COMPLEXITY AND NONSTATIONARITY IN SHORT-TERM NONLINEAR TIME SERIES. NEW METHODS FOR CARDIOLOGICAL DIAGNOSTICS

by

Rafal Ladysz
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Information Technology

Committee:

Vikonomidou	Dr. Vasiliki Ikonomidou, Dissertation Director
	Dr. Jessica Lin, Committee Member
Daniel Carr	Dr. Daniel Carr, Committee Member
	Dr. Ernest Barreto, Committee Member
	Dr. Stephen Nash, Senior Associate Dean
Kenneth S. Ball	Dr. Kenneth S. Ball, Dean, Volgenau School of Engineering
Date: 4/29/2014	Spring Semester 2014 George Mason University Fairfax, VA

The Volgenau School of Engineering

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Rafal Ladysz
Master of Science
Adam Mickiewicz University, 1987

Director: Vasiliki Ikonomidou, Assistant Professor
Electrical and Computer Engineering

Spring Semester 2014
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

Without involvement and kindheartedness of those mentioned below all my efforts would be fruitless and my objective never achieved. First and foremost, my thanks go to my advisor, Dr. Vasiliki Ikonomidou, and my co-advisor, Dr. Ernest Barreto for their professional expertise, paramount patience and tireless devotion combined with friendly and humane attitude. The two other members of my Committee, Dr. Jessica Lin and Dr. Daniel Carr, helped me notice different aspects of my research, and my former advisor, Dr. Daniel Barbará, inspired my interest in the methods applicable to nonlinear-dynamical phenomena. I feel deeply indebted to Dr. Andre Manitius for encouraging me to enroll at GMU, and for his comprehensive help and important advice during my whole academic pursuit. My gratefulness extends to Dr. Paul So and Dr. Tim Sauer (both from GMU) for finding time for me, and Prof. Dr. Jacek Kubica, M.D. with his team from N. Copernicus University for providing me with the experimental data being subject of my doctoral dissertation. Lisa Nolder from the Graduate Student Affairs, was exceptional helpful and kind when I needed her help.

Among my friends who enabled me to go through and successfully complete my study – especially during the last year of it – are Ania & Peter from Vienna, Maria & Mirek from Richmond, NY, and Ela & Szymon from Warsaw. Last but not least I want to thank my beloved wife Justyna for her patience, faith and support excerted for years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE VOLGENAU SCHOOL OF ENGINEERING

Rafal Ladysz, Ph.D.

George Mason University, 2014

Dissertation Director: Dr. Vasiliki Ikonomidou

The cardiac activity can be investigated based on the RR signal – a series of temporal intervals between consecutive heartbeats. The variation of these intervals – called the heart rate variability (HRV) – enables quantitative analysis of functioning of the cardiac control mechanism (Autonomous Nervous System). The mainstream techniques of the HRV analysis are time-consuming and do not identify conditions possibly affecting the HRV, what limits their analytical scope and depth. Hence the demand from cardiological community to develop reliable methods of HRV assessment working in nearly-real time and based on relatively small amount of data, yet at the same time providing nontrivial insight into the cardiac activity. We address this problem by investigating the dynamical changes in nonstationary RR signal, extracted from electrocardiogram recorded in presence of controlled environmental stimuli, including music. For this purpose we introduced the *Sample Entropy*-related methods to quantify complexity in time series, which we applied independently as (i) change detectors and (ii) classification features.

Furthermore, we propose and demonstrate methods of symbolic analysis of the RR signals based on the notion of *Lempel-Ziv complexity*.

Our research has laid the foundation for using novel nonlinear-dynamical statistics implemented as change detection algorithms applicable to the HRV analysis. We have shown that the new methods are sensitive enough to capture effects of subtle stimuli – such as music – on the HRV characteristics, while not compromising robustness to noise and experimental artefacts. Such techniques can find application in a variety of domains beyond cardiology, where the identification of change can be a starting point for the stress detection. In a wider perspective, the methods are potentially applicable to detecting nonstationarity in the systems whose dynamical parameters drift over time.

1. INTRODUCTION

Live organisms generate outputs resulting from interactions which involve a variety of internal and external (i.e. environmental) factors. The effect of these interactions can be found in heartbeats, neural excitations, respiratory functions and other vital activities. Intuitively, one may claim that the higher the complexity of a biological system, the richer the repertoire of adequate responses to the environmental conditions and – consequently – the better the fit-to-survive behavior. Conversely, low complexity and weak behavioral dynamics usually indicate inadequate responsiveness to environmental changes. Indeed, we observe that certain diseases or even healthy aging strongly correlate with less complex physiological outputs [94], which can be explained by compromised or disconnected regulatory mechanisms.

As an illustration, the two plots in Figure 1 represent series of heartbeat intervals – referred to as RR signals - retrieved from two different electrocardiograms. Clearly, the blue (‘*normal*’) signal is much more complex than the red one (‘*patient*’). In this particular example the extreme simplicity of the red plot represents a transplanted (denervated) heart, which does not receive any input from the autonomic nervous system.

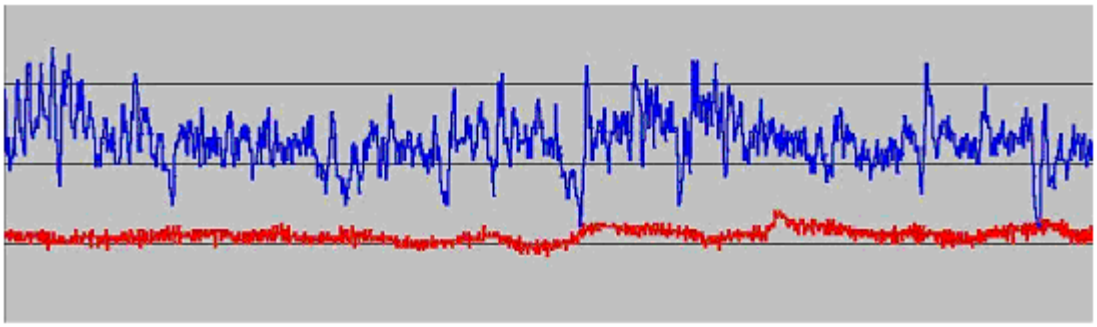


Figure 1. The series of heart beat intervals of normal (blue) and diseased (red) subjects (the values on x- and y-axis are not shown)

These observations would be pointless without the ability to evaluate signal complexity in a precise and objective manner. For this purpose one captures the signal representing the process of interest in living organisms, e.g. the electrocardiogram (ECG), electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG) or photoplethysmogram (PPG). Then, the recording is digitized into an appropriate data format suitable for quantitative, audio or visual analyses. In this research, we use exclusively time series data.

General considerations

The high complexity of physiological signals results from intrinsic for live organisms *concurrency* and *interrelations* of their underlying processes (e.g. blood circulation *and* breathing). These, in turn, entail what we refer to, and introduce further, as dynamical *nonlinearity* and *nonstationarity*. For now, we can intuitively think of them as non-additivity of different physiological influences and adaptive responsiveness to those influences. When observed during sufficiently long period of time, these signals exhibit

also *multiscale* variability and long-range *correlations* [57]. Importantly, their dynamics typically contain both *deterministic* and *stochastic* components [175], whose contributions depend on particular processes: for instance, the cardiovascular dynamics has more regularity than in the dynamics of human gait, what means that the latter is more stochastic (less deterministic) than the former. All these apply to the cardiovascular system, necessitating carefully selected methods and well-prepared data for adequate analysis.

Indeed, within the last two decades numerous *nonlinear-dynamical* and *entropic* methods [177] (whose detailed descriptions are provided in Chapter 2) unveiled interesting facts about the function of the cardiovascular system in a variety of physiological, pharmacological, behavioral and environmental conditions [118]. Among those methods, two entropic regularity measures, *Approximate Entropy* [119] and *Sample Entropy* [127] proved their relevancy for HRV analyses of variety of clinical cases, including but not limited to those mentioned on pages 15 ff.

The methods introduced in this research can be applied to detect to dynamical changes in the RR intervals and targeted for HRV analysis, yet their applicability may extend to domains beyond cardiology as well, what we state in Chapter 6. Below we discuss some typical challenges, opportunities and subtleties the researcher must be aware of and will likely encounter once involved in analyses of complex physiological signals in general and the RR time series in particular.

Linearity vs. nonlinearity. There is abundant evidence that traditional analytical tools – sometimes called *linear* – have limited applicability to physiological signals [84]. The

linear methods work best for purely stochastic and/or (quasi-)periodic processes, whereas most physiological time series are strongly nonlinear and exhibit certain degree of determinism.

Secondly, for capturing effects of *nonlinear dynamics*, it is recommended to use a multidimensional, vector-based approach [92], whereas the methods based on *morphological* similarity (distance) between time series – even though they account for sequential order in the data – operate on scalars (trace plots) rather than vectors (*state space*). To explain the advantage of vector-based approach, we consider two signals: the deterministic-chaotic *Henon* (discrete) map and the *Gaussian* stochastic signal, described by the probability distribution $P(X) = \exp(-0.5((X-\mu)/\sigma)^2)/\sigma(2\pi)^{0.5}$. The concept of a (multidimensional) state space and the *Henon* map needs a brief explanation: The *Henon* map is a function defined on the 2-dimensional real space with two control parameters a and b: $X_{n+1} = 1 - aX_n + Y_n, Y_n = bX_n$. It was initially studied [46] for $(a,b) = (1.4, 0.3)$, i.e. in its chaotic regime (There are also regions of the values of a and b for which the map is intermittent or periodic).

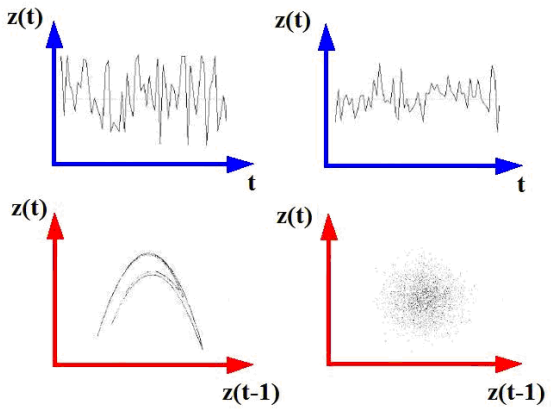


Figure 2. The upper trace plots represent chaotic (left) and stochastic (right) signal, respectively (blue). The lower: Time-delayed phase plane plots representation of the same data (red)

The upper panels of Figure 2 show 1-dimensional trace plots of the *Henon* map (left) and a sequence of normally distributed, randomly generated points (right). In the scalar representation – i.e. using trace plots – these signals look similar, but the visual-only judgment based on the trace plots may be misleading. This is because – in general – most dynamical systems have more than one degree of freedom (i.e. independent variable), regardless of what we are able to observe. Furthermore, we can usually associate only *one* of those variables with the time evolution of the system and – using adequate measurement procedures – acquire as a series of observations. To gain more comprehensive information about the dynamics of the underlying phenomenon, we may need to *reconstruct* the related state space according to Takens' *embedding theorem* [140] (further extended by [174]). The theorem states that a (deterministic) dynamical system can be reconstructed by transforming the sequence of observations into state vectors,

whose components span the system state space. For the reconstruction one needs two *embedding parameters*: dimension (m) and delay (τ). Although no precise and universal formulae for the parameters exist, certain semi-empirical methods can be used for approximation. We implicitly use these parameters when defining the *SampEn*-related statistics: m plays a role of pattern (template) length and τ can be thought of as sampling rate (by default τ = 1). In this sense our methods conform to the multi-dimensional paradigm: we realize this concept by denoting each single observation of the time series of interest as $u(i)$, i=1,2,...,N. Next, we build a series of $N-(m-1)\tau$ m-dimensional vectors $x_m(i)=\left[u(i),u(i+\tau),...,u(i+(m-1)\tau)\right]$, where i=1,...,N-m+1, m=1,...,M is the *embedding dimension* and $\tau \geq 1$ is the *delay* (or lag).

Now we can see the relevancy of the multi-dimensional approach by looking at the lower panels of Figure 2. When plotted in a recurrent manner in 2-dimensional state space: ($x(t)$ vs. $x(t-\tau)$, where τ=1), the signals show a striking difference: The new representation of the *Henon* map exhibits its regularity (which is hidden when represented in one dimension using a trace plot), which stems from its intrinsic determinism, whereas the *Gaussian* signal is purely stochastic (one can only try to predict its values based on the data probability distribution). Clearly, for the *Gaussian* signal the way it is represented dos not matter since it lacks regularity.

The deceptiveness of evaluating nonlinear (physiological) time series just by looking at the trace plots is illustrated by the following example of the four RR signals selected from the PhysioBank (publicly available web-based repository of data addressed to the

biomedical research community). In Figure 3, the red and turquoise lines on the plot look very much alike (having similar mean, variance, trend etc.), yet they represent different classes: *congestive heart failure* and normal sinus rhythm (i.e. '*normal*', respectively, whereas the blue and yellow – although looking fairly different – fall into the same class of 'normal').

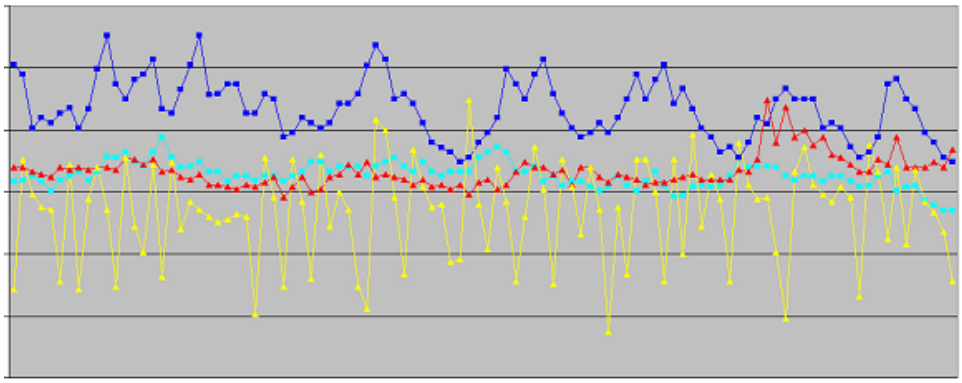


Figure 3. Points colored blue and turquoise represent normal (*healthy*, both marked with squares) cases, while red and yellow points represent cases of CHF (*diseased*, both marked by triangles). The x- and y-axis represent successive indices and durations of the RR intervals, respectively (their values are not shown)

The conclusion is that a visual-only evaluation of the RR trace plots may not be adequate for the problem of identification of cardiac diseases in general, especially in their early stage or when their specific features are blurred (untypical cases). In such circumstances one needs more sophisticated analytical tools, including those based on multi-dimensional and nonlinear-dynamical approach.

Data: the quality and quantity. In spite of the widely spread enthusiasm of early 1990's, certain mainstream nonlinear-dynamical methods, especially those related to chaos (e.g. the largest Lyapunov exponent (L1) and correlation dimension (D2)) – while powerful – are hardly applicable to the HRV analyses. This is because the methods theoretically require signals which are *stationary*, infinitely (or sufficiently) long and *noise-free* – conditions be satisfied by experimental data. Approximations of those requirements may be acceptable in certain situations; nevertheless using the chaos-related methods only is overall risky [90].

The biggest challenge is posed by nonstationarity, which is omnipresent in natural phenomena – particularly in live organisms – and is the results of adaptivity to changing environmentaal conditions (this is oversimplistic but true). In general, nonstationarity severely limits the arsenal of methods to be used for analysis (e.g. spectral and chaos-related), but on the other hand – what we show further – nonstationarity can also be informative about dynamics of the underlying process (given the adequate methods are used). The next problem is related to noise, whose level can be controlled only to certain degree. We demonstrate in chapter 5 when we introduce methods of symbolic analysis that the adverse influence of noise can be mitigated by converting the experimental signal to a sequence of symbols, which can be studied using methods of symbolic dynamics.

The signal length is another factor to take into account. Common sense dictates that '*the more* (data) *the better*'. Intuitively, a long-term signal usually contains more information than its short-term counterpart, given both were registered in comparable circumstances. On the other hand, a variety of technical and cost-related reasons leave the researcher

with only short-term signals available for analysis, with potential problems typical for data shortage. We address this problem indirectly by applying series of stimuli during the ECG recording.

Nonstationarity: intrinsic and stimuli-induced. Our analytical approach is based on the conjecture that by responding to controlled environmental stimuli, the cardiovascular system generates *additional* informational about the perturbed ECG. Such a stimuli-based approach was motivated by abundant experimental evidence [120] supporting the hypothesis that physiological signals generated by healthy organisms feature a higher complexity level than those generated by older and/or diseased organisms. Obviously, we assume that the responses to external stimuli should be reflected in the complexity of the RR signals, especially when measured locally in time when the stimuli were applied. Such a setting opens up the opportunity to analyze the HRV complexity by capturing and quantifying *dynamical changes*, perhaps caused and/or amplified by the stimuli. Exact knowledge of temporal occurrence (onset and duration) of particular stimuli enables comparing different parts of the signals in terms of characteristics we believe are sensitive to the stimuli. What is more, we can evaluate *differences* in the level of those responses exhibited by healthy vs. diseased subjects. This aspect will be thoroughly exploited further in this dissertation.

All these lead to the problem of *nonstationarity*, inherently present in all physiological signals. In physics of *dynamical systems*, an alteration of dynamical properties is called a *parameter change* (e.g. transient chaos or intermittency), where values of those parameters are usually known and controlled during simulation. In the case of

experimental time series, however, their underlying dynamics are in a sense a *black box*, what leaves us with semi-empirical rather than analytical methods. In either case, nonstationarity causes computational difficulties, especially when applying the spectral methods and chaoticity measures or when attempting to *simultaneously* fulfill the requirements for minimum signal length and stationarity. This is because the longer duration of a process, the greater chance its dynamics may change. For instance, during 24 hours of the Holter monitoring, the recording device captures ECG data of whole diurnal activity. During this time physical and physiological conditions drastically change (e.g. sleep vs. physical exercises, stress at work vs. leisure at home). Conversely, selecting relatively short subsequence of the recorded time series – when dynamics of the process do not change significantly – help obtain quasi-stationarity, this however compromises the amount of data. Ultimately, each of the requirements can be satisfied at the cost of another.

Determinism vs. stochasticity. *Coexistence* of stochasticity and determinism is typical for physiological time series. Such a hybrid nature should make the researcher cautious while selecting method(s) for analysis. For instance, the choice of purely *nonlinear/deterministic* or *linear/stochastic* to be used in isolation from other methods in most cases may not be very useful [140]. Particularly important for the methods applied to physiological data is their *robustness to noise* and outliers without compromising sensitivity to (perhaps nonlinear) dynamical characteristics. Entropic methods meet the above criteria.

Complexity. A last remark concerns the meaning of the term *complexity*, used abundantly throughout this dissertation. As a general concept, complexity has different meanings depending on the context it is used. In information theory and computer science the term is related to '(in)*compressibility*' [123]. In physics (of nonlinear dynamical system) complexity is usually quantified by the (topological) entropy and accounts for the (proliferation of) periodic orbits. Other examples can be taken from the areas of corporate management, social networks, biology and physiology, communication and many others. In this study the term denotes (i) *regularity* in data enabling *compression* (quantified by the algorithmic complexity and its *Lempel-Ziv* complexity) and (ii) *prediction* (quantified by the Kolmogorov-Sinai entropy and its *Sample Entropy* implementation).

Functioning and complexity of the human heart

The human heart is a muscle made of self-excitable tissue, functioning in a cyclic manner as follows: It pumps the blood to the lungs to absorb oxygen, then it pumps this oxygen-saturated blood to the organism (via blood vessels). Once the blood supplies the oxygen to organism it returns to the heart (via veins). This cyclic process is synchronized with the heart rate. Heartbeats – the contractions of the heart muscle are controlled by low-voltage electrical signals normally generated the sinoatrial (SA) node – the heart's natural pacemaker. The SA node-generated electrical impulses are spread to both atria and cause them to contract (pump). Another controlling center, the atrioventricular (AV) node, has its role in facilitating contraction of the atria before ventricles contract, thus helping the latter filled with blood. Figure 4 below shows simplified structure of the human heart.

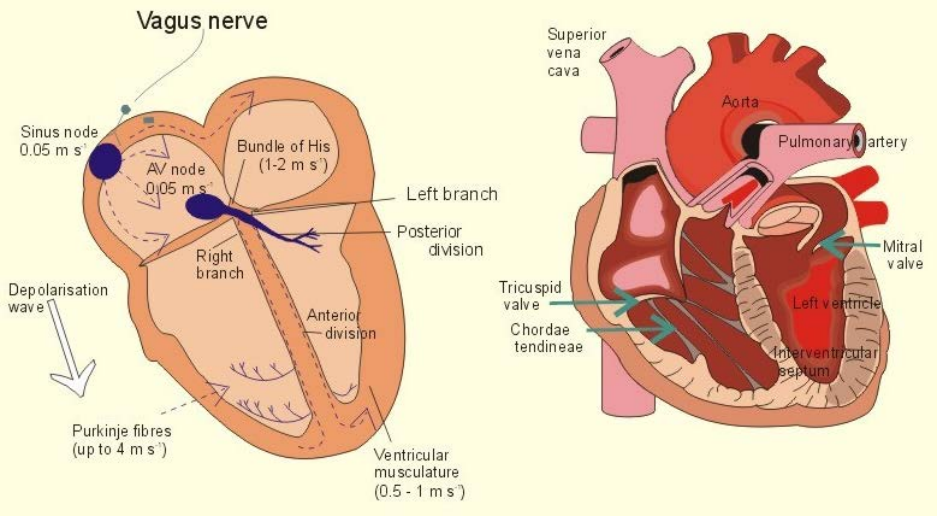


Figure 4. Left: The cardiac conduction system. Right: The anatomy of the four heart chambers (from: Paulev-Zubieta, New Human Physiology, 2nd Edition)

The heart rate is measured in number of heartbeats per minute. The ability of the heart to contract is an intrinsic property of the heart muscle, independent the nervous system. Nevertheless, there are many nerve fibers affecting the cardiac activity. Other physiological factors that determine the heart rate include respiration, blood pressure, chemical messengers (i.e. hormones), age, body temperature, levels of electrolytes (salts) in the blood. At rest, the normal heart rate is determined by dynamic interaction between spontaneous cardiac impulses generated by the SA node and conflicting influences of the two branches of the autonomic nervous system (ANS): the sympathetic and the parasympathetic nervous system. The former is responsible for acceleration of the heart rate, as opposed to decelerating function of the latter. The SA node usually fires between 50 and 100 impulses per minute. However, under certain conditions – for instance during

physical exercises (e.g. jogging) or psychological stimuli (e.g. stress) – the SA node accelerates the heart rate to increase blood supply accordingly to the needs of the organism.

Time series of heartbeat intervals

Among the mainstream cardiac diagnostic techniques the electrocardiogram (ECG) is one of the most widely used. It enables examination of a wide range of cardiac conditions, from minor to life threatening. A special case of the ECG-based study considers the temporal variation between successive heartbeats. This particular technique is called Heart Rate Variability (HRV) analysis, comprising variety of methods and statistics quantitatively characterizing functioning of the ANS, which is responsible for the cardiac regulatory functions.

To conduct HRV analyses, one needs a series of consecutive temporal intervals between adjacent heartbeats. This requires identification of the highest peak in each subsequent ECG cycle (denoted by letter ‘R’ in Figure 5) and can be done automatically using various feature extraction algorithms [112].

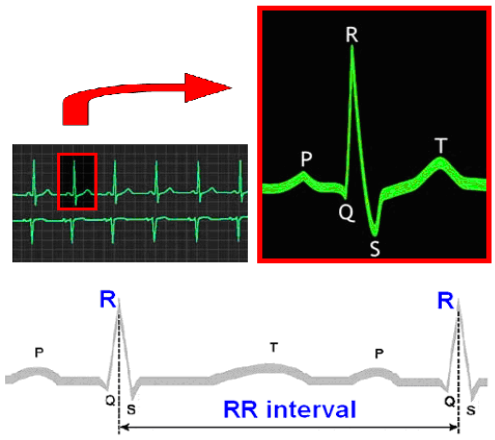


Figure 5. Relationship between ECG and RR signal. Upper graphs (A.D.A.M. Inc.): ECG plot (left) and magnification of selected cycle (right). Lower graph: P, Q, R, S and T denote peaks corresponding to onsets of consecutive states of a heart beat. Temporal distance between two consecutive R-peaks is denoted by τ.

The duration between two adjacent R-peaks – denoted 'RR' – is usually measured in milliseconds. Incorporating the R-peaks into the RR signal is *event-based* (rather than conducted on equal-interval basis). The RR signal is the typical data format used for analyses of the Heart Rate Variability.

Importance of the HRV in cardiac diagnostics

The term *Heart Rate Variability* – acronymed HRV – denotes the variation of the intervals between subsequent heartbeats. The clinical relevance of HRV traces back to 1965 [87]. Since late 1970's the HRV analysis has been recognized as a *non-invasive* research and clinical technique reflecting the activity of the cardiac and autonomic system functioning. Looked at from a wider perspective, HRV enables quantifying the adaptability (of a live organism) to environmental conditions, which is among indicators of the general cardiac health.

As a diagnostic method, the HRV analysis is being routinely applied in variety of clinical applications, including cardiovascular risk stratification (myocardial infarction and angina pectoris, congestive heart failure, malignant arrhythmias and sudden cardiac death, essential hypertension), pulmonary diseases (sleep apnea), risk assessment in neurologic disorders (severe head injury and brain death, acute brainstem stroke, Guillain-Barre syndrome), fetal and neonatal monitoring [50]. An important diagnostic of the HRV analysis is related to diabetes. A possible consequence of the disease is the *diabetic neuropathy* – a type of nerve damage resulting from high blood sugar, associated with altered parasympathetic vs. sympathetic balance, which results in reduced HRV (The latter symptom commensurate with age and severity of the disease).

A comprehensive survey regarding the clinical relevance of HRV, computational methods, technical requirements, algorithmic standards and clinical issues is provided in [87] and references therein.

Motivation and problem statement

The motivation for this research was related to three problems: (i) HRV-based cardiac diagnostics, (ii) nonstationarity detection in time series, (iii) symbolization and symbolic analysis of experimental signals.

(i) The HRV analysis with application of music and physical stimuli. Certain symptoms of the cardiac activity may be emotion- or stress-induced, as well as intrinsically transient and/or infrequent – yet potentially indicating various heart diseases. Hence the necessity of monitoring and capturing evidence of the symptoms, which are happening abruptly and last for a short period of time: those may be not apparent during

standard ECG recording (for instance: transient heart arrhythmias or transient cardiac ischemia). In such situations a technique based on Holter monitor – a small, portable device – has been applied since early 1960's. Capable of continuous monitoring heart activity (ECG) for at least 24 hours and indisputably a well-recognized and powerful technique helpful for the HRV analysis, the 24-hour Holter monitoring has at least three serious drawbacks. First, it is time-consuming as it requires a special apparatus to be carried by a patient all the time during monitoring. Secondly, the device is not able to recognize and evaluate external conditions, which directly surround the patient: These conditions change during the recording session – often drastically and in unknown manner – hence influencing the (recorded) cardiovascular activity. Last but not least, the cardiological diagnoses are often made under time pressure and using only readily available information (e.g. *ad hoc* recorded plain ECG). Hence the urge from cardiological community for HRV assessment methods capable of nearly-real time processing and using relatively small amount of data. This research addresses the demand by proposing novel methods of quantification of changes in dynamical complexity in RR signals up to 35 min., recorded in presence of *controlled environmental stimuli* including music. To date, such methods remain rather unexplored in diagnostic practice. Specifically, we compare how audible stimuli, i.e. music and white noise, perform vs. physical influences (i.e. position change and thermal bath) in affecting the HRV, and how the response (to the stimuli) from healthy participant differs from that of the patients.

(ii) Detection of change point in (nonstationary experimental) time series. Investigating nonstationarity in terms of (detecting) dynamical change points –

potentially indicating nontrivial changes in the underlying dynamics – may provide additional insight into the related process, excluding but not limited to the cardiac diagnostics. In particular case of this research the stimuli-induced nonstationarity enables investigation how different stimuli affect the HRV (in specific group of volunteers) and how different health category groups respond to (specific) stimuli. Additionally, the (identified) change points may be the basis for constructing (alternate) partition to be used for symbolic transformation – as described below.

(iii) Symbolic transformation and analysis. In symbolic representation of time series, a large number of different values of the data can be represented by very few distinct symbols (e.g. 1 and 0), what makes computations simpler and less demanding on computational resources. Whereas advantageous in many respects – including reducing the level of noise (present in the original signal) - the symbolic representation requires data symbolization, being a prerequisite for symbolic analysis. A critical part of symbolization is finding the right partition, which is a non-trivial problem rooted in physics of dynamical systems. A partition is called *generating* if it uniquely preserves all the dynamics of the original (deterministic) data. Finding the generating partition for most dynamical systems is a nontrivial problem. Specifically, for experimental signals of unknown dynamics, the generating partition cannot be found analytically. Nevertheless – based on available observables – one may '*deduce the grammar of the corresponding symbolic dynamics*' [19]. In this research, the problem of partition and symbolization is aproached *semi-empirically*, based on heuristics related to dynamical nonstationarity in experimental signals.

By its very nature, symbolization reduces complexity, thus tends to *underfit* the original data. Although an arbitrarily fixed threshold crossing – usually set as the global median or mean value – seems natural choice for partition, it has intrinsic tendency of excessive ignoring details, what results in poor rendering of the original data [61]. Consequently, the original dynamics is irreversibly lost. To mitigate this problem we propose *alternate* partition composed of *series* of thresholds, each corresponding to local regularity of the data. This preserves the original dynamics with higher fidelity than *one fixed* threshold imposed arbitrarily and globally. Obviously, the more accurately the nonstationarity is detected, the more adequately the alternate partition generates symbolic representation of the original data.

Contribution

The primary objective concerns detecting nonstationarity in experimental signals, with emphasis on cardiac RR time series. To achieve this we developed novel methods for nonstationarity detection, which are based on the notion of algorithmic complexity, i.e. *Sample Entropy*, and data compression, i.e. *Lempel-Ziv complexity*. The methods are implemented as two separate algorithms in Matlab.

The first algorithm – acronymed IRSEG (*Iterative-Recursive Sample Entropy-based Segmentation*) – operates based on the three complexity measures we derived from the *Sample Entropy*: the *static* entropy (*sE*), *dynamic* entropy (*rRnd*) and *delayed* entropy (*rDiv*). The statistics are computed on different time scales of data and evaluated according to predefined ctiteria for change points.

Based on the change points detected by IRSEG, we can construct the *alternate partition* applicable for *dynamically-adaptive* symbolization. Due to its adaptivity, the partition renders the original dynamics more adequately than an arbitrarily selected fixed threshold-crossing. Once symbolized, the signal can be further analyzed as a sequence of (binary) symbols.

Analogously to the case of experimental signals, we introduce three LZC-based methods applicable to symbolic sequences: cRat, cDet, and lDfactor. The methods – being formally introduced in Chapter 5 – have been implemented as the second algorithm, acronymed CARACAL, which operates based on the notion of Lempel-Ziv complexity.

As a special case of this study, we assessed the HRV response to the controlled *environmental stimuli* applied during the ECG recording, what enabled cross-comparison of the responses: from different *health category groups* to *different stimuli*.

The rest of this dissertation is organized as follows. Chapter 2 provides theoretical background of the fundamental concepts based on which we will introduce our methods and algorithms further in this dissertation. These include the notions of the algorithmic complexity, the Kolmogorov-Sinai entropy, non)stationarity and dynamical nonlinearity, partition and complexity of the physiological systems. In Chapter 3 we concentrate on the detection of dynamical change in signals simulated with drifting parameter, for which we introduced the Sample Entropy triplet methods and their algorithmic implementation called IRSEG. Chapter 4 is focused on dynamical change in cardiac signals recorded in presence of external stimuli, which is at core of our study. In this part of the dissertation we describe in detail the experimental setup and cardiac data as well as comparison how

our methods perform vs. well established methods of the HRV analysis. As a follow up of these, in Chapter 5 we present our approach of dynamical change detection in symbolic data, including the dynamically-adaptive partition, Lempel-Ziv complexity and our methods based on it, implemented as the CARACAL algorithm. Finally, Chapter 6 summarizes the methods and results in terms of their advantages and limitations considered with regards to different domains of application and data specificity. The Appendix contains case studies of change detection in particular signals used in the analyses.

2. THEORETICAL BACKGROUND

In this chapter we present the theoretical background for the methods and algorithms used throughout this dissertation. The two aspects of complexity: *regularity* and *predictability*, along with the notions of dynamical nonlinearity, nonstationarity, partition, symbolic transformation and (data) compressibility, are at the core of our study. Although our attention is focused mainly on physiological dynamics – and particularly the heart rate variability – we nevertheless provide a more general context and mathematical foundations of the above concepts. At the end of this chapter we present a brief survey of the mainstream HRV measures, some of which we use further as benchmarks for evaluating our methods of detecting dynamical changes in RR signals.

A. QUANTIFICATION OF COMPLEXITY IN TIME SERIES

There is no consensus on the strict definition of complexity. According to Lucas *et al*., [82] one encounters complexity when '*critically interacting components self-organize to form potentially evolving structures exhibiting a hierarchy of emergent system properties*. While this informal definition implicitly suggests that complexity manifest itself in a spatio-temporal manner, in this study we focus exclusively on the temporal aspect, as we investigate only temporal data.

In the temporal (not to say '*dynamical*') context one may agree that complexity is a '*behavior that is neither completely ordered and predictable nor completely random and unpredictable*' [83]. If so, the class of chaotic phenomena may be an excellent illustration of the complexity, where the most spectacular examples include the Lorenz meteorological model, the Belousov-Zhabotinsky chemical reaction and multi-mode laser excitation. Yet, probably even more illustrative and spectacular are the cases from different domains of physiology, where self-organization, interdependence, emergency and feedback mechanisms richly contribute to the overall complexity of live organisms.

From Occam's razor to algorithmic complexity

William Occam – a scholastic logician who lived during 14th century in England – proposed that '*no more things should be presumed to exist than are absolutely necessary*'. This informal paradigm – usually referred to as the *Occam's razor* – advises pursuing objectives using minimal sufficient resources. Whereas not rigorously stated as a principle to be strictly obeyed, it can be considered as a general heuristic for developing theoretical models. Nowadays, the medieval manifesto has its sophisticated formal implementations in computer science and information theory.

A contemporary formalization of *Occam's razor* can be found in the methods collectively called *Minimum Encoding Length* (MEL). They operate by inductively pursuing the model (i.e. theory or hypothesis) capable of the most compact explanation of data. In the MEL approach we select – out of (perhaps many) competing hypotheses – the simplest one, which still enables achieving the objective. In probabilistic terms this can be expressed as follows: If many hypotheses – each introducing certain risk of error – are

the bases for building a model, then – if the hypotheses (assumptions) do not contribute to a better accuracy of the model – their only ultimate effect might be an increased probability of the model fallacy. Overfitted models are typical examples. The examples of the MEL methods are *Minimum Message Length* (MML) [171] and *Minimum Description Length* (MDL) [130].

Occam's razor was a cornerstone for the seminal concept of *algorithmic complexity* (which we further introduce formally), according to which the most compressible signals – for instance a constant or periodic – are also the least complex, as opposed to uncorrelated stochastic signal (e.g. white noise) which is totally incompressible. Consistently with this compressibility-centric point of view, white noise should be characterized by the highest complexity. Such a conclusion, however, seems to be contradicting our intuitive sense of the complexity: As we show further in this chapter, this difficulty can be partially and indirectly addressed by applying the *surrogate data* test. Further we show how this concept is incorporated into one of our *Sample Entropy*-based methods.

Minimizing description length via data compression

Solomonoff [148] was the first who formalized Occam's razor mathematically in his observation-based *universal inductive inference*. Based on the computed probability distribution, the method predicts an unknown symbol, which directly follows a sequence of already observed (known) symbols. The only assumption made is that the probability distribution of the symbols is *computable*.

