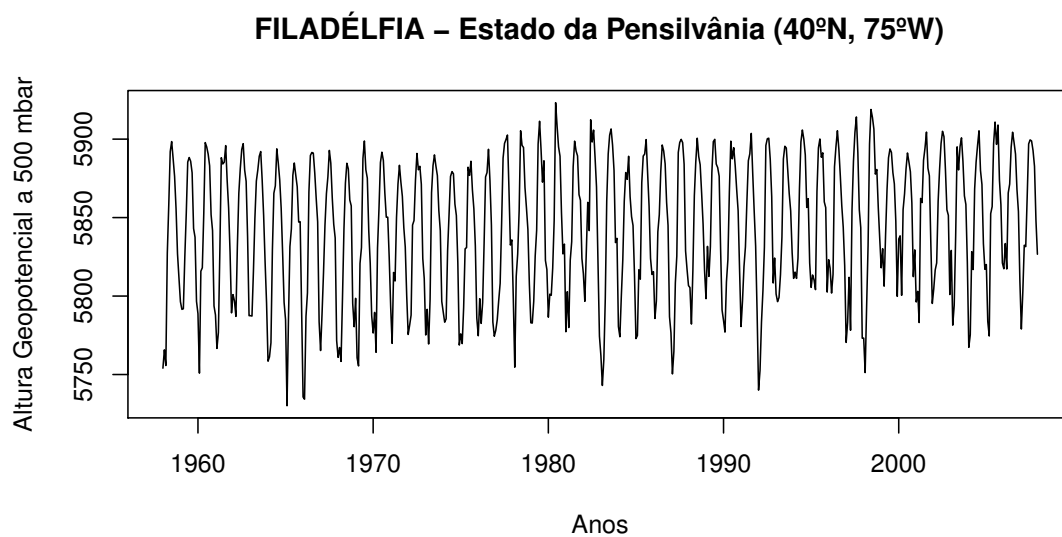


Figura D.8: Altura geopotencial (a 500 milibar) em Filadélfia, no Estado da Pensilvânia (40°N, 75°W) ao longo dos 50 anos em estudo

Bibliografia

- [1] Aires, F., Chédin, A. e Nadal, J-P. (2000). Independent component analysis of multivariate time series: Application to the tropical SST variability. *Journal of Geophysical Research*, **105(D13)**, 17,437-17,455.
- [2] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- [3] Aldrich, C. e Barkhuizen, M. (2003). Process system identification strategies based on the use of singular spectrum analysis. *Minerals Engineering*, **16**, 815-826.
- [4] Allen, A. P. e O'Connor, R. J. (2000). Hierarchical correlates of bird assemblage structure on Northeastern U.S.A. Lakes. *Environmental Monitoring and Assessment*, **62**, 15-37.
- [5] Allen, J. T., Hewett, P. C., Richardson, C. T., Ferland, G. J. e Baldwin, J. A. (2013). Classification and analysis of emission-line galaxies using mean field independent component analysis. *Monthly Notices of the Royal Astronomical Society*, **430(4)**, 3510-3536.
- [6] Allen, M. R. e Smith, L.A. (1994). Investigating the origins and significance of low-frequency modes of climate variability. *Geophysical Research Letters*, **21(10)**, 883-886.
- [7] Allen, M. R. e Smith, L. A. (1996). Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. *Journal of Climate*, **9**, 3373-3404.
- [8] Alonso, F. J., Del Castillo, J. M. e Pintado, P. (2005). Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics*, **38**, 1085-1092.

- [9] Alter, O. e Golub, G. H. (2006). Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. *Proceedings of the National Academy of Science*, **103**(32), 11828-11833.
- [10] Amari, S.-I., Cichocki, A. e Yang, H. H. (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems 8*, 757-763, MIT Press.
- [11] Amari, S.-I. (2000). Estimating functions of independent component analysis for temporally correlated signals. *Neural Computation*, **12**(9), 2083-2107.
- [12] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, **34**, 122-148.
- [13] Anemüller, J., Duann, J.-R., Sejnowski, T. J. e Makeig, S. (2006). Spatio-temporal dynamics in fMRI recordings revealed with complex independent component analysis. *Neurocomputing*, **69**, 1502-1512.
- [14] Ans, B. Héroult, J. e Jutten, C. (1985). Adaptive neural architectures: detection of primitives. *Proceedings of COGNITIVA '85*, 593-597, Paris, France.
- [15] Anton, H. e Rorres, C. (2010). *Elementary Linear Algebra with Supplemental Applications* (10th Edition), John Wiley & Sons, U.S.A..
- [16] Back, A. D. e Weigend, A. S. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, **8**(4), 473-484.
- [17] Baratta, D., Cicioni, G., Masulli, F. e Studer, L. (2003). Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting. *Neural Networks*, **16**, 375-387.
- [18] Barnett, e Hasselmann (1979). Techniques of linear prediction, with application to oceanic and atmospheric fields in the tropical Pacific. *Reviews of Geophysics*, **17**(5), 949-968.
- [19] Bartlett, M. S., Movellan, J. R. e Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, **13**(6), 1450-1464.

