3^{rd} **ed.**

Assessing Complexity of Physiological Interactions

Teresa Sarmento Henriques

PDICSS

PhD PROGRAM
CLINICAL AND HEALTH
SERVICES RESEARCH

ADVISORS:

Luis Filipe Antunes

Cristina Costa- Santos

February 2015

3^{rd} **ed.**

Assessing Complexity of Physiological Interactions

Teresa Sarmento Henriques

PDICSS

PhD PROGRAM
CLINICAL AND HEALTH
SERVICES RESEARCH

ADVISORS:

Luis Filipe Antunes

Cristina Costa- Santos

February 2015

Para os meus avós.

O voo redime.

Acknowledgments

First and foremost I would like to express my special appreciation and thanks to my advisor Luís Antunes, he has been a tremendous mentor for me. The happiness he has for his work was contagious and motivational for me. I would like to thank him for encouraging and allowing me to grow as a researcher. On a personal level, I am thankful for his compassionate supervision and the confidence he placed in me. I would also like to thank Cristina Costa Santos, she took me under her wing and introduced me to a new world of mathematical research leading me to work on diverse exciting projects. Their advice on both research as well as on my career have been priceless. A special thanks to the SisPorto group, Diogo Ayres de Campos, João Bernardes, Antónia Costa, Célia Costa, Hernâni Gonçalves, Paulo Sousa and Inês Nunes for welcomed me to the group since the first day and always generously sharing their time and ideas. I also would like to thank Altamiro Pereira and all the CINTESIS members for unwavering support and encouragement over the years.

I take this opportunity to state my gratitude to Madalena Costa and Ary Goldberger who have been essential in the successful conclusion of this project. To work with them have been a real pleasure to me, with heaps of fun and excitement. Madalena Costa ability to select and approach compelling research problems, her high scientific standards, and her hard work set an example. The joy and enthusiasm Ary Goldberger has for his research was contagious and so motivational for me. Ary have always been patient and encouraging in times of new ideas, have listened to my thoughts and discussions with him frequently led to key insights. Our endless amazing conversations made me learn precious informations (not only the definition of “several” or the number of members in Portuguese Assembly). To Sara Mariani I would like to thank the lovely company and the great mood everyday. It is a pleasure to work in such amazing group. I also would like to acknowledge all the Wyss Institute for Biologically Inspired Engineering members that help me always with the greatest kindness and warmth, especially Paula Fenton.

I will forever be thankful to Matilde Monteiro Soares and to Elisabete Neutel for their friendship that helped me through the school years. I have learned much though our “tête-á-têtes” that I already miss. I cannot say thank enough to Hernâni Gonçalves for the tremendous daily support, who held me in writing and heard my complaints and incented me to strive towards my goal against my laziness (even with the morning rides that did not allow me to sleep a little bit more). I would also like to thank all of my friends that struggle to understand what I was doing but never for a second doubt about me. Thank you all not only for accepting me for who I am but also for believing in me.

Last but not the least, an enormous thanks to my family. Words cannot express how grateful I am to them for all of the sacrifices that they’ve made on my behalf. Thanks to my mother, Inês Sarmento, for guiding me as a person, believing and wanting the best for me. She has taught me to never be afraid of the truth and openly express my thoughts and feelings. And I just wanna be like her when I grow older.

To my father, Manuel Rangel, thank you for the all kinds of measureless support, for encouraging me in all of my pursuits and inspiring me to follow my dreams. He has taught me to be the 'rock' and never complained, at least not loudly. To my siblings, there is no words for describe your love no matter where we are. I will never forget how difficult it was for them to accept my decision to leave. To my new family Lapa I have to thank them for welcoming me, since the beginning, as one of them. Our affection have being growing for eight years and will keep growing for many more. At the end, how can I ever possibly thank to my beloved husband, Paulo Lapa, the most kind-hearted, honest, generous and upright person I have ever known. I will, forever and ever, remember the time he agrees to marry me and change our entire live for that I could pursuit my carrier. He makes me feel special even when times are tough. He has been my pillar of strength through all my ups and downs.

To everyone thank you for your thoughts, well-wishes/prayers, phone calls, e-mails, texts, visits, advice, and being there whenever I needed.

Abstract

The human body is composed of interacting physiological systems whose actions are not always predictable. A healthy system is characterized by the self-adjusting capacity to respond to internal requirements and/or external influences. The analysis of mathematical approaches capable to deal with concepts of complexity, chaos and fractality are of importance to better understand the information contained in physiological signals (nonstationary and nonlinear time series). This notion contradicts that of traditional homeostatic mechanisms of control, whose goal is to maintain a constant baseline. Pathologic states and aging are associated with the breakdown of fractal correlations. There are evidence of complexity loss in lifethreatening conditions, either marked by transitions to periodic dynamics, associated with excessive order or for breakdown of organization, similar to uncorrelated randomness. The output of the mathematical measures can be of practical diagnostic and prognostic use.

The central aspect of this thesis is exploring complexity analysis methodology in medical and decision making research by providing a mathematical background (and software) to deal with the biological data.

One of the most explored physiological interactions, in the past few years, have been the dynamics of the heart rate assessed through the acquisition of the electrocardiogram (ECG) and cardiotocogram (CTG) signals. In **Chapter 2** the main methods used to characterize the variability human heart rate (HRV) are reviewed.

The use of measures based on complexity to assess observer disagreement, almost always present in chaotic scenarios of physiological interactions, is other objective of this thesis (**Chapter 3**). On **Section 3.1** is presented a generalization, for more than two observers, of the information-based measure of disagreement (IBMD). The IBMD, an observer disagreement measure, based of Shannon's notion of entropy, uses logarithms to measures the amount of information contained in the differences between observations. The software created to allow easy assess of observer (dis)agreement measures, a website and a R package, are explored no **Section 3.2**.

Two conceptually different complexity measures: the entropy, a probabilistic approach and the compression, an algorithmic approach are explored on **Chapter 4**. The entropy using the approximate entropy (ApEn) and sample entropy (SampEn) measures and the Kolmogorov complexity through different compressors, as the Lempel-Ziv, the bzip2 and the paq8l, were applied to a set of heart rate signals with the goal of characterizing different pathologies. The first application of these measures in this thesis, **Section 4.1**, was on the analysis of fetal heart rate (FHR). The results presented on this section validate the notion of, although less used, compressors can be effectively used, complementary to entropy indices, to quantify complexity in biological signals.

The multiscale entropy (MSE) approach capture the information "hidden" across multiple spatial and temporal scales obtaining better results in the analysis of dynamics systems and characterizing

pathologies. On **Section 4.2** this approach is modified and the multiscale compression is proposed. In this approach the coarse-grained procedure used to create a new time series for each scale maintains but the calculation of entropy values is replaced by the estimation of the compression rate. This new measure was successfully applied to two data sets: 1) a data set with 40 white-noise and 40 pink-noise time series; 2) a data set of cardiac interbeat interval time series from forty-three congestive heart failure patients and seventy-two healthy subjects.

The low correlation between the two complexity approaches (entropy and compression) obtained in the results of **Sections 4.1 and 4.2** reinforce the importance of considering different measures to assess different physical and physiologic time series. On **Section 4.3** the two complexity approaches are explored challenging complexity from different viewpoints and trying to understand the complementary information that can be derived from both, providing a more comprehensive view of the underlying physiology.

The utility of many of the methods created or used to assess HRV is not exclusive to that purpose. Diabetes mellitus (DM) is one of the world's most prevalent medical conditions affecting tens of millions worldwide. On **Chapter 5** some traditional HRV methods, as Poincaré maps, MSE and detrended fluctuation analysis (DFA) were used with success to assess continuous glucose monitoring (CGM) data acquired during several days. On **Section 5.1** a new visualization tool - *glucose-at-glance* - based on the Poincaré maps and colorized by density, is presented as an easy and quick way to facilitate the assessment of data complexity acquired by CGM systems. On **Section 5.2** the long and short term glucose fluctuation were assessed and described, by using MSE and DFA, as well as the consistent complexity-loss in the CGM time series from patients with diabetes.

In conclusion, this thesis provides several contributions, based on the complexity concept, towards the measurement of disagreement and the evaluation of physiological signal. The confined number of signals in each dataset is a thesis limitation, and thus the obtained results should be supported with further clinical testing. Our findings may be useful in developing and testing mathematical models of physiologic regulation in health and disease.

Resumo

O corpo humano é composto por sistemas fisiológicos interactivos cujas acções nem sempre são previsíveis. Um sistema saudável é caracterizado pela sua adaptabilidade, a capacidade de se auto-ajustar em resposta a influências internas e/ou externas. A anáise de abordagens matemáticas capazes de lidarem com conceitos de complexidade, caos e fractalidade são de extrema importância para melhor entender a informação contida em sinais fisiológicos (séries temporais não-estacionárias e não-lineares). Esta noção contradiz a tradicional noção de homeostase biológica, cujo objectivo é manter um constante equilíbrio. Estados patológicos ou de envelhecimento são associados a perda de correlações fractais. Há evidência de decréscimo de complexidade em situações de risco de vida, quer caracterizadas por dinâmicas periódicas, associadas com excessiva ordem quer com a quebra organizativa similar a aleatoridade não correlacionada. O resultado de medidas matemáticas pode ajudar ao diagnóstico ou prognóstico dessas mesmas condições.

O aspecto central desta tese é explorar o conceito de complexidade no contexto da investigação clínica proporcionando uma base de conhecimentos matemáticos (e *software*) capaz de lidar com dados biológicos.

Uma das interações fisiológicas mais investigadas é a dinâmica da frequência cardica acessivel através da aquisição de electrocardiogramas (ECG) e cardiotocogramas (CTG). No **Capítulo 2** são revistos os métodos mais utilizados, nos últimos anos, para a caracterização da variabilidade da frequência cardiaca (VFC) humana.

O uso de medidas de complexidade para avaliar a concordância entre observadores, quase sempre presente em cenário de interações fisiológicas pertencentes ao domínio caótico, é outro dos objectivos da tese (**Capítulo 3**). Na **Secção 3.1** é apresentada a generalização, para mais do que dois observadores, da medida *information-based measure of disagreement* (IBMD). A medida de discordância IBMD, baseada na noção de entropia proposta por Shannon, utiliza logaritmos para medir a quantidade de informação contida na diferença entre observações. O *software* criado para facilitar o c lculo de medidas de concordância (ou discordãncia) explorado na **Secção 3.2**.

Duas medidas de complexidade conceptualmente diferentes: a entropia, uma abordagem probabilistica e a compressão, uma abordagem algorítmica são investigadas no **Capítulo 4**. A entropia usando as medidas: approximate entropy (ApEn) e sample entropy (SampEn) e a complexidade de Kolmogorov através de diferentes compressores, como são exemplo o Lempel-Ziv, o bzip2 e o paq8l, são aplicados a um conjunto de sinais de frequência cardíaca com o objectivo de caracterizar diferentes patologias. A primeira aplicação destas medidas nesta tese, **Secção 4.1**, foi na análise de batimento cardíaco fetal. Os resultados provenientes deste estudo validaram a noção que, embora menos usados, os compressores devem ser considerados, complementaremente aos índices de entropia, para quantificar sinais biológicos.

O método *Multiscale Entropy* (MSE) consegue avaliar a informação contida nas dinâmicas de sinais fisiológicos "encobertas" nas múltiplas escalas espaciais e temporais, obtendo melhores resultados na

análise de sistemas dinâmicos e na caracterização de algumas patologias. A ideia principal da **Secção 4.2** é extender este método à compressão, ou seja, manter a criação de escalas através de uma metedologia de *coarse-grained* mas de seguida calcular o valor da taxa de compressão em substituição do valor de entropia. Esta nova medida foi aplicada com sucesso em dois conjuntos de dados: 1) 40 séries temporais de ruído branco e 40 séries temporais de ruído fractal; 2) 43 séries temporais de frequência cardíca de pacientes com insuficiência cardíaca congestiva e 72 séries temporais the sujeitos saudáveis.

A baixa correlação entre as tuas abordagens (entropia e compressão) obtidas nos resultados das **Secções 4.1 e 4.2** sugerem que, embora ambas tentem quantificar o conceito de complexidade, cada uma procura diferentes padrões e comportamentos. Estes resultados realçam a importância de considerar diferentes medidas para avaliar diferentes séries temporais. Na **Secção 4.3** os diferentes pontos de vista das duas abordagens são exploradas tentando perceber que informação complementar provém do uso de cada uma das abordagens.

A utilidade de muitos dos métodos criados ou usados para avaliar a variabilidade da frequência cardíaca não é exclusiva desse propósito. Diabetes é uma das doenças com maior prevalência mundial que afecta milhares de pessoas e a sua incidência tem vindo a aumentar a uma taxa alarmante devido à sua associação com a obesidade. No **Capítulo 5** alguns dos métodos de variabilidade de frequência cardíaca, como o mapas de Poincaré, a MSE e a *detrended fluctuation analysis* (DFA) são usado com êxito na avaliação de dados obtidos através de monitorização contínua da glicose (MCG) durante vários dias. Na **Secção 5.1** uma nova ferramenta de visualização - *glucose-at-glance* - baseada nos mapas de Poincaré e colorizado pela densidade, é apresentada como uma forma fácil e rápida de avaliar os dados complexos adquiridos. Na **Secção 5.2** as fluctuações de longo e curtos prazo da glicose são analisadas e descritas, usando os métodos MSE e DFA, assim como a perda de complexidade nos sinais MCG de pacientes com diabetes.

Em conclusão, esta tese contribui com diversas metodologias, baseadas no conceito de complexidade, no sentido de avaliar discordância entre observadores e a dinâmica de sinais fisiológicos. O número reduzido de séries temporais dos diversos conjuntos de dados são uma limitação da tese e portanto os resultados devem ser verificados com futuros testes clínicos. As conclusões aqui presentes poderão ser úteis para desenvolver e testar métodos matemáticos que ajudem a regulação fisiológica na saúde e na doença.

Scientific and Financial Results

The following scientific results are part integral of this thesis:

Published articles:

Teresa Henriques, Luís Antunes, João Bernardes, Mara Matias, Diogo Sato and Cristina Costa-Santos. Information-Based Measure of Disagreement for more than two observers: a useful tool to compare the degree of observer disagreement. BMC Medical Research Methodology. Volume 13, Issue 1, Article 47, March 2013. <http://www.biomedcentral.com/1471-2288/13/47>

Teresa Henriques, Hernâni Gonçalves, Luís Antunes, Mara Matias, João Bernardes and Cristina Costa-Santos. Entropy And Compression: Two Measures Of Complexity. Journal of Evaluation in Clinical Practice. Volume 19, Issue 6, Pages 1101-1106, December 2013.
<http://www.ncbi.nlm.nih.gov/pubmed/23809085>

Teresa Henriques, Medha N. Munshi, Alissa R. Segal, Madalena D. Costa and Ary L. Goldberger. "Glucose-at-a-Glance:" New Method to Visualize the Dynamics of Continuous Glucose Monitoring Data. Journal of Diabetes Science and Technology. Volume 8, Issue 2, Pages 299-306, March 2014.
<http://www.ncbi.nlm.nih.gov/pubmed/24876582>

Madalena D. Costa, Teresa Henriques, Medha N. Munshi, Alissa R. Segal and Ary L. Goldberger. Dynamical Glucometry: Use of Multiscale Entropy Analysis in Diabetes. Chaos. Volume 24, Issue 3, Page 033139, September 2014. <http://www.ncbi.nlm.nih.gov/pubmed/25273219>

Submitted articles:

Teresa Henriques, Luís Antunes, Mara Matias, Cristina Costa-Santos. Facilitating the evaluation of agreement between measurements.

Teresa Henriques, Francisco Mota, João Bernades, Cristina Costa-Santos, Luís Antunes. Compression vs Entropy: on the characterization of the nonlinear features of physiological signal.

Software:

Teresa Henriques, Luís Antunes and Cristina Costa-Santos (2013). obs.agree: An R package to assess agreement between observers. R package version 1.0
<http://cran.r-project.org/web/packages/obs.agree/index.html>

This work was supported by the Portuguese National Science Foundation, though the PhD grant SFRH /BD/70858/2010.

Table of Contents

List of Figures

List of Tables

Acronyms

AF Atrial Fibrillation

ApEn Approximate Entropy

CCC Lin's concordance correlation coefficient

CD Correlation Dimension

CGM Continuous Glucose Monitoring

CHF Congestive Heart Failure

CI Complexity Index

CIDES Department of Health Information and Decision Sciences

CINTESIS Center for Research in Health Technologies and Information Systems

CRAN Comprehensive R Archive Network

CTG Cardiotocogram

DFA Detrended Fluctuation Analysis

DM Diabetes Mellitus

ECG Electrocardiography

EEG Electroencephalography

FCT Fundação para a Ciência e Tecnologia

FD Fractal Dimension

FHR Fetal Heart Rate

FMUP Faculty of Medicine of the University do Porto

HbA1c hemoglobin A1c

HE Hurst Exponent

HR Heart Rate

HRV Heart Rate Variability

IBMD Information-Based Measure of Disagreement

ICC Intraclass Correlation Coefficient

LA Limits of Agreement

LE Lyapunov Exponent

LLE Largest Lyapunov Exponent

MI Mutual Information

MSC Multiscale Compression

MSE Multiscale Entropy

NNT Numerical Noise Titration

PP Poincaré Plot

RAI Raw Agreement Indices

RP Recurrence Plot

RR RR interval

ROC Receiver Operating Characteristic

SampEn Sample Entropy

SD Standard Deviation

SE Shannon Entropy

UAB Umbilical Artery Blood

Introduction

Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less.

Marie Curie

1. Introduction

The human body is composed of several interacting physiological systems whose actions are not always predictable. The traditional notion of physiological systems with homeostatic mechanisms of control [1], whose goal is to maintain a constant baseline, is now contradicted by concepts of chaos, fractality and complexity supporting the theory of a healthy physiological system characterized by a self-adjusting capacity which degrades with pathologic states or aging [2].

In everyday parlance, chaos implies a state of confusion or disorder, a total lack of organization. In science, however, the term chaos [3] refers to a mathematical approach dealing with systems that are fully describable, but which generate random-appearing outputs under certain conditions. Chaos theory deals with the patterns in the time evolution of a system that has sensitivity to initial conditions, i.e., small differences between two sets of initial conditions can lead to huge discrepancies at later times. This mix of rule (determinism) and unpredictability (stochasticity) appears to be pervasive in nature, influencing cloud patterns, ocean currents and the flow of blood through branching blood vessels. Edward Lorenz, an MIT mathematician and meteorologist, and a pioneer in this field, pithily defined chaos as occurring "when the present determines the future, but the approximate present does not approximately determine the future [4]."

A concept closely related to chaos is that of fractality, introduced by Mandelbrot [5–7]. Fractal mathematics can be use to describe many branching or folded forms in biology, both in physiology (e.g., lungs, kidneys, blood vessels, brains, intestines) and in the diversity of life (forest, trees, flowers, plants, certain weather systems) [8–10]. The concept also applies to the output of complex systems that lack a characteristic scale. The main idea of fractality is that patterns look similar (in form or statistically) to themselves on different scales of magnification.

On the other hand, complexity define the amount of structured information of each system. Shannon demonstrated how the information within a signal could be quantified with absolute precision [11] as the amount of unexpected data contained in the message (designated 'entropy'). Subsequently, the Kolmogorov complexity was proposed to quantify information on individual objects as the size of its smallest representation [12]. The Shannon information theory measures the average information from a random source, unlike Kolmogorov complexity that presents a form of absolute information [13].

Current knowledge suggests that the quantification and classification of physiologic signal dynamics contribute to a better understanding of the underlying physiology observed through time course. In the past decades the concern (and consequent improvement) with data collection techniques has been growing, in particular with the acquisition of information during a continuous period of time. The amount of data available is expanding, the quality and the refinement (higher frequency sampling) are

improving and the access to new medical information is now possible. A good example is that of heart rate dynamics assessment through the acquisition of the electrocardiogram (ECG) and cardiotocogram (CTG) signals [14]. Accordingly, the RR intervals, the heart rate (HR) and the fetal heart rate (FHR) are three of the most analyzed signals in research. Concepts of chaos, complexity, fractality, among others have been used as support to numerous nonlinear measures proposed over time [14–16]. Also, several studies indicate a reduction of fractality and complexity in pathological situations [16–24]. These results may lead to improvements in the pathologies assessment and in the treatment monitorization.

The utility of many of the methods created or used to assess heart rate variability (HRV) is not exclusive to that purpose. Diabetes mellitus (DM) is one of the world's most prevalent medical conditions affecting tens of millions worldwide and because of its association with obesity, the incidence of type 2 DM is increasing at an alarming rate. Recently, a continuous glucose monitoring (CGM) technology, with sensors implanted subcutaneously for about a week, has been used in clinical practice. However, the time series derived from CGM recordings remain a largely untapped source of dynamical information.

The difficulty of establishing a "gold standard" measure or even a set of characteristics/values that can describe a pathology or an event, comes, many times, from the disagreement between the observers. Plesk and Greenhalgh [25] conjectured components of health care as belonging to the simple, the complex or the chaotic domains based on the certainty-agreement diagram introduced by Stacy [26], Figure 1.1.

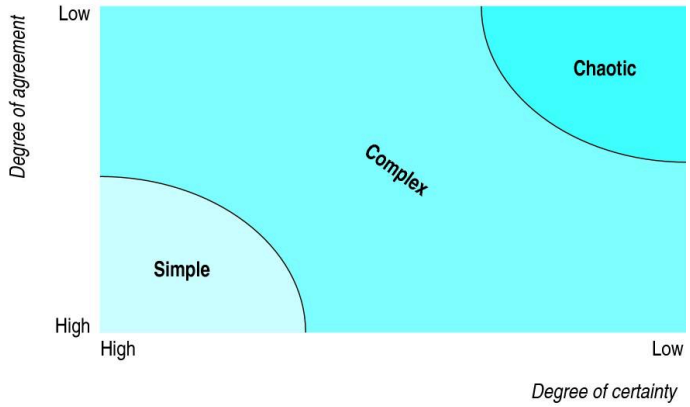

Figure 1.1: The certainty-agreement diagram from [25].

Traditional health problems, with low uncertainty, are generally located in the simple domain. In these scenarios it is reasonable to use linear models and uncomplicated protocols to guide clinical diagnosis and decisions. The complex domain is represented by dynamic systems and their interactions. In this domain the use of non-linear models and mining of patterns may be a helpful way to deal with higher levels of complexity. In the chaos domain (high uncertainty and disagreement) no information is directly visible from the interaction of the systems.

A robust measure of agreement becomes essencial to assess the interpretation of diagnostic tests and clinical decisions specially belonging to the complex or the chaotic domains. Inconsistent results are obtained when assessing observer agreement with most of the measures proposed [27]. In the particular case of continuous variables the limits of agreement (LA) and the intra-class correlation coefficient (ICC) are two of the most used measures despite their limitations. Recently a new measure, the information-based measure of disagreement (IBMD) [28], was introduced as a useful tool for comparing the degree of

observer disagreement. However, the proposed IBMD assesses disagreement between two observers only. A similar problem arises for categorical variables where the observer agreement has been assessed using the kappa statistics [29] and the raw agreement indices (RAI) [30].

1.1 Objectives

The central aspect of this thesis is exploring measures of complexity and fractality in medical research and decision making. In order to do so, more specific aims were proposed.

- as one of the most explored physiological interactions has been the dynamics of the heart rate, this thesis aims to review how human heart rate variability has been characterized;
- other objective of this thesis is to continue to explore the use of measures of complexity to assess observer disagreement almost always present in chaotic scenarios of physiological interactions;
- this thesis also aims to provide the mathematical background to deal with the biological complexity via the exploitation of measures as entropy and compression, in physiological signals as HR and CGM data.

1.2 Outline

Chapter 2 presents a review of the main nonlinear methods used, in the past few years, to characterize human heart rate dynamics, one of the most explored physiological interactions.

As observer disagreement is frequently found in the complex physiological interactions on **Chapter 3** the state of the art of observer agreement measurement is review. A generalization for more than two observers of the IBMD, an observer disagreement measure based of Shannon's notion of entropy, is presented on **Section 3.1**. To help health care professionals and biostatisticians when performing observer agreement studies an intuitive web-based software system was created. The website facilitate an easy application of several statistical agreement measures strategies to their own data. An R package obs.agree was also developed for an easy computation of observer agreement measures. This software package provides an easy way to calculate the RAI for categorical and the IBMD for continuous variables. On **Section 3.2** both platforms are explored and an application of the two functions included in the obs.agree package as well the interpretation of the results obtained are presented.

On **Chapter 4** the differences between the entropy and compression as complexity measures of heart rate dynamics are evaluated. The first application of these measures (**Section 4.1**) was on the analysis of fetal heart rate (FHR). The results presented on this section validate the notion that compressors can be effectively used, complementary to entropy indices, to quantify complexity in biological signals, specifically in FHR analysis. The main ideia on **Section 4.2** is to extend the multiscale entropy approach [22] to compression. The new multiscale compression method maintains the coarse-grained procedure to create a new time series for each scale but the calculation of entropy values is replaced by the estimation of the compression rate. This new measure was successfully applied to two datasets: a dataset with 40 white-noise and 40 pink-noise time series and another dataset of cardiac interbeat interval time series from forty-three congestive heart failure (CHF) patients and seventy-two healthy subjects. The low correlation

between the two complexity approaches (entropy and compression measures) obtained in the results of **Sections 4.1 and 4.2** suggest that although the two measures try to quantify the concept of complexity, each one of them is looking for different patterns/behaviors. These findings reinforce the importance of considering different measures to assess the different physical and physiologic time series. On **Section 4.3** the two complexity approaches are explored challenging complexity from different viewpoints and trying to understand the complementary information that can be derived from both, providing a more comprehensive view of the underlying physiology.

On **Chapter 5** some traditional HRV methods, as Poincaré maps, multiscale entropy (MSE) and detrended fluctuation analysis (DFA) were used with success to determine the different dynamics of glucose fluctuations in health and disease, during several days. On **Section 5.1** a new visualization tool - *glucose-at-glance* - based on the Poincaré maps and colorized by density, is presented as an easy and quick way to facilitate the assessment of data complexity acquired by CGM systems. The long and short term glucose fluctuation were assessed and described, as well as the consistent complexity-loss in the CGM time series from patients with diabetes. These results, presented on **Section 5.2**, support consideration of a new framework, *dynamical glucometry*, to guide mechanistic research and to help assess and compare therapeutic interventions.

Finally, the thesis ends with a general discussion and the conclusion (**Chapter 6**).

Bibliography

[1] Walter B. Cannon. Organization for physiological homeostasis. Physiological Reviews, 9(3), 1929.

[2] Ary L. Goldberger, David R. Rigney, and Bruce J. West. Chaos and fractals in human physiology. Scientific American, 262(2):42–49, 1990.

[3] Edward Ott. Chaos in dynamical systems. Cambridge University Press, Cambridge, United Kingdom, 2002.

[4] Christopher M. Danforth. Chaos in an atmosphere hanging on a wall. `http://mpe.dimacs.rutgers.edu/2013/03/17/chaos-in-an-atmosphere-hanging-on-a-wall/`. Accessed: Jun 2014.

[5] Benoît B. Mandelbrot. Fractals: form, chance and dimension. WH Freeman & Co., San Francisco, CA, USA, 1979.

[6] Benoît B. Mandelbrot. The fractal geometry of nature, volume 173. Macmillan, USA, 1983.

[7] Jean-François Gouyet and Benoît B. Mandelbrot. Physics and fractal structures. Masson Paris, 1996.

[8] James Gleick. Chaos: Making a new science. Viking-Penguin, 1987.

[9] Ary L. Goldberger and Bruce J. West. Fractals in physiology and medicine. The Yale Journal of Biology and Medicine, 60(5):421, 1987.

[10] Julien C.. Sprott. Chaos and time-series analysis, volume 69. Oxford University Press, Oxford, United Kingdom, 2003.

[11] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[12] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information. Problems of Information Transmission, 1(1):1–7, 1965.

[13] Ming Li and Paul M.B. Vitânyi. An introduction to Kolmogorov complexity and its applications. Springer, 2008.

[14] Lewis A. Lipsitz. Physiological complexity, aging, and the path to frailty. Science's SAGE KE, 2004(16):16, 2004.

[15] Ary L. Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. The Lancet, 347(9011):1312–1314, 1996.

[16] Ary L. Goldberger, Chung-Kang Peng, and Lewis A. Lipsitz. What is physiologic complexity and how does it change with aging and disease? Neurobiology of Aging, 23(1):23–26, 2002.

[17] Ary L. Goldberger, Larry J. Findley, Michael R. Blackburn, and Arnold J. Mandell. Nonlinear dynamics in heart failure: implications of long-wavelength cardiopulmonary oscillations. American Heart Journal, 107(3):612–615, 1984.

[18] Lewis A. Lipsitz and Ary L. Goldberger. Loss of "complexity" and aging: potential applications of fractals and chaos theory to senescence. Jama, 267(13):1806–1809, 1992.

[19] Lewis A. Lipsitz. Age-related changes in the "complexity" of cardiovascular dynamics: A potential marker of vulnerability to disease. Chaos, 5(1):102–109, 1995.

[20] Chung-Kang Peng, Shlomo Havlin, Jeffrey M. Hausdorff, Joseph E. Mietus, H. Eugene Stanley, and Ary L. Goldberger. Fractal mechanisms and heart rate dynamics: long-range correlations and their breakdown with disease. Journal of Electrocardiology, 28:59–65, 1995.

[21] David E. Vaillancourt and Karl M. Newell. Changing complexity in human behavior and physiology through aging and disease. Neurobiology of Aging, 23(1):1–11, 2002.

[22] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):021906, 2005.

[23] Madalena Costa, Ionita Ghiran, Chung-Kang Peng, Anne Nicholson-Weller, and Ary L. Goldberger. Complex dynamics of human red blood cell flickering: alterations with in vivo aging. Physical Review E, 78(2):020901, 2008.

[24] Madalena Costa, William T. Schnettler, Célia Amorim-Costa, João Bernardes, Antónia Costa, Ary L. Goldberger, and Diogo Ayres-de Campos. Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia. Early Human Development, 90(1):67–71, 2014.

[25] Paul E. Plsek and Trisha Greenhalgh. The challenge of complexity in health care. BMJ, 323(7313):625–628, 2001.

[26] Ralph D. Stacey. Strategic Management and Organisational Dynamics. Pitman, London, United Kingdom, 1996.

[27] Cristina Costa-Santos, João Bernardes, Diogo Ayres-de Campos, Antónia Costa, and Célia Costa. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. Journal of Clinical Epidemiology, 64(3):264–269, 2011.

[28] Cristina Costa-Santos, Luís Antunes, André Souto, and João Bernardes. Assessment of disagreement: A new information-based approach. Annals of Epidemiology, 20(7):555 – 561, 2010.

[29] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1):37, 1960.

[30] John S. Uebersax. Statistical methods for rater and diagnostic agreement. `http://www.john-uebersax.com/stat/raw.htm`. Accessed: Jun 2014.

Non-Linear Methods

Declare the past, diagnose the present, foretell the future.

Hippocrates

2. Non-Linear Methods

The dynamics of the heartbeat are one of the most analyzed physiological interactions [1]. Many mathematical methods were proposed and have been used for evaluate the heart rate variability. These measures have been successfully applied in research to expand knowledge concerning the cardiovascular dynamics in healthy as well as in pathologic conditions. However, they are still far from clinical medicine.

The review presented in this chapter focus on the most used nonlinear methods to assess the heart rate dynamics based on concepts of chaos, fractality and complexity.

In everyday parlance, chaos implies a state of confusion or disorder, a total lack of organization. In science, however, the term chaos [2] refers to a mathematical approach dealing with systems that are fully describable, but which paradoxically generate random-appearing outputs under certain conditions. Chaos theory deals with the patterns in the time evolution of a system that has sensitivity to initial conditions, i.e., small differences between two sets of initial conditions can lead to huge discrepancies at later times.

The term 'fractal' , first introduced by Mandelbrot [3], is a geometric concept related to, but not synonymous with, chaos [4, 5]. A fractal is an object composed of subunits (and sub-subunits) that resemble the larger scale structure, a property known as self-similarity. The property self-similarity or scale invariance means that the details of the structures are similar, but not necessarily identical, when zooming at different resolutions. A fractal organization is flexible, and breakdown of this scale invariance may lead to a more rigid and less adaptable system with either random or highly correlated behavior of heart rate dynamics [6]. The definition of fractal goes beyond self-similarity per se to exclude trivial self-similarity and include the idea of a detailed pattern repeating itself. As mathematical equations, fractals are usually nowhere differentiable [3]. A key feature of the class of fractals seen in biology is a distinctive type of long-range order [7]. Although fractals are irregular, not all irregular time series are fractal. The self-similarity of the systems fluctuations can be observer when analyzed over different time scale, as seconds, minutes or hours.

The complexity is a property of every system that quantifies the amount of structured information. Shannon demonstrated how the information within a signal could be quantified with absolute precision [8] as the amount of unexpected data contained in the message (designated 'entropy').

2.1 Heart Rate

The electrocardiogram (ECG) records the electrical impulses generated by the polarization and depolarization of cardiac tissue and translating into a waveform used to measure the rate and regularity of the cardiac cycle. A typical ECG tracing of the cardiac cycle (heartbeat) consists of a P wave, a QRS

complex, a T wave, and a U wave (Figure 2.1). The focus of this chapter will be in the QRS complex, in particular, in the methods developed and used to analyze RR interval time series - a set of the time interval between two consecutive R waves (peaks).

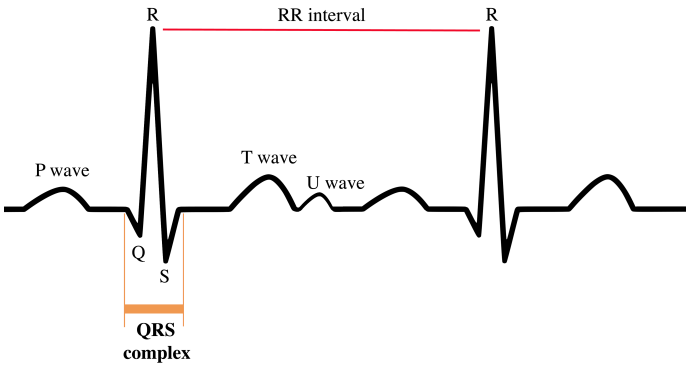

Figure 2.1: Schematic diagram of normal sinus rhythm for a human heart.

The normal sinus rhythm is now known by having a complex behavior that degrades with disease [9]. Two of the most studied heart pathologies are the congestive heart failure (CHF) and the atrial fibrillation (AF). The former is characterized by a period 1-type dynamics while the latter present a random, disorganized temporal structure (Figure 2.2).

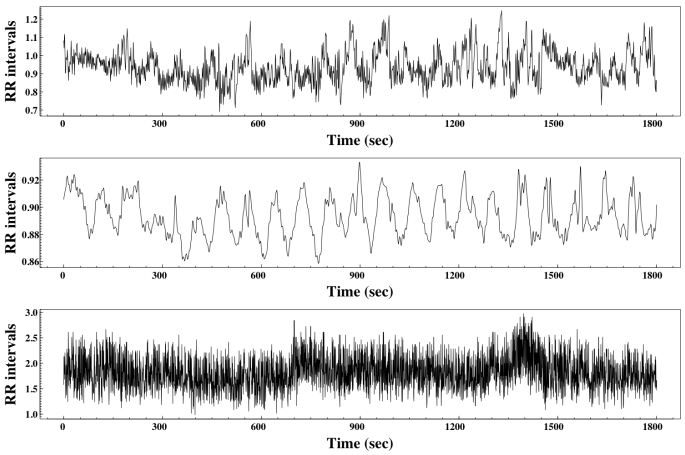

Figure 2.2: Representative RR time series from a healthy subject (top panel); a pathological breakdown of fractal dynamics, leading to single-scale period (CHF) (middle panel) and an uncorrelated randomness (AF) (bottom panel).

Let's start by considering some notation used in this chapter. We define a time series as a set of N consecutive data points $X = \{x(i), i = 1, ..., N\}$, where, in this case, each point will represent the value of a RR interval. From the original time series X, let's define the vectors $X_m^\tau(i)$ as:

$$X_m^\tau(i) = (x(i), x(i+\tau), x(i+2\tau), \ldots, x(i+(m-1)*\tau)) \tag{2.1}$$

with $i = 1, \ldots, K$, where $K = [N - (m-1)*\tau]$, m is the embedding dimension and τ is the embedding lag. The chose of the appropriate embedding parameters is extremely important. The next two sections describe briefly some approaches used to estimate these parameters.