Another attempt to formalize Occam's razor was made by Kolmogorov and Chaitin [68], who introduced the fundamental concept of *algorithmic complexity* – denoted *K*. Formally, *K* defines the complexity of a string *s* as the length *l* of its shortest description *p* obtain from a universal Turing machine *U*: $K(s)=\min\{l(p):U(p)=s\}$. This definition guarantees that *K* is the *lower bound* for any other complexity measures. This definition has two important consequences: (i) *K* is an *objective* and *absolute* (complexity) measure and (ii) *K* is *incomputable* due to its dependence on (specific) Turing machine). The latter necessitates a practical implementation of *K* like the Lempel-Ziv complexity introduced below.

Intuitively, *K*(*regular* data) ≤ *K*(*random*(ized) data). In fact, this inequality is true and shows that any regularity (in data) can be used as a prescription for more compact (i.e. using fewer bits of information) reproduction of the (uncompressed) data. Conversely, in the extreme case of a random string, no compression is possible and the shortest description is the very string specified bit by bit. This shows a close relationship between *K* and data compression, what we illustrate when comparing *K* with the Shannon's entropy *H*.

Lempel-Ziv complexity of symbol sequence

Regardless its incomputability, *K* is the theoretical *lower bound* for any data *compression algorithm*. Furthermore, when a compression algorithm approaches its *optimal encoding* – perhaps after having processed sufficiently long sequence – it approximates the algorithmic complexity with increasing accuracy, thus (asymptotically) tightening the gap

with *K*. This is why compression algorithms are helpful in approximating data complexity.

Data compression has a straightforward objective of uniquely describing the original data by encoding it using less information (number of bits) than the original data consists of. The two main compression methods applied to symbol sequences are those which generate a *statistical model* (e.g. Huffman algorithm [123] and operate using a *dictionary*. Both methods are lossless, fully preserving integrity of the data, so the decompressed data is identical as its original counterpart before the compression.

The essential characteristic of each compression algorithm is *compressibility*, which can be defined as the ratio of lengths of the compressed and original strings. For instance – according the Shannon-McMillan theorem – compressibility of the Huffman encoding approaches its limit when it follows the probability distribution of occurrence of particular symbols (in the original string to be compressed), given the distribution is known (precisely) prior to the compression. Often, however, these probabilities are not known, necessitating *non-parametric* approach, where a statistical model is not required. One of such methods is a *universal compression scheme* introduced by Ziv and Lempel in [176], based on the concept of an *adaptive dictionary*. There are a few algorithmic implementations of this compression scheme (often labeled as LZ-76, LZ-77 and LZ78) – each devised around the concept of the *Lempel-Ziv complexity* (LZC). In this dissertation we will refer to the Lempel-Ziv compression scheme exclusively as to its LZ78 implementation.

For *stationary* and *ergodic* sources and for symbol sequences whose length tends to infinity, LZ78 becomes *asymptotically optimal*, what means it approaches the lower bound defined by *K*. Simply put, the larger the dictionary (having been built up to certain time during compression of a sequence), the more patterns from the remaining (i.e. not yet encoded) part of the sequence being compressed it can reference to.

In practice, however, we deal with sequences of limited length, generated by sources far from stationarity and ergodicity. These have at least two implications. First, short sequences provide little opportunity for the algorithm to approach its optimality (by sufficiently 'learning' while compressing the data), what potentially results in the complexity being *overestimated*. Second, due to dynamical changes in the data (nonstationarity), the compression performance may exhibit – besides random fluctuations – sudden and significant changes, usually increase. This is illustrated in Figure 6, where we can observe sudden increases of the normalized LZC_n:

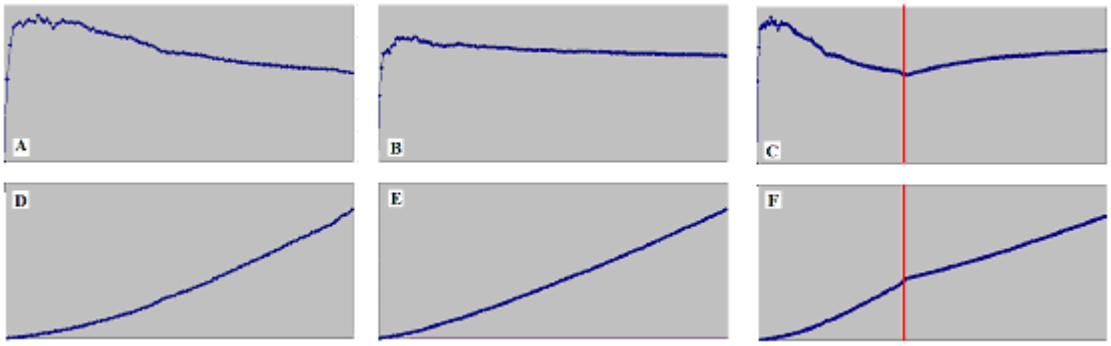


Figure 6. The independent variable represents consecutive instances of adding new word to the dictionary during compressing (left to right): sinusoid, random signal and concatenation of both (each homogenous signal consists of 1000 sample points). Upper plots: Evolution of (normalized) LZC_n for these signals. Lower plots: Evolution of the size of words consecutively added to the compression dictionary. Vertical red bars on plots (C) and (F) denote change point at junction between the concatenated signals point

We address the first problem by using the normalized *LZC* (defined in section B of Chapter 5) by reducing the effect of sequence length (As far as we investigate sequences of identical lengths – what we do – this problem is less serious). Regarding the second problem, we divide the whole sequence into a number of identical segments and perform the segment-wise analysis via sliding window.

Kolmogorov-Sinai (metric) entropy

We now introduce the *Kolmogorov-Sinai entropy* (*K-S*) as a theoretical concept behind the *Approximate Entropy* (*ApEn*) [119] and *Sample Entropy* (*SampEn*) [127]. The *SampEn*, which can be thought of as a regularity measure, illustrates that (un)predictability can be related to complexity.

Approximating the *K-S entropy* for finite and noisy experimental signals

The *K-S* can be defined using the concept of the Shannon *block entropy* H_n for series of words, each comprised of *n* symbols. We define $H_n = \Sigma p^i_n \cdot \log p^i_n$, assuming that each i^{th} word is expected to occur with the probability p^i_n anywhere in the sequence. Let H_n represent the average information necessary to predict a *subsequence* (a word) of length *n*. In such a setting we can iteratively define h_n as a *conditional entropy*: $h_n = H_{n+1} - H_n$ where $h_0 = H_1$, which is the average information necessary to predict the next *symbol*, given the preceding *n* symbols [175]. In the limit, $h = \lim_{n\to\infty} h_n = \lim_{n\to\infty} H_n/n$, what denotes the *Kolmogorov-Sinai entropy* (*K-S*). The *K-S* is a measure of average information needed to predict the next symbol based on the evolution of the system up to occurrence of the symbol. A mathematically rigorous definition of the Kolmogorov-Sinai entropy is found in [1].

The *K-S* is actually an *entropy rate* and as such characterizes the amount of information per unit time needed to completely describe evolution of the system in time. Reliability of the *K-S* estimation increases with length of the sequence of observations and the decreases with the noise level. There is a connection between *K-S* and *SampEn*: they both approximate the *K-S* for finite (short) and perhaps noisy signals. As such, the *ApEn* and *SampEn* can be applied to quantify regularity in experimental time series. In fact, they now belong to the mainstream tools used in the HRV analysis.

Measuring regularity via *Sample Entropy*

Sample Entropy (*SampEn*) estimates the *K-S* for finite-length time series. Since its introduction in 2000, *Sample Entropy* (*SampEn*) has been applied to analyses of variety

of complex physiological time series. One of the main areas of its application is the analysis of heart rate variability (HRV), where methods based on *SampEn* achieved spectacular results. Other examples of the *SampEn* applications include analyses of EEG signals [4], human gait [26], postural sway [128], climate complexity [147], market dynamics [131], chaos communication [55]. An important contribution in cardiology was that from Lake and Moorman [73] who report detecting atrial fibrillation in the context of entropy estimation in very short time series. The list of applications still expands.

Prior to a formal introduction, we provide typical cases of cardiological applications of the *SampEn*, where it has become the *de facto* analytical standard. Since its introduction in 2000 the *SampEn* has been applied to analyses of variety of complex physiological time series. One of its main application areas is the analysis of HRV, where methods based on *SampEn* achieved spectacular results, for instance the quantification of reduced HRV observed as entropy decrease, indicating early symptoms of neonatal sepsis [178]. In a study of regularity mechanisms of the HRV dynamics of premature infants, the emphasis was put on statistical properties of the *SampEn*, crucial for the parameter optimization. The authors found that under general conditions *SampEn* statistics are asymptotically Gaussian.

As a complexity measure of sequential data *SampEn* is particularly suitable for physiological signals because of its (i) robustness to noise and outliers, (ii) sensitivity to temporal order in data, applicability (iii) ranging from determinism to stochasticity and (iv) including both linear and nonlinear processes, (v) little demand on data quantity. Its implementation is simple and with low computational cost. Although not particularly

devised for nonstationarity detection, the sensitivity to the underlying dynamics makes the *SampEn* more adequate quantifier of *regularity variation* in data than strictly statistical measures like the mean, standard deviation or probability distributions.

The *SampEn* originates from the *Approximate Entropy* (*ApEn*) [119] and is estimated based on *three parameters*: the embedding dimension *m* (technically the template length), the tolerance *r* of the pattern matching (technically the pairwise distance between a template and the matched pattern) and the length of the signal *N*. In this paper we adopt an extended $SampEn(m,\tau,r,N)$ definition with the additional parameter of delay time, denoted τ. Figure 7 illustrates the concept for template length m=2:

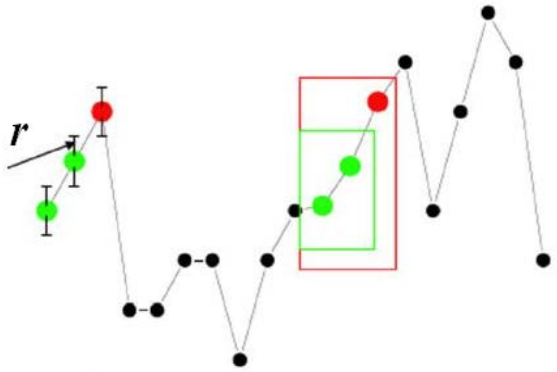


Figure 7. Computational schema of *Sample Entropy* estimation for m=2. The consecutive green dots are components of m-dimensional template vectors, the following red dots are extensions of the templates into (m+1)-dimensional space. The matching tolerance r is shown as error bars

In our case (m=2), the *template* vector of the first two points {1,2} is matched by a vector whose components are points {11,12}, contained within solid green box. Then, a new

template of {1,2,3} is obtained by extending {1,2} with consecutive 3rd point, and now we search for matches in the space of dimensionality incremented to m+1=3. As shown in Figure 7, the points {11,12,13} within red box are the match for the newly created {1,2,3} template vector. This procedure extends to all the data and up to the predefined maximal dimension. The matches are counted separately for templates of each length (dimension), with self-matches discarded – an improvement from its predecessor (*ApEn*). Thus, in a sense, the *SampEn* analyzes regularity in terms of *pattern recurrence* in *sequential* data.

Shannon vs. Kolmogorov approach

We can get better idea of the Kolmogorov complexity K when comparing it with the Shannon entropy (H) – a seminal concept in information theory. The H quantifies *statistical order* in time series, using the data probability distribution according to the formula: $H(P) = \sum_i (P(a_i) log_2 P(a_i))$, where $P(a_i)$ is probability of event a_i (in the formula we arbitrarily choose the base of logarithm equal 2, but in general other values are used as well, for instance 10 and e=2.7182...). Consequently, H achieves the largest value when the data are identically distributed. Furthermore, the H is a *static* measure, insensitive to changes of temporal order in data, what is evident when computing H on randomized (surrogate) data. The sequential insensitivity is one the main differences between Shannon's entropy H and Kolmogorov complexity K (the latter is sensitive to the *sequential order* in data).

Both '*classical*' (Shannon) and '*algorithmic*' (Kolmogorov) approaches assume that complexity can be quantified by the *minimum* number of bits necessary to transmit the *message* or describe the *observation*, often referred to as '*source*' and '*object*', respectively. The main difference is that whereas Shannon's theory considers the methods (of transmitting the message) optimal relative to *a priori* known probability distribution, Kolmogorov's ('deterministic') theory refers to the shortest computer program able to encode and then output a string representing the observation. Simply put, *H* ignores the object and considers only characteristics of the (random) source generating the object (message), which is one possible outcome of the source. *K*, on the other hand, considers only the object (i.e. message or signal) to determine the number of bits in the ultimate compressed version (regardless of how the object arose). In a sense, *K* and *H* can be thought of as two complementary complexity measures related to sequential order and probability distribution of the data, respectively.

According to Shannon's *Coding Theorem* (SCT) [144], the expected code length *L* of a message *C* represented in an alphabet $A = \{a_i, i=1,...N\}$ of size *N*, is given by the formula $L_C \geq H(P)$, where equality holds if and only if $L = -\log_2 P$. Thus, Shannon's entropy *H* of a string is a *limit* for the *lossless compressibility* of *C*.

B. PHYSIOLOGICAL COMPLEXITY AND ITS DYNAMICAL CONTEXT

In biological dynamic systems – and particularly in human physiology – '*complex signals generated by healthy organisms typically manifest at least one of the following dynamical properties* [26]: (1) nonlinearity, i.e. non-additive elements contributing to large effects

due to (perhaps) small perturbations, (2) nonstationarity, which we discuss later, (3) time irreversibility – what takes place when a system operates in a state far from equilibrium and dissipates energy, and – last but not least – (4) multiscale variability. The series of interbeat intervals of a healthy heart is a good example of a physiologic signal displaying all of the above attributes.

Complexity vs. variability

While not synonymous, the notions of *complexity* and *variability* are closely linked: complexity entails variability - measured by the variance and related statistical metrics – although in general the reverse statement is not true. Nevertheless, the *variability* is necessary condition for *nonstationarity* – a quality essential for the change detection. This important aspect will be addressed in detail in section B.

Complexity vs. adaptiveness

Since so far there is no commonly accepted definition of complexity, we adopt a biologically-motivated approach, in which the most complex signals are those generated by organisms which are most *adaptive*. The term 'adaptive' can be thought of as related to generally conceived '*healthy state*'. Noticeably, two signals can have the same degree of statistical variability (i.e., the same global variance and coefficient of variation), but very different complexity properties.

The complexity and adaptivity may be related: Complex systems have rich repertoires of responses (adequate) to the internal and external conditions. Those responses are context- and task-dependent [26]: '*The more adaptive the individual organism, the more complex the signals it produces may be. Complexity degrades with pathology and aging (theory of*

decomplexification) with disease'. Intuitively, the healthy organism is – by average – more adaptive and hence more resilient than a diseased one. This conjecture has become an assertion due to abundant experimental evidence.

Dynamical nonlinearity and nonstationarity of physiological processes

The notions of nonlinearity and nonstationarity are pivotal to our study. Interestingly, these apparently disjoint qualities both contribute to the complexity of analyzed phenomena: Because the casual relations between nonlinearity and nonstationarity often not obvious, explicit separation of their contributions may be hard. We partially address this issue by introducing and applying the *static* and *dynamic* measures of complexity.

Determinism and randomness in the dynamics of natural phenomena

As we have already mentioned, it is typical for physiological time series to manifest simultaneously determinism (or chaos as a special case) and randomness, where the border between these two is rather fuzzy. This makes the strictly linear/stochastic methods irrelevant for scrutinizing the dynamics of the cardiovascular system. On the other hand, selecting (any) nonlinear-dynamical methods 'blindly' is not particularly good. We address the data vs. methods dilemma by conducting *surrogate data* tests for nonlinearity in RR signals.

Surrogate data test (SDT) for nonlinearity in physiological time signals

The method of SDT was first introduced in [162] to detect *nonlinearity* in a time series. In particular, the SDT may be carried out as the first step of chaos identification, because nonlinearity is a *necessary* (but not sufficient) condition for chaotic dynamics.

Each SDT consists of three main elements: (i) a precise statement of the *null hypothesis* and the significance level, (ii) description of how the surrogate *data are generated* and (iii) the *test statistic*(s) to evaluate probability of committing the type-1 error. We will briefly elaborate each of these three parts.

The null hypothesis

It can be simple or composite. The former is more specific, e.g. 'data are generated by *linear, stochastic, Gaussian, stationary*' process, whereas the latter only assumes the process that generates the data belongs to certain category, e.g. 'data are drawn from *some Gaussian*' distribution. Because an explicit and precise definition of the fundamental property of nonlinearity is rather difficult, for practical reasons we regard nonlinearity as the '*absence of linearity*'. From a statistical point of view, the simple null hypothesis, denoted H_0 states that the data are *linear*, yet whether H_0 is rejected or not depends strongly on the second and third elements of SDT.

Objective of SDT is to evaluate H_0 that the data D^0 of interest has required property p_0 (e.g. stationary, linear, stochastic). The surrogate data – denoted D^s, s=1,2,…,N – are generated out of D^0 so that certain properties (statistics) $p_{i\neq 0}$ of D^0 are preserved *also* in D^s. These properties (most often they are the mean value, variance, autocorrelation or power spectra) – whose values are identical for both D^0 and D^s – must remain intact during generating the surrogate data. Conversely, the property p_0 shall be chosen so that it is *sensitive* to the $D^0 \rightarrow D^s$ data modification, hence sufficiently discriminative with

respect to the H_0. Importantly, based on the SDT one may only evaluate the *probability* that H_0 is true: rejecting H_0 is *not* equivalent to accepting H_1.

Surrogate data

In this research we apply two methods of generating surrogate data: the *phase-randomized Fourier-transform* (PRFT) and *random shuffling* (RS) of the sample points. In this dissertation we use the latter. For the original time series $\Theta = \{x_i, i=1,2,...,N\}$ we generate predefined number of the surrogate data sets $\Theta^s = \{x_j, j=\Pi(1,2,...N)\}$, where Π is a random permutation of the set of N indices of the sample points, $s = \{1,2,...S\}$ is the index of surrogate data sets, and S determines the significance level of the SDT: $\alpha = 1/(1+S)$. Throughout this research we consistently maintain the significance level at $1/(1+19) = 0.05$, i.e. we generate 19 surrogate data sets out of each original time series. Table 1 shows how the two algorithms affect the original data:

Table 1. Effects of surrogating data by the RR and PRFT algorithms

method	mean, variance	p. distribution	autocorrelation
RS	retained	retained	destroyed
PRFT	retained	destroyed	retained

Testing statistic

Its selection is crucial for the decision on rejecting H_0. Among those most often used in tests for nonlinearity are: skewness, kurtosis, correlation dimension, Lyapunov exponent, time reversibility and nonlinear prediction error. If we choose statistics that have *low discriminative power* (between the H_0 and H_1 hypotheses), we are at risk of committing a

type-II error, i.e. failure to reject the H_0 when it is actually false. Conversely, choosing statistic with very high power may help prevent false rejection of H_0, but increases risk of the Type I error. Obviously, using a larger data set (with other factors fixed) increases the power of the SDT. Failure to reject H_0 of a linear stochastic process may indicate that (i) the data do not feature nonlinearity or (ii) that the chosen statistic was not sufficiently strong to detect the nonlinearity. Importantly, rejection of H_0 – regardless whether the Type I error took place or not – in most cases does not elucidate where exactly H_0 was violated.

Example

We choose three discriminating statistics sensitive to nonlinearity: *SampEn*, correlation dimension (*D2*) and Shannon entropy *H*. The null hypotheses for RS and PRFT algorithms, respectively, are stated as follows: H_0^{RS} : the data are uncorrelated, iid and linear/stochastic. Rejection of H_0^{RS} requires that nonlinear correlation in original data is significantly higher than in the surrogate data. H_0^{PRFT} : the data are linearly correlated and stochastic. Rejection of H_0^{PRFT} requires presence of nonlinear statistics, e.g. nonlinear correlation. The *SampEn* was computed for parameter configuration m=2, τ=1, r=0.15. For computing *D2* we used the L1D2 algorithm [132] with default parameters. The *H*-entropy was calculated using Matlab code from The Milano Chemometrics and QSAR Research Group. The *Henon* map was simulated with parameters a=1.4, b=0.3 (chaos) and the RR signal was chosen as a 'typical' representative of the RR^C H_0 ('healthy') group. Each signal comprised 1000 sample points. Table 2 shows the test results:

Table 2. Testing H_0^{RS} and H_0^{PRFT}. For clarity: '0' and '1' denote $p \leq 0.05$ and $p > 0.05$, respectively

data	**RR**		**Henon**	
statistic/H_0	H_0^{RS}	H_0^{PRFT}	H_0^{RS}	H_0^{PRFT}
SampEn (*m*=2)	0	0	0	0
D2	1	0	0	0
H	1	0	1	1

Clearly, *SampEn* is the most sensitive and consistently discriminating statistic for both kinds of surrogate data (RS and PRFT) enabling rejection of both null hypotheses.

Nonlinearity and nonstationarity.

In general, we do not know *a priori* whether the data of interest are (non)linear and/or (non)stationary. This uncertainty does not, however, justify excluding the (assumption of) stationarity from H_0 (especially in the case of long and complex signals). What is more, the mainstream linear analytical techniques (e.g. spectral) work under assumptions of both *linear* dynamics and stationarity of data, what makes the above statement even more relevant to analytical practice.

If the data is actually nonstationary – a typical scenario for physiological time series – the assumption of nonstationarity should be incorporated into H_0, otherwise our rejection of H_0 may be due to either nonlinearity *or* nonstationarity. Thus, a valid alternative H_1 may spuriously deduce nonlinearity based on nonstationarity only. Furthermore, most of the *linear statistics* used in SDT may not be able to capture differences between the original and surrogate data.

These problems can be addressed by generating surrogate data in *constrained* manner, for instance shuffling the sample points within series of identical segments, whose size is

sufficiently small to approximate stationarity within each segment. The size of the segments must meet requirement for the minimum amount of sample points for the statistics used in the SDT.

To obtain reliable results, the assumption of nonstationarity in H_0 must be done *explicitly*, hence H_0 should be formulated as a *composite* hypothesis.

Applicability and limitations of nonlinear-dynamical complexity measures

The dynamic nonlinearity of physiological signals origins from *regulatory processes* of feedback-based control mechanisms. In the particular case of the HRV, the most influential regulatory functions of the origin from the Autonomous Nervous Systems (ANS), responsible for so called '*feedback loops*'. The *high gain* and *delay* are the aspects contributing most to the nonlinear dynamics of the HRV. The static linear measures do not capture the HRV dynamics because they are insensitive to the temporal order in data, which contributes to the (dynamical) complexity. This justifies the importance of SDT as a guard against limiting the analyses to linear only techniques.

Essentials of nonlinear dynamics

As opposed to a 'simple' *linear* system – where factors causing instability lead to '*infinite blow-up*' – the nonlinear dynamical (chaotic) system is confined within its attractor [179] (embedded in the reconstructed state space) – a set of all dynamical states the system can be found at. Indeed, linear systems may have only exponentially increasing or periodic solutions, hence any strong irregularities of a process must rely on *irregular input*, which is being processed according to *linear rules* [57].

Historically, until the end of 19th century, most of the natural phenomena were explained in the spirit of reductionism, attempting to explain properties of a system through properties of its elements in isolation. Typically for this approach, the effect was expected to commeasure with its cause, and interactions between components of the system were considered negligible. Simply put in algebraic terms, the '*linear paradigm*' is based on additivity and multiplicativity: $F(x+y) = F(x)+F(y)$ and $G(kx)=kG(x)$, where F, G and x, y are operators and operands, respectively. A wide range of analytical methods and characteristics – including Fourier analysis and linear correlation – were devised accordingly to this (linear dynamical) paradigm, i.e. assuming *linearity* of the observed phenomena. Although mathematically flawless, widely acknowledged and successful in many respects, the methods finally showed their limitations. This happened when it became obvious that almost all natural (and 'interesting') phenomena – particularly those controlled by complex feedback mechanisms – are immanently nonlinear [122]. According to nonlinear dynamical paradigm, small causes can trigger huge effects (what is called *butterfly effect*).

Parameter change

A distinctive feature of many nonlinear dynamical systems is the phenomenon of *bifurcation* [180], alternatively called 'period doubling'. Bifurcation is an example of an abrupt, qualitative change in behavior of the dynamics, occurring when the system parameter changes infinitesimally. A typical example can be *bifurcation* of the logistic map. The left panel of Figure 8 shows the bifurcation diagram – a set of points in 2-dimensional space: for each value of the parameter r there is at least one associated value

of the logistic map. The vertical bars labeled A, B, C and D represent four values of r={2.7, 3.1, 3.5, 3.7}, each corresponding to the analogously labeled four trace plots in the right panel. The simplest of them is plot A, where the map quickly stabilizes at a fixed point. Increasing r we observe consecutive, *gradual* period doubling for B and C (2 and 4, respectively). However, when r increases beyond 3.5 passing C, the *density* of points rapidly increases to finally reach what is called the '*onset of chaos*' approximately at r=3.57 (where the region of period-doubling ends). Typically for nonlinear (chaotic) dynamics, the change is strictly *qualitative* and occurs *abruptly* within very small range of the parameter variability.

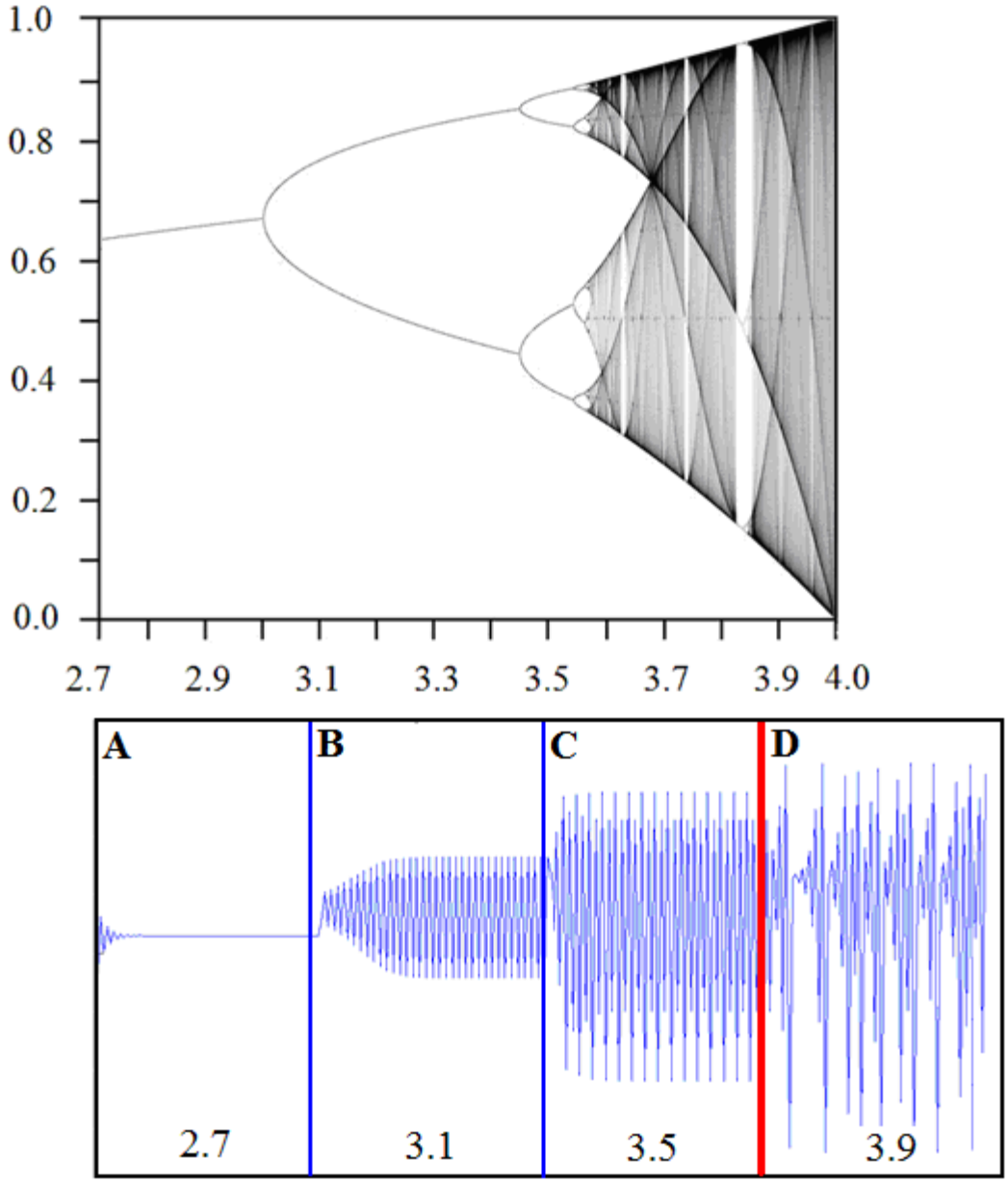


Figure 8. The *logistic* map. Upper panel: bifurcation diagram. Independent axis represents value of the parameter *r* (adopted from Wikipedia). Lower panel: a sequence of four trace plots, each representing the logistic map generated with different value of *r*. The vertical bars show where the parameter changes: the area past the red bar represents chaos

Methodological remarks. We conclude this section with two important rules of thumb. First – as far as the dynamical aspects are concerned – the consequences of scrutinizing nonlinear data using linear models are usually more serious than applying nonlinear models to linear signal. Technically, the latter is analogous to approximating a straight line by a polynomial: as the polynomial degree increases, the sum of residuals approaches zero and the risk of overfitting increases. Conversely, linear approximation of a nonlinear signal may work in most cases only on local scale.

The second rule downplays the role of the chaoticity measures like *D2* and *L1* as the system invariants: they become only the measures of nonlinearity for a particular segment of data [181]. Although using these methods in prediction or classification can still be a good idea given they actually outperform the traditional linear statistics, we shall be aware that measures of chaoticity require sufficient quality and amount of the data. Although in this research we mainly focus on short-term, noisy physiological signals, we actually manage to mitigate the mentioned risk by applying less demanding entropic-based methods and measures.

A common sense dictates the following bottom line of this consideration: Before applying the nonlinear methods it is advisable to reject most trivial case (by running surrogate data test) that the data in fact were generated by linear stochastic process.

C. NONSTATIONARITY IN PHYSIOLOGICAL TIME SERIES

Detecting nonstationarity in experimental time series is important for at least three reasons. First, stationarity should be tested *prior* to application of analytical methods which actually *require stationary* data – otherwise the obtained results may be invalid or meaningless. Secondly, if any of the following applies: the signal of interest is (i) very short, (ii) acquired in specific experimental conditions, (iii) under very limited control – what is often the case of physiological time series – test for nonstationarity is highly recommended as a part of more general procedure called '*sanity check*' of the data. Finally, nonstationarity itself may provide valuable extra information about the underlying dynamics – thus, whether nonstationarity is an obstacle or opportunity – depends vastly on the researcher's point of view and objective.

Dynamical nonstationarity may be caused by *drifting parameters* of the underlying dynamical system – i.e. slow changes of the system properties over sufficiently large time scale. We may or may not, however, know the equation of system evolution in time: in the case of experimental signals usually the latter takes place. If so, nonstationarity may be linked to changing environmental influences, what is typical for the feedback-driven physiological systems. In the rest of this dissertation the latter will be of our primary interest.

Criteria for nonstationarity: dynamical vs. statistical

Theoretically, *stationarity* is a condition when all statistical moments are time independent and finite. In practice, proving stationarity (if feasible at all) is much more difficult than refuting it. All theses is especially truewhen the full description of the

underlying dynamics of the process is not explicitly and/or *a priori* known to the researcher, and his/her access to the data is limited due to the finite observations time. These circumstances actually invalidate the confirmative statement of stationarity, making the above definition of little practical use. Moreover, for objectives of this research, the puristic definition of stationarity and rigorous mathematical formalism are not critical. Instead, we adhere to the following technical 'rules of thumb' [57].

First, the sampling frequency of the signal must be sufficiently high, otherwise the information contained between captured samples is lost (the minimum sampling frequency is formally stated in the *Nyquist–Shannon sampling theorem*). This requirement – while not so obvious – may be compromised due to poor signal acquiring equipment (512 Hz is the minimum sampling frequency in extracting RR signals from ECG). Second, the time series under scrutiny should be much longer than the longest time scale of the temporal evolution of the underlying process. For instance, because 24 hours is the most natural period for majority of physiological activities, many observations in the field of HRV analysis are being carried out for multiplicity of this diurnal cycle. In such a setting, the signal is considered stationary if its properties are time-independent when observed within an *observation period*. Thus, the stationarity of a signal can be evaluated and conclusive only with respect to the predefined (finite) time of observation. As far as the underlying (infinite) process is concerned, however, one is not be able to conclude definitely about its stationarity.

Another important issue is how the problem of stationarity is approached from methodological point of view. '*Nonstationarity makes doubtful application of global*

characteristics of the underlying process' [61], e.g. periodicities, dimensions or autocorrelation. This points to the relevancy of detecting dynamical changes in regularity *locally*, e.g. via minima and maxima of complexity measures computed within predefined segments of the analyzed signal. Such changes are usually referred to as (dynamical) *change points* and considered as indicators of nonstationarity. In this spirit, we use the term *nonstationarity* as an 'observable significant change of certain characteristic'. Furthermore, the change must be quantifiable and temporally localized in the signal.

In general, detection of nonstationarity in time series can be associated with two main categories of criteria: *statistical* (usually the first two statistical moments) and/or *dynamical* (for instance distribution of the nearest neighbors or entropies in different sections of the data). Selecting the right measures to be used in nonstationarity detection is crucial: an arbitrarily chosen statistic, not related to any *natural physical* properties of the process, introduces the risk of irrelevant findings and wrong conclusions. A good example can be analysis of fractional *Brownian motion*, where using variance as the stationarity criterion can be misleading. Indeed, the total variance of the whole signal increases (more precisely: does not decreases) in time, and each segments of the (Brownian) signal contributes to it. However – other things being equal – comparison of the variance from different segments (of equal size) will not show statistically significant difference, since the within-segment variances do not differ substantially among themselves. In such a case the risk of (committing a type-2 error) failing to reject the null hypothesis that 'nonstationarity *does not occur* in the signal' increases. Consequently, assessing *nonstationarity* by comparing values of *static* characteristics like the mean or

variance computed for different parts of the signal may be irrelevant if the characteristics do not contain information about dynamical (physical) properties of the analyzed data. Hence our approach to the problem of (detecting) nonstationarity adheres to the '*dynamical paradigm*'.

Responsiveness of the cardiovascular system to intrinsic and induced stimuli

The factors influencing the HRV fall into two main categories. The first group represents *external factors*, e.g. physical fatigue, sleep apnea, mental stress, smoking (cigarettes), meditation. The second category – the *intrinsic periodic rhythms* – includes respiratory sinus arrhythmia, baroreceptor reflex regulation, thermoregulation, neuroendocrine secretion. This classification – although perhaps somewhat arbitrary – corresponds to the scope and interest of this research. Among our objectives is scrutinizing how the predefined environmental stimuli of different nature perturb the cardiovascular system and influence the HRV. As stated in Chapter 1, the way HRV responds to the stimuli may be highly informative about the different aspects of the cardiac health, what makes the stimuli-related methods potentially valuable in clinical practice as well as more general research. Among interesting case studies we can point to [138], [115], [117], [141], and [63].

A brief survey of the change detection methods and their application

For clarity, we should start with statement that there is a subtle but important distinction between the notions of '*change*' and '*anomaly*' with regards to their detecting in time series data. The former stems from *statistical analysis* where changes in the probability distribution of the analyzed time series data are to be captured together with their

temporal positions. The latter notion origins from the domain of *data mining* and refers to the problem of finding patterns in data that do not conform to expected behavior [22]. In [6] we learn that 'the detection of what is anomalous and what is normal is heavily dependent on the application. The unifying factor across applications is that, in general, anomalies occur only rarely.' In the same work the author states that 'Change point detection can be considered a subtopic in anomaly detection,' where the problem is to detect whether the generating distribution is stable or has changed (abruptly).

From certain perspective, the relation between these two notions may depend on considering the time scale of the phenomenon (time series data) being scrutinized [58]: An *anomaly* – a 'rare' event – usually spans over relatively short part of the time series, as opposed to *change* which can indicate another mode of the signal and whose duration may include many 'cycles'. An illustrating example can be the heart rate of an individual at rest followed by physical exercises: while the latter is expected to increase significantly (i.e. a change in the heart rate occurs), *both* are 'normal' (given the individual is healthy), so there is *no*'anomaly'.

Referring to the above definitions, since now on we focus exclusively on what has been defined as *change detection*, where the term 'change' is often phrased as 'nonstationarity'. The *nonstationarity* detection in time series – a special case of the *segmentation* problem – is nontrivial and addressed by a wide variety of methods, including *Singular Spectrum Analysis* (SSA) [98], adaptive *CUSUM* [5], *Direct Density-Ratio Estimation* [60], recurrence plots [66], space-time index plots [174], cross-correlation sum, nonlinear cross prediction [140], properties of nearest neighbors in phase space and recurrence vs.

sojourn points [37], *Detrended Fluctuation Analysis* [114], *Base-Scale Entropy* [54] – to mention some mainstream examples. Most of those methods operate on vectors in phase space where the underlying system evolves in time, hence are well applicable to nonlinear signals. Other relevant techniques are described in a comprehensive survey and references therein [21].

In our study we present yet another approach to detecting nonstationarity: Motivated by the cardiological community, we propose methods adequate for the heart rate variability (HRV) assessment based on (relatively) small amount of data acquired under *controlled environmental* influences (see section III). The two vital prerequisites imposed on the methods are (i) applicability when the underlying nonlinear dynamics comprise both *stochastic* and *deterministic* components and (ii) robustness to noise and outliers along with moderate level shifts, trends and seasonalities. To extend the applicability beyond the HRV analysis, we additionally require that our methods discern between the *static* and *dynamic* complexity. We achieve these by quantifying variability of *local regularity* in data using statistics based on the *Sample Entropy* (*SampEn*). Thought of as a dynamics-related measure of *complexity*, the *SampEn* estimates the *Kolmogorov-Sinai* entropy [68] of finite-length time series. We demonstrate further how our metods compare vs. some of those referenced in this paragraph.

Regardless of the adopted approach, the change detection can be considered on different levels. On a higher level the problem can be thought of as nonstationarity analysis *per se*, aimed at gaining better interpretation of properties of a complex phenomena. On the lower (technical) level, the change point detection can be referred to as the s*egmentation*

problem. The two important questions to be asked are: (1) *How can we transform the data into a sequence to be segmented*? A popular and appealing approach is dividing the signal into a number of segments – for instance using the sliding window technique – to compute statistic of interest in a segment-wise manner, hence obtaining a series of (consecutive) values. (2) *What is the best number of segments, k, for a given sequence*? This is in certain way analogical to the model selection problem, i.e. choosing the optimal (complexity) *k*.

The (dynamical) changes in time series can vastly differ in terms of their emergence and duration. For instance, abrupt changes of short duration – sometimes referred to as '*burst*' – are typical for neuronal excitation or network intrusion. The long lasting changes can be subdivided into emerging abruptly and gradually, for instance simulating the Henon map with parameter *a* changed from 1.22 to 1.23 and the functioning of cardiovascular system perturbed by the yoga meditation (see Chapter 4, section B), respectively. We cannot provide any convincing example of gradually emerging changes of short duration – such a case seems rather self-contradicting.

Typically, the change detection algorithms proceed in two steps: (i) *scoring* each point in terms of its 'significance' and (ii) *localization*, where each scored candidate is analyzed to temporally identify the change points. Although in many cases these two steps are separate and/or independent, in physiological signals analysis the scoring step is considered more important.

As one could expect, the parametric methods of change-point detection yield superb results when we know the pertaining parameters of the models (i.e. probability

distributions). These methods rely on comparison of probability distributions of the time series data: If there is a statistical significant difference between the past and present distribution, then a change point occurs. On the other hand, if we do not know the parameters – what often is the case – the risk of large estimation error is substantial.

In this study we focus on rather short-term signals, for which density estimation would be rather problematic because of data scarcity. Consequently, in this research we chose non-parametric methods, which do not require any prior knowledge of model parameters and – likewise – do not operate on density estimation. In particular, our methods are based on the variability of local regularity in data, allows them as '*dynamics-sensitive*' and/or '*adaptive*'. The formal description of our approach is outlined in next chapter.

D. METHODS OF THE HRV QUANTIFICATION

For long the cardiovascular signals have been analyzed using exclusively 'traditional' *time*- and *frequency*-domain measures. As we show further, such measures fail to account for important properties related to multi-scale organization and non-equilibrium dynamics, hence the necessity of using quantifiers of the nonlinearity and nonstationarity in the cardiac (physiological) data. Accordingly, the presented research relies heavily on nonlinear-dynamical methods adapted to detect changes in physiological complexity in subjects characterized by different cardiac health.

The following survey outlines advantages and limitations of different mainstream methods used in the HRV analyses. Some of the methods are applied in the analyses

outlined in Chapter 4, of which the measures applicable to nonlinear, nonstationary short-term signals are of special importance in this research.

Linear

Traditionally, the mainstream linear methods in HRV analysis are divided with respect to the domain they operate on. In what follows we present the mainstream of the HRV quantifiers, grouped in four major categories.

Time-domain

This group comprises of the mathematically simplest and straightforward statistics, represented by the first two statistical moments (mean, SD) and the pNNx. The latter is number of successive difference of intervals differing by more than x (usually 50) ms divided by the total number of all RR intervals. The statistics may correspond to the RR measured directly or to the differences between consecutive RR intervals.

Frequency-domain

These characteristics are derived from Fourier analysis. The most often used in the HRV-related research are: LF, which is power in low frequency range, HF – its high frequency counterpart, the ratio LF/HF and VLF – the power in very low frequency range. Popularity of frequency domain-based analysis of HRV is partially due to its capability of discriminating between contributions of both branches of the ANS to the HRV.

Nonlinear dynamical

The nonlinear dynamical methods started paving their way into HRV analyses in early 1990's, partially thanks to seminal works of [40], [35], [51], [11], [162], [55], [59], [26]

to mention just a few. Because the nonlinear methods are still not as widely recognized and applied in field of HRV analyses as their linear counterparts, in the next few paragraphs we describe them in much more detailed and comprehensive way.

Chaotic

The presence of chaos in time series can be detected by measuring the *L1* or *D2* (both non-linear dynamical). The *L1* measures the exponential divergence of initially close trajectories in the phase space (or sensitivity to initial conditions), estimating the amount of chaos in a system. The necessary (but not sufficient) condition for chaos is that $L1<0$. According to the Takens' theorem the computation of *L1* requires infinitely long, noiseless and stationary signal. In practice these requirements can be relaxed to certain extent, nevertheless the measure requires so called 'scaling region' within which its value, i.e. the exponent *L1*, remains approximately linear in log-log coordinates. Another difficulty – both theoretical and computational – is that the *L1* requires carefully chosen embedding parameters, i.e. dimension and time delay, used for reconstructing attractor in phase space. Unfortunately, there is no rigorous mathematical recipe for obtaining these parameters precisely, though existing methods and algorithms give sufficiently good approximations for specific data.

The second measure, *D2*, estimates dimensionality of a set of points in phase space. The *D2* is complementary to *L1* in the sense that whereas the former accounts for a *dynamic* aspect of data, the latter relates to its *geometrical* aspect, i.e. the spatial distribution of points in phase space. Likewise *L1*, *D*2 requires stationarity. The computation of the *D2* was proposed in the Grassberger-Procaccia algorithm. All the difficulties related to *L1*

refer to the *D2* as well, although the latter is much more computationally expensive than *L1*.

Entropic

Entropy-based measures are well applicable to both stochastic and deterministic, as well as linear and nonlinear, dynamical systems which generates the data. They retain robustness in presence of noise without compromising sensitivity to dynamical changes. Furthermore, their algorithmic simplicity and moderate demand on computational resources make them suitable for (near) real-time processing.

The notion of entropy – for instance the seminal concept of Shannon entropy, denoted *H* – is based on uncertainty, which depends on the probability distribution. In particular, for a fixed number of possible states, *H* assumes its maximum for uniform probability distribution (which corresponds to white noise, where each state has the same probability).

The above can be rephrased in terms of a symbolic sequence $S=\{s(i)\}$, $i=1,...$, where each $s(i)$ is a symbol (from a finite alphabet of size m) emitted at time i. In such scenario, $s(i)$ may assume any of its m different values with probability $p_i = p(s_i)$. The average amount of information gained from any ‘measurement‘ corresponding to particular symbol s_i quantifies entropy $H(s_i)$, which is sometime informally referred to as ‘*quantity of surprise* related to the result of a measurement. The *entropy* $H(S)$ is then a measure of uncertainty of the (whole) signal *S*. Now, using an analogy to the symbolic sequence *S*, the *entropy rate*, denoted *h*, can be defined verbally as the average uncertainty of the next symbol of the sequence *S*, given that an arbitrarily large number of symbols have already been seen.

Because of the asymptotical character of this expression, h can be considered as the *intrinsic randomness*, persisting regardless length of the sequence $\{s(i)\}$ taken into account. Hence, h is a measure of unpredictability.

The *Recurrence Period Distribution Entropy,* denoted *RPDE* [80], captures 'repetitiveness' of a signal based on close returns of trajectories in the phase space. The *RPDE* is suitable for capturing nonlinear effects in data. An important strength of this method is its applicability to wide range of data, including periodic, stochastic as well as chaotic-deterministic signals. Furthermore, it does not require the assumption of stationarity. This method has been proved successful in detecting abnormalities in speech-related signals. The *RPDE* ranks analyzed data on the scale from 0 to 1. These numbers correspond to perfect periodicity and randomness, respectively, while chaotic signals are somewhere in the middle of this scale.

The *Permutation Entropy,* denoted *PermEnE* [13] is a complexity measure applicable to any type of time series (chaotic, stochastic, periodic and any experimental signals in between). *PermEn* considers occurrences of orderings – in terms of 'greater' and 'less' – of the values in time series – as opposed to Shannon entropy, which accounts (only) for the distribution of the values. The formal definition $PermEn(n) = -\sum_{\pi} P(\pi)\log_2 P(\pi)$,

where the summation runs over all $n!$ permutations of n-order permutations π. It holds that $0 \leq H(n) \leq \log(n!)$, with the lower bound for any monotonic series of numbers and the upper bound for purely random process. The further is $H(n)$ from its upper bound, the more dynamical regularity is contained in the data.

[16] propose a modification of the *PermEn* by mapping equal values (of observation points) onto the same symbol (rank), enabling more accurate characteristics. This improves accuracy of discerning RR signals recorded for different physiological conditions.

Scaling and fractal

The *Detrended Fluctuation Analysis* (*DFA*) has become a widely accepted technique for detection of long-range correlations in noisy, nonstationary time series and to quantify the scaling properties of time series [182]. The time series of length N is integrated and divided into non-overlapping segments of length $n \leq N$, which are least-square fitted to straight lines, each defining a local (segmentwise) trend y_n, which is subtracted from y, what results in '*detrended*' signal with fluctuations. A log-log plot against the scale (expressed via window size) is a good visualization of the scaling (or self-similarity) parameter α if one exists. In such case the time series has scaling property, and the power low hold. The higher value of α, the more 'regular' time series. This method is especially effective for signals with slowly varying trends. Different values of α indicates various category of signal, e.g. for α=0.5 we have uncorrelated (white) noise, for α=1.0 it is $1/f$ noise and for α=1.5 Brownian noise (random walk). Applied to HRV analysis the *DFA* method computes long-range correlations in RR signal by analyzing fluctuations on different time scales, disregarding trends and nonstationarities in the signal. It was found, for instance, that for different cardiac health conditions: α(CHF) > α(NSR) > α(AF), where CHF, NSR and AF stand for 'congestive heart failure', 'normal sinus rhythm' and 'atrial fibrilation', respectively.

The *Multiscale Entropy*, denoted *MSE* [26], can be used to quantify regularity of time series. The method constructs consecutive coarse-grained signal versions generated iteratively by averaging a successively increasing number of data points in non-overlapping windows. In next step the *SampEn* is calculated for each coarse-grained time series plotted as a function of the scale factor. The *MSE* Taking advantage of computational merits of the *SampEn*, the *MSE* can operate on very short sequences, but is not that robust for longer time scales of the data. From this point of view, the *MSE* complements the *SampEn*, covering many time scales, including long-term time series. The method has been successfully applied to many physiologic time series.

[26] tested the *MSE* method on simulated white and 1/*f* noises. Their results explain the fact that the 1/*f* noise contains complex structures across multiple scales in contrast to the white noise. As to the HRV analysis, they found for single-scale: *MSE*(AF) > *MSE*(NSR) > *MSE*(CHF), and for multi-scale: *MSE*(NSR) > *MSE*(CHF) > *MSE*(AF), what can be compared with analogical relations obtained from the *DFA*.

The Hurst *exponent* (*HEx*) is sometimes referred to as *Rescaled Range Analysis R*/*S*. The two factors used in this analysis are the range *R* (compound difference) and the standard deviation *S* estimated from the time series data. For many natural phenomena the empirical relation $R/S = (c\tau)^h$ holds, where τ is the time span and *h* value of the *HEx*. The *HEx* is estimated by calculating the average rescaled range over multiple regions of the data. Interpretation of the *HEx* can be as follows: $h>0.5$ denotes '*persistent behavior'*, i.e. increases and decreases of the signal value continue to maintain a (increasing or

decreasing, respectively) trend, $h<0$ denotes '*antipersistent behavior*' – minimizing development of any trend and $h=0.5$ describes a pure random walk behavior of the data.

The *HEx* is related to the *fractal dimension* – a measure of the roughness of a line (e.g. coastline). The relationship between the fractal dimension D and the *HEx* is: $D=2-h$. Looked at from another perspective, estimating of the *HEx* helps determine whether the signal is of a random walk type or has certain underlying trends. This is a nontrivial question, because some processes might be thought of as purely random, yet in fact they can be a '*long memory*' processes.

Other nonlinear

The *Nonlinear Cross-Prediction Error* (*NLCPE*) [140] is a test for (dynamical) stationarity. The idea is to use certain statistic $\gamma(S_i, S_j)$ for comparison of two disjoint segments S_i and S_j of the analyzed time series (each long enough for nonlinear prediction). Using the former as the training set (where the nearest neighbors of the latter will be searched) and the latter as test set (for which prediction is being made), one can estimate the prediction error $\gamma(S_i, S_j)$ for each pair (i, j). Whenever the difference between specific $\gamma(S_i, S_j)$ and its average is statistically significant, this indicates nonstationarity (with regard to the computed statistic γ). Suitability of the nonlinear predictability comprises in that the measure is suitable for both strongly nonlinear and stochastic data.

The *Multi-Dimensional Probability Evolution* (*MDPE*) [90] measures how often different regions of the phase space are visited during the reference (also called '*learning*') period. In other words, it determines the distribution of points in different regions of the phase space. By monitoring the distribution over time the method is capable of detecting

changes in the underlying dynamics. It uses a χ-square test to quantify statistical significance of the changes. The method proved sensitive to local changes in EEG signals.

Geometric

The series of RR intervals can be converted into geometric patterns. Among the most frequently used geometric tools for the HRV analysis are the *HRV triangular index* (RRtri), *triangular interpolation* of NN interval (TINN) – where NN denote '*normal RR*' – and *Poincaré plots*.

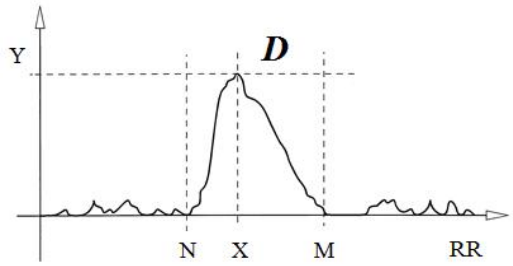


Figure 9. The sample density distribution *D* is expressed as a triangle, which assigns the number of equally long NN intervals to each value of their lengths. The most frequent NN interval of length *X* is assigned *Y*=*D*(*X*), i.e. the maximum of the sample density distribution *D*. The *RRtri* is the value obtained by dividing the area integral of *D* by *Y* [69]

Technically, the *RRtri* is the integral of the density distribution of all NN intervals divided by the maximum of the density distribution, as shown in Figure 9. The TINN is the baseline width of the RR histogram: TINN=$M-N$, where $D(M)=D(N)=0$ and $D(x)>0$ for $N<x<M$.

The *Poincaré plot* – third geometrical measure considered in this study – is a diagram (scattergram) in which each t^{th} RR interval is plotted against the previous $(t\text{-}1)^{th}$ RR interval. Poincaré plots can be interpreted visually (qualitatively) and quantitatively [49]. The shape of the plot can be used to classify the RR signal into one of several classes where the irregular shapes of plots may be classified as nonlinear (some authors consider the Poincaré plot to be based on nonlinear dynamics).

The Poincaré plot is a method of visualization of the beat-to-beat variability. The plot is constructed by pairs of successive RR intervals plotted one against the other. It is often used for complexity quantification of nonlinear behavior in physiological signals, but mostly in the RR time series. The method provides two parameters of the HRV: short- and long-term, called SD1 and SD2, respectively. Variability along the SD2 axis should be substantially larger than along the SD1 axis, as illustrated in Figure 10:

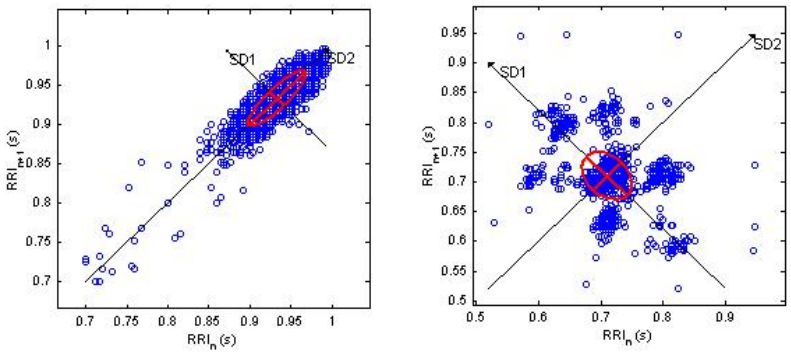


Figure 10. Poincare plots of a healthy (left) and pathological (right) HRV (credit: Kubios toolbox)

Symbolic dynamical

[113] suggest that for many natural processes discrete models might provide more accurate descriptions than time continuous systems, considering symbol sequences as 'accurate and efficient' description of most dynamics. Regardless the above opinion, there is a variety of symbolic dynamical methods applicable for time series analysis (a comprehensive survey of these can be found in [120]). In what follows we present three interesting examples where the analyses are conducted on symbolic sequences.

[42] conduct symbolic analysis of the HRV to evaluate the role of ANS before the onset of major arrhythmias. The method relies on the transformation of the series of RR intervals into patterns consisting of three beats, their classification, and evaluation of their occurrence. The RR time series decomposed into *sequential patterns* consisting of three beats are assigned to any of the three categories: non-variable, variable, and very variable patterns, as depicted in Figure 11.

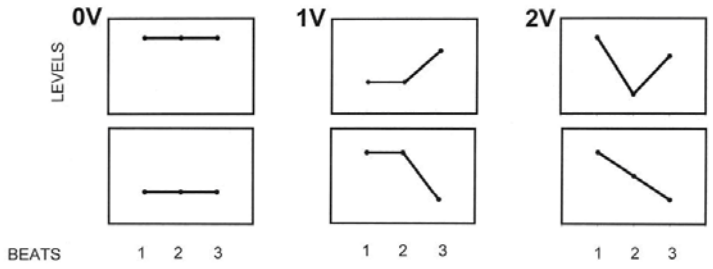


Figure 11. Classification of patterns: the symbols indicate patterns: 0V with no variation, 1V with one variation and 2V with two variations [183]

Sztejzel defines [123] the distance between two sequences of characters in terms of a '*learning process*' analogous to compression of sequential data. Specifically, the authors concentrate on how a compression algorithm 'optimizes' itself at the interface between two qualitatively different sequences (see the cases illustrated in Figure 6 on p. 22 and the related considerations). The authors introduce what they call '*learning function*' which formally quantifies the optimization. As we show in Chapter 3, in our research we adopt this approach and modified it for our purposes of detection of dynamical nonstationarity in RR signals. Reportedly, the method is able to 'recognize which dynamical systems produced a given time sequence'. What is more, the method can be used for segmentation, i.e. detecting changes between homogeneous subsequences of (perhaps) heterogeneous signal in variety of domains of time-series analysis.

3. DYNAMICAL CHANGE IN SIGNALS SIMULATED WITH DRIFTING PARAMETERS

In this chapter we formally introduce our *SampEn*-related methods: *sE, rRnd* and *rDiv* and integrate them into computational framework called IRSEG: Iterative-Recursive Sample Entropy-based seGmentation. The methods are aimed at detecting points in in time series data that mark nontrivial changes of the underlying dynamics, thus acting as candidate *change points*.

A. SAMPLE ENTROPY TRIPLET METHODS

Our method of nonstationarity analysis relies on capturing local changes through three *SampEn*-based statistics, which characterize the complexity in analyzed data. The statistics are: the *static entropy sE,* the *randomness indicator rRnd* and the *divergence indicator rDiv* – referred to as *SampEn*-triplet – used as heuristics for segmentation of the analyzed signal into stationary regions. Specifically, the IRSEG algorithm computes intersegmental differences (of the *SampEn*-triplet) and based on them makes decisions which of the evaluated intersegmental borders are (good candidates for) change points. In what follows we provide a detailed description of the *SampEn* and *SampEn*-triplet.

***Sample Entropy* – formal definition**

Let time series data of *N* sample points have values $u(i)$, where $i = 1,...,N$. We consider *m*-dimensional vectors $\vec{x}_{m,\tau}(i) = \left[u(i), u(i+\tau), ..., u(i+(m-1)\tau)\right]$. Here, the integer τ is a delay-embedding parameter. The number of such vectors that can be constructed is $N-(m-1)\tau$. However, because both $\vec{x}_{m,\tau}(i)$ and $\vec{x}_{m+1,\tau}(i)$ will enter into our statistics calculation, we only use the first $N-m\tau$ vectors so that $\vec{x}_{m+1,\tau}(i)$ is defined. Let $d\left[\vec{x}_{m,\tau}(i), \vec{x}_{m,\tau}(k)\right] \equiv \max\left\{\left|u(i+j)-u(k+j)\right| : j = 0, \tau, 2\tau..., (m-1)\tau\right\}$ be the distance between two such vectors. Using this, we count the number Q of vectors $\vec{x}_{m,\tau}(j)$ that are within a distance r of a chosen template vector $\vec{x}_{m,\tau}(i)$, where we impose $j \neq i$ to eliminate self-matches, and j ranges from 1 to $N-m\tau$. Then, calculating for each template vector

$$B_i^m(\tau, r, N) \equiv \frac{Q}{N-m\tau-1},$$

we average this over all templates to obtain

$$B^m(\tau, r, N) \equiv \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau} B_i^m(\tau, r, N)$$

which is the probability that two sequences, sampled every τ steps, will match for m sequential points. Proceeding analogously with (*m*+1)-dimensional vectors, we define

$$A_i^m(\tau, r, N) \equiv \frac{Q'}{N-m\tau-1},$$

where Q' is the number of vectors $\vec{x}_{m+1,\tau}(j)$ that are within a distance *r* of a template vector $\vec{x}_{m+1,\tau}(i)$, with $j \neq i$ once again ranging from 1 to $N-m\tau$, and

$$A^m(\tau, r, N) \equiv \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} A_i^m(\tau, r, N)$$

which is the probability that two sequences, sampled every τ steps, will match for $m+1$ sequential points. Finally, we define

$$SampEn(m, \tau, r, N) = \lim_{N\to\infty} -\ln \frac{A^m(\tau, r, N)}{B^m(\tau, r, N)}$$

which we approximate for real (finite) data as

$$SampEn(m, \tau, r, N) = -\ln \frac{A^m(\tau, r, N)}{B^m(\tau, r, N)} \qquad (1)$$

For constant data, which are an extreme case of regularity, *SampEn*=0. In general, the value of *SampEn* decreases with increasing regularity in data because (higher) regularity makes the data more predictive. Consequently, having found a match of length *m* for a template of *m* points, we are more likely to find a match of length *m*+1 for the template extended by $(m+1)^{th}$ sample. Based on (1) we introduce three *SampEn* statistics to detect nonstationarity in time series.

Introducing the *SampEn*-triplet (seT)

The three *SampEn*-triplet statistics can be thought of as a compound 'vector statistic', denoted *seT*, which characterizes different aspects of complexity (regularity) in data.

Definition of *sE*: reference entropy (*M* = 1). The *static entropy* is a special case of the *SampEn*: $sE \equiv SampEn(m = 0, \tau = 1, r, N)$. For *m*=0 the template vectors do not exist, hence all the matches in space of dimensionality increased from *m*=0 to *m*=1 involve

single points (rather than *sequences* of points) as the template vectors. Thus, the *sE* quantifies the likelihood that *any two* points are close given *no other information*. This entails *insensitivity* to *sequential order* in data (hence we call it *static*), a feature shared with *Shannon entropy*[1].

Let the time series of interest be denoted $u(i), i = 1, 2\ldots, N.$ Therefore, there are $N\left(N-1\right)/2$ pairs of data points $u(j), u(k)$ with $j \neq k$, $j, k = 1, ..., N$. Let Q^0 be the number of such pairs such that $\left|u(j) - u(k)\right| < r.$ We define the *static entropy* to be

$$sE\left(r, N\right) = -\ln\left(\frac{Q^0}{N\left(N-1\right)/2}\right) \tag{2}$$

where the argument of the logarithm is the likelihood that *any two single* points are within r of each other. Note that in (2) we do not use the embedding parameters *m* and τ, which appear in the general formula for *SampEn* (1). This is because the denominator expresses the total number of possible pairwise matches, *regardless* of the data. Furthermore, the nominator involves only (the number of *r*-close matches among all) the *single* points, i.e. not sequences, hence the *sE* – likewise the *Shannon's entropy*[2] – is *insensitive* to *sequential order* in data.

The *sE* – called '*static entropy*' – is a quantifier complementary to the *dynamic* aspects of complexity, represented by the two other statistics introduced in what follows.

---

[1] *Shannon entropy* is defined as $H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$, where X is a random variable of possible outcomes x_i, occurring with probability $p(x_i)$

[2] *Shannon entropy* is defined as $H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$, where X is a random variable of possible outcomes x_i, occurring with probability $p(x_i)$. Clearly, the H does not depend on the sequence of x_i

Definition of *rRnd*: randomness indicator ($M > 1$). We define the ratio

$$rRnd\left(M,\tau,r,N\right)=\frac{1}{M}\sum_{m=1}^{M}SampEn\left(m,\tau,r,N\right)/sE\left(r,N\right) \tag{3}$$

where the numerator is the averaged *SampEn* computed for $m \le M$. Setting $M = 2$ as the upper bound[3] for computing the *SampEn* in this study, the ratio can be expressed as

$$rRnd\left(2,\tau,r,N\right)=\left[\frac{SampEn\left(1,\tau,r,N\right)+SampEn\left(2,\tau,r,N\right)}{2}\right]/sE \tag{4}$$

The numerator in (4) represents the *dynamic* part of complexity, not accounted by *sE*. As a ratio, the *rRnd* quantifies the *relation* between static and dynamic complexity, which can be used in *comparative* analyses, where the data of interest may feature similar static complexity yet different dynamical patterns or *vice versa* (We will illustrate these cases in Chapter 5). Furthermore, as a normalized measure of 'randomness', *rRnd* $\ll$ 1 denotes high dynamic regularity, as opposed to *rRnd* $\cong 1$ characteristic for random data[4]. Indeed, the total irregularity of random data entails that the patterns in *m*-dimensional space (counted by B^m) are not 'predictive' for their subsequent points in (*m*+1)-dimensional space (counted by A^m). Conversely, even small regularity enforced in data without affecting *sE* reduces the value of the numerator in (4), resulting in reduced value of the *rDiv*.

[3] The *upper bound* for *m* is the highest dimensionality of the space where *template vectors* are constructed

[4] Based on computations involved 100 randomly generated series of uniformly distributed numbers within [0,1] we found $0.98 \le Rnd \le 1.03$

Definition of *rDiv*: divergence indicator ($\tau = 2$). The third statistic we introduce in this paper involves the *time delay* $\tau > 1$. We define the ratio

$$rDiv(r, N) = \frac{SampEn(1, 2, r, N)}{SampEn(2, 1, r, N)} \tag{5}$$

where the *SampEn* in (5) are computed according to (1), but for embedding parameters: $(m = 1, \tau = 2)$ and $(m = 2, \tau = 1)$, respectively, as schematically shown in Figure 12:

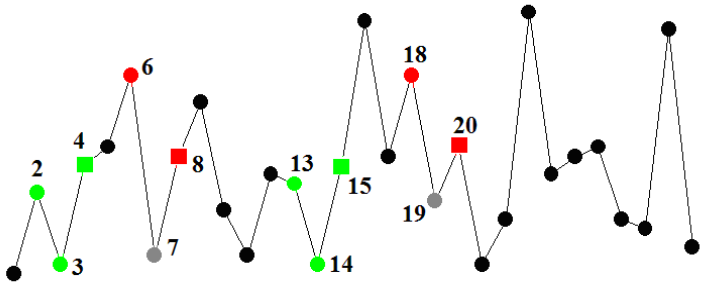


Figure 12. The green and red *circles* represent *templates* for (m=2, τ=1) and (m=1, τ=2), respectively. The corresponding *matches* are represented by the green and red *squares* (labeled 4, 8, 15 and 20. The grey circles labeled 7 and 19 denote sample points in templates for (m=1, τ=2) ignored due to the delay τ=2

The original formula for the *SampEn* (implicitly) assumes $\tau = 1$ and this setting prevails in publications. Yet, some authors use $\tau > 1$: for instance [56] (and references therein) show that (the choice of) the time delay has an important effect on the *SampEn* value, better characterizing the system complexity. Assigning $\tau > 1$ is especially relevant when the data feature strong *short-term linear autocorrelation*, perhaps entailing small $SampEn(\tau = 1)$: setting $\tau > 1$ partially destroys the autocorrelation, thus the regularity in data – if any – can be attribute (also) to nonlinear contributions.

Although no strict rules exist, the most frequently cited settings for delay time are: unity (most frequently reported), the first minimum of the autocorrelation function, and the value optimized for phase space reconstruction where τ is embedding parameter ([55] and references therein). Our choice of $\tau = \{1, 2\}$ seems adequate for comparative study of short-term time series.

Likewise, there are only rough rules of thumb regarding the two other parameters: *m* and *r*, nevertheless the accuracy and confidence of the *SampEn* improve as the number of matches quantified by B^m in formula (1) increases. This can be achieved by choosing shorter templates (decreasing *m*) or larger tolerance (increasing *r*). The price we pay is that small *r* impairs the probability estimates, while large *r* accounts for irrelevant details in data. The tradeoff implies that whereas *m*=1 guarantees more template matches, *m*>1 reveals more dynamics in data. In this study we adhere to $M = 2$ as suggested in literature.

Definition of the *SampEn*-triplet (*seT*). Based on the three introduced statistics we define the *SampleEn*-triplet

$$seT = \left[sE, rRnd, rDiv \right] \tag{6}$$

as a vector-like compound statistic, which is at the heart of the IRSEG computational engine. Specifically, the IRSEG computes the series of segmentwise values *seT*(*i*)=[*sE*(*i*), *rRnd*(*i*), *rDiv*(*i*)], where $i = 1, ..., I$ enumerates consecutive segments and *I* denotes the signal length as (integer) multiplicity of the shortest segment allowed by applicability of

the *SampEn*. Ultimately, the algorithm provides a series of normalized ratios calculated for each *seT* statistic:

$$rSET^s(i) = \min\left(\frac{seT^s(i+1)}{seT^s(i)}, \frac{seT^s(i)}{seT^s(i+1)}\right) \quad (7)$$

where $s = 1, 2, 3$ correspond to *sE*, *rRnd*, *rDiv*, respectively. In this setting, each value of $rSET^s(i) \leq 1$ quantifies the disparity between the values of s^{th} *seT* statistic computed for adjacent t^{th} and $t{+}1^{\text{th}}$ segments. In particular, we define:

$$sEratio(i) = \min\left(\frac{sE(i)}{sE(i+1)}, \frac{sE(i+1)}{sE(i)}\right) \quad (8)$$

$$rRndRatio(i) = \min\left(\frac{rRnd(i)}{rRnd(i+1)}, \frac{rRnd(i+1)}{rRnd(i)}\right) \quad (9)$$

$$rDivRatio(i) = \min\left(\frac{rDiv(i)}{rDiv(i+1)}, \frac{rDiv(i+1)}{rDiv(i)}\right) \quad (10)$$

Clearly, the values of $rSET^s(i)$ range within [0,1]. In what follows we show the *rSET* applied as heuristics for change detection-based segmentation.

The parameters: sensitivity analysis

The IRSEG uses three parameters: *M*, τ, *r* and *W*. The first three are specific to the *SampEn* definition (eq.) In principle, the parameters are user-defined, although in practice *M*=3, τ=2, *r*=0.2, *W*=100 are most frequently reported values. This, however, refers to each single parameter, not necessarily to the joint configuration of them. To assess robustness of our algorithm in terms of how the parameter values affect the computed *seT* statistics we conducted the sensitivity analysis, where at each single step we change the value of one selected parameter, while other parameters remain constant. This kind of

analysis is especially relevant when there are no strict rules on how to precisely choose the parameter values. In a wider perspective, the sensitivity analysis can be considered as a part of parameter tuning or optimization.

We conduct the analysis by varying the three parameters: *M*, *r* and *W* used in computing the *seT* and observe how the *sE*, *rRnd* and *rDiv* respond to the changes. We used two time series: a *Henon* map generated within chaotic regime (a=1.4, b=0.3) and an EEG signal from the 'data set A' available on the homepage of Department of Epileptology, University of Bonn, EEG time series. The delay time τ was not included in the analysis. The results are presented in Table 3:

Table 3. The values of *sE*, *rRnd* and *rDiv* as functions of the *M*, *r* and *W*. Because by its definition the *sE* does not depend on *M*, its values repeat for the same *r* and *W*. The mean value and standard deviation (SD) for each signal-statistic combination is given in the bottom row of the table

M	***r***	***W***	***sE***		***rRnd***		***rDiv***	
			Henon	**EEG**	***Henon***	**EEG**	***Henon***	**EEG**
2	0.1	100	2.721	3.048	1.866	2.500	1.58	1.68
2	0.1	200	2.704	3.087	1.830	2.566	1.30	1.45
2	0.2	100	2.119	2.300	1.481	1.806	1.27	1.13
2	0.2	200	2.133	2.317	1.482	1.829	1.34	1.11
3	0.1	100	2.73	3.029	1.391	2.155	1.82	1.62
3	0.1	200	2.717	3.09	1.360	2.281	1.73	1.39
3	0.2	100	2.121	2.33	1.142	1.536	1.02	1.24
3	0.2	200	2.137	2.32	1.135	1.575	1.32	1.41
mean			2.42	2.69	1.46	2.03	1.42	1.38
SD			0.32	0.40	0.27	0.40	0.27	0.21

B. THE IRSEG ALGORITHM

Let *x* denote an analyzed time series of *N* sample points. Given the lower bound *W* for segment size, the upper bound for the number of segments is determined as $\tilde{T} = \lfloor N / W \rfloor$, each comprising *W* points[5]. We call such a segment *primal* and the lower bound *W* *granularity* (of data). In this study we assume that *x* can be divided into any combination of disjoint segments whose sizes range between *W* and $T \times W$, entailing the number of segments ranging between *T* and 1 (respectively). These two extreme scenarios: 'one large segment' and 'many small segments' represent the initial conditions of two computational procedures described below.

Each *seT* component is computed and analyzed in two ways: (i) iteratively/bottom-up and (ii) recursively/top-down, denoted L2G (local-to-global) and G2L (global-to-local), respectively. The procedures run *independently* of each other and *separately* for each criterion corresponding to different *seT* statistic. In essence, the both IRSEG procedures relie on repeatedly conducted *two-step routine*:

1. (re-)compute the *seT* statistics for *currently segmented* time series being analyzed,
2. update the segmentation according to the criteria of the *seT* statistics mini/max.

The algorithm halts when the procedure L2G or G2L reaches (i) the maximal segment size ($I \times W$) or (ii) the maximal number (*I*) of allowable segments, respectively. A schematic illustration of the L2G and G2L procedures is in figures 13 and 14:

---

[5] Because in general *N* is not divisible by *W*, consequently $\tilde{T} \le T$ and the points beyond $\tilde{T} \times W$ are ignored (if $\frac{N}{W} - \tilde{T} \neq 1$, to make up for $N = T \times W$ one can append the 'missing' points using the data mean values). Regardless of this, for simplicity in this paper we asumee $T = \tilde{T}$

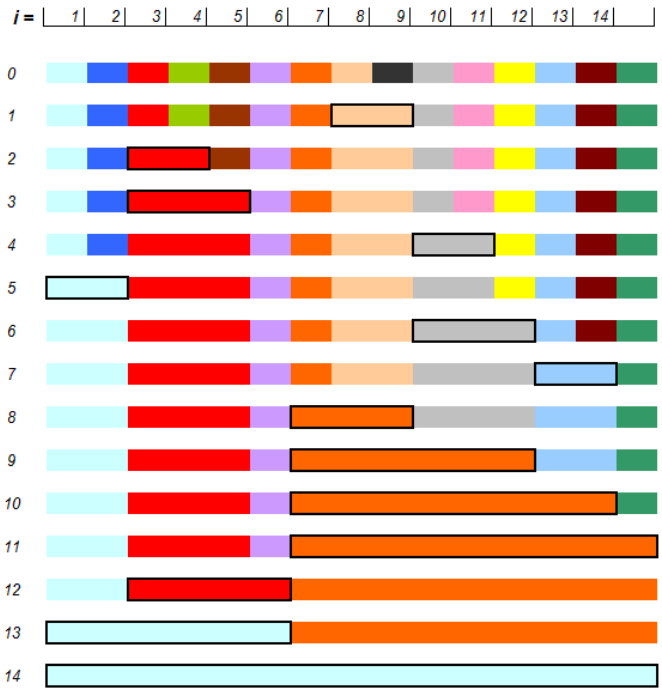


Figure 13. The L2G (bottom-up) processing. The series of numbers t=1,...,14 index consecutive intersegmental borders for 15 segments. The leftmost column of numbers enumerate snapshots of the 14 subsequent iterations. At each step the algorithm identifies the lowest intersegmental difference (associated with each seT component) and by removing the border merges the adjacent segments (The colors are used to better discern between adjacent segments)

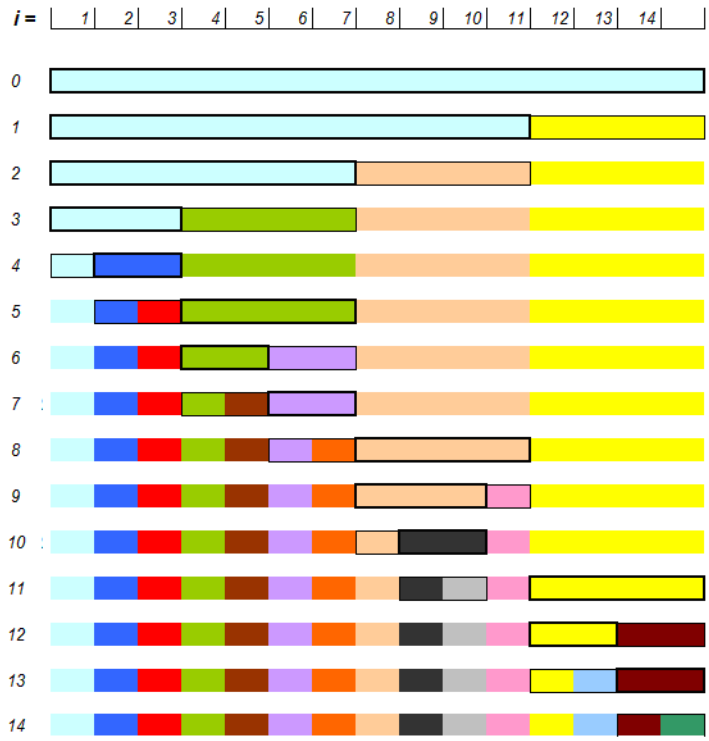


Figure 14. The G2L (top-down) processing. The series of numbers t=1,...,14 index consecutive intersegmental borders (there are 15 segments) and the leftmost column of numbers enumerate snapshots of the 14 subsequent recursion. At each step the algorithm identifies the lowest intersegmental difference (associated with each seT component) and by removing the border merges the adjacent (sequences of) segments

Notably, the segmentations: *final* (initial) of G2L (L2G) and *initial* (final) of L2G (G2L) are identical. Furthermore, the (sequences of) intermediate states of the procedures are determined their initial conditions, yet (in general) not time-symmetrical: the consecutive states of L2G cannot be derived from those of G2L (and *vice versa*) just by reversing their temporal order, hence their *non-redundancy*.

Based on the computations related to the *sE*, *rRnd* and *rDiv* and application of the mini/max criteria, the algorithm generates output whose part is shown in Table 4, containing the following columns:

- crit: criterion related to either of the *sE*, *rRnd*, *rDiv*
- mode: L and G denote bottom-up (L2G) and top-down (G2L), respectively
- *sEratio*, *rRndRatio* and *rDivRatio*: defined according to (8)-(10)

Table 4. Part of output generated by IRSEG: L and G (*mode*) stand for local-2-global and global-2-local, respectively, and *crit* denotes criterion

crit	mode	sEratio	rRndRatio	rDivRatio
1	L	0.94	0.84	0.65
1	L	0.94	0.84	0.65
1	L	0.94	0.84	0.65
...	...	...	...	...
1	L	0.98	0.63	0.74
1	G	0.97	0.90	0.96
...	...	...	...	...
1	G	0.99	1.00	0.74
1	G	0.99	0.65	0.97
2	L	0.94	0.84	0.65
...	...	...	...	...
2	G	0.97	0.90	0.96
...	...	...	...	...
3	L	0.94	0.84	0.65
...	...	...	...	...
3	G	0.97	0.90	0.96
...	...	...	...	...
3	G	0.98	0.88	0.84

The IRSEG evaluates the series of differences of the *sE*, *rRnd* and *rDiv* statistics (*seT*) at each border between adjacent segments. Each of the borders is a *candidate* for a change point. Intuitively, the greater intersegmental difference of the *seT*, the stronger indicator of a change, yet in general such an assumption would be naive because large local differences may be due to outliers or other artefacts – not necessarily related to the dynamics. These is often encountered in experimental (physiological) time series and caused by uncontrollable factors (e.g. side effects of a medicine) or imperfect measurement (involving procedure and/or apparatus). Proceeding in both bottom-up and top-down manner IRSEG accounts for a variety of time scales, thus marginalizing the possible bias due to considering the *seT* statistics only on *local* scale, i.e. considering the differences only between adjacent primals. Technically, this is achieved by segmentation, where the segment sizes are varied dynamically (on the fly), what enables analysis on different time scales.

The bottom-up L2G procedure

The time series is initially divided into $N = T \times W$ primals, where each intersegmental border – a candidate change point – is indexed by its temporal position *t*. At the beginning we set $C(t) = 1,\ t = 1,...,T$, what is equivalent to the segmentation ('dense') up to the data granularity. In each iteration, the L2G identifies $t^s_{\min}$ such that: $C(t^s_{\min}) = 1$ and $rSET^s(t^s_{\min}) = \inf\{rSET^s(t), t = 1,...,T\}$, i.e. the iteration-wise *minimum*. Once identified, we set $C(t^s_{\min}) = 0$, what can be thought of as 'removal' of an intersegmental border at $t^s_{\min}$. This entails merging two adjacent segments of lengths l_1 and l_2 into one segment of

length l_3=l_1+l_2. The L2G proceeds until $\sum_{t=1}^{T} C(t) = 0$, i.e. all the data are merged into one (super)segment.

The top-down G2L procedure

It works analogously as the L2G, but rather than iteratively merging two segments with regards to criteria of minimum values (of the $rSET^s(t)$), it *recursively* divides a currently scrutinized segment into two adjacent – hence disjoint – subsegments based on the criteria of *maximum* values (of the $rSET^s(t)$). At the cut point the value of corresponding $C(t_{\max}) = 0$ is reset to $C(t_{\max}) = 1$. The G2L replicates its analysis within (sub)segments which (i) comprise at least two primals, (ii) are left- (right)-delimited by the least (largest) available $t_{\min}$ ($t_{\max}$) such that $C(t_{\min}) = 1$ ($C(t_{\max}) = 1$), where $t_{\max} \geq t_{\min} + 2$ and (iii) contain an interrupted series of at least one $C(t) = 0$, where $t_{\min} < t < t_{\max}$. The G2L proceeds until $\sum_{t=1}^{T} C(t) = T$, i.e. the data granularity is reached.

Change point evaluation.

The change point candidates CPC (equivalent to intersegmental borders) are characterized by index $t = 1,...,T-1$, which specifies their temporal positions. The indices expressed multiplicities of primal segments covering the signal from its beginning to the CPC(t) of interest. For instance, at t=3 the intersegmental border CPC(3) delimits

the 3rd and 4th segments. Thus, each CPC(t) is uniquely identified regardless advancement of the L2G and G2L analyses.

Prior to providing formal definitions of the two CPC evaluators used in our study, we need to introduce two *weights* which account for (i) disparity of lengths of the compared (adjacent) segments and (ii) predefined importance of the *seT* (*rSET*) statistics. The former is related to each pair of segments adjacent to CPC(t) and defined as the (length) *disparity* (index) $d(t) = \left\lfloor \max(\frac{w(t+1)}{w(t)}, \frac{w(t)}{w(t+1)}) + 0.5 \right\rfloor$, ranging within $[1,...,T-1]$, where $w(t)$ and $w(t+1)$ are the length of segments to the left and to the right of CPC(t), respectively. The index quantifies disparity of lengths of segments being subject of merge and/or cut operations, is an argument of a moderating discrete function $\Phi(d):\ d = 1,...,T \to (0,1)$ defined as follows:

$$\phi = \frac{1}{d};\ \text{for}\ d > 1:\ \phi^1(d) = 1 - \phi(d),\ \Phi(d) = \phi^1(T-d)\ \text{and}\ \Phi(1) = 1.$$

Hence, $\Phi(d)$ is inversely proportional to their corresponding length disparities, and decreases sharply when d approaches T. As such, the function is a moderator reducing contributions from the merge/cut operations with larger values of d involved. This simply means that the 'credibility' of a CPC adjacent to segments whose lengths differ significantly is *weaker* than that of segments with more balanced lengths. The disparity index d is calculated *before merges* in L2G and *after cuts* in G2L, thus always involving two segments.

The *prioritization* of the *rSET* components is achieved by applying the *weigh*t vector $\gamma = [0.33, 0.5, 0.17]$. Accordingly, the *rRndRatio*-related criterion is weighted most

heavily at 50%, whereas the *rDivRatio*-related criterion contributes only 17% to the total evaluation, with *sEratio-related criterion* contributing 33% to the *rSETavg* evaluator. Although the choice of the γ values may seem arbitrary, they reflect emphasis put on the *dynamica*l part of the underlying complexity, contributing two thirds of the total weigh.

Definition of *rSETavg*. The first change point evaluator – denoted *rSETavg* – is defined as $rSETavg(t) = \frac{1}{3}\sum_{crit=1}^{3}\sum_{s=1}^{3}\gamma(s)\times rSET_{crit}^{s}(t)$. In words, it is a γ-weighted, averaged aggregate of the 3x3=9 $rSET^{s=1,2,3}(t)$ values computed for all the criteria. The smaller *rSETavg*(*t*) – which mean poor dynamical similarity on both sides of the C(*t*) – the 'stronger' candidacy of this very C(*t*) as a change point.

Definition of *avgChron*. The second change point evaluator – denoted *avgChron* – characterizes the CPC's *without direct reference* to the *rSET* values (as opposed the *rSETavg*). Let the quantity *mergeAvgChron* be the Φ-adjusted chronological order (an integer from (1,T)) of the merge operations, averaged over all the *rSET*-related criteria and normalized with regards to the maximal chronology, which is *T* (i.e. the number of borders between primals). The c*utAvgChron* is defined analogously but refers the cut operations. We define $avgChron(t) = \left(mergeAvgChron(t)\times cutAvgChron(t)\right)^{0.5}$, which characterizes the CPC(*t*) in terms of the *chronology* of L2G and G2L operations.

Clearly, for the *greater* values of *rSET* the intersegmental differences are more pronounced and the corresponding merges – by average – take place (chronologically) later, being associated with greater temporal indices denoted as $chron = 1,...,T$. Conversely, the cut operations occur earlier, hence are associated with smaller values of

chron indices. A reversed analogy holds for the *smaller* values of *rSET*, linked to earlier occurrences of merges and later occurrences of cuts, which translates into values of *chron* being smaller and larger, respectively.

Ideally, occurrence of a change point $C(t)$ coincides with $\sum_{s=1}^{3} avgChron^{s}(t) = 1$ and 0 otherwise, i.e. depending on whether a change point occurs or not, either *all* or *none* min/max criteria are satisfied at the same time. In practice, however, the sum fluctuates.

Change point identification.

Ultimately we propose certain rules for the change point *identification*. An obvious approach is to define and apply thresholds for the *avgChron* and *rSETavg*. Because in general those statistics do not feature normality, we define the thresholds based on median (rather than mean) and MAD[6] (rather than SD). For a series of numbers S we define the MAD as $mad(S) = median_i\left(\left|S_i - median_j(S_j)\right|\right)$. Specifically, we define threshold $\Theta^{1}_{avgChron} = \max(r \times mad(avgChron(t), slope(avgChron)))$, where $slopeChron_{avgChron} = \frac{1}{T}\left(\max(avgChron(t) - \min(avgChron(t))\right)$. Furthermore, we define two more thresholds: $\Theta^{2}_{avgChron} = \min(1, median(avgChron(t)) + mad(avgChron(t)))$ and $\Theta^{3}_{avgChron} = \min(1, median(avgChron(t)) + 2mad(avgChron(t)))$. Proceeding analogously, we obtain three thresholds $\Theta^{i=1,2,3}_{rSETavg}$ for the $rSETavg(t)$. Finally, let

[6] MAD stands for *median absolute deviation* and plays analogous role to that of standard deviation (SD).

$\Delta_{avgChron}(t) = \min(avgChron(t) - avgChron(t-1), avgChron(t) - avgChron(t+1))$, which evaluates whether *avgChron*(*t*) is a local maximum within direct vicinity of *t*. Again, there is analogous measure $\Delta_{rSETavg}(t)$ corresponding to $rSETavg(t)$.

Having defined the $\Theta^{i=1,2,3}_{evalCPC}$ and $\Delta_{evalCPC}$, where $evalCPC \in \{avgChron, rSETavg\}$ denotes the applied CPC evaluator, we can specify the following *two double conditions* for the change point qualification:

1. '*weak*': $\Delta(t) > \Theta^1_{evalCPC} \wedge avgChron(t) > \Theta^2_{evalCPC}$ and
2. '*strong*': $(1) \wedge avgChron(t) > \Theta^3_{evalCPC}$, where ' $\wedge$ ' denotes logical conjunction.

Loosely speaking, we expect a change point (i) being a 'sharp enough' (local) maximum and (ii) representing a value 'sufficiently' above the pertaining median. These criteria naturally stem from those based on the mean and SD, which are more reliable under condition of normality and/or sufficiently many examples. We cannot guarantee any of these, hence resort to '*non-parametric*' approach. Noticeably, in the definition of $\Theta^1_{evalCPC}$ we used the same tolerance factor *r* as in the *SampEn* definition (1), thus accounting for fluctuations of the $avgChron(t)$ and $rSETavg(t)$ due to noise in the original data.

It remains an open question which of the two CPC evaluators: *rSETavg* and *avgChron*, should be used the 'ultimate' one. The dilemma mostly pertains to the cases when our knowledge of the supposed changes is limited (e.g. physiological data). This leads us to another alternative: Should we use a *single* evaluator – perhaps the 'ultimate' one – or *both*, and if the latter, should *any* (i) or *both* (ii) of the qualifying conditions be met. The choice may depend on whether the priority is *sensitivity* or *specificity*, which favor the

‘lenient’ (i) and ‘stringent’ (ii) approach, respectively. Far from offering conclusive answers, we found – based on a series of comprehensive tests – that approach (i) yields accurate results in most cases. On the other hand, when scrutinizing ‘difficult’ cases (like stationary data with high variability or externally stimulated cardiac signals), approach (ii) helps avoid too many false positives.

C. DATA AND EXPERIMENTAL SETUP

Having introduced all the formalism and modus operandi of our analyses, we finally arrived at presenting all the data processed in our research, together with the parameters used by the introduced and applied algorithms. We will start with the latter.

Parameter optimization

We already considered the parameters used in computing the *SampEn*-triplex statistics to show robustness of our methods to the parameter variability (Chapter 2). Now we provide more detailed analysis of the parameters' influence on particular *seT* statistics. Following loosely analyses reported in [56] we investigate whether they apply to our research.

The IRSEG algorithms

In this chapter we conducted analysis of how the parameters: *M*, *r* and *W* affect the *SampEn*-triplet statistics (the results are presented in Table 3 on p. 54). In this section we broaden our analysis by including the delay parameter τ and considering the parameters *M*, *r*, and *W* as pertaining not just to the *SampEn* (or *SampEn*-triplet) but to the whole IRSEG computational framework.

We analyzed each parameter based on three different time series which can be categorized as *random*, *chaotic* and '*hybrid*' (being a complex mixture of both). Specifically, we use uniformly distributed random numbers between 0 and 1, the *Henon* map generated in chaotic regime and EEG signal (the same as used in for sensitivity analysis, p. 67). For each of the analyzed parameters we apply the same criterion for

'optimality': the *highest accuracy* of change points detection given the knowledge of their temporal positions. When one parameter is tested and its value is discretely varied within predefined range, the values of the remaining parameters are equal to their respective medians (computed for the set of their values within the predefined range).

Similarity threshold *r*

When applied to the *Henon* and RR signals, we observe the highest accuracy for r=0.15, the value widely reported in literature as 'optimal' for *SampEn* in general. For the random signal the value of r has little impact. The results are shown in Table 5.

Table 5. Accuracy as a function of different combinations of (values of) r and time series data

data	**rnd**	**rnd**	**rnd**	***Henon***	***Henon***	***Henon***	**EEG**	**EEG**	**EEG**
r	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**
accuracy	88	91	88	93	100	88	88	97	93

Dimension *M* and delay τ

The highest accuracy was obtained for M=2. Likewise in the previous case of random data, M has little effect on accuracy. This finding should not be surprising in light of the sensitivity analyses conducted for the *SampEn*-triplet methods, when for M=2 was found the most consistent results. The results are shown in Table 6.

Table 6. Accuracy as a function of different combinations of (values of) *M* and time series data

data	**rnd**	**rnd**	**rnd**	***Henon***	***Henon***	***Henon***	**EEG**	**EEG**	**EEG**
M	**2**	**3**	**4**	**2**	**3**	**4**	**2**	**3**	**4**
accuracy	91	91	88	100	88	91	100	88	67

As we expected, the accuracy sharply drops for $\tau > 1$, especially for the nonrandom signals. The explanation may be that the delay greater one sample point reduces regularity, hence the value of *SampEn* is spuriously increased (especially when the original sampling rate of the signal capture is already not sufficient – this applies especially to the RR signals). The results are presented in Table 7.

Table 7. Accuracy as a function of different combinations of (values of) τ and time series data

data	**rnd**	**rnd**	**rnd**	***Henon***	***Henon***	***Henon***	**EEG**	**EEG**	**EEG**
τ	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
accuracy	91	88	79	93	82	67	97	73	67

Segment size *W*

As in the previous cases, the random signal is least affected by parameter variations. In this case, however, the *Henon* and RR signals response to changes of *W* differently: For the former, the highest accuracy was obtained for *W*=200 and for RR signal *W*=100 was

optimal for accuracy. The results are shown in Table 8. A possible explanation may be that in the case of physiological data, the longer analyzed (sub)sequence the more difficult to ensure stationarity.

Table 8. Accuracy as a function of different combinations of (values of) *W* and time series data

data	**rnd**	**rnd**	**rnd**	***Henon***	***Henon***	***Henon***	**EEG**	**EEG**	**EEG**
W	**100**	**150**	**200**	**100**	**150**	**200**	**100**	**150**	**200**
accuracy	97	90	90	100	91	97	93	73	88

In summary, we found IRSEG delivers stable and reproducible results within the tested range of the parameters' values. Actually, the values used in this research are similar (or identical) to those frequently reported in analogous studies.

Data preprocessing – general rules

As a proof of concept we first applied the IRSEG algorithm to *simulated* time series, systematically enforcing the three *types of nonstationarity* via predefined *parameters* and *initial conditions*:

(a) specific signal, e.g. *Henon* map, simulated with *fixed* parameter value: the objective is capturing *transient* dynamics [7] – if any – manifesting itself by evolution from a chaos-like to regular (perhaps periodic) behavior;

[7] This may happen when the initial conditions are sufficiently close to the attractor boundary [48]

(b) specific signal simulated with *varying* parameter value: we concatenate the parameter-specific sequences attempting to identify where the parameters change;

(c) concatenation of *different signals*, e.g. *random*, *logistic* and *Henon*, simulated with fixed values of their respective parameters: the objective is to identify intersegmental borders where the *qualitative changes* of the dynamics take place.

The realization of the (a)-(b) is based on the following seven specifications:

random: series of uniformly distributed i.i.d. numbers within $[x_{\min}, x_{\max}]$, denoted *U*, where x_{min}, x_{max} are the lower and upper bound of the data;

N^{trunc} denotes the Gaussian counterpart of *U* with some points fit into $[x_{\min}, x_{\max}]$;

logistic: $x_{n+1} = 1 - rx_n^2$, with *r* varied, denoted *L*];

Henon: $x_{n+1} = y_n + 1 - ax_n^2, y_{n+1} = bx_n$, with *a* varied and *b* = 0.3, denoted *H* ;

Mackey-Glass: $\frac{dx}{dt} = \beta \frac{x(t-\tau)}{1+(x(t-\tau))^n} - \gamma x(t)$, where $\tau, \gamma, \beta, n > 0$, denoted *MG* ;

Rossler: $\frac{dx}{dt} = -y - z$, $\frac{dy}{dt} = x + ay$, $\frac{dz}{dt} = b + z(x - c)$, denoted *R* ;

sinusoid: $x(n) = \sin(n)$, denoted *S*, where $n \ni \square^+$.

All the simulations start with the pertaining initial values set to 0.5 (e.g. $x(1) = 0.5$), unless stated otherwise. To minimize possible effects due to initial instability, the first 100 sample points of each signal generated with a given parameter value were discarded, with exception of the L^{trans} and H^{trans}, where the instability was in focus. Table 2 shows the simulations parameters. The experimental realization of concatenations and modifications are described in Table 6 and labeled accordingly by *signal ID*.

***Logistic* map.** We followed the guidelines in [179] when selecting different values of parameter *r* used in simulation. Additionally, we analyzed an interesting case of *transient* behavior of the map, simulated with $r = 3.828427127$ and $x(1) = 0.3$.

***Henon* map.** We used four realizations of the *Henon* map, which was simulated using constant value of parameter $b = 0.3$ and varying the parameter *a* as listed in Table 2. Likewise the *logistic* map, *Henon* exhibits *transient* chaotic dynamics for specific combinations of parameters. We observed this effect by simulating the *Henon* map with $(a, b) = (1.3, 0.3)$ and $(x(1), y(1)) = (0.8, 0)$, when after the initial chaotic behavior within approximately 170 points the system settles down into periodicity.

***Mackey-Glass* delayed differential equation (*MG*).** The *MG* is characterized by four parameters: dimension *n*, time delay τ, and two additional parameters: β, and γ. In this paper we follow the guidelines from [84], assuming the initial condition $x(1) = 1$ and maintaining *n*=21, β = 0.2, γ = 0.1 (the values of time delay are listed in Table 9).

Table 9. The parameter values and regimes of three signals based on the *logistic* equation, *Henon* system and *Mackey-Glass* delayed differential equation (the latter discretized using the Runge-Kutta method)

signal ID	parameter *r*	regime	signal ID	parameter *a* (*b*=0.3)	regime	signal ID	parameter τ (*n*=21)	regime
L4	3.58375	periodic	*H3*	1.078	periodic	*MG3*	24	periodic
L5	3.60625	periodic	*H4*	1.18	chaotic	*MG4*	25	periodic
L6	3.828427123	chaotic	*H5*	1.22	chaotic	*MG5*	26	chaotic
L7	3.857	chaotic	*H6*	1.3	transient	*MG6*	27	chaotic

***Rössler* system (*R*).** The *x*-component of the *Rössler* equations was obtained for the parameter set $(a,b,c)=(0.35,0.2,5.7)$. The signal comprises 1000 sample points and is dynamically *homogeneous*, i.e. simulated without any parameter changes, hence we assume *no change points* in the data. Nevertheless, when judged visually based on the trace plot – the signal's halves exhibit clearly different quasi-periodicity.

***Sinusoid*.** The signals – labeled *S*1-*S*4 – were approximated using the frequencies ω, 2ω, 5ω and 10ω, respectively.

Based on the (concatenations of) the described signals we created their combinations and modifications as listed in Table 10, followed by verbal descriptions.

Table 10: The modifications and/or combinations of the *logistic* and *Henon* maps, sinusoids (details in text)

signal ID	description
L^{trans}	*logistic* map: $r=3.828427127$, $x(1)=0.3$
H^{trans}	*Henon* map: $(a,b)=(1.3,0.3)$, $(x(1),y(1))=(0.8,0)$
H^{surrog}	randomly shuffled $H5$
H^{under}	undersamped $H5$
$S^{avg1\&4}$, $S^{avg2\&3}$	the averaged pairs of signals: (*S*1, *S*4) and (*S*2, *S*3)

Compound signals. The H^{surrog} was obtained by *random shuffling*[8] the sample points of (*H5*). In the H^{under} we enforced nonstationarity by varying the *sampling rate* of the simulated *H5*: in its consecutive three subsequences of equal lengths – denoted H^{100}, H^{50}, H^{33} – the sampling rate was 100%, 50%, 33%, respectively.

[8] The *random shuffling* is a method of obtaining *surrogate data* [38]; see description further in this paper

D. RESULTS AND EVALUATION

To evaluate how accurate are our change detection methods we apply the IRSEG algorithm to all the *simulated* (numeric) data described in previous chapter. These data are (i) purely chaotic with drifted parameters and (ii) combination of chaotic, periodic and random. The signals are listed and identified in Table 10.

Change point detection accuracy

The knowledge of where the dynamical changes take place in simulated signals enables evaluation based on more solid ground than when the signal is acquired experimentally (we face this issue in next chapter where conducting analogous evaluation using the cardiac RR signals). Consequently, we consider the evaluation based on the simulated signals as a *baseline* for the performance of our algorithms.

The results for particular sequences are collected in Table 11 and expressed as the *true positives* (*TP*), *false negatives* (*FN*) and *false positives* (*FP*), quantifying the instances of (i) correct change point identification, (ii) failure to identify change point and (iii) mistakenly identifying putative change point which actually is not one. The sum *C=TP+FN* is the '*theoretical*' number of (*true*) change points referred to when computing the *TP*, *FP* and *FN*. The overall (change) *identification accuracy* is quantified by *recall* and *precision*[9] defined as $\alpha = \frac{TP}{TP+FN}$ and $\beta = \frac{TP}{TP+FP}$, respectively.

[9] We use the *recall* and *precision* rather than (the measures of) *sensitivity* and *specificity*, because the latter requires the quantity of *true negatives*, being irrelevant (or problematic) in change point identification

Table 11. The sequences of signals *simulated* in this study (the *nonstationarity type* is described on p. 65)

seq. ID	component signals	nonstationarity type	C	TP	FN	FP
seq1	L^{trans}	*a*	2	2	0	0
seq2	H^{trans}	*a*	1	1	0	0
seq3	*R*	*a*	0	0	0	0
seq4	*L*4, *L*5, *L*6, *L*7	*b*	3	2	1	2
seq5	*L*4, *L*5	*b*	1	1	0	0
seq6	*L*5, *L*6	*b*	1	1	0	1
seq7	*L*6, *L*7	*b*	1	1	0	1
seq8	*H*3, *H*4, *H*5, *H*6	*b*	3	2	1	0
seq9	*H*3, *H*4	*b*	1	1	0	0
seq10	*H*4, *H*5	*b*	1	1	0	0
seq11	*H*5, *H*6	*b*	1	1	0	0
seq12	*MG*3, *MG*4, *MG*5, *MG*6	*b*	3	1	2	0
seq13	*MG*3, *MG*4	*b*	1	1	0	0
seq14	*MG*4, *MG*5	*b*	1	1	0	1
seq15	*MG*5, *MG*6	*b*	1	1	0	0
seq16	*S*1, *S*2, *S*3, *S*4	*b*	3	3	0	1
seq17	*U*, *L*7, *H*8	*c*	2	2	0	0
seq18	*S*1, *L*1, *H*1	*c*	2	2	0	0
seq19	$S^{avg1\&4}$, $S^{avg2\&3}$, $S^{avg1\&4}$	*c*	2	2	0	0
seq20	*H*6, H^{surog}, *U*, N^{trunc}	*c*	3	2	0	1
seq21	H^{100}, H^{50}, H^{33}	*c*	2	2	0	0

The sequences from Table 21 are illustrated in Figure 15 below:

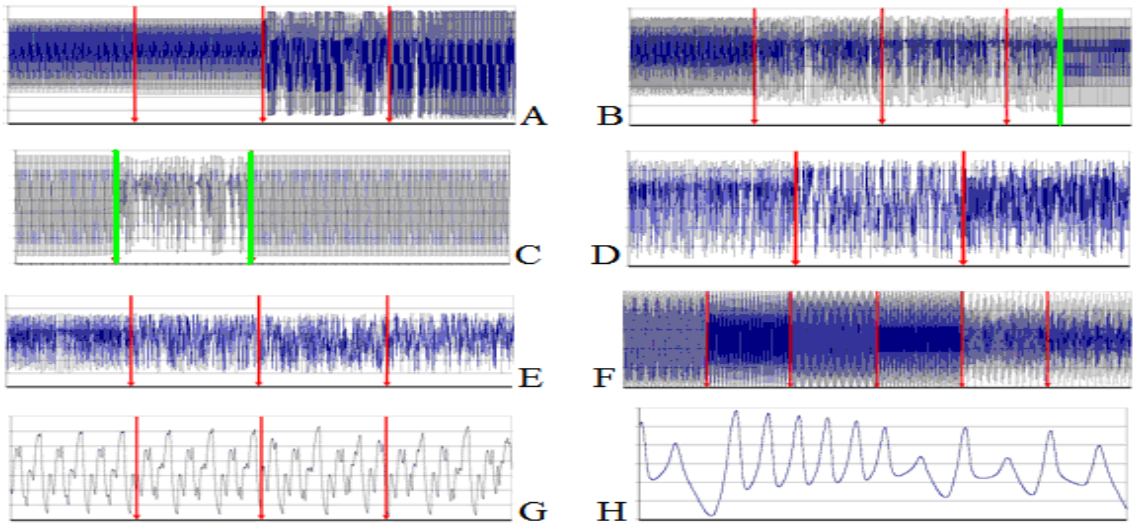


Figure 15. Simulated data. A) *L*4-*L*5-*L*6-*L*7, B: *H*3-*H*4-*H*5-*H*6, C) L^{trans}, D) H^{under}, E) *H5*-H^{sur}-*U*-N^{trunc}, F) *S*1-*S*2-*S*3-*S*4-$S^{avg2\&3}$-$S^{avg1\&4}$, G) *MG*1-*MG*2-*MG*3-*MG*4, and H) *R*. The vertical *red* bars indicate *parameter*-related change points. The vertical *green* lines (L^{trans} and *H*6) indicate the onsets of different behavior due to *transient* dynamics (see description on p. 11). The dependent variable denotes subsequent points of the signals and dependent variable denotes the points' values

The category (a) is represented by the green bars-delimited subsequence on plot C and the last segment – after the red bar – of plot B (the vertical green bar is placed where transient behavior starts). Yet another example of category (a) is shown as a plot R depicting the signal generated by Rossler system (chaotic regime). Contrary to the visual judgment of the signal – where the left and righ parts look different (the oscillations in the left part is more frequent) – there is nevertheless no dynamica change point nor transient and this example was used to test the algorithm precision. The rest of the plots except plot E, which represents category (c), are associated with category (b).

We obtained the highest recall and precision $(\alpha, \beta) = (1,1)$ for the group *a* (blue), where the data is generated based on the same formula (e.g. Henon) and without enforcing

parameter change. If nevertheless – due to the system intrinsic instability – a detectable change takes place, it is often called transient. Because in our cases of L^{trans} and H^{trans} the dynamics drift between chaos and (quasi-)periodicity, detecting such changes should not be challanging. The dynamical regime of the third signal denoted *R* (Rossler) actually do not change, although a visual judgment may suggest otherwise. Nevertheless, IRSEG correctly rejected the midpoint of the signal as a change point, hence not deteriorating its precision. The sequences from group *b* (green) which are different realizations of a specific signal where the dynamic parameter is being changed during simulation, are more challenging for the algorithm because – the parameter change does not incur substantially changed dynamics, what explains the lower performance $(\alpha, \beta) = (0.81, 0.74)$ for this group. The reason for high accuracy achieved on the signals from group *c* (orange) is totally different dynamical properties of those signals – even if the probability distribution of their elements may be close: for this group our algorithm scored $(\alpha, \beta) = (0.91, 0.91)$. The *total score* for all 21 cases is $(\alpha, \beta) = (0.88, 0.81)$. While (α, β) differ *among* the three groups, they commensurate well *within* each group. Scrutinizing the plots in the Appendix where each case is briefly characterized and illustrated may be helpful to capture the strengths of IRSEG as well as where the algorithm should be improved.

Comparison with other techniques

In what follows we provide results obtained from applying four methods (algorithms) to the same data in order to compare their performance with that of our methods. We

selected three different methods applicable in the domain of change detection in time series data (see a brief description of the terms ‘change detection’ vs. ‘anomaly detection’ on p. 36): They are (i) *Cumulative Sum Control Chart* (CUSUM), (ii) *Singular Value Decomposition* (SVA) and (iii) *Nonlinear Cross-Prediction Error* (NLCPE), which was already introduced at the end of Chapter 2.

The CUSUM methods detect small shifts in the mean (or other targer statistic) in the time series data characterizing a process of interest. The evaluation of whether a chenge occured is based on so called CUSUM control charts. Let $u(i)$, i=1,...,N is time series of N sample points. Let S_0, S_i, i=1,...,N, called *cumulative sums*, be defined as follows: S_0=0, $S_i = S_{i-1} + u(i) - \bar{u}$, where $\bar{u}$ is the average of all points $u(i)$ in the analyzed signal. Each time we compute next S_i, we add the difference between the current value and the average to the previous sum [Taylor]. Thus, the CUSUM are cumulative sums of *differences* between the consecutive values and the average. When summed up over the whole series, the last sum S_N = 0. The resulting CUSUM chart wull reflect a random variations of the S_i, centered around zero if there are not (significant) changes in the process. Converseley, when a change occurs, the CUSUM chart points will drift up and/or down, depending on the average which usually varies in time. A sudden change in the CUSUM direction reflects a sudden shift, i.e. the average drift. Conversely, no change in the average is reflected on the CUSUM chart as approximately segments with no or little slope.

The CUSUM charts can be applied to both variables and attributes charts including, for example, ranges and standard deviations. The main shortcoming of this method is its poor detecting isolated abnormal points. Noticeably, the CUSUM charts

make use of all historic data: that is, each value on the cusum is a function of all previous data points [151].

In our study we use the Caterpillar SSA implementation of the SVA-based method, where SSA stands for Singular Spectrum Approximation, which is a generalization of the Principal Component Analysis (PCA). The SSA is is a non-parametric method operating on lagged versions of a single time series variables and is is aimed at identification and extracting information on three 'components' of the data: trends, oscillations and noise. The SSA processing is composed of the following four steps [149], [150]:

Step 1: Embedding, where one-dimensional series is transferred to the higher-dimensional series. In this step the *delayed* trajectory matrix is computed.

Step 2: Singular Value Decomposition (SVD), where the trajectory matrix is decomposed into a sum of rank-one bi-orthogonal elementary matrices. In this step the *eigenvectors* and *eigenvalues* are found.

Step 3: Grouping, where diagonal averaging transfers each obtained matrix into a time series.

Step 4: Reconstruction, based on the computed principal component.

The first two steps can be though of as decomposition whereas the last two steps are the reconstruction.

The two parameters used by the SSA are the window length, and the proportion the elementary matrices are accounted for (grouping). The latter affects separation of the initial time series into trend, periodicity and noise.

In general, the SSA method is useful in finding structure in short time series, detection of structural changes. It is applicable to signals virtually regardless their length, both one-multi-dimensional, stationary and nonstationary, deterministic and stochastic.

In Table 12 we show how our methods compare in terms of the recall and precision of change detection in simulated signals vs. three other methods:

Table 12. The overall recall (R) and precision (P) obtained on the simulated data – comparison of four methods: cross-prediction error, CUSUM, SVD-based SSA and *seT*-based IRSEG

cross-prediction error		**CUSUM**		**SSA (SVD)**		***SampEn*-triplet**		**AVERAGED**	
R	P	R	P	R	P	R	P	R	P
0.85	**0.81**	0.77	0.76	0.84	0.7	**0.88**	**0.81**	0.84	0.77

The *SampEn*-triplet based IRSEG outperforms all three competitors on both recall and precision, but its advantage is more pronounced for precision, which for all the methods is in general lower than the corresponding recall: the is the price paid to maintain (high) sensitivity. Yet, by manipulating the parameters: r, m, W (p. 2) and Θ, Δ (p. 9), the recall-precision relation can be tuned according to a predefined cut-off point.

Selected case studies - evaluation

Table 13 presents case studies of applying different sets of the *SampEn*-triplet to diversified set of signals (their descriptions are given in Table 11 on p. 69):

Table 13. The recall (R) and precision (P) obtained by different components of the *SampEn*-triplet:

	H2		**H_Hsrg_H**		**Hundrsmpl**		**L2**		**L8**		**L9**		**sinusoid3**	
	R	**P**	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**
sE	1.00	1.00	0.50	0.50	0.50	0.50	0.50	0.50	1.00	1.00	1.00	0.50	0.50	0.50
rRnd	0.50	0.50	1.00	1.00	1.00	1.00	0.50	0.50	1.00	0.50	1.00	0.33	1.00	1.00
rDiv	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1.00	0.50	1.00	0.50	1.00	0.67
seT	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.67	1.00	0.50	1.00	1.00

Table 14 shows non-redundancy of using the whole *SampEn*-triplet rather than its components separately: we can notice that applying all the statistics improves the recall in certain cases:

Table 14. The cells highlighted orange show the best performing statistic for signal specified in the leftmost column.

data	**sE**	**rRnd**	**rDiv**	**ALL**	**max**
H2	1	0.5	0.5	1	1
H_Hsrg_H	0.5	1	0.5	1	1
Hundrsmpl	0.5	1	0.5	1	1
L2	0.5	0.5	0.5	1	0.5
L8	1	1	1	1	1
L9	1	1	1	1	1
sinusoids3	0.5	1	1	1	1
overall	**0.7**	**0.9**	**0.7**	**1.0**	0.9

In the Appendix we present a series of results for particular cases of the simulated signals. Here we show just two such illustrations: The first is for *logistic* map (L6 in Table 17) simulated with r=3.828427123 and $x(1)$=0.7 as shown in Figure 16:

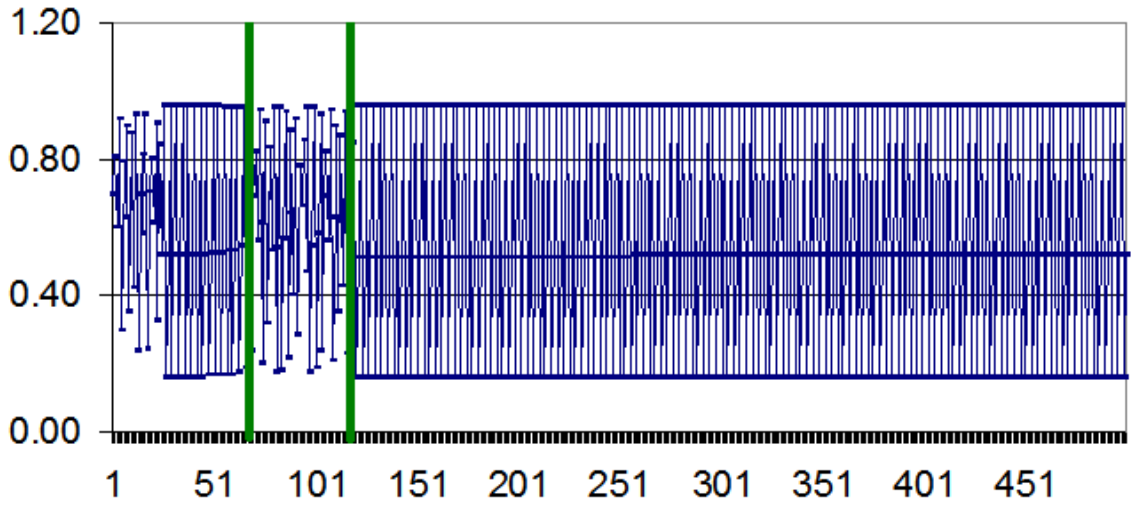


Figure 16. Plot of the *logistic* map: seq. ID=1. The vertical green bars are located at 60 and 110. The dependent variable denotes subsequent points of the signals and dependent variable denotes the points' values

The second example also involves *logistic* map shown on Figure 17, but this time simulated with the parameter r drifting from 3.828 to 3.828427123 and initialized for $x(1)$=0.4 (case 1) and its version concatenated with (uniform) white noise of values ranging from 0 to 15% of the maximum value of original signal (case 2). Whereas the change point – exactly in the middle of the orginal signal (case 1) can be easily captured

visualy, its presence (let alone temporal position) is much less obvious in the noise-concaminated signal (case 2). This example illustrates the robustness to noise of our methods without compromising sensitivity to dynamic changes.

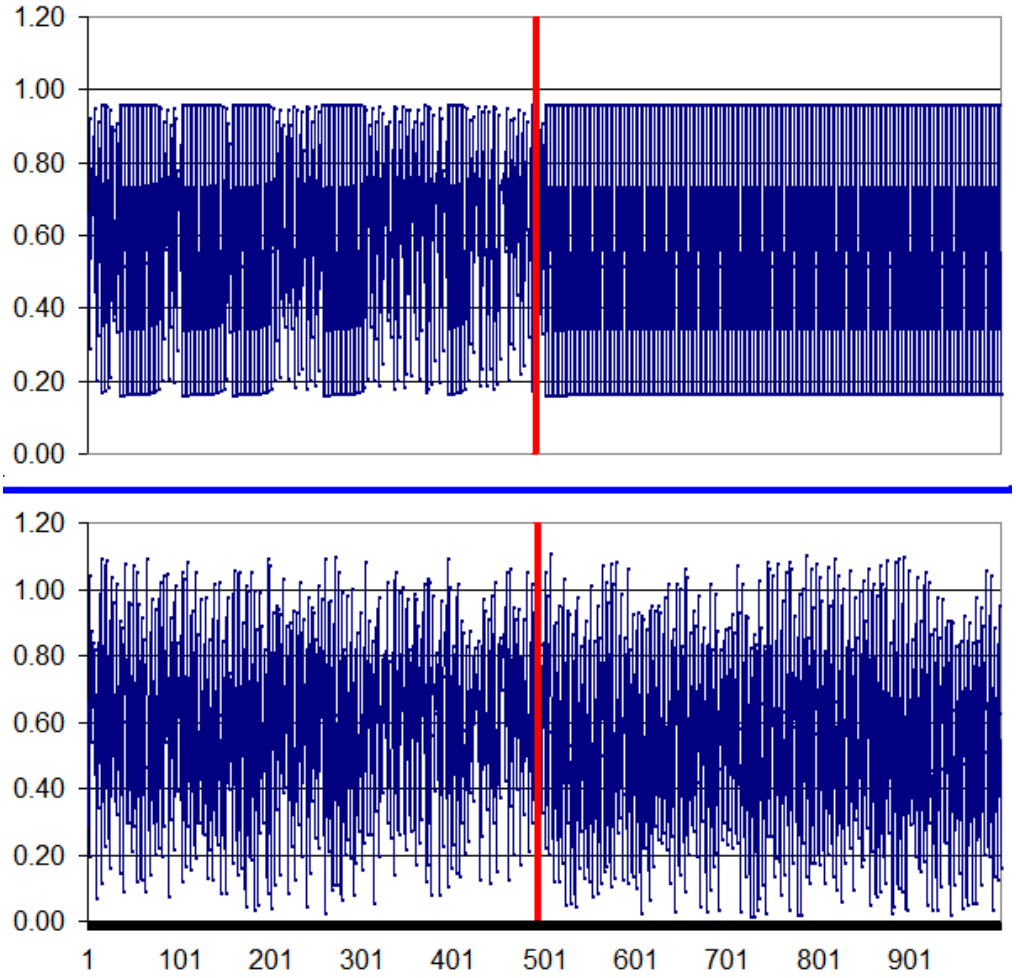


Figure 17. Upper plot: case 1 (original data). Lower plot: case 2 (data with noise). The vertical bars indicate temporal position of dynamical change (identical in both cases).

4. DYNAMIC CHANGE IN CARDIAC SIGNALS RECORDED IN THE PRESENCE OF EXTERNAL STIMULI

Our study of complexity measures originated from the assumption that physiological responses in general, and the ECG or HRV in particular, decrease in complexity as an organism ages or suffers from disease. In this chapter, we will investigate whether the developed measures are able to identify changes in HRV complexity patterns, and whether such changes differ sufficiently between a group of patients and a group of healthy participants in order to have diagnostic value. The study involves ECG data from a total of 143 participants, which were subjected to externally controlled stimuli, and seeks to answer the following two questions:

i. *How do different groups of participants (patients vs. healthy) respond to different external stimuli if measured in terms of different statistics* (considered as HRV complexity measures)*?*

and

ii. *Can different stimuli help identify a participant's health category by affecting different HRV statistics* (considered as classification attributes)*?*

These rely on two completely different computational settings and evaluation methods: In the former case, our aim is to identify change points between different stimuli, and we compute the corresponding recall and precision, using different HRV quantifiers

(selected from the Kubios statistics and our *SampEn*-triplet methods). Then, to address the latter case, we use a classifier to evaluate the discriminative power of different (sets of) HRV quantifiers – now thought of as classification attributes – in discerning between the healthy and patients. The evaluation is expressed as the percentage of correctly classified cases. As we will see, a special emphasis is put on the *SampEn*-triplet and music as the (HRV complexity) statistics and stimulus of primary interest.

A. THE COPERNICUS DATA: APPLYING EXTERNAL STIMULI

Although studies of environmental changes applied in the HRV analysis are rather frequently reported in literature [78], to date we have not encountered audible stimuli (e.g. music and white noise) in this research area[10].

The controlled environmental stimuli – experimental protocol

The data acquisition part of this study took place at the 22nd Military Hospital of Spa and Rehabilitation under supervision of the Dept. of Cardiology and Internal Diseases at Nicolaus Copernicus University (Poland), compliant with an experimental protocol approved by the local Ethics Committee. Our participation in analyzing the data has been approved by the GMU IRB.

[10] Our experiment must not be confused with the technique of 'sonification', which is a sonic representation of the HRV [Ballora], as well as with a wide variety of assessment of the HRV affected by physical excercises

143 volunteers participated in the study after providing informed written consent. After the quality check, we chose data from 121 volunteers for further analysis, which we assigned to one of four disjoint health categories: *healthy* (normal), *diabetic without β*-blocker, *non-diabetic β*-blocker and *diabetic with β*-blocker, denoted *rrH, rrD, rrB* and *rrDB*, respectively. The demographic characteristics for each group, including age, sex, (occurrence of) *ischemic cardiomyopathy* (ICM), *non-ischemic dilated cardiomyopathy* (NI-DCM), previous stroke and/or heart attack are shown in Table 15.

Table 15. Distribution of the participants with regards to age, sex and cardiac problems

health category	**number of cases**	**age average**	**age range**	**male-female** (%)	**heart attack** (%)	**stroke** (%)
rrH	26	61	41-76	39-61	12	0
rrB	48	63	39-84	65-35	96	81
rrD	10	66	49-84	60-40	80	10
rrDB	37	64	45-79	66-34	68	8
TOTAL	**121**	**63.5**	**39-84**	**51-49**	**60%**	**6%**

The participants were fitted with the 'EKG 14C' module (a part of the commercial CARDIV system). Each electrocardiogram acquisition lasted for approximate 35 minutes, during which the participant was subjected to a series of controlled stimuli, as seen in Table 16. Specifically, subjects were asked to:

(a) listen to white noise in sitting position (5 min),

(b) listen to music in sitting position (5 min rock, 5 min classical, 5 min techno),

(c) listen to white noise in sitting position (5 min),

(d) stand up (5 min),

(e) listen to white noise in sitting position (5 min),

(f) undergo a thermal bath, by immersing their arm into water of approximately 10° C for approx. 1 minute.

Table 16 shows the temporal structure of the experiment:

Table 16. Temporal structure of the ECG recording

step	**1**	**2**	**3**	**4**	**5**	**6**
stimulus	white noise	three sequences of different music genres of equal duration	white noise	postural change from sitting to standing	white noise	thermal bath
duration (minutes)	5	15	5	5	5	1

The sequences of heartbeat intervals, denoted RR, were retrieved from ECG recorded in presence of controlled environmental stimuli. The recording sessions were conducted using the CARDIV acquisition system with the sampling frequency 1000 Hz and at resolution 2370 nV/LSB. The R peaks were identified using a modified *Pan & Tompkins*

method [112]. The recorded RR time series were further conditioned regarding the duration: those outside the range of [500,1500] milliseconds or differing from their predecessor by more than 30% were subject of automated preprocessing (i.e. removal, interpolation or splitting). Highly untypical or confusing cases were consulted individually (based on the underlying ECG signal), of which some were discarded from further analyses.

Data preprocessing – general rules

We conducted data preprocessing of the cardiac RR time series as a standard procedure before the analysis. Of special importance for us was adequate outlier removal as well as other artefacts incurred during the signal acquisition. In particular, we searched for all occurrences of RR intervals longer than 1500 or shorter than 500 (ms). In general, we handled such cases by either (i) interpolating a missing R-wave and splitting the spuriously long interval into (usually) two equal R-R intervals or (ii) removing erroneously identified R-wave – i.e. spuriously short interval. These procedures are described in detail in [15]. In the former case of intrapolating, we handle a questionable sample point (x, n) positioned between (x_0, n_0) and (x_1, $n_1 = n_0+1$), and replaced it by inserting the linear interpolation of those, according to the formula: $x = x_0 + (x_1 - x_0)\frac{n - n_0}{n_1 - n_0}$, where x represents the RR interval at temporal position n.

Finally, we plot selected time series – particularly those representing the category of data they belong to. Such illustrations help realize the non-trivial nature of dynamic change detection methods when no indicators of the changes can be captured by eye.

RR signals retrieved from ECG recorded in presence of the stimuli

The descriptive statistics for the RR interval duration for the healthy and (averaged) patients is shown in Table 17:

Table 17. Descriptive statistics of the whole population of RR data sets. All the patients cases: *rrB*, *rrD* and *rrDB* were considered as one group, hence the the statistics were averaged except the minimum and maximum, which was identified globally (for the aggregated group)

STATISTIC	healthy	patients
mean	865	892
median	876	905
mode	883	906
SD	64	57
kurtosis	1.32	3
skewness	-0.22	-1
range	468	466
minimum	653	
maximum	1121	

As we can see from Table 20, the RR signals from the healthy have slightly higher variance and their normality is less compromised than that of the patients. A selection of the cardiac RR signals recorded in presence of external stimuli is shown on Figure 19:

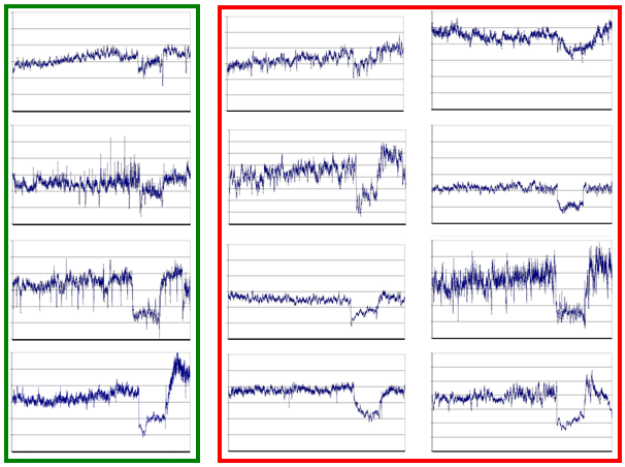


Figure 19. Visualization of selected RR cardiac signals. The left columns of plots illustrate signals from the healthy group and two following columns show signals from the patient group

Reference Point Definition

Based on the structure of the experiment, we defined nine reference points as follows, some of which we use for evaluating the accuracy of change detection:

- s01 = midpoint of the first white noise,
- s10 = onset of (the first genre of) music,
- s11 = onset of the second genre of music,
- s12 = onset of the third genre of music

- s20 = onset of the second white noise,
- s21 = midpoint of the second white noise
- s30 = onset of the postural change,
- s40 = onset of the third white noise,
- s50 = onset of the thermal bath.

In Figure 20 we use an exemplary (stimuli-influenced) RR signal to illustrate how the reference points are related to the stimuli onsets:

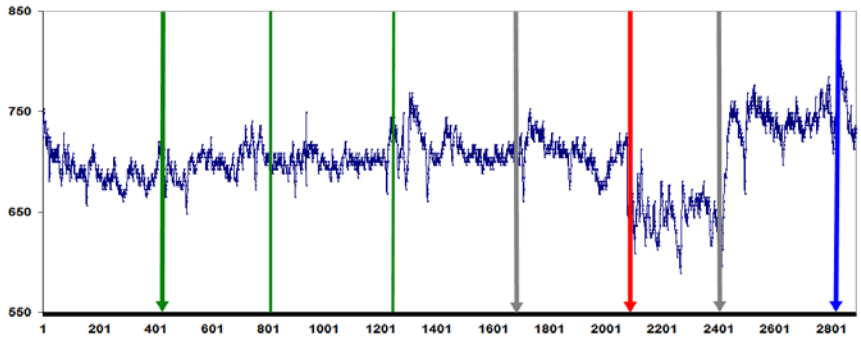


Figure 20. RR signal recorded in the presence of external stimuli, whose onsets are marked by vertical arrows corresponding to: green – music, grey – white nose, red – postural change, blue – thermal bath (two green bars directly following the green arrow denote onsets of different music genres)

Surrogate data tests of the RR signals.

There is abundant evidence [179] that shows that the cardiac regulatory mechanisms operate in highly nonlinear manner. These nonlinearities have their source in complex

nature of the *baroreflex feedback loop* [59], where the *amplifications* and *delays* in the sympathetic branch of the autonomic nervous system cause instabilities which – in a *linear* system – would have led to its physical blow-up to infinity [140]. Since nothing like this happens – by contradiction – the dynamics of cardiovascular system *cannot* be linear/stochastic[11] by nature, hence its adequate description is beyond linear methods (e.g. ARMA, spectral analyses or probability distribution). Regardless of theoretical considerations, we conducted the *surrogate data test* of nonlinearity: For each of blindly selected 20 RR signals from the 'healthy' group we randomly shuffled the samples (within each signal) and repeated this procedure to obtain ensemble of 19 surrogate datasets per signal. Using the *SampEn* for testing (non)linearity we rejected the null hypothesis of '*uncorrelated*, i.i.d., *linear/stochastic* data' at significance level 0.05 (the *linear* methods were *non-discriminant*). This supports conjecture of nontriviality of the RR data and justifies application of nonlinear methods.

B. PHYSIONET DATA: VENTRICULAR TACHYARRHYTHMIA AND MEDITATION

Publicly available data from the PhysioBank resources often serve as a well recognized benchmarks for evaluation of HRV characterization algorithms. We chose two data sets,

[11] Roughly speaking, *stochasticity* can be attributed to two different sources: high-dimensional (perhaps deterministic) dynamics or noise. The former is intrinsic to very complex systems (e.g. neural) whereas the latter can be linked to the measurement process [40]. Hence, in general, we can not decide if what is *perceived as a noise* is not in fact *deeply hidden determinism* (of high-dimensional chaos)

which contain signals registered *before* and *after* an event of interest, hence being suitable for testing our methods of change detection.

ICD-captured Spontaneous Ventricular Tachyarrhythmia Database

We selected 15 cases of episodes of ventricular tachycardia (VT) from the 'Spontaneous Ventricular Tachyarrhythmia Database Version 1.0 from Medtronic, Inc.', [38] for testing our methods.

The RR signals are acquired by detecting consecutive R-waves and the episodes of VT are identified on the fly. The 1024 most recent RR intervals are stored in the implanted cardioverter defibrillators (ICD) buffer. When an episode of VT is detected, the ICD records a snapshot of the RR interval buffer, i.e. the sequence of 1024 RR intervals immediately preceding the VT episode. The sequence ends at the time when the ICD *first* detected the episode (i.e. not at its termination). Before the ICD being interrogated, the most recent RR intervals in the ICD buffer remains available: these can be used as the baseline data to compare with the detected events.

Figure 21 shows plots of two ICD signals: the left half of each represents the 1000 RR intervals directly preceding the VT event, followed by 1000 RR intervals captured.

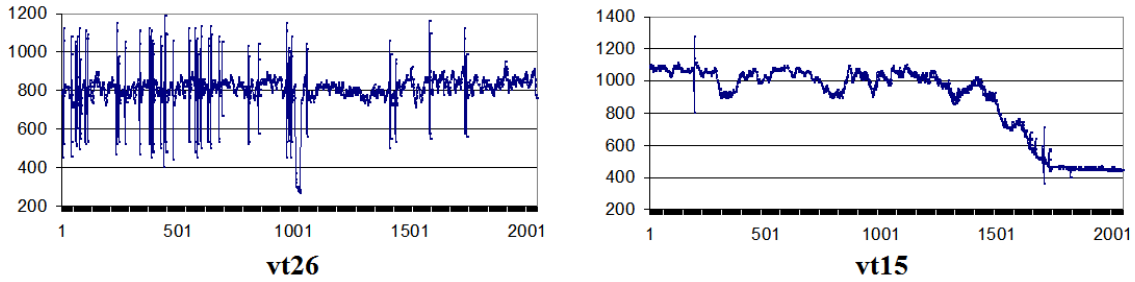


Figure 21. Two cases of VT labeled as vt 26 and vt15 (description in text)

Next we preprocessed the vt26 signal the same way we proceeded with the Copernicus data, what is shown in Figure 22. Next we applied IRSEG to both of these signals (i.e. original and preprocessed vt26), which identified the middle of each signal as putative change point.

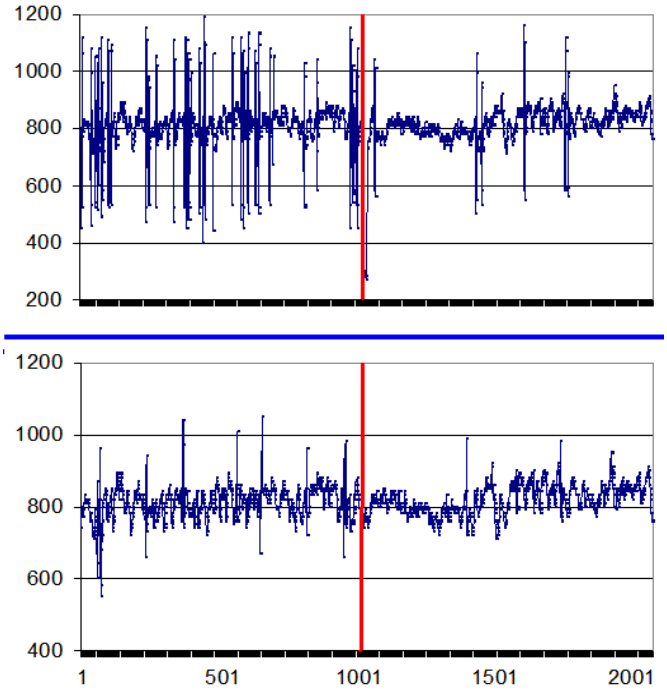


Figure 22. The vt26 before (upper plot) and after (lower plot) preprocessing. The red bar shows temporal position of the VT event captured by ICD apparatus

However, when applied to the vt15 signal, IRSEG failed to detect the midpoint of the signal as the change point.

Exaggerated Heart Rate Oscillations During Two Meditation Techniques

These two data sets are the RR time series recorded *before* and *after* two well-known meditations: Chinese *Tai Chi* (8 double recordings) and *Kundalini Yoga* (four double recordings). To make the signals suitable for our tests, for each subject we concatenated the corresponding pre- and during-meditation heart rate sequences to obtain eight compound signals for the *Tai Chi* data and four compound signals for the *Yoga* data. We

then analyzed the heart rate sequences affected by *Kundalini Yoga* meditation and proceeded analogously for the *Tai Chi* meditation[12].

In this chapter we try to find whether the meditation – considered as an external factor – can be detected using our methods of change detection. Specifically, we are interested in how accurate our methods are in terms of both *sensitivity* and *specificity*. We do not downplay the possibility that such evaluation may suffer from very limited information on the experimental setup, for instance other possible (uncontrolled or not intended) factors possibly having side effects on the HRV (e.g. medicine, ambient noise, emotional stress, diseases). Nevertheless, such an evaluation still makes sense when comparing performance of our methods vs. the mainstream HRV statistics. Then, our quest can be supplemental to those of the authors who 'sought to determine: 1) whether there are any distinctive heart rate dynamics during these practices, and 2) whether meditative states induce a quiescent (less variable) dynamics during a variety of physiological and active (more variable) pattern of autonomic response.'

The plots in figures 23 and 24 illustrate RR signal before (left part, blue dots) and during (right part, red dots) the *Tai Chi* and *Kundalini Yoga* meditation, respectively. The vertical green bar represents the onset of meditation. When evaluated visually, the dynamics of each part differ clearly.

[12] Some of the series were truncated at the size of the shortest signal in a group. Each signal contains the data representing the pre-meditation and meditation part

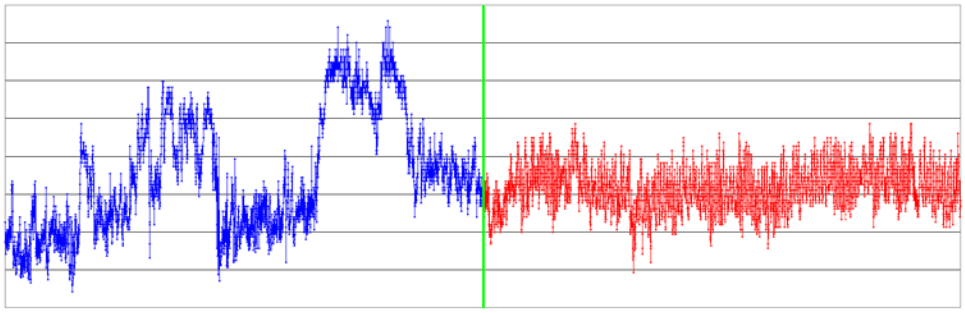


Figure 23. Plot of RR signal recorded before (left part, blue dots) and during (right part, red dots) the *Tai Chi* meditation. The vertical green bar denotes onset of the meditation

Roughly speaking, in both cases the meditation (1) 'stabilizes' the HRV and apparently (2) reduces the average heart rate (i.e. increases the RR intervals). The former effect is stronger pronounced in the case of *Tai Chi*, whereas the latter effect is fairly evident in the case of *Yoga*.

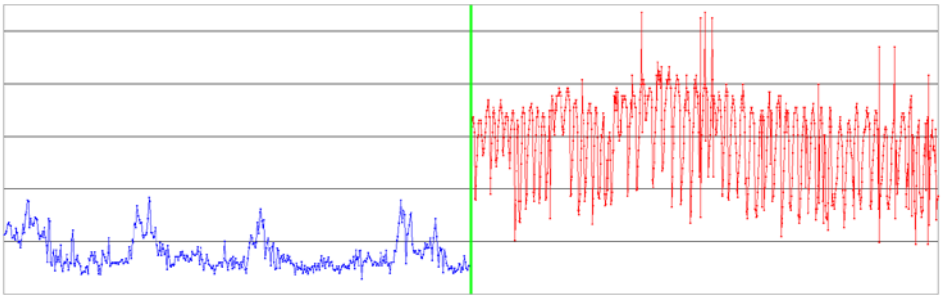


Figure 24. Plot of RR signal recorded before (left part, blue dots) and during (right part, red dots) the *Kundalini Yoga* meditation. The vertical green bar denotes onset of the meditation

C. THE KUBIOS HRV TOOLKIT

The Kubios HRV toolkit is a publicly available software package that was developed by [124]. It computes the mainstream HRV statistics, but can also be applied to other (than RR) time series.

The Kubios HRV is not an open source toolkit, hence user ability to tune parameters is limited. Nevertheless, the user can change the values of most parameters related to frequency-domain analyses. In particular, we use in our study the following (default) parameter values and settings available in Kubios:

- HRV frequency bands :
  - VLF: 0-0.04 Hz,
  - LF : 0.04-0.15 Hz,
  - HF: 0.15-0.4 Hz,
- RR interpolation rate: 4 Hz,
- FFT spectrum estimation:
  - window size: 256 s,
  - window overlap: 50%,
- AR spectrum model order: 16,
- points in frequency domain: 256 points/Hz.

Prior to finalizing the Kubios settings for our analysis, we tested (independently) three window sizes: 128, 256, 512, and three overlaps: 33%, 50% 67%. We applied Kubios to 10 different RR signals from the Copernicus University and computed 13 ‘core’ HRV

statistics. Based on these, we found the Kubios default parameter values most reproducible across the selected signals and HRV statistics.

Because many of the 33 statistics computed by Kubios HRV are highly correlated, we selected 13 that correlate weakly and were subsequently found to be the 'best performers' in our tests for change detection and classification accuracy. Finally, they belong to the mainstream 'gold standards' used in the HRV analyses. Such a selection makes evaluation of our methods more demanding, yet by the same token more convincing given our methods are at least *not* inferior. The selected HRV statistics are those highlighted in Table 18:

Table 18. The Kubios HRV statistics selected for our analysis

time-domain	frequency-domain	NLD and fractal
Mean	VLF peak	Mean Line Length
STD	LF peak	Max Line Length
MeanHR	HF peak	Recurrence Rate
STHR	VLF absolute	Determinism
RMSSD	LF absolute	Shannon
NN50	HF absolute	DFA alpha1
pNN50	VLF relative	DFA alpha2
RRtrindex	LF relative	ApEn
TINN	HF relative	SampEn
	LF normalized	D2
	HF normalized	SD1
	LFHF	SD2

D. RESULTS AND EVALUATION

Likewise in previous chapter – when the subject of our analysis were simulated data - now we apply the IRSEG algorithm to the cardiac RR signals described in sections A and B.

Copernicus data: change detection

As a first step, we applied our methods to the Copernicus data. For evaluating the change detection accuracy we need a strict definition of the *reference points* – analogous to those marked by vertical red bars for simulated signals. The nine *reference* (change) *points*, denoted *RCP*, are described in Table 19, some of which will be used in the accuracy evaluation.

1. evaluation of *recall* and *precision* of the IRSEG algorithm with regards to the applied stimuli considered as change points

Table 19. List of the nine *reference points* and their corresponding r^{TP} factors which determine whether a reference point (RCP) should be considered as a *change point*, thus contributing to TP (true positives)

i	*RCP*	description	r^{TP}
1	*s01*	*midpoint* of the first white noise (directly preceding music)	0
2	*s10*	*onset* of music	1
3	*s11*	(onset of the) *second tercile* of music	1
4	*s12*	(onset of the) *third tercile* of music	1
5	*s20*	*onset* of the second white noise (directly following music)	1
6	*s21*	*midpoint* of second white noise	0
7	*s30*	*onset* of postural change	1
8	*s40*	*onset* of the third white noise (between postural change and thermal bath)	1
9	*s50*	*onset* of thermal bath	1

Let each *RCP*(*i*), where *i*=1,...,9, be characterized by its corresponding $r^{TP}(i)$ and $r^{FP}(i)$, for instance: $RCP(2) = s10$ and $r^{TP}(2) = 1$. Furthermore, if $r^{TP}(i) = 1$ for some i^{th} *RCP*, then classifying it as a change point entails incrementing the sum of *true positives*: $TP = TP + 1$. Conversely, ignoring it (as a change point) entails incrementing the sum of *false negatives*: $FN = FN + 1$. On the other hand, if $r^{TP}(i) = 0$, then (mis)classifying it as a change point increments the sum of *false positives*: $FP = FP + 1$. This way the three statistics (*TP, FP* and *FN*) are maintained (during analysis of a RR signal) and make them final contribution to the related α and β. We arbitrarily decided that:

a) the *midpoints* of the first and second (duration of) *white noise* are assigned $r^{TP} = 0$,

b) the *onset* of each *music tercile* – representing different genre – is assigned $r^{TP} = 1$.

The former implies none expected 'significant' effect in the middle of first white noise, whereas the latter accounts for a wide diversity (among the participants) regarding personal hearing capability, cultural background and preferences (the effect of particular genres on the HRV may differ). This is equivalent to *relaxing* the condition for detecting music. We prevent the music-related *TP* from exceeding unity imposing $s^{TP}_{music} =: 1$ if $\sum_{i=2}^{4} r^{TP}(i) \geq 1$, and 0 otherwise. Hence, the total *TP* score per one RR signal is $0 \leq s^{TP}_{i=1,\ldots,9} \leq 5$.

Based on the above setting, we compute the recall and precision for each health-category group, as shown in Table 20:

Table 20: Accuracy evaluation with respect to *health category* groups where all the stimuli were accounted for. For each group $C=TP+FN=5$

group	recall	precision
all healthy	0.84	0.73
patient *rrD* patients *rrB* patients *rrDB*	0.79 0.72 0.84	0.84 0.78 0.89
all patients	0.79	0.83
ALL CARDIAC	**0.81**	**0.77**

Clearly, the difference between recalls obtained for the healthy and the (all) patients is not negligible (0.84 vs. 0.79, respectively).

Alternatively, for the same population of participants we can present the statistics as *stimulus*-wise, aggregated over all health categories – see Table 21:

Table 21. The recall for the selected *RPC*'s with $r^{TP}=1$

RCP (*TP*=1)	**recall – all subjects**	**recall - healthy**	**recall - patients**
s10, s11, s12 (music)	0.78	0.88	0.76
s20 (white noise)	0.63	0.61	0.64
s30 (postural change)	0.91	0.92	0.89
s40 (white noise)	0.83	0.82	0.83
s50 (thermal bath)	0.88	0.95	0.85
ALL STIMULI	**0.81**	**0.84**	**0.79**

Looking at Table 24 we again observe a difference between recalls of the healthy and patients but now on the level of specific stimuli. These particular results will be used for performance comparison of our methods vs. three other methods in next section.

Classification accuracy and feature ranking

Based on the 17 statistics: 14 computed by the Kubios toolkit, namely Mean, STD, RMSSD, pNN50, trindex, TINN, LF, HF, LF2HF, SD1, REC, *H*, *SampEn*, and *D2*, and three of our methods: *sE, rRnd* and *rDiv,* we conducted a comprehensive analysis of the RR signals comprised of three independent parts:

1. assessing influence of *music*, *white noise-2*, *postural* change and *thermal bath* for the *healthy* and *patients* using t-Test and SVM-based classification;
2. discerning the *healthy* from the *patients* based on the 100 samples from the RR signal which are located directly:
    a. *before* applying music,
    b. *after* applying music,
    c. *before* applying white noise-2,
    d. *after* applying white noise-2,
    e. *before* applying postural change,
    f. *after* applying postural change,
    g. *before* applying thermal bath,
    h. *after* applying thermal bath,

    using SVM-based classification.

3. discerning the *healthy* from the *patients* based on
    a. the part of RR signal *before* music onset (i.e. only white noise accounted for)
    b. the whole signal after the music onset (i.e. all the stimuli accounted for)

    using SVM-based classification

These analyses are conducted using (i) supervised classification based on the Support Vector Machine (SVM) classifier and (ii) attribute selection and/or ranking. We present here only results for selected four 'best performers' from the 14 Kubios HRV statistics, in particular: trindex, TINN, LF2HF and RMSSD, and the best performer of the *SampEn*-triplet, which is *rRnd*.

Classification attribute ranking

Table 24. The (HRV) attribute ranking with regards to H vs. P classification. All the stimuli were used. The ranking method applied for the ranking was InfoGainAttributeEval/Ranker (Weka)

#	attribute	rank
1	rDiv	1.4
2	rRnd	3.6
3	Rrtrindex	4
4	TINN	4
5	pNN50	4.4
6	RMSSD	5.8
7	STD	7
8	LFabs	7.6
9	HFabs	9
10	sE	9.8
11	LF2HF	10.8
12	D2	11.2
13	SampEn	12.6
14	SD1	14.2
15	Shannon	14.6
16	Mean	16

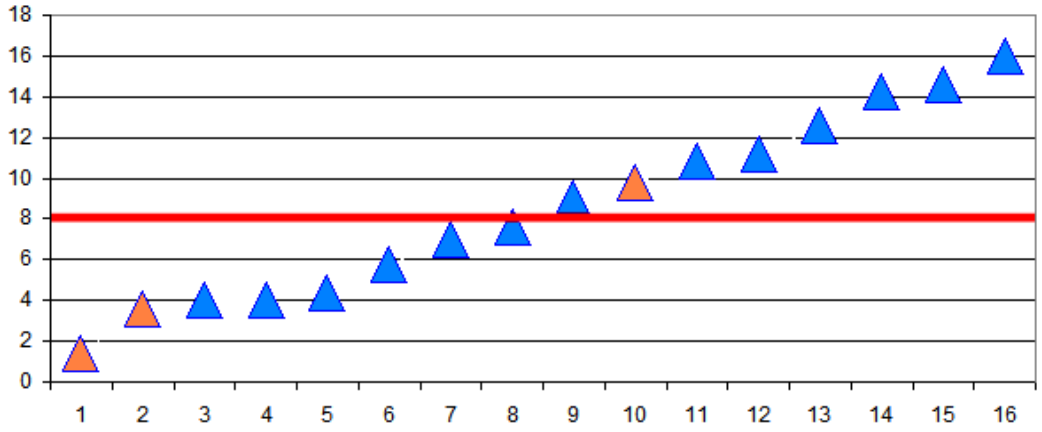


Figure 25. Visual presentation of Table 34

Table 25. Percentage of correctly classified case (H vs. P) for selected (sub)sets of the HRV attributes. Tle rightmost column shows wheather applying all the stimuli gave the accuracy higher than applying only the best (single) stimulus

statistic set	*WN1*	*M123*	*W2*	*PC*	*TB*	*ALL-TB*	*ALL*
K	68.7	65	66.7	72.7	76.3	76	75.7
trindex	71.3	71.3	70.3	76.7	76.3	**79**	76.3
TINN	69.7	71.3	66.7	77	74.3	78	76.3
seT (all)	**73**	**71.7**	71	**77.3**	**75.3**	78	76.7
sE	72.3	69.7	71	78.7	**80**	78	77
rRnd	71.7	70.3	71	75.7	75	78	78.7
rDiv	70.3	70.3	69	77.3	77.7	**79**	**79.3**
K + sE	66.3	64.3	67.3	75.7	77.7	77	74.3
K + rRnd	67.7	65	67	74	77.3	76	76.3
seT + trindex	70	71	**70.3**	76.7	75.7	77	78.7
seT + TINN	71	**71.7**	71.3	76.7	75.3	77	76.3
K + seT	70.3	67.7	65.3	75	78.7	76	77.3

Table 25 uses the following abbreviations:

- WN1: white noise – the first sequence,
- M123: (all) three sequences of music,
- WN2: white noise – the second sequence,
- PC: postural change,
- WN3: white noise – the third sequence,
- TB: thermal bath,
- ALL-TB: all the stimuli besides thermal bath,
- ALL: all stimuli,

t-tests

Independently of the classification we conducted a series of three t-tests (for independent samples).

First, for each participant – regardless his/her health category – we isolated the part of the RR signal before music was applied (i.e. including only the first white noise). Let this stimuli-less subsequence be called $RR0^{H+P}$, and its original counterpart (i.e. the whole RR signal) $RR1^{H+P}$. Having selected *rRnd* as the best performer of our SampEn-triplet and RMSSD, pNN50, trindex, TINN, LF2HF, H and SampEn (not to be confused with our SampEN-triplet) as the best performers from Kubios HRV, we apply each of them to obtain 8+1=9 series of 121 numbers (the number of participants) for each of the modified signals: $RR0^{H+P}$ and $RR1^{H+P}$ (this is *paired* t-test). We found statistically significant difference between the two signals for all but RMSSD and pNN50 HRV statistics. This means that – in general – applying the stimuli (in this considerations we do *not* count the first white noise as a stimulus) *does* influence the HRV significanlty.

In the second step we divide the RR signals with respect to health categories into two groups: H (healthy) and P (patients representing all three cardiac cases: diabetic, β-blocker/nondiabetic and β-blocker/diabetic). In this step we try to test hypothesis that – using only the (stimuli-free) RR0 signals (and more precisely: $RR0^{H}$ and $RR0^{P}$) – there is no difference in the participants' HRV. Using the same set of nine HRV statistics we were able to reject the null hypothesis (of no difference) only for the Shannon entropy and *rRnd* (the latter being part of the *SampEn*-triplet). This test – likewise the one

described in next paragraph – is *independent* t-test (the number of compared samples from each subject group was different).

Finally, the analogous tests based on the the whole stimuli-induced $RR1^{H}$ and $RR1^{P}$ signals show no statistical differences between the two health categories if the HRV was quantified by statistic other than *rRnd*.

Comparison with other methods of change detection: Copernicus data

Having shown (surrogate data test, p. 13) the relevance of applying *nonlinear* methods to our analyses of cardiac data we now demonstrate how the introduced methods are particularly useful for the task of nonstationarity detection in RR signals. As illustrated in Figure 17 (p.78), plots of the RR signals – regardless their health category association – have characteristic *dips coinciding* with the occurrence of *postural change*, followed by well-pronounced fluctuations. However, *prior* to the postural change there is *no* prevailing visual pattern (in general and/or) correlating with the related health category (in particular). Analyses based on the mainstream HRV statistics[13] turned not very helpful in detecting stimuli other than postural change and perhaps thermal bath. On the other hand, our methods applied to the same data show good overall sensitivity to the stimuli, *including music*, as well as robustness to noise – a feature positively impacting the overall accuracy. Although the methods we introduce here are *not exclusively* for detecting the influence of music on the HRV, they nonetheless prove to be relevant when

[13] This refers to another study conducted on the identical cardiac data, but using *machine learning* methods. The term '*mainstream HRV statistics*' refers to those implemented in the HRV Kubios toolkit [49]

(detection of) *subtle* dynamical changes in time series are helpful to get more insight into the underlying process.

In what follows we compare the *seT*-based IRSEG algorithm with the Caterpillar-SSA based on SVD (Singular Value Decomposition). The results presented in Table 26 were aggregated over *all signals* in their related groups (of the healthy and patients). They account for *all* the *stimuli* (considered as change points).

Table 26. The overall recall and precision of change (stimuli) detection for different groups of participants – comparison of four methods. From left to right: nonlinear cross-prediction error, CUSUM, SVD-based Singular Spectrum Analysis (SSA) and *seT*-based IRSEG

DATA	**nonlinear cross-prediction error**		**CUSUM**		**SSA** (SVD)		**IRSEG**		**AVERAGED**	
	R	P	R	P	R	P	R	P	R	P
cardiac *healthy*	0.75	**0.88**	0.71	0.71	0.78	0.7	**0.84**	0.73	0.78	0.76
cardiac *patients*	0.76	**0.91**	0.65	0.83	0.71	0.75	**0.79**	0.83	0.73	0.83
ALL cardiac	0.78	**0.86**	0.66	0.79	0.72	0.74	**0.81**	0.77	0.74	0.79

The IRSEG and Caterpillar-SSA achieved recalls 0.81 and 0.72 when applied to the *cardiac* RR signals, with 0.83 and 0.74 for all the data, respectively. Noticeably, the average response to the stimuli from *healthy* subjects' was *stronger* than that from the patients if quantified by the recall: 0.83 and 0.75, respectively. The precision of the methods is in general lower than the corresponding recall, what is the price paid to maintain (high) sensitivity. Yet, by manipulating the parameters: r, m, W (p. 2) and Θ, Δ (p. 9), the recall-precision relation can be tuned according to a predefined cut-off point.

Focusing exclusively on the *cardiac* data (all health categories), Table 27 shows cross-comparison of the Caterpillar vs. IRSEG and *music* (all three genres) vs. *postural change*. As expected, the influence on the HRV from the latter is easier to capture regardless the method used. Remarkably, the *seT*-based IRSEG outperforms Caterpillar-SSA in average recall being *more sensitive* to each of the stimuli.

Table 27. The *recall* obtained by the *nonlinear cross-prediction error*, *CUSUM*, *SSA* and *seT*-based IRSEG for *music* and *postural change* using the cardiac data

STIMULUS	**nonlinear cross-prediction error**	**CUSUM**	**SSA** (SVD)	**IRSEG**	**AVERAGED**
music (all genres)	0.74	0.61	0.68	**0.78**	0.70
postural change	0.87	**0.92**	0.83	0.91	0.88
ALL stimuli (averaged)	0.79	0.77	0.76	**0.85**	

Comparison with other methods of change detection: Physionet data

Analogously as for the Copernicus data, for the same methods we conducted comparison of their respective accuracies:

Table 28. Results for change detection for RR recorded during Ventricular Tachyarrhythmia episodes

method	recall	precision
IRSEG	0.80	0.71
nonlinear cross-prediction error	0.73	0.62
SSA	0.60	0.67
CUSUM	0.47	0.82

Table 29. Results for change detection for RR recorded during *yoga* and *chi* meditations

yoga			***chi***		
method	**recall**	**precision**	**method**	**recall**	**precision**
IRSEG	1	0.67	IRSEG	0.88	0.69
SSA	1	0.44	SSA	0.63	0.61
CUSUM	1	0.75	CUSUM	0.75	0.80
nonlinear cross-prediction error	1	0.58	nonlinear cross-prediction error	0.92	0.53

As a special case of comparison we use the SAX-based JMotif implementation (credit Pavel Senin, University of Hawai), known as HOT SAX – an algorithm devised for discord and anomaly detection in time series. As we already stated, the anomaly (discord) detection and change detection – although similar in some respect – are by no means the same category of application. Hence the difference in performance between HOT SAX and IRSEG when the both algorithms are applied to cardiac RR signals should not be

surprising. In particular, the HOT SAX obtained the following recall and precision averaged over the whole dataset of 121 RR signals:

k=7, W=50: recall=0.42, precision=0.45 (recall for music only = 0.13),

k=10, W=50: recall=0.51, precision=0.38 (recall for music only = 0.22),

k=7, W=100: recall=0.48, precision=0.48 (recall for music only = 0.21),

k=10, W=100: recall=0.62, precision=0.44 (recall for music only = 0.41),

where k is the predefined number of discords to be identified and W is the window size. For comparison, IRSEG achieved overall recall 0.81 and precision 0.77 (recall for music was 0.78).

Summary

We can summarize our analysis of the RR signals with nonstationarity enforced by environmental stimuli as follows:

1. Detection and temporal localization of the stimuli onsets in the cardiac RR signals is a rather challenging task, partially because of complexity of the human cardiovascular system and – to some extent – individual sensitivity to the stimuli. This applies especially to the 'soft' stimuli, i.e. music (and white noise) and meditations. These two factors combined may cause weakly pronounced effect of the stimuli on HRV and/or the effect being delay (see considerations on complexity of live organisms in chapter 2).
2. The problem of detecting change in HRV influenced by the 'soft' stimuli is better approached using methods which account for *dynamical* rather statistical effects.

Hence our method and – to some extent – the method of cross-prediction error help obtain higher recalls, although sometime at the cost of (lower) precision. The latter is the case of the cross-prediction error-based method: Although devised for analysis of nonlinear dynamical processes, it requires substantially more data (i.e. longer time series) than what we have available for our study. The second factor of rather poor precision of the latter method is its poor robustness to noise, what is one of strong features of our algorithm based on the *SampEn*-triplet.

5. SYMBOLIC ANALYSIS OF DYNAMICAL CHANGE

Symbolic dynamics are based on a coarse-graining of the measurements, where the data are transformed into patterns composed of only a few symbols, e.g. letters from some predefined alphabet. This simplifies the analyzed dynamics, which are now based on a reduced description of symbol sequences. In doing so one loses some amount of detailed information, but at the same time makes the analysis more robust against noise. Another possible advantage of the analysis of symbolic sequences is lower computational cost [72].

A. THE PROBLEM OF ADEQUATE PARTITION FOR SYMBOLIC TRANSFORMATION

To facilitate symbolic analysis, the original (i.e. real number) data must be discretized, i.e. transformed into a sequence of symbols. This highly depends on the choice of quantizing intervals: i.e. how one *partitions* (discretizes) the original data in order to translate real numbers into a sequence of symbols, using a predefined alphabet. In what follows we show that a thoughtful selection of the discretization method may impact the quality of symbolic analysis.

Figure 26 shows an example of symbolic representation of a signal X (red squares) consisting of 100 RR sample points. In this simple case the partition (threshold) is the

median value and the transformation is defined as follows: $y(x) = 1$ if $x \geq \mu(X)$, 0 if $x < \mu(X)$, where $x \in X$, $y \in \{0,1\}$ and μ is median value.

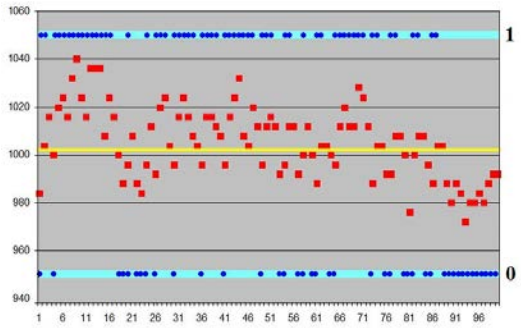


Figure 26. Illustration of symbolic transformation. Red squares denote the original time series, whose values range from 972 to 1040, partitioned with respect to their median equal to 1004 (horizontal yellow bar). Blue circles on light blue strips denote *binary* symbolic representation of the original data using two symbols: '1' and '0'

Concept of the generating partition and its infeasibility for experimental data

The right partition is critical for quality of the symbolic analysis. While intuitively obvious, this requirement is not easy to fulfill. Ideally, each point in the sequence of symbols corresponds to unique point in the state space [2]. This condition is satisfied by a particular kind of partition, called '*generating*', which *preserves all (deterministic) dynamical features of the original data*. The concept of generating partitions simply fails in the case of empirical data because we usually do not know the (differentail) equations which determine evolution of the underlying system. Even for chaotic systems, finding

the generating partition – if feasible at all – is hard, especially for dimensions higher than one. In such circumstances a common practice is to apply the method of *threshold-crossing*, i.e. defining – in a rather arbitrary way – a partition which divides the numeric time series into (two) disjoint sets, each represented by a different symbol. Usually this kind of partition is based on the (global) mean or median of the original data. One obvious reason for the popularity of such an approach is the difficulty of identifying the generating partition. What is more, the threshold crossing is a physically intuitive idea. Although the threshold-crossing method may yield satisfactory results, the price for this (over)simplicity may be high. Costa *et al.* [18] provide a detailed analysis of possible consequences of this approach, arguing that the '*threshold-crossing technique typically yields misleading conclusions about the dynamics generating the data, and therefore one should be extremely cautious when attempting to understand the underlying system from a misrepresented symbolic dynamics*'. Yet, the authors admit that '*There are two reasons for the popularity of the threshold-crossing method: (i) it is extremely difficult to locate the generating partition from chaotic data and (ii) threshold crossing is a physically intuitive and natural idea*'.

Heuristics for the partition construction

In striking opposition to the threshold-crossing approach is, for instance, the recently introduced *base-scale* entropy [75]. The base-scale entropy uses a highly adaptive partitioning, which changes very frequently, i.e. within very short segments of the signal. The method proved to be very robust against noise, nonstationarity and efficient for

physiological signals of short duration. The authors use a '*dynamical adaptive partitioning approach in transforming the time series into a symbol series*' and claim that '*base-scale entropy method was used as a measure to classify physiologic and synthetic heart rate variability series. This method enables analyzing very short, non-stationary, and noisy data*'.

Often, the objective is detecting a qualitative change in temporal data, the absolute values of analyzed statistics might be *less* important than their *relative order* in time. The absolute value (of a variable) may be difficult to interpret out of context. This point of view is expressed in [96] where the authors provide interesting examples of breathing patterns and heart rate variability.

In a sense, the partitioning based on the *base-scale* entropy and (globally fixed) *threshold-crossing* may be perceived as extremes, yet both can contribute to yet another approach, which we call '*alternate partition*'. To construct the alternate partition one needs to identify segments (along the time series) delimited by the dynamical change points – the very same we discussed in two previous chapters in terms of change (point) detection. As we will show further in this chapter, the alternate partition constructed based on (dynamical) change points more preserves the original dynamics of the experimental time series more closely than those based on arbitrarily fixed thresholds.

The Jaynes' *Maximum Entropy* principle

The *Maximum Entropy* principle[14] – denoted *MaxEnt* – is a concept that inspired our method of alternate partition. The *MaxEnt* was introduced mathematically by [53] as a general criterion for inference, recommending that '*of all possible* (probability) *distributions under explicitly known constraints we should choose the one entailing the largest entropy*'.

The *MaxEnt* minimizes the bias due to unjustifiable assumptions about the data by favoring a probability distribution that entails *maximum entropy* (uncertainty). These may be roughly thought of in terms of randomness and/or complexity. In our study we follow the *MaxEnt* paradigm as a *heuristic* for (binary) partition to help minimize the risk of *underfitting* the dynamics in data to be symbolized. More precisely, our objective is two-fold, aimed at an adequate trade off between the two extreme cases: First, when the symbolization *underfits* the original dynamics in data – which is inherent to a fixed partition (possibly due to assumption of data stationarity), and second, when *overfitting* may take place – caused by excessively dense segmentation (perhaps due to spuriously identified or assumed change points). For instance, a uniform fixed partition with the threshold set at the mean (or median) value in general represents the first case. Such an approach can be justified for stationary data (given such information is testable). However, when certain facts (e.g. change points) about the data are known, they should be accounted for and considered as (testable) constraints of the data symbolization. As we show further, partitions comprising more than one threshold – i.e. based on more than

[14] The M*axEnt* was first proposed by Laplace in [*Principle of Insufficient Reason*] and states that '*If there is no reason to prefer one hypothesis over alternatives, simply attribute the same probability to all of them*'

one segment – result in symbol sequences of higher *complexity*, if quantified by entropy-related measures used in next sub-section.

Fixed threshold-crossing vs. alternate partition

In order to quantitatively evaluate the benefit of alternate partition, we selected the following four statistics: (i) normalized Shannon entropy, (ii) normalized Lempel-Ziv complexity, (iii) Lempel-Ziv compression dictionary and (iv) the number of threshold crossings per sequence length (self-descriptive). We consider one of the RR signals from N. Copernicus University (which were the subject of analyses in Chapter 4) composed of 3110 sample points – denoted x^{RR}. The temporal positions of the applied stimuli (music, white noise, postural change, white noise, thermal batch) are 478, 1712, 2102, 2558, 2996 (expressed as the sample numbers). The mean value of the 3110 numbers of x^{RR} is 676 (integer-rounded) and the median is 672. The plot of x^{RR} is shown in figure 27.

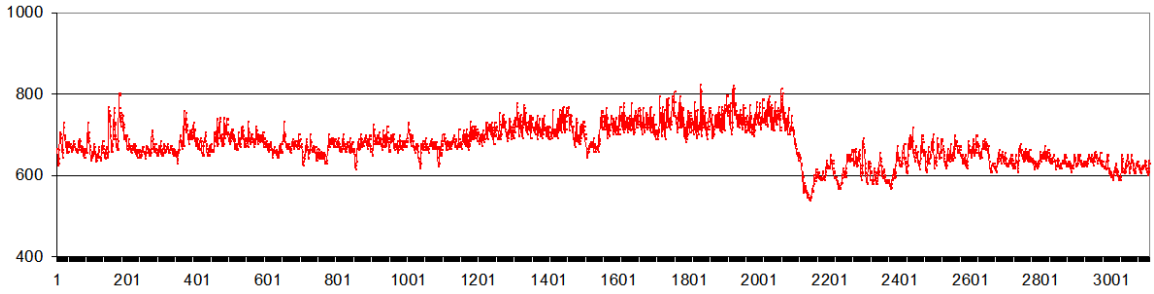


Figure 27. RR signal of 3110 sample points

Applied to the x^{RR} signal, the IRSEG algorithm identified six dynamical change points (CP's) of temporal positions 350, 550, 1500, 1700, 2150 and 2550, which divide the x^{RR}

into seven subsequences: x^{RR}_1=(1,350), x^{RR}_2=(351,550), x^{RR}_3=(551,1500), x^{RR}_4=(1501,1700), x^{RR}_5=(1701,2150), x^{RR}_6=(2151,2550), x^{RR}_7=(2551,3001), each characterized by its respective mean value m^{RR}_i, i=1,...,7. The six CP-based delimiters and the seven mean values $\{m^{RR}_{i=1,...7}\}$ define what we call '*dynamic-adaptive alternate* partition'. In the above settings, symbolization of the whole x^{RR} signal is conducted separately for each x^{RR}_i subsequence by applying its respective (m^{RR}_i-related) threshold crossing condition, perhaps different for each subsequence.

The discretization of x^{RR} signal into sequences of two symbols: '0' and '1' was conducted using two different partitions: (i) the *fixed threshold* crossing around the mean (i.e. 676) constructed on one segment (s=1) and (ii) *alternate* partition constructed on seven segments (s=7) delimited by the six identified dynamical CP's. We did not exceed beyond the latter number of segments because a 'dense' segmentation (where at least certain segments would be short) may cause picking up noise instead of inherent dynamics in the data. More systematical way of showing this could be applying surrogate data test to the original signal and finding at which segment size the symbolization of the original data can not be discerned from symbolization of surrogate data (at predefined significance level).

Figure 28 shows the symbolic sequences obtained using these the described partitions and alphabet size 2 (due to poor resolution the series of color dots may look like a sequences of isolated segments of different lengths). Because the mean- and median-based fixed partitions yield almost identical result and cannot be discerned visually, the latter is not shown.

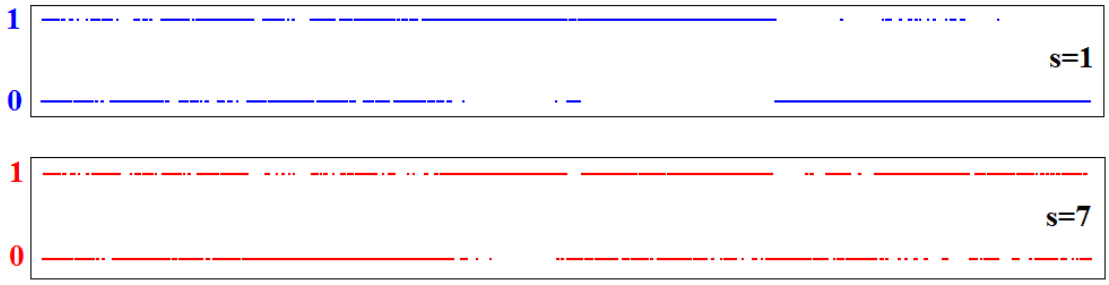


Figure 28 The signal x^{RR} (from figure 27) in symbolized form. Upper plots, blue: fixed threshold crossing around the mean value (number of segments s=1). Middle plots, red: dynamic-adaptive alternate partition constructed based on the (dynamical) change points

Clearly, the higher complexity coincides with more frequent threshold crossings, hence shorter average runs (of each symbol). We can roughly say that – in general - the *alternate* partition (s>1) is *more complex* than its fixed counterpart. Yet, this does not solve the obvious dilemma of what is 'optimal' s or – more generally – how to find *adequate segmentation*. In either case, a '*good*' partition is the one that most effectively reveals the randomness and/or complexity of the original data [14], thus rendering the underlying dynamics more closely. A useful *heuristic* for achieving this goal is maximizing the entropy of the symbolized data. In table 30 we present the three statistics computed on the two symbolic representations (depicted in Figure 28):

Table 30. Statistics characterizing *static* (blue) and *dynamic* (red) aspect of the complexity of the x^{RR} signal symbolized via *fixed* threshold (FT), dynamic-adaptive *alternate* (DA). The rightmost two columns show the ratios: *r1*=statistic(DA)/statistic(FT)

statistic	**fixed part.** (s=1)	**dynamic-adaptive alternate part.** (s=7)	***r***
***LZC**, dictionary size*	238	313	1.32
***LZC**, normalized*	0.92	1.20	1.30
threshold crossings	0.10	0.18	1.80

Clearly, a partition constructed based on a greater number of segments yields a symbolic sequence featuring higher *dynamic* complexities.

Nevertheless, on certain time scale – e.g. within each segment of the signal – the symbolization still depends on fixed threshold crossings. The degree to which a fixed (uniform) partition preserves the original dynamics depends on how the threshold is defined, what illustrates table 6, where the same signal x^{RR} is discretized according to five different thresholds. Out of the five sequences of symbols plotted in figure 29, the threshold corresponding to the mean value: θ=676 generates the most complex symbolic representation, what is confirmed computationally: the results presented in table 6 were obtained analogously as those in table 5 (Note that the cases labeled θ=676 in figure 29 and labeled s=1 in figure 28 are identical).

Figure 29. The signal x^{RR} symbolized via five different fixed partitions with thresholds set at θ=584, 632, 676, 717, 765. The corresponding colors of the plots are: turquoise, blue, black, red and orange

Table 31 Statistics characterizing *static* (Shannon) and *dynamic* (LZC) aspect of the complexity in the RR signal symbolized via *fixed* partition (s=1) set at five different threshold values θ (rounded to two decimals)

statistic	**θ=584**	**θ=632**	**θ=676**	**θ=717**	**θ=765**
***Shannon**, normalized*	0.01	0.06	0.09	0.06	0.01
***LZC**, dictionary size*	98	174	238	218	118
***LZC**, normalized*	0.37	0.65	0.92	0.81	0.44
# of threshold crossings	0.01	0.05	0.10	0.11	0.03

Before introducing the methods of nonstationarity analysis in symbol sequences, we conclude our considerations with two more examples: Figure 30 illustrates the influence of the threshold value on the conditional entropy h_n [121]and the normalized Lempel-Ziv complexity (*LZC*), both obtained on the logistic map generated in chaotic regime. Clearly, the complexity measures assume their maxima when the threshold is set close to the mean value of the data (which is 0.5). The example shows the *LZC* is a consistent evaluator of the complexity as a function of partition (used in symbolization).

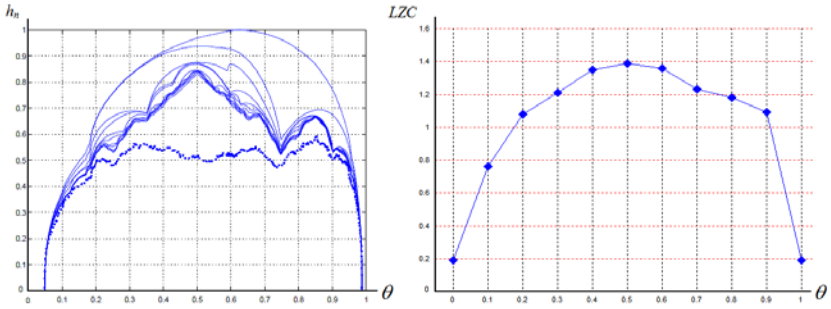


Figure 30. Left panel: conditional entropies h_i, i=1,...,10, for the logistic map as a function of the threshold parameter θ (credit: [121]). Right panel: normalized *LZC* for the same map symbolized via fixed binary partition with thresholds $\theta_i=i/10 - 0.1$, i=1,...,11

The conclusion is that choosing partition which *maximizes entropy* (*complexity*) of the symbolized data given the *testifiable heursitics* (e.g. change points) gives a reasonable basis for symbolization. If no heuristic is available, the approach may rely on applying the *MaxEnt* and controlling the (uniform) segmentation via the surrogate data test.

B. LEMPEL-ZIV (COMPRESSION) COMPLEXITY

The CARACAL – *Compression Dictionary-related Nonstationary Analyses* – is another algorithm we developed for our study. It is an analytical toolbox applicable to binary symbolic sequences. Coupled with IRSEG (as the segmentation-heuristic provider), the CARACAL can be applied as a natural follow up within a broader analytical framework.

Lempel-Ziv compression algorithms

The CARACAL algorithm was developed around the concept of *Lempel-Ziv complexity*, denoted *LZC*, which is implemented as a family of Lempel-Ziv compression algorithms. All of them are adaptive, lossless and based on dictionary which builds up during the data

compressing. Part of the compression is *parsing* a sequence of symbols, denoted *S*. In our research we chose the LZ78 implementation as the computational engine for CARACAL. Before providing algorithmic description of the CARACAL methods, we briefly present the main concept of the LZ78 works using three examples of compressed strings.

The way the dictionary builds up is illustrated in table 33, where the three strings: *S1*, *S2* and *S3* – each comprising 30 symbols (1's and 0's) – are parsed into their respective dictionaries *D1*, *D2* and *D3*. At the bottom row of the table we provide the dictionary sizes, i.e. the numbers of words they consist of (perhaps excluding the uncompressed residuals already parsed). The resulting dictionaries and their sizes are:

D1={0,01,1,10,011,11,110,1101,10,00,000,100}, |D1|=12

D2={1,0,00,01,001,11,110,010,0011,011,10,0110}, |D2|=12

D3={1,0,10,11,00,01,011,100,0110,0111,000}, |D3|=11.

Table 33. String parsing according to the LZ78 compression algorithm. The sequences of symbols in red (in columns D1, D2 and D3) denote consecutively appended words into the dictionaries: they are placed next to the symbol commencing parsed subsequence. The subsequences in green duplicate their counterparts already incorporated to dictionary. Horizontal double lines delimit consecutively parsed subsequences

seq. order	S1	D1	S2	D2	S3	D3
1	0	*0*	1	*1*	1	*1*
2	0	*01*	0	*0*	0	*0*
3	1		0	*00*	1	*10*
4	1	*1*	0		0	
5	1	*10*	0	*01*	1	*11*
6	0		1		1	
7	0	*011*	0	*001*	0	*00*
8	1		0		0	
9	1		1		0	*01*
10	1	*11*	1	*11*	1	
11	1		1		0	*011*
12	1	*110*	1	*110*	1	
13	1		1		1	
14	0		0		1	*100*
15	1	*1101*	0	*010*	0	
16	1		1		0	
17	0		0		0	*0110*
18	1		0	*0011*	1	
19	1	*10*	0		1	
20	0		1		0	
21	0	*00*	1		0	*0111*
22	0		0	*011*	1	
23	0	*000*	1		1	
24	0		1		1	
25	0		1	*10*	0	*000*
26	1	*100*	0		0	
27	0		0	*0110*	0	
28	0		1		0	*011*
29	0	*01*	1		1	
30	1		0		1	
		12		**12**		**11**

From a practical point of view, the algorithm in its original implementation is not practical for strings longer than thousand of symbols because of rapid growth of the dictionary size (i.e. when large subsequences are incorporated). To circumvent this problem, certain optimized implementations of LZ78 have been introduced and successfully applied. Looked at from a wider perspective, the compression algorithms – by measuring entropy[15] – enable estimating 'more sophisticated complexity measures' [14].

The *LZC* is quantified by the number of distinct dictionary words (or patterns) parsed during compression. To mitigate the dependency of *LZC* on length (of the compressed sequence), we define the so called *normalized LZC* : $C_n = N*\log_2 d/d$, where d is the number of the patterns incorporated into the dictionary and N is the sequence length. Although closely related to the concepts of the *source entropy*, in our setting – i.e. with neither knowledge nor assumption of the probability distributions of symbols in the source generating data – the *LZC* characterizes a single sequence (rather than the source), hence its strong affinity to *K*.

C. THE CARACAL ALGORITHM AND THE LZC-TRIPLET METHODS

In what follows we introduce the three methods implemented into the CARACAL computational framework. Analogously to IRSEG, the CARACAL algorithm computes

[15] The higher degree the conditions of ergodicity and infinite sequence length are violated, the less precise is the above statement

three statistics[16]: *Cratio*, *Ccross* and *LDF* – defined further – which are applied to detecting dynamical changes (nonstationarity) in symbol sequences. They do so by analyzing the local variability of the *normalized LZC complexity* related to compression dictionary used in the sequence parsing. Prior to providing their definitions we introduce two concepts related to the compression dictionary.

Generating *native* and *foreign* compression dictionaries

We consider a sequence S comprising two disjoint subsequences S^L and S^R, each containing W symbols: $S = S^L \bigcup S^R$ and $S^L \bigcap S^R = \varnothing$. To each of the subsequences we apply the LZ78 algorithm, and generate what we call a '*precompression dictionary*', denoted D^L and D^R, respectively. Next, we repeat the LZ78 compression but this time – rather than building a dictionary from scratch (as in 'regular' LZ78 procedure) – we apply the nonempty D^L and D^R from the very start of compression. For terseness, let this instance of compression be called '*recompression*'. Thus, with the two subsequences: S^L and S^R, and the related two dictionaries: D^L and D^R, we can conduct *four-way recompression*, each involving different combination of S and D. In particular, we can explore two possible scenarios: (i) using one precompression dictionary to parse two different sequences – a *one-to-many* case, or (ii) using two different precompression dictionaries to parse one sequence – a *many-to-one* case.

We introduce the following formalism. Let $C_n^{i_S} = LZC(S^{i_S})$, where $i_s = \{L, R\}$, denote (normalized) LZC of the S^{i_S} subsequence (i.e. half) of S. Another way of expressing $C_n^{i_S}$

[16] In a sesnse, they can be though of as a *LZC*-triplet – analogously to the term of *SampEn*-triplet as a core of the IRSEG algorithm

is by referring it to the *size* of dictionary D^{i_D} generated during compression: $C_n^{i_S} = c\,|\,D^{i_D}\,|$, where $i_D = i_S$, $c = \lg_2(W)/W$ is a normalizing factor and $|\bullet|$ denotes the size (of the dictionary) as the number of elements (i.e. words incorporated into the dictionary).

Furthermore, let $C_n^{i_S,i_D} = LZC(S^{i_S}\,|\,D^{i_D})$ denote normalized complexity resulting from *recompression* of a sequence S^{i_S} using the *precompression* dictionary D^{i_D}, where i_S, $i_D = \{L, R\}$ correspond to the *left* and *right* halves of *S*. We define a 2x2 complexity matrix Δ^C, representing all the $C_n^{i_S,i_D}$ combinations: $\Delta^C = \begin{bmatrix} C_n^{LL} & C_n^{RL} \\ C_n^{LR} & C_n^{RR} \end{bmatrix}$. For $\Delta^C_{i_S,i_D}$ where $i_S = i_D$ (i.e. elements on the diagonal of Δ^C), the dictionaries D^{i_D} are considered *native* with respect to the data S^{i_S}: These are the cases when $C_n^{i_S=i_D} = LZC(S^{i_S}\,|\,D^{i_D})$ are obtained in *recompressing* S^{i_S} with the *precompression* D^{i_D} built on the same data S^{i_S}. Conversely, for $\Delta^C_{i_S,i_D}$ where $i_S \neq i_D$ the $C_n^{i_S=i_D} = LZC(S^{i_S}\,|\,D^{i_D})$ is determined based on D^{i_D} built from parsing data other than S^{i_S}, hence such dictionaries we call *foreign* (with respect to S^{i_S}). Briefly, dictionary is *native* (with respect to the data it is applied to) when both compression instances are conducted involving the same data, otherwise we deal with a *foreign* dictionary.

Our conjecture is that a dictionary – if applied to its corresponding *native* data – is 'intrinsically more efficient' during *recompression* when compared to its *foreign* counterpart. By 'more efficient' we mean that a (precompression) dictionary – due to rich *diversity* of its content (i.e. the set of patterns) – is capable of parsing the analyzed

sequence with *fewer steps*, because fewer (new) patterns need to be added to perform the task. This is not the case of a dictionary whose content is of poor diversity, entailing limited capability of pattern matching. While the conjecture is not a strict, general rule, we show its applicability in change detection in symbolized sequences.

Similarity measures based on the Lempel-Ziv compression dictionary

We begin with definition of the *Compression Dictionary Similarity* (CoDS). Let $f_d = \max\left\{\frac{C_n^L}{C_n^R}, \frac{C_n^R}{C_n^L}\right\}$, $f_i = \max\left\{\frac{C_n^{RL}}{C_n^{LR}}, \frac{C_n^{LR}}{C_n^{RL}}\right\} = \max\left\{\frac{\Delta_{1,2}^C}{\Delta_{2,1}^C}, \frac{\Delta_{2,1}^C}{\Delta_{1,2}^C}\right\}$ denote two factors contributing to CoDS, which we define $CoDS = \sqrt{f_d^2 + f_i^2}$. The former factor, f_d, is the greater of two reciprocal ratios of the complexities of the *L* and *R* parts of *S*. Because the complexities are computed *directly* based on (single) compression (i.e. without recompression), we index the factor f_d by subscript '*d*'. The index '*i*' of the f_i indicates its *indirect* computation via recompression *foreign* dictionaries. (It is computed analogously as its counterpart but involves the mixed terms of Δ^C). The motivation and explanation of the CoDS is following. (i) We compute the both factors using the maximum of ratios of complexities characterizing adjacent data segments: what counts is to what *degree the complexities differ*, rather than which of them is greater (smaller). Using ratio is less prone to (local) fluctuations than differences, which may blow up in the regions of high complexities thus affecting the variance of the series of segmentwise

complexities. (ii) Roughly speaking, one might consider the two factors[17] f_i and f_d as related to the *quantitative* and *qualitative* aspects of complexity, respectively. The *indirect* factor, f_i, compares the complexities of the two subsequences (L and R) without comparative procedures of recompression. On the other hand, the direct factor f_d relies on mutual relations between the dynamical patterns of the L and R, hence we consider it qualitative: if those patterns are '*similar*', the f_d contributes less to the total CoDS.

When the adjacent subsequences (S^L and S^R) feature *different dynamical patterns* and the subsequence S^L is significantly *more complex* than S^R: $C_n^L \gg C_n^R$, this complexity translates into 'higher efficiency'[18] of D^L as a precompression dictionary, used during recompression of the relatively simple subsequence S^R. Conversely, when D^L is significantly *less complex* than D^R: $C_n^L \ll C_n^R$, this indicates the former is relatively inefficient performer in recompressing S^R, hence the task requires more steps. Although the f_i and f_d in principle are *not independent*, we consider them as two components of the CoDS, hence the formula is analogous to that for vector modulus.

[17] Of course, pointing to those 'two factors' is rather arbitrary point of view. On the other hand, while presence of the 'quantitative' factor is obvious, one should not neglect that there may be two fairly *different patterns of dynamics* in two sequences of symbols, both quantified at the same level

[18] The term 'high efficiency' means that a (precompression) dictionary – due to *rich diversity* of its content (i.e. set of words) – is capable of parsing with *fewer additions* of new words, hence performing the task in fewer steps. This is not the case of a dictionary whose content features *little diversity*, entailing limited capability of matching the patterns

When $CoDS \cong 1$, the dynamics of the compared subsequences (S^L and D^R) are considered *similar*. To identify change points we seek the values of CoDS significantly[19] different from the average $\mu^{CoDS} = \overline{CoDS}$, computed over all the window positions along the whole sequence of symbols. Likewise for the *SampEn* (IRSEG), local maxima – where $CoDS \gg \mu^{CoDS}$ – indicate *change point* candidates.

Definition of Cratio

The first statistics used by CARACAL can be defined using the notation introduced earlier in this chapter, i.e. $C^L = LZC(S^L)$ and $C^R = LZC(S^R)$: *Cratio*=min(C^L/C^R,C^R/C^L), which is simply the normalized ratio of LZC of two adjacent segments. When the value of *Cratio* is close to 1, the complexities of the compared segments are similar. Intuition behind the *Cratio* is that it plays a role of a 'weak' criterion for a change point, or put in other words – a 'preliminary flag' that we may deal with nonstationarity. Nevertheless, by no means it can be considered as a sufficient condition, because dynamically similar sequences may come up as having different values of *Cratio* due to the fact that the compression algorithm may have not gotten close to its (asymptotic) efficiency.

Definition of *Ccross*

Based on the formal notation, $C^{LR} = LZC(S^L \mid D^R)$, $C^{LL} = LZC(S^L \mid D^L)$ and $C^{RL} = LZC(S^R \mid D^L)$. We define $Ccross = C^{LL}C^{RR} - C^{LR}C^{RL}$, and then redefine it to

[19] Because neither the number of segmentwise values (of CoDS or CCTeS) are usually less than 30 and their normality is questionable, in such cases we do not follow the strict criterion based on the statistical significance (e.g. $\alpha \leq 0.05$). In most cases we accept as significant values which differ from μ^{CoDS} by more than one or two standard deviations

obtained only non-positive values: $Ccross = \min(0, Ccross) / (C^{LR} C^{RL})$, where the denominator plays a role of moderating factor. The importance of this statistic is comparable to that of the *rRnd* computed by IRSEG: *Ccross* is (highly) sensitive to dynamical differences between the compared sequences of symbols. For this it can be thought of as a 'strong' indicator of change point.

Defnition of *LDF*: Length Distribution Fraction

This statistics characterizes distributions of the binary symbols in compression dictionary, used in comparing two sequences. Rather than a single number (i.e. scalar value), the *LDF* is a vector as shown in Table 34.

Table 30. The tabularized representation of the results from Figure 26. The quantity ρ represents saturation (explanation in text)

level	actual	max	*LDF*
1	2	2	1
2	4	4	1
3	1	8	0.125
4	1	16	0.0625
5	1	32	0.0313
6	1	64	0.015

The values of *LDF* (in rightmost column of Table 9) denote the fraction of all possible realizations related each level (of length). To illustrate the concept, we use real data of the RR signal from previous example (see the plot D in Figure 31). The resulting symbol sequence is analyzed by CARACAL segment by segment, hence the output is

segmentwise. In figure 31 we show the plot of saturation ρ computed for of all the lengths of parsed words (obtained for one segment of 100 symbols of the RR signal). The values of ρ range from 1 to 0.001.

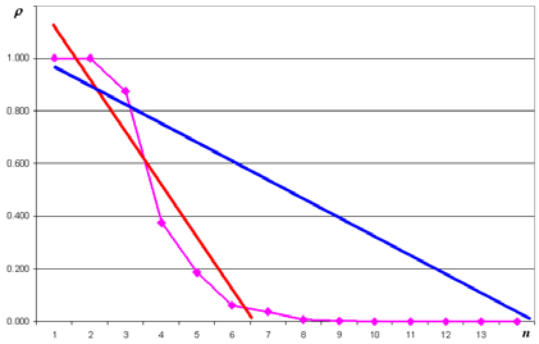


Figure 31. Plot of word length fractions: the axes *l* and *LDF* denote length and fraction, respectively. The blue line represent slope of best fit for all 14 points accounted for. The red line represents the largest (in terms of absolute values) slope obtained when considering the first *n* points only, and varying *n* from 1 to 14

In practice, we compare the *LDF*'s of the adjacent segments. The final quantity we obtain as a measure of *LDF*'s related intersegmental dissimilarity can be defined as a factorial-wise product of the normalized ratios of *LDF*'s computed for the compared segments. The final value of the statistic is raised to the power $P=1/(D-k+1)$, where D is the dimension of the *LDF* vector (see Table 9 and description therein) and k is a user-defined parameter playing a role of a cut-off point (to exclude the factors below a predefined threshold beyond which the subsequent values can be neglected).

D. RESULTS AND CONCLUSIONS

We tested our methods by applying the IRSEG algorithm to the following time series: H100_33_50 (three sequences of Henon chaotic map with three sampling rates), Hsrg_H_Hsrg (three sequences of Henon map, where samples in the middle sequence is randomly shuffled), L7_L8 (concatenation of two logistic maps in chaotic and periodic regime), L7noise_L8noise (as for L7_L8 but with noise added), sinusoid3 (a concatenation of three combinations of sinusoids simulated with different periods), and sinusoid3noise (as for sinusoid3 but with noise added).

The results obtained from these data are shown in Table 35. The first six signals (highlighted blue) are the signals already described in Chapter 3, and the next two examples, d90 and d111, are the cardiac RR signals from the N. Copernicus University described in Chapter 4. The signals whose labels are marked blue were already used in for evluating IRSEG algorithm (which was used in this comparison).

Table 35. Results of symbolic analysis of dynamical change (simulated-symbolized signals)

data	MaxEnt-fixed		IRSEG-based		SAX-based	
	recall	precision	recall	precision	recall	precision
H100_33_50	1	0.67	1	0.67	1	1
Hsrg_H_Hsrg	1	0.67	1	1	1	0.67
L7_L8	1	0.67	1	0.67	1	1
L7noise_L8noise	1	0.5	1	0.67	1	0.5
sinusoid3	1	1	1	1	1	1
sinusoid3noise	1	0.67	1	1	1	1
d90 RR	1	0.63	1	0.71	1	0.83
d111 RR	1	0.56	1	0.71	1	0.63

A word of explanation must be given regarding the above results. First, one should notice that both recall and precision were obtained using three different (numeric) data symbolization: (a) *MaxEnt*-based fixed partition (the *MaxEnt* was introduced in section A of this chapter), (b) based on the change point-related heuristics (computed by IRSEG) and (c) identified by the visual judgment based on the output from the SAX algorithm. Then, the visual judgment was conducted based on SAX run multiple times, with the (parameter of) window size varied as {50,100,*N*}, where *N* is the length of all analyzed time series. We kept the two other parameters fixed: alphabet size=2 and compression rate=1. Although these particular parameter values are rather 'artificial' for the SAX whose objective is to compress the data while maintaining the original morphology as close as possible, these have to match the corresponding parameter values of CARACAL, which does not compress the signal length and operates on alphabet comprising two symbols only.

Nevertheless, we conducted another series of SAX symbolization, varying alphabet size: {2,3,5,10}, window size: {50,100,N}, and compression rate: {1,5,10,100}. The selected examples of results obtained from SAX-based symbolization are included in the Appendix because they deserve additional attention and comments.

Intrinsically symbolic signals: Bernoulli distribution

Our last evaluation of the CARACAL and different symbolization methods are based on the concatenation of five symbolic sequences (each comprising 500 sample points) simulated using different probabilities: p={0.3, 0.4, 0.5, 0.6, 0.7}. In this case we can

express our results in terms of recall and precision assuming each change point coincides with the change of parameter *p* ('theoretically' four change points):

Table 37. Recalls (R) and precisions (P) obtained in analogous way as the results in table 35

Bernoulli data	**MaxEnt**		**IRSEG**		**SAX:** W=50		**SAX:** W=100		**SAX:** W=N	
	R	**P**	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**
	0.75	0.75	0.75	0.75	N/A	N/A	0.75	0.75	0.75	0.75

6. SUMMARY

This research was aimed at three problems: (i) HRV-based cardiac diagnostics, (ii) nonstationarity detection in time series, (iii) symbolization and symbolic analysis of experimental signals

A. The HRV analysis with application of music and physical stimuli

This dissertation was driven by a specific problem in cardiac diagnostics: namely, the applicability of music – along with other environmental stimuli – as a stimulus to gain more information about a patient's condition through HRV analysis. This question, largely driven by our MD collaborators, has a practical context that was outlined in Chapter 1 (p. 14). Technically, we conducted a series of analyses to test whether (i) the influence of music as a subtle, audible stimulus can be detected at all using available HRV measures and (ii) if so, how useful it can be in discerning the healthy subjects from those suffering certain diseases that affect the autonomic / cardiac system – if compared with other (physical) stimuli. Having applied more than 30 mainstream measures – specifically those computed by the Kubios HRV analytical toolkit – we failed to reject null hypothesis that music has no impact on the HRV. Nevertheless, these HRV statistics also failed in discriminating the healthy subjects from the patients, indicating shortcomings in the existing state-of-the-art. Simultaneously, the physical stimuli applied during ECG recording sessions, namely postural change and thermal bath, affected HRV

in statistically significant manner, when tested using certain statistics from those available in Kubios. This finding was a motivation to seek alternative methods which are sufficiently sensitive to detect subtle effects caused by music – as we described in Chapter 1. Having stated a more general problem of detecting (dynamical) nonstationarity in time series, we devised methods based on the notion of *Sample Entropy* which proved sufficiently sensitive to capture the effect of music while maintaining relatively low rate of false alarms. These methods have been introduced as the *SampEn*-triplet whose components are three entropic statistics: the *static* entropy (*sE*), *dynamic* entropy (*rRnd*) and *delayed* entropy (*rDiv*), defined on pp.

Having proof that the features in the *SampEn*-triplet – denoted *seT* – are sensitive enough to detect the presence of our stimuli, as well as to distinguish between healthy participants and patients, we used them as the basis for the IRSEG algorithm, whose aim is to perform change detection. This is the second step in stimulus response analysis: Not only detecting that something has changed in the autonomic system response, but being able to detect when the change took place. While possibly trivial in the context of the featured *Copernicus* experiment, there are two applications of significant interest in this context: (a) the measurement of *delay time* in such a response, and (b) the detection of such change points *in the absence of ground truth*, such as the detection of stressors in post-traumatic stress disorder patients, or the detection of deception in security applications (both of which are currently under investigation).

As a proof of concept, at first we applied IRSEG to analytically defined synthetic time series, described on p. 108 in Table 28, where we enforced nonstationarity by changing a

dynamical parameter while simulating the data. Having generated nonstationary data with full knowledge where a (dynamical) change takes place, we were able to evaluate our methods in terms of recall and precision. The results we obtained on the simulated data using our methods we compared with those obtained from other mainstream methods of change detection.

B. Detection of change point in (nonstationary experimental) time series

Secondly, we applied our methods to two sets of RR signals: (i) recorded during yoga and chi meditation, from PhysioBank and (ii) recorded during application of controlled environmental stimuli – as described in Chapter 4 – from the Copernicus University. Whereas capturing change caused by meditation onset was not very difficult for IRSEG and the two other algorithms, the externally-stimulated signals (Copernicus data) turned to be more challenging. The obtained results show what follows:

i. the influence of music (on the HRV) was capture in 78% of all analyzed cases (for all the participants),
ii. the influence of postural change and thermal batch were capture at the recall 83% and 88%, respectively.

These results surpassed those obtained by two other algorithms based on CUSUM and SSA. More importantly, however, in different feature subset selection and feature ranking methos, the rRnd measure ranked overall as the best discriminator between the healthy and patients (see Table 28 on p. 114), where the each of SampEn-triplet components was tested along the selected mainstream HRV characteristics computed by the Kubios HRV

toolkit. Likewise, rRnd performed better than expected in series of t-tests (see tables on p. 110 and ff.).

Whereas the accuracy level achieved using our methods and approach does not enable using music as the only stimulus for diagnostics, our study is a step towards narrowing the gap between cardiac diagnostics' utility and availability. What is more, our research conclusively addressed the quest by the cardiologists, hence justifies proceeding in this area of research. A natural follow-up will be repeating this research with more rigorously selected participant and better-defined and/or controlled stimuli, perhaps also more diversified ones.

We believe that the methods we developed and used in this study can be adapted to more sophisticated investigations like stress detection.

Secondly, we broadened applicability of our methods to more general problem of change detection in nonstationary time series. For this purpose we generated chaotic, periodic and random signals with enforced nonstationarity and applied our algorithm to evaluate it in terms of recall and precision (of the change point detection). The results are presented in chapters 3 and 4, showing good overall accuracy, especially when compared with the three other methods (nonlinear cross-prediction error, SSA and CUSUM).

Testing our methods on the well-defined signals allows to better understand the role of each of the seT components. For instance, when the compared (sub)sequences feature similar dynamical patterns (see plot 16) yet different probability distribution, the *static*

entropy quantifier *sE* is more sensitive for the change. Conversely – two compared (sub)sequences of similar distribution (e.g. the Henon signal and its surrogate data, see plot 23) can not be efficiently discerned by the means of *sE* (and likewise the Shannon's entropy *H*) becasue what mekes the essential difference between these two signals is intrinsic regularity (or complexity), which can be captured by the dynamics-oriented statistics, i.e. *rRnd* and *rDiv*. In general, however, we consider the *rRnd* measure as 'primary' becasue our understanding of nonstationarity is considered as properties of nearest neighbors in phase space. Finally, the delayed entropy *rDiv* complements the triplet in case when the signal is undersampled (see plot 18) or when a high (linear) autocorrelation may spuriously attributed to linear rather than nonlinear-(dynamical) effects. This is illustrated in the case of the delayed Mackey-Glass equation (see plot 41).

Implementing the *SampEn*-triplet statistics into IRSEG algorithm encountered the problem of picking up (too much of) noise (or artefacts, omnipresent in physiological time series) due to (exclusively) local analysis. This locality was entailed by segmentwise processing, where two 'small' adjacent segments are to be compared. We solved this problem by enforcing two independent processing, called L2G (local-to-global or bottom-up) and G2L (global-to-local or top-down), as described and illustrated in Chapter 3. These double analysis significantly improved quality of our computations (if compared with single analysis conducted using earlier version of the algorithm). Moreover, the two-way analysis (i.e. top-down and bottom-up) can be re-used if instead of *SampEn*-triplet we plug in other methods.

The (dynamical) change points identified using our *SampEn*-based methods can be used in construction of what we call dynamic-adaptive alternate partition. Such a partition addresses a problem of 'optimal' symbolization of numeric time series into sequence of two symbols (this is a special case of more general problem in n-dimensional phase space). As we claim, the alternate partition renders the underlying dynamics more closely than an arbitrarily decided – usually as the mean of median value – fixed threshold crossing.

C. Symbolization and Symbolic transformation and analysis

The purpose of using the CARACAL is four-fold:

i. when the quality of the original (i.e. experimental) numerical signal is poor due to noise – in such a scenario symbolization may be more beneficial than harmful, filtering out the noise (at the expense of irreversible reducing the 'true' information),
ii. when analyzing the numerical data has too high computational demands (and likewise that regarding data storage),
iii. when '*such representations would potentiality allow researchers to avail of the wealth of data structures and algorithms from the text processing and bioinformatics communities*' [77],
iv. finally: when only symbolic data are available for analysis, for instance when we observe a process governed by Bernoulli distribution.

The overall results obtained in this comprehensive research is a good starting basis for further devoloping the presented methods, within and beyond cardiological diagnostics.

Possible areas of application could include those where change point detection is a vital part of analysis, for instance intrusion and anomaly detection [64], [107] or traffic analysis [129].

Summarizing, our work has laid the foundation for the use of (a) novel chaos-based stastistics and (b) novel change point detection algorithms in heart rate variability analysis. We have shown that the new techniques are robust enough to detect the effect of subtle stimuli, such as music, on heart rate variability characteristics, even in the presence of strong noise due to poor recording conditions. Such techniques can find application in a variety of domains, both in the medical and in the security field, where the identification of change points can be a starting point for the detection of stressors; however, proving their applicability for such purposes warrants further, ongoing and future work.

APPENDIX

Case studies - evaluation

Case 1: The *logistic* map. The *logistic* maps are identified in table 17 by seq. ID from 1 to 9. The first seven signals were simulated with fixed value of the parameter r. We pursue our analysis according to consecutive seq. ID's:

1. r=3.828427123, $x(1)$=0.7

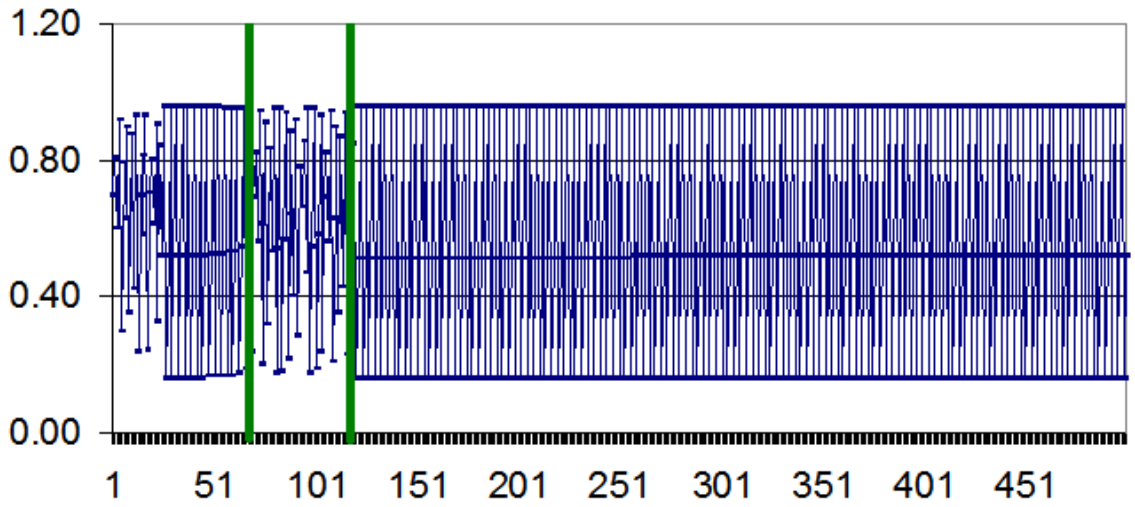


Figure A1. Plot of the *logistic* map: seq. ID=1. The vertical green bars are located at 60 and 110

The IRSEG identified the following (putative) change points 60, 110

2. r=3.828427127, $x(1)$=0.3

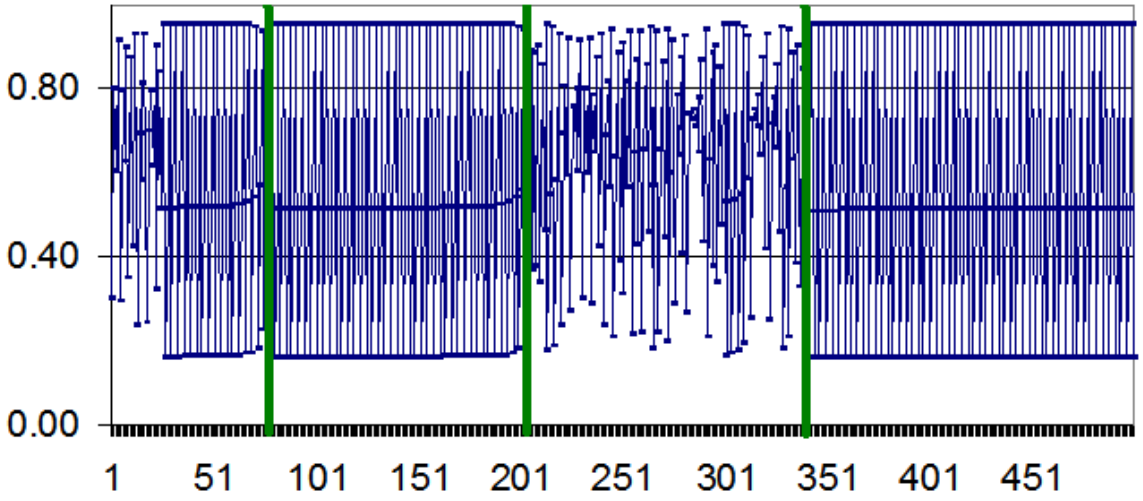


Figure A2. Plot of the *logistic* map: seq. ID=2. The vertical green bars are located at 70, 200 and 340

The IRSEG identified the following (putative) change points 150

3. r=3.828427127

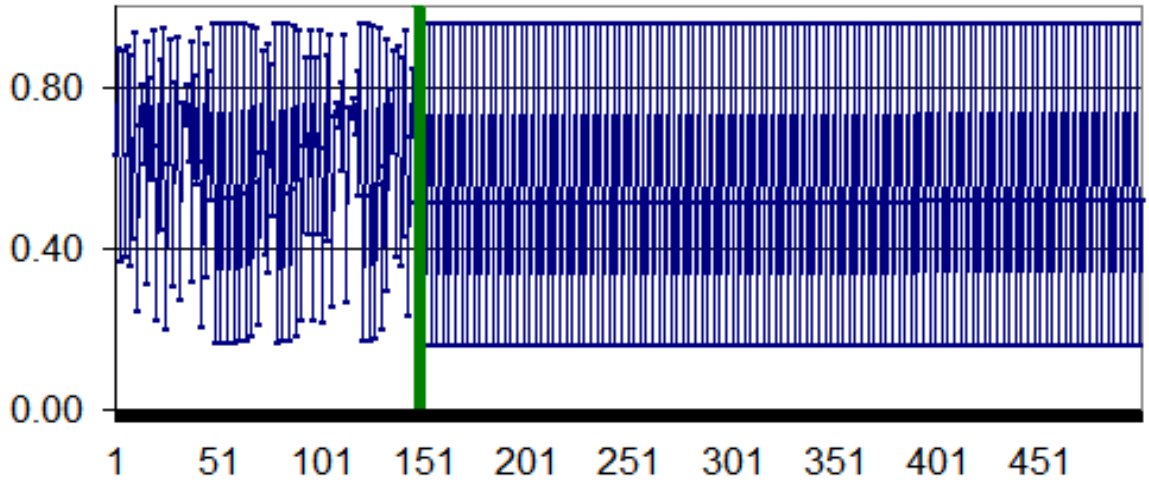


Figure A3. Plot of the *logistic* map: seq. ID=3. The vertical green bars are located at 151

The IRSEG identified the following (putative) change points 151

4. *r*=3.828427127

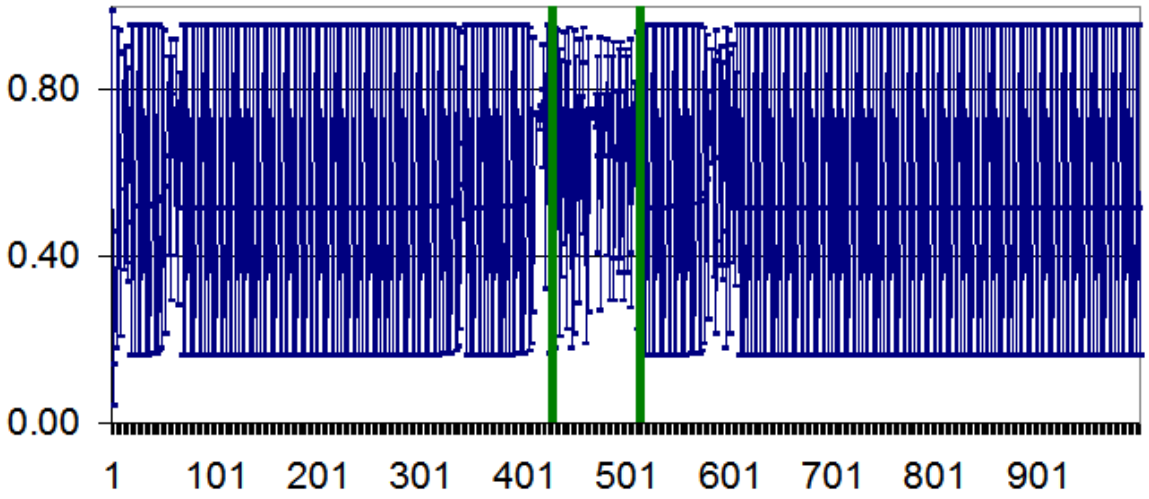


Figure A4. Plot of the *logistic* map: seq. ID=4. The vertical green bars are located at 410

The IRSEG identified the following (putative) change points 410

5. *r*=3.8284271225

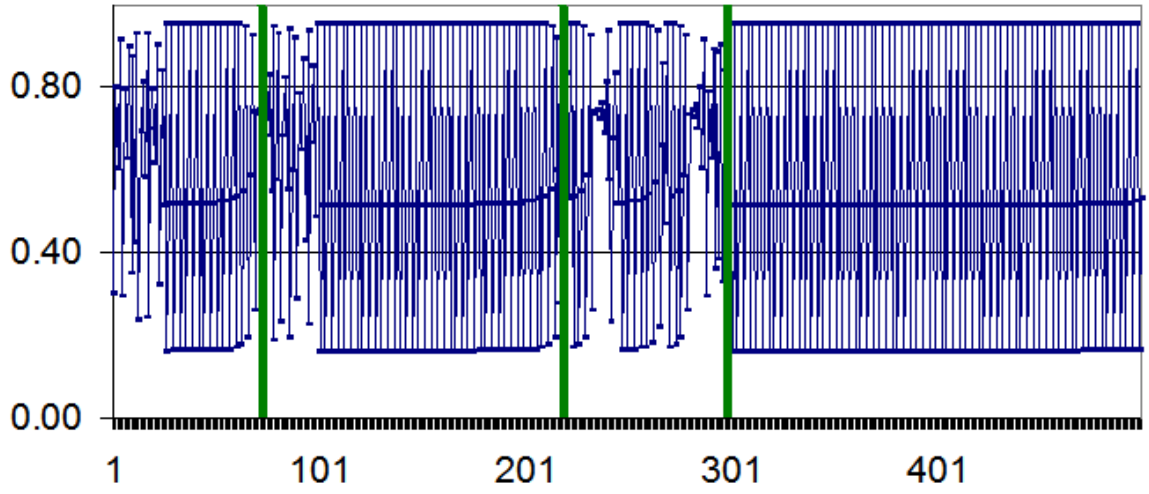


Figure A5. Plot of the *logistic* map: seq. ID=5. The vertical green bars are located at 100, 210, 300

The IRSEG identified the following (putative) change points 100, 200, 300

Case 2: The *Henon* map. The *Henon* maps are identified in table 18 by seq. ID from 10 to 18. The first signal was simulated with fixed value of the parameter *r*, all others involve more than one value of *r*. Analogously as in the case of *logistic* map, we proceed according to sequence of the seq. ID's in table 18:

6. a=1.42207

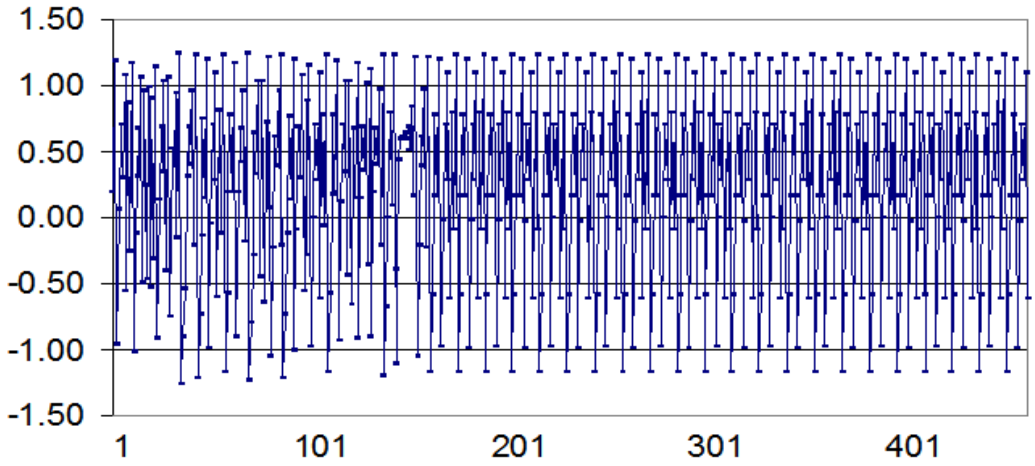


Figure A6. Plot of the *Henon* map: seq. ID=10. The vertical green bars are located at 125

The IRSEG identified the following (putative) change points 200

7. $S1$, $L1$, $H1$

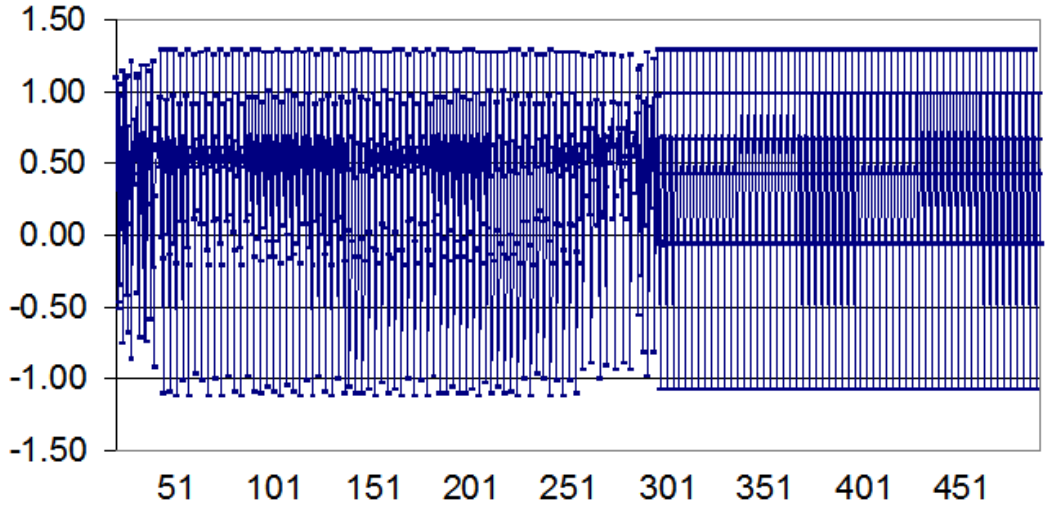


Figure A7. Plot of the *Henon* map: seq. ID=11. The vertical green bars are located at 250

The IRSEG identified the following (putative) change points 300

8. $S^{avg1\&4}$, $S^{avg2\&3}$, $S^{avg1\&4}$

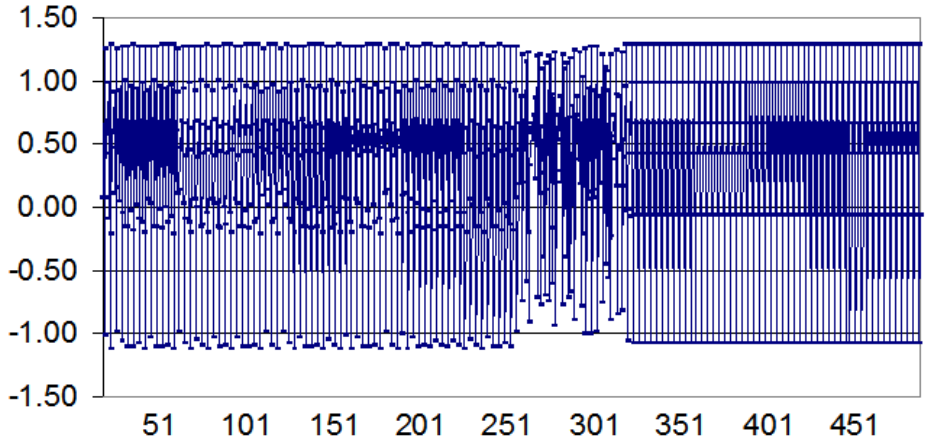
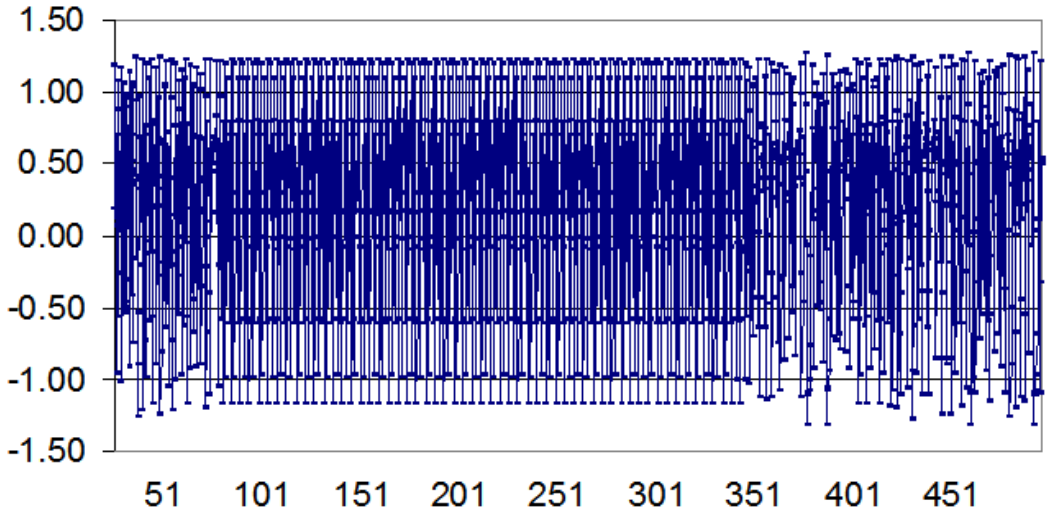


Figure A8. Plot of the *Henon* map: seq. ID=12. The vertical green bars are located at 250, 300

The IRSEG identified the following (putative) change points 300

9. $H6$, H^{surog}, U, N^{trunc}



Figure

fIGURE A9. Plot of the *Henon* map: seq. ID=13. The vertical green bars are located at 70, 370

Case 3: The *Mackey-Glass* equation. The (only) simulation of the *Mackey-Glass* equation is described numerically in table 19 (where it has label '19') and graphically in plot in figure 36. The analogous plot but with delimiters reflecting the change point-based segmentation is chown in figure 52.

10. *Mackey-Glass.*

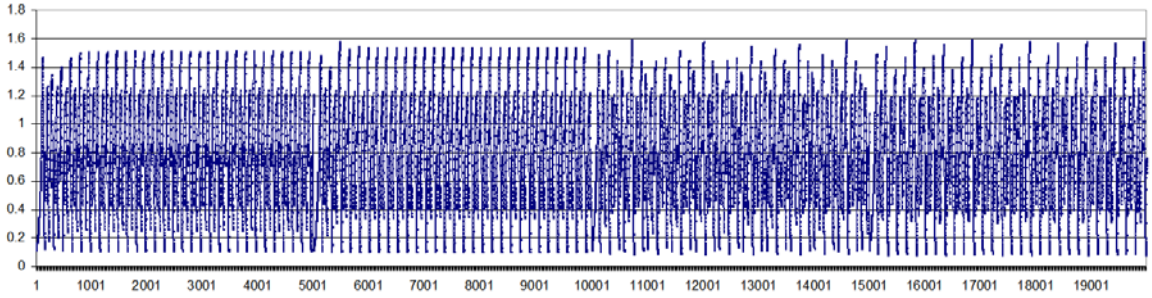


Figure A10. Plot of the four merged MG signals: seq. ID=19. The vertical green bars are located at 1000, 2000, 3000, 4000

The IRSEG identified the following (putative) change points 2000

Case 4: Chaotic and random signals with similar linear properties. To complete evaluation of the IRSEGmethods on simulated data we apply our algorithms to the four signals described in tables 20 (simulation) and 22 (statistical moments).

LULHUHL

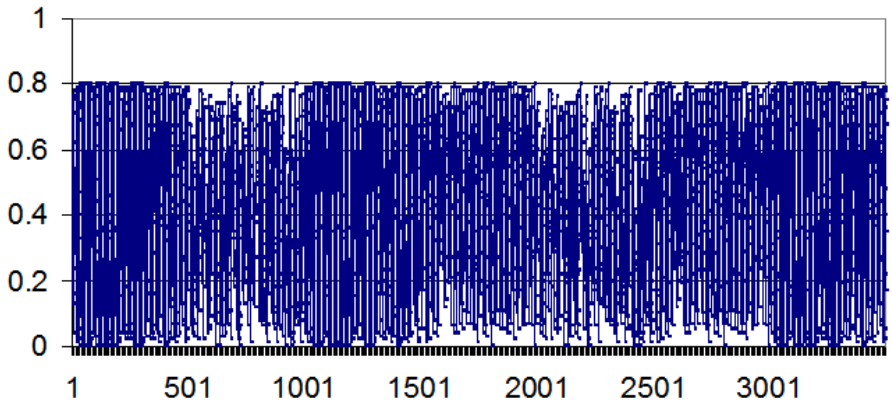


Figure A11. Plot of the all permutations of L, U and H signals, merged into one compound signal: seq. ID=20.

The IRSEG identified the following (putative) change points: 500, 1000, 1500, 2000, 2500, 3000

27. undersampled *Henon* map

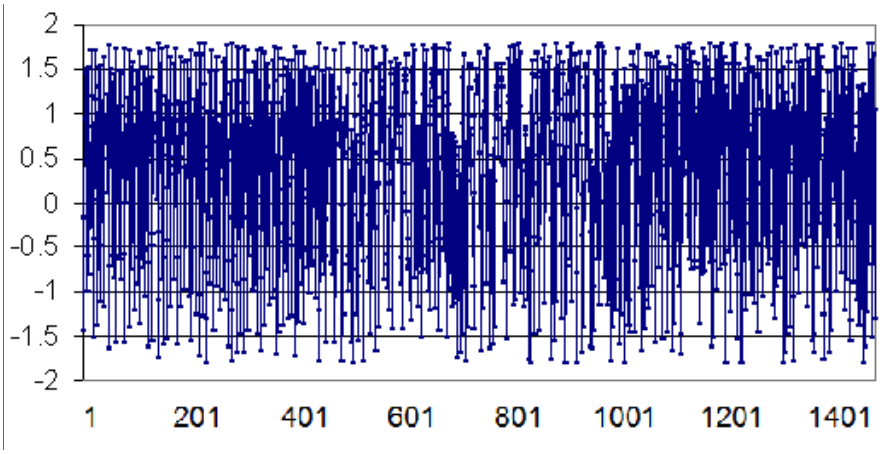


Figure A12. Plot of the undersampled *Henon* map (explanation in text): seq. ID=21 The vertical green bars are located at: none

The IRSEG identified the following (putative) change points 700.

28. X^{sum}

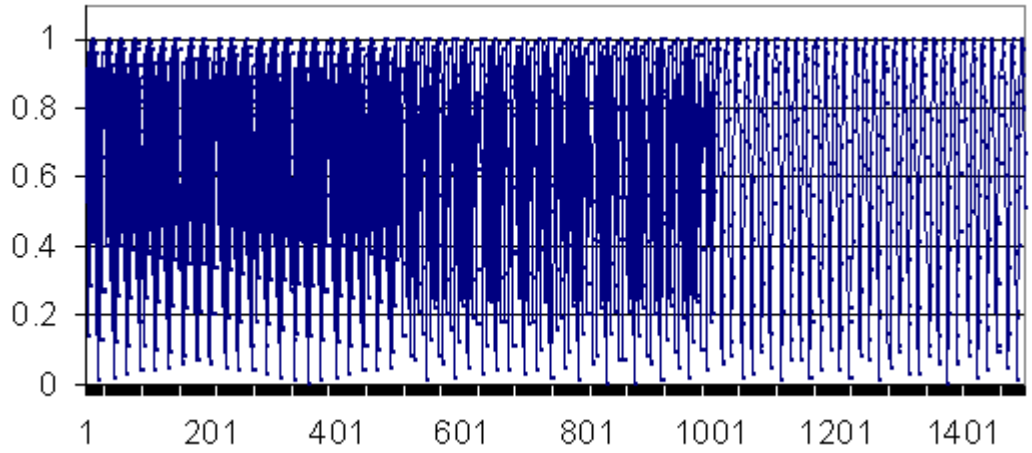


Figure 60. Plot of the three sinusoids simulated with different frequencies: seq. ID=23. The vertical green bars are located at 500, 1000

The IRSEG identified the following (putative) change points 500, 1000.

29. X^{mix}

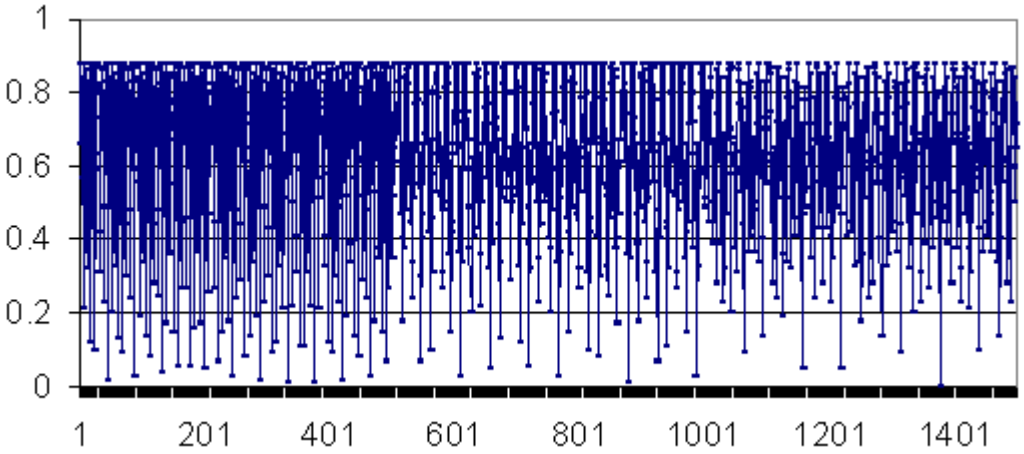


Figure 61. Plot of superimpositions of the three sinusoids plotted in figure 67, seq. ID=24. The vertical green bars are located at 500, 1000.

The IRSEGidentified the following (putative) change points 500, 1000.

30. X^{prod}

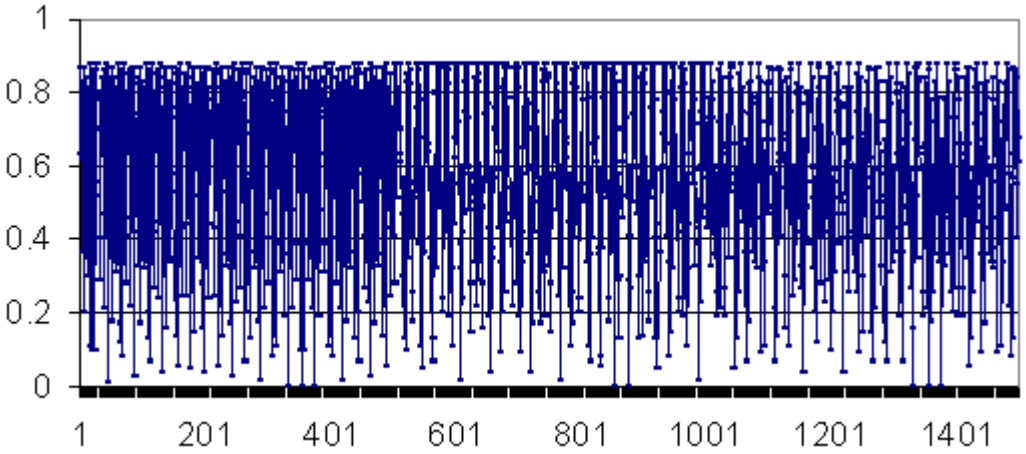


Figure 62. Plot of superimpositions of the three sinusoids plotted in figure 67, seq. ID=25. The vertical green bars are located at 500, 1000.

The IRSEG identified the following (putative) change points 500, 1000.

REFERENCES

[1] (ed.), Mullin. The Nature of Chaos, n.d.

[2] Abarbanel. Analysis of Observed Chaotic Data. Edited by Springer; 1st, 1996.

[3] Analysis of Observed Chaotic Data. First., n.d.

[4] Acharya, and others. "Classification of Cardiac Abnormalities Using Heart Rate Signals." Biol. Eng. Comput., 2004 42 (2004).

[5] "Comprehensive Analysis of Cardiac Health Using Heart Rate Signals." Physiol. Meas 25 (2004).

[6] Akimoto, and others. "Enhanced Open-Loop but Not Closed-Loop Cardiac Baroreflex Sensitivity during Orthostatic Stress in Humans." In: AJP - Regu Physiol November 1 301, no. 5 (2011).

[7] Alcaraz, and others. Study of Sample Entropy Ideal Computational Parameters in the Estimation of Atrial Fibrillation Organization from the ECG, in: Computing in Cardiology 2010;37, n.d.

[8] Alippi. An Adaptive CUSUM-Based Test for Signal Change Detection, in: International Symposium on Circuits and Systems (ISCAS 2006), 21-24 May 2006, Island of Kos, n.d.

[9] Armstrong, and others. "UNDERTOW: Multi-Level Segmentation of Real-Valued Time Series." AAAI 2007 (n.d.): 1842–43.

[10] Asyali. Discrimination Power of Long-Term Heart Rate Variability Measures, n.d.

[11] Ballora, and others. "Heart Rate Sonification: A New Approach to Medical Diagnosis, in: Leonardo." V 37 (February 2004): 41–46.

[12] Benedetto, and others. "Zipping Out Relevant Information, Computing in Science & Engineering" 5 (2003).

[13] Bidlack. "Music from Chaos." In: A J. of Compositional and Theoret. Research in Music, 1992. Vol. 6, No 1 6 (1992): 70–114.

[14] Bolt. “Targeting Control of Chaotic Systems,” n.d.

[15] Bruce, and others. “Sample Entropy Tracks Changes in EEG Power Spectrum With Sleep State and Aging.” In: J. Clin. Neurophysiol. 2009 August; 26, no. 4 (2009).

[16] Chen, and others. Using Random Forest to Learn Imbalanced Data. Dept, 2004.

[17] Cohen, and others. An Algorithm for Segmenting Categorical Time Series into Meaningful Episodes, Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis, n.d.

[18] Costa, and others. Multiscale Entropy Analysis of Human Gait Dynamics. in: Physica A 330, 2003.

[19] “Multiscale Entropy Analysis of Physiologic Time Series.” Phys Rev Lett 2002

[20] Data Mining: Practical Machine Learning Tools and Techniques, n.d.

[21] Deisboeck et al. Complex Systems Sciene in Biomedicine, n.d.

[22] Duin, and others. Object Representation, Sample Size and Dataset Complexity. Edited by in: Basu and others. Data Complexity in Pattern Recognition, Springer, ISBN: 1-84628-171-7, 2006.

[23] Epstein. Shaping Time: Music, the Brain, and Performance, Wadsworth Publishing, n.d.

[24] Feuilloy. Comparison of Feature Selection Methods for Syncope Prediction, 2006 IEEE Congress on Evolutionary Computation, Vancouver, BC, Canada July 16-21, 2006.

[25] Gailly, and others. The Gzip Home Page, n.d. http://www.gzip.org/.

[26] Gao. “Detecting Nonstationarity and State Transitions in a Time Series.” In 2001. Phys. Rev. Vol. 63, 2001.

[27] Govindan, and others. “Revisiting Sample Entropy Analysis, in: Physica A, Statistical Mechanics and Its Applications” 376, no. 0378–4371 (2007): 158–64.

[28] Guyon, and others. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 2003.

[29] “Guzzetti, Symbolic Dynamics of Heart Rate Variability A Probe to Investigate Cardiac Autonomic Modulation. 2005; 112: 465-470 Published Online before Print July 18, 2005,” n.d.

[30] Hall, and others. “Feature Subset Selection: A Correlation Based Filter Approach.” In: Proceedings of the 1997 Int’l Conf. on Neural Information Processing and Intelligent Information Systems, N. Zealand. Ed: Kasabov N. Springer; 855–858. (1997).

[31] “Heart Failure, Wikipedia,” n.d. http://en.wikipedia.org/wiki/Heart_failure.

[32] Hegger, and others. TISEAN, n.d. http://www.mpipksdresden.mpg.de/ ~tisean/Tisean_3.0.1/index.html

[33] Henon. “A Two-Dimensional Mapping with a Strange Attractor.” In Comm. Math. Phys. 50, 50:69–77, 1976.

[34] Hornero, and others. “Nonlinear Analysis of Electroencephalogram and Magnetoencephalogram Recordings in Patients with Alzheimer’s Disease.” In: Phil. Trans. R. Soc. A 28, no. 367 (January 2009): 317.

[35] Huang, and others. A Wrapper for Feature Selection Based on Mutual Information. Int’l Conf, 2006.

[36] “Indian Pacing and Electrophysiology Journal (ISSN 0972-6292), 3(1): 34-40 (2003),” n.d.

[37] Janosi. “Time-Series Analysis of Transient Chaos.” In: Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 1994 Apr;49(4) 2756 (1994).

[38] Jin, and others. “Detecting Dynamical Complexity Changes in Time Series Using the Base-Scale Entropy.” In: Chin. Physics, v 14 (2005).

[39] Junges, and others. Intricate Routes to Chaos in the MackeyGlass Delayed Feedback System. in: Phys Let A 376, 2012.

[40] Kaffashi, and others. The Effect of Time Delay on Approximate and Sample Entropy Calculations. in: Physica D 2008 v 237, 2008.

[41] Kantz. “Quantifying the Closeness of Fractal Measures.” In: Phys. Rev. E 49 (1994): 5091–97.

[42] Kantz, and others. “Nonlinear Time Series Analysis.” Cambridge Univ. Press, ISBN, 1997, 0521653878.

[43] Kaplan. Nonlinearity and Nonstationarity, in: Frontiers of Blood Pressure and Heart Rate Analysis, 1997.

[44] Kawahara. “Sequential Change-Point Detection Based on Direct Density-Ratio Estimation, in: Statistical Analysis and Data Mining” 5, no. 2 (2012): 114127.

[45] Kennel. “Statistical Test for Dynamical Nonstationarity in Observed Time-Series Data.” In: Phys. Rev. E 56, no. 316 (n.d.).

[46] Keogh, and others. Segmenting Time Series: A Survey and Novel Approach, in: Data Mining in Time Series. 57, 2004.

[47] Kiyono, and others. “Phase Transition in a Healthy Human Heart Rate.” Phyical Review Letters 95, 2005, 058101.

[48] Koebbe, and others. “Use of Recurrence Plots in the Analysis of Time-Series Data, in: Proceedings of SFI Studies in the Science of Complexity (...)” XXI (1992): 361378.

[49] Kohavi. Wrappers for Performance Enhancement and Oblivious Decision Graphs, PhD Thesis, Stanford University. Computer Science department, n.d.

[50] Kolmogorov. Entropy per Unit Time as a Metric Invariant of Automorphism, in: Doklady of Russian Academy of Sciences, (1959). 124, n.d.

[51] Kullback. “On Information and Sufficiency.” Annals of Mathematical Statistics 22 (n.d.): 79–86.

[52] Lake, and others. “Accurate Estimation of Entropy in Very Short Physiological Time Series.” In: Am. J. Physiol. Heart. Circ. Physiol. 2011 Jan; 300, no. 1 (2011).

[53] “Sample Entropy Analysis of Neonatal Heart Rate Variability.” In: Am. J. Physiol. Regul. Integr. Comp. Physiol 283 (2002).

[54] Levitin. This Is Your Brain on Music, Plume, n.d.

[55] Li, and others. “Analysis of Nonlinear Dynamics and Detecting Messages Embedded in Chaotic Carriers Using Sample Entropy Algorithm, in: Journal of the Optical Society of America B” 28, no. 1 (2011).

[56] Lin, and others. “A Symbolic Representation of Time Series, with Implications for Streaming Algorithms, in: DMKD’ 03, June 13, 2003, San Diego, CA.” USA. Copyright 2003 ACM, 2003, 1–58113–763–.

[57] “Experiencing SAX: A Novel Symbolic Representation of Time Series, in: Data Mining and Knowledge Discovery” 15 (2007): 107ff.

[58] Little. Recurrence Period Density Entropy, n.d. http://en.wikipedia.org/wiki/Recurrence_period_density_entropy.

[59] Little, and others. “Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection, Biomed Eng Online.” Jun 26 6, no. 1 (2007): 23.

[60] Lucas. Quantifying Complexity Theory, n.d.

[61] Macau. “Understanding the Complexity in Low Dimensional Systems, J. Braz. Soc. Mech. Sci. vol.24 no.4 Rio de Janeiro Nov. 2002,” n.d.

[62] Mackey, and others. Oscillation and Chaos in Physiological Control Systems. in: Science 197, 197AD.

[63] Maestri, and others. “Nonlinear Indices of Heart Rate Variability in Chronic Heart Failure Patients: Redundancy and Comparative Clinical Value, J Cardiovasc Electrophysiol” 18 (March 2007): 1–9.

[64] Makikallio. Analysis of Heart Rate Dynamics by Methods Derived from Nonlinear Mathematics. Clinical applicability and prognostic significance, Acta Universitatis Ouluensis Medica, 1998.

[65] Mallik, Marek. “Heart Rate Variability.” Standards of Measurement, Physiological Interpretation, and Clinical Use, European Heart Journal 17 (1996).

[66] May. Simple Mathematical Models with Very Complicated Dynamics. in: 1976 Nature 261, 1976.

[67] McSharry. Detection of Dynamical Transitions in Biomedical Signals Using Nonlinear Methods, Lecture Notes Artificial Intelligence 3215:483-490, Springer-Verlag. Heidelberg Berlin, n.d.

[68] “The Danger of Wishing for Chaos, Nonlinear Dynamics, Psychology.” And Life Sciences 9, no. 4 (n.d.): 375–97.

[69] McSharry, and others. “Linear and Non-Linear Methods for Automatic Seizure Detection in Scalp Electro-Encephalogram Recordings.” Biol. Eng. Comput. 2002; 40, no. 4 (2002): 447–61.

[70] Meesmann. Heart Rate Variablity and Non Linear Dynamics, n.d.

[71] Meloon, and others. Quantification of Determinism in Music Using Iterated Function Systems. Empirical Studies of the Arts 15, 1997.

[72] Mietus, and others. The pNNx Files: Re-Examining a Widely Used Heart Rate Variability Measure, n.d.

[73] Migliaro, and others. Heart Rate Variability: Short Term Studies Are as Useful as Holter to Differentiate Diabetic Patients from Healthy Subjects, n.d.

[74] Morchen, and others. Optimizing Time Series Discretization for Knowledge Discovery, Proceeding of the Eleventh ACM SIGKDD International Conference on

Knowledge Discovery in Data Mining Table of Contents, Chicago, Illinois, Pages: 660-665, n.d.

[75] Moskvina, and others. “An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection.” In: 2003. Comm. Stat. Simul. Comput 32 (2003).

[76] Motoda, and others. Feature Selection, Extraction and Construction, Invited. Proc, 2002.

[77] Nabors-Oberg, and others. Investigating Possible Development of Short-Term Tolerance Using Spectral Analysis, IEEE ENGINEERING IN MEDICINE AND BIOLOGY. July/August 2002, 2002.

[78] Narayanan, and others. “Unstable Periodic Orbits in Human Cardiac Rhythms.” Phys. Rev E 57 (n.d.): 4594–4603.

[79] “Unstable Periodic Orbits in Human Cardiac Rhythms.” Phys. Rev E 57 (n.d.): 4594–4603.

[80] NIH. What Is an Implantable Cardioverter Defibrillator?, n.d. http://www.nhlbi.nih.gov/health/dci/Diseases/icd/.

[81] Owis, and others. Study of Features Based on Nonlinear Dynamical Modeling in ECG Arrhythmia Detection and Classification, IEEE Transactions on Biomedical Engineering, 49(7):733-736, Jul, 2002.

[82] Study of Features Based on Nonlinear Dynamical Modeling in ECG Arrhythmia Detection and Classification, IEEE Transactions on Biomedical Engineering, 49(7):733-736, Jul, 2002.

[83] Packard, and others. “Geometry from a Time Series.” Phys. Rev. Lett. 45 (1980) 45 (1980): 712–16.

[84] Palazzolo, and others. Entropy Measures of Heart Rate Variation in Conscious Dogs, Am J Physiol Heart Circ Physiol 274: H1099-H1105, 1998.

[85] Pan, and others. “A Real-Time QRS Detection Algorithm.” In: Biom. Eng., IEEE Trans., v. BME-32, Issue 3 (1985): 230–36.

[86] “Park et Al.: Accessing Physiological Complexity of HRV by Using Threshold-Dependent Symbolic Entropy, Journal of the Korean Physical Society, Vol. 44, No. 3, March 2004, Pp. 569-576,” n.d.

[87] Park, and others. “Accessing Physiological Complexity of HRV by Using Threshold-Dependent Symbolic Entropy, Journal of the Korean Physical Society” 44, no. 3 (March 2004): 569–76.

[88] Peng, and others. “Exaggerated Heart Rate Oscillations During Two Meditation Techniques.” In: Int’l J of Card. 70 70 (1999): 101–7.

[89] Mosaic Organization of DNA Nucleotides, 1994.

[90] “Physionet, Http://www.physionet.org/tutorials/hrv/,” n.d. Physionet, http://www.physionet.org/tutorials/hrv/.

[91] Pincus. Approximate Entropy as a Measure of System Complexity, n.d.

[92] Approximate Entropy as a Measure of System Complexity. Proc Natl Acad Sci USA 1991;88:2297-2301, n.d.

[93] Pincus, and others. Physiological Time-Series Analysis: What Does Regularity Quantify? Am J Physiol 1994;266(Heart Circ Physiol):H1643-H1656, n.d.

[94] Porta, and others. “Entropy, Entropy Rate, and Pattern Classification as Tools to Typify Complexity in Short Heart Period Variability Series, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING” 48, no. 11 (2001).

[95] Symbolic Analysis of 24h Holter Heart Period Variability Series: Comparison between Normal and Heart Failure Patients, n.d.

[96] Symbolic Analysis of 24h Holter Heart Period Variability Series: Comparison between Normal and Heart Failure Patients, n.d.

[97] Puglisi. “Data Compression and Learning in Time Sequences Analysis, Physica D: Nonlinear Phenomena, Volume 180, Issues 1-2.” 1 92–107 (June 2003).

[98] Quiroga, Quian, and others. “Kulback-Leibler and Renormalized Entropies: Applications to Electroencephalograms of Epilepsy Patients.” PHYSICAL REVIEW E 62 (2000).

[99] “Kulback-Leibler and Renormalized Entropies: Applications to Electroencephalograms of Epilepsy Patients.” PHYSICAL REVIEW E 62 (2000).

[100] Rajendra. “Heart Rate Variability: A Review.” In: Med Bio Eng Comput 44 (2006).

[101] Richman, and others. “Physiological Time Series Analysis Using Approximate Entropy and Sample Entropy.” In: Am. J. Physiol. 2000; 278, no. 6 (2000).

[102] “Sample Entropy.” In: Numerical Computer Methods. 2004 Vol 384 (2004): 173.

[103] Rigoldi, and others. “Measuring Regularity of Human Postural Sway Using Approximate Entropy and Sample Entropy in Patients with Ehlers-Danlos Syndrome Hypermobility Type.” In: Res. Dev. Disabil. 2013 Feb; 34, no. 2 (2013): 840–46.

[104] “Rissanen, An Introduction to the MDL Principle,” n.d.

[105] Rosenhek. Applying Echocardiography in Heart Failure: How to Provide Meaningful and High-Quality Data Relevant to the Practitioner, American College of Cardiology 52nd Annual Scientific Session, 2003, Chicago, n.d.

[106] Rosenstein, and others. A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets, Physica D 65, 117-134, 1993.

[107] Reconstruction Expansion as a Geometry-Based Framework for Choosing Proper Delay Times, Physica D 73, 82-98, 1994.

[108] Roth. Bringing out the Best Features of Expression Data; Genome Research (Insight/Outlook), 2001.

[109] Rssler. “An Equation for Hyperchaos,” In: Physics Letters 71A (2, 1979.

[110] Saeys, and others. “Feature Selection for Splice Site Prediction: A New Method Using EDA-Based Feature Ranking.” BMC Bioinformatics 2004 5 (2004): 64.

[111] Savi. “Chaos and Order in Biomedical Rhythms.” J. of the Braz. Soc. of Mech. Sci. & Eng. 27 (2005) 27 (2005): 157–69.

[112] Schreiber. “Detecting and Analysing Nonstationarity in a Time Series Using Nonlinear Cross Predictions.” In: Phys. Rev. Lett. 78 78, no. 843 (1997).

[113] “Detecting and Analyzing Nonstationarity in a Time Series Using Nonlinear Cross Predictions.” Phys. Rev. Lett. 78 78 (1997): 843–46.

[114] Schumann, and others. “Potential of Feature Selection Methods in Heart Rate Variability Analysis for the Classification of Different Cardiovascular Diseases, Statistics in Medicine” 21 (n.d.): 2225–42.

[115] Seely, and others. Complex Systems and the Technology of Variability Analysis. Critical Care 2004, 8:R367-R384 (DOI 10, 2004.

[116] Shmulevich, and others. : Measures of Temporal Pattern Complexity, Journal of New Music Research. 29, n.d.

[117] Shuangcheng, and others. Measurement of Climate Complexity Using Sample Entropy, in: Innternational Journal of Climatology, Volume 26, Issue 15, Pages 21312139, 2006.

[118] “Solomonoff, R., A Preliminary Report on a General Theory of Inductive Inference,” n.d.

[119] Song, and others. “Epileptic EEG Signal Analysis and Identification Based on Nonlinear Features.” In: 2012 IEEE Intl Conf. on BIBM, Philadelphia, PA, 2007, ISBN 978–1–4673–2559–2 (2007).

[120] Speidel, and others. How Well Do Practical Information Measures Estimate the Shannon Entropy?, COMMUNICATION SYSTEMS, NETWORKS AND DIGITAL SIGNAL PROCESSING, 5th Int’l Symposium, 2006.

[121] Steuer, and others. Entropy and Optimal Partion for Data Analysis. The European Physical Journal 19, 2001.

[122] Struzik. “Revealing Local Variability Properties of Human Heartbeat Intervals with the Local Effective Holder Exponent.” INS-R0015 June 30 1386–3681 (2000).

[123] Sztajzel. “Swiss Med. Wkly 2004;134:514–522,” n.d.

[124] Tarvainen, and others. “Kubios HRV A Software for Advanced Heart Rate Variability Analysis.” Edited by in: 4th Euro Conf of the Intl Federation for. And Biol. Eng., IFMBE Proceedings Vol 22 (2009).

[125] Tel. “Crossover between the Control of Permanent and Transient Chaos.” In: Int. J. B. Ch. 03 03, no. 757 (1993).

[126] Theiler, and others. “Testing for Nonlinearity in Time Series: The Method of Surrogate Data.” In: Phys. D 58 (1992): 77–94.

[127] Thuraisingham, and others. On Multiscale Entropy Analysis for Physiological Data, n.d.

[128] Tikkanen. “Characterization and Application of Analysis Methodsfor ECG and Time Interval Variability Data.” Acta Universitatis Ouluensis A 323. Oulu: University of Oulu. ISBN 951–42–5214–4 (n.d.).

[129] Turnbull. A Supervised Approach for Detecting Boundaries in Music, Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, n.d.

[130] “UMK,” n.d. http://www.cardiologyonline.com/institutes3.htm.

[131] Urbach. Footprint of Chaos in the Markets, n.d.

[132] Vlachos, and others. On Periodicity Detection and Structural Periodic Similarity. in: Proc, n.d.

[133] Voss, and others. Multiparametric Analysis of Heart Rate Variability Used for Risk Stratification Among Survivors of Acute Myocardial Infarction, Pacing and Clinical Electrophysiology, Volume 21 Issue 1 Page 186-196, 1998.

[134] Nonlinear Dynamics in Cardiovascular Diseases, Proceedings of the Second World Congress on Nonlinear Analysts, Athens, Greece, Pages: 935-941, n.d.

[135] “Wallace et Al.: An Information Measure for Classification.. Computer Journal, Vol 11, No 2, August 1968, Pp 185-194,” n.d.

[136] Wessel, and others. Classifying Simulated and Physiological Heart Rate Variability Signals, n.d.

[137] Entropy Measures in Heart Rate Variability Data. Lecture notes in computer science 2000, 1933.

[138] Short-Term Forecasting of Life-Threatening Cardiac Arrhythmias Based on Symbolic Dynamics and Finite-Time Growth Rates. Phys Rev E, 2000.

[139] “Symbolic Dynamics for Medical Data Analysis, in: Attractors, Signals, and Synergetic (W.” Edited by Klonowski 1 (2002): 45–61.

[140] Wiklund, and others. Short-Term Analysis of Heart-Rate Variability by Adapted Wavelet Transforms, IEEE ENGINEERING IN MEDICINE AND BIOLOGY. September/October 1997, 1997.

[141] Yu, and others. “Detecting Dynamical Nonstationarity in Time Series Data.” In: Chaos 9 865 (1999).

[142] “Space Time-Index Plots for Probing Dynamical Nonstationarity.” In: Phys Lett. A 250 (1998).

[143] Zebrowski, and others. “Entropy, Pattern Entropy and Related Methods for the Analysis of RR Interval Data from 24-H Electrocardiograms.” Phys. Rev. E50 5 (1994).

[144] “Measuring the Complexity of Non-Stationary Time Series - Nonlinear Interpretations of Selected Physiological Processes, Acta Physica Polonica B” 30, no. 8 (August 1999).

[145] “Nonlinear Instabilities and Nonstationarity in Human Heart-Rate Variability.” In: J 6 (2004): 78–83.

[146] Nonlinear Instabilities and Nonstationarity in Human Heart-Rate Variability, Computing in Science and Engineering Archive, Volume 6, Issue 5, Pages: 78-83, n.d.

[147] “Observations and Modeling of Deterministic Properties of Human Heart Rate Variability, Indian Academy of Sciences, in: Pramana” 64, no. 4 (April 2005): 543.

[148] Ziv et Al.: A Universal Algorithm for Sequential Data Compression, IEEE Transactions on Information Theory, Vol. 23, No. 3, Pp. 337–343, n.d.

[149] Polukoshko et al., USE OF “CATERPILLAR” – SSA METHOD FOR ANALYSIS AND FORECASTING OF INDUSTRIAL AND ECONOMIC INDICATORS, Environment. Technology. Resources. Proceedings of the 7th International Scientific and Practical Conference. Volume I1

[150] Wang, A Comparison of Two SSA Approaches, Thesis, University of California, Los Angeles, 2009

[151] Stapenhurst, Mastering Statistical Process Control, ISBN: 978-0-7506-6529-2

BIOGRAPHY

Rafal Ladysz received his M.Sc. in Physics from the Adam Mickiewicz University in Poznan, Poland. Before his arrival to the US in 1990 he worked at the Institute of Low Temperature and Structure Research of Polish Academy of Sciences in Wroclaw, Poland. In the US he focused his interests and professional activity on variety of IT areas, working in New York, Maryland and Virginia. He then enrolled the IT program at George Mason University and received his Ph.D. in Information Technology in 2014. He plans to continue his research in the area of computational medicine in USA and Europe.