- [20] Basak, J., Sudarshan, A., Trivedi, D. e Santhanam M. S. (2004). Weather data mining using independent component analysis. *Journal of Machine Learning Research*, **5**, 239-253.
- [21] Basilevsky, A. e Hum, D. P. J. (1979). Karhunen-Loève analysis of historical time series with an application to plantation births in Jamaica. *Journal of the American Statistical Association*, **74**, 284-290.
- [22] Bell, A. J. e Sejnowski, T. J. (1995a). A non-linear information maximization algorithm that performs blind separation. In *Advances in Neural Information Processing Systems 7*, 467-474, MIT Press, Cambridge, MA.
- [23] Bell, A. J. e Sejnowski, T. J. (1995b). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, **7(6)**, 1129-1159.
- [24] Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. e Moulines, E. (1997). A blind source separation technique based on second order statistics. *IEEE Transactions on Signal Processing*, **45(2)**, 434-444.
- [25] Benlin, X., Fangfang, L., Xingliang, M. e Huazhong, J. (2008). Study on independent component analysis' application in classification and change detection of multispectral images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. **37(B7)**, 871-875.
- [26] Bingham, E. e Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, **10(1)**, 1-8.
- [27] Brandstein, M. S. (1998). On the use of explicit speech modeling in microphone array applications. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'98)*, 3613-3616, Seattle, Washington.
- [28] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Expanded Edition, Holden-Day, San Francisco.
- [29] Brockwell, P. e Davis, R. (1991). *Time Series: Theory and Methods*. 2nd Ed., Springer, New York.

- [30] Broomhead, D. S. e King, G. P. (1986a). Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217-236.
- [31] Broomhead, D. S. e King, G. P. (1986b). On the qualitative analysis of experimental dynamical systems. In S. Sarkar (Ed.). *Nonlinear Phenomena and Chaos*, 113-144. Adam Hilger, Bristol.
- [32] Bryant, E. H. e Atchley, W. R. (1975). *Multivariate Statistical Methods: Within Group Covariation*. Halsted Press, Stroudsburg.
- [33] Buckland, S. T. e Anderson, A. J. B. (1985). Multivariate analysis of Atlas data. In *Statistics in Ornithology*, eds. B. J. T. Morgan and P. M. North, 93-112. Springer-Verlag, Berlin.
- [34] Bumpus, H. C. (1898). The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. *Biological Lectures, Marine Biology Laboratory, Woods Hole*, 11th Lecture, 209-226.
- [35] Cadima, J., Cerdeira, J. O. e Minhoto, M. (2004). Computational aspects of algorithms for variable selection in the context of principal components. *Computational Statistics & Data Analysis*, **47**, 225-236.
- [36] Cadima, J. e Jolliffe, I. T. (1995). Loadings and correlations in the interpretation of principal components, *Journal of Applied Statistics*, **22(2)**, 203-214.
- [37] Cadima, J. e Jolliffe, I. T. (1996). Size- and shape-related principal component analysis. *Biometrics*, **52**, 710-716.
- [38] Cadima, J. e Jolliffe, I. T. (2009). On relationships between uncentred and column-centred principal component analysis. *Pakistan Journal of Statistics*, **25(4)**, 473-503.
- [39] Cangelosi, R. e Goriely, A. (2007). Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*, **2(2)**, doi:10.1186/1745-6150-2-2.
- [40] Cardoso, J.-F. (1989a). Blind identification of independent signals. In *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, USA.

- [41] Cardoso, J.-F. (1989b). Source separation using higher order moments. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'89)*, 2109-2112, Glasgow, UK.
- [42] Cardoso, J.-F. (1997). Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, **4**(4), 112-114.
- [43] Cardoso, J.-F. e Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proceedings-F*, **140**(6), 362-670.
- [44] Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behav. Res.*, **1**, 245-276.
- [45] Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Plenum Press, New York.
- [46] Chen, J.-M. e Harr, P. A. (1993). Interpretation of extended empirical orthogonal function (EEOF) analysis. *Monthly Weather Review*, **121**, 2631-2636.
- [47] Cheung, Y. e Xu, L. (2001). Independent component ordering in ICA time series analysis. *Neurocomputing*, **41**, 145-152.
- [48] Cichocki, A. e Amari, S.-I. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, New York.
- [49] Cichocki, A., Karhunen, J., Kasprzak, W. e Vigário, R. (1999). Neural networks for blind separation with unknown number of sources. *Neurocomputing*, **24**, 55-93.
- [50] Cichocki, A. e Moszczynski, L. (1992). A new learning algorithm for blind separation of sources. *Electronics Letters*, **28**(21), 1986-1987.
- [51] Cichocki, A., Thawonmas, R. e Amari, S. (1997). Sequential blind signal extraction in order specified by stochastic properties, *Electronics Letters*, **33**(1), 64-65.
- [52] Cichocki, A., Unbehauen, R. e Rummert, E. (1994). Robust learning algorithm for blind separation of signals. *Electronics Letters*, **30**(17), 1386-1387.

- [53] Cochran, R. N. e Horne, F. H. (1977). Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments. *Analytical Chemistry*, **49**, 846-853.
- [54] Cohen, S. J. (1983). Classification of 500 mb height anomalies using obliquely rotated principal components. *Journal of Climate and Applied Meteorology*, **22(12)**, 1975-1988.
- [55] Comon, P. (1989). Separation of stochastic processes. *In Workshop on Higher-Order Spectral Analysis*, 174-179, Vail, Colorado, USA.
- [56] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, **36**, 287-314.
- [57] Compagnucci, R. H., Araneo, D. e Canziani, P. O. (2001). Principal sequence pattern analysis: a new approach to classifying the evolution of atmospheric systems. *International Journal of Climatology*, **21**, 197-217.
- [58] Coughlin, K. T. e Tung, K. K. (2004). 11-Year solar cycle in the stratosphere extracted by the empirical mode decomposition method, *Advances in Space Research*, **34**, 323-329.
- [59] Cristescu, R., Ristaniemi, T., Joutsensalo, J. e Karhunen, J. (2000). Delay estimation in CDMA communications using a Fast ICA algorithm. *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, 105-110, Helsinki, Finland.
- [60] de Carvalho, M. e Rua, A. (2014). Nowcasting the US business cycle: singular spectrum analysis at work. *Working Paper* **16**, Banco de Portugal.
- [61] de Leeuw, J. e Mair, P. (2009). Gifi methods for optimal scaling in R: the package homals. *Journal of Statistical Software*, **31(4)**, 1-21.
- [62] De Vos, M., De Lathauwer, L. e Van Huffel, S. (2011). Spatially constrained ICA algorithm with an application in EEG processing. *Signal Processing*, **91**, 1963-1972.
- [63] Delfosse, N. e Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, **45**, 59-83.

