Large-Scale Dimension Densities for Heart Rate Variability Analysis

C Raab[1], N Wessel[1,2], A Schirdewan[2], J Kurths[1]

[1]Institute of Physics, University of Potsdam, Germany
[2]Franz-Volhard-Hospital, Helios-Clinics, Charité, Humboldt University Berlin, Germany

Abstract

We analyse the heart rate variability data (HRV) using the concept of large-scale dimension densities (LASDID). This method enables to analyse very short and non-stationary data, such as HRV, and, hence, also short parts of the data and to look for differences between day and night. The circadian changes in the dimension density enable to an almost completel distinction between real data and computer generated data from CiC 2002 challenge using only one parameter.

Furthermore, we analyse the data of 15 patients with atrial fibrillation (AF), 15 patients with congestive heart failure (CHF), 15 elderly healthy subjects (EH) as well as 18 young and healthy persons (YH). With our method we are able to separate completely the AF group from the others and the CHF patients show significant differences to the young and elderly healthy volunteers.

1. Introduction

Annually, in the United States up to 450,000 people die due to sudden cardiac death [1, 2]. Therefore, an accurate and reliable identification of patients who are at high risk for sudden cardiac death is an important and challenging problem. In this paper we introduce a measures of complexity which may help to solve this problem when applied to HRV. Observational data, such as HRV, often are rather short and may be noisy. Different data analysis techniques to understand complex processes observed in nature[3, 4, 5] were developed. Linear approaches of time series analysis are often not sufficient [6, 7] and most of the nonlinear techniques [8] suffer from the curse of dimensionality. Mostly, there are not enough points in the (often non-stationary) time series to reliably estimate these nonlinear measures. The uncritical application of these methods especially to natural data, therefore, can be very dangerous and often lead to serious pitfalls.

To overcome these difficulties, other measures of complexity have been proposed, such as Renyi entropies, effective measure complexity or ε-complexity [9, 10]. They are mostly basing on symbolic dynamics and are efficient quantities to characterize measurements of natural systems, such as in cardiology [11, 12], cognitive psychology [13] or astrophysics [14]. These methods are often not sufficient for very short data sets, so we focus in this paper on another type of measures of complexity basing on the method of LASDID [15] and apply this methodology to HRV data. LASDID allows to analyse very short data sets, so it is possible to calculate it for short parts of the data and get an overview of the changes in the dimension density inbetween 24 hours.

2. Method

LASDID [15] is estimated with a normalized Grassberger-Procaccia algorithm, which leads to a suitable correction of systematic errors produced by boundary effects in the rather large scales of a system. So it is possible to analyse rather short and non-stationary data.

To calculate the correlation dimension D_2 of a system with the Grassberger-Procaccia algorithm [16], means that the attractor firstly has to be reconstructed by embedding. The embedded time series consists of vectors $\{\vec{x}(t) = (x_1(t), x_2(t), ..., x_m(t))\}$, where m is the embedding dimension. Then one has to calculate the correlation integral $C(r,m) = \frac{1}{N(N-1)} \sum_{i \neq j} \theta(r - |\vec{x}(t_i) - \vec{x}(t_j)|)$, where θ is the Heaviside function and r is the radius around each point within neighbouring points are counted for the correlation sum. D_2 is then defined as

$$D_2 = \lim_{r \to 0} \lim_{m \to \infty} (\mathrm{d} \log C(r,m) / (-\mathrm{d} \log(r))), \quad (1)$$

if this limit exists [16]. Because it is impossible to reach the limit $r \to 0$ in numerical calculations, one has to estimate this dimension from larger distances, i. e. the right hand side of eq. (1) becomes a distant dependent function $D_2(r,m)$. For low-dimensional attractors for small r there often exists a rather large region in $\log_2(r)$ where this $D_2(r,m)$ is nearly constant. This part is referred to as the scaling region [16]. For larger values of r, $D_2(r,m)$ is decreasing because of boundary effects. It has been shown, that with the growing dimension of the attractor the number of data points needed to reach the scaling region is increasing exponentially [8, 15]. If the time series

is too short, one only gets the part of $D_2(r,m)$ with decreasing values. With LASDID we are able to use this part of $D_2(r,m)$ too.