ESTIMATION OF MINIMUM EMBEDDING DIMENSION (m)

For the estimation of the smallest sufficient embedding dimension many algorithms were proposed. One of the most used method is the 'false nearest-neighbors' algorithm proposed by Kennel *et al.* in 1992 [10]. The algorithm identifies the number of 'false nearest neighbors', points that appear to be nearest neighbors of every point in the phase space due to a small embedding dimension value. An appropriate embedding dimension can be determined by examining how the number of false neighbors changes as a function of dimension, i.e., when the number of false nearest neighbors drops to zero, we have embedded the time series into proper dimensional space. The limitation of this method relies on the subjective definition of false neighbor [11]. To overcame this limitation Cao proposed new method [11]. After reconstruct the vectors and the time delay embedding, similar to the idea of the false neighbor method, the quantity $a(i,m)$ is defined as:

$$a(i,m) = \frac{\|X_{m+1}(i) - X_{m+1}(n(i,m))\|}{\|X_m(i) - X_m(n(i,m))\|} \tag{2.2}$$

with $i = 1, \ldots, N - m\tau$, where $\|.\|$ is a measurement of Euclidean distance defined as:

$$\|X_m(k) - X_m(l)\| = max_{0 \leq j \leq m-1} |x(k + j\tau) - x(l + j\tau)| \tag{2.3}$$

$n(i,m)$ is an integer in the range $1 \leq n(i,m) \leq N - m\tau$ such that $X_m(n(i,m))$ is the nearest neighbor of $X_m(i)$ in the m dimensional reconstructed phase space. A quantity $E(m)$, which is the mean value of all $a(i,m)$ is computed as

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a(i,m) \tag{2.4}$$

This averaging process removes the subjectivity involved with fixing threshold and tolerances of the false nearest neighbor method. To investigate its variation from m to $m+1$, E1(m) is define as:

$$E1(m) = \frac{E(m+1)}{E(m)} \tag{2.5}$$

It is found that $E1(m)$ stops changing when it is greater than some value m_o if the series comes from an attractor; then $m_o + 1$ is the minimum embedding dimension. However, in practice it is difficult to resolve whether the $E1(m)$ is slowly increasing or has stopped changing if m is sufficiently large. To solve this problem, another quantity is determined that is useful in distinguishing deterministic signals from stochastic signals

$$E^*(m) = \frac{1}{N - m} \sum_{i=1}^{N-m} |x(i + m\tau) - x(n(i,m) + m\tau)| \tag{2.6}$$

and its variation from m to $m+1$ as

$$E2(m) = \frac{E^*(m+1)}{E^*(m)} \tag{2.7}$$

For random data, future values are independent of past values and $E2(m)$ will be equal to 1 for any m, whereas for deterministic signals $E2(m)$ is related to m, so it cannot be a constant.

TIME DELAY EMBEDDING ESTIMATION (τ)

Various approaches have been proposed for estimation of an appropriate time delay. The most used two are the autocorrelation function and the average mutual information function (AMI) [12]. In the former, the value for which the autocorrelation function ($C(\tau)$) first passes through zero (τ) is searched.

$$C(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} (x(i) - \bar{x})(x(i+\tau) - \bar{x}) \tag{2.8}$$

In the latter, let p_i be the probability to find a time series value in the ith interval of the partition, let $p_{ij}(\tau)$ be the joint probability to find a time series value in the ith interval and a time series value in the jth interval after a time τ, i.e., the probability of transition in τ time from the ith to the jth interval. The average mutual information function is

$$S(\tau) = -\sum_{ij} p_{ij}(\tau) ln \frac{p_{ij}(\tau)}{p_i p_j} \tag{2.9}$$

The value τ chosen is the first one that minimizes the quantity $S(\tau)$.

The AMI seems to be preferred in nonlinear time series analysis since it measures a general dependence of two variables while the autocorrelation function looks for the linear independence.

2.2 Poincaré Plot

The Poincaré plot (PP), also known as a return or delay map, allows assessing the heartbeat dynamics based on a simplified phase-space embedding. The PP is a two-dimensional graphic (scatter plot) in which each RR interval, $x(i)$, is plotted as a function of the previous RR interval, $x(i-1)$. The PP analysis is an emerging quantitative-visual technique, whereby the shape of the plot provides summary information on the behavior of the heart [13, 14]. For a healthy heart, the cloud of points presents a comet shape oriented along the line of identity, the cardiac heart failure dynamics are characterized by a stretched elliptical shaped cloud of pointst also along line of identity, in the AF case the cloud of points present a more circular shape, similar to what happens with white noise time series. (See Figure 2.3)

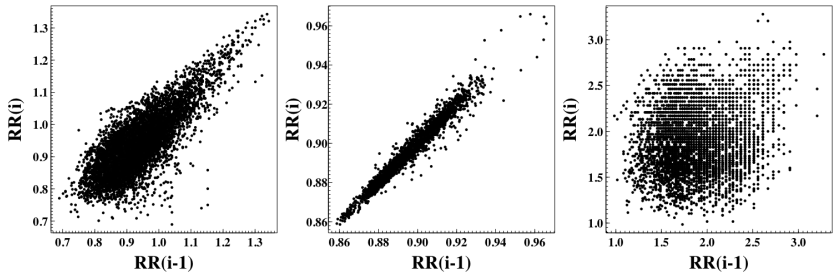

Figure 2.3: Poincaré plot for each of the time series presented in Figure 2.2. The left panel represents a normal sinus rhythm, the middle panel data from a CHF patient and the right panel the AF case. Note that the axis value are different in the three cases.

A number of techniques were developed attempting to quantitatively summarize the plot's geometric appearance. The geometrical descriptors, as the ellipse fitting technique, the histogram techniques and

correlation coefficient, are the most popular in the clinical and HRV literature [15]. The dispersion of points along the line-of-identity is thought to indicate the level of long-term variability and measured by the standard deviation denoted by $SD2$ [16]. On the other hand, the dispersion of points along the perpendicular to the line of identity reflects the level of short-term variability is measured by $SD1$ [14, 17, 18].

The standard deviation of the RR intervals, denoted by SDRR, is often employed as a measure of overall HRV. It is defined as the square root of the variance of the RR intervals, where the mean RR interval is here denoted by E[X].

$$SDRR = \sqrt{E[X^2] - E[X]^2} \tag{2.10}$$

The standard deviation of the successive differences of the RR intervals, denoted by $SDSD$, is an important measure of short-term HRV. It is defined as the square root of the variance of the sequence $\Delta X(i) = x(i) - x(i+1)$

$$SDSD = \sqrt{E[X^2] - E[\Delta X]^2)} \tag{2.11}$$

Note that $E[\Delta X] = 0$ for stationary intervals. This means that $SDSD$ is equivalent to the root-mean-square of the successive differences, denoted $RMSSD$.

The geometric indices obtained by fitting an ellipse to the Poincaré plot are dependent of the standard time domain HRV indices. The width of the Poincaré plot is a linear scaling of the most common statistic used to measure short-term HRV, the $SDSD$ index. In fact, the width of the Poincaré plot correlates with other measures of short-term HRV [14, 19].

$$SD1^2 = Var(\frac{1}{\sqrt{2}}X(i) - \frac{1}{\sqrt{2}}X(i+1)) = \frac{1}{2}Var(\Delta X) = \frac{1}{2}SDSD^2 \tag{2.12}$$

$$SD2^2 = 2SDRR^2 - \frac{1}{2}SDSD^2 \tag{2.13}$$

Two simple generalizations of the Poincaré plot, lagged Poincaré plots and higher-order Poincaré plots, can also be encountered in the literature. In lagged Poincaré plots $x(i)$ is plotting against $x(i+c)$, where c is some small positive value. In general, the plot is still clustered around the line-of-identity. However, the length and width of the plot are altered as the lag is increased. Considering the standard Poincaré plot to be of first order, the second order Poincaré plot is a 3D scatter-plot of the triples $(x(i), x(i+1), x(i+2))$. There are three orthogonal views of the shape of this plot, resulting in 2D projections onto each of the coordinate planes $(x(i), x(i+1))$, $(x(i+1), x(i+2))$ and $(x(i), x(i+2))$. The first two views are equivalent to the standard Poincaré plot and the third is the lag-2 Poincaré plot. This idea can be extended into higher dimensions, with the projections of the plot onto coordinate planes being lagged Poincaré plots. So, an order c Poincaré plot is geometrically described by the set of lagged Poincaré plots up to and including lag c [15].

The Poincaré plot is a powerful tool for graphically representing the summary statistics but also for beat-to-beat structure. Non-geometric techniques, such as scanning parameters [16, 20–23] and image distribution measures [24], are likely to be measuring independent, nonlinear information on the intervals. However, they are not nearly as popular as the “linear” Poincaré plot measures in the literature [15].

2.3 Recurrence Plot Analysis

The recurrence plots (RPs) method was introduced, in 1987 by Eckmann *et al.* [25], to visualize the recurrence of dynamical systems in a phase space. In this method a $K \times K$ recurrence matrix is constructed where the matrix whose elements $RM_{i,j}$ are defined as:

$$RM_{i,j} = \Theta(r - \|X_m^\tau(i) - X_m^\tau(j)\|) \tag{2.14}$$

with $i, j = 1, \ldots, K$, where r is a threshold distance, $\|.\|$ is the Euclidean distance, and $\Theta(.)$ is the Heaviside function. This means that, if two phase space vectors $X_m^\tau(i)$ and $X_m^\tau(j)$ are sufficiently close together then $RM_{i,j} = 1$; otherwise is $RM_{i,j} = 0$. The RP is the representation of the matrix RM as a black (for ones) and white (for zeros) image.

A crucial parameter of a RP is the threshold distance, r. If r is too small, there may be almost no recurrence points and we cannot learn anything about the recurrence structure of the underlying system. On the other hand, if r is chosen too large, almost every point is a 'neighbor' of every other point, which leads to a lot of artifacts [26]. Several criteria for the choice of the distance threshold r have been proposed: a few percent of the maximum phase space diameter [27], a value which should not exceed 10% of the mean or the maximum phase space diameter [28, 29], or a value that ensures a recurrence point density of approximately 1% [30]; another approach is to choose r according to the recurrence point density of the RP by seeking a scaling region in the recurrence point density [30] or to take into account that a measurement of a process is a composition of the real signal and some observational noise with standard deviation σ [31]. One of most used approach uses a fixed number of neighbors for every point of the trajectory, called the fixed amount of nearest neighbors (FAN) [25]. In this approach r_i changes for each state $X_m^\tau(i)$ to ensure that all columns of the RP have the same recurrence density. Using this neighborhood criterion, r_i can be adjusted in such a way that the recurrence rate (REC) has a fixed predetermined value [26].

To quantify different properties of the temporal evolution of a system, such as stability, complexity, and the occurrence of epochs of chaos vs. order in the behavior of the system several measures were developed [28, 32–34] and are known as recurrence quantification analysis (RQA). These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP. The most important structures for RQA are diagonal and vertical lines, briefly described in the following two subsections. Diagonals reflect the repetitive occurrence of similar sequences of states in the system dynamics and express the similarity of system behavior in two distinct time sequences. Verticals result from a persistence of one state during some time interval.

THE PARAMETERS EXTRACTED FROM THE RM BASED ON DIAGONAL LINES

Here are presented five parameters (DET, L, L_{max}, DIV and $Entropy$) used to characterize the information contained in the diagonal line. Considering $P(l)$ as the histogram of the length of diagonal lines.

- Determinism (DET) is the ratio of recurrence points that form a diagonal structure

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{2.15}$$

- Average diagonal line length (L)

$$L = \frac{\sum l = l_m in^N lP(l)}{\sum_{l=l_{min}}^{N} P(l)} \tag{2.16}$$

  Note that the threshold l_{min} excludes the diagonal lines that are formed by the tangential motion of the phase space trajectory. The choice of l_{min} has to take into account that the histogram $P(l)$ can become sparse if l_{min} is too large, and, thus, the reliability of DET decreases.

- Maximal length of a diagonal (L_{max}), or its inverse, the divergence (DIV),

$$L_{max} = max\ l_i{}_{i=1}^{N_l} \text{ and } DIV = \frac{1}{L_{max}} \tag{2.17}$$

  where $N_l = \sum_{l \geq l_{min}} P(l)$ is the total number of diagonal lines.

- *Entropy* refers to the Shannon entropy of the probability $p(l) = \frac{P(l)}{N_l}$ to find a diagonal line of length l.

$$Entropy = -\sum_{l=l_{min}}^{N} p(l) \cdot ln\ (p(l)) \tag{2.18}$$

  The entropy reflects the complexity of the RP in respect of the diagonal lines, e.g. for uncorrelated noise the value of entropy is rather small, indicating its low complexity.

THE PARAMETERS EXTRACTED FROM THE RM BASED ON VERTICAL LINES

Considering that the total number of vertical lines of the length v is given by $P(v)$, the measures based on vertical lines are:

- Laminarity (LAM), the ratio between the recurrence points forming the vertical structures and the entire set of recurrence points, is analogous to DET applied to vertical lines:

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N_R} vP(v)} \tag{2.19}$$

  LAM will decrease if the RP consists of more single recurrence points than vertical structures.

- The average vertical line length (trapping time – TT) estimates the mean time that the system will abide at a specific state or how long the state will be trapped.

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N_R} P(v)} \tag{2.20}$$

The maximal vertical line of the matrix (V_{max}), analogously to the standard measure L_{max} (N_v is the absolute number of vertical lines).

$$V_{max} = max\ (\{v_l\}_{l=1}^{N_v}) \tag{2.21}$$

In contrast to the RQA measures based on diagonal lines, these measures are able to find chaos–chaos transitions [33]. Hence, they allow for the investigation of intermittency, even for rather short and non-stationary data series. Furthermore, since for periodic dynamics the measures quantifying vertical structures are zero, chaos–order transitions can also be identified.

2.4 Fractal Dimension

A fractal dimension (FD) is a statistical index of how detail a pattern changes with the scale at which it is measured. The FD emerges to provide a measure of how much space an object occupies between Euclidean dimensions. The FD of a waveform represents a powerful tool for transient detection. The higher the FD, the more irregular the signal is, i.e., the more self-similar the signal will be.

From the several algorithms available to calculate the FD of a time series, the three most common are the box-counting dimension [35] and the algorithms proposed by Katz [36] and Higuchi [37, 38].

Box Counting Dimension

The main idea of the box counting method is to analyze complex patterns by breaking signal into smaller and smaller pieces, typically "box"-shaped, and analyzing the pieces at each scale. The minimum number of elements of a given size (ε), necessary to fully cover the curve (S), of dimension d is counted (N_ε).

$$N_\varepsilon(S) \sim \frac{1}{\varepsilon^d} \text{ as } \varepsilon \to 0 \tag{2.22}$$

As the size of the element approaches zero, the total area covered by the area elements will converge to the measure of the curve. This way, the FD_B can be estimated via a box-counting algorithm as proposed by Barabasi and Stanley [35] as follows:

$$FD_B = -lim_{\varepsilon \to 0}\ \frac{ln\ N_\varepsilon(S)}{ln\ \varepsilon} \tag{2.23}$$

Katz

The fractal dimension (FD_K) of the waveform representing the time series is estimated using Katz method [36] as follows:

$$FD_K = \frac{log\ L}{log\ d} \tag{2.24}$$

where L is the total length of the curve calculated as the sum of the distance between the successive data points and d is the diameter or planar extent of the curve, estimated as the distance between the first point and the point in the sequence that gives the farthest distance. For the signals that do not cross themselves it can be expressed as $d = max\ (dist(x(1), x(i))), i = 2, \ldots, N$, where $dist(x(i), x(j))$ is the distance between the ith and jth points on the curve.

Higuchi

The Higuchi's method [37, 38] it is a very efficient algorithm to calculate the FD of a curve and it has been increasingly used for the analysis of time series. For a time series expressed by $x(i), i = 1, \ldots, N$, new series X_r^m are obtained as follow:

$$X_r^m = (x(m), x(m+r), x(m+2r), \ldots, x(m + \left[\frac{(N-m)}{r}\right] r)), m = 1, \ldots, r \tag{2.25}$$

where [.] denotes the Gauss notation, m and r are integer that indicate the initial time and the time interval, respectively.

The length of the new series X_r^m is defined as:

$$L_m(r) = \left\{ \left(\sum_{i=1}^{[\frac{N-m}{r}]} |x(m-ir) - x(m+(i-1))| \right) \times \frac{N-1}{[\frac{N-m}{r}]r} \right\} \frac{1}{r} \tag{2.26}$$

The length of the $L(r)$ for the time interval r is obtained by averaging all the subseries lengths $L_m(r)$ that have been obtained for a given r value.

If $L(r)$ is proportional to r^{-D}, the curve describing the shape is fractal-like with the dimension D. Thus, if $L(r)$ is plotted against r, on a double logarithmic scale $(ln\ \frac{1}{r}, ln\ L(r))$, the points should fall on a straight line with a slope equal to $-D$. The coefficient of linear regression of the plot is taken as an estimate of the fractal dimension of the epoch. Applying the above relation implies the proper choice of a maximum value of r for which the relationship $L(r) \propto r^{-D}$ is approximately linear.

2.5 Detrended Fluctuation Analysis

Detrended fluctuation analysis (DFA) quantifies intrinsic fractal-like (short and long-range) correlation properties of dynamic systems [39]. This technique is a modification of root-mean-square analysis of random walks applied to nonstationary signals [40].

First, the time series (of length N) is integrated. Then, the integrated time series is divided into N_n windows of equal length n. In each window of length n, a least-squares line is fitted to the data. The y-coordinate of the straight-line segments are denoted by $y_n(k)$. Next, the integrated time series is detrended, $y_n(k)$, in each window. The root-mean-square fluctuation of this integrated and detrended series is calculated using the equation:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2} \tag{2.27}$$

This computation is repeated over all time scales (box sizes) to characterize the relationship between $F(n)$, the average fluctuation, and the box size, n. Typically, $F(n)$ increases with window size according to $F(n) \propto n^{\alpha}$. The α exponent can be viewed as an indicator of the "roughness" of the original time series: the larger the value of α, the smoother the time series.

If $\alpha \simeq 0.5$, the time series represents uncorrelated randomness (white noise);

if $\alpha \simeq 1$ (1/f noise), the time series has long-range correlations and exhibits scale-invariant properties;

if $\alpha \simeq 1.5$, the time series represents a random walk (Brownian motion).

In this context, 1/f-noise can be interpreted as a compromise or "trade-off" between the complete unpredictability of white noise and the much smoother "landscape" of Brownian noise.

Usually the DFA method involves the estimation of a short-term fractal scaling exponent α_1 and a long-term scaling exponent α_2.

DFA as such is a monofractal method, but multifractal analysis also exists [166]. In the early 1990s an improved multifractal formalism, the wavelet transform modulus maxima (WTMM) method [41, 41–44], has been developed. The multifractal analysis describes signals that are more complex than those fully characterized by a monofractal model, but requires many local and theoretically infinite exponents to fully characterize their scaling properties. The multifractal DFA (MF-DFA) [45, 46] consist in five steps that do not require the modulus maxima procedure. The first two steps are identical to the monofractal DFA procedure. Since the length N of the series is often not a multiple of the considered time scale n, the same procedure is repeated starting from the opposite end. Thereby, the integrated time series, $y_n(k)$ - the fitting polynomial in segment k, was detrend in the $2N_n$ segments obtained by determination of χ^2-functions (variances), for each k from 1 to N_n:

$$F_k^n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y[(k-1)n+i] - y_k(i)]^2} \tag{2.28}$$

while for k from $N_n + 1$ to $2N_n$ the corresponding variances are defined as:

$$F_k^n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y[N-(k-N_n)n+i] - y_k(i)]^2} \tag{2.29}$$

The fitting procedure can be linear, quadratic, cubic, or higher order polynomials (MF-DFAm - the mth order of the MF-DFA) [39, 47, 48]. A comparison of the results for different orders of the MF-DFA allows one to estimate the order of the polynomial segment trends in the time series [45, 48].

The fourth step is the average over all segments to obtain the qth order (the index variable q can take any real value) fluctuation function:

$$F_q(n) = \left[\frac{1}{2N_n}\sum_{k=1}^{2N_n}[F_k^{n\,2}]^{q/2}\right]^{1/q} \tag{2.30}$$

For $q = 2$, the standard DFA procedure is retrieved. We are interested how this q-dependent fluctuation function depends on the time scale n for different values of q hence, stages 2 till 4 must be repeated for several time scales n.

Determine the scaling behavior of the fluctuation functions by analyzing log–log plots $F_q(n)$ versus n for each value of q. If such a scaling exists $ln\ F_q(n)$ will depend linearly on $ln\ n$, with $h(q)$ as the slope. For stationary time series $h(2)$ is identical with the Hurst exponent, HE (see section below) and $h(q)$ is said to be the generalized HE. A monofractal time series is characterized by unique $h(q)$ for all values of q. The generalized HE, $h(q)$, of MF-DFA is related to the classical scaling exponent $\phi(q)$ through the relation

$$\phi(q) = qh(q) - 1. \tag{2.31}$$

A monofractal series with long-range correlation is characterized by linearly dependent q order exponent $\phi(q)$ with a single HE. Multifractal signal have multiple HE and $\phi(q)$ depends non-linearly on q [49].

2.6 Hurst Exponent

Initially defined by Harold Edwin Hurst [50, 51] to develop a law for regularities of the Nile water level, the Hurst exponent (HE) is a dimensionless estimator used to evaluate the self-similarity and the long-range correlation properties of time series. Unlike others nonlinear dynamic features, the HE does not involve state space reconstruction. There is a straightforward relationship between the HE and the fractal dimension FD, given by $FD = E + 1 - HE$, where E is the Euclidean dimension, which for time series is 1 obtaining there relationship $FD = 2 - HE$ [52].

The oldest description of the HE is defined in terms of the asymptotic behavior of the rescaled range (a statistical measure of the variability of a time series) as a function of the time span of a time series as follows:

$$E\left[\frac{R(N)}{S(N)}\right] = CN^H \text{ as } N \to \infty \tag{2.32}$$

where $S(N)$ is the standard deviation, C an arbitrary constant and the range $R(N)$ is defined as the difference between the maximum and the minimum values of a given time series.

There are many algorithms to estimate the HE parameter. The most immediate one is derived from the definition. First, the time series of length N is subdivided into segments of length T. Then the ratio R/S is computed for each segment and the average, for all segment, is calculated. These steps are repeated for several values of T. The HE can be estimated as the slope of the regression line produced by the log-log graph. The HE can also be computed using the periodogram (P_{xx}), which is an approximation of the power spectral density (PSD) [53]. The Bartlett's method [54] (the method of averaged periodograms [55]) or it modification, the Welch method [56] are the most used method for, in practice, estimating power spectra density. As in the previous method the time series of length N is divided in segments of length T. In the Bartlett's method for each segment, the periodogram is computed using the discrete Fourier transform, then the squared magnitude of the result is calculated and divided by T. Average the result of the periodograms above for the data segments. The averaging reduces the variance, compared to the original N point data segment. [54, 57] The Welch method differs since it allows the segments of the time series of each periodogram to overlap. Once the PSD is estimated it is possible to determine β parameter as the slope of the straight line fitted using least squares.

$$log\ ((P_{xx}(f)) = -\beta log\ (f) + C \tag{2.33}$$

And, the computation of HE index is straightforwardly using the relation:

$$\beta = 1 + 2HE \tag{2.34}$$

Other increasingly used approach is related with the DFA method (described in the previous section). The α exponent is related to the HE parameter by [58]: $\alpha = 1 + HE$. The advantages of DFA over conventional methods (periodogram and R/S method) are that of permit the detection of long-range correlations in time series with non-stationarities, and also avoids the spurious detection of apparent

long-range correlations that are an artifact of non-stationarity [7, 40].

The HE may range between 0 and 1 and can indicate:

$0 < HE < 0.5$, that time series has long-range anti-correlations;

$HE = 0.5$, that there is no correlation in the time series;

$0.5 < HE < 1$, that there is long-range correlations in the time series.

$HE = 1$, that the time series is defined self-similar, i.e., it has a perfect correlation between increments. However, the literature on HR variability conventionally uses the term self-similarity even if HE differs from [59].

2.7 Correlation Dimension

To analyze the complexity of a system, usually a transition from the time domain to the phase space is needed. The correlation dimension (CD), one of the most widely used measures of fractal dimension, can be considered as a measure for the number of independent variables needed to define the total system in phase space [60].

For each vector $X_m^1(i)$, the relative number of vectors $X_m^1(j)$ for which $d[X_m^1(i), X_m^1(j)] \leq r$, where r is referred as a threshold tolerance value is computed as:

$$C_m^r(i) = \frac{\text{number of } d[X_m^1(i), X_m^1(j)] \leq r}{N - m + 1} \quad \forall j \tag{2.35}$$

The distance function is defined as:

$$d[X_m^1(i), X_m^1(j)] = \sqrt{\sum_{k=1}^{m} (X_m^1(i,k) - X_m^1(j,k)^2} \tag{2.36}$$

where $X_m^1(i,k)$ and $X_m^1(j,k)$ refer to the kth element of the series $X_m^1(i)$ and $X_m^1(j)$, respectivetly. The probability that two chosen points are close to each other with distance smaller than r, is computed by averaging $C_m^r(i)$ over i:

$$C_m^r = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} C_m^r(i) \tag{2.37}$$

The C_m^r index is computed for increasing values of the embedding dimensions m (usually, here, the embedding dimension varies between 2 and 30 [61]) and the slopes of the log-log plot are determined, obtaining a sequence of $d(m)$. As m increase, $d(m)$ tends to a constant value of saturation, which is the CD value [62]. Grassberger and Procaccia [63] showed that correlation dimension (CD) can be obtained from:

$$CD = \lim_{r \to 0} \lim_{N \to \infty} \frac{log\ C_m^r}{log\ (r)} \tag{2.38}$$

In practice this limit value is approximated by the slope of the regression curve $(log\ (r), log\ C_m^r)$. [64]

Other approach to estimate the CD value is the Levenberg-Marquardt method [65]. The exponential model for the $d(m)$ values is $d(m) = CD(1 - r^{-km})$, where CD and k are the parameters of the curve, where the former represents the asymptotic value of the curve when $m \to \infty$, and the latter is the exponential constant.

2.8 Entropies

Shannon introduced the first notion of entropy (Shannon entropy - SE) to measure how the information within a signal can be quantified with absolute precision as the amount of unexpected data contained in the message [8, 66].

The Renyi entropy, a generalization of the Shannon entropy, is a family of functions of order q (R_q) defined as:

$$R_q = \frac{1}{1-q} ln \sum_i p(x(i))^q \tag{2.39}$$

where $p(x(i))$ is the probability of $X = x(i)$. The particular case when $q = 1$, the Shannon entropy is obtained:

$$SE = -\sum_i p(x(i)) \cdot log \ \ (p(x(i))) \tag{2.40}$$

The conditional entropy [67] (CE) assesses the amount of information carried by the current RR sample when $m-1$ past samples of RR are known. The CE represents the difficulty in predicting the future values based on past values of the same time series. When the future values of RR are completely predictable, given past values, the CE value is 0. If the past values of RR are not helpful to reduce the uncertainty associated with future RR values the value of CE is equal to SE. In the approach introduced by Porta [67] the RR time series is recoded. The RR intervals are sorted into ξ equally spaced bins and the values inside each bin are substituted with an integer ranging from 0 to $\xi - 1$ coding the specific bin, obtaining a new time series $X^\xi = \{X^\xi(i), i = 1, \ldots, N\}$. The patterns are constructed using the technique of the delayed coordinates as $X^\xi_m(i) = (X^\xi(i), X^\xi(i-1), \ldots, X^\xi(i-m+1))$. The conditional entropy (CE) of RR is defined as:

$$CE^m = -\sum_{i=1}^{N-m+1} \left[p\left(X^\xi_m(i)\right) \cdot \sum_{i=2}^{N-m+1} \left[p\left(X^\xi(i)|X^\xi_{m-1}(i-1)\right) \cdot log\left(X^\xi(i)|X^\xi_{m-1}(i-1)\right)\right]\right] \tag{2.41}$$

where $p(X^\xi(i)|X^\xi_{m-1}(i-1))$ is the conditional probability of $X^\xi(i)$ given previous $m-1$ samples.

Since the percentage of patterns found only once grows monotonically towards 100% with m, CE always decreases toward 0 with m independently of the type of RR dynamics. In order to prevent the artificial decrease of the information carried by RR given m past samples, solely related to the shortness of the data sequence, the corrected conditional entropy (CCE) is defined as [67–69]:

$$CCE(m) = CE(m) + SE(1) \cdot perc(X^\xi_m). \tag{2.42}$$

where $X^\xi_m = \{X^\xi_m(i), i = 1, \ldots, N-m+1\}$ are the series of the patterns that can be constructed from X^ξ, $perc(X^\xi_m)$ represents the fraction of patterns found only once in X^ξ_m with $0 \leq perc(X^\xi_m) \leq 1$ and

$$SE(m) = -\sum_i p(X^\xi_m(i)) \cdot log\left(p(X^\xi_m(i))\right) \tag{2.43}$$

The $CCE(m)$ decreases towards 0 only in case that RR is perfectly predictable given past RR values,

it remains constant when past RR values are not helpful to predict future RR, and it exhibits a minimum when past RR values are only partially helpful to predict RR. The minimum of CCE, CCE_{min} represents the minimum amount of information carried by RR given its own past values: the larger this value, the larger the amount of information carried by RR, the smaller the predictability of RR based on its own past values.

The Approximate Entropy ($ApEn$), proposed by Pincus [70], exhibits a good performance in the characterization of randomness even when the data sequences are not very long. In order to calculate the $ApEn$ the new series of vector of length m - embedding dimension - are constructed X_m^1. Similar to CD, for each vector $X_m^1(i)$, the value $C_m^r(i)$, where r is referred as a tolerance value, is computed as:

$$C_m^r(i) = \frac{\text{number of } d[X_m^1(i), X_m^1(i)] \leq r}{N - m + 1} \quad \forall j \tag{2.44}$$

Here the distance function used is defined as:

$$d[X_m^1(i), X_m^1(j)] = max_{k=1,\dots,m}|x(i + k - 1) - x(j + k - 1)| \tag{2.45}$$

Next, the average of the natural logarithm of $C_m^r(i)$ is computed for all i:

$$\Phi_m^r = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} ln(C_m^r(i)) \tag{2.46}$$

Since in practice N is a finite number, the statistical estimate is computed as:

$$ApEn(m, r) = \begin{cases} \Phi_m^r - \Phi_{m+1}^r & \text{for } m > 0 \\ -\Phi_1^r & \text{for } m = 0 \end{cases}$$

The choice of the embedding dimension parameter m was already discussed in the begging of this chapter. In the particular case of the ApEn, the most common value is $m = 2$. Regarding parameter r several approaches are used. Pincus [70, 71] recommends values between the 10% and 25% of the standard deviation of the data, hence obtaining a scale invariant measurement. The approach [72, 73] of choose a fixed r value was also used with success. However, the values of entropy in these case are usually highly correlated with the time series standard deviation. Lu et all [74] showed that $ApEn$ values vary significantly even within the defined range of r values and presented a new method for automatic selection of r that corresponds to the maximum $ApEn$ value.

The Sample Entropy ($SampEn$) was introduced, with the same objective as $ApEn$, to evaluate biological time series, particularly the heart rate time series. The authors highlighted two draw-backs in $ApEn$ properties, stating that 'First, ApEn is heavily dependent on the record length and uniformly lower than expected for short records. Second, it lacks relative consistency. That is, if ApEn of one dataset is higher than that of another, it should, but does not remain higher for all conditions tested' [75]. In order to overcome these limitations, the group proposed a new family of statistics, $SampEn(m, r)$, which, among other differences, eliminates self-matches.

For the $SampEn$ [75] calculation the same parameters defined for the $ApEn$, m and r are required. Considering A as the number of vector pairs of length $m+1$ having $d[X_m^1(i), X_m^1(j)] \leq r$, with $i \neq j$ and

B as total number of template matches of length m. The SampEn is defined as:

$$SampEn = -ln\ \frac{A}{B} \tag{2.47}$$

Traditional entropy-based algorithms quantify the regularity of a time series. The multiscale entropy [76, 77] (MSE) approach, is inspired on Zhang's proposal [78], considers the information of a system's dynamics on different time scales. The multiscale method can be separate in two parts. The first one is the construction of the time series scales: using the original signal, a scale, s, is created from the original time series, through a coarse-graining procedure, i.e, replacing s non-overlapping points by their average. The second step concerns the calculation the value of entropy for each time series scale. The most used entropies in this approach are the $ApEn$ and the $SampEn$. The information of the different time scales is clustered in the complexity index (CI) defined as the area under the MSE curve obtained by plotting the entropy value as a function of scale [76, 77]. This approach extend the entropy concept of regularity to a concept of fractality.

2.9 Lyapunov exponent

The Lyapunov exponent (LE) is a measure of the system dependence to the initial conditions but also quantifies the predictability of the system [79]. A system embedded in an m-dimensional phase space has m Lyapunov exponents. The value of LE increases, corresponding to lower predictability, as the degree of chaos becomes higher; a positive Lyapunov exponent is a strong indicator of chaos [80–82] therefore, the computation only of the largest Lyapunov exponent (LLE) is sufficient for assess chaos. The LLE is most commonly used in nonlinear analysis of physiological signals [83–85].

There are many algorithms available to estimate both LE and LLE [82, 86–91]. The method proposed by Rosenstien et al [88] is one of the most used since it is robust against data length. The algorithm looks for the nearest neighbor of each point, on the trajectory. The distance between two neighboring points at instant $n = 0$ is defined by

$$d_i(0) = min_{X_m^\tau(j)}\ \ \parallel X_m^\tau(j) - X_m^\tau(i) \parallel \tag{2.48}$$

where $\parallel \cdot \parallel$ is the Euclidean norm. This algorithm imposes constraint that nearest neighbors are temporally separated at least by mean period of the time series. The LLE is then estimated as the mean rate of separation of nearest neighbors, i.e., we can write

$$d_j(i) \approx C_j\, e^{\lambda_1 i(\Delta t)} \tag{2.49}$$

where C_j is the initial separation. Taking logarithm on both sides we obtain

$$ln\ (d_j(i)) = ln\ \ C_j + \lambda_1 i(\Delta t) \tag{2.50}$$

it represents a set of approximately parallel lines, where the slope is roughly proportional to the LLE. In practice, the LE is easily and accurately estimated using a least-squares Δt to the "average" line defined

by

$$y(n) = \frac{1}{\Delta t}\langle ln\ d_i(n)\rangle \tag{2.51}$$

where $\langle\cdot\rangle$ denotes the average over all values of i. This last averaging step is the main feature that allows an accurate evaluation of λ even when we have short and noisy data set.

Another widely used method was proposed by Wolf et al. [82] in 1985, where the calculation of the LLE is based on the average exponential rate of divergence or convergence of trajectories which are very close in phase space. Contrary to the Rosenstien method, in the Wolf's method a single nearest neighbor is followed and repeatedly replaced when its separation from the initial trajectory grows beyond a certain threshold. However, this strategy do not take advantage of all the available data [88]. Another disadvantage of Wolf's method is the requirement of an appropriate selection of parameters, not only the time delay and embedding dimension, but also a maximum and minimum scale parameters, an angular size and trajectory evolution time [82, 92].

2.10 Numerical Noise Titration

The Numerical Noise Titration (NNT) [93, 94] method provides a robust numerical test of chaos and a relative measure of chaotic intensity, even in the presence of significant noise contamination. Comparing with the Lyapunov exponent (LE), described above, the LE fails to specifically distinguish chaos from noise and, unless the data series used are longish and free of noise, cannot detect chaos reliably.

The NNT method can be divided into 4 sections: modeling, nonlinear detection (NLD), numerical noise titration and decision tool.

1. Modeling

   In the noise titration method every data segment are first analyzed by using a discrete Volterra autoregressive to calculate the predicted time series y_i^{calc}. Briefly, the Volterra–Wiener algorithm [95] produces a family of polynomial (linear and nonlinear) autoregressive models with varying memory (κ) and dynamical order (d) (i.e. $d = 1$ for linear and $d > 1$ for a nonlinear model), optimally fitted to predict the data.

$$\begin{aligned} y_i^{calc} &= a_0 + a_1 x(i-1) + a_2 x(i-2) + \cdots + a_\kappa x(i-\kappa) + a_{\kappa+1} x(i-1)^2 + \\ &\quad a_{\kappa+2} x(i-1)x(i-2) + \cdots + a_M x(i-\kappa)^d \\ &= \sum_{m=1}^{M-1} a_m z_m(i) \end{aligned} \tag{2.52}$$

   where $M = \frac{(\kappa+d)!}{\kappa!d!}$ is the total dimension and the coefficients a_m are recursively estimated using the Korenberg algorithm [96].

2. Nonlinear detection (NLD)

   The goodness of fit of a model (linear vs. nonlinear) is measured by the normalized residual sum of squared errors:

$$\varepsilon(\kappa, d)^2 = \frac{\sum_{i=1}^{N}(y_i^{calc}(\kappa, d) - x(i))^2}{\sum_{i=1}^{N}(x(i) - \overline{X})^2} \tag{2.53}$$

with $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x(i)$ and $\varepsilon(\kappa, d)^2$ represents a normalized variance of the error residuals. The best linear and nonlinear models is the model that minimizes the Akaike [97] information criterion:

$$C(r) = log\varepsilon(r) + \frac{r}{N} \tag{2.54}$$

where $r \in [1, M]$ is the number of polynomial terms of the truncated Volterra expansion from a certain pair (κ, d).

3. Numerical noise titration

   The NLD is used to measure the chaotic dynamics inherent in the RR series by means of numerical noise titration as follows:

   i If linear, then there is insufficient evidence for chaos.

   ii If nonlinear, it may be chaotic or non-chaotic. To discriminate these possibilities, add a small (<1% of signal power) amount of random white noise to the data and then apply NLD again to the noise corrupted data. If linear, the noise limit (NL) of the data is zero and the signal is non-chaotic.

   iii If nonlinearity is detected, increase the level of added noise and again apply NLD.

   iv Repeat the above step until nonlinearity can no longer be detected when the noise is too high (low signal-to-noise ratio). The maximum noise level (i.e. NL) that can be added to the data just before nonlinearity can no longer be detected, is directly related to the LE.

4. Decision tool.

   According to this numerical titration scheme, $NL > 0$ indicates the presence of chaos, and the value of NL gives an estimate of relative chaotic intensity. Conversely, if $NL = 0$, then the time series may be non-chaotic, but it is also possible that the chaotic component is already neutralized by the background noise. Therefore, the condition $NL > 0$ provides a simple sufficient test for chaos.

