

# Análise da Variabilidade da Frequência Cardíaca em Indivíduos Saudáveis e Doentes

Por  
Cristina Monteiro Pinto

**Orientador:** Doutora Argentina Maria Soeima Leite

**Co-orientador:** Doutor Eduardo José Solteiro Pires

Dissertação submetida à  
UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO  
para obtenção do grau de  
MESTRE  
em Engenharia Biomédica, de acordo com o disposto no  
DR – I série–A, Decreto-Lei n.º 115/2013 de 7 de agosto e no  
Regulamento de Estudos Conducente ao Grau de Mestre da UTAD  
DR, 2.ª série – n.º 133 de 13 de julho de 2016



# Análise da Variabilidade da Frequência Cardíaca em Indivíduos Saudáveis e Doentes

Por  
Cristina Monteiro Pinto

**Orientador:** Doutora Argentina Maria Soeima Leite

**Co-orientador:** Doutor Eduardo José Solteiro Pires

Dissertação submetida à  
UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO  
para obtenção do grau de  
MESTRE  
em Engenharia Biomédica, de acordo com o disposto no  
DR – I série–A, Decreto-Lei n.º 115/2013 de 7 de agosto e no  
Regulamento de Estudos Conducente ao Grau de Mestre da UTAD  
DR, 2.ª série – n.º 133 de 13 de julho de 2016



*Orientação Científica :*

**Doutora Argentina Maria Soeima Leite**

Professora Auxiliar do  
Departamento de Matemática,  
da Escola de Ciências e Tecnologia  
Universidade de Trás-os-Montes e Alto Douro

**Doutor Eduardo José Solteiro Pires**

Professor Auxiliar com agregação do  
Departamento de Engenharias,  
da Escola de Ciências e Tecnologia  
Universidade de Trás-os-Montes e Alto Douro



*"Pessoalmente, sou apenas uma criança a brincar à beira-mar (...) enquanto diante de mim vastos oceanos de verdade permanecem por descobrir."*

*Isaac Newton*

*Dedico, aos meus pais,*

*Manuel e Júlia.*



UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO

Mestrado em Engenharia Biomédica

Os membros do Júri recomendam à Universidade de Trás-os-Montes e Alto Douro a aceitação da dissertação intitulada “ **Análise da Variabilidade da Frequência Cardíaca em Indivíduos Saudáveis e Doentes** ” realizada por **Cristina Monteiro Pinto** para satisfação parcial dos requisitos do grau de **Mestre**.

Dezembro 2018

Presidente: **Doutor Luís José Calçada Torres Pereira,**  
Professor Auxiliar da Escola de Ciências e Tecnologia da  
Universidade de Trás-os-Montes e Alto Douro

Vogais do Júri: **Doutor Miguel Fernando Paiva Velhote Correia,**  
Professor Auxiliar da Faculdade de Engenharia da Universidade  
do Porto

**Doutora Argentina Maria Soeima Leite,**  
Professora Auxiliar da Escola de Ciências e Tecnologia da  
Universidade de Trás-os-Montes e Alto Douro



# Análise da Variabilidade da Frequência Cardíaca em Indivíduos Saudáveis e Doentes

*Cristina Monteiro Pinto*

Submetido na Universidade de Trás-os-Montes e Alto Douro  
para o preenchimento dos requisitos parciais para obtenção do grau de  
Mestre em Engenharia Biomédica

**Resumo** — O estudo da Variabilidade da Frequência Cardíaca (VFC) tem vindo a revelar-se um indicador não invasivo na deteção de patologias cardíacas. Nesta dissertação, a VFC é caracterizada por métodos lineares e métodos não lineares. Os métodos lineares, no domínio da frequência, incluem a modelação autoregressiva, que obtém as componentes espectrais de termo curto do sinal. Devido à complexidade da VFC e à sua dinâmica não linear foram, também, aplicados métodos não lineares, como os modelos autoregressivos de médias móveis integrados fracionariamente com erros autoregressivos condicionalmente heteroscedásticos generalizados, a *detrended fluctuation analysis* e a entropia aproximada. O estudo experimental apresentado nesta dissertação envolve a análise de duas bases de dados independentes através da aplicação dos métodos indicados. A base de dados fornecida pela Universidade de Zaragoza, Espanha, contém os intervalos RR, de 4 árbitros de futebol, em quatro jogos distintos. O tempo, em minutos, dos momentos decisivos, dos momentos menos decisivos mas importantes e dos momentos não decisivos, durante os jogos, são identificados e relevantes para o estudo. A base de dados, disponibilizada pelo MIT/BIH, da Physionet, pertence a 83 registos de Holter, da VFC, relativos a indivíduos saudáveis e a indivíduos com insuficiência cardíaca congestiva. Nesta dissertação obtiveram-se os parâmetros, que caracterizam os níveis de stresse a que um indivíduo está sujeito, assim como possibilitam a discriminação, entre um grupo de indivíduos saudáveis e um grupo de indivíduos doentes. Os parâmetros estatisticamente significativos, dos dois grupos de pacientes são normalizados e utilizados na análise das componentes principais, de forma a resumir a informação presente neste conjunto de parâmetros. Com a finalidade de classificar os grupos de dados, aplica-se a análise discriminante às componentes principais e a todos os parâmetros

normalizados, considerando os classificadores: linear, quadrático e *K-nearest neighbors*.

**Palavras chave:** Variabilidade da frequência cardíaca, registros de Holter, métodos lineares, métodos não lineares, classificação.

# Analysis of Heart Rate Variability in healthy and unhealthy individuals

*Cristina Monteiro Pinto*

Submitted to the University of Trás-os-Montes and Alto Douro  
in partial fulfillment of the requirements for the degree of  
Master of Biomedical Engineering

**Abstract** — The study of Heart Rate Variability (HRV) has becoming a non invasive diagnostic tool in the diagnosis of cardiac pathologies. On this thesis, the HRV is characterized by linear and nonlinear methods. The linear methods, in the frequency domain, include the autoregressive model, which represents the signal short term spectral features. Due to the HRV complexity and nonlinear dynamics, some nonlinear methods are also applied, such as, fractionally integrated autoregressive moving average models with generalized autoregressive conditionally heteroscedastic innovations, the detrended fluctuation analysis and the approximate entropy methods. The experimental study presented in this thesis analyses two independent databases using the cited methods. The first database was provided by the University of Zaragoza in Spain, and it contains the RR Intervals of 4 referees in four different football matches, respectively. The time in minutes during the games, of the decisive, less decisive but important and nondecisive moments, are identified due to their importance for the study. The second database, downloaded from *Physionet* (MIT/BIH), fits 83 Holter registers, of HRV, from healthy individuals and individuals with congestive heart failure. On this master thesis, several parameters were obtained to define the stress levels that each individual underwent in order to set apart between a group of healthy individuals and a group of ill individuals. The parameters shown as statistically relevant from the healthy/unhealthy sets are normalized and used in the principal component analysis to sum up the data present in those parameters. The discriminant analysis is applied on the principal components and all the normalized parameters, with the intent to classify the data set, with the following classifiers: linear, quadratic and K-nearest neighbors.

**Keywords:** Heart rate variability, records of Holter, linear methods, nonlinear methods, classification.

# Agradecimentos

Gostaria de expressar o meu contentamento e agradecimento a todos os cooperantes que tornaram possível o sucesso deste estudo. Um especial obrigado aos meus orientadores, Professora Doutora Argentina Leite e Professor Doutor Eduardo Solteiro Pires pela dedicação, pelo incentivo, pelo apoio e pela partilha, ao longo do trabalho.

A oportunidade de estagiar no exterior, revelou-se uma mais valia ao exigir uma partilha de conhecimentos e ao possibilitar um enriquecimento da dissertação. Por esses motivos, agradeço a todas as entidades envolvidas na instituição acolhedora e na instituição de envio. Aos meus orientadores da Universidade de Zaragoza, Professora Doutora Raquel Bailón e Doutor David Hernando, pelo apoio, disponibilidade e compreensão, à administradora Arantxa Trigo, por ser o pilar e a representação humana da alegria, ao longo dos dois meses passados em Zaragoza e aos colegas do núcleo de estágio, pelo companheirismo, pela simpatia e por tornarem o ambiente de trabalho tão leve e acolhedor.

É difícil expressar por palavras a gratidão que sinto, mas da forma mais fiel possível, gostaria de agradecer aos meus pais e ao meu irmão, eles que são o meu porto seguro, por sempre me encorajarem a ultrapassar barreiras, por nunca desistirem de mim, pela educação e pelo amor partilhado.

Ao meu namorado, pelo ambiente, pela inspiração e pelo amor. Por ser a personificação do estímulo e da paciência nas horas de maior frustração, por ser o meu maior admirador nas horas das conquistas e por me incentivar sempre *a alcançar as*

*estrelas.*

Quero agradecer à minha família, os meus impulsionadores, pelo acompanhamento durante todo o percurso de trabalho e pelo carinho. Em especial, aos meus padrinhos pelas diretrizes fornecidas e pelo incansável apoio.

E por último, mas nem por isso menos importante, aos meus amigos que sempre demonstraram que não são apenas os laços de sangue que formam uma família. Em especial, ao Cláudio Sousa, Jéssica Carina, Lucas Pereira, Natacha Pereira, Nuno Correia, Samuel Pereira e Sandra Dias, que sempre permaneceram fiéis e incansáveis em todas as horas e à Alexandra Marta, ao Pedro Sampaio e à Vera Ribeiro, que aceitaram comigo o desafio de rumar a um país desconhecido, com o intuito de desenvolver a nossa autonomia, conhecimento e experiência na área e pela partilha de angústias, alegrias e resiliência.

Muito obrigada!

UTAD, Vila Real, 30 de outubro de 2018

Cristina Monteiro Pinto

# Índice geral

<b>Resumo</b>	<b>xi</b>
<i>Abstract</i>	<b>xiii</b>
<b>Agradecimentos</b>	<b>xv</b>
<b>Índice de tabelas</b>	<b>xix</b>
<b>Índice de figuras</b>	<b>xxi</b>
<b>Abreviaturas, Notação e Unidades</b>	<b>xxv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação e objetivos . . . . .	2
1.2 Variabilidade da frequência cardíaca . . . . .	2
1.3 Aplicabilidade clínica da VFC . . . . .	5
1.4 Estrutura da dissertação . . . . .	8
<b>2 Métodos lineares na VFC</b>	<b>11</b>
2.1 Métodos no domínio do tempo . . . . .	12
2.2 Métodos no domínio da frequência . . . . .	14
2.3 Aplicação na VFC . . . . .	21
2.4 Comentários finais . . . . .	24
<b>3 Métodos não lineares na VFC</b>	<b>25</b>
3.1 Memória longa e a heteroscedasticidade na VFC . . . . .	26
3.2 Modelação ARFIMA-GARCH . . . . .	27
3.2.1 Ruído branco fracionário . . . . .	28
3.2.2 Modelos ARFIMA . . . . .	31
3.2.3 Modelos GARCH . . . . .	33
3.2.4 Modelos ARFIMA-GARCH . . . . .	34

3.2.5	Estimação de processos ARFIMA-GARCH . . . . .	36
3.2.6	Aplicação em dados de Holter . . . . .	38
3.3	Método <i>Detrended Fluctuation Analysis</i> . . . . .	41
3.3.1	Introdução ao método . . . . .	41
3.3.2	Expoentes de escala na VFC . . . . .	43
3.3.3	Análise de registos de Holter . . . . .	44
3.4	Entropia . . . . .	46
3.4.1	Entropia aproximada . . . . .	47
3.4.2	Aplicação em registos de Holter . . . . .	49
3.5	Comentários finais . . . . .	51
<b>4</b>	<b>Análise experimental</b>	<b>53</b>
4.1	Introdução . . . . .	54
4.2	Análise multivariada e classificação . . . . .	56
4.3	Resultados e discussão . . . . .	58
4.3.1	Base de dados dos Árbitros de Futebol . . . . .	59
4.3.2	Base de dados de Saudáveis/Doentes . . . . .	65
4.4	Conclusões . . . . .	81
<b>5</b>	<b>Considerações finais e trabalho futuro</b>	<b>85</b>
	<b>Referências bibliográficas</b>	<b>89</b>
<b>A</b>	<b>Anexos</b>	<b>101</b>

# Índice de tabelas

2.1	Descrição dos índices da VFC, no domínio do tempo: estatísticos e geométricos. . . . .	13
2.2	Índices da VFC, no domínio da frequência. . . . .	14
3.1	Expoentes de escala usando o método DFA. . . . .	43
4.1	Parâmetros obtidos pela aplicação dos métodos estudados, na análise experimental dos registos da VFC. . . . .	55
4.2	Medidas de desempenho do modelo de classificação. . . . .	58
4.3	Matriz de confusão. . . . .	58
4.4	Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N1. . . . .	62
4.5	Estimativa global (média $\pm$ desvio padrão) dos parâmetros obtidos pela aplicação dos métodos, relativamente aos grupos de indivíduos saudáveis e indivíduos doentes. Utilização do teste Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos para os dois grupos. . . . .	73
4.6	Desempenho de 3 classificadores, para todos os parâmetros estatisticamente significativos normalizados e para as componentes principais, através das medidas de <i>accuracy</i> , sensibilidade, especificidade e do método da validação cruzada, para $k = 10$ e $k = \text{LOO}$ . . . . .	77

4.7	Estimativa global (média $\pm$ desvio padrão) dos parâmetros obtidos através da aplicação dos métodos, referentes aos grupos de indivíduos saudáveis adultos e indivíduos saudáveis idosos. Utilização do teste estatístico Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos, para os dois grupos. . . . .	78
4.8	Estimativa global (média $\pm$ desvio padrão) dos parâmetros obtidos pela aplicação dos métodos, referentes aos grupos de indivíduos saudáveis homens e indivíduos saudáveis mulheres. Utilização do teste estatístico Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos, para os dois grupos. . . . .	80
A.1	Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N2. . . . .	102
A.2	Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N3. . . . .	103
A.3	Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N4. . . . .	103

# Índice de figuras

1.1	Sinal ECG: respetivas ondas eletrocardiográficas (onda P, complexo QRS, onda T) e intervalo RR. . . . .	3
1.2	(a) Tacograma de um indivíduo saudável, registo longo e não estacionário composto por 101160 batimentos, com início às 9h20m e (b) o respetivo tacograma, mas decomposto em faixas de 10 mil batimentos. . . . .	4
2.1	(a) e (b) Tacograma com 1024 batimentos de um indivíduo saudável, durante o período do dia e durante o período da noite, respetivamente. Em (c) e (d) espetros não paramétricos do período do dia e da noite, com marcação das baixas e altas frequências e em (e) e (f) espetros paramétricos do período do dia e do período da noite, com marcação das baixas e altas frequências. . . . .	22
2.2	(a) Tacograma de um indivíduo saudável com duração aproximada de 24 horas; (b)-(d) seguimento ao longo do tempo dos parâmetros no domínio do tempo; (e)-(f) seguimento das componentes espectrais, usando a modelação AR. . . . .	23
3.1	(a) Segmento curto com 1024 batimentos, do registo longo, de um indivíduo saudável, (c) resíduos ARFIMA e (e) resíduos ARFIMA <sup>2</sup> . Em (b), (d) e (f) estão representadas as correspondentes ACF. . . . .	27

3.2	(a) Registo curto da VFC, com 1024 batimentos, (b) Registo RR curto, Sem Memória Longa (Sinal RR SML) e (c) Representação gráfica do espectro de AR, do Sinal RR SML e do modelo não paramétrico, periodograma, do Sinal RR SML. . . . .	39
3.3	Tacograma de um indivíduo saudável, ao longo de 24 horas, e respetivos parâmetros ARFIMA-GARCH. . . . .	40
3.4	Tacograma de um indivíduo saudável, referente a dois segmentos curtos: (a) segmento com 1024 batimentos, que estima os (b) expoentes de escala $\alpha_{LF}$ e $\alpha_{HF}$ e (c) segmento curto com 4096 batimentos, que estima o (d) expoente de escala $\alpha$ . . . . .	44
3.5	(a) Tacograma de um registo de Holter de 24 horas, pertencente a um indivíduo saudável. Representação dos parâmetros obtidos pela segmentação combinada com o método DFA: (b) $\alpha_{LF}$ , (c) $\alpha_{HF}$ e (d) $\alpha$ para o Sinal RR e (e) $\alpha_{LF}$ e (f) $\alpha_{HF}$ para o Sinal RR sem componente de memória longa (Sinal RR SML). . . . .	45
3.6	Tacograma de um indivíduo saudável (segmentos curtos com 1024 batimentos), durante (a) período diurno e (b) período noturno, com indicação dos correspondentes valores de entropia aproximada. . . . .	50
3.7	Tacograma de um registo de Holter, de um indivíduo saudável e descrição detalhada do parâmetro $ApEn$ , ao longo do tempo. . . . .	51
4.1	Registo típico de um árbitro de futebol N1. A vermelho representam-se os momentos decisivos (MD), a amarelo os momentos menos decisivos mas importantes (MDI) e a verde os momentos não decisivos (MND). . . . .	59
4.2	Momento decisivo do Registo N1, aos 20 minutos. Segmento curto de cinco minutos para: (a) o Sinal RR, (c) o Sinal RR SML, (e) os Resíduos ARFIMA( $p,d,0$ ) e (g) os Resíduos ARFIMA <sup>2</sup> ( $p,d,0$ ). Função de autocorrelação para: (b) o Sinal RR, (d) o Sinal RR SML, (f) os Resíduos ARFIMA( $p,d,0$ ) e (h) os Resíduos ARFIMA <sup>2</sup> ( $p,d,0$ ). (i) Espectro dos dados e (j) espectro paramétrico e não paramétrico. . . . .	61

4.3	Resultados dos parâmetros estudados, para todas as amostras dos quatro árbitros. A representação simbólica está relacionada com o respetivo árbitro: Círculo - Registo 1, Quadrado - Registo 2, Losango - Registo 3 e Triângulo - Registo 4. As cores representam os três momentos (MD, MDI e MND) e seguem o mesmo padrão utilizado no estudo de um registo típico. (a) Componente espectral $LF_{AR}$ , (b) parâmetro de memória longa, $d$ , (c) expoente de escala, do Sinal RR, $\alpha_{LF RR}$ e (d) expoente de escala, do Sinal RR, com ausência de memória longa, $\alpha_{HF SML}$ . . . . .	64
4.4	Representação do Sinal RR e dos parâmetros, no domínio do tempo, dos registos de Holter de (a)-(d) um indivíduo saudável N21 e de (e)-(h) um indivíduo doente CHF1. . . . .	66
4.5	Representação do Sinal RR e obtenção das componentes espectrais, da modelação AR, ao longo do tempo, para (a)-(c) um indivíduo saudável N21 e para (d)-(f) um indivíduo doente CHF1. . . . .	67
4.6	Descrição do Sinal RR e do comportamento dos parâmetros ARFIMA-GARCH, de (a)-(f) um sujeito saudável N21 e de (g)-(i) um paciente doente CHF1, ao longo do tempo. . . . .	68
4.7	Representação do Sinal RR e obtenção dos expoentes escalares, ao aplicar o método DFA, para o referido sinal, de (a)-(d) um paciente saudável N21 e de (e)-(h) um paciente doente CHF1. . . . .	70
4.8	Caraterização do Sinal RR SML e dos expoentes escalares do Sinal RR SML, dos dois registos típicos estudados: (a)-(c) indivíduo saudável N21 e (d)-(f) indivíduo com patologia CHF1. . . . .	71
4.9	Apresentação dos resultados da aplicação da Entropia Aproximada, para um registo de (a) e (b) um sujeito saudável N21 e de (c) e (d) um sujeito doente CHF1, ao longo de 24h. . . . .	72
4.10	Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos Saudáveis (S) e indivíduos Doentes (D). . . . .	74
4.11	Gráfico com 7 componentes principais. . . . .	76

4.12	Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos saudáveis Adultos (A) e indivíduos saudáveis Idosos (I). . . . .	79
4.13	Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos do sexo Masculino (M) e indivíduos do sexo Feminino (F). . . . .	81
A.1	Representação dos registos dos árbitros, ao longo do tempo: (a) Registo N2, (b) Registo N3 e (c) Registo N4. . . . .	102
A.2	Representação gráfica dos parâmetros $HF_{AR}$ , componentes espectrais ARFIMA, parâmetro ARCH, parâmetro GARCH, $\alpha_{HF RR}$ , $\alpha$ , $\alpha_{LF SML}$ e da $ApEn$ , para os quatro jogadores de futebol. A representação simbólica está relacionada com cada jogador: Círculo - Registo 1, Quadrado - Registo 2, Losango - Registo 3 e Triângulo - Registo 4. A vermelho estão representados os MD, a amarelo os MDI e a verde os MND. . . . .	104

# Abreviaturas, Notação e Unidades

## Lista de Abreviaturas

<b>Sigla</b>	<b>Expansão</b>
ACC	<i>Accuracy</i>
ACF	Função de autocorrelação
ACP	Análise de componentes principais
AR	Modelos autoregressivos
ARMA	Modelos autoregressivos de médias móveis
ARFIMA	Modelos autoregressivos de médias móveis integrados fracionariamente
ARFIMA-GARCH	Modelos ARFIMA com erros GARCH
ARIMA	Modelos autoregressivos integrados de médias móveis
ApEn	Entropia aproximada
CHF	Insuficiência Cardíaca Congestiva
DFA	<i>Detrended fluctuation analysis</i>
DFT	Transformada discreta de Fourier
DNA	Ácido desoxirribonucleico
ECG	Eletrocardiograma
fBm	Movimento Browniano fracionário
FC	Frequência cardíaca

<b>Sigla</b>	<b>Expansão</b>
FFT	Transformada rápida de Fourier
GARCH	Modelos autoregressivos condicionalmente heteroscedásticos generalizados
HF	Alta frequência
KNN	<i>K-nearest neighbor</i>
LDA	Análise linear discriminante
LF	Baixa frequência
LWE	Estimador local de Whittle
MA	Modelos de médias móveis
Média RR	Média dos intervalos RR
pNN50	Proporção das diferenças entre intervalos RR sucessivos, maiores que 50 ms
PSD	Densidade espectral de potência
QDA	Análise linear quadrática
RMSSD	Raiz quadrada da média da soma dos quadrados das diferenças entre intervalos RR sucessivos, em ms
SDNN	Desvio padrão de todos intervalos RR
SENS	Sensibilidade
SNA	Sistema nervoso autônomo
SNP	Sistema nervoso parassimpático
SNS	Sistema nervoso simpático
SPEC	Especificidade
TIRR	Interpolação triangular do histograma de intervalos RR
VLF	Frequência muito baixa

## Notação e Unidades

<b>Abreviatura</b>	<b>Significado(s)</b>
$d$	Parâmetro de memória longa
$f(\omega)$	Função de densidade espectral
$f_{AG}(\omega)$	Função de densidade espectral de um processo ARFIMA( $p, d, q$ )-GARCH( $r, s$ )
$f_{ARFIMA}(\omega)$	Função de densidade espectral de um processo ARFIMA( $p, d, q$ )
$f_{ARMA}(\omega)$	Função de densidade espectral de um processo ARMA( $p, q$ )
$f_{SM}(\omega)$	Função de densidade espectral de um processo de termo curto
$h$	<i>lag</i>
$n$	Dimensão do registo
$x(t)$	Processo estocástico discreto no tempo
$\rho(h)$	Função de autocorrelação
$\Gamma(\cdot)$	Função gama
$\alpha$	Expoente escalar de memória longa
$\alpha_{HF}$	Expoente escalar de memória curta
$\alpha_{LF}$	Expoente escalar de memória curta



# Capítulo 1

## Introdução

O elevado risco das doenças cardíacas tem vindo a criar um interesse pela procura de métodos não invasivos, que permitem a sua deteção. Através da introdução da análise de sinais cardiovasculares foi investigada a relação existente entre o Sistema Nervoso Autónomo (SNA) e a mortalidade estimulada por irregularidades cardiovasculares. O aumento da atividade simpática e a diminuição da atividade parassimpática revelaram-se bons indicadores na investigação referida. Esta conjuntura foi encontrada em várias patologias, o que levou ao desenvolvimento de marcadores quantitativos que representam a influência da atividade autónoma no coração, como a Variabilidade da Frequência Cardíaca (VFC).

A análise da VFC é bastante complexa, exibindo características não estacionárias, dependência de termo longo e heteroscedasticidade. Com o auxílio de metodologias a nível interdisciplinar é possível descrever a VFC sob várias circunstâncias, como atividades do quotidiano e patologias. A presente investigação utiliza metodologias, lineares e não lineares, para determinar e caracterizar os registos da VFC, em grupos de indivíduos saudáveis e indivíduos com patologia.

## 1.1 Motivação e objetivos

A análise da VFC tem contribuído em diversas aplicações clínicas, nomeadamente no diagnóstico não invasivo de doenças cardiovasculares (Malik *et al.*, 2004).

A análise da VFC, em registos de Holter, é fundamental como meio auxiliar de diagnóstico e de estratificação de risco de patologias cardíacas. As séries de Holter são tipicamente longas, correspondendo a registos de 24 horas (aproximadamente 100 mil batimentos cardíacos), que permitem identificar diversas patologias, não detetáveis nos registos de rotina do Eletrocardiograma (ECG), nomeadamente as arritmias malignas (ESC/NASPE Task Force, 1996).

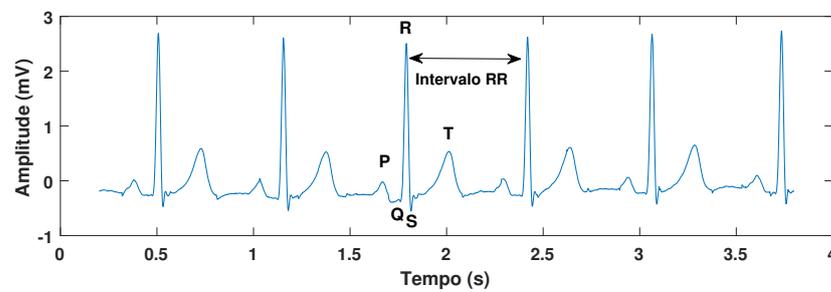
A motivação desta dissertação rege-se pela análise da VFC, através da aplicação de métodos lineares e não lineares, com o intuito de encontrar meios não invasivos de deteção de patologias. O comportamento dos parâmetros obtidos neste estudo revela-se uma ferramenta clínica capaz de discriminar indivíduos saudáveis e indivíduos com patologia e demonstrar a influência da atividade autónoma na VFC.

O objetivo desta dissertação consiste em determinar os parâmetros de caracterização da VFC, que permitem discriminar os dados dos indivíduos saudáveis e dos indivíduos com patologia. Assim, para a análise da VFC, serão estudados métodos tradicionais e métodos baseados em medidas não lineares. As metodologias descritas serão aplicadas em duas bases de dados, uma constituída por registos de atividade, de 2 horas, referentes a quatro árbitros de futebol e outra constituída por registos de Holter, de 24 horas, referentes a indivíduos saudáveis e a indivíduos com insuficiência cardíaca congestiva.

## 1.2 Variabilidade da frequência cardíaca

As variáveis cardiovasculares, como a Frequência Cardíaca (FC), a pressão arterial e a forma das ondas do complexo QRS num ECG (Figura 1.1), apresentam variabilidade entre batimentos.

A VFC é um fenômeno fisiológico medido pela variação do intervalo RR, sendo R o pico máximo do complexo QRS, do sinal ECG e RR o intervalo entre ondas R sucessivas (Cui, 2013). As flutuações na FC demonstram uma sincronia com a respiração: aumentam no decorrer da inspiração e diminuem durante a expiração (Billman, 2011). Na Figura 1.1 estão representadas as deflexões positivas e negativas do ECG, em relação à linha isoeletrica. A onda eletrocardiográfica P, corresponde à despolarização auricular e as ondas do complexo QRS, correspondem à despolarização ventricular, sendo ambas originadas pelas contrações do coração. A onda T corresponde à repolarização ventricular, responsável pelo relaxamento cardíaco.

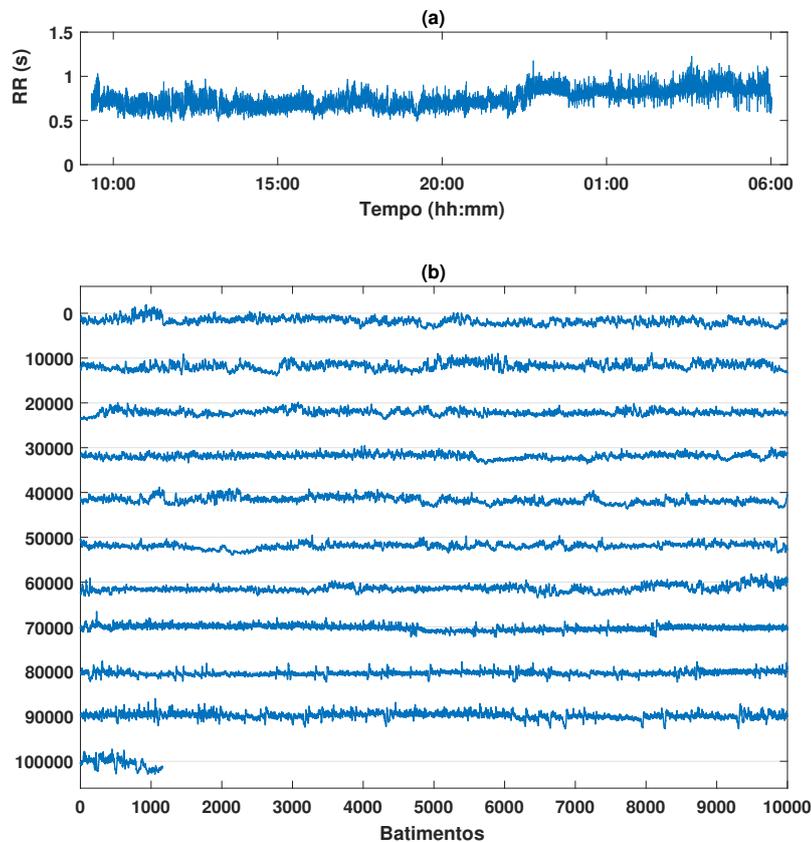


**Figura 1.1** – Sinal ECG: respectivas ondas eletrocardiográficas (onda P, complexo QRS, onda T) e intervalo RR.

As séries longas da VFC ostentam características não estacionárias complexas, com existência de pontos de inflexão, valores em falta e *outliers* (batimentos incorretos) e exibição de dependência de termo longo e heteroscedasticidade. As oscilações dos sinais cardiovasculares refletem a interação entre as perturbações cardiovasculares, promovidas pela variabilidade do ambiente pessoal interno e externo e a resposta dinâmica do sistema de regulação cardiovascular. A análise desta relação constitui um método não invasivo, que avalia a atividade autônoma, sob determinadas condições fisiológicas e viabiliza a associação entre estados patológicos e a integridade do sinal (Acharya *et al.*, 2006).

Na caracterização da VFC usa-se o tacograma (série discreta de intervalos RR sucessivos, do ECG) por ser o sinal cardiovascular mais simples. O tacograma de um

indivíduo saudável permite identificar o respetivo padrão da VFC. Através da Figura 1.2.(a), pode observar-se que, durante o período noturno, os valores do intervalo RR são superiores e exibem uma maior variabilidade, em comparação com o período diurno (Leite *et al.*, 2013). Este tacograma permite ilustrar o tipo de variação global. No entanto, para uma melhor compreensão das suas características não estacionárias, representa-se na Figura 1.2.(b) o mesmo tacograma decomposto em faixas de 10 mil batimentos, onde são notórias as várias diferenças de ritmo.



**Figura 1.2** – (a) Tacograma de um indivíduo saudável, registo longo e não estacionário composto por 101160 batimentos, com início às 9h20m e (b) o respetivo tacograma, mas decomposto em faixas de 10 mil batimentos.

O uso aparentemente fácil da derivação da VFC, converteu-a numa ferramenta clínica adequada, nomeadamente na área da biomédica. O desenvolvimento de

dispositivos que potencializam a medição automática da VFC facilitou a tarefa clínica e investigativa dos cardiologistas. Porém, a interpretação de determinadas medidas da VFC é complexa, e muitas vezes menosprezada, originando conclusões incorretas e extrapolações infundadas (Billman, 2011; Acharya *et al.*, 2006; ESC/NASPE Task Force, 1996).

### 1.3 Aplicabilidade clínica da VFC

O coração é um órgão do sistema circulatório, responsável por grande parte das funções que sustentam a vida. Apresenta células com ritmo próprio, maioritariamente controladas pelo SNA e capazes de gerar potenciais de ação. A função primária do SNA é manter a homeostase, através da regulação das variáveis cardiovasculares e da pressão arterial (Orini *et al.*, 2017).