The large-scale dimension density $\rho_{\mathrm{ls}}(r,m)$ is defined by normalizing the dimension density $D_2(r,m)/m$ of all coordinates m of the embedded system to the dimension density $D_2(r,1)$ of one coordinate of this system [15]:

$$\rho_{\mathrm{ls}}(r,m) = D_2(r,m)/(mD_2(r,1)). \tag{2}$$

With this normalization we get a plateau for large scales r yielding an estimate of ρ_{ls}. The advantage of LASDID is that it is possible to estimate it from rather short and non-stationary time series. So we can cut every RR-interval time series in M shorter pieces. For every of this short pieces we calculate the large-scale dimension density via eq. 2. This leads to a time series of $\rho_{\mathrm{ls}}(t)$. For this time series we calculate the mean value $\mu_{\rho_{\mathrm{ls}}}$ by

$$\mu_{\rho_{\mathrm{ls}}} = \frac{1}{M}\sum_{i=1}^{M}\rho_{\mathrm{ls}}(t_i), \tag{3}$$

the standard deviation $\sigma_{\rho_{\mathrm{ls}}} = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(\rho_{\mathrm{ls}}(t_i)-\mu_{\rho_{\mathrm{ls}}})^2}$ and the coefficient of variation $cv_{\rho_{\mathrm{ls}}}$ by

$$cv_{\rho_{\mathrm{ls}}} = \sigma_{\rho_{\mathrm{ls}}}/\mu_{\rho_{\mathrm{ls}}}. \tag{4}$$

For the calculation of LASDID we use an embedding-dimension of $m = 4$ and a delay of $\tau = 1$. But the results are qualitatively the same with embedding dimensions between $m = 4, ..., 8$ and delay times between $\tau = 1, ..., 5$.

3. Data

Physiological data very often show complex structures which cannot be simply described and, therefore, their interpretation is difficult. For the HRV data we are analysing in this paper, it is well known that a metronomic heart rate is pathological - the healthy heart is influenced by multiple neural and hormonal factors that result in variations in RR intervals. Even after three decades of study, new techniques continue to reveal properties of the time series of RR intervals. Moreover, the simulation of such time series is still extremely sophisticated and PhysioNet [17] and Computers in Cardiology 2002 organized a challenge to improve the momentary understanding of cardiovascular regulation. The aim of the first part of this challenge was to construct simulations of the RR interval time series spanning a full 24 hours with sufficient verisimilitude to be taken as real. In a second part a blind classification of a mixed set of real and simulated RR interval time series shall be performed.

In this paper, we reanalyze the 46 time series from the second part of this challenge using LASDID to test whether new information in RR interval variation can be revealed. Therefore, the first intention of this contribution is to sketch our way of discriminating both types of time series using LASDID.

The second intention of this paper is to demonstrate a possible application for risk stratification. Therefore, we analyze the data of 15 patients with atrial fibrillation (AF) (15 male, age: 67 ± 12), of 15 patients with congestive heart failure (CHF) (11 male, 4 female, age: 56 ± 11), of 15 elderly healthy subjects (10 male, 5 female, age: 50 ± 9) as well as of 18 young healthy persons (13 female, 5 male, age: 34 ± 8). The data of the CHF patients and the young healthy subjects are available from Physionet [17]. After pre-processing [18], we calculate LASDID and compare it with standard time and frequency domain parameters as well as parameters based on symbolic dynamics which have been recently successfully applied to other cardiological problems [11, 19]. The following HRV parameters are calculated from the time series: meanNN, the mean value of normal beat-to-beat intervals; sdNN, the standard deviation of intervals between two normal; rmssd, the root mean square of successive RR-intervals; and pNN50, the percentage of RR-interval-differences greater than 50 ms. Additionally, in the frequency domain the normalised low-frequency (LFn) the ratio LF/HF are estimated. Finally, HRV is analyzed by methods of nonlinear dynamics, especially symbolic dynamics [12, 19]: FWSHANNON, the Shannon entropy of the word distribution and POLVAR10, a measure to detect intermittently decreased HRV.