- [64] Delorme, A., Sejnowski, T. e Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, **34**, 1443-1449.
- [65] Delorme, A., Palmer, J., Onton, J., Oostenveld, R. e Makeig, S. (2012). Independent EEG components are dipolar. *PLoS ONE*, **7(2)**, e30135. doi:10.1371/journal.pone.0030135.
- [66] Draghici, S., Graziano, F., Kettoola, S., Sethi, I. e Towfic, G. (2003). Mining HIV dynamics using independent component analysis. *Bioinformatics*, **19(8)**, 981-986.
- [67] Du, Q., Kopriva, I. e Szu, H., (2006). Independent-component analysis for hyperspectral remote sensing imagery classification. *Optical Engineering*, **45(1)**, 017008:1-13.
- [68] Elsner, J. B. e Tsonis, A. A. (1996). *Singular Spectrum Analysis. A New Tool in Time Series Analysis*. Plenum Press, New York.
- [69] Engreitz, J. M., Daigle, B. J. Jr., Marshall, J. J. e Altman, R. B. (2010). Independent component analysis: mining microarray data for fundamental human gene expression modules. *Journal of Biomedical Informatics*, **43(6)**, 932-944.
- [70] Escoufier, Y. (1986). A propos du choix des variables en analyse des données. *Metron*, **44**, 31-47.
- [71] Faria, P., Fregni, F., Sebastião, F., Dias, A. I. e Leal, A. (2012). Feasibility of focal transcranial DC polarization with simultaneous EEG recording: preliminary assessment in healthy subjects and human epilepsy. *Epilepsy & Behavior*, **25(3)**, 417-425.
- [72] Farmer, S. A. (1971). An investigation into the results of principal component analysis of data derived from random numbers. *Statistician*, **20(4)**, 63-72.
- [73] Feichtinger, H. G. e Strohmer, T. (1997). *Gabor Analysis and Algorithms*. Birkhauser.
- [74] Ferré, L. (1995). Selection of components in principal component analysis: A comparison of methods. *Computational Statistics & Data Analysis*, **19(6)**, 669-682.
- [75] Ferreira, A. (2003). *Aplicação da Análise em Componentes Independentes na Compressão de Imagem*. Tese de Mestrado, Instituto Superior Técnico da Universidade Técnica de Lisboa.

- [76] Ferreira, A. e Figueiredo, M. (2006). On the use of independent component analysis for image compression. *Signal Processing: Image Communication*, **21(5)**, 378-389.
- [77] Focardi, S. M. e Fabozzi, F. J. (2004). *The Mathematics of Financial Modeling and Investment Management*. Wiley, New Jersey.
- [78] Forbes, K. e Fiume, E. (2005). An efficient search algorithm for motion data using weighted PCA. *Eurographics/ACM SIGGRAPH Symposium on Computer Animation*.
- [79] Frisch, R. (1929). Correlation and scatter in statistical variables. *Nordic Statist. J.*, **8**, 36-102.
- [80] Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453-467.
- [81] Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F. e Yiou, P. (2002). Advanced spectral methods for climatic time series. *Reviews of Geophysics*, **40(1)**, 3.1-3.41.
- [82] Ghil, M. e Mo, K. C. (1991) Intraseasonal oscillations in the global atmosphere. Part I: Northern hemisphere and tropics. *Journal of the Atmospheric Sciences*, **48(5)**, 752-779.
- [83] Ghil, M. e Taricco, C. (1997). Advanced spectral analysis methods. In *Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Models*. G. C. Castagnoli e A. Provenzale (Eds.), Società Italiana di Fisica, Bologna, & IOS Press, Amsterdam, 137-159.
- [84] Ghil, M. e Vautard, R. (1991). Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, **350**, 324-327.
- [85] Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, Chichester.
- [86] Gillett, N. P., Graf, H. F. e Osborn, T. J. (2003). Climate change and the North Atlantic Oscillation. In *the North Atlantic Oscillation: Climatic significance and environmental impact*. James W. Hurrell, Yochanan Kushnir, Geir Ottersen e Martin Vsibeck (edit.), American Geophysical Union, 193-209.

- [87] Girshick, M. A. (1936). Principal components. *Journal of the American Statistical Association*, **31**, 519-528.
- [88] Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, **10**, 203-224.
- [89] Golub, G. e Van Loan, C. (1996). *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD.
- [90] Golyandina, N. E. e Korobeynikov, A. (2013). Basic singular spectrum analysis and forecasting with R. *Computational Statistics & Data Analysis*, **71**, 934-954.
- [91] Golyandina, N. E., Nekruktin, V. V. e Zhigljavsky, A. A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall, Boca Raton.
- [92] Golyandina, N. E. e Osipov, E. (2007). The "Caterpillar"- SSA method for analysis of time series with missing values. *Journal of Statistical Planning and Inference*, **137(8)**, 2642-2653.
- [93] Golyandina, N. E. e Stepanov, D. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. *Proceedings of the 5th Workshop on Simulation*, 293-298, St. Petersburg, Russia.
- [94] Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [95] Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325- 338.
- [96] Gower, J. C. e Hand, D. J. (1996). *Biplots*. Chapman and Hall, London.
- [97] Groth, A. e Ghil. M. (2011). Multivariate singular spectrum analysis and the road to phase synchronization. *Physical Review E*, **84(3)**, 036206.
- [98] Guerreiro, M. (2007). *Análise Diferenciada das Componentes Independentes de Potenciais Evocados P3a e P3b*. Tese de Mestrado, Universidade do Algarve.