2.11 Symbolic Dynamics

The concept of symbolic dynamics goes back to Hadamard [98] and allows a simplified description of the dynamics of a system with a limited amount of symbols. For HRV analysis, the underlying theoretical concept is used in a rather pragmatic way. The main idea is to encode, according to some transformation rules, RR intervals and their changes into a few symbols of a certain alphabet. Subsequently, the dynamics of that symbol string are quantified, providing more global information regarding heart rate dynamics. Two techniques introduced, by Voss et al. (1996) [99] and Porta et al. (2001) [69], are the most used ones.

According to the symbolic dynamics approach described by Voss et al., the series of RR intervals are transformed into an alphabet of 4 symbols: 0, 1, 2, 3, depending on how much single RR intervals differ from the mean. The transformation rules proposed is presented in next equation, where μ is the mean RR intervals and α is a special scaling parameter (usually equal to 0.1). In this transformation, the symbols ‘‘0’’ and ‘‘2’’ indicate a small, ‘‘1’’ and ‘‘3’’ encode a large difference from the mean.

$$0 : \mu < x(i) \leq (1 + \alpha) \cdot \mu \tag{2.55}$$

$$1 : (1 + \alpha) \cdot \mu < x(i) < \infty \tag{2.56}$$

$$2 : (1 + \alpha) \cdot \mu < x(i) \leq \mu \tag{2.57}$$

$$3 : 0 < x(i) \leq (1 + \alpha) \cdot \mu \tag{2.58}$$

This method studies the probability distribution of words with three successive symbols from the alphabet to characterize symbol strings. In this way, one obtains 64 (4^3) different word types (bins). There are several parameters that characterize symbolic strings, here we described:

- forbidden words (FORBWORD) – the number of word types that occur with a probability less than 0.001; A high number of forbidden words reflect a reduced dynamic behavior in time series, and vice versa.
- Measures of complexity
  - Shannon entropy – Shannon entropy computed over all word types: a measure of word-type distribution complexity;
  - Renyi entropy 0.25 – Renyi entropy with a weighting coefficient of 0.25 computed over all word-types, predominately assessing the words with low probability;
  - Renyi entropy 4 – Renyi entropy with a weighting coefficient of 4 computed over all word-types, predominantly assessing words with high probabilities.
- wpsum - wpsum02 is measured as the percentage of words consisting of the symbols '0' and '2' only and the wpsum01 is the percentage of words containing only the symbols '1' and '2'. According to the meaning of the symbols, high values for wpsum02 indicate low complexity of heart rate time series, while high wpsum13 indicates higher complexity.

Voss *et al* also developed a modified approached of SymD for low or high variability. In this approach the RR intervals time series are transformed in a symbolic string using a simplified alphabet consisting only of symbols '0' or '1', where the symbol '0' stands for a difference between two successive beats lower than a special limit and the symbol '1' represents those cases where the difference exceeds this special limit. As time limits, 2, 5, 10, 20, 50, and 100 ms have been proposed, however, the limit 10 ms is the most used one since has been shown to be most useful according to hierarchical cluster and stepwise discriminant function analyses [100]. The low variability parameter is measured as the probability of occurrence of sequences containing six consecutive marks of "0" (*plvar*) whereas the high variability parameter (*phvar*) is calculated as the probability of sequences of six consecutive marks of "1". Taken together, in this model an increase of ''000000'' sequences, resulting in increased values of *plvar*, and a decrease in ''111111'' sequences, leading to reduced values of *phvar*, indicate reduced system complexity.

In the symbolic analysis according to the approach described in Porta et al (2001) [69], briefly, the RR series was first transformed into a sequence of symbols using a coarse graining approach based on a uniform quantization procedure. The full range of the series was spread over ξ symbols, with a resolution given by

$\frac{X_{max}-X_{min}}{\xi}$, where X_{max} and X_{min} are the maximum and the minimum of the series. After quantization, the RR series became a sequence $X^{\xi} = \{X^{\xi}(i), i = 1, \ldots, N\}$ of integer values ranging from 0 to $\xi-1$ where the X^{ξ} series are transformed into a ξ^m subseries $X^{\xi}_m(i) = (X^{\xi}(i), X^{\xi}(i-1), \ldots, X^{\xi}(i-m+1))$ with $i = m, \ldots, N$, using the technique of the delayed coordinates [101]. The values of m and ξ have to be small in order to avoid too large a number of possible patterns: for applications over short data sequences (250-300 samples), the best compromise has been shown to be $\xi = 6$ and $m = 3$ (216 possible patterns) [69]. To reduce the number of patterns without losing information, all the patterns were grouped without any loss into four families according to the number and types of variations from one symbol to the next one. The pattern families were as follows:

1. patterns with no variation (0 V, all the symbols were equal);

2. patterns with one variation (1V, two consecutive symbols were equal and the remaining one is different);

3. patterns with two like variations (2 LV, the three symbols formed an ascending or descending ramp);

4. patterns with two unlike variations (2 UV, the three symbols formed a peak or a valley).

The indexes 0V%, 1V%, 2 LV% and 2 UV% are computed as the percentages of occurrence (number of times that a pattern $X^{\xi}_m(i)$ belonging to a specific family was found and divided by $N-m+1$ (multiplied by 100) of these families. Since the sum of all symbolic parameters is equal to 100% (i.e., 0V% + 1V% + 2LV% + 2UV% =100%), 0V% and 2UV% can increase or decrease at the expense of 1V% and 2LV% [102].

2.12 Mutual Information

The mutual information (MI) measure is widely used to describe the mutual dependence of the two variables. In heart rate studies the MI reflects the probability of finding a given time-series value in one interval and the probability of finding the same value in another interval after the delay time τ. In the case of independence between the two MI is zero and greater otherwise.

Two new time series are constructed as $Y(i) = x(i+\tau)$ and $Z(i) = (x(i), x(i+\tau), \ldots, x(i+(m-1)\tau))$. The mutual information function (MIF) [103, 104] depend on four parameters: the length of the original time series, N; the embedding dimension, m; the distance parameter r and the delay parameter τ. From MIF we can simply get the (average) uncertainty on $x(i+\tau)$ remaining if $(x(i), ..., x(i+(m-1)\tau))$ is known.

$$MIF_{YZ}(m, r, N, \tau) = SE_Y - SE_{Y|Z} = SE_Y + SE_Z - SE_{YZ} = \sum_{ij} s_{ij} log_2 \frac{s_{ij}}{p_i q_j} \tag{2.59}$$

where s_{ij} represents the joint probability distribution and the corresponding marginal distributions are $p_i = \sum_j s_{ij}$ and $q_i = \sum_i s_{ij}$ of the y– and z–series, respectively.

Bibliography

[1] Lewis A. Lipsitz. Physiological complexity, aging, and the path to frailty. Science's SAGE KE, 2004(16):16, 2004.

[2] Edward Ott. Chaos in dynamical systems. Cambridge University Press, Cambridge, United Kingdom, 2002.

[3] Benoît B. Mandelbrot. The fractal geometry of nature. Macmillan, USA, 1983.

[4] Ary L. Goldberger, David R. Rigney, and Bruce J. West. Science in pictures: Chaos and fractals in human physiology. Scientific American, 262:42–49, 1990.

[5] Bruce J. West. Fractal physiology, volume 2. Oxford University Press, Oxford, United Kingdom, 1994.

[6] Ary L. Goldberger and Bruce J. West. Fractals in physiology and medicine. The Yale Journal of Biology and Medicine, 60(5):421, 1987.

[7] Chung-Kang Peng, Shlomo Havlin, H. Eugene Stanley, and Ary L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos: An Interdisciplinary Journal of Nonlinear Science, 5(1):82–87, 1995.

[8] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[9] Ary L. Goldberger, David R. Rigney, and Bruce J. West. Chaos and fractals in human physiology. Scientific American, 262(2):42–49, 1990.

[10] Matthew B. Kennel, Reggie Brown, and Henry D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A, 45(6):3403, 1992.

[11] Liangyue Cao. Practical method for determining the minimum embedding dimension of a scalar time series. Physica D: Nonlinear Phenomena, 110(1):43–50, 1997.

[12] Harry L. Fraser, Andrew M .and Swinney. Independent coordinates for strange attractors from mutual information. Physical Review A, 33(2):1134, 1986.

[13] Mary A. Woo, William G. Stevenson, Debra K. Moser, Robert B. Trelease, and Ronald M. Harper. Patterns of beat-to-beat heart rate variability in advanced heart failure. American Heart Journal, 123(3):704–710, 1992.

[14] Peter W. Kamen, Henry Krum, and Andrew M. Tonkin. Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. Clinical Science, 91(Pt 2):201–208, 1996.

[15] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. New insights into the relationship between poincare plot geometry and linear measures of heart rate variability. In Engineering in

Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, volume 1, pages 526–529. IEEE, 2001.

[16] Mikko P. Tulppo, Timo H. Makikallio, Timo E. Takala, Tapio Seppanen, and Heikki V. Huikuri. Quantitative beat-to-beat analysis of heart rate dynamics during exercise. American Journal of Physiology-Heart and Circulatory Physiology, 271(1):H244–H252, 1996.

[17] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Poincaré plot interpretation using a physiological model of hrv based on a network of oscillators. American Journal of Physiology-Heart and Circulatory Physiology, 283(5):H1873–H1886, 2002.

[18] Peter W. Kamen and Andrew M. Tonkin. Application of the poincaré plot to heart rate variability: a new measure of functional status in heart failure. Australian and New Zealand Journal of Medicine, 25(1):18–26, 1995.

[19] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? Biomedical Engineering, IEEE Transactions on, 48(11):1342–1347, 2001.

[20] Maurice E. Cohen, Donna L. Hudson, and Prakash C. Deedwania. Applying continuous chaotic modeling to cardiac signal analysis. Engineering in Medicine and Biology Magazine, IEEE, 15(5):97–102, 1996.

[21] Giovanni D'Addio, D. Acanfora, Gian D. Pinna, Roberto Maestri, Giuseppe Furgi, C. Picone, and Franco Rengo. Reproducibility of short-and long-term poincare plot parameters compared with frequency-domain hrv indexes in congestive heart failure. In Computers in Cardiology, pages 381–384. IEEE, 1998.

[22] Heikki V Huikuri, Tapio Seppänen, M. Juhani Koistinen, K.E. Juhani Airaksinen, M.J. Ikäheimo, Agustin Castellanos, and Robert J. Myerburg. Abnormalities in beat-to-beat dynamics of heart rate before the spontaneous onset of life-threatening ventricular tachyarrhythmias in patients with prior myocardial infarction. Circulation, 93(10):1836–1844, 1996.

[23] Hélène Otzenberger, Claude Gronfier, Chantal Simon, Anne Charloux, Jean Ehrhart, François Piquard, and Gabrielle Brandenberger. Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. American Journal of Physiology, 275:H946–H950, 1998.

[24] Katerina Hnatkova, Xavier Copie, Anne Staunton, and Marek Malik. Numeric processing of lorenz plots of rr intervals from long-term ecgs: comparison with time-domain measures of heart rate variability for risk stratification after myocardial infarction. Journal of Electrocardiology, 28:74–80, 1995.

[25] Jean-Pierre Eckmann, S. Oliffson Kamphorst, and David Ruelle. Recurrence plots of dynamical systems. Europhysics Letters, 4(9):973, 1987.

[26] Norbert Marwan, M. Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. Physics Reports, 438(5):237–329, 2007.

[27] Gabriel M. Mindlin and R. Gilmore. Topological analysis and synthesis of chaotic time series. Physica D: Nonlinear Phenomena, 58(1):229–242, 1992.

[28] Joseph P. Zbilut and Charles L. Webber Jr. Embeddings and delays as derived from quantification of recurrence plots. Physics Letters A, 171(3):199–203, 1992.

[29] Matthew Koebbe and Gottfried Mayer-Kress. Use of recurrence plots in the analysis of time-series data. In Santa fe Institute Studies In The Sicence Of Complexity - Proceedings, volume 12, pages 361–361. Citeseer, 1992.

[30] Joseph P. Zbilut, José-Manuel Zaldivar-Comenges, and Fernanda Strozzi. Recurrence quantification based liapunov exponents for monitoring divergence in experimental data. Physics Letters A, 297(3):173–181, 2002.

[31] Marco Thiel, M. Carmen Romano, Jürgen Kurths, Riccardo Meucci, Enrico Allaria, and F. Tito Arecchi. Influence of observational noise on the recurrence quantification analysis. Physica D: Nonlinear Phenomena, 171(3):138–152, 2002.

[32] Joseph P. Webber Jr, Charles L .and Zbilut. Recurrence quantification analysis of nonlinear dynamical systems. Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences, pages 26–94, 2005.

[33] Norbert Marwan, Niels Wessel, Udo Meyerfeldt, Alexander Schirdewan, and Jürgen Kurths. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. Physical Review E, 66(2):026702, 2002.

[34] Charles L. Webber Jr and Joseph P. Zbilut. Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology, 76(2):965–973, 1994.

[35] Albert-László Barabási. Fractal concepts in surface growth. Cambridge University Press, Cambridge, United Kingdom, 1995.

[36] Michael J. Katz. Fractals and the analysis of waveforms. Computers in Biology and Medicine, 18(3):145–156, 1988.

[37] Tomoyuki Higuchi. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena, 31(2):277–283, 1988.

[38] Tomoyuki Higuchi. Relationship between the fractal dimension and the power law index for a time series: a numerical investigation. Physica D: Nonlinear Phenomena, 46(2):254–264, 1990.

[39] Chung-Kang Peng, Sergey V. Buldyrev, Shlomo Havlin, Michael Simons, H. Eugene Stanley, and Ary L. Goldberger. Mosaic organization of dna nucleotides. Physical Review E, 49(2):1685, 1994.

[40] Chung-Kang Peng, Shlomo Havlin, Jeffrey M. Hausdorff, Joseph E. Mietus, H. Eugene Stanley, and Ary L. Goldberger. Fractal mechanisms and heart rate dynamics: long-range correlations and their breakdown with disease. Journal of Electrocardiology, 28:59–65, 1995.

[41] Jean-François Muzy, Emmanuel Bacry, and Alain Arneodo. The multifractal formalism revisited with wavelets. International Journal of Bifurcation and Chaos, 4(02):245–302, 1994.

[42] Plamen Ch. Ivanov, Luís A. N. Amaral, Ary L. Goldberger, Shlomo Havlin, Michael G. Rosenblum, Zbigniew R. Struzik, and H. Eugene Stanley. Multifractality in human heartbeat dynamics. Nature, 399(6735):461–465, 1999.

[43] Jean-François Muzy, Emmanuel Bacry, and Alain Arneodo. Wavelets and multifractal formalism for singular signals: application to turbulence data. Physical Review Letters, 67(25):3515–3518, 1991.

[44] Luís A.N. Amaral, Plamen Ch. Ivanov, Naoko Aoyagi, Ichiro Hidaka, Shinji Tomono, Ary L. Goldberger, H. Eugene Stanley, and Yoshiharu Yamamoto. Behavioral-independent features of complex heartbeat dynamics. Physical Review Letters, 86(26):6026, 2001.

[45] Jan W. Kantelhardt, Eva Koscielny-Bunde, Henio H. A. Rego, Shlomo Havlin, and Armin Bunde. Detecting long-range correlations with detrended fluctuation analysis. Physica A: Statistical Mechanics and its Applications, 295(3):441–454, 2001.

[46] Jan W. Kantelhardt, Eva Zschiegner, Stephan A .and Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications, 316(1):87–114, 2002.

[47] SM Ossadnik, Sergey V. Buldyrev, Ary L. Goldberger, Shlomo Havlin, Rosario N. Mantegna, Chung-Kang Peng, Michael Simons, and H. Eugene Stanley. Correlation approach to identify coding regions in dna sequences. Biophysical Journal, 67(1):64–70, 1994.

[48] Armin Bunde, Shlomo Havlin, Jan W. Kantelhardt, Thomas Penzel, Jörg-Hermann Peter, and Karlheinz Voigt. Correlated and uncorrelated regions in heart-rate fluctuations during sleep. Physical Review Letters, 85(17):3736, 2000.

[49] Yosef Ashkenazy, Shlomo Havlin, Plamen Ch. Ivanov, Chung-Kang Peng, Verena Schulte-Frohlinde, and H. Eugene Stanley. Magnitude and sign scaling in power-law correlated time series. Physica A: Statistical Mechanics and its Applications, 323:19–41, 2003.

[50] Harold E. Hurst, Robert P. Black, and Y. M. Simaika. Long-term storage: an experimental study. Constable, 1965.

[51] Harold E. Hurst. Long-term storage capacity of reservoirs. Transactions of the American Society of Engineers, 116:770–808, 1951.

[52] Hans E. Schepers, Johannes HGM Van Beek, and James B. Bassingthwaighte. Four methods to estimate the fractal dimension from self-affine signals (medical application). Engineering in Medicine and Biology Magazine, IEEE, 11(2):57–64, 1992.

[53] Arthur Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terrestrial Magnetism, 3(1):13–41, 1898.

[54] Maurice S. Bartlett. Smoothing periodograms from time series with continuous spectra. Nature, 161(4096):686–687, 1948.

[55] Shlomo Engelberg. Digital signal processing: an experimental approach. Springer Publishing Company, Incorporated, 2008.

[56] Peter D. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics, 15(2):70–73, 1967.

[57] Maurice S. Bartlett. Periodogram analysis and continuous spectra. Biometrika, 37(1-2):1–16, 1950.

[58] Frank G. Borg. Review of nonlinear methods and modelling. arXiv preprint physics/0503026, 2005.

[59] Ary L. Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. The Lancet, 347(9011):1312–1314, 1996.

[60] Caroline Bogaert, Frank Beckers, Dirk Ramaekers, and André E. Aubert. Analysis of heart rate variability with correlation dimension method in a normal population and in heart transplant patients. Autonomic Neuroscience, 90(1):142–147, 2001.

[61] Stefano Guzzetti, Maria G. Signorini, Chiara Cogliati, Silvia Mezzetti, Alberto Porta, Sergio Cerutti, and Alberto Malliani. Non-linear dynamics and chaotic indices in heart rate variability of normal subjects and heart-transplanted patients. Cardiovascular Research, 31(3):441–446, 1996.

[62] Mingzhou Ding, Celso Grebogi, Edward Ott, Tim Sauer, and James A. Yorke. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? Physica D: Nonlinear Phenomena, 69(3):404–424, 1993.

[63] Peter Grassberger and Itamar Procaccia. Characterization of strange attractors. Physical Review Letters, 50(5):346–349, 1983.

[64] Raúl Carvajal, Niels Wessel, Montserrat Vallverdú, Pere Caminal, and Andreas Voss. Correlation dimension analysis of heart rate variability in patients with dilated cardiomyopathy. Computer Methods and Programs in Biomedicine, 78(2):133–140, 2005.

[65] Jorge J. Moré. The levenberg-marquardt algorithm: implementation and theory. In Numerical Analysis, pages 105–116. Springer, 1978.

[66] Claude E Shannon. Communication theory of secrecy systems*. Bell System Technical Journal, 28(4):656–715, 1949.

[67] Alberto Porta, Giuseppe Baselli, D. Liberati, Nicola Montano, C. Cogliati, Tomaso Gnecchi-Ruscone, Alberto Malliani, and Sergio Cerutti. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. Biological Cybernetics, 78(1):71–78, 1998.

[68] Alberto Porta, Stefano Guzzetti, Nicola Montano, Massimo Pagani, V Somers, Alberto Malliani, Giuseppe Baselli, and Sergio Cerutti. Information domain analysis of cardiovascular variability signals: evaluation of regularity, synchronisation and co-ordination. Medical and Biological Engineering and Computing, 38(2):180–188, 2000.

[69] Alberto Porta, Stefano Guzzetti, Nicola Montano, Raffaello Furlan, Massimo Pagani, Alberto Malliani, and Sergio Cerutti. Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series. Biomedical Engineering, IEEE Transactions on, 48(11):1282–1291, 2001.

[70] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[71] Steven M. Pincus and Ary L. Goldberger. Physiological time-series analysis: what does regularity quantify? American Journal of Physiology, 266:H1643–H1643, 1994.

[72] Joaquim P. Marques de Sá. Characterization of fetal heart rate using approximate entropy. In Computers in Cardiology, pages 671–673. IEEE, 2005.

[73] Filipe Magalhaes, Joaquim P. Marques de Sá, João Bernardes, and Diogo Ayres-de Campos. Characterization of fetal heart rate irregularity using approximate entropy and wavelet filtering. In Computers in Cardiology, pages 933–936. IEEE, 2006.

[74] Sheng Lu, Xinnian Chen, Jørgen K. Kanters, Irene C. Solomon, and Ki H. Chon. Automatic selection of the threshold value for approximate entropy. Biomedical Engineering, IEEE Transactions on, 55(8):1966–1972, 2008.

[75] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[76] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[77] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):021906, 2005.

[78] Yi-Cheng Zhang. Complexity and 1/f noise. a phase space approach. Journal de Physique I, 1(7):971–977, 1991.

[79] Jean-Pierre Eckmann, S. Oliffson Kamphorst, David Ruelle, and S. Ciliberto. Liapunov exponents from time series. Physical Review A, 34(6):4971–4979, 1986.

[80] Henry D.I. Abarbanel, Reggie Brown, John J. Sidorowich, and Lev Sh. Tsimring. The analysis of observed chaotic data in physical systems. Reviews of Modern Physics, 65(4):1331, 1993.

[81] Simon Haykin and Xiao Bo Li. Detection of signals in chaos. Proceedings of the IEEE, 83(1):95–122, 1995.

[82] Alan Wolf, Harry L. Swift, Jack B .and Swinney, and John A. Vastano. Determining lyapunov exponents from a time series. Physica D: Nonlinear Phenomena, 16(3):285–317, 1985.

[83] U. Rajendra Acharya, N. Kannathal, Ong W. Sing, Luk Y. Ping, and TjiLeng Chua. Heart rate analysis in normal subjects of various age groups. BioMedical Engineering OnLine, 3(1):24, 2004.

[84] Vikram K. Yeragani, Rao K.A. Radhakrishna, Kalpathi R. Ramakrishnan, and S.H. Srinivasan. Measures of lle of heart rate in different frequency bands: a possible measure of relative vagal and sympathetic activity. Nonlinear Analysis: Real World Applications, 5(3):441–462, 2004.

[85] Vikram K. Yeragani, M. Ramesh Radhakrishna, Rao K.A .and Smitha, Robert B. Pohl, Richard Balon, and K. Srinivasan. Diminished chaos of heart rate time series in patients with major depression. Biological Psychiatry, 51(9):733–744, 2002.

[86] Keith Briggs. An improved method for estimating liapunov exponents of chaotic time series. Physics Letters A, 151(1):27–32, 1990.

[87] Reggie Brown. Calculating lyapunov exponents for short and/or noisy data sets. Physical Review E, 47(6):3962, 1993.

[88] Michael T. Rosenstein, James J. Collins, and Carlo J. De Luca. A practical method for calculating largest lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena, 65(1):117–134, 1993.

[89] Masaki Sano and Yasuji Sawada. Measurement of the lyapunov spectrum from a chaotic time series. Physical Review Letters, 55(10):1082, 1985.

[90] Shinichi Sato, Masaki Sano, and Yasuji Sawada. Practical methods of measuring the generalized dimension and the largest lyapunov exponent in high dimensional chaotic systems. Progress of Theoretical Physics, 77(1):1–5, 1987.

[91] Xubin Zeng, Richard Eykholt, and Roger A. Pielke. Estimating the lyapunov-exponent spectrum from short time series of low precision. Physical Review Letters, 66(25):3229, 1991.

[92] Jürgen Fell and Peter E. Beckmann. Resonance-like phenomena in lyapunov calculations from data reconstructed by the time-delay method. Physics Letters A, 190(2):172–176, 1994.

[93] Mauricio Barahona and Chi-Sang Poon. Detection of nonlinear dynamics in short, noisy time series. Nature, 381(6579):215–217, 1996.

[94] Chi-Sang Poon and Mauricio Barahona. Titration of chaos with added noise. Proceedings of the National Academy of Sciences, 98(13):7107–7112, 2001.

[95] Chi-Sang Poon and Christopher K. Merrill. Decrease of cardiac chaos in congestive heart failure. Nature, 389(6650):492–495, 1997.

[96] Michael J. Korenberg. Identifying nonlinear difference equation and functional expansion representations: the fast orthogonal algorithm. Annals of Biomedical Engineering, 16(1):123–142, 1988.

[97] Hirotugu Akaike. A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6):716–723, 1974.

[98] Jacques Hadamard. Les surfaces à courbures opposées et leurs lignes géodésique. Journal de Mathématiques Pures et Appliquées, 4:27–73, 1898.

[99] Andreas Voss, Juergen Kurths, H.J. Kleiner, A Witt, Niels Wessel, P. Saparin, K. J. Osterziel, R. Schurath, and Rainer Dietz. The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. Cardiovascular Research, 31(3):419–433, 1996.

[100] Andreas Voss, Katerina Hnatkova, Niels Wessel, Juergen Kurths, Andre Sander, Alexander Schirdewan, A John Camm, and Marek Malik. Multiparametric analysis of heart rate variability used for risk stratification among survivors of acute myocardial infarction. Pacing and Clinical Electrophysiology, 21(1):186–196, 1998.

[101] Floris Takens. Detecting strange attractors in turbulence. In Dynamical systems and turbulence, pages 366–381. Springer, 1981.

[102] Alberto Porta, Eleonora Tobaldini, Stefano Guzzetti, Raffaello Furlan, Nicola Montano, and Tomaso Gnecchi-Ruscone. Assessment of cardiac autonomic modulation during graded head-up tilt by symbolic analysis of heart rate variability. American Journal of Physiology-Heart and Circulatory Physiology, 293(1):H702–H708, 2007.

[103] Henry D.I. Abarbanel and Jerry P. Gollub. Analysis of observed chaotic data. Physics Today, 49(11):86–88, 2008.

[104] Dirk Hoyer, Bernd Pompe, Ki H Chon, Henning Hardraht, Carola Wicher, and Ulrich Zwiener. Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales. Biomedical Engineering, IEEE Transactions on, 52(4):584–592, 2005.

Generalized Measure for Observer Agreement

Measure what can be measured, and make measurable what cannot be measured.

Galileo Galilei

3. Generalized Measure for Observer Agreement

Disagreement on the interpretation of diagnostic tests and clinical decisions remains an important problem with potential research, clinical and medico-legal consequences [1]. In the past years some measures were proposed [2–7], as also the use of more than one measure [8], trying to quantify this notion but problems arise when comparing the degree of observer agreement among different methods, populations or circumstances. The inconsistent results obtained when assessing observer agreement, as the particular case of the two widely used measures, the limits of agreement (LA) and the intra-class correlation coefficient (ICC) [9], appears to be the main reason of the lack of a "gold-standard" measure. Recently a new measure, the information-based measure of disagreement (**IBMD**) [10], based on the amount of information contained in the differences between two observations was introduced. However, it was proposed to assess disagreement between only two observers. The generalized measure to include more than two observers is presented on the first section of chapter [11].

On **Section 3.2** two platforms to help health care professionals and biostatisticians when performing observer agreement studies are presented. An intuitive web-based software system was created to facilitate an easy application of several statistical agreement measures strategies to their own data. A R package **obs.agree** was also built to provide an easy way to assess agreement (or disagreement) among multiple measurements for the same subject by different observers. The package includes two different measures considered adequate to assess the agreement among multiple observers allowing their comparability across populations: the Raw Agreement Indices (RAI) [3] for categorical data and the IBMD for continuous data.

Of note, the work developed in this chapter resulted in a publish article, a website, a R package and an article submitted.

Bibliography

[1] Cristina Costa-Santos, Altamiro C. Pereira, and João Bernardes. Agreement studies in obstetrics and gynaecology: inappropriateness, controversies and consequences. BJOG: An International Journal of Obstetrics & Gynaecology, 112(5):667–669, 2005.

[2] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1):37, 1960.

[3] John S. Uebersax. Statistical methods for rater and diagnostic agreement. http://www.john-uebersax.com/stat/raw.htm. Accessed: Jun 2014.

[4] John J. Bartko. The intraclass correlation coefficient as a measure of reliability. Psychological Reports, 19(1):3–11, 1966.

[5] Lawrence Lin, A. Samad Hedayat, and Wenting Wu. A unified approach for assessing agreement for continuous and categorical data. Journal of Biopharmaceutical Statistics, 17(4):629–652, 2007.

[6] J. Martin Bland and Douglas G. Altman. Agreement between methods of measurement with multiple observations per individual. Journal of Biopharmaceutical Statistics, 17(4):571–582, 2007.

[7] Philip Schluter. A multivariate hierarchical bayesian approach to measuring agreement in repeated measurement method comparison studies. BMC Medical Research Methodology, 9(1):6, 2009.

[8] Ronir R Luiz and Moyses Szklo. More than one statistical strategy to assess agreement of quantitative measurements may usefully be reported. Journal of Clinical Epidemiology, 58(3):215–216, 2005.

[9] Cristina Costa-Santos, João Bernardes, Diogo Ayres-de Campos, Antónia Costa, and Célia Costa. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. Journal of Clinical Epidemiology, 64(3):264–269, 2011.

[10] Cristina Costa-Santos, Luís Antunes, André Souto, and João Bernardes. Assessment of disagreement: A new information-based approach. Annals of Epidemiology, 20(7):555–561, 2010.

[11] Teresa Henriques, Luís Antunes, João Bernardes, Mara Matias, Diogo Sato, and Cristina Costa-Santos. Information-based measure of disagreement for more than two observers: a useful tool to compare the degree of observer disagreement. BMC Medical Research Methodology, 13(1):1–6, 2013.

Information-Based Measure of Disagreement for more than two observers: a useful tool to compare the degree of observer disagreement

Teresa Henriques, Luis Antunes, João Bernardes,
Mara Matias, Diogo Sato and Cristina Costa-Santos

BMC Medical Research Methodology
Volume 13, Issue 1, Article 47
March 2013

3.1 Information-Based Measure of Disagreement for more than two observers: a useful tool to compare the degree of observer disagreement

Abstract

Background: Assessment of disagreement among multiple measurements for the same subject by different observers remains an important problem in medicine. Several measures have been applied to assess observer agreement. However, problems arise when comparing the degree of observer agreement among different methods, populations or circumstances.

Methods: The recently introduced information-based measure of disagreement (IBMD) is a useful tool for comparing the degree of observer disagreement. Since the proposed IBMD assesses disagreement between two observers only, we generalized this measure to include more than two observers.

Results: Two examples (one with real data and the other with hypothetical data) were employed to illustrate the utility of the proposed measure in comparing the degree of disagreement.

Conclusion: The IBMD allows comparison of the disagreement in non-negative ratio scales across different populations and the generalization presents a solution to evaluate data with different number of observers for different cases, an important issue in real situations.

A website for online calculation of IBMD and respective 95% confidence interval was additionally developed. The website is widely available to mathematicians, epidemiologists and physicians to facilitate easy application of this statistical strategy to their own data.

Background

As several measurements in clinical practice and epidemiologic research are based on observations made by health professionals, assessment of the degree of disagreement among multiple measurements for the same subjects under similar circumstances by different observers remains a significant problem in medicine. If the measurement error is assumed to be the same for every observer, independent of the magnitude of quantity, we can estimate within-subject variability for repeated measurements by the same subject with the within-subject standard deviation, and the increase in variability when different observers are applied using analysis of variance[1]. However this strategy is not appropriate for comparing the degree of observer disagreement among different populations or various methods of measurement. Bland and Altman proposed a technique to compare the agreement between two methods of medical measurement allowing multiple observations per subject [2] and later Schluter proposed a Bayesian approach [3]. However, problems arise when comparing the degree of observer disagreement between two different methods, populations or circumstances. For example, one issue is whether during visual analysis of cardiotocograms, observer disagreement in estimation of the fetal heart rate baseline in the first hour of labor

is significantly different from that in the last hour of labor when different observers assess the printed one-hour cardiotocography tracings. Another issue that remains to be resolved is whether interobserver disagreement in head circumference assessment by neonatologists is less than that by nurses. To answer to this question, several neonatologists should evaluate the head circumference in the same newborns under similar circumstances, followed by calculation of the measure of interobserver agreement, and the same procedure repeated with different nurses. Subsequently, the two interobserver agreement measures should be compared to establish whether interobserver disagreement in head circumference assessment by neonatologists is less than that by nurses.

Occasionally, intraclass correlation coefficient (ICC), a measure of reliability, and not agreement [4] is frequently used to assess observer agreement in situations with multiple observers without knowing the differences between the numerous variations of the ICC [5]. Even when the appropriate form is applied to assess observer agreement, the ICC is strongly influenced by variations in the trait within the population in which it is assessed [6]. Consequently, comparison of ICC is not always possible across different populations. Moreover important inconsistencies can be found when ICC is used to assess agreement [7].

Lin's concordance correlation coefficient (CCC) is additionally applicable to situations with multiple observers. The Pearson coefficient of correlation assesses the closeness of data to the line of best fit, modified by taking into account the distance of this line from the 45-degree line through the origin [8–13]. Lin objected to the use of ICC as a way of assessing agreement between methods of measurement, and developed the CCC. However, similarities exist between certain specifications of the ICC and CCC measures. Nickerson, C. [14] showed the asymptotic equivalence among the ICC and CCC estimators. However, Carrasco and Jover [15] demonstrated the equivalence between the CCC and a specific ICC at parameter level. Moreover, a number of limitations of ICC, such as comparability of populations and its dependence on the covariance between observers, described above, are also present in CCC [16]. Consequently, CCC and ICC to measure observer agreement from different populations are valid only when the measuring ranges are comparable [17].

The recently introduced information-based measure of disagreement (IBMD) provides a useful tool to compare the degree of observer disagreement among different methods, populations or circumstances [18]. However, the proposed measure assesses disagreement only between two observers, which presents a significant limitation in observer agreement studies. This type of study generally requires more than just two observers, which constitutes a very small sample set.

Here, we have proposed generalization of the information-based measure of disagreement for more than two observers. As sometimes in real situations some observers do not examine all the cases (missing data), our generalized IBMD is set to allow different numbers of examiners for various observations.

Methods

IBMD among more than two observers

A novel measure of disagreement, denoted 'information-based measure of disagreement' (IBMD), was proposed [18] on the basis of Shannon's notion of entropy [19], described as the average amount of information contained in a variable. In this context, the sum over all logarithms of possible outcomes of the variable is a valid measure of the amount of information, or uncertainty, contained in a variable

[19]. IBMD, use logarithms to measures the amount of information contained in the differences between two observations. This measure is normalized and satisfies the flowing properties: it is a metric, scaled invariant with differential weighting [18].

N was defined as the number of cases and x_{ij} as observation of the subject i by observer j. The disagreement between the observations made by observer pair 1 and 2 was defined as:

$$IBMD = \frac{1}{N}\sum_{i=1}^{N} log_2(\frac{|x_{i1} - x_{i2}|}{max\{x_{i1}, x_{i2}\}} + 1) \tag{3.1}$$

We aim to measure the disagreement among measurements obtained by several observers, allowing different number of observations in each case. Thus, maintaining N as the number of cases, we consider $M_i, i = 1, .., N$, as the number of observations in case i.

Therefore considering N vectors, one for each case, $(x_{11}, ..., x_{1M1}), ..., (x_{N1}, ..., x_{NMN})$ with non-negative components, the generalized information-based measure of disagreement is defined as:

$$IBMD = \frac{1}{\sum_{i=1}^{N}\binom{M_i}{2}} \sum_{i=1}^{N}\sum_{j=1}^{M_i-1}\sum_{k=j+1}^{M_i} log_2(\frac{|x_{ij} - x_{ik}|}{max\{x_{ij}, x_{ik}\}} + 1) \tag{3.2}$$

with the convention $\frac{|0-0|}{max\{0,0\}} = 0$.

This coefficient equals 0 when the observers agree or when there is no disagreement, and increases to 1 when the distance, i.e. disagreement among the observers, increases.

The standard error and confidence interval was based on the nonparametric bootstrap, by resampling the subjects/cases with replacement, in both original and generalized IBMD measures. The bootstrap uses the data from a single sample to simulate the results if new samples were repeated over and over. Bootstrap samples are created by sampling with replacement from the dataset. A good approximation of the 95% confidence interval can be obtained by computing the $2.5th$ and $97.5th$ percentiles of the bootstrap samples. Nonparametric resampling makes no assumptions concerning the distribution of the data. The algorithm for a nonparametric bootstrap is as follows [20]:

1. Sample N observations randomly with replacement from the N cases to obtain a bootstrap data set.
2. Calculate the bootstrap version of IBMD.
3. Repeat steps 1 and 2 a B times to obtain an estimate of the bootstrap distribution.

For confidence intervals of 90–95 percent B should be between 1000 and 2000 [21, 22]. In the results the confidence intervals were calculated with B equal to 1000.

Software for IBMD assessment

Website

We have developed a website to assist with the calculation of IBMD and respective 95% confidence intervals [23]. This site additionally includes computation of the intraclass correlation coefficient (ICC). Lin's concordance correlation coefficient (CCC) and limits of agreement can also be measured when considering only two observations per subject. The website contains a description of these methods.

PAIRSetc software

PAIRSetc [24, 25], a software that compares matched observations, provide several agreement measures, among them the ICC, the CCC and the 95% limits of agreement. This software is constantly updated with new measures introduced on scientific literature, in fact, a coefficient of individual equivalence to measure agreement, based on replicated readings proposed in 2011 by Pan et al. [26, 27] and IBMD, published in 2010, were already include.

Examples

Two examples (one with real data and the other with hypothetical data) were employed to illustrate the utility of the IBMD in comparing the degree of disagreement.