4. Results

First we use the method of LASDID to compare time series of real ECG data with those of simulated data. We subdivide every time series in pieces of 1000 heart beats and calculate ρ_{ls}. This leads to time series with fluctuating values of ρ_{ls} which are analyzed by calculating the mean value $\mu_{\rho_{\mathrm{ls}}}$, the standard deviation $\sigma_{\rho_{\mathrm{ls}}}$ and the coefficient of variation $cv_{\rho_{\mathrm{ls}}}$. For real data we find values of $\mu_{\rho_{\mathrm{ls}}}$ between 0.5 to 0.7, whereas simulated data ranges between 0.4 to 0.9, only half of the models generated data which also ranges between 0.5 to 0.7. Values near one indicate a rather stochastic behaviour of the heart rate, values near zero mean deterministic heart beats.

Furthermore real data shows stronger fluctuations in the time series of LASDID, i.e. the values of $\sigma_{\rho_{\mathrm{ls}}}$ are higher for real data ($\sigma_{\rho_{\mathrm{ls}}}$ from 0.09 to 0.17 for real data against $\sigma_{\rho_{\mathrm{ls}}}$ from 0.02 to 0.11 for simulated data) representing circadian varibility changes. The best discrimination result, however, we get with the coefficient of variation $cv_{\rho_{\mathrm{ls}}}$. It makes it possible to distinguish between real and simulated data by using only one parameter. Almost all simulated time series can be detected with this method (see fig 1).

All calculations also have been done with more or less

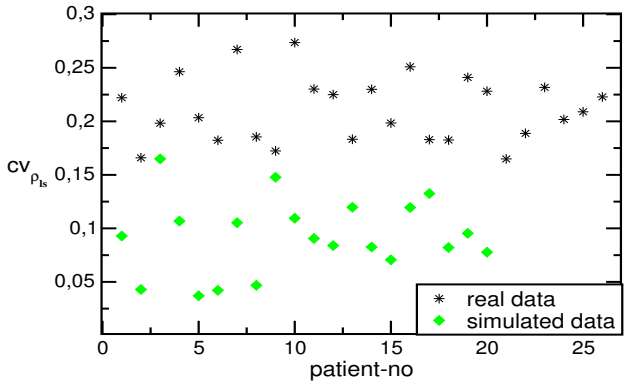

Figure 1. A comparison of $cv_{\rho_{1s}}$ (Eq.4) of real data and simulated data shows higher values for real data.

heart beats per interval. For less than 500 heart beats, ρ_{1s} can not be calculated reliably. Intervals of 2000 heart beats give almost the same results as intervals of 1000 heart beats, for longer intervals more and more information about the circadian changes gets lost.

The records of the real data always started and ended in the morning, so it is possible to distinguish between day and night. In the following we used the first 5 hours of the 24 hours ECGs as day interval and the hours 17 to 22 as night interval. For real data we find higher values of $\mu_{\rho_{1s}}$ for the night for most of the records (day: $\mu_{\rho_{1s}} = 0.546 \pm 0.056$; night: $\mu_{\rho_{1s}} = 0.628 \pm 0.069$; P for day vs. night < 0.001). But only few of the simulated data sets show differences between two different time intervals.

The second intention of this paper was to demonstrate a possible application for risk stratification. Therefore, we compared the data of different pathologies and healthy subjects. For patients with atrial fibrillation (AF) we find values of $\mu_{\rho_{1s}}$ near one which indicates almost stochastic heart beats. The coefficient of variation $cv_{\rho_{1s}}$ for this patients is very low (see tab. 1). This means, the AF-group separates completely from the others. Elderly Patients with congestive heart failuer (CHF) show higher values of $cv_{\rho_{1s}}$. The highest values we find for elderly healthy patients (EH).

Table 1. The four different groups of patients are AF (Atrial Fibrillation), CHF (Congestive heart Failure), EH (Elderly Healthy) and YH (Young Healthy). They have different mean values of $\mu_{\rho_{1s}}$ (Eq.3) and $cv_{\rho_{1s}}$ (Eq.4) (* $p < 0.001$ vs. AF group, $^\diamond$ $p < 0.05$ vs. CHF group, $^\nabla$ $p < 0.05$ vs. EH group).