- [99] Ham, F. M. e Faour, N. A. (1999). Infrasound signal separation using independent component analysis. *Proceedings of the 21st Seismic Research Symposium*, 133-140, Las Vegas, Nevada.
- [100] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.
- [101] Hannachi, A., Jolliffe, I. T., Stephenson, D. B. e Trendafilov, N. T. (2006). In search of simple structures in climate: simplifying EOFs. *International Journal of Climatology*, **26**, 7-28.
- [102] Hannachi, A., Jolliffe, I. T. e Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: a review. *International Journal of Climatology*, **27**, 1119-1152.
- [103] Hannachi, A., Unkel, S., Trendafilov, N. T. e Jolliffe, I. T. (2009). Independent component analysis of climate data: a new look at EOF rotation. *Journal of Climate*, **22**, 2797-2812.
- [104] Harman, H. H. (1976). *Modern Factor Analysis*. 3rd ed., Chicago: University of Chicago Press.
- [105] Harmel, R. D. e Smith, P. K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology*, **337**, 326-336.
- [106] Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, **5**, 239-257.
- [107] Hassani, H., Heravi, S. e Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, **25**, 103-118.
- [108] Hassani, H., Heravi, S. e Zhigljavsky, A. (2012). Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, **32(5)**, 395-408.
- [109] Hasselman, K. (1988). PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93**, 11,05-11,021.

- [110] Hérault, J. e Ans, B. (1984). Circuits neuronaux à synapses modifiables: décodage de messages composites par apprentissage non supervisé. *Comptes Rendus de l'Académie des Sciences*, **299(III-13)**, 525-528.
- [111] Hérault, J., Jutten, C. e Ans, B. (1985). Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. *In Actes du Xème colloque GRETSI*, 1017-1022, Nice, France.
- [112] Himberg, J. e Hyvärinen, A. (2001). Independent component analysis for binary data: an experimental study. *In Proceedings of the Third International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2001)*, 552-556, San Diego, California.
- [113] Horel, J. D. (1981). A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Monthly Weather Review*, **109**, 2080-2092.
- [114] Horn, R. A. e Johnson, C. R. (1985a). *Matrix Analysis - Vol. I*. Cambridge University Press, Cambridge.
- [115] Horn, R. A. e Johnson, C. R. (1985b). *Matrix Analysis - Vol. II*. Cambridge University Press, Cambridge.
- [116] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, **24**, 417-441, 498-520.
- [117] Hotelling, H. (1936). Simplified calculation of principal components. *Psychometrika*, **1**, 27-35.
- [118] Hoyer, P. O. e Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, **11(3)**, 191-210.
- [119] Huang, D.-Y. e Tseng, S.-T. (1992). A decision procedure for determining the number of components in principal component analysis. *J. Statist. Plan. Inf.*, **30**, 63-71.
- [120] Huang, D.-S. e Mi, J.-X. (2007). A new constrained independent component analysis method. *IEEE Transactions on Neural Networks*, **18(5)**, 1532-1535.

- [121] Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics*, **13(2)**, 435-475.
- [122] Hurrell, J. W. (1995). Decadal Trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, **269(5224)**, 676-679.
- [123] Hurri, J., Hyvärinen, A. e Oja, E. (1997). Wavelets and natural image statistics. *Proceedings of the 10th Scandinavian Conference on Image Analysis*, 13-18, Lappeenranta, Finland.
- [124] Hyvärinen, A. (1997). A family of fixed-point algorithms for independent component analysis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, **5**, 3917-3920, Munich, Germany.
- [125] Hyvärinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. In *Advances in Neural Information processing Systems*, **10**, 273-279.
- [126] Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, **10(3)**, 626-634.
- [127] Hyvärinen, A. (1999b). Fast independent component analysis with noisy data using gaussian moments. In *Proceedings of the International Symposium on Circuits and Systems*, 57-61, Orlando, Florida.
- [128] Hyvärinen, A. (1999c). Survey on independent component analysis. *Neural Computing Surveys*, **2**, 94-128.
- [129] Hyvärinen, A. (1997). One-unit contrast functions for independent component analysis: A statistical analysis. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing VII*, 388-397, Amelia Island, Florida.
- [130] Hyvärinen, A. e Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9(7)**, 1483-1492.
- [131] Hyvärinen, A. e Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, **13(4-5)**, 411-430.

- [132] Hyvärinen, A., Hoyer, P. O. e Oja, E. (2001a). Image denoising by sparse code shrinkage. Haykin, S. e Kosko, B. Editores, *Intelligent Signal Processing*. IEEE Press, 554-568.
- [133] Hyvärinen, A., Karhunen, J. e Oja, E. (2001b). *Independent Component Analysis*. John Wiley & Sons, Inc., U.S.A..
- [134] Ikeda, S. e Murata, N. (1999). A method of ICA in time-frequency domain. *Proceedings of the 1th International Workshop on Independent Component Analysis and Signal Separation (ICA '99)*, 365-370, Aussois, France.
- [135] Ilin, A., Valpola, H. e Oja, E. (2005). Semibind source separation of climate data detects *El Niño* as the component with the highest interannual variability. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN'2005)*, 1722-1727, Montréal, Québec, Canada.
- [136] Jackson, J. E. (1991). *A User's Guide to Principal Components*. Wiley, New York.
- [137] James, C. J. e Hesse, C. H. (2005). On the use of spectrally constrained ICA applied to single-channel ictal EEG recordings within a dynamical embedding framework. *Proceedings of the 27th IEEE EMBS Annual International Conference*, 956-959, Shanghai, China.
- [138] Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. *Applied Statistics*, **16**, 225-236.
- [139] Johansson, D., Lindgren, P. e Berglund, A. (2003). A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. *Bioinformatics*, **19**(4), 467-473.
- [140] Jolliffe, I. T. (1972). Discarding variables in a principal component analysis 1: Artificial data. *Applied Statistics*, **21**, 160-173.
- [141] Jolliffe, I. T. (1987). Rotation of principal components: Some comments. *Journal of Climatology*, **7**, 507-510.
- [142] Jolliffe, I. T. (1989). Rotation of ill-defined principal components. *Applied Statistics*, **38**, 139-147.