A gymnast's performance is evaluated by a jury according to rulebooks, which include a combination of the difficulty level, execution and artistry. Let us suppose that a new rulebook has been recently proposed and subsequently criticized. Some gymnasts and media argue that disagreement between the jury members in evaluating the gymnastics performance with the new scoring system is higher than that with the old scoring system, and therefore oppose its use. To better understand this claim, consider a random sample of eight judges evaluating a random sample of 20 gymnasts with the old rulebook, and a different random sample of 20 gymnasts with the new rulebook. In this case, each of the 40 gymnasts presented only one performance based on pre-defined compulsory exercises, and all eight judges simultaneously viewed the same performances and rated each gymnast independently, while blinded to their previous medals and performances. Both scoring systems ranged from 0 to 10. The results are presented in Table 3.1.1.

Visual analysis of the maternal heart rate during the last hour of labor can be more difficult than that during the first hour. We believe that this is a consequence of the deteriorated quality of signal and increasing irregularity of the heart rate (due to maternal stress). Accordingly, we tested this hypothesis by examining whether in visual analysis of cardiotocograms, observer disagreement in fetal heart rate baseline estimation in the first hour of labor is lower than that in the last hour of labor when different observers assess printed one-hour cardiotocography tracings. To answer this question, we evaluated the disagreement in maternal heart rate baseline estimation during the last and first hour of labor by three independent observers.

Specifically, the heart rates of 13 mothers were acquired, as secondary data collected in Nélio Mendonça Hospital, Funchal for another study, during the initial and last hour of labor, and printed. Three experienced obstetricians were asked to independently estimate the baseline of the 26 one-hour segments. Results are presented in Table 3.1.2. The study procedure was approved by the local Research Ethics Committees and followed the Helsinki declaration. All women who participate in the study gave informed consent to participate.

Table 3.1.1: Performance of 40 gymnasts, 20 evaluated by eight judges using the old rulebook and 20 by the same judges using the new rulebook

ID gymnast	Rulebook	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	Judge 7	Judge 8
1	Old	7.10	7.20	7.00	7.70	7.10	7.10	7.00	7.30
2	Old	9.30	9.70	8.90	9.60	8.60	9.50	9.60	9.70
3	Old	8.90	8.80	8.10	9.30	8.50	8.10	7.60	8.70
4	Old	8.00	8.10	7.30	8.70	7.50	8.70	7.40	9.50
5	Old	9.10	9.00	8.20	9.00	8.20	9.50	7.80	8.00
6	Old	9.10	9.20	8.30	9.10	7.90	8.90	9.00	9.20
7	Old	8.90	9.00	7.70	9.00	8.00	9.40	8.00	7.70
8	Old	8.30	8.70	8.10	8.90	7.80	9.20	7.80	9.30
9	Old	9.30	9.40	8.20	9.40	8.80	9.30	9.20	9.80
10	Old	9.40	9.80	9.40	9.70	9.10	10.00	9.30	9.60
11	Old	7.70	8.70	7.60	9.00	7.70	8.50	7.70	7.70
12	Old	9.20	9.70	8.50	9.60	8.60	9.90	9.70	7.40
13	Old	7.40	7.30	7.10	7.90	7.10	7.40	7.00	7.50
14	Old	8.40	8.90	7.40	8.60	7.80	8.10	7.40	8.90
15	Old	7.40	7.60	7.10	8.10	7.20	7.60	7.10	8.80
16	Old	9.80	9.90	9.20	9.80	9.30	10.00	9.40	9.60
17	Old	9.60	9.60	9.50	9.80	9.10	9.90	9.40	9.90
18	Old	9.60	9.80	9.50	9.80	8.80	9.90	9.80	9.20
19	Old	8.50	9.20	7.80	9.30	7.90	9.00	7.70	9.70
20	Old	7.10	9.50	8.80	9.40	8.50	9.60	7.90	8.50
21	New	6.50	8.20	6.60	9.80	7.50	7.80	6.10	5.10
22	New	7.00	9.70	7.60	9.60	8.30	6.90	6.70	8.60
23	New	7.50	8.60	6.60	7.80	9.50	8.10	6.20	7.60
24	New	8.50	9.00	8.10	7.00	8.30	9.40	6.70	8.00
25	New	9.70	8.10	7.50	6.80	7.70	8.60	8.30	7.40
26	New	8.00	9.10	7.40	9.30	8.30	9.70	6.00	9.90
27	New	7.80	9.70	7.00	9.70	8.70	10.00	9.60	9.50
28	New	9.30	7.90	8.20	7.80	6.30	7.40	6.10	7.20
29	New	7.10	9.80	8.10	9.50	6.30	9.40	8.90	6.50
30	New	8.90	9.30	7.90	6.80	8.20	9.10	7.90	6.80
31	New	9.30	9.80	8.80	6.60	8.50	9.80	7.40	9.90
32	New	7.90	8.20	6.70	9.40	7.60	6.10	7.40	7.10
33	New	7.60	8.50	6.40	8.50	9.20	7.80	6.20	9.40
34	New	8.60	8.90	6.50	9.00	7.70	9.10	6.50	7.10
35	New	8.80	7.20	8.80	9.30	8.40	9.30	6.90	8.60
36	New	8.40	9.30	7.50	8.70	7.90	9.60	7.90	7.90
37	New	7.50	8.00	7.20	8.40	7.40	7.20	9.10	9.20
38	New	9.70	9.80	9.50	9.80	9.00	9.90	9.40	9.60
39	New	8.50	9.20	8.70	9.30	7.00	9.70	8.30	8.00
40	New	7.30	8.70	7.20	8.10	7.30	7.30	7.10	7.20

Table 3.1.2: Estimation of baseline (bpm) in 26 segments of 13 traces (13 segments corresponding to the initial hour of labor and 13 to the final hour of labor) by three obstetricians

Mother	ID Segment	Obstetrician 1	Obstetrician 2	Obstetrician 3
1	Initial hour	80	80	80
2	Initial hour	65	66	70
3	Initial hour	65	66	70
4	Initial hour	63	67	65
5	Initial hour	82	83	85
6	Initial hour	75	76	75
7	Initial hour	80	81	85
8	Initial hour	84	85	80
9	Initial hour	100	102	105
10	Initial hour	82	82	80
11	Initial hour	67	65	70
12	Initial hour	75	74	87
13	Initial hour	70	70	70
1	Last hour	78	75	75
2	Last hour	90	90	100
3	Last hour	70	67	70
4	Last hour	70	65	65
5	Last hour	87	87	90
6	Last hour	72	73	75
7	Last hour	75	75	75
8	Last hour	100	98	100
9	Last hour	110	108	110
10	Last hour	103	103	100
11	Last hour	80	80	100
12	Last hour	98	100	100
13	Last hour	70	70	65

Results

Hypothetical data example

Using IBMD in the gymnast's evaluation, we can compare observer disagreement and the respective confidence interval (CI) associated with each score system.

The disagreement among judges was assessed as IBMD = 0.090 (95% CI = [0.077; 0.104]) considering the old rulebook and IBMD = 0.174 (95% CI = [0.154; 0.192]) with new rulebook. Recalling that the value 0 of the IBMD means no disagreement (perfect agreement), these confidence intervals clearly indicate significantly higher observer disagreement in performance evaluation using the new scoring system, compared with the old system.

Real data example

The disagreement among obstetricians in baseline estimation, considering the initial hour of labor, was IBMD = 0.048 (95% CI = [0.036; 0.071]), and during the last hour of labor, IBMD = 0.048 (95% CI = [0.027; 0.075]). The results indicate no significant differences in the degree of disagreement among observers between the initial and last hour of labor.

Discussion

While comparison of the degree of observer disagreement is often required in clinical and epidemiologic studies, the statistical strategies for comparative analyses are not straightforward.

Intraclass correlation coefficient is several times used in this context, however sometimes without careful in choosing the correct form. Even when the correct form of ICC is used to assess agreement, its dependence on variance does not always allow the comparability of populations. Other approaches to assess observer agreement have been proposed [28–33], but comparative analysis across populations is still difficult to achieve. The recently proposed IBMD is a useful tool to compare the degree of disagreement in non-negative ratio scales [18], and its proposed generalization allowing several observers overcomes an important limitation of this measure in this type of analysis where more than two observers are required.

Conclusions

IBMD generalization provides a useful tool to compare the degree of observer disagreement among different methods, populations or circumstances and allows evaluation of data by different numbers of observers for different cases, an important feature in real situations where some data are often missing.

The free software and available website to compute generalized IBMD and respective confidence intervals facilitates the broad application of this statistical strategy.

Competing interests

There are any non-financial competing interests (political, personal, religious, ideological, academic, intellectual, commercial or any other) to declare in relation to this manuscript.

Authors' contributions

TH, LA and CCS have made substantial contributions to article conception and design. They also have been involved the analysis and interpretation of data and they draft the manuscript. JB have been involved in article conception and he revised the manuscript critically for important intellectual content. MM and DS were responsible for the creation of the software and also were involved in the data analysis. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Paula Pinto from Nélio Mendonça Hospital, Funchal, who allowed us to use the maternal heart rate dataset collected for her PhD studies, approved by the local Research Ethics Committees. This work was supported by the national science foundation, Fundação para a Ciência e Tecnologia, through FEDER founds though Programa Operacional Fatores de Competitividade – COMPETE through the project CSI2 with the reference PTDC/EIA-CCO/099951/2008, through the project with the reference PEST-C/SAU/UI0753/2011 and though the PhD grant with the reference SFRH /BD/70858/2010.

Bibliography

[1] J. Martin Bland. An introduction to medical statistics. Oxford University Press, third edition, 2000.

[2] J. Martin Bland and Douglas G. Altman. Agreement between methods of measurement with multiple observations per individual. Journal of Biopharmaceutical Statistics, 17(4):571–582, 2007.

[3] Philip Schluter. A multivariate hierarchical bayesian approach to measuring agreement in repeated measurement method comparison studies. BMC Medical Research Methodology, 9(1):6, 2009.

[4] H. De Vet. Observer Reliability and Agreement. John Wiley & Sons, Ltd, 2005.

[5] Patrick E. Shrout and Joseph L. Fleiss. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, January 1979.

[6] Reinhold Muller and Petra Buttner. A critical discussion of intraclass correlation coefficients. Statistics in Medicine, 13(23-24):2465–2476, 1994.

[7] Cristina Costa-Santos, João Bernardes, Diogo Ayres-de Campos, Antónia Costa, and Célia Costa. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. Journal of Clinical Epidemiology, 64(3):264–269, 2011.

[8] Huiman X. Barnhart, Michael J. Haber, and Lawrence I. Lin. An overview on assessing agreement with continuous measurements. Journal of Biopharmaceutical Statistics, 17(4):529–569, 2007.

[9] Josep L. Carrasco, Tonya S. King, and Vernon M. Chinchilli. The concordance correlation coefficient for repeated measures estimated by variance components. Journal of Biopharmaceutical Statistics, 19(1):90–105, 2009.

[10] Tonya S. King, Vernon M. Chinchilli, and Josep L. Carrasco. A repeated measures concordance correlation coefficient. Statistics in Medicine, 26(16):3095–3113, 2007.

[11] Lawrence I. Lin, A. Samad Hedayat, and Wenting Wu. A unified approach for assessing agreement for continuous and categorical data. Journal of Biopharmaceutical Statistics, 17(4):629–652, 2007.

[12] Lawrence I. Lin. A concordance correlation coefficient to evaluate reproducibility. Biometrics, pages 255–268, 1989.

[13] Lawrence I. Lin, A. Samad Hedayat, Bikas Sinha, and Min Yang. Statistical methods in assessing agreement: Models, issues, and tools. Journal of the American Statistical Association, 97(457):257–270, 2002.

[14] Carol A.E. Nickerson. A note on" a concordance correlation coefficient to evaluate reproducibility". Biometrics, pages 1503–1507, 1997.

[15] Josep L. Carrasco and Lluis Jover. Estimating the generalized concordance correlation coefficient through variance components. Biometrics, 59(4):849–858, 2003.

[16] Greg Atkinson and Alan Nevil. Comment on the use of concordance correlation to assess the agreement between two variables. Biometrics, 53:775–777, 1997.

[17] Lawrence I. Lin and Vernon Chinchilli. Rejoinder to the Letter to the Editor from Atkinson and Nevill. Biometrics, 53(2):777–778, 1997.

[18] Cristina Costa-Santos, Luís Antunes, André Souto, and João Bernardes. Assessment of disagreement: A new information-based approach. Annals of Epidemiology, 20(7):555–561, 2010.

[19] Claude E. Shannon. The mathematical theory of communication. 1963. MD Computing: Computers in Medical Practice, 14(4):306, 1997.

[20] James Carpenter and John Bithell. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. Statistics in Medicine, 19(9):1141–1164, 2000.

[21] Bradley Efron and Robert J. Tibshirani. An Introduction to the Bootstrap. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.

[22] Anthony C. Davison and David V. Hinkley. Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[23] Cristina Costa-Santos. Ibmd on-line calculator. url = http://disagreement.med.up.pt. Accessed: Jun 2014.

[24] Joseph H. Abramson. Winpepi (pepi-for-windows): computer programs for epidemiologists. Epidemiologic Perspectives & Innovations, 1(1):6, 2004.

[25] Joseph H. Abramson. Winpepi updated: computer programs for epidemiologists, and their teaching potential. Epidemiologic Perspectives & Innovations, 8(1):1, 2011.

[26] Yi Pan, Michael Haber, and Huiman X. Barnhart. A new permutation-based method for assessing agreement between two observers making replicated binary readings. Statistics in Medicine, 30(8):839–853, 2011.

[27] Yi Pan, Michael Haber, Jingjing Gao, and Huiman X. Barnhart. A new permutation-based method for assessing agreement between two observers making replicated quantitative readings. Statistics in Medicine, 31(20):2249–2261, 2012.

[28] Ronir R. Luiz, Antonio J.L. Costa, Pauline L. Kale, and Guilherme L. Werneck. Assessment of agreement of a quantitative variable: a new graphical approach. Journal of Clinical Epidemiology, 56(10):963–967, 2003.

[29] Katherine L. Monti. Folded empirical distribution function curves-mountain plots. The American Statistician, 49(4):342–345, 1995.

[30] Hui Quan and Weichung J. Shih. Assessing reproducibility by the within-subject coefficient of variation with random effects models. Biometrics, pages 1195–1203, 1996.

[31] Lawrence I. Lin. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. Statistics in Medicine, 19(2):255–270, 2000.

[32] Geòrgia Escaramís, Carlos Ascaso, and Josep L. Carrasco. The total deviation index estimated by tolerance intervals to evaluate the concordance of measurement devices. BMC Medical Research Methodology, 10(31):1–12, 2010.

[33] Huiman X. Barnhart, Andrzej S. Kosinski, and Michael J. Haber. Assessing individual agreement. Journal of Biopharmaceutical Statistics, 17(4):697–719, 2007.

Facilitating the evaluation of agreement between measurements

Teresa Henriques, Luís Antunes,
Mara Matias, Cristina Costa-Santos

SUBMITTED

3.2 Facilitating the evaluation of agreement between measurements

Abstract

Background: Several measures have been applied to assess agreement in epidemiologic studies. However, problems arise when comparing the degree of observer agreement among different methods, populations or circumstances.

Objective: The project endeavors to create an intuitive Web-based software system, available to mathematicians, epidemiologists and physicians that facilitate easy application of several statistical agreement measures strategies to their own data. An R package **obs.agree** was also developed for an easy computation of observer agreement measures.

Methods: In order to make the website easily available in several platforms (from mobile devices as well as regular laptops) we used the responsive design approach. To implement this design, HTML5 and the library bootstrap which includes both CSS and javascript modules for graphic response were used. We use the programing language R to implement two recent functions for measuring agreement: the raw agreement indices (RAI) to categorical data and information-based measure of disagreement (IBMD) to continuous data, and made it available as an R package.

Results: There is now widely available a website where mathematicians, epidemiologists and physicians can easily evaluate and compare the results of several measures of agreement. There is also available, from the Comprehensive R Archive Network, a new R package **obs.agree** to assess the agreement among multiple measurements for the same subject by different observers.

Conclusions: Comparison of the degree of observer disagreement is often required in clinical and epidemiologic studies. However, the statistical strategies for comparative analyses are not straightforward and software for RAI and IBMD assessment is lacking. The website and package have the potential to help health care professionals and biostatisticians when performing observer agreement studies, as it provides an easy way to calculate raw agreement indices to categorical data and information-based measure of disagreement to continuous variables.

Introduction

Assessment of agreement among multiple measurements for the same subject by different observers under similar circumstances remains an important problem in medicine and epidemiology. Several measures have been applied to assess observer agreement based on the data type.

When assessing agreement on discrete data the Cohen's Kappa coefficient is one of the most widely used (there are more than one R package that includes this measure). It was proposed by Cohen [1] as the proportion of chance-expected disagreements which do not occur, or alternatively, it is the proportion of agreement after chance agreement is removed from consideration. However, it has some inherent limitations, in particular not always very low values of Kappa reflect low rates of overall agreement since Kappa coefficient is affected by prevalence and by imbalance in the table's marginal totals [2, 3]. Uebersax

presented the generalized raw agreement indices (RAI) [4, 5] overcoming the concerns with the Kappa measure.

Considering continuous data, Bland and Altman [6] proposed a technique to compare the agreement between two methods of medical measurement allowing multiple observations per subject; and later Schluter [7]proposed a Bayesian approach. However, problems rise when comparing the degree of observer agreement among different methods, populations or circumstances. The Intraclass correlation coefficient (ICC) is a measure of reliability, and not agreement. However, it is often used for the assessment of observer agreement in situations with multiple observers [8]. The ICC is strongly influenced by variations in the trait within the population in which it is assessed. Consequently, comparison of this value is not possible across different populations [9]. Lin objected to the use of ICC as a way of assessing agreement between methods of measurement, and developed the Lin's concordance correlation coefficient (CCC). However, a number of limitations of ICC, such as comparability of populations, are also present in CCC [10]. Recently a new approach was introduced, the information-based measure of disagreement (IBMD) [11, 12] that assesses disagreement allowing different number of observers for different cases. The IBMD allows comparison of the disagreement in non-negative ratio scales of different populations.

We developed a website [13] to assist with the calculation of several agreement measures, namely: the IBMD and respective 95% confidence intervals, the ICC, CCC and limits of agreement can also be measured. The website contains a description of these methods.

During the development of the site we build R implementations of the RAI and of IBMD measures, now available in the R package **obs.agree** . These two measures are adequate to assess the agreement among multiple observers allowing their comparability across populations.

Methods

Disagreement website

The purpose of the website [13] is to make widely available to mathematicians, epidemiologists and physicians, and to facilitate easy application, several statistical agreement measures strategies to their own data. To achieve this goal one of the central points of the website was to have a calculator available so users could actually see how the measures works, and even compare it's results to other known measures such as CCC or ICC. One of the requirements for the website was that it should be usable from mobile devices as well as regular laptops, this was the main reason why we used the responsive design approach. To implement this design, HTML5 and the library bootstrap [14] which includes both CSS and javascript modules for graphic response were used. The website's backend is implemented in PHP5. Since the site is composed of a series of static content pages and a single interactive page were the calculator itself is, to decrease the amount of duplicate code the PHP template engine Smarty [15] was used. In Illustration 1 we present the responsive design for various display aspects.

Illustration 1: Responsive Design for various display aspects. TopLeft:320x480, Top right:786x1024, bottom: 1280x600

One of the concerns in any website with user input is the validation of the data entered. Given its importance, we decided to implement the validation both on the frontend and the backend. The reason for this double validation is because the backend functionalities could potentially be used directly

Table 3.2.1: Data for measuring agreement between two raters on dichotomous ratings

		Rater2	
		+	−
Rater1	+	a	b
	−	c	d

through some library (like urllib from python), so even though the data passes the frontend's validations it is processed again at the backend. The frontend validation is done by javascript functions, and provides the user with an immediate response to the data as it's being entered, for example with hints as to why a particular calculator cannot be used for the data set entered. The backend validation is done by PHP functions before the data is accepted and passed into each of the calculators chosen by the user. Since this is an open website and we could never guarantee the complete anonimization of the data entered, so we chose to use a clear communication. All data is sent to the server in clear text and no security mechanisms like SSL certificates are used, and so no data at all is permanently recorded.

obs.agree package

Several R (R Development Core Team 2012) packages are available on the Comprehensive R Archive Network (CRAN) [16] to evaluate observer agreement. In this section we describe the package **obs.agree** [17], which includes two measures adequate to assess the agreement among multiple observers allowing their comparability across populations. The raw agreement indices (RAI) are more suitable for categorical data, whereas the information-based measure of disagreement (IBMD) is more adequate for continuous data.

RAI is a group of measures composed by an overall agreement index and specific agreement indices for each rating categories, considering the proportion of agreement between two raters on dichotomous ratings as shown in Table 3.2.1.

The most frequently used overall proportion of agreement index is simply $P_O = \frac{a+d}{N}$. The proportions of specific agreement for each category, positive (PA) and negative (NA) are [18]:

$$PA = \frac{2a}{2a+b+c} \text{ and } NA = \frac{2d}{2d+b+c}.$$

The generalized case of the raw agreement indices allow any number of raters, making polytomous ratings (either ordered category or purely nominal), with potentially different numbers of raters for each case.

Let C be the number of rating categories. Define $n_{jk} = n_{1k}, n_{2k}, ..., n_{Ck}$ as the number of ratings of category j in case k, with $j = 1, ..., C$, for all $k = 1, ..., K$ and $n_k = \sum_{j=1}^{C} n_{jk}$ the number of observers of case k. The number of total agreements (S_j) and possible agreements ($Sposs_j$) in each category j are

$$S_j = \sum_{k=1}^{K} n_{jk}(n_{jk} - 1) \text{ and } Sposs_j = \sum_{k=1}^{K} n_{jk}(n_k - 1).$$

Thus, the specific agreement index to a category j is the total number of agreements on category j

divided by the total number of opportunities for agreement on category j,

$$Ps_j = \frac{S_j}{Sposs_j}.$$

The overall agreement index, the proportion of observed agreement regardless of category, (P_O) is the total number of actual agreements, $O = \sum_{j=1}^{C} S_j$, divided by the total number of possible agreements, $O_{poss} = \sum_{k=1}^{K} n_k(n_k - 1)$,

$$P_O = \frac{O}{O_{poss}}.$$

Another description and examples can be found in [5].

IBMD, was proposed in [11] based on Shannon's notion of entropy [19]. Its main intuition is to consider the logarithm of the differences between two observations as a measure of the disagreement between the observers. This measure is normalized and satisfies the following properties: it is a metric, scaled invariant with differential weighting.

The IBMD measures the disagreement among measurements obtained by several observers, allowing different number of observations in each case. It is appropriate for ratio-scale variables with positive values and ranges from 0 (no disagreement) to 1. Consider N as the number of cases, and $M_i, i = 1, \cdots, N$, as the number of observations in case i. Therefore considering N vectors, one for each case,$(x_{11}, \cdots, x_{1M_1}), \cdots, (x_{N1}, \cdots, x_{NM_N})$ with non-negative values, the generalized information-based measure of disagreement is defined as:

$$IBMD = \frac{1}{\sum_{i=1}^{N} \binom{M_i}{2}} \sum_{i=1}^{N} \sum_{j=1}^{M_i-1} \sum_{k=j+1}^{M_i} log_2\left(\frac{|x_{ij} - x_{ik}|}{max\{x_{ij}, x_{ik}\}} + 1\right) \tag{3.3}$$

with the convention $\frac{|0-0|}{max\{0,0\}} = 0$. This coefficient is 0 when the observers agree or when there is no disagreement, and increases to 1 when the distance, i.e. disagreement among the observers, increases.

The standard error and confidence interval was based on the nonparametric bootstrap, by resampling the subjects/cases with replacement, in both original and generalized IBMD measures. The bootstrap uses the data from a single sample to simulate the results as if new samples were repeated over and over. Bootstrap samples are created by sampling with replacement from the dataset. A good approximation of the 95% confidence interval can be obtained by computing the $2.5th$ and $97.5th$ percentiles of the bootstrap samples. Nonparametric resampling makes no assumptions concerning the distribution of the data. The algorithm for a nonparametric bootstrap is as follows [20]:

1. Sample N observations randomly with replacement from the N cases to obtain a bootstrap data set.

2. Calculate the bootstrap version of IBMD.

3. Repeat steps 1 and 2 B times to obtain an estimate of the bootstrap distribution.

For confidence intervals of $90 - 95$ percent B should be between 1000 and 2000 [21, 22]. In the package, the confidence interval is calculated, by default, with B equal to 1000.

Results

Disagreement website

At the initial web-page the user can insert the data either directly in the specific box or upload a txt file. Each tuple of observations must be in a different row and the observations can be space or tab-delimited, an example is given in order to facilitate this process. After the data is introduced the user can choose which measure to calculate besides the IBMD, namely he can choose between One-way ICC, Two-way ICC, CCC and Bland and Altman. Notice that, if the data is not fitted to a given measure its name will appear strikethrough and will not be given as a valid option. The user can press the calculate button, the data is then sent to the backend, included in the request will be the calculators to be applied to the data. The data goes through some preprocessing to ensure it's compatible with the calculators indicated in the request. If the data is confirmed to be valid it is written to a temporary file. Any calculator that rejects the data will output an error message that explains why the data was not valid. If the data is valid the temporary file is used in a script dynamically generated that calls the R implementation of the calculator. Then, finally, the result of the R routines implementing each measure will be presented in the standard R format.

obs.agree package

The observer agreement package, **obs.agree** , includes two functions for measuring agreement: RAI to categorical data and IBMD to continuous data. It can be used for multiple raters and multiple readings cases. As required for any package in the Comprehensive R Archive Network the **obs.agree** [17] package includes documentation with some examples on how to calculate the measures.

Categorical data: cardiotocographic traces classification

The package **obs.agree** includes an artificial data set `ctg`, in the form of a matrix containing 151 cardiotocographic traces classified as patologic (3), suspect (2) or normal (1) by 18 clinicians with different level of experience: 6 interns (E1), 6 clinicians (E2) and 6 experts (E3). Nine of them (3 interns, 3 clinicians and 3 experts) classified the traces based on a guideline (GL1) different from the other nine (GL2).

The first step is to load the package.

```
R> library(obs.agree)
```

Then load the dataset `ctg`.

```
R> data(ctg)
```

Each row of the matrix corresponds to a cardiotocographic trace classified by 18 clinicians (columns of the matrix). The first 9 observers classified according to guideline 1 (GL1) and the last 9 according to guideline 2 (GL2).

```
R> colnames(ctg)
[1] "GL1_E1_O1" "GL1_E1_O2" "GL1_E1_O3" "GL1_E2_O1" "GL1_E2_O2" "GL1_E2_O3"
[7] "GL1_E3_O1" "GL1_E3_O2" "GL1_E3_O3" "GL2_E1_O1" "GL2_E1_O2" "GL2_E1_O3"
[13] "GL2_E2_O1" "GL2_E2_O2" "GL2_E2_O3" "GL2_E3_O1" "GL2_E3_O2" "GL2_E3_O3"
```

To assess if following one guideline the clinicians agree more than with the other, just compute the RAI to each guideline.

```
 R> RAI(ctg[,1:9]) #Guideline 1

$Subjects
[1] 151
$Observers
[1] 9
$Overall_agreement
       value     ( 2.5% - 97.5% )
1 0.5123596 0.4809971 0.5466795
$Categories
[1] 1 2 3
$Specific_agreement
       value     ( 2.5% - 97.5% )
1 0.5197044 0.4504645 0.5867110
2 0.4115406 0.3714993 0.4531956
3 0.6045769 0.5477483 0.6605998

R> RAI(ctg[,10:18]) #Guideline 2

$Subjects
[1] 151
$Observers
[1] 9
$Overall_agreement
       value    ( 2.5 % - 97.5 % )
1 0.6196469 0.5858345 0.6535072
$Categories
[1] 1 2 3
$Specific_agreement
       value    ( 2.5 % - 97.5 % )
1 0.6055118 0.5325463 0.6689757
2 0.6183014 0.5664239 0.6677420
3 0.6342293 0.5643657 0.6952517
```

The results show that, considering a significance level of 5%, with guideline 2 there is significantly more overall observer agreement than with guideline 1. With guideline 2 there is also significantly more observer agreement in the category 2 (suspect) than with guideline 1. In other words, with guideline 1 if an observer rates a tracing as suspect the probability that another observer will also rate that tracing as suspect is 0.62 with a 95% confidence interval between 0.57 and 0.67. With guideline 2, if an observer rates a tracing as suspect, the probability that another observer will also rate that tracing as suspect is lower (0.41 with a 95% confidence interval between 0.37 and 0.45).

We can also examine if, using guideline 1, the experts agree more than the interns do. In this case, consider a 90% confidence interval.

```
R> RAI(ctg[,7:9],0.9) #experts using GL1

$Subjects
[1] 151
$Observers
[1] 3
$Overall_agreement
       value      ( 5 % - 95 % )
1 0.5122494 0.4657684 0.55902
$Categories
[1] 1 2 3
$Specific_agreement
       value      ( 5 % - 95 % )
1 0.5500000 0.4691222 0.6271914
2 0.3754045 0.3154125 0.4342984
3 0.6148867 0.5365760 0.6790361

R> RAI(ctg[,1:3],0.9) #interns using GL1

$Subjects
[1] 151
$Observers
[1] 3
$Overall_agreement
       value      ( 5 % - 95 % )
1 0.5122494 0.4676841 0.5607199
$Categories
[1] 1 2 3
$Specific_agreement
       value      ( 5 % - 95 % )
1 0.5865922 0.5150606 0.6503687
```

```
2 0.3131673 0.2462568 0.3751591
3 0.6254826 0.5453785 0.6963772
```

These results show that, considering a significance level of 10%, with Guideline 1 the observer agreement among experts is not significantly different from the observer agreement among interns in the tracings classified as normal, suspect or pathological.

Continuous data: Gymnasts new rulebook

Consider the following situation: a jury evaluates a gymnast's performance according to a given rulebook. Suppose that a new rulebook has been recently proposed and subsequently criticized. Some gymnasts and media argue that disagreement among the jury members in evaluating the gymnastics performance with the new scoring system, or rulebook, is higher than that with the old scoring system, and therefore oppose to its use. The scores used to assess the gymnastics performance are a continuous variables. The package **obs.agree** includes a data set `gymnasts` with the scores of 20 gymnasts evaluated, by a set of 8 judges, according to the old rulebook and the scores of other 20 gymansts evaluated by the same 8 judges but using the new rulebook. Using IBMD, one can compare observer disagreement and the respective confidence interval (CI) associated with each score system.

We begin by loading the package.

```
R> library(obs.agree)
```

Then load the dataset gymnasts which has 40 gymnasts and 8 judges.

```
R> data(gymnasts)
```

Data should be in a data frame with 40 rows and 9 columns with column names "rulebook", "judge1",..., "judge8".

```
R> head(gymnasts)
```

	Rulebook	Judge.1	Judge.2	Judge.3	Judge.4	Judge.5	Judge.6	Judge.7	Judge.8
1	Old	7.1	7.2	7.0	7.7	7.1	7.1	7.0	7.0
2	Old	9.3	9.7	8.9	9.6	8.6	9.5	9.6	9.7
3	Old	8.9	8.8	8.1	9.3	8.5	8.1	7.6	8.7
4	Old	8.0	8.1	7.3	8.7	7.5	8.7	7.4	9.5
5	Old	9.1	9.0	8.2	9.0	8.2	9.5	7.8	8.0
6	Old	9.1	9.2	8.3	9.1	7.9	8.9	9.0	9.2

Call the function in its simplest form `IBMD(data)`. In this case the confidence interval will be, by default, 95%. To compare the two rulebooks run:

```
R> IBMD(gymnasts[1:20,2:9])
$subjects
[1] 20
$observers
[1] 8
$IBMD
        value     ( 2.5 % - 97.5 % )
1 0.09269632 0.07838681 0.1064475

R> IBMD(gymnasts[21:40,2:9])
$subjects
[1] 20
$observers
[1] 8
$IBMD
       value   ( 2.5 % - 97.5 % )
1 0.1740529 0.1526705 0.1923382
```

The gymnasts' opposition to the new scoring systems was supported by the results; actually the disagreement among judges was significantly higher when the judges use the new scoring system to evaluate the gymnast's performance than when the judges use the old scoring system.

Discussion

Often health researchers need to assess the agreement among multiple measurements for the same subject by different observers under similar circumstances, or a new method with an existing one, to evaluate if these methods can be used interchangeably or the new method can replace the established one. Bland and Altman proposed the use of a graphical method to plot the difference scores of two measurements against the mean for each subject. The intuition is that if the new method agrees sufficiently well with the old, the old may be replaced. There are several published clinical studies evaluating agreement between two measurements using Bland and Altman analysis. The original Bland and Altman paper has more than 11 500 citations, a convincing evidence of its importance in medical research. The main reason for this impact is its simplicity of evaluation and interpretation. However, problems appear when comparing the degree of observer agreement among different methods, populations or circumstances. To overcome these problems the raw agreement indices (RAI) to categorical data and information-based measure of disagreement (IBMD) measures have been introduced. Nevertheless, the statistical strategies for comparative analyses are not straightforward and software for RAI and IBMD assessment is lacking.

The website and package introduced in this paper have the potential to help health care professionals and biostatisticians when performing observer agreement studies, as it provides an easy way to calculate raw agreement indices to categorical data and information-based measure of disagreement to continuous variables. Given its simplicity we anticipate that it may also become an important and widely used tool within health researchers.

Conclusion

Comparison of the degree of observer disagreement is often required in clinical and epidemiologic studies. However, the statistical strategies for comparative analyses are not straightforward and software for RAI and IBMD assessment is lacking. Both RAI and IBMD provide useful tools to compare observer disagreement among different methods, populations or circumstances and allow evaluation of data by different numbers of observers for different cases, an important feature in real situations where some data are often missing. RAI are a very important descriptive statistics, composed by two type of measures: the overall proportion of agreement and the proportions of specific agreement for each rating categories. The IBMD measure, recently introduced in [11] and generalized in [12], is based on Shannon's notion of entropy [19] considering the logarithm of the differences between two observations as a measure of the disagreement between the observers.

The website and package have the potential to help health care professionals and biostatisticians when performing observer agreement studies, as it provides an easy way to calculate raw agreement indices to categorical data and information-based measure of disagreement to continuous variables.

Bibliography

[1] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1):37, 1960.

[2] Anthony J. Viera and Joanne M. Garrett. Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine, 37(5):360–363, May 2005.

[3] Alvan R. Feinstein and Domenic V. Cicchetti. High agreement but low kappa: I. the problems of two paradoxes. Journal of Clinical Epidemiology, 43(6):543 – 549, 1990.

[4] John S. Uebersax. Diversity of decision-making models and the measurement of interrater agreement. Psychological Bulletin, 101(1):140–146, 1987.

[5] John S. Uebersax. Statistical methods for rater and diagnostic agreement. `http://www.john-uebersax.com/stat/raw.htm`. Accessed: Jun 2014.

[6] J. Martin Bland and Douglas G. Altman. Agreement between methods of measurement with multiple observations per individual. Journal of Biopharmaceutical Statistics, 17(4):571–582, 2007.

[7] Philip Schluter. A multivariate hierarchical bayesian approach to measuring agreement in repeated measurement method comparison studies. BMC Medical Research Methodology, 9(1):6, 2009.

[8] H. De Vet. Observer Reliability and Agreement. John Wiley & Sons, Ltd, 2005.

[9] Reinhold Muller and Petra Buttner. A critical discussion of intraclass correlation coefficients. Statistics in Medicine, 13(23-24):2465–2476, 1994.

[10] Greg Atkinson and Alan Nevil. Comment on the use of concordance correlation to assess the agreement between two variables. Biometrics, 53:775–777, 1997.

[11] Cristina Costa-Santos, Luís Antunes, André Souto, and João Bernardes. Assessment of disagreement: A new information-based approach. Annals of Epidemiology, 20(7):555 – 561, 2010.

[12] Teresa Henriques, Luís Antunes, João Bernardes, Mara Matias, Diogo Sato, and Cristina Costa-Santos. Information-based measure of disagreement for more than two observers: a useful tool to compare the degree of observer disagreement. BMC Medical Research Methodology, 13(1):1–6, 2013.

[13] Cristina Costa-Santos. Disagreement. `http://www.dcc.fc.up.pt/disagreement/`. Accessed: Oct 2014.

[14] Bootstrap. `http://getbootstrap.com/`. Accessed: Oct 2014.

[15] Smarty: Template engine. `http://www.smarty.net/`. Accessed: Oct 2014.

[16] R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria. `http://www.R-project.org/`. Accessed: Oct 2014.

[17] Teresa Henriques, Luís Antunes, and Cristina Costa-Santos. obs.agree: An R package to assess agreement between observers. R package version 1.0. `http://CRAN.R-project.org/package=obs.agree`. Accessed: Oct 2014.

[18] Joseph L. Fleiss, Bruce Levin, and Myunghee C. Paik. Statistical Methods for Rates and Proportions. Wiley Series in Probability and Statistics. Wiley, 2003.

[19] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[20] James Carpenter and John Bithell. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. Statistics in Medicine, 19(9):1141–1164, 2000.

[21] Bradley Efron and Robert J. Tibshirani. An Introduction to the Bootstrap. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1994.

