Similarity-based approaches for the analysis and prediction of physiological time series

Teresa Raquel Corga Teixeira da Rocha

UNIVERSITY OF COIMBRA
Department of Informatics Engineering
Faculty of Sciences and Technology

PhD thesis submitted to the
Faculty of Sciences and Technology of University of Coimbra

Coimbra, August 2012

Advisor

Prof. Doutor Jorge Manuel Oliveira Henriques

Assistant Professor
Informatics Engineering Department
Faculty of Sciences and Technology of University of Coimbra

"The best way to predict the future
is to create it"

Peter Drucker (1909-2005)
Writer and management consultant

Acknowledgments

I would not have been able to successfully complete this thesis without the aid and support of countless people over the past four years.

Firstly, I would like to thank my supervisor Professor Jorge Henriques who was my mentor and source of inspiration. I am very grateful for his complete availability during the course of the thesis, offering me innumerable advices, patiently supervising me, and always guiding me in the right direction. These few words are not sufficient to express my great gratitude.

I also acknowledge the precious help of other research group members, namely Professors Bernardete Ribeiro, Paulo de Carvalho and Paulo Gil, for their valuable ideas and suggestions. Their contributions made a significant difference to the overall result of my work.

For all the support I have received in terms of clinical knowledge, I wish to express my sincere thanks to Doctor João Morais, Chief of Cardiology Division at the Centro Hospitalar de Leiria-Pombal. Without his medical expertise the conducted research would not have been feasible.

I am very grateful to the Coimbra Institute of Engineering/ Polytechnic Institute of Coimbra (ISEC/IPC), HeartCycle EU project (FP7-216695) and CISUC (Centre for Informatics and Systems of University of Coimbra) for their crucial support, without which it would have been impossible to bring the work to a successful conclusion.

I want to thank all my colleagues of DEIS/ISEC in general, but particularly my friends Simão Paredes and Cristina Chuva, my true "comrades in arms". With them I shared joys, sorrows, anxieties and hopes throughout this journey. With them, I arrived to the finish line. I will never forget our "happy talks".

Finally, I address my deepest gratitude to my all family for their love, support, and patience over the last few years. To my daughter Laura, my husband Jorge, my mother and father, my brothers and sisters in law, my nieces and nephews, a big thank you for helping me to overcome the difficult moments and regain my joy and sense of humour.

Abstract

Cardiovascular disease (CVD), a general name for a wide diversity of diseases, disorders and conditions that affect the heart and often the blood vessels, is the largest cause of death in the European Union. Since it is well known that heart health is linked to behaviour and lifestyle, the focus should be on prevention. In the context of preventive medicine, telemonitoring solutions are making a huge impact by enabling remote patient monitoring for the healthy and for those requiring management of chronic diseases.

One of the projects that address CVDs management by means of telemonitoring is HeartCycle, a European Integrated Project (FP7-216695) that aims at researching, developing and clinically validating innovative solutions for this purpose. Particularly, the goal of HeartCycle is to improve the quality of care for coronary artery disease (CAD) and heart failure (HF) patients.

Integrated in the third workpackage (*WP3-Multi-parametric Analysis and Decision Support*), the present thesis is centred on the development of specific clinical applications, which target cardiovascular conditions identified as relevant for the CAD/HF management, such as ischemia, arrhythmias and hypertension, based on the analysis and processing of the electrocardiogram (ECG) and blood pressure (BP) signals daily collected by home telemonitoring. Namely, investigation is made on techniques for the diagnosis of the referred conditions, and for the analysis of future trends of these signals enabling the early detection of critical events.

Specifically, this thesis presents methodologies for similarity detection and prediction in biosignal time series, which are mainly founded on the representation of signals as linear combinations of a set of orthogonal basis and on the time-frequency analysis of those signals.

Particularly, it proposes a new strategy for diagnosing ischemia comprising a measure for evaluating the ST deviation based on the time-frequency analysis of the ECG through the Wigner-Ville transform, and the use of Hermite basis functions to capture the most relevant morphologic characteristics of the QRS complex. This methodology was tested using the European Society of Cardiology ST-T public database, and the relevant results achieved, namely a sensitivity of 96.7% and a positive predictivity of 96.2%, confirmed its potential.

Additionally, a new similarity measure based on a combination of the wavelet transform with the Karhunen-Loève transform for temporal patterns detection in biosignal time series, mainly to support prediction methodologies, was developed. The respective validation was performed by quantitatively comparing the proposed measure with other three common measures through the use of data from a public dataset of Physionet (MIMIC-II) and from a private telemonitoring platform (TEN-HMS). The obtained

results confirm that the proposed similarity is particularly appropriate to deal with noise, trends and signals that are not perfectly aligned in time. Moreover, an iterative implementation allows for its efficient computational implementation.

In terms of predictive strategies two approaches are explored.

The first, based on generalized regression neural networks integrated into a multi-model structure is designed for the accurate prediction of time series future values. It was applied in the prediction of acute hypotensive episodes (AHE) and validated in the context of the 2009 Physionet/Computers in Cardiology Challenge using data from MIMIC-II dataset. A correct prediction of 10 out of 10 AHE for test set A and of 37 out of 40 AHE for test set B was achieved, corresponding to the best results of all entries in the two events of the challenge.

The main advantage of the second approach is that it does not require the development of a model. It exploits the multi-resolution analysis provided by the wavelet transform to estimate future evolution trend of biosignals, based on the trend evolution of similar historic signals. Its validity was demonstrated by the comparison with other common predictive methodologies. It was employed in the evaluation of the hypertension risk using data from TEN-HMS and MyHeart studies. The obtained results, in terms of Sensitivity-Specificity, were of 84.2%-75.5% and of 85.7%-91.8%, respectively, for the TEN-HMS and the MyHeart datasets, confirming the capability of the approach in this type of application.

Resumo

A doença cardiovascular (DCV), nome genérico para uma grande diversidade de doenças, distúrbios e condições que afectam o coração e muitas vezes os vasos sanguíneos, é a maior causa de morte na União Europeia. Como é um facto aceite que a saúde do coração está intimamente ligada ao comportamento e estilo de vida das pessoas, a aposta deve ser na prevenção. No contexto da medicina preventiva, as soluções de telemonitorização estão a ter um enorme sucesso, ao permitirem a monitorização remota e consequente gestão de indivíduos que padecem de doenças crónicas.

Um dos projectos que tem como alvo a gestão de doenças cardiovasculares através de soluções de telemonitorização é o HeartCycle, um projecto europeu integrado (FP7-216695), que visa pesquisar, desenvolver e validar soluções clinicamente inovadoras para esta finalidade. Particularmente, o objectivo do HeartCycle é melhorar a qualidade dos cuidados de saúde dos pacientes de doença arterial coronária (DAC) e insuficiência cardíaca (IC).

Integrada na terceira *workpackage* – WP3 (*Multi-parametric Analysis and Decision Support*), a presente tese centra-se no desenvolvimento de aplicações clínicas específicas que têm como alvo condições cardiovasculares identificadas como relevantes para a gestão de DAC/IC, tais como a isquemia, arritmias e hipertenaão, principalmente com base na análise e processamento dos sinais de electrocardiograma (ECG) e pressão arterial (PA), adquiridos diariamente em casa do paciente através de telemonitorização. Nomeadamente, é feita investigação em técnicas que permitam o diagnóstico das condições cardiovasculares referidas, bem como a previsão da evolução futura dos mencionados sinais de modo a possibilitar a detecção precoce de situações críticas.

Especificamente, esta tese apresenta metodologias para a detecção de similaridades e predição em séries temporais, que são fundamentalmente baseadas na representação dos sinais como combinações lineares de um conjunto de bases ortogonais, e na análise tempo-frequência dos mesmos.

Em particular é proposta uma nova estratégia para o diagnóstico de isquemia. Esta envolve uma medida para avaliar o desvio do segmento ST baseada na análise de tempo-frequência do ECG através da transformada de Wigner-Ville, bem como o uso de funções base de Hermite para capturar as características morfológicas mais pertinentes do complexo QRS. Esta metodologia foi testada recorrendo à base de dados pública *European Society of Cardiology* ST-T. Os resultados relevantes alcançados, correspondendo a uma sensibilidade de 96.7% e um valor positivo preditivo de 96.2%, confirmaram o seu potencial.

Além disso, é desenvolvida uma nova medida de similaridade com base numa combinação das transformadas Wavelet e Karhunen-Loève, para a detecção de padrões (templates) em bio-sinais, com o objectivo de servir de suporte às metodologias de predição. A respectiva validação foi realizada através da comparação quantitativa da medida proposta com outras três medidas, através da utilização de dados de uma base de dados pública da Physionet (MIMIC II) e de uma plataforma de telemonitorização privada (TEN-HMS). Os resultados obtidos confirmam que a medida proposta é particularmente apropriada para lidar com ruído, tendências e sinais não alinhados no tempo. A implementação de um esquema iterativo permite reduzir consideravelmente a complexidade do algortimo.

Em termos de estratégias preditivas são exploradas duas abordagens.

A primeira, baseada em redes neuronais regressivas integradas numa estrutura de multi-modelos, é projectada para a previsão precisa de valores futuros de uma série temporal. Foi aplicada na predição de episódios agudos de hipotensão (EAH) e validada no contexto da 9ª edição do Physionet/Computers in Cardiology Challenge, usando a base de dados MIMIC-II da Physionet disponibilizada para efeitos do concurso. A previsão correcta de 10 EAH dos 10 existentes no conjunto de teste A, e de 37 EAH dos 40 existentes no conjunto de teste B, conduziu à obtenção do primeiro lugar da classificação nos dois eventos do desafio.

A vantagem da segunda abordagem é não utilizar explicitamente um modelo. Esta, explorando a análise de multi-resolução proporcionada pela transformada Wavelet, tem como objectivo estimar a tendência de evolução futura dos bio-sinais com base nas tendencias de sinais similares identificados no histórico. A validade desta abordagem foi comprovada através da sua comparação com outras metodologias de predição. A mesma foi aplicada na avaliação do risco de hipertensão e validada utilizando dados de duas plataformas de telemonitorização privada (TEN-HMS e MyHeart). Os resultados obtidos em termos de sensibilidade-especificidade foram de 84.2%-75.5% e de 85.7%-91.8%, respectivamente para os dados dos estudos TEN-HMS e MyHeart, confirmando as potencialidades da abordagem neste tipo de aplicação.

Table of Contents

x

Acronyms

ABP	Arterial blood pressure
AC	Accuracy
AHE	Acute hypotensive episode
AIC	An information criterion
ANN	Artificial neural network
ANOVA	Analysis of variance
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
ART	Adaptive resonance theory
ARX	Autoregressive with exogenous input
AVP	Average of patterns
BIM	Bioimpedance monitor
BP	Blood pressure
CA	Cardiac arrest
CAD	Coronary artery disease
CORC	Pearson's correlation coefficient
CV	Cardiovascular
CVD	Cardiovascular disease
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DSS	Decision support system
DTW	Dynamic time warping
DWT	Discrete wavelet transform
ECG	Electrocardiogram
EEG	Electroencephalogram
ERP	Event related potential
ESC	European Society of Cardiology
EU	European union
FFNN	Feed-forward neural network
FFT	Fast Fourier transform
FN	False negative
FP	False positive

FPE	Final prediction error
FT	Fourier transform
GARCH	Generalized autoregressive conditional heteroscedasticity
GPT	Glutamate pyruvate transaminase
GRNN	Generalized regression neural network
HD	Haemodialysis
HF	Heart failure
HR	Heart rate
ICG	Impedance cardio thoracic
ICU	Intensive care unit
IFN	Interferon
KLT	Karhunen-Loève transform
kNN	k-nearest neighbour
LMS	Least means square
MAP	Mean arterial blood pressure
MAPE	Mean absolute percentage error
MIMIC-II	Multiparameter Intelligent Monitoring in Intensive Care Databases
MIMO	Multiple input multiple output
MLNN	Multi-layer neural network
MRA	Multi-resolution analysis
MSE	Mean squared error
NAR	Non-linear autoregressive
NARMA	Non-linear autoregressive moving average
NARMAX	Non-linear autoregressive moving average with exogenous input
NRMSE	Normalized root mean squared error
PCA	Principal component analysis
PCA-indexing	Piecewise constant approximation indexing
PDA	Personal digital assistant
PLR	Piecewise linear representation
PVC	Premature ventricular contraction
RBF	Radial basis function
RNN	Recurrent neural network
SC	Subtractive clustering
SE	Sensitivity
SISO	Single input single output
SP	Specificity

SSE	Sum of squared errors
STFT	Short-time Fourier transform
SVD	Singular value decomposition
SVM	Support vector machine
SVR	Support vector regression
SWK	Similarity measure based on wavelet decomposition+KLT
TDNN	Time-delay neural network
TEN-HMS	Trans-European Network Homecare Monitoring Study
TN	True negative
TP	True positive
VF	Ventricular fibrillation
VT	Ventricular tachycardia
WMM	Wavelet multi-resolution prediction scheme
WP	Workpackage
WVD	Wigner-Ville distribution

List of Figures

Chapter 4

List of Tables

Chapter 4

1. Introduction

1.1 Motivation

Cardiovascular disease (CVD), a general name for a wide diversity of diseases, disorders and conditions that affect the heart and often the blood vessels, is a major cause of disability and premature death throughout the world. Cardiovascular diseases can take many forms, such as raised blood pressure (hypertension), cardiac arrhythmias (abnormal heart rhythms), coronary artery disease, heart attack (myocardial infarction), heart failure, and cerebrovascular disease (stroke).

According to data of the European Commission (Health-EU), CVDs are the largest cause of death in the European Union (EU) and account for approximately 40% of deaths (2 million) per year. In financial terms, the charges with the EU health care systems related to CVDs have been estimated to be just under € 110 billion (in 2006), which represent a cost per capita of € 223 per year, about 10% of the total health care expenditure across the EU.

As it is well known, the heart health is linked to behaviour and lifestyle. Therefore, the focus should be on prevention, recognized by clinical professionals as the best method to avoid diseases from happening. Modern medicine mainly based on medical treatment, hospital care, and invasive surgery, transformed itself in a postmodern one, where the goal is to prevent illness, and to prolong and improve the quality of people's life. In fact, the emphasis on therapeutic dimension has moved to the preventive one. The key idea is that to prevent is better than to treat. In fact, preventive medicine is more cost effective, that is, able to obtain the same outcome in terms of health, with smaller costs, less pain, and within a short time.

In the context of preventive medicine, telemonitoring solutions are making a huge impact by enabling remote patient monitoring for the healthy and for those requiring management of chronic diseases. In that sense, European countries have implemented new and more efficient health care models, focused on the prevention of chronic diseases and their consequences. The use of information and communication technologies, and specially, home telemonitoring, allows extending the health care outside the hospital by virtual medical visits, combining the tracking of the patients with a cost reduction. Moreover, the information collected through the continuous monitoring of patients during long periods allows for advances in the diagnosing of a disease, for the description

of its evolution and for the prediction of possible complications, including the early prevention of the occurrence of severe events that may require hospitalization. Despite these advantages, the effectiveness of telemonitoring became more evident when mobile communications and wireless technologies were introduced. Originally, medical telemonitoring systems consisted of home telecare units that send biosignals or alarms to the medical team from a PC connected to the fixed telephony network. Resulting from the progress in the field of smart medical sensors, together with the expansion of wireless and mobile communications, a new type of monitoring systems based on the use of wireless sensors and smartphones or PDAs (Personal Digital Assistants), for forwarding the data received by the sensors to the monitoring points to be stored, reviewed and analysed by healthcare providers, has come up. More recently, telemonitoring systems perform data processing on the patient side, generating alarms and transmitting the corresponding data to the professional staff, whenever an abnormal situation is detected. In this context, personal Health (pHealth) systems are a new and fast growing concept. The patient is at the centre of the health delivery process and, through remote monitoring and management applications, pHealth systems aim the continuity of care at all levels of health care delivery. Several potential benefits may be achieved using this new paradigm, ranging from better adjustment of pharmacological treatments, to higher patient survival rates.

One of the projects that address CVDs management by means of telemonitoring is HeartCycle, a European Integrated Project (FP7 – 216695) that aims at researching, developing and clinically validating innovative solutions for this purpose. Particularly, the goal of HeartCycle is to improve the quality of care for coronary artery disease (CAD) and heart failure (HF) patients. This can be achieved by using personalized systems for monitoring their condition at home, and involving them in the daily management of their disease, as well as by developing mechanisms to automatically report relevant monitoring data back to clinicians so that they can prescribe personalized therapies and lifestyle recommendations. The project assumes that the appropriate remote management of such patients at home is a promising solution to both the delivery of innovative healthcare, and the reduction of future healthcare cost. Considered as its major challenges are the detection of critical health status trends in time, the appropriate decision support for professionals to react to such trends, and supporting the motivation of patients to be compliant to treatment regimens and to adopt a beneficial lifestyle.

Figure 1.1 depicts a schematic diagram of the care system to be developed under HeartCycle project. According to it, the system contains a patient loop interacting directly with the patient to support the daily treatment that shows the health development, including treatment adherence and effectiveness. The idea is that being motivated, compliance will increase, and health will improve. The system also contains a professional loop involving medical professionals, for example, alerting to revisit the care

plan. Finally, the patient loop is connected with hospital information systems, to ensure optimal and personalised care.

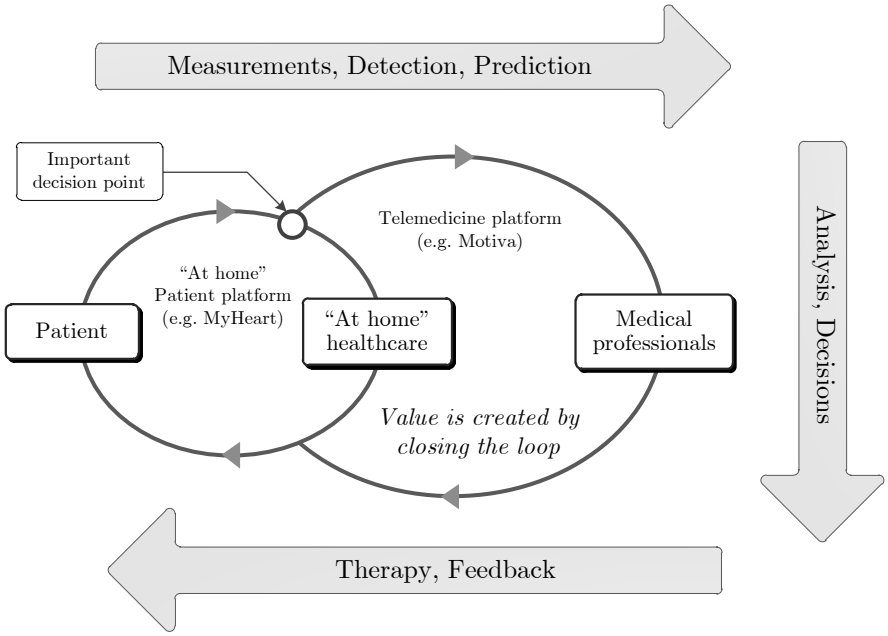


Figure 1.1 - Schematic diagram of the HeartCycle[1]*project.*

In technical terms, the HeartCycle systems comprise innovative sensors for home use by patients, such as unobtrusive and easy-to-use wearable sensor technologies, and devices for communication and information exchange, such as smartphones or TV via set top boxes. The collected information on the health status is sent to medical backend systems where sophisticated algorithms analyse the patient data to detect alarms or alerts. This is achieved by multi-parametric monitoring and analysis of vital signs and other measurements such as lab results and symptom questionnaires. Furthermore, decision support algorithms provide recommendations in case of alarms and alerts.

Among the various workpackages in which the overall work to be carried out under the project is divided, the third (WP3) – *Multi-parametric Analysis and Decision Support*, aims to provide patients and professionals with the essential information for an optimal management of heart failure and coronary artery disease. The main focus of interest is the development of algorithms able to assess relevant cardiovascular conditions (such as arrhythmias, ischemia, and hypertension) that allow to evaluate the patient's cardiovascular health status. Fundamentally, the WP3 module consists of two main components: *i*) Algorithms for evaluating the cardiovascular status of a patient, and a *ii*) Decision support system, which using the results of these algorithms, is responsible for

[1] www.heartcycle.eu

the implementation of strategies describing the expert knowledge to support clinical decisions.
Integrated in WP3, the present thesis is centred on the development of specific clinical applications, which target cardiovascular conditions identified as relevant for the CAD/HF management by the clinical partners, mainly based on the analysis and processing of the electrocardiogram (ECG) and blood pressure (BP) signals daily collected by home telemonitoring.
For this effect, investigation is made on techniques for diagnosis able to detect the referred conditions, and prediction methodologies for the trend analysis of these signals and consequent early detection of their critical evolution. From the clinicians' viewpoint these applications represent valuable tools both in terms of diagnosis and prognosis, which allow for an effective management of patients. In effect, together with other information sources (such as symptoms and biomarkers), they contribute for the evaluation of patient's cardiovascular health status by detecting the occurrence of specific conditions (such as ischemia and hypertension), and thus providing the medical staff with an idea of the patient's general state. Furthermore, the clinicians have access to the individual conditions assessment that may use as support in diagnosis formulation. On the other hand, by analysing the evolution trend of individual vital signs, the developed applications can generate alerts in face of dangerous progressions of the same, enabling clinicians to timely intervene (for example, adjusting medication) so as to avoid situations of extreme risk for the patient.

As conclusion, the use of new telemonitoring solutions, together with adequate diagnosis and prediction methodologies is of extreme significance for the conception of early prevention systems, providing professionals with adequate tools to evaluate and diagnose specific cardiovascular conditions. Several benefits might be achieved using this new health delivery paradigm, ranging from the prompt detection of forthcoming clinical severe conditions to better adjustment of pharmacological treatments. The awareness that these methodologies represent a central part of the professional loop, namely in helping medical decision making, contributing to the success of the whole care system of HeartCycle, is the major stimulus for the work to develop in the present thesis.

1.2 Hypotheses

Hospitals, healthcare centres, and clinics, accumulate numerous datasets with patients data, namely physiological time series routinely collected in intensive care units by the continuous monitoring of patients, or resulting from telemonitoring systems. On the other hand, given the population aging and the growth of health expenditures, the need for high quality and efficient healthcare, both at home and in hospital, is becoming more and more important. One way to achieve this is by taking advantage of the existing data to develop algorithms for the automatic evaluation of clinical conditions and their progression, that allow for the prompt detection of forthcoming clinical severe events and, thus, for the prompt intervention of clinicians in order to avoid critical situations for the patients.

From the clinical perspective, it is assumed that similarities between physiological temporal series can be used to detect and predict specific clinical conditions that contribute to the global cardiovascular assessment of the patient. For this effect, it is recognized that a relevant source of information could be the patient's historical or, as an alternative, the historical of other patients that have experimented similar behaviours in their health data.

Consequently, the research to be conducted under this thesis is founded on three main hypotheses.

- Firstly, it starts from the assumption that the cardiovascular status can be characterized based on the evaluation of specific cardiovascular conditions, such as hypertension, myocardial ischemia, arrhythmias, desynchronization, and pulmonary oedema, which are defined through literature or by clinical expertise. In fact, this hypothesis reflects the way the clinician assesses CV status, which is centred on well-known specific CV conditions, such as those referred above, evaluated from signals, symptoms, and laboratory results.
- On the other hand, it is assumed that characteristic patterns in biosignals (such as electrocardiogram and blood pressure) common to patients with a similar disease may have diagnostic/prognostic value in terms of clinical assessment. Therefore, algorithms to search similarities in biosignal time series may help in identifying the presence of a certain cardiovascular condition.
- Finally, it is presupposed that the estimation of biosignals future evolution trend by predictive methodologies, based on current and past measurements taken from historical data of a group of patients (including the patient under study), may help clinicians in the management of a cardiovascular disorder.

These hypotheses will be tested under HeartCycle project through the development of algorithms for the diagnosis and prognosis of relevant cardiovascular conditions with impact on the CV status. Two of these conditions, ischemia and hypertension, will be

particularly addressed by this thesis, although others such as ventricular arrhythmias and desynchronization have also been implemented by the candidate but not included in this document. Thereby, through the integration of the referred algorithms into the professional loop of HeartCycle, the cardiovascular health status of a patient may be evaluated based on their results and other sources of information, namely symptoms and lab results. On the other hand, the diagnosis of ischemia and the prediction of hypertension risk will allow testing the other two hypothesis, involving similarities searching and evolution trend analysis in blood pressure signals.

1.3 Objectives

The objectives outlined for the present thesis can be divided into scientific and clinical ones. In effect, in scientific terms the following goals are identified:

- Development of techniques for similarity searching in a set of biosignal time series to detect the occurrence of particular patterns, which will serve two purposes: helping diagnosis if patterns are characteristic of specific cardiovascular conditions, or being the starting point of predictive methodologies. These techniques should be efficient in view of the possible long duration of biosignal time series, and able to deal with trends (one the main characteristics of biosignals to be addressed in this thesis). For this end, the representation of signals in terms of a reduced set of orthogonal basis, and methodologies that exploit their time-frequency characteristics, will be investigated.
- Implementation of predictive strategies to be applied in the forecast of biosignal time series. For this end, two approaches will be explored: the first, involving the development of models for the accurate prediction of time series future values; the second, consisting in algorithms able to estimate the evolution trend of biosignals. In both approaches the first step corresponds to a similarity analysis procedure performed by the previously referred technique. Solutions based on neural network multi-models and multi-resolution analysis will be considered.

In clinical terms the main goal is to assist the clinical staff in the management of the HF and CAD patients by including the developed algorithms in the decision support system. Particularly, the support to the clinicians will consist in:

- Helping them in the correct diagnosis of cardiovascular conditions that are relevant for the cardiovascular health status of the patients, like ischemia and hypertension.
- Enabling them to timely intervening in order to prevent risk situations for the patients, by predicting the evolution of vital signs such as blood pressure.

1.4 Contributions

The present thesis basically provides methodologies for similarity search and prediction in biosignal time series, which are mainly founded on the representation of signals as linear combinations of a set of orthogonal basis, and on the time-frequency analysis of those signals. The specific contributions are:

- A new strategy for diagnosing ischemia, which comprises a measure for evaluating the ST deviation based on the time-frequency analysis of the ECG through the Wigner-Ville transform, and the use of Hermite basis functions to capture the most relevant morphologic characteristics of the QRS complex and T wave. This methodology will be tested using the European Society of Cardiology ST-T public database.
- A new technique for similarity searching in biosignal time series, which involves a new similarity measure and an efficient indexing scheme, based on a reduced set of basis resulting from the combination of the wavelet with the Karhunen-Loève transforms. This is particularly adequate to deal with biosignal trends. This technique, applied in the detection of patterns in biosignal time series, may be used to support classification, clustering and, mainly, predictive methodologies. The validation of the proposed similarity measure will be performed using MIMIC-II and TEN-HMS datasets.
- A new predictive methodology based on a multi-model scheme of neural networks, which allows for the accurate forecast of time series future values. This methodology, validated in the context of the 2009 Physionet/Computers in Cardiology challenge, in the prediction of acute hypotensive episodes using data from MIMIC-II dataset, obtained the best results of that challenge.
- A new approach for the prediction of biosignal time series based on evolution trends, without involving explicit modelling, through which a rough estimative of the biosignal's future values is possible. This will be applied in the prediction of blood pressure signals, namely in the assessment of hypertension risk, and validated using TEN-HMS and MyHeart datasets.

Furthermore, the developed work will be integrated in a clinical decision support system, to assist health professionals in the management of heart failure patients, monitored by a telemonitoring system. This decision support system, developed by Philips Research Eindhoven, implements a rule-based system fundamentally defined from clinical evidence knowledge and from specific clinical guidelines.

1.5 Thesis organization

The present thesis is organized into five chapters.

Following the general introduction in this chapter, Chapter 2 deals with methods that detect similarities in long time series data. It starts by presenting an overview over the most common similarity measures and indexing algorithms and it shows the clinical relevance of similarity searching methods when applied to biosignals. The rest of the chapter is dedicated to the introduction of a new similarity measure based on the wavelet and Karhunen-Loève transforms, and of an innovative and efficient indexing scheme.

Chapter 3 addresses the research and implementation of methodologies for time series prediction, proposing two strategies: the first based on neural network multi-models aiming to achieve an accurate prediction of future time series values, and the second based on the multi-resolution analysis provided by wavelet transform targeting the estimation of the time series future evolution trend. This chapter starts by justifying the clinical relevance of the predictive methodologies and then presents their underlying theoretical background. Following, the proposed methodologies are explained in detail and, finally, some conclusions are drawn.

Chapter 4 presents and discusses the results related with the methodologies proposed through the thesis. Regarding similarity measures, a comparison in terms of the sensitivity of several approaches (including the proposed one) is, in a first phase, performed. Subsequently, this chapter exploits the capacity of the developed methodologies for both the diagnosis and prognosis of cardiovascular conditions, applying them to the specific cases of ischemia detection, hypotensive episodes and hypertension risk prediction.

In Chapter 5, a general assessment of the research outcomes in face of the goals established in the first chapter is performed, followed by conclusions and possible directions of future work.

2. Similarity Detection in Time Series

This chapter deals with methods that effectively detect patterns in long time series data. A new similarity measure based on wavelet and Karhunen-Loève transforms is introduced. Basically, a data transformation procedure is carried out by means of a set of orthogonal wavelet basis functions, which is then reduced to an optimal subset through the application of the Karhunen-Loève transform. As result, the time series is efficiently described by a linear combination of a reduced set of wavelet basis with the corresponding coefficients reflecting the main dynamic patterns of the biosignal. These coefficients are later employed to derive a temporal index that enables to effectively identify similarities between a particular template and a time series. An iterative implementation allows for the reduction of the computational complexity of the method. When applied to biosignals, the main assumption is that similar physiological temporal patterns can be used in the characterization and evolution analysis of a patient status.

2.1 Introduction

Large amounts of time series data are regularly produced in multiple application areas, such as industry, finances, and medicine. If available, automatic analysis tools of such data will be, unquestionably, an important instrument for professionals working in these areas. Among the several research lines, the discovery of relationships between time series is a central issue, recognized as a decisive investigation area. Moreover, given the vast quantity and length (time duration) of such time series, the development of similarity detection methods that are not only effective but also efficient, is a fundamental requisite (Hetland, 2004), (Park *et al.*, 2000), (Agrawal and Srikant, 1995), (Agrawal *et al.*, 1993). On the other hand, given that similarity detection methods support other time series subjects, such as clustering, classification or prediction, their study is, definitely, of general importance.

The problem of similarity between biosignal time series aims at finding whether different time series present, or not, a similar behaviour. A generalization of the similarity assessment is the similarity indexing problem, which intends to determine the subsequences of a time series that are similar to a given template. In this framework, one of the main scientific goals of the present thesis is the development of algorithms able to find segments of a time series (subsequences) that present the same dynamics of a given temporal template. The main clinical hypothesis is that characteristic patterns of biosignals and their interrelationships, common to patients with similar disease progressions, may have prognostic value in the medical assessment process. In effect, these algorithms may help in finding groups of patients with similar clinical behaviour, thus, in the identification of temporal patterns that may be suggestive of different clinical events. In terms of predictive algorithms, to be detailed in the next chapter, the objective is to estimate significant clinical trends or outcomes, based on current physiological measurements. The predictions can be useful in simple future events detection, in alerts generation when integrated in clinical decision support systems, as well as in helping clinicians to optimize care plans for critically-ill patients.

The biosignals under consideration are assumed to be regularly collected by means of monitoring applications, such as in intensive care units (ICU) or through home telemonitoring. In particular, the use of new continuous telemonitoring solutions and specialized processing methodologies, offer professionals the adequate information for evaluating cardiovascular conditions and symptoms progression that enable the prompt detection of forthcoming severe clinical disorders.

Of particular importance is the management of cardiovascular (CV) diseases, the leading group of conditions that cause death worldwide, which include coronary artery disease (CAD) and heart failure (HF) disorders. In this specific domain, HeartCycle intends to improve the life quality of patients with CAD or HF, by monitoring their condition and involving them in the daily management of their disease (Reiter and Maglaveras, 2009). Integrated in HeartCycle, the third workpackage (WP3) is the responsible for the research and development of models to assess some specific cardiovascular conditions. Basically, these models assume that the referred conditions can be characterized based on the analysis of biosignals collected during daily home monitoring process. Examples of these CV conditions are arrhythmias (Rocha *et al.*, 2008), (Henriques *et al.*, 2008) myocardial ischemia (Rocha *et al.*, 2010), hypertension episodes (Rocha *et al.*, 2011) and congestion.

In face of a collection of biosignal time series, the first step to be accomplished is to determine whether time series present similar behaviours. This topic, the assessment of similarity between time series, is a central concept in knowledge discovery and consists of evaluating the similitude between two different time series, by means of a measure of similarity or dissimilarity. The indexing problem aims at finding the sequences in a time series that exhibit temporal correspondences with a given template. In general, the similarity problem is intrinsically related to the indexing problem. In fact, simple

similarity measures usually allow for an easy derivation of index values, while more sophisticated similarity measures cause the indexing problem to be more complex.

Significant advances have been made in the development of methods for the detection of similarities in time series. From those that have been proposed, two main groups of algorithms can be identified: time domain and transform-based methods. The former work directly with the raw signals (eventually with some pre-processing) and the main goal is to derive a measure (scalar) based on the comparison of the original time series. Euclidean distance for signals with the same length and dynamic time warping (DTW) technique for signals with different lengths, are well-known examples of such algorithms (Liao, 2005), (Li *et al.*, 2002). The second group includes, among others, the Fourier transform, the singular value decomposition, the Karhunen-Loève transform, the wavelet decomposition, the Wigner-Ville transform, and the multi-dimensional scaling.

In the context of this work, the time-frequency analysis methods included in the second group of algorithms assume a particular interest. The transformation of a signal from one-dimensional time domain to two-dimensional time-frequency domain can reveal more details of the referred signal and, thus, help to characterize it (Carreño and Vuskovic, 2005), (Liu and Motoda, 1998), (Aussem *et al.*, 1998), (Cohen, 1995). Another important advantage of these methods consists in their capability to deal with the nonstationarities that are typical of biosignal time series.

In particular, the wavelet transform of a time series is investigated in this work. Basically, this transform produces features that describe properties of the time sequence both at various locations and at numerous time granularities (detail levels), which is extremely important when dealing with the similarity assessment problem. Moreover, the wavelet transform is a dimensionality reduction technique that, consequently, improves indexing performance. In fact, indexing applied to a long time series is usually inefficient mainly due to the very high dimensionality of the search. Therefore, through the description of the time series in a lower dimensional space, efficient indexing solutions can be achieved. That is the case of the wavelet transform, in which the time series is described by a set of coefficients that correspond to the contribution of each wavelet basis. These coefficients, which reflect the dynamical patterns of the time series, are employed to effectively compute the similarities between a particular template and the given time series.

In alternative to the wavelet basis functions, another basis sets may be used to represent time series, such as Gaussian kernels, polynomials basis, Hermite basis, etc. In fact, although wavelets are of general application, there are particular situations where other basis are more appropriate to capture signal-specific features. In truth, the choice of the correct basis can significantly increase the ability of an algorithm to identify the meaningful characteristics of a signal.

For instance, in the case of algorithms for detecting ischemic episodes, the isoelectric line and the ST segment deviation point (J point) are typically determined based on the QRS complex wave. In this situation, the use of a basis set able to generate features that

adequately describe the QRS morphology, makes possible an accurate estimation of these points of interest. In that sense, Hermite basis functions are appropriate. Moreover, in this particular case, the determination of the subsequences (QRS complex) in the time series (ECG) is a well-known problem, easily solved by specific methods. As a consequence, the search for the QRS complexes is not relevant (achieved with a regular R peak detector), being the main goal to identify subtle variations in their morphology. Supported in this principle, a strategy for the automatic detection of ischemic episodes will be presented in Chapter 4. A measure for evaluating the ST deviation based on the time-frequency analysis of the ECG through the Wigner-Ville transform, and the use of Hermite basis functions to capture the most relevant morphologic characteristics of the QRS complex, will be the key points of the proposed methodology.

Apart from the referred advantages in terms of similarity detection, the motivation to apply wavelet transform is also founded in the time series prediction, namely in situations where a given template has been observed in a given physiological signal. In this case, a two steps methodology is considered. In the first phase, a set of conditions similar to a current one (template) are searched using the proposed indexing technique. Since it is hypothesized that patients present similar physiological temporal patterns prior to a specific clinical event, the similarity assessment is performed in historical data sets of several patients. In a second phase, the most representative similar patterns are employed in the predictive approach that uses wavelet decomposition to forecast the future of the physiological under consideration (prediction will be the focus of the next chapter).

The structure of this chapter is as follows: in section 2, an overview over the most common similarity measures and indexing algorithms for time series is made. Section 3 introduces the wavelet transform method and points out the important properties that make it suitable for the approach presented in this thesis. Section 4 is concerned with the proposed similarity measure and indexing algorithms themselves. Finally, in section 5, some conclusions are drawn.

2.2 Clinical Relevance

As previously referred, one of the main scientific goals of the present thesis is the development of algorithms able to find segments of a time series that present the same dynamics of a given temporal template. This development is done under the clinical hypothesis that characteristic patterns of biosignals and their interrelationships, common to patients with similar disease progressions, may have prognostic value in the medical assessment process. Consequently, these algorithms may help in finding groups of patients with similar clinical behaviour or, equivalently, in the identification of temporal patterns that may be suggestive of different clinical events.

This idea has been explored by several authors which works demonstrate the existence of a relationship between temporal patterns (extracted from biosignal time series) and clinical events or conditions, confirming the relevance that the similarity search process assumes in clinical contexts.

In effect, Lehman *et al.*, (2008), Saeed (2007), Saeed and Mark (2006), identified similar physiological patterns in hemodynamic time series from intensive care units patients. They considered that the similarities between different patient's time series could have a meaningful physiologic interpretation in the detection of impending hemodynamic deterioration, and could be of potential use in clinical decision-support systems. In fact, they hypothesized that hemodinamically unstable patients could have similar physiological patterns prior to severe decompensation.

On the other hand, Hirano and Tsumoto (2005), Hirano and Tsumoto (2002), developed a method for finding similar patterns in temporal sequences and demonstrated the usefulness of the method on chronic hepatitis datasets, containing long time series data on laboratory examinations of patients with hepatitis B and C. They found some interesting patterns in their study. In effect, on glutamate pyruvate transaminase (GPT) sequences, they identified patterns that could represent the effectiveness of interferon (IFN) treatment. On platelet count sequences they found that if IFN treatment was ineffective, the count kept decreasing following the progress of liver fibrosis. On the contrary, if the IFN treatment was effective the platelet count started increasing.

In the neurology field, Gandhi and Green (2006) investigated the feasibility of a pattern recognition based approach using time-frequency representations of event related potentials (ERPs) of the electroencephalograms (EEGs) for the early diagnosis of Alzheimer's disease. In effect, numerous studies established a strong relationship between patterns of the ERPs and mental ability.

In turn, Noren *et al.* (2009) presented a framework for pattern discovery in large patient record repositories and demonstrated its usefulness through a set of examples from a collection of over two million patient records in the United Kingdom. The identified patterns included temporal relationships between drug prescriptions and medical events suggestive of persistent and transient risks of adverse events, possible beneficial effects of

drugs, periodic co-occurrence, and systemic tendencies of patients to switch from one medication to another.

Regarding renal diseases, Montani *et al.* (2006), and Leonardi *et al.* (2007), developed an auditing system to assess the performance of Haemodialysis (HD) treatments, which resorts to temporal data mining techniques for discovering relationships between the time patterns of the data automatically collected during HD sessions and the performance outcomes. On the other hand, they used case based reasoning to retrieve similar time series within HD data in order to evaluate critical patterns. The overall strategy showed to be appropriate for knowledge discovery and critical patterns similarity assessment on real patients' HD data, helping clinicians in improving their understanding of patients' behaviour.

In cardiology domain, Manocha and Singh (2011) made a review of the algorithms developed for ischemia detection from ECG, the mostly and commonly recorded signal in the patient monitoring and examination process. In effect, ischemia causes a decreased blood flow to muscle tissues of the heart which is manifested in ECG by elevated or depressed ST segment and T wave flattening or inversion. It is referred in literature that these abnormal patterns are suggestive of myocardial ischemia. As a result, despite of using different approaches, all the methodologies assume that detection of ischemia can be achieved by analysing the ST-T complex of the ECG.

Still in the field of cardiology, Crespo (2002), Wang *et al.* (2010), hypothesised that changes in arterial blood pressure signal morphology could be detected prior to the onset of hypotension. In that sense, they carried out a study in order to quantify, examine, and assess the statistical sensitivity of some of the possible changes in morphology. In turn, Frolich and Caton (2002) published a study in which they stated that baseline heart rate could be predictive of obstetric spinal hypotension. In fact, they concluded that higher baseline heart rate, possibly reflecting a higher sympathetic tone, could be a useful parameter to predict postspinal hypotension.

2.3 Background

2.3.1 Time Series Similarity Measure and Indexing

Essentially, a discrete time series $X(t) = \{x(1),\ \ldots,\ x(t),\ \ldots,\ x(N)\}$ is a sequence of real numbers $x(t) \in \mathbb{R}$, assumed to be collected at regular intervals over a period of time (N instants), as illustrated in Figure 2.1.

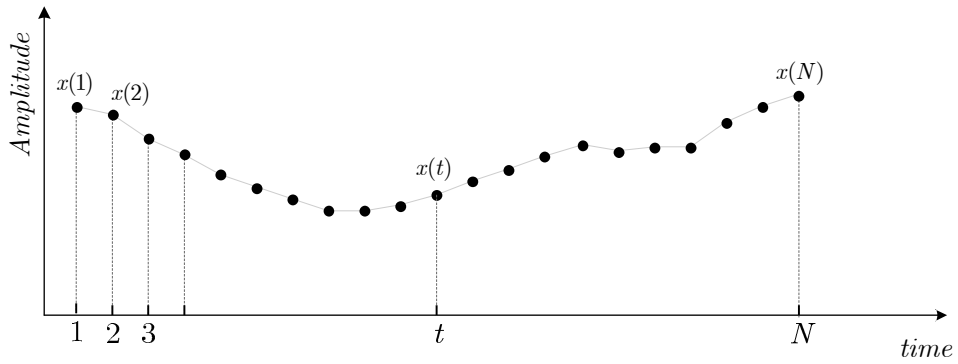


Figure 2.1 - Discrete time series definition.

The majority of the algorithms that have been proposed for the assessment of similarity between time series can be generally grouped in two main sets: time domain and transform-based methods (Liao, 2005), (Alonso *et al.*, 2003), (Keogh *et al.*, 2001). In addition, the latter can be sub-divided as feature extraction and model-based methods. Figure 2.2 depicts these different techniques: time domain, feature-based and model-based approaches.

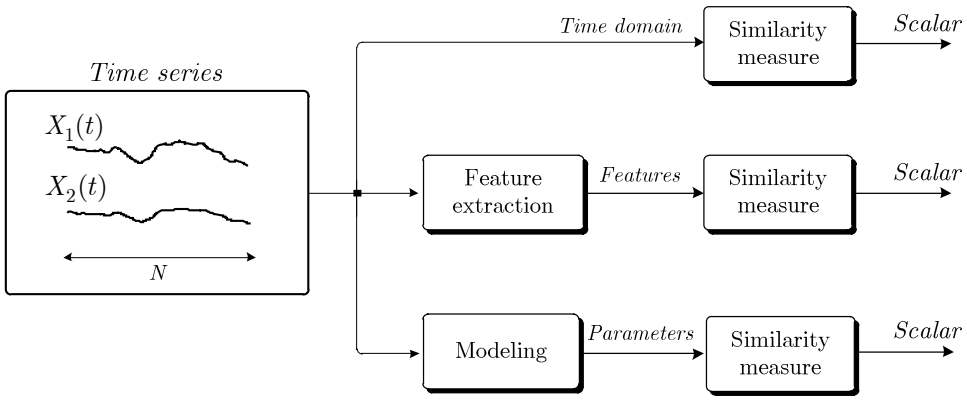


Figure 2.2 - Time series similarity approaches.

The first group of methods (time domain) works directly with the raw time series data, possibly with some pre-processing procedure, thus, they are also called raw-data-based techniques. In turn, the second group firstly derives a set of features or model parameters of lower dimension from the raw time series data and, then, use these features or parameters to calculate the similarity measure. The basic principle of transform-based methods is to convert data in a way that allows to extract additional characteristics, otherwise difficult to detect.

An issue related with the similarity assessment is the similarity indexing (illustrated in Figure 2.3), in which the main goal is to identify the subsequences in a time series that are similar to a specific template. Given a time series $T(t)$ of length T, $T(t) = \{t(1), ..., t(T)\}$, and a template $X(t)$ of length N, $X(t) = \{x(1), ..., x(N)\}$, the objective is to detect the subsequences $Y(t)$ of length N ($N < T$), that are similar to $X(t)$.

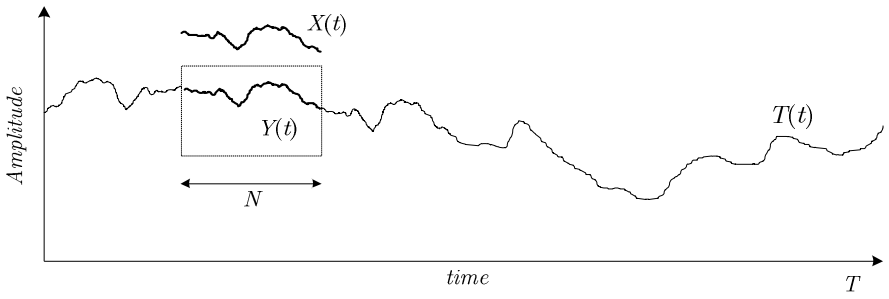


Figure 2.3 - Similarity indexing problem.

The problem of similarity indexing can be extended and, instead of performing the search in only one time series, a historic collection of S time series $\mathbf{S}(t) = \{T_1(t),\ T_2(t),\ \ldots\ ,\ T_S(t)\}$ can be considered. In this context, the subsequence indexing problem consists of finding the particular time series in the historic $\mathbf{S}(t)$ which subsequences are similar to $X(t)$.

2.3.2 Time Domain Methods

The simplest algorithms for computing a similarity metric between time series are the time domain approaches, such as the Euclidean distance, root mean square distance, Minkowski distance and correlation coefficient values. Given two discrete time series of length N, $X_1(t) = \{x_1(1), ..., x_1(N)\}$ and $X_2(t) = \{x_2(1), ..., x_2(N)\}$, the Euclidean distance, $D_E \in \mathbb{R}_0^+$, between them, is defined based on their individual values, $x_1(t)$ and $x_2(t)$, as:

$$D_E\left(X_1(t), X_2(t)\right) = \sqrt{\sum_{t=1}^{N}\left(x_1(t) - x_2(t)\right)^2} \tag{2.1}$$

The Minkowski distance, $D_M \in \mathbb{R}_0^+$, is a generalization of the Euclidean distance, defined as (2.2), where q is a positive integer.

$$D_M\Big(X_1(t), X_2(t)\Big) = \sqrt[q]{\sum_{t=1}^{N}\big(x_1(t) - x_2(t)\big)^q} \tag{2.2}$$

The linear correlation coefficient, $CC \in \mathbb{R}$, sometimes referred as Pearson's correlation coefficient, which evaluates the strength of a linear relationship between two variables, is another typical time domain approach for similarity estimation. It is computed by (2.3), where $\overline{X_1}$ and $\overline{X_2}$ are, respectively, the average values of time series $X_1(t)$ and $X_2(t)$.

$$CC = \frac{\sum_{t=1}^{N}(x_1(t) - \overline{X_1})\ (x_2(t) - \overline{X_2})}{\sqrt{\sum_{t=1}^{N}(x_1(t) - \overline{X_1})^2}\ \sqrt{\sum_{t=1}^{N}(x_2(t) - \overline{X_2})^2}} \tag{2.3}$$

$$\overline{X_i} = \frac{1}{N}\sum_{t=1}^{N} x_i(t), \quad i = 1,2 \tag{2.4}$$

This correlation value (CC) is a number in the range $[-1,1]$ that, as referred, assesses the degree of association between two time series. In case of $CC = +1$, a perfect positive linear relationship (correlation) occurs; in case of $CC = -1$, a perfect negative linear relationship (anti-correlation) occurs; in case of $CC = 0$ (zero), there is no correlation between signals (uncorrelated).

Despite their recognized value, some limitations of the Euclidean distance and correlation measures have been identified by several authors. In effect, Shasha and Zhu (2004) showed that the Euclidean distance is not a flexible similarity measure between time series. First of all, it can not be applied to time series of different lengths, even though these time series are similar to each other. In fact, this distance metric, as well as the correlation measures, assumes that both sequences have the same length and are uniformly sampled. To overcome these constraints, modifications have been introduced based on the principle of dynamic time-warping (DTW), where signals are "stretched" or "compressed" so that they can be compared (Keogh *et al.*, 2001).

On the other hand, two time series can be similar even though they have very different baselines (mean values) or amplitudes. Euclidean distance algorithms in particular, and most time series similarity metrics in general, assume that signals are aligned so that "similar" signals will have analogous dynamics at the same points in time. Windowing and segmentation techniques are examples of alternatives that have been developed to divide a signal into a set of subsequences, which allow for a greater flexibility in matching time series by means of shifting operations (Hetlan, 2004).

Moreover, the use of the linear correlation coefficient is limited to cases where there exists a linear transformed relationship (Bernhard and Darbellay, 1999). Since every real-world time series, such as clinical biosignals, present non-linear dependencies, measurements of non-linear relationships have been therefore developed. Mutual information, originated from the Shannon's theory of communication, has been proposed as a global effective similarity measure for time series (Bernhard and Darbellay, 1999). Mutual information is derived from the computation of the entropy of a variable $X(k)$, $H\big(X(k)\big)$, defined as the uncertainty about $X(k)$. For discrete distributions the entropy is given by (2.5).

$$H\big(X(k)\big) = -\sum_{k=1}^{N} p_k \log p_k \tag{2.5}$$

In the previous equation the variable $X(k)$ is a discrete random variable (time series) defined in the range specified by (2.6), and p_k is the probability defined according to (2.7).

$$X(k) = \Big\{\, x(k) \,|\, k = 1,\ 2,\ ...,\ N \,\Big\} \tag{2.6}$$

$$p_k = P\Big(\, X(k) = x(k) \,\Big) \tag{2.7}$$

Considering discrete time variables $X_1(k)$ and $X_2(k)$, defined as $X(k)$, equation (2.6), the so-called joint entropy $H\big(X_1(k), X_2(k)\big)$, simplified to $H\big(X_1, X_2\big)$, is given by equation (2.8).

$$H(X_1, X_2) = -\sum_{x_1 \in X_1} \sum_{x_2 \in X_2} p(x_1, x_2) \log p(x_1, x_2) \tag{2.8}$$

In the expression above, $p(x_1,x_2)$ is the joint probability density function. Given these definitions, the mutual information between the two time series is obtained through (2.9).

$$I(X_1, X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \tag{2.9}$$

Based on mutual information, standard (normalized) similarity measures have been derived (Cheong, 2004). One of such measures is the global correlation coefficient defined as (2.10), where the D_{MI} is a value in the range $[0,1]$.

$$D_{MI} = \sqrt{1 - e^{-I(X_1, X_2)}} \tag{2.10}$$

This measure, that is able to capture the global dependence of X_1 and X_2, is composed of both linear and non-linear components.

In any case, the explicit use of time domain methods, such as Euclidean distances and linear and non-linear correlations, is usually a very demanding computational task (order of $O(N^2)$). In fact, a time series of length N can be seen as a tuple in the N-dimensional space. Consequently, working directly in this space is inefficient given its high dimensionality. Therefore, a possible alternative is the use of dimensionality reduction techniques, that map the N dimension time series into a lower M dimensional space (ideally, $M \ll N$). In addition to this, other advantages have been recognized to the transform-based techniques. In fact, the direct time domain comparison between two time series can be complex, since the discrimination of the relevant characteristics may be not straightforward. This shortcoming can be overcome by transforming the time series into a different space where any similarity matching is more meaningful. Another direct benefit relates to noise reduction. Since time series data are normally deteriorated by noise, one positive consequence of dimensionality reduction is noise attenuation, which can considerably contribute to improve the similarity search process.

2.3.3 Transform-based Methods

As mentioned, the transform-based techniques map high-dimensional vectors (time series) into lower-dimensional ones (features or parameters). Since the data transformation reduces the dimensionality of the original signal, it facilitates the development and use of algorithms for matching similar time series. The main goal is to extract specific characteristics of data that can reveal valuable information of the time series, otherwise hidden or inaccessible.

Numerous transforms have been introduced and several possible classification approaches have been proposed to categorize them. In one of these, transforms can be simply classified as quantitative or qualitative techniques (Bravi *et al.*, 2011). All the methods belonging to the first category make use of specific mathematical models to transform the time series. The second group, composed of the qualitative transforms, aims to quantize the time series without imposing any particular models. Because quantization is a many-to-few mapping, where a large set of input values is mapped into a smaller set, its potential advantage is to make easier handling problems associated with noise in the data and make the analysis more robust to artefacts. Conversely, the loss of potentially valuable information is, probably, the major drawback.

On the other hand, from the data mining perspective, data dimensionality reduction techniques can be classified in two groups: feature extraction and feature selection (Liu and Motoda, 1998). Feature extraction approaches (that include feature-based and model-based methodologies) derive a set of features (or parameters) from the original time series through some procedure or transform (Wyse *et al.*, 1980). Ideally, these features should be representative of the main characteristics of the original data. Feature selection is a process that selects a subset of the extracted features.

Numerous feature extraction algorithms have been proposed. Examples of these are the Karhunen-Loève transform (KLT), also known as singular value decomposition (SVD) or principal component analysis (PCA), the discrete Fourier transform (DFT), the discrete wavelet transform (DWT), the piecewise linear representation (PLR), (Keogh and Pazzani, 1998), and the piecewise constant approximation indexing (PCA-indexing), (Keogh and Pazzani, 2000), (Keogh *et al.*, 2001).

In parallel, numerous modelling approaches have been presented. Namely, the power of autoregressive (AR) and autoregressive moving average models (ARMA), linear or non-linear, has been extensively exploited in this context. Of the non-linear methods, neural networks have become very popular. Many different types of neural networks, such as multi-layer perceptron and radial basis function, have been proven to be universal function approximators, which make them attractive for classification and for time series modelling and prediction. Fuzzy approaches have also been valid alternatives for modelling purposes, mainly given their interpretability and their capacity to express and incorporate human reasoning.

Given that feature extraction procedures, by means of some quantitative transform method, are the most common approaches for time series analysis, they represent a central focus of this work. As a consequence, some of them will be subsequently described.

Karhunen-Loève transform

Among the proposed feature extraction techniques, Karhunen-Loève transform also known as Singular Value Decomposition and Principal Component Analysis in statistical literature, is one of the most effective algorithms, guaranteeing a minimal reconstruction error (in the sense that it minimizes the total mean squared error). Basically, a time series is mapped into an orthogonal space, where variables composing the base are orthogonal to each other. Using this space representation, time series can be approximated by a reduced set of basis, by discarding the variables with lower energy. The key idea is that the coefficients associated to these fundamental basis, can be employed to support the similarity search process, avoiding the use of the signal itself.

Given a time-series signal $X(t) \in \mathbb{R}^{1,N}$, the goal is to determine a set of J orthogonal basis, $\varphi_j(t) \in \mathbb{R}^{1,N}$, $j = 1,\ldots,J$, and the respective coefficients, $\{c_1, c_2, \ldots, c_J\}$, such that the signal $X(t)$ can be described as (2.11).

$$X(t) = \sum_{j=1}^{J} c_j \ \varphi_j(t) \tag{2.11}$$

Fundamentally, these orthogonal basis functions are obtained as the eigenvectors (also known as principal components) of a covariance matrix $R \in \mathbb{R}^{N,N}$, composed of all possible N-dimensional basis. From this matrix, N eigenvectors, $v_k \in \mathbb{R}^{1,N}$, and the respective eigenvalues, λ_k ($k = 1,\ldots,N$), are obtained by solving (2.12).

$$R\ v_k^T = \lambda_k\ v_k \tag{2.12}$$

The eigenvectors, $\{v_1, v_2, \ldots, v_N\}$, form an orthogonal base. The vectors to be employed as basis can be established by considering the descending order of magnitude of the corresponding eigenvalues λ_k, that is,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_J \geq \ldots \geq \lambda_N \tag{2.13}$$

Assuming this order, the optimal set of basis (eigenvectors) is obtained by taking the first J coefficients, $\{\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_J\}$, containing most of the information about the signal to be represented. The main aspect is that KLT expansion is optimal in the sense that it minimizes the mean squared error approximation. This means that the best approximation of a signal $X(t)$ can be achieved through a linear combination of the J basis that corresponds to the highest J eigenvalues.

When applied in time-series indexing, KLT presents a time complexity $O(MN^2)$, where M is the number of time series in a dataset collection and N is the length of each time series.

Hermite transform

Similarly to the KLT, the Hermite transform decomposes a signal as a linear combination of orthogonal basis functions. However, unlike the former that is signal dependent, Hermite transform has the advantage to be signal independent, since the set of basis functions are predefined and do not require any prior knowledge of the dataset. This reason, coupled with its ability in capturing specific morphologic changes using a low number of basis functions, has led to the employment of the Hermite expansion, namely to identify subtle variations in a signal (Lagerholm, 2000). These properties will be explored in the detection of ischemic episodes, where the main goal is to identify variations in the morphology of the QRS complex in an electrocardiogram. In fact, given the similarity of some of the Hermite basis functions with the QRS complex, ischemia detection is a typical application where they can be successfully employed to capture small morphologic alterations.

Basically, the Hermite functions represented by $H^n(t,l)$, where t denotes time and l a scaling factor, form an orthonormal base of $L^2(\mathbb{R})$, the space of square-integrable functions. They can be determined as the product of a Gaussian by the Hermite polynomials with some normalization constants (Fedoryuk, 2001), as described by (2.14).

$$H^n(t,l) = \frac{1}{\sqrt{n!2^n\sqrt{\pi}l}} e^{\frac{-t^2}{2l^2}} P^n(t/l) \tag{2.14}$$

In the previous equation, $P^n(t/l)$ represents the Hermite polynomial of order n, and l a scaling factor that allows width adjusting.

The Hermite polynomials at a time instant t, can be determined by the following recursive relations:

$$\begin{aligned} P^o(t) &= 1 \\ P^1(t) &= 2t \\ P^n(t) &= 2tP^{n-1}(t) - 2(n-1)P^{n-2}(t) \end{aligned} \tag{2.15}$$

As an example, Figure 2.4 shows the first six Hermite functions ($n = 0,1,\ldots,5$), considering the scaling factor l=3.

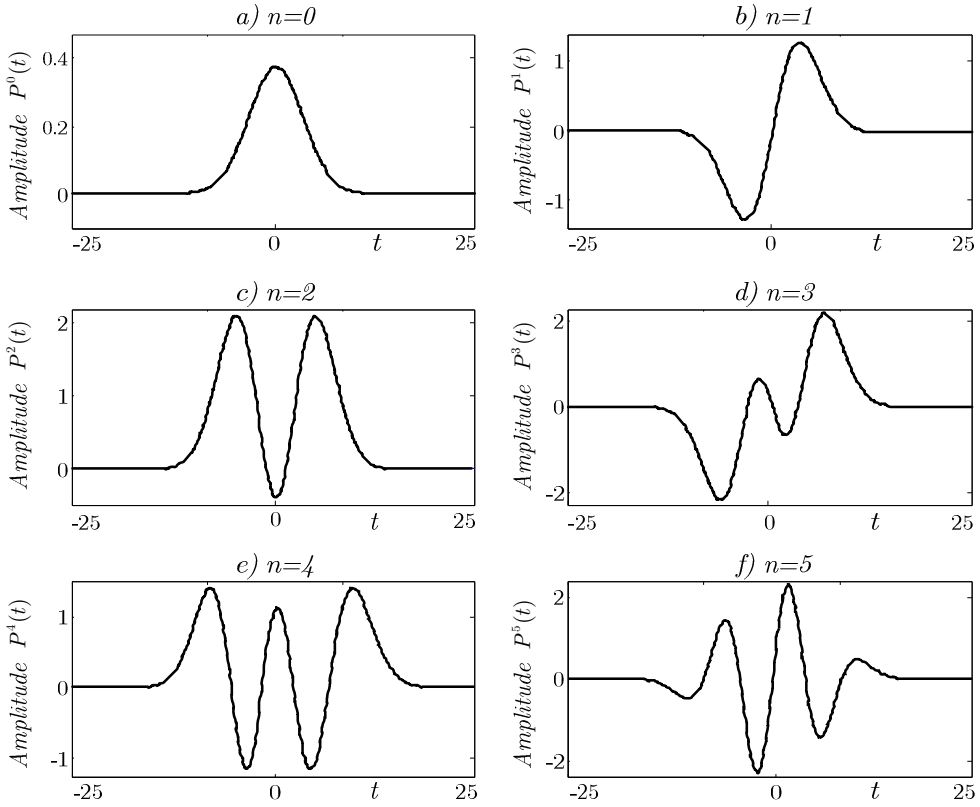


Figure 2.4 - The first six Hermite functions (order n=0 to n=5, and scaling factor l=3).

In the context of similarity assessment, the goal is to transform a high-dimensional vector (time series), $X(t)$, into a lower-dimensional space, by describing it as a linear combination of Hermite basis functions $H^j(t,l)$, with coefficients c_j (similarly to the representation in (2.11)), as described by equation (2.16).

$$X(t) = \sum_{j=0}^{J-1} c_j \; H^j(t,l) \tag{2.16}$$

The reduced set of coefficients, c_j, that reflect the main dynamics of the signal, can then be straightforwardly employed to support the assessment of similarity.

Discrete Fourier transform

The discrete Fourier transform (DFT) is, definitely, one of the most powerful signal processing techniques. Basically, the DFT decomposes a sequence of values (time series) into components of different frequencies, enabling to identify similarities between time series. Moreover, fast computational algorithms, such as the fast Fourier transform (FFT), can compute the DFT coefficients in $O(N \log N)$ times, being N the length of the time series (computing DFT through the Fourier definition takes $O(N^2)$ arithmetical operations).

The original work of Agrawal *et al.* (1993) utilizes the DFT to perform dimensionality reduction on the data. They suggest the use of the DFT for feature extraction, arguing that most of the real signals need only the first few Fourier coefficients to approximate them. Then, by means of methods in the transformed space (frequency domain), it is possible to improve (speed and accuracy) the similarity searching process. Faloutsos *et al.* (1994) extended the work of Agrawal to perform subsequence similarity indexing using a discrete Fourier transformation. The scheme considers the mapping of each time series into a small set of multi-dimensional rectangles in feature space. Subsequently, these rectangles are easily indexed using traditional spatial access methods, like R-tree. A sliding window over the data sequence is used to extract features. Moon *et al.* (2002) used a generalized window procedure to reduce the false negatives of Faloutsos method. The algorithm, called general match, divides data sequences into generalized sliding windows and the query sequence into generalized disjoint windows, thus achieving no false dismissal.

Wavelet transform

The discrete wavelet transform (DWT) is a powerful and efficient technique for decomposing, denoising, compressing and analysing time series signals. Unlike DFT, which takes the original signal from the time domain and transforms it into the frequency domain, DWT transforms the time series from the time domain into time-frequency domain. It has the time-frequency localization property, which means that it is able to give locations in both time and frequency. Basically, the DWT breaks a signal into several time-frequency components enabling the extraction of features suitable for signal analysis, namely, for similarities evaluation. Moreover, if some particular type of wavelets are considered (such as the Haar wavelet), it is possible to achieve linear time complexity, $O(N)$, which makes DWT more efficient than DFT.

Given these properties, wavelet transform has been extensively used as a feature extraction methodology (Carreño and Vuskovic, 2005), (Percival and Walden, 2000). Huhtala *et al.* (1999) proposed a wavelet transformation to produce a natural set of features in order to search similarities in aligned time series. The time series can have different vertical positions, scales, and overall trends but still presenting related features at the same locations. Chan and Fu (1999) proposed the use of the Haar wavelet transform for time series indexing, showing that Euclidean distance is preserved in the Haar transformed domain and that no false dismissals occur. Additionally, they

concluded that Haar transform can outperform DFT and that the method is able to accommodate vertical shift of time series. Gilbert *et al.* (2003) presented techniques for computing small space representation of massive data streams using wavelet-based approximations that consist of specific linear projections of underlying data. By means of various linear projections of the data and using them to provide point wise and range sum estimation of data stream, the method uses small amount of space and per-item time, as well as provides accurate representation of data. In the same context, Chan *et al.* (2003) used the Haar wavelet for time-series classification and showed performance improvement over DFT. Popivanov and Miller (2002) proposed an algorithm using the Daubechies wavelet for efficient similarity over time series.

2.4 Time-Frequency Transforms

The emergence of the time-frequency transforms was a consequence of the limitations that have been recognized to the Fourier transform. These techniques are applicable in situations where the time localization of the spectral components of a signal is useful, since they allow for the representation of a signal in terms of time and frequency, simultaneously, unlike the Fourier transform that only gives information about what frequency components exist.

Two classes of time-frequency transforms can be considered. In the first class, the transforms decompose the signal into basic components, well localized in time and in frequency (also known as linear time-frequency representations). The short-time Fourier and Wavelet transforms are examples of these types of methodologies. In the second class, the approach is to develop a joint function of time and frequency, known as time-frequency distribution that can describe the energy density of a signal in both time and frequency, simultaneously. The Wigner-Ville transform belongs to this second class.

Essentially, the Fourier transform uses a sum of sine and cosine functions of different wavelengths to represent a given time series signal. Since sine and cosine are global periodic functions, they are inherently non-local. In fact, as they extend over the entire real axis, any change at a particular point in the time domain has an effect on the entire signal. Thus, it is assumed that the frequency content of the function is stationary along the time axis. This limitation has led to the development of the Short Time Fourier transform, in which data is divided into several windows along the time axis and the Fourier transform is taken for each window separately. Since it is supposed that the signal inside a window is stationary, the size of the window is very important for obtaining accurate results. The drawback of this technique is that the window size is fixed.

As referred, the development of time-frequency methods, namely the Wigner-Ville and wavelet transforms, allowed overcoming the referred limitations. In particular, wavelet transform, by means of a scalable windowing technique, enables to adjust window size depending on if the focus is on frequency resolution or on time resolution. It makes use of long time intervals if the interest is on precise low-frequency information, or short time intervals if the goal is to capture high-frequency information. One major advantage of the wavelet transform is its capacity to analyse a localized area in a large signal. In fact, in opposition to other signal analysis techniques, it allows that some aspects of data, as trends, breakdown points, and self-similarity, can be revealed.

2.4.1 Short-time Fourier Transform

Fourier transform

The Fourier transform (FT) is the oldest frequency domain technique available for analysing time series data. It consists in a mathematical tool that transforms a given signal from the time domain to the frequency domain. According to the Fourier theory, a signal may be decomposed in several sinusoidal components (also known as basis functions), possibly with different amplitudes, phases and frequencies. For a continuous signal, $X(t)$, the Fourier transform is given by:

$$X(f) = \int_{-\infty}^{+\infty} X(t)\ e^{-j2\pi ft} dt \tag{2.17}$$

In the previous equation $X(f)$ is the Fourier transform of the signal $X(t)$ and $e^{-j2\pi ft} = \cos(2\pi ft) + j\sin(2\pi ft)$ gives the frequency components of the signal.

In the case $X(t)$ is a discretely sampled signal or time series, the discrete version of the Fourier transform (DFT) is computed as:

$$X(f) = \sum_{t=-\infty}^{+\infty} X(t)\ e^{-j2\pi ft} \tag{2.18}$$

As referred before, the limitation of the Fourier transform is that it only gives information about what frequency components exist in the signal but no indications about where those spectral components appear in time. Therefore, it is not a suitable technique for non-stationary signals, whose frequency content changes with time.

Short-time Fourier transform

In an attempt to overcome the limitation presented by the Fourier transform, a windowed version of the FT was developed which was called Short-Time Fourier transform (STFT). Basically, the STFT extracts several sections of the signal to be analysed by means of a sliding window along the time axis. If the time window is sufficiently narrow, each extracted section can be considered as stationary so that Fourier transform can be applied. Each resultant frequency spectrum shows the frequency content during a short time, and, so, the successive spectra show the evolution of frequency content with time. The STFT of a continuous signal $X(t)$ is a function of two variables, time and frequency, and is calculated by:

$$X(\tau, f) = \int_{-\infty}^{+\infty} X(t)\ g(t-\tau)\ e^{-j2\pi ft} dt \tag{2.19}$$

In the previous equation, $g(t)$ is a window function and $X(\tau, f)$ is the Fourier transform of $X(t)g(t-\tau)$, a complex function that represents the phase and magnitude of the signal over time and frequency.

As previously stated, the weakness of the Short Time Fourier transform is the time window fixed length. In fact, the shorter the window is, the higher the time resolution is. Nevertheless, this is usually accompanied by a poor frequency resolution. For a long window, the frequency resolution is high, but the time resolution is low. This phenomenon reflects the Heisenberg's uncertainty principle. Therefore, there is always a trade-off between time resolution and frequency resolution in STFT.

One possible solution to overcome this drawback is to use time-frequency approaches, namely Wigner-Ville and Wavelet transforms.

2.4.2 Wigner-Ville Distribution

Following a different approach, the Wigner-Ville transform (or distribution) is an energy distribution methodology that fundamentally describes a signal, simultaneously in time (t) and in frequency (f), by considering the autocorrelation function. The Wigner-Ville distribution (WVD) $W_X(t,f)$ of a continuous time signal $X(t)$, is defined as (2.20),

$$W_X(t,f) = \int_{-\infty}^{\infty} X(t+\frac{\tau}{2})\ X^*(t-\frac{\tau}{2})\ e^{-j2\pi f\tau}\ d\tau \tag{2.20}$$

where $r(\tau) = X(t+\frac{\tau}{2})X^*(t-\frac{\tau}{2})$ is the instantaneous autocorrelation function and the operator $(*)$ indicates the conjugate operation.

In an analogy to the STFT, in the case of the WVD the window is basically a shifted version of the signal itself. In fact, the WVD is obtained by comparing the information of the signal with its own information at other times and frequencies.

The related discrete time transform $W_X(nT,f)$ is given by equation (2.21).

$$W_{\mathrm{X}}(nT,f) = 2T\sum_{k=-L}^{L} X(n+k)\ X^*(n-k)\ w(k)\ w^*(-k)\ e^{-j4\pi f\ k} \tag{2.21}$$

In the previous equation, T represents the sampling period and w is a sliding window, symmetrical and with finite-length duration, verifying $w(kT)=0$ for $abs(k)>L$. This relationship defines the discrete WVD at the time origin. At any other point in time, the discrete WVD can be obtained by shifting the signal $X(t)$, so that is mapped on the time origin.

This distribution satisfies several desirable mathematical properties. In particular, the WVD is always real-valued, it preserves time and frequency shifts and has the ability to provide a high-resolution representation in both time and frequency for non-stationary signals. Additionally, the Wigner-Ville transform presents a constant resolution. Thus, specifically in the low frequencies, this transform preserves a good time-frequency resolution, while not presenting serious artefacts as in higher frequencies.

Although WVD does not suffer from interaction between time and frequency resolutions, it presents some other undesired properties, such as cross-term interferences. Because WVD is a linear combination of auto and cross correlation terms, each pair of signal components creates one additional cross-term in the spectrum, which introduces difficulties in obtaining the desired time-frequency representation. One way of lowering cross-term interference is to apply a low-pass filter to the WVD, but such smoothing reduces its frequency resolution (Qian and Chen, 1996). Another alternative is to use the WVD of the analytic signal, which avoids all cross-terms associated with negative frequency components.

Considering a real time signal, the equivalent analytic signal can be obtained by adding to the real signal, its Hilbert transform as the imaginary part, as described by equation (2.22).

$$X_a(t) = X(t) + j\ H\big[X(t)\big] \tag{2.22}$$

In the above equation, $X(t)$ represents the real signal, $X_a(t)$ denotes the equivalent analytic one, and $H[\cdot]$ corresponds to the Hilbert transform. This definition has a simple interpretation in the frequency domain, since $X_a(t)$ is a single-sided Fourier transform where the negative frequency values have been removed and the strictly positive ones have been doubled. However, the analytic function differs from the original signal in several ways. For example, its instantaneous properties may substantially diverge from those of the original signal (Martin and Flandrin, 1985). In fact, the cross-terms must be present to guarantee the notable properties of the WVD. Thus, there is trade-off between the interferences and the desirable properties of the WVD.

2.4.3 Wavelet Transform

The general idea behind wavelet transform is very similar to the Fourier analysis in the sense that both approximate a signal by means of a set of basis functions that, in the case of wavelet transform, are called wavelets. In mathematical terms, wavelets are orthogonal bases consisting of small waves. In comparison to the sinusoidal waves of the Fourier transform, which are smooth and of infinite length, the wavelets are irregular in shape and compactly supported (they have non-zero values only in a finite time interval). These characteristics make wavelets an ideal tool for analysing signals of non-stationary nature. In fact, their irregular shape makes them suitable to analyse signals with discontinuities or sharp changes, while their compact support enables the temporal localisation of signal features.

It was referred that the Fourier transform has no capacity to associate features in the frequency domain with their time location, since an alteration in the frequency domain causes changes throughout the time axis. On the contrary, the wavelet transform allows

the localisation in both the time and the scale domain through translations and dilations of a wavelet function $\psi(t)$, called the mother wavelet (Akay, 1997).

It is important to refer that scale is inversely related with frequency. In fact, a low scale corresponds to a high frequency, while a high scale corresponds to a low frequency.

Continuous wavelet transform

The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by a set of basis functions, or wavelets, that are scaled, shifted versions of the wavelet function, or mother wavelet, $\psi(t)$. Wavelets, represented by $\psi_{a,b}(t)$, can be generated as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \ \psi\left(\frac{t-b}{a}\right) \tag{2.23}$$

In the previous equation, the parameter $a \in \mathbb{R}^+$ reflects the scale, that is, the width of a particular basis function, and is defined as $(1/\textit{frequency})$; the parameter $b \in \mathbb{R}$ specifies the translated position along time axis and the factor $(\frac{1}{\sqrt{a}})$ guarantees that wavelets are normalized, that is, at every scale all wavelets have the same energy. Therefore, continuous wavelet transform of $X(t)$ can be defined as:

$$W(a,b) = \int X(t) \ {\psi_{a,b}}^*(t) \ dt \tag{2.24}$$

where the character $(*)$ denotes the complex conjugation. To summarize, the continuous wavelet transform is calculated by continuously shifting a continuously scalable function over a signal and calculating the correlation between them. This process represents a considerable amount of work and originates data redundancy since collected information is more than enough for reconstructing the original signal.

Discrete wavelet transform

To overcome the problem of data redundancy, discrete wavelets were introduced. These are not continuously scalable and translatable but can only be scaled and translated in discrete steps. They can be defined by:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{{s_o}^j}} \ \psi\left(\frac{t - k\tau_o {s_o}^j}{{s_o}^j}\right) \tag{2.25}$$

Although they are called discrete wavelets, usually they are continuous functions by sections. In the previous equation, j and k are integers, $s_o > 1$ is a fixed dilation step and τ_o is a translation factor that depends on s_o.

The wavelet discretization originates the segmentation of the time-scale space into discrete intervals. Usually scales and positions are chosen to be based on powers of two $(s_o = 2$ and $\tau_o = 1)$, resulting in a dyadic time-scale grid sampling. In this case, from

(2.25), the wavelet basis functions originated from dilations and translations of the mother function $\psi(t)$, form an orthonormal set represented by (2.26).

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \; \psi\left(\frac{t - k \; 2^j}{2^j}\right) \tag{2.26}$$

In the previous equation, j and k are integers that control the wavelet dilation and translation, respectively. Typically this equation (2.26) is presented in a equivalent form, as follows (2.27).

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}} \; \psi\left(2^{-j} t - k\right) \tag{2.27}$$

Using this set of orthonormal basis functions, $\psi_{j,k}(t)$, it is possible to describe a time series signal as (2.28), where $c_{j,k}$ are the coefficients corresponding to each wavelet basis $\psi_{j,k}(t)$.

$$X(t) = \sum_j \sum_k \; c_{j,k} \psi_{j,k}(t) \tag{2.28}$$

Multi-resolution analysis

For the practical implementation of the DWT, an efficient method was developed by Mallat (1989), which uses a technique called quadrature mirror filters (a pair of high and low-pass filters). Mallat showed that using this multi-resolution analysis (MRA) technique, any discrete wavelet transform could be efficiently performed by means of a pyramid algorithm. In fact, wavelet transform based on multi-resolution algorithm has linear computational complexity of $O(N)$ (clearly more efficient than the discrete Fourier transform, which has a complexity of order $O(N^2)$).

In MRA, firstly the original signal $X(t) = \{x(1), \; \ldots, \; x(t), \; \ldots\}$ is separately passed through a highpass filter ($H_f = \left[h_0, \; \ldots, \; h_{n_f - 1}\right]$), as well as through a lowpass filter ($L_f = \left[l_0, \; \ldots, l_{n_f - 1}\right]$). The parameter n_f represents the filter length, that is, the number of non-zero coefficients, and the parameters h_i and l_i represent the filter coefficients.

From this procedure two sequences are generated: *i*) by convolving the signal $X(t)$ with the lowpass filter (L_f), equation (2.29), a sequence $A_1(t)$ is obtained; *ii*) by convolving the signal $X(t)$ with the highpass filter (H_f), equation (2.30), a sequence $D_1(t)$ is achieved.

$$A_1(t) = \sum_k \; l_{t-k} \; x(t) \tag{2.29}$$

$$D_1(t) = \sum_k \; h_{t-k} \; x(t) \tag{2.30}$$

The application of the referred filters to a signal (vector) with N elements originates two vectors with $N/2$ elements each: one containing the data smoothed by the low-pass filter ($A_1(t)$), which represents the signal approximation, and the other holding the removed detail information ($D_1(t)$). Successively applying such filters to the approximation, the subsequent scales are generated, following a decomposition tree scheme, as illustrated in Figure 2.5. In this way, a multi-level decomposition process can be achieved (with L levels, being $L = \log_2(N)$) where, at each level J, the approximation ($A_J(t)$) is successively decomposed into approximation ($A_{J+1}(t)$) and detail ($D_{J+1}(t)$) components.

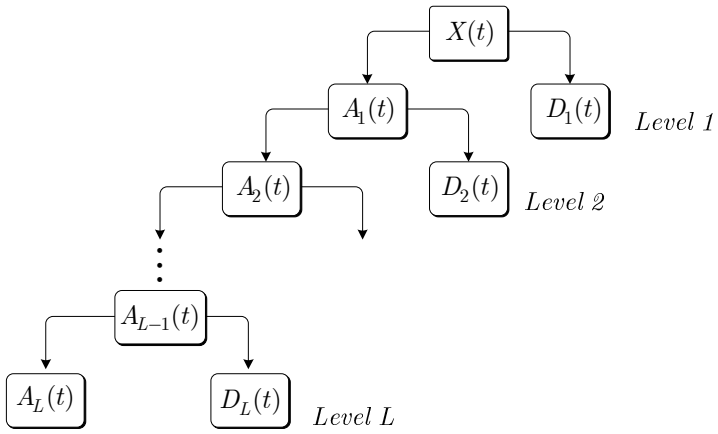


Figure 2.5 - Wavelet decomposition tree.