- [143] Jolliffe, I. T. (1995). Rotation of principal components: Choice of normalization constraints. *Journal of Applied Statistics*, **22**, 29-35.
- [144] Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd Edition). Springer-Verlag, New York.
- [145] Jolliffe, I. T. e Uddin, M. (2000). The simplified component technique. An alternative to rotated principal components. *Journal of Computat. Graph. Statist.*, **9**, 689-710.
- [146] Jones, M. C. e Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society, Series A* , **150(1)**, 1-36.
- [147] Joutsensalo, J. e Ristaniemi, T. (1998). Learning algorithms for blind multiuser detection in CDMA downlink. *Proceedings of the 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 267-270, Boston, Massachusetts.
- [148] Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V. e Sejnowski, T. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, **10**, 894-900, MIT Press.
- [149] Jutten, C. e Karhunen, J. (2003). Advances in nonlinear blind source separation. *Proceedings of the 4th International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 245-256, Nara, Japan.
- [150] Jutten, C. e Taleb, A. (2000). Source separation: from dusk till dawn. *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, 15-26, Helsinki, Finland.
- [151] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, **20**, 141-151.
- [152] Kalnay, E., kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. e Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, **77**, 437-470.

- [153] Karhunen, J. (1996). Neural approaches to independent component analysis and source separation. *Proceedings of the 4th European Symposium on Artificial Neural Networks (ESANN'96)*, 249-266, Bruges, Belgium.
- [154] Keppenne, C. L. e Ghil, M. (1992). Adaptive filtering and prediction of the southern oscillation index. *Journal of Geophysical Research*, **97D18**, 20,449-20,454.
- [155] Kiers, H. A. L. (1993). A comparison of techniques for finding components with simple structure. In *Multivariate Analysis: Future Directions 2*, eds. C. M. Cuadras and C. R. Rao, 67-86. North-Holland, Amsterdam.
- [156] Kimoto, M., Ghil, M. e Mo, K. C. (1991). Spatial structure of the extratropical 40-day oscillation. In *Proceedings of the 8th Conference on Atmospheric and Oceanic Waves and Stability*, 115-116, Amer. Meteorol. Soc., Boston.
- [157] Kiviluoto, K. e Oja, E. (1998). Independent component analysis for parallel financial time series. *Proceedings of ICONIP'98*, **2**, 895-898.
- [158] Kolenda, T. e Hansen, L. K. (1999). Independent components in text. *Advances in Independent Component Analysis*, 235-256.
- [159] Krause, P., Boyle, D. P. e Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, **5**, 89-97.
- [160] Kroonenberg, P. M. (1983). *Three-Mode Principal Component Analysis*. DSWO Press, Leiden.
- [161] Kumar, S., Desai, J.V., Mukherjee, S. e Das, P. K. (2013). Suitability of independent component analysis in digital image forgery detection. *International Journal of Engineering and Technology*, **5(1)**, 226-231.
- [162] Labitzke, K. e Matthes, K. (2005). Eleven-year solar cycle variations in the atmosphere: observations, mechanisms and models. *The Holocene*, **13(3)**, 311-317.
- [163] Lacoume, J.-L. e Ruiz, P. (1988). Sources identification: a solution based on cumulants. *IEEE ASSP Workshop*, Mineapolis, USA.

- [164] Lappalainen, H., Giannakopoulos, X., Honkela, A. e Karhunen, J. (2000). Nonlinear independent component analysis using ensemble learning: experiments and discussion. *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, 351-356, Helsinki, Finland.
- [165] Lavado, N. (2012). *Extensões via Splines da Análise em Componentes Principais*. Tese de Doutorado, ISCTE - Instituto Universitário de Lisboa.
- [166] Legates, D. R. e McCabe, G.J. Jr. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, **35(1)**, 233-241.
- [167] Lima, C., Silva, C., Tavares, A. e Oliveira, J. (2003). Blind source separation by independent component analysis applied to electroencephalographic signals. *International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, 3, Firenze.
- [168] Lombard, A., Zheng, Y., Buchner, H. e Kellermann, W. (2011). TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis. *IEEE Transactions on Audio, Speech and Language Processing*, **19(6)**, 1490-1503.
- [169] Lu, W. e Rajapakse, J. C. (2003). Eliminating indeterminacy in ICA. *Neurocomputing*, **50**, 271-290.
- [170] Lu, W. e Rajapakse, J. C. (2005). Approach and applications of constrained ICA. *IEEE Transactions on Neural Networks*, **16(1)**, 203-212.
- [171] Lu, C.-J., Lee, T.-S. e Chiu, C.-C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, **47(2)**, 115-125.
- [172] Maciel, C. D. e Pereira, W. C. A. (2000). RF ultrasound echo decomposition using singular-spectrum analysis. *Acoustical Imaging*, **24**, 101-106, Editado por Hua Lee, Kluwer Academic/Plenum Publishers.

- [173] Makeig, S., Bell, A. J., Jung, T.-P. e Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, **8**, 145-151, MIT Press.
- [174] Makino, S., Araki, S., Mukai, R. e Sawada, H. (2004). Audio source separation based on independent component analysis. *Proceedings of the 2004 International Symposium on Circuits and Systems (ISCAS '04)*, **5**, 668-671.
- [175] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11(7)**, 674-693.
- [176] Mandel, J. (1972). Principal components, analysis of variance and data structure. *Statistica Neerlandica*, **26**, 119-129.
- [177] Manduchi, R. e Portilla, J. (1999). Independent component analysis of textures. *Proceedings of the 7th IEEE International Conference on Computer Vision*, **2**, 1054-1060.
- [178] Maroco, J. (2007). *Análise Estatística com Utilização do SPSS* (3ª ed.) Edições Sílabo, Lisboa.
- [179] Marques, G. (2002). Análise em componentes independentes baseada na informação mútua. *Conferência Científica e Tecnológica em Engenharia: "O Saber do Passado e o Desafio do Futuro"*, Lisboa, Portugal.
- [180] Matsuoka, K., Ohoya, M. e Kawamoto, M. (1995). A neural net for blind separation of nonstationary signals. *Neural Networks*, **8(3)**, 411-419.
- [181] McCullagh, P. (1995). *Tensor Methods in Statistics*. Chapman & Hall, Reprint.
- [182] Meulman, J. (1986). *A Distance Approach to Nonlinear Multivariate Analysis*. DSWO Press, Leiden.
- [183] Meulman, J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformation of variables. *Psychometrika*, **57**, 539-565.
- [184] Mielke, P. W. Jr. e Berry, K. J. (2001). *Permutation Methods: A Distance Function Approach*. Springer-Verlag, New York, 352.

- [185] Mikhael, W., Ranganathan, R. e Yang, T. (2010). Complex adaptive ICA employing the conjugate gradient technique for signal separation in time-varying flat fading channels. *Circuits Systems and Signal Processing*, **29(3)**, 469-480.
- [186] Minhoto, M. (2009). *Seleção de Variáveis em Estatística Multivariada*. Tese de Doutorado, Universidade Técnica de Lisboa.
- [187] Molgedey, L. e Schuster, H. G. (1994). Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, **72(23)**, 3634-3637.
- [188] Morrison, D. F. (1976). *Multivariate Statistical Methods*. 2nd Ed., Tokyo: McGraw-Hill Kogakusha.
- [189] Mote, P. W., Clark, H. L., Dunkerton, T. J., Harwood, R. S. e Pumphrey, H. C. (2000). Intraseasonal variations of water vapor in the tropical upper troposphere and tropopause region. *Journal of Geophysical Research*, **105(D13)**, 17457-17470.
- [190] Murteira, B., Müller, D. e Turkman, K. (1993). *Análise de Sucessões Cronológicas*. McGraw-Hill, Lisboa.
- [191] Nadal, J.-P., Korutcheva, E. e Aires, F. (2000). Blind source separation in the presence of weak sources. *Neural Networks*, **13(6)**, 589-596.
- [192] Nascimento, J. e Dias, J. (2005). Does independent component analysis play a role in unmixing hyperspectral data?. *IEEE Trans. on Geoscience and Remote Sensing*, **43(1)**, 175-187.
- [193] Nash, J. E. e Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology*, **10(3)**, 282-290.
- [194] Nguyen, H. e Zheng, R. (2011). Binary independent component analysis with OR mixtures. *IEEE Transactions of Signal Processing*, **59(7)**, 3168-3181.
- [195] Oja, E., Kiviluoto, K. e Malaroui, S. (2000). Independent component analysis for financial time series. *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, 111-116.

- [196] Oliveira, I. (2003). *Correlated Data in Multivariate Analysis*. Ph.D. Thesis in Statistics at the University of Aberdeen, Scotland, United Kingdom.
- [197] Ollila, E., Oja, H. e Koivunen, V. (2008). Complex-valued ICA based on a pair of generalized covariance matrices. *Computational Statistics and Data Analysis*, **52**, 3789-3805.
- [198] Orihara, Y., Kamogawa, M., Nagao, T. e Uyeda, S. (2009). Independent component analysis of geoelectric field data in the northern Nagano, Japan. *Proceedings of the Japan Academy, Series B*, **85(9)**, 435-442.
- [199] Pajunen, P. (1998). Blind source separation using algorithmic information theory. *Neurocomputing*, **22**, 35-48.
- [200] Parmar, S. e Unhelkar, B. (2010). Performance comparisons of ICA algorithms to DS-SS-CDMA detection. *Journal of Telecommunications*, **1(1)**, 53-56.
- [201] Patterson, K., Hassani, H., Heravi, S. e Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics*, **38(10)**, 2183-2211.
- [202] Pearlmutter, B. A. e Parra, L. C. (1996). A context-sensitive generalization of ICA. *International Conference on Neural Information Processing (ICONIP'96)*, 151-157, Springer.
- [203] Pearson, K. (1901). On lines and planes of close fit to system of points in space. *Philosophical Magazin (6)*, **2**, 559-572.
- [204] Penedo, M. (2012). *Análise por Componentes Independentes dos Exames de EEG/IfRM de Epilepsia*. Tese de Mestrado, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- [205] Pereira, H. (2012). *Reconhecimento de Artefactos Relativos à Fala em Registos de EEG*. Tese de Mestrado, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- [206] Pereira, J. (2003). *Análise em Componentes Independentes na Separação Cega de Fontes Coincidentes no Espectro*. Tese de Mestrado, Faculdade de Ciências / Faculdade de Engenharia da Universidade do Porto.