[22] Anthony C. Davison and David V. Hinkley. Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

Complexity Measures Applied to Heart Rate Signal

The only way to control chaos and complexity is to give up some of that control

Gyan Nagpal, Talent Economics: The Fine Line Between Winning and Losing the Global War for Talent

4. Complexity Measures Applied to Heart Rate Signal

Conventional linear mathematical methods fail to describe biological systems belonging to the complex and chaotic domains. Different approaches based on chaos and complexity concepts have been successfully considered. Shannon demonstrated how the information within a signal could be quantified with absolute precision as the amount of unexpected data contained in the message (designated entropy) [1]. On the other hand, Kolmogorov complexity was proposed to quantify information on individual objects as the size of its smallest representation [2]. In this chapter we explore the applicability of these two conceptually different complexity measures: entropy, a probabilistic approach and compression, an algorithmic approach. The entropy using the approximate entropy (ApEn) [3] and sample entropy (SampEn) [4] and the Kolmogorov complexity through different compressors, as the Lempel-Ziv [5], bzip2 [6] and paq8l [7], were applied to a set of heart rate signals with the goal of characterizing different pathologies.

The first section of this chapter addresses the challenge of assessing the dynamics of fetal heart rate (FHR) signals from fetuses born with neonatal acidemia using nonlinear indices. The analysis of a FHR signal represents a noninvasive tool for evaluating the fetal condition. The information obtained may provide important auxiliary information to clinicians in the diagnostic of the fetal state and help in consequent clinical decisions in life-threatening situations, particularly in the minutes prior to delivery. It is shown in this section that compressors can be effectively used as an alternative (or complementarily) to the widely used entropy measures to quantify complexity in biological signals.

The complex fluctuations present on physiological signals contain information about the underlying dynamics across multiple spatial and temporal scales. The Multiscale Entropy (MSE) [8] approach was proposed to enhance these multiscale features of a time series. Briefly, the multiscale entropy comprise two steps: 1) the creation of a new time series for each scale through a coarse-grained procedure applied to the original time series; 2) the calculation of the entropy value of each of the coarse-grained time series. In the second section of this chapter an extension of the multiscale approach to compression is proposed. This new approach maintains the coarse-grained procedure but replace the latter step by a compression measure.

The low correlation between the two complexity approaches obtained in the results of the two sections previously described, suggest that each one of them is looking for different patterns/behavior. On the third section of this chapter the complexity is explored from different viewpoints and the properties of the different approaches is investigated.

Of note, the work developed in this chapter resulted in two published articles.

Bibliography

[1] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[2] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information. Problems of Information Transmission, 1(1):1–7, 1965.

[3] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[4] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[5] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. Information Theory, IEEE Transactions on, 23(3):337–343, 1977.

[6] Julian Seward. bzip2 and libbzip2. url=http://www.bzip.org. Accessed: Jun 2014.

[7] Matt Mahoney. Data compression programs. `http://mattmahoney.net/dc/index.html`. Accessed: Jun 2014.

[8] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

Entropy And Compression: Two Measures Of Complexity

Teresa Henriques, Hernâni Gonçalves, Luís Antunes,
Mara Matias, João Bernardes and Cristina Costa-Santos

Journal of Evaluation in Clinical Practice
Volume 19, Issue 6, Pages 1101–1106
December 2013

4.1 Entropy And Compression: Two Measures Of Complexity

Abstract

Rationale, aims and objectives Traditional complexity measures are used to capture the amount of structured information present in a certain phenomenon. Several approaches developed to facilitate the characterization of complexity have been described in the related literature. Fetal heart rate (FHR) monitoring has been used and improved during the last decades. The importance of these studies lies on an attempt to predict the fetus outcome, but complexity measures are not yet established in clinical practice. In this study, we have focused on two conceptually different measures: Shannon entropy, a probabilistic approach, and Kolmogorov complexity, an algorithmic approach. The main aim of the current investigation was to show that approximation to Kolmogorov complexity through different compressors, although applied to a lesser extent, may be as useful as Shannon entropy calculated by approximation through different entropies, which has been successfully applied to different scientific areas.

Methods To illustrate the applicability of both approaches, two entropy measures, approximate and sample entropy, and two compressors, paq8l and bzip2, were considered. These indices were applied to FHR tracings pertaining to a dataset composed of 48 delivered fetuses with umbilical artery blood (UAB) pH in the normal range (pH $\geq$ 7.20), 10 delivered mildly acidemic fetuses and 10 moderate-to-severe acidemic fetuses. The complexity indices were computed on the initial and final segments of the last hour of labour, considering 5- and 10-minute segments.

Results In our sample set, both entropies and compressors were successfully utilized to distinguish fetuses at risk of hypoxia from healthy ones. Fetuses with lower UAB pH presented significantly lower entropy and compression indices, more markedly in the final segments.

Conclusions The combination of these conceptually different measures appeared to present an improved approach in the characterization of different pathophysiological states, reinforcing the theory that entropies and compressors measure different complexity features. In view of these findings, we recommend a combination of the two approaches.

Introduction

Researchers and clinicians recognize that many unsolved medical problems, as the prediction of fetal outcome, are due to the application of conventional mathematics methods to describe biological complex systems [1]. More recently, a different approach based on non-linear dynamics, chaos and complexity has been considered, which recognizes irregularity, subjectivity and uncertainty as intrinsic and fundamental [2].

Complexity is a property of every system that quantifies the amount of structured information. Shannon demonstrated how the information within a signal could be quantified with absolute precision [3] as the amount of unexpected data contained in the message (designated 'entropy'). Subsequently, the Kolmogorov complexity was proposed to quantify information on individual objects as the size of its

smallest representation [4]. The Shannon information theory measures the average information from a random source, unlike Kolmogorov complexity that presents a form of absolute information [5].

Increasing use of entropy in medicine has accompanied theoretical improvements over the years. In 1991, Pincus suggested the use of Approximate Entropy (ApEn) to classify complex systems [6]. Following its application to quantify the creation of information in a time series, ApEn has since been used in the analysis of a wide range of signals, such as electroencephalography (EEG) and electrocardiography (ECG). The sample entropy (SampEn) concept appeared later in 2000 with the objective of reducing the ApEn bias [7], and was subsequently employed in the analysis of biomedical signals. In 2004, Ohmeda [8] developed an anaesthesia EEG analyser with an entropy module that calculates the characteristics of the biosignal with analysis of time-frequency balanced spectral entropy. The index, based on EEG and electromyography (EMG) activity (Entropy), is an indicator of the hypnotic effects of propofol, thiopental, isoflurane, sevoflurane and desflurane [9]. In 2002, Costa et al. [10] proposed the multiscale entropy (MSE) technique applicable in the analysis of physiologic time series.

Compression is a measure of system complexity, but has been used to a lower extent in the analysis of biomedical signals. Many types of compressors exist. Traditionally, the different available compressors are divided in two classes, specifically, lossless and lossy compressors. With lossless compression, every bit of data originally in the file remains after the file is uncompressed, whereas in lossy compression, the file is reduced by permanently eliminating certain information, particularly redundant information. The Lempel-Ziv is a compression algorithm introduced in 1977 [11], used as a measure of complexity in EEG and ECG. Jean-Loup Gailly and Mark Adler subsequently developed gzip [12], which is a combination of Lempel-Ziv and Huffman coding [13]. Julian Seward developed bzip2 using the Burrows–Wheeler block sorting text compression algorithm[14] and Huffman coding. PAQ is a series of lossless data compressors that uses a context-mixing algorithm. PAQ8L was released on 8 March 2007 by Matt Mahoney [15].

Cardiotocography is a technical means of recording the heart rate of the fetus (FHR) and uterine contractions, and is widely used for fetal monitoring during labour. The value of cardiotocography lies in its ability to predict newborn outcomes. Timely prediction of newborn outcome in the intrapartum period, that is, immediately before delivery, may lead to a decrease in perinatal mortality and morbidity. The FHR is measured in beats per minute and an external sensor attached on the mother's abdomen acquires it. FHR alterations are used to evaluate the fetal health condition and allow the obstetrician to intervene to prevent potential compromise and irreversible damage. However, studies have shown significant differences in inter- and intra-observer in FHR analysis and interpretation [16–18]. On the other hand, new signal processing and pattern recognition techniques have paved the way towards automated approaches.

One of the most widely accepted measures of newborn outcome is umbilical artery blood (UAB) pH, as it represents an active measure of fetal oxygenation. A low UAB pH indicates the presence of acidemia occurring during labour and delivery, presenting higher risk of perinatal death or neurological injuries from hypoxia [19].

In the particular case of fetal heart rate (FHR) tracings, several entropy measures have been widely used to detect different pathologies. ApEn and SampEn statistics are the most used entropy metrics, followed more recently by MSE. Despite the successful application of Lempel-Ziv and gzip compressors to FHR tracings to detect pathologies, compressors have been used only to a limited extent in the analysis of biological signals to date [20, 21].

In the current study, we aimed to show that compression can be effectively applied as an alternative measure of complexity to the widely used entropy in biological signals. An example of FHR tracings was applied to demonstrate the utility of these non-linear indices [22].

Methods

Complexity is a property of every system that quantifies the amount of structured information. The quantification of information within a signal may be achieved using entropy or compression [7–9]. In this section, we briefly describe the entropy and compression approaches and their related indices as well as the dataset composed by FHR tracings used to evaluate the two approaches.

Entropy

In 1948, Shannon introduced the first approach for measuring 'information' [3]. This new measure, known as Entropy (Shannon entropy), attempts to determine how random a message is expected to be within a given distribution, and takes into account the minimum number of bits to transmit a message from a random source of known characteristics through an error-free channel (Grunwald, P. D. & Vitanyi, P. M. B unpublished data).

Let X be a random variable, taking values in Y with distribution $P(X = x) = p_x$. The Shannon entropy of the random variable X is given by

$$H(X) = \sum_{x \in Y} p_x log \frac{1}{p_x}$$

The logarithm is base 2, so that entropy is measured in bits and $\lim_{x \to 0} logx = 0$; and thus, traditionally it is conventioned that $0log0 = 0$.

With the goal of 'quantification of the amount of regularity in (heart rate) time-series data', Pincus presented the ApEn statistic in 1991 [6]. It can be estimated through the family of statistics $ApEn(N, m, r)$, given N points, and is approximately equal to the negative average natural logarithm of the conditional probability that two sequences similar for m points remain similar within tolerance, r, at the next point. Accordingly, a low ApEn value is associated with high degree of regularity.

SampEn was introduced in 2000 with the same objective as ApEn by Richman and Moorman [7] to evaluate biological time series, particularly heart rate. The authors highlighted two draw-backs in ApEn properties, stating that 'First, ApEn is heavily dependent on the record length and uniformly lower than expected for short records. Second, it lacks relative consistency. That is, if ApEn of one dataset is higher than that of another, it should, but does not remain higher for all conditions tested' [7]. In order to overcome these limitations, the group proposed a new family of statistics, SampEn (m, r), which, among other differences, eliminates self-matches.

Regarding the necessary parameters to the evaluation of ApEn and SampEn, Pincus and Goldberger [23] concluded that $m = 2$, r values between 0.1 and 0.25 of the standard deviation and N value of 10^m, or preferably 30^m, will yield statistically reliable and reproducible results. In the particular case of FHR analysis, segments of 5 or 10 minutes (i.e. 600 or 1200 points for 2 Hz sampling) are adequate and can make a difference in sensitivity and specificity [22], since in smaller segments some patterns may not be detected, whereas in larger segments some patterns may get lost.

Compression

In 1958 [24], Kolmogorov connected the dynamic systems theory with information theory. Based on his work, Sinai [25] introduced the Kolmogorov–Sinai entropy that generalizes the probabilistic definition of entropy, allowing the use of entropy in time-series data.

The 'algorithm information theory' was proposed years later independently by Solomonoff in 1960/1964 [26, 27], Kolmogorov in 1965 [4] and Chaitin in 1969 [28], which considers the amount of information in an object as the length of the smallest description of the object. The Kolmogorov complexity attempts to answer how 'random' an individual object is in terms of the number of bits necessary to describe it.

In [29] Kolmogorov claimed 'The ordinary definition of entropy uses probability concepts, and thus does not pertain to individual values, but random values, i.e. probability distributions within a given group of values'.

The Kolmogorov complexity of a finite object is defined as the length of its shortest effective description (Gr̈unwald, P. D. & Vitányi, P. M. B unpublished data). This measure is not computable, but can be easily approximated using compressors. Good results were obtained using this approach in different areas, including languages tree, genomics, optical character recognition, literature [30], music [31], computer virus and Internet traffic analysis [32] and FHR anomaly detection [21]. In fact, Keogh [33] showed that when clustering heterogeneous data and anomaly detection in time sequences, the compression approach outperforms every known data-mining method.

In 1952, Huffman [13] developed an algorithm to use a short bit stream for characters that appear more often. In 1977, Ziv and Lempel introduced the Lempel-Ziv algorithm based on 'the concept of encoding future segments of the source output via maximum-length copying from a buffer containing the recent past output' [11]. Recently, the Lempel-Ziv algorithm has been widely used in the field of medicine. However, numerous compressors can be used with the same purpose.

The main point of data compression is the encoding of information using fewer bits than the original data. Gailly and Adler created the first version of gzip [12], representing a combination of Lempel-Ziv and Huffman coding. Seward developed bzip2 [14] in 1996, which was more effective than Lempel-Ziv and gzip, but considerably slower. A similar compressor was successfully used in 2006 to cluster FHR [21]. PAQ8 [15] represents a series of lossless compressors with the world's highest compression ratio. PAQ8L, based on Dynamic Markov compression [34], was released in 2007. We believe that these compressors can be successfully used in the medical field as well.

Example of fetal heart rate

Most clinical decisions in the intrapartum period, in countries with the best maternal and perinatal indicators, are strongly based on FHR monitoring [16, 18, 19]. However, conventional visual analysis of FHR recordings suffers from unacceptable intra- and inter- observer variation [17–19]. To overcome this shortcoming, computer systems for bedside analysis of FHR recordings have been developed [35]. These systems have provided important progresses in FHR monitoring, but there is still room for their improvement, namely using methods for complexity analysis of FHR variability [20, 22], which remains one of the most challenging tasks in the analysis of FHR recordings [17, 19, 36].

In this study, we analysed 68 FHR intrapartum traces consecutively selected from a pre-existing database of term singleton gestations, with at least 60 minutes of tracing. Of the 68 cases, 48 delivered

Table 4.1.1: Median, first quartile (Q_1) and third quartile (Q_3) of complexity measures of fetal heart rate tracings from moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses in the final 5-minute segments

	MSA		MA		N		
	Median	(Q_1, Q_3)	Median	(Q_1, Q_3)	Median	(Q_1, Q_3)	P
Entropy							
ApEn(2,0.1)	0.585	(0.525, 0.733)	0.738	(0.686, 0.774)	0.682	(0.555, 0.739)	0.207
ApEn(2,0.15)	0.496	(0.291, 0.738)	0.642	(0.561, 0.792)	0.607	(0.490, 0.702)	0.304
ApEn(2,0.2)	0.351	(0.251, 0.553)	0.582	(0.469, 0.795)	0.516	(0.420, 0.627)	**0.044**
SampEn(2,0.1)	0.476	(0.325, 0.658)	0.598	(0.540, 0.985)	0.541	(0.402, 0.615)	0.149
SampEn(2,0.15)	0.309	(0.172, 0.636)	0.459	(0.403, 0.632)	0.434	(0.320, 0.549)	0.338
SampEn(2,0.2)	0.231	(0.172, 0.307)	0.369	(0.308, 0.637)	0.341	(0.256, 0.404)	**0.036**
Compression							
paq8l	234.0	(211.0, 279.0)	355.0	(306.0, 393.0)	335.0	(293.5, 372.5)	**0.009**
bzip2	283.5	(270.0, 382.0)	444.0	(404.0, 501.0)	426.5	(362.5, 488.0)	**0.017**

Further details on the complexity measures may be found in the Entropy and Compression sections. Boldfaced numerals correspond to P values < 0.05.

Table 4.1.2: Median, first quartile (Q_1) and third quartile (Q_3) of complexity measures of fetal heart rate tracings from moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses in the final 10-minute segments

	MSA		MA		N		
	Median	(Q_1, Q_3)	Median	(Q_1, Q_3)	Median	(Q_1, Q_3)	P
Entropy							
ApEn(2,0.1)	0.533	(0.376, 0.611)	0.837	(0.727, 0.930)	0.751	(0.604, 0.877)	**0.003**
ApEn(2,0.15)	0.392	(0.298, 0.541)	0.687	(0.667, 0.907)	0.613	(0.506, 0.739)	**0.002**
ApEn(2,0.2)	0.328	(0.243, 0.403)	0.593	(0.553, 0.662)	0.516	(0.408, 0.613)	**0.001**
SampEn(2,0.1)	0.342	(0.234, 0.397)	0.682	(0.534, 0.815)	0.572	(0.375, 0.678)	**0.009**
SampEn(2,0.15)	0.215	(0.152, 0.316)	0.498	(0.425, 0.770)	0.422	(0.286, 0.522)	**0.002**
SampEn(2,0.2)	0.194	(0.114, 0.238)	0.398	(0.349, 0.489)	0.317	(0.226, 0.406)	**0.002**
Compression							
paq8l	511.0	(452.0, 537.0)	648.5	(570.0, 692.0)	592.0	(515.5, 647.5)	**0.019**
bzip2	658.5	(616.0, 719.0)	814.0	(726.0, 850.0)	716.0	(667.5, 814.0)	**0.032**

Further details on the complexity measures may be found in the Entropy and Compression sections. Boldfaced numerals correspond to P values < 0.05.

fetuses with UAB pH in the normal (N) range (pH $\geq$ 7.20), 10 delivered with UAB pH between 7.10 and 7.20, mildly acidemic (MA) fetuses and 10 moderate-to-severe acidemic (MSA) fetuses with UAB pH $\leq$ 7.10. All traces were resampled at a frequency of 2 Hz after pre-processing, based on an algorithm described in previous studies. A more detailed description of the data and pre-processing algorithm is presented in [22].

Results

The non-linear methods, entropy (ApEn and SampEn) and compression (paq8l and bzip2), were calculated in the first and last 5 and 10 minutes of the tracings. The median and interquartile (IQ) range, as well as the statistical significance of the non-parametric Kruskal–Wallis test, are presented in Tables 4.1.1 and 4.1.2.

The mean running times for entropy (ApEn), paq8l compressor and bzip2 compressor in 10-minute segments were 4.2, 0.9 and 0.006 seconds, while those in 5-minute segments were 1.2, 0.5 and 0.003 seconds, respectively.

Table 4.1.3: Comparison of the complexity measures computed in the initial and final minutes of fetal heart rate tracings with respect to moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses

	Initial versus final 5 minutes			Initial versus final 10 minutes		
	MSA	MA	N	MSA	MA	N
Entropy						
ApEn(2,0.1)	0.074	**0.028**	**<0.001**	**0.050**	**0.017**	**<0.001**
ApEn(2,0.15)	0.386	0.203	**<0.001**	**0.028**	0.386	**<0.001**
ApEn(2,0.2)	0.285	0.386	**<0.001**	**0.017**	**0.028**	**<0.001**
SampEn(2,0.1)	0.059	0.114	**<0.001**	**0.005**	**0.009**	**<0.001**
SampEn(2,0.15)	0.285	0.169	**<0.001**	**0.005**	0.203	**<0.001**
SampEn(2,0.2)	0.203	0.203	**<0.001**	**0.005**	**0.022**	**<0.001**
Compression						
paq8l	0.114	0.878	0.272	0.878	0.541	**0.024**
bzip2	0.575	0.333	**0.001**	0.241	0.139	**<0.001**

Values provided in the table correspond to P values of the Mann–Whitney test.
Further details on the complexity measures may be found in the Entropy and Compression sections.
Boldfaced numerals correspond to P values < 0.05.

Considering the outcome MSA and the last 5-minute segments, entropy measures tend to have less discriminatory power (lower areas under the ROC curve ranging from 0.584 to 0.705) than compressors (areas under the ROC curve between 0.773 and 0.797). Considering the last 10-minute segments, entropy measures tend to have greater discriminatory power (higher areas under the ROC curve ranging from 0.797 to 0.855) than compressors (areas under the ROC curve between 0.676 and 0.699).

A comparison between the initial and final 5 minutes of FHR tracings was performed for each complexity measure and each group of fetuses (MSA, MA and N). The same procedure was performed for comparison between the initial and final 10 minutes of FHR tracings, and significance results of the non-parametric Mann-Whitney test are presented in Table 4.1.3.

As expected, SampEn was highly correlated with ApEn, with the correlation coefficient ranging between 0.928 and 0.955 (for similar values of r). Moreover, the two compressors were highly correlated with a correlation coefficient of 0.888. However, the correlation values between the considered compression and entropy metrics were lower, ranging between 0.316 and 0.414.

Figure 4.1.1 depicts a non-linear relationship between entropy (ApEn) and compression measures (paq8l). The correlation coefficient between ApEn (2,0.15) and paq8l for all cases was 0.414, and the same measure for the MSA, MA and N groups were 0.648, 0.219 and 0.214, respectively. Similar results were obtained with other possible combinations in the final 5- and 10-minute segments. Despite both being complexity measures, the poorcorrelation between them shows that entropy and compression are using algorithms that seek and emphasize different characteristics and patterns of each time series. Therefore, the combination of both measures may improve the classification of a FHR tracing.

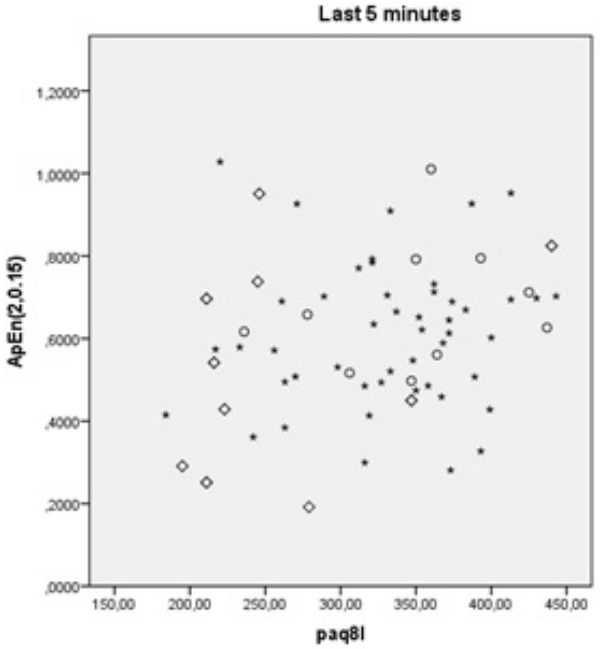

Figure 4.1.1: Scatterplot of indices ApEn(2,0.15) and paq8l for the final 5-minute segments, comparing normal fetuses ($*$), mildly acidemic fetuses ($\bullet$) and moderate-severe acidemic fetuses ($\diamond$).

Conclusions

We have shown that entropy and compression measures allow for the characterization of different pathophysiological conditions of the fetus - distinguishing fetuses at risk of hypoxia from their healthy counterparts and between different stages of labour - through the analysis of the FHR signal.

The use of compression as a measure of complexity has rarely been applied in the analysis of biological signals. However, we have shown that compressors can be effectively used as an alternative to the widely used entropy measures to quantify complexity in biological signals.

By using entropy and compression approaches, one can quantify different features of a system complexity, as shown by the low/moderate correlations between the entropy and compression measures. Accordingly, further research is required in order to study the different complexity features captured by either entropy or compression, as well as in exploring different combinations of the two strategies.

The small computational time associated with both measures, particularly the compressors, allows for considering their inclusion in existing FHR monitoring systems. In this way, the information on the fetus complexity obtained from the FHR signal may provide important auxiliary information to clinicians, in the diagnostic of the fetal state and help in consequent clinical decisions.

Acknowledgements

This work was supported by the National Science Foundation, Fundação para a Ciência e Tecnologia (FCT), through FEDER funds through Programa Operacional Factores de Competitividade - COMPETE through the project CSI2 with the reference PTDC/ EIA-CCO/099951/2008, through the project with the reference PEST-C/SAU/UI0753/2011 and through the PhD grant with the reference

SFRH/BD/70858/2010. Hernâni Gonçalves is financed by a post-doctoral grant (SFRH/BPD/69671/2010) from FCT, Portugal.

Bibliography

[1] Colin J. Alexander. Complexity and Medicine: The Elephant in the Waiting Room. Nottingham University Press, 2010.

[2] Vivian S. Rambihar. Science, evidence, and the use of the word scientific. The Lancet, 355(9216):1730, 2000.

[3] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[4] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information'. Problems of Information Transmission, 1(1):1–7, 1965.

[5] Ming Li and Paul Vitâanyi. An introduction to Kolmogorov complexity and its applications. Springer, 2008.

[6] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[7] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[8] Hanna Viertiö-Oja, V Maja, M Särkelä, P Talja, N Tenkanen, H Tolvanen-Laakso, M Paloheimo, A Vakkuri, A Yli-Hankala, and P Meriläinen. Description of the entropy algorithm as applied in the datex-ohmeda s/5 entropy module. Acta Anaesthesiologica Scandinavica, 48(2):154–161, 2004.

[9] Antti J. Aho, A Yli-Hankala, L-P Lyytikäinen, and V Jäntti. Facial muscle activity, response entropy, and state entropy indices during noxious stimuli in propofol–nitrous oxide or propofol–nitrous oxide–remifentanil anaesthesia without neuromuscular block. British Journal of Anaesthesia, 102(2):227–233, 2009.

[10] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[11] Jacob Ziv and Abraham Lempel. A universal algorithm for sequential data compression. Information Theory, IEEE Transactions on, 23(3):337–343, 1977.

[12] Jean loup Gailly and Mark Adler. The gzip home page. = http://www.gzip.org. Accessed: Jun 2014.

[13] David A. Huffman. A method for the construction of minimum-redundancy codes. Proceedings of the IRE, 40(9):1098–1101, 1952.

[14] Julian Seward. bzip2 and libbzip2. =http://www. bzip. org. Accessed: Jun 2014.

[15] Matt Mahoney. Data compression programs. `http://mattmahoney.net/dc/index.html`. Accessed: Jun 2014.

[16] Dermot MacDonald, Adrian Grant, Margaret Sheridan-Pereira, Peter Boylan, and Iain Chalmers. The dublin randomized controlled trial of intrapartum fetal heart rate monitoring. American Journal of Obstetrics and Gynecology, 152(5):524 – 539, 1985.

[17] João Bernardes, Antónia Costa-Pereira, Diogo Ayres-de Campos, Geijn Herman P. van, and Luís Pereira-Leite. Evaluation of interobserver agreement of cardiotocograms. International Journal of Gynecology & Obstetrics, 57(1):33–37, 1997.

[18] Diogo Ayres-de Campos, Joã Bernardes, Antónia Costa-Pereira, and Luís Pereira-Leite. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. BJOG: An International Journal of Obstetrics & Gynaecology, 106(12):1307–1310, 1999.

[19] João Bernardes and Diogo Ayres-de Campos. The persistent challenge of foetal heart rate monitoring. Current Opinion in Obstetrics and Gynecology, 22(2):104–109, 2010.

[20] Manuela Ferrario, Maria G. Signorini, and Giovanni Magenes. Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. Medical and Biological Engineering and Computing, 47(9):911–919, 2009.

[21] Cristina Costa Santos, João Bernardes, Paul Vitányi, and Luís Antunes. Clustering fetal heart rate tracings by compression. In Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on, pages 685–690. IEEE, 2006.

[22] Hernâni Gonçalves, Ana Paula Rocha, Diogo Ayres-de Campos, and João Bernardes. Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery. Medical and Biological Engineering and Computing, 44(10):847–855, 2006.

[23] Steven M. Pincus and Ary L. Goldberger. Physiological time-series analysis: what does regularity quantify? American Journal of Physiology, 266:H1643–H1643, 1994.

[24] Andrey N. Kolmogorov. A new metric invariant of transitive dynamical systems and lebesgue space automorphisms. In Dokl. Acad. Sci. USSR, volume 119, pages 861–864, 1958.

[25] Yaha G. Sinai. On the concept of entropy of a dynamical system. In Dokl. Akad. Nauk. SSSR, volume 124, pages 768–771, 1959.

[26] Ray J. Solomonoff. A formal theory of inductive inference. part i. Information and Control, 7(1):1–22, 1964.

[27] Ray J. Solomonoff. Formal theory of inductive inference, part ii. Information and Control, 7(2):224–254, 1964.

[28] Gregory J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. Journal of the ACM (JACM), 16(1):145–159, 1969.

[29] Andrey N. Kolmogorov. Logical basis for information theory and probability theory. Information Theory, IEEE Transactions on, 14(5):662–664, 1968.

[30] Rudi Cilibrasi and Paul Vitányi. Clustering by compression. Information Theory, IEEE Transactions on, 51(4):1523–1545, 2005.

[31] Rudi Cilibrasi, Paul Vitanyi, and Ronald De Wolf. Algorithmic clustering of music. In Web Delivering of Music, 2004. WEDELMUSIC 2004. Proceedings of the Fourth International Conference on, pages 110–117. IEEE, 2004.

[32] Stephanie Wehner. Analyzing worms and network traffic using compression. Journal of Computer Security, 15(3):303–320, 2007.

[33] Eamonn Keogh, Stefano Lonardi, and Chotirat A. Ratanamahatana. Towards parameter-free data mining. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 206–215. ACM, 2004.

[34] Gordon V. Cormack and R. Nigel Horspool. Data compression using dynamic markov modelling. The Computer Journal, 30(6):541–550, 1987.

[35] Inês Nunes, Diogo Ayres-de Campos, Catarina Figueiredo, and João Bernardes. An overview of central fetal monitoring systems in labour. Journal of Perinatal Medicine, 41(1):93–99, 2013.

[36] Diogo Ayres-de Campos, Joã Bernardes, Anneke Kwee, Michelle Westerhuis, and Gerard H.A. Visser. Computer quantification of short-term variability as an adjunct to fetal electrocardiographic monitoring. BJOG: An International Journal of Obstetrics & Gynaecology, 114(11):1445–1446, 2007.

Multiscale Compression: an Effective Measure of Individual Complexity

4.2 Multiscale Compression: an Effective Measure of Individual Complexity

Abstract

Complexity is a property that reflects the amount of structured information in a given system. Kolmogorov complexity and Shannon entropy are conceptually different complexity measures. The main aim of this work is to show that, for dynamic signals, the use of a multiscale approach combined with an approximation to Kolmogorov complexity through different compressors may be as useful as the multiscale entropy calculated by approximation through different entropies. We explore these approaches in two data sets: a data set with 40 white-noise and 40 pink-noise time series and another data set of cardiac interbeat interval time series from forty-three congestive heart failure (CHF) patients and seventy-two healthy subjects. The dispersion value between the two complexity measures shows the importance of considering different measures to different physical and physiologic time series.

Lead Paragraph

Plesk and Greenhalgh [1] conjectured elements of health care as belonging to simple, complex and chaotic domains. Traditional health evidence lies in the simple domain (high agreement and certainty) usually associated to predictable linear actions, evidence based guidelines. In the chaos domain (low certainty and agreement) no information is visible from the interaction of the systems; however, there may be emergence and creativity and one should look deeply for intrinsic patters. Several measures have been developed trying to measure the information contained in physical and physiological systems belonging to the complex and chaotic domains. The multiscale approach is based on the observation that the output of complex systems generally shows structures with long-range correlations on spatial and temporal scales. These multiscale features are ignored by conventional complexity measures. In this work we compare two conceptually different measures of complexity, in conjunction with the multiscale approach: the entropy and the Kolmogorov complexity.

Introduction

Solomonoff [2, 3], Kolmogorov [4] and Chaitin [5, 6] independently defined the complexity of an object as the length of the shortest program that produces the object. This individual complexity measure is usually known as Kolmogorov complexity or algorithmic information theory. A binary string that contains many long runs of the same bit will be highly compressible and therefore have a low Kolmogorov complexity, whereas a binary string that is generated using fair coin flips will have a high Kolmogorov complexity (with high probability). Practical applications of Kolmogorov complexity (in the sense of real, runnable computer programs) have emerged only in recent years. By approximating the Kolmogorov complexity by realworld compressors, one can transform the theoretical notion into applications that work better than one would expect. The Shannon entropy [7] of a random variable X is a measure of its

average uncertainty. It is the smallest number of bits required, on average, to describe x, the output of the random variable X. Several approximations to the entropy, the Approximate (ApEn) [8] and Sample (SampEn) [9] entropy, has been traditionally used to quantify the amount of information/complexity carried by a physiological signal such as heart rate and its deterministic dynamics.

The Multiscale Entropy (MSE) [10], consider the information contained in multiple scales obtained better results in the analysis of dynamics systems. Summarily, the SampEn is calculated for each coarse-grained time series derived from the original time series. So, multiscale entropy is looking for the sufficient statistic using as model the different scales. Kolmogorov complexity and Shannon entropy are conceptually different measures. However it is a well knows result that, for any recursive probability distribution, the expected value of Kolmogorov complexity equals its Shannon entropy, up to a constant. The main aim of this work is to show that, for dynamic signal, the use of multiscale approach combined with an approximation to Kolmogorov complexity through different compressors may be as useful [11] as the multiscale entropy calculated by approximation through different entropies, which has been successfully applied to different scientific areas [10, 12].

Methods

It is well know that Kolmogorov complexity is not computable, however we can approximate it by using standard compressors. The main point of data compression is the encoding of information using fewer bits than the original data. Several compressors were developed in last 40 years, each one with distinct properties suitable for different data sets. In this work we use two lossless compressors: the gzip compressor [13] one of the most used everyday compressor and paq8l [14], one of the best compressors at this moment, based on Dynamic Markov compression.

Multiscale entropy (MSE) was introduced in 2002 [10], the key innovation was to representing the information of a system's dynamics on different time scales. The multiscale method can be separate in two parts. The first one is the construction of the time series scales: using the original signal, a scale, s, is created from the original time series, through a coarse-graining procedure, i.e, replacing s points by their average. The second step concerns the calculation the value of entropy for each time series scale. In this work we use the Sample Entropy (SampEn) with the parameters $m = 2 and r = 15\%$ of the standard deviation of the signal. The complexity index (CI) is the area under the MSE curve obtained by plotting the entropy value as a function of scale. More details can be found in [10, 12].

We extend this idea to other measure of information namely compression, i.e., we consider different scales however we calculate the compress ratio [1] of the reduced time series instead of computing its entropy, thereby introducing the multiscale compression (MSC). For this purpose we use the compressors gzip and paq8l.

In the first section of results we study the behavior of multiscale compression in a set of white and pink noise time series. The white noise is a randomly generated signal that has equal power in any band of a given bandwidth. The pink noise, also called "1/f noise", is a signal with a frequency spectrum linear in logarithmic space, i.e., the power spectral density is inversely proportional to the frequency. These time series are completely orthogonal, and this is the main reason why we start our study of multiscale

[1] Naturally, as the size of the time series decrease the size of its compression also decreases. The most natural way to avoid this problem is to consider the compression rate (CR = Size Compress file / Size Original file) instead of the value of compression, i.e., the percentage of compression.

compression with them. Then we analyze the differences between subjects with congestive heart failure and healthy subjects using the same data-set as in [12].

Results

White and Pink Noise

We created 40 normalized time series, with 2500 point each, for each type of noise. Figure 4.2.1 has an example of each. Figure 4.2.2 displays the compression rate at different scales in both time series. The

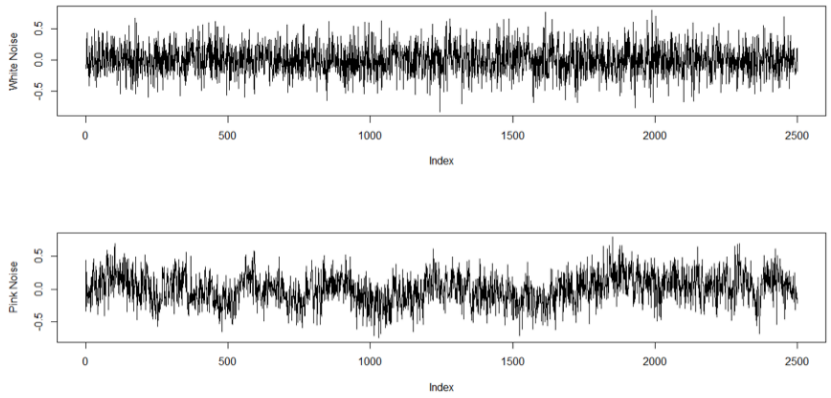

Figure 4.2.1: White and Pink noise examples

value of compression rate, for both compressors, for the white noise (blue circles in the figure) decreases through the scales while the behavior for pink noise (red stars) with both compressors the value increases when the scale factor increases but not in different ways. The results of MSE corroborate the fact that pink noise contains complex structures across all scales, unlike white noise (figure 4.2.3).

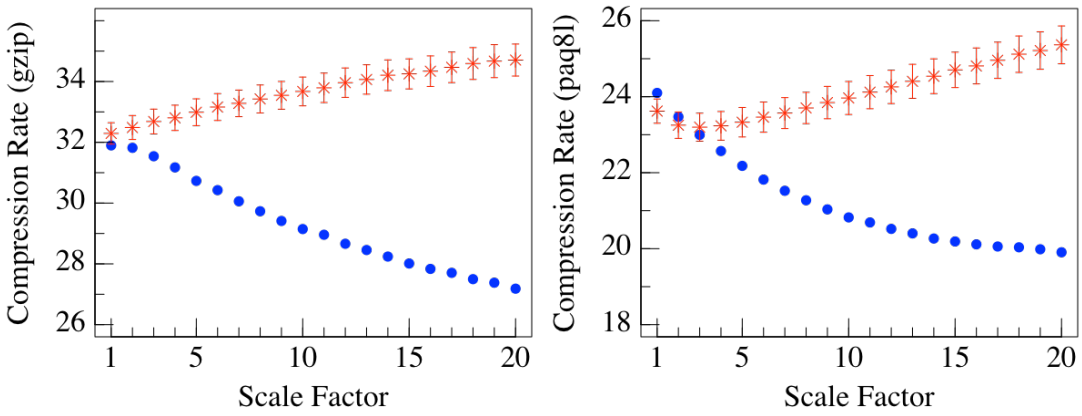

Figure 4.2.2: Multiscale compression (MSC) analysis of white and pink noise. Each panel display the results of gzip and paq8l, respectively. The blue circles represent the white noise and the red stars represent the pink noise.Values are presented as means ± 2*standard error.