Using this procedure, the original signal can be described, at any level J of decomposition, as a function of the current approximation, $A_J(t)$, and the current and previous detail components, $D_J(t)$, $D_{J-1}(t)$, ..., $D_1(t)$.

$$X(t) = f\Big(A_1(t),\, D_1(t) \Big) \tag{2.31}$$

$$X(t) = f\Big(A_2(t),\, D_2(t),\, D_1(t) \Big) \tag{2.32}$$

$$X(t) = f\Big(A_J(t),\, D_J(t),\, D_{J-1}(t),\, ..., D_1(t) \Big) \qquad J = 1,\ldots,L \tag{2.33}$$

The reconstruction of the signal can be achieved by, at every level, upsamplig the signals by two, passing them through reconstruction filters and, finally, adding them.

This hierarchical decomposition via filter banks essentially describes a signal in terms of basis. In fact, the decomposition procedure can be achieved by means of wavelet and scaling functions, respectively, $\psi(t)$ and $\phi(t)$. The first, known as mother wavelet effectively represents the detail and high-frequency parts of time series, while the latter, known as father wavelet, represents the smooth and low-frequency components. Just as the wavelet functions (equation (2.27)), scaling functions $\phi(t)$ can be derived by (2.34).

$$\phi_{j,k}(t) = 2^{-\frac{j}{2}}\ \phi\left(2^{-j}t - k\right) \tag{2.34}$$

According to (2.33), the signal $X(t)$ can be rewritten at any level J of decomposition, using scaling and wavelet basis functions, as follows:

$$X(t) = c_{0,0}\phi_{0,0}(t) + \sum_{k=0}^{2^J-1} d_{J,k}\psi_{J,k}(t) + \sum_{k=0}^{2^{J-1}-1} d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_{k=0}^{2^0-1} d_{0,k}\psi_{0,k}(t) \tag{2.35}$$

where $c_{J,k}$, $d_{J,k}$, ...,$d_{1,k}$ are the wavelet transform coefficients, and $\phi_{j,k}(t)$, $\psi_{j,k}(t)$ are the basis functions, respectively scaling and wavelets. Since the basis are orthonormal, the coefficients $c_{J,k}$ and $d_{J,k}$, ...,$d_{1,k}$ can be easily computed by the inner product of the signal with the appropriate function, according to (2.36) and (2.37).

$$c_{j,k} = \sum_t x(t)\ \phi_{j,k}(t) \tag{2.36}$$

$$d_{j,k} = \sum_t x(t)\ \psi_{j,k}(t) \tag{2.37}$$

Mother and father functions

The starting point of the wavelet decomposition is the so-called father wavelet (or scaling function), $\phi(t)$, from where mother wavelet (or wavelet function), $\psi(t)$, can be derived. As referred, the former represents the smooth and low-frequency components, while the latter represents the detail and high-frequency parts of the time series. Therefore, father wavelet can be used for the identification of trend components and mother wavelet for identifying deviations from the trend. According to the definition, the scaling and wavelet functions verify the relationship:

$$\int \phi(t)\ dt = 1\,,\ \int \psi(t)\ dt = 0 \tag{2.38}$$

Moreover, the scaling function must obey to the dilation equation (2.39), where l_k represents the scaling filter coefficients, as well as to equation (2.34).

$$\phi(t) = \sqrt{2}\ \sum_k l_k\ \phi(2t - k) \tag{2.39}$$

For the mother wavelet function $\psi(t)$, a similar relationship is obtained from the father wavelet, where h_k represents the wavelet filter coefficients.

$$\psi(t) = \sqrt{2}\ \sum_k h_k\ \phi(2t - k) \tag{2.40}$$

Since basis functions are orthonormal, the coefficients l_k and h_k, corresponding to lowpass filter ($L_f = \left[l_0, \ \ldots, l_{n_f-1}\right]$) and to highpass filter ($H_f = \left[h_0, \ \ldots, \ h_{n_f-1}\right]$), can be computed as follows:

$$l_k = \frac{1}{\sqrt{2}} \sum_t \ \phi(t) \ \phi(2t-k) \tag{2.41}$$

$$h_k = \frac{1}{\sqrt{2}} \sum_t \ \psi(t) \ \phi(2t-k) \tag{2.42}$$

Moreover, these coefficients are related through equation (2.43).

$$h_k = (-1)^k l_{n_f-k-1}, \ k = 0, 1, \ldots, n_f - 1 \tag{2.43}$$

Firstly, the scaling function is chosen to preserve the area prerequisite (2.38), known as stability condition. The following equality can thus be derived for scaling coefficients:

$$\sum_{k=0}^{n_f-1} l_k = \sqrt{2} \tag{2.44}$$

Secondly, the convergence of wavelet expansion requires that (2.45) is verified, where $m = 0,1,2, \ \ldots, \frac{n_f}{2} - 1$.

$$\sum_{k=0}^{n_f-1} (-1)^k k^m l_k = 0 \tag{2.45}$$

Subsequently, the orthogonality of wavelets is guaranteed by (2.46).

$$\sum_{k=0}^{n_f-1} l_k l_{k+2m} = 0 \tag{2.46}$$

Finally, the scaling function has to be orthogonal, thus, equation (2.47) should be verified.

$$\sum_{k=0}^{n_f-1} l_k^2 = 1 \tag{2.47}$$

Typical wavelets

Although wavelets can present considerably different configurations, they all share the same basic structure. In fact, generically a wavelet basis consists of a father wavelet, $\phi(t)$, that represents the smooth baseline trend, and a mother wavelet, $\psi(t)$, that is dilated and translated to generate different levels of detail. Several families of wavelets have proven to be especially useful in various applications. They differ with respect to orthogonality, smoothness and other related properties, such as vanishing moments or size of the support. Among others, examples of well-known wavelets are the Haar,

Daubechies, Symmlets and Coiflets wavelet families. Figure 2.6 shows some of these wavelets, their corresponding mother and father functions. The Haar wavelet is a square wave and, contrasting with the others, it is not continuous. It is the only one that is simultaneously orthogonal and symmetric.

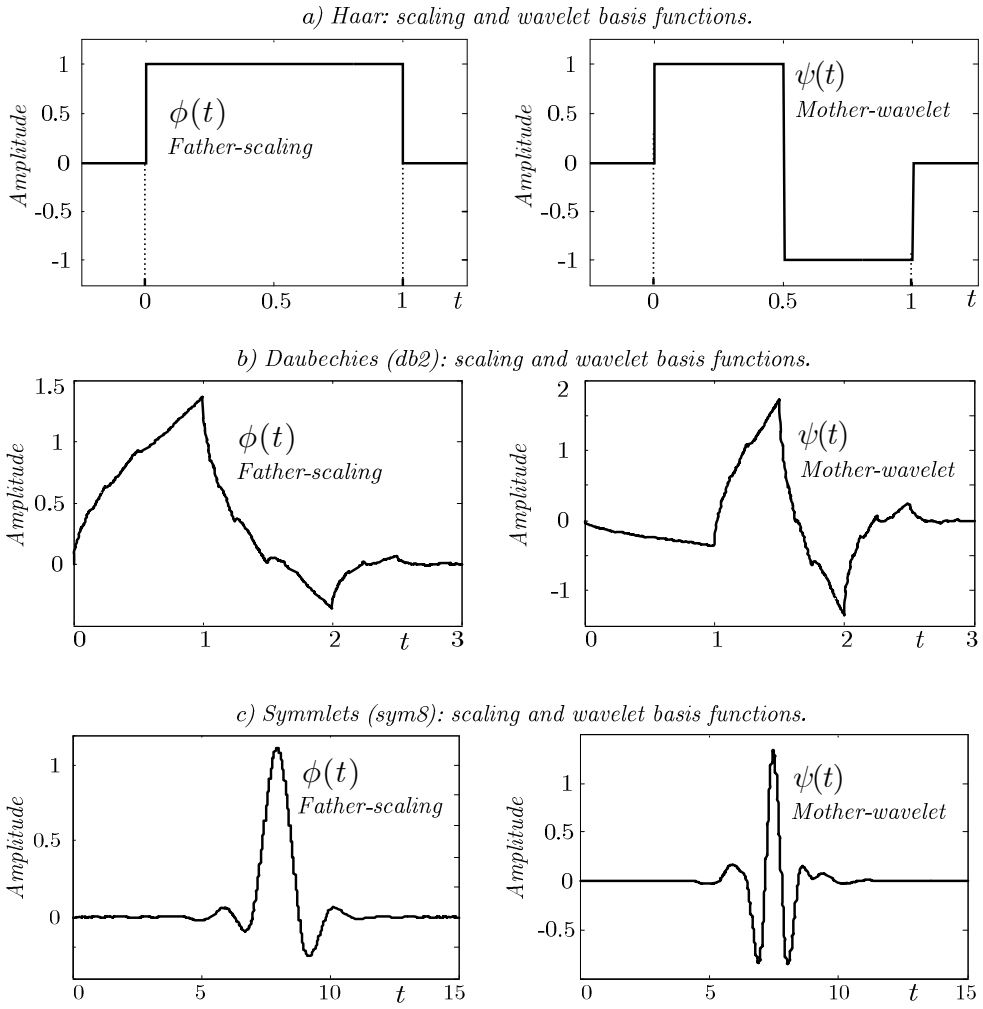


Figure 2.6 - Examples of orthogonal wavelets (mother and father basis functions).

i. Haar wavelet

The filter coefficients of the Haar wavelet ($n_f = 2$, $m = 0$), verifying equations (2.44), (2.45) and (2.47), are the following (Li *et al.*, 2002):

$$\begin{cases} l_0 + l_1 = \sqrt{2} & \textit{stability} \\ l_0 - l_1 = 0 & \textit{convergence} \\ l_0^2 + l_1^2 = 1 & \textit{scaling orthogonality} \end{cases}$$

The solution (unique) is given by

$$l_0 = l_1 = \frac{1}{\sqrt{2}} \tag{2.48}$$

From (2.43) it is possible to compute the corresponding wavelet coefficients:

$$h_0 = \frac{1}{\sqrt{2}} \qquad h_1 = -\frac{1}{\sqrt{2}} \tag{2.49}$$

The Haar scaling function satisfies (2.39), while wavelet function verifies the equation (2.40). In particular, for $k = 0,1$ and using the values of $l_0,\ l_1,\ h_0,\ h_1$, it follows:

$$\phi(t) = \phi(2t) + \phi(2t-1) \tag{2.50}$$

$$\psi(t) = \phi(2t) - \phi(2t-1) \tag{2.51}$$

The solution for this recurrence is the Haar scaling function (2.52), and the corresponding wavelet function (2.53), as shown in Figure 2.6*a*).

$$\phi(t) = \begin{cases} 1 & if\ \ t \in [0,1] \\ 0 & otherwise \end{cases} \tag{2.52}$$

$$\psi(t) = \begin{cases} 1 & t \in [0,\frac{1}{2}[\\ -1 & t \in [\frac{1}{2},1[\\ 0 & otherwise \end{cases} \tag{2.53}$$

ii. **Daubechies (db2) wavelet**

The filter coefficients of the Daubechies wavelet (db2) ($n_f = 4\,,\ m = 0,\ 1$), result from the following set of equations (Li *et al.*, 2002):

$$\begin{cases} l_0 & +l_1 & +l_2 & +l_3 & = \sqrt{2} & & stability \\ l_0 & -l_1 & +l_2 & -l_3 & = 0 & (m=0) & convergence \\ & -l_1 & +2l_2 & -3l_3 & = 0 & (m=1) & convergence \\ l_0 l_2 & +l_1 l_3 & & & = 0 & & wavelet\ \ orthogonality \\ l_0^2 & +l_1^2 & +l_2^2 & +l_3^2 & = 1 & & scaling\ \ orthogonality \end{cases} \tag{2.54}$$

The solution for the scaling coefficients is

$$\begin{aligned} l_0 &= \frac{1+\sqrt{3}}{4\sqrt{2}}; \qquad & l_1 &= \frac{3+\sqrt{3}}{4\sqrt{2}} \\ l_2 &= \frac{3-\sqrt{3}}{4\sqrt{2}}; \qquad & l_3 &= \frac{1-\sqrt{3}}{4\sqrt{2}} \end{aligned} \tag{2.55}$$

From (2.43), the resulting wavelet coefficients are:

$$\begin{aligned} h_0 &= l_3; \qquad & h_1 &= -l_2 \\ h_2 &= l_1; \qquad & h_3 &= -l_0 \end{aligned} \tag{2.56}$$

The corresponding scaling and wavelet functions are illustrated in Figure 2.6*b*).

2.4.4 Contractive Property

One of the first works in the domain of time-series similarity was presented by Agrawal *et al.* (1993). They proposed the use of DFT for feature extraction and introduced an indexing mechanism for similarity search, designated by F-Index. This mechanism was used as the framework for a majority of the subsequent works, being meanwhile generalized for subsequence similarity index search by Faloutsos *et al.* (1994). Basically, by introducing the F-Index, Agrawal *et al.*, (1993) proved that to guarantee no false dismissals, the distance measure of any two objects (time series), $X(t)$ and $Y(t)$, and their corresponding transforms, $\Im(X(t))$ and $\Im(Y(t))$, must satisfy the following lower bounding lemma:

$$D_{feature}\Big(\Im(X(t)),\Im(Y(t))\Big) \leq D_{time}\Big(X(t),Y(t)\Big) \tag{2.57}$$

This property, referred to as the contractive property of the transform $\Im(\cdot)$, ensures no false dismissals. The key idea is that the computed distance based on the transformed features may have false positives but no false negatives (or false dismissals).

One of the main requirements when applying the general wavelet transform to similarity search problems, is the ability to preserve the contractive property and, thus, to guarantee no false dismissals. Fukanaga (1990) presented a proof that the Euclidean distance is preserved for the class of orthonormal transforms. The Haar wavelet, as well as other wavelets, belongs to the class of orthonormal transforms. Therefore, based on this principle, Chan and Fu (1999) showed that the contractive property can be established for the Haar wavelet. Later, Popivanov and Miller (2002) showed that a wider class of wavelet transforms can be used to support similarity search, extending the contractive property for a class of bi-orthonormal wavelets.

To summarize, the application of the Haar wavelet transform ensures the preservation of the Euclidean distance between any two time-series in the transformed space, which is an extremely important property to support dimension reduction of time series data. In effect, it guarantees that no qualified time sequence will be rejected or, in other words, that no false dismissal occur when searching for similarities in time series.

2.5 Proposed Similarity Measure and Indexing Scheme

2.5.1 Introduction

The matching of similar patterns in long sequences is an important topic in time series analysis field. Some transform-based methods, such as the Fourier transform, can be used for dimension reduction, but can not provide any features specific for a given local. One of the main hypotheses to be explored in this work is motivated by the wavelet transform property of time-frequency localization. In fact, contrasting with other methods, wavelet transform enables to distinguish characteristics of the sequence both at various locations and at varying time granularities. Since the basis functions are local in both time and frequency, description of a time series is possible by the analysis of the weights associated with the relevant basis.

This work introduces a similarity measure based on the natural set of features generated by the wavelet transform (coefficients of the basis functions), which reflect the fundamental dynamics of a time series. Through a dimension reduction performed by the Karhunen-Loève transform (KLT), an optimal number of wavelet basis is obtained. The coefficients corresponding to the reduced set of basis are the origin of the comparison scheme. Basically, by the referred reduction procedure, an effective description of the time series in the most representative regions (local wavelets basis) is achieved, enabling a practical comparison between time sequences. In addition to the locality, the proposed wavelet scheme makes possible to obtain a rough estimation of the signal (behaviour) at a predefined degree.

In this particular framework, the specific Haar wavelet can be efficiently used in the proposed similarity search scheme. Several studies have addressed the Haar wavelet transform as a dimensionality reduction technique within the context of similarity detection, showing that it captures the shape of time series better than Fourier transform, (Chan and Fu, 1999), (Wu *et al.*, 2000). On the other hand, a major drawback that has been pointed to Haar wavelet relates with the basis functions that are not smooth, that is, are not continuously differentiable. As result, this wavelet approximates any signal by a ladder-like structure that may be not adequate for smooth functions. However, in the particular case of the present thesis, the final goal is to identify the main characteristics of the signal, that is, the main trends or behaviour (possibly in some specific time regions). Thus, this inconvenience is not significant.

Moreover, although taking as the starting point the Euclidean distance as the similarity measure, the proposed scheme is able to circumvent some of its major weaknesses: *i*) high computational complexity, *ii*) not applicable to time series that are not aligned in time (dissimilar dynamics at the same points in time); *iii*) not adequate if signals have different baselines (mean values); *iv*) limited to cases where there exists a linear relationship. In effect, the proposed methodology provides a significant complexity decrease by implementing a reduction procedure that results in an optimal reduced set

of orthogonal basis. Additionally, the rough estimation of the signal achieved by the Haar wavelet transform, allows to define a similarity measure based on the signal trend, thus enabling, to some extent, the comparison of signals not aligned in time. Another benefit is the inherent ability of wavelets to deal with the non-stationarity and non-linearity present in the signals. Furthermore, another positive consequence of the proposed wavelet decomposition is its capacity to cope with the presence of noise in the signal. In effect, since only an approximation of the signal is used, a noise reduction is naturally achieved, which significantly contributes to improve the similarity searching process.

Main steps

The Figure 2.7. depicts the main steps involved in the proposed similarity measure and indexing scheme.

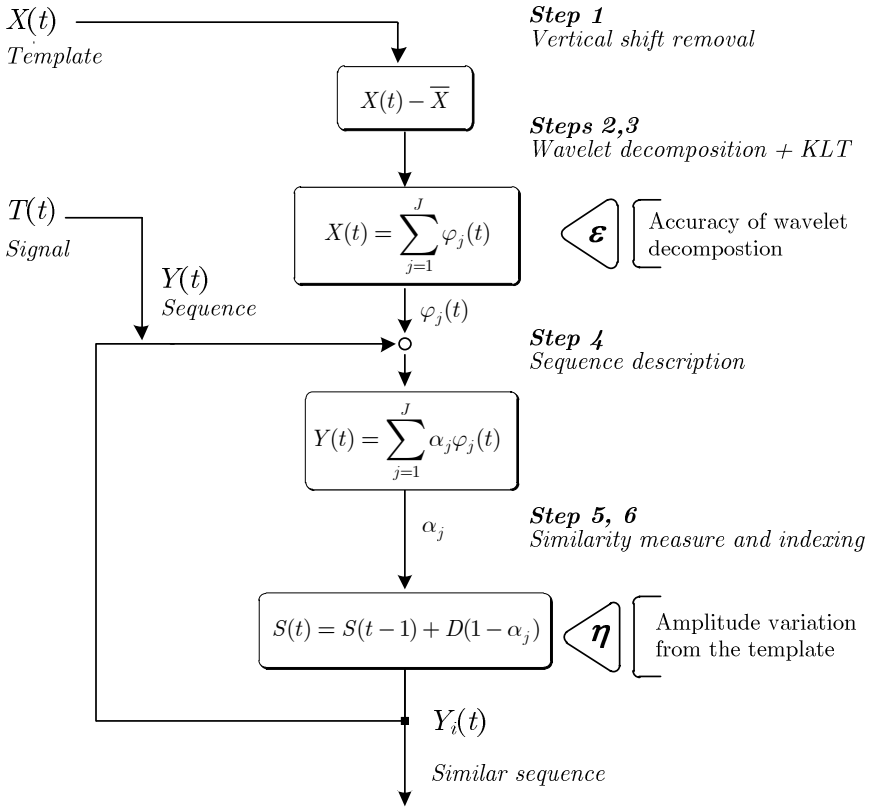


Figure 2.7 - Proposed similarity measure and indexing scheme.

Basically, the following steps are accomplished:

- **Step 1. Vertical shift removal**: To guarantee that similarity assessments are independent of variations in the vertical position, a vertical shift removal procedure is employed.
- **Step 2. Wavelet decomposition of the template:** In this step, the decomposition of the template (pattern) to be compared with the time series is achieved by means of a set of orthogonal (or orthonormal) wavelet basis.
- **Step 3. Optimal dimension reduction (KLT):** Based on the localisation property of the wavelet basis, those that significantly reflect the dynamical patterns of the template are chosen to compose the reduced set of basis. In the case of basis are orthonormal, a simple selection of highest coefficients is required; in the case of basis are orthogonal, a Karhunen-Loève transform is applied.
- **Step 4. Sequence description**: The subsequence of the signal to be compared with the template, is described by means of the previous reduced set of basis. It is important to refer that this description does not involve a wavelet decomposition, but a simple computation of coefficients.
- **Step 5. Similarity measure**: The coefficients obtained for the template and subsequence description in terms of the reduced set of basis, are employed to derive a similarity measure. They allow the interpretation as an evolution trend, as well as a relative comparison between the amplitudes of the template and subsequence.
- **Step 6. Subsequence indexing**: Based on the previous similarity measure and using the particular Haar wavelet, an efficient iterative similarity indexing algorithm is proposed, enabling to reduce the computational complexity.

Finally, the proposed one-dimensional time series similarity measurement is extended to multi-dimensional time series.

Pre-defined parameters

In the previous scheme there are two key parameters that have to be previously established. These parameters, $\varepsilon \in \mathbb{R}^+$ and $\eta \in \mathbb{R}^+$ correspond to:

- ε : **accuracy of wavelet decomposition**. This parameter controls the approximation error by determining the number of basis to be considered in the template decomposition. Moreover, the selected basis define the most significant spatial localizations of the signal, allowing the interpretation in terms of its evolution trend.
- η : **difference in amplitude between a given signal and the template**. While the first parameter (ε) characterizes the time evolution of the signal, this second parameter defines when two signals that present the same behaviour are or are not considered as similar. To this aim, it establishes a limit to the difference (ratio) of signals' amplitudes.

2.5.2 Step 1: Vertical Shift Removal

The starting point for the establishment of the similarity measure proposed in this work is the Euclidean distance, equation (2.1). From the definition, two time series with the same length (N), where N is a power of 2 ($N=2^n, n\in\mathbb{N}$), are similar if their Euclidean distance, $D_E(\cdot)$, is lower than a given threshold $\zeta\in\mathbb{R}^+$.

$$D_E\Big(X_1(t),X_2(t)\Big)=\sqrt{\sum_{t=1}^{N}\big(x_1(t)-x_2(t)\big)^2}<\zeta \tag{2.58}$$

The proposed scheme, assumes that similarity should be insensitive to changes in the vertical position of the time series. As stated by Chan *et al.* (2003), signals that are shifted (that is, that have a different mean) can not be considered similar using the Euclidean definition, (2.58). Therefore, according to the same authors, a simple vertical shift similarity, $D_V(\cdot)$, is proposed as follows.

$$D_V\Big(X_1(t),X_2(t)\Big)=\sqrt{\sum_{t=1}^{N}\Big(\big(x_1(t)-\overline{X_1}\big)-\big(x_2(t)-\overline{X_2}\big)\Big)^2}<\zeta \tag{2.59}$$

Consequently, the template signal $X(t)\in\mathbb{R}^{1,N}$ is in a first step modified as (2.60), where $\overline{X}$ is the mean value of the template $X(t)$, obtained according to equation (2.4).

$$X(t)=X(t)-\overline{X} \tag{2.60}$$

Thus, neglecting their vertical offsets, any two time sequences are said to be vertical shift similar if the Euclidean distance is less than or equal to a threshold, equation (2.59). This definition can give a better estimation of the similarity between two time sequences. In effect, if analogous trends exist at two completely different levels, signals can be regarded as similar.

2.5.3 Step 2: Wavelet Decomposition of the Template

In a second step, the template signal $X(t)$ is decomposed using a DWT process. Basically, the DWT decomposes a time series in terms of an approximation of the original sequence, plus a set of details. The main trend of the input sequence is preserved in the approximation part, while the localized changes are kept in the detail parts. Assuming that the length of the signal is N, and considering the L level of decomposition, such that $L=\log_2(N)$, the original signal can be reconstructed as described by (2.61).

$$X(t)=c_{0,0}\phi_{0,0}(t)+\sum_{j}\sum_{k}d_{j,k}\psi_{j,k}(t) \tag{2.61}$$

Therefore, the signal $X(t)$ can be represented as a linear combination of N basis functions, $\phi_{0,0}(t)$ and $\psi_{j,k}(t)$, respectively, approximation and detail functions.

Number of detail basis

Considering L levels of decomposition, Figure 2.5, equation (2.35), the number of detail basis $\psi_{j,k}(t)$ and corresponding coefficients is $(N-1)$, as shown in Figure 2.8.

	$L = \log_2 N$	$X(t)$	$length = 2^L = N$
Level L	$J = L-1$	$\sum_{k=0}^{2^J-1} d_{J,k}\psi_{J,k}(t) \rightarrow \left[\psi_{L-1,0},\ \psi_{L-1,1},\ \ldots,\ \psi_{L-1,2^{L-1}-1}\right]$	$basis = 2^{L-1} = \frac{N}{2}$
$\vdots$		$\vdots$	$\vdots$
Level 2	$J = 1$	$\sum_{k=0}^{2^J-1} d_{J,k}\psi_{J,k}(t) \rightarrow \left[\psi_{1,0},\ \psi_{1,1}\right]$	$basis = 2^1 = 2$
Level 1	$J = 0$	$\sum_{k=0}^{2^J-1} d_{J,k}\psi_{J,k}(t) \rightarrow \left[\psi_{0,0}\right]$	$basis = 2^0 = 1$

Figure 2.8 - Number of detail basis in a wavelet decomposition.

Thus, considering all L levels of decomposition, the number of basis involved is $\left(\frac{N}{2}+\frac{N}{4}+\ldots+2+1\right)$. This sequence represents a geometric progression of ratio 2, with a number of terms $L = \log_2 N$, and with the first term equal to 1. Therefore, the total number of terms can be computed by (Roger, 1993):

$$1\frac{2^{\log_2 N}-1}{2-1} = N-1$$

Number of approximation basis

Since vertical shift is considered, the mean value of $X(t)$ is zero, $\overline{X} = 0$, according to the definition (2.60). Therefore, since the first coefficient of equation (2.61) reflects the mean value of the transformed signal $X(t)$, it follows that $c_{0,0} = 0$.

As result, a signal can be written as a linear combination of the detail basis according to (2.62), which uses a simplified notation.

$$X(t) = \sum_{j=1}^{N-1} d_j\ \psi_j(t) \tag{2.62}$$

In the previous equation, $d_j \in \mathbb{R}$ are scalars (weights) and $\psi_j(t) \in \mathbb{R}^{1,N}$ are orthonormal basis (wavelets). In fact, the number of basis is $(N-1)$ and not N, since $c_{0,0} = 0$. However, in order to maintain the generality, it is considered that the coefficient $c_{0,0}$ can be nonzero.

Moreover, to achieve a more generic approach, the last description is extended to orthogonal basis as well. In order to simplify the computation of the similarity measurement to be presented in section 2.55, it is assumed the signal generically described as (2.63).

$$X(t) = \sum_{j=1}^{N} \varphi_j(t) \tag{2.63}$$

In the previous equation it is considered that the basis functions $\varphi_j(t) \in \mathbb{R}^{1,N}$ are orthogonal (not necessarily orthonormal) and that incorporate the respective coefficients. As a consequence, the template can be fundamentally expressed as a '*sum of the basis*'.

Using matrix notation the equation (2.64) is obtained, where $\Gamma \in \mathbb{R}^{1,N}$ is a weight vector composed of coefficients, in this case a vector of ones, $\Gamma = [1,\ 1,\ \ldots,\ 1]$, and the matrix $\Phi(t) \in \mathbb{R}^{N,N}$ is composed of the orthogonal wavelet basis, described by (2.65).

$$X(t) = \Gamma\ \Phi(t) \tag{2.64}$$

$$\Phi(t) = \begin{bmatrix} \psi_{L-1,0}(t) \\ \vdots \\ \psi_{L-1,2^{L-1}-1}(t) \\ \\ \vdots \\ \\ \psi_{1,0}(t) \\ \psi_{1,1}(t) \\ \\ \psi_{0,0}(t) \\ \phi_{0,0}(t) \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \frac{N}{2} & \textit{basis (detail)} \\ \\ \vdots & \\ \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} 2 & \textit{basis (detail)} \\ \\ \left.\begin{matrix} \end{matrix}\right\} 1 & \textit{basis (detail)} \\ \left.\begin{matrix} \end{matrix}\right\} 1 & \textit{basis (approximation)} \end{matrix} \tag{2.65}$$

A simplified representation can be used, denoting the basis as $\varphi_j(t),\ \ j = 1,\ldots,N,\ \ t = 1,\ldots,N$, where j identifies each wavelet and t denotes time.

$$\Phi(t) = \begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \\ \vdots \\ \varphi_j(t) \\ \vdots \\ \varphi_N(t) \end{bmatrix} = \begin{bmatrix} \varphi_1(1) & \varphi_1(2) & \ldots & \varphi_1(t) & \ldots & \varphi_1(N) \\ \varphi_2(1) & \varphi_2(2) & \ldots & \varphi_2(t) & \ldots & \varphi_2(N) \\ \vdots & \vdots & & \vdots & & \vdots \\ \varphi_j(1) & \ldots & \ldots & \varphi_j(t) & \ldots & \varphi_j(N) \\ \vdots & \vdots & & \vdots & & \vdots \\ \varphi_N(1) & \varphi_N(2) & \ldots & \varphi_N(t) & \ldots & \varphi_N(N) \end{bmatrix} \tag{2.66}$$

2.5.4 Step 3: Optimal Basis Reduction

In order to obtain a reduced set of wavelet basis, a Karhunen-Loève transform is implemented, which is able to approximate a signal with a pre-defined accuracy. The combination of wavelet decomposition with KLT analysis enables to attain an efficient representation that reduces the number of wavelet basis and preserves the main (local) characteristics of the signal. This aspect is of fundamental importance when establishing similarity measures.

i. In case the basis are orthonormal

If the basis are orthonormal, the optimal approximation to the original signal in terms of the l^2 norm is straightforwardly achieved maintaining the largest J wavelet coefficients (Karim and Adeli, 2002). Let $X(t)$ be a signal described by a set of N orthonormal basis, $\psi_j(t)$, with coefficients $\{d_1,d_2,...,d_N\}$, according to (2.67).

$$X(t) = \sum_{j=1}^{N} d_j \ \psi_j(t) \tag{2.67}$$

Let $\sigma(\cdot)$ be a permutation of the indices $\{1,2,...,N\}$ and $\widehat{X}(t)$ the signal that results from the first J coefficients ($J < N$) of the permutation $\sigma(\cdot)$, represented by (2.68).

$$\widehat{X}(t) = \sum_{j=1}^{J} d_{\sigma(j)} \ \psi_{\sigma(j)}(t) \tag{2.68}$$

Proposition 2.1

Sorting the coefficients in descending order of magnitude and selecting the first J, provides the best l^2 norm approximation error, which is obtained by the following equation (2.69), where the operator $\| \ . \ \|_2$ represents de referred norm.

$$\left\| X(t) - \widehat{X}(t) \right\|_2^2 = \ < X(t) - \widehat{X}(t), X(t) - \widehat{X}(t) > \tag{2.69}$$

Proof (Liao, 2005)

Considering the definitions of $X(t)$ and $\widehat{X}(t)$ by equations (2.67) and (2.68), it follows that:

$$\left\| X(t) - \widehat{X}(t) \right\|_2^2 = \left\langle \sum_{j=J+1}^{N} d_{\sigma(j)} \ \psi_{\sigma(j)}(t), \sum_{k=J+1}^{N} d_{\sigma(k)} \ \psi_{\sigma(k)}(t) \right\rangle \ ^{1} = \tag{2.70}$$

$$= \sum_{j=J+1}^{N} \sum_{k=J+1}^{N} d_{\sigma(j)} d_{\sigma(k)} \left\langle \psi_{\sigma(j)}(t), \psi_{\sigma(k)}(t) \right\rangle = \sum_{j=J+1}^{N} \left(d_{\sigma(j)} \right)^2 \tag{2.71}$$

---

[1] The operator $< a,b >$ denotes the dot product between the vectors a and b, that is, $< a,b >= \sum_i a(i) \ b(i)$

In equation (2.70), the terms corresponding to $j=1,...,J$ are not considered since they are common to $X(t)$ and $\widehat{X}(t)$ and, thus, eliminated by the subtraction of the signals. On the other hand, in equation (2.71) the dot product between the basis is 0 (zero) when $j \neq k$ (cross terms), or equal to 1, in case $j=k$, since they are orthonormal.

Therefore, in order to minimize the approximation error, the best choice for the permutation is the one that considers the highest coefficients in the decreasing order of magnitude. This means that permutation should verify the following relationship, where $\left\|d_{\sigma(j)}\right\|_2^2 = (d_{\sigma(j)})^2$ represents the square of the l^2 norm of the coefficient $d_{\sigma(j)}$.

$$\left\|d_{\sigma(1)}\right\|_2^2 \geq \left\|d_{\sigma(2)}\right\|_2^2 \geq \ldots \geq \left\|d_{\sigma(J)}\right\|_2^2 \ldots \geq \left\|d_{\sigma(N)}\right\|_2^2 \tag{2.72}$$

Following this idea, choosing the J orthonormal basis is straightforward. In fact, given a predefined level of precision ε (accuracy of approximation), an optimal basis reduction can be achieved selecting the highest coefficients so that the level of precision error ε is attained.

ii. In case the basis are orthogonal

Theoretically, when decomposing a function in terms of a set of basis, the best solution in terms of the mean squared error (MSE) is achieved with a Karhunen-Loève base of orthogonal eigenfunctions (Clifford *et al.*, 2006). Basically, given a collection of N-dimensional vectors, they are projected onto a J-dimensional subspace, where $J<N$, maximizing the variances in the chosen dimensions.

Considering a set of N-dimensional basis, an efficient eigenbasis to represent a signal $X(t)$ requires the fewest eigenvectors needed to approximate $X(t)$ to a desired accuracy (level of error). These basis are designated as KLT basis functions, obtained as the eigenvectors (also known as principal components) of a covariance matrix.

In this particular case, the covariance matrix $R \in \mathbb{R}^{N,N}$ is composed of the wavelet basis $\varphi_j(t)$, matrix (2.66), in the following form:

$$R = \frac{1}{N-1} \Phi(t)\ \Phi(t)^T \tag{2.73}$$

From this matrix, N eigenvectors $\upsilon_k \in \mathbb{R}^{1,N}$ and the respective eigenvalues λ_k $(k=1,...,N)$ are obtained by solving (2.74).

$$R\ \upsilon_k^T = \lambda_k\ \upsilon_k \tag{2.74}$$

The eigenvectors $\{\upsilon_1, \upsilon_2, ..., \upsilon_N\}$ form an orthogonal basis. The vectors $\varphi_j(t)$, to be employed as basis functions, can be established by considering the descending order of magnitude of the corresponding eigenvalues λ_k, that is,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_J \geq \ldots \geq \lambda_N \tag{2.75}$$

As result, an hybrid method combining wavelet with Karhunen-Loève transforms is considered. The objective is to obtain a description of a signal $X(t)$ using the most representative wavelet basis. By combining these two approaches, it is possible to reduce the dimension of the underlying space and hence to achieve an efficient scheme to describe data. In fact, the global aim is to merge the key features of the KLT analysis with the significant features of the wavelet decomposition. The idea is that the eigen-analysis provides the most important features of the signal while the wavelet decomposition gives the particular local features of the same.

iii. Dimension (level of decomposition): parameter ε

Generally, choosing the appropriate dimension of the features in the transformed space is a challenging problem. In the particular case of the wavelet transform, the dimension of the features is typically determined by the level of decomposition. Several studies have been conducted to determine the optimal level of decomposition of a wavelet transform, mainly based on energy and/or entropy concepts. For example, Coifmain and Wickerhauser (1992) proposed algorithms based on Shannon entropy for best basis selection, which permits efficient compression of signals; Sang *et al.*, (2010) proposed a similar method to determine the decomposition level based on the wavelet energy entropy.

However, following the exposed approach this problem can be easily circumvented. In fact, by means of the proposed scheme, the level of decomposition in the wavelet transform procedure is converted into the selection of the wavelets basis that allow a given level of accuracy in the representation of the signal. Moreover, using this approach, an optimal feature extraction technique that efficiently reduces the data into a lower-dimensional space is obtained, which preserves the main properties (localization) of the original time series.

To summarize, the procedure for selecting the J basis is the following: given a predefined level of precision $\varepsilon \in \mathbb{R}^+$ (accuracy of approximation), an optimal basis reduction can be achieved choosing the highest coefficients, in case wavelets are orthonormal, or the highest eigenvalues, in case wavelets are orthogonal.

1. In case of orthonormal basis, where $< \psi_i, \psi_j > \begin{cases} 0 \ if \ i \neq j \\ 1 \ if \ i = j \end{cases}$, the next steps are applied:

- Compute the wavelet coefficients of the original signal, $\{d_1, d_2, ..., d_N\}$.
- Sort the coefficients in decreasing order of magnitude to obtain a permutation such as: $\left\|d_{\sigma(1)}\right\|_2^2 \geq \left\|d_{\sigma(2)}\right\|_2^2 \geq ... \geq \left\|d_{\sigma(J)}\right\|_2^2 ... \geq \left\|d_{\sigma(N)}\right\|_2^2$.
- Consider the first coefficients of the permutation (the highest) such that

$$\frac{\sum_{j=1}^{J}\left\|d_{\sigma(j)}\right\|_2^2}{\sum_{j=1}^{N}\left\|d_{\sigma(j)}\right\|_2^2} \geq \varepsilon \tag{2.76}$$

- Select the basis corresponding to the first coefficients of the permutation as the fundamental basis.

2. In case of orthogonal basis, where $<\varphi_i,\varphi_j> \begin{cases} 0 \ \ if \ \ i \neq j \\ \varphi_i\varphi_j^T \ \ if \ \ i = j \end{cases}$, the next steps are applied:

- Determine the covariance matrix R and the respective eigenvalues.
- Sort the eigenvalues in decreasing order of magnitude: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_J \geq \ldots \geq \lambda_N$.
- Consider the first eigenvalues of the decomposition (the highest) such that

$$\frac{\sum_{j=1}^{J}\lambda_j}{\sum_{j=1}^{N}\lambda_j} \geq \varepsilon \tag{2.77}$$

- Select the basis corresponding to the first eigenvalues of the decomposition as the fundamental basis.

iv. Conclusion

Wavelet and Karhunen-Loève transforms provide different and complementary ways to represent information. The basis vectors for KLT analysis come from the eigenvectors of a covariance matrix, thus representing global information. On the other hand, wavelets are local in nature and, consequently, make possible the revelation of significant local characteristics of the signal.

Thus, in conclusion, by merging the KLT with the wavelet decomposition, the original signal $X(t)$ can be described as a reduced set of J basis, denoted by $\widehat{X}(t)$.

$$\widehat{X}(t) = \sum_{j=1}^{J}\varphi_j(t) \tag{2.78}$$

Moreover, using the KLT procedure it is possible to decrease the l^2 norm approximation error as the number of basis increases, that is, the following relationship is verified.

$$\begin{aligned} \left\|X(t)-\varphi_1(t)\right\|_2^2 \geq \left\|X(t)-\left(\varphi_1(t)+\varphi_2(t)\right)\right\|_2^2 \geq \ldots \\ \ldots \geq \left\|X(t)-\sum_{j=1}^{J}\varphi_j(t)\right\|_2^2 \geq \ldots \geq \left\|X(t)-\sum_{j=1}^{N}\varphi_j(t)\right\|_2^2 = 0 \end{aligned} \tag{2.79}$$

Preliminary discussion

Figure 2.9 illustrates this approximation process for a signal $X(t)$ with length $N = 64$. Using the Haar wavelet decomposition and establishing a predefined level of precision (threshold ε), an optimal basis reduction is achieved. In this particular case, setting $\varepsilon = 0.92$ (roughly meaning that the signal should be approximated with a level of confidence of 92%), three basis were obtained considering the previously presented procedure. As referred, these basis are not determined based on a specific level of decomposition, but are chosen as those that best represent the signal (in terms of the l^2 norm approximation error) among all decomposition levels ($L = \log_2 N = 6$).

As can be seen in this figure, using only three basis the main characteristics of the signal are captured. Although it is possible to obtain small approximation errors (by increasing the threshold ε and, therefore, the number of basis), it should be noted that this is not the main focus of the proposed scheme. Effectively, the final goal is to identify the main characteristics of the signal, that is, the main trends or behaviour, as it will be discussed latter.

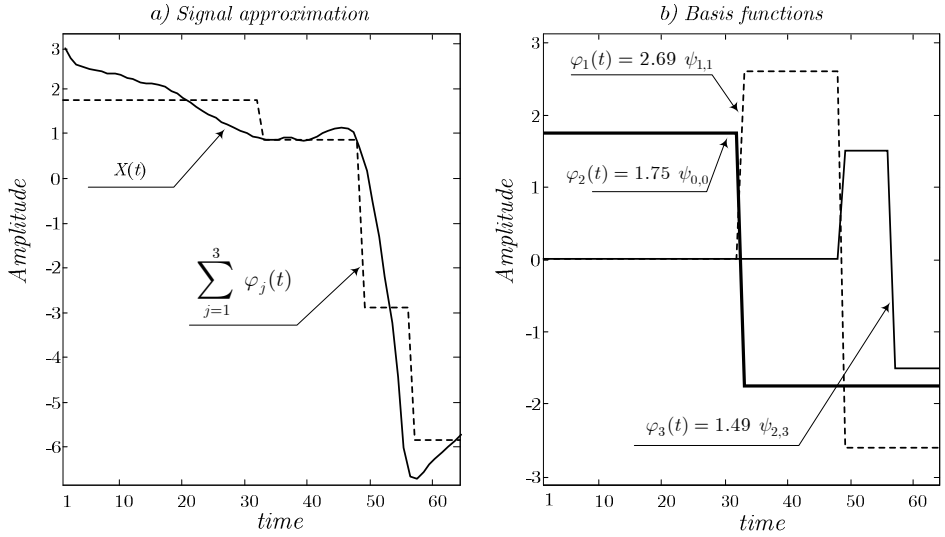


Figure 2.9 - Signal approximation using the Haar wavelet decomposition.

In this particular case, the most important basis is $\varphi_1(t) = 2.69\ \psi_{1,1}$ (largest coefficient). This means that the highest variation in the signal occurs at the instants $[33,64]$, as can be confirmed by Figure 2.9*a*). Moreover, the second basis, $\varphi_2(t) = 1.75\ \psi_{0,0}$, is the one that reflects the contribution of all signal. Since the coefficient has a positive value, it can be concluded that the signal presents a global positive variation, that is, the mean of the first half, instants $[1,32]$, is higher than the mean of the second half, instants $[33,64]$. Finally, similar conclusion can be taken for the third basis, $\varphi_3(t) = 1.49\ \psi_{2,3}$. In fact, in the corresponding time region, instants $[49,64]$, the signal presents a significant variation from a higher to a lower value and, thus, the coefficient has a positive sign.

2.5.5 Step 4: Similarity Measure

The proposed similarity measure is based on the Euclidean distance, equation (2.58). However, it is computed from the coefficients of the reduced set of wavelet basis, which reflect the main dynamic patterns of the time series, instead of directly from the original signals. Using this strategy, a simple but effective assessment of similarity between a particular template and a time series, is achieved.

i. Signal description

Given a signal, $Y(t) \in \mathbb{R}^{1,N}$, to be compared with the template, $X(t) \in \mathbb{R}^{1,N}$, the first step consists in describing it as a linear combination of the orthogonal basis functions, $\varphi_j(t)$, that were used to describe the template.

Proposition 2.2

A time series signal $Y(t) \in \mathbb{R}^{1,N}$ can be described by means of the orthogonal basis set, $\varphi_j(t) \in \mathbb{R}^{1,N}$, according to equation (2.80).

$$Y(t) = \sum_{j=1}^{N} \alpha_j \ \varphi_j(t) \tag{2.80}$$

The coefficients $\alpha_j \in \mathbb{R}$ always exist and are given by (2.81).

$$\alpha_j = \frac{< Y(t), \varphi_j(t) >}{< \varphi_j(t), \varphi_j(t) >} \tag{2.81}$$

In case the basis are orthonormal, the last equation can be straightforwardly simplified to (2.82), since $< \varphi_j(t), \varphi_j(t) >= 1$.

$$\alpha_j = < Y(t), \varphi_j(t) > \tag{2.82}$$

Proof

Let us define the basis (wavelet functions) using matrix $\Phi(t)$, equation (2.66). Representing the coefficients to be determined as $\Omega = [\alpha_1 \ \ \alpha_2 \ \ \ldots \ \ \alpha_N]$, the signal (vector) $Y(t)$ can be written as (2.83), similarly to (2.64). In order to simplify the comprehension, the parameter (t) is omitted from now.

$$Y = \Omega \ \Phi \tag{2.83}$$

Considering a trivial operation, it can be shown that the coefficients result in (2.84), where the matrix $\Phi^{\dagger}$ denotes the pseudo-inverse of matrix Φ.

$$\Omega = Y \ \Phi^T \left(\Phi \Phi^T \right)^{-1} = Y \Phi^{\dagger} \tag{2.84}$$

Since the basis are orthogonal, the relation (2.85) is verified.

$$\sum_{i,j} \varphi_i \varphi_j = \begin{cases} < \varphi_i, \varphi_i > & i = j \\ \\ 0 & otherwise \end{cases} \tag{2.85}$$

Thus, $\Phi\Phi^T \in \mathbb{R}^{N,N}$ is a diagonal matrix with elements $< \varphi_i, \varphi_i >$.

$$\Phi\Phi^T = \begin{bmatrix} < \varphi_1, \varphi_1 > & & & \\ & < \varphi_2, \varphi_2 > & & \\ & & \ddots & \\ & & & < \varphi_N, \varphi_N > \end{bmatrix} \tag{2.86}$$

The respective inverse, $(\Phi\Phi^T)^{-1}$, is straightforwardly obtained. It is also a diagonal matrix with elements $\dfrac{1}{< \varphi_j, \varphi_j >}$. Moreover, since by definition $< \varphi_j, \varphi_j > \neq 0$, this inverse always exists.

$$\left(\Phi\Phi^T\right)^{-1} = \begin{bmatrix} \dfrac{1}{< \varphi_1, \varphi_1 >} & & & \\ & \dfrac{1}{< \varphi_2, \varphi_2 >} & & \\ & & \ddots & \\ & & & \dfrac{1}{< \varphi_N, \varphi_N >} \end{bmatrix} \tag{2.87}$$

The product of $(Y\ \Phi^T)$ provides a matrix of dimension $(1, N)$ defined as:

$$Y\Phi^T = \left[\ < Y, \varphi_1 >, \ \ < Y, \varphi_2 >, \ \ \ldots \ \ < Y, \varphi_N > \ \right] \tag{2.88}$$

Finally, from (2.87) and (2.88), the coefficients can be obtained by (2.89).

$$\Omega = Y\ \Phi^T \left(\Phi\Phi^T\right)^{-1} = \left[\ \frac{< Y, \varphi_1 >}{< \varphi_1, \varphi_1 >}, \ \ \frac{< Y, \varphi_2 >}{< \varphi_2, \varphi_2 >}, \ \ \ldots \ \ \frac{< Y, \varphi_N >}{< \varphi_N, \varphi_N >} \right] \tag{2.89}$$

Note that if the basis are orthonormal, $< \varphi_i, \varphi_i > = 1$, and the previous equation results in (2.90).

$$\Omega = Y\ \Phi^T \left(\Phi\Phi^T\right)^{-1} = \left[\ < Y, \varphi_1 >, \ \ < Y, \varphi_2 >, \ \ \ldots \ \ , < Y, \varphi_N > \right] \tag{2.90}$$

Proposition 2.3

It is possible to relate the squared Euclidean distance between the signals $X(t)$ and $Y(t)$ using the obtained coefficients, Ω , as follows:

$$D_E{}^2(X,Y) = \Big([\Gamma - \Omega]\ \Phi \Big)\Big([\Gamma - \Omega]\ \Phi\ \Big)^T \tag{2.91}$$

Proof

The proof is immediate.

$$D_E{}^2(X,Y) = \sum_{j=1}^{N}\big(\varphi_j - \alpha_j\varphi_j\big)^2 = \sum_{j=1}^{N}\Big(\ (1-\alpha_j)\varphi_j\Big)^2 = \sum_{j=1}^{N}\Big((1-\alpha_j)\varphi_j\Big)\Big((1-\alpha_j)\varphi_j\Big)^T \tag{2.92}$$

Or, in the matrix form:

$$D_E{}^2(X,Y) = \Big(\ [\Gamma - \Omega]\ \Phi\ \Big)^2 = \Big(\ [\Gamma - \Omega]\ \Phi\Big)\Big(\ [\Gamma - \Omega]\ \Phi\Big)^T$$

ii. Reduced representation using KLT

As explained, wavelet and KL transforms enable to obtain a reduced representation of the template $X(t)$. One important property to be verified when dealing with dimension reduction is the contractive property. In fact, it is of fundamental importance to ensure that no false dismissals occur.

Proposition 2.4

Consider two signals, $X(t) \in \mathbb{R}^{1,N}$ and $Y(t) \in \mathbb{R}^{1,N}$, and the Euclidean distance between them, equation (2.1). If they are described according to (2.63) and (2.80), the dimension reduction, considering the combined wavelet and KL transforms procedure (denoted here by $\Im_{KW}(\cdot)$), ensures the contractive property, that is,

$$D_{feature}\Big(\ \Im_{KW}(X), \Im_{KW}(Y)\ \Big) \le D_{time}(X,Y) \tag{2.93}$$

where

$$\Im_{KW}(X) = \sum_{j=1}^{J}\ \varphi_j(t) \tag{2.94}$$

and

$$\Im_{KW}(Y) = \sum_{j=1}^{J}\ \alpha_j\ \ \varphi_j(t) \tag{2.95}$$

Proof

By definition, from the wavelet representation it follows that:

$$X(t)-Y(t)=\sum_{j=1}^{N}\varphi_j(t)-\sum_{j=1}^{N}\alpha_j\varphi_j(t) \tag{2.96}$$

Or, equivalently:

$$X(t)-Y(t)=\sum_{j=1}^{N}\left(1-\alpha_j\right)\ \varphi_j(t) \tag{2.97}$$

Thus, the following relation is obtained:

$$\left(\ X(t)-Y(t)\ \right)^2=\left[\sum_{j=1}^{N}\ \left(1-\alpha_j\right)\varphi_j(t)\right]^2=\sum_{j=1}^{N}\ \left[\left(1-\alpha_j\right)\varphi_j(t)\right]^2\quad\Leftrightarrow \tag{2.98}$$

$$\Leftrightarrow\quad\left(\ X(t)-Y(t)\right)^2=\left[\ (1-\alpha_1)\varphi_1(t)\right]^2+\ldots+\left[\ (1-\alpha_J)\varphi_J(t)\right]^2+\ldots+\left[\ (1-\alpha_N)\varphi_N(t)\right]^2$$

In the previous equation, the square of the sum is equal to the sum of the squares, since the terms corresponding to the product of $\varphi_j(t)\varphi_k(t)^T$ is 0 (zero for $j\neq k$), as a consequence of the orthogonality of the basis.

Since all terms are positive, if only the first J terms are considered, the result is:

$$\left(\ X(t)-Y(t)\right)^2\geq\left[\ (1-\alpha_1)\varphi_1(t)\right]^2+\ldots+\left[\ (1-\alpha_J)\varphi_J(t)\right]^2 \tag{2.99}$$

Thus, it follows that:

$$D_{feature}\left(\ \Im_{KW}(X(t)),\Im_{KW}(Y(t))\right)\leq D_{time}\left(\ X(t),Y(t)\ \right)$$

iii. Proposed similarity measure: parameter η

As shown, it is possible to consider only the first terms of the wavelet representation, ensuring that no false dismissals occur. The first J terms of the error, denoted as $D_{KW}(X(t),Y(t))$, can be written as (2.100).

$$D_{KW}(X,Y)=(1-\alpha_1)^2{\varphi_1}^2+\ldots+(1-\alpha_J)^2{\varphi_J}^2 \tag{2.100}$$

Both terms, $(1-\alpha_j)^2$ and ${\varphi_j}^2$, are positive scalars. In fact:

- being α_j a real coefficient, $(1-\alpha_j)^2$ is a real positive scalar;
- in turn, ${\varphi_j}^2=<\varphi_j(t),\varphi_j(t)>=\varphi_j(t)\varphi_j(t)^T=\sum_t\varphi_j(t)^2$ is also a positive scalar.

On the other hand, from eigenvalues and correlation matrix definitions, and assuming the decreasing order of the basis, (2.101) is verified.

$${\varphi_1}^2\geq{\varphi_2}^2\geq\ldots\geq{\varphi_J}^2\geq\ldots\geq{\varphi_N}^2 \tag{2.101}$$

As result, the error represented by equation (2.100) can be interpreted as a function of two terms:

- The first set of terms, $(1-\alpha_j)^2$, represents the difference in amplitude between the template (weight=1) and the signal to be compared (weight=α_j) in terms of each local basis;
- The second set of terms, ${\varphi_j}^2$, can be seen as the importance of each amplitude difference (from the highest to the lowest magnitude).

Given this idea, a similarity measure between the template $X(t)$ and the sequence $Y(t)$, can be basically defined by the distance between the two vectors of coefficients $\Gamma=[1,\ 1,\ ...,\ 1]$ and $\Omega=[\alpha_1,\alpha_2,...,\alpha_J]$.

$$D(X,Y) \ \simeq \ D(\Gamma,\Omega) \tag{2.102}$$

This distance is computed as an Euclidean distance by (2.103) .

$$D(\Gamma,\Omega)=\sqrt{\ (1-\alpha_1)^2+(1-\alpha_2)^2+\ldots+(1-\alpha_J)^2} \tag{2.103}$$

Finally, to transform this distance measure into a similarity measure, the following conversion is implemented.

$$S(X,Y)=e^{-D(\Gamma,\Omega)} \tag{2.104}$$

This enables to obtain a normalized measure in the interval $[0,1]$, which makes easy to identify the signals that best match the template. A value near zero indicates a very low similarity (high value of distance $D(X,Y)$), while a value of one corresponds to the maximum similarity (distance $D(X,Y)=0$).

In conclusion, using the described procedure, two time series signals are similar if equation (2.105) is verified, where $\boldsymbol{\eta}\in\mathbb{R}^+$ is a positive scalar.

$$S(X,Y)=e^{-D(\Gamma,\Omega)}\leq\boldsymbol{\eta} \tag{2.105}$$

As referred, the aim is to capture the main dynamic characteristics of the signal, that is, the main trends or behaviour of the signal (possibly in some specific time regions) and not to obtain an optimal value of the error in the least-squares sense.

An example of this aspect was depicted in Figure 2.9, where a signal $X(t)$ is described using only three basis (and, therefore, the similarity measure only uses three coefficients). In fact, the key point of this scheme is that a signal can be viewed as composed of a set of relevant local components (wavelet basis).

Additionally, the proposed similarity measure can be easily interpreted. In fact, the distance $(1-\alpha_j)^2$ can be seen as the difference in amplitude between the template $X(t)$ and the signal $Y(t)$, in a specific localization (the local wavelet basis $\varphi_j(t)$). Thus, for a specific local (wavelet basis support), several situations may occur:

- $\alpha_j > 1$: means that the local amplitude of the signal is higher than the template's amplitude;
- $\alpha_j = 1$: implies that local amplitudes of the template and the signal are equal;
- $0 < \alpha_j < 1$: indicates that the local amplitude of the signal is lower than the amplitude of the template;
- $\alpha_j < 0$: suggests that the local amplitude of the signal is the opposite of the template's amplitude.

Thus, independently of its specific value, a positive coefficient ($\alpha_j > 0$) reveals that signal and template present the same behaviour, that is, the same evolution trend. In case of a negative value ($\alpha_j < 0$), it means that signal and template have opposite trends (behaviour). This result is very interesting and important, since the behaviour of a signal, when compared with a template, can be simple and effectively determined by taking into account the signs of the coefficients (α_j).

In effect, one of the hypothesis that is explored in this work is the fact that two signals are similar if, and only if, their coefficients have the same sign. In qualitative terms, it can be stated that, in this situation, they present the same behaviour. Therefore, a simplification in the complexity of the similarity search algorithm can be easily achieved by automatically discarding signals that do not present coefficients with the same sign. Then, in a subsequent phase, only for signals verifying this criterion, the similarity measure is computed to explicitly quantify the proximity between signals (using the threshold $\boldsymbol{\eta}$).

Preliminary discussion

The following two examples illustrate this idea, considering the template $X(t)$ depicted in Figure 2.9*a*), and its description by means of three basis, according to (2.106).

$$X(t) = \sum_{j=1}^{3} \varphi_j(t) = \varphi_1(t) + \varphi_2(t) + \varphi_3(t) \tag{2.106}$$

In the first example, a signal $Y_1(t)$, illustrated in Figure 2.10 and described as follows, is considered.

$$Y_1(t) = \sum_{j=1}^{3} \alpha_j \ \varphi_j(t) = 0.29\varphi_1(t) + 0.14\varphi_2(t) + 0.49\varphi_3(t) \tag{2.107}$$

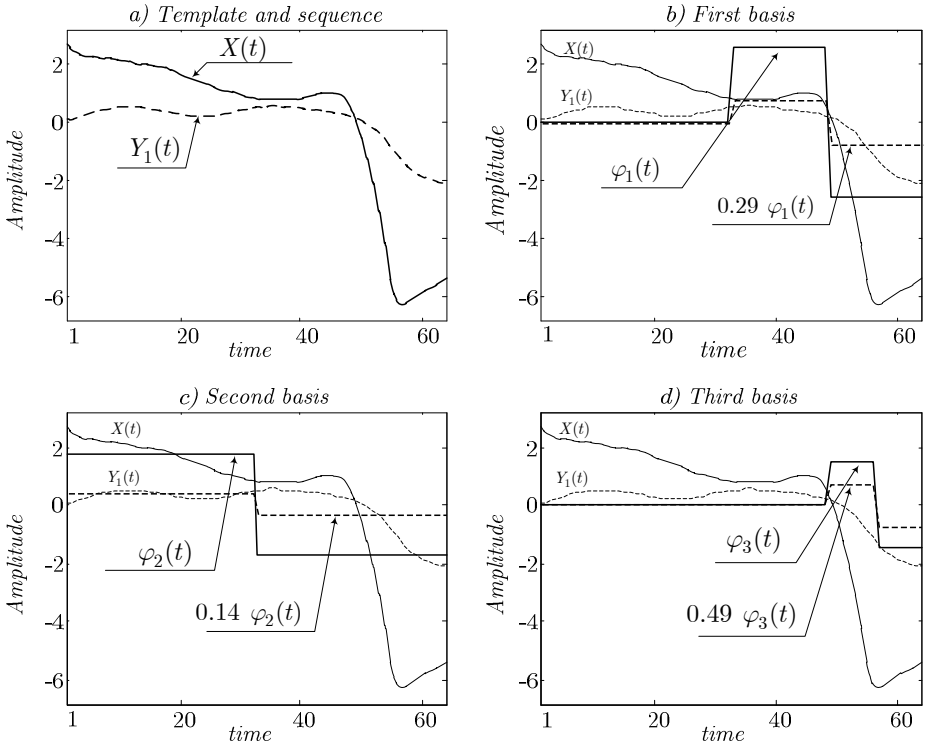


Figure 2.10 - Similarity measure between time series: same behaviour.

All coefficients α_j, $j = 1,2,3$ are positive, thus having the same sign as the ones of the template (all equal to 1). From this simple statement, it can be concluded that template and signal present the same temporal trend. Observing, for example, the second term (basis), $0.14\ \varphi_2(t)$, it can be verified that, although different, globally (for all instants $[1,64]$) signal and template present the same behaviour (the mean of the first half is higher than the mean of the second half). Moreover, the signal presents an average amplitude lower than that of the template ($\alpha_2 = 0.14$).

In a second example, the template is compared with a signal $Y_2(t)$, described as follows and illustrated in Figure 2.11.

$$Y_1(t) = \sum_{j=1}^{3} \alpha_j\ \varphi_j(t) = -0.93\varphi_1(t) - 0.49\varphi_2(t) + 0.21\varphi_3(t) \tag{2.108}$$

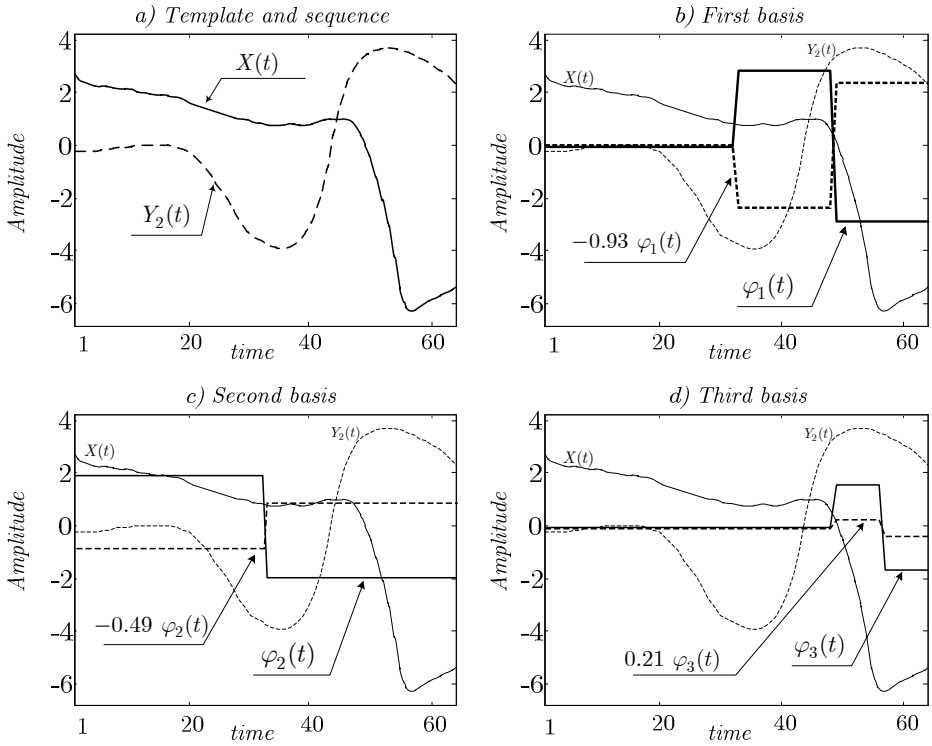


Figure 2.11 - Similarity measure between time series: opposite behaviour.

Two of the coefficients are negative and a third is positive. Observing, for example, the second term, $-0.49\ \varphi_2(t)$, it can be concluded that, globally (for all instants $[1,64]$) signal and template present opposite behaviours: in the corresponding time region, the template decreases (average derivative is negative) and the signal increases (average derivative is positive). Comparing their amplitudes, it can be observed that signal's amplitude is lower than that of the template (approximately 0.5). The positive coefficient corresponds to the third basis, thus, to time instants $[49,64]$. In fact, both signals have a "similar" behaviour: globally, in this region, both signals decrease. Additionally, the respective coefficient (0.21) reveals the high difference in amplitude between the signal and template, as it can be observed.

In conclusion, the positive or negative value of the local basis coefficients (parameters α_j) can approximately characterize the behaviour, that is, the main trend of the signals under comparison. Therefore, the similarity between two signals can be simply assessed by taking into account the signs of these coefficients (α_j). The key principle is that, qualitatively, two signals are similar if they present the same behaviour, or equivalently, if their coefficients have the same signs. This means that the proposed similarity assessment can be applied even in cases in which signals are not perfectly aligned in time. In effect, since the similarity measure is based on the signals' trend and it is not directly dependent on the dynamics at the same points in time, it makes possible the comparison of signals not aligned in time.

2.5.6 Step 5: Similarity Indexing

i. Similarity index

The indexing procedure proposed in the present work, makes use of a windowing scheme to compute the similarity between the template $X(t) \in \mathbb{R}^{1,N}$ and the signal being analysed $T(t) \in \mathbb{R}^{1,T}$ ($T > N$), as illustrated in Figure 2.3. The similarity measure is estimated for each segment, $Y(t) \in \mathbb{R}^{1,N}$, thus, a set of $(T - N)$ similarity measures have to be computed. First, each segment $Y(t)$ is described using the reduced set of J basis functions derived from the decomposition of the template $X(t)$, as following:

$$Y(t) = \sum_{j=1}^{J} \alpha_j \ \varphi_j(t) \tag{2.109}$$

It should be noted that this technique has the advantage of being efficient. In fact, the wavelet decomposition procedure is implemented only once for the template $X(t)$. The coefficients α_j are obtained through a simple matrix multiplication, by means of a pseudo-inverse formulation, equation (2.84). Moreover, the pseudo-inverse, depending on the basis, is calculated just at the beginning and not during the computation of the similarity measure for each sequence.

On the other hand, when using a reduced set of basis, the computation of the pseudo-inverse is highly simplified. From equation (2.87), if only J basis are used, the pseudo-inverse $\Phi^{\dagger} \in \mathbb{R}^{N,J}$ is obtained according to equation (2.110).

$$\Phi^{\dagger} = \Phi^T \left(\Phi\Phi^T\right)^{-1} = \left[\frac{\varphi_1}{<\varphi_1,\varphi_1>} \quad \frac{\varphi_2}{<\varphi_2,\varphi_2>} \quad \ldots \quad \frac{\varphi_J}{<\varphi_J,\varphi_J>}\right] \tag{2.110}$$

Ons/su

(2.81), the J coefficients α_j are then easily obtained by the following equation.

$$\alpha_j = Y\Phi^{\dagger} = < Y(t), \frac{\varphi_j(t)}{<\varphi_j(t),\varphi_j(t)>} > \tag{2.111}$$

Finally, based on these coefficients, the similarity measure for each subsequence is computed using (2.104). In conclusion, the assessment of the similarity between a given template $X(t)$ and each subsequence $Y(t)$ of the signal $T(t)$, is carried out by means of the following procedure.

Pre-processing (computed only once)

- Decompose the template $X(t)$ as a linear combination of a set of wavelet basis
- Reduce this wavelet basis set to the most significant J basis (fixed by the parameter ε)
- Determine the pseudo-inverse from the reduced set of basis

For each subsequence (computed $T-N$ times)

- Calculate coefficients α_j that enable to describe each subsequence $Y(t)$ by means of a linear combination of the reduced set of basis
- Compute the similarity measure, $S(X,Y)$, based on the α_j coefficients
- Save the index of subsequence $Y(t)$ if the similarity measure $S(X,Y) \leq \eta$

ii. Iterative implementation

The previous similarity indexing algorithm can be formulated as an iterative procedure. In fact, taking into account that the basis are fixed and present a compact support, the index of similarity can be computed using an iterative scheme that significantly decreases the computational complexity of the method.

Proposition 2.5

Consider a time series of length N, $Y(t)=\{y(t),y(t+1),\ldots,y(t+N-1)\}$. Also consider a set of basis $\varphi_j(t)$, from where it is possible to define the pseudo-inverse matrix, equation (2.110), composed of elements $K_j(t) \in \mathbb{R}^{1,N}$.

$$K_j(t)=\frac{\varphi_j(t)}{<\varphi_j(t),\varphi_j(t)>} \tag{2.112}$$

Assuming that the coefficient $\alpha_j(t)$ corresponding to the wavelet basis $\varphi_j(t)$, equation (2.111), is known for the time instant t, the coefficient for the next time instant $t+1$, $\alpha_j(t+1)$, can be iteratively computed as (2.113), with $Y(t+1)=\left\{y(t+1),\ldots,y(t+N)\right\}$.

$$\alpha_j(t+1)=\alpha_j(t)+f\left(K_j(t),Y(t+1)\right) \tag{2.113}$$

Proof

Without loss of generality, the Haar wavelet is considered, being the results easily extended to other type of wavelets. Figure 2.12 depicts a generic Haar wavelet basis $\varphi_j(t)$, with length N, and a signal $Y(t)$.

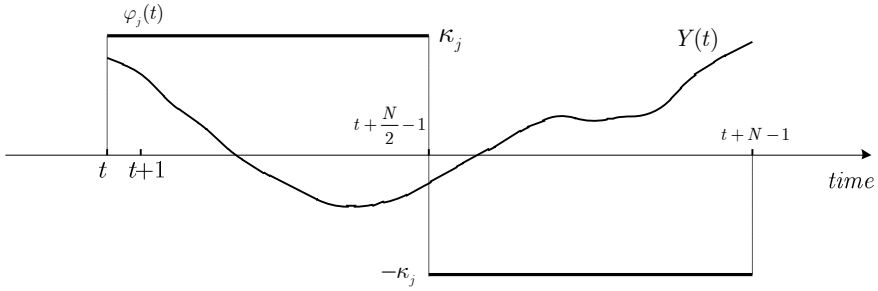


Figure 2.12 - Similarity indexing: iterative implementation.

For this example, the element $K_j(t)$ of the pseudo-inverse matrix is defined as (2.114).

$$K_j(t) = [\ \underbrace{\kappa_j \ \ldots \ \kappa_j}_{t=0,\ldots,\frac{N}{2}-1} \ \ \underbrace{-\kappa_j \ldots \ -\kappa_j}_{t=\frac{N}{2},\ldots,N-1}\ ]^T \tag{2.114}$$

At instant time t, the coefficient $\alpha_j(t)$ is given by the product of the signal $Y(t)$, by the respective column of the pseudo-inverse, $K_j(t)$. Thus, the coefficient $\alpha_j(t)$, equation (2.111), can be computed as :

$$\begin{aligned}\alpha_j(t)=&\kappa_j\left(y(t)+y(t+1)+\ldots+y(t+\frac{N}{2}-1)\right)\\&-\kappa_j\left(y(t+\frac{N}{2})+\ldots+y(t+N-1)\right)\end{aligned} \tag{2.115}$$

The coefficient $\alpha_j(t+1)$, computed at instant $(t+1)$, is determined by

$$\begin{aligned}\alpha_j(t+1)=&\kappa_j\left(y(t+1)+y(t+2)+\ldots+y(t+\frac{N}{2})\right)\\&-\kappa_j\left(y(t+\frac{N}{2}+1)+\ldots+y(t+N)\right)\end{aligned} \tag{2.116}$$

Therefore, the following relationship is straightforwardly obtained from equations (2.115) and (2.116).

$$\alpha_j(t+1)=\alpha_j(t)+\kappa_j\left(-y(t)-y(t+N)+2\ y(t+\frac{N}{2})\right) \tag{2.117}$$

As a result, instead of computing the coefficients at each time instant using the pseudo-inverse formulation, they can be iteratively obtained by means of equation (2.117). This way, each coefficient only depends on the wavelet amplitude (κ_j), and on the first, last, and middle values of the signal under analysis, $Y(t)$.

In conclusion, the dimensionality reduction method (wavelet decomposition), that can be seen as relatively costly, is applied only once. In turn, the similarity measure to be calculated for each subsequence $Y(t)$, therefore $T-N$ times, is based on coefficients iteratively computed, enabling to obtain a global very efficient scheme.

2.5.7 Complexity analysis

The efficiency of an algorithm can be assessed following several approaches, namely in terms of the execution time (time complexity) and of the amount of memory required (space complexity). Here, only the time complexity is addressed and, in particular, the necessary number of arithmetic operations.

Considering the similarity indexing process, the main goal of the present section is to determine the number of operations involved in the computation of the similarity between a template $X(t) \in \mathbb{R}^{1,N}$ and a signal $T(t) \in \mathbb{R}^{1,T+N}$. Note that, without loss of

correctness, it is assumed that the length of the signal $Y(t)$ is $(T+N)$, instead of (T), to simplify the results. Consequently, a set of T individual similarity measures have to be computed.

According to the proposed algorithm, three parameters determine the number of operations: *i)* N, the length of the template $X(t)$; *ii)* T, the length of the signal $Y(t)$, typically ($T \gg N$); *iii)* J, the number of wavelet basis used in the representation of the signal.

The number of operations required to implement the proposed approach is compared with the number demanded by the Euclidean distance approach. No distinction is made regarding the type of operations involved, namely additions, subtractions, multiplications and divisions.

Euclidean distance based similarity indexing

This similarity indexing strategy computes the Euclidean distance, ${D_E}^2(X,Y)$, between the template $X(t)$ and a subsequence $Y(t) \in \mathbb{R}^{1,N}$ of the signal $T(t)$.

$$ {D_E}^2(X,Y) = \sum_{t=1}^{N} \Big(X(t) - Y(t) \Big)^2 \tag{2.118} $$

It is straightforward to write (2.118) as (2.119).

$$ {D_E}^2(X,Y) = \Big(X(t_1) - Y(t_1) \Big)\Big(X(t_1) - Y(t_1) \Big) + \ldots + \Big(X(t_N) - Y(t_N) \Big)\Big(X(t_N) - Y(t_N) \Big) \tag{2.119} $$

Two operations are needed at each time instant t_i, $\Big(X(t_i) - Y(t_i) \Big)\Big(X(t_i) - Y(t_i) \Big)$, namely one subtraction and one multiplication. Moreover, a total of $(N-1)$ additions must be performed to obtain the distance ${D_E}^2(X,Y)$.

In conclusion, the number of operations, $nop(N)$, for computing the distance (2.118), is given by (2.120).

$$ nop(N) = 3N - 1 \tag{2.120} $$

Considering the computation through all the signal $T(t)$, the total number of operations result in (2.121).

$$ nop(N,T) = T(3N-1) \tag{2.121} $$

Wavelet based similarity indexing

As referred, the similarity indexing method proposed in the present work involves two main phases. In the first, computed only once, the template $X(t)$ is described by means of a reduced set of basis and a pseudo-inverse is determined from this set. Then, in a second phase, each subsequence $Y(t)$ is described by means of the obtained reduced set of basis, being the respective coefficients iteratively computed using the pre-determined pseudo-inverse values. The obtained coefficients are employed to evaluate a distance measure, similarly to equation (2.118).

Step 1: Pre-processing (computed only once)

i) Decompose the template $X(t)$ as a linear combination of a set of wavelet basis
ii) Reduce the wavelet basis set to the most significant J basis (established by the parameter ε)
iii) Determine the pseudo-inverse from the reduced set of basis.

i. Decomposition of the template

As depicted in Figure 2.7 (the main steps of the similarity scheme), the process starts by describing the template $X(t) \in \mathbb{R}^{1,N}$ as the sum of a orthonormal Haar wavelet basis set, $\psi_j(t) \in \mathbb{R}^{1,N}$, equation (2.122), where the k_j values are the coefficients associated with each basis $\psi_j(t)$.

$$X(t) = \sum_{j=1}^{N} k_j \ \psi_j(t) \tag{2.122}$$

For a template with length N, N basis have to be considered, equation (2.62).

Although fast implementations of the Haar wavelet decomposition have been proposed for computing the N coefficients (Kopenkov, 2008), a conventional implementation is assumed here (Chang and Piau, 2007). Basically, a set of $2(N-1)$ operations of type $(x+y)/2$ or $(x-y)/2$ is performed, resulting from the successive decomposition of the signal in details and approximations. As a consequence, the number of operations is given by (2.123).

$$nop(N) = 4(N-1) \tag{2.123}$$

ii. Selection of the reduced set of basis

The reduction of the number of basis from N to J, implies that the error equation is verified (2.124), where $\widehat{X}(t)$ is the approximated signal described by means of these J basis, (2.125).

$$e^2(t) = \left(X(t) - \widehat{X}(t) \right)^2 < \varepsilon \tag{2.124}$$

$$\widehat{X}(t) = \sum_{j=1}^{J} k_j \ \psi_j(t) \tag{2.125}$$

Since the basis $\psi_j(t) \in \mathbb{R}^{1,N}$ are orthonormal (Haar wavelet basis), the determination of the reduced set of basis only involves the identification of the highest coefficients. As a consequence, a sorting operation of an array of length N is required (the array composed of the coefficients). A typically sorting method is assumed, such as the quicksort, resulting in $N \log_2(N)$ operations (Sedgewick, 1978).

$$nop(N) = N \log_2 (N) \tag{2.126}$$

After having sorted the coefficients, the adequate number of basis has to be determined. Two phases are required: *i*) describe the approximated signal using these basis (2.125) and *ii*) compute the error between the actual and the approximated signals (2.124).

Considering only one basis, $j=1$, for the first phase N operations (multiplications) are necessary. For the second phase, $3N-1$ operations are necessary, similarly to the Euclidean strategy, equation (2.120). Considering J basis, $j=J$, it is possible to extend these results to $\Big(JN+(J-1)N\Big)$ and $(3N-1)$, respectively for the first and the second phases. Since the determination of the adequate number of basis involves all the computations from $j=1$ to $j=J$, the total number of operations is determined by the sum of operations corresponding to the phases 1 and 2.

	Phase 1		**Phase2**
$j=1$	N	$+0$	$3N-1$
$j=2$	$2N$	$+N$	$3N-1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$j=J$	JN	$+(J-1)N$	$3N-1$

As a consequence, for the first phase the total of operations is given by (2.127).

$$\begin{aligned} nop(N,J) &= N(1+2+\cdots+J)+N\Big(0+1+\cdots+(J-1)\Big) \\ nop(N,J) &= \frac{JN}{2}(J+1)+\frac{JN}{2}(J-1)=J^2N \end{aligned} \tag{2.127}$$

In turn, for the second phase, the total of operations is given by

$$nop(N,J)=J(3N-1) \tag{2.128}$$

iii. Pseudo-inverse computation

The third step demands the computation of the pseudo-inverse elements $K_j(t)\in\mathbb{R}^{1,N}$, according to equation (2.112).

$$K_j(t)=\frac{\varphi_j(t)}{<\varphi_j(t),\varphi_j(t)>}$$

For the inner product, $(2N-1)$ operations are performed. The coefficient $K_j(t)$ is obtained through N additional divisions. The total number of operations for the J basis is given by (2.129).

$$nop(N,J)=J(3N-1) \tag{2.129}$$

Moreover, using the proposed iterative algorithm, the coefficients α_j are, at the initial instant, obtained through the product of the subsequence $Y(t)$ by the pseudo-inverse matrix $\Phi^{\dagger}$, equation (2.111). Consequently, $J(2N-1)$ operations are required.

As conclusion, for the step 1 of the proposed method, the total number of operations is obtained from equations (2.123), (2.126), (2.127), (2.128), and (2.129), resulting in equation (2.130).

$$\begin{aligned} nop(N,J) &= 4(N-1)+N\log_2 N+J^2N+J(3N-1)+J(3N-1)+J(2N-1) \\ nop(N,J) &= (4N-1)+N\log_2 N+J^2N+J(8N-1) \end{aligned} \tag{2.130}$$

Step 2: For each subsequence (computed T times)

i) Calculate coefficients α_j that enable to describe each subsequence $Y(t)$ by means of a linear combination of the reduced set of basis

ii) Compute the similarity measure, $S(X,Y)$, based on the α_j coefficients.

i. Subsequence description using the reduced set of basis

The computation of coefficients α_j can be determined by means of an iterative procedure. In fact, for updating each coefficient, a set of five operations is needed (three additions/subtractions and two multiplications), equation (2.117).

$$\alpha_j(t+1)=\alpha_j(t)+\kappa_j\left(-y(t)-y(t+N)+2\ y(t+\frac{N}{2})\right)$$

For the J coefficients, the number of operations is given by (2.131).

$$nop(J)=5J \tag{2.131}$$

ii. Wavelet based distance

Finally, the obtained coefficients are used to compute a distance from where the proposed similarity measure is evaluated, according to (2.132).

$$D^2(\Gamma,\Omega)=(1-\alpha_1)^2+(1-\alpha_2)^2+\ldots+(1-\alpha_J)^2 \tag{2.132}$$

It is straightforward to conclude that the number of operations is given by (2.133).

$$nop(J)=3J-1 \tag{2.133}$$

Considering the computation through all the signal $T(t)$, the total number of operations is given by (2.134).

$$nop(N,J,T)=\underbrace{4(N-1)+N\log_2 N+J^2N+J(8N-1)}_{Step\,1}+\underbrace{T(8J-1)}_{Step\ 2} \tag{2.134}$$

Euclidean vs. Wavelet based similarity indexing

Finally, in order to compare the complexity of the approaches, two assumptions are made. Basically, both the parameters T and J, respectively the length of the signal and the number of basis, can be described as a function of N (the length of the template), according to equations (2.135) and (2.136).

$$T = nN,\ n \in N \tag{2.135}$$

$$J = m\log_2 N,\ m \in N \tag{2.136}$$

The first assumption is acceptable given that T is typically larger than N. On the other hand, a reduced number of basis is usually adequate for representing the template, since a rough estimation of the signal is enough in terms of the proposed similarity assessment strategy. Thus, for a template of length N a number of basis equal to $J = \log_2 N$ is commonly appropriated. This assumption can be confirmed by the example presented in section 2.5.5, where it is possible to describe a template of size $N = 32$ with only three basis ($J = 3$).

In conclusion, for the Euclidean distance the complexity is of order $O(N^2)$.

$$nop(N) = nN(3N-1) \tag{2.137}$$

Regarding the wavelet approach, the complexity is of order $O(N(\log_2 N)^2)$ and $O(N\log_2 N)$, respectively for the first and the second steps.

$$\begin{aligned} nop(N) = & \underbrace{4(N-1) + N\log_2 N + \left(m\log_2 N\right)^2 N + m\log_2 N(8N-1)}_{Step\ 1} + \\ & + \underbrace{nN(8m\log_2 N - 1)}_{Step\ 2} \end{aligned} \tag{2.138}$$

The Figure 2.13 illustrates the variation of the parameters N, T and J, and the corresponding effect on the total number of operations. The different values considered for each parameter are:

- N: $N = \{2,4,8,16,32,64,128\}$
- T: $n = \{1,2,4,8,16,32,64,128\}$ resulting in
  $$T = \{32,64,128,256,512,1024,2048,4096\}$$
- J: $m = \{1,2,3,4,5,6\}$ leading to $J = \{5,10,15,20,25,30\}$.

The default values adopted are $N = 32$, $T = 512$ and $J = 5$.

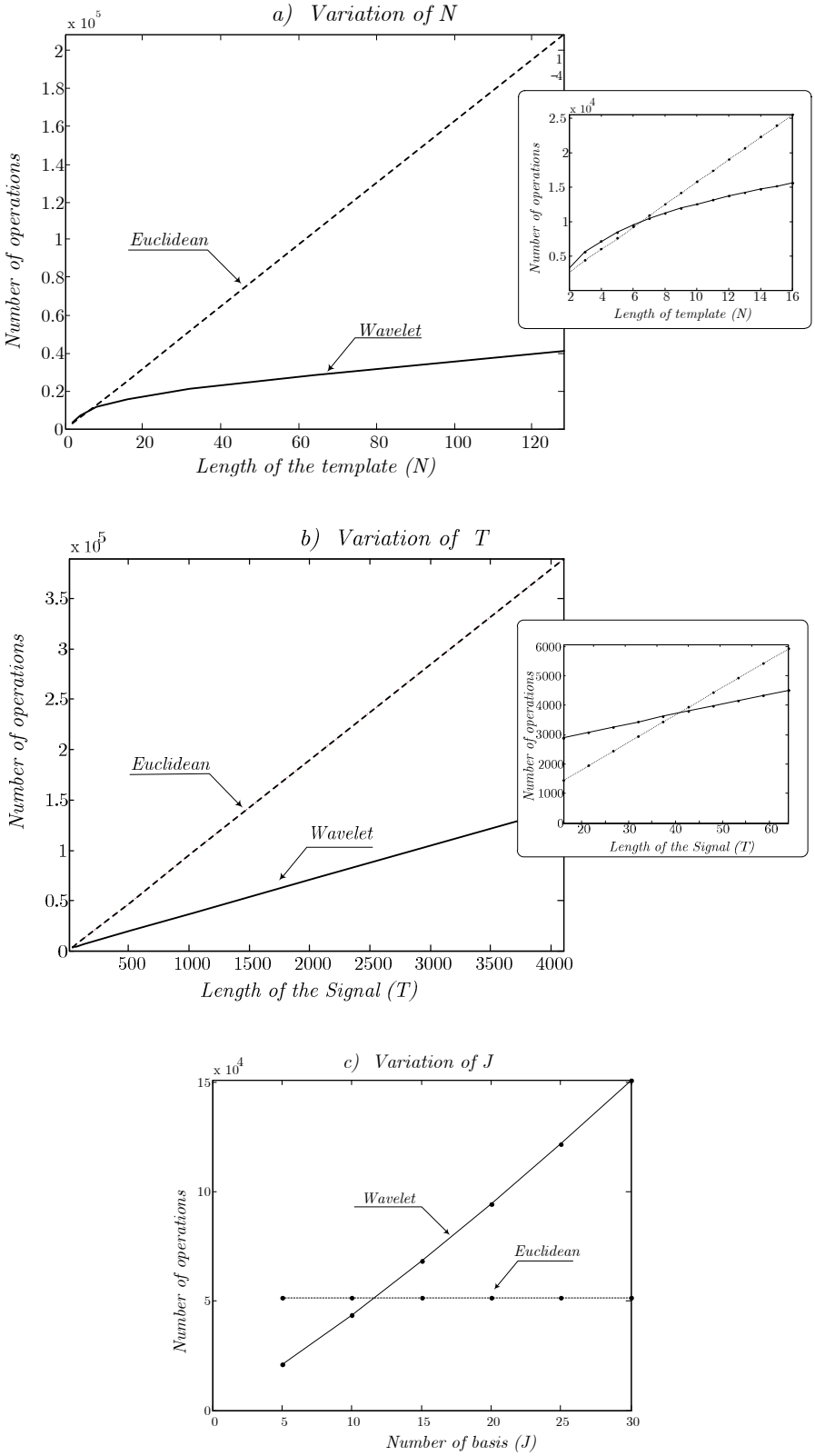


Figure 2.13 - Effect of variations in N, T and J on the number of operations.

As can be observed in Figure 2.13*a*), the wavelet approach is clearly superior (in terms of the number of operations) for larger values of N. Since the step 1 of the proposed wavelet strategy requires a high number of operations, equation (2.130), for small values of N the Euclidean methodology is advantageous. However, for approximately $N > 6$ $(T = 512\,,\ J = 5)$, the situation is reversed.

In the case of T variation, depicted in Figure 2.13*b*), a similar conclusion can be taken. In fact, for approximately $T > 40$ $(N = 32\,,\ J = 5)$, the proposed strategy presents a number of operations manifestly inferior as T increases.

With respect to the variation in the number of basis (J), the number of operations required by the Euclidean approach remains constant, since it does not depend on J, equation (2.121). In turn, for approximately $J > 12$ $(N = 32\,,\ T = 512)$ the high number of operations required does not favour the wavelet proposed approach.

In conclusion, the proposed strategy is especially advantageous when the sizes of the template $X(t)$ and of the signal $T(t)$ are greater than given values, that are smaller than those usually employed in practice.

2.5.8 Multi-dimensional Time Series

The proposed scheme can be straightforwardly extended to a multi-dimensional time series framework.

Consider a template, $\mathbf{X}(t) \in \mathbb{R}^{S,N}$, composed of a set of S individual time series with length N, $X_i(t) \in \mathbb{R}^{1,N}$, as described by (2.139).

$$\mathbf{X}(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_S(t) \end{bmatrix} \tag{2.139}$$

Also consider a time series, $\mathbf{T}(t) \in \mathbb{R}^{S,T}$, composed of a set of S individual time series with length T, $T_S(t) \in \mathbb{R}^{1,T}$, as described by (2.140).

$$\mathbf{T}(t) = \begin{bmatrix} T_1(t) \\ T_2(t) \\ \vdots \\ T_S(t) \end{bmatrix} \tag{2.140}$$

A multi-dimensional similarity measure, $S_M \in \mathbb{R}^{S,1}$, between the template $\mathbf{X}(t)$ and a subsequence of the time series $\mathbf{T}(t)\,,\ \mathbf{Y}(t) \in \mathbb{R}^{S,N}$, can be defined as:

$$\mathbf{S}_M = \begin{bmatrix} S_M^1(t) = D\big(X_1(t), Y_1(t)\big) \\ S_M^2(t) = D\big(X_2(t), Y_2(t)\big) \\ \vdots \\ S_M^S(t) = D\big(X_S(t), Y_S(t)\big) \end{bmatrix} \tag{2.141}$$

The global similarity, $S_{Mglobal} \in \mathbb{R}^+$, can be simply defined by a weighted sum of the distances, as follows:

$$S_{Mglobal} = \sum_{i=1}^{S} \gamma_i \; S_M^i(t) \tag{2.142}$$

The parameter γ_i denotes the weight factor for the i^{th} similarity measure. In this way, two multi-dimensional sequences are said to be similar if the sum of individual distances is within a user-defined threshold, $\boldsymbol{\eta}_{global} \in \mathbb{R}^+$ (a positive scalar).

$$S(\mathbf{X}(t), \mathbf{Y}(t)) = \mathbf{S}_{Mglobal} \leq \boldsymbol{\eta}_{global} \tag{2.143}$$

2.5.9 Applications of the Similarity Measure

The aim of the previous research (described in the last section) was to propose an efficient technique for indexing time series similarity, based on a distance measure between the coefficients of a description in terms of wavelet basis. Although the main goal is the detection of similar conditions to support detection and prediction tasks, other several data mining applications, namely classification and clustering, can be based on this similarity measure.

Basically, data mining refers to the overall process of discovering useful knowledge in the data. Classification and clustering are employed as important knowledge discovery tools: the goal of clustering is to identify organization in an unlabelled dataset by arranging data into homogeneous groups; in turn, classification assigns input data to one or more pre-specified classes, based on the extraction of significant similar attributes or features.

In effect, the characterization of clinical conditions can be linked with a multitude of patients' factors, in the form of attributes or features, reflected in the collected physiological time series. These features, such as the similarity measure previously described, can then be used to characterize a clinical condition, thus producing valuable information for clustering and classification applications.

Clustering

Clustering categorizes data into groups that are not typically predefined, but defined by the data itself. In effect, the organization in groups is based on some evaluation of similarity between time series, such that the similarity measure within the same group is minimized and the dissimilarity measure between different groups is maximized. As a

result, it is commonly designated as unsupervised learning. The clustering process is, therefore, carried out by determining similarities among data, based on a set of characteristic features.

The majority of algorithms for clustering analysis have been explicitly developed to work with static data, Kavitha and Punithavalli (2010). In effect, the high dimensionality and very high feature correlation that characterize dynamic time series data present some difficult challenges. Han and Kamber (2001) proposed five categories to classify the existent static clustering methods: *i*) partitioning, *ii*) hierarchical, *iii*) density-based, *iv*) grid-based, and *v*) model-based methods.

Essentially, given a set of N unlabelled data tuples, a partitioning method creates k partitions of the data, being this number a priori defined. Each partition represents a cluster, with each cluster containing at least one object. Examples of those methods are the well-known k-means clustering, in case the partitions are crisp, or the fuzzy c-means, otherwise (Bezdek, 1981). Similarity measures, such as the Euclidean distance, are employed to evaluate similarities between groups of data. A hierarchical clustering organizes data instances into a tree of clusters (Fernandez and Gomez, 2008). Thus, a hierarchical representation is achieved, which structure is more informative than the unstructured set of clusters obtained by flat clustering (which includes partitioning methods). This type of clustering does not require the a priori definition of the number of clusters and generally involves two types of methods: top-down or bottom-up. The last, treat each object as an independent cluster and then successively merge (or agglomerate) similar pairs of clusters until all clusters have been merged into a single cluster containing all objects. Density-based methods, such as the subtractive clustering (Kriegel *et al.*, 2011), apply a local density cluster criterion. Clusters are considered as regions in the data space in which the objects are dense, and are separated by regions of low density objects. These regions may have an arbitrary shape and the points inside a region may be arbitrarily distributed. In opposition, grid-based methods simply divide the object space into a finite number of cells that form a grid structure (Ilango and Mohan, 2010). Thus, this clustering approach differs from the other algorithms in the sense that it is concerned not with the data points but with the space where the data objects exist. Finally, model-based methods assume a model for each of the clusters and attempt to best fit the data to the assumed model. In this case, the issues of selecting an adequate clustering method and determining the correct number of clusters, are reduced to model selection problems. Gaussian models are examples of powerful clustering models that have been reported in many applications (Lourme, 2010).

In parallel, many different algorithms have been proposed for clustering dynamic time series data (Antoniadis *et al.*, 2003). Nevertheless, a possible approach is to transform the dynamic problem into a static one, so that the static algorithms can be applied. The key idea is to convert the original time series data into a feature vector of lower dimension or into a number of model parameters, and then apply a conventional clustering algorithm to them.

Subtractive clustering using the proposed similarity measure

Using the proposed similarity measure, a set of features (namely the coefficients of wavelet basis as previously described) can be used to derive a metric that enables to assess similarity between time series, from where clustering techniques can then be applied.

Furthermore, among the several static algorithms, density-based approaches have a great potential in the context of this work. One of the main hypothesis consists in the identification of patterns in the historic dataset that are similar to the current template. Using a density-based strategy, where clusters are considered as regions of the data space in which the objects are dense, it is possible to identify the areas in the historic dataset where the similarity is more evident. In particular, the Subtractive Clustering method (SC) presents some advantages that are decisive in the present work (to be presented in Chapter 3): the cluster centres are a subset of the actual data, the method has the capacity to naturally deal with outliers and to discard redundant data, as well as to cope with noisy data.

The subtractive clustering method was originally proposed by Chiu (1994), as a basis for a fast and robust algorithm for identifying fuzzy models. This method is a modification of the mountain clustering (Yager and Filev, 1993) which was also used to identify the number of fuzzy rules. Basically, the differences between the two methods are in the approach they use to estimate potential values and the influence of a neighbouring data point.

The subtractive clustering method assumes that *i*) each data point is a potential cluster centre; *ii*) a data point with more neighbouring data will have a higher opportunity to become a cluster centre than points with fewer neighbouring data.

Based on these assumptions, a density measure of surrounding data points assessing the potential value for each data point, $PV(p_i)$, is calculated as follows:

$$PV(p_i) = \sum_{j=1}^{n} exp\left(\frac{-4\left\|p_i - p_j\right\|^2}{R_a^2}\right) \quad i = 1,\ldots,n \tag{2.144}$$

The variables p_i and p_j represent data points, n is the number of data points, and R_a is a positive constant that allows to define the concept "neighbourhood". According to this definition, equation (2.145), data outside this range have little influence on the potential.

After the potential of every data point has been computed, the data point with the highest potential is chosen as the first cluster centre. Let C_1 be the location of the first cluster centre and PV_1 the corresponding potential value. The potential of the remaining data points p_i is then updated by:

$$PV(p_i) = PV(p_i) - PV_1\, exp\left(\frac{-4\left\|p_i - P_1\right\|^2}{R_b^2}\right) \tag{2.145}$$

The parameter R_b is a positive constant such that ($R_b > R_a$). This constant is the radius defining the neighbourhood that will have considerable reduction in potential. Moreover, to avoid obtaining closely spaced cluster centres, R_b is greater than R_a. By means of this approach, the data points near the identified cluster will have greatly reduced potential and, as a consequence, it is unlikely to be selected as the next cluster centre.

Normally, to simplify the process, the parameters R_a and R_b are related trough a parameter designated as the Squash Factor (SF), defined by (2.146).

$$SF = \frac{R_b}{R_a} \tag{2.146}$$

The process involving equations (2.145) and (2.146), continues until no further cluster centre is found. To decide whether to stop this process, two parameters are taken into account: the accept ratio and the reject ratio. These parameters, together with the influence range and squash factor, set the four criteria for the selection of cluster centres. For example, the first criteria states that, if the ratio of the potential value of the current data point to the original first cluster centre is larger than the accept ratio, then the current data point is chosen as a cluster centre. Therefore, the larger the value of the accept ratio, the fewer the number of chosen cluster centres.

The Figure 2.14 illustrates the application of such similarity measure to the clustering of an ECG signal. Essentially, a template with length $N = 64$ was chosen, approximately corresponding to a QRS complex. Based on this template, four groups of time series sequences presenting identical similarity measures were determined, as shown in Figure 2.14. By observing this figure, it is clear the capacity of the used approach in identifying homogeneous sequences of time series.

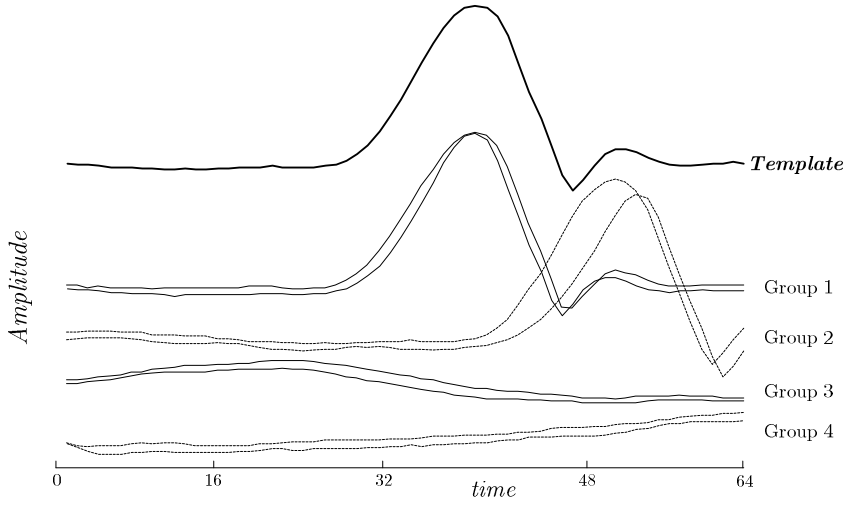


Figure 2.14 - Partitioning clustering of ECG time series.

Classification

Classification is possibly the most relevant application of the data mining techniques. Examples of implementations involving classifiers include image and pattern recognition, spam filtering, medical diagnosis and the classification of time series events. Although analogous to clustering, classification maps input data into predefined classes and, since the classes are determined prior to the data analysis, it is often referred to as supervised learning. A well-known type of classification problem is the pattern recognition, where an input pattern is classified in one of the several classes, based on its similarity with these predefined classes.

Among the most popular methods in time series classification, the following can be emphasized (Spiegel, *et al.*, 2001) (Lucas, 2010) *i*) instance-based learning classifiers; *ii*) decision trees; *iii*) Bayes classifiers and *iv*) neural networks.

Instance based learning algorithms consist of storing a set of training examples (training dataset) and when a new instance is encountered, a set of similar related instances is retrieved from memory and used to classify the query instance (target function). One of the most well-known instance-based learning algorithm is the k-nearest neighbour (kNN). In this method, an object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours (where k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbour. Decision trees classifiers organize a series of data into a tree structure, classifying instances based on decision nodes or decision rules. The basic idea is to break up a complex dataset into a union of several simpler groups, thus providing a solution that is often easier to interpret. Decision trees can be translated into a set of rules by creating a separated rule for each path, from the root to a leaf in the tree. However, rules can also be directly induced from a training data using a variety of rule-based algorithms (Gupta and Toshniwal, 2011). Rather than a deterministic classification, probabilistic/statistical learning algorithms give the probability of an instance to belong to a given class. An important classifier in this domain is the naïve Bayes (Tsymbal, 2003). It assumes a particular configuration of a Bayesian network, which is composed of a directed acyclic graph with only one parent (unobserved node) and several children (observed nodes). Finally, a neural network consists of a set of basic units (neurons) which can be arranged in layers that are combined to form the network. The architecture of a neural network is defined by the number of neurons and by the way they are connected. Basically, for a given input, a neural network is trained to provide a desired output. Given their properties, namely the capacity to learn from data examples, the ability to generalize and the universal approximation, neural networks have been extensively applied to solve classification problems in several domains (industry, medicine, etc.). In effect, neural networks are one of the most successful classifiers, and assume a key importance in this work. Classification and prediction schemes involving neural networks will be presented in the next chapter.

Prediction

In opposition to the classification problems, the main goal of the prediction methods is to forecast a future state, rather than categorize a current one. Time series prediction algorithms typically involve regression analysis, where the forecast of future values is based on historical trends and statistics. For example, the prediction of acute hypotensive episodes can be implemented through the development of predictive models using an historical of blood pressure signals.

One of the major goals of the present work, in the context of prediction, is the forecasting of time series in situations where similar conditions/patterns have been observed. To this aim, a generic methodology consisting of two main phases is proposed, which will be detailed in the next chapter. In the first phase, a similarity analysis procedure is carried out, between the current time series signal and the historical dataset. From this analysis, the most similar conditions are identified and the proper prediction models derived. Subsequently, in a second phase, these models are employed to the current time series signal (template) to predict its future evolution, as illustrate in Figure 2.15. In effect, the investigation of prediction schemes is the main subject of the next chapter.

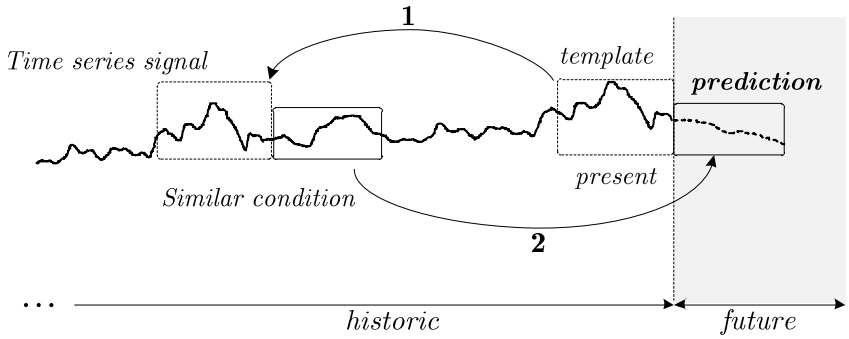


Figure 2.15 - Prediction of the time series future evolution.

2.6 Conclusions

The major goal of the present chapter was to propose methods for similarity measuring and indexing in physiological time series. In this sense, techniques able to effectively detect similarities in time series data were investigated, namely those founded on time-frequency transforms. Methodologies, such as the Wigner-Ville and wavelet transforms were reviewed, given their ability to represent time and frequency, simultaneously. On the other hand, the transformation of a time series by means of orthogonal spaces enabling to approximate it through a set of basis, was explored. In the context of this work, Karhunen-Loève, wavelet, and Hermite basis functions, assumed a special relevance.

A similarity measure based on the natural set of features generated by the wavelet transform (coefficients of the basis functions), was introduced supported on five propositions.

Propositions 2.1, 2.2 and 2.3 - Through a reduction dimension performed by the Karhunen-Loève transform (KLT), an optimal reduced number of wavelet basis is obtained and the corresponding coefficients are the origin of the comparison scheme.

Basically, by the referred reduction procedure, an effective description of the time series in the most representative regions (local wavelets basis) is achieved, providing a practical comparison between time sequences. Additionally, the used wavelet decomposition scheme makes possible a rough estimation of the signal (behaviour or trend) to a predefined degree. Therefore, a simple and interpretable characterization of time series similarity is attained, mainly based on the identification of the trends.

Proposition 2.4 – Based on the description of the time series by means of the reduced set of basis, a similarity measure is defined ensuring that no false dismissals occur.

Proposition 2.5 - With respect to the search of subsequences in a time series (similarity indexing) an iterative formulation was proposed, supported on the properties of wavelets. In fact, taking into account the wavelet compact support, the assessment of similarity between the template and each subsequence can be done based on coefficients computed by means of an iterative scheme, which leads to a significant decrease in the number of operation required. Consequently, a very efficient similarity indexing procedure is accomplished.

Although taking as the starting point the Euclidean distance as a similarity measure, the proposed strategy is able to circumvent some of its major weaknesses that are relevant in the clinical context. In effect, it provides a significant computational complexity decrease by implementing a reduction procedure that results in an optimal reduced set of orthogonal basis (in the sense of least square error). Additionally, through the rough estimation of the signal achieved by orthogonal wavelet transform, the similarity

measure is based on the signal trend and, thus, may provide the comparison of signals not aligned in time. On the other hand, it benefits from the ability of wavelets in dealing with the non-stationarity and non-linearity that are typical of biosignal time series, as well as from their capacity in coping with the presence of noise in the signal, which significantly contributes to improve the similarity searching process.

Several known orthogonal wavelets could be used in the context of the proposed scheme, namely Haar and Daubechies wavelets. In fact, both have a compact support, fundamental to the iterative similarity indexing procedure previously described in this chapter. However, the Haar wavelet was selected due to its fast and easy computation. The fact that it is not adequate to deal with smooth functions (because of its ladder-like approximation), is not fundamental for the present work. In effect, the final goal is to identify the main characteristics of the signal, that is, its main trends or behaviour (possibly in some specific time regions), and not to obtain an optimal value of the error in the least-squares sense.

As mentioned in the course of the chapter, a simple way to verify the similarity between two time series consists in analysing the coefficients resulting from their description in terms of the reduced set of basis: "*two time series are similar if they present the same behaviour or, equivalently, if their coefficients present the same sign*". By means of a windowing technique, this procedure is applied to compare a template with each subsequence of the signal being analysed in order to derive the similarity index. Thus, since the template is represented by positive coefficients (all ones) a subsequence is similar to the template if it is also described by means of positive coefficients. The behaviour of the signals, that is, the granularity of the trends, is determined by the parameter ε, which implicitly defines the number of wavelet basis. On the other hand, the question of when two signals presenting the same behaviour are or are not similar, is addressed by the parameter η, by basically evaluating the relative difference between their local amplitudes. Finally, taking into account that the Haar wavelet basis are fixed and present a compact support, the computational complexity of the similarity indexing procedure can be decreased by using an iterative scheme for calculating the coefficients of the sequence being compared with the template.

In conclusion, a time series similarity indexing scheme based on an optimal basis description obtained from the combination of wavelet decomposition with Karhunen-Loève transform was presented in this chapter. This scheme has the potential to be an effective and appropriate tool in identifying physiological patterns in biosignal time series, which is one of the central points of this work. Furthermore, the time series description by means of the derived reduced set of basis (coefficients) can directly support clustering and classification problems.

The next chapters will address the exploitation of the proposed similarity measure in identifying clinical patterns in time series, namely as a basis for predictive methodologies.

3. Time Series Prediction

This chapter addresses the research and implementation of methodologies for time series prediction, mainly to support the early detection of critical events. Anchored in the developed similarity measure for the selection of patients who display similar behaviours in their physiological time series, two main strategies are proposed.

The first strategy consists in a prediction scheme based on multi-models. Regression neural network structures, previously trained with a set of representative patterns established by the similarity analysis procedure, are chosen to integrate the multi-models scheme, given their approximation and generalization properties. Contrasting with auto-regressive model structures, multi-models do not recursively use model outputs as inputs for step ahead predictions. As result, prediction errors are not propagated over the forecast horizon and future values can be accurately estimated.

The second strategy does not involve an explicit model. It is based on the multi-resolution analysis of the historic similar time series, resulting from the previous similarity analysis procedure. The wavelet decomposition disaggregates time series into different components, allowing an adequate separation of the main trends from the time series. Using a distance-based measurement optimization process, the most appropriate decomposition levels are identified and combined to directly provide an estimation of the time series future evolution.

3.1 Introduction

It is unquestionable that no one knows the future, but there is a lot of benefits in deriving predictions to form a picture of what the future could be, even if this prediction is imprecise. By definition, prognosis is a medical term used to describe the likely outcome of the future health status of a patient and/or the probable evolution of his health indicators (Langlois, 2002). Prognosis is, therefore, one of the key components involved in the care delivery process and patient management, and almost all clinical decisions are based on it. In effect, since a current decision will become operational at some point in the future, it should be, definitely, based on forecasts of future conditions.

On the other hand, it is recognized that a relevant source of information for supporting prognosis may be the patient's historical of similar health status or, as an alternative, the historical of other patients that experimented similar behaviours in their health data. Thus, the development of automatic systems able to estimate future values of health status based on its past values, is of central importance in this framework. Moreover, these systems might reveal underlying relationships between temporal patterns and the onset of significant clinical conditions that are not evident or hidden in the data.

This chapter addresses the development of methodologies that may be used to predict specific clinical conditions, based on the temporal trajectories of related physiological signals and/or on inter-patient similarities. The main clinical goal is to provide the clinicians with the adequate information for the near-term prognosis and for future projections of those physiological time series and clinical conditions. This way, clinicians can better assess the near-term impact of their decisions and of potential events that may affect the patient. From the technical perspective, the main purpose is the research on advanced methodologies of time series processing for the development of prediction strategies. Thus, besides similarity measurements for identifying patients who exhibit similar trends in their physiological time series data (presented in the last chapter), methodologies able to project patient data into the future are the key research topic of the current chapter.

A significant amount of research has been carried out during the last decades addressing the prediction of time series in general, and of biosignals, in particular. Among them, linear regression methods, such as autoregressive and autoregressive moving average structures have been the most used in practice (Makridakis *et al.,* 1998). The well-established theory of linear models and the availability of effective algorithms for the estimation of corresponding parameters have justified their success. However, linear models are usually inadequate for clinical time series, since, in practice, almost all are non-linear to some extent. Therefore, the theoretical development of non-linear methods, together with the evolution of computer processing speed and data storage, has motivated the introduction of non-linear schemes for prediction tasks. Of the non-linear methods, neural networks became very popular mainly due to their universal approximation properties. Many different types of neural networks, such as time delay and recurrent neural networks, have been proven to be effective for time series modelling in general, and for biosignals forecasting, in particular.

On the other hand, an important prerequisite for the successful application of the non-linear techniques, such as the neural networks, is the stationarity of the involved data (Fryzlewicz *et al.*, 2003). In most clinical cases, an assumption of global stationarity can not be considered. In effect, usually a physiological time series is a combination of several separated sources, such as long-term trends and short-term cyclic occurrences. Thus, the development of strategies that overcome the aforementioned problems assumes a central importance. Of particular interest for this work are the time-frequency analysis methods, which can produce a good local representation of the

signal in both the time and frequency domains, offering an appropriate framework to deal with the non-stationarities that characterize biosignal time series.

Among time-frequency methods, wavelet transform was recently introduced mainly for addressing pattern detection and modelling tasks in time series. Although the wavelet transform itself is not a forecasting methodology, it may be incorporated in hybrid prediction schemes involving the multi-resolution decomposition of signals. In this context, some schemes combining wavelet transformation with forecasting methods have been proposed.

In the present work two alternatives are proposed for the prediction of biosignals. The first employs neural networks structures, previously trained using the similar patterns that resulted from the similarity analysis process, in a multi-model prediction scheme to derive accurate values for the future values. The second, which does not involve the explicit development of a model, employs the wavelet decomposition of those similar patterns to capture the main trends of the biosignals, enabling to provide a rough estimation of the time series future values. The structure of this chapter is as follows: in section 2, the clinical relevance of predictive approaches is justified. Section 3 addresses the theoretical background of predictive methodologies, namely system modelling and time series prediction topics. Taking into account the goals of the present work, it also introduces neural network- and wavelet transform-based prediction methods. Section 4 is devoted to the proposed methodologies. Basically, the two strategies are introduced: the neural-network multi-model scheme and the wavelet multi-decomposition scheme. Finally, in section 5, some conclusions are drawn.

3.2 Clinical Relevance

Hippocrates, the father of medicine, considered prognosis as a major concept of medicine. In general, prognosis means foreseeing, predicting, or estimating the probability or risk of future conditions, such as in weather and economic contexts. Particularly in medicine, prognosis commonly relates to the probability or risk of an individual to develop a particular state of health over a specific time.

In recent years there has been an increasing research in the area of clinical prediction, which is justified by the benefits that this can bring both to patients and physicians. Two types of research can be distinguished: on diagnostic prediction and on prognostic prediction. The first investigates the ability of variables in predicting the presence or absence of a specific condition. In this category, it can be referred the work developed under this thesis concerning the diagnosis of ischemia based on the ST segment deviation, and on changes in morphology of the QRS complex and T wave (which will be presented in chapter 4). The second investigates the ability of variables in predicting future outcomes, such as the risk of developing an hypertension condition (also developed under this thesis and presented in the next chapter).

The following examples aim to demonstrate the importance of the prediction methodologies in the medical field. Given the particular interest of the cardiology area to the present work, the selected cases relate to heart diseases.

For instance, a heart attack, a serious medical emergency in which the supply of blood to the heart is suddenly blocked, can seriously damage the heart muscle. The main problem is that the misunderstanding of the symptoms and early warning signs can result in a long recovery, or even death. In fact, often minor and very treatable heart attack progresses to a critical stage resulting in death due to apathy and negligence. A simple treatment of taking an aspirin and a low dose of nitroglycerin tablet can open arteries and save a life, but must be timely administered. Consequently, it is easily understood that the development of automatic methodologies to recognize early signs and predict the occurrence of an heart attack, is of the major importance to increase the survival rate. Some works, such as the one of Patil and Kumaraswamy (2009), addressed this problem, proposing a methodology for the extraction of significant patterns (blood pressure range, cholesterol range or heart rate greater than a given threshold, abnormal ECG, etc.) from the heart disease data warehouses, and using them in the prediction of heart attack through an intelligent approach.

Another problem of concern is the occurrence of acute hypotensive episodes (AHE) in intensive care units, which endanger the lives of patients that come to depend on the prompt intervention of clinicians. However, procedures used to treat such events are usually invasive and aggressive. Therefore, a prediction system able to identify an

imminent event brings a significant benefit to timely support non-invasive and preventive treatments. In recent years, the development of medical monitoring technology and of signal analysis and processing methods, led to the emergence of numerous approaches for the automatic forecasting of AHE, based on the trend analysis of vital signs that are probably related to the occurrence of such episodes (arterial blood pressure, heart rate, and oxygen saturation). Mention may be made to the works of Wang *et al.* (2010) and Rocha *et al.* (2011).

The prediction and prevention of sudden cardiac arrest is another great challenge of the contemporary cardiology. Particularly, thousands of children experience cardiac arrest (CA) events every year in paediatric intensive care units and most of these children die. Moreover, the risk of a child to suffer a CA in this environment, is 10 times higher than for people in standard hospital beds. Once more, the implementation of methods with the potential to bring arrest prediction to the paediatric intensive care environment, possibly allowing for intervention that can save lives and prevent disabilities, is of vital importance. In that sense, Kennedy and Turley (2011) proposed a six step method for creating clinical prediction models using time series data, which is able to characterize deterioration that often precedes cardiac arrests.

Ventricular arrhythmias, which are abnormal heart rhythms originated in the lower chambers of the heart, represent another serious problem of the cardiology domain. They include ventricular tachycardia (VT) and ventricular fibrillation (VF) that are both life threatening. In fact, ventricular fibrillation is one of the main causes of sudden cardiac death in the western world. It causes the heart to beat chaotically, making it unable to pump blood. VF is usually preceded by VT, which is another type of arrhythmia that also constitutes a medical emergency. It is vital for the patient to receive immediate medical intervention when either VF or VT occurs. Therefore, the development of methods able to predict their occurrence, even a few seconds in advance, can potentially save lives. In this context, Rocha *et al.* (2008) developed a methodology to detect the occurrence of VT and VF and also to characterize their evolution.

3.3 Background

Included in the area of dynamic systems, time series prediction is a research topic intrinsically linked to the systems modelling field (Box and Jenkins, 1976), (Pandit and Wu, 1983). As a consequence, the definition of the model structure, the determination of the respective parameters as well as the establishment of the appropriate validation procedures, should be taken into consideration. In this context, linear autoregressive (AR) and autoregressive moving average (ARMA) models have been applied with relative success to time series modelling, both in direct and indirect schemes (Gang *et al.*, 2008).

The linear autoregressive models are basically linear representations assuming that data are stationary, thus, with a limited ability to capture non-stationarities and non-linearities of data, so common in clinical contexts. Therefore, the research for alternative models, able to deal with non-linearities and non-stationarities has been of central importance in the forecasting framework. Examples of such methodologies are the extended non-linear autoregressive moving average model (NARMA) (Chen *et al.,* 1989), along with other approaches such as artificial neural networks (Zhang *et al.*, 1998), fuzzy systems (Vaidehi, 2008), and phase space reconstruction techniques (Camilleri, 2004). In particular, artificial neural networks (ANNs) have shown a great ability in modelling and forecasting non-linear and non-stationary time series, due to their innate non-linear properties and flexibility for modelling (Haykin, 2008).

Moreover, several types of transforms have also been applied for time series forecasting, such as principal component analysis (Hiden *et al.,* 1999), independent component analysis (Roberts *et al.,* 2004), Fourier transform (Schoukens and Pintelon, 1991) and wavelet transform-based methods (Yao *et al.,* 2000), (Chong, 2009) , (Cao *et al.*, 1995), (Soltani, 2002). Due to the notable approximation properties of wavelets (Vetterli *et al.*, 2000), they have been introduced to describe non-linear dynamical systems. Moreover, they allow for the representation of the signals in both time and frequency domains (Yevgeniy *et al.*, 2005), while preserving time information in the transformed variables. For that reason, they seem to be ideal for time series forecasting. In effect, several approaches have been introduced for time series prediction based on wavelet transform together with autoregressive models (Soltani *et al.,* 2000), Kalman filters and neural networks.

This subsection starts by reviewing the system modelling process applied to time series, including the respective steps involved (structure identification, parameter estimation, and validation). Then, the main methodologies for time series forecasting are presented. Finally, given their relevance to the present work, neural network-based and wavelet transform-based time series predictions are described.

3.3.1 System Modelling

Introduction

A model is a mathematical representation of a system, such as a physical, a mechanical, an economical, a biological or a physiological system. Essentially, it is an analogy used to help understand a system, and to make predictions about how this system will behave. System modelling concerns the precise description of the model, that is, it aims the developing and characterization of the mathematical model that captures the system's dynamic (the term *dynamic* indicates that the process presents a temporal dependent behaviour). A simplified input-output block diagram of a dynamic system is depicted in Figure 3.1.

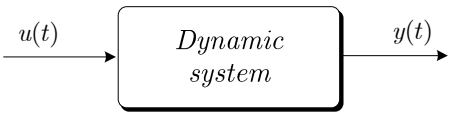


Figure 3.1 - Dynamic system definition.

The variable $u(t) \in \mathbb{R}^{n_u}$ and $y(t) \in \mathbb{R}^{n_y}$, represent, respectively, the input and the output of the system at a discrete time instant t, and the parameters n_u and n_y, define, respectively, the number of inputs and outputs of the system. Typically, the outputs are the variables of interest and, the inputs, the variables that can be manipulated, mainly to influence the outputs. The input and the output variables can be scalars or vectors: in the first case, $n_u = n_y = 1$, the system is designated as Single Input Single Output (SISO) system. In the case of several inputs and several outputs, the system is designated as Multiple Input Multiple Output (MIMO) or multivariable system.

When developing a model several purposes can be addressed, namely:

- *Prediction* - the main goal consists in predicting future values of some variables. Typical applications are, for example, the prediction of future outcomes, early detection of future adverse events (e.g. hypotensive episodes) or the prediction of medication effects;
- *Controller design* - the developed model provides a base for the design of a controller. Examples of such are model-based controllers of mechanical ventilation to automatically guide the ventilator during clinical anaesthesia (Michel and Struys, 2006);
- *Fault diagnosis* - can be viewed as the process of linking symptoms to causes, similar to medical diagnosis. Thus, the goal of a model in the context of fault diagnosis may be to match patterns of a particular sensor measurements to specific devices malfunctions (sensor errors), or to distinguish specific events (symptoms and clinical conditions) to generate specific alarms;

- *Estimation of non-measurable variables* - the development of a model can be useful when it is not possible to directly measure some variables of the system. In effect, the existence of a model can circumvent such difficulties by providing an estimation of these non-accessible variables.

Two main approaches are typically followed for model developing: *i*) analytic modelling and *ii*) experimental modelling (data-driven).

In the first case, the model is built founded on prior knowledge and physical insight of the system to be modelled. An example of such is the physiological model of the cardiac system, attempting to relate and quantify the effects of the major interacting parameters on the cardiac function, involving biology and engineering concepts, such as diffusion, active ionic transport, membrane channels and receptors, cellular metabolism, energy consumption, etc. In general, those models are described by a set of complex differential or difference equations, respectively for the continuous and discrete time cases. However, most of the times there is an insufficient knowledge (or even no knowledge) about the detailed interactions of the system and the only accessible information are the past measurements collected from that system. In this case, an experimental modelling may be carried out, known as black-box modelling (Ljung, 1999), in contrast to the analytical approach, also called white-box modelling.

On the other hand, there are circumstances where the nature of the process may difficult, or even prevent, a precise mathematical description. However, if domain expertise is available, other alternatives can be followed to generate a model. Assuming the pre-existence of expert knowledge based on clinical experience, observations, and deductions from clinical practice (such as guidelines obtained from clinical studies and statistical analysis, medical literature, etc.), deductive or knowledge-driven approaches can be applied. Typically, in this case, qualitative models can be formulated, where the simplest form is a rule-based structure that makes use of "*if-then-else*" rules to describe the system's behaviour. Although these rules are usually created by human experts, other methodologies can be applied to process data and generate them, such as genetic algorithms and rule induction techniques (Krajnak and Xue, 2006).

In this work, the modelling task is mainly oriented for prediction purposes, that is, for the forecasting of time series future values. Moreover, the models to be developed are based on historic data, namely biosignals regularly collected from patients. Thus, an experimental approach (data-driven) is assumed here. Without loss of generality, the description followed in the present document concerns SISO systems.

Considering the historic clinical data and using a similarity searching algorithm, a set of similar conditions are in a first step identified. Based on these similar conditions, an experimental model (data-based modelling) is built and employed, in a second phase, to forecast specific variables (time series). The estimated evolution of these variables are then used to predict future outcomes, such as the risk of developing an hypertension condition.

System identification

Formally, the system identification problem can be established based on five factors (Ljung, 1999), (Sodeström and Stoica, 1989): a system to be identified P, an experimental dataset Z, a particular structure M with parameters θ, an estimation method for computing those parameters, and a validation criterion J. This can be mathematically described by the equation (3.1).

$$y(t) = \Upsilon\Big(Z, M(\theta) \Big) \tag{3.1}$$

It is assumed that there exists a mapping, $\Upsilon(\cdot)$, able to describe the dynamics of the system. Thus, it is possible to establish a relationship between the current output $y(t)$, and the observed dataset Z, based on a given structure $M(\theta)$. The variable θ represents the parameter set that characterizes the particular structure $M(\cdot)$.

The validity of this approach was mathematically established by Takens (1981), who proved that, for a noise free system and for an infinite precision in the measurements of the system output, it is possible to exactly reconstruct the internal dynamics of that system by means of a regression formulation. In particular, the observed data dimension should verify $N > 2n+1$, being n the dimension of the state vector, and N the dimension of the observed data. Actually, the theorem of Takens only considers systems without external inputs, that is, it is applied to autonomous time series. The extension of that theorem to input-output systems was later presented by Casdagli (1991).

In practice, due to uncertainties, unknown external variations and the use of a finite dataset, the identification process is reduced to the estimation of the true function $\Upsilon(\cdot)$. As result, equation (3.1) is commonly replaced by (3.2).

$$\hat{y}(t) = f\Big(Z, M(\hat{\theta})\Big) \tag{3.2}$$

The variable $\hat{y}(t)$ represents the estimated value of the real output $y(t)$, and the function $f(\cdot)$ is an approximation of the true function $\Upsilon(\cdot)$. The variable $\hat{\theta}$ corresponds to the estimation of the true parameters θ that characterize the system.

Modelling phases

Assuming the existence of a dataset Z, the development of a model consists of the following main steps (Pintelon and Schoukens, 1991):

i) structural identification: which seeks to assign a particular model, M, from the set of possible candidates;

ii) parameter estimation: which, according to a certain criterion J, computes the values of the parameters θ;

iii) validation: which intends to test the correctness of the derived mathematical model, namely with new sets of data.

i. Structural identification (M)

This is certainly one of the most important phases in the identification procedure. There are several possible factors that characterize the candidate models to be used, namely: input-output or state space structures, the type and number of external inputs, and the functional relationships between the inputs and outputs, particularly, linear and non-linear mappings.

Linear models

The simplest structure that may be assumed for the true function, $\Upsilon(\cdot)$, is the one that linearly relates the current output with the previous observed values. Equation (3.3) describes a generic AutoRegressive with eXogenous input model (ARX) (Ljung, 1999).

$$\begin{aligned}\hat{y}(t) = b_1 u(t-1) + b_2 u(t-2) + \ldots + b_{nb} u(t-n_b) + \\ + a_1 y(t-1) + a_2 y(t-2) + \ldots + a_{na} y(t-n_a) + \zeta(t)\end{aligned} \tag{3.3}$$

The variable $\hat{y}(t)$ represents the estimated value for the actual output, $y(t)$, and the parameters a_i and b_i are weights associated, respectively, with the past outputs and the past inputs. The constants n_a and n_b are the orders of the respective polynomials, that is, the number of past inputs and past outputs considered as representative and included in the model. The variable $\zeta(t)$ represents an unknown disturbance described as a white noise, that is, a sequence of serially uncorrelated random variables with zero mean and finite variance.

Linear models have been widely applied to system modelling in various fields. The simplicity of the parameters' estimation procedure, associated with the existent strong theory, justify its success, Landau *et al.* (1997). Furthermore, it is possible to provide an interpretation of linear models and they can be fully characterized by their transfer function or impulse response.

Non-linear models

Although satisfactory results have been achieved in numerous practical applications, the modelling capacities of linear models are sometimes inadequate to accurately describe a system. In fact, since most of the real applications are inherently non-linear, a linear model is, in these situations, clearly inappropriate. A more general description is the Non-linear AutoRegressive with eXogenous input model (NARX), given by (3.4), which is an extension of the linear ARX model to the non-linear case (Leontaritis and Billings, 1985), (Chen and Billings, 1989).

$$\hat{y}(t) = f\Big(u(t-1),\ldots,u(t-n_b), y(t-1),\ldots,y(t-n_a) \Big) + \zeta(t) \tag{3.4}$$

The function $f(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^{ny}$ ($d = n_a + n_b$), is a non-linear function which purpose is to approximate the true one, $\Upsilon(\cdot)$.

Another factor that characterizes a given structure concerns the type of signals involved in it. The dynamic behaviour of a system may be mathematically described by a deterministic form or by a stochastic process, where there is an additional disturbance term representing the system noise. In equations (3.3) and (3.4), this disturbance term is the white noise, $\zeta(t)$. The introduction of noise means that the output will not be an exact function of the available information, but includes a random term. In this work, only the deterministic part is addressed. Consequently, it is assumed that the unknown contribution is of secondary importance with respect to the system output and, thus, an adequate prediction of $y(t)$ can be achieved using the past data.

General model

Covering linear and non-linear models, a general formulation for the input-output modelling problem is usually described as follows (3.5).

$$\hat{y}(t) = f\left(\vartheta(t-1), \hat{\theta}\right) \tag{3.5}$$

As a consequence, three partial components should be considered:

i) The regression vector, $\vartheta(t) \in \mathbb{R}^d$ ($d = n_a + n_b$), composed of past observations as described in (3.6).

$$\vartheta(t-1) = \left[\, u(t-1),\ldots u(t-n_b), y(t-1),\ldots y(t-n_a) \,\right] \tag{3.6}$$

ii) The non-linear function, $f(\cdot)$, mapping the regressor space into the output space, $f(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}^{ny}$.

iii) The parameters, $\hat{\theta} \in \mathbb{R}^{n\theta}$, that characterize the specific mapping, being n_θ the number of parameters.

In this context (general modelling), an unified overview of the mathematical structure of non-linear models was presented by Sjoberg *et al.* (1994) and (1995). They proposed a general black-box model to infer an approximation to the function $\Upsilon(\cdot)$, by means of a basis function parameterization, equation (3.7).

$$\hat{y}(t) = \Upsilon\left(\vartheta(t-1), \hat{\theta}\right) = \sum_j \hat{\theta}_j \; g_j(\vartheta(t-1)) \tag{3.7}$$

The non-linear function, $g_i(\cdot): \mathbb{R}^d \rightarrow \mathbb{R}$, is the basis function, where d is the dimension of the regression vector space. The variable $\hat{\theta}_j$ represents the j^{th} component of the vector $\hat{\theta}$, a set of weight factors. For example, using this representation, the linear model, equation (3.1), is a particular case of equation (3.7). In effect, the basis function is defined by the regression vector, $g_j(\vartheta(t)) = \vartheta(t)$, defined by (3.6), and the parameter vector in the linear model is given by $\hat{\theta} = \left[\; b_1,\ldots,b_{nb} \;\; a_1,\ldots,a_{na} \right]$.

According to this general formulation, (3.7), the modelling problem is essentially converted into the determination of the basis function $g_j(\vartheta(t))$. A possible approach is a mother basis function parameterization $\chi(\cdot)$, given by (3.8).

$$g_j(\vartheta(t)) = \chi\big(\vartheta(t), \beta_j, \gamma_j\big) \tag{3.8}$$

The variables β_j and γ_j denote parameters of different nature: usually β_j is related to a scale operation and γ_j to some position or translation. Following this common approach, Sjoberg *et al.* (1994) and (1995), were able to express some typical model structures: sigmoidal neural networks, radial basis networks, fuzzy models, wavelets, B-splines, among others.

State Space Models

An alternative to the input-output formulation previously described, is a state space scheme, as presented in equation (3.9).

$$\begin{aligned} x(t) &= f\Big(x(t-1), u(t-1) \Big) \\ \hat{y}(t) &= h(\, x(t)\,) \end{aligned} \tag{3.9}$$

The variable $x(t) \in \mathbb{R}^n$ denotes the state vector at time t, typically assumed to be inaccessible for measurement. The mappings $f(\cdot) : \mathbb{R}^{n+nu} \rightarrow \mathbb{R}^n$ and $h(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{ny}$ are non-linear functions. State space models can be used either as knowledge based models, if enough prior knowledge about the physics of the process is available (analytical models), as well as black-box models (experimental models, thus following a data-driven approach).

Some advantages were pointed to the use of non-linear state space models (Rivals and Personnaz, 1995) , (Rivals, 1995). Firstly, they can describe a larger class of dynamical systems than the input-output models; only the outputs, $y(t)$, have desired values, while the internal state variables, $x(t)$, are not imposed. This aspect makes state space models more flexible, that is, with the ability to describe more complex input-output behaviour. Secondly, state space models require lower order models and, consequently, smaller number of past observations (that is, smaller number of regressors), than the input-output models to achieve the same level of accuracy.

On the other hand, for the non-linear case, different state space representations might display the same input-output behaviour. Moreover, it is always possible to rewrite a non-linear input-output model in a state space form, while, in opposition, an input-output model globally equivalent to a given state space model might not exist (Leontaritis and Billings, 1985). Another of the advantages of the state space representation lies on the approach followed for MIMO systems that is similar to the one followed for SISO systems, which does not happen in the case of systems represented by input-output models. Finally, these models allow for the employment of well-established algorithms, like the Kalman filter methods.

ii. Parameter estimation

Once selected the model structure and the optimal dimension of the data observations, the set of parameters, θ, has to be found. This subject (parameter estimation) is commonly known as learning or training in the computational intelligence field, such as in the case of neural networks models. Following a supervised approach, a training dataset Z, consisting of N examples of past observations and corresponding target outputs $y(t)$, is required, as described in equation (3.10).

$$Z = \begin{bmatrix} \Phi(t) & Y(t) \end{bmatrix} \tag{3.10}$$

The variable $\Phi(t)$ is composed of N regression vectors

$$\Phi(t-1) = \begin{bmatrix} \vartheta(t-1) \dots \vartheta(t-N) \end{bmatrix}^T$$

and the variable $Y(t)$ is composed of the corresponding N observed outputs

$$Y(t-1) = \begin{bmatrix} y(t-1) \dots y(t-N) \end{bmatrix}^T .$$

A typical approach for determining the optimal values of the parameters involves the minimization of a quadratic cost function, defined as the summation of the N errors, equation (3.11), the so-called sum of squared errors (SSE) criterion.

$$J(\hat{\theta}) = \frac{1}{N} \sum_{\tau=t-N+1}^{t} \left[e^T(\tau) \; e(\tau) \right] \tag{3.11}$$

The term $e(t) \in \mathbb{R}^{ny}$ defines the identification error at time instant t, given by (3.12).

$$e(t) = y(t) - \hat{y}(t, \hat{\theta}) \tag{3.12}$$

The output $\hat{y}(t, \hat{\theta})$ represents the model output given the determined parameters $\hat{\theta}$. Regarding the minimization of the criterion, equation (3.11), no analytical solution is usually available. Therefore, the minimization has to be done by some numerical iterative procedure, such as the basic gradient descent method or other advanced methods as the Levenberg-Marquardt algorithm. Another successful alternatives are the nature-inspired methods, which is the case of genetic algorithms.

iii. Model Validation

The parameter estimation step intends to minimize the modelling error criterion based on the dataset. However, from an application point of view, the modelling error itself is not enough for assessing the quality of the derived model. A more relevant measure of quality for a trained model is its generalization ability, that is, the capacity to provide good predictions on examples that were not used in the training set (Ljung, 1999).

Furthermore, another measure of interest is the model complexity. If the chosen model is too simplistic, it will not be flexible enough to describe the dynamics of the system, leading to the so-called underfitting situation. On the other hand, if the chosen model is too complex, the excessive degrees of freedom will allow the model to fit not only the original data but also additional noise. This situation is known as overfitting. Both cases will lead to a large generalization error, (Haykin, 2008) and (Bishop, 1995). For this problem, Akaike (1974) suggested a compromise between the best approach and the smallest possible number of parameters, proposing the well-known Akaike information criterion (AIC), which measures the relative goodness of fitting of a particular model.

3.3.2 Time Series Prediction

System modelling addresses the development of a model able to estimate the current output values $\hat{y}(t)$, based on previous information. In the specific case of time series modelling, only a regression of the variable $y(t)$, itself, is typically involved and no exogenous input $u(t)$ is considered. Therefore, in this situation, equations (3.5) and (3.6) are simplified, originating equations (3.13) and (3.14), respectively.

$$\hat{y}(t) = f\left(\vartheta(t-1), \hat{\theta}\right) \tag{3.13}$$

$$\vartheta(t-1) = \left[\, y(t-1), \ldots y(t-n_a) \,\right] \tag{3.14}$$

Time series prediction is unquestionably related with the modelling procedure. In effect, the model obtained in the modelling phase is usually employed in the prediction of future values $\hat{y}(t+p)$, this time using past and current information, as illustrated in Figure 3.2.

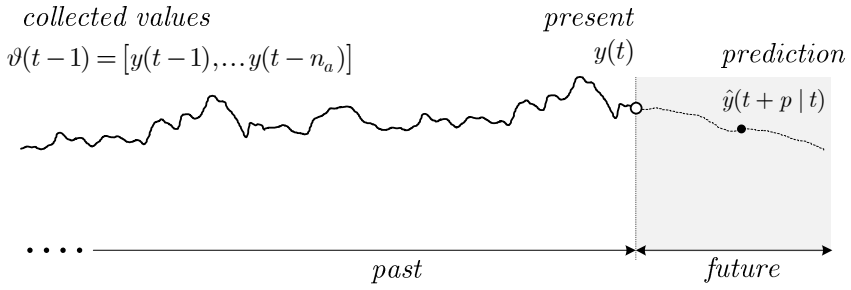


Figure 3.2 - Prediction of future instants.

According to the value of p, two different cases are considered in the prediction of the value $y(t+p)$: one-step ahead and multi-step ahead predictions.

i) *One-step ahead prediction* ($p=1$): this is the most common case and, in general, is referred to as short-term prediction. Basically, the model derived in the modelling phase (such as a generic non-linear model (3.4)) is employed, using the past observations, $\vartheta(t-1)$, and the current information, $y(t)$, to predict the next output $\hat{y}(t+1\,|\,t)$, according to (3.15).

$$\begin{aligned} \hat{y}(t+1\,|\,t) &= f(\ \vartheta(t), \theta\ ) \\ \vartheta(t) &= \Big[\ y(t), y(t-1) \ldots y(t-(n_a-1))\ \Big] \end{aligned} \tag{3.15}$$

ii) *Multi-step ahead prediction* ($p>1$): also called long-term prediction is, unlike the short-term approach, typically faced with several uncertainties originated from various sources. For instance, the accumulation of prediction errors by using a single model, makes the multi-step prediction a challenging problem. Two approaches can be implemented: the direct and the iterative.

- *Iterative or recursive approach*: for the prediction of a future instant $\hat{y}(t+p\,|\,t)$, a one-step ahead model is iteratively applied during p times, being the current predictions fed back to the model in order to obtain the next values;
- *Direct approach*: p different predictors are independently trained, each one corresponding to each p step-ahead value.

Recursive versus direct approaches

The recursive (iterative) strategy seems to be the most intuitive and simple. Only one model is involved (such as a generic non-linear model $f(\cdot)$ presented in last section, equation (3.4)) and, basically, the approach recursively uses the predicted values as inputs to predict the next ones. This situation is illustrated in Figure 3.3.

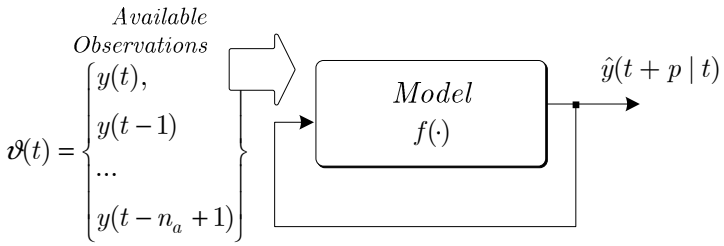


Figure 3.3 - Prediction using a recursive strategy.

In more detail, the same model $f(\cdot)$ used to one-step ahead prediction, equation (3.15), is used to predict the two-step ahead value, equation (3.16).

$$\hat{y}(t+2\,|\,t) = f\Big(\hat{y}(t+1\,|\,t), y(t), \ldots y(t-(n_a-2))\Big) \tag{3.16}$$

In the last equation, the predicted value $\hat{y}(t+1\,|\,t)$ is used instead of the true value $y(t+1)$, which is unknown. Using this scheme, it is straightforward to extend this formulation to multi-step ahead predictions in an iterative way, as described in (3.17).

$$\hat{y}(t+p\,|\,t) = f\Big(\,\hat{y}(t+p-1\,|\,t),\hat{y}(t+p-2\,|\,t),\ldots,y(t),y(t-1),\ldots\Big) \tag{3.17}$$

Although simple, the use of the predicted values as inputs typically deteriorates the accuracy of the predictions. Using this representation, the value $\hat{y}(t+1\,|\,t)$ is an input of the $f(\cdot)$ mapping, which is used for the prediction of the value $\hat{y}(t+2\,|\,t)$. Since the value $\hat{y}(t+1\,|\,t)$ is an estimation of the actual output $y(t+1)$, prediction errors are propagated and long-term forecasts can not be usually accurately performed by means of autoregressive models. In effect, the error for the p^{th} step prediction is the accumulation of the errors of the previous $(p-1)^{th}$ steps. Thus, in general, the longer the forecasting horizon is, the larger the accumulated errors are, and the less accurate the iterative method is.

Another alternative for the long-term prediction is the direct approach. According to this, the estimation of a future value $\hat{y}(t+p\,|\,t)$ is directly performed from the available observations, being an independent model employed for each sampling instant within the prediction horizon. Using this representation, the output at the next instant, $y(t+1)$, can be estimated using a specific mapping $f_1(\cdot)$, given by (3.18), similarly to the iterative scheme.

$$\hat{y}(t+1\,|\,t) = f_1\Big(\,y(t),y(t-1),\ldots,y(t-(n_a-1)\,\Big) \tag{3.18}$$

However, the prediction of the output at instant $t+2$, $\hat{y}(t+2\,|\,t)$, is achieved using the same formulation. Thus, an independent mapping $f_2(\cdot)$, has to be found, equation (3.19).

$$\hat{y}(t+2\,|\,t) = f_2\ \Big(\,y(t),y(t-1),\ldots,y(t-(n_a-2)\,\Big) \tag{3.19}$$

Generically, the prediction of the output at a specific future time instant p, can be expressed by the generic equation (3.20), by means of a particular mapping $f_p(\cdot)$.

$$\hat{y}(t+p\,|\,t) = f_p\Big(\,y(t),y(t-1),\ldots\,\Big) \tag{3.20}$$

As result, if a time series has to be predicted over a future horizon of length P, P distinct regression models have to be derived. This type of prediction, known as a multi-models approach, is illustrated in Figure 3.4. In this scheme, different types of models and different regression vectors can be employed to each individual model.

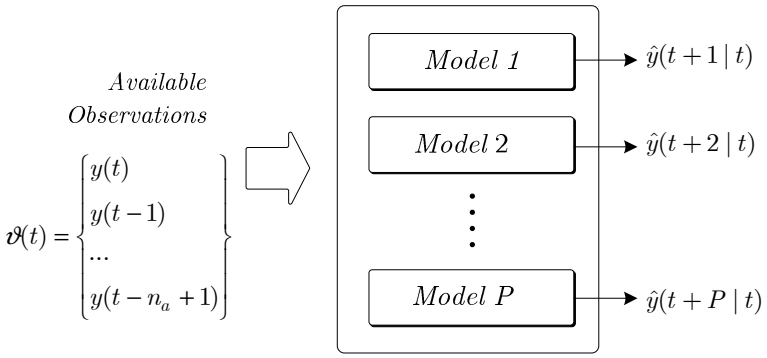


Figure 3.4 - The multi-models scheme for the prediction of P future instants.

Thanks to this structure, predictions over a forecast horizon do not depend on previous predictions, but only on information available at the current instant. As a consequence, the multi-models approach allows for more accurate estimations (Lawrynczuk, 2008). However, usually the direct strategy increases the complexity of the model. In effect, the analytical representation of the non-linear functions $f_p(\cdot)$ may be complex, even for a system described by a very simple model, which can be seen as one of the main drawbacks of this approach.

In conclusion, both the recursive and the direct approaches can be used to perform multi-step ahead predictions of time series. In theory, long-term predictions can be obtained from a short-term predictor by simply applying it multiple times (steps) in an iterative way. Using the direct approach, predictions can be directly obtained by the established models, in a way that is similar to one-step ahead predictions.

Although it is possible to compare the iterative and the direct approaches for a given specific problem, it is difficult to achieve generic theoretical conclusions regarding which one should be used. In fact, the decision about a particular scheme is, in general, an experimental decision (Marcellino *et al.*, 2005). These authors carried out a large-scale empirical study, comparing both iterative and direct predictions schemes using data of 170 U.S. macroeconomic time series variables, monthly available from 1959 to 2002. Rather than narrowing it on individual series, this study considered the larger question of whether the iterative or direct forecasts were more accurate, on average, for the population of the U.S. macroeconomic time series. Assuming that the time series dynamics is accurately captured by the model, the iterative prediction scheme is, in most cases, more efficient than the direct approach. Examples of such systems are the sunspot prediction problem. Weigend *et al.* (1992) showed, in this context, that the direct predictions produced by a neural network model were significantly worse than the iterated predictions. However, on the other hand, direct schemes are, theoretically, more robust when dealing with model uncertainties and misspecifications (Marcellino *et al.*, 2005). For specific types of models, the iterative approach may be less accurate than the direct one, if the process model is not correctly specified (Ing, 2003).

Nevertheless, the main problem of the iterative prediction is that errors accumulate at each step. Even in face of small prediction errors at the beginning of the horizon, they accumulate and propagate, often resulting in large overall prediction errors that are difficult to estimate. The situation is even worse for non-linear complex systems that are difficult to correctly characterize. Thus, some authors pointed that the direct approach could give a better prediction than the iterative method for non-linear time series, Atiya *et al.* (1999).

Time series forecasting methodologies

Time series forecasting has been a very active area of research and, as a result, numerous methodologies have been introduced, which can be generally divided into the following groups:

- *Simple or analytical approaches*: are easy to implement and usually provide robust results. Examples are the moving average and the exponential smoothing approaches (Gardner, 2006), (Makridakis and Hibon, 2000).
- *Regressive approaches*: are based on regressive models that employ past values of the inputs and outputs (Box and Jenkins, 1976). They are typically associated with traditional or classic methodologies, such as the linear systems theory, thus, based on a strong mathematical support.
- *Advanced methods*: include recent transforms and computational intelligence methods. They provide a base for the conception, design and application of methodologies, suitable to deal with uncertainties and imprecisions. They consist in the association of a set of computational techniques based on fuzzy logic, neural networks, support vector machines and genetic algorithms. When applied to modelling problems, they show an high potential and their success is unquestionable for the solution of practical problems where traditional methods present difficulties, especially in the presence of complexities, non-linearities and uncertainties.

Although the above methodologies can be individually employed, they can also be combined in hybrid schemes that integrate different methods in the same structure, in order to profit the potential of each one. In fact, the combination of different forecast methodologies can improve the global forecast accuracy by exploiting their different strengths, while also compensating their limitations.

1. Simple or analytical approaches

This type of models aims to represent the variable to be predicted, $\hat{y}(t+1)$ (one-step ahead), as a time function of the observations at the past and current instants. Examples are the random walk, moving average, single exponential smoothing, and polynomial regression approaches.

Random walk: this algorithm simply predicts the next value of the time series by considering it equal to the current value, thus by (3.21).

$$\hat{y}(t+1) = y(t) \tag{3.21}$$

Moving average: in this case, the predicted value is simply determined by the arithmetic mean of the last M observations, equation (3.22). Although basic, this method provides acceptable results in the situation of very simple systems.

$$\hat{y}(t+1) = \frac{1}{M} \sum_{k=t-M+1}^{t} y(k) \tag{3.22}$$

Single exponential smoothing: the predicted value is calculated (smoothed) by adjusting the previous forecast by the error it produced, according to (3.23). The parameter λ, a constant smoothing factor between 0 and 1, controls the degree of this adjustment, where a higher value of λ faster reduces older observations.

$$\hat{y}(t+1) = \lambda\ y(t) + (1-\lambda)\ \hat{y}(t) \tag{3.23}$$

Polynomial regression: this model fits a polynomial to the time series by regressing time series indices with time series values. The equation (3.24) shows the example of a polynomial regression of order three, where the α_i are the parameters of the regression model.

$$\hat{y}(t+1) = \alpha_0 + \alpha_1\ t + \alpha_2\ t^2 + \alpha_3\ t^3 \tag{3.24}$$

2. Regressive methods

As mentioned is the last section, regression is the study of relationships among variables, a principle that supports modelling and predictions tasks. By means of this approach, the future output value is estimated based on previous and current observations. Box and Jenkins (1976) presented a general formulation to describe autoregressive integrated moving average model (ARIMA) and its variations, such as autoregressive, moving average, and autoregressive moving average. In effect, a general tool for modelling and forecasting time series is achieved by an ARIMA model. This structure is described by the notation $ARIMA(n_a, d, n_c)$ and consists of the following three main terms:

i) *Autoregressive component*: $AR(n_a)$ - denotes the autoregressive part of the time series $y(t)$, equation (3.25), where the output depends on the time-lagged values of the signal itself, weighted by parameters a_i.

$$\hat{y}(t+1) = a_1 y(t) + \ldots + a_{na} y(t - n_a + 1) \tag{3.25}$$

ii) *Moving average component:* $MA(n_c)$ - consists in the moving average of the error series $e(t)$, equation (3.26). It can be described as a regression using the past error values of the series, weighted by parameters c_i.

$$\hat{y}(t+1) = c_1 e(t) + \ldots + c_{nc} e(t - n_c + 1) \tag{3.26}$$

The error term, $e(t)$, corresponds to unknown factors of the series at time instant t that can not be explained by the past values.

iii) *Degree of differentiation*: $I(d)$ - defines the degree of differencing involved in the model. Differencing can be used to cope with non-stationarity in a time series, by calculating the change between each observation and its predecessor. The first difference of a time series, $dy(t)$, is given by (3.27).

$$dy(t) = y(t) - y(t-1) \tag{3.27}$$

As proposed by the Box-Jenkins formalism, the build-up of a general ARIMA model requires a series of well-defined steps.

The first step corresponds to the identification of the model. It involves defining the structure (AR, MA or ARMA) and the order of the model (parameters d, n_a and n_c). To this aim, autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis are typically employed (Tran and Reed, 2004). An alternative for identifying ARMA models is to use of a goodness-of-fit statistic. In this approach, a set of candidate models are fit, and goodness-of-fit statistics are computed in order to appropriately penalize excessive complexity. Akaike's Final Prediction Error (FPE) and Information Criterion (AIC) are two common statistical measures of goodness-of-fit of an $ARMA(n_a, n_c)$ model. Usually the fitting process is guided by the principle of parsimony by which the best model is the simplest possible model - the model with the fewest parameters - that adequately describes the data. The second step is to estimate the coefficients of the model. Among several approaches, these can be estimated using the maximum likelihood or the least-squares regression methods. The third step consists in checking the model. This step, also called verification, aims to ensure that the residuals of the model are random (uncorrelated in time), and that the estimated parameters are statistically significant. Typically this phase involves a residual analysis, checking the significance of residual autocorrelations. This way, approximate confidence intervals can be estimated. A different approach for evaluating the randomness of the ARMA residuals is the Portmanteau statistic, or Q statistics, an objective diagnostic measure of white noise in a time series.

Although linear regression models are useful, general non-linear regressive models can be also formulated. Engle (1982) introduced a non-linear autoregressive conditional heteroscedasticity model, or ARCH(p) model, for modelling the changing volatility. The non-linear term is the variance of the disturbance. Latter, Bollerslev (1986) presented an extension of the ARCH model, the GARCH or generalized ARCH model. In turn, Chen and Billings (1989) proposed a general non-linear autoregressive moving average (NARMA) model, based on which, prediction can be formulated as (3.28).

$$\hat{y}(t+1) = f\Big(\ y(t),\ldots,y(t-n_a+1),\ e(t),\ldots,e(t-n_c+1)\ \Big) \tag{3.28}$$

where $f(\cdot)$ is some non-linear function, similarly to model (3.4). In some cases, the function $f(\cdot)$ does not depend on the lagged noise signals and, thus, the NARMA model is reduced to the NAR mode, equation (3.29).

$$\hat{y}(t+1) = f\Big(\ y(t),,\ldots y(t-(n_a+1))\ \Big) \tag{3.29}$$

3. Advanced methods

Recent advances in signal processing and in non-linear time series analysis have greatly improved the power of forecasting techniques. In fact, many types of functions and approaches based on non-parametric regression have been investigated to approximate the unknown function, which include polynomials, neural networks, radial basis functions, support vector machines, neuro-fuzzy basis functions, and wavelet transform. These approaches demonstrated superior performance in many areas of application, Fan and Yao (2003).

Given their importance in the present work, neural network and wavelet transform based models will be described in the following sections.

Neural networks

Neural networks are one of the methodologies belonging to the computational intelligence area that most frequently and successfully have been used for forecasting purposes. The universal approximation and generalization properties, as well as the ability to adjust online their parameters, allow neural networks to meet two of the main challenges for which conventional modelling techniques present serious limitations (Principe *et al.,* 2000) (Haykin, 2008): *i*) generality and precision in modelling problems and *ii*) adaptation capabilities to time-varying dynamics.

Wavelet transform

In reality, wavelet transform itself is not a true forecasting technique. In its simplest form, it transforms a signal into different levels of resolution, enabling time and frequency localizations. By means of this property, it is possible to capture low and high frequency features of a signal at successive decomposition levels. By increasing the level of decomposition, it is possible to generate smooth signal components that can be easily predictable. From the combination of individual predictions, independently performed at each scale, it is possible to reconstruct the prediction of the original signal through the wavelet inverse transform.

3.3.3 Neural Network-based Time Series Prediction

Original artificial neural networks basically represented a static mapping, just relating an input with the corresponding output. As a result, these structures were mainly oriented to classification problems, and not adequate to address dynamic modelling purposes. To achieve this goal, two main approaches were introduced in order to incorporate time (dynamics) into the network structure: delayed information and recurrent networks. Time-Delay neural networks (TDNN) that introduce time delays in the connections of feedforward networks, were successfully applied in non-linear system identification (Lin *et al.*, 1995), (Yazdizadeh and Khorasani, 2000), (Yazdizadeh and Khorasani, 2002). Dynamic or recurrent neural networks (RNN) were first introduced by Hopfield (1982) and were developed in other works (Gupta *et al.*, 2003), (Becerikli *et al.*, 2003), (Bambang, 2007). Due to their intrinsic ability to incorporate time (involving dynamic elements and internal feedback), RNN structures have advantages with respect to static neural networks for modelling dynamic processes. Moreover, RNN are in a standard form and present a low order compact structure, making them ideal candidates to be incorporated in prediction schemes (Gil *et al.*, 2002).

Multi-layer neural networks as universal approximators

Multi-layer neural networks (MLNN), also known as feedforward neural networks (FFNN), result from perceptrons joined together in multiple layers. Figure 3.5 depicts a multi-layer neural network with four layers, where the second and the third are known as hidden or inner layers. The first one is the input layer and the last is the output layer. The activation functions in the hidden layers are usually sigmoidal, while a linear activation function is used in the output layer.

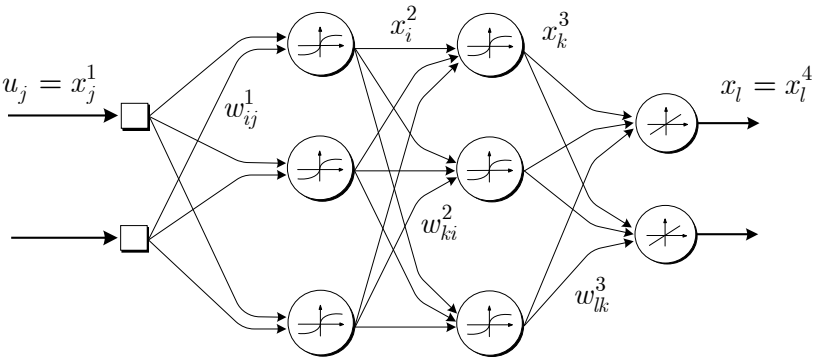


Figure 3.5 - Multi-layer neural network with 2 hidden layers.

For a multi-layer structure with L layers, the input is the first one $(l=1)$, the output is the last one $(l=L)$, being the remaining $L-2$, the hidden layers. Each layer l has N_l neurons, and x_i^l is the output of the neuron i at the layer l.

Thus, following these definitions, for the layers $l = 2,\ldots,L$ it is possible to write the equation (3.30).

$$\begin{aligned} a_i^l &= \sum_{j=1}^{N_l-1} w_{ij}^{l-1}\, x_j^{l-1} + b_i^{l-1} \\ x_i^l &= \sigma(a_i^l) \end{aligned} \tag{3.30}$$

In the previous equation, w_{ij}^{l-1} represents the weight of the input x_i^{l-1}, b_i^{l-1} denotes a bias, and σ is the activation function.

Considering a three layer neural network, Cybenko (1989), Funahashi (1989), Hornik *et al.* (1990), and Hornik (1991), showed that, if the hidden layer has an enough number of neurons (with sigmoidal activation functions) and if a suitable training algorithm is used, it would be possible to approach, with an arbitrary degree of accuracy, a continuous non-linear function. However, other structures such as trigonometric series, polynomial expansions, and splines, exhibit the same property. Therefore, the question is why did neural networks attract so much interest.

One of the main reasons concerns the curse of dimensionality. Barron (1993) partially showed that the parameterization using neural networks is advantageous with regard to other expansions, especially for high dimension spaces.

Concretely, for a particular class of functions, $f(\cdot)\colon \mathbb{R}^d \rightarrow \mathbb{R}$, being d the dimension of the input space, Barron showed that the approximation error provided by a three layer MLNN would be possible to characterize by (3.31).

$$\left\| f(x) - NN(x) \right\| < \varsigma \left(\frac{C_f}{M^{1/2}} \right) \tag{3.31}$$

In the previous expression, $NN(x)$ represents the output of the neural model given the input x, and $\|\cdot\|$ defines a quadratic norm. The parameter ς is a positive constant, C_f is the first momentum of the Fourier Transform of function $f(\cdot)$ (a finite value) and M is the number of hidden neurons. On the other hand, he showed that there is no linear combination of M functions, including polynomials, trigonometric series or splines, such that their output $\hat{f}(\cdot)$ can provide an approximation error lower than the value given by (3.32).

$$\left\| f(x) - \hat{f}(x) \right\| < \varsigma \left(\frac{C_f}{M^{1/d}} \right) \tag{3.32}$$

Consequently, in the last case, as the dimension of the input space increases, the advantages of neural models become clear.

Modelling dynamic systems with neural networks

The previous configuration represents a static mapping, particularly adequate for classification problems. In order to incorporate dynamics into the structure basically two main approaches can be followed: *i*) time-delay neural networks (TDNN), and recurrent neural networks (RNN).

Time delay neural networks

In the first approach (TDNN), the temporal dependencies are obtained by modifying the static neural structure through the introduction of delays (considering past observations by means of the operator q^{-1}), as given by equation (3.33) and illustrated in Figure 3.6. The variable $y_n(t)$ denotes the output of the neural network structure $NN(\cdot)$.

$$y_n(t) = NN\Big(u(t-1),\ldots,u(t-n_b),y_n(t-1),\ldots,y_n(t-n_a) \Big) \tag{3.33}$$

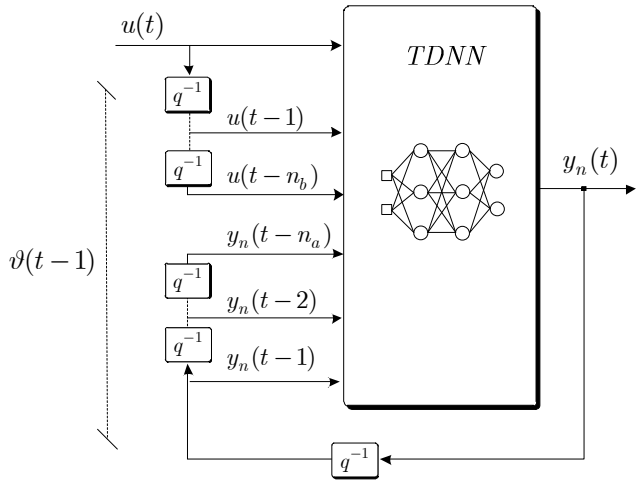


Figure 3.6 - Time-delay neural network.

According to the particular neural network structure $NN(\cdot)$, distinct models can be obtained, such as sigmoidal neural networks, radial basis networks and support vector regressions.

Recurrent neural networks

In case of recurrent neural networks, the dynamics is naturally introduced in the neurons, thus enabling to create an internal state of the network that allows it to exhibit dynamic temporal behaviour. Due to their inherent ability to incorporate time, recurrent neural networks are particularly suitable for modelling non-linear dynamic systems. Moreover, RNN are in a standard form and present a low order compact structure, making them ideal candidates to be incorporated into prediction schemes (Gil *et al.,*

2002). A typical recurrent neural network is described by equation (3.34) and shown in Figure 3.7.

$$\begin{aligned} x_n(t) &= \sigma\left(A\ x_n(t-1) + B\ u(t-1)\right) \\ y_n(t) &= C\ x_n(t) \end{aligned} \tag{3.34}$$

The vector $x_n(t) \in \mathbb{R}^n$ is the output of the hidden neural layer, also known as the network hyper-state, $y_n(t) \in \mathbb{R}^{n_y}$ is the network output, $u(t) \in \Re^{n_u}$ represents the input, and $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n_u}$, and $C \in \mathbb{R}^{n_y \times n}$, are interconnection matrices. The activation function is defined by the function $\sigma(\cdot)$.

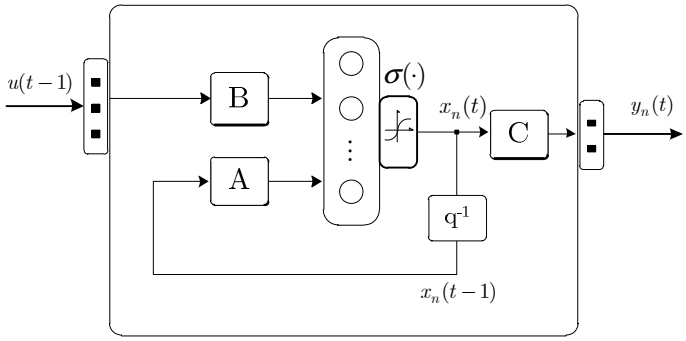


Figure 3.7 - Recurrent neural network.

Support vector regression

A particular structure of a time delay neural network is the support vector regression (SVR). The SVR consists in an extension of the Support Vector Machines (SVM) usually applied to classification problems, for the regression case. Both SVR and SVM are based on the statistical learning theory developed by Vapnik (1998). The usage of a non-linear kernel transformation to map the input samples into a higher dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive feature of these methods.

Assuming the following formulation for the prediction problem,

$$\hat{y}(t+1) = SVR\Big(\ y(t), \ldots, y(t-n_a+1)\ \Big) = f\left(D(t)\right) \tag{3.35}$$

the regression consists in estimating a function $f(\cdot)$ such that $\hat{y}(t+1) = f\left(D(t)\right)$, where $\hat{y}(t+1)$ is the predicted value, and $D(t) \in \mathbb{R}^{1,n_a}$ is a vector composed of current and past values used in the prediction. Considering a given training dataset $\left\{(D_1, y_1), (D_2, y_2), \ldots, (D_n, y_n)\right\}$ with n input/output data pairs (D_i, y_i), two different regression scenarios can be considered: linear and non-linear.

Linear regression

In the case of linear regression, $f(\cdot)$ is a linear function described by

$$f(D_i) = \langle w, D_i \rangle + b \tag{3.36}$$

where $w \in \mathbb{R}^{1,n_a}$, $\langle .,. \rangle$ denotes the dot product operator, and b represents a bias.

Firstly, the desired $f(\cdot)$ function should be flat, which implies that w should be small. Thus, flatness is ensured by the minimization of $\|w\|^2 = \langle w, w \rangle$. In addition, $f(\cdot)$ must approximate all desired predictions y_i within a precision of a given value ε. However, to account for prediction errors higher than ε, a soft margin is considered by introducing slack variables ξ, ξ^* into the problem. Then, the optimization problem to solve is formulated as (3.37) and (3.38).

$$min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3.37}$$

subject to

$$\begin{aligned} & y_i - \langle w, D_i \rangle - b \leq \varepsilon + \xi_i \\ & \langle w, D_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{3.38}$$

The constant $C > 0$ represents the trade-off between the flatness of $f(\cdot)$ and the quantity of tolerated errors higher than ε. This corresponds to dealing with the ε-insensitive loss function expressed by (3.39).

$$|\xi|_\varepsilon = max\left\{0, |\xi| - \varepsilon\right\} \tag{3.39}$$

Formulations (3.37) and (3.38) correspond to the classical model of the linear SVR. However, a computationally more efficient alternative consists in the dual formulation using Lagrange multipliers (Smola and Schölkopf, 2004).

The dual problem can be described as

$$max \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle D_i, D_j \rangle \\ -\varepsilon\frac{1}{2}\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{3.40}$$

subject to

$$\begin{aligned} & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C \quad i = 1,..,n \end{aligned} \tag{3.41}$$

The result is the optimal value for w given by

$$w^* = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\, D_i \tag{3.42}$$

and the corresponding regression function can be described by

$$f(D) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\, \langle D_i, D \rangle + b \tag{3.43}$$

Non-linear regression

In the case of non-linear regression, the main idea is to map the training vectors D_i into a feature space F and then perform the linear SVR algorithm.

Let $z = \phi(D)$, where $\phi(\cdot)$ is a mapping of appropriate dimensions. In the feature space F, the function $f(z)$ is linear, that is, $f(z) = \langle w, z \rangle + b$. As referred, the algorithm is based on the dot product between patterns D, hence the explicit knowledge of $\phi(\cdot)$ is not necessary. Rather, only the dot product $\langle \phi(D_i), \phi(D_j) \rangle$, represented by the kernel function $K(D_i, D_j) = \langle \phi(D_i), \phi(D_j) \rangle$, is required. This step corresponds to the kernel trick, and the corresponding optimization problem is

$$max \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(D_i, D_j) \\ -\varepsilon \frac{1}{2} \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i\, (\alpha_i - \alpha_i^*) \end{cases} \tag{3.44}$$

subject to

$$\begin{aligned} &\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \\ &0 \le \alpha_i, \alpha_i^* \le C \quad i = 1,..,n \end{aligned} \tag{3.45}$$

Therefore, the optimal value for w is

$$w^* = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\, \phi(D_i) \tag{3.46}$$

and the corresponding regression function can be described by

$$f(D) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\, K(D_i, D) + b \tag{3.47}$$

For the non-linear case several kernel functions have been proposed, having to satisfy the Mercer's condition (for more details please refer to Vapnik, (1998)). The choice of a kernel, which depends on the application, usually is made on a trial and error basis. A prior knowledge about the linearity of the data leads to a linear kernel or, in the

non-linear case, it helps to select the appropriate kernel. However, the usual strategy for kernel selection consists in the observation of the prediction performance on validation data. The most commonly used are the linear, the polynomial, the sigmoid and the radial basis function (RBF) or Gaussian.

Given its good performance in many different non-linear regressions and classification applications, the RBF is one of the most widespread kernels. In particular, $K(D_i, D_j) = exp(-\gamma \left\| D_i - D_j \right\|^2)$, where the scalar γ is known as the width of the Gaussian kernel.

3.3.4 Wavelet-based Time Series Prediction

Introduction

Research works, as well as practical results, have shown that wavelet transform is a powerful tool for time series analysis (Nason and Sachs, 1999). As mentioned in the last chapter, the discrete wavelet transform (DWT) is a mathematical tool that basically describes a time series in terms of a set of orthogonal basis functions and respective coefficients.

Initially, most research around wavelet transform focused on the detection of patterns and transient events (Priestley, 1996), (Morrettin, 1997), (Gao, 1997), (Percival and Walden, 2000), mainly due to its capacity to provide a local representation in both time and frequency domains. Other common application was in the support of compression methodologies, since the description of the signal can be highly concentrated in the most representative scales. In fact, it was shown that the DWT can effectively compress a wide range of signals, since a large proportion of the coefficients can be set to zero without appreciable loss of information. On the other hand, due to the notable approximation property of wavelets (Chui, 1992), (Daubechies 1992), (Meyer, 1987), (Mallat, 1989), and to their inherent capacity to deal with non-stationary data and to cope with heterogeneous and transient behaviour, wavelet-based approaches were also introduced with the purpose of modelling and forecasting non-linear dynamical systems.

Although the wavelet transform is not a true forecasting technique, when combined with common prediction methods, it can improve forecasting results. In effect, Bjorn (1995) used the discrete wavelet transform to decompose a signal into several scales and considered that each of the extracted scales could be viewed as a stationary time series, which could be individually modelled, analysed, and predicted. Ramsey (2002) stated that improvements could be achieved by decomposing the series to be forecasted into its timescale components, by means of the wavelet transform, and devising appropriate forecasting strategies for each of them. He justified the approach by the fact that the wavelet transform allows to separate time series data into trend, seasonal fluctuations and noise, thus enabling to isolate the local from the global. And, since prediction targeted the permanent components of the series rather than the strictly local events,

with the separation of these two effects it would be possible to improve forecasts by eliminating the contamination from strictly local non-recurrent events. He referred works using ARIMA models and neural networks (Aussem *et al.*, 1998) for the forecasting of individual components. For all the cases, the final forecast of the complete series could be obtained by adding up the individual forecasts.

Trends extraction

A general time series $X(t)$ can be seen as composed of some individual components, equation (3.48): a smooth changing function or the so-called trend $X_t(t)$, a function with a known period (such as daily, weekly), referred to as a seasonal component $X_s(t)$, and a random stochastic component $X_\xi(t)$, sometimes called the noise process.

$$X(t) = X_t(t) + X_s(t) + X_\xi(t) \tag{3.48}$$

Although there is no consistent accepted definition for trend, it is usually regarded as a non-random (deterministic) smooth function representing near-term/long-term movement or systematic variations in a time series. As an example, Craigmile and Percival (2002) define a trend as "*a tendency to increase (or decrease) steadily over time*" or to "*fluctuate in periodic manner*". Kendall (1973) states that the essential idea of trend is that "*it shall be smooth*", and Chatfield (2000) defines trend as "*a long-term change in the mean level*".

In the clinical field, a trend can be defined as a consistent, unidirectional change in the value of a variable (biosignal) (Varadharajan, 2004). Such concept strongly relates with near-term/long-term patient's status evolution, providing important insights for the patient prognostic. In effect, the analysis of the evolution of some health variables is an important way of monitoring a subject's physiological status, which may help clinicians to early assess potential critical events, as well as to project the impact of their decisions. As a result, the recognition of clinically significant trends in monitored signals plays an important role in clinical applications.

Currently there are available many methods for trend extraction, which differ in their complexity and interpretability, as well as in the algorithms and methodologies they use. Given their characteristics, the application of wavelets has been recognized as an adequate tool for time series analysis and, in particular, for trend extraction. In effect, using the wavelet transform, two types of coefficients are obtained, namely wavelet and scaling coefficients. The first are related to changes in averages over specific scales, whereas scaling coefficients are associated with averages on a specified scale. Since the scale that is associated with scaling coefficients is usually rather large, the information captured by these coefficients agrees well with the notion of trend. Thus, the key idea behind trend analysis with wavelets, is the association of the scaling coefficients with the trend.

Wavelet-based methodologies

In recent times, there has been an increased interest in developing hybrid schemes, merging wavelet decomposition techniques with simple regressive models, as well as with other advanced computational intelligence approaches, such as neural networks and neuro-fuzzy systems (Bodyanskiy and Vynokurova, 2012). In parallel, the explicit development of specific wavelet neural networks structures have also been exploited by other authors. In fact, wavelet approaches for non-linear dynamical systems prediction can be usually classified in two major categories: *i*) multi-resolution wavelet decomposition and *ii*) wavelet neural networks.

i. Multi-resolution wavelet decomposition

The wavelet transform provides an efficient multi-resolution decomposition of a time series into a local representation of the signal in both time and frequency domains, so that wavelet-transformed data improves the capability of a forecasting model by capturing useful information at various resolution levels.

The basic idea of the multi-scale decomposition is to decompose the data into trend and irregular components. Thus, the forecasting of stationary data and of non-stationary data can be separately achieved. In effect, by decomposing the non-stationary time series of non-linear systems into different components, it is possible to achieve a better separation of the general trend terms. The application of the most suitable prediction methods (for example, linear regressive structures or non-linear neural networks methods) to the components under different resolutions, has the potential to produce more accurate prediction results. Figure 3.8 illustrates this idea.

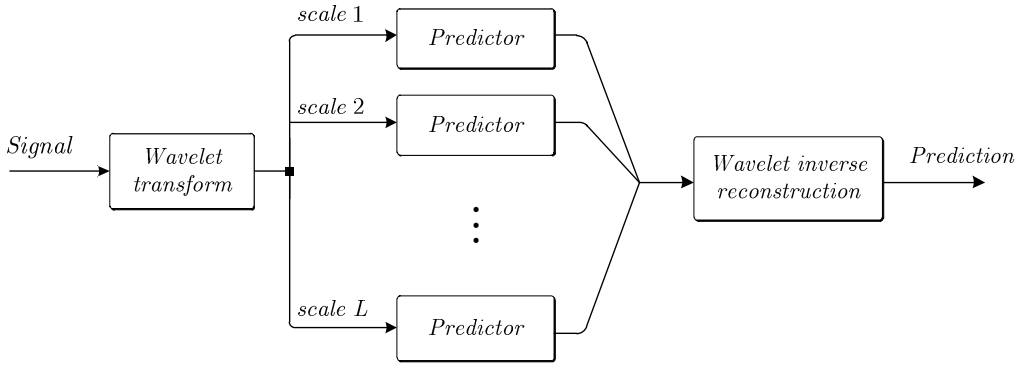


Figure 3.8 - Prediction using a multi-resolution decomposition scheme.

Concerning the predictors, several approaches have been proposed, which differ, among others, in:

- The representation they assume (input-output or state space methods);

- The specific structure (regressive, neural networks, fuzzy systems);
- How to achieve multi-step predictions (direct or iterative approaches);
- The way the parameters of the prediction model are computed (such as simple least squares methods, backpropagation methods and genetic algorithms).

One of the pioneering works was presented by Aussem *et al.* (1998). They proposed a strategy to improve neural network prediction accuracy, based on the idea of combining predictions at varying resolution levels. Specifically, a dynamical recurrent neural network was trained at each resolution scale with the temporal-recurrent backpropagation algorithm. They proposed the employment of the particular *à-trous* wavelet transform (redundant transform where the decimation is not carried out) and they applied the strategy to the prediction of sunspot series.

A similar hybrid model, combining wavelet analysis with artificial neural networks, was suggested by other authors (Gan *et al.*, 2005), (Loh, 2003). Basically, wavelets were used to decompose the original time series into a certain number of sub-time series that were separately predicted employing artificial neural network structures (typically time delayed neural networks). The short and long term prediction capabilities were validated in several time series examples, revealing that the suggested scheme could improve the forecast accuracy and prolong the prediction horizon, when compared with common neural networks based approaches.

Kaboudan (2005) followed a similar wavelet neural network strategy for the prediction of sunspot series. Original time series was first transformed using the Haar wavelet, and for the neural networks training, he proposed the use of genetic algorithms as an alternative to the back-propagation learning method.

Pan and Wang (1998) introduced a new wavelet-based estimator that combined state space models with wavelet transform, mainly exploiting wavelet-Kalman filter methods to deal with non-stationary processes. Basically, the coefficients of the wavelet estimator were formulated as a stochastic process, so that the Kalman filter could be applied in their estimation. Zheng *et al.* (2001) followed the same idea, proposing to assume the wavelet coefficients as the state variables of a state space model, possible to estimate by the recursive extended Kalman filter algorithm. In the same context, Zhao *et al.* (2006) proposed the use of unscented Kalman filters, a powerful non-linear estimation technique that was originated from the standard Kalman filter theory.

Yousefi *et al.* (2005) suggested the use of different predictors according to the characteristics of the signals obtained at each decomposition level. In effect, signal decomposed into different levels presents different behaviours. For a highest level (smooth signal) a spline fit was applied to extend the signal. For the lower detail levels (showing higher frequencies) trigonometric functions, such as a sine wave, seemed to be more appropriate for obtaining an extension of those parts of the signal.

Zhang *et al.* (2001) introduced a mechanism to select the optimal length of the time window applied for determining the number of regressions in a wavelet neural network

short-term forecast scheme. In a first phase, a multi-resolution decomposition was performed, using an *à-trous* algorithm to guarantee the shift-invariance of the wavelet transform. In a second phase, a Bayesian mechanism for automatic relevance determination, enabling the selection of relevant inputs for the separate multi-layer perceptron models, was carried out. In effect, when applying neural networks to time series forecasting, it is important to decide on the appropriate size of the time-window of inputs. This is similar to a regression problem, in which there are many possible input variables, some of which may be less relevant or even irrelevant to the prediction of the output variable. The automatic relevance determination provides a way for choosing the length of past windows to fed the neural networks. The proposed strategy was validated against simulation data, involving four different financial datasets.

Popoola (2007) presented a wavelet hybrid scheme, incorporating multi-scale wavelet decomposition into a set of fuzzy-rule bases. The approach employed a shift invariant wavelet transform, designated as maximal overlap discrete wavelet transform (Percival and Walden, 2000), and a fuzzy-rule base was created to separately predict each wavelet decomposition. To generate a global forecast, the prediction results of individual wavelet decompositions were directly combined using the linear reconstruction property of the wavelet multi-resolution analysis. The proposed approach was validated in the forecasting of non-stationary financial time series.

Similarly to the previously referred work, Rajaee *et al.* (2010) investigated the prediction of suspended sediment load by the conjunction of neuro-fuzzy with wavelet analysis. In the proposed model, observed time series of river discharge and suspended sediment load were decomposed at different scales by wavelet transform. Then, total effective time series of discharge and suspended sediment load were fed as inputs to the neuro-fuzzy model for the prediction of suspended sediment load, one day in advance. Results showed that the performance of the wavelet analysis and neuro-fuzzy model was better in the prediction rather than the neuro-fuzzy model itself.

Motivated by the successful applications of wavelet approaches in non-linear dynamical modelling, especially by the flexibility and adaptability of the wavelet NARMAX models, Wei and Billings (2006) introduced a class of multi-resolution wavelet models to derive a direct predictor for accurate multi-step predictions of non-linear time series. The results were later extended by the same authors, also in the context of long-term predictions, which presented a new direct predictor involving multi-resolution wavelet transform combined with regressive models. They showed that improved predictions could be obtained using the new approach based on three simulation examples.

Renaud *et al.* (2003), Renaud *et al.* (2005) and Benaouda *et al.* (2006), developed an hybrid scheme combining wavelet and regressive models. However, instead of using a separate prediction for each decomposition level, a different idea was explored. The prediction was based on a small number of wavelet coefficients (the most relevant), obtained from the decomposition at different scales. Assuming a time series $Y(t)$, collected from the past to the current time instant t, the detail coefficients, $d_{i,j}$, and the

approximation coefficients, $a_{i,j}$, to be used in the prediction of a particular value $y(t+1)$, are obtained by (3.49).

$$\begin{gathered} d_{j,t-2^j(k-1)} \qquad k=1,2,\ldots,A_j \qquad j=1,2,\ldots,L \\ a_{L,t-2^L(k-1)} \qquad k=1,2,\ldots,A_j \end{gathered} \tag{3.49}$$

The Figure 3.9 illustrates these values for a decomposition of level $L=4$, order $A_j=2$ and $N=2^L=16$. For the prediction of $\hat{y}(t+1)$, the wavelet coefficients at level $j=1$ are $(d_{1,t},\ d_{1,t-2})$; the wavelet coefficients at level $j=2$ are $(d_{2,t},\ d_{2,t-4})$; the wavelet coefficients at level $j=3$ are $(d_{3,t},\ d_{3,t-8})$; the wavelet coefficients at level $j=4$ are $(d_{4,t},\ d_{4,t-16})$. The coefficients corresponding to the approximation component are $(a_{4,t},\ a_{4,t-16})$.

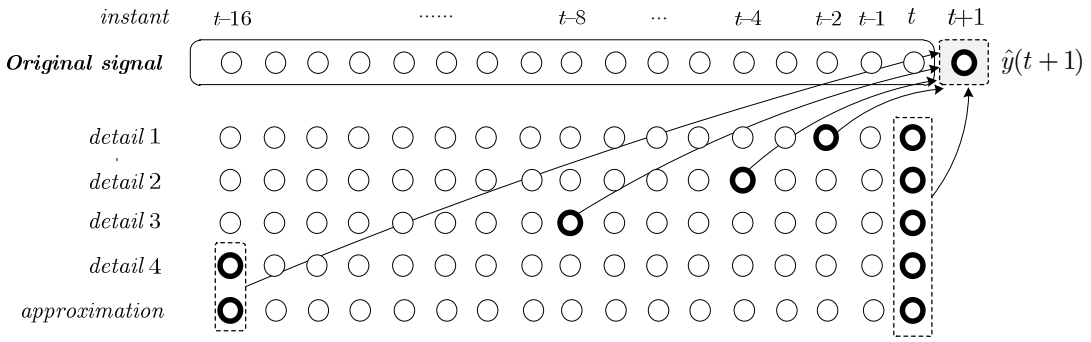


Figure 3.9 - Wavelet coefficients used for the prediction of the next value.

In this scheme, only ten coefficients are used for the prediction, which reflect detail and low-resolution information. The choice of this small set of lagged wavelet coefficients can be justified by two main reasons: completeness and parsimony (Renaud *et al.*, 2005). Completeness in the sense that the selected coefficients correspond to orthogonal basis for all the past values. So, for large A_j, these coefficients are just enough to express all the information of the series. Parsimony in the sense that the minimum number of coefficients to ensure a correct prediction is considered.

By means of a non-decimated Haar (*à-trous*) algorithm for wavelet decomposition and using the mentioned lagged wavelet coefficients as the inputs, Renaud *et al.* (2003) developed a linear wavelet prediction model, known as multi-scale autoregressive model, and also introduced a non-linear predictive model using neural network structures.

Thus, for the linear case, equation (3.50) is verified,

$$\hat{y}(t+1) = \sum_{j=1}^{L} \sum_{k=1}^{A_j} c_{j,k} d_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} c_{L+1,k} a_{L,t-2^L(k-1)} \tag{3.50}$$

where j is the number of levels ($j = 1,2,\ldots,L$), A_j is the order of the regressive model ($k = 1,2,\ldots,A_j$), $d_{j,k}$ represents the wavelet detail coefficients, $a_{j,k}$ represents the approximation coefficients, and $c_{j,k}$ denotes the regressive coefficients.

For the non-linear case, a neural network model was proposed, as described by (3.51).

$$\hat{y}(t+1) = \sum_{m=1}^{M} w_m \sigma \left(\sum_{j=1}^{L} \sum_{k=1}^{A_j} c_{j,k} d_{j,t-2^j(k-1)} + \sum_{k=1}^{A_j} c_{L+1,k} a_{L,t-2^L(k-1)} \right) \tag{3.51}$$

In the previous equation, $\sigma(\cdot)$ is an activation function in the hidden layer, which usually is a sigmoid logistic, and w_m is the weight between a neuron (M neurons) and the output $\hat{y}(t+1)$. To obtain the optimal weights (parameters), evolutionary programming algorithms or gradient-based back-propagation were proposed by the authors.

ii. Wavelet Neural Networks

Real systems are typically time-varying, which makes the application of regressive schemes (linear or non-linear) a difficult task. In effect, assuming the N observed past values of a given time series, $y(t-n)$, $n = 1,\ldots,N$, a typical regression model addresses the prediction by building a mapping $f(\cdot)$, such that future values are based on current and past observations, $y(t+p) = f(y(t), y(t-1),\ldots)$. However, it is not easy to express how $y(t-1)$ influences $y(t)$ and, simultaneously, using the same mapping, how $y(t-N-1)$ influences $y(t-N)$. To overcome this limitation, Ling and Wu (2005) proposed artificial process neuron model structures. The major characteristic that distinguishes the process neuron from the traditional artificial neuron, is that inputs and corresponding weights can be time-varying functions. A wavelet process neuron model, proposed by Gang *et al.* (2008), is a structure that merges process neuron models with wavelet functions, as illustrated in Figure 3.10.

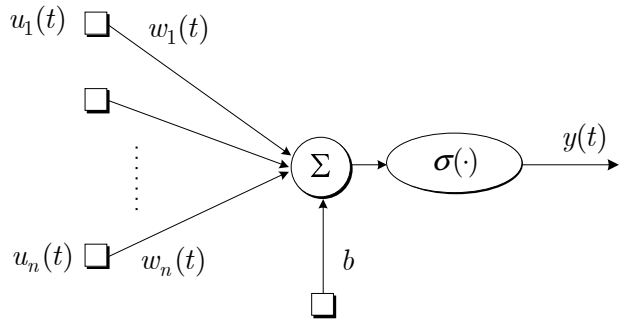


Figure 3.10 - Diagram of a wavelet process neuron model.

The wavelet process neuron model is composed of four main components: inputs, weights, an activation unit and the output. The inputs $u_i(t)$ and the connection weights $w_i(t)$ of the wavelet process neuron, are continuous time-varying functions. The

activation function $\sigma(\cdot)$, can be a generic function related with wavelet decomposition, which provides the structure with the capability of handling two dimensional information of time and space simultaneously. The output $y(t)$ of the wavelet process neuron model is given by (3.52), where the variable b is a bias term.

$$y(t) = \sigma\left(\sum_{i=1}^{n}\int_{0}^{T} w_i(t)u_i(t)dt - b\right) \tag{3.52}$$

Using a neural network composed of several wavelet process neurons, Gang *et al.* (2008) proposed a predictive scheme able to deal with the mentioned time-varying limitations. Figure 3.11 depicts a wavelet neural network composed of n neurons. The first layer is the input layer, which has only one node. The second layer is the hidden layer, which is composed of n wavelet process neurons. The last layer is the output layer, which also has only one node.

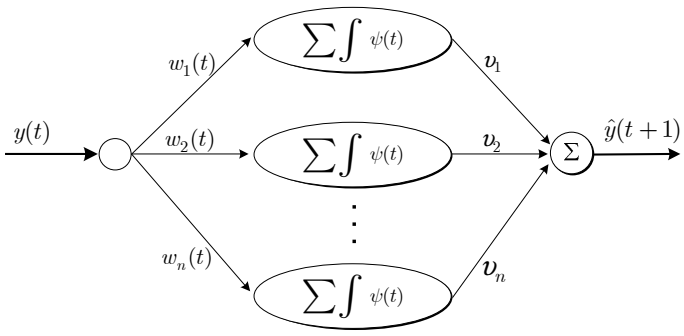


Figure 3.11 - Architecture of the wavelet neural network predictive scheme.

Assuming a time series $y(t-n)$, $n=0,1,...,N$, the estimation of the next output $\hat{y}(t+1)$, based on the wavelet neural network, can be expressed as (3.53).

$$\hat{y}(t+1) = \sum_{i=1}^{n} \upsilon_i \psi\left(\int_{0}^{T} w_i(t)y(t)dt - b_i\right) \tag{3.53}$$

In the previous equation, $w_i(t)$ denotes the connection weight function between the i^{th} wavelet process neuron in the hidden layer and the input unit $y(t)$. The parameter υ_i identifies the connection weights between the i^{th} wavelet process neuron in the hidden layer and the output unit $\hat{y}(t+1)$. Finally, $\psi(\cdot)$ is the wavelet basis function of the wavelet process neurons in the hidden layer.

This method and the corresponding learning algorithm were successful applied by Gang *et al.* (2008) to Mackey–Glass time series prediction. Moreover, results showed that the wavelet neural network has a faster convergence and presents a higher prediction accuracy when compared with the typical multi-layer neural networks.

3.3.5 Non-decimated Wavelet Transform

Shift variance

The result of a discrete wavelet transform typically assumes a hierarchical form to represent the information content, as it was illustrated in section 2.4.2, Figure 2.4. Such representation (hierarchical tree) is originated by the process of *decimation* or the retaining of one sample out of every two. The major advantage of decimation is that just enough information is kept to allow the exact reconstruction of the input data, thus it is ideal for compression applications. On the other hand, from the down-sampling operation, a significant potential problem occurs: a shift-variant transformation is obtained by the process.

Shift-variance is, by definition, "a *phenomenon of not necessarily matching the shift of the one-level DWT with the one-level DWT of the same shift of a data sequence"* (Chui, 1987). In other words, due to shift-variance, small shifts in the input signal may cause large changes in the wavelet coefficients. That is, a translated version of a signal does not correspond to the translated version of the wavelet transform of that signal.

Unquestionably, shift-invariance is important in many applications, namely in modelling and prediction tasks. In the context of prediction, wavelet decomposition should be ideally performed in such a way that the wavelet coefficients (for each level) at a time instant t should not be influenced by the behaviour of future instants. As a consequence, wavelet coefficients at a time instant t should be calculated from signal samples at instants before or equal to t, but never after.

The simplest way to avoid shift-variance is to skip the down-sampling operation and this is, in effect, the key principle that differentiates the normal from the so-called redundant or non-decimated wavelet transforms. The non-decimation is inherently a redundant scheme, as the output of each decomposition level contains the same number of samples as the input. In fact, a redundant transform of a N-length input time series originates, for each of the decomposition levels, a N-length resolution scale (the same length as the original signal). Although the drawback of the larger storage requirements, by means of this procedure it is possible to relate information at each resolution scale with the same point in time.

To achieve shift-invariant wavelet transforms, several modifications have been independently proposed and a number of different names have been suggested for the same principle. The original implementation was in the form of the *à-trous* (with holes) algorithm (Starck *et al.*, 2007), (Shensa, 1992), which refers to the insertion of zeros in the filters. Other variations are the non-decimated discrete wavelet transform, the overcomplete discrete wavelet transform (Zaciu *et al.*, 1999), the shift-invariant discrete wavelet transform (Lang *et al.*, 1995), the discrete wavelet frames (Unser *et al.*, 1998), and the maximal overlap discrete wavelet transform (Percival and Walden, 2000).

The *à-trous* wavelet transform

Among the several redundant discrete wavelet transforms (Aussem *et al.,* 1998), the *à-trous* exploits the redundant information by eliminating the decimation effect to generate approximations and details. Applying the *à-trous* wavelet transform to a signal $X(t)$, the respective scaling coefficients, at different scales, are computed by equations (3.54) and (3.55) (Renaud *et al.*, 2005).

$$A_0(t) = X(t) \tag{3.54}$$

$$A_j(t) = \sum_{k=-n_f}^{n_f} L_f(k) A_{j-1}(t + 2^{j-1}k) \tag{3.55}$$

The variable j identifies the decomposition level ($j = 1,\ldots,L$), $L_f(\cdot)$ is a low pass filter with compact support, and the parameter n_f is related with the filter length. The increase in the distances between the points, (2^{j-1}), explains the designation of *à-trous* (from the French, with holes), since it defines the information that is discarded/taken into account in the computation of the transform. The wavelet coefficients at scale j, $D_j(t)$, can be obtained by taking the difference of the successive smoothed version of the signal as described by (3.56).

$$D_j(t) = A_{j-1}(t) - A_j(t) \tag{3.56}$$

The vector composed of the coefficients $[D_1, D_2, \ldots, D_L, A_L]$ represents the *à-trous* wavelet transform of the signal up to the level L. Moreover, the original signal can be reconstructed as a linear combination of the wavelet and scaling coefficients, equation (3.57).

$$X(t) = A_L(t) + \sum_{j=1}^{L} D_j(t) \tag{3.57}$$

The approximation term $A_L(t)$ represents a smooth trend of the original signal $X(t)$, while the terms D_j, $j = 1,\ldots,L$, provide the details of the signal, which capture small features in the data.

The computational complexity of the *à-trous* wavelet algorithm, considering a fixed number of scales (L), is of $O(N)$, where N is the length of the signal.

The Haar *à-trous* wavelet transform

Among the several alternatives in terms of the *à-trous* wavelets, the simplest is the Haar one (Zheng *et al.*, 1999), which uses the following filter, equation (3.58).

$$L_f = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{3.58}$$

The Haar *à-trous* wavelet transform ensures that at any time instant t, the information after t is never used to compute the scaling and wavelet coefficients. Moreover, in opposition to typical wavelets, which require that the length of the signals is in the form $N = 2^n$, where N defines the length and $n \in \mathbb{N}$ is an integer, the *à-trous* transform does not have any restriction regarding the length of the signal.

Using the referred simple filter, (3.58), equations (3.54) and (3.55) can be simplified. In effect, considering the first decomposition level, the approximation and detail are obtained as follows, equation (3.59).

$$\begin{aligned} A_0(t) &= X(t) \\ A_1(t) &= \frac{1}{2}\big(A_0(t) + A_0(t-1)\big) \\ D_1(t) &= A_0(t) - A_1(t) \end{aligned} \tag{3.59}$$

This result can be easily extended to any decomposition level, as shown in equation (3.60), which establishes a straightforward implementation of the Haar *à-trous* wavelet transform.

$$\begin{aligned} A_{j+1}(t) &= \frac{1}{2}\Big(A_j(t) + A_j(t-2^j)\Big) \\ D_{j+1}(t) &= A_j(t) - A_{j+1}(t) \end{aligned} \tag{3.60}$$

Figure 3.12 shows the time steps that are used to compute the wavelet coefficients at the different scales, considering a decomposition level of $L = 4$. As can be observed, a wavelet coefficient at a time instant t is computed from the signal samples at instants before or equal to t, but never after.

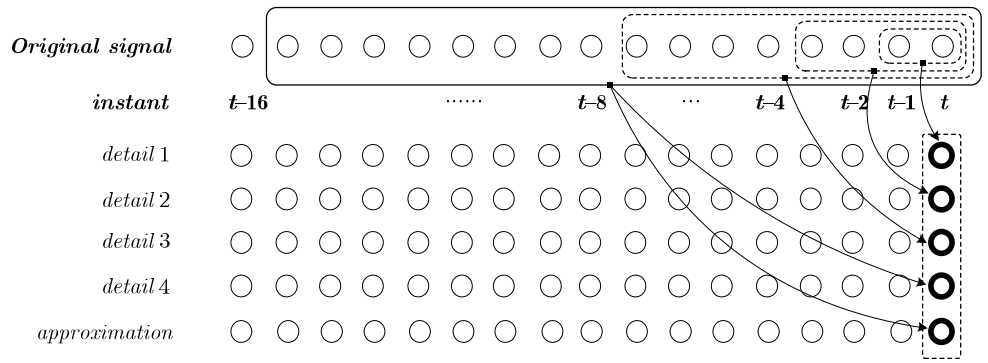


Figure 3.12 - Time steps used to compute the à-trous wavelet coefficients at the different scales.

3.4 Proposed Prediction Strategies

3.4.1 Main approach

Prediction and diagnosis methodologies have been intensively studied in multiple areas, including medicine. Examples of these are neural networks structures and fuzzy systems, as well as numerous other methodologies involving signal processing techniques (Shankaracharya *et al.*, 2010), (Ephzibah and Sundarapandian, 2012).

Particularly in the clinical field, it is usually very difficult to derive accurate mathematical models that take into account all involved variables and respective relationships, as a result of the complex interaction between multiple physiological variables. Furthermore, a fundamental aspect to consider in the development of such models, is the availability of data to compute the necessary parameters (training process). In effect, given the complexity of the dynamics and behaviour of most physiological systems, it is essential a representative dataset (both in quantity and in quality) to capture the diversity and variability of such systems. Even in case the dataset is available, the selection of the most representative examples may be not a straightforward process.

On the other hand, clinical professionals frequently have to make decisions based on their experience and knowledge, mainly by observing temporal trends of multiple signals. In effect, for complex systems, prediction and diagnosis approaches can be interpreted as a pattern recognition problem, very often not requiring explicit system modelling.

From the exposed, two main prediction strategies will be exploited in this work:

> *i*) The first approach assumes the development of accurate models for the prediction of physiological time series. It is based on a multi-model scheme assuming the existence of a suitable dataset for the required training process.
>
> *ii*) The second approach does not involve the explicit development of a model. It aims the estimation of the time series evolution trend, based on similar conditions retrieved from a historic dataset. The strategy is based on a wavelet multi-resolution scheme, where the most representative decomposition levels are selected to project the future trend.

In both schemes the starting point is the similarity analysis process introduced in the previous chapter. In the first strategy (multi-model scheme), the sequences in the historical dataset that present a dynamics similar to the current template are selected to compose the adequate dataset for the training of the neural models. In the second (trend estimation), these sequences are directly employed to predict the current time series future evolution, without the use of an explicit model.

To perform the similarity analysis an approach combining wavelet decomposition with the Karhunen-Loève transform was introduced in Chapter 2. Considering a template $X(t) \in \mathbb{R}^{1,N}$, the similarity analysis process originates a set of M similar conditions (patterns), $\mathbf{X}(t) \equiv \left\{ X_m(t) \in \mathbb{R}^{1,N} \right\}$, $m = 1,..,M$, from where the corresponding subsequent P future values, $\mathbf{Y}(t) \equiv \left\{ Y_m(t) \in \mathbb{R}^{1,P} \right\}$, are straightforwardly obtained (known past values).

$$X_m(t) \rightarrow Y_m(t) \qquad m = 1,\ldots,M$$

such that

$X_m(t)$ *is similar to* $X(t)$

The global set of patterns, $\mathbf{Z}(t) \in \mathbb{R}^{M,N+P}$, is therefore composed of two components $\mathbf{X}(t)$ and $\mathbf{Y}(t)$, in the form:

$$\mathbf{Z}(t) = [\ \mathbf{X}(t) \quad \mathbf{Y}(t)\ ] \tag{3.61}$$

Figure 3.13 illustrates the concept behind the proposed prediction approach. Basically, the known "*future*" evolution of the selected patterns, $\mathbf{Y}(t) \equiv \left\{ Y_m(t) \right\}$, can be used in a prediction mechanism to estimate the evolution of the current template, $\hat{Y}(t)$.

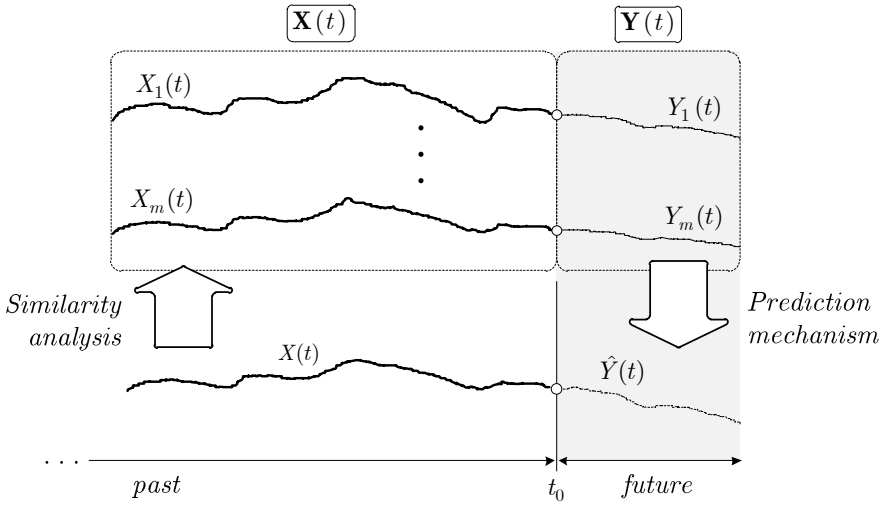


Figure 3.13 - Generic approach of the proposed prediction methodology.

1. Neural network multi-models scheme

The first approach addresses the forecast of time series through a predictive strategy based on a multi-models scheme using regression structures. Among regression models, neural networks have shown considerable capabilities to learn and to generalize from non-linear environments, enabling to capture the fundamental data dynamics. Additionally, since an independent neural sub-model is used for each sampling instant,

and does not depend on previous predictions, simple neural structures can be employed. Particularly in the present work, the capacity of the Generalized Regression Neural Networks will be exploited for modelling and prediction purposes. Figure 3.14 depicts the proposed multi-models scheme. As can be observed, it is mainly composed of three distinct phases: *i*) similarity analysis to identify patients who display a similar behaviour in their physiological time series; *ii*) multi-models training based on the time series retrieved from such patients (patterns), $\mathbf{X}(t) \equiv \{X_m(t)\}$ and $\mathbf{Y}(t) \equiv \{Y_m(t)\}$; *iii*) feeding the set of trained models with the current patient signal segment (template), $X(t)$, in order to estimate its future values, $\widehat{Y}(t)$.

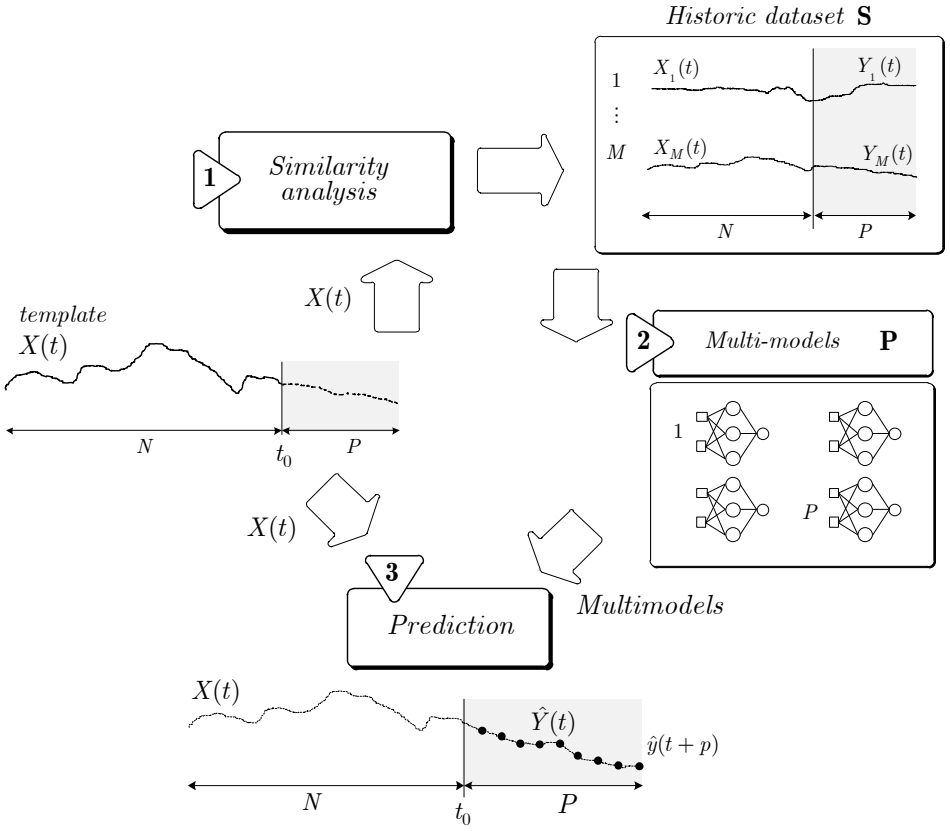


Figure 3.14 - Proposed multi-model prediction methodology.

i. Direct approach versus multi-models scheme

As mentioned, to derive long-term predictions two main approaches can be employed: the iterative and the direct (multi-models). Theoretically, long-term predictions can be obtained from a short-term predictor, by simply applying a one-step ahead predictor, iteratively. Direct approaches are applied to a specific future instant, in a way that is similar to the computation of the one step prediction. However, a different model is required for predicting each instant (multi-models scheme).

The iterative approach is often more adequate in cases where the fundamental dynamics is correctly captured by a short-term model (typically, a one-step ahead model). On the other hand, the main advantage of the direct method is the fact that direct forecasts are usually more robust in case of model misspecifications. In effect, by using a specific model for each future time instant, prediction errors are not accumulated and long-term predictions can be accurately estimated.

Given the exposed, a direct approach will be explored in this work for predicting the future values of the current signal $X(t)$, that is, $Y(t)$. Basically, a neural network multi-models scheme, where a different model $f_p(\cdot)$ $(p = 1,..,P)$ is derived for every future time instant $t_p = t_0 + p$, is employed in the estimation of $\hat{y}(t_p)$. Fundamentally, each model makes use of a time series regression formulation using neural networks, as described in (3.62).

$$\hat{y}(t_0 + p) = NN_p\Big(y(t_0), y(t_0 - 1), ...\Big) \tag{3.62}$$

ii. Neural networks to be incorporated into the multi-model scheme

The selection of the most adequate forecasting model to be part of the multi-model scheme is, definitely, a complex issue. In effect, several aspects have to be considered in this process, such as linear versus non-linear approaches, and time domain versus frequency domain approaches. A common solution consists in linear prediction models, based on time series regression analysis. However, this solution generally only produces reasonable prediction results for short-time forecasts and for linear time series signals. As it is known, biosignals are inherently non-linear and non-stationary, thus, the application of such methods is questionable. Recent advances in signal processing and in non-linear time series modelling have contributed to improve the potential of the prediction techniques. Approaches based on non-parametric regression, artificial neural networks, wavelet decomposition, among others, have showed to be adequate in many areas where linear regression methods have faced some difficulties.

Neural networks models are universal approximators and, theoretically, they can deal with any complex mapping, whether linear or non-linear. Thus, due to their powerful capacity in capturing non-linear mappings, high accuracy for learning and good robustness properties, they have been widely used in several areas, namely, for classification, modelling and prediction tasks. Moreover, they also have the capacity to learn the behaviour of poorly-understood phenomena and systems where the dependency between inputs and outputs are too complex to be mathematically described. Additionally, the prediction results of a well-trained neural network are usually accurate.

Given the exposed reasons, neural networks were chosen to integrate the multi-model structure.

2. Wavelet multi-resolution scheme

The second scheme exploits the wavelet transform to estimate the trend of a time series. As mentioned, wavelet transform is not a true forecasting technique itself. It provides a formal method to de-noise, de-trend, and decompose a complex time series, capturing useful information at various resolution levels, so that the capacity of a forecasting model can be improved.

Figure 3.15 depicts this global multi-resolution scheme. It is composed of three main distinct phases: *i*) similarity analysis to identify patients who display similar behaviour in their physiological time series; *ii*) multi-resolution decomposition of the time series retrieved from such patients, $\mathbf{X}(t) \equiv \left\{ X_m(t) \in \mathbb{R}^{1,N} \right\}$ and $\mathbf{Y}(t) \equiv \left\{ Y_m(t) \in \mathbb{R}^{1,N} \right\}$; *iii*) projection of the current patient data (template), $X(t)$, into the future, $\widehat{Y}(t)$, by combining the optimal decomposition levels of the historic patterns $\mathbf{Y}(t)$.

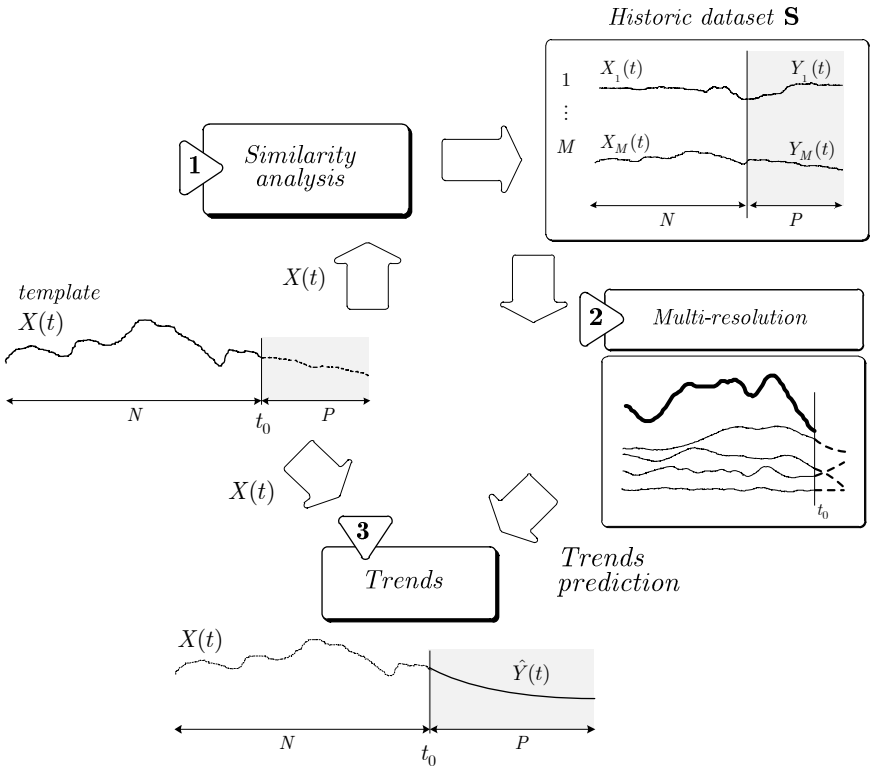


Figure 3.15 - Proposed wavelet multi-resolution prediction methodology.

The forecasting process using the wavelet transform typically involves the independent prediction of each decomposition level, through regressive models such as ARIMA or neural networks. Subsequently, the individual predictions are combined to derive the global time series prediction. Following a different approach, the methodology proposed in the present work does not explicitly involve a model. In effect, it is based on the wavelet decomposition of the similar historical patterns, in order to derive an optimal future trend for the template. To achieve this goal two main steps are required.

i. Representative trends

The first step involves the wavelet decomposition of the similar historic time series signals. Then, at each decomposition level, the obtained decompositions are combined to obtain a representative trend. The subtractive clustering method is used for this purpose, given its properties, such as robustness to outliers and the capability to deal with the concept of data density. Moreover, the clustering process is carried out based on the reduced set of basis derived from the similarity analysis (coefficients) to represent each decomposition level.

ii. Optimal trends

The second step aims at the selection of a subset from the representative trends, designated as optimal trends, to be used in the prediction of the current time series. To achieve this goal, an optimization process involving the minimization of a set of distance-based measurements is proposed. From this process, it is possible to quantify the aptness of each individual representative trend to integrate the optimal subset. In particular, the similarity measure proposed in Chapter 2 is employed, comprising the original signal $X(t)$, the respective wavelet decomposition, and the representative trends.

Finally, the resulting optimal trends are then straightforwardly extended to the future and aggregated to derive the global trend, $\widehat{Y}(t)$.

3.4.2 Scheme 1: Neural Network Multi-models

1. Prediction scheme based on neural network multi-models

Assuming a direct scheme (multi-models), the prediction process implies the construction of a different model for each prediction instant. Thus, each future value at a time instant p, $y(t+p)$, has to be independently estimated in the form given by (3.63).

$$\hat{y}(t+p \mid t) = f_p\Big(y(t), y(t-1), ..., y(t-n+1)) \Big) \tag{3.63}$$

The mapping $f_p(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}$ identifies the particular model for the p^{th} instant. The variable n identifies the order of the model, that is, it corresponds to the number of past observations that are employed by the prediction model. Using a simplified representation, equation (3.63) can be rewritten as (3.64), where $\vartheta(t)$ defines the available information at time instant t, and $\hat{\theta}_p$ are the parameters that characterize each specific neural mapping.

$$\begin{aligned} &\hat{y}(t+p \mid t) = f_p(\ \vartheta(t),\ \hat{\theta}_p) \\ &\vartheta(t) = \Big[\ y(t), y(t-1) \ldots y(t-n+1) \Big] \end{aligned} \tag{3.64}$$

2. Neural-network regression models

As previously referred, each regression sub-model $f_p(\cdot)$ is implemented by means of a neural network model. Basically, the goal is to build a mapping such that for an input $\vartheta(t)$ the output $\hat{y}(t+p \mid t)$ is obtained.

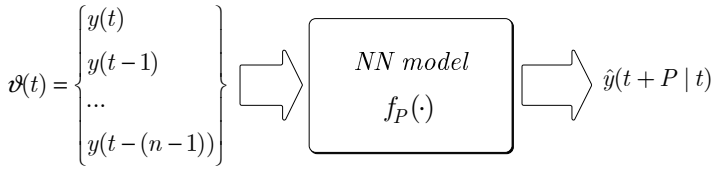


Figure 3.16 - Neural network regression model.

Since each neural network model can be independently trained by means of a standard backpropagation algorithm, common neural network structures can be easily implemented, such as multi-layer neural networks, radial basis networks, etc.

In particular, generalized regression neural networks (GRNN), a type of radial basis function networks, were the elected for this work. A GRNN model can be seen as a normalized radial basis function (RBF) network, in which there is a hidden unit centred at every training case. These RBF units are called *kernels* and, usually, are probability density functions, such as Gaussian functions. The weights from the hidden to output layer are just the target values, so the output is simply a weighted average of the target values of the training cases that are close to the given input case. As a consequence, the only parameters to be learned are the widths of the RBF units (Bauer, 1995).

Figure 3.17 depicts an example of a particular static GRNN, consisting of an input with three dimensions, $\vartheta \in \mathbb{R}^{3,1}$ and one output, $y(\vartheta_i)$. Furthermore, four training pairs $\{\vartheta_j, y_j\}$ $j = 1,..,4$, and four hidden neurons, $h_d(\cdot)$, $d = 1,...,4$, are considered.

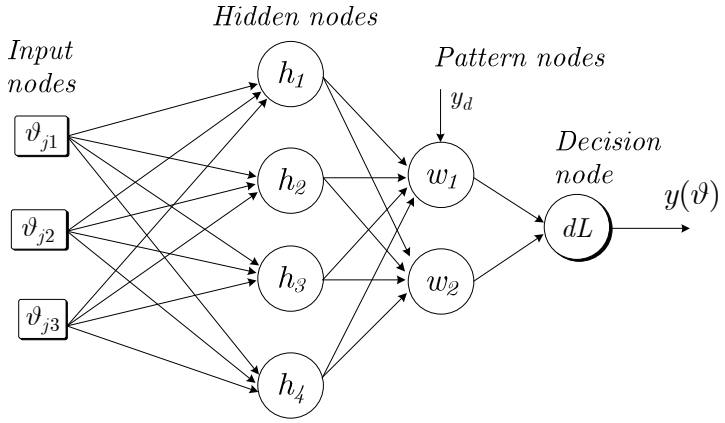


Figure 3.17 - Example of a GRNN architecture.

Receiving the vector of input values ϑ_j from the input layer, the hidden layer $h_d(\cdot)$ computes the Euclidean distance between the input vector and the neuron's centre of the kernel ϑ_d, to obtain the RBF kernel function, $\sigma_d(\vartheta_d - \vartheta_j)$. The resulting values are passed to the numerator and denominator neurons in the pattern layer, respectively $w_1(\cdot)$ and $w_2(\cdot)$. This second unit (denominator) adds the values coming from each of the hidden neurons $\sigma_d(\cdot)$, while the first unit (numerator) adds those values multiplied by the actual target value (y_d). Finally, the decision layer, dL, divides the value in the numerator unit by the value in the denominator unit to derive the predicted value, $y(\vartheta_j)$. Mathematically, a GRNN can be described by (3.65).

$$y(\vartheta_j) = \frac{w_1(\cdot)}{w_2(\cdot)} = \frac{\sum_{j=1}^{NT} y_d \ \sigma_d(\vartheta_d - \vartheta_j)}{\sum_{j=1}^{NT} \sigma_d(\vartheta_d - \vartheta_j)} \tag{3.65}$$

The vector ϑ_j represents the current input, $\sigma_d(\cdot)$ identifies the particular kernel radial basis function, y_d is the target output corresponding to the input ϑ_d, and NT is the total number of training pairs $\{\vartheta_d, y_d\}$, $d = 1,..,NT$. In case a Gaussian radial basis function is used, $\sigma_d(\cdot)$ is defined according to equation (3.66), where the parameter λ defines the kernels width.

$$\sigma_d(\cdot) = exp^{\frac{-(\vartheta_d - \vartheta_j)^T (\vartheta_d - \vartheta_j)}{2\lambda^2}} \tag{3.66}$$

The principal advantages of GRNNs are their aptness for smooth function approximation, their ability to predict the behaviour of systems based on few training samples and their interpolation properties between training samples (Bauer, 1995). Moreover, they enable a fast learning and, for specific applications, they are often more accurate than multi-layer perceptron networks. These properties make the GRNN a very useful tool to perform predictions (Rutkowski, 2004). On the other hand, like kernel methods, they suffer from the curse of dimensionality dilemma requiring more memory space to store the trained model.

3. Multi-models training

Once the type of model is selected (GRNN), the next aspect to consider is the definition of the required training data. According to the defined strategy, this data can be obtained as a result of the similarity analysis procedure.

On the other hand, the selection of a specific sub-model structure involves the characterization of the function $f_P(\cdot)$ and of the number of past observations to be considered by the model. Typically the dimension of the training dataset, $\{\vartheta_d, y_d\}$, $d = 1,..,D$, determines the number of hidden neurons. In the application of the GRNN structure to the particular problem of acute hypotensive episodes (AHE)

prediction (Section 4.4), the number of previous instants considered by each model (designated as the *order*), together with the time period before the starting of the forecast window (designated as *size*), determine the dimension of the dataset and, therefore, the number of hidden layers. These parameters, *size* and *order*, as well as the prediction scheme, are illustrated in Figure 3.18.

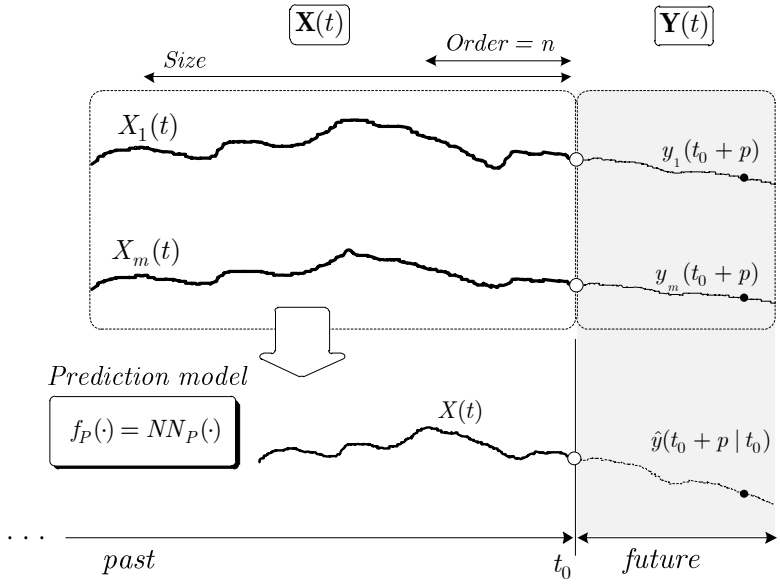


Figure 3.18 - Model 1: Neural network multi-model scheme.

In order to train each neural network model, $NN_p(\cdot)$, $p = 1,..,P$, the available information $\mathbf{X}(t)$ (set of M time series similar to the current signal $X(t)$), as well as the respective future values, $\mathbf{Y}(t)$, are used. Considering the first instant ($t_1 = t_0 + 1$), that is, one-step ahead prediction, each training data pair is composed of n values (*order*) obtained from the similar templates, $\mathbf{X}(t)$ (from instant ($t_0 - n + 1$) to instant t_0), and the respective future values $\mathbf{Y}(t)$ (at time instant t_1), as represented in (3.67).

$$\left\{ \mathbf{X}(t_0, t_0 - 1, ..., t_0 - n + 1),\ \mathbf{Y}(t_1) \right\} \tag{3.67}$$

The training dataset can be extended considering other data pairs, using available information before instant t_0 . Thus, a set of training data pairs (of dimension *size*), can be used.

$$\left\{ \mathbf{X}(t_0 - s, ..., t_0 - n - s + 1),\ \mathbf{Y}(t_1 - s) \right\} \quad s = 0, ..., size - n \tag{3.68}$$

In general, for training a neural model $NN_p(\cdot)$ to predict time instant $(t_0 + p)$, information from patterns $\mathbf{X}(t)$ (from instant ($t_0 - n - s + 1$) to instant t_0), and the

respective future values $\mathbf{Y}(t)$ at time instant $t_p = t_0 + p$, are used. Therefore, the training dataset is composed of the following data pairs, (3.69).

$$\left\{ \mathbf{X}(t_0 - s, \cdots, t_0 - n - s + 1),\ \mathbf{Y}(t_p - s) \right\} \quad s = 0,...,size - n \tag{3.69}$$

4. Prediction with neural-network multi-models

Once obtained the P neural models, $NN_p(\cdot)$, $p = 1,..,P$, the prediction of the future P values of the unknown signal $Y(t)$, that is, $\hat{y}(t_0 + p \mid t_0)$, is straightforwardly performed by (3.64), considering the particular time instant $t = t_0$, as given by (3.70).

$$\begin{aligned} &\hat{y}(t_0 + p \mid t_0) = NN(\ \vartheta(t_0), \hat{\theta}_p\ ) \\ &\vartheta(t_0) = \left[\ x(t_0), x(t_0 - 1) \ldots x(t_0 - (n-1))\ \right] \end{aligned} \tag{3.70}$$

The parameters $\hat{\theta}_p \in \mathbb{R}^{n\theta}$ characterize the specific mapping, being n_θ the number of parameters.

The main assumption of this strategy is that the dynamics of similar situations in the past can be used to predict the future. The current situation is expressed by the time series $X(t)$, and the dynamics of the past is captured by means of the neural network models since they reflect the dynamics of the time series $X_m(t)$ and respective evolution $Y_m(t)$. These models are used to estimate the future values of the current time series $X(t)$, therefore, to obtain an estimation $\hat{Y}(t)$ of the signal $Y(t)$.

3.4.3 Scheme 2: Wavelet Multi-resolution

The Figure 3.19 depicts the main steps involved in the proposed multi-resolution prediction scheme, namely: *i*) template decomposition; *ii*) representative trends; *iii*) optimal trends; *iv*) trend prediction.

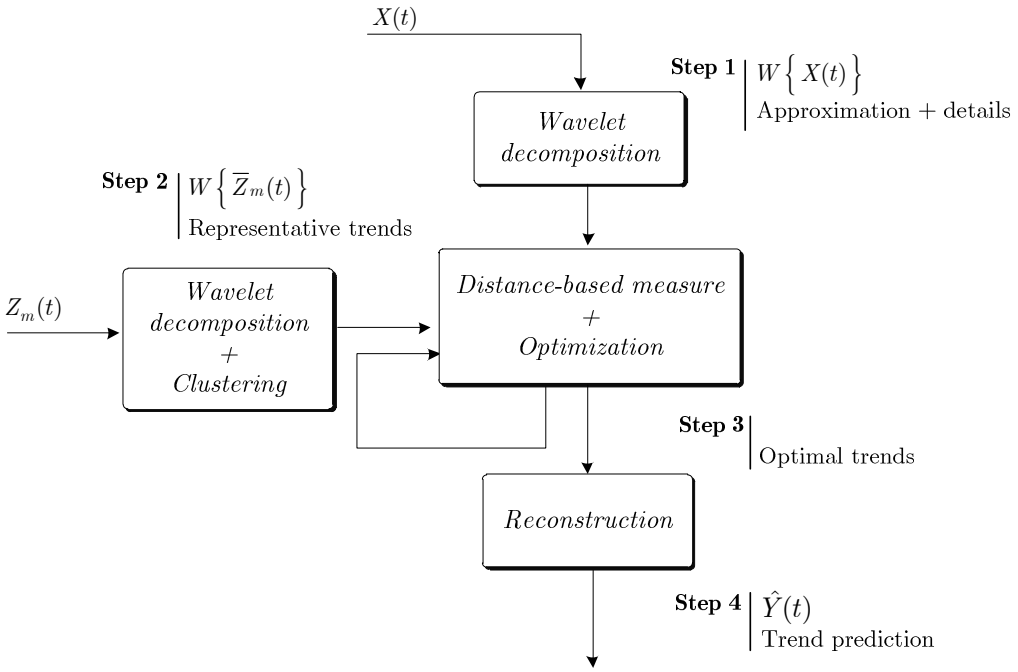


Figure 3.19 - Steps involved in the multi-resolution prediction methodology.

Step 1. Template decomposition

In the first step, the template $X(t)$ is decomposed using the Haar *à-trous* wavelet transform (subsection 3.3.5). The Figure 3.20 illustrates this procedure for a template with a duration of $N = 32$. The level of decomposition considered was $L = 5$, thus, the details $d^l X(t)$ for scales 1 to 5, plus the smooth trend $a^5 X(t)$, are presented.

The original time series can be reconstructed at each instant, by employing the corresponding value at every scale. This is strongly related with the shift invariance, that is, with the possibility to relate information at each resolution scale with the same point in time. The Figure 3.20 exemplifies this aspect, where at time instant $t_0 = 32$, the signal $X(t)$ can be obtained as the sum of details $d^l X(t)$, for scales 1 to 5, plus the smooth trend $a^5 X(t)$, as given by equation (3.71).

$$x(t_0) = d^1 x(t_0) + d^2 x(t_0) + d^3 x(t_0) + d^4 x(t_0) + d^5 x(t_0) + a^5 x(t_0) \tag{3.71}$$

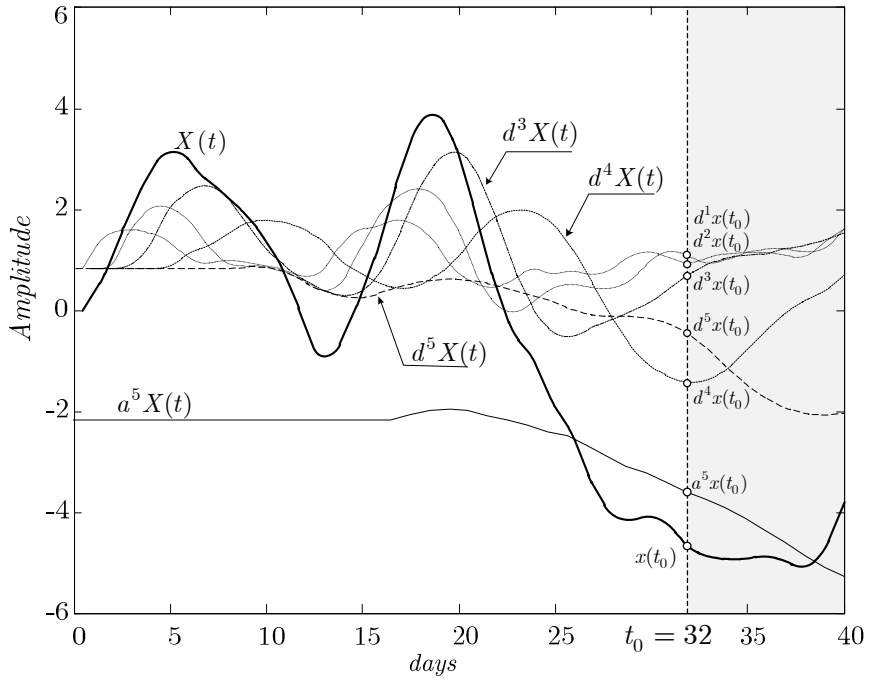


Figure 3.20 - Template decomposition using the wavelet à-trous transform.

The value $x(t_0)$ represents the particular value of $X(t)$ at time instant t_0, while $d^l x(t_0)$ and $d^5 x(t_0)$, respectively denote the values of the details $d^l X(t)$ and approximation $a^5 X(t)$ at the same instant. For a general case, given a time series $X(t) \in \mathbb{R}^{1,N}$, its wavelet decomposition is represented by $W\{X\} \in \mathbb{R}^{L+1,N}$, according to (3.72).

$$W\{X\} = \left\{ d^l X(t), a^L X(t) \right\},\ l = 1,..,L \tag{3.72}$$

Step 2. Representative trends

The second step involves the determination, at each decomposition level, of the most representative trends from the historic signals retrieved by the similarity analysis procedure. In a first phase, all the identified historic signals are decomposed using the *à-trous* wavelet. Then, in a second phase, the most representative decomposition at each level is determined through a clustering process.

i. Wavelet decomposition of the similar time series

Similarly to the template decomposition, each pattern $Z_m(t)$ that composes the set of M analogous historic signals is decomposed according to (3.73).

$$Z_m(t) = a^L Z_m(t) + \sum_{l=1}^{L} d^l Z_m(t), \qquad t = 1..,N+P \tag{3.73}$$

The variables $a^L Z_m(t)$ and $d^l Z_m(t)$, $l = 1,\ldots,L$, represent the approximation and the details, respectively. Similarly to equation (3.72), this wavelet decomposition can be represented as (3.74).

$$W\{Z_m\} = \left\{d^l Z_m(t), a^L Z_m(t)\right\},\ l = 1,..,L \tag{3.74}$$

It is important to note that, in this case, the decomposition is extended to the "*future*" (time instants from $N+1$ to $N+P$), thus, $W\{Z_m\} \in \mathbb{R}^{L+1,N+P}$.

Although the decomposition at each level has been performed considering a set of similar time patterns, they are, naturally, different. As can be seen in Figure 3.21, for the same decomposition level ($L = 5$), the obtained signals $d^5 Z_m(t)$ present variations and do not perfectly correspond to the decomposition of the template $d^5 X(t)$. In this example, the number of patterns considered is $M = 10$, the length of the template $X(t)$ is $N = 32$, and the forecast horizon is $P = 8$.

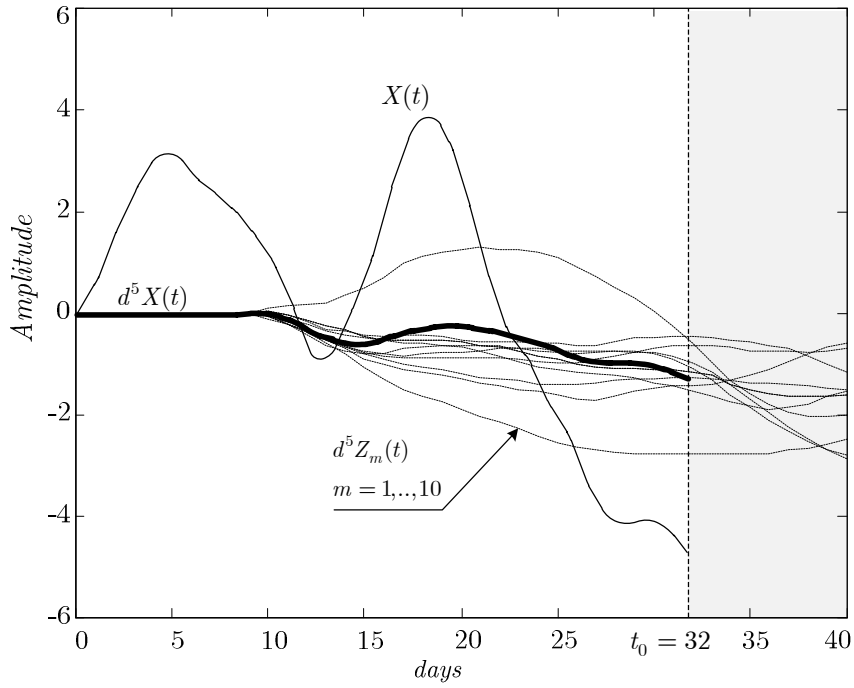


Figure 3.21 - Wavelet decomposition of patterns at level l=5.

As result, the next step consists in the extraction of the most representative trend at each decomposition level, that is, the trend with major potential to contribute for consistent future predictions.

ii. Extraction of representative trends

In order to obtain the representative trend at each level $l = 1,...,L$, a clustering strategy is employed by means of the subtractive method introduced in subsection 2.5.8, thus according to (3.75) and (3.76).

$$d^l\overline{Z}(t) = clustering\left\{d^l Z_m(t)\right\}, \quad m = 1,..,M \tag{3.75}$$

$$a^L\overline{Z}(t) = clustering\left\{a^L Z_m(t)\right\}, \; m = 1,..,M \tag{3.76}$$

In equations (3.75) and (3.76), $d^l\overline{Z}(t) \in \mathbb{R}^{1,N+P}$ and $a^L\overline{Z}(t) \in \mathbb{R}^{1,N+P}$ denote, respectively, the representative details and approximation.

In order to apply the subtractive clustering method for the extraction of representative trends, two main assumptions are considered: *i*) the subtractive clustering is not directly employed to the time series; *ii*) only a cluster has to be computed.

To address the first assumption, the M time series at each decomposition level are previously described by means of a common representation using the wavelet transform/KLT scheme proposed in Chapter 2.

Consequently, each of the M signals (respective details and approximation) can be written as a linear combination of a set of basis, according to (3.77) and (3.78).

$$d^l Z_m(t) = \sum_{j=1}^{J} \alpha_{j,m}^{l} \; \varphi_j^l(t) \;, l = 1,\ldots,L \tag{3.77}$$

$$a^L Z_m(t) = \sum_{j=1}^{J} \beta_{j,m} \; \varphi_j^{L+1}(t) \tag{3.78}$$

In the previous equations, $\varphi_j^l(t) \in \mathbb{R}^{1,N+P}$ defines a set of J basis (wavelets) at the decomposition level l, and $\alpha_{j,m}^{l}$ are the corresponding coefficients. In turn, the variables $\beta_{j,m}$ represent the expansion coefficients for the approximation level. Using this reduced representation, the decomposition $W\left\{Z_m\right\}$ can be defined by a set of coefficients $\Phi\left\{Z_m\right\} \in^{L+1,J}$.

$$\begin{aligned} W\left\{Z_m\right\} \approx \Phi\left\{Z_m\right\} = \\ = \left[\alpha_{1,m}^1, \alpha_{2,m}^1, ..., \alpha_{J,m}^1, ..., \alpha_{1,m}^L, \alpha_{2,m}^L, ..., \alpha_{J,m}^L, \beta_{1,m}, \beta_{2,m}, ..., \beta_{J,m}\right] = \\ = \left[\Phi\{Z_m^1\}, \Phi\{Z_m^2\}, ..., \Phi\{Z_m^L\}, \Phi\{Z_m^{L+1}\}\right] \end{aligned} \tag{3.79}$$

The result of the clustering process (only the first cluster) are the vectors of coefficients,

$$\Phi\left\{ d^l\overline{Z}(t)\right\} = clustering\left\{\Phi\left\{\mathbf{Z}^l\right\}\right\} \; l = 1,...,L \tag{3.80}$$

$$\Phi\left\{a^L\overline{Z}(t)\right\} = clustering\left\{\Phi\left\{\mathbf{Z}^{L+1}\right\}\right\} \tag{3.81}$$

where $\Phi\left\{\mathbf{Z}^l\right\}=\left\{\Phi\{Z_1^l\},...,\Phi\{Z_M^l\}\right\}$, and $\Phi\left\{\mathbf{Z}^{L+1}\right\}=\left\{\Phi\{Z_1^{L+1}\},...,\Phi\{Z_M^{L+1}\}\right\}$.

These coefficients are finally employed to determine the representative trend at level l, $d^l\overline{Z}(t)$, equation (3.82), where $\Phi\left\{d^l\overline{Z}(t)\right\}_j$ is the j^{th} element of the vector $\Phi\left\{d^l\overline{Z}(t)\right\}$ and $\Phi\left\{a^L\overline{Z}(t)\right\}_j$ is the j^{th} element of the vector $\Phi\left\{a^L\overline{Z}(t)\right\}$.

$$d^l\overline{Z}(t)=\sum_{j=1}^{J}\Phi\left\{d^l\overline{Z}(t)\right\}_j\varphi_j^l(t)\;\; l=1,...,L \tag{3.82}$$

$$a^L\overline{Z}(t)=\sum_{j=1}^{J}\Phi\left\{a^L\overline{Z}(t)\right\}_j\varphi_j^{L+1}(t) \tag{3.83}$$

Figure 3.22 presents the computed representative trend, $d^l\overline{Z}(t)$, obtained for the example illustrated in Figure 3.21, thus considering $l=5$. Note that in this figure the signal $d^lX(t)$, $l=5$, was vertical shifted to allow a more clear visualization.

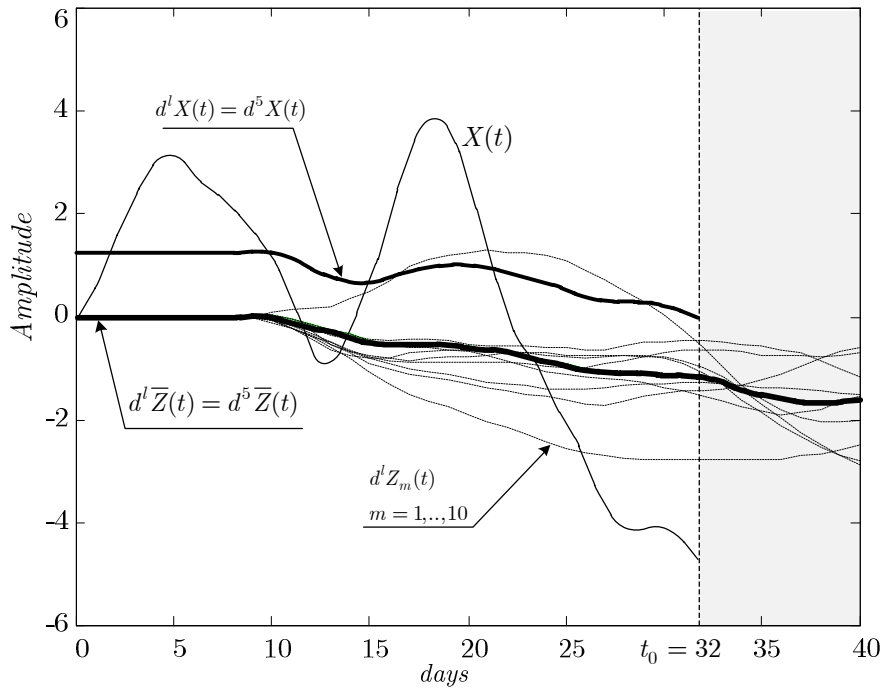


Figure 3.22 - Extraction of the representative trend for the decomposition level l=5.

Using the representative trends of the several decomposition levels, $d^l\overline{Z}(t)$, $l=1,...,L$, and the approximation $a^L\overline{Z}(t)$, the future instants of the template, $\widehat{Y}(t)$, $t\in[N+1,N+P]$, can be easily obtained.

$$\widehat{Y}(t) = a^L \overline{Z}(t) + \sum_{l=1}^{L} d^l \overline{Z}(t) \qquad t = N+1,..,N+P \tag{3.84}$$

However, this simple approach naturally has some problems. In effect, the patterns are not perfectly similar to the template, and the employed ones may present future unpredictabilities, inconsistencies, etc., thus resulting in inappropriate future estimations. Consequently, in many cases it is not possible to attain good and reliable predictions with this methodology.

Step 3. Optimal trends

In the third step, the set of representative trends is reduced to an optimal set, that is, to a set of trends that have the potential to contribute to a consistent prediction.

In effect, the proposed methodology selects a subset of the most appropriate representative trends (optimal trends) to be considered in the prediction process, by minimizing a set of distance-based measures. The main goal is to derive a quantitative measure that allows to assess the probability of a representative trend to contribute to a correct estimation.

It is important to remind that no explicit model is involved in this methodology, and that its goal is not to perform an accurate prediction, but to obtain a reasonable estimation of the future trend.

i. Distance-based measures

The distance-based measures are computed for each decomposition level $l = 1,..,L+1$, where $L+1$ stands for the approximation, using:

- The template $X(t) \in \mathbb{R}^{1,N}$
- The corresponding wavelet decomposition at the l level, $d^l X(t) \in \mathbb{R}^{1,N}$
- The wavelet decomposition of the similar patterns at the same level, $d^l Z_m(t) \in \mathbb{R}^{1,N+P}$, $m = 1,..,M$
- The corresponding clustering, that is, the representative trends $d^l \overline{Z}(t) \in \mathbb{R}^{1,N+P}$.

Based on these signals (depicted in Figure 3.22), a set of distance-based measures θ_i^l are computed as follows:

$$\theta_1^l = S(X(t), d^l X(t)), \qquad t = 1,\ldots,N \tag{3.85}$$

$$\theta_2^l = S(d^l X(t), d^l \overline{Z}(t)), \qquad t = 1,\ldots,N \tag{3.86}$$

$$\theta_3^l = mean\left[S(d^l X(t), d^l Z_m(t))\right], \qquad t = 1,\ldots,N;\ m = 1,\ldots,M \tag{3.87}$$

$$\theta_4^l = exp\left[-std\left[S(d^l X(t), d^l Z_m(t))\right]\right], \qquad t = 1,\ldots,N;\ m = 1,\ldots,M \tag{3.88}$$

$$\theta_5^l = mean\left[S(d^l\overline{Z}(t), d^l Z_m(t))\right], \qquad t = N+1\ldots, N+P;\ m = 1,\ldots, M \tag{3.89}$$

$$\theta_6^l = exp\left[-std\left[S(d^l\overline{Z}(t), d^l Z_m(t))\right]\right], \qquad t = N+1,\ldots, N+P;\ m = 1,\ldots, M \tag{3.90}$$

As result, a vector composed of six measures is obtained, where $\theta_i \in [0,1]$.

$$\Theta^l = \left[\ \theta_1^l, \theta_2^l, \ldots, \theta_R^l\right]\ _{R=6} \tag{3.91}$$

The similarity measure $S(X_1(t), X_2(t)) \in [0,1]$ is defined in section 2.5.5, equation (2.105). The operators $mean(\cdot)$ and $std(\cdot)$ denote the mean and standard deviation operators, respectively.

These measurements can be qualitatively justified as follows:

- The first measure, θ_1^l, quantifies the similarity between the original template and the respective decomposition (at l level). Therefore, it assesses the proximity between the original template and the current decomposition. If they are similar it means that the current decomposition level is appropriated to be included in the determination of the future trend (they present the same behaviour).
- The second measure, θ_2^l, provides a quantification of how the decomposition of historic patterns (representative trend) is similar to the decomposition of the template at the same level. Basically, it measures the appropriateness of the current representative trend (they present a similar dynamics).
- The third and fourth measures, θ_3^l and θ_4^l, quantify the variation of the patterns, that is, how homogenous the historic patterns are during the time instants corresponding to the template. The more homogenous they are, the greater is the probability of their future evolution to be consistent.
- Finally, the fifth and sixth measures, θ_5^l and θ_6^l, quantify the discrepancy of the patterns, that is, how homogenous the historic patterns are through the forecast horizon. The main idea is that a homogenous group increases the likelihood to achieve a reliable future prediction.

ii. Selection of the optimal trends

The optimization strategy, through the minimization of a set of distance-based measures θ_i^l, selects the most appropriate representative trends in order to increase the correctness of the future trend. Therefore, a decision regarding the inclusion or exclusion of the representative trend in the global trend, is taken according to the values of these measures.

In order to combine these parameters into a single index $p(\Theta^l)$, a conjunction operator is employed. Examples of such operator are the *product*$(\cdot)$ and *minimum*$(\cdot)$. Assuming that the *product*$(\cdot)$ is employed, the index for each representative trend is obtained according to:

$$p(\Theta^l) = \prod_{i=1}^{R} \theta_i^l \quad l = 1,...,L+1 \tag{3.92}$$

However, since more than a combination of the representative trends can be considered, an aggregation operator is used to aggregate the possible combinations. Examples of such operator are the *sum*$(\cdot)$ or the *maximum*$(\cdot)$. If two levels L_i and L_j are combined by the *maximum*$(\cdot)$ operator, the global index $p(\Theta_\sigma)$ is obtained according to (3.93),

$$p(\Theta_\sigma) = \max\left\{p(\Theta^{L_i}), p(\Theta^{L_j})\right\} \tag{3.93}$$

where the subscript σ denotes the combination of levels L_i and L_j.

It is important to emphasize that the process involves not only each representative trend individually, but also takes into account the combinations of the several representative trends. In this case, since various decompositions can be simultaneously combined, equation (3.93) is modified to reflect this aspect.

$$p(\Theta) = \max\left\{p(\Theta\sigma_1),...,p(\Theta\sigma_i),...,p(\Theta\sigma_n)\right\} \tag{3.94}$$

In equation (3.94) the variable σ_i denotes each of the possible level combinations, resulting from the operator $C(nL, nN)$ (combinations of nL taken nN at a time).

$$C(nL, nN) = \frac{(nL)!}{(nL - nN)!\ nN!} \tag{3.95}$$

In particular, considering the decomposition levels $L = [3,4,5,6]$ (the last three decomposition levels and the approximation), a total of 15 combinations (σ_i, $i = 1,..,15$) are possible. Namely:

$$C([3,4,5,6],1) \rightarrow \{3\},\{4\},\{5\},\{6\} \equiv \sigma_1,\sigma_2,\sigma_3,\sigma_4$$
$$C([3,4,5,6],2) \rightarrow \{3,4\},\{3,5\},\{3,6\},\{4,5\},\{4,6\},\{5,6\} \equiv \sigma_5,\sigma_6,\sigma_7,\sigma_8,\sigma_9,\sigma_{10}$$
$$C([3,4,5,6],3) \rightarrow \{3,4,5\},\{3,4,6\},\{3,5,6\},\{4,5,6\} \equiv \sigma_{11},\sigma_{12},\sigma_{13},\sigma_{14}$$
$$C([3,4,5,6],4) \rightarrow \{3,4,5,6\} \equiv \sigma_{15}$$

The use of only the last three decomposition levels in the combination process can be justified by the results of the wavelet decomposition. In effect, the last levels of the decomposition present the slowest dynamics (levels 3, 4, 5, 6), while the low level details

(levels 1 and 2) present highest variations and noise. Thus, it makes sense to discard these first levels for the prediction task.

Step 4. Trend prediction

Finally, the optimal trends are combined to obtain the trend prediction corresponding to the template $X(t)$, as (3.96)

$$\widehat{Y}(t) = a^{\sigma}\,\overline{Z}(t) + \sum d^{\sigma}\,\overline{Z}(t) \qquad t = N+1,..,N+P \tag{3.96}$$

where the subscript σ denotes the representative trends identified by the minimization process of the distance-based measures.

3.5 Conclusions

The major goal of the present chapter was to propose methodologies for the prediction of time series. Although the methods have been developed in the clinical context, they can be viewed as of general purpose, suitable to be applied to any type of temporal series.

The starting point for these prediction strategies is the similarity analysis described in Chapter 2. By means of this, signals presenting a dynamics similar to the current time series, are retrieved from the historic and then employed to support the prediction of that current time series. The main clinical hypothesis is that patterns in physiological time series of other patients with similar disease progressions or intra-patient occurrences, may have prognostic value in estimating future clinical evolution of a current condition.

Based on the similarities between physiological time series, two main strategies have been proposed. The first involves the development of accurate models for the prediction of the time series. It is based on a multi-model scheme assuming the existence of a suitable dataset (derived from the similarity analysis procedure) for the required training process. From among regression models, generalized regression neural network structures were chosen to integrate the multi-model scheme, given their approximation and generalization properties. The strategy employs a direct approach and, therefore, an independent model (multi-models arrangement) has to be trained for each sampling instant within the prediction horizon.

The second approach does not involve the explicit development of a model. The methodology consists in the estimation of a trend for the future, supported on the "*future*" evolution of similar historic patterns. The strategy is based on a wavelet decomposition, where the most representative trends are extracted at each decomposition level. Then, a set of distance-based measures able to assess the likelihood of each representative trend is introduced. From an optimization process a subset of these trends are selected and aggregated to derive the required time series trend.

4. Results

This chapter presents and discusses the results related with the algorithms and methodologies proposed through this thesis.

Partially integrated in the HeartCyle project, the main focus of the present work is the development of algorithms able to detect and predict specific cardiovascular conditions, fundamentally based on the electrocardiogram and blood pressure signals. The central efforts are on time series similarity search methodologies, in order to enable the detection of such conditions, and on predictive techniques to be applied in the trend analysis of biosignals and early detection of future events.

Regarding similarity measures, a comparison in terms of the sensitivity of several approaches (including the proposed one) is, in a first phase, performed. The results highlight the importance of knowing the biosignal characteristics, as well as if a greater or lower sensitivity to variations is desired, when selecting a similarity measure. Moreover, these results justify the application of the proposed one.

Then, a methodology for the detection of ischemic episodes is presented and the rest of the chapter exploits the capacity of the developed predictive methodologies, together with the similarity measure, in the prediction of specific cardiovascular events. For this purpose, neural networks multi-models and wavelet decomposition are applied in the forecast of hypotensive episodes and in the estimation of hypertension risk.

4.1 Introduction

Integrated in the Heartcycle project, the final goal of the third workpackage (*WP3*), *Multi-parametric Analysis and Decision Support*, is to provide patients and professionals with the essential information for an optimal management of cardiovascular diseases, namely heart failure and coronary artery disease. The main topic of research is the development of models and algorithms able to assess relevant cardiovascular conditions, as well as globally evaluate patient's health status condition based on a multi-parametric analysis, involving the continuous update of measurements, parameters, and symptoms, collected during daily home monitoring process. From the clinical perspective, the development of such models will assist a decision support system in the conception of recommendations for therapy and care plans management. Moreover, they will be valuable to increase patient's safety through the prompt detection of those specific

cardiovascular conditions and in the generation of alerts in face of dangerous evolutions of important cardiovascular parameters.

The approach followed in the WP3–*Multi-parametric Analysis and Decision Support*, as well as its connections with other workpackages, is outlined in Figure 4.1.

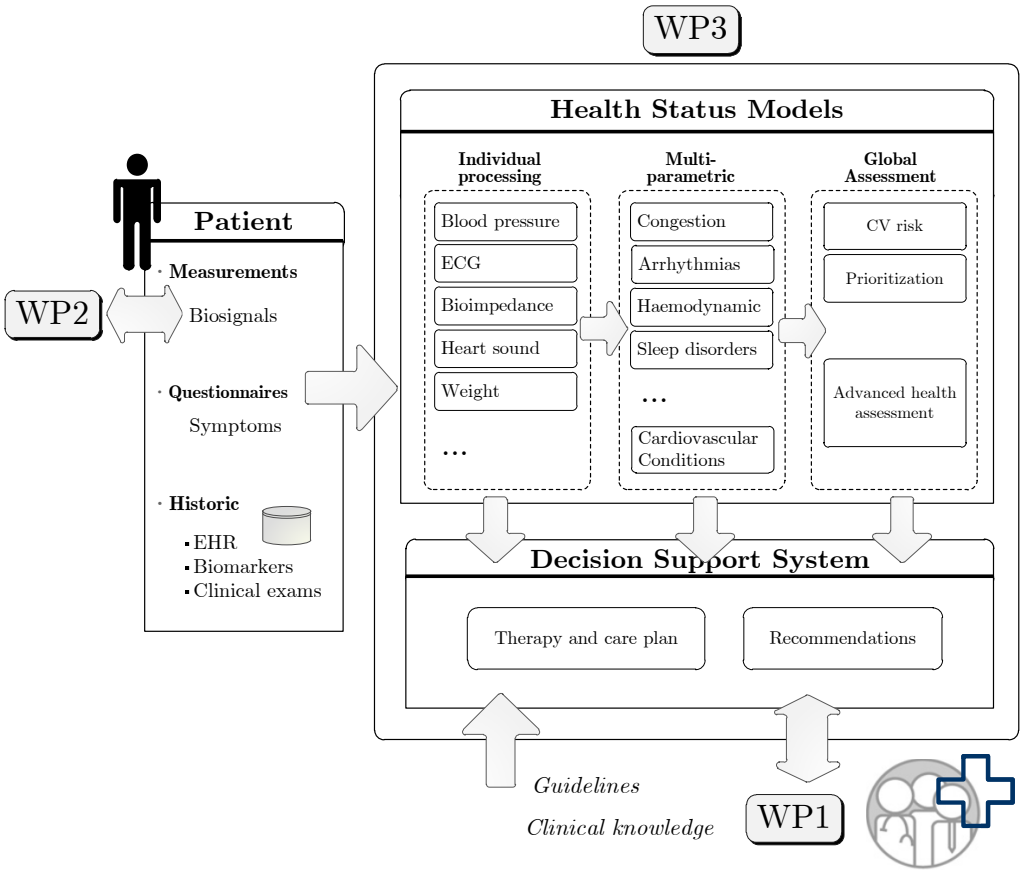


Figure 4.1 - Heartcycle-WP3: Multi-parametric Analysis and Decision Support System.

The WP3 module consists of two main components: *i*) Health status models, aiming at the development of algorithms for assessing the cardiac status of a patient and *ii*) Decision support system, responsible for the implementation of strategies describing the expert knowledge to support clinical decisions. The algorithms for cardiac health status are based on patient information, in particular on data regularly acquired by sensor devices provided by WP2–*Sensors and Parameter Extraction*. The decision support system includes results of these algorithms and clinical knowledge directly provided by WP1-*Coordination of the medical and technical expertise*.

4.1.1 Multi-parametric Analysis and Decision Support System

Decision support system

As a component of the decision support system (DSS), the *Therapy and Care plan* module provides suggestions for the management of therapies and care plans, namely recommendations for new prescriptions and indications for adjusting medication doses. These recommendations are based on patient data analysis, namely on cardiovascular conditions and advanced health status, in combination with explicit clinical knowledge (medical expertise), as well as with guidelines provided by WP1. In this way, the process of making recommendations related to treatment is under the control of clinical experts. Moreover, two kinds of indications are provided by the therapy and care plan modules: the first focus the professional, the second concentrate on the patient.

There is also a module *Recommendations*, which does not relate to treatment but supports further investigations. For example, in the case of arrhythmias, an implantable defibrillator can be recommended to the patient.

Health status models

The *Health Status Models* module aims the identification of specific parameters and conditions related with patient's cardiovascular health status. These include innovative health features, derived from advanced data analysis algorithms, which make use of patient databases and innovative sensors, major cardiovascular conditions, as well as high-level executive summary of patient's cardiovascular status.

A central part of the health status models, fundamental to the decision support system, consists of algorithms providing an *Advanced Health Assessment* of patient status. In effect, these algorithms offer an advanced form of decision support, given that they use, as inputs, high-level cardiovascular parameters and conditions statements, which are derived from different sources of information such as, daily measurements (electrocardiogram, blood pressure, heart sound, weight, etc), symptoms, biomarkers (cholesterol, BNP-B type natriuretic peptide, etc), and extracted parameters from biosignals analysis (heart rate, ST elevation, cardiac output, etc). In particular, WP2– *Sensors and Parameter Extraction*, is the workpackage responsible for providing solutions for sensing biosignals using textile and wearable sensors.

The health status models to develop may include individual analysis of biosignals, as well as other measurements. Moreover, several sources of information can be simultaneously processed in a multi-parametric analysis. In effect, three main procedures can be considered in the development of these models:

- **1.** *Individual processing*: where signals are independently analysed in order to extract relevant features. For example, deviations from a given threshold, to generate a simple alert, can be integrated here. Determination of the heart rate from the ECG, involving R peaks detection, is another example of such processing.

- **2.** *Multi-parametric analysis*: where the information coming from different sources (sensors, symptoms) is merged in order to define relevant clinical conditions/applications. Basically, it is assumed that these conditions can be characterized based on the analysis of biosignals and on their dynamic evolution. Examples of these are congestion, haemodynamic status, sleep disorders, pulmonary oedema, etc.
- **3.** *Global assessment models*: consider the combination of multi-parametric models (clinical conditions), as well as results from individual signals analysis, to summarize the clinical condition of a patient. Examples of such global assessment models are the advanced health assessment module, the cardiovascular risk assessment module, the prioritization of alerts/patients and the detection of worsening heart failure.

Clinical conditions/applications

In order to accomplish the advanced decision support system, several clinical applications, that is, cardiovascular conditions, centred on the patient needs and relevant for cardiac disease management, have been identified by WP1 (clinical expertise). These applications are supported on innovative sensors, including *i*) the bed sensor (ballistrocardiography signal) that provides information about the heart rate, respiration, and movements during patient sleep, to enable night arrhythmias and sleep disorders analysis; *ii*) the BIM sensor (bioimpedance monitor) for congestion detection and; *iii*) the Sensatron device (acquires heart sounds, photoplethymographic and electrocardiogram signals) for estimating cardiac output and system vascular resistance in the hemodynamic status application.

On the other hand, in addition to the innovative sensors, other clinical applications make use of further inputs, such as symptoms from questionnaires and data from electronic health record, as well as information related with standard sensors, namely ECG, weight, and blood pressure.

4.1.2 Specific Clinical Applications

Following this direction, the key challenge of the present thesis in the context of HeartCycle was the development of specific clinical applications based on the analysis of the ECG and blood pressure signals. These goals should be achieved by means of research on advanced algorithms for biosignal analysis, as well as on computational intelligence approaches for modelling and prediction purposes. The main effort was on time series similarity methodologies, in order to enable the detection of specific cardiovascular conditions, and on predictive techniques to be applied in the trend analysis of the biosignals and early detection of future events.

As result, the investigation conducted in this work was oriented in two main directions:

***i*) ECG analysis**: The development of techniques for similarity detection in biosignal time series, particularly in the ECG signal, which enable to detect specific cardiovascular conditions (e.g. ventricular arrhythmias, ischemic episodes and ventricular dyssynchrony). Methods that exploit the time-frequency characteristics of the signals were investigated and described in Chapter 2, and the corresponding results will be presented and discussed in section 4.2-Similarities Detection, and in section 4.3-Ischemia Detection.

***ii*) Trend analysis**: Development of predictive methods to be applied in the trend analysis of biosignals, in particular blood pressure, and consequent prediction of future hypertension risk. For this purpose, wavelet multi-resolution was researched in Chapter 3 and its application will be presented and discussed in section 4.5-Trend Prediction of BP Signals.

ECG analysis

In particular, some relevant cardiovascular conditions have been investigated in the context of heart failure (HF) and coronary artery disease (CAD): *i*) ventricular arrhythmias assessment; *ii*) dyssynchrony and heart block; *iii*) detection of ischemic events.

i. Ventricular arrhythmias

In the context of cardiovascular problems, the ventricular arrhythmias assume a very important role since they can lead to situations of severe complexity and risk. Particularly, ventricular fibrillation (VF) is potentially fatal, being considered the main cause of sudden cardiac death. Moreover, ventricular arrhythmias evolve from simple premature ventricular contractions, which are in most situations benign, to ventricular tachycardia (VT) and finally to critical ventricular fibrillation episodes. Therefore, the development of methodologies able to detect the occurrence of these arrhythmias is of extreme importance for the conception of early prevention systems.

The work proposed in this framework was centred on a non-linear dynamic signal processing approach. Based on the phase space reconstruction of the electrocardiogram (Rocha *et al.*, 2008), some features were extracted. Features from current and previous time instants were provided to a dynamic neural network classifier, enabling arrhythmias detection.

ii. Dyssynchrony and heart block

The main goal of this clinical application was to assess the severity of the heart dyssynchrony. Asynchronous condition has serious effects on ventricular pump function, leading to prolonged contraction period, thus to the reduction of the ejection fraction. If properly diagnosed, dyssynchrony can be treated (by means of cardiac resynchronization

therapy) to increase survival, as well as to improve symptoms in patients with congestive heart failure.

Using parameters extracted from the ECG, namely from the QRS complex analysis, in conjunction with impedance cardiography and heart sound signals (collected using Sensatron system), the severity of the heart dyssynchrony was addressed.

iii. Ischemic events

In patients of coronary artery disease, coronary arteries become narrowed by atherosclerosis, restricting the supply of blood and oxygen to the heart. Ischemia can be silent, without evidence of symptoms, or characterized by chest pain also known as angina pectoris. A severe and sudden blockage of coronary arteries causing a prolonged lack of blood supply to the heart, may lead to a myocardial infarction (heart attack) due to cellular necrosis. Moreover, myocardial ischemia is the pathological substrate to originate serious abnormal heart rhythms (arrhythmias), which can cause fainting or, frequently, sudden death. For the exposed reasons, its early diagnosis and treatment is of great importance to improve patient's health. In effect, if blood supply of heart muscle is timely re-established, myocardial ischemia can be reversed, cellular necrosis limited and all complications avoided.

For ischemia detection, a strategy was developed using the time-frequency analysis of the ECG to characterize ST segment deviation events, and Hermite basis functions for T wave and QRS complex morphology characterization.

The approach proposed for the detection of ischemic events is presented in section 4.3-Ischemia detection.

Trend analysis

High blood pressure or hypertension, is among the top most factors associated with cardiovascular diseases. In fact, uncontrolled and prolonged elevation of blood pressure can result in a multiplicity of alterations in the myocardial structure, coronary vasculature, and conduction system of the heart, which can lead to the development of left ventricular hypertrophy, coronary artery disease, myocardial infarction, cardiac arrhythmias, heart failure, among others (Riaz and Ahmed, 2012).

This module aims at the development of predictive methods to estimate biosignals future values. In particular, the research addresses the trend analysis, in order to predict if blood pressure signal of a given patient evolves towards hypertension values or, on the contrary, is maintaining or decreasing to normal values.

4.1.3 Detection of Acute Hypotensive Episodes

During the course of this thesis, in 2009, Physionet in cooperation with Computing and Cardiology proposed its annual challenge[1] which goal was to predict which patients in a specific dataset would experience an acute hypotensive episode (AHE) beginning within a forecast window. The fundamental purpose was the prediction of blood pressure signals, strongly related with the objectives of the present thesis (trend prediction of BP signals). However, according to the challenge definition of an AHE, future accurate values for BP signals were required. Therefore, to take into account these constraints (specific dataset and accurate values), a particular strategy based on a multi-model scheme involving neural networks was proposed. Its description was introduced in Chapter 3 and the corresponding results will be presented in section 4.4 - Prediction of Acute Hypotensive Episodes.

4.1.4 Implementation and Databases

All the algorithms, the accesses to the different datasets and respective validations were implemented using the Matlab software[2]. Moreover, for the implementation and training of support vector regression models the libsvm[3] framework was employed.

In order to test and validate the developed algorithms, public and private datasets were used. With respect to public datasets, there are available databases of ECG and other physiological recordings for distinct purposes (arrhythmias, ischemia, etc.). Of special relevance is the Physionet[4], possibly the largest and best organized source of information currently available. The private datasets are the result of the TEN-HMS study (Cleland *et al.*, 2005) and MyHeart project[5].

In particular, the following databases were used for validation purposes in the present thesis:

Ventricular arrhythmias

- **MIT-DB[6]**: The Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmias Database (48 records, 35 minutes each)
- **CU-DB[7]**: The Creighton University Sustained Ventricular Arrhythmia database (35 records, 8 minutes each)

[1] physionet.org/challenge/2009/
[2] www.mathworks.com/products/matlab/
[3] www.csie.ntu.edu.tw/~cjlin/libsvm/
[4] www.physionet.org
[5] www.hitech-projects.com/euprojects/myheart/
[6] www.physionet.org/physiobank/database/mitdb/
[7] physionet.org/physiobank/database/cudb/

Similarity measure analysis (section 4.2)

- **MIMIC-II**[8] - Multiparameter Intelligent Monitoring in Intensive Care, is a multi-parametric long-term dataset comprising ICU patients. It contains high temporal resolution data including lab results, electronic documentation, and bedside monitor trends and waveforms.
- **TEN-HMS** (Cleland *et al.*, 2005) - The Trans-European Network Homecare Monitoring Study (TEN-HMS) is a home telemonitoring consisting of self-measuring of weight, blood pressure, heart rate and rhythm.

Ischemic episodes (section 4.3)

- **ESC ST-T DB**[9]: The European Society of Cardiology ST-T database (90 records, 2 hours each)

Hypotension episodes (section 4.4)

- **MIMIC-II**[8]

Trend analysis of blood pressure (section 4.5)

- **TEN-HMS**
- **MyHeart**[5] - Telemonitoring system enabling to monitor vital body signs with wearable technology. This system was used in a clinical observational study carried out with 148 patients, daily measuring weight, blood pressure, heart rate, breathing rate and bioimpedance.

[8] mimic.physionet.org/

[9] www.physionet.org/physiobank/database/edb/

4.2 Similarity Measure Analysis

4.2.1 Introduction

One of the main goals of this work was the development of effective methods to evaluate the similarity between historic signals collected from a telemonitoring system (blood pressure, heart rate, respiration rate, etc.) and a particular template (e.g., sudden increase in blood pressure). On the other hand, the proposed measure should be able to deal with some aspects inherent to clinical environments, namely with noise artefacts, with signals that are not perfectly aligned in time and with signals that present similar trends (analogous behaviour).

The major objective of this section is to provide a quantitative comparison between some similarity measures, including the one proposed in Chapter 2. To this aim, specific variations in time and in amplitude were induced in the template, in order to evaluate the sensitivity of the referred measures to those different alterations (Lhermitte *et al.*, 2011). For this purpose, four similarity measures were considered: *i*) Euclidean distance, *ii*) linear correlation coefficient, *iii*) Fourier based measure and, finally, *iv*) Wavelet-KLT based similarity measure (developed in this thesis). Experimental and simulation results were conducted using: *i*) blood pressure signals collected in intensive care units (MIMIC-II dataset) and; *ii*) blood pressure and *iii*) heart rate signals collected by a telemonitoring system (TEN-HMS dataset).

The diversity in terms of the sensitivity of the several approaches, highlights the importance of understanding the biosignal characteristics before selecting a given similarity measure. In particular, the proposed similarity measure that combines the Haar wavelet with the Karhunen-Loève transforms, is especially adequate to deal with noisy signals, with signals that are not aligned in time, as well as with trends.

This section is organized as follows: firstly, in subsection 4.2.2, some similarity measures are presented and some specific time series variations are introduced. Then, in subsection 4.2.3, the performance of the similarity measures is assessed using the aforementioned datasets and, finally, in subsection 4.2.4, some conclusions are drawn.

4.2.2 Time Series Similarities and Variations

The capacity of a given similarity measure to detect variations in time series trajectories depends on several aspects. First of all, the times series behaviour, such as the temporal correlation between the data and seasonal components, influences the performance of any similarity detection strategy. This type of correlation occurs when observations in a time series are correlated with different temporal observations. In biosignals, this is typical of cyclical systems, which present a periodic behaviour. Examples of these are

the ECG and the circadian rhythms, in particular daily variations of blood pressure. On the other hand, the characteristics of a specific similarity measure (time domain approach, transform-based, etc.), as well as the template variations, decisively determine the similarity measure performance.

Time series similarity measures

The performance of four similarity measures is assessed: *i*) Euclidean distance, *ii*) linear correlation coefficient, *iii*) Fourier based measure and, *iv*) the proposed wavelet-KLT based similarity measure. These can be grouped in two major categories *i*) time-domain approaches, that are computed directly from the time series data (Euclidean distance, linear correlation coefficient), *ii*) transformed-based approaches, computed indirectly with parameters extracted from the original time series data (Fourier based and the proposed similarity measures).

i. Euclidean distance

Given two discrete time series of length N, $X_1(t) = \{x_1(1),\ ...,\ x_1(N)\}$ and $X_2(t) = \{x_2(1),\ ...,\ x_2(N)\}$, the Euclidean distance, $D_E \in \mathbb{R}_0^+$, between the two time series is defined based on their individual values $x_1(t)$ and $x_2(t)$ as:

$$D_E\Big(X_1(t), X_2(t)\Big) = \sqrt{\sum_{t=1}^{N}\big(x_1(t) - x_2(t)\big)^2} \tag{4.1}$$

The main advantage of the Euclidean distance is that it is easy to compute and to interpret.

ii. Linear correlation – Pearson's coefficient

The linear correlation coefficient, $CC \in \mathbb{R}$, sometimes referred as Pearson's correlation coefficient, quantifies the degree of linear relationship between two time series. It is a time-domain approach, computed by (4.2).

$$CC\Big(X_1(t), X_2(t)\Big) = \frac{\sum_{t=1}^{N}\Big(x_1(t) - \overline{X_1}\Big)\ \Big(x_2(t) - \overline{X_2}\Big)}{\sqrt{\sum_{t=1}^{N}(x_1(t) - \overline{X_1})^2}\ \sqrt{\sum_{t=1}^{N}(x_2(t) - \overline{X_2})^2}} \tag{4.2}$$

In the previous equation, $\overline{X_1}$ and $\overline{X_2}$ are, respectively, the average values of the time series $X_1(t)$ and $X_2(t)$, obtained as follows:

$$\overline{X_i} = \frac{1}{N}\sum_{t=1}^{N} x_i(t) \qquad i = 1,2$$

This correlation value (CC) is a scalar in the range $[-1, 1]$. In case of $CC = 1$, a perfect positive linear relationship (correlation) occurs; in case of $CC = -1$, a perfect negative linear relationship (anti-correlation) occurs; in case of $CC = 0$ (zero), there is no correlation between the signals (uncorrelated).

iii. Fourier based distance

The Fourier transform (FT) allows the decomposition of a time series into periodic signals in the frequency domain. If the signal is real (typical in most of the applications), half of the data is redundant. In effect, since in time domain the imaginary part of the signal is zero, in frequency domain both real and imaginary parts of the spectrum are symmetric. Discrete cosine transform (DCT), derived from discrete Fourier transform (DFT), generates the real spectrum of a real signal avoiding redundant data. Using DCT, a discrete time series of length N, $X(t) = \{x(1), \ldots, x(N)\}$, can be described by a set of scaled cosine waves with amplitude C_k, as defined in (4.3).

$$X(t) = \rho(t) \sum_{k=1}^{N} C_k \cos\left(\frac{\pi(2k-1)(t-1)}{2N}\right) \quad t = 1,\ldots,N \tag{4.3}$$

In the previous equation, the parameter $\rho(t)$ is a normalization coefficient defined by (4.4).

$$\rho(t) = \begin{cases} \frac{1}{\sqrt{N}} & t = 1 \\ \sqrt{\frac{2}{N}} & 2 \le t \le N \end{cases} \tag{4.4}$$

Based on the DCT coefficients (C_k, $k = 1,\ldots N$) a distance measure using the first m coefficients (the most relevant) can be implemented, by means of equation (4.5). The variables $C_k^{X_1}$ and $C_k^{X_2}$ represent, respectively, the DCT coefficients for the time series $X_1(t)$ and $X_2(t)$.

$$D_F\left(X_1(t), X_2(t)\right) = \sqrt{\sum_{k=1}^{m} (C_k^{X_1} - C_k^{X_2})^2} \tag{4.5}$$

Mathematically, the Fourier based distance is equivalent to the Euclidean distance in case all set of cosine waves are considered ($m = N$). Therefore, the difference between the distances D_F and D_E is a result of reducing the number of cosine waves ($m < N$). The main advantage of FT distance measure is the possibility to distinguish signals with a specific period. This is particularly relevant for detecting periodic patterns in biosignals, namely respiration cyclic signals. Due to the dimensionality reduction, the capacity to deal with noisy signals is another advantage recognized by the Fourier distance approach.

iv. Wavelet-KLT based distance

As previously referred in Chapter 2, the similarity measure proposed in this thesis is founded on the distance between the coefficients that result from the description of the signals being compared in terms of a reduced set of basis, determined by the wavelet+KLT procedure.

Firstly, a time series (template), $X_1(t)$, is described using a set of wavelet basis $\varphi_j(t)$, as (4.6).

$$X_1(t) = \sum_{j=1}^{J} \varphi_j(t) \tag{4.6}$$

Then, a second time series (to compare with the template), $X_2(t)$, is also described using the same set of basis $\varphi_j(t)$, as (4.7).

$$X_2(t) = \sum_{j=1}^{J} \alpha_j \ \varphi_j(t) \tag{4.7}$$

Considering those descriptions, equations (4.6) and (4.7), a simple distance measure can be established by using the corresponding coefficients, as (4.8).

$$D_W\left(X_1(t), X_2(t)\right) = \sqrt{\sum_{j=1}^{J} (1-\alpha_j)^2} \tag{4.8}$$

As in the Fourier decomposition, Wavelet-KLT based distance is equivalent to the Euclidean distance in case all set of basis are considered ($J = N$). One of the major advantages of the proposed distance measure is its interpretability. In fact, regardless of their exact value, if coefficients of equation (4.7) are positive ($\alpha_j > 0$), that means that signals $X_2(t)$ and $X_1(t)$ present the same behaviour, that is, the same evolution or trend, with respect to the particular base, $\varphi_j(t)$. In case of a negative value ($\alpha_j < 0$), it means that the signal and template have opposite trends, with respect to the specific base, $\varphi_j(t)$.

Similarity measures computed from distance measures

Equations (4.1), (4.5) and (4.8) define a distance between two time series. Based on those distances, normalized similarity measures, $S_i(\cdot)$, are defined as (4.9), with $i = E,F,W$ (respectively, Euclidean, Fourier and wavelet based cases).

$$S_i\Big(X_1(t), X_2(t)\Big) = e^{-D_i\left(X_1(t), X_2(t)\right)} \tag{4.9}$$

The similarity measure (S_i) is a scalar in the range [0, 1]. In case $D_i = 0$ (time series are equal), the similarity measure is $S_i = 1$. On the other hand, the similarity measure decreases as the distance increases (when $D_i \to \infty$, $S_i \to 0$).

In the case of the correlation coefficient, equation (4.2), a similarity measure in the range $[0,\ 1]$, can be directly implemented by means of equation (4.10).

$$S_C\Big(X_1(t),X_2(t)\Big)=\begin{cases} CC\Big(X_1(t),X_2(t)\Big) & if \quad CC\Big(X_1(t),X_2(t)\Big)>0 \\ 0 & if \quad CC\Big(X_1(t),X_2(t)\Big)\leq 0 \end{cases} \tag{4.10}$$

Based on this approach, all similarity measures are normalized in the range $[0,\ 1]$, therefore, possible to compare.

$$S_i\Big(X_1(t),X_2(t)\Big)\in[0,1] \qquad i=E,C,F,W \tag{4.11}$$

Template variations

An example of a biosignal template $X(t)$, designated as baseline, is illustrated in Figure 4.2. In this particular case, it is characterized by a peak shape behaviour, during about one month (more precisely, 32 days). It is important to note that this is just an example and that the shape of the template can assume any other form. In fact, it is randomly selected from the original time series.

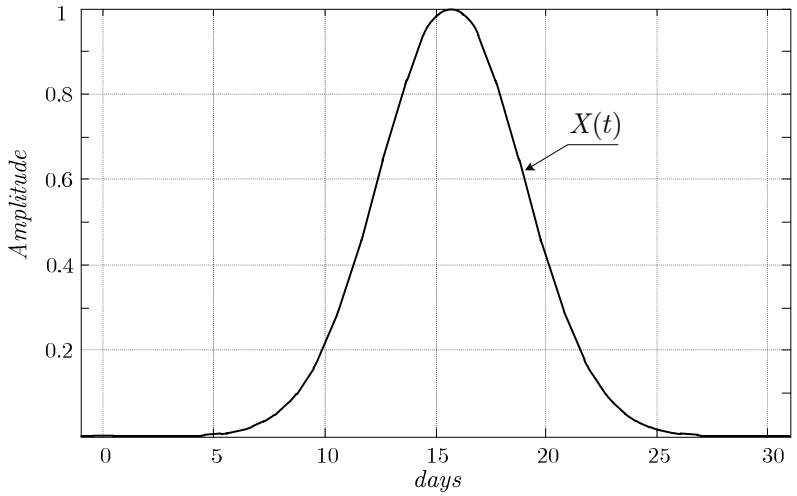


Figure 4.2 - Template or baseline.

Although several types of changes can affect the baseline, three main types are considered here: amplitude, time and shape variations.

1) **Amplitude variations** basically result in changes in the difference of amplitudes between the baseline, $X(t)$, and the modified time series, $Y(t)$. Amplitude scaling, $Y_{as}(t)$, occurs when the baseline peak is stretched or compressed along the y-axis (Figure 4.3a). Amplitude translation, $Y_{at}(t)$, occurs when time series is obtained by shifting the baseline along the y-axis (Figure 4.3b).

2) **Time effects** are the result of modifying the time location of the peak or the form of the baseline, along the time axis. For the baseline template, $X(t)$, illustrated in Figure 4.2, time scaling, $Y_{ts}(t)$, results from stretching or compressing, in time, the width of the peak (Figure 4.3*c*). Another typical variation is time translation, $Y_{tt}(t)$, occurring when the time series is shifted in time, without changing the baseline form (Figure 4.3*d*).

3) **Shape effects** take place when the behaviour of the baseline is changed. Here, two types of variations affecting the global shape of the baseline are considered. The first one, trends variation, $Y_{tr}(t)$, results in the alteration of the baseline trajectory (trend) (Figure 4.3*e*). The second, noise variation, $Y_{no}(t)$, consists in introducing a random disturbance in the baseline (Figure 4.3*f*)

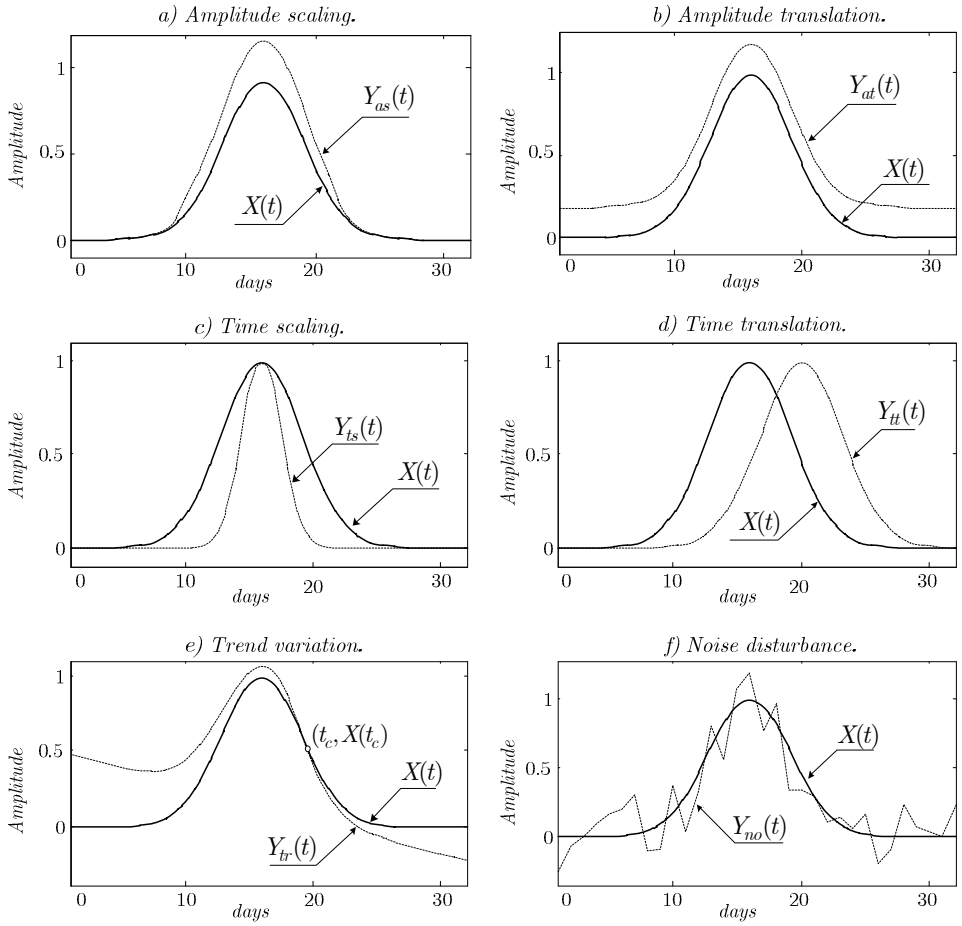


Figure 4.3 - Variations in the baseline.
a) amplitude scaling, b)amplitude translation, c) time scaling, d) time translation, e) trend variation, f) noise disturbance.

The several variations, $Y_{var}(t)$, with $var = \{as, at, ts, tt, tr, no\}$, illustrated in Figure 4.3, are generated from the baseline, $X(t)$, through equations (4.12) to (4.17).

$$Y_{as}(t) = X(t) + \upsilon_{as} B(t) \tag{4.12}$$

$$Y_{at}(t) = X(t) + \upsilon_{at} \tag{4.13}$$

$$Y_{ts}(t) = X(\upsilon_{ts} t - \tau) \tag{4.14}$$

$$Y_{tt}(t) = X(t - \upsilon_{tt}) \tag{4.15}$$

$$Y_{tr}(t) = rotate(X(t),\ \upsilon_{tr},\ t_c) \tag{4.16}$$

$$Y_{no}(t) = X(t) + \zeta(t) \tag{4.17}$$

The parameters υ_{as}, υ_{at}, υ_{ts}, υ_{tt}, υ_{tr} are scalars, the variables τ and t_c are time instants and the variable $\zeta(t)$ denotes a white noise signal, with normal distribution $N(\mu, \sigma^2) = N(0,1)$.

The amplitude scaling, equation (4.12), is obtained by adding to the baseline, a similar signal $B(t)$ (peak shape) affected by a factor υ_{as}, where υ_{as} determines how strong the scaling effect is. The amplitude translation, (4.13), is straightforwardly achieved by introducing an offset value, υ_{at}. The time scaling, equation (4.14), compresses the baseline by a factor υ_{ts}, being the signal centred on $t = \tau$, and the time translation, equation (4.15), introduces a time shift υ_{tt} to the baseline. The trend variation, equation (4.16), is obtained by applying a $rotate(\cdot)$ operator. This rotates the baseline, $X(t)$, υ_{tr} degrees, around a random point $(t_c, X(t_c))$. Finally, the signal disturbed by noise, equation (4.17), results from introducing random noise into the observations.

4.2.3 Results

Datasets

To quantify the sensitivity of the four similarity measures, three different types of signals belonging to two datasets were used.

Firstly, arterial blood pressure signals available in the context of the Physionet/Computers in Cardiology Challenge 2009 were employed. These records, included in the MIMIC-II database[8], are composed of 110 signals, each one with 10 hours of duration. Template signals, $X(t)$, with a duration of $N = 64$ minutes (approximately one hour), were randomly selected from each one of the 110 available blood pressure signals and subsequently modified according to the equations (4.12)-(4.17). These signals will be designated as BP-M (Blood Pressure – MIMIC).

From the TEN-HMS dataset (Cleland *et al.*, 2005), that consists of signals daily acquired during a telemonitoring study, two signals were considered: systolic blood pressure and heart rate. Template signals, $X(t)$, with a duration of $N = 32$ days

(approximately one month), were randomly selected from each one of the available signals (50 records corresponding to 50 patients) and modified according to the equations (4.12)-(4.17). These signals will be designated as BP-T, HR-T, respectively for Blood Pressure and Heart Rate (TEN-HMS dataset).

Baseline variations

The sensitivity of the four similarity measures was assessed in the presence of the referred variations υ_{var}, $var = \{as, at, ts, tt, tr, no\}$. The amount of variations introduced depended on the parameters υ_{var}, as detailed in Table 4.1. Moreover, in order to obtain a similar effect in the different amplitude variations, scaling parameters υ_{var} were normalized by a factor $range(\cdot)$. This factor was computed for each specific baseline, as defined in (4.18).

$$range\Big(X(t)\Big) = max\Big(X(t)\Big) - min\Big(X(t)\Big) \tag{4.18}$$

For each group of variations the same number of elements (21) was considered. As an example, the notation $\upsilon_{as} = [0 : 0.1 : 2]$, shown in Table 4.1, means that the minimum and maximum value of υ_{as} is, respectively, 0 and 2, with an increment of 0.1 (a total of 21 elements).

Table 4.1

Variations of the parameters in the different experiments.

	BP-M, BP-T, HT-T	Description
Amplitude scaling	$\upsilon_{as} = [0 : 0.1 : 2] \times range$	υ_{as} : scaling effect
Amplitude translation	$\upsilon_{at} = [0 : 0.1 : 2] \times range$	υ_{at} : scaling effect
Time scaling	$\upsilon_{ts} = [1 : 0.15 : 4]$; $\tau = \dfrac{N}{2}$	υ_{ts} : scaling effect N : length of baseline $X(t)$
Time translation	$\upsilon_{tt} = [0 : 1 : 20]$	υ_{tt} : time delay
Trend variation	$\upsilon_{tr} = [0 : 1.25 : 25]^{o}$; $t_c \in [0, N]$	υ_{tr} : rotation angle (degrees) t_c : random instant
Noise disturbance	$\upsilon_{no} = [0 : 0.05 : 1] \times range$	υ_{no} : standard deviation White noise: $N(\mu, \sigma^2) = N(0, \upsilon_{no}^2)$

Similarity measures

The similarity measures based on the Euclidean distance and on the linear correlation, were implemented as defined in (4.1) and (4.2).

On the other hand, for the Fourier based similarity measure, the first $m=4$ coefficients ($C_k^{X_1}$ and $C_k^{X_2}$ $k=1,...,4$) were considered, equation (4.5) (corresponding approximately to an accuracy of 90%). With respect to the number of wavelet basis, $\varphi_j(t)$, equations (4.6) and (4.7), a pre-defined accuracy of $\varepsilon=0.92$ was established. As result, in average, the number of basis used to represent the baseline, $X(t)$, was $J=4$.

Assessment of the similarity measures

Figure 4.4, Figure 4.5, and Figure 4.6, illustrate the effect of amplitude, time and shape variations on the similarity measures S_i, $i=E,C,F,W$ (Euclidean, Correlation, Fourier and Wavelet), respectively for the BP-M, BP-T and HR-T time series.

Based on the obtained results, it can be concluded that similarity values are very close when computed using the three different types of signals. In fact, for the three cases that were investigated (blood pressure-intensive care unit, blood pressure-telemonitoring system and heart rate-telemonitoring system), the evolutions of the four similarity measures, when considering the referred baseline variations, are analogous.

i. BP-M: Blood pressure-MIMIC dataset

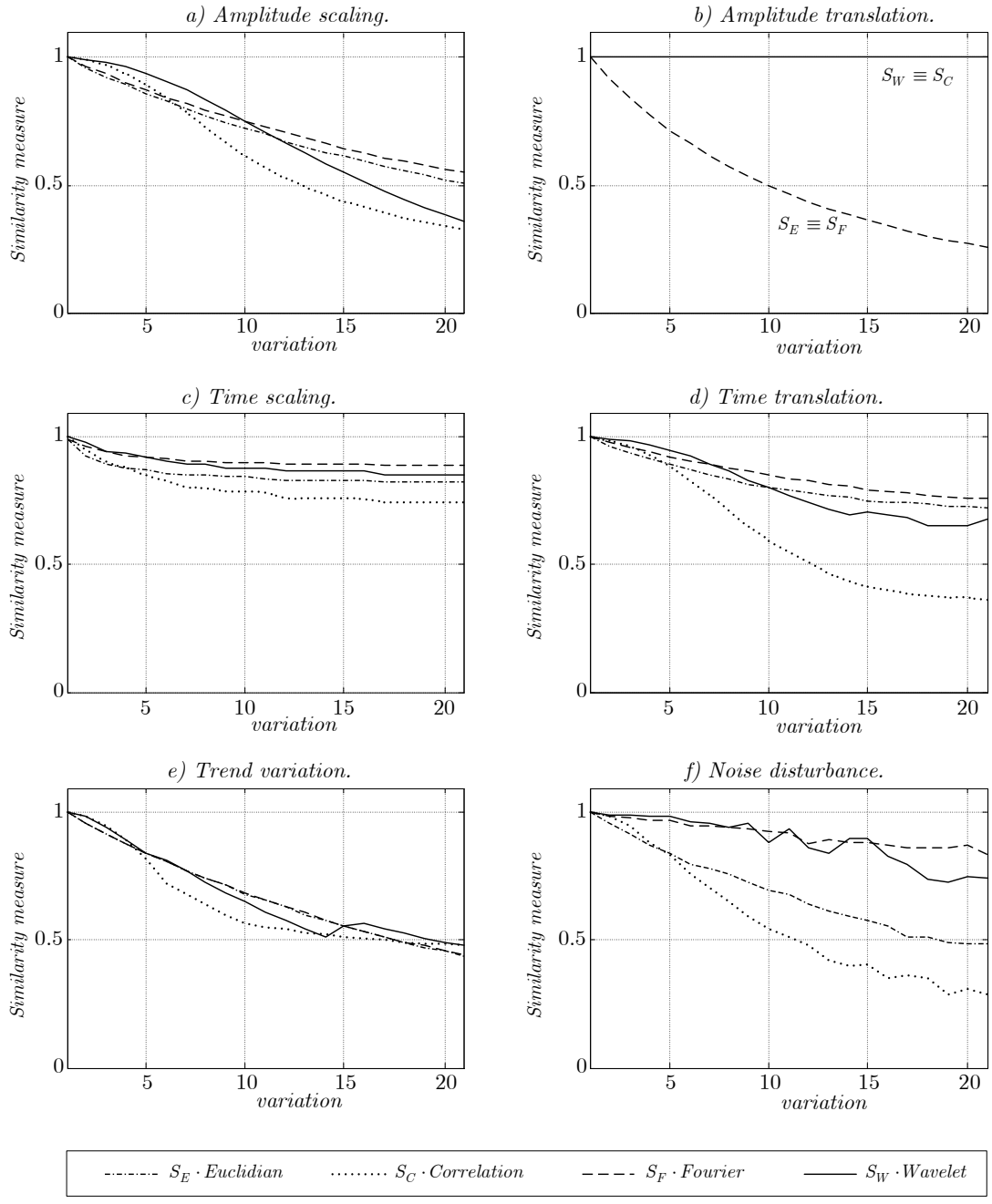


Figure 4.4 - Blood Pressure (MIMIC): Effect of baseline variation on similarity measures (Euclidean, Correlation, Fourier and Wavelet-KLT).

a) Amplitude scaling, b) Amplitude translation, c) Time scaling, d) Time translation, e) Trend variation, f) Noise disturbance.

ii. BP-T: blood pressure-TEN-HMS dataset

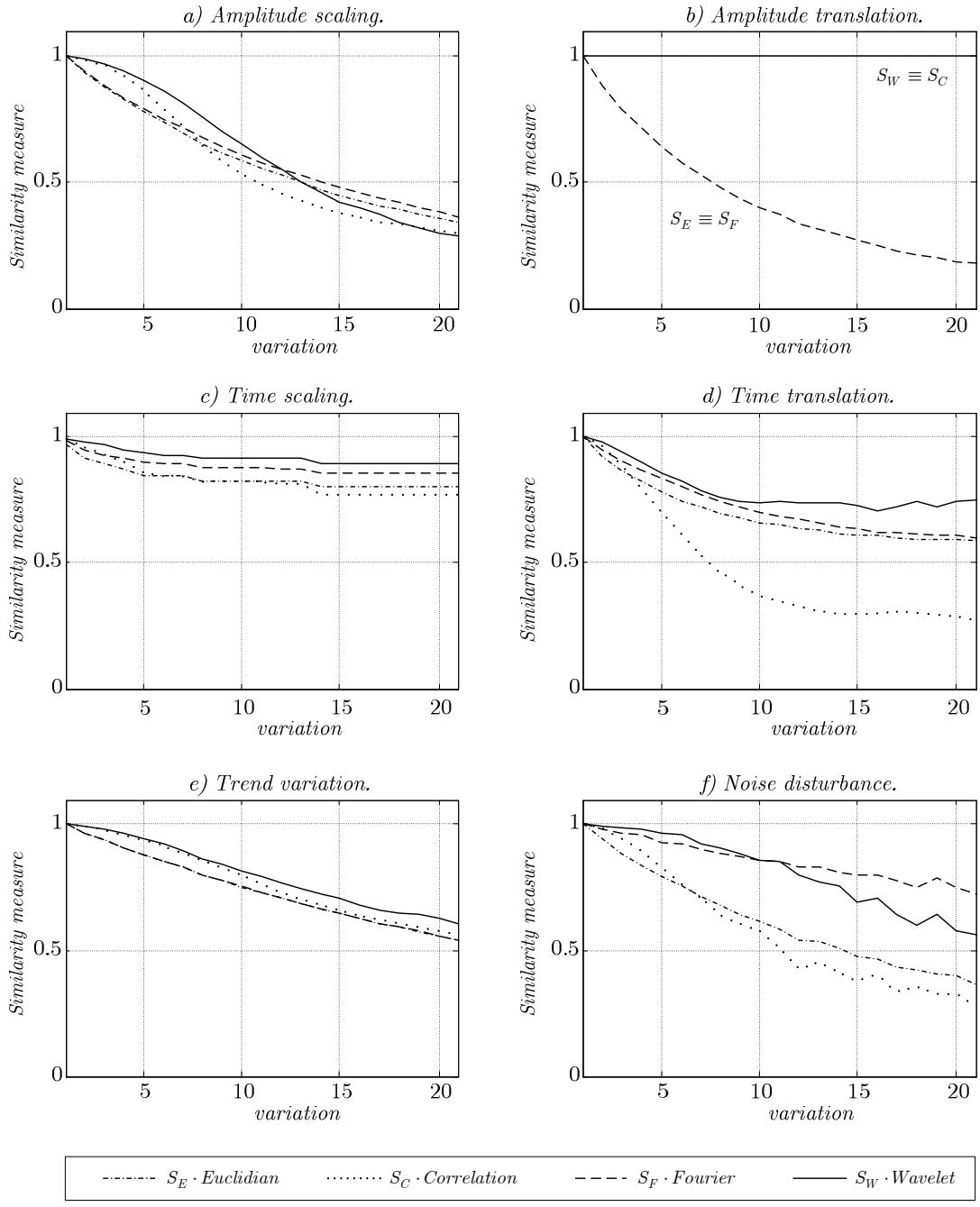


Figure 4.5 - Blood Pressure (TEN-HMS): Effect of baseline variation on similarity measures (Euclidean, Correlation, Fourier and Wavelet-KLT).

a) Amplitude scaling, b) Amplitude translation, c) Time scaling, d) Time translation, e) Trend variation, f) Noise disturbance.

iii. HR-T: Heart rate-TEN-HMS dataset

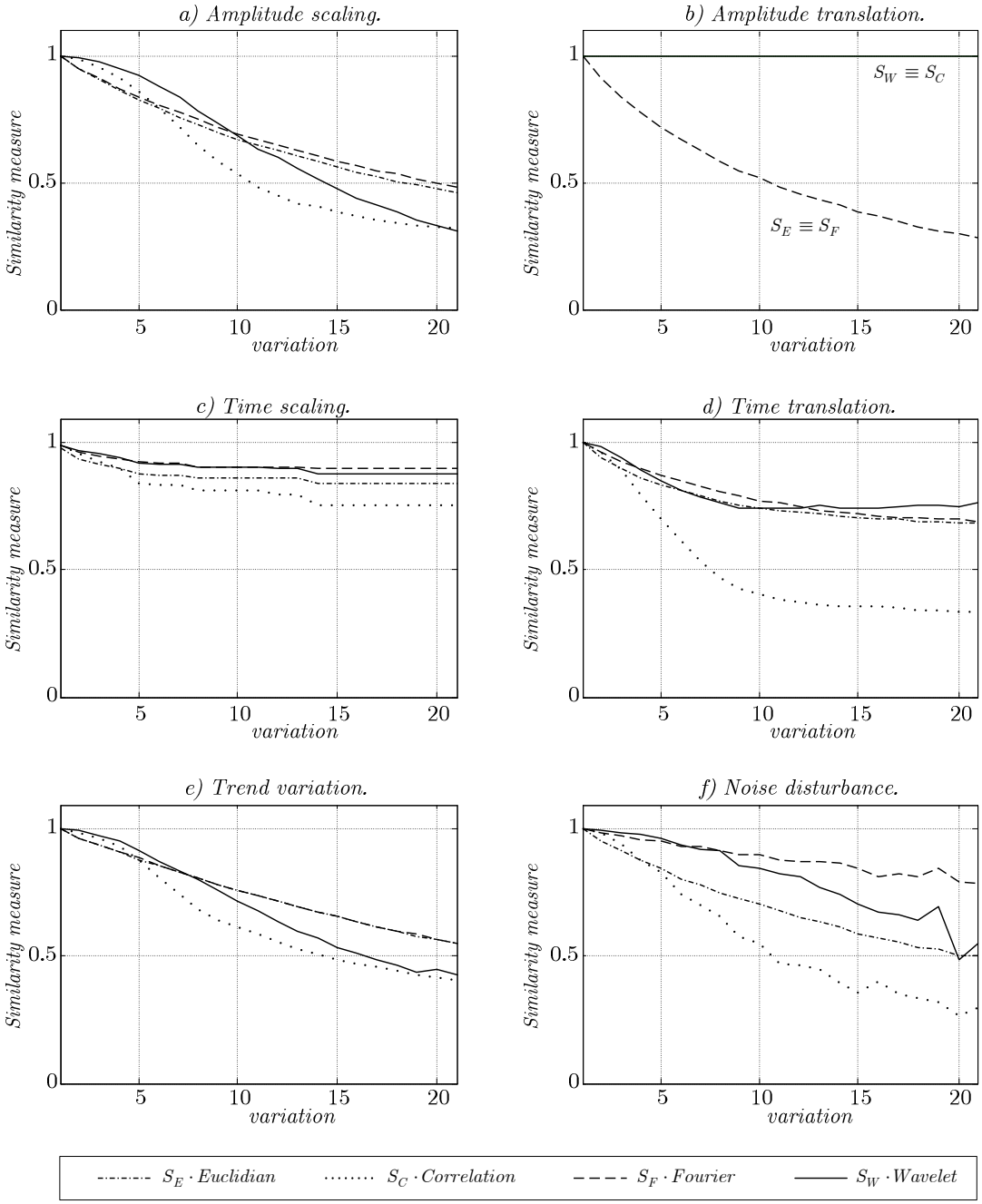


Figure 4.6 - Heart rate (TEN-HMS): Effect of baseline variation on similarity measures (Euclidean, Correlation, Fourier and Wavelet-KLT).

a) amplitude scaling, b)amplitude translation, c) time scaling, d) time translation, e) trend variation, f) noise disturbance.

Moreover, as expected, it can be observed that similarity values decrease when the level of baseline variation is increased. However, also as expected, the sensitivity to these changes is not equivalent for all the similarity measures: *i*) Euclidean based, *ii*) Linear correlation based, *iii*) Fourier based and *iv*) Wavelet-KLT decomposition based.

i. Euclidean distance based similarity measure: S_E

The Euclidean distance based similarity measure, S_E, is sensitive to all variations introduced in the baseline, exhibiting a global decrease as the degree of variations increases. That can be justified by the fact that S_E directly quantifies the difference between the raw data (baseline and the respective variations). As result, Euclidean based similarity measure reveals a higher sensitivity to all the baseline modifications.

ii. Correlation based similarity measure: S_C

The similarity measure S_C, based on the linear correlation, quantifies the temporal relationship and does not account for the difference between time series. Therefore, it reveals the expected lower sensitivity when amplitude effects are introduced. Firstly, amplitude translations do not affect correlation similarity (Figures 4.4*b*, 4.5*b*, 4.6*b*, always equal to 1). On the other hand, for the amplitude scaling case (Figures 4.4*a*, 4.5*a*, 4.6*a*), correlation measure presents lower sensitivities when facing small variations in the baseline. However, for larger values of amplitude scaling variations, the decreasing in the similarity measure is visible. In case of time scaling, time translation or noise effects, the correlation based similarity measure is very sensitive (when compared with the other methods). This is particularly noticeable for noise disturbances (Figures 4.4*f*, 4.5*f*, 4.6*f*) and time translations (Figures 4.4*d*, 4.5*d*, 4.6*d*). The last introduce a lag between the original baseline and corresponding modified time series values. As result, correlation between them is severely affected, justifying the low similarity values observed.

iii. Fourier based similarity measure: S_F

The Fourier based similarity measure, S_F, is, in general, sensitive to all introduced changes. Moreover, excluding noise disturbances, its results are analogous to S_E, with a comparable sensitivity to amplitude scaling (Figures 4.4*a*, 4.5*a*, 4.6*a*), amplitude translations (Figures 4.4*b*, 4.5*b*, 4.6*b*) and trend variation (Figures 4.4*e*, 4.5*e*, 4.6*e*). However, it presents a slight lower sensitivity to time scaling (Figures 4.4*c*, 4.5*c*, 4.6*c*) and time translation (Figures 4.4*d*, 4.5*d*, 4.6*d*). Furthermore, as expected, it exhibits a lower sensitivity to noise disturbances (Figures 4.4*f*, 4.5*f*, 4.6*f*). In effect, time series description as Fourier series permits noise reduction by considering a small set of cosine basis (on the assumption that noise is only contained in the high frequency Fourier components).

iv. Wavelet-KLT based similarity measure: S_W

The Wavelet-KLT based similarity measure is identical to S_E when all the basis are included. Consequently, the differences between S_W and S_E are the result of reducing the number of basis (J) considered in the time series description, equation (4.8).

The S_W measure shows, in general, the lowest sensitivity values, especially when small variations are introduced in the baseline. As the level of variations increases, the S_W measure tends to present a behaviour comparable to the other similarity measures. These aspects (low sensitivity for low levels of variations and high sensitivity for high level of variations) are particularly observed in case of amplitude scaling (Figures 4.4*a*, 4.5*a*, 4.6*a*), time translation (Figures 4.4*d*, 4.5*d*, 4.6*d*), trend variation (Figures 4.4*e*, 4.5*e*, 4.6*e*) and noise disturbances (Figures 4.4*f*, 4.5*f*, 4.6*f*).

The robustness (low sensitivity) of the S_W measure for small degree variations, reveals the importance of an adequate selection of the number of wavelet basis. In effect, by means of the wavelet-KLT decomposition scheme, the main characteristics of the baseline are captured employing a reduced set of basis. When variations of small magnitude are introduced in the baseline, the sensitivity of wavelet-KLT is still low, since part of the variability due to the introduced effects is still captured by the same basis.

Additionally, in the same way as Fourier decomposition, wavelet-KLT scheme presents low sensitivity to noise disturbances, given the possibility to approximate the baseline by a small set of wavelet basis.

The results achieved for the specific similarity measures, clearly show the importance of understanding the time series characteristics before selecting any specific similarity measure. Effectively, the dynamics of the particular template (baseline), as well as the possible level of variations affecting this template, definitely determine the value of the similarity measure. Therefore, it is not reasonable to conclude or claim that a similarity measure is better than the others. In effect, if the objective is to detect small level of baseline variations, high sensitivity measures are desirable; on the other hand, if the goal is to have robust similarity measures (less affected by baseline variations), low sensitivity measures are more adequate.

The proposed wavelet-KLT similarity measure

Given the previous results, and in the context of the present work (signals acquired by telemonitoring systems), the proposed wavelet-KLT similarity measure is justified. In effect, it is particularly appropriate to assure low sensitivity values in the presence of small levels of baseline variations. As result, it shows an adequate behaviour to deal with the evaluation of similarity between historic signals and a particular template (baseline). Basically, if the signals are analogous or are almost analogous, the similarity measure presents comparable results. Otherwise (a large variation occurs in relation to the template) the similarity measure tends to considerably decrease.

As mentioned before, one of the characteristics of the proposed similarity measure is its interpretability and simplicity. In fact, regardless the exact values of the coefficients α_j (that result from the wavelet-KLT decomposition), it is possible to straightforwardly assess the similarity between two time series. By definition, for the baseline $X(t)$, these coefficients are fixed and equal to one, that is, $\alpha_j = 1$. The signal $Y(t)$, obtained by

introducing variations in the baseline, is described using the same set of wavelet basis, weighted by coefficients α_j . As result, it can be stated that:

- **1**. In case all coefficients are positive ($\alpha_j > 0, \ \forall_j$), thus having the same sign as the baseline coefficients (all equal to 1), the two signals present the same behaviour (trend).
- **2**. In case one of the coefficients is negative ($\alpha_j < 0$), it means that the two signals present opposite trends for the corresponding basis (φ_j).

Figure 4.7 illustrates this idea. For all the experiments with the three types of signals (BP-M, BP-T and HR-T, resulting in 110+50+50=210 experiments), the average number of coefficients that present negative sign is depicted. These results correspond to the six variations considered (amplitude scaling, amplitude translation, time scaling, time translation, trend variation, and noise disturbance).

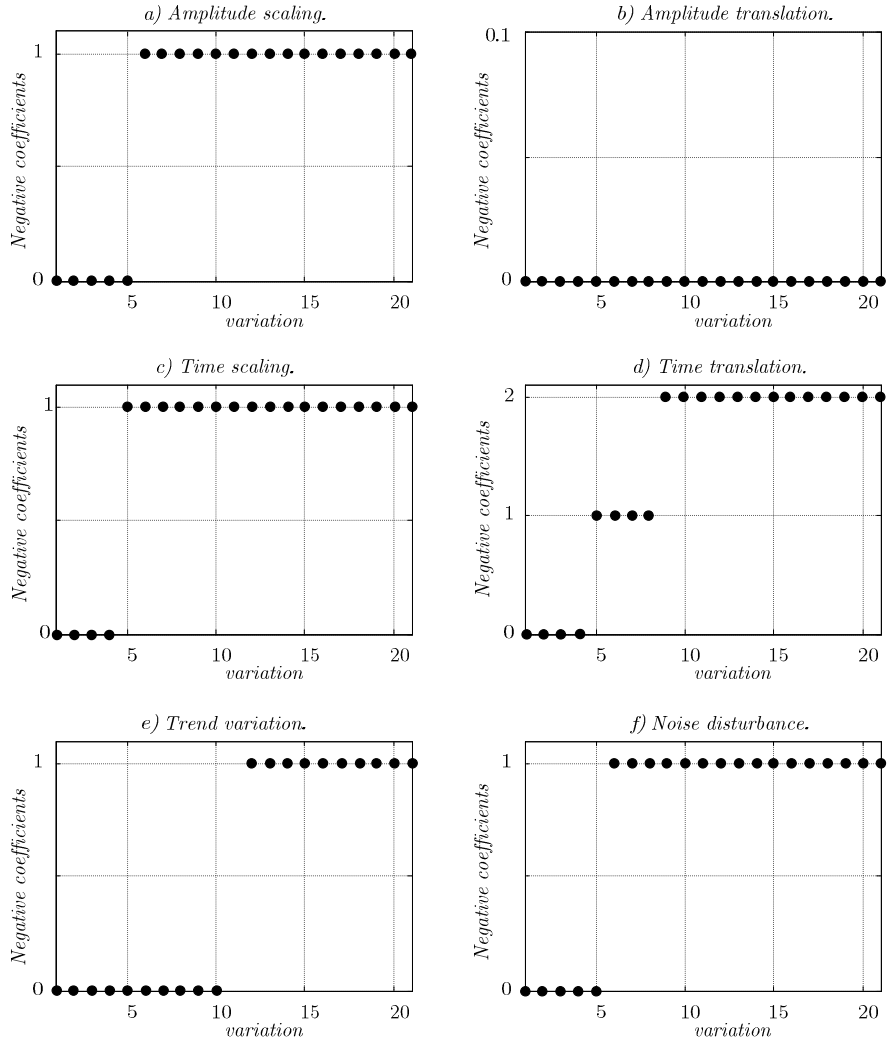


Figure 4.7 - Negative coefficients present in the wavelet representation of modified baseline.
a)amplitude scaling, b)amplitude translation, c)time scaling, d)time translation, e)trend variation, f)noise disturbance.

From the figure, it may be concluded that, with the exception of amplitude translation, there are no negative coefficients when small variations are introduced in the baseline. As the level of variations increases, the number of coefficients with negative value (reflecting an opposite behaviour, when compared with the baseline) increases. For example, for the first four variations in the time scale, $\upsilon_{ts}(i),\ \ i=1,...,4$, the baseline and the modified baseline exhibit coefficients with the same sign. As result, it can be stated that they present the same behaviour in this range of variation. However, for a level of variation higher than the fifth, $\upsilon_{ts}(i)$, $\mathrm{i}\geq 5$, there exists a negative coefficient. Therefore, in the time region corresponding to the basis that has a negative coefficient, the behaviour of the baseline and of the modified baseline, are opposite.

Thus, by means of this approach (sign of coefficients), a binary similarity strategy (yes or no) can be implemented. Basically, it depends on the number of wavelet basis used to describe the template (baseline), that is essentially determined by the parameter ε , subsection 2.5.4, equations (2.76) and (2.77). Nevertheless, if the objective is to quantify the similarity measure, a quantitative measure has to be used (equations (4.8) and (4.9)) and compared with a predefined threshold η , subsection 2.5.5, equation 2.105.

The average values ($\overline{S}_{var}$, $var=\{as,at,ts,tt,tr,no\}$) obtained by the wavelet-KLT similarity measure, for the three types of signals (210 experiments) are illustrated in the Figure 4.8 and detailed in Table 4.2. Additionally, the associated 95% confidence intervals ($\pm\mu$), indicating the reliability of the estimated values, are also presented. In this table only the even similarity values are shown ($\upsilon(i),\ \ i=0,2,4,..,20$).

Table 4.2

Average values of wavelet-KLT similarity measure and respective 95% confidence intervals.

		υ**(2)**	υ**(4)**	υ**(6)**	υ**(8)**	υ**(10)**	υ**(12)**	υ**(14)**	υ**(16)**	υ**(18)**	υ**(20)**
Amplitude scaling	$\overline{S_{as}}-\mu$	0.99	0.95	0.87	0.77	0.67	0.57	0.48	0.41	0.35	0.29
	$\overline{S_{as}}$	**0.99**	**0.96**	**0.88**	**0.79**	**0.70**	**0.61**	**0.52**	**0.45**	**0.39**	**0.34**
	$\overline{S_{as}}+\mu$	1.00	0.97	0.90	0.81	0.72	0.64	0.56	0.49	0.44	0.39
Amplitude translation	$\overline{S_{at}}-\mu$	nd	nd	Nd	nd	nd	nd	nd	nd	nd	nd
	$\overline{S_{at}}$	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	$\overline{S_{at}}+\mu$	nd	nd	Nd	nd	nd	nd	nd	nd	nd	nd
Time scaling	$\overline{S_{ts}}-\mu$	0.98	0.94	0.90	0.89	0.88	0.86	0.84	0.84	0.83	0.83
	$\overline{S_{ts}}$	**0.98**	**0.95**	**0.92**	**0.91**	**0.90**	**0.90**	**0.88**	**0.88**	**0.88**	**0.88**
	$\overline{S_{ts}}+\mu$	0.98	0.96	0.94	0.93	0.93	0.93	0.92	0.92	0.92	0.92
Time translation	$\overline{S_{tt}}-\mu$	0.99	0.91	0.84	0.78	0.74	0.70	0.69	0.67	0.67	0.67
	$\overline{S_{tt}}$	**0.99**	**0.92**	**0.86**	**0.80**	**0.76**	**0.74**	**0.73**	**0.72**	**0.72**	**0.71**
	$\overline{S_{tt}}+\mu$	0.99	0.93	0.87	0.82	0.79	0.77	0.76	0.76	0.76	0.76
Trend variation	$\overline{S_{tr}}-\mu$	0.99	0.93	0.85	0.78	0.70	0.63	0.56	0.54	0.50	0.48
	$\overline{S_{tr}}$	**0.99**	**0.94**	**0.87**	**0.80**	**0.73**	**0.66**	**0.60**	**0.59**	**0.55**	**0.52**
	$\overline{S_{tr}}+\mu$	1.00	0.95	0.89	0.82	0.75	0.70	0.64	0.63	0.59	0.57
Noise disturbance	$\overline{S_{no}}-\mu$	0.99	0.97	0.94	0.90	0.84	0.79	0.76	0.70	0.62	0.56
	$\overline{S_{no}}$	**1.00**	**0.98**	**0.95**	**0.92**	**0.87**	**0.83**	**0.80**	**0.74**	**0.66**	**0.60**
	$\overline{S_{no}}+\mu$	1.00	0.99	0.97	0.94	0.89	0.86	0.84	0.78	0.71	0.65

nd – Not defined.

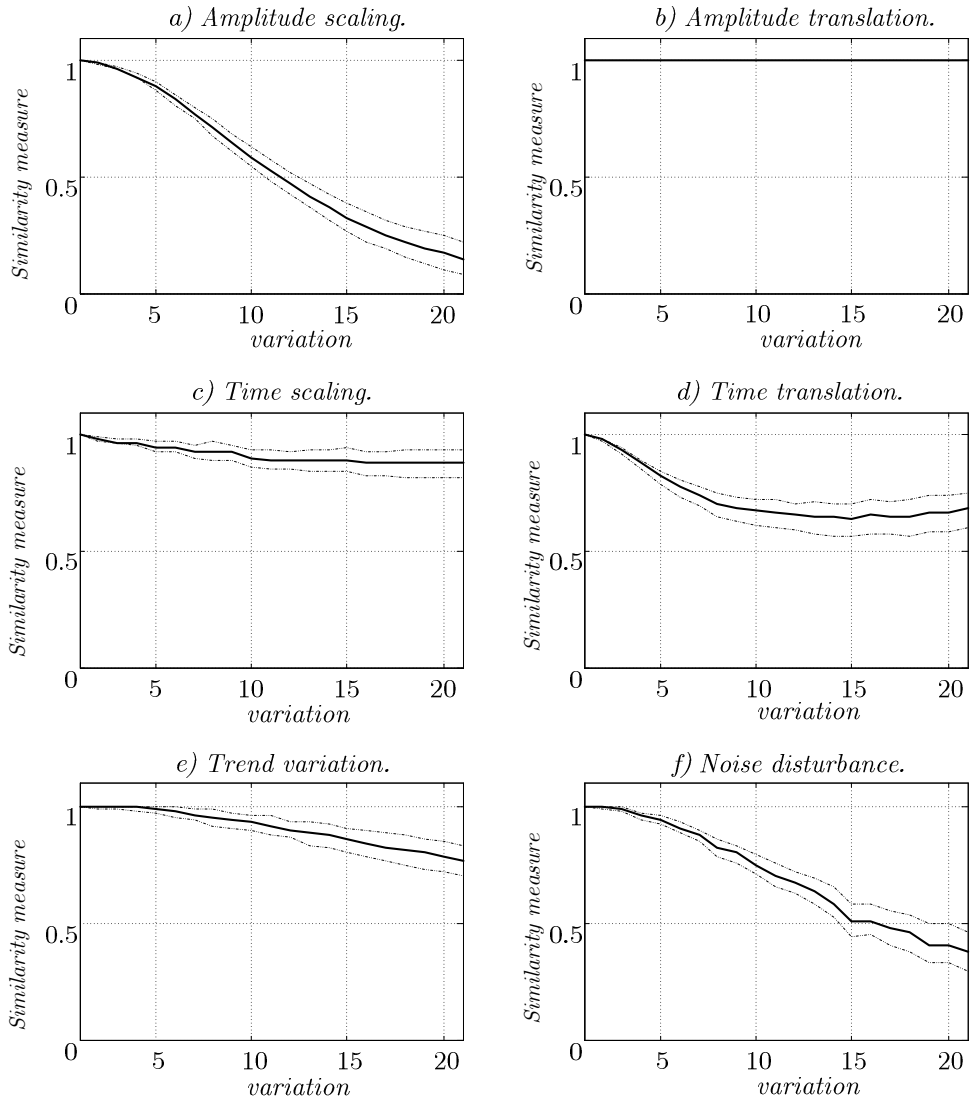


Figure 4.8 - Average values of wavelet-KLT similarity measure and respective 95% confidence intervals. a) amplitude scaling, b)amplitude translation, c) time scaling, d) time translation, e) trend variation, f) noise disturbance.

4.2.4 Conclusions

The major objective of this section was to provide a quantitative comparison of some common similarity measures. In particular, the results of four similarity measures were compared: *i*) Euclidean distance, *ii*) linear correlation coefficient, *iii*) Fourier based measure and *iv*) Wavelet-KLT based similarity measure (the measure proposed in this thesis).

To evaluate the sensitivity of the referred measures when applied under distinct circumstances, specific variations in time and in amplitude were induced in a template. The results showed that those measures present different sensitivities to the induced variations. In particular, the proposed similarity measure that combines the Haar wavelet with the Karhunen-Loève transforms, is especially adequate to deal with noisy signals, with signals that are not aligned in time, as well as with trends.

4.3 Ischemia Detection

In this section a strategy for the automatic detection of ischemic episodes is presented. A measure for ST deviation based on the time-frequency analysis of the ECG and the use of a reduced set of Hermite basis functions for T wave and QRS complex morphology characterization, are the key points of the proposed methodology.

Usually ischemia manifests itself in the ECG signal by ST segment deviation or by QRS complex and T wave changes in morphology. These effects might occur simultaneously. Time-frequency methods are especially adequate for the detection of small transient characteristics hidden in the ECG, such as ST segment alterations. A Wigner-Ville transform based approach is proposed to estimate the ST shift. To characterize the alterations in the T wave and QRS morphologies, each cardiac beat is described by expansions in Hermite functions. These demonstrated to be suitable to capture the most relevant morphologic characteristics of the signal. A lead dependent neural network classifier considers, as inputs, the ST segment deviation and the Hermite expansion coefficients. The ability of the proposed method in ischemia episodes detection is evaluated using the European Society of Cardiology ST-T database. A sensitivity of 96.7% and a positive predictivity of 96.2%, reveal the capacity of the proposed strategy to perform ischemic episodes identification.

This section is organized as follows: the first subsection, 4.3.1, introduces the problem and its clinical relevance. In subsection 4.3.2, the proposed methodology is described, in subsection 4.3.3, validation results using the European Society of Cardiology (ESC) ST-T database are presented, and, finally, in subsection 4.3.4, some conclusions are drawn.

4.3.1 Introduction

The World Health Organization estimates that 17.3 million people died of cardiovascular diseases in 2008, representing 30% of all global deaths. Out of these, 7.3 million were due to coronary artery disease (CAD) (WHO, 2011)[10]. As one of the leading causes of death worldwide, this cardiovascular condition represents a focus of international interest. On the other hand, the use of new monitoring technologies and specialized processing based on wearable information technology, provide professionals with the adequate information for the evaluation of cardiovascular conditions and symptoms progression, enabling the early detection of forthcoming clinical severe conditions (Reiter and Maglaveras, 2009).

This section focuses coronary artery disease and, in particular, the development of algorithms for myocardial ischemia detection.

[10] www.who.int/mediacentre/factsheets/fs317/en/index.html

In CAD, coronary arteries become narrowed by atherosclerosis, restricting the supply of blood and oxygen to the heart. This deprivation may originate a cardiac disorder called myocardial ischemia, which can be silent, without evidence of symptoms, or it might be characterized by chest pain also known as angina pectoris. A severe and sudden blockage of coronary arteries causing a prolonged lack of blood supply to the heart, may lead to a myocardial infarction due to cellular necrosis. Moreover, myocardial ischemia is the pathological substrate to originate serious abnormal heart rhythms (arrhythmias), which can cause fainting or frequently sudden death. Hence, it is observed that early diagnosis and treatment of CAD is of primary importance to avoid serious consequences for patient's health, treatment success and quality of life. In effect, if blood supply of the heart muscle is timely re-established, myocardial ischemia can be reversed, cellular necrosis limited and severe complications avoided.

The analysis of the electrocardiogram's (ECG) characteristics, namely the ST segment deviation as well as the QRS complex and T wave morphologies, are determinant for accurate detection of ischemic episodes (Barill, 2003).

The automatic diagnosis of myocardial ischemia based on the ECG signal usually involves two phases: ischemic beat classification and ischemic episode identification. In the first phase, each cardiac beat is labelled as normal or ischemic and, in the second phase, sequential ischemic beats are appropriately grouped in order to identify ischemic episodes.

In the context of ischemic beat detection and ischemic episodes identification using the ECG, several methodologies have been developed. Time, frequency and time-frequency domain analysis techniques (Akselrod *et al.*, 1987), (Benhorim *et al.*, 1996), (Badilini *et al.*, 1992), (Garcia *et al.*, 2000), (Ranjith *et al.*, 2003), (Milosavljevic and Petrovic, 2006) have been successfully applied for feature extraction and analysis. Some authors have explored the projection onto different sets of basis functions for feature extraction. In this context, principal component analysis (PCA) and Karhunen-Loève transform (KLT) (Castells *et al.*, 2007), (Pang *et al.*, 2005), (Afsar *et al.*, 2008), have been extensively utilized, while a few number of works have used discrete Hermite functions (Gopalakrishnan *et al.*, 2004). The classification stage has been tackled using different approaches. For instance, artificial neural networks based methods (Maglaveras *et al.*, 1998), (Papaloukas *et al.*, 2002), (Mohebbi and Moghadam, 2007), have been proposed. Other authors favour rule-based (Papaloukas *et al.*, 2001), (Andreao *et al.*, 2004) and fuzzy rule (Vila *et al.*, 1997), (Exarchos *et al.*, 2007) approaches. In terms of time and frequency domain analysis techniques, Akselrod *et al.*, (1987), proposed the first method for the direct analysis of the ST segment. It is based on a single measure of the magnitude of the point located 104 ms after the R peak. Another method was proposed by Benhorim *et al.*, (1996), in which two points, considered as the start and end points of the ST segment, are calculated depending on the RR interval of each beat. Badilini *et al.*, (1992), presented an algorithm that uses statistic variables, extracted from the frequency distributions of ST displacements, to discriminate between normal and ischemic ambulatory ECG recordings. Furthermore, ischemic episodes are identified by

using a cluster technique. Garcia *et al.,* (2000), applied an adaptive amplitude threshold method to the root mean square series of differences between ST-T complex (or ST segment) and an average pattern segment, to detect ischemic episodes. Ranjith *et al.* (2003), employed a wavelet transform to determine ECG characteristic points from which ST segment deviation and T wave amplitude measures are obtained and used to detect ischemic episodes. Milosavljevic and Petrovic, (2006), proposed the use of wavelets for extracting myocardial ischemia characteristic patterns, which are obtained through different decomposition scales. ST deviation is calculated for each beat and the number of ST deviations is correlated with the time of consecutive appearances in order to distinguish normal from ischemic ECGs.

Regarding methods based on neural networks, Maglaveras *et al.* (1998), introduced an adaptive backpropagation neural network to identify ischemic beats. In this approach, ischemic episodes classification is achieved by analysing a sequence of classified beats. Mohebbi and Moghadam, (2007), also proposed a beat classification method based on an adaptive backpropagation neural network. Papaloukas *et al.* (2002), employed a feed-forward neural network (trained using a Bayesian regularization method) as a beat classifier, which was integrated into a four-stage procedure for the detection of ischemic episodes.

With respect to PCA and KLT approaches, Castells *et al.* (2007) reviewed the application of principal component analysis techniques for the detection of myocardial ischemia and abnormalities in ventricular repolarization. Pang *et al.* (2005) utilized Karhunen-Loève transform parameters extracted from ST-T complex and a measure of the ST segment deviation to detect ischemia by means of an adaptive neuro-fuzzy logic classifier. In turn, Afsar *et al.* (2008) used Karhunen-Loève transform to reduce ST segment data together with an ensemble of lead-specific neural networks classifiers to detect ST segment deviation episodes. In terms of Hermite functions based methods, Gopalakrishnan *et al.* (2004) used ECG expansion in discrete Hermite functions for a real-time monitoring of ischemic changes. Namely, the first fifty Hermite coefficients are applied as inputs to a committee neural network classifier, trained to identify ischemic beats.

Regarding rule-based methods, Papaloukas *et al.* (2001), proposed a strategy to detect ECG changes suggestive of ischemia using a rule-based expert system. Specifically, the system is able to distinguish between episodes of ST segment deviation and T wave changes. Andreao *et al.* (2004), presented an ischemia detection system that uses a hidden Markov model approach for online beat detection and segmentation, and a rule-based classifier for ischemic episodes detection, derived from some heuristic rules defined by cardiologists.

Vila *et al.* (1997), developed an intelligent monitoring system supported on fuzzy set theory, which uses three electrocardiographic leads and one invasive cardiovascular pressure signal in real time to detect ischemic episodes. Exarchos *et al.* (2007), proposed a methodology to create fuzzy expert systems for ischemic beats detection that involves a set of rules extraction using a decision three.

Despite of the many works that have been developed in the context of ischemia automatic detection, the results achieved in terms of Sensitivity and Positive Predictivity can yet be improved. Thus, the search for better results is an incentive for further investigation.

In the present section a new methodology for automatic detection of ischemic episodes is proposed considering the ST segment deviation, and the T wave and QRS morphology variations. In effect, it is known that variations in the ST segment are not always associated with ischemia. For example, sudden changes in QRS morphology can reflect shifts in the electrical axis and ventricular depolarization of the heart, which usually causes considerable alterations in the ST segment level (Castells *et al.*, 2007). Thus, taking into account the QRS morphology, it is expected to improve the detection of true ischemic beats. A new measure of the ST deviation based on the time frequency analysis of the signal, and the expansion onto Hermite basis functions to capture the T wave and QRS complex morphologies, are the key points of the proposed strategy.

4.3.2 Proposed Methodology

Figure 4.9 depicts the schematic diagram of the methodology followed in this work. The input consists of a discrete ECG signal, which is passed through a set of pre-processing stages for noise reduction, fiducial points identification, premature ventricular contractions (PVCs) elimination and baseline removal. Following this, the algorithm involves two processing steps: firstly, each individual beat is classified as normal or ischemic. For this end, features based on the ST deviation, and on the T wave and QRS complex morphologies, are considered. Secondly, ischemic episodes detection is performed using a sliding window procedure. The detailed algorithms are introduced in the next subsections.

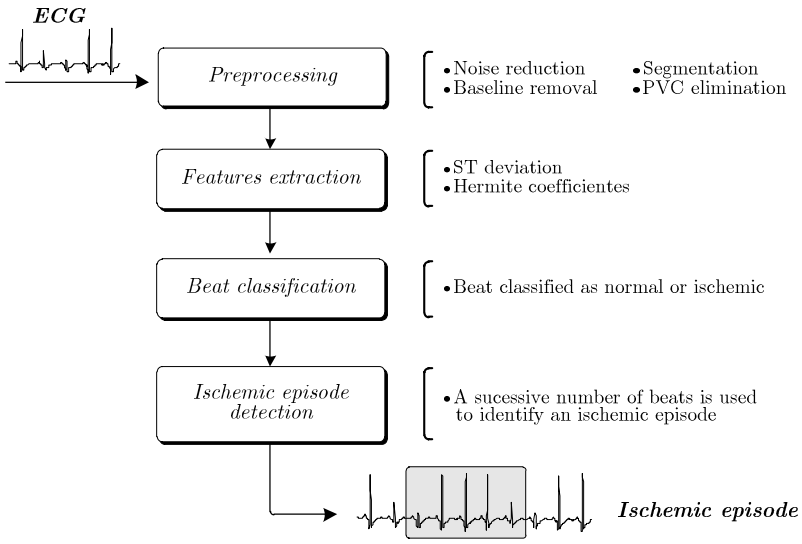


Figure 4.9 - Proposed ischemic episode detection methodology.

Pre-processing

The first stage of pre-processing is concerned with noise reduction. This is achieved by applying a low pass filter to the ECG signal. In particular, a 4^{th} order Butterworth low-pass filter, with a cut-off frequency of 40 Hz, is employed for this purpose.
Afterwards, a segmentation algorithm is used in order to identify the beginning, the peak and the end of each ECG characteristic wave (P, Q, R, S and T). The applied algorithm is based on the work introduced by Sun (Sun, 2001). After the segmentation stage, PVCs are detected and removed from the signal. The algorithm implemented for this task is adapted from Couceiro *et al.* (2008).

The final pre-processing stage consists of baseline wander removal. Baseline wander presence increases the difficulty of the ECG analysis, especially while assessing ST segment deviation. Since the spectrum of baseline wander and the low frequency components of the ECG usually overlap, baseline removal using filtering can cause significant distortion of important clinical information, particularly, ST segment alterations. An effective baseline removal approach has been proposed by Wolf (Wolf, 2004). This method does not require the isoelectric level determination and preserves the low frequency ECG information. Originally, the method considered the average of the distances between consecutive R peaks to split the signal into cardiac cycles. As it is illustrated in Figure 4.10, based on the segmentation procedure previously mentioned, Wolf's method is modified to consider as starting and ending points the start of the P wave (P_{begin}) and the end of the T wave (T_{end}), respectively. The average of the first and last N_{cs} cardiac cycle samples (in this work $N_{cs} = 5$) is used to define a first order polynomial. Fundamentally, baseline shift is approximated by this first order polynomial, being the baseline removal procedure completed by subtracting this baseline shift from the original cardiac cycle signal.

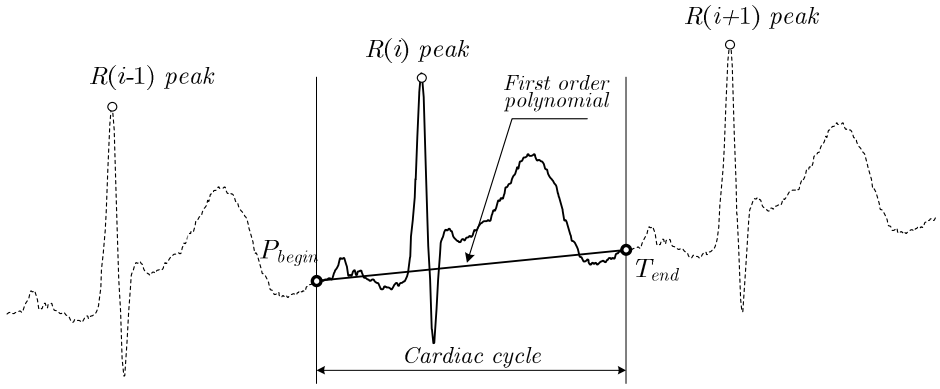


Figure 4.10 - Baseline removal.

Features extraction

The approach followed here, assumes that variations in the T wave and QRS complex morphologies, and the ST segment shift estimation, can be used to discriminate ischemic from non-ischemic episodes.

i. ST segment deviation

The ST segment deviation is assessed considering two different approaches. In the first, the ST deviation is evaluated based on the heart rate and on the R peak location. This information can be easily obtained by means of any ECG segmentation algorithm. Mainly for this reason, this is a very simple and practical method, guarantying robustness, even in the presence of noise and artifacts. However, since it basically depends on the R peak location and not on the ECG waves morphology, this method does not guarantee accurate results. On the other hand, the second approach, based on the time frequency analysis is able to explicitly capture the transient characteristics of the ECG waves. Given these properties, this method is ideal to estimate the ST deviation being, however, more sensitive to noise and artifacts. The strategy followed in this work aims to take advantage of both approaches by providing accurate ST estimation in the case of noise free signals, while ensuring satisfactory results in the presence of artifacts.

ST segment deviation based on R peak location

Through a correlation analysis procedure, three algorithms for ST shift estimation (Akselrod *et al.*, 1987), (Pang *et al.*, 2005) and (Taddei *et al.*, 1992) have been implemented, compared and validated using the ESC ST-T database. In view of the obtained results, the method proposed by Pang *et al.* (2005), was chosen for this task. In this method, the ST segment deviation is evaluated in a point that depends on the heart rate and on the R peak location, according to Table 4.3.

Table 4.3

ST deviation – measuring point.

Heart Rate (bpm)	Measuring point
< 100	$\mathrm{R}_{peak} + 120$ ms
100 ~ 110	$\mathrm{R}_{peak} + 112$ ms
110 ~ 120	$\mathrm{R}_{peak} + 104$ ms
>120	$\mathrm{R}_{peak} + 100$ ms

ST segment deviation based on time-frequency analysis

In general, time-frequency methods are applied to provide a more detailed view of the time distribution of the spectral components that constitute a signal. In particular, it is recognized that time-frequency methods are especially adequate for the detection of small transient characteristics hidden in the ECG, such as ST segment alterations. Thus, this work proposes a new approach for the estimation of ST deviation based on a time-frequency analysis, in particular using the Wigner-Ville transform that it offers a good balance between time and frequency resolutions.

Figure 4.11 illustrates the Wigner-Ville transform, corresponding to the cardiac cycle shown in Figure 4.12a).

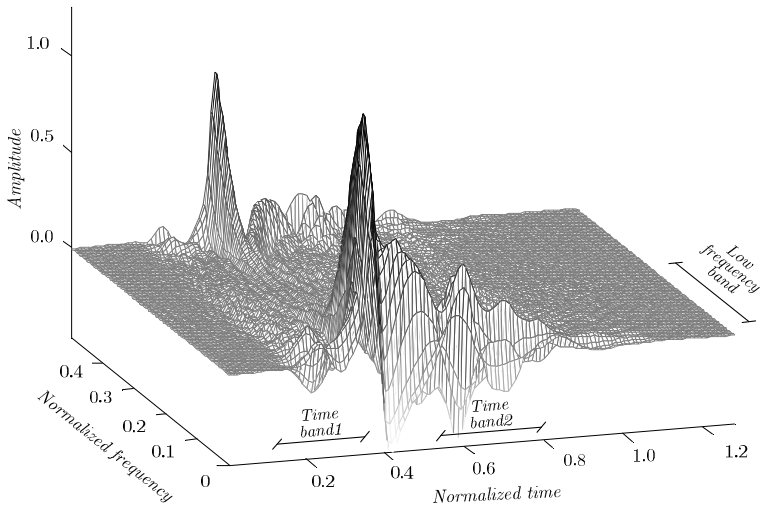


Figure 4.11 - Wigner-Ville transform of a cardiac beat.

Through the Wigner-Ville transform, a time-frequency representation of each heart beat is performed. Then, using this time-frequency map, frequency characteristics of specific time regions can be analysed. In particular, for the ST shift estimation, the difference between the ST segment and the isoelectric line has to be computed. Therefore, two main points have to be identified: the isoelectric point and the J point. From the clinical perspective, these two points correspond to stable regions, where the heart's electrical activity is constant. Moreover, the isoelectric point is located between the end of the P wave and the onset of the QRS complex; the J point is to be found after the end of the QRS complex and before the begin of the T wave.

One of the hypothesis assumed in this work is that the energy content of the Wigner-Ville distribution can be employed to identify these points. Thus, in a first step, two time bands were considered (time band 1 and time band 2), as shown in Figure 4.11 and Figure 4.12*a*). The time band 1 corresponds to the time period between the end of P wave and the onset of the QRS complex; the time band 2 corresponds to the time period between the end of QRS complex and the onset of the T wave.

Then, in a second phase, for those time bands, the respective high frequency components of the Wigner-Ville transform were analysed. In particular, it is assumed that the sum of the absolute values of these frequency components in each time band (time band 1 and time band 2) can identify, respectively, the isoelectric point and the J point.

Figure 4.12*b*) illustrates this idea, depicting the corresponding high frequency components of the heart beat shown in Figure 4.12*a*). These frequency components correspond to values in the range [0.5,1.0], considering normalized frequency values in the interval [0.0,1.0].

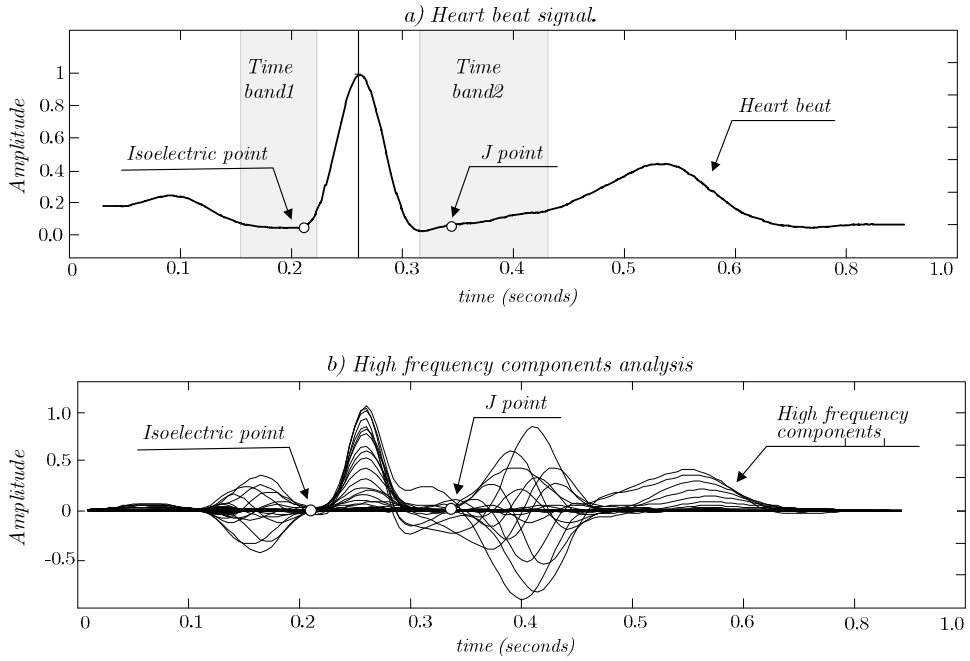


Figure 4.12 - Cardiac cycle and respective high frequency components.

ii. QRS complex and T wave characterization

As mentioned, ischemia may induce morphology alterations in the T wave and in the QRS complex. In order to characterize these changes in morphology, each cardiac beat is represented in a space spanned by a limited number of Hermite basis functions. Basically, using the expansion in Hermite functions method, the signal of interest is decomposed into a linear combination of orthonormal basis functions, which coefficients can be used as features in the classification process, just as with the Principal Component Analysis technique (Castells *et al.*, 2007). However, the former has the advantage to be patient independent, since the set of basis functions are predefined (Figure 2.4, subsection 2.3.3) and do not require any prior knowledge of the data set. This reason, coupled with its ability in capturing the relevant morphology changes using a low number of basis functions, led to the choice of the expansion in Hermite functions methodology for using in the present work.

Coefficients of the expansion in Hermite functions

The Hermite functions form an orthonormal base of $L^2(R)$, the space of integrable functions. They can be determined as the product of a Gaussian by the Hermite polynomials with some normalization constants (Clifford *et al.*, 2006), as described in equation (4.19) (subsection 2.3.3, equation (2.14)).

$$H^n(t,l) = \frac{1}{\sqrt{n!2^n\sqrt{\pi}\,l}} e^{\frac{-t^2}{2l^2}} P^n(t/l) \tag{4.19}$$

In the previous equation, $P^n(t/l)$ represents the Hermite polynomial of order n, with l as a scaling factor (allows width adjusting). The Hermite polynomials can be recursively determined, see subsection 2.3.3, equation (2.15).

In order to approximate as close as possible the shape of each beat to the shape of the Hermite functions, while using a reduced number of coefficients, each cardiac cycle was divided in two segments: *Segment*1 is defined from the end point of the P wave until the J point and *Segment*2 is defined from the J point until the end of the T wave. Thus, two expansions in Hermite functions were actually carried out for each cardiac beat. The goal was to describe each discrete signal segment, $Y(t) \in \mathbb{R}^{1,N}$, as (4.20).

$$\hat{Y}(t) = \sum_{j=0}^{m-1} c_j \; H^j(t,l) \tag{4.20}$$

In the previous equation, $\hat{Y}(t) \in \mathbb{R}^{1,N}$ stands for the estimated signal segment, m defines the number of basis functions and c_j correspond to the expansion coefficients. The last ones can be obtained by a criterion error, consisting of minimizing the sum squared error, as follows:

$$E(c_j) = \sum_t \left[Y(t) - \sum_{j=0}^{m-1} c_j \; H^j(t,l) \right]^2 \tag{4.21}$$

Consider a given signal $Y(t) \in \mathbb{R}^{1,N}$, and a matrix formed by the Hermite functions, $H \in \mathbb{R}^{m,N}$, defined as (4.22).

$$H = \left[H^0, \; H^1, \; \ldots, \; H^{m-1} \right]^T \tag{4.22}$$

The vector of coefficients, $C \in \mathbb{R}^{m,1}$, defined according to (4.23), can be obtained by using a least square error formulation, thus, through a pseudo inverse computation, as presented in equation (4.24).

$$C = \left[c_0, \; c_1, \; \ldots, \; c_{m-1} \right]^T \tag{4.23}$$

$$C = \left(HH^T \right)^{-1} HY^T = H^{\dagger} Y \tag{4.24}$$

One of the assumptions in this work is that the expansion coefficients C can reflect the changes in the ECG morphology induced by ischemia and, therefore, be adequate to represent the second set of features used in beat classification process. Figure 4.13 depicts a real cardiac beat and the corresponding approximation using its expansion in Hermite functions. The Hermite coefficients were evaluated for each segment (each one resampled to 64 samples) using *m=6*.

The scaling factors for *Segment*1 and *Segment*2, were, respectively, $l = 5$ and $l = 8$, determined in order to minimize the approximation error. For each segment, the resulting coefficients were C1 and C2 presented below.

$$\begin{aligned} C1 &= [\ 0.9835 \quad -0.1700 \quad -0.2926 \quad 0.1298 \quad 0.0595 \quad -0.0840\ ] \\ C2 &= [\ 0.1631 \quad -0.0041 \quad -0.0451 \quad 0.0294 \quad -0.0107 \quad 0.0188\ ] \end{aligned} \tag{4.25}$$

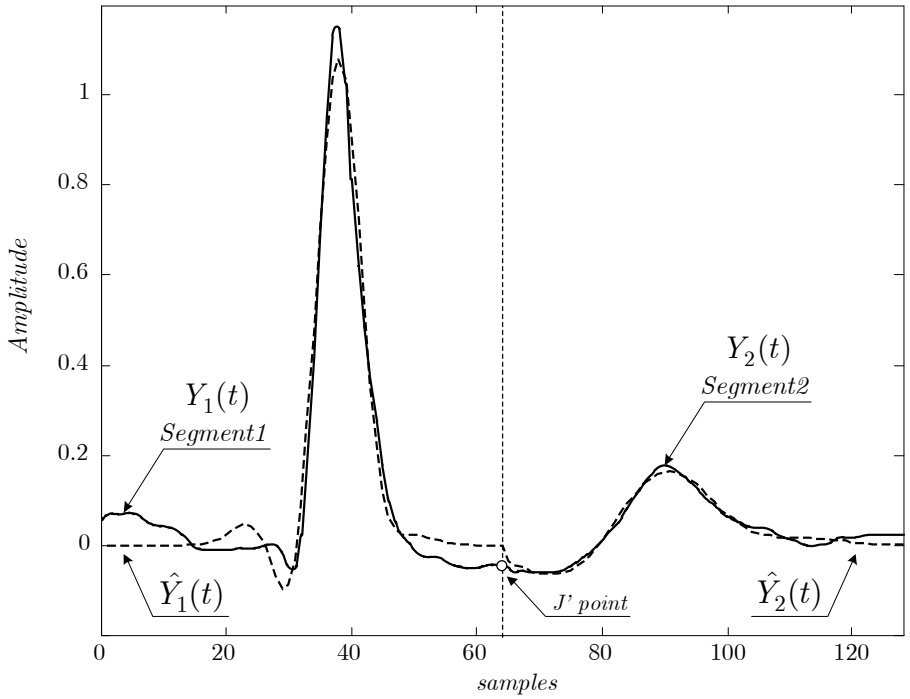


Figure 4.13 - Approximation of a cardiac beat using expansion in Hermite functions.

It is important to highlight that it is not fundamental to obtain very low approximation errors. In fact, the underlying idea is that the approximated signal and, indirectly, the Hermite coefficients, have the capacity to capture the most relevant morphologic characteristics of the signal.

In order to determine the adequate number of Hermite basis functions an experiment was carried out, which results are presented in Figure 4.14. In effect, for a selected set of representative cardiac beats, several expansions were done considering different numbers of basis functions and the respective errors calculated (as the difference between the real signal and the estimated one). Figure 4.14 presents the mean error obtained from the selected set of signals, which is separately indicated for *Segment*1 and *Segment*2. From the analysis of both error curves it can be concluded that a total of 6 Hermite basis functions is a choice that leads to an acceptable error at the same time that guaranties a number of coefficients sufficiently low.

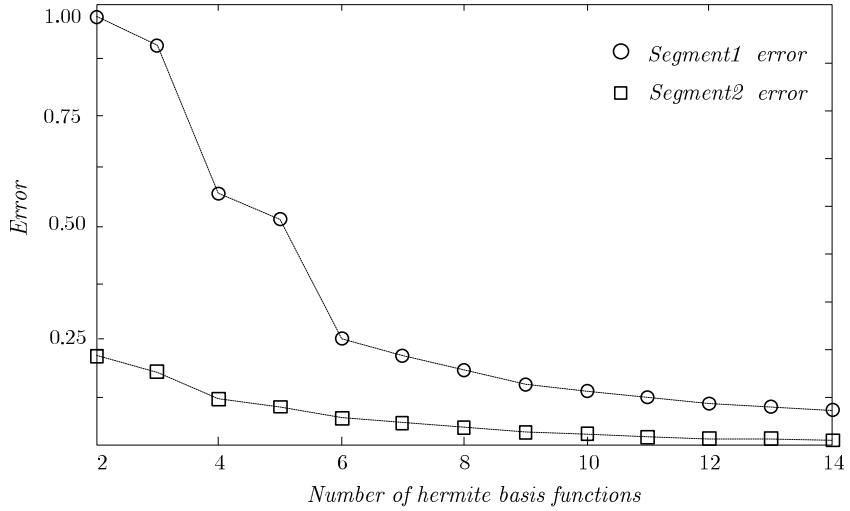


*Figure 4.14 - Approximation of Segment*1 *and Segment 2 using Hermite functions.*

To analyse how changes in the ECG morphology are reflected in the Hermite coefficients, a simulated cardiac beat was created: the first segment presenting a deep Q wave and the second segment exhibiting an inverted T wave. Figure 4.15 shows this simulated cardiac beat and the respective Hermite approximation. The corresponding Hermite coefficients are C1 and C2 presented below, with the same scaling factors as before.

$$\begin{aligned} C1 &= [\ 0.8561 \quad 0.3626 \quad \text{-}0.2008 \quad 0.0496 \quad \text{-}0.1550 \quad 0.1959\ ] \\ C2 &= [\ \text{-}0.3369 \quad \text{-}0.2261 \quad 0.0349 \quad 0.0955 \quad \text{-}0.0562 \quad \text{-}0.0092\ ] \end{aligned} \tag{4.26}$$

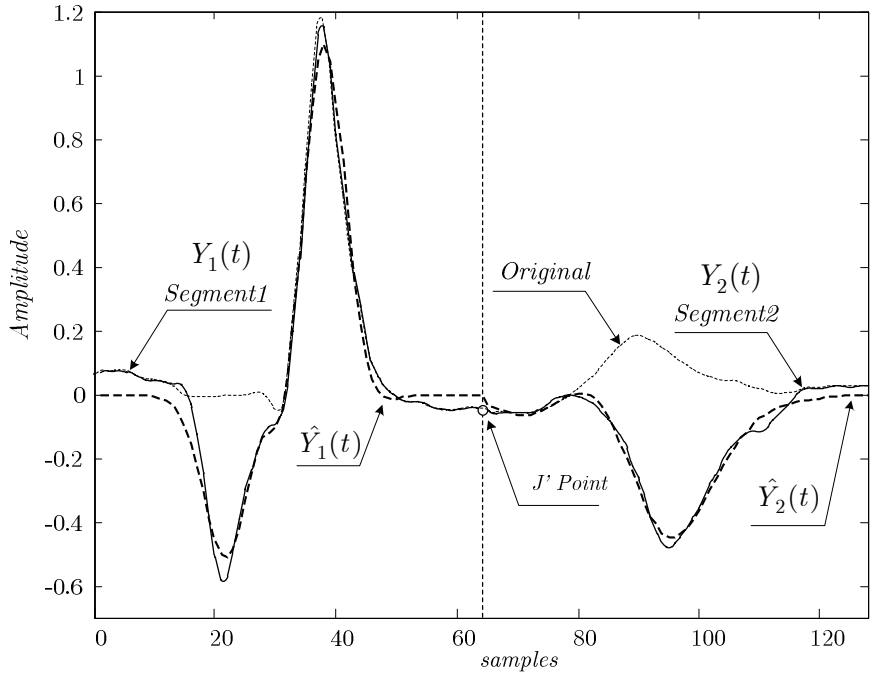


Figure 4.15 - Approximation of a cardiac beat using Hermite expansion.

For these two particular situations, the coefficients revealed the morphology changes in the ECG. In fact, for the first segment, the major variation occurred in the second coefficient that assumed a value of -0.1700 for the original signal and of 0.3626 for the simulated one. This disparity reflects the Q wave variation, mainly corresponding to the Hermite function observed in Figure 2.4 of subsection 2.3.3, with n=1. For the second segment, the major variation was verified in the first coefficient, which took a value of 0.1631 in the case of the original segment and of -0.3369 in the case of the simulated one. Clearly, this difference is related to the T wave inversion, as observed in Figure 2.4 of subsection 2.3.3, with n=0. These results confirm that the QRS complex and T wave morphologies can be characterized by a relatively small number of Hermite functions and that the corresponding coefficients have the potential to be used as features for ischemic beats identification.

Classification

The first classification strategy considered two classifiers: one to deal with ST elevation and other to manage ST depression. However, since the morphology of the ECG waves depends on the specific ECG acquisition lead, the results achieved with this approach were not significant when compared to the ones presented in literature. Therefore, to deal with the particularities of each lead configuration, a lead dependent classification system was the chosen solution. As a result, a specific classifier is implemented for each lead.

Given their properties, neural networks have been recognized as a powerful tool for pattern classification problems, especially when applied to numeric data classification. In the context of ischemic beats classification, neural networks have been extensively applied with significant performance results (Maglaveras *et al.*, 1998), (Papaloukas *et al.*, 2001), (Mohebbi and Moghadam, 2007). Due to their universal approximation nature, low complexity and excellent results achieved in similar classification tasks, neural networks were used in this work.

Figure 4.16 depicts the main modules of the classification scheme.

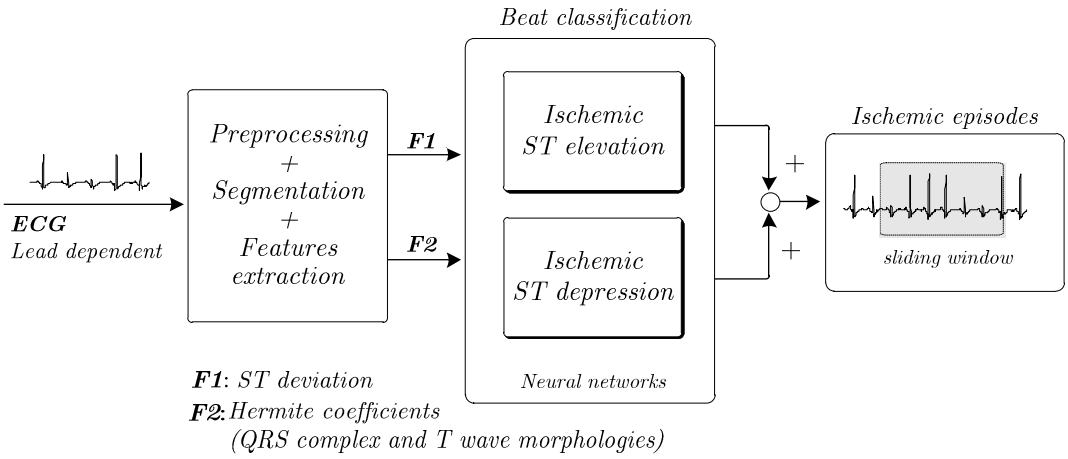


Figure 4.16 - Proposed classification scheme.

i. Beat Classification

For classifier selection several experiments were carried out with different types of neural networks. The chosen strategy consists of two independent Feed-Forward Neural Networks (FFNNs) (Figure 3.7, subsection 3.3.3.) for each lead: the first classifies the nature of the ST elevation of each beat and the second distinguishes beats with ST depression from others. After beat classification a sliding window with size of 40 beats is applied to each FFNN output signal in order to eliminate isolated misclassified beats. At the end, the outputs from both networks (elevation and depression) are combined by an OR operation.

Episode Detection

Ischemic episodes detection involves two steps: first a sliding window procedure is applied to the entire ECG signal. The window's length is set to 40 beats. It is considered as an ischemic episode if more than 50% of the beats are classified as ischemic. In a second phase, the classification done in the previous step is reviewed and episodes with a separation of less than 40 beats are merged.

4.3.3 Results

For algorithm validation purposes, the European Society of Cardiology ST-T database was used (Taddei, *et al.*, 1992), (Goldberger *et al.*, 2000). This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects for which myocardial ischemia was diagnosed or suspected. Each record is two hours in duration and contains two signals from 8 different leads (V1, V2, V3, V4, V5, MLI, MLIII and D3). These signals are sampled at 250 Hz. From the 90 complete records of this database, 48 records are freely available and were used in this work.

To assess the quality of the proposed algorithms, sensitivity (SE) and positive predictivity (PP) have been evaluated, according to equations (4.27) and (4.28), respectively.

$$SE = \frac{TP}{TP + FN} \tag{4.27}$$

$$PP = \frac{TP}{TP + FP} \tag{4.28}$$

In the equations above, TP (true positives) represents the annotated beats/episodes in the database that were identified by the algorithms, FN (false negatives) corresponds to the annotated beats/episodes that were not detected and, finally, FP (false positives) denotes the number of beats/episodes that were not annotated in the database, but that were incorrectly identified by the algorithms.

Features extraction

The extracted features in each cardiac cycle were related with ST segment deviation as well as with the QRS and T wave morphology changes. ST deviation was evaluated using both approaches described in subsection 4.3.2. In turn, each cardiac beat segment (Segment 1 and Segment 2) was approximated by a linear combination of the first six Hermite functions (order 0 to 5). Taking into account that the expansion of each segment originated 6 coefficients, a total of 14 features were determined for each cardiac beat: 2 features related to the ST deviation and 12 Hermite coefficients. Subsequently, a moving average filter of order 10 was applied to all the features.

To validate the potential of the referred features in discriminating normal from ischemic beats, a linear correlation analysis procedure took place. Therefore, the correlation coefficients between these parameters and the classification values in the ESC ST-T database were computed. Table 4.4 presents the average correlation coefficients and the respective standard deviations for all the signals considered in this study, where ST1 and ST2 represent the ST deviation calculated using the two different approaches mentioned before, and the Hj denote the Hermite coefficients from order 0 to 5.

Table 4.4

Features correlation analysis.

Correlation coefficient	ST Deviation		Hermite coefficients for Segment1						Hermite coefficients for Segment2					
	ST1	*ST2*	*H0*	*H1*	*H2*	*H3*	*H4*	*H5*	*H0*	*H1*	*H2*	*H3*	*H4*	*H5*
Average	0.63	0.64	0.30	0.34	0.41	0.40	0.37	0.35	0.35	0.41	0.28	0.46	0.48	0.63
Standard deviation	0.21	0.23	0.25	0.22	0.27	0.26	0.24	0.26	0.26	0.26	0.20	0.24	0.23	0.22

The values presented in Table 4.4 confirm the potential of the features in question.

The Figure 4.17, Figure 4.18, and Figure 4.19, illustrate this correlation analysis performed on the e0103 record, channel 2, lead MLIII, which is composed of 6997 beats. In order to simplify the visualization, only some of the features are depicted. Thus, Figure 4.17 depicts the ST deviation obtained using the Wigner-Ville approach. Figure 4.18 shows the first three Hermite coefficients corresponding to *Segment*1 and Figure 4.19 presents the first three Hermite coefficients corresponding to *Segment*2. For this particular example, the correlation coefficients obtained are presented in Table 4.5.

Table 4.5

Correlation analysis for the e0103 record.

	ST Deviation	Hermite coefficients for Segment1			Hermite coefficients for Segment2		
	ST2	*H0*	*H1*	*H2*	*H0*	*H1*	*H2*
Correlation Coefficient	0.82	0.53	0.49	0.79	0.83	0.11	0.44

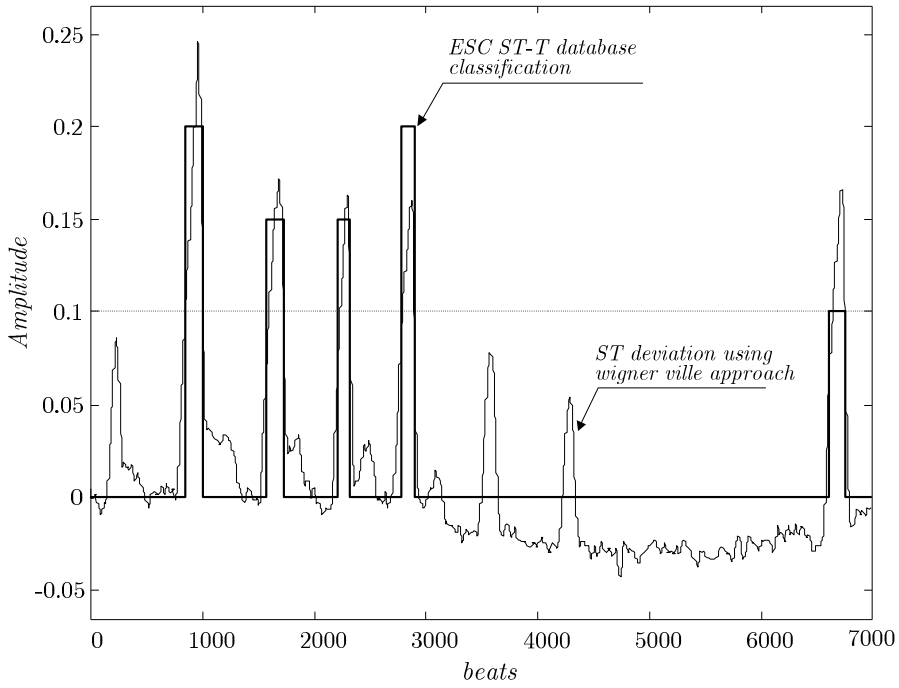


Figure 4.17 - ST deviation.

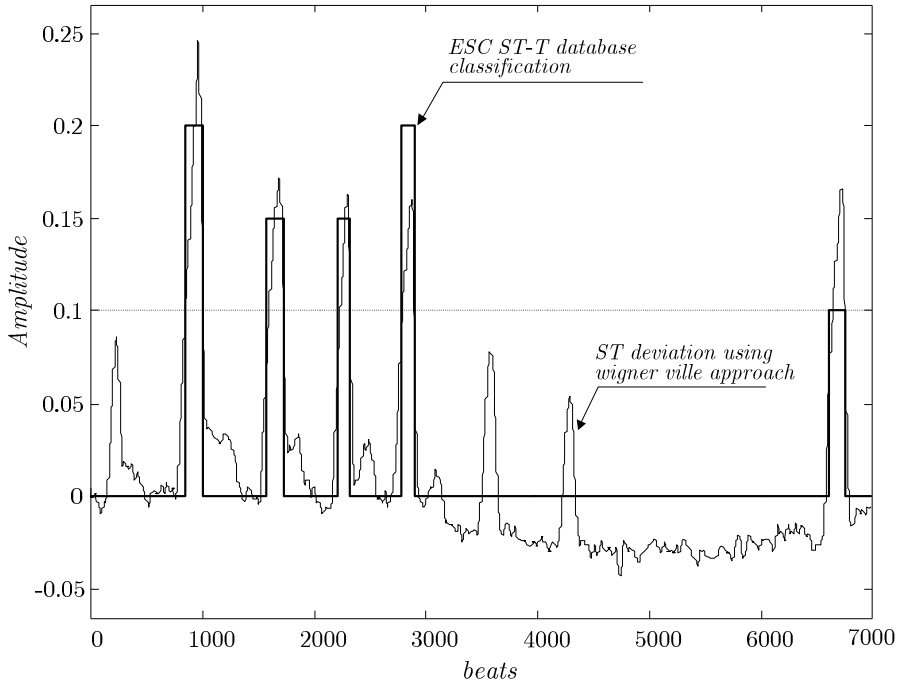


*Figure 4.18 - First three coefficients of Hermite expansion: Segment*1.

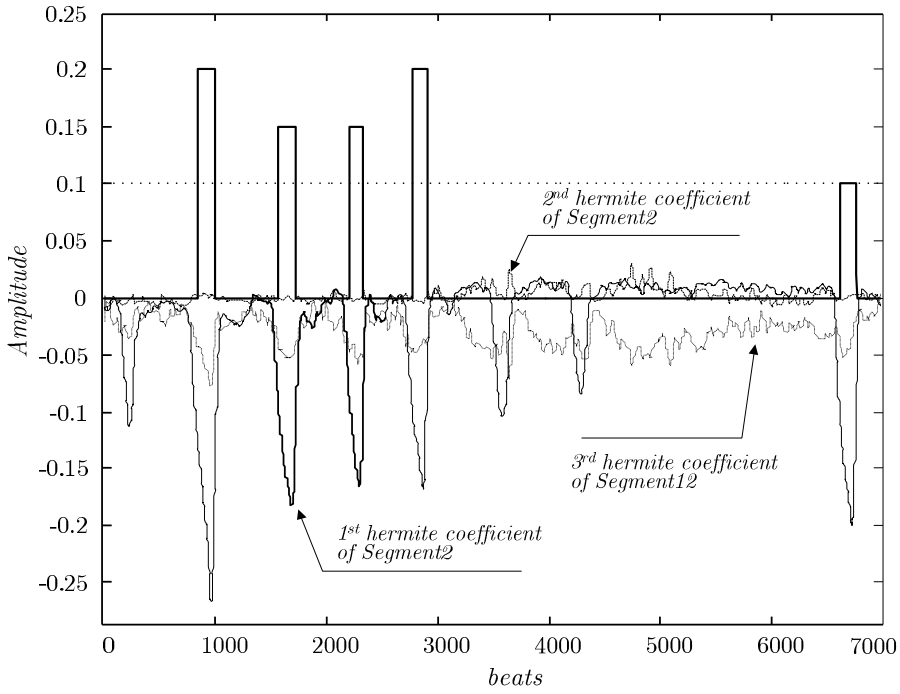


Figure 4.19 - First three coefficients of Hermite expansion: Segment2.

The correlation coefficients presented in Table 4.5, as well as the respective visualization in Figure 4.17, Figure 4.18, and Figure 4.19, demonstrate the discrimination effectiveness of the selected features.

Training and validation

Dataset

Regarding training and validation, data subsets from the 48 freely available signals of the ESC ST-T database were selected according to each lead. Each ECG record was split into two sub-records, that is one from channel 1 and another one from channel 2, originating 96 signals for training and validation. The specific classifier for each lead, was trained and validated using signals contained in the database for that lead type. Moreover, after the pre-processing phase, some of the cardiac cycles were removed and were considered neither for training nor for validation purposes. In this category are PVCs and noisy beats.

To validate beat classification, 81 of the 96 available signals were utilized. In effect, some signals containing annotations that were not considered consistent were discarded. An example of an ambiguous situation is illustrated in Figure 4.20. The first part of the figure (Figure 4.20a) shows a four seconds section of the e0207 record, channel 1, lead V5, starting at index 1620900. The second part (Figure b) shows a four seconds section of the e0303 record, channel 2, lead V5, starting at index 1107300.

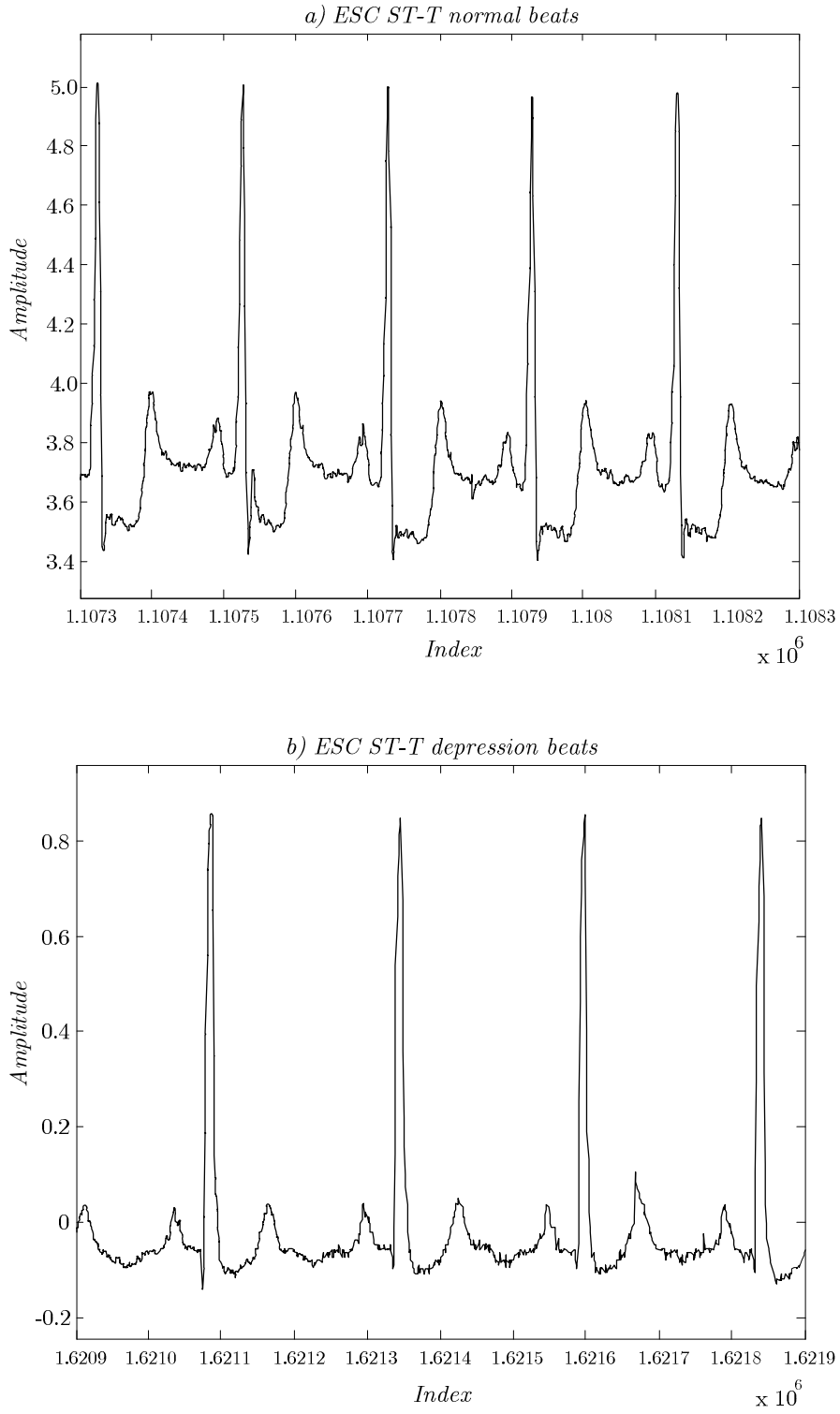


Figure 4.20 - Non consistent beat classification examples.

According to the ESC ST-T database classification, the first signal (Figure 4.20-a)) was annotated as normal, while the second (Figure 4.20-b)) was considered as having ST depression. However, as can be observed, this classification contradicts what one might expect.

Table 4.6 presents the exact number of cardiac cycles used for each lead (after PVCs, noisy and non-consistent beat elimination). The discarded signals, for each lead, were: e0207 for lead MLI, e0109, e0121, e0609 and e0613 for lead MLIII, e0403 for lead V1, e0415 and e0603 for lead V2, e0119, e012 and e0161 for lead V4, and e0207, e0213, e0303 and e0405 for leadV5.

Table 4.6

Training and validation dataset.

Lead	Nº of signals in database	Cardiac cycles in database	Nº of signals considered	Cardiac cycles considered
V1	5	33554	4	30548
V2	8	49704	6	35110
V3	3	14487	3	14487
V4	19	134872	16	108107
V5	27	192249	23	162747
MLI	8	63851	7	56990
MLIII	25	159142	21	133477
D3	1	1465	1	1465
Total	**96**	**649324**	**81**	**542931**

Training and Validation

For training purposes only a small portion of representative signals (30 beats before and after the annotated episodes transitions) were applied. As already mentioned, beat classification was lead dependent and was carried out by means of two FFNNs per lead. Considering the 8 different leads (V1, V2, V3, V4, V5, MLI, MLIII and D3) present in the ESC ST-T database, a total of 16 neural networks were utilized. A neural architecture composed by two hidden layers (sigmoid tangent activation functions) was considered. The number of hidden neurons was experimentally determined and the parameters (weights and bias) that characterize all the FFNNs were trained using the Levenberg Marquardt algorithm.

Table 4.7 presents, for each lead, the architecture of each classifier. The notation used is the number of neurons corresponding to [*inputs*, *first layer*, *second layer*, *output*].

Table 4.7

Lead NN Structure.

Lead	NN structure
V1	[14, 5, 3, 1]
V2	[14, 5, 3, 1]
V3	[14, 5, 3, 1]
V4	[14, 6, 4, 1]
V5	[14, 7, 3, 1]
MLI	[14, 3, 3, 1]
MLIII	[14, 5, 3, 1]
D3	[14, 5, 3, 1]

For ischemic episode validation, beat sequences of annotated and identified episodes were compared. If the beginning and the end of them matched within a defined tolerance (40 beats) then episode detection was considered as successful. Otherwise, it was considered as unsuccessful. In Figure 4.21, a representative example of ischemic episodes identification by the proposed algorithms is presented using the e0103 record. In fact, the manifest overlap between the annotated episodes and the ones identified by the algorithms, testifies the ability of the methodology to perform the intended detection task.

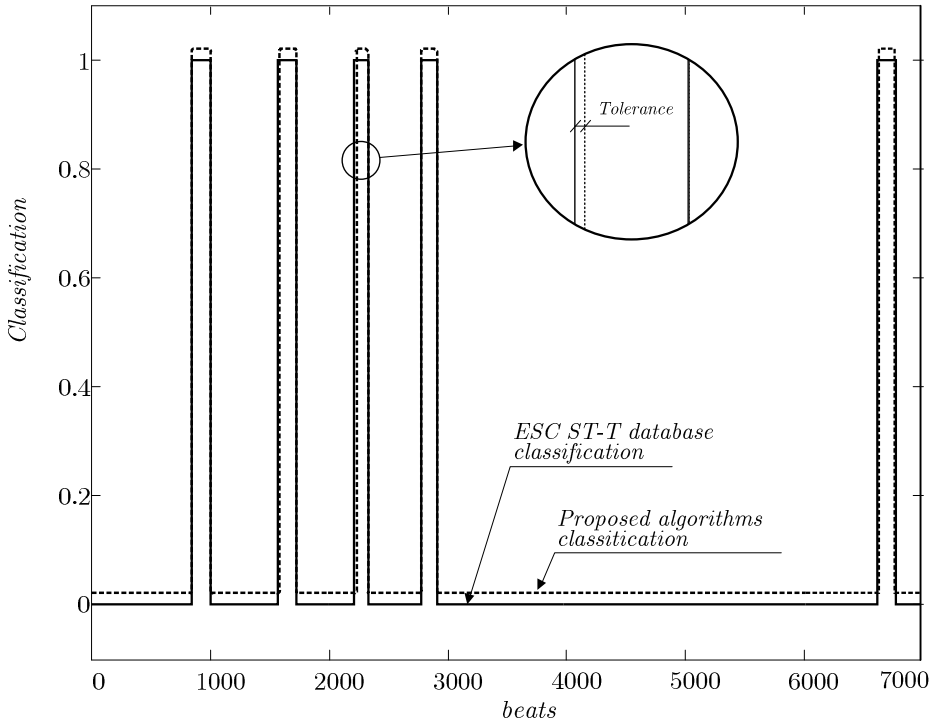


Figure 4.21 - Ischemic episodes validation (e0103 record).

Results and discussion

The results achieved by the proposed algorithms for ischemic beat classification and ischemic episode detection are presented in Table 4.8 and Table 4.9.

Table 4.8

Beat classification performance.

Lead	Nº of Signals	Nº of Beats	FFNN Neg.		FFNN Pos.	
			SE	PP	SE	PP
V1	4	30548	100.0%	100.0%	-	-
V2	6	35110	99.6%	99.9%	100.0%	100.0%
V3	3	14487	100.0%	99.3%	-	-
V4	16	108107	94.6%	92.8%	100.0%	100.0%
V5	23	162747	95.7%	97.5%	100.0%	100.0%
MLI	7	56990	99.2%	98.7%	99.6%	99.3%
MLIII	21	133477	100.0%	100.0%	96.4%	97.0%
D3	1	1465	-	-	100.0%	100.0%
Total	**81**	**542931**	**98.4%**	**98.3%**	**99.3%**	**99.3%**

Table 4.9

Episodes detection performance.

Lead	Episodes	TP	FP	FN	SE	PP
V1	5	5	0	0	100.0%	100.0%
V2	5	5	0	0	100.0%	100.0%
V3	2	2	0	0	100.0%	100.0%
V4	34	32	2	4	88.9%	94.1%
V5	38	35	3	6	85.4%	92.1%
MLI	5	5	1	0	100.0%	83.3%
MLIII	32	32	0	0	100.0%	100.0%
D3	1	1	0	0	100.0%	100.0%
Total	**122**	**117**	**6**	**10**	**96.7%**	**96.2%**

As can be observed in the previous tables, the global sensitivity and positive predictivity reached average values of 96.7% and 96.2%, respectively.

Despite the importance of evaluating the performance of the proposed algorithms with respect to other state of the art methods, it is realized that this is a very difficult task due to the fact that different data sets are applied to derive the reported results in many of these papers. Badilini *et al.* (1992), uses a private database. Although many authors use the ESC ST-T database to evaluate their algorithms, it is observed that some (e.g. (Milosavljevic and Petrovic, 2006), (Gopalakrishnan *et al.*, 2004)) do not specify the number of records (although the latter refers the number of ischemic/normal beats used). Others, e.g. Papaloukas *et al.* (2002) use all the available records of the database but with reviewed annotations. In other studies, as those reported in (Mohebbi and Moghadam, 2007), (Exarchos *et al.*, 2007), only a small set of records are utilized (5 and 10, respectively) for validation purposes. On the other hand, some authors evaluate their methods using all database records with the original annotations. This is the case of Pang (Pang *et al.*, 2005) and Vila (Vila *et al.*, 1997) (ischemic episodes detection), where SE varies from 81.3% to 83.0% and PP ranges from 74.7% to 75.0%. For ST segment deviation episodes, Garcia *et al.* (2000) and Papaloukas *et al.* (2002), report for SE between 84.7% and 92.0%, while PP ranges from 86.1% to 93.8%. Another group of authors base their evaluation on the 48 records freely available. This is the case of the method for ischemic episodes detection introduced by Andreao *et al.* (2004), that achieves a SE of 83.0% and a PP of 85.0%. The algorithm reported by (Afsar *et al.*, 2008) for ST segment deviation episodes achieves a SE of 90.8% and a PP of 89.2%. For obvious reasons, it would be unfair to compare the current work with the group that used the entire records of the database as they would be in disadvantage. In fact, the most suitable studies for comparison are actually those reported by Andreao *et al.* (2004) and by Afsar *et al.* (2008) since the data set used for their evaluation is basically the same as the one applied in the present study. Consequently, it can be concluded that the results presented in Table VII significantly improve the results achieved by the methods reported by these authors, both in terms of sensitivity (96.7%), as well as in terms of positive predictivity (96.2%).

It should be stressed that all modules of the algorithm have been designed to operate on short signal windows (40 beats). This is an important aspect, since it does not require significant durations of ECG to perform ischemia characterization. Since the proposed detection scheme is based on neural network models, the classification process is very fast. For these reasons, and given the importance that the early detection of ischemic events assumes in the management of CAD patients, the described methodology was included in the telemonitoring system of the HeartCycle project.

4.3.4 Conclusions

In this section a strategy for ischemic episode detection was proposed. The methodology consists of two main steps: first, each individual beat is classified as normal or ischemic, considering features based on the ST deviation, and on the T wave and QRS complex morphologies. To handle the particularities of each lead configuration, a lead dependent classification scheme is implemented using two FFNNs per lead, specifically designed to deal with ST elevation and ST depression, respectively. In the second stage of the algorithm, ischemic episodes detection is performed based on a sliding window procedure.

The most innovative aspects are the new approach for accurate ST shift and isoelectric point estimation based on the time-frequency analysis, and the ECG beat morphology effective characterization using the expansion in Hermite functions.

The methodology's potential was confirmed by using the European Society of Cardiology ST-T database. In fact, the achieved results (sensitivity of 96.7% and positive predictivity of 96.2%) are relevant when compared with similar works reported in literature.

4.4 Prediction of Acute Hypotensive Episodes

This section deals with the application of neural network multi-models to the prediction of adverse acute hypotensive episodes (AHE) occurring in intensive care units (ICU).

A generic methodology consisting of two phases is considered. In the first phase, a similarity analysis between the current blood pressure time signal (template) and a collection of historical blood pressure signals is carried out. From this procedure, the most similar signals are determined and the respective prediction neural models, previously trained, selected. Then, in a second phase, the multi-model structure is employed to predict the future evolution of current blood pressure signal, enabling the detection of AHE.

The effectiveness of the methodology was validated in the context of the 10^{th} Physionet/Computers in Cardiology Challenge - Predicting Acute Hypotensive Episodes, applied to a specific set of blood pressure signals, available in MIMIC-II database. A correct prediction of 10 out of 10 AHE for event 1 and of 37 out of 40 AHE for event 2 was achieved, corresponding to the best results of all entries in the two events of the challenge.

4.4.1 Introduction

Hypotension, a clinical condition characterized by abnormal low blood pressure values, is one of the recurrent situations occurring in intensive care units. Among the most frequent events, acute hypotensive episodes (AHE) are particularly critical, since they may result in irreversible organ damage and, eventually, death (Piccini and Nilsson, 2006). As a consequence, the characterization of such episodes is of fundamental importance in the management of critical ill patients. In fact, when promptly detected, it is possible to improve the clinical decision concerning which intervention is more appropriated for each specific condition (sepsis, myocardial infarction, cardiac arrhythmia, pulmonary embolism, haemorrhage, dehydration, or any of a wide variety of other causes of hypovolemia, insufficient cardiac output, or vasodilatory shock). Additionally, early detection of AHE will give professionals enough time to select a more effective treatment, without exposing the patient to additional risks of delaying therapy. Therefore, the development of methodologies able to detect not only the presence of this condition but also to predict its occurrence, is of extreme importance concerning appropriate clinical interventions. Moreover, since the interventions to treat such events are usually invasive and aggressive, a prediction system that could identify an imminent episode would be a significant benefit to timely support non-invasive and preventive treatments.

It is clinically accepted that if there exists enough patient's clinical information, then a prediction system for hypotensive episodes, over a specific time period, can be developed.

Typically this information is based on the medical record, such as clinical history, laboratory tests and medications, as well as on information extracted from physiological vital signals, such as electrocardiogram, blood pressure and respiration. In this context, Singla *et al.* (2006) showed the correlation between some independent variables and the development of early hypotension episodes. These variables included age, sex, body mass index, history of hypertension, diabetes mellitus, anaemia, heart rate, systolic and diastolic blood pressure. Similarly, Lin *et al.* (2008) studied the association of specific variables with the increasing risk of hypotensive episodes, namely weight, height, American Society of Anaesthesiologists physical status, surgical category (orthopaedics, plastic surgery, general surgery, obstetrics, and urology) and systolic blood pressure. Based on these variables, Lin *et al.* (2008)proposed a logistic regression model to assess the risk of developing an hypotensive episode.

In practice, the development of automatic solutions for hypotensive episodes prediction has explored the correlation between patient's clinical information collected in real-time, such as arterial blood pressure (ABP), heart rate (HR) and oxygen saturation (SO2), and the onset of the hypotensive episode. In particular, Bassale (2001) proposed the use of parametric and non-parametric methods to analyse and characterize ABP before hypotensive episodes. He concluded that ABP variability and shape features have the potential to predict such events. Crespo *et al.* (2002) also suggested the use of changes in the ABP morphology occurring immediately before an episode of hypotension. They proposed the variance of the ABP signal and the variance of the wave slope as the most relevant features to consider when predicting AHE. Lehman *et al.* (2008) presented a similarity-based searching and pattern matching algorithm, applicable to classification and forecasting tasks. Using real physiological measurements, they employed the methodology to forecast hypotensive episodes in intensive care units. Also, Saeed (2006) and Saeed *et al.* (2007) and introduced a new temporal similarity metric, based on a transformation of time series data into an intuitive symbolic representation. They used wavelet decomposition to characterize time series dynamics at multiple time scales. Their algorithm was employed to identify similar physiological patterns in hemodynamic time series from ICU patients, with potential to be used in the detection of imminent hemodynamic deterioration. Frolich *et al.* (2002) suggested the use of baseline HR as a significant predictor of obstetric spinal hypotension. Basically, they showed that higher baseline HR could be a useful parameter to predict postspinal hypotension. Using spectral analysis of HR and ABP variability, Pelosi *et al.* (1999) identified precursors of hypotensive episodes during renal dialysis. Also using frequency analysis techniques, Reich *et al.* (2005) investigated the correlation between HR variability analysis and hypotension events. Chamchad *et al.* (2004) found a significant correlation between non-linear HR variability dimension analysis and the presence of hypotension, occurring after spinal anaesthesia for caesarean delivery. Hanss *et al.* (2005) also concluded that HR variability analysis could be used to predict the occurrence of hypotension during spinal anaesthesia. In particular, they investigated the ratio of low to high frequency peaks of the HR variability power spectrum (LF/HF) for the prediction of hypotension events after spinal anaesthesia, in the specific cases of pregnant women and elderly men.

Mancini *et al.* (2008) showed the potential of SO2 short-term variability in anticipating the hypotension onset. Recently, Lee and Mark (2010) investigated the existence of discriminatory patterns in ICU data that could be indicative of impending hypotensive episodes. Based on neural networks, they proposed a binary classification scheme (normotensive vs. hypotensive) and an estimation strategy for predicting future mean blood pressure values.

This work proposes the forecast of acute hypotensive episodes through the development of predictive multi-models, applicable to the mean ABP (MAP) time-series signal. To achieve this goal, a generic methodology consisting of two main phases is considered. In the first phase, a similarity analysis procedure is carried out between the current MAP signal and a representative set of historical MAP evolution trends. The most similar ones are identified and the corresponding predictive neural multi-models, previously trained using those historical signals, selected. In the second phase, these models are employed to the current MAP signal to predict its future evolution and, therefore, the detection of an AHE occurrence. Basically, the prediction methodology consists of a multi-model scheme using neural network structures. Multi-models do not recursively use model outputs as inputs for step ahead predictions, therefore, prediction errors are not propagated and long-term predictions can be accurately estimated. Among regression models, neural networks have shown considerable capabilities to learn and to generalize from non-linear environments, enabling to capture the fundamental data dynamics. Moreover, multi-models can be trained by means of simple standard backpropagation algorithms. In fact, since an independent neural sub-model is used for each sampling instant and does not depend on previous predictions, a static training algorithm, as the referred backpropagation, can be employed.

The effectiveness of the proposed approach was validated in the context of 2009 Physionet/Computers in Cardiology challenge - Predicting Acute Hypotensive Episodes. The data for training and validation purposes was obtained from MIMIC-II dataset[8] that includes data before and during the prediction horizon. The forecast was made using the trained neural multi-model structure, only considering the information available before the forecast period. The occurrence of an AHE within the forecast window (one hour) was assessed according to AHE definition (Moody and Lehman, 2009). A sensitivity of 94.74% and a specificity of 93.55% revealed the effectiveness of the strategy, which obtained the best results of the challenge[11].

This section is organized as follows: in subsection 4.4.2. the 2009 Physionet/Computers in Cardiology challenge is presented. In subsection 4.4.3., it is described the general methodology for the prediction task as well as how it can be used to address the challenge. In subsection 4.4.4., results using MIMIC-II dataset are presented and discussed, in the context of the Physionet/Computers in Cardiology challenge. Finally, in subsection 4.4.5., some conclusions are drawn.

---

[11] www.physionet.org/challenge/2009/final-scores

4.4.2 The 2009 Physionet/Computers in Cardiology Challenge

Challenge goal

The 2009 challenge was the tenth in the annual series of open challenges hosted by Physionet in cooperation with Computers in Cardiology Conference. The goal of that year's challenge was to predict which patients in the available dataset (MIMIC-II) would experience an acute hypotensive episode beginning within the forecast window, motivated by the possibility of improving care and survival of those patients (Moody and Lehman, 2009).

The challenge dataset included, among other data, time series of mean arterial blood pressure (MAP) at one-minute intervals. Each sample of the series was an average of the blood pressure measured in the radial artery over the previous minute. Given such a time series, an AHE was defined, for the purposes of the challenge, as any period of 30 minutes, or more, during which at least 90% of the MAP measurements were at or below 60 mmHg. The forecast window was defined as the one-hour period immediately following a specified time instant t_0 (Figure 4.22). In the test sets, all data following t_0 was withheld and the forecast should be made using the available information before t_0 .

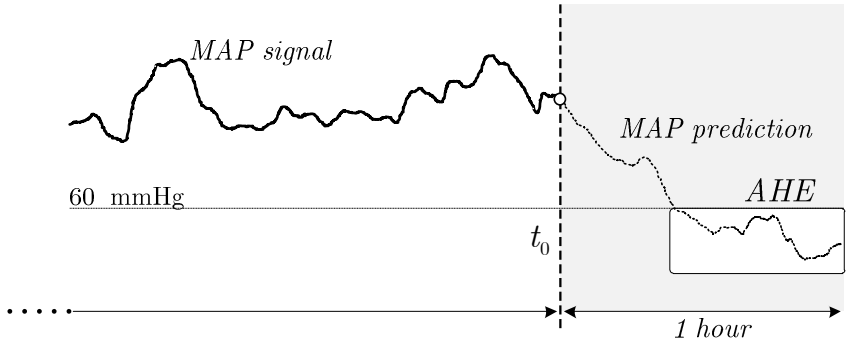


Figure 4.22 - 2009 Physionet/Computers in Cardiology challenge goal.

MIMIC-II project

Data used in this challenge was collected and contributed to Physionet by the MIMIC-II project (*Multi-parameter Intelligent Monitoring for Intensive Care*), a Bioengineering Research Partnership funded by the US National Institutes of Health and its National Institute of Biomedical Imaging and Bioengineering, with additional support from Philips Medical Systems. The MIMIC-II project has collected data from about 30,000 ICU patients to date, including recorded physiological signals and time series, as well as accompanying clinical data such as interventions performed in the ICU, laboratory tests, observations and medication (Saeed *et al.*, 2002). Basically, this information is organized

in two main databases: *numerics record* that contains time-series data and *clinical records* that deals with clinical information (observations, medication, etc.). The intent is that a MIMIC-II record should be sufficiently detailed to allow its use in studies that otherwise would require access to an ICU, e.g., for basic research in intensive care medicine, or for development and evaluation of diagnostic and predictive algorithms for medical decision support.

Training and test datasets

The 2009 challenge dataset consisted of selected patient records from the MIMIC-II database. In the training set, the records included all available data before and after instant t_0 . In the test sets the records were truncated at t_0 , being the data after this instant unknown during the period of the challenge and made available only after its conclusion. The records that were chosen for the challenge dataset included, at a minimum: *i*) at least 12 hours of data before t_0 , and at least one hour of data after t_0 ; *ii*) ECG and arterial blood pressure (ABP) signals sampled at 125 Hz; *iii*) time series of vital signs sampled once per minute (in the training set) and once per second (in the test sets). These comprised heart rate and mean systolic and diastolic ABP. The majority of the records included a variety of additional vital signs time series, most often containing respiration rate and oxygen saturation.

In particular, the training set consisted of 60 records (with data after t_0) belonging to two groups: H and C. Records in group H contained an episode of acute hypotension beginning during the forecast window (the one-hour period following t_0), while records in group C contained no AHE within the forecast window. Within group H, 15 records belonged to subgroup H1 - patients who received pressor medication, and 15 belonged to subgroup H2 - patients who did not receive pressor medication. Within group C, 15 records belonged to subgroup C1 - patients with no documented AHE at any time during their hospital stay, and 15 records belonged to subgroup C2 - patients who had AHE before or during the forecast window.

The validation set consisted of two datasets, A and B. The test set A comprised 10 records, excluding data after t_0 . From these, 5 records were from subgroup H1 (AHE in subjects receiving pressors), and 5 were from subgroup C1 (no AHE in subjects receiving pressors). The test set B consisted of 40 records, also excluding data after t_0 . Between 10 and 16 from these belonged to group H, and between 24 and 30 belonged to group C.

Challenge events

The challenge comprised two events. The event 1, using A dataset, focused on patients who were receiving pressor medication and aimed to distinguish between two groups of ICU patients: the ones who would experience an acute hypotension episode and those who would not. The event 2, using B dataset, addressed the broad question of predicting an AHE in the general population.

4.4.3 Methodology

The proposed scheme is depicted in Figure 3.14, subsection 3.4.1. The detection of a future acute hypotensive episode (AHE) is carried out by means of neural network multi-models trained using mean arterial blood pressure signals (MAP).

The input consists of a MAP signal available before t_0 (template), the instant where the forecast period starts. From a similarity analysis procedure, between the template and a set of MAP signals, representative of historical MAP evolution trends, the most similar ones are identified. After that, the corresponding neural multi-models, previously trained using those historical representative signals, are selected and employed to predict the future evolution of the template, from instant t_0 until the end of the forecast window. Finally, the occurrence of an AHE is straightforwardly determined.

This section starts by reviewing the general multi-model approach for prediction purposes. Then, the neural network models structure and their incorporation into the multi-model scheme are presented. Finally, the prediction of AHE based on this strategy is addressed.

Multi-models

Using a multi-model strategy, one independent sub-model is employed for each sampling instant within the prediction horizon. Consequently, future predictions do not depend on previous predictions, allowing to obtain more accurate estimations. In general, a specific future time instant P can be expressed by equation (4.29), being $f_P(\cdot)$ a mapping such that $f_P : \mathbb{R}^n \rightarrow \mathbb{R}$.

$$\hat{y}(t+P \mid t_0) = f_P\Big(y(t_0), y(t_0-1), ..., y(t_0-(n+1)) \Big) \tag{4.29}$$

Thanks to this structure, predictions over a forecast horizon do not depend on previous predictions, but only on information available at the current instant k, $\varphi(k)$. On the other hand, one independent model ($f_i(\cdot)$) has to be used for each sampling instant within the prediction horizon. As result, if P future instants have to be predicted, P distinct regression models have to be derived.

Neural-network regression models

Each regression sub-model ($f_i(\cdot)$) is here implemented by means of a neural network model. In particular, generalized regression neural networks (GRNN), a type of radial basis function networks, are considered. As mentioned in section 3.4.2 the principal advantages of GRNN are their aptness for smooth function-approximation, their ability to predict the behaviour of systems based on few training samples and their interpolation properties between training samples (Bauer, 1995). They enable a fast learning and are often more accurate than multi-layer perceptron networks. On the other hand, although multi-models are used for long-range prediction, each neural network can be trained by means of a *standard backpropagation* algorithm. Actually, as referred, the dimension of the training dataset, $\{\mathbf{X}(t), \mathbf{Y}(t)\}$ $t = 1,..,M$, predetermines

the number of hidden neurons (*NP*). Thus, the training of a GRNN only involves the estimation of the kernels width, λ. In the application of the GRNN structure to the particular problem of AHE prediction, the number of previous instants considered by each model (designated here as the *order*) together with the time period before the starting of the forecast window (designated here as *size*) determine the dimension of the dataset, therefore the number of the hidden layers. These *size* and *order* parameters will be detailed in subsection 4.1.1.

4.4.4 Application to the Prediction of AHE

The prediction of an AHE is based on a set of multi-models trained using mean arterial blood pressure signals (MAP), obtained from MIMIC-II *numerics record* dataset (H and C datasets). No information from *clinical records* was used.

The input consists of a discrete MAP signal (sampled once per minute) considering the information available before t_0. This template passes through a set of pre-processing techniques, namely to deal with missing information, noise reduction and normalization. Then, a similarity analysis procedure is carried out between the processed template and a set of MAP signals, representative of historical MAP evolution trends. From this similarity analysis the most similar signals are identified and the corresponding multi-models, previously trained, are selected. These specific neural multi-models are then employed to predict the future evolution of the template, from instant t_0 until the end of the forecast window (one hour). Finally, an AHE is identified if, according to the challenge definition, at least 90% of the MAP predicted signal is at or below 60 mmHg during a period of 30 minutes or more.

Pre-processing

Firstly, a pre-processing procedure is applied to the original MAP signal. This involves resampling the raw signals, deal with missing values and perform a noise reduction, as illustrated in Figure 4.23.

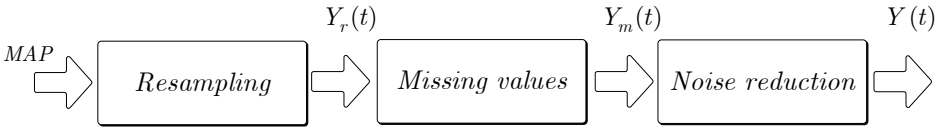


Figure 4.23 - Pre-processing stages.

In the first phase all MAP signals are resampled to 1 sample per minute, resulting in the signal $Y_r(t)$. Thus, for signals presenting a sampling rate of 1 Hz (A and B testing datasets) the average of the blood pressure measured in each 60 samples is considered.

To deal with missing values a simple procedure is carried out. In case of missing values in the MAP signal, a first order linear interpolation is performed, enabling to obtain the

signal $Y_m(t)$. This process uses the values in the limits of the missing interval, e.g., the last available value (on the left) and the first available value (on the right).

For noise reduction, a simple first order filter is used, considering a pole at 0.8, as described in (4.30).

$$Y(t) = \frac{0.2\ q^{-1}}{1-0.8\ q^{-1}} Y_m(t) \tag{4.30}$$

Finally, it is assumed the same duration, equal to T=11 hours, for all the signals $Y(t)$. Specifically, periods of 10 hours before t_0, and 1 hour after t_0 are considered.

Figure 4.24 illustrates the pre-processing procedure for the particular signal H1_#1 (training set group H1 signal #1, *h1_a40439*). As can be seen, missing values, approximately between instants 320 and 380 minutes, and between 480 and 580 minutes, are replaced using the described technique.

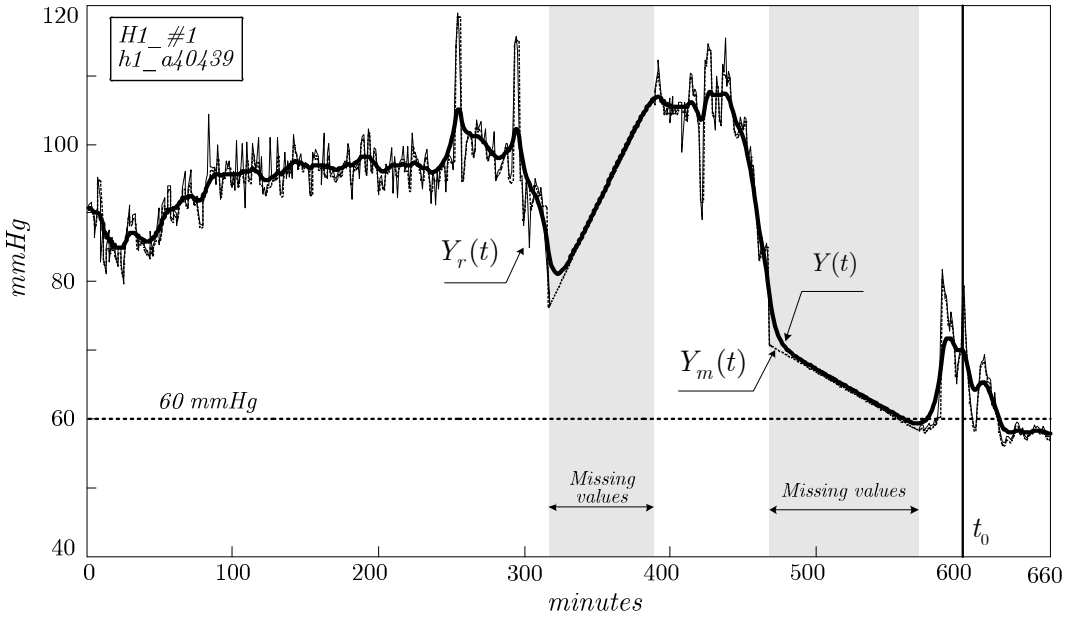


Figure 4.24 - Pre-processing phases: $Y_r(t)$: *resampling of MAP signal (1sample per minute);* $Y_m(t)$: *replacing of missing values;* $Y(t)$: *filtered signal.*

MAP signals and GRNN multi-models

To identify the representative MAP signals, a historical dataset composed of past and future tendencies has to be considered. In the proposed strategy, the dataset consists of 59 training records (H and C), available in Physionet/CinC challenge [11]. Actually, one signal was excluded (C2_#4, *c2_40234*), since it presented a significant discontinuity in the neighbourhood of the instant t_0, which caused serious difficulties during the training phase. Therefore, a matrix of M signals, $Y \in \mathbb{R}^{M \times T}$, with $M = 59$ and $T = 660$ minutes, is defined.

$$Y = \begin{bmatrix} Y^1 \\ Y^2 \\ \dots \\ Y^M \end{bmatrix} \qquad with \quad Y^i \in \mathbb{R}^{1\times T}, i = 1...M \tag{4.31}$$

The pre-processing phase, described in the last section, is applied to each Y^i signal.

To address future predictions, each of these time series signals (H and C) is modelled using the GRNN multi-model approach. These models are trained using past and future information (before and after t_0). Moreover, to reduce the number of sub-models, each GRNN structure is trained to deal with 15 step ahead predictions. As result, given the forecast period (1 hour), 4 neural sub-models are considered for each MAP representative signal (f_1, f_2, f_3, f_4). Consequently, the total number of models is 59×4=236.

Prediction of MAP signals

Given a new MAP testing signal, truncated at time instant t_0 , the forecast is done based on previously trained GRNN multi-models. To select the specific multi-models, a similarity analysis procedure takes place. Basically, in a first stage, the strategy proposed in Chapter 2 is employed to assess the similarity between new MAP data and stored MAP signals.

From this analysis, a vector composed of similarity measures, $S_M \in \mathbb{R}^{1\times M}$, is computed for a specific period of time (*size*), starting before the forecast window until instant t_0 (*size* parameter will be introduced in the next section). The signal $X(t)$ represents the new MAP signal, with dimension $T' = 10\ hours$, $X(t) \in \mathbb{R}^{1\times T'}$. As referred, the Y^i represents the i^{th} signal ($i = 1,...,M$) of the matrix Y.

In a second stage, the MAP signals that present the highest similarity measure are selected. In particular, being $S_M \in \mathbb{R}^{1\times M}$ the vector composed of these similarities (sorted in descending order), the first m signals ($m < M$) are selected, such that equation (4.32) is verified.

$$\frac{\sum_{i=1}^{m} S_M(i)}{\sum_{i=1}^{M} S_M(i)} > \boldsymbol{\lambda} \tag{4.32}$$

The parameter $\boldsymbol{\lambda}$ is a pre-defined scalar, aiming to only select the most relevant multi-models (with the highest similarities). Therefore, using this approach a variable number of multi-models is selected for each particular MAP signal.

Finally, the occurrence of an AHE is assessed according to its definition (Moody and Lehman, 2009), considering the predicted MAP signal $Y_p(t)$ over the forecast window.

For the prediction of the MAP signal, a weighted average of the predictions performed by the m multi-models is computed, as described by equation (4.33).

$$Y_p(t) = \frac{\sum_{i=1}^{m} S_m(i) \times Y^i(t)}{\sum_{i=1}^{m} S_m(i)} \qquad t = t_0 + 1, ..., t_0 + 60 \tag{4.33}$$

4.4.5 Results

This section focuses on the validation results of the proposed strategy. Firstly the concepts of *size* and *order*, essential to the MAP modelling through GRNN, are introduced, as well as information about the neural networks training process. Then, AHE prediction results are presented in detail and discussed.

GRNN multi-models

GRNN size and order

When modelling each MAP signal, the selection of the *order* and the *size* is of particular importance. The parameter *size* is defined as the period before the starting of the forecast window, from where information is used for training purposes. The parameter *order* defines the previous instants considered by each model. In order to estimate these parameters an optimization procedure was carried out, by means of the minimization of the least square prediction error over the forecast window.

$$\min_{size,order} \sum_{t=t_0+1}^{t=t_0+60} \left(\mathrm{Y}(t) - Y_p(t) \right)^2 \tag{4.34}$$

Variables $Y(t)$ and $Y_p(t)$ define, respectively, the actual and the predicted MAP signal. The referred minimization procedure was performed considering different values for the *order* and *size* parameters, specifically in the ranges $order \in [60,\ 90]$ and $size \in [120,\ 180]$, with increments of 10 minutes.

Figure 4.25 depicts the histogram regarding the *order* and *size* parameters obtained for all test datasets (H and C datasets, 60 signals). As can be seen, the predominance of the models can be described by (*order*, *size*)=(60, 170) and (*order*, *size*)=(80, 140). The statistical analysis of mean and standard deviation, results in the following values: mean(*size*)=152 minutes, std(*size*)=15 minutes; mean(*order*)=70 minutes, std(*order*)=9 minutes. Based on these results, in can be concluded that the evolution of the MAP signals can be characterized, on average, using the past 70 minutes (*order*), and the models should be trained using the past 2,5 hours (*size*).

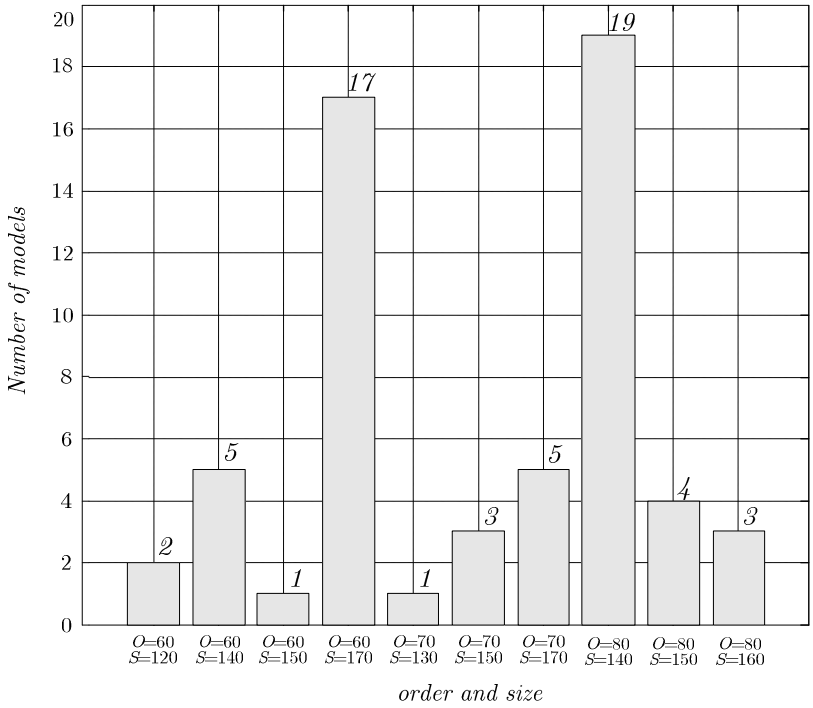


Figure 4.25 - Size and order parameters, estimated for training signals (H and C dataset).

GRNN training

The GRNN structures were defined and trained using the *newgrnn* function [2], available in Matlab toolbox. Basically, the training of a GRNN is performed in a single step (no backpropagation of error is involved). In particular, the training comprises the determination of kernels widths. Figure 4.26 presents the training results for the particular testing record H1_#4 (*h1_a40834*). For this specific signal, the *size* and *order* values are, respectively, 140 and 80 minutes.

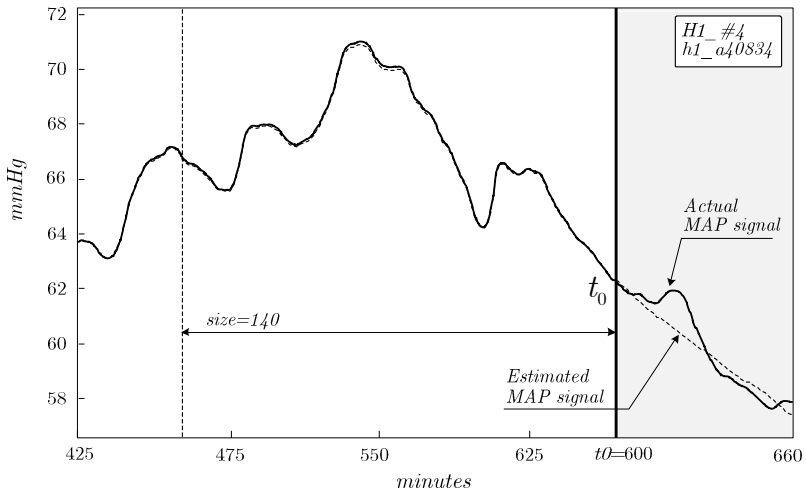


Figure 4.26 - GRNN training: testing signal H1_#4 (h1_*a40834).*

It is important to stress that the neural network multi-models predict future behaviour of signals over the whole prediction horizon, only using information available before the starting of the forecast window (instant t_0).

AHE prediction in the challenge context

Using the present strategy, test datasets available in the Physionet/CinC challenge (10 records of A dataset, and 40 records of B dataset) were used for validation purposes. Firstly, each of these 50 signals was compared with the MAP representative signals, considering a specific period of *size* minutes before instant t_0 (which depends on the specific signal). The corresponding GRNN models, determined from the similarity analysis procedure, were used to predict future MAP values.

Figure 4.27 shows the prediction over the forecast window for the particular signal A_#10 (*a_110bnm*). For the determination of models to be employed, a value of $\lambda = 0.15$ (tolerance) was considered in equation (4.32). From this computation, 4 signals (m=4) were obtained. These signals correspond to the training records C2_#9 (*c2_a40329*), H1_#9 (*h1_a41835*), C2_#8 (*c2_a40306*) and C2_#10 (*c2_a40355*), with similarity measures of 0.5370, 0.5126, 0.5031 and 0.4997, respectively.

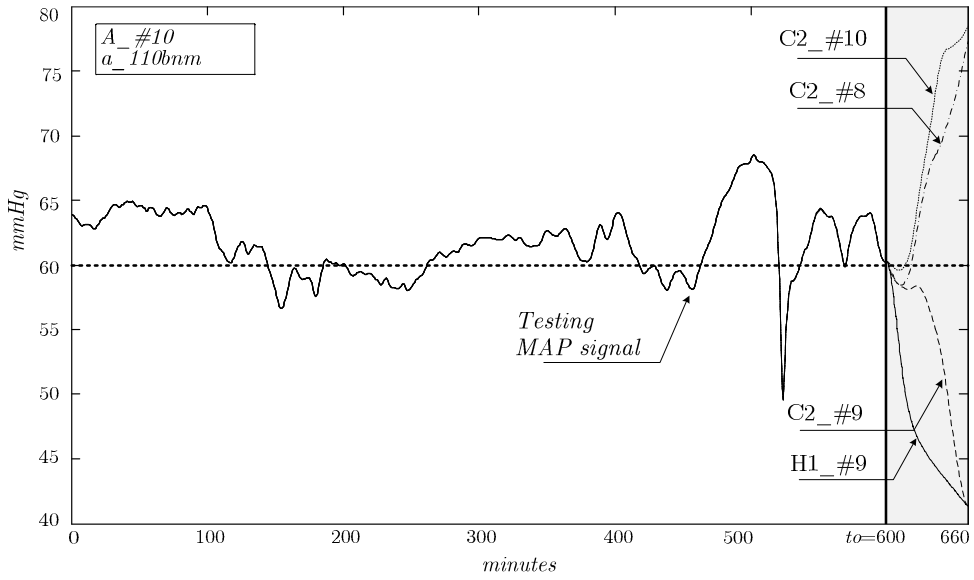


*Figure 4.27 - Prediction of MAP signal A_#10 (*a_110bnm*) using neural multi-models.*

The final MAP predicted signal is computed as the weighted average of all four estimated predictions, equation (4.33), being the identification of an episode straightforwardly performed using the definition of AHE. In this work, an AHE is considered to occur if in a period of 20 minutes or more, at least 90% of the MAP measurements are at or below 60 mmHg (instead of the 30 minutes originally defined by the challenge). This reduction is mainly due to the pre-processing phase, since it

introduces some delay in the signal evolution, as well as it produces a smoothness of the original signal. Figure 4.28 shows the prediction of the specific MAP signal A_#10 (*a_ 110bnm*) over the forecast horizon, resulting from the weighted average of the mentioned signals. For this particular case, using the previous definition an AHE is identified.

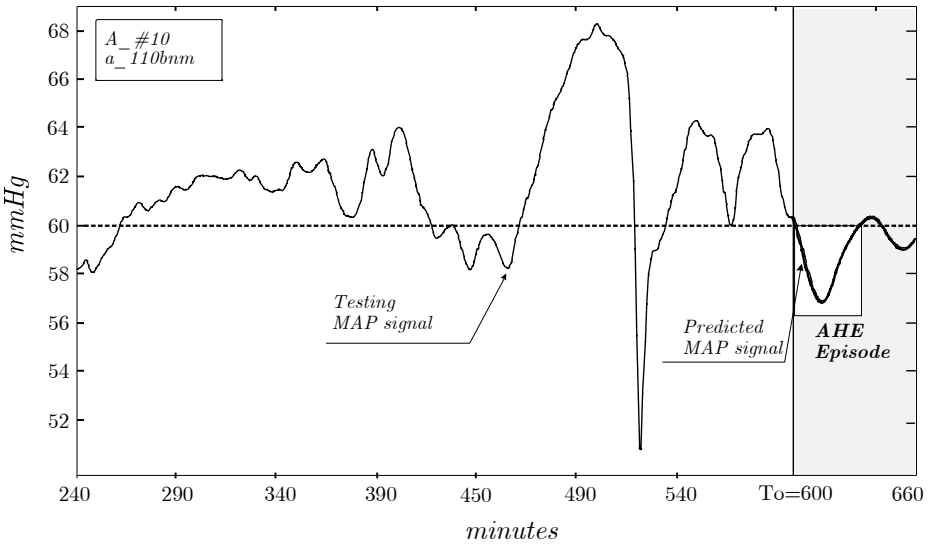


Figure 4.28 - MAP prediction and AHE identification; testing signal A_ #10 (a_ 110bnm).

Table 4.10 presents the global results for the 2009 Physionet/Computers in Cardiology challenge. As can be observed, for the event 1 (using dataset A), 5 AHE were identified corresponding to the signals {1, 2, 4, 9, 10}. For the event 2 (using dataset B), 15 episodes were identified.

The proposed methodology achieved a correct prediction of 10 out of 10 AHE for event 1, and of 37 out of 40 AHE for event 2, which were the best results of all entries in the two events of the challenge.

Table 4.10

AHE detection.

	AHE detected
Dataset A	1, 2, 4, 9, 10
Dataset B	2, 3, 5, 7, 9, 14, 17, 18, 22, 23, 25, 26, 34, 38, 39

Discussion of results

The obtained results can be described in terms of sensitivity, specificity and accuracy. Sensitivity (SE), equation (4.35), gives the percentage of actual AHE that were correctly identified; specificity (SP), equation (4.36), gives the percentage of AHE that did not occur and were correctly identified, and accuracy (AC), equation (4.37), gives the total percentage of correct predictions.

$$SE = \frac{TP}{TP + FN} \tag{4.35}$$

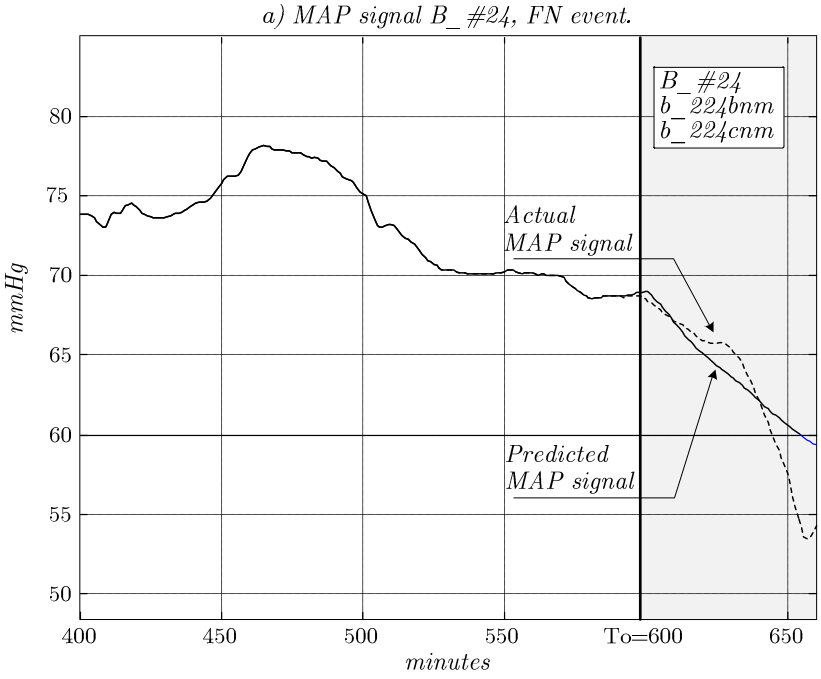
$$SP = \frac{TN}{TN + FP} \tag{4.36}$$

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.37}$$

The variables TP, TN, FP and FN define, respectively, true positive, true negative, false positive and false negative events detected.

In the validation phase, three incorrect predictions were made: one FN event {B_#26} and two FP events, {B_#5, B_#24}. As result, the global sensitivity, specificity and accuracy values were, respectively, SE=94.74%, SP=93.55% and AC=94.00%.

Figure 4.29 shows the predicted results, as well as the actual MAP signals for the AHE incorrectly classified (the last were made available only after the conclusion of the challenge).



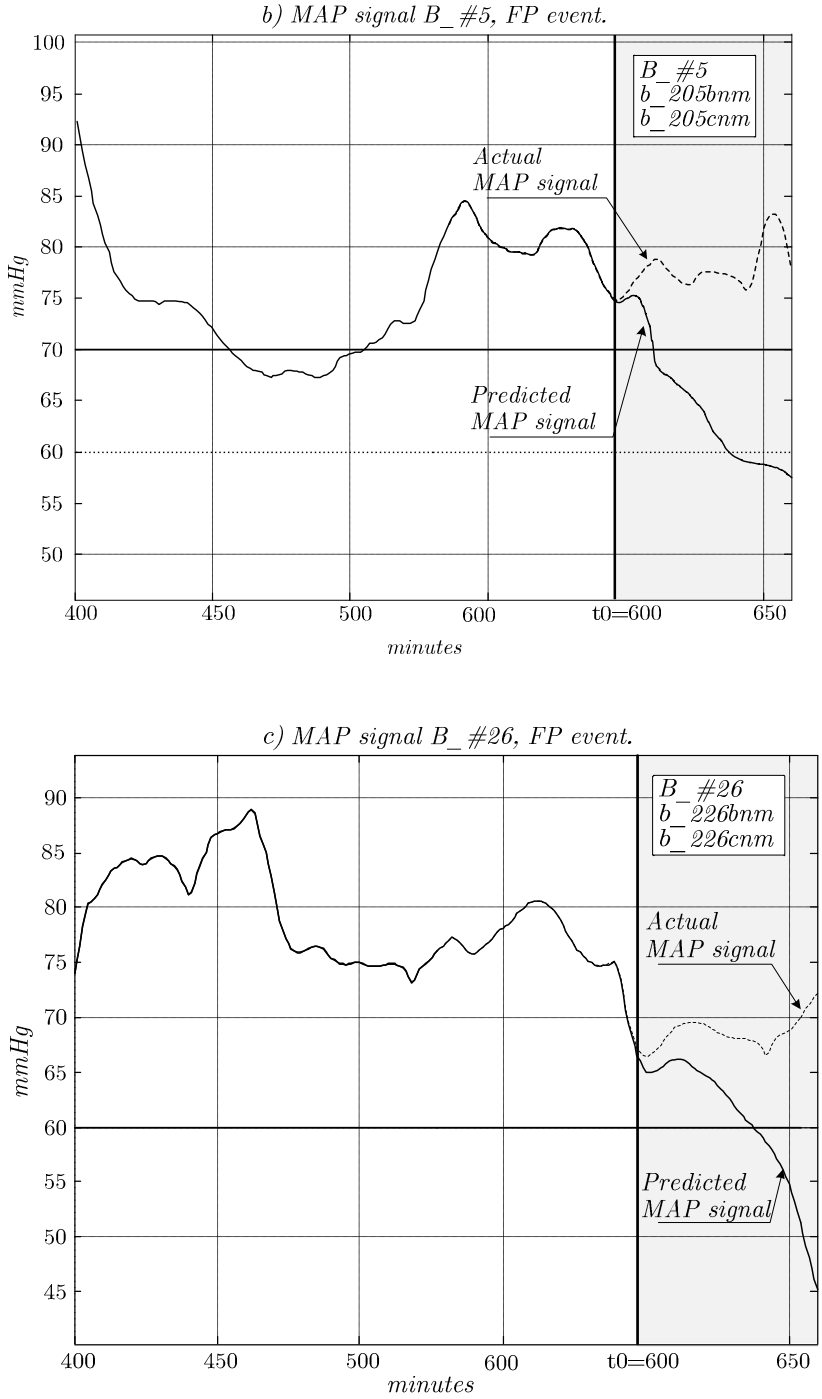


Figure 4.29 - MAP signals incorrectly classified: predicted and actual values.

Concerning the FN episode (record B_#24), as can be seen in Figure 4.29*a*), actual MAP signal presents a sudden drop approximately at instant (t_0 +35) minutes. Although the multi-model strategy was able to capture the evolution trend of the signal, it was not fast enough to identify the AHE event.

Regarding the FP episodes (records B_#5 and B_#26), there is not an obvious justification for the verified situations. Since the proposed strategy only uses MAP history and no clinical information (observations, medication), possibly for these particular cases other sources of information (such as clinical record data) should be considered in order to achieve a correct prediction. On the other hand, given that the prediction scheme is based on the set of representative signals, the evolutions of these particular templates can be interpreted as uncommon behaviours.

In conclusion, although the obtained results are relevant, the experiments performed have suggested that additional clinical information, such as medication, should be considered in a future implementation.

4.4.6 Conclusions

In this section, generalized regression neural network models, integrated into a multi-model structure, were proposed to address time-series prediction over a forecast horizon. Although this is a generic technique may be with potential to be employed in different areas, it was validated in the prediction of acute hypotensive episodes. The methodology consists of two steps: in the first, a similarity analysis procedure is carried out between the current signal and a representative set of historical evolution trends signals. The most similar ones are identified and the corresponding prediction neural models, previously trained using those historical signals, selected. In the second step this multi-models structure is employed to the current signal to predict its future evolution.

Applied to the mean arterial blood pressure (MAP) time-series, considered in the 2009 Physionet/Computers in Cardiology challenge, the referred strategy allowed to adequately capture MAP evolution and, consequently, to detect the occurrence of hypotensive episodes. In this context, a correct prediction of 10 out of 10 AHE for event 1, and of 37 out of 40 AHE for event 2, was achieved (SE=94.74% and SP=93.55%), enabling to obtain the best results of all entries in the two events of the challenge.

4.5 Trend Prediction of Blood Pressure Signals

High blood pressure or hypertension, one of the leading public health problems, is among the top most factors associated with cardiovascular diseases. In fact, uncontrolled and prolonged elevation of blood pressure can lead to a multiplicity of alterations in the myocardial structure, coronary vasculature, and conduction system of the heart, which can lead, among others, to the development of left ventricular hypertrophy, coronary artery disease, myocardial infarction, cardiac arrhythmias and heart failure (Who, 2011).

On the other hand, hypertension is an important modifiable risk factor, in particular for coronary artery disease and congestive heart failure (Perry, 1998), (Kannel, 1996). Thus, if blood pressure signals are continuously available, the analysis of their evolution may help clinicians in the management of the patient, namely in making decisions and in the adjustment of the care plan (Mancia *et al.*, European guidelines for the management of arterial hypertension, 2007), (Sierra and Sierra, 2008).

According to Mancia *et al.* (2007), being cardiovascular diseases a dynamic process, an early detection of hypertensive conditions and, consequently, adequate cardiovascular risk stratification, appears to be one of the most important strategies in order to prevent the development of CV diseases. In truth, the analysis of the future evolution of blood pressure can be of valuable importance considering that the earlier the treatment the better the prognosis of the patient (Chughtai and Peixoto, 2003).

Therefore, an important procedure to help reducing the burden of hypertension-related diseases is providing patients with adequate mechanisms for managing blood pressure (Bosworth and Oddone, 2004), (Marchiando and Elston, 2003), Bosworth and Oddone (2004). In this context, telemedicine solutions play a decisive role, such as the ones offered by HeartCycle.

The present section focuses on the analysis of blood pressure (BP) signals daily collected by means of a telemonitoring platform. The main goal is to assess the effectiveness of the developed predictive strategy to estimate the evolution trend of such signals, namely in case the patient is facing a potential hypertension condition. To achieve this objective, signals collected during the TEN-HMS and MyHeart studies are employed in two set of experiments.

The first group assesses the capacity of the proposed wavelet multi-resolution scheme in the trend prediction of blood pressure signals. Moreover, the performance of this scheme is compared with other typical prediction strategies, namely a linear regression model, the autoregressive integral moving average model - ARIMA, and two non-linear regression models, the generalized regression neural network - GRNN and the support vector regression – SVR.

The second set of experiments selects patients with blood pressure values in a critical range (around the threshold of hypertension), and uses the previously estimated trend to determine the risk of hypertension. Specifically, the goal is to evaluate whether during the following week the blood pressure signal of a given patient evolves towards hypertension values or, on the contrary, is maintaining or decreasing to normal values[12].

The remainder of this section is organized as follows. In subsection 4.5.1, some relevant definitions and information about the hypertension condition are introduced, as well as a brief presentation of the datasets (TEN-HMS and MyHeart) used in the experiments. In subsection 4.5.2, the effectiveness of the proposed wavelet multi-resolution scheme to estimate trends is assessed and compared with other common methodologies. In turn, subsection 4.5.3 shows the application of the predicted trends in checking the evolution of blood pressure signals to hypertension values. Finally, in subsection 4.5.4, some conclusions are outlined.

4.5.1 Introduction

The precise definition of hypertension is difficult to find and, therefore, the threshold for this condition should be considered flexible, based on the level and profile of the patient's cardiovascular risk (Mancia *et al.*, European guidelines for the management of arterial hypertension, 2007). For example, a blood pressure value may be considered as unacceptably high for patients in high risk state, but still acceptable for low risk patients.

On the other hand, the diagnosis of hypertension is made by measuring blood pressure and checking if levels are above the normal limit. However, a high value alone is not synonymous of disease. In fact, an hypertensive individual has necessarily to present elevated levels in a series of evaluations. Nevertheless, it is the responsibility of the clinician to diagnose the disease, that is, to decide if the individual suffers from hypertension not only based on blood pressure measurements, but also on the patient's history.

Typically in clinical practice, the minimum hypertension levels are defined as 140/90 mm Hg (systolic BP/diastolic BP) in awake, and 125/80 mm Hg during sleep. In case of ambulatory blood pressure, these values are considered slightly lower, as shown in Table 4.11. It should be noted that for the present work only systolic blood pressure signals were considered.

[12] www.min saude.pt/portal/conteudos/enciclopedia+da+saude/doencas/doencas+do+aparelho+circulatorio/hipertensao+arterial.htm

Table 4.11

Ambulatory blood pressure vs. clinic blood pressure thresholds.

Hypertension	Clinical blood pressure (mmHg)	Ambulatory blood pressure (mmHg)
Daily	140/90	135/85
Nigh time	125/80	120/75

The selection of a forecasting horizon is an essential step when predicting time-series data. From a clinical perspective, this period should be longer enough to allow a timely intervention in order to avoid an undesirable outcome. From the prediction perspective, the period should be as shorter as possible, since trends in data may not persist for too long. Taking into account these aspects, a forecast period of approximately one week (eight days, $P = 8$) was stipulated. In terms of the length of the template, that is, the past information used in the prediction, the value suggested by the clinical partners was about a month ($N = 32$). As described in subsection 4.5.2, this number was confirmed to be adequate through a set of experiments. Moreover, the values of N and P are powers of two, to simplify the wavelet transform operations.

TEN-HMS dataset

The Trans-European Network Homecare Monitoring Study (TEN-HMS) was designed to assess whether home based telemonitoring could reduce morbidity and mortality in patients with heart failure, compared with usual care or regular telephone contact. In this study, a total of 426 patients with a recent admission due to HF and left ventricular ejection fraction $<40\%$, were assigned randomly to home telemonitoring (168), nurse telephone support (173) and usual care (85). Particularly, home telemonitoring consisted of twice-daily self-measuring of weight, blood pressure, heart rate and rhythm, with automated devices linked to a cardiology centre, during the period of two years.

For the present work, a subsection of the complete dataset containing data from 83 patients was made available. In terms of strategy validation, only blood pressure signals were employed. Furthermore, only patients for whom there were BP measurements in, at least, 150 days (5 months) were selected for this purpose, resulting in a total of 51 patients.

MyHeart dataset (FP7- IST-2002-507816)

The MyHeart vision is of a home telemonitoring system that follows the health of the heart failure patient, enabling intervention when appropriate. This is done by monitoring vital body signs with wearable technology, processing the measured data and giving recommendations (when appropriate) to the patient and professional users of the system. Using the measured data to give user feedback, the system "closes the loop" of measurement and therapy.

This system was used in a clinical observational study carried out with 148 patients from six clinical centres in Germany and Spain. The trial had an enrolment phase of 9 months with 12 months of patient follow up. During the clinical study patients were requested to daily measure weight, blood pressure, and, using a vest, HR and bioimpedance, as well as heart rate, breathing rate and activity during the night by means of a bed sensor. Moreover, they were requested to complete each day two questionnaires of symptoms and mood/general well-being. From the 148 patients recruited, 102 (69%) were considered analyzable, that is, with more than 30 days of telemonitoring measurements.

Data pre-processing

Firstly, a pre-processing procedure is applied to the original BP signals collected through the telemedicine platform. The result of this procedure is a historic dataset of blood pressure time series signals, with a value per day (sampling rate of 1 day).

This process involves: *i*) averaging of values (in case two measurements have been performed in the same day), *ii*) dealing with missing values and, *iii*) noise reduction. To deal with missing values a simple procedure is carried out, analogous to the one presented in subsection 4.4.4. Fundamentally, a simple first order linear interpolation is performed. However, if the number of consecutive missing values exceeded five days the corresponding patient was discarded. For noise reduction, a simple first order filter was used. Additionally, the wavelet transform incorporated in the prediction scheme naturally deals with noisy signals.

4.5.2 Analysis of Evolution Trend

In a first group of experiments, the capability of the proposed wavelet scheme in predicting trends of blood pressure signals is assessed and compared with other regression models, namely ARIMA, GRNN and SVR. As mentioned, the main strategy is supported on the idea that similar patterns in the historic dataset can be used in the prediction of the current trend. In fact, when considering a given time instant in the past, it is possible to access its past and "future". The hypothesis to test here is that this information (past and "future") can be helpful in the prediction of the current signal.

1. Basics

The experimental process runs as follows (illustrated in Figure 4.30):

- **Templates**: for each patient in the dataset (a total of 51 patients), a template, that is, a signal with length $N = 32$, $X(t) \in \mathbb{R}^{1,N}$, is randomly generated;
- **Regression models**: the current template is used to estimate the ARIMA, GRNN and SVR parameters. These models are then used to iteratively predict future instants (as described in equation (3.17), subsection 3.3.2);

- **Similarity-based prediction**: from all patients in the historic dataset, the M most similar patterns, $Z_m(t) \in \mathbb{R}^{1,N+P}$, $m = 1,...,M$, are identified and used in the proposed strategy to predict the trend evolution $\widehat{Y}(t)$, over the forecast horizon ($P = 8$).

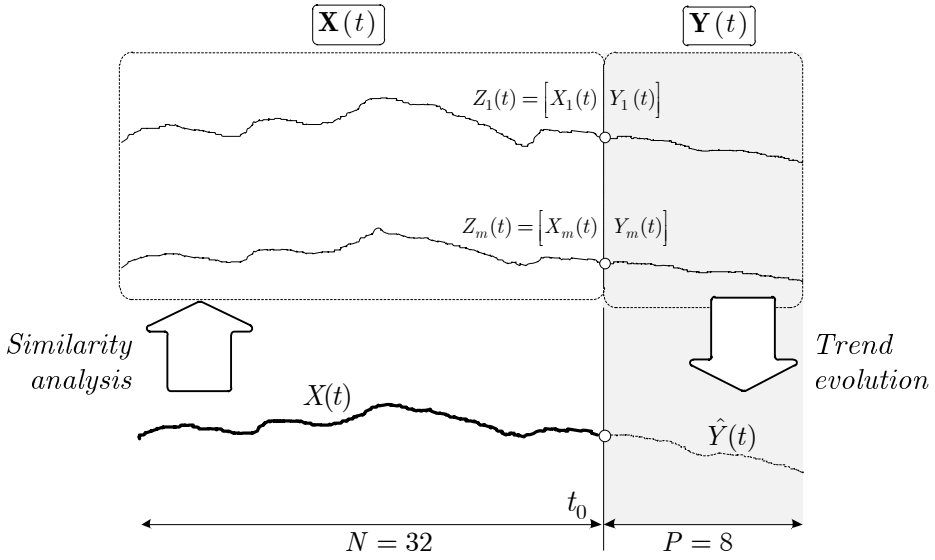


Figure 4.30 - Trend analysis strategy.

2. Predictive methods

Five methods were applied and compared in the prediction of future BP signals. The choice of the first method, ARIMA, is due to its sound theoretical basis and the numerous research publications based on it. In effect, the ARIMA model is, in theory, the most general class of linear models for forecasting a time series and has dominated many areas of time series prediction, including medicine. However, being a linear model, the approximation of ARIMA to complex non-linear problems may not be adequate. On the other hand, using neural networks, non-linear relationships embedded in the data can be captured. Moreover, neural networks are well known for their notable capabilities of approximation. Therefore, the second method to be implemented employs a neural network structure, in particular, a generalized regression neural network (GRNN). Support vector machines (SVMs) are support vector networks commonly used for classification. They present a structure similar to radial basis neural networks, being trained with a learning algorithm derived from statistical learning theory. The learning forms the key feature of a SVM, guaranteeing superior generalization properties. For regression purposes the corresponding structures are the support vector regression (SVR), which were included here. The fourth method simply considers the average value of signals $Y_m(t)$, as an estimation for the prediction of $\widehat{Y}(t)$. Finally, the last method implements the wavelet multi-resolution strategy proposed in subsection 3.4.3.

i. Auto regressive integral moving average model (ARIMA)

As mentioned in subsection 3.3.2, the build-up of a general ARIMA model requires a series of well-defined steps, namely: *i*) model identification; *ii*) estimation of the coefficients of the model and; *iii*) verification of the model.

The first stage involves the identification of the order of the model $ARIMA(n_a, d, n_c)$. The parameters n_a, d and n_c identify, respectively, the number of autoregressive terms, the degree of differencing and the number of lagged forecast errors in the prediction equation.

Degree of differencing (d): the first step is to determine the order of differencing needed to make the time series stationary. In other words, if the time series shows a strong trend (increase or decrease), then the process is clearly not stationary. The differencing operator can modify this by making the mean, variance, and autocorrelation constant over time. In practice, one or two degrees of differencing are often enough to reduce a non-stationary time series to an apparent stationary one. The estimation of the differencing order is performed by evaluating the autocorrelation of the time series. Typically, the correct degree of differencing is the lowest order that yields a time series fluctuating around a well-defined mean value and whose autocorrelation function decays fairly rapid to zero. From the experiments carried out with the blood pressure signals, considering $d=1$ and $d=2$, the value $d=1$ was selected. After been differencing, the ARIMA model results in an ARMA model, as described by (4.38), where the variable $dy(t)$ is the differencing operator defined as $dy(t) = y(t) - y(t-1)$.

$$dy(t) + a_1 dy(t-1) + ... + a_{na} dy(t-n_a) = e(t) + c_1 e(t-1) + ... + c_{nc} e(t-n_c) \tag{4.38}$$

Note that differencing a time series and computing the autocorrelation analysis can be done with the *diff*$(\cdot)$ and *autocorr*$(\cdot)$ Matlab commands, respectively.

Regressive and moving average orders (n_a and n_c): after a time series has been made stationary by differencing, the next step is to determine the AR or MA orders, respectively, n_a and n_c. The examination of the autocorrelation and partial autocorrelation functions of the differenced series, is used in the estimation of these parameters. Basically, a set of rules is tested (Tran and Reed, 2004), taking into account the way the autocorrelation and the partial autocorrelation functions decay over time. Experiments using blood pressure signals considering $n_a = 1,2,3$ and $n_c = 1,2,3$, suggested $n_a = n_c = 2$. As result, the ARIMA structure is ARIMA(2,1,2). The autocorrelation and partial autocorrelation analysis was done, respectively, with the *autocorr*$(\cdot)$ and *parcorr*$(\cdot)$ Matlab commands.

Estimating the coefficients of the model*:* once the model is established, the respective parameters can be estimated using a least mean square approach. In practice, the estimation of parameters was carried out with the *armax*$(\cdot)$ Matlab command. Moreover, to ensure that only models corresponding to stable predictors are tested, the Matlab algorithm performs a stability test of the predictor.

As conclusion, for all the templates, the structure was predetermined and given by ARIMA(2,1,2). However, for each template, a set of parameters $\{a_1, ..., a_{na}, c_1, ..., c_{nc}\}$ was computed.

ii. Generalized regression neural networks (GRNN)

Generalized regression neural networks, are similar to radial basis function networks. In fact, GRNN structures can be seen as normalized radial basis function networks, in which there is a hidden unit centred at every training case. These units are called "kernels" and, usually, are probability density functions, such as Gaussian functions. The weights from the hidden to output layer are just the target values, so the output is simply a weighted average of the target values of the training cases, close to the given input case. As a consequence, the only parameters to be learned are the widths of the units.

In the experiments using the blood pressure signals, the width of the kernels was experimentally determined as $\lambda = 0.2$. The *newgrnn*$(\cdot)$ MatLab command was used to implement this neural model. Moreover, a different neural network had to be trained for each template.

iii. Support vector regressions (SVR)

The implementation of the SVR prediction method was done through the *libsbm* framework[13], with a new model being trained for each template. The template (training data) was previously normalised in the range [−1,1]. The parameters for the training were (see subsection 3.3.3):

- *type of SVR:* epsilon-SVR
- *kernel type*: radial basis function
- *with of kernel function* (γ): 0.5
- cost (C): 1
- *tolerance of termination criterion* (ε): 0.001

iv. Average of the patterns (AVP)

Basically, the main hypothesis of the proposed prediction scheme is that future evolution of similar conditions identified in the historic dataset, support the prediction of the current condition. As a result, the forecasting capability of the weighted average, $\overline{Y}(t)$, of these identified patterns, computed as (4.39), was assessed. The variable $Y_m(t)$ defines each similar time series signal, and S_m is the corresponding similarity measure.

[13] www.csie.ntu.edu.tw/~cjlin/libsvm/

$$\overline{Y}(t) = \frac{\sum_{m=1}^{M} S_m Y_m(t)}{\sum_{m=1}^{M} S_m} \quad t = N+1,...,N+P \tag{4.39}$$

v. Proposed wavelet multi-resolution prediction scheme (WMM)

The last approach put into practice the proposed wavelet strategy, considering the following parameters (subsection 3.4.3):

Similarity analysis

- $N = 32$, $P = 8$, where N and P denote, respectively, the time intervals before and after the current time instant
- Number of patterns retrieved from the historic dataset, $M = 5$

Wavelet multi-resolution analysis

- Wavelet decomposition level, $L = 5$

Selection of the optimal trends

- Number of decompositions considered in the optimal trend selection $l = 3,4,5,6$ (the details are the levels $l = 3,4,5$; the approximation is the level $l = 6$);
- The first two levels of detail ($l = 1,2$) were neglected;
- Conjunction and aggregation operators were, respectively, the *maximum*$(\cdot)$ and the *product*$(\cdot)$ operators.

3. Prediction metrics

The accuracy of the forecasting methods was determined in terms of four performance metrics: *i*) the proposed similarity measure based on the wavelet decomposition+KLT (SWK), (4.40); *ii*) the Pearson's correlation coefficient (CORC), (4.41); *iii*) the normalised root mean squared error (NRMSE), (4.42) and *iv*) the mean absolute percentage error (MAPE).

$$SWK = S\left(Y(t), \widehat{Y}(t)\right) \quad t = N+1,...,N+P \tag{4.40}$$

$$CORC = \frac{\sum_{t=N+1}^{N+P} \left(Y(t) - \overline{Y}\right) \left(\widehat{Y}(t) - \overline{\widehat{Y}}\right)}{\sqrt{\sum_{t=N+1}^{N+P} \left(Y(t) - \overline{Y}\right)^2} \sqrt{\sum_{t=N+1}^{N+P} \left(\widehat{Y}(t) - \overline{\widehat{Y}}\right)^2}} \tag{4.41}$$

$$NRMSE = \frac{1}{P}\frac{\sum_{t=N+1}^{N+P}\left(Y(t)-\widehat{Y}(t)\right)^2}{\sum_{t=N+1}^{N+P}\left(Y(t)-\overline{Y}\right)^2} \tag{4.42}$$

$$MAPE = \frac{1}{P}\sum_{t=N+1}^{N+P}\left|\frac{Y(t)-\widehat{Y}(t)}{Y(t)}\right| \tag{4.43}$$

In the previous equations, $Y(t)$ is the actual BP value, $\widehat{Y}(t)$ is the forecasted BP, $\overline{Y}$ and $\overline{\widehat{Y}}$ are, respectively, the means of the actual and the estimated signals.

The metrics NRMSE and MAPE were transformed to $NRMSE = \exp(-\kappa_N\, NRMSE)$ and $MAPE = \exp(-\kappa_M\, MAPE)$, in order to guarantee that their values are in the range [0,1]. The parameters κ_N and κ_M are adequate constants, respectively, $\kappa_N = 0.25$ and $\kappa_M = 10$.

4. Comparison of prediction methods

Friedman test

The correct validation is a decisive and fundamental procedure when comparing the results of several methods. Among the available parametric and nonparametric tests, the Friedman test is a nonparametric one that enables to perform multiple comparisons in experimental studies. This test (Friedman, 1937), (Friedman, 1940) is equivalent to ANOVA and is particularly adequate for machine learning studies when the assumptions (independency, normality and homoscedasticity) do not hold or are difficult to verify for a parametric test (Garcia *et al.*, 2010).

The objective of the Friedman test is to determine if it is possible to conclude, from a set of results, that there is a difference among the several methods. The first step in calculating the test statistics is to convert the original results in ranks. Thus, for each problem (experiment), the methods are separately ranked, where the best performing method should have the rank of 1, the second best the rank of 2, etc. The Friedman test then compares the average ranks R_j of each method, computed as (4.44).

$$R_j = \frac{1}{n}\sum_i r_i^j \tag{4.44}$$

The variable r_i^j denotes the rank of the j^{th} of k algorithms on the i^{th} of n data experiments. Under the null hypothesis, which states that "*Ho*: *all the algorithms behave similarly and thus their ranks* R_j *should be equal*", the Friedman statistics, equation

(4.45), is distributed according to χ_F^2, with $k-1$ degrees of freedom, when n and k are big enough (as a rule of a thumb, $n>10$ and $k>5$).

$$\chi_F^2 = \frac{12n}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{4.45}$$

From the computation of the corresponding $p-value$, the null hypothesis can be or not rejected at a given level of significance.

Nemenyi test

The Nemenyi test enables a parwise comparison of the methods, based on the average ranks computed in the Friedman test. Basically, by means of the Nemenyi test, two methods are significantly different at $\alpha=1\%$, $\alpha=5\%$, or $\alpha=10\%$, if their average ranks differ at least the critical value CD, computed as (4.46).

$$CD_\alpha = q_\alpha\sqrt{\frac{k(k+1)}{6n}} \tag{4.46}$$

The possible values for parameter q_α are listed in Table 4.12, depending on the number of involved methods (k).

Table 4.12

Critical values for the two-tailed Nemenyi test.

#methods (k)	2	3	4	5	6	7	8	9	10
q_1	2.576	2.913	3.113	3.255	3.364	3.452	3.526	3.590	3.646
q_5	1.960	2.344	2.569	2.728	2.850	2.948	3.031	3.102	3.164
q_{10}	1.645	2.052	2.291	2.460	2.589	2.693	2.780	2.855	2.920

For example, considering five methods to be compared ($k=5$), the Nemenyi test table is obtained as depicted in Table 4.13, where R_j denotes the average rank of each method.

Table 4.13

Nemenyi test with k=5.

#methods (k)	2	3	4	5
1	R_2-R_1	R_3-R_1	R_4-R_1	R_5-R_1
2		R_3-R_2	R_4-R_2	R_5-R_2
3			R_4-R_3	R_5-R_3
4				R_5-R_4

Finally, the comparison of each value ($R_j - R_i$) with the obtained critical value CD_α enables to decide about how significantly different two methods are. Basically, a method (j) is significantly better at a level α than a method (i), if the relationship $(R_j - R_i > CD_\alpha)$ is verified.

5. Experiments with the TEN-HMS dataset

i. Variation in the number of the patterns (M)

The first group of experiments was performed in order to estimate the most adequate number of patterns to consider in the prediction strategy. To this aim, this number was successively increased from $M = 5$ to $M = 30$ with increments of 5 units, resulting in $M = \{ 5, 10, 15, 20, 25, 30\}$. For each value of M, a total of 140 experiments were performed. Figure 4.31 depicts the corresponding box-plots, with respect to the four performance metrics.

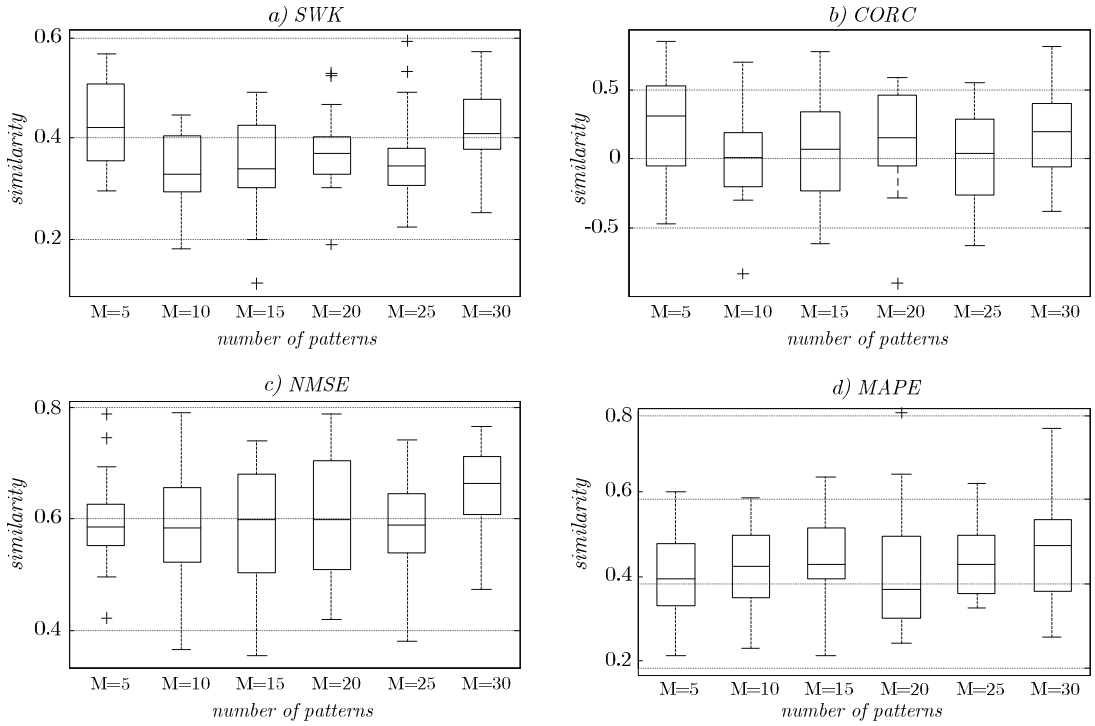


Figure 4.31 - Variation in the number of patterns: M={5, 10, 15, 20, 25, 30}.

As can been observed from the figure, the decision regarding the optimal number of patterns to be chosen is not clear. Observing all the similarity measures, it is not possible to identify an experiment that achieves superior results in comparison with the others. Moreover, for the same experiment, different variations occur for the different metrics. For example, in the particular case of M=5, higher variations occur in the SWK metrics and small variations are present in the NRMSE metrics.

i. Comparison of the number of patterns using the SWK metrics

The Table 4.14 compares the six possible values of the number of patterns, where the values in each row result from the average of 7 experiments (total of 20 groups from the 140 experiments). The values in the parentheses are used in the computation of the Friedman test to rank the methods.

Table 4.14

Comparison of the number of patterns (M) – SWK metrics.

	M=5	**M=10**	**M=15**	**M=20**	**M=25**	**M=30**
Group1	0.562 (6)	0.299 (2)	0.113 (1)	0.367 (3)	0.389 (4)	0.473 (5)
Group2	0.470 (6)	0.420 (5)	0.201 (1)	0.319 (3)	0.377 (4)	0.288 (2)
Group3	0.426 (3)	0.449 (5)	0.311 (1)	0.531 (6)	0.323 (2)	0.436 (4)
Group4	0.332 (2)	0.318 (1)	0.492 (5)	0.524 (6)	0.355 (3)	0.373 (4)
Group5	0.297 (2)	0.366 (4)	0.240 (1)	0.354 (3)	0.492 (6)	0.481 (5)
Group6	0.362 (2)	0.414 (4)	0.398 (3)	0.467 (5)	0.318 (1)	0.572 (6)
Group7	0.468 (6)	0.318 (1)	0.455 (5)	0.372 (4)	0.372 (3)	0.350 (2)
Group8	0.550 (6)	0.448 (5)	0.375 (3)	0.377 (4)	0.286 (1)	0.344 (2)
Group9	0.389 (4)	0.401 (5)	0.333 (2)	0.189 (1)	0.534 (6)	0.385 (3)
Group10	0.547 (6)	0.388 (3)	0.349 (2)	0.445 (5)	0.299 (1)	0.397 (4)
Group11	0.321 (2)	0.293 (1)	0.383 (4)	0.337 (3)	0.593 (6)	0.509 (5)
Group12	0.558 (6)	0.313 (3)	0.301 (2)	0.402 (5)	0.337 (4)	0.252 (1)
Group13	0.350 (4)	0.292 (1)	0.319 (3)	0.406 (5)	0.295 (2)	0.428 (6)
Group14	0.393 (5)	0.210 (1)	0.376 (4)	0.363 (3)	0.327 (2)	0.485 (6)
Group15	0.333 (3)	0.388 (4)	0.492 (6)	0.308 (1)	0.332 (2)	0.412 (5)
Group16	0.454 (4)	0.338 (3)	0.476 (5)	0.303 (2)	0.224 (1)	0.530 (6)
Group17	0.568 (6)	0.425 (4)	0.308 (1)	0.323 (2)	0.359 (3)	0.446 (5)
Group18	0.422 (6)	0.182 (1)	0.303 (3)	0.384 (4)	0.29 (2)	0.387 (5)
Group19	0.379 (5)	0.195 (1)	0.294 (2)	0.334 (3)	0.356 (4)	0.408 (6)
Group20	0.467 (5)	0.213 (1)	0.479 (6)	0.398 (3)	0.417 (4)	0.383 (2)
Average rank	**4.45**	**2.75**	**3.00**	**3.55**	**3.05**	**4.20**

From the average of ranks, the value $\chi_F^2 = 13.77$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected at a significance level of 5% ($p - value = 0.0171$). Then, the Nemenyi test was performed to compare the number of patterns, based on the computed average ranks, which results are presented in Table 4.5. The performance of the methods is significantly different at levels of 1%, 5% and 10%, corresponding to critical values of $CD_1 = 1.9902$, $CD_5 = 1.6861$ and $CD_{10} = 1.5317$, respectively.

Table 4.15

Nemenyi test – SWT metric.

#patterns	M=10	M=15	M=20	M=25	M=30
M=5	-1.70 **	-1.45	-0.90	-1.40	-0.25
M=10		0.25	0.80	0.30	1.45
M=15			0.55	0.05	1.20
M=20				-0.50	0.65
M=25					1.15

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

From the Table 4.15 it can be observed that only the case $M=5$ is superior to the $M=10$ (at the level of 5%). No other cases of superiority are verified.

ii. Comparison of the number of patterns using all metrics

The Table 4.16*a*) and Table 4.16*b*) summarize the average ranks and the respective values of qui-square and *p-value*.

Table 4.16

Comparison of the number of patterns (M) – all metrics.

	M=5	M=10	M=15	M=20	M=25	M=30
SWK	4.45	2.75	3.00	3.55	3.05	4.20
CORC	4.10	3.20	3.40	3.50	3.00	3.80
NRMSE	3.35	3.20	3.55	3.65	3.05	4.20
MAPE	3.25	3.50	3.35	3.05	3.45	4.40

a) Average ranks.

	χ_F^2	$p-value$
SWK	13.77	0.0171
CORC	4.57	0.4704
NRMSE	4.74	0.4481
MAPE	6.29	0.2794

b) Qui-square and p-value.

From the Table 4.16*b*) it can be observed that only for the SWK measure the null hypothesis can not be rejected (and, in this case, only $M=5$ is superior to $M=10$ at the level of 5%). Consequently, it can be concluded that the selection of the number of patterns is not decisive for the performance of the prediction strategy. Thus, mainly due to simplicity reasons, a reduced number of patterns was selected. This number ($M=5$) will be employed as the basis for the next experiments.

On the other hand, these results are understandable. In effect, the prediction strategy is mainly based on the identification of the representative trends, obtained from a

clustering process that considers the patterns retrieved from the historic dataset. Thus, by means of the referred clustering process, a representative trend corresponds to a region where the patterns are more concentrated (the subtractive clustering is a density based method). As a result, every time new patterns are added to the clustering process, they do not affect significantly the determination of the centre (assuming that the similarity with respect to the template decreases as the number of the pattern to be compared increases). This aspect is strongly related with the robustness of the clustering process: outliers, that is, patterns with lower similarities with the template, tend to be neglected.

ii. Variation in the number of days before the current instant (N)

As referred, from the clinical perspective, the use of one month before the current time instant is considered reasonable to perform the desired prediction (one week). These experiments aim to investigate the adequateness of this value. For the effect, the number of days before the current instant was chosen to be $N = \{\ 8, 16, 32, 64\}$ (a power of 2 to simplify the wavelet decomposition process). For each value of N, a total of 140 experiments were performed. The Figure 4.32 depicts the corresponding box-plots with respect to the four performance measures.

The performance of the methods is significantly different at the levels of 1%, 5% and 10%, corresponding to critical values of $CD_1 = 1.2709$, $CD_5 = 1.0488$ and $CD_{10} = 0.9353$, respectively.

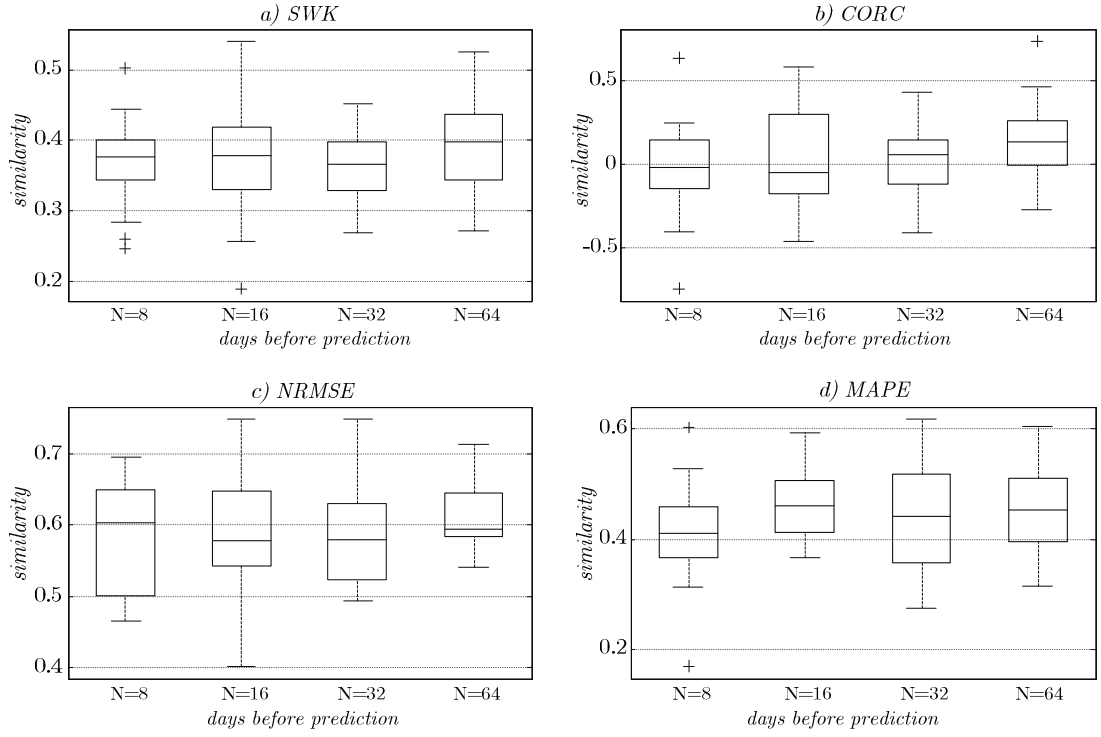


Figure 4.32 - Variation in the number of days before prediction: N={8, 16, 32, 64}.

From the analysis of Figure 4.32, the decision about the most adequate number of previous days is not obvious. In fact, the median values obtained for each metric are similar, even though variations are different.

i. Comparison of the number of previous days using the SWK metrics

The Table 4.17 compares the four N values, where each row results from the average of 7 experiments (total of 140 experiments).

Table 4.17

Comparison of SWK among the different number of previous days (N).

	N=8	**N=16**	**N=32**	**N=64**
Group1	0.444 (4)	0.351 (2)	0.267 (1)	0.371 (3)
Group2	0.368 (2)	0.516 (4)	0.363 (1)	0.392 (3)
Group3	0.350 (3)	0.282 (2)	0.277 (1)	0.360 (4)
Group4	0.501 (4)	0.440 (2)	0.324 (1)	0.489 (3)
Group5	0.245 (1)	0.374 (2)	0.400 (3)	0.522 (4)
Group6	0.393 (3)	0.357 (1)	0.383 (2)	0.409 (4)
Group7	0.298 (1)	0.328 (2)	0.451 (3)	0.464 (4)
Group8	0.352 (1)	0.463 (4)	0.420 (3)	0.412 (2)
Group9	0.403 (2)	0.409 (4)	0.349 (1)	0.405 (3)
Group10	0.381 (3)	0.378 (2)	0.396 (4)	0.290 (1)
Group11	0.337 (1)	0.538 (4)	0.365 (2)	0.465 (3)
Group12	0.372 (2)	0.300 (1)	0.398 (4)	0.392 (3)
Group13	0.374 (3)	0.368 (2)	0.328 (1)	0.451 (4)
Group14	0.407 (3)	0.420 (4)	0.363 (2)	0.342 (1)
Group15	0.374 (2)	0.414 (4)	0.378 (3)	0.344 (1)
Group16	0.444 (4)	0.351 (2)	0.267 (1)	0.371 (3)
Group17	0.368 (2)	0.516 (4)	0.363 (1)	0.392 (3)
Group18	0.350 (3)	0.282 (2)	0.277 (1)	0.360 (4)
Group19	0.501 (4)	0.440 (2)	0.324 (1)	0.489 (3)
Group20	0.245 (1)	0.374 (2)	0.400 (3)	0.522 (4)
Average rank	**2.45**	**2.50**	**2.25**	**2.80**

From the average of ranks, the value $\chi_F^2 = 1.86$ was obtained, equation (4.45), with a $p-value$ of 0.602. As result, the null hypothesis "*Ho: all the methods behave similarly*" could not be rejected.

ii. Comparison of the number of previous days using all metrics

The Table 4.18*a*) and Table 4.18*b*) summarize the average ranks and the respective values of qui-square and *p-value*.

Table 4.18

Comparison of the number of previous days (N) – all metrics.

	N=8	N=16	N=32	N=64
SWK	2.45	2.50	2.25	2.80
CORC	2.30	2.55	2.55	2.60
NRMSE	2.50	2.35	2.40	2.75
MAPE	2.05	2.80	2.55	2.60

a) Average ranks.

	χ_F^2	$p-value$
SWK	1.86	0.602
CORC	0.66	0.882
NRMSE	1.14	0.767
MAPE	3.66	0.306

b) Qui-square and p-value.

From Table 4.18*b*) it is clear that the null hypothesis has to be accepted for all the metrics. In effect, it was not possible to find, with a reasonable degree of confidence, a value that contradicts the suggested number of previous days ($M = 32$). Therefore, this number was accepted.

iii. Comparison of the prediction methods (ARIMA, GRNN, SVR, AVP, WMM)

This set of experiments aimed at the comparison of the proposed prediction method against the other strategies. For the effect, a total of 300 experiments were performed ($M = 5$, $N = 32$, $P = 8$). The Figure 4.33 depicts the box-plot resulting from the comparisons.

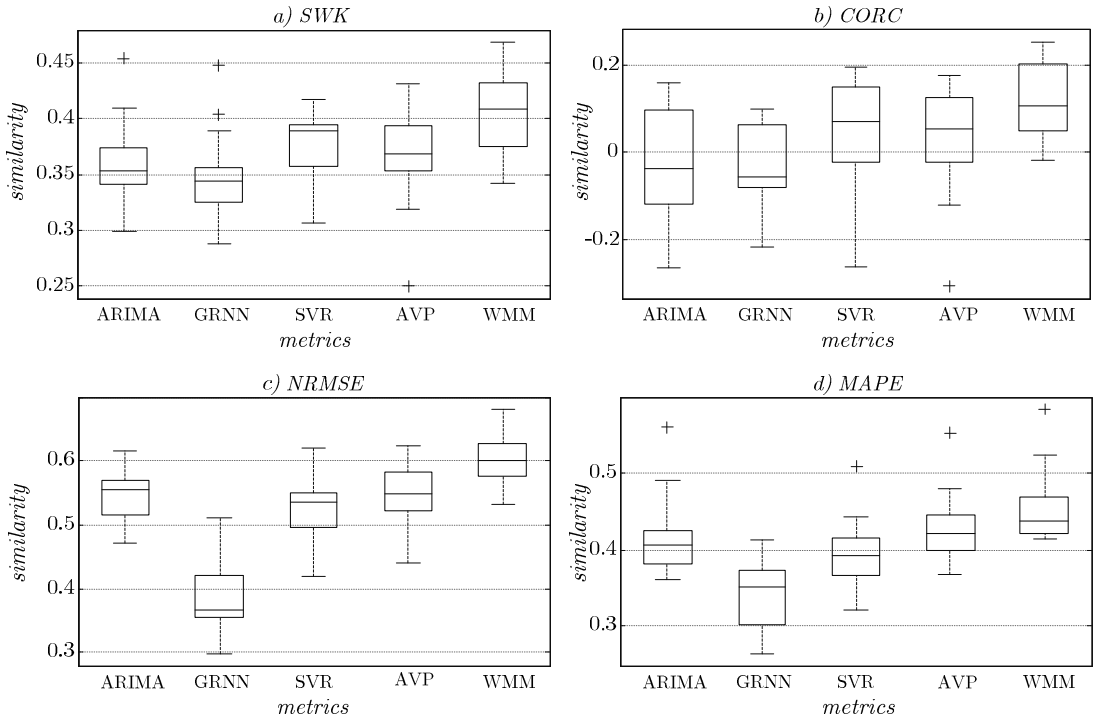


Figure 4.33 - Comparison of the prediction methods (SWK, CORC,NRMSE, and MAPE metrics) using TEN-HMS dataset.

From the analysis of Figure 4.33 and, in global terms, it appears that the proposed method is superior to the others. In effect, the wavelet based prediction method (WMM) presents the highest median for all the metrics showing, however, a higher variability for some of these metrics.

Three of the methods (ARIMA, GRNN, and SVR) compute the prediction based on an iterative approach: a one-step ahead model is iteratively applied during P times, being the current predictions used by the model in order to obtain the next forecast. The last two methods (AVP and WMM) do not involve the explicit computation of a model, thus, they are, to some extent, similar to a direct approach. This fact can justify why GRNN and ARIMA present poor results for some metrics (namely SWK and NRMSE). Additionally, since the training of the respective parameters is only based on the data from the template ($N = 32$), generalization problems can occur. On the other hand, the SVR method partially mitigates this drawback through its particular training process (probably the main advantage recognized for SVR).

In order to accurately compare the predictive methods, the Friedman test was implemented, considering the four metrics.

i. Comparison of the methods using the SWK metric

The Table 4.19, corresponding to the Figure 4.33*a*), compares the five methods according to the SWK metrics, where the values in each row result from the average of groups of 15 experiments (total of 300 experiments). The ranks in parentheses are used in the computation of the Friedman test to evaluate the methods.

Table 4.19

Comparison of SWK among the five prediction methods (TEN-HMS).

	ARIMA	**GRNN**	**SVR**	**AVP**	**WMM**
Group1	0.341 (1)	0.403 (4)	0.407 (5)	0.352 (2)	0.368 (3)
Group2	0.365 (4)	0.300 (1)	0.306 (2)	0.345 (3)	0.386 (5)
Group3	0.410 (5)	0.354 (2)	0.332 (1)	0.355 (3)	0.369 (4)
Group4	0.355 (2)	0.292 (1)	0.396 (4)	0.383 (3)	0.422 (5)
Group5	0.347 (1)	0.350 (2)	0.393 (4)	0.380 (3)	0.430 (5)
Group6	0.358 (2)	0.351 (1)	0.387 (3)	0.417 (4)	0.448 (5)
Group7	0.351 (2)	0.325 (1)	0.383 (3)	0.431 (4)	0.468 (5)
Group8	0.346 (3)	0.321 (1)	0.364 (4)	0.370 (5)	0.342 (2)
Group9	0.355 (4)	0.347 (2)	0.351 (3)	0.250 (1)	0.391 (5)
Group10	0.341 (2)	0.348 (3)	0.393 (5)	0.318 (1)	0.362 (4)
Group11	0.298 (1)	0.340 (2)	0.392 (3)	0.411 (4)	0.462 (5)
Group12	0.330 (1)	0.339 (2)	0.417 (5)	0.393 (4)	0.389 (3)
Group13	0.368 (4)	0.326 (1)	0.347 (2)	0.356 (3)	0.424 (5)
Group14	0.392 (5)	0.325 (1)	0.377 (3)	0.361 (2)	0.380 (4)
Group15	0.346 (1)	0.447 (5)	0.365 (2)	0.367 (3)	0.411 (4)
Group16	0.337 (1)	0.388 (2)	0.389 (3)	0.393 (4)	0.434 (5)
Group17	0.452 (5)	0.340 (1)	0.402 (2)	0.406 (3)	0.449 (4)
Group18	0.337 (2)	0.288 (1)	0.391 (4)	0.364 (3)	0.421 (5)
Group19	0.390 (4)	0.357 (2)	0.329 (1)	0.385 (3)	0.404 (5)
Group20	0.378 (4)	0.366 (3)	0.398 (5)	0.323 (1)	0.347 (2)
Average rank	**2.70**	**1.90**	**3.20**	**2.95**	**4.25**

From the average of ranks, the value $\chi_F^2 = 23.24$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p - value = 0.0001$).

In the next phase, the Nemenyi test was performed to compare the methods based on the computed average ranks, which results are presented in Table 4.20. The performance of the methods is significantly different at the levels of 1%, 5% and 10%, corresponding to critical values of $CD_1 = 1.6275$, $CD_5 = 1.3640$ and $CD_{10} = 1.2300$, respectively.

Table 4.20

Nemenyi test (SWK metrics - TEN-HMS).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	-0.80	0.50	0.25	1.55 **
GRNN		1.30 *	1.05	2.35 ***
SVR			-0.25	1.05
AVP				1.30 *

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

From the table, it can be concluded that the proposed WMM method outperforms ARIMA, GRNN and AVP predictors, respectively at the levels of 5%, 1%, and 10%. In turn, at a level of 10%, the SVR outperforms the GRNN method.

ii. Comparison of the methods using the CORC metric

The Table 4.21, corresponding to the Figure 4.33b), compares the five methods using the CORC with the same groups of experiments.

Table 4.21

Comparison of CORC among the five prediction methods (TEN-HMS).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	-0.164 (1)	0.095 (3)	0.150 (5)	-0.116 (2)	0.112 (4)
Group2	-0.016 (3)	-0.152 (1)	-0.105 (2)	-0.014 (4)	0.062 (5)
Group3	0.040 (2)	0.022 (1)	0.091 (4)	0.073 (3)	0.103 (5)
Group4	0.109 (4)	-0.159 (1)	0.040 (3)	0.150 (5)	0.015 (2)
Group5	-0.141 (1)	-0.069 (2)	0.177 (4)	-0.017 (3)	0.190 (5)
Group6	0.130 (3)	0.079 (1)	0.099 (2)	0.178 (4)	0.253 (5)
Group7	-0.097 (1)	-0.071 (2)	0.153 (3)	0.169 (4)	0.191 (5)
Group8	-0.076 (1)	-0.056 (2)	-0.019 (3)	0.086 (5)	0.059 (4)
Group9	0.027 (4)	0.101 (5)	0.008 (3)	-0.307 (1)	-0.020 (2)
Group10	-0.062 (2)	0.048 (5)	-0.023 (3)	-0.121 (1)	-0.006 (4)
Group11	-0.266 (1)	-0.047 (2)	0.196 (4)	0.177 (3)	0.235 (5)
Group12	-0.179 (2)	-0.217 (1)	0.169 (5)	0.106 (4)	0.081 (3)
Group13	0.090 (4)	-0.059 (2)	-0.140 (1)	-0.002 (3)	0.248 (5)
Group14	-0.061 (2)	-0.114 (1)	0.067 (5)	-0.021 (3)	0.036 (4)
Group15	-0.157 (1)	0.098 (4)	-0.049 (2)	0.073 (3)	0.209 (5)
Group16	-0.091 (1)	0.030 (2)	0.075 (3)	0.148 (4)	0.228 (5)
Group17	0.162 (4)	-0.084 (1)	0.122 (3)	0.035 (2)	0.195 (5)
Group18	0.093 (3)	-0.069 (1)	0.163 (4)	0.075 (2)	0.164 (5)
Group19	0.101 (5)	0.088 (4)	-0.262 (1)	-0.023 (2)	0.073 (3)
Group20	0.098 (5)	-0.076 (1)	-0.012 (4)	-0.047 (2)	-0.018 (3)
Average rank	**2.50**	**2.10**	**3.20**	**3.00**	**4.20**

From the average of ranks, the value $\chi_F^2 = 30.32$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p - value = 0.0004$). In the next phase, the Nemenyi test was performed to compare the methods based on the computed average ranks, which results are presented in Table 4.22.

Table 4.22
Nemenyi test (CORC metrics - TEN-HMS).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	-0.40	0.70	0.50	1.70***
GRNN		1.10	0.90	2.10***
SVR			-0.20	1.00
AVP				1.20

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

It can be concluded that the proposed WMM method outperforms the ARIMA and the GRNN predictors at the level of 1%.

iii. Comparison of the methods using the NRMSE metrics

The Table 4.23, compares the five methods according to the *NRMSE* metrics.

Table 4.23
Comparison of NRMSE among the five prediction methods (TEN-HMS).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.554 (4)	0.512 (1)	0.533 (3)	0.517 (2)	0.578 (5)
Group2	0.519 (4)	0.299 (1)	0.428 (2)	0.441 (3)	0.560 (5)
Group3	0.560 (4)	0.414 (1)	0.494 (2)	0.546 (3)	0.574 (5)
Group4	0.473 (2)	0.334 (1)	0.535 (3)	0.586 (4)	0.637 (5)
Group5	0.557 (2)	0.453 (1)	0.581 (3)	0.623 (4)	0.674 (5)
Group6	0.615 (4)	0.358 (1)	0.552 (2)	0.566 (3)	0.667 (5)
Group7	0.531 (3)	0.311 (1)	0.518 (2)	0.575 (4)	0.609 (5)
Group8	0.578 (4)	0.463 (1)	0.541 (2)	0.605 (5)	0.574 (3)
Group9	0.473 (3)	0.345 (1)	0.491 (4)	0.473 (2)	0.600 (5)
Group10	0.544 (3)	0.402 (1)	0.549 (4)	0.518 (2)	0.579 (5)
Group11	0.507 (2)	0.359 (1)	0.582 (4)	0.581 (3)	0.681 (5)
Group12	0.511 (2)	0.364 (1)	0.620 (4)	0.625 (5)	0.610 (3)
Group13	0.511 (4)	0.406 (1)	0.419 (2)	0.464 (3)	0.539 (5)
Group14	0.572 (4)	0.363 (1)	0.551 (3)	0.549 (2)	0.597 (5)
Group15	0.539 (3)	0.398 (1)	0.537 (2)	0.545 (4)	0.615 (5)
Group16	0.564 (3)	0.494 (1)	0.565 (4)	0.524 (2)	0.635 (5)
Group17	0.559 (4)	0.370 (1)	0.520 (2)	0.549 (3)	0.619 (5)
Group18	0.569 (4)	0.353 (1)	0.542 (3)	0.537 (2)	0.592 (5)
Group19	0.590 (4)	0.427 (1)	0.474 (2)	0.589 (3)	0.602 (5)
Group20	0.586 (5)	0.357 (1)	0.497 (2)	0.525 (3)	0.534 (4)
Average rank	**3.40**	**1.00**	**2.75**	**3.10**	**4.75**

From the average of ranks, the value $\chi_F^2 = 58.36$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p-value = 6.4e-12$). In the next phase, the Nemenyi test was performed to compare the methods based on the computed average ranks, which results are presented in Table 4.24.

Table 4.24

Nemenyi test (NRMSE metrics - TEN-HMS).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	-2.40***	-0.65	-0.30	1.35*
GRNN		1.75***	2.10***	3.75***
SVR			0.35	2.00***
AVP				1.65***

** at a significance level of 10%, ** at a significance level of 5%, *** at a significance level of 1%*

From the Table 4.24, it can be concluded that the proposed WMM method outperforms GRNN, SVR and AVP predictors at a levels of 1%. It also outperforms the ARIMA method at a significance level of 10%.

iv. Comparison of the methods using the MAPE metrics

The Table 4.25, compares the five methods according to the MAPE metrics.

Table 4.25

Comparison of MAPE among the five prediction methods (TEN-HMS).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.361 (1)	0.394 (3)	0.416 (4)	0.392 (2)	0.422 (5)
Group2	0.372 (4)	0.303 (1)	0.330 (2)	0.369 (3)	0.420 (5)
Group3	0.423 (4)	0.341 (1)	0.376 (2)	0.420 (3)	0.436 (5)
Group4	0.389 (2)	0.363 (1)	0.400 (3)	0.452 (4)	0.472 (5)
Group5	0.405 (3)	0.362 (1)	0.396 (2)	0.433 (4)	0.477 (5)
Group6	0.560 (4)	0.411 (1)	0.510 (2)	0.554 (3)	0.584 (5)
Group7	0.388 (3)	0.264 (1)	0.377 (2)	0.421 (4)	0.436 (5)
Group8	0.428 (3)	0.380 (1)	0.422 (2)	0.447 (5)	0.440 (4)
Group9	0.382 (2)	0.281 (1)	0.387 (3)	0.404 (4)	0.430 (5)
Group10	0.380 (2)	0.368 (1)	0.387 (3)	0.405 (4)	0.446 (5)
Group11	0.400 (2)	0.347 (1)	0.415 (3)	0.479 (4)	0.524 (5)
Group12	0.360 (2)	0.272 (1)	0.418 (4)	0.427 (5)	0.415 (3)
Group13	0.407 (4)	0.386 (3)	0.356 (1)	0.376 (2)	0.424 (5)
Group14	0.422 (5)	0.325 (1)	0.389 (2)	0.403 (3)	0.413 (4)
Group15	0.362 (3)	0.301 (1)	0.351 (2)	0.376 (4)	0.422 (5)
Group16	0.409 (3)	0.405 (2)	0.403 (1)	0.421 (4)	0.456 (5)
Group17	0.416 (4)	0.300 (1)	0.354 (2)	0.394 (3)	0.419 (5)
Group18	0.451 (4)	0.327 (1)	0.409 (2)	0.441 (3)	0.467 (5)
Group19	0.437 (4)	0.356 (2)	0.320 (1)	0.435 (3)	0.447 (5)
Group20	0.490 (5)	0.363 (1)	0.443 (2)	0.465 (3)	0.477 (4)
Average rank	**3.20**	**1.30**	**2.25**	**3.50**	**4.75**

From the average of ranks, the value $\chi_F^2 = 54.44$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p-value = 4.2e-11$). In the next phase, the Nemenyi test was performed to compare the methods based on the computed average ranks, which results are presented in Table 4.26.

Table 4.26
Nemenyi test (MAPE metrics - TEN-HMS).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	-1.90***	-0.95	0.30	1.55**
GRNN	0	0.95	2.20***	3.45***
SVR	0	0	1.25*	2.50***
AVP	0	0	0	1.25*

** at a significance level of 10%, ** at a significance level of 5%, *** at a significance level of 1%*

From the Table 4.26, it can be concluded that the proposed WMM method outperforms GRNN and SVR at a level of 1%, the ARIMA at a level of 5% and the AVP at a level of 10%.

v. Comparison of the prediction methods using all metrics

The Table 4.27*a*) and Table 4.27*b*) summarize the average ranks and the respective values of qui-square and *p-value*.

Table 4.27
Comparison of the prediction methods–all metrics (TEN-HMS).

	ARIMA	GRNN	SVR	AVP	WMM
SWK	2.70	1.90	3.20	2.95	4.25
CORC	2.50	2.10	3.20	3.00	4.20
NRMSE	3.40	1.00	2.75	3.10	4.75
MAPE	3.20	1.30	2.25	3.50	4.75

a) Average ranks.

	χ_F^2	$p-value$
SWK	23.24	0.0001
CORC	20.32	0.0004
NRMSE	58.36	6.4e-12
MAPE	54.44	4.2e-11

b) Qui-square and p-value.

From the analysis of Table 4.37*b*) the null hypothesis has to be rejected for all the metrics. Moreover, from the previous comparison using individual metrics, it can be concluded that the proposed method is globally superior to the others. In effect, considering the values of Table 4.20, Table 4.22, Table 4.24, and Table 4.26, it is clear

that the wavelet based prediction method (WMM) presents, in most cases, higher performances.

iv. A step-by-step example

The Figure 4.34 illustrates the prediction methodology applied to the signal $X(t)$. The variable $X_1(t)$ identifies the section of the signal $X(t)$ before the time instant $t_0 = 32$ (template), and $X_2(t)$ the respective future (assumed to be unknown). The variable $Y(t)$ represents the result of the prediction method.

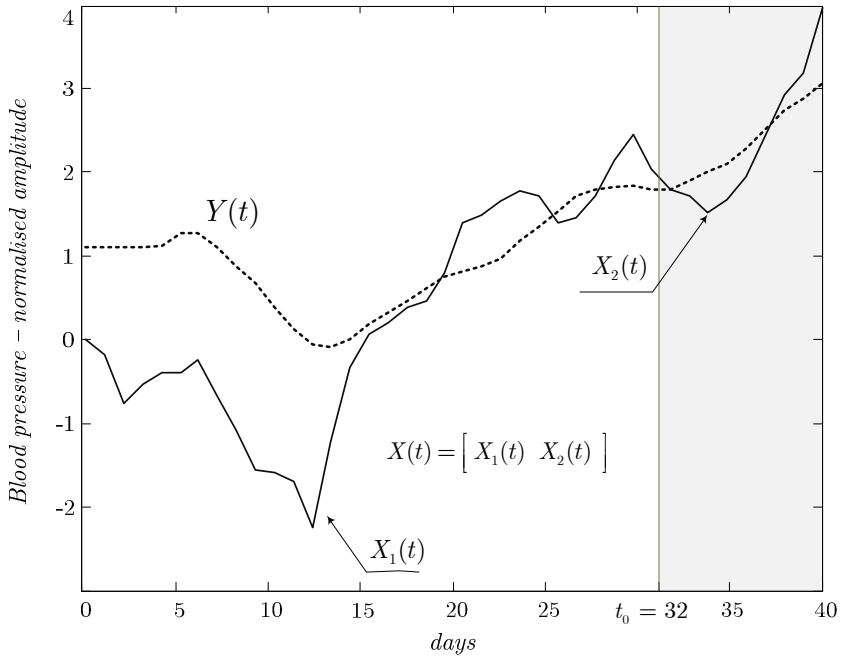


Figure 4.34 - Example of the prediction method result.

The Figure 4.35 depicts, for the approximation and details (level $l = 1,2,3,4,5$), the decomposition of the template $X_1(t)$, the decomposition of the patterns, $aZ_m(t)$, $m = 1,..,5$ and $d^l Z_m(t)$, $m = 1,..,5$, and the respective representative trends, $a\overline{Z}(t)$ and $d^l \overline{Z}(t)$.

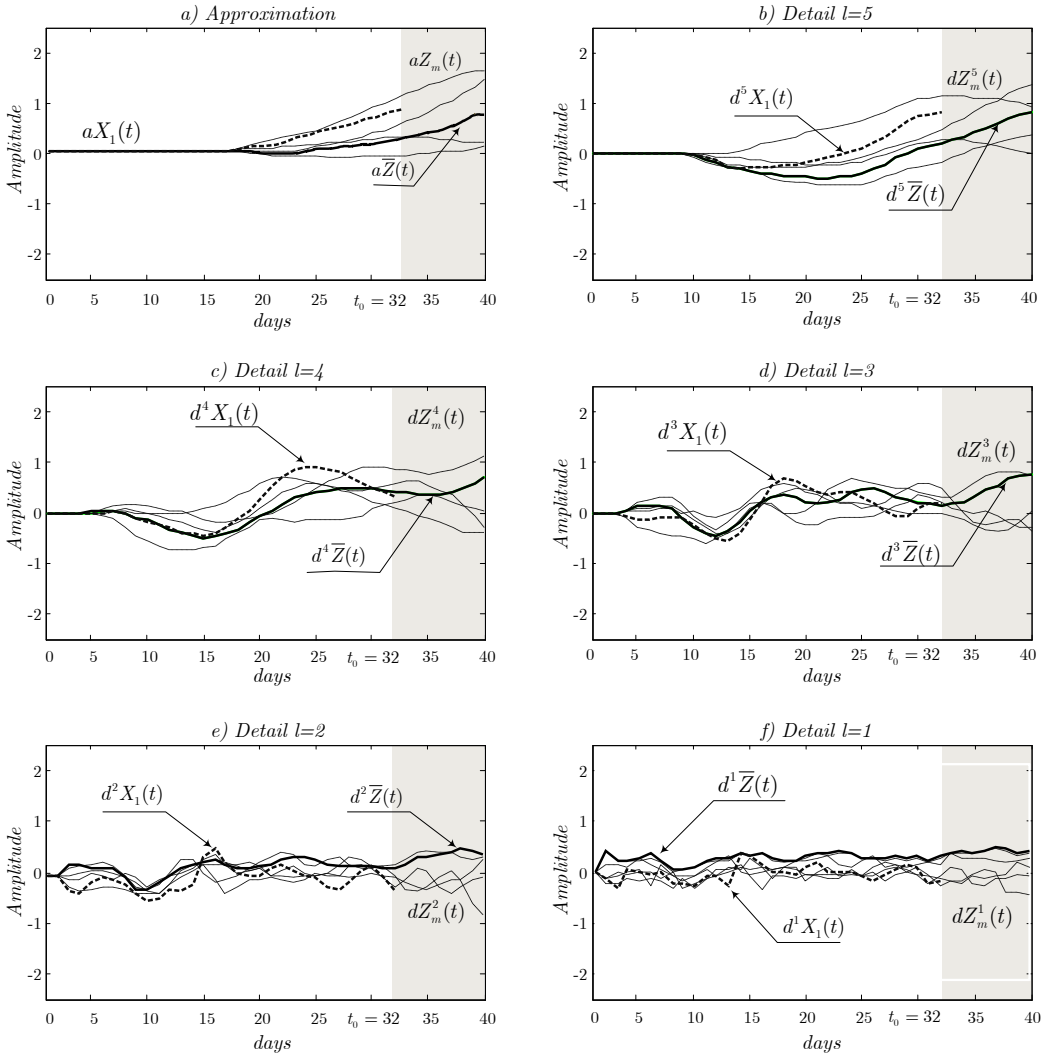


Figure 4.35 - Decomposition of the template and of the patterns.

From the optimization process using a set of distance-based measures (subsection 3.4.3) the levels l=3 and l=5 were selected for the computation of $Y(t)$. In the combination process only the approximation and the last three detail levels (l=3,4,5) were used, since they present a smooth evolution, and thus are more adequate to represent trends.

6. Experiments with the MyHeart dataset

Comparison of the prediction methods (ARIMA, GRNN, SVR, AVP, WMM)
Similarly to the tests conducted using the TEN-HMS dataset, a set of experiments was carried out based on the MyHeart dataset for comparing the different predictive strategies. A total of 500 experiments were performed ($M = 5$, $N = 32$, $P = 8$) and the respective comparison results are depicted in Figure 4.36.

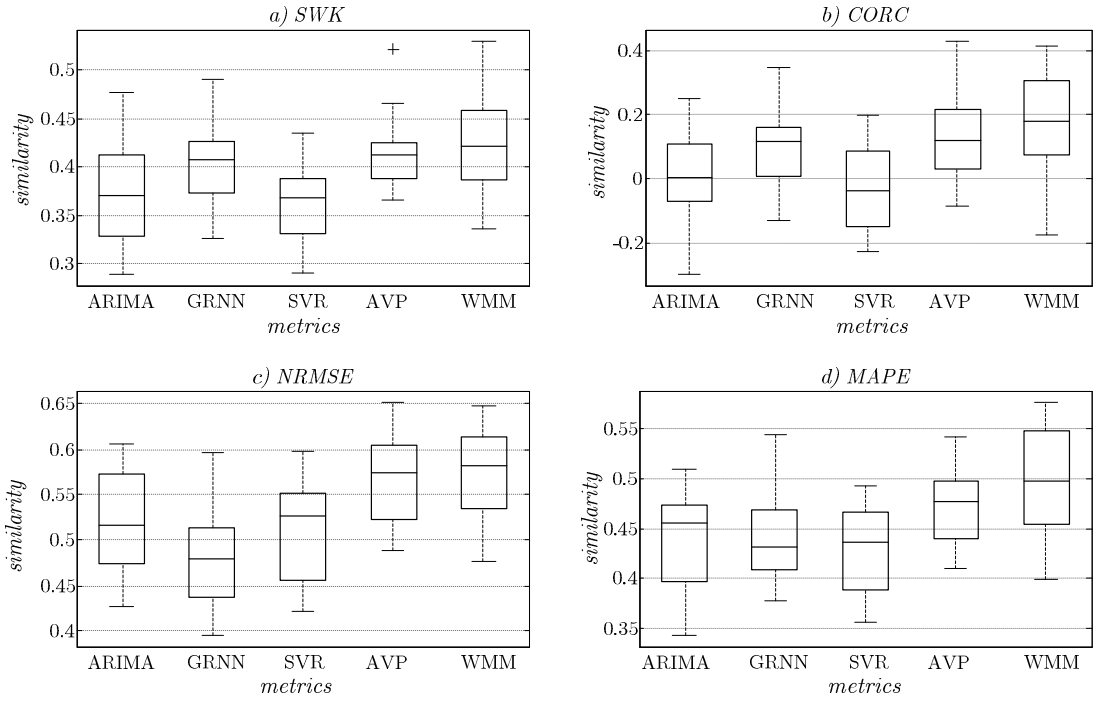


Figure 4.36 - Comparison of the predictive methods (SWK, CORC, NRMSE, and MAPE metrics) using MyHeart dataset.

From the analysis of Figure 4.36 and, in global terms, it appears that the proposed method is slightly superior to the others. In effect, the wavelet based prediction method (WMM) presents the highest median for all the metrics presenting, however, a higher variability for some of these metrics.

i. Comparison of the methods using the SWK metric

The Table 4.28, corresponding to the Figure 4.36*a*), compares the five methods according to the SWK metrics, where the values in each row result from the average of groups of 25 experiments (total of 500 experiments).

Table 4.28

Comparison of SWK among the five prediction methods (MyHeart).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.442 (4)	0.419 (2)	0.391 (1)	0.420 (3)	0.443 (5)
Group2	0.291 (1)	0.374 (2)	0.433 (4)	0.521 (5)	0.422 (3)
Group3	0.381 (4)	0.406 (5)	0.354 (2)	0.369 (3)	0.344 (1)
Group4	0.327 (1)	0.414 (4)	0.362 (2)	0.414 (3)	0.416 (5)
Group5	0.383 (2)	0.410 (4)	0.359 (1)	0.417 (5)	0.386 (3)
Group6	0.360 (2)	0.414 (3)	0.296 (1)	0.422 (4)	0.481 (5)
Group7	0.403 (2)	0.431 (3)	0.397 (1)	0.453 (4)	0.530 (5)
Group8	0.361 (2)	0.330 (1)	0.383 (3)	0.424 (5)	0.416 (4)
Group9	0.476 (4)	0.490 (5)	0.292 (1)	0.389 (2)	0.424 (3)
Group10	0.417 (5)	0.346 (3)	0.327 (1)	0.364 (4)	0.335 (2)
Group11	0.420 (3)	0.456 (4)	0.358 (1)	0.389 (2)	0.473 (5)
Group12	0.327 (2)	0.326 (1)	0.394 (5)	0.390 (4)	0.354 (3)
Group13	0.405 (4)	0.400 (3)	0.381 (1)	0.423 (5)	0.384 (2)
Group14	0.453 (4)	0.451 (3)	0.373 (1)	0.374 (2)	0.453 (5)
Group15	0.322 (1)	0.370 (4)	0.336 (2)	0.386 (5)	0.359 (3)
Group16	0.391 (2)	0.489 (5)	0.290 (1)	0.444 (3)	0.463 (4)
Group17	0.330 (1)	0.402 (3)	0.375 (2)	0.402 (4)	0.421 (5)
Group18	0.289 (1)	0.362 (2)	0.375 (3)	0.465 (5)	0.435 (4)
Group19	0.339 (1)	0.412 (4)	0.434 (5)	0.381 (2)	0.389 (3)
Group20	0.335 (2)	0.394 (3)	0.311 (1)	0.410 (4)	0.479 (5)
Average rank	**2.40**	**3.20**	**1.95**	**3.70**	**3.75**

From the average of ranks, the value $\chi_F^2 = 20.44$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p - value = 0.0004$).

The Nemenyi test was performed to compare the methods based on the computed average ranks, which results are presented in Table 4.29. The performance of the methods is significantly different at the levels of 1%, 5% and 10%, corresponding to critical values of $CD_1 = 1.6275$, $CD_5 = 1.3640$ and $CD_{10} = 1.2300$, respectively.

Table 4.29

Nemenyi test (SWK metrics - MyHeart).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	0.80	-0.45	1.30 *	1.35 *
GRNN		-1.25 *	0.50	0.55
SVR			1.75 ***	1.80 ***
AVP				0.05

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

From the table, it can be concluded that the proposed WMM method outperforms ARIMA and SVR, respectively at the levels of 10% and 1%. In turn, at a level of 10%, the GRNN outperforms the SVR method. The AVP method behaves similarly to the WMM.

ii. Comparison of the methods using the CORC metric

The Table 4.30, corresponding to the Figure 4.36b), compares the five methods using the CORC metrics with the same groups of experiments.

Table 4.30

Comparison of CORC among the five prediction methods (MyHeart).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.066 (2)	0.109 (3)	0.025 (1)	0.117 (4)	0.197 (5)
Group2	-0.297 (1)	-0.130 (2)	0.158 (3)	0.331 (5)	0.309 (4)
Group3	0.018 (2)	0.143 (5)	0.040 (4)	0.022 (3)	-0.032 (1)
Group4	-0.192 (1)	0.151 (3)	0.092 (2)	0.205 (4)	0.234 (5)
Group5	0.015 (2)	0.127 (4)	-0.209 (1)	0.234 (5)	0.081 (3)
Group6	-0.041 (2)	0.149 (3)	-0.203 (1)	0.169 (4)	0.364 (5)
Group7	0.094 (2)	0.172 (3)	-0.120 (1)	0.359 (4)	0.415 (5)
Group8	-0.068 (2)	-0.091 (1)	0.138 (5)	0.127 (3)	0.131 (4)
Group9	0.252 (4)	0.347 (5)	-0.199 (1)	0.018 (2)	0.200 (3)
Group10	0.087 (5)	-0.027 (3)	-0.099 (1)	-0.084 (2)	0.077 (4)
Group11	0.117 (3)	0.220 (5)	-0.014 (1)	0.111 (2)	0.163 (4)
Group12	-0.223 (1)	-0.028 (2)	0.033 (3)	0.153 (5)	0.069 (4)
Group13	0.137 (3)	0.140 (4)	-0.175 (1)	0.157 (5)	0.020 (2)
Group14	0.146 (3)	0.211 (4)	0.077 (2)	0.015 (1)	0.305 (5)
Group15	-0.062 (3)	0.043 (5)	-0.103 (2)	0.032 (4)	-0.173 (1)
Group16	0.167 (2)	0.288 (3)	-0.229 (1)	0.319 (4)	0.329 (5)
Group17	-0.065 (1)	0.060 (3)	0.096 (4)	0.024 (2)	0.247 (5)
Group18	-0.199 (1)	-0.033 (3)	-0.094 (2)	0.431 (5)	0.160 (4)
Group19	-0.074 (1)	0.047 (3)	0.195 (5)	0.083 (4)	0.008 (2)
Group20	-0.007 (2)	0.088 (3)	-0.061 (1)	0.101 (4)	0.341 (5)
Average rank	**2.15**	**3.35**	**2.1**	**3.60**	**3.80**

From the average of ranks, the value $\chi_F^2 = 21.24$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p-value = 0.0003$). The Table 4.31 presents the Nemenyi test results.

Table 4.31
Nemenyi test (CORC metrics - MyHeart).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	1.20	-0.05	1.45 **	1.65 ***
GRNN		-1.25 *	0.25	0.45
SVR			1.50 **	1.70 ***
AVP				0.20

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

The results of the Table 4.31 are comparable to the ones of Table 4.29 and, thus, also the conclusions that can be inferred.

iii. Comparison of the methods using the NRMSE metrics

The Table 4.32 compares the five methods according to the *NRMSE* metrics.

Table 4.32
Comparison of NRMSE among the five prediction methods (MyHeart).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.543 (3)	0.551 (4)	0.524 (1)	0.536 (2)	0.594 (5)
Group2	0.427 (1)	0.438 (2)	0.545 (3)	0.599 (5)	0.547 (4)
Group3	0.476 (3)	0.438 (1)	0.460 (2)	0.512 (4)	0.519 (5)
Group4	0.471 (1)	0.529 (2)	0.598 (4)	0.601 (5)	0.583 (3)
Group5	0.562 (4)	0.475 (1)	0.544 (2)	0.618 (5)	0.558 (3)
Group6	0.495 (2)	0.518 (3)	0.453 (1)	0.577 (4)	0.621 (5)
Group7	0.523 (3)	0.486 (1)	0.503 (2)	0.651 (5)	0.635 (4)
Group8	0.567 (3)	0.419 (1)	0.549 (2)	0.576 (4)	0.581 (5)
Group9	0.606 (4)	0.572 (3)	0.422 (1)	0.509 (2)	0.630 (5)
Group10	0.580 (5)	0.449 (1)	0.533 (3)	0.553 (4)	0.532 (2)
Group11	0.526 (3)	0.510 (2)	0.449 (1)	0.565 (4)	0.648 (5)
Group12	0.449 (2)	0.402 (1)	0.576 (4)	0.629 (5)	0.534 (3)
Group13	0.592 (5)	0.505 (1)	0.530 (2)	0.572 (3)	0.582 (4)
Group14	0.594 (4)	0.463 (1)	0.588 (3)	0.489 (2)	0.618 (5)
Group15	0.505 (3)	0.463 (1)	0.554 (4)	0.579 (5)	0.476 (2)
Group16	0.590 (2)	0.596 (3)	0.439 (1)	0.608 (4)	0.608 (5)
Group17	0.451 (3)	0.396 (1)	0.438 (2)	0.489 (4)	0.515 (5)
Group18	0.469 (1)	0.484 (2)	0.495 (3)	0.652 (5)	0.577 (4)
Group19	0.507 (2)	0.503 (1)	0.591 (5)	0.535 (3)	0.537 (4)
Group20	0.480 (2)	0.404 (1)	0.493 (3)	0.498 (4)	0.592 (5)
Average rank	**2.80**	**1.65**	**2.45**	**3.95**	**4.15**

From the average of ranks, the value $\chi_F^2 = 35.12$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p-value = 4.3e-7$). The corresponding Nemenyi test is presented in Table 4.33.

Table 4.33
Nemenyi test (NRMSE metrics - MyHeart).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	-1.15	-0.35	1.15	1.35 *
GRNN		0.80	2.30 ***	2.50 ***
SVR			1.50 **	1.70 ***
AVP				0.20

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

From the Table 4.33, it can be concluded that the proposed WMM method outperforms ARIMA, GRNN and SVR, respectively at the levels of 10%, 1% and 1%. The AVP method outperforms the GRNN and SVR, respectively at the levels of 1% and 5%.

iv. Comparison of the methods using the MAPE metrics

The Table 4.34, compares the five methods according to the MAPE metrics.

Table 4.34
Comparison of MAPE among the five prediction methods (MyHeart).

	ARIMA	GRNN	SVR	AVP	WMM
Group1	0.470 (1)	0.512 (4)	0.483 (2)	0.494 (3)	0.554 (5)
Group2	0.400 (1)	0.424 (2)	0.454 (3)	0.494 (5)	0.472 (4)
Group3	0.451 (3)	0.440 (2)	0.425 (1)	0.503 (4)	0.540 (5)
Group4	0.393 (1)	0.442 (2)	0.480 (3)	0.486 (4)	0.490 (5)
Group5	0.424 (3)	0.409 (2)	0.377 (1)	0.447 (5)	0.433 (4)
Group6	0.353 (1)	0.421 (3)	0.364 (2)	0.440 (4)	0.468 (5)
Group7	0.460 (2)	0.479 (3)	0.448 (1)	0.542 (4)	0.573 (5)
Group8	0.504 (4)	0.472 (2)	0.464 (1)	0.478 (3)	0.505 (5)
Group9	0.505 (2)	0.545 (4)	0.408 (1)	0.508 (3)	0.578 (5)
Group10	0.510 (4)	0.464 (2)	0.458 (1)	0.510 (5)	0.483 (3)
Group11	0.422 (3)	0.408 (2)	0.355 (1)	0.463 (4)	0.518 (5)
Group12	0.358 (1)	0.390 (2)	0.492 (5)	0.485 (4)	0.458 (3)
Group13	0.476 (5)	0.467 (4)	0.406 (1)	0.459 (3)	0.449 (2)
Group14	0.480 (3)	0.415 (1)	0.483 (4)	0.427 (2)	0.561 (5)
Group15	0.416 (4)	0.408 (3)	0.399 (1)	0.441 (5)	0.400 (2)
Group16	0.470 (2)	0.536 (5)	0.377 (1)	0.477 (3)	0.508 (4)
Group17	0.464 (2)	0.465 (3)	0.449 (1)	0.521 (4)	0.553 (5)
Group18	0.343 (1)	0.377 (3)	0.376 (2)	0.439 (5)	0.421 (4)
Group19	0.387 (1)	0.401 (2)	0.409 (3)	0.411 (4)	0.423 (5)
Group20	0.468 (3)	0.424 (1)	0.469 (4)	0.440 (2)	0.545 (5)
Average rank	**2.35**	**2.60**	**1.95**	**3.80**	**4.30**

From the average of ranks, the value $\chi_F^2 = 32.12$ was obtained, equation (4.45). As result, the null hypothesis "*Ho: all the methods behave similarly*" was rejected with a high significance level ($p-value = 1.8e-6$). The Nemenyi test was performed and the obtained results are presented in Table 4.35.

Table 4.35
Nemenyi test (MAPE metrics - MyHeart).

#methods (k)	GRNN	SVR	AVP	WMM
ARIMA	0.25	-0.40	1.45 **	1.95 ***
GRNN	0	-0.65	1.20	1.70 ***
SVR	0	0	1.85 ***	2.35 ***
AVP	0	0	0	0.50

* *at a significance level of 10%,* ** *at a significance level of 5%,* *** *at a significance level of 1%*

From the Table 4.35, it can be concluded that the proposed WMM method outperforms ARIMA, GRNN and SVR, at a level of 1%. In turn, the AVP method outperforms SVR and ARIMA methods, respectively, at a level 1% and 5%.

v. Comparison of the prediction methods using all metrics

The Table 4.36*a*) and Table 4.36*b*) summarize the average ranks and the respective values of qui-square and *p-value*.

Table 4.36
Comparison of the prediction methods–all metrics (MyHeart).

	ARIMA	GRNN	SVR	AVP	WMM
SWK	2.40	3.20	1.95	3.70	3.75
CORC	2.15	3.35	2.10	3.60	3.80
NRMSE	2.80	1.65	2.45	3.95	4.15
MAPE	2.35	2.60	1.95	3.80	4.30

a) Average ranks.

	χ_F^2	$p-value$
SWK	20.44	0.0004
CORC	21.24	0.0003
NRMSE	35.12	4.3e-7
MAPE	32.12	1.8e-6

b) Qui-square and p-value.

From the analysis of Table 4.36*b*) the null hypothesis has to be rejected for all the metrics. Moreover, from the previous comparison using individual metrics, it can be concluded that the proposed method is globally superior to the ARIMA, GRNN and SVR, but comparable with the AVP method.

4.5.3 Evaluation of hypertension risk

Experiments

The present group of experiments is particularly applied to patients whose blood pressure values are in a critical range (around the threshold of hypertension). The main goal is to employ the trend prediction results to assess a patient hypertension risk. Specifically, the aim is to determine whether during the following week the blood pressure signal of that patient evolves towards hypertension values or, on the contrary, is maintaining or decreasing to normal values. The Figure 4.37 illustrates this idea.

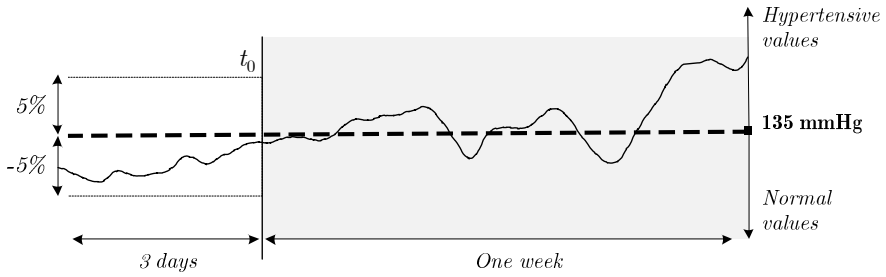


Figure 4.37 - Assessment of the hypertension risk.

The procedure starts by identifying the patients that have recently shown blood pressure values in a critical range, more specifically, that have presented blood pressure values in the range $[-5\%,+5\%]$ of the limit value of 135 mmHg during 3 consecutive days.

Then, for those patients, the blood pressure values of the following week are predicted. According to the percentage of values that are above the limit threshold (135 mmHg), the risk of the patient is assessed: if the percentage is higher than 75%, the patient is considered to be at risk of developing an hypertension condition; in the other case (less than 75%), the patient is considered to have no hypertension risk.

Results with the TEN-HMS dataset

The effectiveness of the proposed strategy was tested by selecting, from a set of 600 random templates, the ones that verified the referred requirement (to be in the critical range). In effect, 68 verified this condition: in 19 cases the patient presented risk of developing a hypertension condition, and in 49 cases the patient revealed no risk.

The Table 4.37 shows the discrimination capability of the method.

Table 4.37
Confusion matrix (TEN-HMS dataset).

		Actual class	
		No risk	*In risk*
Predicted class	*No risk*	37	3
	In risk	12	16

To quantify the validity of the method, the sensitivity (SE) and specificity (SP) were determined, resulting in a SE of 84.2% and a SP of 75.5%.

Although it is not possible to compare these results with other works, considering that the prediction involved fully random templates, the obtained SE and SP values are very satisfactory. In effect, these metrics demonstrate the potential of the trend prediction strategy.

Results with the MyHeart dataset

Regarding the MyHeart dataset, the same 500 experiments performed in the last section (4.5.2), were used for validating the hypertension risk evaluation strategy. From these, 95 exhibited values in the critical range: in 21 cases the patient presented risk of developing an hypertension condition, and in 74 cases the patient revealed no risk.

The Table 4.38 shows the discrimination capability of the method.

Table 4.38
Confusion matrix (MyHeart dataset).

		Actual class	
		No risk	*In risk*
Predicted class	*No risk*	68	3
	In risk	6	18

The sensitivity and specificity were, respectively, of 85.71% and 91.89%, demonstrating the potential of the trend prediction strategy.

4.5.4 Conclusions

This section focused the validation of the wavelet multi-resolution based prediction scheme using real data collected by two telemonitoring platforms (TEN-HMS and MyHeart).

Firstly, the strategy was compared with other prediction methods, namely a linear regression model, the autoregressive integral moving average model - ARIMA, two non-linear regression models, the generalized regression neural network - GRNN and the support vector regression – SVR, and the average value of the patterns - AVP. From the non-parametric Friedman and Nemenyi tests, it was possible to validate the main goal of the developed methodology: the capability to predict trends.

In a second step, this capability was tested in the assessment of the hypertension risk in patients whose blood pressure values were in a critical range (around the threshold of hypertension). Particularly, the goal was to evaluate whether during the following week blood pressure signals would evolve towards hypertension values.

4.6 Conclusions

In this chapter the methodologies and algorithms proposed through this thesis were tested and validated. Partially integrated in the HeartCyle project, the main focus of the present work was the development of strategies able to detect and predict specific cardiovascular conditions, fundamentally based on the electrocardiogram and blood pressure signals. The central effort was on time series similarity search methodologies, in order to enable the detection of such conditions, and on predictive techniques to be applied in the trend analysis of biosignals and early detection of future events.

Regarding similarity measures, a comparison of several approaches (including the proposed one) was performed in section 4.2. To evaluate the sensitivity of the referred measures when applied to distinct situations, specific variations in time and in amplitude were induced in a template. The results highlighted the importance of knowing the biosignal characteristics, as well as if a greater or lesser sensitivity to variations is desired, when selecting a similarity measure. In particular, the experiments revealed that the proposed similarity measure that combines the Haar wavelet with the Karhunen-Loève transforms, is especially adequate to deal with noisy signals, with signals that are not aligned in time, as well as with trends.

In the section 4.3 the strategy for ischemic episodes detection was presented and validated. The central aspects are the approach for accurate ST shift and isoelectric point estimation based on the time-frequency analysis, and the ECG beat morphology effective characterization using the expansion in Hermite functions. The effectiveness of the methodology was confirmed by using the European Society of Cardiology ST-T database. The achieved results (sensitivity of 96.7% and positive predictivity of 96.2%) were relevant when compared with similar works reported in literature.

The sections 4.4 and 4.5 investigated the application of the strategies proposed for physiological time series prediction. In the section 4.4 the neural network multi-model scheme was employed to the prediction of adverse acute hypotensive episodes (AHE) occurring in intensive care units (ICU). The effectiveness of the methodology was validated in the context of the 10th Physionet/Computers in Cardiology Challenge - Predicting Acute Hypotensive Episodes, applied to a specific set of blood pressure signals available in MIMIC-II database. A correct prediction of 10 out of 10 AHE for event 1 and of 37 out of 40 AHE for event 2 was achieved, corresponding to the best results of all entries in the two events of the challenge.

Finally, in section 4.5, the forecast capacity of the wavelet multi-resolution scheme was validated. In a first phase it was compared with other four methodologies (ARIMA, GRNN, SVR, AVP), having revealed a superior performance in the trend prediction of blood pressure. This was particularly demonstrated by the statistical analysis of

correlation coefficient metrics using Friedman and Nemenyi tests. In a second phase, the strategy was successfully applied in the evaluation of hypertension risk, that is, in estimating if BP signals evolve towards a critical range of values.

5. Conclusions and Perspectives

This thesis aimed at the development of specific clinical applications addressing cardiovascular conditions identified as relevant for the CVD management by the HeartCycle project clinical partners, mainly based on the analysis and processing of the electrocardiogram and blood pressure signals.

Specifically, the main goals of this work were: *i*) The development of techniques for similarity searching in a set of biosignal time series to detect the occurrence of particular patterns which would serve two purposes: helping diagnosis if patterns were characteristic of specific cardiovascular conditions, or being the starting point of predictive methodologies. These techniques should be efficient in view of the possible long duration of the time series, and able to deal with trends, one the main characteristics of biosignals to be addressed in this thesis. *ii*) The implementation of predictive strategies to be applied in the forecast of biosignal time series.

In this chapter, an overview of the main findings of this research is presented and discussed, and some suggestions for future work are also introduced.

5.1 Main research findings

Contributing to the achievement of the first objective, two methodologies for accessing similarities in time series were introduced, both based on the representation of signals as linear combinations of a set of orthogonal basis functions. While the first was specifically projected for helping in ischemia diagnosis, the second was mainly developed to support the predictive methods later referred.

In effect, a strategy for ischemic episodes detection was presented, which exhibited, as the most innovative aspects, the new approach for accurate ST shift estimation based on the time-frequency analysis through the Wigner-Ville transform, and the ECG beat morphology effective characterization using the expansion in Hermite functions. Basically, in order to characterize the changes in T wave and in QRS complex morphologies due to ischemia, each cardiac beat was represented by a linear combination of orthogonal Hermite basis functions, which coefficients were used as features in the classification process. The last was carried out by two independent feed-forward neural networks for each ECG lead, which classify heart beats as having

ST elevation, ST depression, or as normal. Finally, a sliding window applied to the entire ECG signal allowed ischemia episodes identification. A comparison of this methodology with similar works reported in literature confirmed its potential by the relevant results achieved, namely a sensitivity of 96.7%, and a positive predictivity of 96.2%.

A more generic methodology consisting in a time series similarity measure and indexing scheme based on an optimal basis description obtained from the combination of the Haar wavelet decomposition with the Karhunen-Loève transform, was also proposed. The main characteristics of the referred similarity measure are its capacity to deal with trends, as well as its interpretability and simplicity, since the assessment of similarity between two time series is done by simply analysing the coefficients resulting from their description in terms of the reduced set of Haar basis. In fact, two time series may be considered as similar if their coefficients present the same sign. On the other hand, a very efficient procedure for similarity indexing based on an iterative formulation was also introduced. The potential of this scheme to be an effective and appropriate tool for identifying physiological patterns in biosignal time series, was tested by quantitatively comparing the proposed measure with other three measures: the Euclidean distance, the linear correlation coefficient and a Fourier based measure. The comparison study consisted in the assessment of similarity between a baseline signal (template) and some variations (in amplitude, time and shape) induced in that signal. Although from the achieved results it was not reasonable to conclude that one similarity measure was better than the others, it was found that the proposed one was particularly appropriate to assure low sensitivity values in the presence of small levels of baseline variations. Consequently, that it was adequate to deal with the evaluation of similarity between historic signals and a particular template, the main purpose for which it was designed.

Regarding the second objective, two predictive methodologies were introduced.

The first used generalized regression neural networks integrated into a multi-model structure for the accurate prediction of a time-series over a forecast horizon. It was employed and validated in the context of the 2009 PhysioNet/Computers in Cardiology challenge, where the goal was to predict acute hypotensive episodes occurring in intensive care units. The methodology consisted of two steps: in the first, a similarity analysis procedure was carried out between the current signal (template) and a historical dataset. From this, the most similar signals (patterns) were identified and the corresponding prediction neural models, previously trained using those historical signals, selected. In the second step, the multi-model structure was employed to the current signal to forecast its future evolution. The performed predictions corresponded to the best results of all entries in the two events of the challenge.

The second scheme exploited the multiresolution analysis provided by the wavelet transform for a rough estimative of a biosignal time series future values, based on evolution trends, without involving explicit modelling. The global multiresolution scheme comprised three main distinct phases. Firstly, by means of a similarity analysis

procedure applied to a historical dataset, patients who displayed in their physiological time series a behaviour similar to the patient's current data (template), were identified. Then, a multiresolution decomposition of the time series retrieved from such patients (historical patterns) took place. Finally, the future trend of the patient's current data (template) was estimated, by combining the optimal decomposition levels of the historical patterns. The effectiveness of the developed method in predicting trends was evaluated, by comparing it with other three predictive methodologies, namely ARIMA, GRNN, and SVR. From the non-parametric Friedman and Nemenyi tests, in particular from the correlation coefficient analysis, the efficacy of the proposed method for the intended purpose was demonstrated. In a following stage, this was validated in the assessment of the hypertension risk in patients whose blood pressure values were in a critical range (around the threshold of hypertension). Specifically, the goal was to evaluate whether during the following week blood pressure signals would evolve towards hypertension values. The obtaining results using TEN-HMS and MyHeart datasets confirmed the potential of the proposed strategy in this type of application.

5.2 Future work

Although generic, the methodologies researched in this work were limited to specific biosignals and particular cardiovascular conditions. In effect, ventricular arrhythmias, desynchronization, ischemia, hypotension and hypertension were addressed, based on the analysis and processing of individual biosignals, namely the electrocardiogram and the blood pressure.

Given the demonstrated potential of these methodologies, a natural direction of future work will be their extension to other conditions, eventually involving the multi-parametric analysis of several measurements. In effect, these aspects will be investigated in a near future under the HeartWays project - *Advanced Solution for Supporting Cardiac Patients in Rehabilitation* (FP7-315659), planned to start in October, 2012.

On the other hand, based on the individual measurements/parameters acquired during the daily monitoring process, and taking into account the collected historic dataset, personalized dynamic cardiovascular risk models can also be a topic of future research.

Additionally, for each individual source of information (time series signals) relevant characteristic parameters and/or trend estimations can be achieved by the developed approaches. As result, these can be directly used to provide valuable indications and generate useful alarms regarding relevant cardiovascular conditions, as well as to predict critical future evolutions. For example, simple alarms can be triggered in case of abnormal blood pressure values and, by means of a predictive scheme that estimates future values of blood pressure, the occurrence of future hypertension events can be anticipated.

6. References

Agrawal et al. (1993); R. Agrawal, C. Faloutsos, A. Swami; *Efficient similarity search in sequence databases*; Proceedings of the 4th Conference on Foundations of Data Organization and Algorithms, 30, pp 69-84.

Agrawal and Srikant (1995); R. Agrawal, R. Srikant; *Mining sequential patterns*; Proceedings of the 7th International Conference on Data Engineering, pp 3-14.

Akaike (1974); H. Akaike; *A new look at the statistical model identification*; IEEE Transactions on Automatic Control 19 (6): 716–723.

Akay (1997); M. Akay; *Wavelets for biomedical signal processing*; Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE, 6, 2688- 2691.

Akselrod et al. (1987); S. Akselrod, M. Norymberg, L. Peled, E. Karabelnik, M. Green; *Computerised analysis of st segment changes in ambulatory electrocardiograms*; Medical and Biological Engineering and Computing, vol. 25, pp. 513-519.

Alonso et al. (2003); F. Alonso, J. Valente, M. Martinez, L. Montes; *Discovering similar patterns for characterizing time series in a medical domain*; Knowledge and Information Systems 5: 183–200.

Anderson (1972); J. Anderson; *A simple neural network generating and interactive memory*; Mathematical BioSciences, 14, 197-220.

Andreao et al. (2004); R. Andreao, B. Dorizzi, J. Boudy, J. Mota; *St-segment analysis using hidden markov model beat segmentation: application to ischemia detection*; Computers in Cardiology, pp. 381-384.

Antoniadis et al. (2003); A. Antoniadis, X. Brossat, J. Cugliari, J. Poggi; *Clustering functional data using wavelets*; Methodology Applications, Report number:RR-7515.

Atiya et al. (1999); A. Atiya, S. El-Shura, S. Shaheen, S. El-Sherif; *A comparison between neural-network forecasting techniques—case study: river flow forecasting*; IEEE Transactions on Neural Networks, 10, pp. 402–409.

Aussem et al. (1998); A. Aussem, J. Campbell, F. Murtagh; *Wavelet-based feature extraction and decomposition strategies for financial forecasting*; , International Journal of Computational Intelligence in Finance, 6, pp. 5-12.

Badilini et al. (1992); F. Badilini, M. Merri, J. Benhorin, A. Moss; *Beat-to-beat quantification and analysis of st displacement from holter ecgs: a new approach to ischemia detection*; Computers in Cardiology, pp. 179-182.

Bambang (2007); R. Bambang; *Nonlinear active noise control using ekf-based recurrent fuzzy neural networks*; International Journal of Hybrid Intelligent Systems archive, Vol. 4, 4.

Barill (2003); T. Barill; *The six second ecg: a practical guidebook to basic ecg interpretation*; Chapter 4, Nursecom Educational Technologies.

Barron (1993); A. Barron; *Universal approximation bounds for super positions of a sigmoidal function*; IEEE Transactions on Information Theory, 39, 3, 930-945.

Bassale (2001); J. Bassale; *Hypotension prediction and arterial blood pressure variability*; Technical report.

Bauer (1995); M. Bauer; *General regression neural network for technical use*; Master's Thesis, University of Wisconsin-Madison.

Becerikli et al. (2003); Y. Becerikli, A. Konar, T. Samad; *Intelligent optimal control with dynamic neural networks*; Elsevier Science Ltd, Neural Networks archive, Volume 16 , 2.

Benaouda et al. (2006); D. Benaouda, O. Renaud, J. Starck, F. Murtagh; *Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting*; Neurocomputing, Volume 70, 139 - 154.

Benhorim et al. (1996); J. Benhorim, F. Badilini, A. Moss, W. Hall, M. Merri, W. Zareba; *New approach to detection of ischemic-type st segment depression*; Noninvasive Electrocardiology, chapter 19, pp. 345-355, W.B. Saunders Co, London.

Bernhard and Darbellay (1999); H. Bernhard, G. Darbellay; *Performance analysis of the mutual information function for nonlinear and linear processing*; IEEE International Conference on Acoustics, Speech and Signal Processing (USA), 3, 1297-1300.

Bezdek (1981); J. Bezdek; *Pattern recognition with fuzzy objective function algoritms*; Plenum Press, New York.

Bishop (1995); C. Bishop; *Neural networks for pattern recognition*; Clarendon Press.

Bjorn (1995); V. Bjorn; *Multiresolution methods for financial time series prediction*; Proc. IEEE/IAFE 1995 Conf. Computational Intelligence for Financial Eng.

Bodyanskiy and Vynokurova (2012); Y. Bodyanskiy, O. Vynokurova; *Hybrid adaptive wavelet-neuro-fuzzysystem for chaotic time series identification*; Information Sciences, Article in Press (10 pp).

Bollerslev (1986); T. Bollerslev; *Generalized autoregressive conditional heteroskedasticity*; Journal of Econometrics, 31:307-327.

Bosworth and Oddone (2004); H. Bosworth, E. Oddone; *Telemedicine and hypertension*; JCOM, Vol. 11, No. 8 517-522.

Bosworth and Oddone (2004); H. Bosworth, Z. Oddone; *Telemedicine and hypertension*; JCOM, Vol. 11, No. 8 August.

Box and Jenkins (1976); G. Box, G. Jenkins; *Time series analysis: forecasting and control*; San Francisco, CA: Holden-Day.

Bravi et al. (2011); A. Bravi, A. Longtin, A. Seely; *Review and classification of variability analysis techniques with clinical applications*; BioMedical Engineering OnLine 10:9.

Camilleri (2004); M. Camilleri; *Forecasting using non-linear techniques time series analysis: an overview of techniques and main issues*; Computer Science Annual Research workshop, 19-28.

Cao et al. (1995); L. Cao, Y. Hong, G. Fang; *Predicting chaotic time series with wavelet networks*; PhysicaD, 85, pp. 225–238.

Carpenter and Grossberg (1987); G. Carpenter, S. Grossberg; *A massively parallel architecture for a self-organizing neural pattern recognition*; Computer Vision, Graphics and Image Processing, 37, 54-115.

Carreño and Vuskovic (2005); R. Carreño, M. Vuskovic; *Wavelet transform moments for feature extraction from temporal signals*; ICINCO'05, pp 71-78.

Casdagli (1991); M. Casdagli; *State space reconstruction in presence of noise*; Physica D, 51, 52-98.

Castells et al. (2007); F. Castells, P. Laguna, L. Sornmo, A. Bollmann, J. Roid; *Principal component analysis in ecg signal processing*; EURASIP Journal on Advances in Signal Processing, issue 1, pp. 98-98.

Chamchad et al. (2004); D. Chamchad, V. Arkoosh, J. Horrow; *Using heart rate variability to stratify risk of obstetric patients undergoing spinal anesthesia*; Anesthesia Analg, 99: 1818–21.

Chan and Fu (1999); K. Chan, A. Fu; *Efficient time series matching by wavelets*; Proceedings of International Conference on Data Engineering (ICDE '99), Sydney, 126-133.

Chan et al. (2003); K. Chan, A. Fu, T. Clement; *Haar wavelets for efficient similarity search of time-series: with and without time warping*; IEEE Trans. on Knowledge and Data Engineering, 15(3):686-705.

Chang and Piau (2007); P, Chang, P. Piau; *Modified Fast and Exact Algorithm for Fast Haar Transform*; Proceedings of the World Academy of Science, Engineering and Technology, V. 26:509-512.

Chatfield (2000); C. Chatfield; *Time-series forecasting*; Chapman & Hall/Crc.

Chen and Billings (1989); S. Chen, S. Billings; *Recursive prediction error parameter estimation for non-linear models*; International Journal of Control, 29, 569-594.

Chen et al. (1989); S. Chen, S. Billings, W. Luo; *Orthogonal least squares methods and their application to non-linear system identification'*; International Journal of Control, 50, pp. 1873–1896.

Cheong (2004); C. Cheong; *Clustering and predicting stock prices using advanced wavelet methods and neural networks*; Master Thesis, Department of Information Technology, Master of Science in Information Technology, Malaysian University of Science and Technology.

Chiu (1994); S. Chiu; *Fuzzy model identification based on cluster estimation*; Journal of Intelligent & Fuzzy Systems 2: 267–278.

Chong (2009); T. Chong; *Financial time series forecasting using improved wavelet neural network*; Master's Thesis.

Chughtai and Peixoto (2003); I. Chughtai, A. Peixoto; *Ambulatory blood pressure monitoring: a review of its clinic and prognostic relevance*; Hospital Physician, March 2003.

Chui (1992); C. Chui; *An introduction to wavelets*; Boston: Academic Press.

Chui (1987); C. Chui; *Wavelets: a mathematical tool for signal analysis*; Society for Industrial Mathematics.

Cleland et al. (2005); J. Cleland, A. Louis, A. Rigby, U. Janssens, A. Balk; *Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death*; J Am Coll Cardiol, 45:1654-1664.

Clifford et al. (2006); G. Clifford, F. Azuaje, P. McSharry; *Advanced methods and tools for ecg data analysis*; Engineering in Medicine & Biology, Artech House Inc.

Cohen and Grossberg (1983); G. Cohen, S. Grossberg; *Absolute stability of global pattern formation and parallel memory storage by competitive neural networks*; IEEE Transactions on Systems, Man and Cybernetics, 13, 5, 815-826.

Cohen (1995); L. Cohen; *Time-frequency analysis*; Englewood Cliffs, NJ, Prentice-Hall.

Coifman and Wickerhauster (1992); R. Coifman, M Wickerhauster; *Entropy based algorithms for best basis selection*; IEEE Trans. Inform. Theor. 1992, 38, 713-718.

Couceiro et al. (2008); R. Couceiro, P. Carvalho, J. Henriques, M. Antunes; *On the detection of premature ventricular contractions ;*; EMBC-08, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, August 20-24.

Craigmile and Percival (2002); P. Craigmile, D. Percival; *Wavelet-based trend detection and estimation*; Entry in the Encyclopedia of Environmetrics, edited by A. El-Shaarawi and W. W. Piegorsch, Chichester, UK: John Wiley & Sons.

Crespo (2002); C. Crespo; *Precursors in the arterial blood pressure signal to episodes of acute hypotension in sepsis*; Proceedings of the 16th International EURASIP Conference BIOSIGNAL, 16; 206-8.

Cybenko (1989); G. Cybenko; *Approximation by superposition of a sigmoidal function*; Mathematics of Control, Signals and Systems,2, 3, 303-314.

Daubechies (1992); I. Daubechies; *Ten lectures on wavelets*; Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Emgle (1982); R. Emgle; *Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation*; Econometrica 50 (1982): 987-1008.

Engle (1982); R. Engle; *Co-integration and error-correction: representation, estimation and testing*; Econometrica, 50, 987-1008.

Ephzibah and Sundarapandian (2012); E. Ephzibah, V. Sundarapandian; *A neuro fuzzy expert system for heart disease diagnosis*; Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.1.

Exarchos et al. (2007); T. Exarchos, M. Tsipouras, C. Exarchos, C. Papaloukas, D. Fotiadis, L. Michalis; *A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree*; Artificial Intelligence in Medicine, vol. 40, issue 3, pp. 187-200.

Faloutsos et al. (1994); C. Faloutsos, M. Ranganathan, Y. Manolopoulos; *Fast subsequence matching in time-series databases*; In Proc. of the ACM SIGMOD, pp 419-429, Minneapolis, USA.

Fan and Yao (2003); J. Fan, Q. Yao; *Nonlinear time series: nonparametric and parametric methods*; Springer-Verlag, New York.

Fedoryuk (2001); M. Fedoryuk; *Hermite functions*; Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4.

Fernandez and Gomez (2008); A. Fernandez, S. Gomez; *Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms*; Journal of Classification 25: 43–65.

Friedman (1937); M. Friedman; *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*; Journal of the American Statistical Association 32 674–701.

Friedman (1940); M. Friedman; *A comparison of alternative tests of significance for the problem of m rankings*; Annals of Mathematical Statistics 11, 86–92.

Frolich and Caton (2002); A. Frolich, D. Caton; *Baseline heart rate may predict hypotension after spinal anaesthesia in prehydrated obstetrical patients*; Can J Anaesth, 49; 185-9.

Fryzlewicz et al. (2003); P. Fryzlewicz, V. Bellegem, R. Sachs; *Forecasting non-stationary time series by wavelet process modelling*; Inst. Statist. Math. Vol. 55, No. 4, 737–764.

Fukanaka (1990); K. Fukanaka; *Introduction to statistical*; Pattern Recognition. Academic Press.

Funahashi (1989); K. Funahashi; *On the approximate realization of continuous mappings by neural networks*; Neural Networks, 2, 3, 183-192.

Gan et al. (2005); X. Gan, Y. Liu, F. Austin; *A prediction method for time series based on wavelet neural networks*; Computational Intelligence and Security, Lecture Notes in Computer Science, 3801: 902-908.

Gandhi and Green (2006); H. Gandhi, D. Green; *Stacked generalization for early diagnosis of alzheimer's disease*; Proceedings of the 28th IEEEEMBS Annual International Conference, New York City, USA, Aug 30-Sept 3.

Gang and Yang (2008); D. Gang, L. Yang; *Time series prediction using wavelet process neural network*; Chinese Physics B, 17:6.

Gang et al. (2008); D. Gang, Z. Shi, L. Yang; *Time series prediction using wavelet process neural network*; Chinese Physics B, Vol 17 No 6, 1674-1056.

Gao (1997); H. Gao; *Choice of thresholds for wavelet shrinkage estimate of the spectrum*; J. Time Series Anal., 18, 231-251.

Garcia et al. (2000); J. Garcia, L. Sornmo, S. Olmos, P. Laguna; *Automatic detection of st-t complex changes on the ecg using filtered rms difference series: application to ambulatory ischemia monitoring*; Transactions on Biomedical Engineering, vol. 47, n° 9, pp. 1195-1201.

Garcia et al. (2010); S. Garcia, A. Fernández, J. Luengo, F. Herrera; *Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power*; Information Sciences: an International Journal archive, Volume 180 Issue 10, May, 2010, pp 2044-2064.

Gardner (2006); E. Gardner; *Exponential smoothing: the state of the art—part ii*; International Journal of Forecasting 22 (4) 637–666.

Gil et al. (2002); P. Gil, J. Henriques, P. Carvalho, A. Dourado, R. Ramos; *Adaptive Neural Model Based Predictive Control of A Solar Power Plant*; IJCNN, International Join Conference on Neural Networks; Honolulu, USA, May, 12-17.

Gilbert et al. (2003); A. Gilbert, Y. Kotidis, S. Muthukrishnan, M. Strauss; *One-pass wavelet decompositions of data streams*; IEEE Transactions on knowledge and data engineering, vol. 15, no. 3, may.

Goldberger et al. (2000); A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley; *Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals*; Circulation, 101, e215-e220.

Gopalakrishnan et al. (2004); R. Gopalakrishnan, S. Acharya, D. Mugler; *Real time monitoring of ischemic changes in electrocardiograms using discrete hermite functions*; 26th Annual International Conference of the Engineering in Medicine and Biology Society, pp. 438 - 441.

Granger and Andersen (1978); W. Granger, A. Andersen; *An introduction to bilinear time series models*; RecenzeVandenhoeck und Ruprecht.

Grossberg (1976); S. Grossberg; *Adaptive pattern classification and universal recoding: parallel development and coding of neural feature detectors*; Biological Cybernetics,23, 121-134.

Gupta and Rao (2003); M. Gupta, D. Rao; *Dynamic neural units with application to the control of unknown linear systems*; Journal of Intelligent and Fuzzy Systems, 1,1, 73-92.

Gupta and Toshniwal (2011); P. Gupta, D. Toshniwal; *Performance comparison of rule based classification algorithms*; International Journal of Computer Science & Informatics, Volume-I, Issue-II, pp 38-42.

Habetha (2006); J. Habetha; MyHeart - A new approach for remote monitoring and management of cardiovascular diseases; Conf Proc IEEE Eng. Med. Biol Soc (2006).

Han and Kamber (2001); J. Han, M. Kamber; *Data mining: concepts and technique2*; Morgan Kaufmann, San Francisco, pp 346–389.

Hanss et al. (2005); R. Hanss, B. Bein, T. Ledowski.; *Heart rate variability predicts severe hypotension after spinal anesthesia for elective cesarean delivery*; Anesthesiology, 102:1086-93.

Haykin (2008); S. Haykin; *Neural networks and learning machines*; Third Edition. Prentice Hall.

Hebb (1949); D. Hebb; *The organization of the behaviour*; John Wiley & Sons.

Henriques et al. (2008); J. Henriques, P. Carvalho, M. Harris, M. Antunes, R. Couceiro, M. Brito, R. Schmidt; *Assessment of arrhythmias for heart failure management*; Phealth08, Valencia, May 21-23.

Hetland (2004); M. Hetland; *A survey of recent methods for efficient retrieval of similar time sequences*; Data Mining in Time Series Databases, World Scientific.

Hiden et al. (1999); H. Hiden, M. Willis, M. Tham, G. Montague; *Linear principal components analysis using genetic programming*; Computers and Chemical Engineering, 23: 413-425.

Hirano and Tsumoto (2002); S. Hirano, S. Tsumoto; *Mining similar temporal patterns in long time-series data and its application to medicine*; IEEE International Conference on Data Mining (ICDM'02), Maebashi City, Japan, ISBN: 0-7695-1754-4.

Hirano and Tsumoto (2005); S. Hirano, S. Tsumoto; *Automated discovery of chronological patterns in long time-series medical datasets*; International Journal of Intelligent Systems, Volume 20, Issue 7, pp 737–757.

Hopfield (1982); J. Hopfield; *Neural networks and physical systems with emergent collective computational abilities*; National Academy of Sciences, 79, 2554-2558, USA.

Hornik et al. (1990); K. Hornik, M. Stinchcombe, H. White; *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*; Neural Networks,3, 551-560.

Hornik (1991); K. Hornik; *Approximation capabilities of multilayer feedforward networks*; Neural Networks, 4, 251-157.

Huhtala et al. (1999); Y. Huhtala, J. Kärkkäinen, H. Toivonen; *Mining for similarities in aligned time series using wavelets*; Data Mining and Knowledge Discovery: Theory, Tools, and Technology, SPIE Proceedings Series, Orlando, FL, vol. 3695, pp. 150-160.

Ilango and Mohan (2010); M. Ilango, V. Mohan; *A survey of grid based clustering algorithms*; International Journal of Engineering Science and Technology Vol. 2(8), 3441-3446.

Ing (2003); C. Ing; *Multistep prediction in autoregressive processes'*; Econometric Theory, 19, pp. 254–279.

Kaboudan (2005); M. Kaboudan; *Wavelets in multi-step-ahead forecasting*; The 16th IFAC World Congress, Prague, July 4.

Kannel (1996); W. Kannel; *Blood pressure as a cardiovascular risk factor: prevention and treatment*; JAMA;275:1571–6.

Karim and Adeli (2002); A. Karim, H. Adeli; *Incident detection algorithm using wavelet energy representation of traffic patterns*; Journal Of Transportation Engineering / May/June, 128:3, 232.

Kavitha and Punithavalli (2010); V. Kavitha, M. Punithavalli; *Clustering time series data stream – a literature survey*; International Journal of Computer Science and Information Security, Vol. 8, No. 1, April 2010.

Kendall (1973); M. Kendall; *Time series*; London: Charles Griffin.

Kennedy and Turley (2011); C. Kennedy, J. Turley; *Time series analysis as input for clinical predictive modeling: modeling cardiac arrest ina pediatric icu*; Theoretical Biology and Medical Modelling, 8:40.

Keogh and Pazzani (2000); E. Keogh, M. Pazzani; *Scaling up dynamic time warping for datamining applications*; KDD Boston, MA, 2000, pp. 285-289.

Keogh et al. (2001); E. Keogh, K. Chakrabarti, S. Mehoritra, M. Pazzani; *Locally adaptive dimensionality reduction for indexing large time series databases*; Proc. SIGMOD, Santa Barbara, California, 151-162.

Keogh and Pazzani (1998); E. Keogh, M. Pazzani; *An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback*; Proceedings of the 4th International Conference of Knowledge Discovery and Data Mining, pp 239- 241.

Keogh et al. (2000); E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra; *Dimensionality reduction of fast similarity search in large time series databases*; Journal of Knowledge and Information System, 3:263-286.

Kohonen (1972); T. Kohonen; *Correlation matrix memories*; IEEE Transactions On Computers, 21, 353-359.

Kopenkov (2008); V. Kopenkov; Efficient algorithms of local discrete wavelet transform with Haar-like bases; Pattern Recognition and Image Analysis V. 18, N. 4, 654-661.

Krajnak and Xue (2006); M. Krajnak, J. Xue; *Optimizing fuzzy clinical decision support rules using genetic algorithms*; Conf Proc IEEE Eng Med Biol Soc 06; 1:5173-6.

Lagerholm (2000); M. Lagerholm; *Clustering ecg complexes using hermite functions and self-organizing maps*; IEEE transactions on biomedical engineering, vol. 47, no. 7.

Landau et al. (1997); L. Landau, R. Lozano, M. Msaad; *Adaptive control: communications and control engineering*; Springer Verlag.

Lang et al. (1995); M. Lang, J. Guo, C. Odegard, C. Burrus, O. Wells; *Nonlinear processing of a shift invariant dwt for noise reduction*; Proc. SPIE Wavelet Applications II, vol. 2491, H. H. Szu, Ed., Orlando, FL, pp 640–651.

Langlois (2002); J. Langlois; *Making a diagnosis*; Chapter 10 in Fundamentals of clinical practice: Mark B. Mengel, Warren Lee Holleman, Scott A. Fields. 2nd edition. p.204.

Lawrynczuk (2008); M. Lawrynczuk; *Computationally efficient nonlinear predictive control based on rbf neural multi-models*; 18th International Conference on Artificial Neural Networks.

LeCun (1985); Y. LeCun; *Une procédure d'apprendissage pour réseau à seuil assymétrique*; A la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neuronsciences, Paris, 599-604.

Lee and Mark (2010); J. Lee, R. Mark; *An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care*; BioMedical Engineering OnLine 9.1, 62.

Lehman et al. (2008); L. Lehman, M. Saeed, G. Moody, R. Mark; *Similarity-based searching in multi-parameter time series databases*; Computers in Cardiology, 35; 653-6.

Leonardi et al. (2007); A. Leonardi, S. Montani, L. Portinale; *Cbr for temporal abstractions configuration in haemodyalisis*; Proc. Case-Based Reasoning in the Health Sciences Workshop, International Conference on Case Based Reasoning (ICCBR), Belfast, August 2007.

Leontaritis and Billings (1985); I. Leontaritis, S. Billings; *Input-output parametric models for non-linear systems, part i: deterministic non-linear systems, part ii: stochastic non-linear systems*; International Journal of Control, 41, 2, 303-344.

Lhermitte et al. (2011); S. Lhermitte, C. Verbesselt, W. Verstraeten, P. Coppin; *A comparison of time series similarity measures for classification and change detection of ecosystem dynamics*; Remote Sensing of Environment 115, 3129–3152.

Li et al. (2002); T. Li, Q. Li, S. Shu, M. Ogihara; *A survey on wavelet applications in data mining*; ACM SIGKDD Explorations Newsletter , Volume 4 (2), Association for Computing Machinery.

Liang and Wu (2005); J. Liang, X. Wu; *Worm harm prediction based on segment procedure neural networks*; RSKT'06 Proceedings of the First international conference on Rough Sets and Knowledge Technology, Pages 383-388 .

Liao (2005); T. Liao; *Clustering of time series data - a survey*; Pattern Recognition 38, 1857–1874.

Lin et al. (2008); C. Lin, J. Chiu, M. Hsieh, M. Mok, Y. Li, H. Chiu; *Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks*; Computer methods and programs in biomedicine, 92, 193-197.

Lin et al. (1995); D. Lin, J. Dayhoff, P. Ligomenides; *Trajectory production with the adaptive time-delay neural network*; Network, Vol.8, No.3, pp. 447-461.

Lindsay and Cox (2004); D. Lindsay, S. Cox; *Improving the reliability of decision tree and naive bayes learners*; Proc. of the 4th ICDM, IEEE 459–462.

Liu and Motoda (1998); H. Liu, H. Motoda; *Feature extraction, construction and selection: a data mining perspective*; Kluwer Academic, Boston.

Ljung (1999); L. Ljung; *System identification: theory for the user*; Prentice Hall, 2nd edition.

Loh (2003); R. Loh; *Time series forecast with neural network and wavelet techniques*; Master Thesis; School of Information Technology and Electrical Engineering , Department of Electrical and Computer Engineering, University of Queensland.

Lourme (2010); A. Lourme; *Simultaneous gaussian model-based clustering for samples of multipleorigins*; Biometrics , 1–25.

Lucas (2010); D. Lucas; *Classification techniques for time series and functional data*; PhD thesis, Universidad Carlos III de Madrid, Departamento de Estadisticaica.

Maglaveras et al. (1998); N. Maglaveras, T. Stamkopoulos, C. Pappas, M. Strintzis; *An adaptive backpropagation neural network for real-time ischemia episodes detection: development and performance analysis using european st-t database*; Transactions on Biomedical Engineering, vol. 45, issue 7, pp. 805-813.

Makridakis et al. (1998); S. Makridakis, C. Wheelwright, R. Hyndman; *Forecasting, methods and applications*; 3rd ed. New York:Wiley.

Makridakis and Hibon (2000); S. Makridakis, S. Hibon; *The m3-competition: results, conclusions and implications*; International Journal of Forecasting 16 (4) (2000) 451–476.

Mallat (1989); S. Mallat; *A theory for multi-resolution signal decomposition: the wavelet representation*; IEEE Transactions on Pattern Analysis and Machine Intelligence, 11:674-693.

Mancia et al. (2007); G. Mancia, G. Backer, A. Dominiczak, R. Cifkova, R. Fagard, G. Germano, G. Grassi, A. Heagerty, E. Kjeldsen, S. Laurent; *2007 guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the european society of hypertension (esh) and of the european society of cardiology (esc)*; Eur Heart J. 2007 Jun;28(12):1462-536.

Mancini et al. (2008); L. Mancini, L. Corazza, D. Cannarile, M. Soverini, S. Cavalcanti, S. Cavani, A. Fiorenzi, A. Santoro; *Short term variability of oxygen saturation during hemodialysis is a warning parameter for hypotension appearance*; Computers in Cardiology, 35:881-883.

Manocha and Singh (2011); A. Manocha, L. Singh; *An overview of ischemia detection techniques*; International Journal of Scientific & Engineering Research, Volume 2, Issue 11, November-2011 1, ISSN 2229-5518.

Marcellino et al. (2005); M. Marcellino, J. Stock, M. Watson; *A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series*; ; CEPR Discussion Paper No. 4976.

Marchiando and Elston (2003); J. Marchiando, M. Elston; *Automated ambulatory blood pressure monitoring: clinical utility in the family practice setting*; American Family Physician, 67, 11, pp 2343-2350.

Martin and Flandrin (1985); W. Martin, P. Flandrin; *Wigner-ville spectral analysis of nonstationaryprocesses*; IEEE transactions on acoustics, speech, and signal processing, vol. Assp-33, no. 6, december.

Mcculloch and Pitts (1943); W. Mcculloch, W. Pitts; *A logical calculus of the ideas immanent in nervous activity*; Bulletin of Mathematical Biophysics, 9, 127-147.

Meyer (1987); Y. Meyer; *Uncertainty principle, hilbert bases, and algebras of operators*; Seminaire Bour-Baki, 145-146 , 209.

Michel and Struys (2006); M. Michel, M. Struys; *Closed loops in anaesthesia; best practice & research clinical*; Anaesthesiology, Volume 20, Issue 1, pp 211-220.

Milosavljevic and Petrovic (2006); N. Milosavljevic, A. Petrovic; *St segment change detection by means of wavelets*; 8th Seminar on Neural Network Applications in Electrical Engineering, NEUREL.

Minsky and Papert (1969); M. Minsky, S. Papert; *Perceptrons*; MIT Press, Cambrigde, USA.

Mohebbi and Moghadam (2007); M. Mohebbi, H. Moghadam; *Real-time ischemic beat classification using backpropagation neural network*; Signal Processing and Communications Applications, pp. 1-4.

Montani et al. (2006); S. Montani, L. Portinale, G. Leonardi, G. Bellazzi, R. Bellazzi; *Case-based retrieval to support the treatment of end stage renal failure patients*; Artificial Intelligence in Medicine 37 (2006) 31-42.

Moody and Lehman (2009); G. Moody, L. Lehman; *Predicting acute hypotensive episodes: the 10th annual physionet/computers in cardiology challenge*; Vol 36, Computers in Cardiology.

Moon et al. (2002); Y. Moon, K. Whang, W. Han; *General match: a subsequence matching method in time-series databases based on generalized windows*; SIGMOD, pp. 382-393.

Morrettin (1997); P. Morrettin; *Wavelets in statistics*; Resenhas, 3, 211-272.

Nason and Sachs (1999); G. Nason, R. Sachs; *Wavelets in time series analysis*; Phil. Trans. R. Soc. Lond. A, 357, 2511-2526.

Noren et al. (2009); G. Noren, J. Hopstadius, A. Bate, K. Star, I. Edwards; *Temporal pattern discovery in longitudinal electronic patient records*; Data Mining and Knowledge Discovery archive, Volume 20 Issue 3, May 2010, pp 361 - 387.

Pan and Wang (1998); Z. Pan, X. Wang; *A stochastic nonlinear regression estimator using wavelets*; Computational Economics, 11, 89-102.

Pandit and Wu (1983); S. Pandit, S. Wu; *Time series and system analysis with applications*; New York: Wiley & Sons.

Pang et al. (2005); L. Pang, I. Tchoudovski, M. Braecklein, K. Egorouchkina, W. Kellermann, A. Bolz; *Real time heart ischemia detection in the smart home care system*; 27th Annual International Conference of the Engineering in Medicine and Biology Society, pp. 3703-3706.

Papaloukas et al. (2002); C. Papaloukas, D. Fotiadis, A. Likas, L. Michalis; *An ischemia detection method based on artificial neural networks*; Artificial Intelligence in Medicine, vol. 24, issue 2, pp. 167-178.

Papaloukas et al. (2001); C. Papaloukas, D. Fotiadis, A. Likas, A. Liavas, L. Michalis; *A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms*; Med Biol Eng Comput. Jan, 39(1): pp 105-112.

Park et al. (2000); S. Park, W. Chu, J. Yoon, C. Hsu; *Efficient searches for similar subsequences of different lengths in sequence databases*; Proc. of the International Conference of Data Engineering, 23-32.

Parker (1987); D. Parker; *Optimal algorithms for adaptive networks: second order back propagation, second order direct propagation and second order hebbian learning*; IEEE 1st International conference on Neural Networks, 2, 592-600.

Patil and Kumaraswamy (2009); S. Patil, Y. Kumaraswamy; *Intelligent and effective heart attack prediction system using data mining and artificial neural network*; European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4, pp.642-656.

Pelosi et al. (1999); G. Pelosi, M. Emdin, C. Carpeggiani, M. Morales, M. Piacenti, P. Dattolo, T. Cerrai, A. Macerata, A. L'abbate, Q. Maggiore; *Impaired sympathetic response before intradialytic hypotension: a study based on spectral analysis of heart rate and pressure variability*; Clinical Science 96, 23-31.

Percival and Walden (2000); D. Percival, A. Walden; *Wavelet methods for time series analysis*; Cambridge University Press.

Perry and Roccella (1998); H. Perry, E. Roccella; *Conference report on stroke mortali- ty in the southeastern united states*; Hypertension;31: 1206–15.

Piccini and Nilsson (2006); J. Piccini, K. Nilsson; *The osler medical handbook*; 2nd Edition, Chapter 20-Hypotension and Shock; R. Scott Stephens, S. Haldar and C. Wiener; Elsevier, Saunders.

Pintelon and Schoukens (1991); R. Pintelon, J. Schoukens; *Identification of linear systems: a practical guideline to accurate modelling*; Pergamon Press.

Popivanov and Miller (2002); I. Popivanov, R. Miller; *Similarity search over time-series data using wavelets*; Proc. of the 18th International Conference on Data Engineering, 212.

Popoola (2007); A. Popoola; *Fuzzy-wavelet method for time series analysis*; PhD thesis, Department of Computing School of Electronics and Physical Sciences , University of Surrey.

Priestley (1996); M. Priestley; *Wavelets and time-dependent spectral analysis*; J. Time Series Analysis, 17, 85-104.

Principe et al. (2000); J. Principe, E. Euliano, W. Lefebvre; *Neural and adaptive systems: fundamentals through simulation*; John Wiley & Sons.

Qian and Chen (1996); S. Qian, D. Chen; *Joint time-frequencyanalysis - methods and applications*; Prentice-Hall, Upper Saddle River, NJ, p. 101-131.

Rajaee et al. (2010); T. Rajaee, S. Mirbagheri, V. Nourani, S. Alikhani; *Prediction of daily suspended sediment load using wavelet and neuro- fuzzy combined model*; Int. J. Environ. Sci. Tech., 7 (1), 93-110.

Ramsey (2002); J. Ramsey; *Wavelets in economics and finance: past and future*; Stud Nonlinear Dynam Economet, 6 (3), pp. 1–27.

Ranjith et al. (2003); P. Ranjith, P. Baby, P. Joseph; *Ecg analysis using wavelet transform: application to myocardial ischemia detection*; ITBM-RBM, vol. 24, issue 1, pp. 44-47.

Reich et al. (2005); D. Reich, S. Hossain, M. Krol; *Predictors of hypotension after induction of general anesthesia*; Anesthesia Analg, 101:622– 8.

Reiter and Maglaveras (2009); H. Reiter, N. Maglaveras; *Heartcycle: compliance and effectiveness in hf and cad closed-loop management*; EMBC-09, 31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, USA, September 2-6, pp 299 - 302.

Renaud et al. (2005); O. Renaud, J. Starck, F. Murtagh; *Wavelet-based combined signal filtering and prediction*; IEEE Transactions SMC, Part B, Volume 35, Issue 6, 1241 - 1251.

Renaud et al. (2003); O. Renaud, J. Starck, F. Murtagh; *Prediction based on a multiscale decomposition*; International Journal of Wavelets, Multiresolution and Information Processing, Vol. 1, No. 2, pp. 217-232.

Riaz and Ahmed (2012); K. Riaz, A. Ahmed; *Hypertensive heart disease*; Department of Internal Medicine, Section of Cardiology, Wright State University, http://emedicine.medscape.com/article/162449-overview.

Rivals (1995); I. Rivals; *Modélisation et commande de processus par réseaux de neurones: application au pilotage dun véhicule autonome*; PHd thesis, Université Paris 6.

Rivals and Personnaz (1995); I. Rivals, L. Personnaz; *Black-box modelling with state-space neural networks*; Neural Adaptive Control Technology, Chap. 8, 237-263.

Roberts et al. (2004); S. Roberts, E. Roussos, R. Choudrey; *Hierarchy, priors and wavelets: structure and signal modeling using ica*; Signal Processing, 84: 283-297.

Rocha et al. (2008); T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Antunes; *Phase space reconstruction approach for ventricular arrhythmias characterization*; EMBC-08, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, pp 5470-5473.

Rocha et al. (2010); T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Harris, J. Morais, M. Antunes; *A lead dependent ischemic episodes detection strategy using hermite functions*; Biomedical Signal Processing and Control 5, 271-281.

Rocha et al. (2011); T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Antunes; *A wavelet-based approach for time series pattern detection and events prediction applied to telemonitoring data*; EMBC2011 - International Conference Engineering in Medicine and Biology Society, pp 6037- 6040.

Rocha et al. (2009); T. Rocha, S. Paredes, P. Carvalho, J. Henriques, M. Harris, J. Morais; *Ischemia detection in the context of a cardiovascular status assessment tool*; EMBC - 2009, 31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, USA, Septembert, 1-5.

Rocha et al. (2011); T. Rocha, S. Paredes, P. Carvalho, J. Henriques; *Prediction of acute hypotensive episodes by means of neural network multi-models*; Comput Biol Med. 2011 Oct;41(10):881-90. Epub 2011 Sep 6.

Roger (1993); B. Roger; *Proofs without words: exercises in visual thinking*; The Mathematical Association of America, Washington, DC.

Rosenblatt and Rosenblatt (1958); F. Rosenblatt, F. Rosenblatt; *The perceptron: a probabilistic model for information storage and organization in the brain*; Psychological Review, 65, 386-408.

Rumelhart et al. (1986); D. Rumelhart, G. Hinton, R. Williams; *Learning internal representations by error propagation*; PDP Research Group, Parallel Distributed Processing MIT Cambridge, USA.

Rutkowski (2004); L. Rutkowski; *Generalized regression neural networks intime-varying environment*; IEEE Transactions on neurall networks, Vol. 15, N. 3.

Saeed et al. (2002); M. Saeed, C. Lieu, G. Raber, R. Mark; *Mimic ii: a massive temporal icu patient database to support research in intelligent patient monitoring*; Computers in Cardiology, 29:641-644.

Saeed and Mark (2006); M. Saeed, R. Mark; *A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations*; AMIA Annu Symp Proc. 2006; 679–683.

Saeed (2007); M. Saeed; *Temporal pattern recognition in multiparameter icu data*; Doctoral dissertation, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Sang et al. (2010); Y. Sang, D. Wang, J. Wu; *Entropy-based method of choosing the decomposition level in wavelet threshold de-noising*; Entropy,12, 1499-1513.

Schoukens and Pintelon (1991); J. Schoukens, R. Pintelon; *Identification of linear systems: a practical guideline to accurate modeling*; Pergamon Press.

Sedgewick (1978); R. Sedgewick*; Implementing Quicksort programs*; Communications of the ACM CACM Homepage archiveV.21, Issue 10, 847-857.

Shankaracharya et al. (2010); D. Shankaracharya, D. Odedra, S. Samanta, A. Vidyarthi; *Computational intelligence in early diabetes diagnosis: a review*; Rev Diabet Stud. 2010 Winter; 7(4): 252–262.

Shasha and Zhu (2004); D. Shasha, Y. Zhu; *High performance discovery in time series: techniques and case studies*; New York: Springer.

Shensa (1992); M. Shensa; *The discrete wavelet transform: wedding the "à trous" and mallat algorithms*; IEEE Trans. Signal Proc, vol. 40, no. 10, 2464-2482.

Sierra and Sierra (2008); C. Sierra, A. Sierra; *Early detection and management of the high-risk patient with elevated blood pressure*; Health and Risk Management 2008:4(2) 289–296289.

Singla et al. (2006); D. Singla, K. Suneet, A. Singh, T. Kau, S. Gupta; *Risk factors for development of early hypotension during spinal anaesthesia*; J Anaesth Clin Pharmacol, 22(4) : 387-393 387.

Sjoberg et al. (1995); J. Sjoberg, Q. Zhang, L. Ljung, A. Bienveniste, B. Deylon, B. Glorennec, D. Halmarsson, A. Juditsky; *Non-linear black box modelling in system identification: a unified overview*; Automatica, 31, 12, 1691-1724.

Sjoberg et al. (1994); J. Sjoberg, K. Hjalmarsson, L. Ljung; *Neural networks in system identification*; Proc. 10th Symposium on System Identification, 1, 167-190.

Smola and Schölkopf (2004); A. Smola, B. Schölkopf; *A tutorial on support vector regression*; Statistics and Computing, 14:199–222.

Sodestrom and Stoica (1989); T. Sodestrom, P. Stoica; *System identification*; Prentice Hall International.

Soltani et al. (2000); S. Soltani, D. Boichu, P. Simar, S. Canu; *The long term-memory prediction by multi-scale decomposition*; Signal Processing, 80: 2195-2205.

Soltani (2002); S. Soltani; *On the use of wavelet decomposition for time series prediction*; Neurocomputing, 48, pp. 267–277.

Spiegel et al. (2011); S. Spiegel, S. Spiegel, J. Gaebler, A. Lommatzsch, E. Luca, S. Albayrak; *Pattern recognition and classifiation formultivariate time series*; Sensor KDD San Diego, CA, USA.

Starck et al. (2007); J. Starck, J. Fadili, F. Murtagh; *The undecimated wavelet decompositionand its reconstruction*; IEEE transactions on image processing, vol. 16, no. 2, february.

Sun (2001); Y. Sun; *Arrhythmia recognition from electrocardiogram using non-linear analysis and unsupervised clustering techniques*; PhD Thesis, School of Electrical & Electronic Engineering.

Taddei et al. (1992); A. Taddei, G. Distante, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, C. Marchesi; *The european st-t database: standard for evaluating systems for the analysis of st-t changes in ambulatory electrocardiography*; European Heart Journal 13, pp. 1164-1172.

Takens (1981); F. Takens; *Detecting strange attractors in turbulence*; Dynamical Systems and Turbulence, 898, 366-381.

Tran and Reed (2004); N. Tran, D. Reed; *Automatic arima time series modeling for adaptive i/o prefetching*; IEEE transactions on parallel and distributed systems, vol. 15, no. 4.

Tsymbal (2003); T. Tsymbal; *Ensemble feature selection with the simple Bayesian classification*; Information fusion, Vol. 4, Issue.2, pp. 87-100.

Unser et al. (1998); M. Unser, P. Thevenaz, A. Aldrounbi; *Shift-orthogonal wavelet bases*; IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 46, NO. 7, JULY 1998.

Vaidehi (2008); V. Vaidehi; *A prediction system based on fuzzy logic*; Proceedings of the World Congress on Engineering and Computer Science WCECS 08, October 22 - 24, San Francisco, USA.

Vapnik (1998); V. Vapnik; *Statistical learning theory*; Wiley, 1st edition.

Varadharajan (2004); C. Varadharajan; *A wavelet-based system for event detection in online real-time sensor data*; Submitted in partial fulfillment of the requirements for the degree of Master of Science, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Vetterli et al. (2010); M. Vetterli, J. Kovacevic, V. Goya; *Fourier and wavelet signal processing*; Report, E´cole Polytechnique F´ed´erale de Lausanne and University of California, Berkeley.

Vila et al. (1997); J. Vila, J. Presedo, M. Delgado, S. Barro, R. Ruiz, F. Palacios; *Sutil: intelligent ischemia monitoring system*; International Journal of Medical Informatics, vol. 47, issue 3, pp. 193-214.

Wang et al. (2010); Z. Wang, L. Lai, D. Xiong, X. Wu; *Study on predicting method for acute hypotensive episodes based on wavelet transform and support vector machine*; 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010).

Wei and Billings (2006); H. Wei, S. Billings; *Long term prediction of non-linear time series using multiresolution*; International Journal of Control, 79: 6, 569 — 580.

Weigend et al. (1992); A. Weigend, B. Huberman, D. Rumelhart; *Predicting the future: a connectionist approach*; International Journal of Neural Systems, 1, pp. 193–209.

Werbos (1974); P. Werbos; *Beyond regression: new tools for prediction and analysis in the behavioural sciences (phd)*; Harvard University.

WHO (2011); C. WHO; *World health organization, fact sheet n^{o}317, September.*

Widrow and Hoff (1960); B. Widrow, M. Hoff; *Adaptive switching circuits*; IRE WESCON Convention Record,4, 96-104, NY.

Widrow and Smith (1963); B. Widrow, F. Smith; *Pattern-recognizing control systems*; Computer and Information Sciences, Spartan, Washington USA.

Wolf (2004); A. Wolf; *Automatic analysis of electrocardiogram signals using neural networks (in portuguese)*; PUC-Rio, Ms. Thesis, n^o 0210429/CA2004.

Wu et al. (1996); D. Wu, D. Agrawal, A. Abbadi, A. Singh, T. Smith; *Efficient retrieval for browsing large image databases*; Proc. of the fifth international conference on Information and knowledge management, 11-18.

Wyse et al. (1980); N. Wyse, R. Dubes, A. Jain; *A critical evaluation of intrinsic dimensionality algorithms*; Pattern Recognition in Practice, pp 415-425. Morgan Kaufmann Publishers, Inc, San Mateo, CA.

Yager and Filev (1993); R. Yager, D. Filev; *Learning of fuzzy rules by mountain clustering*; Conference, Boston, 246-254, 1993.

Yao et al. (2000); S. Yao, Y. Song, L. Zhang, X. Cheng; *Wavelet transform and neural networks for short-term electrical load*; Energy Conversion and Management, 41: 1975-1988.

Yazdizadeh and Khorasani (2000); A. Yazdizadeh , K. Khorasani ; *Identification of a two-link flexible manipulator using adaptive time delay neural networks*; IEEE Transactions on Systems, Man and Cybernetics, Part B, Vol.30 , No. 1, pp. 165-172.

Yazdizadeh and Khorasani (2002); A. Yazdizadeh , K. Khorasani ; *Adaptive time delay neural network structures for nonlinear system identification*; Neurocomputing , Vol.47, pp. 207-240.

Yevgeniy et al. (2005); B. Yevgeniy, M. Lamonova, O. Vynokurova; *An adaptive learning algorithm for a wavelet neural network*; Expert Systems, 22: 235-140.

Yousefi et al. (2005); S. Yousefi, I. Weinreich, D. Reinarz; *Wavelet-based prediction of oil prices*; Chaos, Solitons and Fractals 25, 265–275.

Zaciu et al. (1999); R. Zaciu, C. Lamba, C. Burlacu, G. Nicula; *Image compression using an overcomplete discrete wavelet transform*; IEEE Trans. Consum. Electron., vol. 42, no. 3, pp. 800–807.

Zhang et al. (2001); B. Zhang, R. Coggins, M. Jabri, D. Dersch, B. Flower; *Multiresolution forecasting for futures trading using wavelet decompositions*; IEEE transactions on neural networks, vol. 12, no. 4.

Zhang et al. (1998); G. Zhang, E. Patuwo, Y. Michael; *Forecasting with artificial neural networks: the state of the art*; International Journal of Forecasting, Volume 14, Issue 1, 1 March 1998, pp 35–62.

Zhao et al. (2006); A. Zhao, C. Zhang, B. Huan, Z. Cai, Q. Wang; *Application of unscented kalman filter for non-linear estimation in deformation monitoring*; 3rd IAG / 12th FIG Symposium, Baden, May 22-24.

Zheng et al. (1999); G. Zheng, J. Starck, J. Campbell, F. Murtagh; *Multiscale transforms for filtering financial data streams*; Journal of Computational Intelligence in Finance, Vol. 7, pp. 18-35.

Zheng et al. (2001); Y. Zheng, L. Zhiping, D. Tay; *State-dependent vector hybrid linear and nonlinear arma modeling: applications*; Circuits, Systems, and Signal Processing; Vol. 20, N. 5, 575-597.