O SNA ramifica-se em Sistema Nervoso Simpático (SNS) e em Sistema Nervoso Parassimpático (SNP). O SNS inerva o coração, as veias sanguíneas, os rins e a medula da suprarrenal e possui um papel predominante na regulação cardiovascular, em particular, na inervação do nóculo sinusal do coração. Permite que os nervos simpáticos aumentem a inclinação da despolarização diastólica durante o potencial de ação, resultando no aumento da frequência cardíaca. O aumento da atividade simpática, proporciona o aumento do efeito de contração do miocárdio e consequentemente o aumento do volume sistólico. Com a possibilidade de modular a frequência cardíaca e o volume sistólico, os nervos autónomos acionam um mecanismo que ajusta o débito cardíaco rapidamente. Estudos recentes permitiram relacionar a atividade simpática com as oscilações de baixa frequência, na repolarização ventricular. Com efeitos paradoxais, a existência de resposta fisiológica vagal e a descarga simpática culminam, respetivamente, na diminuição e no aumento do débito cardíaco (Rizas *et al.*, 2014).

O nervo vago é responsável pela inervação parassimpática do coração, dos pulmões, do fígado, do esófago, do estômago, do intestino delgado, da metade proximal do

cólon, da vesícula biliar, do pâncreas, dos rins e das porções superiores dos ureteres. Em repouso, a ativação da resposta parassimpática predomina sobre a simpática. A estimulação parassimpática diminui a frequência cardíaca e a força de contração, assim como inibe o SNS e hiperpolariza diretamente o nódulo sinusal. Estudos desenvolvidos recentemente (Orini *et al.*, 2017; Lopes *et al.*, 2013) mostraram que a estimulação direta do nervo vagal diminui a frequência cardíaca, a pressão e a contração ventricular esquerda, além de prolongar a duração do potencial de ação ventricular.

O desenvolvimento de marcadores quantitativos da atividade autônoma cardíaca, como a padronização da VFC permitem compreender as oscilações entre os intervalos RR dos batimentos cardíacos, que representam as mudanças resultantes da atuação autônoma sobre o comportamento da frequência cardíaca, em indivíduos saudáveis e não saudáveis.

A VFC tem sido alvo de vários estudos, em diferentes áreas da saúde, com o propósito de a relacionar com determinadas patologias. De seguida, apresentam-se algumas áreas, onde o estudo da VFC tem demonstrado que o comportamento funcional do corpo humano é refletido no sinal cardíaco.

**Insuficiência Cardíaca Congestiva:** A insuficiência cardíaca é a incapacidade que o coração tem em bombear o sangue, eficientemente, para as células. Uma circulação sanguínea anormal, acarreta danos nos rins que recebem menos sangue e conseqüentemente filtram menos fluído da circulação para a urina. O fluído extra acumula-se em diversos tecidos, originando a congestão. A insuficiência cardíaca congestiva pode ser diagnosticada por técnicas não invasivas, baseadas na análise do intervalo RR (Mahajan *et al.*, 2017; Isler *et al.*, 2015b; Shahbazi e Asl, 2015).

**Enfarte do miocárdio:** O enfarte do miocárdio é uma das doenças que apresenta predominância da atividade simpática, redução do controlo cardíaco parassimpático (Acharya *et al.*, 2006) e diminuição na VFC (ESC/NASPE Task Force, 1996). Como conseqüência, o sistema de controlo simpático diminui a fibrilação ventricular, que em casos extremos acarreta danos severos (Acharya *et al.*, 2006; Lopes *et al.*, 2013).

**Danos cerebrais e depressão:** As alterações cíclicas normais na FC são restringidas na existência de danos cerebrais (Lowensohn *et al.*, 1977) e depressão (Carney *et al.*, 2001). A importância da análise da VFC em diagnósticos psiquiátricos é manifestada pela simplicidade da detecção do desequilíbrio simpático-vagal.

**Arritmia cardíaca:** A introdução dos métodos não lineares na análise da VFC potenciou uma compreensão mais exata do sistema dinâmico e uma consequente detecção e classificação da arritmia cardíaca (Acharya *et al.*, 2006). Os modelos lineares Autoregressivos (AR) permitiram categorizar a arritmia cardíaca, em 6 classes (Ge *et al.*, 2002).

**Diabetes:** Os diabetes podem causar disfunções autônomas severas e autenticar sintomas incapacitantes. Wheeler e Watkins (1973) foram os pioneiros na descrição da variabilidade entre batimentos, durante a respiração profunda, em casos de pacientes com neuropatia diabética. Outros estudos centraram-se na comparação entre os testes de função autônoma cardíaca e os índices da VFC. É usual uma diminuição da atividade cardíaca parassimpática em pacientes diabéticos, antes dos sintomas clínicos da neuropatia se manifestarem (Pfeifer *et al.*, 1982; Singh *et al.*, 2000; Villareal *et al.*, 2002).

**Falhas renais:** Estudos sobre falhas renais têm sido baseados na função autônoma (Ewing e Winney, 1975; Zoccali *et al.*, 1982) e na análise espectral da VFC (Axelrod *et al.*, 1987). A análise da função autônoma revelou danos no SNP e a análise espectral detetou uma redução acentuada na potência espectral da FC, tanto a nível simpático como a nível parassimpático (Zoccali *et al.*, 1982). Lerma *et al.* (2004) mostraram que, comparando indivíduos saudáveis e pacientes com falha crônica renal, a potência de alta frequência é superior nos indivíduos saudáveis e a potência de baixa frequência é superior nos pacientes com falha renal.

**Recém-nascidos:** A asfixia no nascimento induz alterações nas altas frequências (Rosenstock *et al.*, 1999). Segundo Watkins *et al.* (1996), os recém-nascidos, com pouco peso e com uma VFC menor no primeiro dia, apresentam um diagnóstico propenso a hemorragia intraventricular neonatal, sete dias após o parto. Os bebês e

as crianças com morte encefálica apresentam menor variabilidade que um conjunto de controlo e que uma população com síndrome de desconforto respiratório. Os bebés normais sustentam ritmos cardíacos complexos e não lineares, enquanto que os recém-nascidos com morte cerebral são desprovidos do controlo de *feedback* não linear (Rosenstock *et al.*, 1999). Estudos como o de Emese *et al.* (2000) revelaram que os recém-nascidos masculinos apresentam uma menor variação da FC do que os femininos.

**Idade/Género:** A desregulação do SNA potenciada pela idade, género, ansiedade, stresse, obesidade e causas orgânicas contribui para o aparecimento de patologias cardiovasculares. A idade é um dos principais fatores de risco nas doenças cardiovasculares. A VFC reduzida está relacionada com o aumento da morbidade e da mortalidade cardiovascular, em idosos. A análise do sistema cardiovascular de idosos, revela uma mudança significativa na modulação autónoma da FC, preferencialmente na vertente simpática. A dependência entre a idade e a VFC pode ser causada pelas modificações cardiovasculares, do ponto de vista estrutural. Tomando o sexo feminino como exemplo, a VFC é mais ativa em jovens mulheres.

Estudos mostraram a dominância do tom vagal, no controlo autónomo do coração das mulheres e, em contraste, uma preponderância de respostas mediadas pelo SNS, nos homens (Orini *et al.*, 2017). As diferenças encontradas entre jovens mulheres e jovens homens podem ser causadas pelas divergências hormonais, culminando na alteração do balanço simpático-vagal. No envelhecimento, o desaparecimento do fator género, pode ser explicado pela reestruturação hormonal, possivelmente devido à menopausa na mulher e à andropausa no homem (Orini *et al.*, 2017). De modo sumariado em indivíduos saudáveis, a VFC diminui com a idade e tem efeito superior no homem (Bonnemeier *et al.*, 2003; Acharya *et al.*, 2006).

## 1.4 Estrutura da dissertação

Esta dissertação é dividida em duas partes principais: o enquadramento teórico (Capítulo 2 e 3) e a análise experimental (Capítulo 4). O Capítulo 2 sublinha os

princípios fundamentais teóricos associados aos métodos tradicionais, no domínio do tempo e no domínio da frequência. O método paramétrico no domínio da frequência, utilizando modelos AR é enfatizado na caracterização da VFC, ao viabilizar uma descrição adequada dos dados. Com o intuito de obter uma descrição mais simples do registo longo, não estacionário aplica-se a análise sequencial recorrendo à segmentação, decompondo o registo em segmentos curtos, aproximadamente estacionários e com dimensão fixa. Este capítulo descreve ainda a aplicação dos modelos, em registos reais da VFC.

Os métodos tradicionais apenas caracterizam a memória curta do sinal e não têm capacidade de descrever os fenómenos não lineares da VFC, influenciados pela complexidade das interações hemodinâmicas, eletrofisiológicas, humorais, assim como, pelas regulações nervosas centrais e autónomas. As referidas limitações exigem uma abordagem que permita ultrapassar essa conjuntura. No Capítulo 3 referem-se os métodos não lineares. Os modelos Autoregressivos de Médias Móveis Integrados Fracionariamente com erros Autoregressivos Condicionalmente Heteroscedásticos Generalizados (ARFIMA-GARCH) são uma extensão dos modelos AR, incorporando a modelação conjunta de estruturas de correlação de termo curto e de termo longo, assim como a modelação de características associadas à dependência da variabilidade instantânea, nas observações passadas. Permite, também, estimar e remover a memória longa, conduzindo a uma satisfatória descrição dos parâmetros de memória curta e permite estimar a volatilidade. O método não paramétrico, *Detrended Fluctuation Analysis* (DFA), é uma técnica de deteção de correlação de termo curto e de termo longo em dados não estacionários (Leite *et al.*, 2013), que permite descrever a flutuação dos resíduos de regressão. A Entropia Aproximada (*ApEn*) é uma medida independente, utilizada na quantificação da complexidade de um sistema, através de uma série temporal (Kaffashi *et al.*, 2008).

O Capítulo 4 reporta uma síntese dos resultados do estudo empírico, que surgem da aplicação dos modelos concetuais em análise, a dois conjuntos de registos individualizados. Considera-se uma base de dados com os intervalos RR, de quatro árbitros, durante quatro partidas de futebol. Os momentos temporais decisivos e os momentos

temporais menos decisivos mas importantes para o jogo foram recolhidos, permitindo uma análise comparativa dos parâmetros obtidos para esses momentos, relativamente aos parâmetros obtidos para os momentos não decisivos. Considerando todos os árbitros e todos os momentos definidos, formulam-se os resultados globais dos parâmetros mais conclusivos.

Uma segunda base de dados, exhibe os registos de Holter de dois tipos de pacientes: saudáveis e com Insuficiência Cardíaca Congestiva (CHF). Um registo típico de um elemento de cada grupo retrata detalhadamente o comportamento dos parâmetros obtidos pela aplicação dos métodos em estudo. Os resultados globais possibilitam a caracterização dos conjuntos de indivíduos saudáveis e com patologia, durante o período de tempo registado. O teste de Wilcoxon-Mann-Whitney determina quais os parâmetros que apresentam diferenças estatisticamente significativas entre os grupos, a um nível de significância de 5%. Ainda na análise experimental, os parâmetros estatisticamente significativos, dos grupos de indivíduos saudáveis e doentes são normalizados, pelas médias e desvio padrão e utilizados na Análise das Componentes Principais (ACP), com o intuito de resumir a informação presente neste conjunto de parâmetros. Aplica-se a análise discriminante a todos os parâmetros normalizados e ao novo conjunto de parâmetros, considerando três classificadores: linear, quadrático e *K-nearest neighbors*.

Por último, o Capítulo 5 expõe as considerações finais relevantes do estudo realizado e inclui algumas perspetivas de trabalho futuro.

A aplicação dos métodos, o tratamento estatístico dos dados e a classificação foram executados em ambiente Matlab, utilizando as *toolboxes* de Processamento de Sinal, de Econometria e de Estatística.

# Capítulo 2

## Métodos lineares na VFC

Este capítulo destaca os métodos convencionais utilizados na análise da VFC: métodos no domínio do tempo e métodos no domínio da frequência. Relativamente aos primeiros, são determinados os índices estatísticos e os índices geométricos do registo contínuo em análise, que traduzem as flutuações ao longo do tempo dos ciclos cardíacos. O estudo no domínio de frequência está relacionado com a análise da densidade espectral, que decompõe um sinal dependente do tempo em componentes sinusoidais.

A análise espectral permite estudar a distribuição da potência total, em função da frequência, em registos finitos de dados estacionários. Independentemente dos métodos utilizados (métodos paramétricos ou métodos não paramétricos), apenas são obtidas estimativas da Densidade Espectral de Potência (PSD), dos sinais originais.

Neste capítulo apresentam-se os métodos tradicionais aplicados a séries temporais. A Secção 2.1 descreve os conceitos relativos aos métodos no domínio do tempo e respetivos parâmetros. A Secção 2.2 expande aos métodos no domínio da frequência, que englobam a modelação AR e as componentes espectrais associadas à atividade do SNA. Por último, os referidos modelos são aplicados a registos curtos e registos longos da VFC, na Secção 2.3.

## 2.1 Métodos no domínio do tempo

Os métodos no domínio do tempo são computacionalmente simples de aplicar e permitem determinar a FC em qualquer instante de tempo. No entanto, estes métodos, demonstram uma fraca tendência em discriminar as contribuições do SNS e do SNP, na VFC (Acharya *et al.*, 2006).

Num registo ECG calculam-se as medidas no domínio do tempo, a partir das quais se determina a dispersão da duração dos intervalos entre complexos QRS. Os índices no domínio do tempo são determinados e processados matematicamente e podem ser divididos em dois grupos: índices estatísticos e índices geométricos (ESC/NASPE Task Force, 1996; Orini *et al.*, 2017).

### Índices Estatísticos

A duração do registo da FC é diretamente proporcional à complexidade das medidas estatísticas calculadas. Assim, em registos de 24 horas, as medidas estatísticas são mais precisas e também mais complexas. Os principais índices estatísticos obtidos são a média e o desvio padrão do período cardíaco (ESC/NASPE Task Force, 1996).

Os índices estatísticas, indicados na Tabela 2.1, podem ser obtidos da análise do ECG ou de registos da VFC. Estes índices podem ser divididos em duas classes. A primeira representa as medidas obtidas diretamente do intervalo RR ou da frequência cardíaca instantânea. Destacam-se o valor médio dos intervalos RR e o desvio padrão dos intervalos RR (SDRR). O SDRR reflete as componentes cíclicas responsáveis pela variabilidade. A segunda classe engloba as medidas derivadas das diferenças entre intervalos RR. As mais usadas são: o RMSDD, o NN50 e o pNN50, definidos na Tabela 2.1. Estas medidas são altamente correlacionadas (ESC/NASPE Task Force, 1996).

### Índices geométricos

São obtidos essencialmente da função de densidade de probabilidade do intervalo RR. A série de intervalos RR pode ser convertida num padrão geométrico. Têm

**Tabela 2.1** – Descrição dos índices da VFC, no domínio do tempo: estatísticos e geométricos (ESC/NASPE Task Force, 1996; Orini *et al.*, 2017).

Variáveis (unidades)	Descrição
<b>Índices Estatísticos</b>	
Média RR (ms)	Valor médio dos intervalos RR
SDRR (ms)	Desvio padrão dos intervalos RR
SDARR (ms)	Desvio padrão das médias dos intervalos RR, em todos os segmentos de 5 minutos do registo longo
RMSSD (ms)	Raiz quadrada da média da soma dos quadrados das diferenças, entre intervalos RR sucessivos
pNN50 (%)	Proporção das diferenças entre intervalos RR sucessivos, superiores a 50 ms
<b>Índices Geométricos</b>	
Índice triangular	Número total de intervalos RR dividido pela altura do histograma de todos os intervalos RR, medidos numa escala discreta
TIRR (ms)	Interpolação triangular do histograma de intervalos RR

sido utilizadas diversas aproximações, com o objetivo de derivar as medidas da VFC, através deste processo (ESC/NASPE Task Force, 1996). Este método é independente da análise qualitativa da série de intervalos RR e exige um determinado número de intervalos para construir uma abordagem geométrica. Registos com mais de 20 minutos fornecem uma análise adequada.

Existem três abordagens gerais dos métodos geométricos:

- A medida básica do padrão geométrico é convertida numa medida da VFC;
- Uma forma matemática definida interpola o padrão geométrico, com a finalidade de posterior utilização dos seus parâmetros;
- A forma geométrica é classificada em várias categorias baseadas em padrões relativos a diferentes classes da VFC.

As variáveis geométricas estão definidas na Tabela 2.1. O *índice triangular* apresenta elevada correlação com o SD e é insensível a artefactos e a batimentos ectópicos. Não exige um pré-processamento dos sinais. Em comum o TIRR e o *índice triangular* expressam a VFC, medida durante 24 horas e são mais influenciados pelas baixas do que pelas altas frequências (ESC/NASPE Task Force, 1996).

## 2.2 Métodos no domínio da frequência

A análise da VFC de termo curto, no domínio da frequência, baseada na função da densidade espectral do tacograma, é um método em constante evolução na área da medicina (Spangl e Dutter, 2012).

O objetivo dos métodos no domínio da frequência é representar os sinais da VFC em função da PSD, através de métodos paramétricos e não paramétricos. Entre as componentes espectrais, estão as de muito baixa frequência (VLF), as de baixa frequência (LF) e as de alta frequência (HF), apresentadas na Tabela 2.2.

**Tabela 2.2** – Índices da VFC, no domínio da frequência (Orini *et al.*, 2017).

Variáveis (unidades)	Descrição
VLF (ms <sup>2</sup> )	Potência de muito baixa frequência $f \leq 0.04$ Hz
LF (ms <sup>2</sup> )	Potência de baixa frequência $f \in (0.04 - 0.15)$ Hz
LF (u.n.)	Potência LF em unidades normalizadas: $\left(\frac{LF}{Potência_{Total-VLF}}\right)$
HF (ms <sup>2</sup> )	Potência de alta frequência $f \in (0.15 - 0.4)$ Hz
HF (u.n.)	Potência HF em unidades normalizadas: $\left(\frac{HF}{Potência_{Total-VLF}}\right)$
LF/HF	Razão LF (ms <sup>2</sup> ) / HF (ms <sup>2</sup> )

As componentes espectrais representam determinadas funções no sistema cardiológico. A componente HF corresponde à modulação respiratória e é um indicador da atuação do nervo vago no coração, significando que varia com as modulações do sistema nervoso parassimpático. A componente LF é influenciada pelo sistema simpático e pelo parassimpático. A razão LF/HF pode ser considerada um marcador do balanço simpático-vagal (Acharya *et al.*, 2006; Evrengül *et al.*, 2005). É evidente que a interpretação fisiológica das componentes é mais problemática, quando se transfere de registos de período curto para registos de período longo (aproximadamente 24 horas). Recentemente, com o intuito de obter uma medição mais correta das componentes, sugeriu-se que a análise espectral da VFC seja restrita a registos de termo curto, sob determinadas condições (Orini *et al.*, 2017).

### Estimação Espetral

Se a função de autocorrelação,  $\rho(h)$ , de um processo estacionário  $x(t)_{t \in \mathbb{Z}}$  satisfaz a condição (Shumway, 2011):

$$\sum_{h=-\infty}^{\infty} |\rho(h)| < \infty, \quad (2.1)$$

então a **Função de autocorrelação**, pode ser representada por:

$$\rho(h) = \int_{-\pi}^{\pi} e^{i\omega h} f(\omega) \, d\omega, \quad h = 0, \pm 1, \pm 2, \dots, \quad (2.2)$$

e a **Função de densidade espectral** pode ser definida como sendo a transformada de Fourier da função de autocorrelação:

$$f(\omega) = \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\omega h}, \quad -\pi \leq \omega \leq \pi. \quad (2.3)$$

Quando a condição (2.1) é satisfeita, a função de autocorrelação e a função de densidade espectral contêm a mesma informação. Porém, enquanto a primeira expressa a informação em termos de *lags* (atrasos. Períodos de tempo entre eventos), a segunda apresenta a informação em termos de ciclos.

A análise espectral pode ser executada com base em aproximações paramétricas ou aproximações não paramétricas.

### Estimação espectral não paramétrica

A caracterização dos dados no domínio da frequência e o cálculo da PSD podem ser deduzidos por métodos não paramétricos, como a Transformada Rápida de Fourier (FFT) e o periodograma.

### Transformada de Fourier

A transformada de Fourier representa, no domínio da frequência, a conversão de uma função matemática no domínio do tempo. A frequência de amostragem da

transformada de Fourier é designada de Transformada Discreta de Fourier (DFT). Sendo  $x(1), \dots, x(n)$  um dado sinal, define-se a DFT por (Shumway, 2011):

$$d(\omega_j) = n^{-\frac{1}{2}} \sum_{t=1}^n x(t) e^{-2\pi i \omega_j t},$$

para  $j = 0, 1, \dots, n-1$ , onde  $\omega_j = \frac{j}{n}$  são as frequências de Fourier ou frequências fundamentais. O estudo de Cooley e Tukey (1965) mostrou que se  $n$  for uma potência de base 2, a DFT pode ser obtida pela FFT (Cui, 2013). Por outras palavras, determinados sinais podem ser calculados como DFT em amostras de frequência, usando o algoritmo FFT. Seguidamente, apresenta-se o periodograma que pode ser definido como o quadrado do módulo da DFT.

### Periodograma

O periodograma é uma técnica amplamente utilizada na análise de séries temporais. Apresenta múltiplas vantagens, tais como, a deteção de periodicidades ocultas e a estimação dos parâmetros sinusoidais (amplitude e frequência). A partir dos máximos locais do periodograma obtêm-se estimações mais exatas das frequências (Li, 2010).

Sendo  $x(1), \dots, x(n)$  um dado sinal, define-se por periodograma a expressão (Shumway, 2011)

$$I(\omega_j) = |d(\omega_j)|^2, \quad (2.4)$$

onde  $j = 0, 1, 2, \dots, n-1$ .

Para  $j = 0$ ,  $I(0) = n\bar{x}^2$ , em que  $\bar{x}$  é a média amostral.

Para  $j \neq 0$ , pode-se escrever a DFT da seguinte forma:

$$d(\omega_j) = n^{-\frac{1}{2}} \sum_{t=1}^n [x(t) - \bar{x}] e^{-2\pi i \omega_j t}.$$

Assim para  $j \neq 0$ ,

$$\begin{aligned}
I(\omega_j) &= n^{-1} \sum_{t=1}^n \sum_{s=1}^n [x(t) - \bar{x}][x(s) - \bar{x}] e^{-2\pi i \omega_j (t-s)} \\
&= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} [x(t+|h|) - \bar{x}][x(t) - \bar{x}] e^{-2\pi i \omega_j h}, \\
&= \sum_{h=-(n-1)}^{n-1} \hat{\rho}(h) e^{-2\pi i \omega_j h}
\end{aligned} \tag{2.5}$$

onde  $h = t - s$ .

A Equação (2.5) é um estimador de  $f(\omega)$  (Equação (2.3)), sendo possível considerar o periodograma como um estimador da função de densidade espectral de  $x(t)$ .

### Estimação espectral paramétrica

Os métodos paramétricos permitem controlar a complexidade do modelo, ao limitar o número de parâmetros utilizados na estimação espectral (Kozat e Singer, 2014).

### Modelos ARMA

O Modelo Autoregressivo de Médias Móveis, de ordem  $p$  e  $q$ ,  $\text{ARMA}(p, q)$ , baseia-se na hipótese de que um processo estocástico estacionário, em tempo discreto,  $x(t)_{t \in \mathbb{Z}}$ , e pode ser descrito segundo a resposta de um sistema linear e invariante no tempo ao ruído branco,  $\varepsilon(t)_{t \in \mathbb{Z}}$ , com média zero e variância  $\sigma_\varepsilon^2$ ,

$$x(t) = \sum_{i=1}^p \phi_i x(t-i) + \sum_{j=1}^q \theta_j \varepsilon(t-j) + \varepsilon(t), \tag{2.6}$$

onde  $\phi_1, \dots, \phi_p$  e  $\theta_1, \dots, \theta_q$  são parâmetros ARMA.

Usando o operador de atraso  $B$ ,  $Bx(t) = x(t-1)$ , o modelo  $\text{ARMA}(p, q)$  pode ser escrito do seguinte modo (Klepsch *et al.*, 2017; Shumway, 2011):

$$\phi(B)x(t) = \theta(B)\varepsilon(t),$$

em que  $\phi(B) = 1 - \sum_{k=1}^p \phi_k B^k$  e  $\theta(B) = 1 + \sum_{k=1}^q \theta_k B^k$  são polinómios.

A Função de densidade espectral do processo ARMA( $p, q$ ) é dada por (Shumway, 2011):

$$f_{\text{ARMA}}(\omega) = \sigma_\varepsilon^2 \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \quad -\pi \leq \omega \leq \pi, \quad (2.7)$$

onde  $\sigma_\varepsilon^2$  é a função de densidade espectral de  $\varepsilon(t)$ .

Os Modelos ARMA apresentam os seguintes casos específicos:

1. Modelo Autoregressivo de ordem  $p$ , AR( $p$ ), onde se assume  $\theta_k = 0$ , para  $k = 1, \dots, q$ ;
2. Modelo de Média Móvel de ordem  $q$ , MA( $q$ ), onde se assume  $\phi_k = 0$ , para  $k = 1, \dots, p$ .

## Modelos AR

A notação AR( $p$ ) refere-se ao Modelo Autoregressivo de ordem  $p$  e é dado por (Liu *et al.*, 2017; Klepsch *et al.*, 2017; Shumway, 2011):

$$x(t) = \sum_{i=1}^p \phi_i x(t-i) + \varepsilon(t), \quad (2.8)$$

ou

$$\phi(B)x(t) = \varepsilon(t),$$

onde  $\phi_1, \dots, \phi_p$  são parâmetros AR e  $\varepsilon(t)$  é um processo de ruído branco com média zero e variância  $\sigma_\varepsilon^2$ . São necessárias limitações nos valores dos parâmetros para que o modelo AR permaneça estacionário. Neste caso,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  é um polinómio, de modo que as raízes de  $\phi(B) = 0$  estejam fora do círculo unitário.

Na estatística e no processamento de sinal, um modelo AR é uma representação de um determinado processo aleatório, ao descrever processos que variam no tempo (Liu *et al.*, 2017).

A **Função de autocorrelação** do processo  $x(t)_{t \in \mathbb{Z}}$ , que satisfaz a Equação (2.8) é dada por (Shumway, 2011):

$$\rho(h) = C_1 B_1^{-|h|} + \dots + C_p B_p^{-|h|}, \quad h \in \mathbb{Z},$$

onde  $B_1, \dots, B_p$  são as raízes do polinómio  $\phi(B)$  e  $C_1, \dots, C_p$  são constantes. É de notar que a função de autocorrelação decai exponencialmente para zero, quando a variável  $h$  aumenta. Esta relação é uma característica típica dos processos de memória curta.

Parzen (1983) introduziu os estimadores espectrais autoregressivos sumariados por:

$$f_{\text{AR}}(\omega) = \frac{\sigma_\varepsilon^2}{|\phi(e^{-i\omega})|^2}, \quad \pi \leq \omega \leq \pi. \quad (2.9)$$

A estimação da função de densidade espectral de  $\text{AR}(p)$ , resume-se à estimação dos parâmetros  $\phi_1, \phi_2, \dots, \phi_p$  e de  $\sigma_\varepsilon^2$ , que podem ser obtidos a partir das equações de Yule-Walker (Mikosch *et al.*, 1995), com resolução através do algoritmo de Durbin-Levinson (Shumway, 2011; Levinson, 1946; Durbin, 1960).

## Modelos MA

A notação  $\text{MA}(q)$  refere-se ao Modelo de Média Móvel, de ordem  $q$ :

$$x(t) = \sum_{j=1}^q \theta_j \varepsilon(t-j) + \varepsilon(t), \quad (2.10)$$

ou

$$x(t) = \theta(B)\varepsilon(t),$$

onde  $\theta_1, \dots, \theta_q$  são parâmetros de MA e  $\varepsilon(t)$  é um processo de ruído branco. O modelo  $\text{MA}(q)$  suaviza a série de tempo, produzindo ciclos e definindo uma tendência, mesmo quando os dados originais são eventos aleatórios independentes, com média fixa. Essa característica diminui a sua utilidade como mecanismo de controlo (Liu *et al.*, 2017).

Neste trabalho, considerou-se a modelação AR, dado ser a metodologia mais usada em registos da VFC. Em relação à ordem  $p$ , foi usado o valor  $p = 12$ , já que é um valor vulgarmente usado na literatura para a caracterização espectral AR, da VFC (ESC/NASPE Task Force, 1996).

### **Análise paramétrica vs. análise não paramétrica**

Uma comparação geral entre as características paramétricas e as não paramétricas permite detetar as limitações de cada análise e encontrar forma de as compensar. Estudos na área indicaram, as seguintes vantagens relativamente à estimação paramétrica: espectros mais suaves, que diferenciam claramente as componentes de frequência e conseqüentemente um cálculo mais assertivo (Stavrinou *et al.*, 2014); aumento da resolução espectral sem comprometer a resolução temporal (Kadado *et al.*, 1994); em condições de baixo ruído, as abordagens paramétricas e não paramétricas são similares, no entanto, em condições de ruído elevado, as paramétricas indicam uma análise mais completa; valores significativos mais fortes (Gordon *et al.*, 2013). As principais desvantagens baseiam-se na possibilidade de existir um desvio entre o sinal modulado e o sinal da VFC original, com aparecimento de ruído que oculta as diferenças das frequências do sinal (Stavrinou *et al.*, 2014) e a incapacidade de rastrear modelos criticamente amortecidos ou descontínuos (Dziedziech *et al.*, 2018).

A análise não paramétrica apresenta as seguintes vantagens: ajuste automático dos parâmetros, de modo a otimizar o seu desempenho; computacionalmente mais eficazes (Gordon *et al.*, 2013); simplicidade do algoritmo utilizado, como é o caso da FFT; velocidade de processamento; utilidade demonstrada quando o conhecimento prévio do sistema é limitado e os dados apresentam boa qualidade. De um ponto de vista de inferioridade, os métodos não paramétricos não são recomendados em casos de excitação aleatória e em sinais curtos corrompidos por ruído (Dziedziech *et al.*, 2018).

Recentemente, têm sido desenvolvidos estudos (Dziedziech *et al.*, 2018) que indicam as vantagens de uma abordagem combinada. A fusão das características paramétricas

e não paramétricas resultam num método melhorado, que engloba as vantagens e compensa as desvantagens sumariadas.

## 2.3 Aplicação na VFC

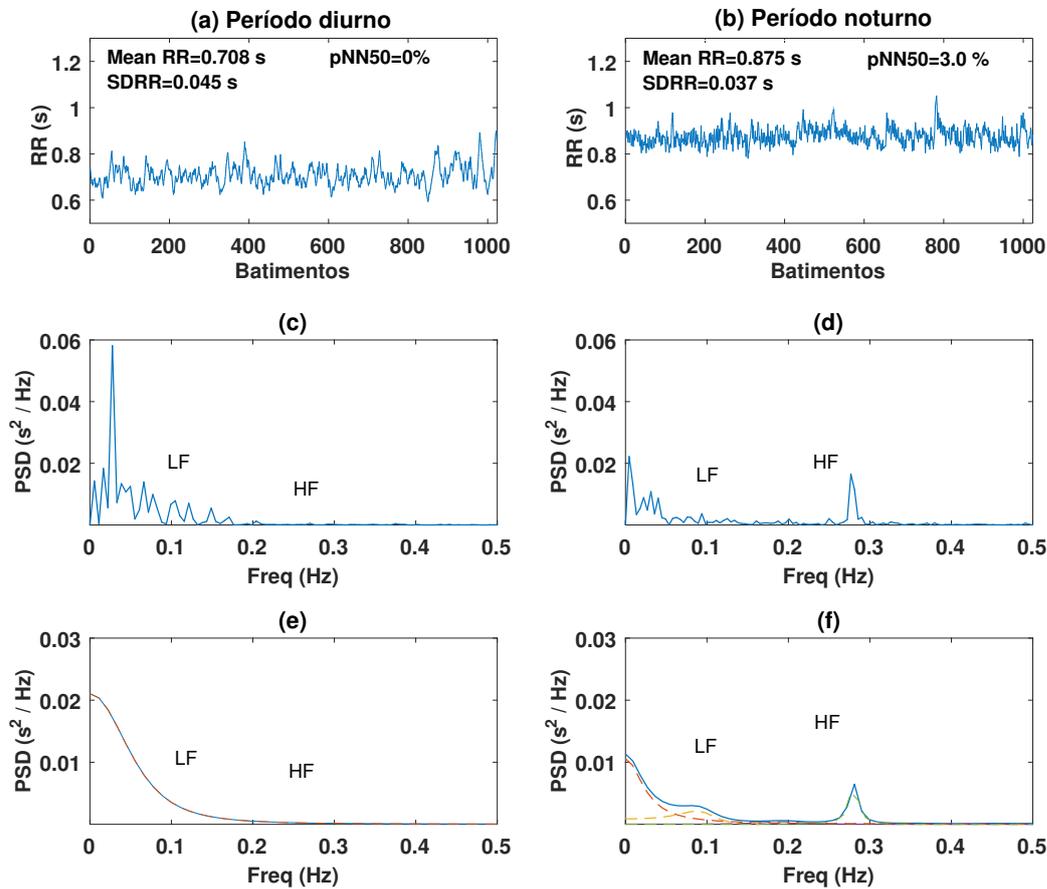
Esta secção ilustra a aplicação dos métodos no domínio do tempo e no domínio da frequência, em registos da VFC. Considera-se um registo longo não estacionário, de um indivíduo saudável, com duração aproximada de 24 horas (com 101160 batimentos) e dois segmentos com 1024 batimentos, aproximadamente estacionários retirados do registo longo, durante os períodos do dia e da noite.