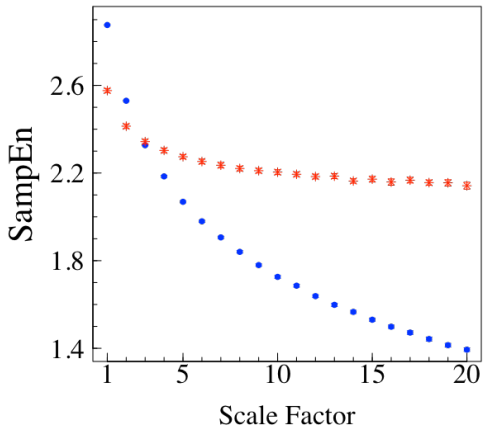

Figure 4.2.3: Multiscale entropy (MSE) - SampEn(2,0.10*sd) - analysis of white and pink noise. The blue circles represent the white noise and the red stars represent the pink noise.

Congestive Heart Failure vs Healthy subjects

In this section the MSC is applied to cardiac interbeat interval time series from 43 congestive heart failure (CHF) patients and 72 healthy subjects, using the same data-set as in [10, 12]. We start analyzing the sleeping periods. As figure 4.2.4 shows the compression rate of both interbeat interval time series

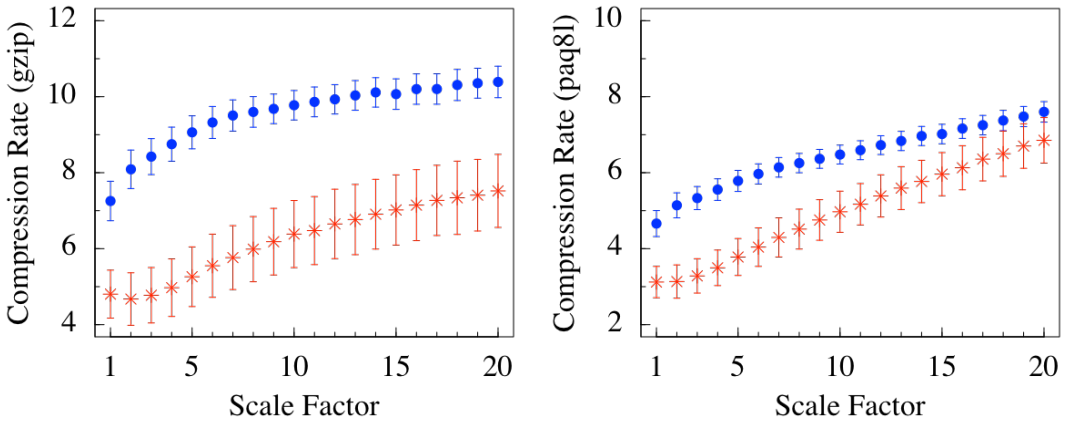

Figure 4.2.4: Multiscale compression rate, using gzip and paq8l, analysis of interbeat interval time series during sleep derived from CHF patients and healthy subjects. Values are given as means ± 2*standard error. The blue circles represent the healthy subjects and the red stars represent the CHF patients.

increases a little, similar to pink noise, suggesting the presence of complex structures in these time series. In [12] the sleeping period was also analyzed with one difference that was the division of healthy subjects in young and elderly. In figure 4.2.5 the different of behavior in interbeat interval time series from healthy and from CHF patients is clear.

The dynamic associated to the two methods, MSC and MSE, is similar however the MSC presents a better separation between the two groups. The complexity indexes for entropy ($C_{SampEn}I$) and for different compressors ($C_{compressor}I$), for the 20 scales, are significantly different for the two groups ($p - values < 0.005$ and < 0.001, respectively). The Pearson correlation coefficient between these two measures ($C_{SampEn}I$ and $C_{compressor}I$) is 0.34 (significantly different from 0, $p - value < 0.001$).

Considering the 24 hours recordings we performed the same analysis.

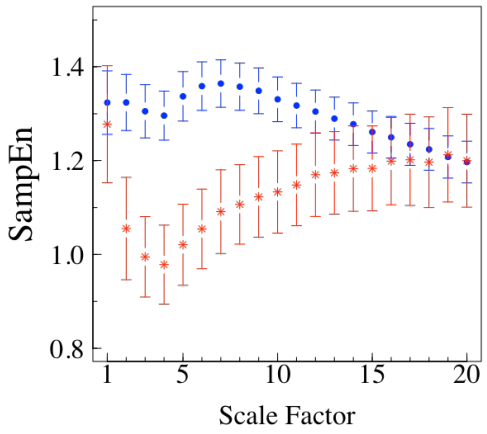

Figure 4.2.5: Multiscale SampEn(2,0.15*sd) analysis of interbeat interval time series from healthy subjects (blue circles) and from CHF subjects(red stars). Values are given as means ± 2*standard error.

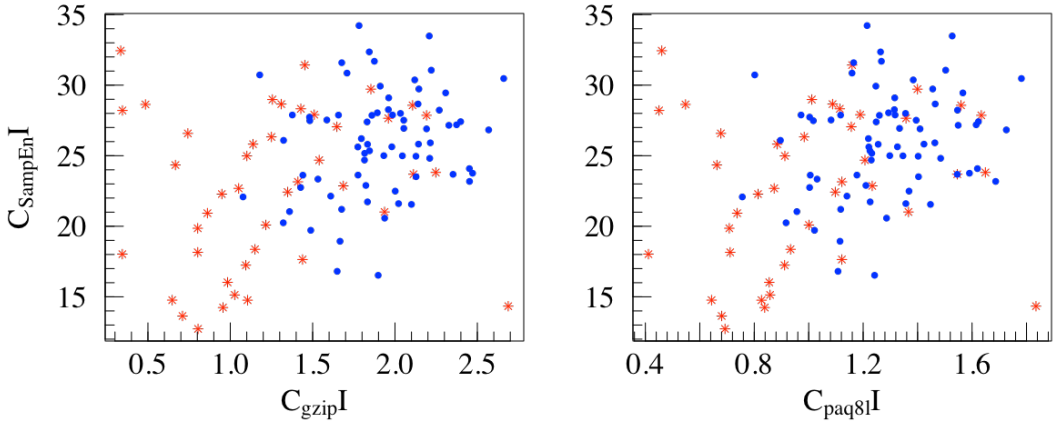

Figure 4.2.6: Dispersion between the complexity index computed using the MSE curve or the MSC curve for each one of the compressors. The Pearson correlation coefficient between these two measures is 0.34 (significantly different from 0, $p - value < 0.001$). The blue circles represent the healthy subjects and the red stars represent the CHF patients.

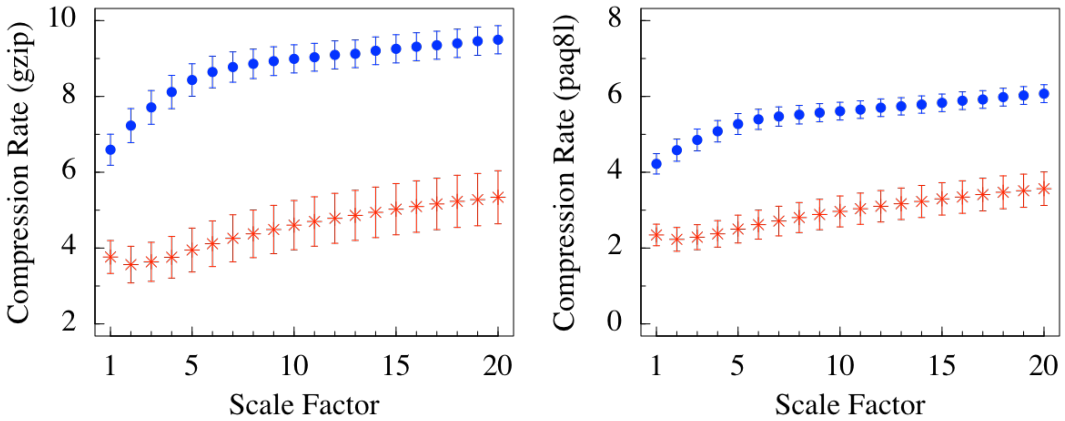

Figure 4.2.7: Multiscale compression (MSC) gzip and paq8l analysis of interbeat interval time series derived from CHF patients and healthy subjects for 24h. Values are given as means ± 2*standard error. The blue circles represent the healthy subjects and the red stars represent the CHF patients.

Analyzing figure 4.2.7 one can see a good separation and different dynamics for initial scales among the two groups, as in the sleeping periods.

In [12], these traces were also analyzed and compared with Atrial Fibrillation (AF) subject. Figure 4.2.8 show that the entropy also has different behavior across scale for each group. However, the distinction between the two groups is worst.

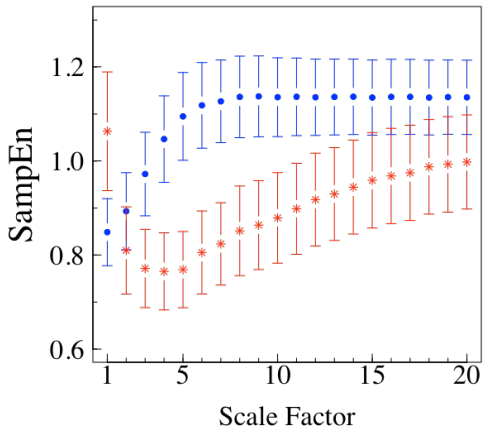

Figure 4.2.8: Multiscale analysis of interbeat interval time series from healthy subjects, CHF subjects and from AF subjects. [12]

Once more the dynamics associated with MSC and MSE presents a better separation between the two groups. The complexity indexes for entropy ($C_{SampEn}I$) and for different compressors ($C_{compressor}I$), for the 20 scales, are significantly different for the two groups ($p-values < 0.005$ and < 0.001, respectively). See figure 4.2.9.

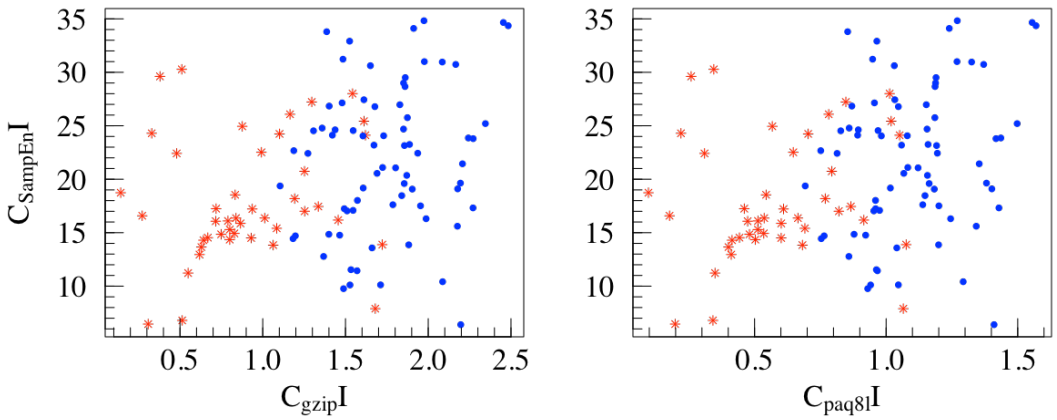

Figure 4.2.9: Dispersion graph between the CI compute from the MSE curve and from the MSC curve for different compressors. The Pearson correlation coefficient between these two measures ($C_{SampEn}I$ and $C_{compressor}I$) is 0.32 (significantly different from 0, $p-value < 0.001$).

Discussion

In this work, it is not our purpose to present a new method but join two successful ones: the coarse-grained method used in MSE and compression as a measure of complexity. The multiscale analysis has

been successfully applied in the analysis of dynamic systems specially when associated with entropy. Compression is also a useful complexity measure, measuring different properties, the combination of the two can be specifically important in medicine where there are many dynamic systems whose behavior is unknown.

In correlated and uncorrelated noises the compression rate and the entropy have similar behavior across the scales. This behavior, support the idea that both are complexity measures and that the multiscale approach, independently of the complexity measure used, can be very useful to distinguish data. The pink noise with the entropy measure remains constant across the scales while the compression rate increases. Contrary to white noise behavior that, in both measures, as expected after a coarse-graining procedure, decreases while scales increases.

The multiscale approach allows the analysis of features present in time series structures with long-range correlations on multiple spatial and temporal scales. The MSE and MSC methods show that healthy dynamics are the more complex, compatible with the concept that physiologic complexity is fundamentally related to the adaptive capacity of the organism. On the other hand, pathologic dynamics are associated with either increased regularity and decreased of variability, sustained a breakdown of long-range correlations and loss of information [15]. Although both methods present a similar dynamic and the same tendency (higher values of CI in healthy subjects) the distinguish power and the dispersion values are quite different between the two complexity measures. This shows the importance of considering different measures to different physical and physiologic time series.

Bibliography

[1] Paul E. Plsek and Trisha Greenhalgh. The challenge of complexity in health care. Bmj, 323(7313):625–628, 2001.

[2] Ray J. Solomonoff. A formal theory of inductive inference. part i. Information and Control, 7(1):1–22, 1964.

[3] Ray J. Solomonoff. Formal theory of inductive inference, part ii. Information and Control, 7(2):224–254, 1964.

[4] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information'. Problems of information transmission, 1(1):1–7, 1965.

[5] Gregory J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. Journal of the ACM (JACM), 16(1):145–159, 1969.

[6] Gregory J. Chaitin. On the simplicity and speed of programs for computing infinite sets of natural numbers. Journal of the ACM (JACM), 16(3):407–422, 1969.

[7] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[8] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[9] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[10] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[11] Teresa Henriques, Hernâni Gonçalves, Luís Antunes, Mara Matias, João Bernardes, and Cristina Costa-Santos. Entropy and compression: two measures of complexity. Journal of Evaluation in Clinical Practice, 19(6):1101–1106, 2013.

[12] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):021906, 2005.

[13] Jean loup Gailly and Mark Adler. The gzip home page. `http://www.gzip.org`. Accessed: Jun 2014.

[14] Matt Mahoney. Data compression programs. `http://mattmahoney.net/dc/index.html`. Accessed: Jun 2014.

[15] Ary L. Goldberger, Luis A.N. Amaral, Jeffrey M. Hausdorff, Plamen C. Ivanov, Chung-Kang Peng, and H. Eugene Stanley. Fractal dynamics in physiology: alterations with disease and aging. Proceedings of the National Academy of Sciences, 99(suppl 1):2466–2472, 2002.

Compression vs Entropy: on the characterization of the nonlinear features of physiological signal

Teresa Henriques, Francisco Mota , João Bernades,
Cristina Costa-Santos, Luís Antunes

SUBMITTED

4.3 Compression vs Entropy: on the characterization of the nonlinear features of physiological signal

Abstract

Conventional linear mathematic methods fail to describe biological systems belonging to the complex and chaotic domains. Different approaches based on chaos and complexity concepts have been successfully considered. In this work we addresses the challenge of assessing the dynamics of fetal heart rate (FHR) signals from fetuses born with neonatal acidemia using nonlinear indices. The analysis of a FHR signal represents a noninvasive tool for evaluating the fetal condition. The information obtained may provide important auxiliary information to clinicians, in the diagnostic of the fetal state and help in consequent clinical decisions in life-threatening situations, particularly in the minutes prior to delivery.

The most widely used complexity measures in the literature are approximations to the entropy, namely Sample Entropy or Approximate Entropy. In this work we challenge complexity from a different viewpoint, namely approximation to the mathematical sound notion of Kolmogorov complexity. One of the main objective of this work is also to understand the complementary information that can be derived from These different approaches, providing a more comprehensive view of the underlying physiology.

We start by studying the usefulness of approximations of entropy and Kolmogorov complexity to deal with the biological complexity of FHR. We achieve it, showing that the correlations between the entropy values and the compression ratio are not significative different of 0. This is a strong evidence that despite the two approaches quantify the complexity of a trace they probe different characteristics of this. Finally we give some strong evidence regarding the characteristics of the FHR traces that each of the complexity measures explore, namely short and long term variability.

Introduction

The human body is constituted by several interacting and auto-adjusting physiological systems whose actions one can not always predict. The behavior of the physiological systems determining the functioning of the body is based on a complex adjustment of various variables according to internal requirements and external influences. Any physiological mechanism evolves along well-assigned temporal scales. The study of this complex physiology needs mathematical approaches capable to deal with concepts of information, complexity. In addition, there is a growing amount of data that indicate a reduction of complexity during pathological situations, suggesting that the use of these complexity measures may characterize pathology.

In this paper we study the usefulness of approximations of entropy and Kolmogorov complexity to deal with the biological complexity. This approach challenges complexity from different viewpoints and tries to understand the complementary information that can be derived from different approaches, providing a more comprehensive view of the underlying physiology.

To approximate the Komogorov complexity we use some well known compressors, namely lzma, bzip2

and paq8l and to approximate the theoretical valued of the entropy we use the Sample Entropy ($SampEn$), an improvement of the Approximate Entropy ($ApEn$), designed to evaluate biological time series, particularly the heart rate signal.

To better ilustrate our results we use a dataset of 128 traces of fetal heart rate (FHR). Cardiotocography is used to assess the fetal wellbeing, by the record of the fetal heart rate and the uterine contractions. Computer analysis of cardiotograms provides quantifying parameters that are difficult to assess by the human eye, and overcomes low agreement among clinicians in FHR analysis. Carditotocography seems to be a clinical procedure inserted in a complex system, with a non-linear behavior.

Methods

Entropy

Information theory quantifies the a priori uncertainty about the results of an experiment. Classical information theory originated in Shannon's paper "A mathematical theory of communication"[1], where the author defined the notion of entropy and showed that it corresponded to the amount of information associated with any given statistical event.

The Shannon entropy of a random variable X is defined as:

$$H(X) = -\sum_i p(x(i)) \cdot log\ \ (p(x(i))) \tag{4.1}$$

where $p(x(i)) = P\{X = x(i)\}$.

Traditionally, when studying physiological signals, one can not compute the exact value of the entropy as we do not know the underlying distribution, however we can approximate it. To estimate the Shannon entropy – $H(X)$ – of a distribution X, in general, requires the knowledge of the entire distribution X, that could potentially require exponentially many samples from x.

The Approximate Entropy ($ApEn$), proposed by Pincus [2], tries to approximate of the entropy and exhibits a good performance in the characterization of randomness (or regularity) of physiological data. In order to calculate the $ApEn$, a new series of vector of length m is constructed as follows: $X_i = (x(i), x(i+1), ..., x(i+m-1)), i = 1, \ldots, N-m+1$. For each vector X_i, the value $C_i^m(r)$ is computed as:

$$C_i^m(r) = \frac{\text{number of } d[X_i, X_j] \leq r}{N-m+1} \quad \forall j \tag{4.2}$$

where r is referred as a tolerance value to be pre-define and the distance function used is defined as:

$$d[X_i, X_j] = max_{k=1,\ldots,m}|x_{i+k-1} - x_{j+k-1}| \tag{4.3}$$

Next, the average of the natural logarithm of $C_i^m(r)$ is computed for all i:

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} ln(C_i^m(r)) \tag{4.4}$$

Since in practice N is a finite number, the statistical estimate is computed as:

$$ApEn(m,r) = \begin{cases} \Phi^m(r) - \Phi^{m+1}(r) & \text{for } m > 0 \\ -\Phi^1(r) & \text{for } m = 0 \end{cases}$$

The Sample Entropy ($SampEn$) [3] was introduced, as an improvement of the $ApEn$ measure, to evaluate biological time series, particularly the heart rate signal. For the $SampEn$ calculation the same parameters defined for the $ApEn$, m and r are required. Considering A as the number of vector pairs of length $m+1$ having $d[X_i, X_j] \leq r$, with $i \neq j$ and B as total number of template matches of length m. The SampEn is defined as:

$$SampEn = -ln\ \frac{A}{B} \tag{4.5}$$

In other words the SampEn is the negative natural logarithm of the conditional probability that templates of a certain length (m) that match pointwise, within a tolerance threshold (r), also match if the template increases one point (length $m+1$).

The multiscale entropy [4, 5](MSE) approach, extend the entropy concept of regularity to a concept of fractality considering the information of a system s dynamics on different time scales. The multiscale method comprise two parts: 1) the construction of the time series scales: using the original signal, a scale, s, is created through a coarse-graining procedure by replacing s consecutive points by their average; 2) the computation of the entropy index for each time series scale. The MSE curve is obtained by plotting the entropy value as a function of scale.

Kolmogorov Complexity

Algorithmic information theory was proposed in the 60's, independently by Kolmogorov, Solomonoff and Chaitin [6–9]. This quantity is now known as Kolmogorov complexity and is defined as the length of the shortest program that can produce a given string x, denoted as $K(x)$. Unlike entropy, this quantity depends exclusively on the string, and not on the probability with which it is sampled from some given distribution. As such, Kolmogorov complexity measures the intrinsic information in a given string which for some application is more useful than entropy.

We saw that to evaluate the entropy one in general requires knowledge of the entire distribution and so typically we approximate it. The Kolmogorov complexity seems even worse, as it also well known to be uncomputable. However in practice, one can often estimate $K(x)$ reasonably well by compressed file size, when x is fed to a standard compression algorithm such as gzip.

In 2004/2005, a new method for data clustering using compression was introduced in [10], based on a similar method first used in mitochondrial genome phylogeny [11]. It is feature-free, there are no parameters to tune, and no domain-specific knowledge is used. Good results where obtained applying it on different areas: languages tree, genomics, optical character recognition, literature [10], music [12], and computer virus and internet traffic analysis [13]. This method is extremely powerful and general. It can mine patterns in completely different areas, there are no domain-specific parameters to set, and it does not require any background knowledge. Its performance in clustering heterogeneous data and anomaly detection in time sequences has been shown to outperform that of every known data-mining method in a massive recent study [14].

Physiological Signal: An Application

The assessment of fetal heart rate (FHR) variability is extremely important in obstetric care to assure the fetal well-being. Some pathologies as fetal hypoxia, congenital heart anomalies and fetal tachycardia can cause decreased variability [15]. Traditionally, in this area, two measures of variability are used. The short-term variability (STV) a measure of the oscillations of the FHR between consecutive points and the long-term variability (LTV) representing slower oscillation in heart rate.

We consider a data set composed by 182 FHR traces from healthy fetus collected between the 36^{th} and 38^{th} gestational week. Prematurity decreases variability, however variability should be normal after 32 weeks [16]. The dataset was collected in Hospital de São João, Porto, Portugal and then analyzed by Omniview-SisPorto®, a central fetal monitoring viewing system, producing some clinical features. In this work we focus our attention on the percentage of abnormal STV and LTV. In the Omniview-SisPorto®, an abnormal STV point is defined whenever the difference between two adjacent FHR values are less than 1 bpm. On the other hand, a point with abnormal LTV is identified when the difference between maximum and minimum values of a sliding 60-sec window centered on it does not exceed 5 bpm [17]. However, there is no universal consensus on these definitions.

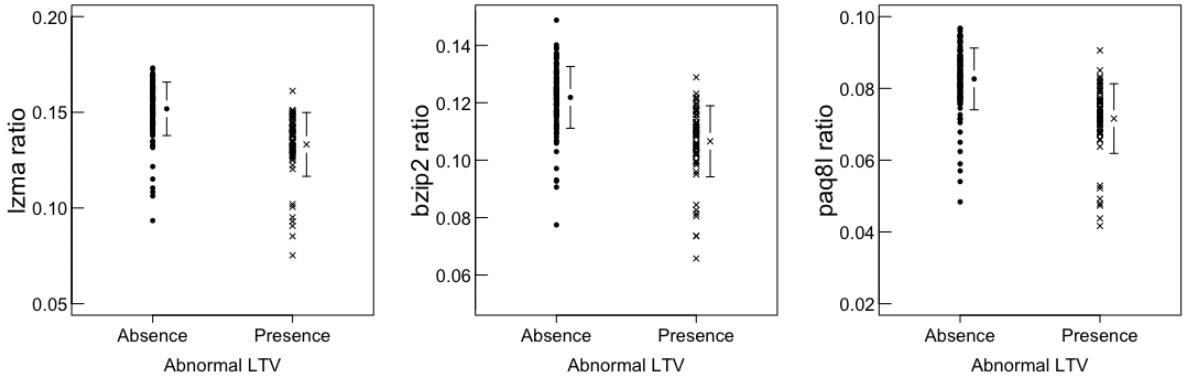

Figure 4.3.1: Compression Rate using three different compressors (A) lzma, (B) bzip2 and (C) paq8l for the fetal heart rate time series without (∘) and with (×) abnormal LTV groups. The complexity rate is significantly (p-value $<<$ 0.001) lower in the group with presence of abnormal LTV comparing with the other group. Symbols with error bars represent mean and SD values, respectively.

Using nonlinear method such as SampEn and the compression rate we can significantly (p-values < 0.005 and $<< 0.001$, respectively) distinguish fetus with or without percentage of long term variability (Figures 4.3.1 and 4.3.2).

All the 182 FHR traces presented a percentage of abnormal STV higher than 50%. The six complexity indices, based on entropy and compression, are significantly higher (p-values for compressors $<< 0.001$ and for SampEn < 0.001, respectively) for tracings in which this percentage is higher than 80 (Figures 4.3.3 and 4.3.4).

Despite the two approaches quantify the complexity of a trace they probe different characteristics of the data. The correlations between the entropy values and the compression ratio are not significative different of 0. In figures 4.3.5 and 4.3.6 we display the scatter-plot between the SampEn ($m = 2, r = 0.20$) and the three compression ratios. We used SampEn as it optimizes the separation between the groups,

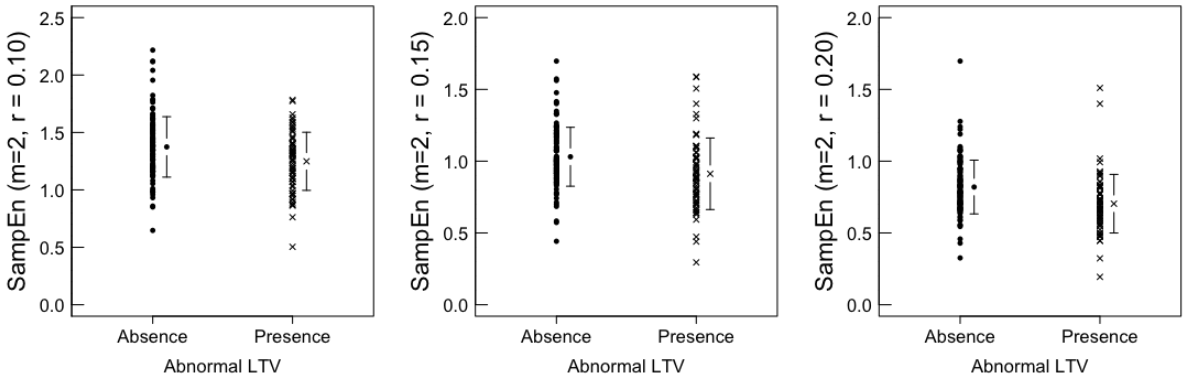

Figure 4.3.2: SampEn using $m = 2$ three r thresholds (A) $r = 0.10$, (B) $r = 0.15$ and (C) $r = 0.20$ for the fetal heart rate time series without ($\circ$) and with ($\times$) abnormal LTV groups. The entropy in the pathologic group was significantly (p-value < 0.005) lower in the group with presence of abnormal LTV comparing with the other group. Symbols with error bars represent mean and SD values, respectively.

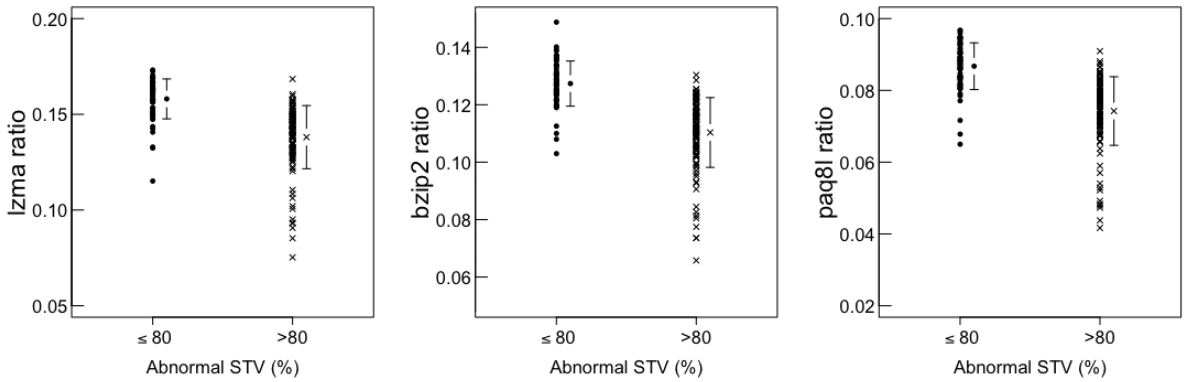

Figure 4.3.3: Compression Rate using three different compressors (A) lzma, (B) bzip2 and (C) paq8l for the fetal heart rate time series without ($\circ$) and with ($\times$) abnormal STV groups. The complexity rate is significantly (p-value $<< 0.001$) lower in the group with presence of abnormal STV $> 80\%$ comparing with the other group. Symbols with error bars represent mean and SD values, respectively.

similar arguments where use to choose the threshold for (r) values.

The absence of correlation between the two approaches raised the central question for this work: which features of the FHR traces are SampEn and the compressors exploring? So, one should look for a better understanding of these approaches behavior in the assessment of correlations and in the presence of noise. To do so, for each fetal heart time series we create two new time series: a *noisier* adding white noise (with mean 0 and standard deviation 1) and a *randomized* by shuffling the order of all the points of the original time series. For each of the new time series we compute the six complexity measures.

Examining the results of figures 4.3.7 and 4.3.8 it seems that time series with noise or random have higher complexity than the original FHR time series. However, analyzing the results in different time scales by applying the multiscale approach, combined with all six different measure, to the three sets of time series we observe that the entropy decreases for the *noisier* and *randomized* time series.

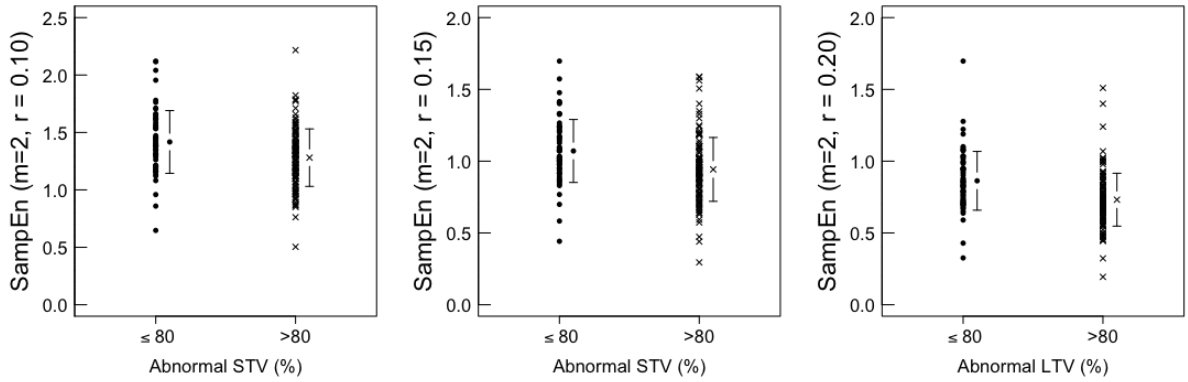

Figure 4.3.4: SampEn using $m = 2$ three r thresholds (A) $r = 0.10$, (B) $r = 0.15$ and (C) $r = 0.20$ for the fetal heart rate time series without ($\circ$) and with ($\times$) abnormal STV groups. The entropy in the pathologic group was significantly (p-value < 0.001) lower in the group with presence of abnormal STV $> 80\%$ comparing with the other group. Symbols with error bars represent mean and SD values, respectively.

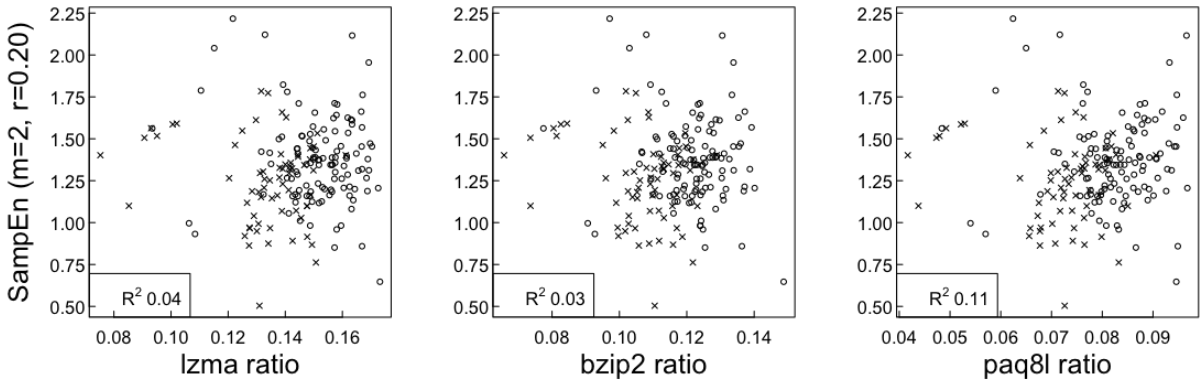

Figure 4.3.5: Dispersion between the SampEn (r = 0.20) and each one of the complexity rates (A) lzma, (B) bzip2 and (C) paq8l. The Pearson correlation coefficient between the measures was not significant. The solid circles represent the healthy group (absence of abnormal LTV) and the crosses represent the ones with presence of abnormal LTV.

Discussion

In the classification of short and long term variability of the FHR traces both complexity measures successfully distinguish the two groups however the compression measures give the best separation. The non significant correlation between the entropy and compression approaches suggest that each one can quantify different features of a system. These findings reinforce the importance of considering different measures assessing physiologic time series.

The SampEn measure has been widely used to assess the complexity in physiological data. Despite that, a limitation of the entropy measures relates to the fact that they are parametric measures, and thus their results vary according to the parameters chosen (vector length m, tolerance value r and number of points of the time series). Actually, if r is chosen as a fixed value the results obtained are highly correlated with the time series amplitude (SD). To assess the data correlations independently of the time

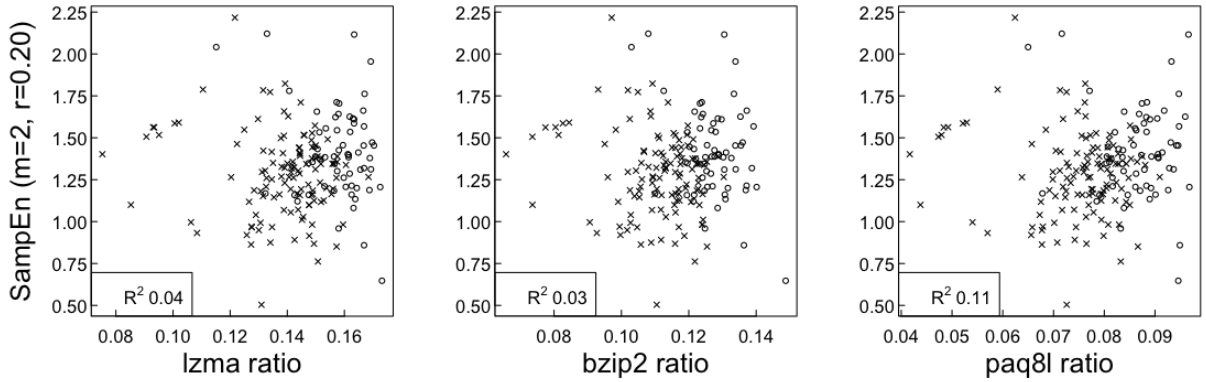

Figure 4.3.6: Dispersion between the SampEn (r = 0.20) and each one of the complexity rates (A) lzma, (B) bzip2 and (C) paq8l. The Pearson correlation coefficient between the measures was not significant. The solid circles represent the group with percentage of abnormal STV lower or equal to 80% and the crosses represent the ones with percentage of abnormal STV higher than 80%.

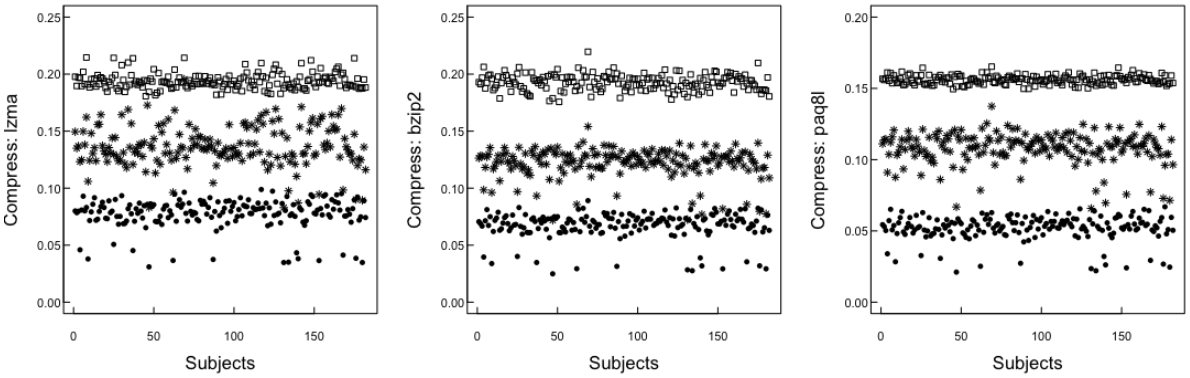

Figure 4.3.7: Compression Rate using three different compressors (A) lzma, (B) bzip2 and (C) paq8l for the original fetal heart rate time series (•), for the *noisier* time series (□) and for the *randomized* time series (∗).