Group	$\mu_{\rho_{1s}}$	$cv_{\rho_{1s}}$
AF	0.968 ± 0.021	0.024 ± 0.013
CHF	$0.651 \pm 0.125^*$	$0.168 \pm 0.053^*$
EH	$0.563 \pm 0.042^{*\diamond}$	$0.209 \pm 0.028^{*\diamond}$
YH	$0.6062 \pm 0.0392^{*\nabla}$	$0.185 \pm 0.021^{*\nabla}$

Table 2. Correlations coefficients r (p-value) between LASDID and HRV parameters (* $p < 0.001$, $^\diamond$ $p < 0.01$, $^\nabla$ $p < 0.05$).

	$\mu_{\rho_{1s}}$	$cv_{\rho_{1s}}$
meanNN	0.053	0.107
sdNN	0.227	0.152
rmssd	$0.509^\diamond$	0.059
pNN50	$0.510^\diamond$	0.046
LF/HF	-0.607^*	-0.226
LFn	-0.735^*	-0.163
FWSHANNON	-0.659^*	0.037
POLVAR10	$-0.553^\diamond$	-0.145

This means, low values of $cv_{\rho_{1s}}$ indicate a higher risk of heart disease. For the healthy persons we again find higher values of $\mu_{\rho_{1s}}$ for the night, but not in patients with congestive heart failure (EH: day 0.54 ± 0.05, night 0.61 ± 0.05, $p = 0.002$; YH: day 0.57 ± 0.05, night 0.67 ± 0.07, $p < 0.001$; CHF: day 0.65 ± 0.13, night 0.66 ± 0.12, $n.s.$). Thus, finding no circadian differences in $\mu_{\rho_{1s}}$ is also a pathological sign.

In order to investigate the physiological correlates for LASDID we performe a correlation analysis. Pearson correlation coefficients between different HRV parameters and $\mu_{\rho_{1s}}$ and $cv_{\rho_{1s}}$ are given in table 2. Mean heart rate (inversely related to meanNN) as well as sdNN, the standard deviation of the time series, does not correlate with $\mu_{\rho_{1s}}$ and $cv_{\rho_{1s}}$. For rmssd, the root mean square of successive differences, however, we see a significant relation to $\mu_{\rho_{1s}}$, i.e. short term respiratory induced oscillation in HRV plays an important role for LASDID. The highest correlation we find for the normalized low frequency band around 0.1 Hz to $\mu_{\rho_{1s}}$, demonstrating that the Mayer waves having the strongest influence for estimating LASDID. Interstingly, $cv_{\rho_{1s}}$ did not show any significant relation to HRV parameters.

5. Conclusions

We presented a way of discriminating the 46 simulated and physiological HRV time series from the 2002 Computers in Cardiology challenge [20]. In a previous paper [21] we used three different parameter which are based on the distribution of RR-intervals, the circadian beat-to-beat variability as well as the beat-to-beat dynamics. Using cut-offs for these parameters, both time series groups could be discriminated completely. The cut-offs were subjectively chosen based on the knowledge of the normal ranges of the used parameters. Moreover, it was an act of instinct which parameter to choose first. To the best of our knowledge, until today there was no single parameter for the complete separation of the considered groups. Using the concept of

LASDID, a nearly perfect classification was performed for the first time. Only one of the simulated time series (no. 4) was falsely classified as a real one. This time series showed a comparable number of degrees of freedom (number of modes) as compared to real data and this number showed a circadian dependence. The modes, however, were choosen too rigid - one can easily detect this time series as an artificial one from its frequency spectrum. The averaged LASDID $\mu_{\rho_{1s}}$, characterizing the number of independent modes (the working regulatory circuits) generating the heart rate data, are statistically different between real and simulated data. The circadian variation of the number of independent modes $cv_{\rho_{1s}}$, however, enables a nearly perfect discrimination between physiological and artificial data. Real heart rate data are characterized by circadian variability changes due to different mechanisms. At daytime there are influences from physical or mental stress, food intake - in the night you should have no stress, however, there are significant differences in the sleep stages, too. No simulation in this data base was able to model all these effects.

In the second part of this paper we demonstrate its potentials for risk stratification. Patients with atrial fibrillation show averaged large-scale dimension densities near to one and can be completely discriminated from the other groups. In addition, the group of the elderly healthy subjects is statistically different in $\mu_{\rho_{1s}}$ to the congestive heart failure group. Interestingly, the young healthy volunteers are not statistically different to the CHF group. This is due to the fact that HRV decreases with age, here the number of modes $\mu_{\rho_{1s}}$ decreases too (see YH vs. EH in Tab.1). In the CHF group $\mu_{\rho_{1s}}$ is increased compared to elderly healthy subjects. This means the number of independent modes increases due to the disease - possible explanations are ventricular ectopy or pulsus alternans. For the circadian variation of $cv_{\rho_{1s}}$ the same penomena can be detected: Patients with AF do not show circadian variations and the young healthy group is inbetween the CHF and the elderly healthy group.

Finally, looking at the correlation of LASDID to standard HRV parameters and finding no statistical significant relation for $cv_{\rho_{1s}}$ demonstrates the independence of our approach which may be important also for clinical risk stratification.

References

[1] Alberte C, Zipes DP. Use of nonantiarrhythmic drugs for prevention of sudden cardiac death. J Cardiovasc Electrophysiol 2003;14:87 – 95.

[2] Barron HV, Lesh MD. Autonomic nervous system and sudden cardiac death. J Am Coll Cardiol 1996;27(5):1053–1060.

[3] Glass L. Synchronization and rhythmic processes in physiology. Nature 2001;410:277–284.

[4] Schafer C, Rosenblum MG, Kurths J, Abel HH. Heartbeat synchronized with ventilation. Nature 1998;392:239–40.

[5] Marvel KB. Astrophysics. a stellar performance. Nature 2001;411(6835):252–252.

[6] Goldberger AL, et al. Nonlinear dynamics in sudden cardiac death syndrome: Heart rate oscillations and bifurcations. Experientia 1988;44:983–987.

[7] Glass L, Kaplan D. Time series analysis of complex dynamics in physiology and medicine. Med Prog Technol 1993; 19(3):115–128.

[8] Kantz H, Schreiber T. Nonlinear Time Series Analysis. Cambridge: Cambridge University Press, 1997.

[9] Wackerbauer R, et al. A comparative classification of complexity measures. Chaos Solitons Fractals 1994;4(4):133–173.

[10] Rapp PE, et al. Effective normalization of complexity measurements for epoch length and sampling frequency. Physical Review E 2001;64:16209.

[11] Kurths J, et al. Quantitative analysis of heart rate variability. Chaos 1995;5:88–94.

[12] Voss A, et al. The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. Cardiovasc Res 1996;31:419–433.

[13] Engbert R, et al. Symbolic dynamics of physiological synchronisation: examples from bimanual movements and cardiorespiratory interaction. Nonlin Anal Theo Meth Appl 1997;30:973–984.

[14] Schwarz U, et al. Analysis of solar spike events by means of symbolic dynamics methods. Astron Astrophys 1993; 277:215–224.

[15] Raab C, Kurths J. Estimation of large-scale dimension densities. Physical Review E 2001;64(1).

[16] Grassberger P, Procaccia I. Characterization of strange attractors. Physical Review Letters 1983;50:346–349.

[17] Goldberger A, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 2000;101:E215.

[18] Wessel N, et al. Nonlinear analysis of complex phenomena in cardiological data. Herzschr Elektrophys 2000;11:159–173.

[19] Wessel N, Schirdewan A, Kurths J. Intermittently decreased beat-to-beat variability in congestive heart failure. Phys Rev Lett 2003;91:119801.

[20] http://www.physionet.org/challenge/2002/. Computers in Cardiology 2002;29.

[21] Wessel N, et al. Classifying simulated and physiological heart rate variability signals. Computers in Cardiology 2002;29:133–135.

Address for correspondence:

Corinna Raab
University of Potsdam, Institute of Physics,
PF 601553, D-14415 Potsdam, Germany
corinna@agnld.uni-potsdam.de