**Análise de registos curtos** Os segmentos curtos com 1024 batimentos, de um indivíduo saudável, no período diurno e no período noturno estão representados nas figura 2.1.(a) e 2.1.(b), assim como os parâmetros no domínio do tempo.

Da interpretação da figura verifica-se que os parâmetros são mais elevados durante o período da noite, quando comparados com os do período diurno. O espectro não paramétrico calculado pelo periodograma é representado nas figura 2.1.(c) e 2.1.(d). O espectro paramétrico de AR é representado nas figura 2.1.(e) e 2.1.(f) e é obtido pela aplicação do modelo de Yule-Walker, com resolução através do algoritmo de Durbin-Levison e com ordem do modelo  $p = 12$ , selecionado pelo critério de Akaike (Akaike, 1974). Nos espectros são ilustradas as componentes mais elevadas, correspondentes às altas e baixas frequências. A componente espectral VLF situa-se na banda de frequência inferior a 0.04 Hz, a LF está entre (0.04 – 0.15) Hz e os valores frequenciais (0.15 – 0.4) Hz correspondem à componente HF. De acordo com a informação recolhida na literatura, a componente HF está associada ao SNP, enquanto a LF é estimulada pela atividade do SNP e do SNS. Na análise do espectro não paramétrico, durante a noite, visualiza-se um aumento significativo das componentes espectrais. Durante o dia a componente HF não se evidencia, como durante a noite.

A análise paramétrica de AR representa uma descrição dos resultados da análise

não paramétrica suavizados e permite a identificação de cada componente espectral, através da respetiva frequência central.

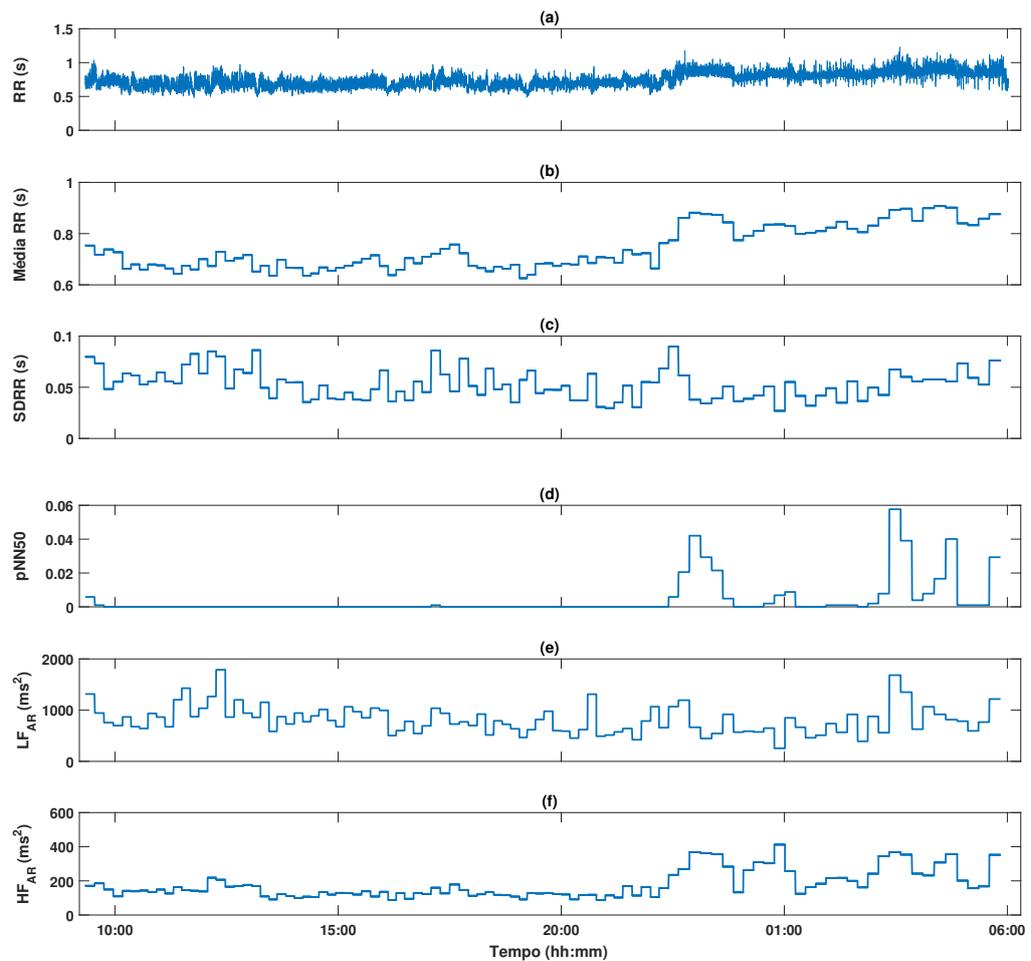


**Figura 2.1** – (a) e (b) Tacograma com 1024 batimentos de um indivíduo saudável, durante o período do dia e durante o período da noite, respetivamente. Em (c) e (d) espectros não paramétricos do período do dia e da noite, com marcação das baixas e altas frequências e em (e) e (f) espectros paramétricos do período do dia e do período da noite, com marcação das baixas e altas frequências.

### Análise de registos longos

Uma forma simplificada de descrever registos longos não estacionários, pode ser obtida a partir da análise sequencial: o registo longo é decomposto em segmentos curtos, aproximadamente estacionários. A cada segmento curto pode ser aplicado um método usual para o estudo da VFC, nomeadamente os métodos no domínio do tempo e os métodos no domínio da frequência.

A Figura 2.2 representa o registo longo de um indivíduo saudável, com início às 9h20 e término às 6h00 do dia seguinte, e a descrição dos parâmetros no domínio do tempo e das componentes espectrais do modelo AR, ao longo do tempo. A média RR e o pNN50 apresentam valores elevados durante o período noturno.



**Figura 2.2** – (a) Tacograma de um indivíduo saudável com duração aproximada de 24 horas; (b)-(d) seguimento ao longo do tempo dos parâmetros no domínio do tempo; (e)-(f) seguimento das componentes espectrais, usando a modelação AR.

A componente espectral HF apresenta variação circadiana, mais elevada durante a noite, verificando-se uma predominância da atividade do SNP. A componente espectral LF não apresenta variações distinguíveis, entre o período diurno e o noturno.

## 2.4 Comentários finais

Os métodos lineares apresentados neste capítulo permitem caracterizar registos longos não estacionários, considerando os índices estatísticos dos métodos no domínio do tempo e considerando a modelação AR. A análise no domínio do tempo é mais simples de aplicar, porém, não permite discriminar a influência do SNA, na VFC o que encaminha para a análise de registos, segundo a modelação AR. Na modelação AR, os registos longos são sujeitos a uma segmentação, obtendo-se segmentos curtos e aproximadamente estacionários.

Os resultados da aplicação dos métodos tradicionais, no domínio do tempo e no domínio da frequência, em registos de Holter apresentam valores mais elevados durante o período noturno. Em registos estacionários de termo curto, a interpretação fisiológica no domínio da frequência é mais pormenorizada e informativa, quando comparada à no domínio do tempo. As componentes espectrais, LF e HF, estão associadas ao funcionamento do SNS e SNP e do SNP, respetivamente. Os modelos AR apenas permitem caracterizar a memória curta dos dados. No entanto, os espectros da VFC apresentam nas baixas frequências uma componente elevada, sendo uma característica típica dos processos de memória longa.

## Capítulo 3

# Métodos não lineares na VFC

Neste capítulo abordam-se metodologias alternativas às lineares, as metodologias não lineares. Os métodos apresentados no Capítulo 2 não se mostram adequados no estudo da complexidade e da dinâmica não linear da VFC. A introdução dos métodos não lineares solucionou o problema, mostrando uma nova abordagem das alterações do sistema cardiovascular sobre várias condições fisiológicas e patológicas. Estas aproximações diferem das medidas convencionais de análise da VFC, pois quantificam as propriedades do sinal, ao invés de analisar apenas a amplitude e a frequência (Schulz e Voss, 2017) e fornecem uma análise da VFC de termo longo, complexa e estendida. Os índices não lineares apresentam ainda vantagem em termos de estratificação de risco (Paradisi e Allegrini, 2017). Enquanto os métodos do Capítulo 2 analisam a componente de termo curto, neste capítulo introduz-se a análise do termo de memória longa (Secção 3.1), caracterizado pelos métodos não lineares. A Secção 3.2 dá especial atenção à modelação Autoregressiva de Médias Móveis Integradas Fracionariamente (ARFIMA), com erros Autoregressivos Condicionalmente Heteroscedásticos Generalizados (GARCH), os designados ARFIMA-GARCH. Estes modelos são uma extensão dos modelos AR e permitem descrever as componentes de termo curto e de termo longo, assim como as características associadas à dependência da volatilidade. Os métodos não lineares *Detrended Fluctuation Analysis* (DFA) e Entropia Aproximada (*ApEn*) estudam, nomeadamente,

a dinâmica e a complexidade não linear da VFC e são apresentados, respetivamente, nas Secções 3.3 e 3.4. Ao longo das várias secções apresentam-se os resultados da aplicação dos métodos não lineares, em registos de Holter, de indivíduos saudáveis.

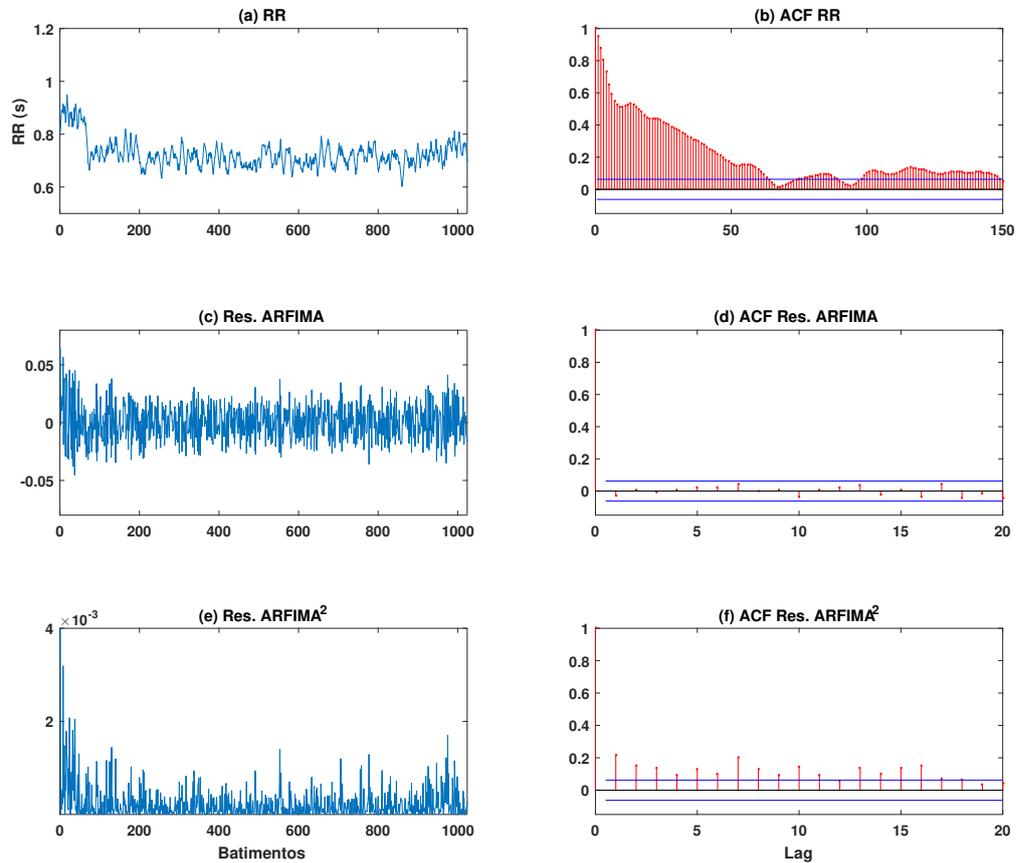
### 3.1 Memória longa e a heteroscedasticidade na VFC

A evidência empírica sobre a memória longa remonta a [Hurst \(1951\)](#) no campo da hidrologia ([Pérez e Ruiz, 2001](#)). Sem especificar a memória longa, o estudo deste conceito na economia foi proposto por [Granger e Joyeux \(1980\)](#), ao relacionarem a forma espectral e as variáveis das séries temporais económicas ([Haldrup e Valdés, 2017](#)). [Kobayashi e Musha \(1982\)](#) foram os pioneiros na observação da característica de memória longa na análise da VFC. A memória longa ilustra uma estrutura não linear, o que torna os métodos tradicionais ineficientes em explicar a essência da variável estudada ([Xiu e Jin, 2007](#)).

Uma série temporal de memória longa é um processo estacionário, que permite detetar a dependência existente entre o valor corrente e os valores em diferentes momentos temporais da série ([Sadai et al., 2016](#)). Pode ser caracterizada pela Função de Autocorrelação (ACF), que decai de forma hiperbólica. Este decaimento é lento quando comparado com o existente na memória curta. A característica de memória longa é ilustrada na Figura 3.1. Considerando em (a) um tacograma de um indivíduo saudável (referente a um segmento com 1024 batimentos retirado do registo longo representado na Figura 1.2) e em (b) a função de autocorrelação correspondente. Na Figura 3.1.(f), as séries dos resíduos ARFIMA<sup>2</sup> apresentam correlações significativas, sendo esta uma característica típica de registos com heteroscedasticidade.

Com a finalidade de estudar a memória longa, foram introduzidos os modelos ARFIMA. Da aplicação da modelação ARFIMA nos registos da VFC resultam séries residuais que não revelam correlações significativas, como é ilustrado na Figura 3.1 considerando em (c) a série residual e em (d) a correspondente função de autocorrelação. A modelação ARFIMA revela-se assim adequada na descrição dos

registos da VFC, no entanto, omite a não linearidade da descrição matemática destas séries temporais o que pode acarretar consequências indesejáveis.



**Figura 3.1** – (a) Segmento curto com 1024 batimentos, do registo longo, de um indivíduo saudável, (c) resíduos ARFIMA e (e) resíduos ARFIMA<sup>2</sup>. Em (b), (d) e (f) estão representadas as correspondentes ACF.

## 3.2 Modelação ARFIMA-GARCH

Séries temporais não estacionárias, que exibem dependência da memória longa, surgem em diversos campos de estudo. Para ultrapassar a limitação da não estacionaridade, [Box et al. \(1976\)](#) propuseram o Modelo Autoregressivo Integrado de Média Móvel (ARIMA( $p, d, q$ )). Este modelo aplica ao parâmetro de memória longa  $d$  um

número inteiro que estabelece o nível de diferenças necessárias para transformar a série temporal não estacionária numa série estacionária. Porém, assim como outros métodos convencionais, não abrange o estudo da característica de memória longa e tornou-se essencial a investigação de métodos que a incluíssem (Coli *et al.*, 2005). Na década de 80, foi proposto por Granger e Joyeux (1980) e Hosking (1981), o modelo ARFIMA( $p, d, q$ ) para modelar dependências de termo longo (Zevallos e Palma, 2013). Sendo uma generalização dos métodos ARIMA, os métodos ARFIMA diferenciam-se destes ao estenderem o parâmetro  $d$  a valores reais. Com o desígnio de obter uma caracterização alternativa da variância condicional, são propostos os modelos ARFIMA-GARCH, que detalham as componentes de termo curto e de termo longo, assim como as características associadas à dependência da variabilidade instantânea (Leite *et al.*, 2013).

### 3.2.1 Ruído branco fracionário

O modelo ARFIMA( $0, d, 0$ ) também conhecido como Ruído Branco Fracionário foi proposto independentemente por Granger e Joyeux (1980) e Hosking (1981). É considerado uma analogia em tempo discreto do Movimento Browniano Fracionário ( $fBm$ ) e é equivalente ao processo de Ruído Gaussiano Fracionário. Esta relação é de fácil deteção, visto que  $fBm$  pode ser designado como um processo auto-selhante Gaussiano com índice  $H$ , que pode tomar valores entre 0 e 1, com incrementos estacionários. O estudo de Mandelbrot e Ness (1968) permitiu definir os referidos processos não estacionários em tempos contínuos. Os modelos ARFIMA, contrariamente a  $fBm$  (que considera apenas a dependência de termo longo), permitem a modelação separada das dependências de termo curto e de termo longo (Ellis e Wilson, 2004).

O processo  $x(t)_{t \in \mathbb{Z}}$ , para  $d \in \mathfrak{R}$  é designado de ARFIMA( $0, d, 0$ ), quando denotado pela seguinte equação às diferenças:

$$(1 - B)^d x(t) = \varepsilon(t), \quad (3.1)$$

onde  $\varepsilon(t)$  é o processo de ruído branco, com média zero e variância  $\sigma_\varepsilon^2$ . O operador

diferença  $(1 - B)^d$  com  $d$  fracionário é definido pelo desenvolvimento binomial:

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j, \quad (3.2)$$

com  $\pi_0 = 1$ ,

$$\pi_j = \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} = \prod_{k=1}^j \frac{k - 1 - d}{k}, \quad j \in \mathbb{N}, \quad (3.3)$$

e a função  $\Gamma(u)$  é representada por:

$$\Gamma(u) \begin{cases} \int_0^{\infty} t^{u-1} e^{-t} dt, & u > 0 \\ \infty, & u = 0 \\ u^{-1} \Gamma(1 + u), & u < 0 \end{cases} .$$

### Propriedades do modelo ARFIMA(0,d,0) (Hosking, 1981)

(1) Para valores  $d < \frac{1}{2}$ ,  $x(t)_{t \in \mathbb{Z}}$  é um processo estacionário e pode ser representado por um processo de médias móveis de ordem infinita, dado por:

$$x(t) = (1 - B)^{-d} \varepsilon(t) = \left( \sum_{j=0}^{\infty} \psi_j B^j \right) \varepsilon(t) = \sum_{j=0}^{\infty} \psi_j \varepsilon(t - j),$$

onde  $\psi_0 = 1$ ,

$$\psi_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} = \prod_{k=1}^j \frac{h - 1 + d}{h}, \quad j \in \mathbb{N}.$$

Utilizando a expressão de *Stirling*,

$$\Gamma(u) \sim \sqrt{2\pi} e^{-u+1} (u - 1)^{u-\frac{1}{2}}, \quad u \rightarrow \infty, \quad (3.4)$$

aos coeficientes  $\psi_j$ , obtém-se:

$$\psi_j \sim \frac{j^{d-1}}{\Gamma(d)} \quad j \rightarrow \infty.$$

(2) Para valores  $d > -\frac{1}{2}$ ,  $x(t)_{t \in \mathbb{Z}}$ , é um processo autoregressivo invertível e pode ser representado por um processo de ordem infinita, dado por:

$$(1 - B)^d x(t) = \left( \sum_{j=0}^{\infty} \pi_j B^j \right) x(t) = \sum_{j=0}^{\infty} \pi_j x(t - j) = \varepsilon(t), \quad (3.5)$$

com  $\pi_j$  definido segunda a Equação (3.3). Aplicando novamente a fórmula de *Stirling* (3.4) aos coeficientes  $\pi_j$ , obtém-se:

$$\pi_j \sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad j \rightarrow \infty.$$

(3) Para valores  $-\frac{1}{2} < d < \frac{1}{2}$ ,  $x(t)_{t \in \mathbb{Z}}$  é um processo ARFIMA(0,  $d$ , 0) estacionário e invertível, que apresenta as seguintes propriedades:

### Função de Densidade Espetral

Num processo  $x(t)_{t \in \mathbb{Z}}$ , a função de densidade espectral é representada por:

$$f(\omega) = \sigma_\varepsilon^2 |1 - e^{-i\omega}|^{-2d} = \sigma_\varepsilon^2 \left| 2 \sin \left( \frac{\omega}{2} \right) \right|^{-2d}, \quad -\pi \leq \omega \leq \pi. \quad (3.6)$$

E como  $\sin \frac{\omega}{2} \sim \frac{\omega}{2}$ , para  $\omega \rightarrow 0$ , obtém-se a seguinte aproximação:

$$f(\omega) \sim \sigma_\varepsilon^2 |\omega|^{-2d}, \quad \omega \rightarrow 0. \quad (3.7)$$

### Função de Autocorrelação

Num processo  $x(t)_{t \in \mathbb{Z}}$ , a função de autocorrelação é representada por:

$$\rho(h) = \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)}, \quad (3.8)$$

e aplicando a fórmula de *Stirling*, obtém-se:

$$\rho(h) \sim \frac{\Gamma(1-d)}{\Gamma(d)} |h|^{2d-1}, \quad h \rightarrow \infty. \quad (3.9)$$

Considerações relevantes sobre o modelo ARFIMA(0,d,0) são indicadas seguidamente (Hosking, 1981):

- No intervalo  $d \in ]-\frac{1}{2}, \frac{1}{2}[$ , o processo ARFIMA(0,d,0) é estacionário e invertível e os coeficientes de representação autoregressiva e média móvel decaem hiperbolicamente;
- No intervalo  $d \in ]0, \frac{1}{2}[$ , o processo ARFIMA(0,d,0) é estacionário e apresenta memória longa. Em relação aos valores de  $\rho(h)$  são positivos e convergem hiperbolicamente e lentamente para zero. A densidade espectral tende para infinito, em frequências  $\omega$  próximas de zero;
- No intervalo  $d \in ]-\frac{1}{2}, 0[$  e para  $h > 0$ , o processo  $x(t)$  é estacionário com memória curta. Relativamente aos valores de  $\rho(h)$ , são negativos e convergem hiperbolicamente para zero, com  $\sum |\rho(h)| < \infty$ ;
- No intervalo  $d = 0$ , o processo ARFIMA(0,0,0) é um processo de ruído branco.

### 3.2.2 Modelos ARFIMA

O modelo ARFIMA(p,d,q) pode ser entendido como uma generalização do modelo ARFIMA(0,d,0). Para que  $x(t)_{t \in \mathbb{Z}}$  possa ser denominado por processo ARFIMA, é necessário que a série das diferenças fracionárias  $(1-B)^d x(t)$  seja um processo ARMA. Quando satisfaz a condição,  $x(t)_{t \in \mathbb{Z}}$  é um modelo ARMA integrado fracionalmente, logo, ARFIMA. A série temporal  $x(t)_{t \in \mathbb{Z}}$  pode ser denominada por processo ARFIMA(p,d,q) quando satisfaz a equação às diferenças (Xiu e Jin, 2007):

$$\phi(B)(1-B)^d x(t) = \theta(B)\varepsilon(t), \quad (3.10)$$

onde os polinómios  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  e  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  representam, respetivamente, as características autoregressivas e as de médias móveis. Todas as raízes de  $\phi(B)$  e  $\theta(B)$  estão fora do círculo unitário (Liu *et al.*, 2017) e referem-se à estrutura de correlação de termo curto. Se  $d \in ] -\frac{1}{2}, \frac{1}{2}[$ , o processo ARFIMA( $p, d, q$ ) é estacionário e invertível.

De seguida, são enumeradas algumas propriedades do modelo AFIRMA( $p, d, q$ ), tais como:

### Função de densidade espectral

Seja  $x(t)_{t \in \mathbb{Z}}$  um processo estacionário e invertível, definido na Equação (3.10). A sua função de densidade espectral é dada por (Reisen *et al.*, 2001),

$$f_{\text{ARFIMA}}(\omega) = f_{\text{SM}}(\omega) |1 - e^{-i\omega}|^{-2d}, \quad -\pi \leq \omega \leq \pi, \quad (3.11)$$

$$\text{onde } f_{\text{SM}}(\omega) = \sigma_\varepsilon^2 \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \quad (3.12)$$

é a função de densidade espectral referente às correlações de termo curto (*short memory*), neste caso dos processos ARMA (Reisen *et al.*, 2001). E ainda:

$$f_{\text{ARFIMA}}(\omega) \sim \sigma_\varepsilon^2 \frac{|\theta(1)|^2}{|\phi(1)|^2} |\omega|^{-2d}, \quad \omega \rightarrow 0. \quad (3.13)$$

### Função de autocorrelação

(Hosking, 1981) derivou as expressões exatas da função de autocorrelação para os modelos ARFIMA e provou que numa situação em que  $0 \leq d < \frac{1}{2}$ , existe uma constante positiva  $C_\rho$ , tal que:

$$\rho(h) \sim C_\rho |h|^{2d-1}, \quad h \rightarrow \infty. \quad (3.14)$$

Retira-se ainda que a função de autocorrelação decai hiperbolicamente e não é absolutamente somável.

Neste modelo podem indicar-se as seguintes considerações relevantes:

- Tal como no modelo ARFIMA(0,  $d$ , 0), no intervalo  $d \in ]0, \frac{1}{2}[$ , o processo  $x(t)_{t \in \mathbb{Z}}$  é estacionário com memória longa;
- Para valores  $d < \frac{1}{2}$  o processo  $x(t)_{t \in \mathbb{Z}}$  é estacionário quando todas as raízes de  $\phi(B) = 0$ , estão fora do círculo unitário;
- Para valores  $d > -\frac{1}{2}$ , o processo  $x(t)_{t \in \mathbb{Z}}$  é invertível quando todas as raízes de  $\theta(B) = 0$ , estão fora do círculo unitário.

Uma revisão dos processos ARFIMA pode ser consultada em [Box \*et al.\* \(1976\)](#), [Brockwell e Davis \(1991\)](#), [Beran \(1991\)](#) e [Doukhan \*et al.\* \(2003\)](#).

### 3.2.3 Modelos GARCH

Para descrever a volatilidade, tendo em atenção os efeitos heteroscedásticos de uma série temporal são aplicados os modelos Autoregressivos Condicionalmente Heteroscedásticos Generalizados (GARCH). Os Modelos GARCH, propostos por [Bollerslev \(1986\)](#), são uma extensão dos modelos ARCH, propostos por [Engle \(1982\)](#), e incluem a informação da dependência da componente da variância condicional  $\sigma_\varepsilon^2$ , não só do passado do processo, mas também do seu próprio passado ([Gajda \*et al.\*, 2018](#)).

O processo  $\varepsilon(t)_{t \in \mathbb{Z}}$  diz-se GARCH( $r, s$ ) quando satisfaz as condições:

$$\varepsilon(t) = \sigma_\varepsilon(t)z(t) \text{ e } \sigma_\varepsilon^2(t) = u_0 + \sum_{j=1}^r u_j \varepsilon^2(t-j) + \sum_{l=1}^s v_l \sigma_\varepsilon^2(t-l), \quad (3.15)$$

onde  $z(t)$  é uma sucessão de variáveis aleatórias, independentes e identicamente distribuídas, com média nula e variância igual a 1,  $u_0 > 0$  e  $u_1, \dots, u_r, v_1, \dots, v_s \geq 0$ . Se  $\sum_{j=1}^r u_j + \sum_{l=1}^s v_l < 1$ , então o modelo GARCH( $r, s$ ) é estacionário ([Bollerslev, 1986](#)). Esta condição permite demonstrar que ([Leite, 2008](#)):

$$\sigma_\varepsilon^2 = \text{var}[\varepsilon(t)] = \frac{u_0}{1 - \sum_{j=1}^r u_j - \sum_{l=1}^s v_l}$$

Este trabalho restringe-se ao Modelo Garch( $r,s$ ) com  $r=1$  e  $s=1$ , que são valores correntemente referidos na literatura (Baillie e Chung, 1996; Hu e Tsoukalas, 2006):

$$\varepsilon(t) = \sigma_\varepsilon(t)z(t) \text{ e } \sigma_\varepsilon^2(t) = u_0 + u_1 \varepsilon^2(t-1) + v_1 \sigma_\varepsilon^2(t-1),$$

composto por apenas três parâmetros  $u_0$ ,  $u_1$  e  $v_1$ , onde  $u_1$  é referido como parâmetro de estimação ARCH e  $v_1$  é referido como parâmetro de estimação GARCH.

Para o modelo GARCH indicam-se as seguintes considerações relevantes:

- O modelo GARCH( $r,s$ ), reduz-se ao modelo ARCH( $r$ ), quando  $s=0$ ;
- Designa-se por processo integrado GARCH( $r,s$ )(IGARCH), quando  $\sum_{j=1}^r u_j + \sum_{l=1}^s v_l = 1$ .

Uma revisão dos processos ARCH, GARCH e IGARCH encontra-se disponível em Hansen (1995) e Tsay (2005).

### 3.2.4 Modelos ARFIMA-GARCH

Os modelos ARFIMA, como referido anteriormente, descrevem a média condicional dos dados. Porém a VFC apresenta também mudanças na variância ao longo do tempo, com períodos de acentuada variabilidade seguidos de períodos de estabilidade, sugerindo heteroscedasticidade (variação da variância). Bollerslev (1986) desenvolveu os modelos GARCH, resultado da extensão do modelo Autoregressivo Condicionalmente Heteroscedástico (ARCH), proposto por Engle (1982). Leite *et al.* (2007), estudaram a modelação conjunta de características de memória longa e heteroscedasticidade da VFC e acoplaram os modelos ARFIMA e GARCH, considerando os modelos não lineares ARFIMA-GARCH.

O processo  $x(t)_{t \in \mathbb{Z}}$  diz-se ARFIMA( $p,d,q$ )-GARCH( $r,s$ ) quando satisfaz a equação às diferenças (Leite, 2008):

$$\phi(B)(1-B)^d x(t) = \theta(B)\varepsilon(t), \quad (3.16)$$

com

$$\varepsilon(t) = \sigma_\varepsilon(t)z(t) \text{ e } \sigma_\varepsilon^2(t) = u_0 + \sum_{j=1}^r u_j \varepsilon^2(t-j) + \sum_{l=1}^s v_l \sigma_\varepsilon^2(t-l), \quad (3.17)$$

onde  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  e  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  são polinômios em  $B$ , tais que todas as raízes de  $\phi(B)$  e  $\theta(B)$  estão fora do círculo unitário.  $B$  é o operador de atraso e  $(1-B)^d$  é o operador diferença definido em (3.2) e (3.3),  $z(t)$  corresponde à sucessão de variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância 1,  $u_0 > 0, u_1, \dots, u_r, v_1, \dots, v_s \geq 0$  e  $\sum_{j=1}^r u_j + \sum_{l=1}^s v_l < 1$ .

Na Equação (3.16) o parâmetro  $d$  determina o comportamento de memória longa na média, enquanto os parâmetros  $p$  e  $q$  e os coeficientes de  $\phi(B)$  e  $\theta(B)$  permitem modelar as propriedades de termo curto. A Equação (3.17) descreve a variância condicional do processo variável no tempo, inversamente aos modelos tradicionais. O parâmetro  $u_1$  caracteriza a propriedade de memória curta, na volatilidade, e o parâmetro  $v_1$  caracteriza a persistência na volatilidade (Almeida *et al.*, 2017).

Para valores de  $d = 0$ , obtém-se o modelo ARMA( $p, q$ )-GARCH( $r, s$ ). Ao considerar  $r = s = 0$ , a modelação ARFIMA( $p, d, q$ )-GARCH( $r, s$ ) reduz-se à modelação ARFIMA( $p, d, q$ ). Tendo ainda em atenção as duas últimas condições e adicionando valores de  $q = 0$ , obtém-se o modelo ARFIMA( $p, 0, 0$ ), isto é, o modelo AR( $p$ ) usual no estudo da VFC (Leite, 2008).

### Função de densidade espectral

A função da densidade espectral dos processos ARFIMA-GARCH com  $-\frac{1}{2} < d < \frac{1}{2}$ , é baseada na combinação do espectro AR( $p$ ) (2.9) com o espectro ARFIMA( $p, d, q$ ) (3.11), isto é,

$$f_{AG}(\omega) = f_{SM}(\omega) |1 - e^{-i\omega}|^{-2d}, \quad -\pi \leq \omega \leq \pi, \quad (3.18)$$

onde  $f_{SM}(\omega)$ , definido na Equação (3.12), corresponde à função de densidade espectral, relativa às correlações de termo curto, processo ARMA( $p, q$ )-GARCH( $r, s$ ), com a equação  $\sigma_\varepsilon^2 = \frac{u_0}{1 - \sum_{j=1}^r u_j - \sum_{l=1}^s v_l}$ . Para valores de  $\frac{1}{2} \leq d < 1$ , a função (3.18) designa-se por *pseudo-densidade espectral* e o processo ARFIMA-GARCH é não estacionário.