- [207] Pham, D.-T., Garrat, P. e Jutten, C. (1992), Separation of a mixture of independent sources through a maximum likelihood approach. *Proceedings EUSIPCO*, 771-774.
- [208] Pike, E.R., McWhirter, J.G., Bertero, M. e de Mol, C. (1984). Generalized information theory for inverse problems in signal processing. *IEEE Proc.* **131**, 660-667.
- [209] Pinto da Costa, J., Alonso, H. e Roque, L. (2011). A weighted principal component analysis and its application to gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **8(1)**, 246-252.
- [210] Plaut, G. e Vautard, R. (1994). Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. *Journal of the Atmospheric Sciences*, **51(2)**, 210-236.
- [211] Preisendorfer, R. W. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, Amsterdam.
- [212] Puga, A. T. (2000). *Análise e Compressão de Imagem Usando Componentes Independentes, Simétricas e Principais*. Tese de Doutorado, Faculdade de Engenharia da Universidade do Porto.
- [213] Ramsay, J. O., ten Berge, J. e Styán, G. P. H. (1984). Matrix correlation. *Psychometrika*, **49(3)**, 403-423.
- [214] Ramsay, J. O. e Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.
- [215] Ranganathan, R. e Mikhael, W. B. (2007). A novel interference suppression technique employing complex adaptive ICA for time-varying channels in diversity wireless QAM receivers. *Circuits and Systems, ISCAS 2007. IEEE International Symposium*, 101-104, 27-30.
- [216] Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A*, **26**, 329-358.
- [217] Rasmusson, E. M., Wang, X. e Ropelewski, C. F. (1990). The biennial component of ENSO variability. *Journal of Marine Systems*, **1**, 71-96.
- [218] Rencher, A. C. (1995). *Methods of Multivariate Analysis*. Wiley, New York.

- [219] Reyment, R. A. e Jöreskog, K. G. (1993). *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press, Cambridge.
- [220] Richman, M. B. (1986). Rotation of principal components. *Journal of Climatology*, **6**, 293-335.
- [221] Richman, M. B. (1987). Rotation of principal components: A reply. *Journal of Climatology*, **7**, 511-520.
- [222] Ristaniemi, T. e Joutsensalo, J. (1999). On the performance of blind source separation in CDMA downlink. *Proceedings of the 1th International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, 437-441, Aussois, France.
- [223] Robert, P. e Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV coefficient. *Applied Statistics*, **25(3)**, 257-265.
- [224] Robison, E. A. (1982). A historical perspective of spectrum estimation. *Proceedings of the IEEE*, **70**, 885-907.
- [225] Rodrigues, P. C. e de Carvalho, M. (2008). Monitoring calibration of the singular spectrum analysis method. *Proceedings of the CompStat 2008, 18th International Conference on Computational Statistics* (P. Brito Ed.). Physica-Verlag, 955-964, Porto, Portugal.
- [226] Rodrigues, P. C. e Lima, A. T. (2009). Analysis of an European union election using principal component analysis. *Statistical Papers*, **50**, 895-904, doi:10.1007/s00362-009-0264-2.
- [227] Ropelewski, C. F. e Jones, P. D. (1987). An extension of the Tahiti-Darwin southern oscillation index. *Monthly Weather Review*, **115**, 2161-2165.
- [228] Salgado, D. R. e Alonso, F. J. (2006). Tool wear detection in turning operations using singular spectrum analysis. *Journal of Materials Processing Technology*, **171**, 451-458.
- [229] Schoellhamer, D. H. (2001). Singular spectrum analysis for time series with missing data. *Geophys. Res. Lett*, **28**, 3187-3190.
- [230] Sebastião, F. e Oliveira, I. (2009). Estudo de séries temporais na análise em componentes principais e na análise em componentes independentes. *Actas do XVI Congresso Anual da Sociedade Portuguesa de Estatística*, Edições da SPE, 633-640, Vila Real, Portugal.

- [231] Sebastião, F. e Oliveira, I. (2013). Independent component analysis for extended time series in climate data. *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications. Studies in Theoretical and Applied Statistics*, 427-436.
- [232] Sebastião, F. e Oliveira, I. (2015). Vítimas mortais em acidentes de viação em Portugal Continental: componentes principais e previsão. *XVIII Jornadas de Classificação e Análise de Dados*, Vila Real (aceite para publicação).
- [233] Simão, N. (2010). *Análise Agregada da Eficácia das Políticas de Segurança Rodoviária em Portugal*. Dissertação de Mestrado. Instituto Superior Técnico da Universidade Técnica de Lisboa.
- [234] Singer, A. e Coifman, R. R. (2008). Non-linear independent component analysis with diffusion maps. *Applied and Computational Harmonic Analysis*, **25**, 226-239.
- [235] Somers, K. M. (1986). Multivariate allometry and removal of size with principal component analysis. *Systematic Zoology*, **35**, 359-368.
- [236] Somers, K. M. (1989). Allometry, isometry and shape in principal component analysis. *Systematic Zoology*, **38**, 169-173.
- [237] ter Braak, C. J. F. (1983). Principal components biplots and alpha and beta diversity. *Ecology*, **64**, 454-462.
- [238] Thurstone, L. L. (1931). Multiple factor analysis. *Psychol. Rev.*, **38**, 406-427.
- [239] Timmerman, M. E. e Kiers, H. A. L. (2000). Three-mode principal component analysis: choosing the numbers of components and sensitivity to local optima. *Brit. J. Math. Stat. Psychol.*, **53**, 1-16.
- [240] Tong, L., Liu, R.-W., Soon, V. C. e Huang, Y.-F. (1991). Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, **38(5)**, 499-509.
- [241] Torkkola, K. (1999). Blind separation for audio signals - Are we there yet?. *Proceedings of the 1th International Workshop on Independent Component Analysis and Blind Signal Separation (ICA '99)*, 239-244, Aussois, France.

- [242] Trenberth, K. E. (1984). Signal versus noise in the southern oscillation. *Monthly Weather Review*, **112**, 326-332.
- [243] Trendafilov, N. T. e Unkel, S. (2011). Exploratory factor analysis of data matrices with more variables than observations, *Journal of Computational and Graphical Statistics*, **20(4)**, 874-891.
- [244] Tseng, Y.-H. e Tsai, D.-M. (2006). Using Independent component analysis based process monitoring in TFT-LCD manufacturing. *Journal of the Chinese Institute of Industrial Engineers*, **23(3)**, 262-267.
- [245] Unkel, S., Hannachi, A., Trendafilov, N. T. e Jolliffe, I. T. (2011). Independent component analysis for three-way data with an application from atmospheric science. *Journal of Agricultural, Biological, and Environmental Statistics*, **16(3)**, 319-338.
- [246] Üzümcü, M., Frangi, A. F., Reiber, J. e Lelieveldt, B. (2003). Independent component analysis in statistical shape models. *Medical Imaging. Proceedings of SPIE*, **5032**, 375-83.
- [247] van den Dool, H., Saha, S. e Johansson, Å. (2000). Empirical orthogonal teleconnections. *Journal of Climate*, **13**, 1421-1435.
- [248] van den Dool, H. (2007). *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press.
- [249] Varadi, F., Pap, J. M., Ulrich, R. K., Bertelo, L. e Henney, J.C. (1999). Searching for signal in noise by random-lag singular spectrum analysis. *The Astrophysical Journal*, **526**, 1052-1061.
- [250] Vautard, R. e Ghil, M. (1989). Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35**, 395-424.
- [251] Vautard, R., Yiou, P. e Ghil, M. (1992). Singular spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, **58**, 95-126.
- [252] Vigário, R. (1997). Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. Clin. Neurophysiol.*, **103(3)**, 395-404.

- [253] Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R. e Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. *Advances in Neural Information Processing Systems*, **10**, 229-235, MIT Press.
- [254] Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M. e Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, **47(5)**, 589-593.
- [255] von Storch, H. e Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.
- [256] Wallace, J. M. e Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, **109**, 784-812.
- [257] Watterson, I. G. (1996). Non-dimensional measures of climate model performance. *International Journal of Climatology*, **16**, 379-391.
- [258] Weare, B. C. e Nasstrom, J. S. (1982). Examples of extended empirical orthogonal function analyses. *Monthly Weather Review*, **110**, 481-485.
- [259] Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences* (2nd Edition). Academic Press. USA.
- [260] Willmott, C. J. (1981). On the validation of models. *Physical Geography*, **2**, 184-194.
- [261] Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, **63**, 1309-1313.
- [262] Willmott, C. J. (1984). On the evaluation of model performance in physical geography. *Spatial Statistics and Models*, edited by G. L. Gaile and C. J. Willmott, D. Reidel, Norwell, Mass., 443-460.
- [263] Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J. e Rowe, C. M. (1985). Statistics for the evaluation of model performance. *Journal of Geophysical Research*, **90(C5)**, 8995-9005.
- [264] Willmott, C. J., Robeson, S. M. e Matsuura, K. (2012). Short communication: A refined index of model performance. *International Journal of Climatology*, **32**, 2088-2094.

- [265] Wu, E. H. C. e Law, R. J. B. (2010). Data mining for hotel occupancy rate: An independent component analysis approach. *Journal of Travel and Tourism Marketing*, **27**(4), 426-438.
- [266] Yang, H. H. e Amari, S.-I. (1997). Adaptive on-line learning algorithms for blind separation - maximum entropy and minimum mutual information. *Neural Computation*, **9**(7), 1457-1482.
- [267] Yao, F., Coquery, J. e Cao, K-A. L. (2012). Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. *BMC Bioinformatics*, **13**:24.
- [268] Yeredor, A. (2000). Blind separation of gaussian sources via second-order statistics with asymptotically optimal weighting. *IEEE Signal Processing Letters*, **7**(7), 197-200.
- [269] Yiou, P., Baert, E. e Loutre, M. F. (1996). Spectral analysis of climate data. *Surveys in Geophysics*, **17**, 619-663.
- [270] Yiou, P., Sornette, D. e Ghil, M. (2000). Data-adaptive wavelets and multi-scale singular-spectrum analysis. *Physica D*, **142**, 254-290.
- [271] Youssef, T., Youssef, A. M., LaConte, S. M., Hu, X. P. e Kadah, Y. M. (2003). Robust ordering of independent components in functional magnetic resonance imaging time series data using canonical correlation analysis. *Proc. SPIE*, **2**, 5031-5037.
- [272] Ypma, A. e Pajunen, P. (1999). Rotating machine vibration analysis with second-order independent component analysis. *Proceedings of the 1th International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, 37-42.
- [273] Yu, Z. (2003). Late Quaternary dynamics of tundra and forest vegetation in the southern Niagara Escarpment, Canada. *New Phytologist*, **157**, 365-390.
- [274] Yue, H. e Tomoyasu, M. (2004). Weighted principal component analysis and its applications to improve FDC performance. *43rd IEEE Conference on Decision and Control*, 4262-4267.

- [275] Zhang, H., Guo, C., Shi, Z. e Feng, E. (2008). A new constrained fixed-point algorithm for ordering independent components. *Journal of Computational and Applied Mathematics*, **220**, 548-558.
- [276] Zhang, M., Tong, P., Wang, W., Geng, J. e Du, Y. (2011). Proteomic profile analysis and biomarker discovery from mass spectra using independent component analysis combined with uncorrelated linear discriminant analysis. *Chemometrics and Intelligent Laboratory Systems*, **105**, 207-214.
- [277] Zhao, Y., Yu, Z., Chen, F., Ito, E. e Zhao, C. (2007). Holocene vegetation and climate history at Hurleg Lake in the Qaidam Basin, northwest China. *Review of Palaeobotany and Palynology*, **145**, 275-288.