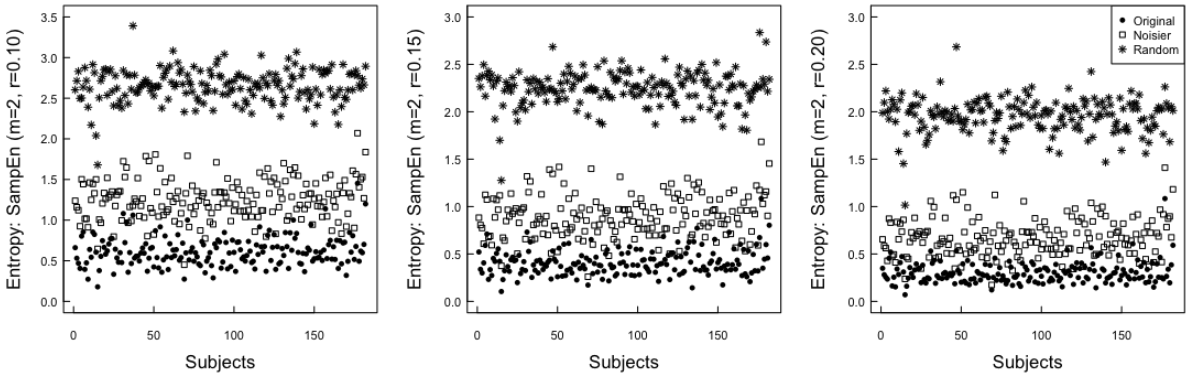

Figure 4.3.8: SampEn with $m = 2$ and three r thresholds (A) $r = 0.10$, (B) $r = 0.15$ and (C) $r = 0.20$ for the original fetal heart rate time series (•), for the *noisier* time series (□) and for the *randomize* time series (∗).

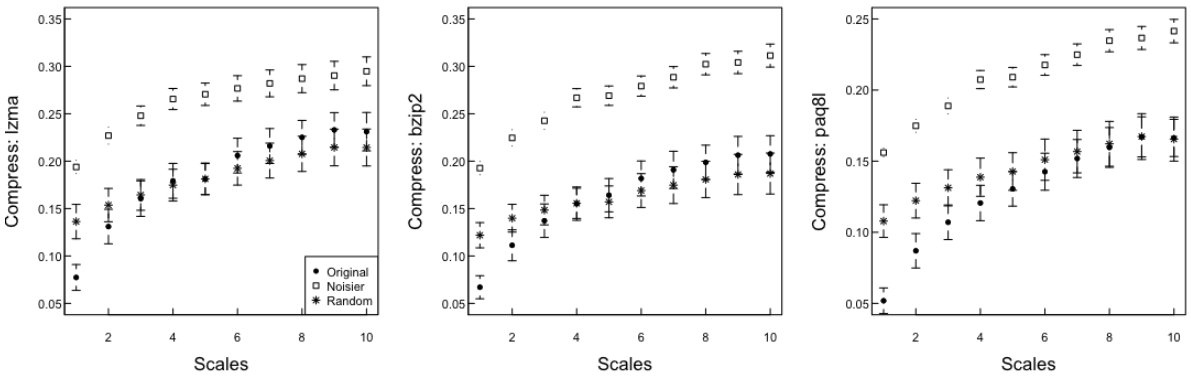

Figure 4.3.9: Multiscale compression rate using three different compressors (A) lzma, (B) bzip2 and (C) paq8l for the original fetal heart rate time series (•), for the *noisier* time series (□) and for the *randomize* time series (∗). Symbols with error bars represent group mean and SD values, respectively.

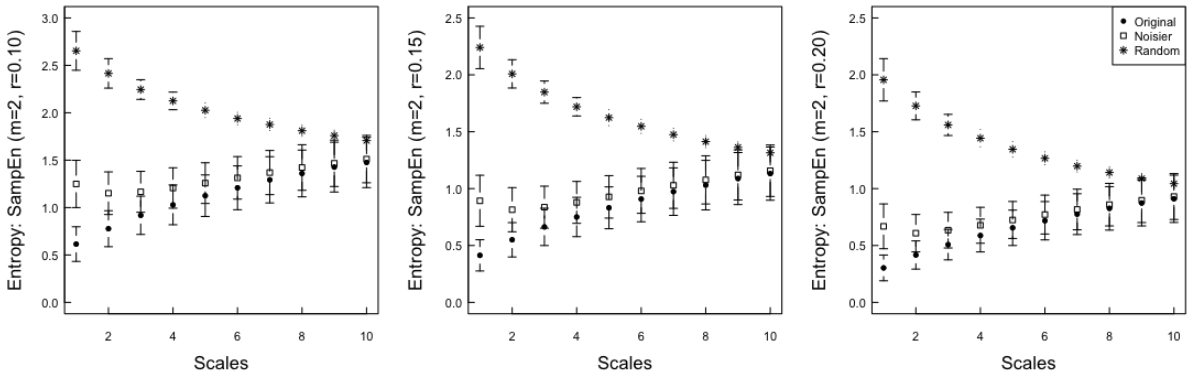

Figure 4.3.10: Multiscale SampEn using $m = 2$ three r thresholds (A) $r = 0.10$, (B) $r = 0.15$ and (C) $r = 0.20$ for the original fetal heart rate time series (•), for the *noisier* time series (□) and for the *randomize* time series (∗). Symbols with error bars represent group mean and SD values, respectively.

series amplitude, a common choice is to use the r value as a percentage of the standard deviation of the time series. This choice has the advantage of being equivalent to normalizing the time series to unit variance and zero mean *à prior*, but the results still depend on the percentage r chosen.

The dictionary (number of unique values) of the time series to be analyze is of extremely importance in the use of these measures. The entropy, due to the use of threshold r allow a certain flexibility when comparing two signals. However, attention is specially needed when computing entropy in discrete time series. In compression measures a small difference in the decimal values can produce totally different final results, as we showed when introducing noise to the original time series. To reduce the impact of this limitation we round the values of all time series to three decimal places.

The presence of noise in a fractal-like time series can reproduce misleading results. By using entropy measures, if the noise is superior to the threshold (r) chosen, the values obtained are higher than the ones from the original time series. Using the multiscale approach, in a few scales the noise is "eliminated" and the results of both time series became similar. In contrast, using compression measures the values obtained when analyzing the noisier time series after introduced white noise with SD = 1 to the original ones are approximately 10% (0.10) superior. When doing the multiscale approach both time series present

the same behavior but the difference between the two slightly decreases with the decrease of noise in both time series.

The process of randomized a time series eliminate the correlations between consecutive values. The compression ratio values, on the first scale, are higher for the randomized than the original time series. Using the multiscale approach, the behavior of the two set of data appears to be different. The randomized time series maintains or slightly increase through the scales (similar to fractal noise) while the original time series increases their value through the scales. The entropy values for randomized time series are much higher than the original and the *noisier* time series. However, this values are not representative of higher complexity of uncorrelated time series associated with healthier systems. The entropy value of the randomized time series decreases when the scales increases similar to the white noise behavior.

We believe that the differences in the behavior of the two measures relates with presence of noise is due to the fact of the SampEn is looking for short term correlations in contrast with the compressors that probe short and long term fluctuations.

Conclusion

The values of Kolmogorov complexity, an algorithmic approach, approximated by compressors were compared to entropy ones, a probabilistic approach, by using the SampEn measure. The findings presented on this section validate the notion that compressors, despite less frequent, can be effectively used, complementary to entropy indices, to quantify complexity in biological signals, specifically in FHR analysis.

The results obtained suggest that while the entropy measure probe essentially the short term correlations (by using $m = 2$) the compressors also assess the long term correlations. The use of multiscale entropy combine with these measures improve the performance of both allowing to more accurately examine the fractal dynamics present in the physiological signals and to distinguish the "real" dynamics from the presence of noise.

The strengths and limitation of both measure should be take into consideration either before applying or when interpreting the results.

Bibliography

[1] Claude E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.

[2] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[3] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[4] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[5] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):021906, 2005.

[6] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information'. Problems of Information Transmission, 1(1):1–7, 1965.

[7] Ray J. Solomonoff. A formal theory of inductive inference. part i. Information and Control, 7(1):1–22, 1964.

[8] Ray J. Solomonoff. Formal theory of inductive inference, part ii. Information and Control, 7(2):224–254, 1964.

[9] Gregory J. Chaitin. On the length of programs for computing finite binary sequences. Journal of the ACM (JACM), 13(4):547–569, 1966.

[10] Rudi Cilibrasi and Paul Vitányi. Clustering by compression. Information Theory, IEEE Transactions on, 51(4):1523–1545, 2005.

[11] Ming Li, Jonathan H. Badger, Xin Chen, Sam Kwong, Paul Kearney, and Haoyong Zhang. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. Bioinformatics, 17(2):149–154, 2001.

[12] Rudi Cilibrasi, Paul Vitányi, and Ronald De Wolf. Algorithmic clustering of music based on string compression. Computer Music Journal, 28(4):49–67, 2004.

[13] Stephanie Wehner. Analyzing worms and network traffic using compression. Journal of Computer Security, 15(3):303–320, 2007.

[14] Eamonn Keogh, Stefano Lonardi, and Chotirat A. Ratanamahatana. Towards parameter-free data mining. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 206–215. ACM, 2004.

[15] Interpreting FHR Patterns. Interpretation of the electronic fetal heart rate during labor. American Family Physician, 59(9):2487–2500, 1999.

[16] J Kurse. Electronic fetal monitoring during labor. The Journal of Family Practice, 15(1):35–42, 1982.

[17] Diogo Ayres-de Campos, João Bernardes, António Garrido, Joaquim Marques-de Sá, and Luís Pereira-Leite. Sisporto 2.0: a program for automated analysis of cardiotocograms. Journal of Maternal-Fetal and Neonatal Medicine, 9(5):311–318, 2000.

Non-Linear Models To Assess Continuous Glucose Monitoring Data

Learn from yesterday, live for today, hope for tomorrow.
The important thing is to not stop questioning.

Albert Einstein, Relativity: The Special and the General Theory

5. Non-Linear Models To Assess Continuous Glucose Monitoring Data

The applicability of many of the methods proposed or used to characterize heart rate variability are not solely for that purpose and may be adapted to the analysis of other time series.

Diabetes mellitus (DM) is one of the world's most prevalent medical conditions affecting tens of millions worldwide. Recently a continuous glucose monitoring (CGM) technology, used in the management of patients with DM, provides serial measures of glucose levels for about a week [1]. Contemporary management focuses on lowering mean blood glucose values toward a normal range, but largely ignores the dynamics of glucose fluctuations [2]. The time series derived from CGM recordings remain a largely untapped source of dynamical information.

In this chapter some of the well defined methods, commonly used on heart rate analysis, were used to explore and analyze these "new" CGM time series. On the first section a new method, termed *glucose-at-a-glance*, is presented to visualize and analyze CGM outputs that may facilitate the clinical assessment of short and long-term glucose variability. The new method is based on density delay maps (Poincaré plots) [3] and enhance by adding a color scheme that represents different levels of density of the data points.

On the second section we go a step forward and using multiscale entropy (MSE) [4] analysis, we quantified the complexity of the temporal structure of the CGM time series from a group of elderly subjects with type 2 DM and age-matched controls. We also show that the fluctuations in CGM values sampled every 5 minutes are not random.

The findings support extending the conventional notion of a discrete therapeutic target to considerating of a new framework, *dynamical glucometry*, which should enhance complexity of glucose fluctuations and not just lower mean and variance of blood glucose levels.

Of note, the work developed in this chapter resulted in two published articles.

Bibliography

[1] American Diabetes Association. Standards of medical care in diabetes. Diabetes Care, 77 Suppl 1:S14–S80, 2014.

[2] American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care, 37 Suppl 1:S81–S90, 2014.

[3] Peter W. Kamen and Andrew M. Tonkin. Application of the poincaré plot to heart rate variability: a new measure of functional status in heart failure. Australian and New Zealand Journal of Medicine, 25(1):18–26, 1995.

[4] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

"Glucose-at-a-Glance:" New Method to Visualize the Dynamics of Continuous Glucose Monitoring Data

Teresa Henriques,Medha N. Munshi, Alissa R. Segal,
Madalena D. Costa and Ary L. Goldberger

Journal of Diabetes Science and Technology
Volume 8, Issue 2, Pages 299-306
March 2014

5.1 "Glucose-at-a-Glance:" New Method to Visualize the Dynamics of Continuous Glucose Monitoring Data

Abstract

The standard continuous glucose monitoring (CGM) output provides multiple graphical and numerical summaries. A useful adjunct would be a visualization tool that facilitates immediate assessment of both long- and short-term variability. We developed an algorithm based on the mathematical method of delay maps to display CGM signals in which the glucose value at time t_i is plotted against its value at time t_{i+1}. The data points are then color-coded based on their frequency of occurrence (density). Examples of this new visualization tool, along with the accompanying time series, are presented for selected patients with type 2 diabetes and non-diabetic controls over the age of 70 years. The method reveals differences in the structure of the glucose variability between subjects with a similar range of glucose values. We also observe that patients with comparable hemoglobin A1c (HbA1c) values may have very different delay maps, consistent with marked differences in the dynamics of glucose control. These differences are not accounted by the amplitude of the fluctuations. Furthermore, the delay maps allow for rapid recognition of hypo- and hyperglycemic periods over the full duration of monitoring or any subinterval. The glucose-at-a-glance visualization tool, based on colorized delay maps, provides a way to quickly assess the complex data acquired by CGM systems. This method yields dynamical information not contained in single summary statistics, such as HbA1c values, and may also serve as the basis for developing novel metrics of glycemic control.

Introduction

Continuous glucose monitoring (CGM), used in the management of patients with diabetes mellitus, provides serial measures of glucose levels. The standard CGM report includes multiple graphical and numerical summaries. We introduce a new method, termed *glucose-at-a-glance* to visualize and analyze CGM outputs that may facilitate the clinical assessment of short and long-term glucose variability.

The new method is based on *density delay maps*, which display the value of a variable at time t_i versus its value at time t_{i+1}. Up to the present, the primary biomedical application of traditional delay maps has been in the research analysis of heart rate time series [1–4], where they are referred to as Poincaré plots. In addition, two parameters of delay maps, quantifying the local and global time series' standard deviations (abbreviated SD1 and SD2), have been proposed for the analysis of CGM data [5, 6]. Here, we adapt and expand the delay map approach in new directions to allow for visualization of CGM data by adding a color scheme that represents different levels of density of the data points. To our knowledge, such colorized delay maps have not been previously used to help display and summarize CGM data.

Complementary to hemoglobin A1c (HbA1c) measurements, the gold standard in assessing recent glycemic control, and to currently used CGM statistic summaries, the density delay maps provide rapidly

accessible information about actual glucose dynamics. This information relates both to the temporal "structure" of serum glucose variability and the duration of periods of hypo/hyperglycemia.

Methods

Clinical data

To illustrate the visualization method, we used previously acquired CGM data from elderly subjects over the age of 70 years without diabetes (unpublished data) and with type 2 diabetes, who had been enrolled in clinical studies by the Joslin Geriatric Diabetes Research Group. The glycemic status of the diabetic subjects varied widely, as reflected in their HbA1c values. The CGM data were obtained using the iPro^{TM} system version 1 or 2 (Medtronic, Inc., Minneapolis, MN) set at a sample rate of 1 measurement every 5 minutes. The studies had been approved by the Institutional Review Board at the Joslin Diabetes Center [7, 8].

Colorized delay map

The algorithm for constructing the "glucose-at-a-glance" plots comprises 2 basic sequential steps: (1) constructing a delay map, and (2) color coding this map.

Delay Map Construction

The CGM data, used here consists of glucose measurements sequentially acquired at 5-minute intervals. The delay map is simply a plot of the i^{th} glucose value versus its $(i+1)^{th}$ value.

Delay Map Colorization

Each data point in the delay map (representing two consecutive CGM measurements) is assigned a color according to its density, calculated using a standard nonparametric technique [9, 10].
In the implementation used here, the color spectrum (given by the vertical bar on the right side of the graphs) ranges from a dark red-brown to a dark blue, where the former represents the most frequently occurring pairs of glucose values (G_i, G_{i+1}) and the latter the least frequently occurring ones. Additional technical details are provided in the Appendix, including an explanation of how these maps can be used to calculate the percentage of time that consecutive glucose values are within a given range (sub-region) of the delay map.

The delay map also provides insight into the structure of the variability of the CGM measurements. For example, uncorrelated outputs, such as white noise, yield delay maps with the appearance of circularly, symmetric scatter plots. In contrast, the delay maps of correlated output variables representing processes with feedback control show more complex patterns, as described below.

Results

To illustrate the basic principles of this methodology, we present original CGM time series and their respective colorized delay maps. For demonstration purposes, we show examples of data obtained from

the following individuals: 2 non-diabetic subjects (Fig. 5.1.1), 3 patients with diabetes with HbA1c values of 9.4% (Fig. 5.1.2), and 3 patients with diabetes with HbA1c values of 7.1% (Fig. 5.1.3).

The maps from both the non-diabetic subjects and the patients with diabetes have a stretched elliptical shape, a finding indicating that a given glucose value is followed (or preceded) by one of similar magnitude. The width of the ellipse measured perpendicularly to the diagonal line reflects the amplitude of the short-term (5 min in these cases) glucose fluctuations. Delay maps for measures 10 min far apart would have a slightly larger width. In fact, the width will increase as the time delay between consecutive glucose values expands.

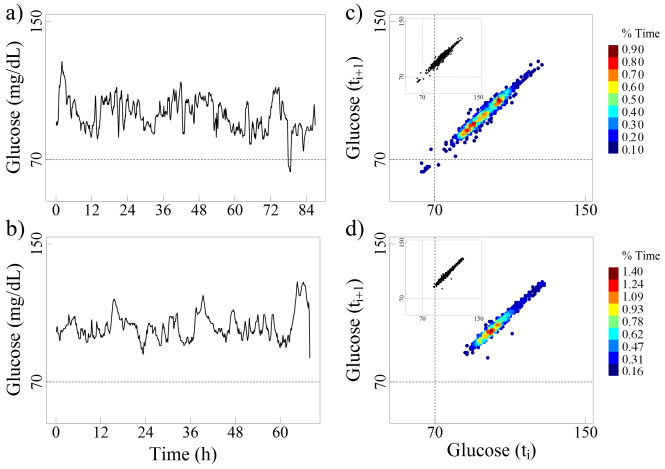

Figure 5.1.1: The left panels (a and b) present the glucose time series for 2 non-diabetic elderly subjects (82 and 76 years, respectively). The right panels (c and d) present their respective colorized delay maps, where the brown color indicates the most frequent pairs of glucose values and the blue color the least frequent ones. The insets display the traditional monochromatic delay maps.

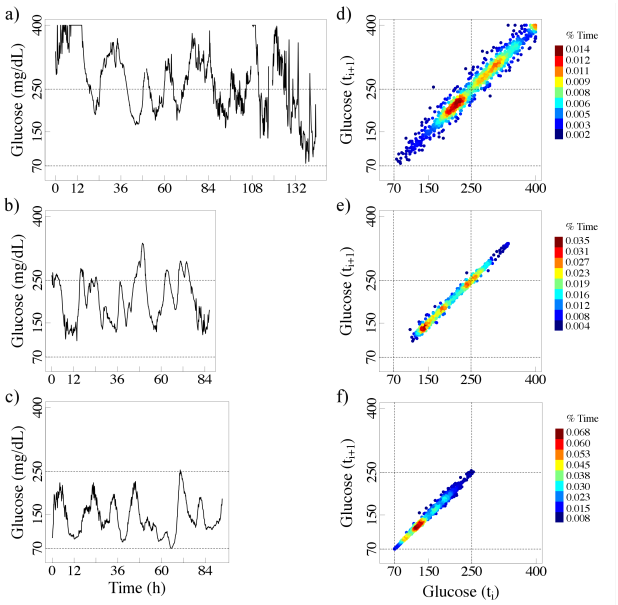

Figure 5.1.2: The left panels (a, b and c) show the glucose time series for three patients (76, 72 and 72 yrs, respectively) with 9.4% HbA1c values. The right panels (d, e and f) show their colorized delay maps.

To illustrate what the typical shape of the delay map is for a random sequence, we randomized the time

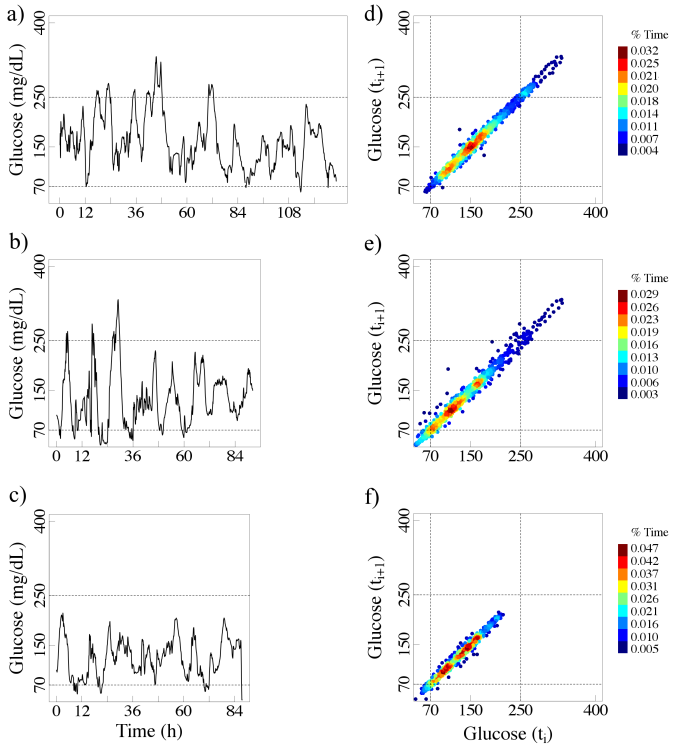

Figure 5.1.3: The left panels (a, b and c) present the glucose time series for three patients (73, 77 and 73 yrs), all with 7.1% HbA1c values. The right (d, e and f) panels present their colorized delay maps.

series from healthy subjects by shuffling the order of the glucose values. The time series for the randomized signals and their respective colorized density maps are presented in Fig. 5.1.4. Note the dramatic change in the shape of the delay map that becomes much more circular. This change is consistent with the fact that a given value is likely followed or preceded of another of (unrelated) magnitude.

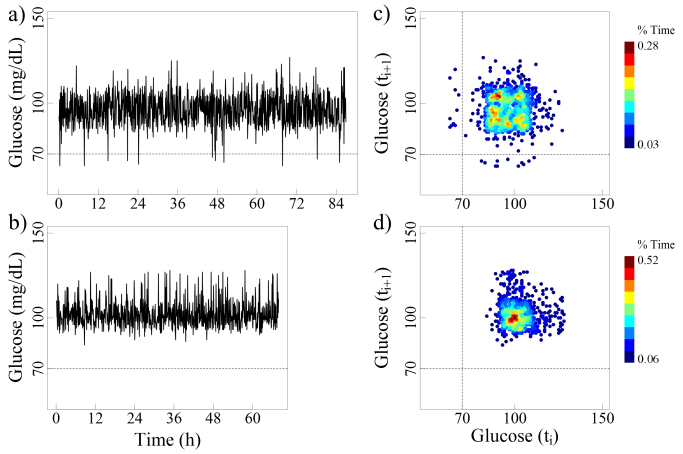

Figure 5.1.4: The left panels (a and b) show the randomized glucose time series values for the 2 non-diabetic subjects, shown in Figure 5.1.1. The right panels (c and d) show their colorized delay maps.

In the “real-world” examples shown here, as expected, the glucose values of healthy subjects fluctuate within a relatively narrow range (50-150 mg/dL). The delay maps for these healthy subjects show small,

well-circumscribed zones of increased density, representing "preferred" glucose values (Fig. 5.1.1). The glucose time series for the patients with type 2 diabetes present larger elliptical patterns, covering higher ranges of values compared to their healthy counterparts. Selected examples of the effect of noise on these delay maps are also presented in the Appendix.

Furthermore, patients with comparable HbA1c values can exhibit very different glucose fluctuation patterns. Note that in Figs. 5.1.2 and 5.1.3, the axes of the colorized delay maps cover a wider range of values than those observed in healthy subjects (Fig. 5.1.1). By way of comparison we present, in Fig. 5.1.5, the delay maps for the same two non-diabetic subjects using the wider axes (50 - 400 mg/dL).

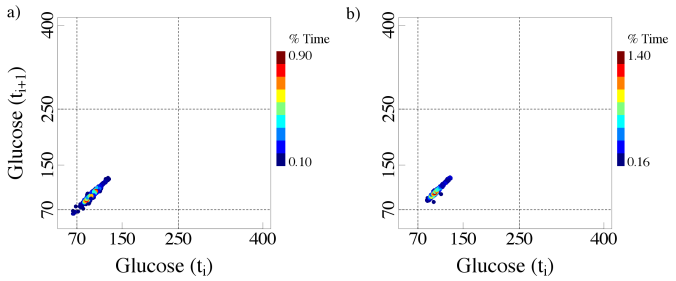

Figure 5.1.5: Colorized delay maps of time series of two non-diabetic subjects. Note that the difference between these panels and those presented in Figure 5.1.1 is the use of expanded axes ranges (50-400 mg/dL).

An apparent limitation of the density delay map method, as described here, is the fact that it does not give information about the time of occurrence of any given point (representing a consecutive pair of values). To add such information, a "point-and-click" adjunct can be incorporated which maps any point of interest in the delay map onto the original colorized time series (Fig. 5.1.6).

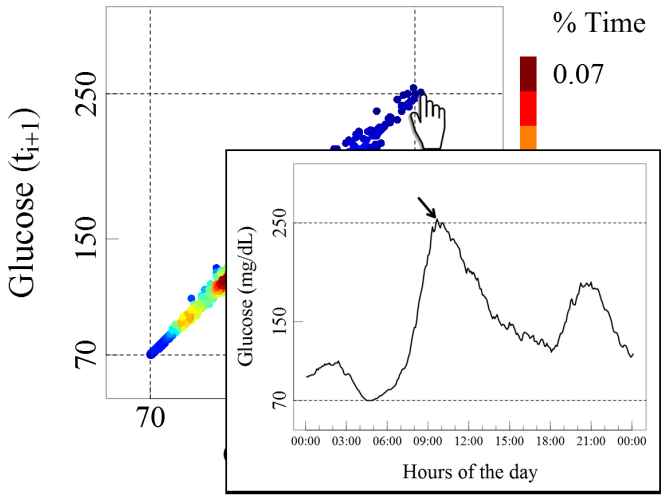

Figure 5.1.6: Example of a dynamic view of one of the colorized delay maps (Figure 5.1.3f) showing how the point-and-click option can be used to link any point on the delay map to its location on the original CGM time series.

Discussion

The “glucose-at-a-glance” visualization tool is a new way to display the complex, frequently sampled data acquired by CGM systems. The motivation is to enhance and facilitate assessment of glucose dynamics. Of note, the colorization based on the frequency of occurrence of sequential glucose values is of key importance in enriching the information provided by monochromatic delay maps (insets, Fig. 5.1.1). As described above, the latter have been widely used by investigators probing heartbeat dynamics [2–4] and suggested for exploring CGM data [5, 6]. To our knowledge, however, neither the colorized density maps, nor the point-and-click adjunct connecting these maps with the glucose time series (Fig. 5.1.6), have been previous introduced.

The analysis presented here, based on the density delay map method, shows that the differences in the glucose dynamics of non-diabetic subjects and patients with diabetes are encoded both in the amplitude of the analyte fluctuations and their temporal structures. In particular, the colorized delay maps of non-diabetic subjects show relatively small brown-yellow zones corresponding to sustained periods of stable glucose levels. In contrast, the patients with diabetes often show a single or multiple enlarged “smeared out brown-yellow zone indicating the absence of a stable baseline or the presence of multimodal instabilities, such that the glucose values appear to oscillate between different “attractors. Finally, this new visualization tool provides information complementary to the HbA1c values. As discussed above, the range and the structure of glucose variability may be very different for patients with comparable HbA1c values. The clinical implications of this graphically depicted instability remain to be determined.

Future Directions

The translational utility of the colorized delay map (“glucose-at-a-glance) method as a general visualization tool in both types 1 and 2 diabetes mellitus will require clinical testing. The method may also inform more basic work on models of glucose control in health and disease, since the output of such models should replicate the graphical representations shown here. We note that for this presentation, we used the shortest delay provided by a commercial system, corresponding to a sampling rate of 5-min. One can use longer delays depending on clinician preferences. For this demonstration, we also used the entire time series. However, the method can be applied to any segment of interest (e.g., daytime glucose values) provided that a reasonable number of points (of the order of 50 or more) is available. We did not have access to longitudinal studies. However, we anticipate that the colorized density delay map will change over time, depending on therapeutic interventions, diet, and so forth. Finally, the use of this class of density delay maps to develop and test new quantitative metrics of variability also requires prospective evaluation.

Conclusions

The “glucose-at-a-glance visualization tool, which is based on colorized delay (Poincar) maps, provides a way to facilitate the assessment of complex data acquired by CGM systems. This method yields dynamical information not contained in single summary statistics, such as HbA1c values, and may serve as the basis for developing novel metrics and models of glycemic control.

Appendix

Color-coding algorithms

In this implementation, we arbitrarily divided the interval of the density values into 9 equally spaced bins, each of which is assigned a different color according to a preselected chromatic scheme. In each figure the values next to the color bar legend represent the upper density value for each bin. The delay maps presented in this article were created in R [11] using the functions *smoothScatterCalcDensity* and *densCols* from the open source software package *grDevices*. The number of bins used was 250, and the bandwidth was calculated using the *smoothScatterCalcDensity* default formula.

Delay map: sub-region analysis

One can also quantify properties of selected subregions of the delay maps. As an example, consider two consecutive glucose values, 129 and 153 mg/dL, corresponding to points A and B in Figure 5.1.7a. The value 129 occurs 7 times and the value 153 occurs 9 times. Given that the original recording comprised 1000 data points, the relative-frequency of these values is 0.007 and 0.009, respectively. However, these values occur sequentially (Figure 5.1.7b) only 1 time, resulting in a relative-frequency of 0.001 (dark blue) for pair (A, B). To calculate the amount of time that pairs of consecutive measurements occur in a given region, we simply add the frequency of occurrence of each of the points that lie in that preselected region of interest. For example, the percentage of time that the glucose values of the subject depicted in Figure 5.1.7c is in the highlighted region (>250 mg/dL) is 22. As a more general example, consider that there were a region X in Figure 5.1.7c with 10 red-brown data points, 5 green, and 2 yellow. Taking into consideration the color bar legend with the percentage time for individual points, the estimated percentage time in that region X would be $(10 * 0.035) + (5 * 0.019) + (2 * 0.023) = 0.491$.

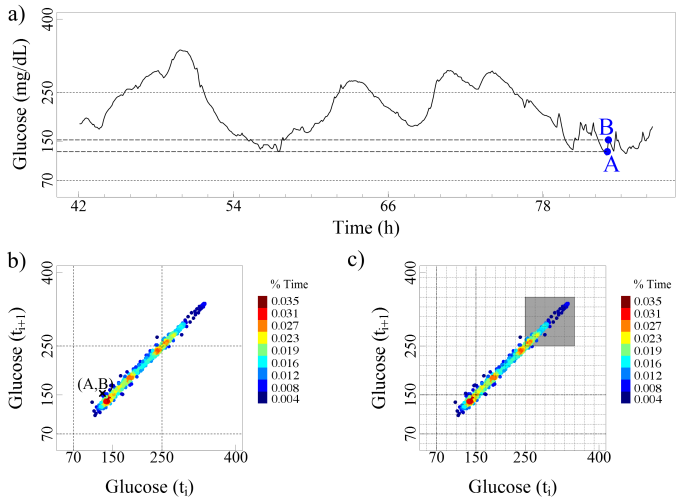

Figure 5.1.7: The top panel (a) display a part of the time series from one of the patient with diabetes (same as shown in Figure 5.1.2b). Two consecutive points, A and B are selected for illustration purpose. The bottom left panel (b) presents the delay map for the entire time series for the same subject and shows the location of the pair (A, B). The proposed method also allows for computation of the percentage of time that two consecutive glucose measurements are within a given selected range (e.g., gray sub-region presented in the bottom right panel).

Effect of Noise on Delay Maps

To help analyze the effect of noise of any source on the delay map method, we show examples of delay plots for one healthy subject and for two patients with diabetes before and after adding Gaussian white noise (mean = 0 and variance = 5 mg/dL) to the original data (Fig. 5.1.8). This degree of noise has only a minimal effect on the plots.

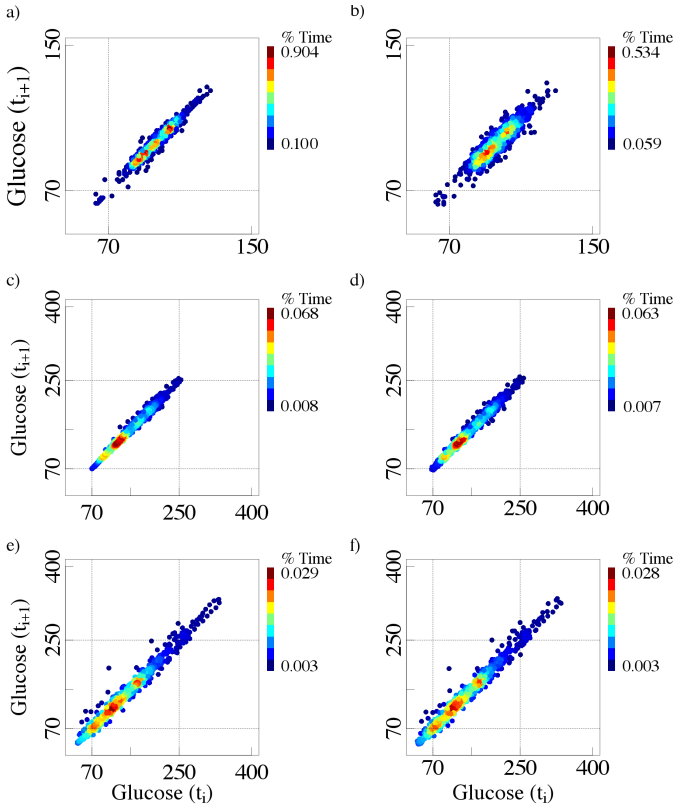

Figure 5.1.8: The panels on the left show the delay maps for the original CMG data and those on the right display the delay maps for the CMG data with white noise (mean=0 and variance=5 mg/dL). Each row represents a different subject. The plots in the first row were constructed using data from one of the healthy subjects (presented in Figure 5.1.1a). the plots in the other two rows were constructed using data from patients with diabetes (same as those shown in Figures 5.1.2c and 5.1.3b). The added noise does not have a prominent effect on the graphs.

Abbreviations

CGM, continuous glucose monitoring; HbA1c, hemoglobin A1c.

Authors' Note

MDC and ALG are joint senior authors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Portuguese National Science Foundation (grant SFRH/BD/70858/2010; TH). We also gratefully acknowledge support from the Wyss Institute for Biologically Inspired Engineering (ALG and MDC); the G. Harold and Leila Y. Mathers Charitable Foundation (ALG and MDC); the James S. McDonnell Foundation (MDC); and the National Institutes of Health (grants K99/R00 AG030677 [MDC] and R01GM104987 [ALG]).

Bibliography

[1] Holger Kantz and Thomas Schreiber. Nonlinear time series analysis, volume 7. Cambridge University Press, 2004.

[2] Peter W. Kamen and Andrew M. Tonkin. Application of the poincaré plot to heart rate variability: a new measure of functional status in heart failure. Australian and New Zealand Journal of Medicine, 25(1):18–26, 1995.

[3] Mary A. Woo, William G. Stevenson, Debra K. Moser, Robert B. Trelease, and Ronald M. Harper. Patterns of beat-to-beat heart rate variability in advanced heart failure. American Heart Journal, 123(3):704–710, 1992.

[4] Peter W. Kamen, Henry Krum, and Andrew M. Tonkin. Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. Clinical Science, 91(Pt 2):201–208, 1996.

[5] Boris P. Kovatchev, William L. Clarke, Marc Breton, Kenneth Brayman, and Anthony McCall. Quantifying temporal glucose variability in diabetes via continuous glucose monitoring: mathematical methods and clinical application. Diabetes Technology & Therapeutics, 7(6):849–862, 2005.

[6] William Clarke and Boris Kovatchev. Statistical tools to analyze continuous glucose monitor data. Diabetes Technology & Therapeutics, 11(S1):S–45, 2009.

[7] Medha N. Munshi, Alissa R. Segal, Emmy Suhl, Courtney Ryan, Adrianne Sternthal, Judy Giusti, Yishan Lee, Shane Fitzgerald, Elizabeth Staum, Patricia Bonsignor, Laura Desrochers, Richard McCartney, and Katie Weinger. Assessment of barriers to improve diabetes management in older adults a randomized controlled study. Diabetes care, 36(3):543–549, 2013.