### Função de autocorrelação

Num processo  $x(t)_{t \in \mathbb{Z}}$  estacionário e invertível, a função de autocorrelação para o modelo ARFIMA( $p, d, q$ )-GARCH( $r, s$ ) (Almeida *et al.*, 2017) é dada por,

$$\rho(h) \sim C|h|^{2d-1}, \quad \text{para } h \rightarrow \infty \quad \text{com } C \neq 0.$$

### Considerações relevantes sobre o modelo ARFIMA-GARCH

Se  $-\frac{1}{2} < d < \frac{1}{2}$ , todas as raízes de  $\phi(B)$  e  $\theta(B)$  estão fora do círculo unitário e  $\sum_{j=1}^r u_j + \sum_{l=1}^s v_l < 1$ , então o processo ARFIMA( $p, d, q$ )-GARCH( $r, s$ ), satisfazendo (3.16) e (3.17) é estacionário e invertível.

### 3.2.5 Estimação de processos ARFIMA-GARCH

#### Estimação de um modelo ARFIMA(0, $d$ , 0)

Partindo de uma base de dados, o primeiro passo na identificação é investigar a estrutura de dependência dos respectivos dados. Nos modelos ARFIMA, a força de dependência é refletida pelo parâmetro de memória  $d$  (Burnecki e Sikora, 2017). Existem dois processos de estimação principal: a estimação paramétrica e a estimação semi-paramétrica. A primeira corresponde ao método de máxima verosimilhança. No segundo processo os estimadores podem ser construídos usando métodos locais ou globais, não sendo necessário a especificação total do modelo. Nos métodos locais, apenas as frequências que estão numa vizinhança de zero são usadas. Por outro lado, nos métodos globais são utilizadas todas as frequências (Leite, 2008).

Atualmente, existem várias metodologias para estimar o parâmetro de memória longa, nomeadamente (Burnecki e Sikora, 2017; Leite, 2008; Reisen *et al.*, 2001):

#### 1. Estatística R/s clássica

Hurst (1951) foi pioneiro no desenvolvimento de um algoritmo, que permite estimar

a memória longa de um processo:

$$\log(R/s)_n = c \cdot n^H,$$

onde  $R/s$  é a estatística do intervalo de escala, medida ao longo de um índice de tempo  $n$ ,  $c$  é uma constante e  $H$  o expoente de Hurst (Mandelbrot e Wallis, 1969). Através do estatístico  $(R/s)_n$  é possível estimar o expoente de Hurst que caracteriza a série, com (Mohamed, 2016):

$$\log(R/s)_n = \log(c) + H \log(n).$$

Considerando uma amostra do conjunto de dados  $x(t) : t = 1, 2, \dots, n$ , de um processo estacionário de memória longa, o termo  $(R/s)_n$  é calculado de acordo com a equação:

$$(R/s)_n = \frac{\max_{1 \leq k \leq n} \sum_{t=1}^k [x(t) - \bar{x}] - \min_{1 \leq k \leq n} \sum_{t=1}^k [x(t) - \bar{x}]}{s}, \quad (3.19)$$

onde  $\bar{x}$  e  $s$  são, respectivamente, a média e o desvio padrão do conjunto  $n$ .

O coeficiente de Hurst estima o parâmetro de diferenciação fracional,  $d$ , tendo em conta a seguinte equação:

$$d = H - \frac{1}{2}.$$

## 2. Estimador local de Whittle (LWE)

A ideia inicial de usar métodos locais foi proposta por Geweke e Porter-Hudak (1983). Posteriormente, Kunsch (1987) propôs o estimador semi-paramétrico Gaussiano ou estimador local de Whittle (LWE). O estimador semi-paramétrico que apenas usa a forma paramétrica da função espectral numa vizinhança de origem é dado por:

$$f(\omega) \sim G\omega^{-2d}, \quad \omega \rightarrow 0,$$

onde  $0 < G < \infty$  e  $-\frac{1}{2} < d < \frac{1}{2}$ .

O método local de Whittle permite estimar o parâmetro  $d$  minimizando a seguinte função objetivo:

$$\log \left\{ \frac{1}{m} \sum_{j=1}^m \omega_j^{2d} I(\omega_j) \right\} - \frac{2d}{m} \sum_{j=1}^m \log \omega_j,$$

com  $I(\omega_j)$  definido em (2.4). Estudos de Robinson (1995) e Velasco (1999) demonstraram que o estimador LWE é consistente para valores de  $-\frac{1}{2} < d < \frac{1}{2}$ . O valor de  $m$  (número de frequências na vizinhança de zero) é importante para uma estimação apropriada. A literatura apresenta:  $m = n^\eta$ , com  $\eta \in [\frac{1}{2}, \frac{13}{20}]$ .

### Estimação de um modelo ARFIMA( $p, d, q$ )-GARCH( $r, s$ )

Para estimar os parâmetros do modelo ARFIMA( $p, d, q$ )-GARCH( $r, s$ ) definido em (3.16) e (3.17) pode ser utilizado o método de máxima verosimilhança aproximada, descrito por Baillie e Chung (1996) e Ling e Li (1997). Estudos de Robinson e Henry (1999), Henry (2001) e Nielsen e Frederiksen (2005) usaram simulações, demonstrando que para valores de  $-\frac{1}{2} < d < \frac{1}{2}$ , o estimador local de Whittle é robusto nos processos com erros condicionalmente heteroscedásticos.

O procedimento para a estimação dos parâmetros dos modelos ARFIMA-GARCH é deveras semelhante ao usado para modelos ARFIMA:

1. Estimar o parâmetro  $d$  num dado registo  $x(t)_{t \in \mathbb{Z}}$ ;
2. Obter o registo filtrado  $y(t) = (1 - B)^d x(t)$ ;
3. Estimar os parâmetros do modelo ARMA( $p, q$ )-GARCH( $r, s$ ) no registo filtrado  $y(t)$ .

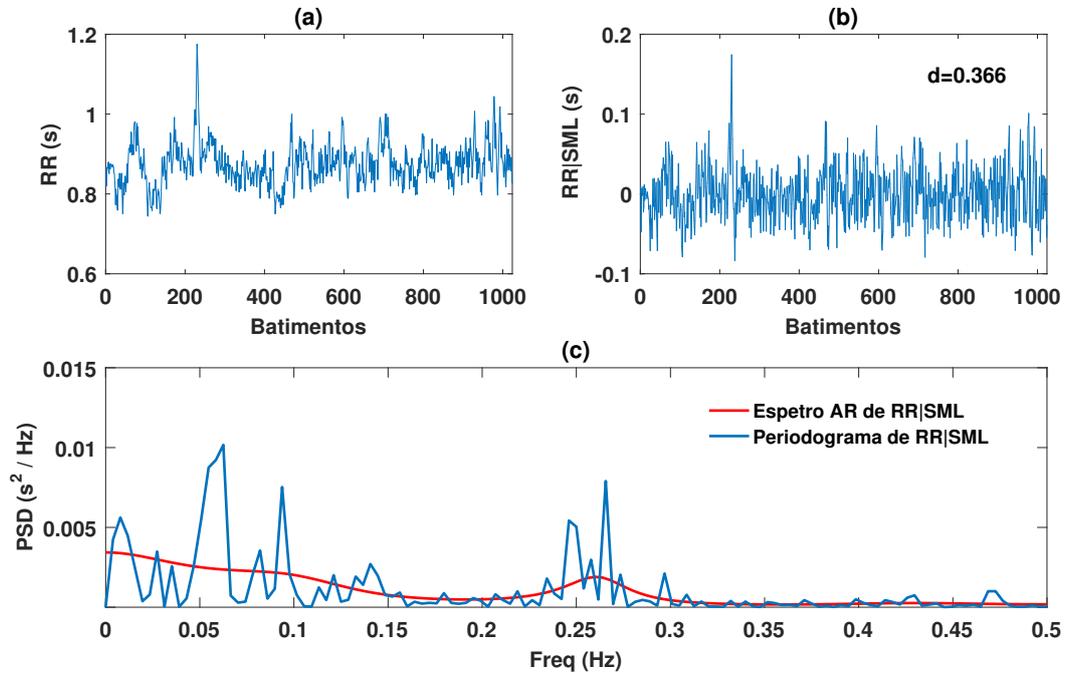
Relativamente ao passo 1, o parâmetro  $d$  é estimado considerando o estimador LWE. No passo 3, para estimar os parâmetros do modelo ARMA-GARCH é utilizado o método de máxima verosimilhança (Hannan, 1973), com a *toolbox* de Econometria do MATLAB.

### 3.2.6 Aplicação em dados de Holter

#### Análise de registos curtos

Como referido anteriormente, na VFC a modelação ARFIMA-GARCH apresenta

vantagens relativamente à modelação tradicional AR, uma vez que deteta e remove a memória longa (Figura 3.2).



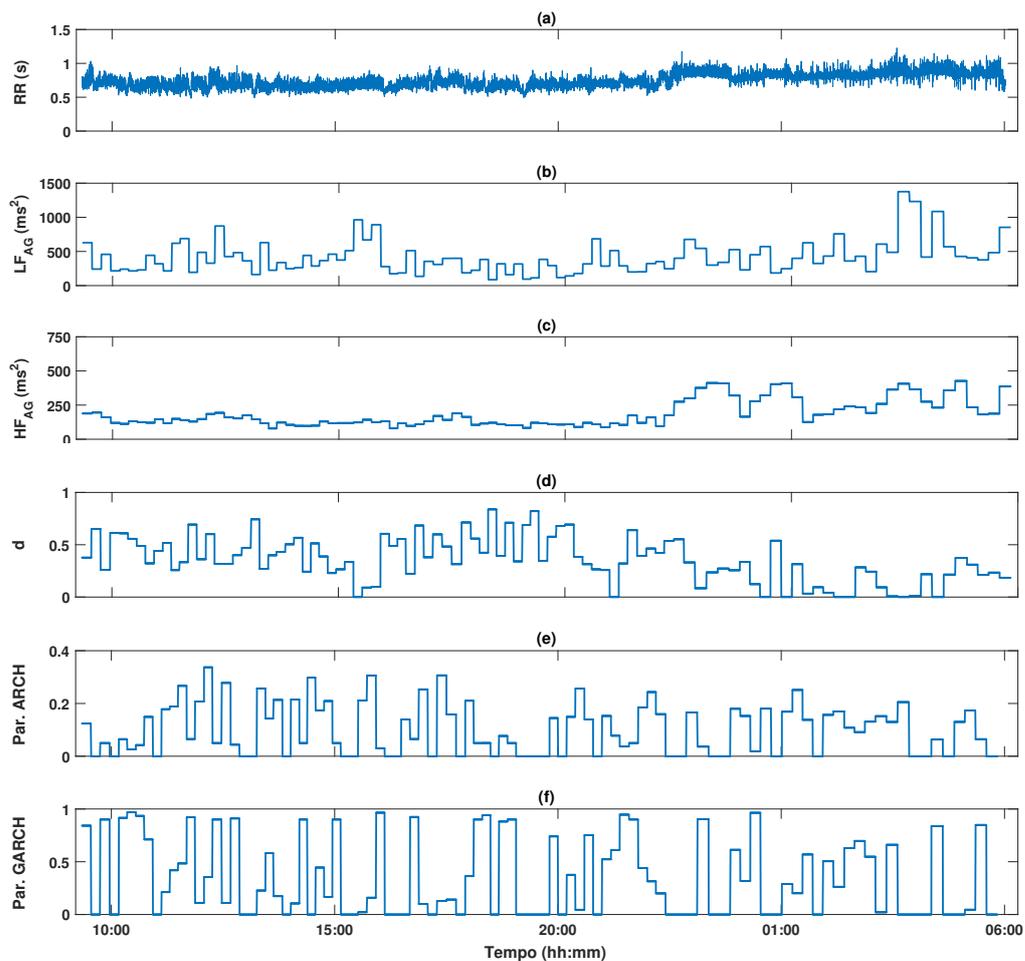
**Figura 3.2** – (a) Registo curto da VFC, com 1024 batimentos, (b) Registo RR curto, Sem Memória Longa (Sinal RR|SML) e (c) Representação gráfica do espectro de AR, do Sinal RR|SML e do modelo não paramétrico, periodograma, do Sinal RR|SML.

A remoção promove a fiabilidade dos resultados de  $LF_{AG}$  e  $HF_{AG}$  (componentes espectrais, de memória curta). O modelo permite ainda calcular a volatilidade do sinal.

Na Figura 3.2.(a) exemplifica-se o tacograma de um segmento curto, com dimensão igual a 1024 batimentos (aproximadamente estacionário), retirado do registo longo ilustrado na Figura 1.2. O parâmetro  $d$  dita a presença de memória longa. Para o caso ilustrado,  $d = 0.366$  o que indica que o registo tem memória longa. A estimação do parâmetro de memória longa,  $d$ , potencia a aplicação de um filtro que retira a componente de termo longo, obtendo-se o sinal representado na Figura 3.2.(b). No espectro denotam-se duas deflexões positivas, uma nas baixas frequências

(LF:0.04 – 0.15Hz) e outra nas altas frequências (HF:0.15 – 0.4Hz). Utilizou-se uma análise sequencial, em que os registos longos são decompostos em segmentos curtos (com 1024 batimentos).

### Análise de registos longos



**Figura 3.3** – Tacograma de um indivíduo saudável, ao longo de 24 horas, e respetivos parâmetros ARFIMA-GARCH.

Na Figura 3.3.(a) representa-se um registo de Holter, referente a um indivíduo saudável (registo longo e não estacionário, com 101160 batimentos). Pretende-se obter uma descrição adequada de registos longos e não estacionários, para isso o

registro é sujeito a uma segmentação combinada com o modelo ARFIMA-GARCH, de modo a obter segmentos com dimensão igual a 1024, aproximadamente estacionários. Com a remoção da memória longa, a descrição das componentes espectrais ARFIMA é mais concisa. Nas figuras 3.3.(b) e 3.3.(c) denotam-se valores superiores durante o período noturno, para as componentes espectrais. O parâmetro  $d$  é mais elevado entre as 19h00 e as 20h00 e menor durante a noite.

Os parâmetros ARCH e GARCH representados, respetivamente, nas figuras 3.3.(e) e 3.3.(f), apresentam bastante variabilidade para o caso estudado.

### 3.3 Método *Detrended Fluctuation Analysis*

A análise não paramétrica por *Detrended Fluctuation Analysis* (DFA) é uma metodologia aplicada em sistemas dinâmicos e complexos. Quando aplicada na análise da VFC permite interpretar os intervalos RR, ao quantificar as características de memória longa, isto é, ao detetar a presença (fractal) da memória da lei da potência, no sinal (Paradisi e Allegrini, 2017; Piskorskia *et al.*, 2018; Alvarez-Ramirez *et al.*, 2005).

#### 3.3.1 Introdução ao método

O método escalar DFA foi originalmente desenvolvido por Peng *et al.* (1995) para descrever a estrutura mosaico da sequência de DNA. É geralmente implementado com o intuito de avaliar as características dos dados, como são exemplo, as sequências de DNA, a dinâmica da frequência cardíaca, os registos meteorológicos e as séries temporais financeiras (Alvarez-Ramirez *et al.*, 2005). O principal objetivo do método DFA é obter a correlação estrutural das flutuações, que correspondem a elevado ruído, resultando, após a remoção da *tendência*, na sua representação através de um polinómio. Apresenta, relativamente aos métodos tradicionais, a vantagem de analisar as séries temporais não estacionárias, considerando a componente de termo curto e de termo longo (Piskorskia *et al.*, 2018).

A relação entre o DFA e a VFC é a base de muitas pesquisas com o intuito de avaliar o risco de mortalidade em vários grupos (Hsin *et al.*, 2010).

### Estimação do método DFA

Considerando um processo  $x(1), \dots, x(n)$ , o procedimento para obter a estimação usando o método DFA pode ser resumido através dos seguintes passos (Absil *et al.*, 1999):

1. Calcular o integral indeterminado de  $x$ :

$$y(l) = \sum_{i=1}^l [x(i) - \bar{x}],$$

onde  $\bar{x}$  representa a média amostral;

2. Dividir  $y(l)$  em janelas de dimensão  $k$ . Geralmente não é possível dividir exatamente os  $n$  pontos da série em janelas de dimensão  $k$ . Assim, para cada valor, designa-se  $\tilde{n}$  o maior múltiplo de  $k$ , inferior ou igual a  $n$ ;
3. Para cada janela de dimensão  $k$ , calcular a tendência linear local, ao ajustar uma reta pelo método dos mínimos quadrados. Denota-se por  $y_k(l)$  a coordenada de reta ajustada;
4. Subtrair a tendência linear local  $y_k(l)$  de cada segmento à serie  $y(l)$ ;
5. Para cada  $k$ , determinar a raiz quadrada da média dos quadrados dos desvios, ou seja, a função DFA:

$$F(k) = \sqrt{\frac{1}{\tilde{n}} \sum_{l=1}^{\tilde{n}} [y(l) - y_k(l)]^2}; \quad (3.20)$$

6. Para determinar o coeficiente  $\alpha$ : calcular o declive da reta ajustada aos valores de  $\log F(k)$ , em função de  $\log k$ .

Para valores elevados de  $k$ , o expoente de escala com valor  $\alpha = \frac{1}{2}$  representa dados sem características de memória longa. Por outro lado, valores de  $\alpha > \frac{1}{2}$  indicam a presença de memória longa. Relaciona-se o expoente de escala  $\alpha$  com o parâmetro  $d$  da modelação ARFIMA-GARCH, através da fórmula  $\alpha = d + \frac{1}{2}$ .

### 3.3.2 Expoentes de escala na VFC

Para estimar a memória longa  $\alpha$ , usando o DFA, em registos de VFC, de 24 horas (aproximadamente 100 mil batimentos) têm sido considerados vários intervalos para as escalas de tempo  $k$ :  $20 \leq k \leq 10000$  (Peng *et al.*, 1995),  $20 \leq k \leq 4000$  (Struzik *et al.*, 2006) e  $128 < k < 4096$  (Cerutti *et al.*, 2007). Leite *et al.* (2007) considerou segmentos com 4096 batimentos e  $64 \leq k \leq 1024$ , para estimar a memória longa. Embora o método DFA tenha sido desenvolvido para estimar a memória longa, também tem sido considerado na determinação de expoentes referentes à memória curta. Peng *et al.* (1995) estima o expoente escalar em duas escalas de tempo diferentes:  $\alpha_1$  para o intervalo  $4 \leq k \leq 16$  e  $\alpha_2$  para o intervalo  $16 \leq k \leq 64$ . Em outras escalas:  $\alpha_{\text{HF}}$  para o intervalo  $4 \leq k \leq 11$  e  $\alpha_{\text{LF}}$  para o intervalo  $12 \leq k \leq 32$ .

**Tabela 3.1** – Expoentes de escala usando o método DFA.

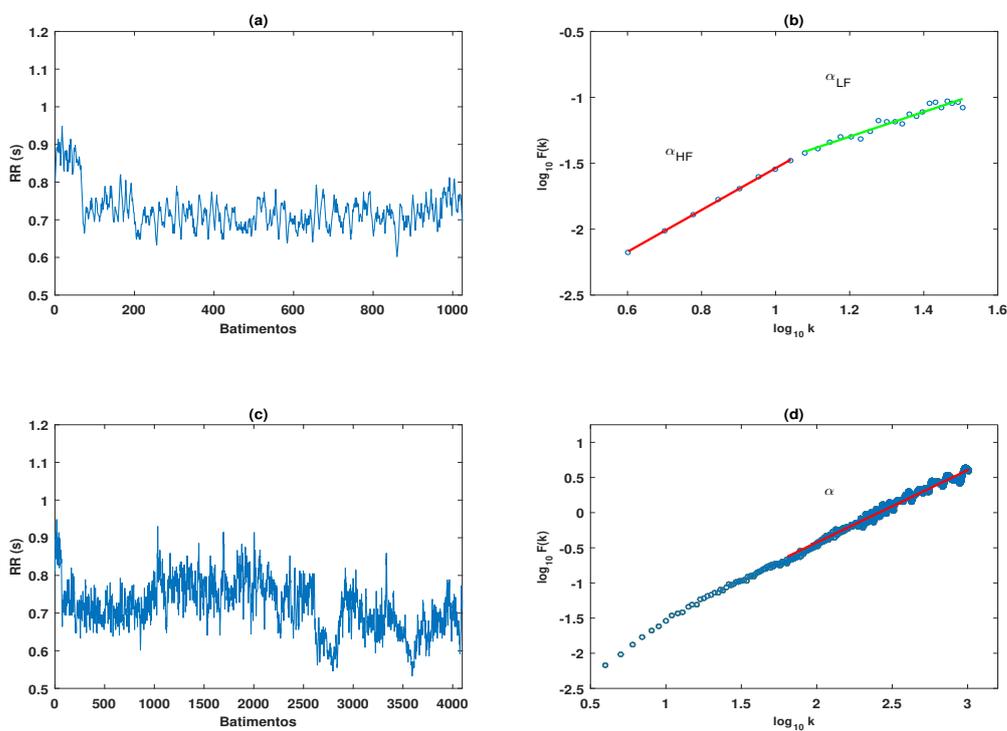
Expoentes escalares	Memória	Escalas de Tempo	Dimensão dos segmentos
<b>Sinal RR</b>			
$\alpha_{\text{HF}}$	Curta	$4 \leq k \leq 11$	1024
$\alpha_{\text{LF}}$	Curta	$12 \leq k \leq 32$	1024
$\alpha$	Longa	$64 \leq k \leq 1024$	4096
<b>Sinal RR SML</b>			
$\alpha_{\text{HF} SML}$	Curta	$4 \leq k \leq 11$	1024
$\alpha_{\text{LF} SML}$	Curta	$12 \leq k \leq 32$	1024

Neste trabalho, para além do cálculo dos expoentes referentes à memória curta,  $\alpha_{\text{LF}}$  e  $\alpha_{\text{HF}}$ , foram ainda calculados os mesmos expoentes nos sinais sem componente de memória longa (Sinal RR|SML), que se designam por  $\alpha_{\text{LF}|SML}$  e  $\alpha_{\text{HF}|SML}$ . A memória longa é removida usando o filtro  $(1-B)^d$  RR, onde o  $d$  é estimado através do método LWE. Os expoentes de escala usados neste estudo são sumariados na Tabela 3.1.

### 3.3.3 Análise de registos de Holter

A aplicação do método DFA na análise da VFC, de um indivíduo saudável, é ilustrada nas figuras 3.4 e 3.5, nomeadamente, em registos curtos e num registo de 24 horas.

#### Análise de registos curtos

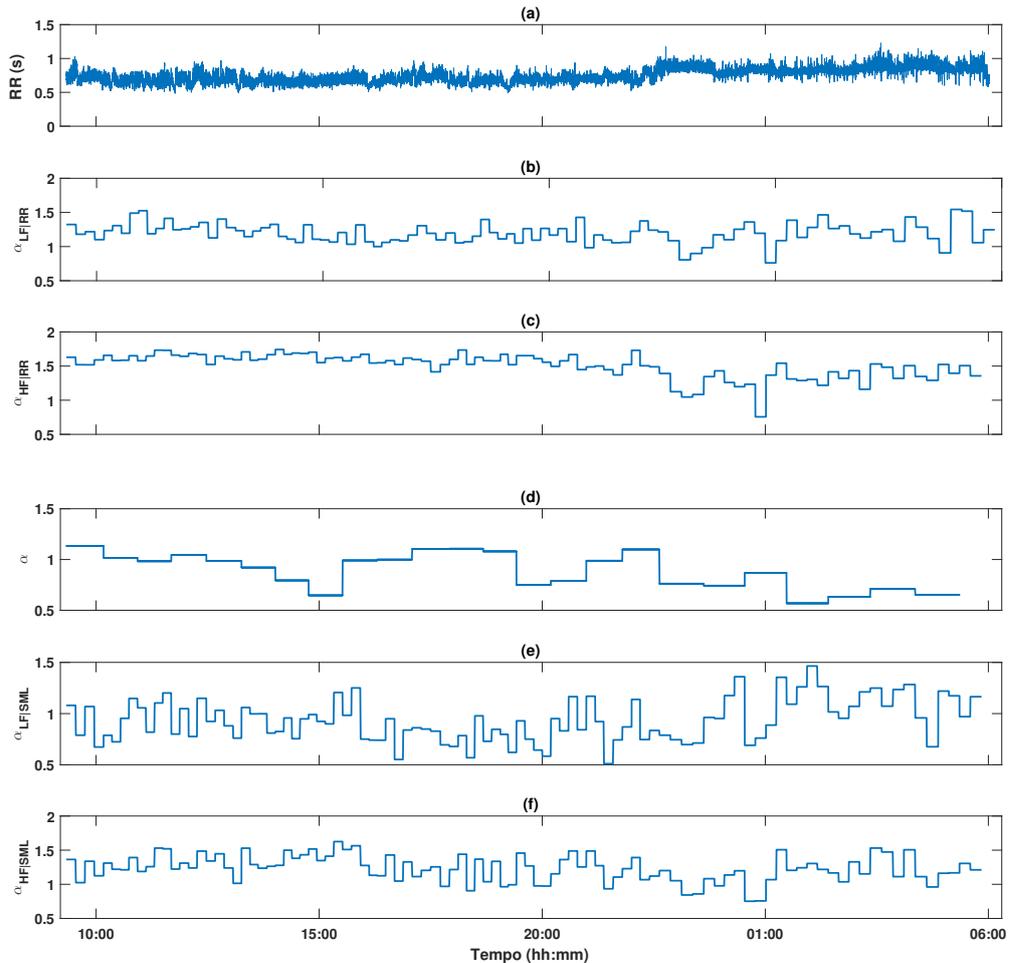


**Figura 3.4** – Tacograma de um indivíduo saudável, referente a dois segmentos curtos: (a) segmento com 1024 batimentos, que estima os (b) expoentes de escala  $\alpha_{LF}$  e  $\alpha_{HF}$  e (c) segmento curto com 4096 batimentos, que estima o (d) expoente de escala  $\alpha$ .

A Figura 3.4 ilustra dois segmentos curtos, retirados de um registo longo, referentes a um indivíduo saudável. A aplicação do método DFA no segmento curto com uma dimensão igual a 1024, é apresentada na Figura 3.4.(b) e estima os expoentes escalares de memória curta,  $\alpha_{LF}$  e  $\alpha_{HF}$ . A escala de tempo  $k$  referente a  $\alpha_{HF}$  é

$4 \leq k \leq 11$  e relativamente a  $\alpha_{LF}$  é  $12 \leq k \leq 32$ . Verifica-se ainda um ponto de mudança, para  $k = 12$ . A estimação do expoente de escala de memória longa  $\alpha$  é exibida na Figura 3.4.(d), para um segmento curto com dimensão igual a 4096.

### Análise de registos longos



**Figura 3.5** – (a) Tacograma de um registo de Holter de 24 horas, pertencente a um indivíduo saudável. Representação dos parâmetros obtidos pela segmentação combinada com o método DFA: (b)  $\alpha_{LF}$ , (c)  $\alpha_{HF}$  e (d)  $\alpha$  para o Sinal RR e (e)  $\alpha_{LF}$  e (f)  $\alpha_{HF}$  para o Sinal RR sem componente de memória longa (Sinal RR|SML).

O registo de Holter de um indivíduo saudável é descrito através dos expoentes escalares da Figura 3.5. O registo longo é sujeito a uma segmentação combinada

com o método DFA, obtendo-se segmentos curtos com dimensão fixa.

A análise dos expoentes escalares de memória curta, do Sinal RR (figuras 3.5.(b) e 3.5.(c)) e do Sinal RR|SML (figuras 3.5.(e) e 3.5.(f)) é obtida em segmentos curtos com 1024 batimentos. Por outro lado, o expoente escalar de memória longa (Figura 3.5.(d)) é descrito em segmentos curtos com 4096 batimentos (dimensão mínima para uma estimativa adequada do expoente escalar). Da interpretação da figura verifica-se que o expoente  $\alpha$  atinge valores mínimos durante o período noturno, o que entra em concordância com o comportamento do parâmetro de memória longa  $d$ . Os expoentes  $\alpha_{\text{HF|RR}}$  e  $\alpha_{\text{HF|SML}}$  apresentam também valores mais baixos durante a noite.  $\alpha_{\text{LF|RR}}$  não apresenta variação significativa quer no período do dia quer no período da noite, o que não se verifica para  $\alpha_{\text{LF|SML}}$  que atinge o seu valor máximo a partir da 1h00.

## 3.4 Entropia

A entropia é considerada uma importante grandeza da Física e tem vindo a ser cada vez mais relevante na quantificação da VFC. A entropia de Shannon, a entropia condicional, a entropia aproximada e a entropia amostral são aproximações utilizadas frequentemente na estimativa da entropia. A avaliação da entropia em múltiplas escalas de tempo foi proposta com o intuito de quantificar a complexidade da VFC em diferentes regimes de tempo (Baumert *et al.*, 2013).

O conceito de entropia foi originalmente definido por Clausius (1865), para representar o processo de transformação e a irreversibilidade dos processos cíclicos. A partir da extensão da equação de Clausius, Kondepudi e Prigogine (1998) chegaram à fórmula que define a entropia termodinâmica. Esta fórmula pode ser utilizada não só sob condições de equilíbrio como também em condições de não equilíbrio (Takada *et al.*, 2015). Posteriormente, na mecânica, foi introduzido o conceito da entropia de Boltzmann, que representa a função relevante quando todos os estados microscópicos têm a mesma energia, ou de uma forma mais geral, quando todos os estados têm a

mesma probabilidade de ocorrer. A entropia de Gibbs ou entropia estatística, traduz a alteração dos métodos cinéticos em combinatórios. O conceito de conjuntos de Gibbs deu origem a técnicas de cálculo das propriedades de equilíbrio. A entropia de Shannon tem origem na teoria da informação e quantifica o valor esperado de informação (Takada *et al.*, 2015). Esta última inspirou a entropia de Kolmogorov-Sinai que permite obter uma medida da taxa de informação referente ao sistema, ao longo do tempo. Contudo, os limites existentes na sua formulação não permitem determinar o seu valor para séries temporais curtas. Assim, surgiram alternativas como, por exemplo, a Entropia Aproximada (*ApEn*) (Pincus, 1991).

### 3.4.1 Entropia aproximada

A entropia aproximada é muito usada na deteção de mudanças em sistemas complexos e no estudo da natureza de séries temporais. Revelou-se uma métrica de grande importância em diversas áreas, incluindo na análise do intervalo RR, ao quantificar a regularidade do ciclo cardíaco e ao interpretar com precisão o ECG, o EEG e os sinais respiratórios.

#### Estimação da *ApEn*

Dada uma série de tempo discreta  $x(t) = x(1), x(2), \dots, x(n)$ , de dimensão  $n$ , a *ApEn* é determinada segundo dois parâmetros de entrada (Molina-Picó *et al.*, 2011): o parâmetro de tolerância  $r$  e dimensão  $m$ . A sequência de entrada  $x(t)$  é dividida numa série de subsequências  $X_m(i)$  que correspondem a parcelas de comprimento  $m$  da sequência de entrada, iniciada em  $x(i)$ .

$$X_m(i) = [x(i), x(i+1), \dots, x(i+m-1)], \quad 1 \leq i \leq n-m+1, \quad (3.21)$$

A distância  $d$  é definida entre duas subsequências,  $X_m(i)$  e  $X_m(j)$ , como o máximo absoluto da diferença entre as respetivas amostras:

$$d[X_m(i), X_m(j)] = \max_{0 \leq k \leq m-1} |x(i+k) - x(j+k)|. \quad (3.22)$$

A partir da Equação (3.22) são formulados:

$$B_i(r) = \text{número de } j : d[X_m(i), X_m(j)] \leq r$$

$$A_i(r) = \text{número de } j : d[X_{m+1}(i), X_{m+1}(j)] \leq r, \quad (1 \leq j \leq n - m + 1),$$

onde  $B_i(r)$  representa o número de subsequências de dimensão  $m$ , cuja distância não excede o limiar  $r$  e  $A_i(r)$  contabiliza as subsequências de comprimento  $m + 1$ .

Já com os logaritmos de  $B_i^m(r) = \frac{1}{n-m+1} B_i(r)$  e  $A_i^m(r) = \frac{1}{n-m} A_i(r)$  formulados, isto é:

$$\phi^m(r) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \log B_i^m(r)$$

$$\phi^{m+1}(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} \log A_i^m(r),$$

a entropia determina a média de ambos e calcula a sua diferença. A fórmula de entropia desenvolvida por Eckmann e Ruelle (Zurek *et al.*, 2012) é dada por:

$$ER = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]. \quad (3.23)$$

Para superar algumas limitações da entropia ER, quando aplicada a conjuntos de dados experimentais ruidosos, Pincus (1991) fixou os valores de  $r$  e  $m$ . A  $ApEn$  é dada pela seguinte equação:

$$ApEn(m, r) = \lim_{n \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]. \quad (3.24)$$

A estimativa da  $ApEn$  para um determinado conjunto de dados longos de  $n$ -pontos finitos é representada por:

$$ApEn(m, r, n) = \phi^m(r) - \phi^{m+1}(r). \quad (3.25)$$

A  $ApEn$  quantifica a probabilidade (logarítmica), ou seja, entre os padrões semelhantes de tamanho  $m$ , calcula a percentagem de padrões que continuará similar para  $m + 1$ . A  $ApEn$  apresenta valores positivos, e a identificação entre *regularidade* e *irregularidade* é demonstrada pelos baixos e altos valores da entropia. Assim, valores de  $ApEn$  próximos de zero, correspondem a elevadas probabilidades dos padrões continuarem próximos, o que indica séries com elevada regularidade (baixa

complexidade). Por outro lado, valores elevados correspondem a sequências irregulares (elevada complexidade). De forma sucinta, o valor de  $ApEn$  é inversamente proporcional à regularidade da sequência (Pincus, 1991).

Considerações relevantes sobre a medida da entropia aproximada são apresentadas em seguida.

A  $ApEn$  é uma medida utilizada para a quantificação da regularidade de um sinal. Em particular, quando o valor de  $ApEn$  é (Restrepo *et al.*, 2014):

- próximo de zero o sinal contém sequências regulares, como um sinal sinusoidal;
- elevado ( $ApEn = 2$ , para  $m = 2$ ) o sinal tem sequências independentes, como ruído branco.

Segundo a literatura, sobre dados VFC, é recomendada a seleção dos parâmetros  $m=2$  e o valor de  $r$  num alcance entre  $(0.1 - 0.25)$  (Hansen, 1995) do desvio padrão da série temporal (neste caso dos intervalos RR) e o comprimento de registro  $n$  maior que 800 batimentos, por exemplo  $n=1024$  batimentos. Uma seleção inadequada pode conduzir a resultados inconsistentes de métricas regulatórias (Molina-Picó *et al.*, 2011).

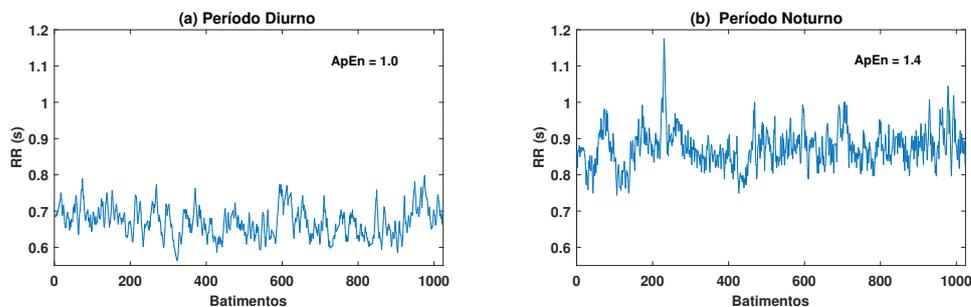
Nesta dissertação os valores dos parâmetros da  $ApEn$  utilizados, são:  $m = 2$ , tolerância,  $r = 0.2\sigma_{RR}$ , e para o conjunto de dados de indivíduos saudáveis e pacientes com patologia a dimensão dos segmentos curtos é  $n = 1024$ .

### 3.4.2 Aplicação em registros de Holter

O método não linear da entropia aproximada é aplicado em registros curtos (segmentos curtos, de 1024 batimentos, aproximadamente estacionários) e em registros longos da VFC com cerca de 24 horas (101160 batimentos).

### Análise dos registos curtos

A Figura 3.6 reflete a aplicação do método da Entropia Aproximada em dois períodos distintos, retirados de um registo de 24 horas, referentes a um indivíduo saudável. Com efeito, na Figura 3.6.(a) representa-se um tacograma (com 1024 batimentos), durante o período diurno e na Figura 3.6.(b) um tacograma, com a mesma dimensão, durante o período noturno. Verifica-se que o período diurno tem valores de  $ApEn = 1.0$ , enquanto que o período noturno apresenta valor de  $ApEn = 1.4$ . Estes resultados, demonstram uma maior complexidade no período noturno, quando comparado com o período diurno.



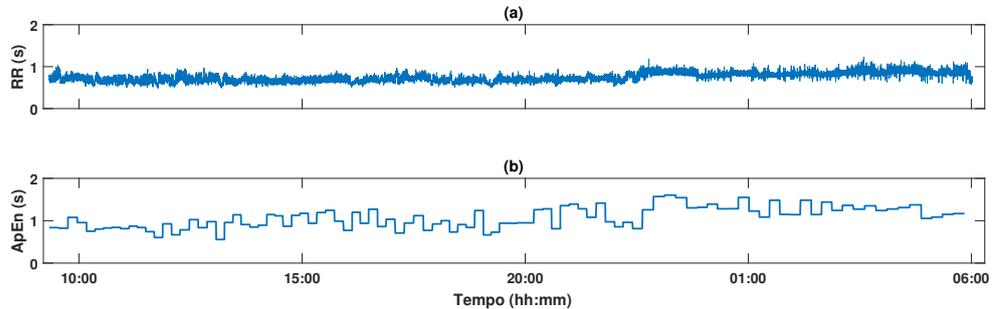
**Figura 3.6** – Tacograma de um indivíduo saudável (segmentos curtos com 1024 batimentos), durante (a) período diurno e (b) período noturno, com indicação dos correspondentes valores de entropia aproximada.

### Análise dos registos longos

A  $ApEn$  combinada com a segmentação é aplicada em registos longos. O registo longo é dividido em segmentos curtos com 1024 batimentos e é determinado o valor da  $ApEn$  em cada segmento.

Na Figura 3.7.(a) representa-se um tacograma referente a um indivíduo saudável, registo longo com 101160 batimentos e em 3.7.(b) o seguimento ao longo do tempo da entropia aproximada. A Figura 3.7 sublinha as informações retiradas da Figura 3.6: valores mais elevados durante o período noturno, o que revela regularidade

superior durante o período do dia.



**Figura 3.7** – Tacograma de um registo de Holter, de um indivíduo saudável e descrição detalhada do parâmetro  $ApEn$ , ao longo do tempo.

### 3.5 Comentários finais

Os recentes modelos ARFIMA-GARCH caracterizam estruturas conjuntas de termo curto e de termo longo e removem a memória longa do sinal, obtendo uma melhor descrição das componentes espectrais LF e HF. Introduziram a variância condicional do passado de um processo, mas também do seu próprio passado. Os segmentos curtos analisados a partir deste modelo, têm uma dimensão de batimentos igual a 1024. O método DFA combinado com a segmentação, nos registos longos, estima os expoentes de escala de memória curta, baseados em segmentos curtos, de dimensão mínima igual a 1024 e estima o expoente de escala de memória longa, baseado em segmentos curtos, de dimensão mínima igual a 4096. Esta constatação permite deduzir que o cálculo da memória longa é mais eficaz para os modelos ARFIMA-GARCH, pois analisam segmentos curtos de menor dimensão, logo mais descritivos. O método da Entropia Aproximada estuda a regularidade do sistema não linear. Valores de  $ApEn$  próximos de zero representam sinais pouco complexos e valores próximos de dois representam sinais complexos.



## Capítulo 4

### Análise experimental

A análise experimental da presente dissertação tem como objeto de estudo duas bases de dados, onde se aplicam os métodos lineares e não lineares, referidos nos capítulos anteriores. O primeiro caso refere-se à análise de sinais cardíacos de árbitros, no decorrer de partidas de futebol. São comparados os momentos determinantes do jogo com os momentos não decisivos, através dos parâmetros obtidos pela aplicação dos métodos, e são posteriormente associados aos níveis de stresse dos sujeitos. O segundo caso é baseado na análise de registos de Holter, de indivíduos saudáveis e de indivíduos com Insuficiência Cardíaca Congestiva (CHF). Depois de implementadas as metodologias estudadas, efetuou-se uma avaliação dos resultados individuais e globais obtidos, determinando o contributo dos respetivos modelos e as características de cada grupo. O Capítulo 4 apresenta as bases de dados em estudo e a descrição dos parâmetros utilizados na análise desses dados. A Secção 4.2 aborda conceitos da análise estatística, tais como: o teste de Wilcoxon-Mann-Whitney, a Análise das Componentes Principais (ACP) e a Classificação. Os resultados relativos às duas bases de dados são apresentados na Secção 4.3, assim como a discussão dos mesmos.

## 4.1 Introdução

A secção experimental deste trabalho visa a análise dos registos da VFC de dois conjuntos de dados individualizados. Num estudo primordial são analisados os registos de quatro árbitros de futebol em atividade, no decorrer de quatro partidas de futebol (110 minutos, com intervalo expetável e descontos), respetivamente. Na concretização de um segundo estudo foram coletados intervalos RR, pertencentes a dois grupos: indivíduos com ritmo cardíaco normal e indivíduos com insuficiência cardíaca congestiva. São indicadas em seguida as características dos conjuntos de dados.

**A base de dados de árbitros de futebol** inclui 4 registos de intervalos RR de árbitros semi-profissionais, do sexo masculino, com idades compreendidas entre os 25 e os 30 anos. O ritmo cardíaco dos árbitros foi medido, ao longo da partida, por um pulsómetro (Polar V800). Os momentos relevantes para o jogo são anotados quando ocorrem golos, golos *fantasma*, penáltis, não marcação de penáltis, expulsões, possíveis expulsões e protestos dos jogadores, da equipa técnica e/ou dos adeptos. Para cada registo foram identificados três momentos: os momentos decisivos (MD), os momentos menos decisivos mas importantes (MDI) e os momentos não decisivos (MND). Os dados foram revistos e corrigidos manualmente. Os registos analisados foram facultados pela Universidade de Zaragoza, Espanha.

**A base de dados de pacientes Saudáveis/Doentes** é constituída por um conjunto de 83 registos de Holter, da VFC, referentes a dois grupos de indivíduos. A hora de início dos registos não é conhecida. Os dados utilizados estão disponíveis na base de dados do MIT/BIH, da PhysioNET (<http://www.physionet.org>). Os grupos são:

- **Grupo N(*nsr2db*)**: 54 registos de intervalos RR de pacientes com ritmo sinusal normal (31 homens e 23 mulheres, com idades entre os 28.5 e os 76 anos). Os batimentos foram automaticamente analisados, com revisão e correção manual;

- **Grupo CHF (*chf2db*):** 29 registos de intervalos RR de pacientes com insuficiência cardíaca congestiva (8 homens, 2 mulheres e 19 incógnitos, com idades entre os 34 e os 79 anos). Os batimentos foram automaticamente analisados, com revisão e correção manual.

Os métodos lineares e não lineares, reportados nos capítulos anteriores, foram aplicados às bases de dados apresentadas. Com o intuito de obter uma descrição adequada dos registos longos e não estacionários recorre-se à segmentação do sinal (decomposição do registo longo em segmentos curtos, aproximadamente estacionários e com dimensão fixa). Os segmentos curtos são caracterizados, segundo os parâmetros sumariados na Tabela 4.1.

**Tabela 4.1** – Parâmetros obtidos pela aplicação dos métodos estudados, na análise experimental dos registos da VFC.

<b>Domínio do Tempo</b>	Média RR	Média do Sinal RR
	SDRR	Desvio padrão do Sinal RR
<b>AR</b>	pNN50	% intervalos RR sucessivos, que diferem mais do que 50 ms
	$LF_{AR}$	Componente espectral, para baixas frequências (0.04 – 0.15) Hz
<b>ARFIMA - GARCH</b>	$HF_{AR}$	Componente espectral, para altas frequência (0.15 – 0.4) Hz
	$LF_{AG}$	Componente espectral, para baixas frequências (0.04 – 0.15)Hz
	$HF_{AG}$	Componente espectral, para altas frequência (0.15 – 0.4) Hz
<b>GARCH</b>	$d$	Parâmetro de memória longa
	$u_1$	Parâmetro ARCH
<b>DFA</b>	$v_1$	Parâmetro GARCH
	$\alpha_{HF}$	Expoente escalar de memória curta, para $4 \leq k \leq 11$
	$\alpha_{LF}$	Expoente escalar de memória curta, para $12 \leq k \leq 32$
	$\alpha$	Expoente escalar de memória longa, para $64 \leq k \leq 1024$
	$\alpha_{HF SML}$	Expoente escalar de memória curta, para $4 \leq k \leq 11$ , de RR SML
<b>Entropia</b>	$\alpha_{LF SML}$	Expoente escalar de memória curta, para $12 \leq k \leq 32$ , de RR SML
	$ApEn$	$m = 2$ e $r = 0.2\sigma_{RR}$

Na base de dados dos árbitros de futebol, para cada momento em estudo, considera-se um intervalo de 5 minutos (2.5 minutos antes do momento selecionado e 2.5 minutos após). A escolha da dimensão do intervalo de tempo é justificada pela dimensão mínima exigida, de 512 batimentos, na obtenção das componentes AR, parâmetros ARFIMA-GARCH e expoentes de escala de memória curta. Por outro

lado, na análise da base de dados Saudáveis/Doentes utilizaram-se segmentos curtos de dimensão igual a 1024 batimentos. Na obtenção do parâmetro de memória longa  $\alpha$ , do método DFA utilizaram-se segmentos com 4096 batimentos.

## 4.2 Análise multivariada e classificação

A utilização de um determinado teste estatístico depende do tipo de hipóteses a serem testadas, como a distribuição normal e a homogeneidade da variância, das amostras em estudo. A utilização de testes não paramétricos, como o teste Wilcoxon-Mann-Whitney, é justificável quando essas hipóteses não se verificam. Este teste consiste na substituição de elementos de dois conjuntos, combinados e ordenados (do menor para o maior, independentemente do conjunto original), numa única amostra. O teste estatístico soma os números ordinais da amostra e indica qual o conjunto com maiores e menores valores. A confirmação de que dois conjuntos são originários da mesma distribuição, provém do resultado probabilístico do teste. Para valores de probabilidade superiores a 5%, as amostras provém da mesma distribuição e *a hipótese nula não é rejeitada*. Para valores de probabilidade menores ou iguais a 5%, *a hipótese nula é rejeitada* e considera-se que as amostras são originárias de diferentes distribuições (diferença estatisticamente significativa ao nível de 5%) (John e Priebe, 2007; Wijnand e van de Velde, 2000; Bucchianico, 1999). Neste estudo, os parâmetros estatisticamente significativos entre os dois grupos de pacientes são as características consideradas na análise das componentes principais e na classificação.

Para uma otimização da classificação é necessário harmonizar as escalas e amplitudes. Esta harmonização pode ser suprida pela normalização das características utilizadas, ao adaptar numericamente as escalas de duas grandezas distintas. Neste trabalho, recorreu-se às médias e desvio padrão para normalizar os valores dos parâmetros estatisticamente significativos.

### Análise das Componentes Principais

Com o objetivo de analisar estruturas de correlação, Hotelling (1933) utilizou a ACP, inicialmente utilizada em métodos computacionais. A ACP é um modelo de análise estatística multivariada que transforma linearmente, um conjunto de variáveis inicialmente correlacionadas entre si, num outro conjunto, substancialmente menor, de variáveis não correlacionadas, ou seja, as componentes principais. As componentes principais permitem reter a informação mais importante dos dados (Hongyu *et al.*, 2016; Martinez *et al.*, 2005) e, adicionalmente, podem ser utilizadas em análises, tais como, na classificação de dados.

### Algoritmos de Classificação

**Análise Linear Discriminante (LDA):** o método tradicional da LDA pode ser utilizado na redução do tamanho da amostra, bem como na respetiva classificação, ou seja, analisa as diferenças de dois conjuntos de dados, de acordo com uma ou mais características. O LDA associa termos discriminantes, consoante as combinações lineares das características independentes, a cada amostra, de modo a classificar o tipo de resposta, em relação à característica estudada. Assume que os conjuntos têm covariância idêntica (Isler *et al.*, 2015a).

**Análise Quadrática Discriminante (QDA):** conceito semelhante ao LDA, diferindo na assunção de que as distribuições de cada conjunto têm a mesma covariância. Graficamente, utiliza funções quadráticas discriminantes, que fornecem um limite curvo mais efetivo, com diferentes formas e orientações (Zhang, 1997).

**Classificador *K-Nearest Neighbour* (KNN):** prevê a classe da amostra em questão, baseando-se nos  $K$  vizinhos mais próximos. As amostras são colocadas num espaço dimensional, conforme os valores das características correspondentes usadas na classificação. A estimação dos KNN é baseada na distância entre as amostras, calculada através da *distância Euclidiana*. O objeto de interesse é classificado consoante a classe maioritária dos KNN (Isler e Kunta, 2007; Isler *et al.*, 2015a; Shahbazi e Asl, 2015).

A **Validação cruzada** permite avaliar a eficácia dos classificadores estudados. A base de dados, com  $n$  amostras, é dividida aleatoriamente em  $k$  grupos, onde  $k - 1$  grupos são utilizados para treino e o restante é utilizado para teste. O processo é treinado  $k$  vezes (Isler *et al.*, 2015a). A análise da validação cruzada, que neste estudo utiliza  $k = 10$  e  $k = n$  (ou *Leave-one-out*, LOO), considera ainda as medidas de **Accuracy** (ACC), de **Sensibilidade** (SENS) e de **Especificidade** (SPEC) descritas na Tabela 4.2.

Tabela 4.2 – Medidas de desempenho do modelo de classificação.

Medida	Sigla	Fórmula
<i>Accuracy</i>	ACC	$\left(\frac{TP+TN}{TP+FN+TN+FP}\right) 100$
Sensibilidade	SENS	$\left(\frac{TP}{TP+FN}\right) 100$
Especificidade	SPEC	$\left(\frac{TN}{TN+FP}\right) 100$
TP: Número de Verdadeiros Positivos FN: Número de Falsos Negativos TN: Número de Verdadeiros Negativos FP: Número de Falsos Positivos		

A matriz de confusão ilustrada na Tabela 4.3 relaciona o conjunto de dados a que cada amostra pertence e o conjunto de dados da previsão da pertença de cada amostra, de acordo com o método de classificação (Son *et al.*, 2012).

Tabela 4.3 – Matriz de confusão.

Matriz de Confusão	Conjunto Preditivo	
	Conjunto Real	TP
FP		TN

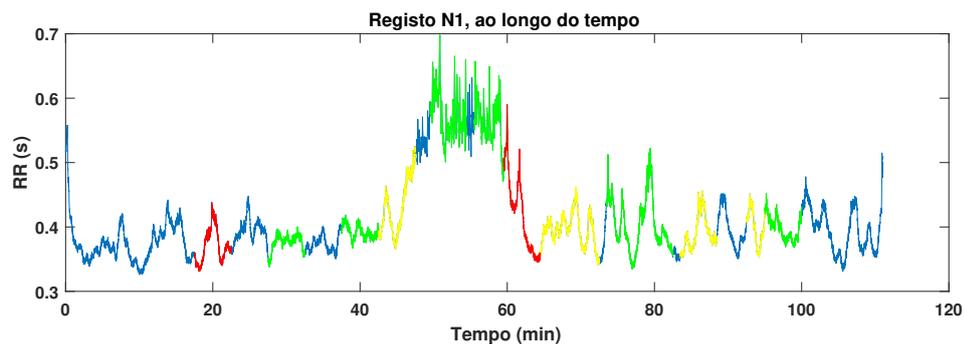
## 4.3 Resultados e discussão

Nesta secção apresenta-se a análise dos registos típicos e a análise dos resultados globais, após a aplicação dos modelos descritos nos capítulos anteriores. Relativa-

mente aos resultados globais, da aplicação dos métodos no conjunto de dados Saudáveis/Doentes, consideram-se os valores médios e respetivo desvio padrão dos parâmetros calculados, ao longo de 24 horas. Nesta base de dados, utilizou-se ainda o teste não paramétrico de Wilcoxon-Mann-Whitney, com o objetivo de obter os parâmetros estatisticamente significativos a utilizar, posteriormente, na classificação.

### 4.3.1 Base de dados dos Árbitros de Futebol

Neste trabalho são analisados, os quatro registos, de quatro partidas de futebol (com cerca de duas horas cada). Na Figura 4.1 estão representados os momentos decisivos (MD), os momentos menos decisivos mas importantes (MDI) e os momentos não decisivos (MND) do indivíduo N1, durante o jogo.



**Figura 4.1** – Registo típico de um árbitro de futebol N1. A vermelho representam-se os momentos decisivos (MD), a amarelo os momentos menos decisivos mas importantes (MDI) e a verde os momentos não decisivos (MND).

#### Segmento curto do registo típico N1

Na Figura 4.2.(a) é representado um segmento curto do indivíduo N1. O segmento tem 5 minutos, com o momento decisivo, ocorrido aos 20 min. É estimado o parâmetro  $d$  e aplicado um filtro de remoção da componente de memória longa no Sinal RR, obtendo-se o Sinal RR|SML (Figura 4.2.(c)). Posteriormente, calculam-se

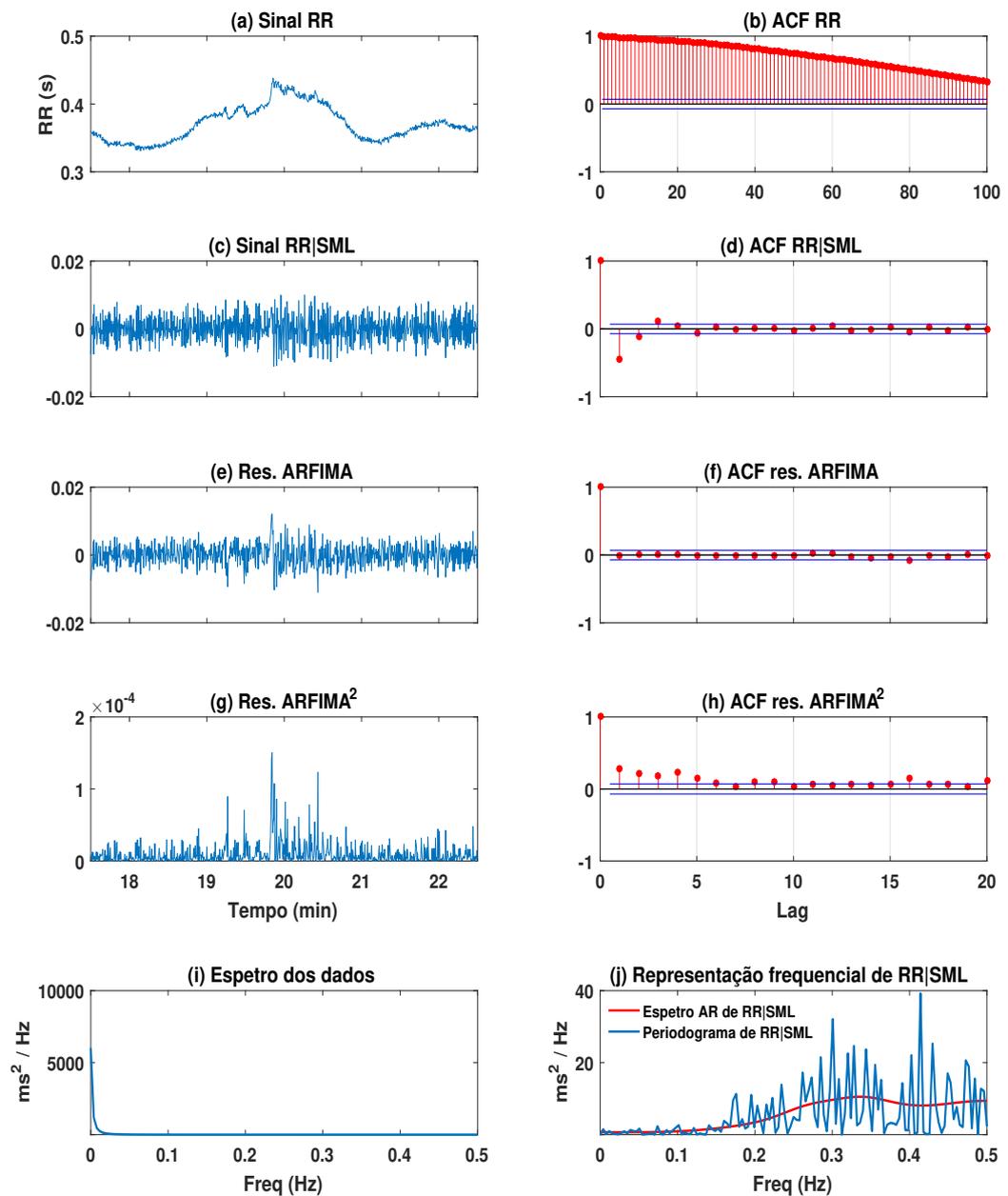
os resíduos ARFIMA( $p,d,0$ ) (Figura 4.2.(e)) e os resíduos ARFIMA<sup>2</sup>( $p,d,0$ ) (Figura 4.2.(g)).

A Função de Autocorrelação permite obter a informação sobre o grau de correlação significativo da variável, atuando como uma medida de memória do processo. O surgimento de um decaimento, lento e hiperbólico, traduz uma dependência elevada da variável de memória longa. Nas figuras 4.2.(b), 4.2.(d), 4.2.(f) e 4.2.(h) estão representadas as funções de autocorrelação, do momento exposto anteriormente, para o Sinal RR, para o Sinal RR|SML, para os Resíduos ARFIMA( $p,d,0$ ) e para os Resíduos ARFIMA<sup>2</sup>( $p,d,0$ ). Observam-se as seguintes características:

- Existência de elevada memória longa, no Sinal RR, já que para 100 *lags* ainda se verifica uma correlação da variável;
- Ao aplicar o filtro de remoção de memória longa ao Sinal RR, constata-se uma diminuição de correlação (apresenta correlação de termo curto, memória curta);
- É visível a inexistência de qualquer tipo de dependência ao longo dos resíduos, sugerindo que o modelo ARFIMA é adequado;
- Inversamente aos resíduos, os resíduos<sup>2</sup> apresentam correlação, pressupondo que a aplicação do modelo GARCH no presente caso é oportuna.

Para validar os resultados temporais acima descritos apresentam-se, nas figuras 4.2.(i) e 4.2.(j), os resultados no domínio da frequência: o espectro dos dados, o espectro paramétrico AR e o periodograma referentes ao segmento curto. Observa-se:

- Um pico elevado nas frequências muito baixas (característica típica de sinais com memória longa), no espectro dos dados;
- O espectro AR e o periodograma, do sinal após a remoção da memória longa, apresentam as componentes LF e HF com mais clareza.



**Figura 4.2** – Momento decisivo do Registo N1, aos 20 minutos. Segmento curto de cinco minutos para: (a) o Sinal RR, (c) o Sinal RR|SML, (e) os Resíduos ARFIMA( $p,d,0$ ) e (g) os Resíduos ARFIMA<sup>2</sup>( $p,d,0$ ). Função de autocorrelação para: (b) o Sinal RR, (d) o Sinal RR|SML, (f) os Resíduos ARFIMA( $p,d,0$ ) e (h) os Resíduos ARFIMA<sup>2</sup>( $p,d,0$ ). (i) Espetro dos dados e (j) espectro paramétrico e não paramétrico.

## Resultados do registo típico N1

Em ambiente Matlab foram obtidos os parâmetros espectrais da modelação AR, os parâmetros da modelação ARFIMA-GARCH, os expoentes do método DFA e a entropia aproximada, para os segmentos curtos com cinco minutos de duração. Na Tabela 4.4 estão representados os parâmetros correspondentes a cada momento MD, MDI e MND, do indivíduo N1.

**Tabela 4.4** – Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N1.

Registo N1	min	Sinal RR											Sinal RR SML	
		LF <sub>AR</sub> (ms <sup>2</sup> )	HF <sub>AR</sub> (ms <sup>2</sup> )	LF <sub>AG</sub> (ms <sup>2</sup> )	HF <sub>AG</sub> (ms <sup>2</sup> )	$d$	$u_1$	$v_1$	$\alpha_{LF}$	$\alpha_{HF}$	$\alpha$	$ApEn$	$\alpha_{LF}$	$\alpha_{HF}$
MD	20	27	4	0.3	1	1.18	0.23	0.53	1.44	0.94	1.68	0.34	0.18	0.23
	62	185	38	3	6	1.09	0.31	0.45	1.54	1.61	1.93	0.14	0.36	0.47
MDI	45	122	31	2	4	1.08	0.18	0.55	1.34	1.25	1.95	0.27	0.39	0.43
	65	29	6	0.1	1	1.44	0.00	0.00	1.57	0.92	1.19	0.60	0.05	0.16
	70	55	10	0.3	1	1.29	0.10	0.82	1.58	1.15	1.52	0.39	0.15	0.20
	86	48	10	1	2	1.21	0.23	0.44	1.25	1.32	1.49	0.35	0.31	0.29
	95	36	7	1	2	1.16	0.05	0.94	1.62	1.14	1.2	0.53	0.21	0.29
MND	30	23	4	0.3	1	1.10	0.00	0.00	1.51	0.87	1.78	0.82	0.16	0.29
	40	19	3	1	1	0.89	0.00	0.00	1.38	0.98	1.47	0.90	0.47	0.18
	52	644	39	209	30	0.46	0.05	0.90	1.13	1.79	1.50	0.68	0.88	0.29
	58	478	61	47	24	0.79	0.19	0.59	1.41	1.69	0.97	0.53	0.88	1.49
	76	115	16	3	4	0.97	0.34	0.29	1.74	1.52	1.50	0.27	0.56	1.17
	80	127	21	1	3	1.11	0.08	0.91	1.30	1.58	1.50	0.29	0.59	0.32
	97.5	40	9	1	2	1.03	0.10	0.17	1.57	1.08	1.47	0.68	0.25	0.31

É possível verificar que:

- Os valores de frequência para o Sinal RR são menores do que o previsto. Tendo em conta o estudo executado anteriormente, estes resultados seguem a apologia de que a média do Sinal RR é demasiado baixa e os valores do parâmetro de memória longa  $d$  são superiores ao habitual;
- Os valores do parâmetro  $d$ , são mais elevados nos MD do que nos MND, podendo relacionar-se com a influência do stresse no ritmo cardíaco;
- No final do jogo o parâmetro LF diminui e o valor mais elevado, durante o jogo, ocorre aos 62 minutos, para MD. Os valores máximos apresentados ocorrem no intervalo para LF e também para HF;

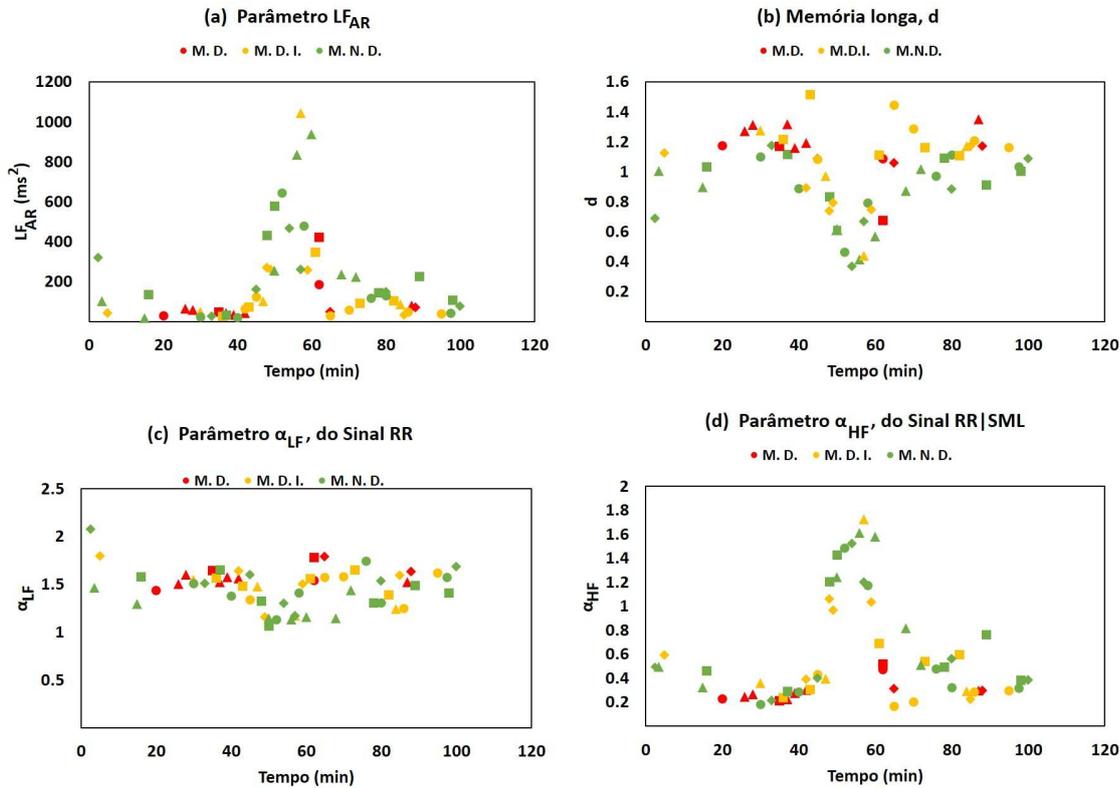
- Em concordância com o parâmetro  $d$ , o expoente de escala  $\alpha$  apresenta valores mais baixos durante o intervalo. O valor mais elevado ocorre para MDI (apesar de ser o resultado previsto, a dimensão do registo utilizado é baixa, para uma obtenção fiável de  $\alpha$ );
- O parâmetro ARCH apresenta, em geral, valores mais elevados para os MD;
- Os expoentes escalares de memória curta do Sinal RR e do Sinal RR|SML são mais elevados para os MND, na segunda parte do jogo;
- Nos MND o árbitro em análise apresenta valores superiores de  $ApEn$  em relação aos MD e aos MDI, na primeira parte do jogo.

### Resultados globais

Os resultados globais, esquematizados na Figura 4.3, apresentam alguns parâmetros dos métodos lineares e não lineares, descritos anteriormente. Para uma melhor interpretação agrupam-se os valores dos parâmetros obtidos, para os vários momentos da partida, de todos os registos (caraterizados por símbolos diferentes).

Da interpretação da figura, retiram-se as seguintes conclusões:

- A potência de baixa frequência está associada à estimulação do SNS e do SNP, como referido anteriormente. Tendo em conta os resultados globais constata-se que as decisões mais importantes tomadas pelo árbitro geram uma alteração no SNA, já que, durante o jogo e para o mesmo árbitro, os valores de MD são superiores aos de MND, especialmente na primeira parte do jogo, num mesmo intervalo de tempo. O indivíduo 4 apresenta a componente LF mais elevada do gráfico, para um MDI;
- No parâmetro  $d$  é visível uma divergência entre os MD e os MND. Não há uma diferença significativa entre os MD e os MDI, porém os MND apresentam valores do parâmetro  $d$  mais baixos do que os restantes, para o mesmo intervalo de tempo. É possível notar uma diminuição dos valores do parâmetro  $d$ , entre



**Figura 4.3** – Resultados dos parâmetros estudados, para todas as amostras dos quatro árbitros. A representação simbólica está relacionada o com respectivo árbitro: Círculo - Registo 1, Quadrado - Registo 2, Losango - Registo 3 e Triângulo - Registo 4. As cores representam os três momentos (MD, MDI e MND) e seguem o mesmo padrão utilizado no estudo de um registo típico. (a) Componente espectral  $LF_{AR}$ , (b) parâmetro de memória longa,  $d$ , (c) expoente de escala, do Sinal RR,  $\alpha_{LF|RR}$  e (d) expoente de escala, do Sinal RR, com ausência de memória longa,  $\alpha_{HF|SML}$ .

os instantes 45 min e os 57 min. Isso deve-se ao início e fim do intervalo, obtendo-se o valor mínimo de  $d = 0.4$ . Independentemente do tipo de momento, durante o período de atividade os valores de  $d$  são mais elevados do que durante o período de descanso. Adicionalmente, no parâmetro  $\alpha_{LF|RR}$  é visível uma pequena diminuição durante o intervalo (inferior a 1.5), relativamente ao jogo;

- De um ponto de vista geral, o parâmetro  $\alpha_{LF|RR}$  nos MD apresenta valores superiores a 1.5. Por outro lado, os MND têm valores menores e os MDI

situam-se entre os valores de MD e os de MND;

- O expoente  $\alpha_{HF|RR}$  apresenta valores menores para os MD, quando comparados com os valores dos MND. Os valores são mais baixos durante a partida, ocorrendo um crescimento, durante o intervalo, e um decréscimo no início da segunda parte;

Os parâmetros  $HF_{AR}$ ,  $LF_{AG}$ ,  $HF_{AG}$ , parâmetro ARCH, parâmetro GARCH,  $\alpha_{HF|RR}$ ,  $\alpha$ ,  $\alpha_{LF|SML}$  e  $ApEn$  podem ser objeto de consulta, na secção Anexo.

### 4.3.2 Base de dados de Saudáveis/Doentes

#### Registos típicos dos grupos N e CHF

Nesta secção apresenta-se o comportamento detalhado, dos vários parâmetros, ao longo de aproximadamente 24 horas, para os indivíduos saudáveis e com patologia.

Com efeito, é considerado um registo típico de um paciente saudável (Figura 4.4.(a) homem de 59 anos, com 103914 batimentos, N21) e um registo típico de um paciente com CHF (Figura 4.4.(e) homem de 55 anos, com 112310 batimentos, CHF1). Na Figura 4.4 representam-se também os vários parâmetros obtidos no domínio do tempo. A análise da figura permite retirar as seguintes conclusões:

- A média do intervalo RR e o parâmetro SDRR apresentam valores superiores e maior variabilidade, para o indivíduo saudável;
- Comparando ambos os casos, verifica-se que o parâmetro pNN50 é consideravelmente mais baixo para o paciente doente;
- No paciente saudável, os valores mais elevados dos parâmetros do domínio do tempo coincidem, ao longo do tempo (podendo esse ser o período do sono. Como verificado no Capítulo 2 e em (Almeida, 2015), os valores são superiores no período noturno).

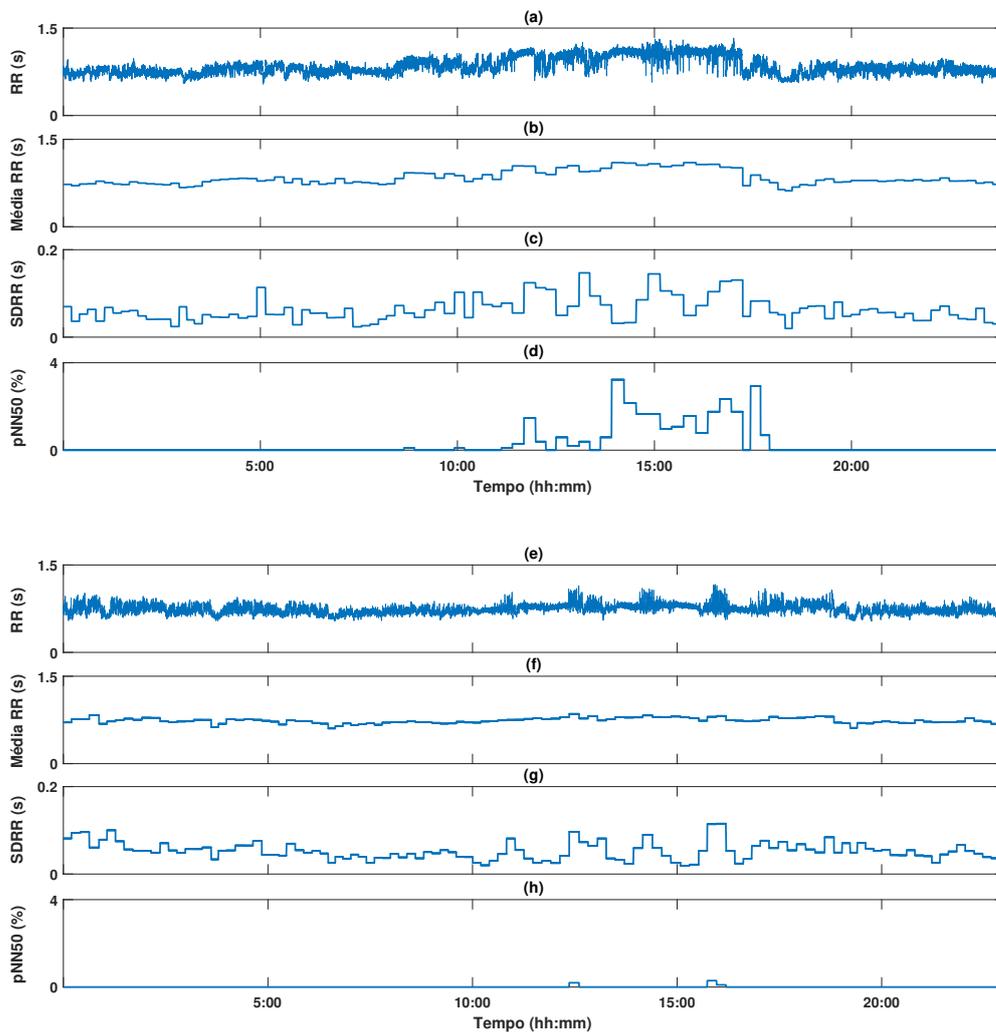
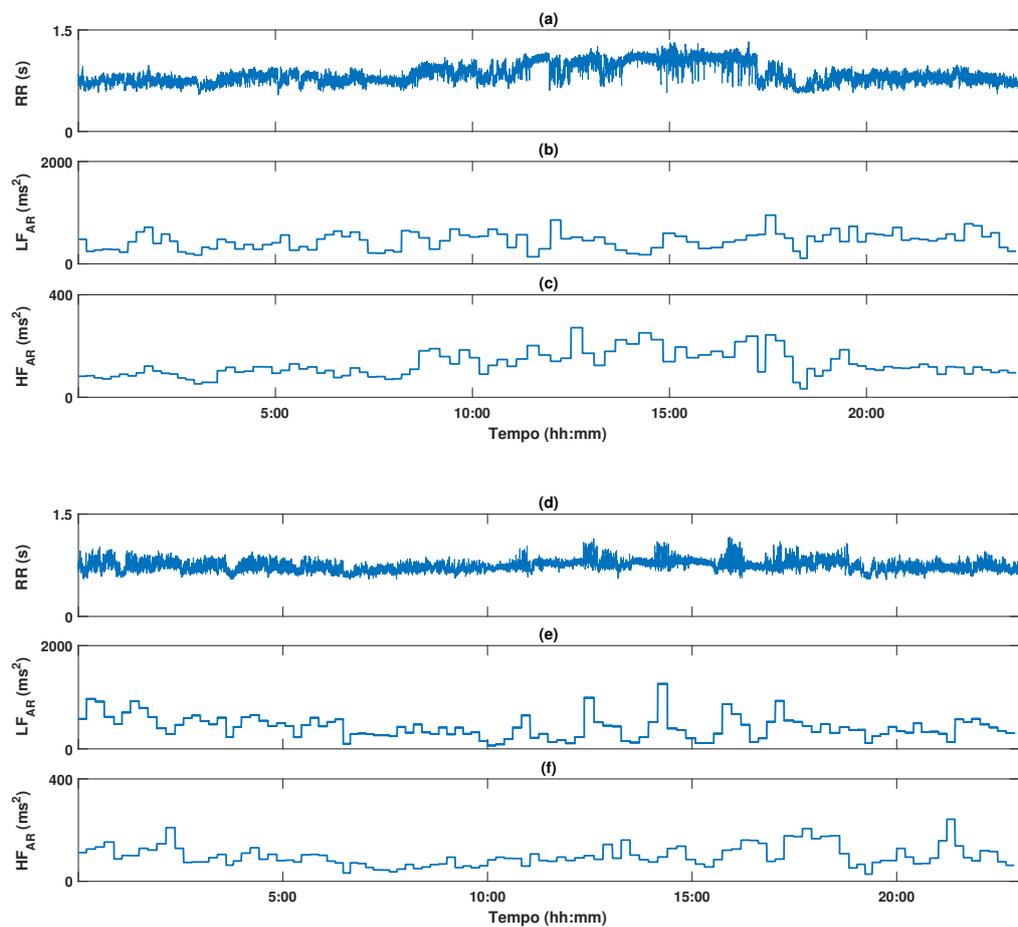


Figura 4.4 – Representação do Sinal RR e dos parâmetros, no domínio do tempo, dos registos de Holter de (a)-(d) um indivíduo saudável N21 e de (e)-(h) um indivíduo doente CHF1.

Na Figura 4.5 representam-se as componentes espectrais, da modelação AR aplicadas nos dois casos típicos. É possível verificar que:

- A diferença do parâmetro LF no N21 e no CHF1 não é nítida visivelmente, ao contrário do parâmetro HF que é superior para N21. Através do cálculo do valor médio de LF constata-se que o parâmetro é ligeiramente superior para N21 ( $452.5 \text{ ms}^2$ ), quando comparado com CHF1 ( $417.8 \text{ ms}^2$ );



**Figura 4.5** – Representação do Sinal RR e obtenção das componentes espectrais, da modelação AR, ao longo do tempo, para (a)-(c) um indivíduo saudável N21 e para (d)-(f) um indivíduo doente CHF1.

- Os valores mais elevados de HF, para N21, ocorrem em simultâneo com os valores mais elevados do Sinal RR (como referido anteriormente pode estar relacionado com a hora do sono). Denota-se ainda bastante variabilidade nesta componente.

Da aplicação da modelação ARFIMA-GARCH, nos registos típicos, obtiveram-se os resultados apresentados na Figura 4.6.

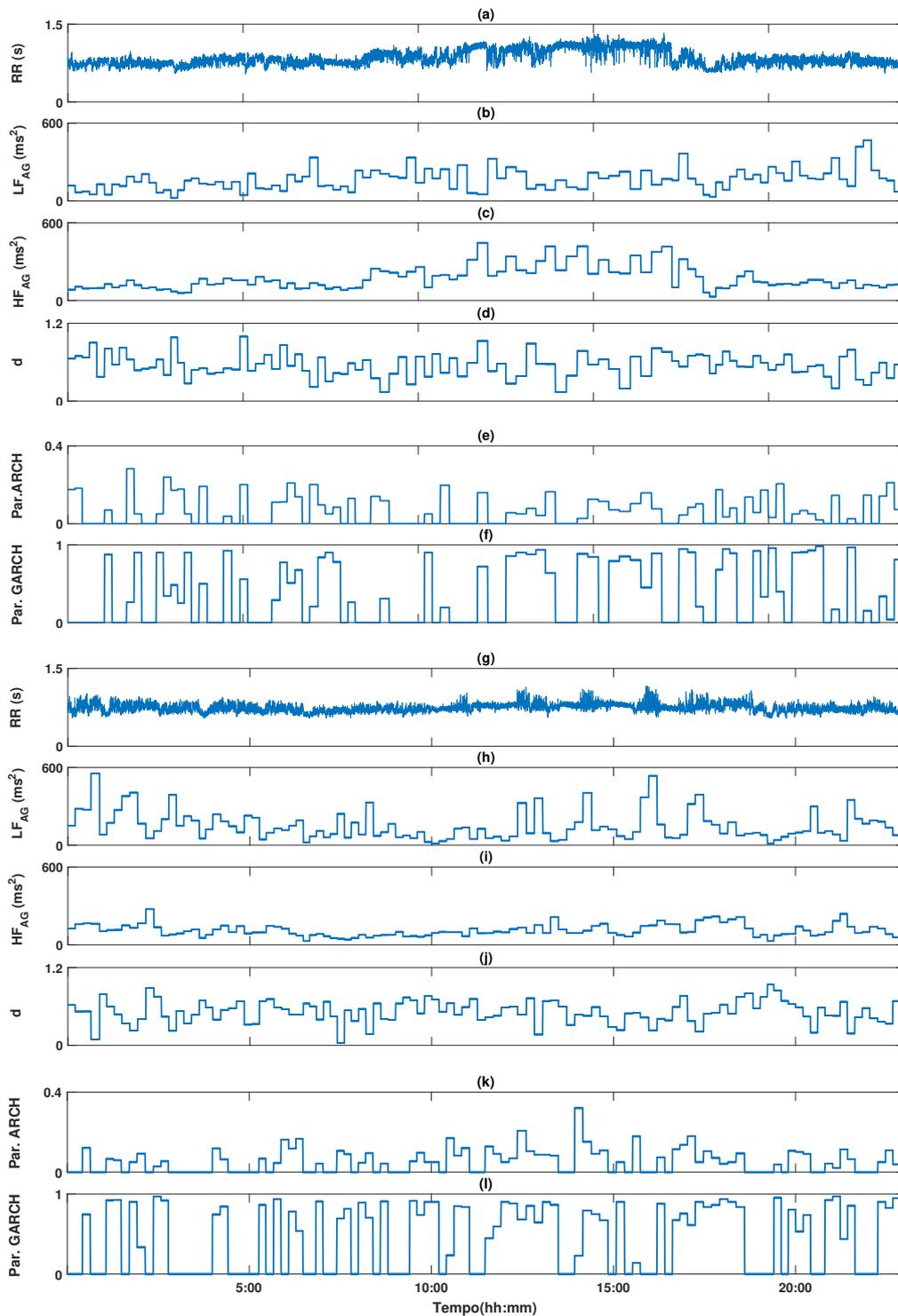


Figura 4.6 – Descrição do Sinal RR e do comportamento dos parâmetros ARFIMA-GARCH, de (a)-(f) um sujeito saudável N21 e de (g)-(l) um paciente doente CHF1, ao longo do tempo.

Verifica-se que:

- No que diz respeito às componentes espectrais LF e HF, estas apresentam valores mais elevados para o indivíduo saudável;
- Os valores do parâmetro de memória longa  $d$ , variam ao longo do tempo, indicando características de multifractalidade. O valor médio de  $d$  é 0.55 para o registo N21 e 0.52 para o registo CHF1;
- Os parâmetros ARCH e GARCH apresentam valores médios mais elevados para o paciente doente. Para o indivíduo saudável os valores são, respetivamente, 0.033 e 0.195 e para o indivíduo doente são 0.053 e 0.441.

O método DFA foi aplicado em dois sinais, de cada registo típico: ao Sinal RR e ao sinal após a remoção da memória longa, Sinal RR|SML. Os resultados obtidos estão representados, respetivamente, nas figuras 4.7 e 4.8.

A partir das figuras pode observar-se que:

- Ao conjugar a informação relativa aos parâmetros de memória curta,  $\alpha_{LF}$  e  $\alpha_{HF}$  (obtidos ao longo das 24 horas) verifica-se que o Sinal RR e o Sinal RR|SML apresentam valores mais elevados para CHF1;
- O parâmetro de memória longa  $\alpha$ , varia ao longo do tempo e apresenta valores superiores para N21. Estes resultados entram em concordância com os obtidos para o parâmetro  $d$ .

A aplicação do método não linear, entropia aproximada, nos registos do indivíduo saudável e do indivíduo doente são descritos na Figura 4.9. Comparando os dois registos, constata-se que o parâmetro  $ApEn$  não é distinguível visivelmente. Em concordância, os valores médios de  $ApEn$  são bastantes próximos: 1.00 para N21 e 0.92 para CHF1.

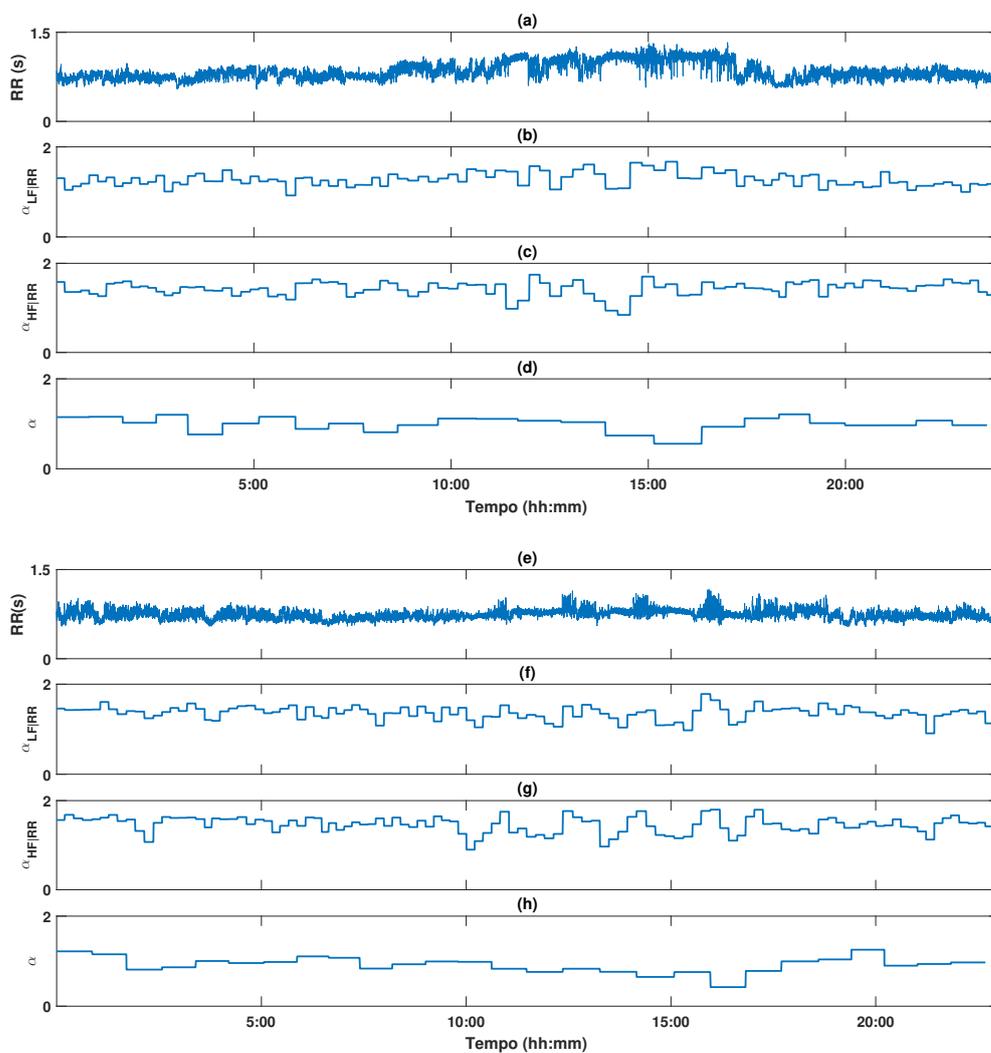
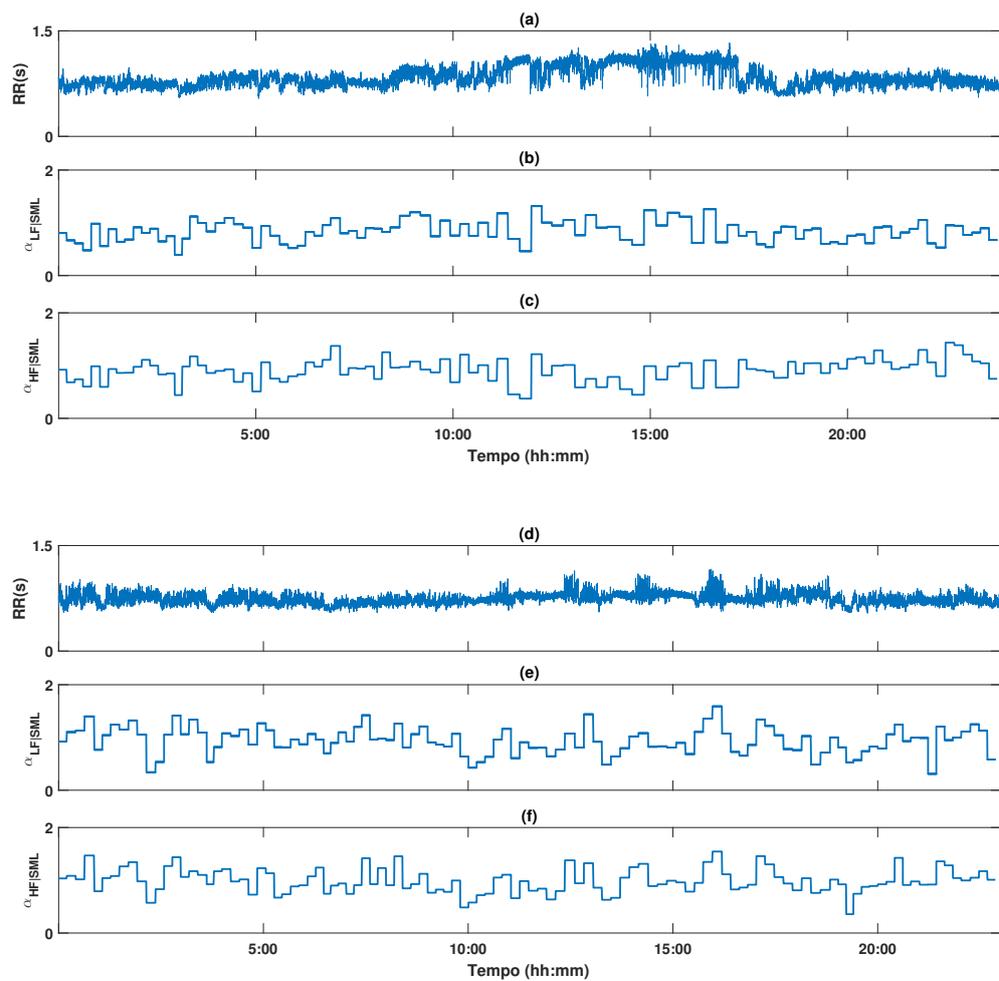


Figura 4.7 – Representação do Sinal RR e obtenção dos expoentes escalares, ao aplicar o método DFA, para o referido sinal, de (a)-(d) um paciente saudável N21 e de (e)-(h) um paciente doente CHF1.



**Figura 4.8** – Caracterização do Sinal RR|SML e dos expoentes escalares do Sinal RR|SML, dos dois registos típicos estudados: (a)-(c) indivíduo saudável N21 e (d)-(f) indivíduo com patologia CHF1.

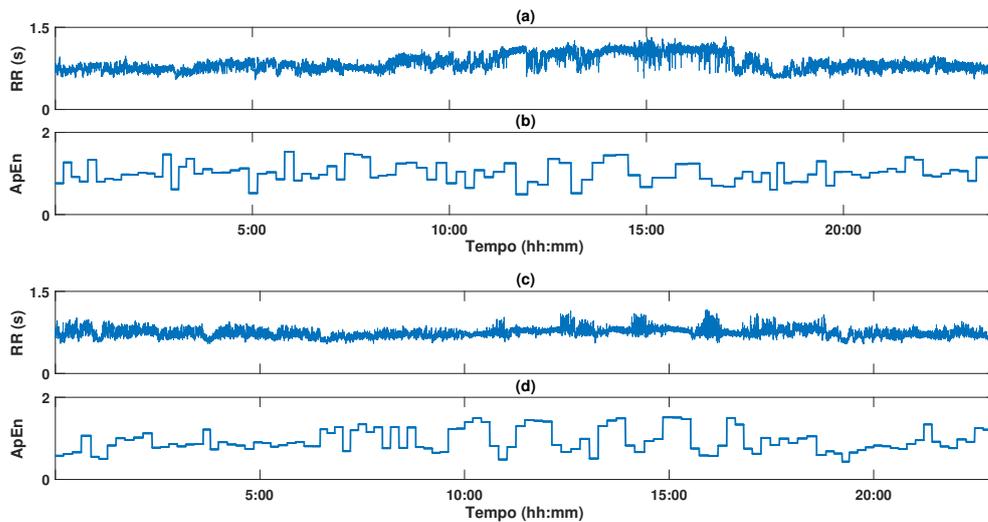


Figura 4.9 – Apresentação dos resultados da aplicação da Entropia Aproximada, para um registo de (a) e (b) um sujeito saudável N21 e de (c) e (d) um sujeito doente CHF1, ao longo de 24h.

### Resultados Globais dos grupos N e CHF

Descentralizando os registos típicos e abrangendo os registos a um estudo global, obtiveram-se os valores médios e o respetivo desvio padrão para os vários parâmetros (Figura 4.10 e Tabela 4.5). O objetivo passa por caracterizar e discriminar o grupo de 54 indivíduos saudáveis e o grupo de 29 pacientes doentes, ao longo de 24 horas. Um estudo adicional baseia-se na análise da VFC, em função do fator idade e do fator género. Avaliando os valores médios dos parâmetros aplicados nos dois grupos de indivíduos verifica-se que:

- O parâmetro no domínio do tempo SDRR, o parâmetro de memória longa  $d$ , o parâmetro ARCH, o parâmetro GARCH, o expoente de escala  $\alpha_{HF|RR}$  e o parâmetro  $ApEn$  apresentam valores mais elevados para o grupo dos indivíduos doentes. O resultado do parâmetro  $d$  está de acordo com [Baillie \*et al.\* \(2009\)](#) e [Rocha \*et al.\* \(2016\)](#);
- No grupo dos indivíduos saudáveis existe uma discrepância nas componentes espectrais de baixa frequência,  $LF_{AR}$  e  $LF_{AG}$ , relativamente ao grupo dos

**Tabela 4.5** – Estimativa global (média  $\pm$  desvio padrão) dos parâmetros obtidos pela aplicação dos métodos, relativamente aos grupos de indivíduos saudáveis e indivíduos doentes. Utilização do teste Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos para os dois grupos.

Metodologias	Parâmetros	Saudáveis	Doentes	p-value
	N.º de Registos	54	29	
	Idade	61.4 $\pm$ 11.6	55.3 $\pm$ 11.6	
<b>Domínio do Tempo</b>	Média RR (s)	0.779 $\pm$ 0.078	0.688 $\pm$ 0.086	< <b>0.001</b>
	SDRR (s)	0.138 $\pm$ 0.030	0.224 $\pm$ 0.800	< <b>0.001</b>
	pNN50 (%)	0.430 $\pm$ 0.820	0.150 $\pm$ 0.350	<b>0.001</b>
<b>AR</b>	LF <sub>AR</sub> (ms <sup>2</sup> )	497 $\pm$ 262	210 $\pm$ 238	< <b>0.001</b>
	HF <sub>AR</sub> (ms <sup>2</sup> )	101 $\pm$ 55	56 $\pm$ 39	< <b>0.001</b>
<b>ARFIMA</b>	LF <sub>AG</sub> (ms <sup>2</sup> )	215 $\pm$ 129	82 $\pm$ 111	< <b>0.001</b>
	HF <sub>AG</sub> (ms <sup>2</sup> )	118 $\pm$ 72	63 $\pm$ 46	< <b>0.001</b>
-	d	0.546 $\pm$ 0.063	0.607 $\pm$ 0.128	<b>0.033</b>
<b>GARCH</b>	$u_1$	0.077 $\pm$ 0.027	0.081 $\pm$ 0.055	0.500
	$v_1$	0.351 $\pm$ 0.078	0.418 $\pm$ 0.094	<b>0.001</b>
<b>DFA</b>	$\alpha_{LF RR}$	1.462 $\pm$ 0.126	1.160 $\pm$ 0.284	< <b>0.001</b>
	$\alpha_{HF RR}$	1.238 $\pm$ 0.059	1.301 $\pm$ 0.123	<b>0.001</b>
	$\alpha$	1.047 $\pm$ 0.067	1.028 $\pm$ 0.110	0.468
	$\alpha_{LF SML}$	0.809 $\pm$ 0.066	0.705 $\pm$ 0.255	0.093
	$\alpha_{HF SML}$	0.990 $\pm$ 0.199	0.670 $\pm$ 0.279	< <b>0.001</b>
<b>Entropia</b>	$ApEn$	0.980 $\pm$ 0.105	1.068 $\pm$ 0.120	<b>0.001</b>

indivíduos doentes, que apresenta valores mais baixos. As componentes espectrais de alta frequência, HF<sub>AR</sub> e HF<sub>AG</sub> também são mais elevados no grupo dos indivíduos saudáveis. Estes resultados corroboram outros estudos como os de [Leite e Rocha \(2013\)](#) e de [Almeida \(2015\)](#);

- Os valores obtidos para o expoente de escala,  $u_1$ ,  $v_1$  e  $\alpha$ , não são os previstos pela literatura, podendo estar relacionado com fatores externos, como a idade ou o género;
- De acordo com [Leite \(2008\)](#),  $\alpha_{HF|SML}$  apresenta valores mais elevados no grupo dos indivíduos saudáveis. O mesmo comportamento ocorre para este caso nos expoentes,  $\alpha_{LF|RR}$ ,  $\alpha_{LF|SML}$  e  $\alpha_{HF|SML}$ .

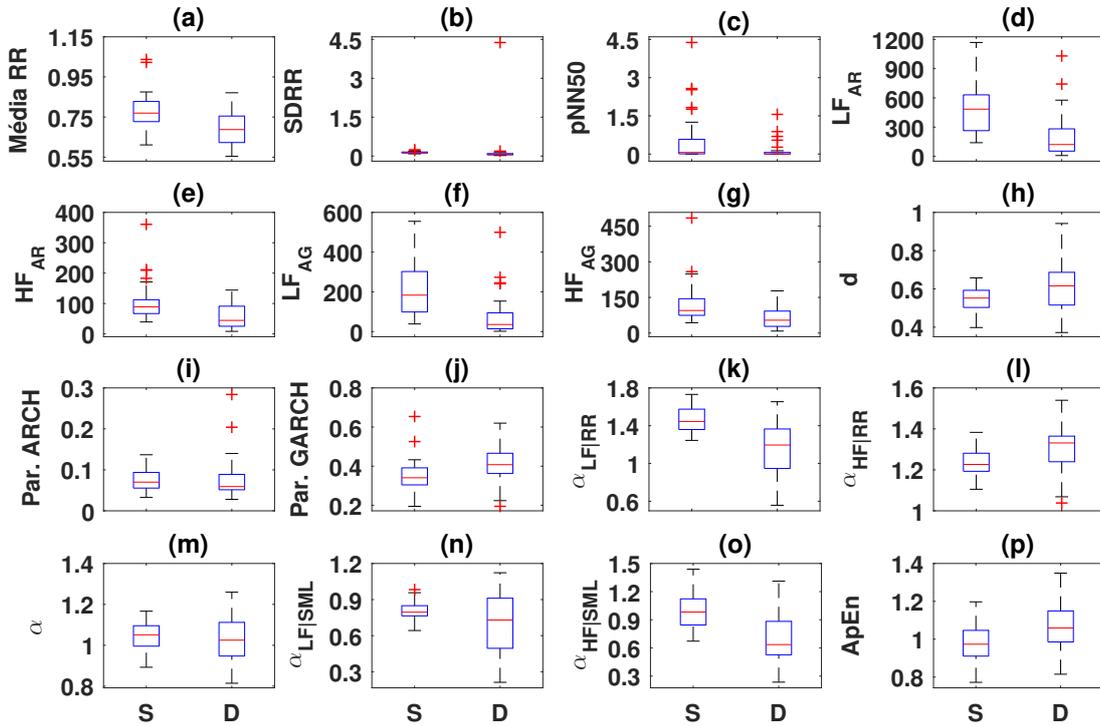


Figura 4.10 – Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos Saudáveis (S) e indivíduos Doentes (D).

Tendo ainda em atenção, os registos típicos representados entre as figuras 4.4 e 4.9 e os valores da Tabela 4.5 e Figura 4.10, observam-se as seguintes características:

- Concordância relativa aos valores mais elevados do parâmetro  $pNN50$ , das componentes espectrais, da modelação AR e da modelação ARFIMA-GARCH para o grupo de indivíduos saudáveis;
- O registo do indivíduo N21 apresenta o valor do parâmetro  $d$  e  $ApEn$  mais elevado do que o indivíduo CHF1. Adicionalmente, os valores de  $\alpha_{LF|RR}$  e  $\alpha_{HF|SML}$  apresentam valores mais elevados para CHF1, tendências não seguidas pela tabela;
- Concordância relativa aos valores mais elevados dos expoentes  $\alpha$  e  $\alpha_{HF|RR}$  para o grupo de indivíduos doentes.

A utilização do teste de Wilcoxon-Mann-Whitney permite analisar as diferenças estatísticas entre o conjunto de dados dos indivíduos saudáveis e dos indivíduos doentes, ao longo de 24 horas. Os parâmetros que diferem nos dois grupos, ao nível de significância de 5% são nomeadamente, os do domínio do tempo Média RR, SDRR e pNN50, as componentes espectrais  $LF_{AR}$ ,  $HF_{AR}$ ,  $LF_{AG}$  e  $HF_{AG}$ , o parâmetro  $d$ , o parâmetro GARCH, os expoentes de escala  $\alpha_{LF|RR}$ ,  $\alpha_{HF|RR}$  e  $\alpha_{HF|SML}$  e a  $ApEn$ .

### Classificação

A Classificação permite identificar a que conjunto de dados pertence a amostra, considerando as características conhecidas. Neste estudo utilizam-se os parâmetros que apresentam diferenças estatisticamente significativas entre os grupos, determinados através do teste de Wilcoxon-Mann-Whitney: Média RR, SDRR, pNN50,  $LF_{AR}$ ,  $HF_{AR}$ ,  $LF_{AG}$ ,  $HF_{AG}$ ,  $d$ , parâmetro GARCH,  $\alpha_{LF|RR}$ ,  $\alpha_{HF|RR}$ ,  $\alpha_{HF|SML}$  e  $ApEn$ . Antes da aplicação da ACP e dos três classificadores (LDA, QDA, KNN), os parâmetros são sujeitos a uma normalização, pelas médias e desvio padrão. Na Figura 4.11 é representada a aplicação da ACP, nos 13 parâmetros estatisticamente significativos. Observa-se que as primeiras 5 componentes principais apresentam uma boa discriminação dos dois grupos e explicam 86% do total da variância. Por outro lado, a primeira componente considera 46%, as 3 primeiras 71% e as sete primeiras explicam 96%, do total da variância. Neste estudo são utilizadas 3 e 5 componentes principais (ACP<sub>3</sub> e ACP<sub>5</sub>), para explicar o total da variabilidade das características.

A validação cruzada verifica o desempenho de cada classificador, recorrendo à matriz de confusão obtida. Os classificadores são aplicados nos dados normalizados: com todos os parâmetros (TPN) e com 3 e 5 componentes principais (pela ACP). Na validação cruzada são utilizados os valores  $k = 10$  e  $k = \text{LOO}$ .

A aplicação dos classificadores em todos os parâmetros e nas componentes principais, é apresentada na Tabela 4.6, juntamente com a validação dos valores obtidos para cada classificador. A *accuracy*, a sensibilidade e a especificidade, representadas na Tabela 4.2, são três das medidas que avaliam o desempenho do classificador. Da interpretação da Tabela 4.6 é possível retirar as seguintes conclusões:

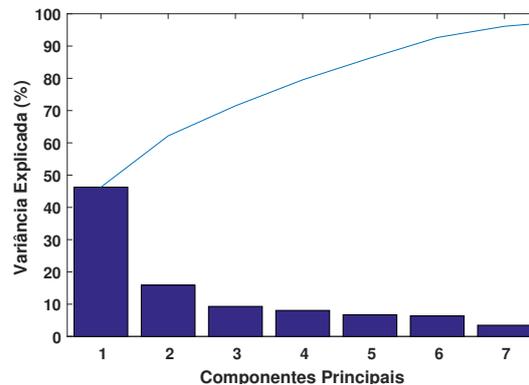


Figura 4.11 – Gráfico com 7 componentes principais.

- De um modo geral, o classificador que apresenta melhores resultados, é o classificador QDA, aplicado em todos os parâmetros estatisticamente significativos normalizados. A validação deste tipo de classificador é a única que atinge a casa dos 90%, para LOO;
- Analisando as medidas *accuracy*, sensibilidade e especificidade, obtidas para cada classificador, observa-se que os valores de *accuracy*, considerando 3 e 5 componentes principais são superiores para o classificador KNN;
- O classificador KNN apresenta ainda valores de especificidade máximos, porém falha na sensibilidade, com valores muito inferiores, quando comparados com os dos restantes classificadores. Este último pressuposto sublinha o estudo de [Isler et al. \(2015a\)](#);
- Todos os parâmetros normalizados, para os classificadores LDA e QDA, discriminam melhor as amostras entre os dois grupos, quando comparados com as ACP<sub>3</sub> e ACP<sub>5</sub>. Na validação cruzada quando utilizada  $k = 10$  e  $k = \text{LOO}$  os valores resultantes são relativamente próximos e geralmente superiores para LOO.

Como descrito no capítulo introdutório desta dissertação existem determinados fatores que induzem a VFC, nomeadamente a idade e o género de um sujeito. Com

**Tabela 4.6** – Desempenho de 3 classificadores para todos os parâmetros estatisticamente significativos normalizados e para as componentes principais, através das medidas de *accuracy*, sensibilidade, especificidade e do método da validação cruzada, para  $k = 10$  e  $k = \text{LOO}$ .

Modelos		ACC (%)	SENS (%)	SPEC (%)	Val. Cruzada (%)	
LDA	ACP <sub>3</sub>	84.3	75.7	88.9	k=10	82.4 ± 0.8
					LOO	81.9
	ACP <sub>5</sub>	89.2	75.9	96.3	k=10	86.5 ± 1.2
					LOO	86.8
	TPN	91.2	79.3	98.2	k=10	84.7 ± 1.4
					LOO	84.3
QDA	ACP <sub>3</sub>	85.5	79.3	88.9	k=10	82.4 ± 1.4
					LOO	83.1
	ACP <sub>5</sub>	89.2	75.9	100	k=10	87.3 ± 1.4
					LOO	86.8
	TPN	92.8	82.8	98.2	k=10	90.2 ± 1.1
					LOO	90.4
KNN	ACP <sub>3</sub>	86.8	62.1	100	k=10	85.6 ± 1.2
					LOO	85.5
	ACP <sub>5</sub>	91.6	69.0	100	k=10	87.4 ± 1.2
					LOO	88.0
	TPN	85.5	58.6	100	k=10	85.3 ± 0.7
					LOO	85.5

o intuito de averiguar de que forma estes fatores influenciam a VFC, são calculados os parâmetros dos métodos estudados ao longo da dissertação, para a base de dados de indivíduos saudáveis (54 registos).

### Análise da VFC em função da idade

A idade é um fator que influencia o funcionamento do sistema cardiovascular. Para determinar a influência que detém na VFC, são analisados registos de indivíduos saudáveis adultos e indivíduos saudáveis idosos. A Figura 4.12 e a Tabela 4.7 apresentam os resultados globais (média ± desvio padrão) dos parâmetros dos métodos aplicados, para indivíduos com idades entre os 28.5 e os 64 anos (adultos) e com idade superior a 65 anos (idosos).

**Tabela 4.7** – Estimativa global (média  $\pm$  desvio padrão) dos parâmetros obtidos através da aplicação dos métodos, referentes aos grupos de indivíduos saudáveis adultos e indivíduos saudáveis idosos. Utilização do teste estatístico Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos, para os dois grupos.

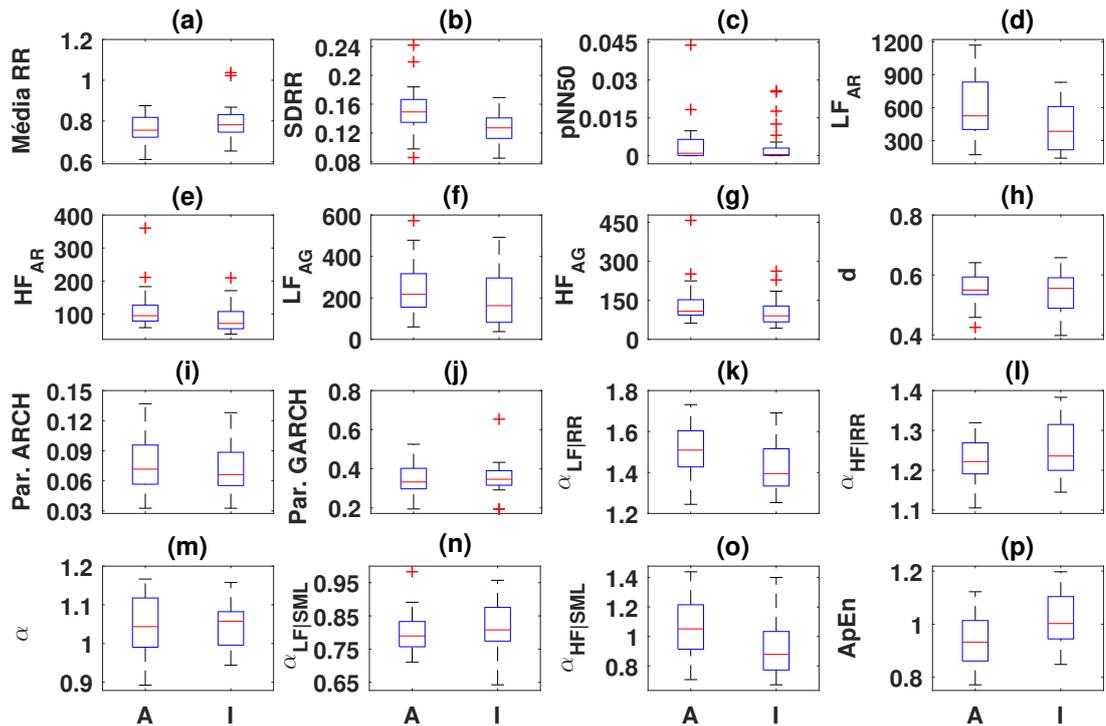
Metodologias	Idade	Adultos	Idosos	p-value
	N.º de Registos	26	28	
	Idades (anos)	54.1 $\pm$ 13.0	68.1 $\pm$ 3.6	
Domínio do Tempo	Média RR (s)	0.762 $\pm$ 0.069	0.688 $\pm$ 0.083	0.194
	SDRR (s)	0.151 $\pm$ 0.034	0.127 $\pm$ 0.021	<b>0.003</b>
	pNN50 (%)	0.48 $\pm$ 0.91	0.150 $\pm$ 0.74	0.337
AR	LF <sub>AR</sub> (ms <sup>2</sup> )	589 $\pm$ 281	210 $\pm$ 213	<b>0.022</b>
	HF <sub>AR</sub> (ms <sup>2</sup> )	116 $\pm$ 63	87 $\pm$ 43	<b>0.027</b>
ARFIMA	LF <sub>AG</sub> (ms <sup>2</sup> )	239 $\pm$ 132	187 $\pm$ 125	0.107
	HF <sub>AG</sub> (ms <sup>2</sup> )	134 $\pm$ 82	106 $\pm$ 55	0.057
-	d	0.548 $\pm$ 0.056	0.544 $\pm$ 0.070	0.867
GARCH	$u_1$	0.078 $\pm$ 0.028	0.072 $\pm$ 0.027	0.406
	$v_1$	0.340 $\pm$ 0.073	0.352 $\pm$ 0.083	0.511
DFA	$\alpha_{LF RR}$	1.504 $\pm$ 0.124	1.423 $\pm$ 0.118	<b>0.011</b>
	$\alpha_{HF RR}$	1.225 $\pm$ 0.050	1.250 $\pm$ 0.065	0.200
	$\alpha$	1.047 $\pm$ 0.077	1.045 $\pm$ 0.055	0.904
	$\alpha_{LF SML}$	0.801 $\pm$ 0.059	0.817 $\pm$ 0.073	0.276
	$\alpha_{HF SML}$	0.936 $\pm$ 0.187	0.924 $\pm$ 0.190	<b>0.011</b>
Entropia	$ApEn$	0.980 $\pm$ 0.100	1.021 $\pm$ 0.096	<b>0.005</b>

Verifica-se que:

- Os indivíduos adultos apresentam valores mais elevados para os parâmetros do domínio do tempo Média RR, SDRR e pNN50, para as componentes espectrais do modelo AR e do modelo ARFIMA-GARCH, para o parâmetro ARCH e para os expoentes escalares de memória curta  $\alpha_{LF|RR}$  e  $\alpha_{HF|SML}$  (Almeida, 2015);
- Os valores dos parâmetros de memória longa,  $d$  e  $\alpha$  são mais elevados para o grupo de adultos, adversamente ao estudo de Leite (2008). A  $ApEn$  apresenta o mesmo comportamento, não corroborando o estudo de Almeida (2015) o que pode estar relacionado com a existência de mais de uma dezena de adultos

com idades entre os 60 e os 64 anos.

- Os parâmetros estatisticamente significativos são nomeadamente o SDRR, o  $LF_{AR}$ , o  $HF_{AR}$ , o  $\alpha_{LF|RR}$ , o  $\alpha_{HF|SML}$  e a  $ApEn$ .



**Figura 4.12** – Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos saudáveis Adultos (A) e indivíduos saudáveis Idosos (I).

### Análise da VFC em função do Género

Analisando a modulação da VFC, em função do género, são considerados os parâmetros das metodologias lineares e não lineares analisadas, para o grupo dos homens (31 registos) e para o grupo das mulheres (23 registos).

Na Figura 4.13 e na Tabela 4.8 ilustram-se os resultados globais, de cada parâmetro para cada grupo, permitindo comparar valores e determinar a influência do género na VFC.

**Tabela 4.8** – Estimativa global (média  $\pm$  desvio padrão) dos parâmetros obtidos pela aplicação dos métodos, referentes aos grupos de indivíduos saudáveis homens e indivíduos saudáveis mulheres. Utilização do teste estatístico Wilcoxon-Mann-Whitney, com o intuito de determinar os parâmetros estatisticamente significativos, para os dois grupos.

Metodologias	Género N.º de Registos	Masculino 31	Feminino 23	p- <i>value</i>
<b>Domínio do Tempo</b>	Média RR(s)	0.775 $\pm$ 0.088	0.782 $\pm$ 0.064	0.910
	SDRR(s)	0.138 $\pm$ 0.023	0.139 $\pm$ 0.039	0.594
	pNN50 (%)	0.34 $\pm$ 0.62	0.55 $\pm$ 1.0	0.766
<b>AR</b>	LF <sub>AR</sub> (ms <sup>2</sup> )	526 $\pm$ 250	457 $\pm$ 278	0.252
	HF <sub>AR</sub> (ms <sup>2</sup> )	96 $\pm$ 45	107 $\pm$ 67	0.594
<b>ARFIMA</b>	LF <sub>AG</sub> (ms <sup>2</sup> )	227 $\pm$ 124	193 $\pm$ 138	0.205
	HF <sub>AG</sub> (ms <sup>2</sup> )	113 $\pm$ 56	129 $\pm$ 86	0.558
-	<i>d</i>	0.532 $\pm$ 0.068	0.564 $\pm$ 0.052	0.117
<b>GARCH</b>	<i>u</i> <sub>1</sub>	0.078 $\pm$ 0.029	0.072 $\pm$ 0.025	0.512
	<i>v</i> <sub>1</sub>	0.339 $\pm$ 0.058	0.356 $\pm$ 0.099	0.773
<b>DFA</b>	$\alpha_{LF RR}$	1.496 $\pm$ 0.131	1.416 $\pm$ 0.106	<b>0.031</b>
	$\alpha_{HF RR}$	1.230 $\pm$ 0.060	1.248 $\pm$ 0.058	0.315
	$\alpha$	1.022 $\pm$ 0.062	1.078 $\pm$ 0.056	<b>0.002</b>
	$\alpha_{LF SML}$	0.819 $\pm$ 0.068	0.796 $\pm$ 0.064	0.144
	$\alpha_{HF SML}$	1.042 $\pm$ 0.211	0.920 $\pm$ 0.160	<b>0.037</b>
<b>Entropia</b>	<i>ApEn</i>	0.967 $\pm$ 0.114	0.997 $\pm$ 0.090	0.290

Observa-se que:

- Os parâmetros no domínio do tempo são inferiores para o grupo dos indivíduos saudáveis do sexo masculino e não discriminam, a um nível de significância de 5%, os dois grupos estudados;
- As componentes espectrais LF e o parâmetro ARCH apresentam valores mais elevados nos homens, enquanto que as componentes espectrais HF e o parâmetro GARCH apresentam valores superiores no grupo das mulheres. Os parâmetros não são estatisticamente significativos;
- As componentes de memória longa *d* e  $\alpha$  são superiores para o grupo das mulheres saudáveis, corroborando o estudo de Almeida (2015). O expoente escalar  $\alpha$  permite discriminar o grupo das mulheres e dos homens;

- Os expoentes de escala de memória curta, do Sinal RR e do Sinal RR|SML são superiores para os grupos dos homens, com exceção do  $\alpha_{HF|RR}$ . Os expoentes  $\alpha_{LF|RR}$  e  $\alpha_{HF|SML}$  são estatisticamente significativos, a um nível de 5%.

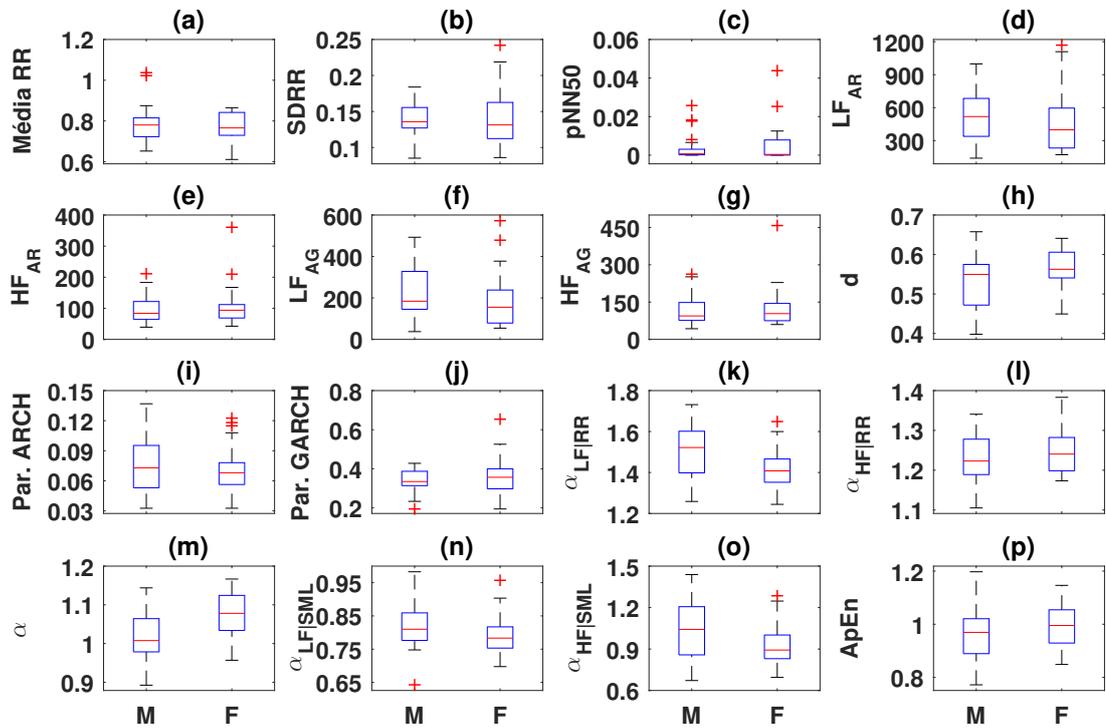


Figura 4.13 – Diagrama de blocos, dos parâmetros obtidos pela aplicação dos métodos estudados, para o conjunto de dados de indivíduos do sexo Masculino (M) e indivíduos do sexo Feminino (F).

## 4.4 Conclusões

A análise da VFC, de duas bases de dados, através dos métodos estudados, permitiu retirar diversas conclusões relevantes. Relativamente aos conceitos:

**Parâmetro de memória longa,  $d$ :** É o que melhor discrimina os momentos decisivos e não decisivos, da base de dados dos árbitros de futebol. Altos níveis de stress influenciam a memória longa, apresentando valores superiores para os

momentos decisivos. Da análise da base de dados Saudáveis/Doentes verifica-se que a componente de memória longa é mais elevada para os doentes, quando comparados com os saudáveis;

**Influencia do stresse:** Denota-se que para momentos de maior stresse (MD e MDI), a componente LF é superior em relação aos MND. A potência de baixa frequência é estimulada pela atividade do SNS e do SNP. Após a tomada de uma decisão, os valores de LF muitas vezes diminuem, o que pode estar relacionado, com o facto de ser uma decisão clara ou derivado de outros fatores (como deixar de correr após a decisão). Sendo assim, um aumento do parâmetro após a tomada da decisão tem um efeito paradoxal.

**Volatilidade:** A variância condicional calculada, possibilita a identificação de diferenças de ritmo cardíaco acentuadas e a identificação de zonas de transição. O estudo da base de dados de pacientes Saudáveis/Doentes, demonstrou que o parâmetro ARCH e o parâmetro GARCH aumentam para o grupo dos doentes.

**Restantes parâmetros estudados:** A componente espectral LF está associada ao SNS e SNP e apresenta valores mais elevados, para o grupo de indivíduos saudáveis e para a tomada de decisões, dos árbitros de futebol, já que estão sujeitos a maiores níveis de stresse, pela tomada de uma boa ou má decisão para o jogo. O parâmetro HF influenciado pelo SNP também apresenta o mesmo comportamento, quando observados os grupos de indivíduos saudáveis e doentes. Os expoentes escalares, de memória curta são superiores para os saudáveis (excetuando  $\alpha_{HF|RR}$ ), quando comparados com os doentes, contrariamente à entropia aproximada que é superior para os indivíduos com patologia.

**Teste estatístico:** Deteta quais as características que discriminam estatisticamente dois conjuntos de dados, a um nível de significância de 5%. Na base de dados Saudáveis/Doentes, os parâmetros estatisticamente significativos são: Média RR, SDRR, pNN50,  $LF_{AR}$ ,  $HF_{AR}$ ,  $LF_{AG}$ ,  $HF_{AG}$ ,  $d$ , parâmetro GARCH,  $\alpha_{LF|RR}$ ,  $\alpha_{HF|RR}$ ,  $\alpha_{HF|SML}$  e  $ApEn$ . No estudo da influência da idade na VFC, os parâmetros estatisticamente significativos são: SDRR,  $LF_{AR}$ ,  $HF_{AR}$ ,  $\alpha_{LF|RR}$ ,  $\alpha_{HF|SML}$  e  $ApEn$  e para o fator

gênero, os parâmetros estatisticamente significativos são:  $\alpha_{LF|RR}$ ,  $\alpha$ ,  $\alpha_{HF|SML}$

**Classificação:** Determina a (%) de discriminação do conjunto de amostras, através da aplicação de três classificadores. O classificador que melhor desempenho revela, em termos globais, é o QDA quando utiliza os treze parâmetros normalizados. Informação correlacionada com a implementação da validação cruzada e com a sensibilidade.

**Gênero e Idade:** Estes fatores influenciam o sistema cardiovascular. A aplicação dos métodos estudados em Homens/Mulheres e Adultos/Idosos permitiu comparar e discriminar os parâmetros, consoante os grupos, determinando a sua influência na VFC. A componente espectral LF é mais elevada para o grupo dos homens, revelando elevada influência da atividade do SNS e do SNP, por outro lado a componente espectral HF é superior para o grupo das mulheres expondo uma predominância da atividade do SNP.



## Capítulo 5

# Considerações finais e trabalho futuro

A análise da VFC tem-se revelado uma técnica não invasiva, de detecção de doenças cardíacas. A literatura refere inúmeros estudos da caracterização dos dados da VFC através da aplicação de métodos lineares tradicionais, quer no domínio do tempo quer no domínio da frequência. Nos exemplos de métodos no domínio da frequência enquadram-se os métodos paramétricos, como a modelação AR, e os métodos não paramétricos, como o periodograma. A modelação AR permite analisar a memória curta do sinal RR.

Uma abordagem mais recente impulsionou o estudo das características de memória longa e da dinâmica não linear, presente nos dados da VFC. Com o objetivo de ultrapassar as limitações existentes, nos estudos da memória longa, é aplicada a modelação ARFIMA-GARCH. Os modelos não lineares ARFIMA-GARCH são uma extensão dos modelos AR, que incorporam a modelação conjunta de estruturas de correlação de termo curto e de termo longo e a modelação de características associadas à dependência da volatilidade. Os métodos não lineares possuem a vantagem de descrever as interações não lineares, existentes nos dados da VFC. O método não linear DFA permite descrever a flutuação dos resíduos de regressão e o método não linear da *ApEn* permite quantificar a complexidade de um sistema.

Para analisar os registos longos e não estacionários da VFC, foram aplicados os métodos lineares no domínio do tempo e uma segmentação combinada com os métodos no domínio da frequência e os métodos não lineares analisados. Como objeto de estudo, foram usadas duas bases de dados distintas.

A primeira base de dados, foi fornecida pela Universidade de Zaragoza, Espanha, e é composta por intervalos RR, de quatro árbitros, durante quatro respetivas partidas de futebol. Foram coletados os momentos decisivos de cada árbitro, os momentos menos decisivos mas importantes e os momentos não decisivos. A aplicação de parâmetros analisados anteriormente, nos momentos mencionados permitiu concluir que:

- Os valores mais baixos do parâmetro  $d$  correspondem aos momentos não decisivos do jogo e ao período de descanso. Estes períodos traduzem-se num relaxamento dos árbitros. Quando, durante a partida, há uma decisão importante a ser tomada, a frequência cardíaca e os níveis de stresse aumentam, assim como o valor do parâmetro  $d$ ;
- O parâmetro LF exhibe valores mais elevados durante os momentos decisivos. Segundo a literatura, a atividade do sistema nervoso simpático e do sistema nervoso parassimpático revela uma predominância de LF. Quando o árbitro está sujeito a maiores níveis de stress, o SNS responde ao organismo. Esta conjuntura explica os valores de LF.

Analisou-se uma outra base de dados, com 83 registos de Holter, pertencentes a indivíduos saudáveis e a indivíduos com insuficiência cardíaca congestiva. Os registos foram recolhidos da base de dados do MIT/BIH, da Physionet. Dos resultados, é possível evidenciar que:

- Os parâmetros no domínio do tempo Média RR, SDRR e pNN50, as componentes espectrais  $LF_{AR}$ ,  $HF_{AR}$ ,  $LF_{AG}$  e  $HF_{AG}$ , o parâmetro de memória longa  $d$  e as medidas não lineares  $v_1$ ,  $\alpha_{LF|RR}$ ,  $\alpha_{HF|RR}$ ,  $\alpha_{LF|SML}$  e  $ApEn$  permitem discriminar entre indivíduos saudáveis e doentes;

- A VFC é dependente dos fatores idade e género, como comprovado no estudo realizado. No grupo dos Adultos/Idosos o parâmetro SDRR, as componentes espectrais de AR (LF e HF), os expoentes de escala,  $\alpha_{LF|RR}$  e  $\alpha_{HF|SML}$  e a  $ApEn$  são estatisticamente significativos. No grupo de dados Homens/Mulheres, os parâmetros que permitem discriminar os dois grupos são respetivamente  $\alpha_{LF|RR}$ ,  $\alpha$  e  $\alpha_{HF|SML}$ .

A implementação das características estatisticamente significativas, nos três tipos de classificadores utilizados (LDA, QDA e KNN) revelou que, em termos globais, o classificador que apresenta melhores resultados é o QDA, quando aplicado a todos os parâmetros normalizados.

Este estudo sublinha que a caracterização da VFC, através de métodos lineares e não lineares, permite detetar doenças do fórum cardiológico. Determinados parâmetros apresentam um papel de mediadores discriminativos, entre indivíduos saudáveis e doentes, revelando um contributo a nível científico e clínico.

Como trabalho futuro sugere-se:

- Utilização de outras extensões do modelo ARFIMA na análise da VFC, como o modelo ARFIMA-EGARCH;
- Expansão da aplicação dos expoentes de escala, aos quadrados dos resíduos ARFIMA, a denominada volatilidade;
- Utilização de redes neuronais ou máquinas de vetores de suporte (*support vector machine*) para discriminar os pacientes.



# Referências bibliográficas

- Absil, P.-A., Sepulchre, R., Bildge, A., e Gérard, P. (1999). Nonlinear analysis of cardiac rhythm fluctuations using DFA method. *Physica A*, 271:235–244. [42](#)
- Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., e Suri, J. S. A. (2006). Heart rate variability: review. *Med. Biol. Eng. Comput.*, 44:1031–1051. [3](#), [5](#), [6](#), [7](#), [8](#), [12](#), [14](#)
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):761–723. [21](#)
- Almeida, R., Dias, C., Silva, M. E., e Rocha, A. P. (2017). *ARFIMA-GARCH Modeling of HRV: Clinical Application in Acute Brain Injury*, chapter 17. Springer. [35](#), [36](#)
- Almeida, S. A. A. (2015). Analysis of heart rate variability linear and nonlinear methods. M.S. thesis, Universidade de Trás-os-Montes e Alto Douro, Portugal. [65](#), [73](#), [78](#), [80](#)
- Alvarez-Ramirez, J., Rodriguez, E., e Echeverría, J. (2005). Detrending fluctuation analysis based on moving average filtering. *Physica A*, 354:199–219. [41](#)
- Axelrod, S., Lishner, M., Oz, O., e Ravid, J. B. M. (1987). Spectral analysis of fluctuations in heart rate: an objective evaluation of autonomic nervous control in chronic renal failure. *Nephron*, 45:202–206. [7](#)
- Baillie, R. T., Cecen, A. A., e Erkal, C. (2009). Normal heartbeat series are

- nonchaotic, nonlinear, and multifractal: new evidence from semiparametric and parametric tests. *Chaos*, 19:028503. [72](#)
- Baillie, R. T. e Chung, C. F. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics*, 11:23–40. [34](#), [38](#)
- Baumert, M., Voss, A., e Javorka, M. (2013). Compression based entropy estimation of heart rate variability on multiple time scales. In *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, volume 50, pages 37–40. [46](#)
- Beran, J. (1991).  $M$  estimators of location for Gaussian and related processes with slowly decaying serial correlations. *Journal of the American Statistical Association*, 86(415):704–708. [33](#)
- Billman, G. E. (2011). Heart rate variability: a historical perspective. *Frontiers in Physiology*, 2:86. [3](#), [5](#)
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327. [33](#), [34](#)
- Bonnemeier, H., Richardt, G. H., Potratz, J., Wiegand, U. K. H., Brandes, A., Kluge, N. Y., e Katus, H. A. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: differing effects of aging and gender on heart rate variability. *Journal of cardiovascular electrophysiology*, 14(8):791–9. [8](#)
- Box, G. E. P., Jenkins, G. M., e Reinsel, G. C. (1976). *Time Series Analysis Forecasting and Control*. Wiley, 4 edition. [27](#), [33](#)
- Brockwell, P. J. e Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, NY. [33](#)
- Bucchianico, A. D. (1999). Combinatorics, computer algebra and the Wilcoxon-Mann-Whitney test. *Journal of Statistical Planning and Inference*, 79(2):349–364. [56](#)

- Burnecki, K. e Sikora, G. (2017). Identification and validation of stable ARFIMA processes with application to UMTS data. *Chaos, Solitons and Fractals*, pages 1–11. [36](#)
- Carney, R. M., Blumenthal, J. A., Stein, P. K., Watkins, L., Catellier, D., Berkman, L. F., Czajkowski, S. M., O'Connor, C., Stone, P. H., e Freedland, K. E. (2001). Depression, heart rate variability, and acute myocardial infarction. *Circulation*, page 104:2024. [7](#)
- Cerutti, S., Esposti, F., Ferrario, M., Sassi, R., e Signorini, M. G. (2007). Long-term invariant parameters obtained from 24-h Holter recordings: a comparison between different analysis techniques. *Chaos*, 17(1):015108–9. [43](#)
- Clausius, R. E. (1865). Über Verschiedene für die Anwendung bequemere formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physik und Chemie*, 125:353–400. [46](#)
- Coli, M., Fontanella, L., e Granturco, M. (2005). Parametric estimation for ARFIMA models via spectral methods. *Statistical Methods and Applications*, 14:11–27. [28](#)
- Cooley, J. W. e Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19 (90):297–301. [16](#)
- Cui, T. (2013). Spectrum analysis of heart rate variability (HRV). [3](#), [16](#)
- Doukhan, P., Oppenheim, G., e Taqqu, M. S. (2003). *Theory and applications of long-range dependence*. Birkouster. [33](#)
- Durbin, J. (1960). The fitting of time-series models. *International Statistical Institute (ISI)*, 28(3):233–244. [19](#)
- Dziedziech, K., Czop, P., Staszewski, W. J., e Uhl, T. (2018). Combined non-parametric and parametric approach for identification of time-variant systems. *Mechanical Systems and Signal Processing*, 103:295–311. [20](#)
- Ellis, C. e Wilson, P. (2004). Another look at the forecast performance of ARFIMA models. *International Review of Financial Analysis*, 13:63–81. [28](#)

- Emese, N., Hajnalka, O., Bardos, G., e Molnar, P. (2000). Gender related heart rate differences in human neonates. *Pediatr. Res.*, 47(6):778–780. [8](#)
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1008. [33](#), [34](#)
- ESC/NASPE Task Force (1996). Heart rate variability, standards of measurement, physiological interpretation and clinical use. *Circulation*, 93:1043–1065. [2](#), [5](#), [6](#), [12](#), [13](#), [20](#)
- Evrengül, H., Tanriverdi, H., Dursunoglu, D., Kaftan, A., Kuru, O., Unlu, U., e Kilic, M. (2005). Time and frequency domain analyses of heart rate variability in patients with epilepsy. *Epilepsy Research*, 63:131–139. [14](#)
- Ewing, D. J. e Winney, R. (1975). Autonomic function in patients with chronic renal failure on intermittent haemodialysis. *Nephron*, 15 (6):424–429. [7](#)
- Gajda, J., Bartnicki, G., e Burnecki, K. (2018). Modeling of water usage by means of ARFIMA-GARCH processes. *Physica A: Statistical Mechanics and its Applications, Elsevier*, 512(C):644–657. [33](#)
- Ge, D., Srinivasan, N., e Krishnan, S. M. (2002). Cardiac arrhythmia classification using autoregressive modeling. *BioMed. Eng. Online*, 1:1–5. [7](#)
- Geweke, J. e Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238. [37](#)
- Gordon, S. M., Franaszczuk, P. J., Hairston, W. D., Vindiola, M., e McDowell, K. (2013). Comparing parametric and nonparametric methods for detecting phase synchronization in EEG. *Journal of Neuroscience Methods*, 212:247–258. [20](#)
- Granger, C. W. e Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–29. [26](#), [28](#)
- Haldrup, N. e Valdés, J. E. V. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199:1–11. [26](#)

- Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, 10:130–145. [38](#)
- Hansen, B. E. (1995). Time series analysis : James D. Hamilton 1994. *Econometric Theory*, 11(3):625–630. [34](#), [49](#)
- Henry, M. (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis*, 22(3):293–316. [38](#)
- Hongyu, K., Martins, V. L., e Oliveira Junior, G. J. (2016). Principal component analysis: theory, interpretations and applications. *E and S Engineering and Science*, 5(1):83–90. [57](#)
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176. [28](#), [29](#), [31](#), [32](#)
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *The Journal Educational Psychology*, 24:498–520. [57](#)
- Hsin, H. T., Yang, C. Y., Yeih, D. F., Shieh, J. S., e Li, A. H. (2010). The detrended fluctuation analysis of acute-phase heart-rate variability in acute coronary syndromes - a pilot study. *Int. J. Cardiol*, 140:252–255. [42](#)
- Hu, M. Y. e Tsoukalas, C. (2006). Conditional volatility properties of sleep-disordered breathing. *Computers in Biology and Medicine*, 36(3):303–12. [34](#)
- Hurst, H. E. (1951). Methods for storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, 116:770–799. [26](#), [36](#)
- Isler, Y. e Kunta, M. (2007). Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*, 37(10):1502–1510. [57](#)
- Isler, Y., Narin, A., e Ozer, M. (2015a). Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing. *Measurement Science Review*, 15(4):196–201. [57](#), [58](#), [76](#)

- Isler, Y., Narin, A., e Ozer, M. (2015b). Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing congestive heart failure. *Meas. Sci. Rev.*, 15(4):196–201. [6](#)
- John, M. e Priebe, C. E. (2007). A data-adaptive methodology for finding an optimal weighted generalized Mann-Whitney-Wilcoxon statistic. *Computational Statistics and Data Analysis*, 51(9):4337–4353. [56](#)
- Kadado, T., Maulik, D., e Chakrabarti, S. (1994). Comparison of parametric and nonparametric spectral estimation of continuous Doppler ultrasound shift waveforms. In *IEEE Proc. Digit. Sig. Process WS*, volume 6, pages 145–148. [20](#)
- Kaffashi, F., Foglyano, R., Wilson, C. G., e Loparo, K. A. (2008). The effect of time delay on approximate and sample entropy calculations. *Physica D*, 237:3069–3074. [9](#)
- Klepsch, J., Kluppelberg, C., e Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics*, 1:128–149. [17](#), [18](#)
- Kobayashi, M. e Musha, T. (1982). 1/f fluctuation of heartbeat period. *IEEE Transactions on Biomedical Engineering*, 29:456–457. [26](#)
- Kondepudi, D. e Prigogine, I. (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*, chapter 3-4. Wiley. [46](#)
- Kozat, S. S. e Singer, A. C. (2014). *Academic Press Library in Signal Processing: Signal Processing Theory and Machine Learning*, chapter 11. Academic Press Books. [17](#)
- Kunsch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika*, 74(3):517–524. [37](#)
- Leite, A. e Rocha, A. P. (2013). Análise da variabilidade da frequência cardíaca em indivíduos saudáveis, doentes com insuficiência cardíaca e doentes transplantados. *Motricidade*, 9(4):54–63. [73](#)

- Leite, A., Rocha, A. P., e Silva, M. E. (2013). Beyond long memory in heart rate variability: An approach based on fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *CHAOS*, 23:023103. [4](#), [9](#), [28](#)
- Leite, A., Rocha, A. P., Silva, M. E., Gouveia, S., Carvalho, J., e Costa, O. (2007). Long-range dependence in heart rate variability data: ARFIMA modelling vs. detrended fluctuation analysis. *Proc. Comput. Cardiol*, 34:21–24. [34](#), [43](#)
- Leite, A. M. S. (2008). *Contribuições para a Modelação da Variabilidade da Frequência Cardíaca*. Ph.D. thesis, Faculdade de Ciências da Universidade do Porto. [33](#), [34](#), [35](#), [36](#), [73](#), [78](#)
- Lerma, C., Minzoni, A., Infante, O., e José, M. V. (2004). A mathematical analysis for the cardiovascular control adaptations in chronic renal failure. *Artif. Organs.*, 28 (4):398–409. [7](#)
- Levinson, N. (1946). The wiener (root mean square) error criterion in filter design and prediction. *Journal of Mathematics and Physics*, 26:261–78. [19](#)
- Li, T.-H. (2010). A robust periodogram for high-resolution spectral analysis. *Signal Processing*, 90:2133–2140. [16](#)
- Ling, S. e Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *American Statistical Association*, 92(439):1184–1193. [38](#)
- Liu, K., Chen, Y.-Q., e Zhang, X. (2017). Na evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) programs. *Axioms*, 6 (2):1:16. [18](#), [19](#), [32](#)
- Lopes, P. F., Oliveira, M. I. B., André, S. M. S., Nascimento, D. L. A., Silva, C., Rebouças, G., Filho, T. F. N. J. B. A., e Medeiros, H. J. (2013). Aplicabilidade clínica da variabilidade da frequência cardíaca. *Rev. Neurocienc.*, 21(4):600–603. [6](#)

- Lowensohn, R. I., Weiss, M., e Hon, E. H. (1977). Heart-rate variability in brain-damaged adults. *Lancet*, 1:626–628. [7](#)
- Mahajan, R., Viangteeravat, T., e Akbilgic, O. (2017). Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics. *International Journal of Medical Informatics*, 108:55–63. [6](#)
- Malik, M., Hnatkova, K., Batchvarov, V., Gang, Y., P, S., e Camm, A. J. (2004). Sample size, power calculations, and their implications for the cost of thorough studies of drug induced QT interval prolongation. *Pacing Clin. Electrophysiology*, 27:1659–1669. [2](#)
- Mandelbrot, B. B. e Ness, J. W. V. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10 (4):422–437. [28](#)
- Mandelbrot, B. B. e Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5):967–988. [37](#)
- Martinez, W. L., Martinez, A. R., e Solka, J. L. (2005). *Exploratory data analysis with MATLAB*, chapter 2. Computer Science and data analysis series. [57](#)
- Mikosch, T., Gadrich, T., Kliippelberg, C., e Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, 23(1):305–326. [19](#)
- Mohamed, R. A. H. (2016). Using ARFIMA models in forecasting the total value of traded securities on the Arab Republic of Egypt. *IJRRAS*, 27(1). [37](#)
- Molina-Picó, A., Cuesta-Frau, D., Aboy, M., Crespo, C., Miró-Martínez, P., e Oltra-Crespo, S. (2011). Comparative study of approximate entropy and sample entropy robustness to spikes. *Artificial Intelligence in Medicine*, 53:97–106. [47](#), [49](#)

- Nielsen, M. O. e Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24:405–443. [38](#)
- Orini, M., Barbieri, R., Nardelli, M., Scilingo, E. P., e Valenza, G. (2017). *Complexity and Nonlinearity in Cardiovascular Signals*, chapter 1. Springer. [5](#), [6](#), [8](#), [12](#), [13](#), [14](#)
- Paradisi, P. e Allegrini, P. (2017). *Complexity and Nonlinearity in Cardiovascular Signals*, chapter 66. Springer. [25](#), [41](#)
- Parzen, E. (1983). Time series ARMA model identification by estimating information. In *Proc. Interface of Computer Science and Statistics*. [19](#)
- Peng, C.-K., Havlin, S., Stanley, H. E., e Golberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5:82–87. [41](#), [43](#)
- Pérez, A. e Ruiz, E. (2001). Modelos de memória larga para séries económicas y financieras. *Investigaciones económicas*, 26 (3):395–445. [26](#)
- Pfeifer, M. A., Cook, D., Brodsky, J., Tice, D., Reenan, A., Swedine, S., e al. (1982). Quantitative evaluation of cardiac parasympathetic activity in normal and diabetic man. *Diabetes*, 3:339–45. [7](#)
- Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA*, 88 (6):2297–2301. [47](#), [48](#), [49](#)
- Piskorskia, J., Kosmidera, M., Mieszkowskia, D., Krauzec, T., Wykretowicz, A., e Guzik, P. (2018). Properties of asymmetric detrended fluctuation analysis in the time series of RR intervals. *Physica A: Statistical Mechanics and its Applications*, 491:347–360. [41](#)
- Reisen, V., Abraham, B., e Lopes, S. (2001). Estimation of parameters in ARFIMA processes: a simulation study. *Comm. Statist.*, 30 (4):787–803. [32](#), [36](#)

- Restrepo, J. F., Schlotthauer, G., e Torres, M. E. (2014). Maximum approximate entropy and R threshold: A new approach for regularity changes detection. *Physica A*, 409:97–109. [49](#)
- Rizas, K. D., Nieminen, T., Barthel, P., Zrn, C. S., Kahonen, M., Viik, J., Lehtimaki, T., Nikus, K., Eick, C., Greiner, T. O., Wendel, H. P., Seizer, P., Schreieck, J., Gawaz, M., Schmidt, G., e Bauer, A. (2014). Sympathetic activity-associated periodic repolarization dynamics predict mortality following myocardial infarction. *The Journal of clinical investigation*, 124(4):1770–80. [5](#)
- Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661. [38](#)
- Robinson, P. M. e Henry, M. (1999). Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory*, 15:299–336. [38](#)
- Rocha, A., Leite, A., e Silva, M. E. (2016). Volatility leveraging in heart rate: health vs disease. *Computing in Cardiology*, 43:25–28. [72](#)
- Rosenstock, E. G., Cassuto, Y., e Zmora, E. (1999). Heart rate variability in the neonate and infant: analytical methods, physiological and clinical observations. *Acta Paediatrica*, 88:477–482. [7](#), [8](#)
- Sadaei, H. J., Enayatifar, R., Guimaraes, F. G., Mahmud, M., e Alzamil, Z. A. (2016). Combining ARFIMA models and fuzzy time series for the forecast of long memory time series. *Neurocomputing*, 175:782–796. [26](#)
- Schulz, S. e Voss, A. (2017). *Complexity and Nonlinearity in Cardiovascular Signals*, chapter 2. Springer. [25](#)
- Shahbazi, F. e Asl, B. M. (2015). Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability. *Computer Methods and Programs in Biomedicine*, 122(2):191–198. [6](#), [57](#)

- Shumway, R. H. (2011). *Time Series Analysis and Its Applications: With R Examples*, chapter 4. Springer. [15](#), [16](#), [17](#), [18](#), [19](#)
- Singh, J. P., Larson, M. G., ODonell, C. J., Wilson, P. J., Tsuji, H., Lyod-Jones, D. M., e Levy, D. (2000). Association of hyperglycemia with reduced heart rate variability: the Framingham heart study. *Am. J. Cardiol*, 86:309–312. [7](#)
- Son, C.-S., Kim, Y.-N., Kim, H.-S., Park, H.-S., e Kim, M.-S. (2012). Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *Journal of Biomedical Informatics*, 45(5):999–1008. [58](#)
- Spangl, B. e Dutter, R. (2012). Analyzing short-term measurements of heart rate variability in the frequency domain using robustly estimated spectral density functions. *Computacional Statistics and Data Analysis*, 56:1188–1199. [14](#)
- Stavrinou, M. L., Sakellaropoulos, G. C., Trachani, E., Sirrou, V., Polychronopoulos, P., Nikiforidis, G., e Chroni, E. (2014). Methodological issues in the spectral analysis of the heart rate variability: Application in patients with epilepsy. *Biomedical Signal Processing and Control*, 13:1–7. [20](#)
- Struzik, Z. R., Hayano, J., Soma, R., Kwak, S., e Yamamoto, Y. (2006). Aging of complex heart rate dynamics. *IEEE Transactions on Biomedical Engineering*, 53(1):89–94. [43](#)
- Takada, A., Conradt, R., e Richet, P. (2015). Residual entropy and structural disorder in glass: A review of history and na attempt to resolve two apparently conflicting views. *Jornal of Non-Crystalline Solids*, 429:33–44. [46](#), [47](#)
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley, 2 edition. [34](#)
- Velasco, C. (1999). Gaussian semiparametric estimation of nonstationary time series. *Journal of Time Series Analysis*, 20:87–127. [38](#)
- Villareal, R. P., Liu, B. C., e Massumi, A. (2002). Heart rate variability and cardiovascular mortality. *Curr. Atheroscler. Rep.*, 4 (2):120–127. [7](#)

- Watkins, T. W., Horns, K. M., Harvadi, D. G., Sohacy, R., e Milley, P. J. W. J. R. (1996). Heart rate variability: relationship to IVH in VLBW neonates (abstract). *Pediatric Res.*, 39:251A. [7](#)
- Wheeler, T. e Watkins, P. J. (1973). Cardiac denervation in diabetes. *Br Med J*, 4:584–586. [7](#)
- Wijnand, H. P. e van de Velde, R. (2000). Mann-Whitney/Wilcoxon's nonparametric cumulative probability distribution. *Computer Methods and Programs in Biomedicine*, 63(1):21–28. [56](#)
- Xiu, J. e Jin, Y. (2007). Empirical study of ARFIMA model based on fractional differencing. *Physica A*, 377:138–154. [26](#), [31](#)
- Zevallos, M. e Palma, W. (2013). Minimum distance estimation of ARFIMA processes. *Computacional Statistics and Data Analysis*, 52:242–256. [28](#)
- Zhang, M. Q. (1997). Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci.*, 94(2):565–568. [57](#)
- Zoccali, C., Ciccarelli, M., e Maggiore, Q. (1982). Defective reflex control of heart rate in dialysis patients: evidence for an afferent autonomic lesion. *Clin. Sci.*, 63:285–292. [7](#)
- Zurek, S., Guzik, P., Pawlak, S., Kosmider, M., e Piskorski, J. (2012). On the relation between correlation dimension, approximate entropy and sample entropy parameters, and a fast algorithm for their calculation. *Physica A*, 391:6601–6610. [48](#)

# Apêndice A

## Anexos

Nos anexos são representados os registros dos restantes três árbitros N2, N3 e N4, descritos ao longo do tempo. Adicionalmente, são identificados os momentos MD, MDI, MND. Os valores dos parâmetros obtidos pela aplicação dos métodos estudados, para os respectivos árbitros, são representados nas seguintes tabelas. Os resultados globais com todos os valores das componentes espectrais  $HF_{AR}$ ,  $LF_{AG}$ ,  $HF_{AG}$ , dos parâmetros ARCH e GARCH, dos expoentes de escala  $\alpha_{HF|RR}$ ,  $\alpha$  e  $\alpha_{LF|SML}$  e da  $ApEn$  são também expostos nos anexos, para todos os momentos (MD, MDI, MND) e para todos os árbitros (N1, N2, N3, N4).

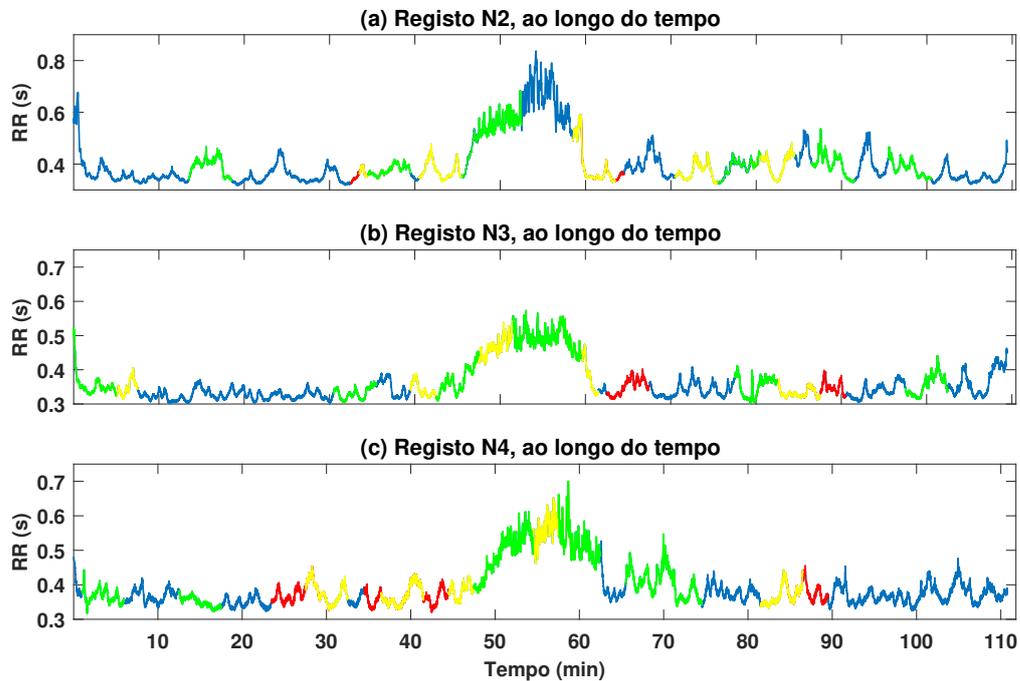


Figura A.1 – Representação dos registos dos árbitros, ao longo do tempo: (a) Registo N2, (b) Registo N3 e (c) Registo N4.

Tabela A.1 – Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N2.

Registo N2		Sinal RR											Sinal RR SML	
		LF <sub>AR</sub> (ms <sup>2</sup> )	HF <sub>AR</sub> (ms <sup>2</sup> )	LF <sub>AG</sub> (ms <sup>2</sup> )	HF <sub>AG</sub> (ms <sup>2</sup> )	<i>d</i>	<i>u</i> <sub>1</sub>	<i>v</i> <sub>1</sub>	$\alpha_{LF}$	$\alpha_{LF}$	$\alpha$	<i>ApEn</i>	$\alpha_{LF}$	$\alpha_{HF}$
M.D.	35	45	8	0.3	1	1.167	0.000	0.000	1.646	1.134	1.530	0.441	0.234	0.213
	62	420	108	2	1	0.674	0.288	0.150	1.783	1.840	1.218	0.175	1.005	0.521
M.D.I.	36	24	3	0.3	1	1.215	0.000	0.000	1.567	1.169	1.668	0.466	0.197	0.239
	43	73	14	0.3	2	1.513	0.333	0.246	1.484	1.634	1.610	0.212	0.193	0.305
	61	346	43	1	2	1.112	0.160	0.820	1.559	1.814	0.826	0.103	0.623	0.692
	73	91	14	1	2	1.159	0.292	0.222	1.649	1.624	1.206	0.242	0.432	0.540
	82	103	21	2	3	1.105	0.083	0.899	1.388	1.686	1.515	0.243	0.558	0.596
M.D.N.	16	135	20	3	2	1.031	0.209	0.310	1.580	1.702	1.553	0.256	0.696	0.461
	37	30	4	1	0.4	1.113	0.000	0.000	1.648	1.232	1.410	0.469	0.311	0.288
	48	429	99	24	23	0.830	0.132	0.860	1.325	1.795	1.579	0.260	0.640	1.206
	50	578	72	130	45	0.612	0.000	0.000	1.064	1.730	0.705	0.774	0.663	1.429
	78	143	27	2	3	1.088	0.241	0.258	1.307	1.654	1.416	0.272	0.629	0.493
	89	226	29	5	4	0.912	0.134	0.827	1.487	1.919	1.416	0.202	0.755	0.763
	98	105	13	2	2	1.003	0.337	0.379	1.411	1.567	1.884	0.272	0.478	0.381

**Tabela A.2** – Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N3.

Registo N4 min		Sinal RR											Sinal RR SML	
		LF <sub>AR</sub> (ms <sup>2</sup> )	HF <sub>AR</sub> (ms <sup>2</sup> )	LF <sub>AG</sub> (ms <sup>2</sup> )	HF <sub>AG</sub> (ms <sup>2</sup> )	<i>d</i>	<i>u</i> <sub>1</sub>	<i>v</i> <sub>1</sub>	$\alpha_{LF}$	$\alpha_{HF}$	$\alpha$	<i>ApEn</i>	$\alpha_{LF}$	$\alpha_{HF}$
M.D.	65	47	5	1	1	1.056	0.163	0.444	1.790	1.258	0.969	0.354	0.397	0.311
	88	69	7	1	1	1.167	0.238	0.407	1.633	1.393	1.340	0.491	0.418	0.292
M.D.I.	5	41	4	1	1	1.124	0.050	0.914	1.794	1.358	1.422	0.450	0.577	0.591
	42	64	6	1	1	0.889	0.059	0.907	1.637	1.301	1.184	0.330	0.814	0.390
	48	267	47	17	9	0.737	0.197	0.355	1.317	1.740	1.657	0.308	0.808	1.058
	49	259	43	20	10	0.789	0.164	0.683	1.155	1.696	1.442	0.455	0.823	0.963
	59	257	50	16	9	0.745	0.122	0.852	1.499	1.697	1.230	0.169	0.916	1.033
85	31	5	0.2	1	1.169	0.050	0.900	1.595	1.069	1.372	0.628	0.172	0.223	
M.D.N.	2.5	319	49	2	1	0.688	0.060	0.897	2.073	1.444	1.042	0.186	0.904	0.491
	33	24	4	0.1	1	1.173	0.042	0.913	1.509	1.117	1.783	0.403	0.196	0.213
	45	159	23	1	1	1.086	0.149	0.114	1.596	1.569	1.361	0.233	0.537	0.397
	54	464	41	125	26	0.367	0.273	0.503	1.298	1.835	0.935	0.767	1.106	1.520
	57	258	39	38	18	0.666	0.238	0.506	1.171	1.683	1.215	0.633	0.872	1.199
	80	147	16	5	4	0.883	0.142	0.828	1.536	1.625	1.673	0.346	0.893	0.559
100	74	6	1	1	1.084	0.074	0.898	1.681	1.456	1.155	0.266	0.540	0.381	

**Tabela A.3** – Resultados da aplicação da modelação AR, da modelação ARFIMA-GARCH, do método DFA e da Entropia Aproximada, num registo típico N4.

Registo N4 min		Sinal RR											Sinal RR SML	
		LF <sub>AR</sub> (ms <sup>2</sup> )	HF <sub>AR</sub> (ms <sup>2</sup> )	LF <sub>AG</sub> (ms <sup>2</sup> )	HF <sub>AG</sub> (ms <sup>2</sup> )	<i>d</i>	<i>u</i> <sub>1</sub>	<i>v</i> <sub>1</sub>	$\alpha_{LF}$	$\alpha_{LF}$	$\alpha$	<i>ApEn</i>	$\alpha_{LF}$	$\alpha_{HF}$
M.D.	26	62	11	0.4	2	1.267	0.107	0.148	1.501	1.108	1.646	0.517	0.155	0.244
	28	57	11	0.4	2	1.312	0.147	0.164	1.601	1.206	1.460	0.482	0.132	0.261
	37	41	8	0.2	1	1.316	0.138	0.199	1.519	0.864	1.831	0.817	0.068	0.222
	39	31	7	1	2.	1.155	0.050	0.900	1.570	0.900	1.602	0.514	0.139	0.272
	42	40	10	1	3	1.189	0.050	0.900	1.558	1.093	1.692	0.528	0.162	0.298
	87	79	17	1	3	1.348	0.050	0.900	1.520	1.175	1.520	0.643	0.156	0.295
M.D.I.	30	48	12	1	3.	1.274	0.142	0.372	1.549	1.150	1.384	0.477	0.180	0.356
	47	100	19	3	4.	0.968	0.074	0.195	1.478	1.230	1.554	0.542	0.504	0.393
	57	1041	70	346	52	0.400	0.000	0.000	1.170	1.908	1.065	0.600	0.926	1.722
	84	84	23	1	3	1.167	0.198	0.302	1.243	1.122	1.511	0.518	0.210	0.291
M.D.N.	3.5	101	9	2	3	1.004	0.255	0.076	1.460	1.388	1.395	0.644	0.586	0.490
	15	17	5	0.3	1	0.896	0.000	0.000	1.295	0.760	1.132	0.937	0.112	0.318
	50	252	61	40	20	0.613	0.050	0.900	1.141	1.665	1.562	0.315	0.730	1.240
	56	832	90	414	61	0.410	0.050	0.900	1.131	1.849	1.346	0.695	0.832	1.611
	60	936	82	134	43	0.565	0.289	0.413	1.154	1.851	0.759	0.540	0.943	1.578
	68	236	43	16	13	0.871	0.195	0.526	1.142	1.485	1.316	0.514	0.791	0.813
72	222	33	8	7	1.016	0.274	0.303	1.439	1.603	1.394	0.311	0.577	0.505	

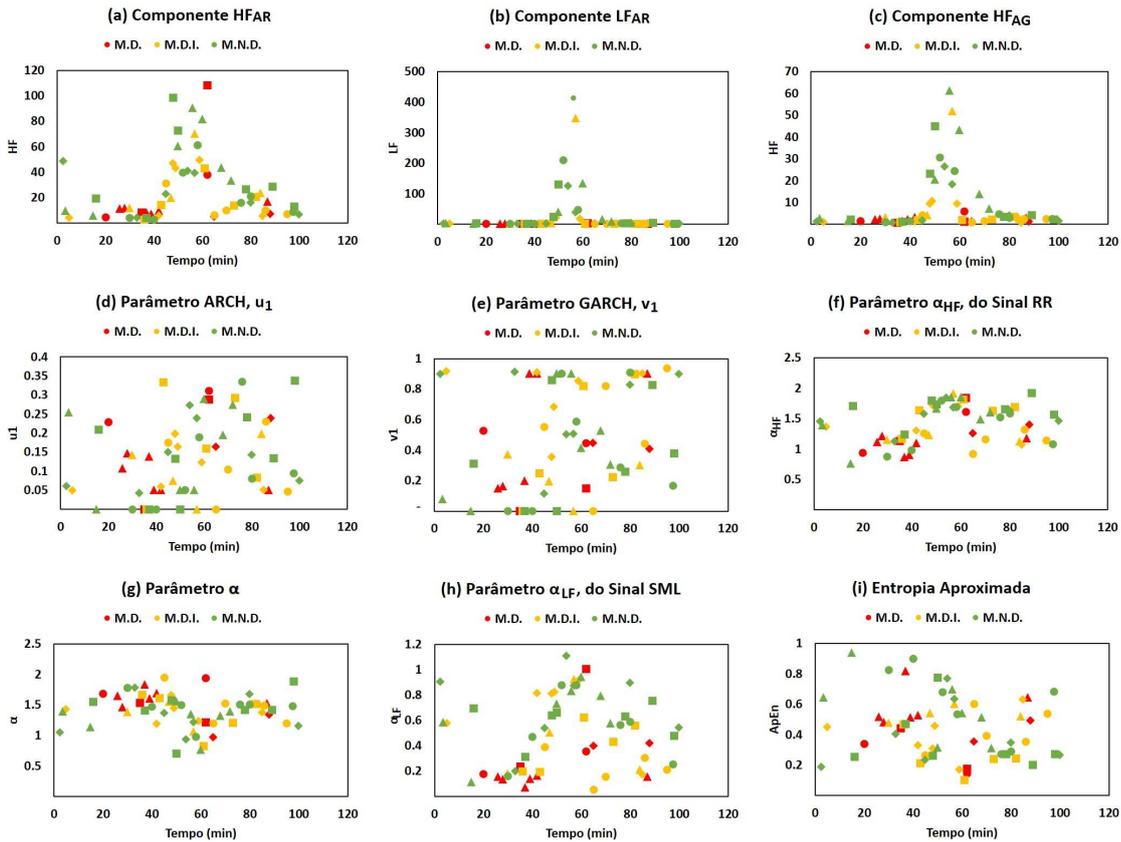


Figura A.2 – Representação gráfica dos parâmetros  $HF_{AR}$ , componentes espectrais ARFIMA, parâmetro ARCH, parâmetro GARCH,  $\alpha_{HF|RR}$ ,  $\alpha$ ,  $\alpha_{LF|SML}$  e da  $ApEn$ , para os quatro jogadores de futebol. A representação simbólica está relacionada com cada jogador: Círculo - Registo 1, Quadrado - Registo 2, Losango - Registo 3 e Triângulo - Registo 4. A vermelho estão representados os MD, a amarelo os MDI e a verde os MND.