[8] Medha N. Munshi, Alissa R. Segal, Emmy Suhl, Elizabeth Staum, Laura Desrochers, Adrianne Sternthal, Judy Giusti, Richard McCartney, Yishan Lee, Patricia Bonsignore, and Katie Weinger. Frequent hypoglycemia among elderly patients with poor glycemic control. Archives of Internal Medicine, 171(4):362, 2011.

[9] Matt P. Wand. Fast computation of multivariate kernel estimators. Journal of Computational and Graphical Statistics, 3(4):433–445, 1994.

[10] Matt P. Wand and M. Chris Jones. Kernel smoothing, volume 60. Crc Press, 1995.

[11] The R Development Core Team. R: A language and environment for statistical computing, 2009.

Dynamical Glucometry: Use of Multiscale Entropy Analysis in Diabetes

Madalena D. Costa, **Teresa Henriques**,Medha N. Munshi,
Alissa R. Segal and Ary L. Goldberger

Chaos
Volume 24, Issue 3, Page 033139
September 2014

5.2 Dynamical Glucometry: Use of Multiscale Entropy Analysis in Diabetes

Abstract

Diabetes mellitus (DM) is one of the world's most prevalent medical conditions. Contemporary management focuses on lowering mean blood glucose values toward a normal range, but largely ignores the dynamics of glucose fluctuations. We probed analyte time series obtained from continuous glucose monitor (CGM) sensors. We show that the fluctuations in CGM values sampled every 5 minutes are not random. Next, using multiscale entropy (MSE) analysis, we quantified the complexity of the temporal structure of the CGM time series from a group of elderly subjects with type 2 DM and age-matched controls. We further probed the structure of these CGM time series using detrended fluctuation analysis (DFA). Our findings indicate that the dynamics of glucose fluctuations from control subjects are more complex than those of subjects with type 2 DM over time scales ranging from about 5 min to 5 h. These findings support consideration of a new framework, *dynamical glucometry*, to guide mechanistic research and to help assess and compare therapeutic interventions, which should enhance complexity of glucose fluctuations and not just lower mean and variance of blood glucose levels.

Patients with type II diabetes mellitus (DM) need to closely monitor their blood glucose levels to manage the disease and help avoid its associated problems. Continuous glucose monitoring (CGM) systems were introduced into clinical practice approximately a decade ago. Using a sensor inserted under the skin, these systems provide glucose lev- els in tissue fluid at 5 min intervals for up to a week. The information extracted from these recordings for clinical use is primarily based on measures of the mean and range of glucose variability. One opportunity to improve our understanding and therapy of this highly prevalent disease is by looking for information encoded in the dynamical structure of glucose fluctuations. In this study, using techniques from statistical physics and nonlinear dynamics, we probed the multiscale complexity of glucose fluctuations obtained with CGM devices from elderly subjects with type II DM and from non-diabetic control subjects. We found that the glucose fluctuations in the patients with DM were significantly less complex compared to those in control subjects. Furthermore, we found that this loss of complexity could not be accounted for simply by changes in the average or range of glucose values. These findings are of interest because they raise the possibility of using information in glucose time series ("dynamical glucometry") in both laboratory and clinical settings. In particular, these findings support the concept that diagnostic and therapeutic assessments should not only take into account the mean and variance of glucose values but also measurement of properties related to their fluctuations over multiple time scales.

Introduction

Blood glucose control, central to physiologic fuel regulation, involves the complex interplay of hormones (e.g., insulin and glucagon), food intake, muscle activity, liver and fat cell function, among multiple other factors [1]. Type 2 diabetes mellitus (DM) is marked by hyperglycemia due to tissue resistance to insulin and to relative insufficiency of insulin secretion. The disease affects tens of millions worldwide. Complications include stroke, heart attack, blindness, limb loss, neurologic dysfunction and kidney failure. Because of its association with obesity, the incidence of type 2 DM is increasing at an alarming rate. Current management relies on diet and exercise as well as drugs that lower blood sugar toward a normal range, including insulin therapy [2]. However, type 2 DM remains a major challenge in terms of prevention and effective treatment.

The most widely used way of guiding day-to-day therapy is with blood glucose measurements obtained by finger sticks. In addition, the hemoglobin A1c (HbA1c) blood test provides information about glucose levels over the preceding several months [1, 2]. A more recent and minimally invasive modality uses continuous glucose monitoring (CGM) technology, with sensors implanted subcutaneously for about a week [2]. However, the time series derived from CGM recordings remain a largely untapped source of dynamical information. In clinical practice, the CGM indices of interest are primarily based on calculations of the percentage of time above and below given thresholds, range and average values of blood glucose. In a small number of research studies [3–6], which investigated glucose dynamics using detrended fluctuation analysis (DFA), no differences were noted between diabetic and non-diabetic subjects on relatively short (5-60 min) time scales.

This communication advances an alternative statistical physics approach to the analysis of CGM time series, primarily based on a complexity measure termed multiscale entropy [7]. The overall goal of this study is to extract information about the nature of physiologic vs. pathologic glucose control from CGM recordings. The specific hypotheses are that: 1) the noisy-appearing fluctuations in blood glucose levels in non-diabetic control subjects are complex, not random, and exhibit correlations over multiple time scales and; 2) blood glucose fluctuations are less complex in patients with type 2 DM compared with non-diabetic subjects.

Methods

We retrospectively analyzed CGM data from 18 patients with type 2 DM and 12 elderly control subjects without DM, enrolled in studies (approved by the Institutional Review Board) conducted at the Joslin Diabetes Center by the Geriatric Diabetes Research Group (Table 5.2.1). The CGM data were obtained using the iPro™ system version 1 or 2 (Medtronic, Inc., Minneapolis, MN), set at a sampling rate of one measurement every 5 min. The duration of the recordings varied between two and four days, therefore, the time series comprised between approximately 550 and 1100 CGM values. All patients with diabetes were on an insulin regimen with/without additional oral anti-diabetic agents. Fig. 5.2.1 shows selective CGM time series from a control subject and a patient with type 2 DM.

Results, unless otherwise stated, are reported as means ± standard deviation (SD). Group values were compared using a two-tailed, non-parametric (Mann-Whitney) test. Statistical significance was set at a p-value < 0.05. Coefficients of determination, r^2, derived from the linear regression of the

Variables	Non-diabetic	Type 2 DM
n	12	18
Gender (M/F)	3/9	6/12
HbA1c (%)	5.6 ± 0.2	6.8 ± 0.8
Age (yrs)	79.3 ± 5.0	76.0 ± 5.0
BMI (lb/in^2)	25.8 ± 6.93	31.0 ± 7.0

Table 5.2.1: Characteristics of control subjects and patients with type 2 DM. Data are expressed as mean ± SD.

dynamical variables (complexity index and α-exponent) against standard clinical measures of glucose control (HbA1C, means and SD of CGM time series) were computed.

Results

To assess whether the short-term (5-min) fluctuations in CGM time series were solely attributable to uncorrelated noise (instrumental and other), we first compared the structure of the time series of the differences between consecutive glucose values, termed first difference time series, with that of their shuffled surrogate time series [8] (Fig. 5.2.2). If the former were less entropic than the latter, one would infer that the structure of the original first difference time series was not random and, consequently, that there was information in the short-term CGM fluctuations.

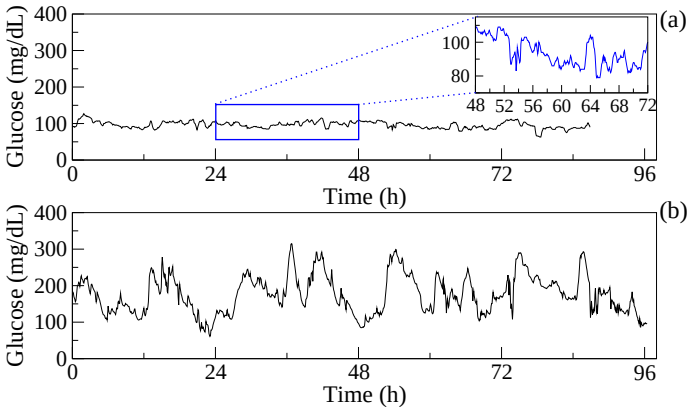

Figure 5.2.1: Time series of blood glucose derived from the continuous glucose monitoring (CGM) recordings of a non-diabetic control subject (a) and a patient with type 2 diabetes (b). The inset is a magnification of a 24-h data segment.

Thus, for each first difference time series, we generated 100 surrogate time series, obtained by shuffling the order of the data points in the original series. To quantify entropy, i.e., the degree of irregularity of the time series, we used the sample entropy (SampEn) algorithm [9]. SampEn is a conditional probability measure quantifying the likelihood that if a sequence of m consecutive data points matches (within a tolerance r) another sequence of the same length, then the two will still match when their length increases from m to $m+1$ data points. Here, SampEn was calculated with standard parameter values [9] $m = 2$ and $r = 0.15$. Regular and/or periodic time series have (theoretically) zero entropy and uncorrelated random time series have maximum entropy, whose value depends on the time series length [9].

Of note, using a percentage of a time series' SD as the r value for the computation of SampEn is equivalent to normalizing the time series to unit variance and zero mean prior to the computation of

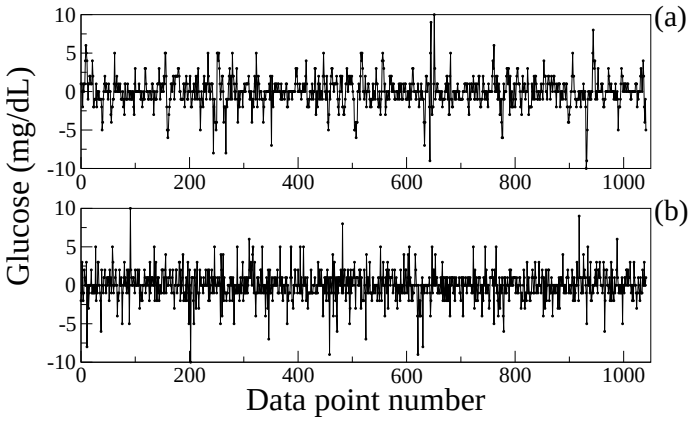

Figure 5.2.2: Time series of the differences between consecutive blood glucose values derived from the CGM recording of a control subject (a) and a surrogate signal (b) obtained by shuffling the temporal order in which these difference values appear.

SampEn. Therefore, differences in SampEn are not attributable to differences in the mean or SD of the signals.

Figure 5.2.3 shows that for each control subject, the 100 shuffled surrogate time series were consistently more entropic than the original first difference time series. Qualitatively similar results were obtained for the first difference time series of patients with type 2 DM.

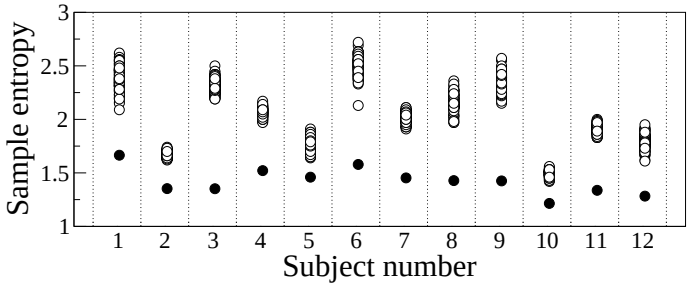

Figure 5.2.3: Sample entropy (SampEn) values for the first difference time series (solid circles) of non-diabetic control subjects and their shuffled surrogate (open circles) time series (n=100/subject).

These findings support the hypothesis that the short-term fluctuations in CGM values are not random. Accordingly, we next investigated the question of whether the glucose fluctuations over a range of time scales were more complex in control vs. diabetic subjects. For this purpose, we employed the MSE method [7, 10–12], which quantifies the degree of irregularity of a time series over a range of time scales. Briefly, the method comprises three steps: (1) use the original signal to derive a set of time series representing the system's dynamics on different time scales through a coarse-graining procedure, (2) calculate sample entropy (SampEn) [9] for each coarse-grained time series, and (3) compute a complexity index, defined here as the sum of SampEn values over a pre-defined range of scales. The coarse-grained time series for scale i is obtained by averaging the data points inside consecutive non-overlapping windows of length i slided over the original time series. The MSE curves are obtained by plotting the SampEn values for the coarse-grained time series as a function of scale. In this study, given the length of the recordings and the fact that glucose values were recorded every 5 min, we explored scales 1 to 6, where scale 1 corresponds to 5 min and scale 6 to 30 (5×6) min. With $m = 2$ ($m+1 = 3$), the SampEn values of the coarse-grained time series for scale 6, probe glucose fluctuation patterns that range from 60 (2×30) to 90 (3×30) min.

The complexity index derived from the MSE analysis of the glucose dynamics on scales 1 to 6 (Fig. 5.2.3) was significantly higher ($p < 0.004$) for the non-diabetic than the diabetic group (Fig. 5.2.5, left panel). On time scales ranging from 1 to 12, the complexity index remained significantly higher ($p < 0.007$) for the non-diabetic than the diabetic group. Since the number of data points in the coarse-grained time series for scale 12 is only half of that in the coarse-grained time series for scale 6, we used a less restrictive r value (0.25) in this analysis to avoid spuriously high entropy values due to the occurrence of none or very few matches. The results obtained are independent of differences between the two groups in the mean and SD values of the CGM time series.

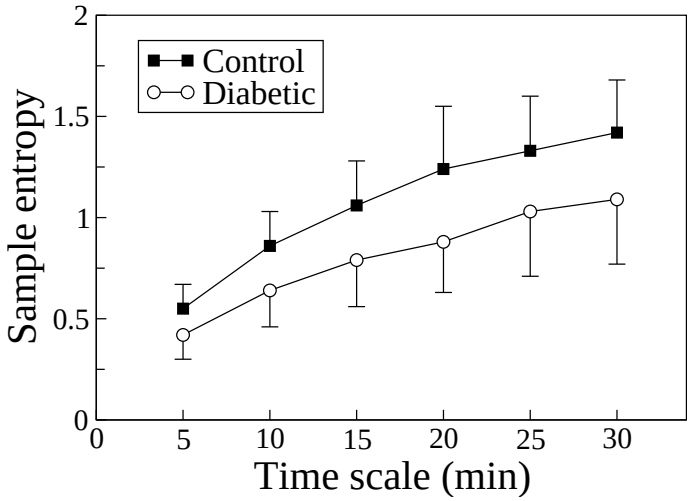

Figure 5.2.4: Multiscale entropy (MSE) analysis of glucose time series for the non-diabetic (n=12) and diabetic (n=18) groups. Symbols represent the mean values of entropy for each group at each scale and the bars represent the standard deviation (SD).

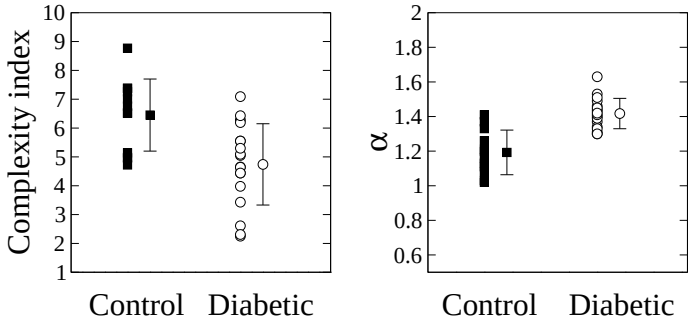

Figure 5.2.5: Complexity index and DFA α-exponent (time scales $>$ 1 h) for the non-diabetic and diabetic groups. Symbols with error bars represent mean and SD values, respectively.

To further characterize the dynamics of CGM time series and, in particular, to investigate their long-range correlation properties, we computed the fractal scaling exponents using the detrended fluctuation analysis (DFA) method [13]. Briefly, the algorithm, a modified root mean square analysis of a random walk, quantifies the relationship between $F(n)$, the root mean square fluctuation of an integrated and detrended time series, and the observation window size, n. $F(n)$ increases with window size according to $F(n) \propto n^{\alpha}$. If $\alpha = 0.5$, there are no correlations and the signal is white noise, i.e., uncorrelated randomness. If $0.5 < \alpha < 1$, the signal exhibit long-range temporal correlations. For the special case when $\alpha \simeq 1$ (1/f noise), the time series exhibits scale-invariant properties. If $\alpha = 1.5$, the time series represents a random walk (brown noise). α values ≥ 1.5 are associated with strong trends ("black noise").

For time scales shorter than one hour, there were no significant difference between the α-exponents of CGM time series from diabetic (1.87 ± 0.29) and non-diabetic subjects (1.86 ± 0.22). These results are consistent with those previously reported [3, 4] and indicate that the structure of the CGM time series on short time scales is very smooth and characterized, primarily, by monotonically increasing or decreasing trends. In contrast, we found that for time scales larger than one hour, (Fig. 5.2.5, right panel), the α-exponents were significantly ($p < 0.0001$) lower (1.19 ± 0.13), and closer to 1, in control subjects than in patients with type 2 DM (1.42 ± 0.09), respectively. These results indicate that the dynamics of glucose control in non-diabetic subjects over the specified range is fractal, i.e., time-invariant, whereas in this group of subjects with type 2 DM it is not different from brown noise.

Next, we investigated the relationship between the two dynamical variables of interest, the complexity index and the DFA α-exponent, and three clinical (static) measures of glucose control, HbA1c, mean and SD values of CGM time series (Table 5.2.2). We found significant correlations between higher glucose complexity indexes and lower (i.e., better) values of standard clinical measures of diabetes control (lower HbA1c and lower mean and SD of CGM glucose levels). We also found significant correlations between higher DFA α-exponents (i.e., those closer to 1.5 or brown noise) and poorer indexes of clinical glucose control. Alternatively stated, the DFA α-values closer to 1 were associated with better clinical control.

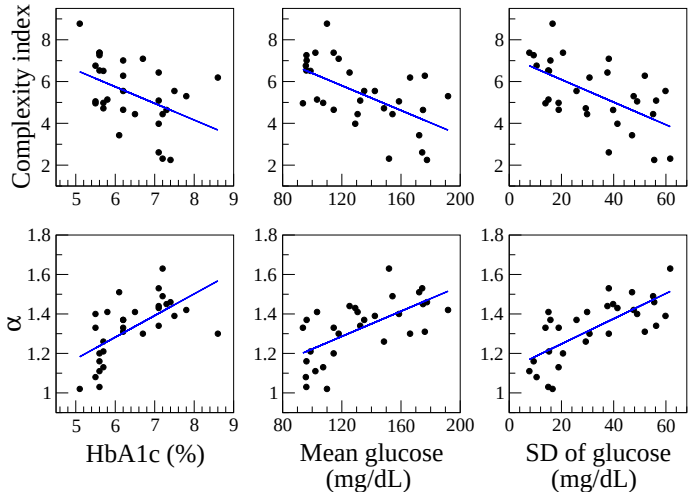

Figure 5.2.6: Linear regression of dynamical (complexity index (scales 1 to 6) and DFA α-exponent for time scales > 1 h) against clinical (HbA1c, mean and SD values of the CGM time series) measures of glucose control. Statistical summaries are presented in Table 5.2.2.

Variables	m	r^2	p value
CI vs HbA1c	-0.796	0.19	1.5×10^{-2}
CI vs mean	-0.030	0.33	8.3×10^{-4}
CI vs SD	-0.054	0.34	6.6×10^{-4}
α vs HbA1C	0.110	0.39	2.3×10^{-4}
α vs mean	0.003	0.41	1.3×10^{-4}
α vs SD	0.006	0.52	$< 10^{-5}$

Table 5.2.2: Slope m, coefficient of determination, r^2, and the two-tail p value [14] derived from the linear regression of the dynamical variables, complexity index and DFA α-exponent for time scales > 1 h, against three clinical measures, HbA1c, mean and SD values of the CGM time series.

Discussion

Our study is notable for a number of reasons. We find that the short-term fluctuations in serum glucose are not white noise (uncorrelated randomness), as shown by the analysis summarized in Fig. 5.2.3. Previous attempts to analyze the dynamical structure of the CGM time series on time scales < 1h using DFA failed to discriminate between diabetic and non-diabetic groups. This apparent lack of sensitivity may be due to the confounding influence of certain types of nonstationarities on DFA computations [15].

We report, for the first time to our knowledge, that the complexity of the CGM time series is significantly higher for the non-diabetic than the diabetic subjects, using MSE analysis. From a mathematical point of view these results indicate that the degree of irregularity of the CGM time series is higher for the non-diabetic than the diabetic group, across a physiologically wide range of time scales ($5--90$ min). The reduction in complexity with diabetes cannot be attributed to higher CGM values or higher SD of the CGM time signals. From a biologic point of view, these results, albeit part of a small, retrospective analysis, are consistent with the complexity-loss framework of disease and aging [10–12, 16], that subtends the notion that a breakdown in underlying physiologic control mechanisms results in less complex, i.e., either more random or more periodic, output signals.

Assessment of the nature of the correlations in the CGM time series, using DFA, on time scales larger or equal to one-hour, confirms and extends previous findings [3] of a significant increase in the α-exponent towards that of brown noise (1.5) with type 2 DM. These results are consistent with a loss of fractality in the dynamics of CGM signals. Furthermore, our findings are based on 2-4 day recordings, not solely on a single day of CGM data as previously reported [3].

Dynamical and conventional assays of glucose control are complementary. Higher complexity indices and fractal exponents (close to 1) were correlated with more salutary clinical measures of both short and long term glucose control, i.e., lower glucose mean and SD values and lower HbA1c. However, the correlations were modest, consistent with the fact that dynamical and the widely used clinical measures are likely not redundant.

Current therapeutic strategies focus almost exclusively on glucose lowering by searching for molecular "targets" involved in increasing sensitivity to insulin or insulin release by the pancreas. Our findings support extending the conventional notion of a discrete (pathway or signaling) therapeutic target to include the overall control "system as target." For example, two interventions may reduce blood glucose to a similar extent; however, one may do so by restoring healthy multiscale dynamics, the other may fail to do so. We would anticipate that the former might be preferable than the latter and that *dynamical glucometry*, defined here as the mathematical analysis, derived from statistical physics and complex systems, will help uncover "hidden" information in serum glucose time series of basic and clinical value.

We note that CGM time series obtained with greater precision and with higher sampling frequency than those available here may provide more information about glucose control in health and disease. The sampling rate employed in current clinical CGM studies obviates the possibility of probing glucose dynamic on scales shorter that 5 min, which earlier studies suggested are physiologically important [17, 18].

Future prospective studies may also help address additional questions, including: are serum glucose fluctuations in younger healthy subject more complex than in healthy elderly? Is loss of complexity associated with increased comorbidity related to multiorgan dysfunction? The concomitant use of continuous

heart rate monitoring in assessing diabetic therapeutics might add additional information about neuroautonomic function. Multimodal serial measurements of insulin and other molecular regulators involved in physiologic fuel control could also contribute important information.

Finally, our findings may be useful in developing and testing mathematical models of physiologic fuel regulation in health and disease.

Acknowledgments

This work was supported by the Portuguese National Science Foundation, though the PhD grant SFRH /BD/70858/2010. We also gratefully acknowledge support from the Wyss Institute for Biologically Inspired Engineering (ALG and MDC); the G. Harold and Leila Y. Mathers Charitable Foundation (ALG and MDC); the James S. McDonnell Foundation (MDC); the National Institutes of Health grants K99/R00 AG030677 (MDC) and R01GM104987 (ALG); and the Graetz Foundation (MM).

Bibliography

[1] American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care, 37 Suppl 1:S81–S90, 2014.

[2] American Diabetes Association. Standards of medical care in diabetes. Diabetes Care, 77 Suppl 1:S14–S80, 2014.

[3] Hitomi Ogata, Kumpei Tokuyama, Shoichiro Nagasaka, Akihiko Ando, Ikuyo Kusaka, Naoko Sato, Akiko Goto, Shun Ishibashi, Ken Kiyono, Zbigniew R. Struzik, and Yoshiharu Yamamoto. Long-range negative correlation of glucose dynamics in humans and its breakdown in diabetes mellitus. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology, 291(6):R1638–43, 2006.

[4] Hitomi Ogata, Kumpei Tokuyama, Shoichiro Nagasaka, Akihiko Ando, Ikuyo Kusaka, Naoko Sato, Akiko Goto, Shun Ishibashi, Ken Kiyono, Zbigniew R. Struzik, and Yoshiharu Yamamoto. Long-range correlated glucose fluctuations in diabetes. Methods of Information in Medicine, 46(2):222–6, 2007.

[5] Naomune Yamamoto, Yutaka Kubo, Kaya Ishizawa, Gwang Kim, Tatsumi Moriya, Toshikazu Yamanouchi, and Kuniaki Otsuka. Detrended fluctuation analysis is considered to be useful as a new indicator for short-term glucose complexity. Diabetes Technology & Therapeutics, 12(10):775–783, 2010.

[6] Hitomi Ogata, Kumpei Tokuyama, Shoichiro Nagasaka, Takeshi Tsuchita, Ikuyo Kusaka, Shun Ishibashi, Hiroaki Suzuki, Nobuhiro Yamada, Kumiko Hamano, Ken Kiyono, Zbigniew R. Struzik, and Yoshiharu Yamamoto. The lack of long-range negative correlations in glucose dynamics is associated with worse glucose control in patients with diabetes mellitus. Metabolism, 61(7):1041–1050, 2012.

[7] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[8] James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. Physica D, 58(1?4):77–94, 1992.

[9] Joshua S Richman and J Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–2049, 2000.

[10] Madalena Costa, Ionita Ghiran, Chung-Kang Peng, Anne Nicholson-Weller, and Ary L. Goldberger. Complex dynamics of human red blood cell flickering: alterations with in vivo aging. Physical Review E, 78:20901–1–4, 2008.

[11] Leopoldo C. Cancio, Andriy I. Batchinsky, William L. Baker, Corina Necsoiu, José Salinas, Ary L. Goldberger, and Madalena Costa. Combat casualties undergoing lifesaving interventions have decreased heart rate complexity at multiple time scales. Journal Critical Care, 28:1093–8, 2013.

[12] Madalena Costa, William T. Schnettler, Célia Amorim-Costa, João Bernardes, Antónia Costa, Ary L. Goldberger, and Diogo Ayres-de Campos. Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia. Early Human Development, 90:67–71, 2014.

[13] Chung-Kang Peng, Shlomo Havlin, H. Eugene Stanley, and Ary L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos, 5(1):82–87, 1995.

[14] Patrick Wessa. Linear regression graphical model validation (v1.0.7) in free statistics software (v1.1.23-r7), office for research development and education, 2012.

[15] Kun Hu, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. Effect of Trends on Detrended Fluctuation Analysis. Physical Review E, 64:011114, 2001.

[16] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71:021906, 2005.

[17] George E. Anderson, Yusuf Kologlu, and Constantin Papadopoulos. Fluctuations in postabsorptive blood glucose in relation to insulin release. Metabolism, 16(7):586 – 596, 1967.

[18] Arthur Iberall, M. Ehrenberg, Arthur S. Cardon, and M. Simenhoff. High frequency blood glucose oscillations in man. Metabolism, 17(12):1119 – 1121, 1968.

General Discussion and Conclusions

The scientist is not a person who gives the right answers,
he's one who asks the right questions.

Claude Lévi-Strauss

6. General Discussion and Conclusions

Clinical research is in constant evolution. The development of new technologies brought new data, new data brought new information, new information raised questions and answering these questions can help clinicians with patient management. The importance of this thesis relies in the latter step. The access to physiological control systems information, such as blood pressure, heart rate (HR), brain electrical activity and gait, measured on a moment-to-moment basis - that is, the access to the continuous behavior of physiological systems rather than their average value over some time period - led to the establishment of new mathematical concepts and methods allowing the detection of inherent dynamics of these systems.

Numerous nonlinear measures were proposed over time to analyze physiological time series fluctuations exploring concepts of chaos, complexity (randomness and correlation), fractality (self-similarity) and time irreversibility (symmetry), among others [1–3]. Previous studies consistently proved the loss of complexity [2, 4–10] and alterations of the fractal dynamics [11–16] in aging and disease.

The focus of this thesis is on the concept of Complexity, in particular in the two complexity measures: entropy and compression applied to physiological time series analysis. The interdisciplinarity of applying complexity theory approaches to biomedicine should be supported with the idea that no single measure should be used to assess the physiologic systems dynamics. Indices obtained from nonlinear methods combined with those of traditional linear ones (time- and frequency-domain), can provide useful information to a clinician leading to a better patient management.

The disagreement between clinicians (observers), aggravated by the growing information available, can be a setback when assessing a new index/measure to describe a pathology or an event. For this reason, the use of the appropriate measures to assess observer agreement is of extremely importance. In this work two measures of agreement, the information-based measure of disagreement (IBMD) [17, 18] and the raw agreement indices (RAI) [19], are proposed to evaluate both continuous and categorical data, respectively. Also two software platforms, a website an a R package were developed to facilitate the measures estimation.

One of the most studied physiological dynamics are the cardiovascular dynamics [3]. The HR fluctuations have been subject of a number of studies, as their relates with aging and disease. With that in mind, we compared the well known and described measures of entropy: the approximate entropy (ApEn) [20] and the sample entropy (SampEn) [21] with the "new" compression ratio metric, applied to human (fetal and adult) heart rate time series (**Chapter 4**).

Among the existent family of entropy measures, the ApEn and the SampEn, proposed to quantify the predictability of HR time series, are two of the most used in the analysis of physiological time series. In fact, ApEn is fundamentally a "regularity" statistics, does not quantify the fractal scaling behavior of a time series [1]. The SampEn, presented as an improving of the ApEn measure, was considered

the most adequate entropy measure for biological systems [22]. A limitation of these measures relates to the fact that they are parametric measures, and thus their results vary according to the parameters chosen (vector length m, tolerance value r and number of points of the time series - see **Chapter 2** for details). Actually, if r is chosen as a fixed value the results obtained are highly correlated with the time series amplitude (standard deviation). To assess the data correlations independently of the time series amplitude, a common choice is to use the r value as a percentage of the standard deviation of the time series. This choice has the advantage of being equivalent to normalizing the time series to unit variance and zero mean à priori, but the results still depend on the percentage r chosen. The number of points of a time series to compute the SampEn should be larger than approximately 750 [21]. This can be a limitation when analyzing time series with low frequency sampling. Attention is also needed when comparing results from time series with total number of points of different order of magnitude.

In this work, we showed that both entropy and compression measures allow for the characterization of different pathological conditions supporting the notion of complexity decreasing in disease. Another main finding of this thesis is that of using entropy and compression approaches can quantify different features of a system complexity despite both try to quantify the time series complexity.

The use of compression as a measure of complexity has rarely been applied in the analysis of biological signals. However, compressors can be effectively used as an alternative to the widely used entropy measures to quantify complexity in biological signals [23]. The use of compressors has the advantages of not require a minimum of data and can be compute easily and efficiently. Other advantage of using compressors is the possibility of quantifying complexity in another type of data as sound, image and video. The best compressor to assess the dynamics of physiological time series is yet to be established. Future studies should address the questions of which are the strengths and limitations of each compressor and if their applicability is preserved in different types of data. During this work we also tried to use a measure known as logical depth that combines the size of the compression with the running time of the compressor. The results were not promising so we did not explore this line further.

One of the problems of applying, in a clinical context, the measures widely studied in other scientific fields is owed to the existing noise on the everyday clinical data. The multiscale entropy approach (MSE) [24] probes long-range correlations of the time series adding information to the entropy measure transforming a “regularity” metric into a fractal complexity metric. Although this approach was originally applied combined with entropy, the application goes far beyond and can be extended to several others measures. On **Section 4.2** the multiscale approach was successfully combined with measures of compression in the assessment of HR dynamics. The multiscale approach, due to its property of capturing the information across multiple spatial and temporal scales, in particular due to the coarse-grained procedure, can work as a filter and in a few scales some noise is reduced and the values obtained by the measures are “adjusted”. The multiscale approach permit to distinguish noisy data from data collected from a patient with a pathology or from a healthy subject. The number of scales depends on the number of data points from the original time series. In the case of entropy, the coarse-grained time series obtained in the maximum scale should have more than 200 points. This can be a limitation of MSE when analyzing short time series.

The utility of the nonlinear methods presented or used to assess heart dynamics is not, generally, restricted to the analysis of these time series. On **Chapter 5** some methods previously used to explore the heart beat dynamics were successfully applied to the analysis of continuous glucose monitoring (CGM)

time series. The increase of information available to the clinicians led to a growing concern with the rapid and easy visualization and assessment of the data available. Here was proposed a new visualization method based on delay maps and enhanced by adding color representing levels of density [25]. Also in this chapter the results of exploring the metrics of MSE and detrended fluctuation analysis (DFA) applied to CGM data are consistent with the complexity-loss framework of disease and aging. The *method's universality* have to be tested carefully and the results should be thoroughly interpreted. Future prospective studies may also help to address additional questions of the methods applicability in different physiological time series.

In conclusion, this thesis provides several contributions towards the measurement of disagreement and the evaluation of complexity in physiological signal. The confined number of signals in each dataset is a thesis limitation, and thus the obtained results should be supported with further clinical testing. Finally, our findings may be useful in developing and testing mathematical models of physiologic regulation in health and disease. The degree of complexity in physiological dynamics should be considered as a therapeutic target in addition to the current ones.

Bibliography

[1] Ary L. Goldberger. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. The Lancet, 347(9011):1312–1314, 1996.

[2] Ary L. Goldberger, Chung-Kang Peng, and Lewis A. Lipsitz. What is physiologic complexity and how does it change with aging and disease? Neurobiology of Aging, 23(1):23–26, 2002.

[3] Lewis A. Lipsitz. Physiological complexity, aging, and the path to frailty. Science's SAGE KE, 2004(16):pe16, 2004.

[4] Ary L. Goldberger, Larry J. Findley, Michael R. Blackburn, and Arnold J Mandell. Nonlinear dynamics in heart failure: implications of long-wavelength cardiopulmonary oscillations.

[5] Lewis A. Lipsitz and Ary L. Goldberger. Loss of "complexity" and aging: potential applications of fractals and chaos theory to senescence. Jama, 267(13):1806–1809, 1992.

[6] Lewis A. Lipsitz. Age-related changes in the "complexity" of cardiovascular dynamics: A potential marker of vulnerability to disease. Chaos: An Interdisciplinary Journal of Nonlinear Science, 5(1):102–109, 1995.

[7] David E. Vaillancourt and Karl M. Newell. Changing complexity in human behavior and physiology through aging and disease. Neurobiology of Aging, 23(1):1–11, 2002.

[8] Madalena Costa, Ary L Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2):021906, 2005.

[9] Madalena Costa, Ionita Ghiran, Chung-Kang Peng, Anne Nicholson-Weller, and Ary L. Goldberger. Complex dynamics of human red blood cell flickering: alterations with in vivo aging. Physical Review E, 78(2):020901, 2008.

[10] Madalena D Costa, William T. Schnettler, Célia Amorim-Costa, João Bernardes, Antónia Costa, Ary L. Goldberger, and Diogo Ayres-de Campos. Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia. Early Human Development, 90(1):67–71, 2014.

[11] Ary L. Goldberger, Valmik Bhargava, Bruce J. West, and Arnold J. Mandell. On a mechanism of cardiac electrical stability. The fractal hypothesis. Biophysical Journal, 48(3):525–528, 1985.

[12] Ary L. Goldberger and Bruce J. West. Fractals in physiology and medicine. The Yale Journal of Biology and Medicine, 60(5):421, 1987.

[13] Mark J. Woyshville and Joseph R. Calabrese. Quantification of occipital EEG changes in Alzheimer's disease utilizing a new metric: the fractal dimension. Biological Psychiatry, 35(6):381–387, 1994.

[14] Jeffrey M. Hausdorff, Susan L. Mitchell, Renee Firtion, Chung-Kang Peng, Merit E. Cudkowicz, Jeanne Y. Wei, and Ary L. Goldberger. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. Journal of Applied Physiology, 82(1):262–269, 1997.

[15] Ary L. Goldberger, Luis A.N. Amaral, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Chung-Kang Peng, and H. Eugene Stanley. Fractal dynamics in physiology: alterations with disease and aging. Proceedings of the National Academy of Sciences, 99(suppl 1):2466–2472, 2002.

[16] Bruce J. West. Fractal physiology and chaos in medicine, volume 16. World Scientific, 2013.

[17] Cristina Costa-Santos, Luís Antunes, André Souto, and João Bernardes. Assessment of disagreement: A new information-based approach. Annals of Epidemiology, 20(7):555 – 561, 2010.

[18] Teresa Henriques, Luís Antunes, João Bernardes, Mara Matias, Diogo Sato, and Cristina Costa-Santos. Information-based measure of disagreement for more than two observers: a useful tool to compare the degree of observer disagreement. BMC Medical Research Methodology, 13(1):1–6, 2013.

[19] John S. Uebersax. Statistical methods for rater and diagnostic agreement. `http://www.john-uebersax.com/stat/raw.htm`. Accessed: Jun 2014.

[20] Steven M. Pincus. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences, 88(6):2297–2301, 1991.

[21] Joshua S. Richman and J. Randall Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, 2000.

[22] Robert L. Williams and Ismet Karacan. Sleep disorders: diagnosis and treatment. Wiley Medical Publication. Wiley, 1978.

[23] Teresa Henriques, Hernâni Gonçalves, Luís Antunes, Mara Matias, João Bernardes, and Cristina Costa-Santos. Entropy and compression: two measures of complexity. Journal of Evaluation in Clinical Practice, 19(6):1101–1106, 2013.

[24] Madalena Costa, Ary L. Goldberger, and Chung-Kang Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6):068102, 2002.

[25] Teresa Henriques, Medha N. Munshi, Alissa R. Segal, Madalena D. Costa, and Ary L. Goldberger. "glucose-at-a-glance" new method to visualize the dynamics of continuous glucose monitoring data. Journal of Diabetes Science and Technology, page 1932296814524095, 2014.

U.PORTO

FMUP FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO