Large-Scale Dimension Densities for Heart Rate Variability Analysis

Corinna Raab,[1] Niels Wessel,[1] Alexander Schirdewan,[2] and Jürgen Kurths[1]

[1]*Institute of Physics, University of Potsdam, Germany*[*]

[2]*Franz-Volhard-Hospital, Helios-Clinics,*

Charité, Humboldt University Berlin, Germany

(Dated: December 16, 2005)

Abstract

In this work, we reanalyse the HRV data from the 2002 Computers in Cardiology (CiC) Challenge using the concept of large-scale dimension densities and additionally apply this technology to data of healthy persons and of patients with cardiac diseases. The large-scale dimension density is estimated from the time series using 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. In this way, it is possible to analyse very short, non-stationary and unfiltered data, such as HRV. Moreover, this method allows us to analyse short parts of the data and to look for differences between day and night. The circadian changes in the dimension density enable us to distinguish almost completely between real data and computer generated data from CiC 2002 challenge using only one parameter. In the second part we analysed 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 ($\mu_{\rho_{\mathrm{ls}}} = 0.97 \pm 0.02$) group from the others and, especially during daytime, the CHF patients show significant differences to the young and elderly healthy volunteers (CHF: 0.65 ± 0.13, EH: 0.54 ± 0.05, YH: 0.57 ± 0.05, $p < 0.05$ for both comparisons). Moreover, for the CHF patients we find no circadian changes in $\mu_{\rho_{\mathrm{ls}}}$ (day: 0.65 ± 0.13, night: 0.66 ± 0.12, n.s.) in contrast to healthy controls (day: 0.57 ± 0.05, night: 0.67 ± 0.07, $p = 0.00004$). Correlation analysis showed no statistical significant relation between standard HRV and circadian LASDID demonstrating a possibly independent application of our method for clinical risk stratification.

INTRODUCTION

Annually, in the United States up to 450,000 people die due to sudden cardiac death [1–3]. 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 measure of complexity which may help to solve this problem when applied to heart rate variability (HRV) data. Observational data, such as HRV, often are rather short and may be noisy. Different data analysis techniques to understand complex processes observed in nature [4–6] were developed. Linear approaches of time series analysis are often not sufficient [7, 8] and most of the nonlinear techniques [9, 10] 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, ε-complexity or renormalized entropy [11, 12]. They are mostly basing on symbolic dynamics and are efficient quantities to characterize measurements of natural systems, such as in cardiology [13–15], cognitive psychology or astrophysics [16–18]. These methods are often not sufficient for very short data sets. For short data sets the method of point correlations has been introduced [19], but the dimension is estimated from a short part of the classical correlation dimension at small scales where no scaling region can be found for short data sets. In this paper we focus on another type of measures of complexity basing on the method of large-scale dimension densities (LASDID) [20] 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.

The paper is organized as follows: First, we give a short overview of the method of large-scale dimension densities. Next, we describe the data used for this study. Then, we apply this technique to HRV data and show the ability to distinguish between real and simulated data. Finally, we analyse HRV data of AF patients and CHF patients in comparison to healthy persons.

METHOD OF LARGE-SCALE DIMENSION DENSITIES (LASDID)

LASDID [20] 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 [21], 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 $C(r, m)$ by

$$C(r, m) = \frac{1}{N(N-1)} \sum_{i \neq j} \theta(r - |\vec{x}(t_i) - \vec{x}(t_j)|) \tag{1}$$

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

$$D_2 = \lim_{r \to 0} \lim_{m \to \infty} (\mathrm{d} \log C(r, m) / (-\mathrm{d} \log(r))), \tag{2}$$

if this limit exists [21]. 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. (2) 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 [21]. For larger values of r, $D_2(r, m)$ is decreasing because of boundary effects, for small distances the dimension is fluctuating rather irregularly due to the finite amount of data. 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 [10, 20, 22]. 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.

We have recently introduced the large-scale dimension density $\rho_{\mathrm{ls}}(r, m)$ [20] which 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:

$$\rho_{\mathrm{ls}}(r, m) = D_2(r, m) / (m D_2(r, 1)). \tag{3}$$

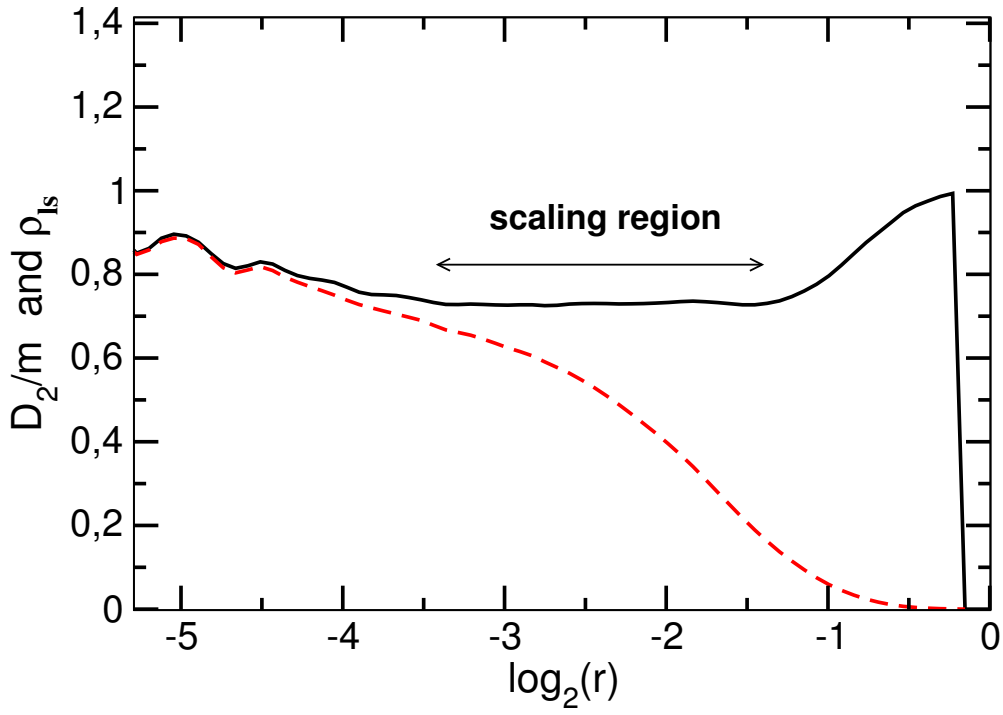

FIG. 1: Comparison of LASDID (solid line) with the Grassberger-Procaccia algorithm(dashed line) calculated for HRV data. With LASDID we get a plateau for scales between $1/2$ and $1/10$ of the attractor diameter, corresponding to $log_2(r) = -1$ to -3.4. For the calculation we used only 2000 RR-intervals, which is not enough to find a scaling region with the Grassberger-Procaccia algorithm. The data was embedded with $\tau = 1$ and embedding dimension $m = 4$.

This normalization is the main point of our new approach and leads to a surprisingly well-expressed plateau for large scales r yielding an estimate of $\rho_{\rm ls}$. In fig. 1 the normalized curve is shown and compared with the original Grassberger-Procaccia algorithm. The large scaling region of the normalized curve enables to estimate a reliable value of $\rho_{\rm ls}$ by averaging all values of this region.

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. To reduce very large RR-intervals which sometimes occur because of measurement errors in the unfiltered data which we use in the second part of the paper, it is necessary to transform the data to a gauss distribution. Then, for every of this short and transformed pieces, we calculate the large-scale dimension density

$\rho_{\mathrm{ls}}(r,m)$ via the basic eq. 3 and estimate ρ_{ls} from the plateau. This leads to a time series of $\rho_{\mathrm{ls}}(t)$. For this time series we calculate further measures of complexity: the mean value $\mu_{\rho_{\mathrm{ls}}}$ by

$$\mu_{\rho_{\mathrm{ls}}} = \frac{1}{M}\sum_{i=1}^{M}\rho_{\mathrm{ls}}(t_i), \tag{4}$$

the standard deviation $\sigma_{\rho_{\mathrm{ls}}}$ by

$$\sigma_{\rho_{\mathrm{ls}}} = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(\rho_{\mathrm{ls}}(t_i) - \mu_{\rho_{\mathrm{ls}}})^2} \tag{5}$$

and the coefficient of variation $cv_{\rho_{\mathrm{ls}}}$ by

$$cv_{\rho_{\mathrm{ls}}} = \sigma_{\rho_{\mathrm{ls}}}/\mu_{\rho_{\mathrm{ls}}}. \tag{6}$$

As shown in [20] the large-scale dimension density is decreasing with increasing embedding dimension m. But in this work our main intention is to compare data of different groups of patients, that means not the absolute value of the dimension density is important but the comparison of them, i.e. here ρ_{ls} and the derived measures of complexity $\mu_{\rho_{\mathrm{ls}}}$, $\sigma_{\rho_{\mathrm{ls}}}$ and $cv_{\rho_{\mathrm{ls}}}$ have to be understood as relative measures. 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 $\tau = 1, ..., 5$. Finally, approximations of the large-scale dimension $m\rho_{\mathrm{ls}}$ and the large-scale dimension density ρ_{ls} are made with embedding dimensions up to $m = 200$. Group summaries are expressed as mean value $\pm$ standard deviation. Statistical analysis was performed via Mann-Whitney U test and Pearsons correlation coefficients where appropriate. In all tests, the criterion for statistical significance is $p < 0.05$.

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 (see fig.2), 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

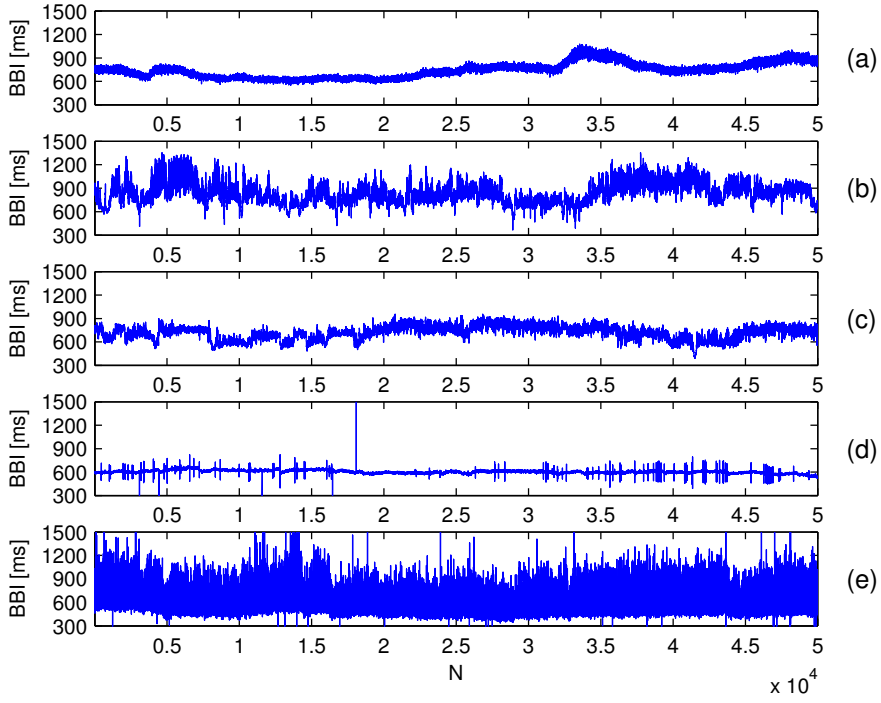

FIG. 2: Representative beat-to-beat intervals (RR-intervals) from simulations (a, time series 34 from Computers in Cardiology challenge 2002), from a young and healthy volunteer (b), from an elderly healthy volunteer (c), from a patient with congestive heart failure (CHF) (d) as well as from a patient with atrial fibrillation (AF) (e).

techniques continue to reveal properties of the time series of RR-intervals. **Hier verstehen wir nicht ganz, wie Sie das mit 'focus on problems' gemeint hatten.** Moreover, the simulation of such time series is still extremely sophisticated and PhysioNet [23] 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 reanalyse 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 study whether these

both types of time series can be discriminated by LASDID parameters.

The second intention of this paper is to demonstrate a possible application for risk stratification of cardiac diseases. Therefore, we analyse the 24 hours HRV 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 original 24 hours ECG recordings were digitized at 128 samples per second with standard Holter devices, and the beat annotations were obtained by automated analysis with manual review and correction. The data of the CHF patients and the young healthy subjects are available from Physionet [23]. We calculate LASDID with the unfiltered data 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 [13, 24, 25]. The following HRV parameters are calculated from the filtered time series [26] [**?**]: 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 analysed by methods of nonlinear dynamics, especially symbolic dynamics [14, 27]: FWSHANNON, the Shannon entropy of the word distribution and POLVAR10, a measure to detect intermittently decreased HRV. At last, we use LASDID to estimate dimensions of HRV data with high embedding dimensions.

RESULTS

Separation of real and simulated data

First we use the method of LASDID to compare time series of real ECG data with those of simulated data. (see fig.2a,b). We subdivide every time series in pieces of an equal amount of heart beats and calculate $\rho_{\mathrm{ls}}(r, m)$ (Eq.3). After estimating the dimension from the plateau at large scales, this leads to a time series with fluctu-

ating values $\rho_{\mathrm{ls}}(t)$ which are analysed by calculating the mean value $\mu_{\rho_{\mathrm{ls}}}$ (Eq.4), the standard deviation $\sigma_{\rho_{\mathrm{ls}}}$ (Eq.5) and the coefficient of variation $cv_{\rho_{\mathrm{ls}}}$ (Eq.6). To find the best length of the short pieces all calculations have been done with different amounts of heart beats. For less than 500 heart beats ρ_{ls} can not be calculated reliably. The region of the plateau becomes to short because the part with the fluctuating values, which usually exists for small scales is shifted to larger scales and cuts off the plateau. For pieces of 1000 heart beats the plateaus are not cut and we get almost the same results as with intervals of 2000 heart beats. But for pieces longer than 2000 heart beats more and more information about the circadian changes gets lost. So the following calculations are done with 1000 heart beats per piece of RR-interval.

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 variability 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. 3).

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 time between 8:00 a.m. and 1:00 p.m. as day interval and the time from 1:00 a.m. to 6:00 a.m. as night interval. For real data we find higher values of $\mu_{\rho_{\mathrm{ls}}}$ for the night for most of the records (day: $\mu_{\rho_{\mathrm{ls}}} = 0.546 \pm 0.056$; night: $\mu_{\rho_{\mathrm{ls}}} = 0.628 \pm 0.069$). According to the Mann-Whitney U-test this difference between day and night is significant (P for day vs. night < 0.001). But only few of the simulated data sets show differences between two different time intervals.

Interestingly, always two datasets of the simulated data have been generated with the same model. These pairs do not differ much in $\mu_{\rho_{\mathrm{ls}}}$ which enables to assign the data with lower $\mu_{\rho_{\mathrm{ls}}}$ to a single model. For data with higher $\mu_{\rho_{\mathrm{ls}}}$ always two models come into question (see fig. 4).

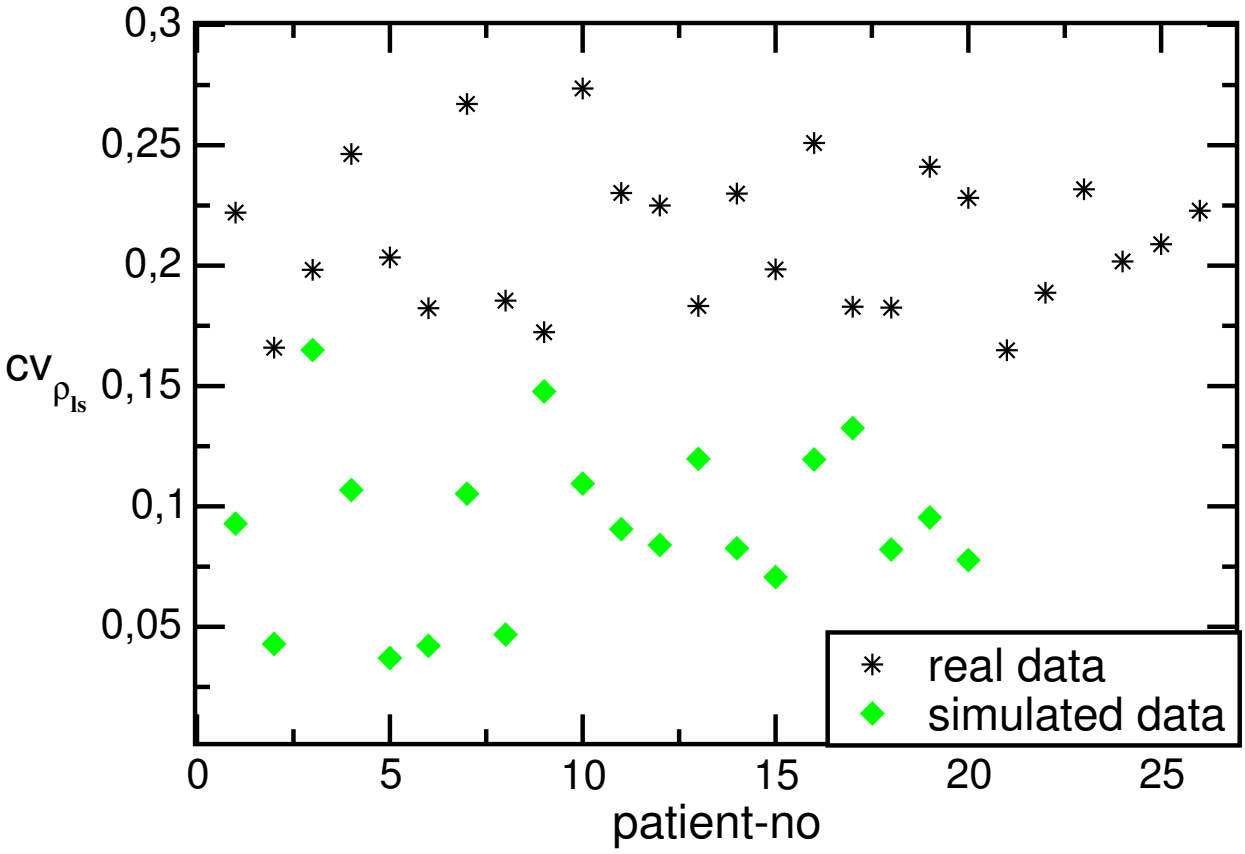

FIG. 3: A comparison of the coefficients of variation $cv_{\rho_{\rm ls}}$ (Eq.6) of real data and simulated data shows higher values for real data.

Risk stratification of cardiac diseases

The second intention of this paper was to demonstrate a possible application for risk stratification of cardiac diseases. Therefore, we compare the data of different pathologies and healthy subjects. For patients with atrial fibrillation (AF) we find values of $\mu_{\rho_{\rm ls}}$ near one which indicates almost stochastic heart beats. The coefficient of variation $cv_{\rho_{\rm ls}}$ for this patients is very low (see fig. 5 and tab. I). This means, the AF-group separates completely from the others. Elderly patients with congestive heart failure (CHF) show higher values of $cv_{\rho_{\rm ls}}$. The highest values we find for elderly healthy patients (EH) (see fig. 5 and tab. I). This means, low values of $cv_{\rho_{\rm ls}}$ indicate a higher risk of heart disease. For the healthy persons we again find higher values of $\mu_{\rho_{\rm ls}}$ for the night, but not in patients with congestive heart failure (see tab. II). Thus, finding no circadian differences in $\mu_{\rho_{\rm ls}}$ is also a pathological sign.

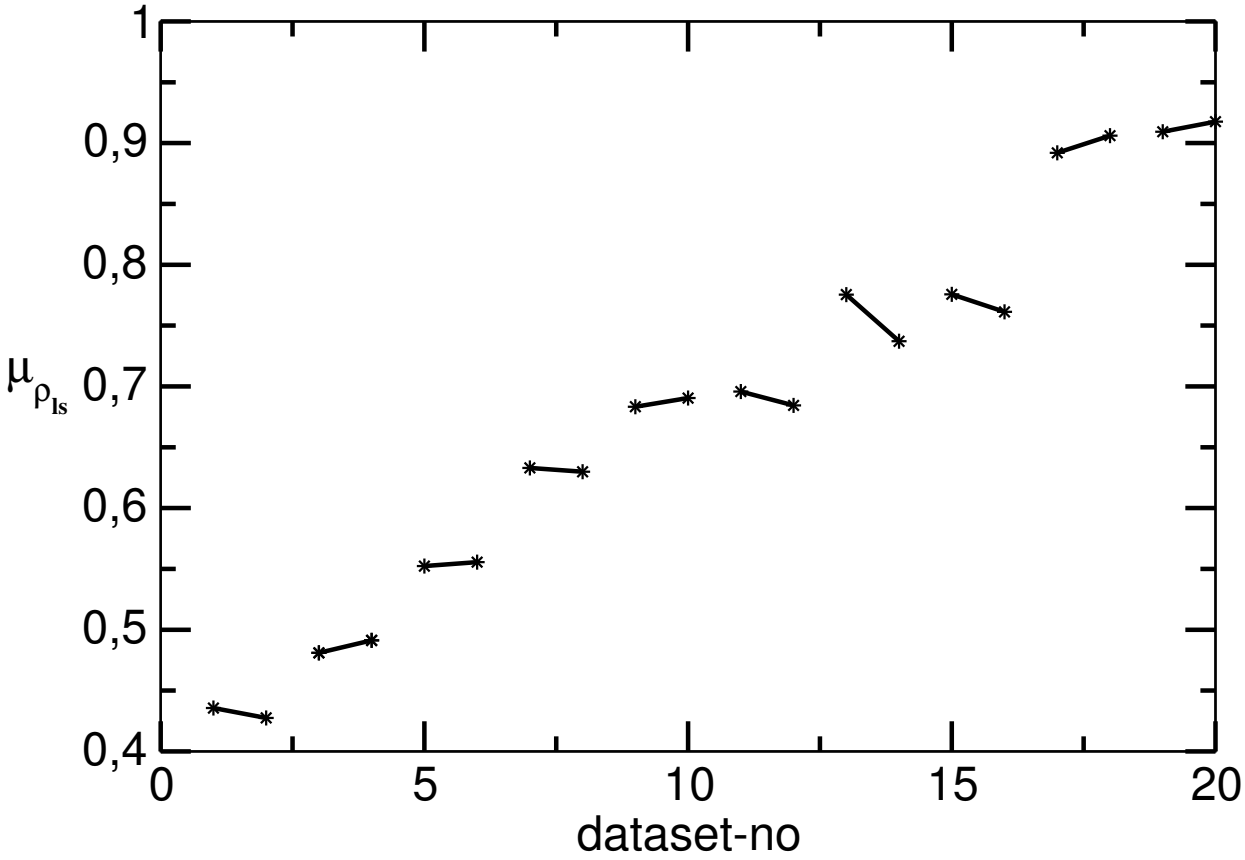

FIG. 4: $\mu_{\rho_{\rm ls}}$ of the simulated data. Always two datasets have been generated with the same model and are connected with lines.

A closer look at the dependency of the mean value of the age of the patients shows that $\mu_{\rho_{\rm ls}}$ is decreasing for increasing age (see fig. 6). This can be interpreted as a decreasing of the number of degrees of freedom in the underlying processes with age. This result agrees with other studies that have found decreasing dimensionality of heart beats with age, described by the group of Goldberger, using detrended fluctuation analysis [28, 29] and by Yoshikawa, calculating Lyapunov dimensions [30].

For standard analyses it is necessary to exclude artefacts and premature beats from the HRV data to make it for instance possible to estimate spectra reliably. To see, how sensitive LASDID is to this filtering, we pre-processed [26] the data and calculated LASDID again. For healthy persons we find almost no differences in $\mu_{\rho_{\rm ls}}$, $\sigma_{\rho_{\rm ls}}$ and $cv_{\rho_{\rm ls}}$. For most of the AF-patients $\mu_{\rho_{\rm ls}}$ is decreasing and $cv_{\rho_{\rm ls}}$ increasing respectively. This means, some of the random processes in the heart beats of AF-patients are filtered out. For most of the CHF-patients we find no differences between filtered

Group	$\mu_{\rho_{\rm ls}}$	$\sigma_{\rho_{\rm ls}}$	$cv_{\rho_{\rm ls}}$
AF	0.968 ± 0.021	0.023 ± 0.012	0.024 ± 0.013
CHF	$0.651 \pm 0.125^{*}$	$0.105 \pm 0.027^{*}$	$0.168 \pm 0.053^{*}$
EH	$0.563 \pm 0.042^{*\diamond}$	$0.120 \pm 0.022^{*}$	$0.209 \pm 0.028^{*\diamond}$
YH	$0.6062 \pm 0.0392^{*\nabla}$	$0.112 \pm 0.016^{*}$	$0.185 \pm 0.021^{*\nabla}$

TABLE I: 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 $\mu_{\rho_{\rm ls}}$ (Eq.4), $\sigma_{\rho_{\rm ls}}$ (Eq.5) and $cv_{\rho_{\rm ls}}$ (Eq.6) (* $p < 0.001$ vs. AF group, $^{\diamond}$ $p < 0.05$ vs. CHF group, $^{\nabla}$ $p < 0.05$ vs. EH group).

Group	$\mu_{\rho_{\rm ls}}$ day	$\mu_{\rho_{\rm ls}}$ night	p day vs. night
EH	0.54 ± 0.05	0.61 ± 0.05	0.002
YH	0.57 ± 0.05	0.67 ± 0.05	< 0.001
CHF	0.65 ± 0.13	0.66 ± 0.12	$n.s.$

TABLE II: Comparison of the day and night values of $\mu_{\rho_{\rm ls}}$ for healthy persons (EH and YH) with patients with congestive heart failure (CHF).

and unfiltered data, but for patient number two, six and fifteen $\mu_{\rho_{\rm ls}}$ is higher and $cv_{\rho_{\rm ls}}$ is lower for the unfiltered data (CHF2: $\mu_{\rho_{\rm ls}} = 0.682$ vs. 0.596; $cv_{\rho_{\rm ls}} = 0.154$ vs. 0.128; CHF6: $\mu_{\rho_{\rm ls}} = 0.803$ vs. 0.720; $cv_{\rho_{\rm ls}} = 0.099$ vs. 0.067; CHF15: $\mu_{\rho_{\rm ls}} = 0.565$ vs. 0.543; $cv_{\rho_{\rm ls}} = 0.204$ vs. 0.191) A closer look at the data shows that these three patients have lots of ventricular premature beats which make filtering almost impossible. Because of that also important HRV information is filtered out by pre-processing and it becomes more difficult to separate the CHF-patients from the healthy persons. But, filtering out ventricular premature beats is not changing the dimensionality of the data. The CHF patients three and eight also have ventricular premature beats, but not as much as the other three patients. Here only the ventricular premature beats are filtered and no differences occur in the results. On the other hand, in the data of patient CHF4 there are lots of errors resulting of technical problems, and they do not influence the unfiltered results. So it is another advantage of LASDID that unfiltered data can be used and a loss of information resulting from pre-processing can be

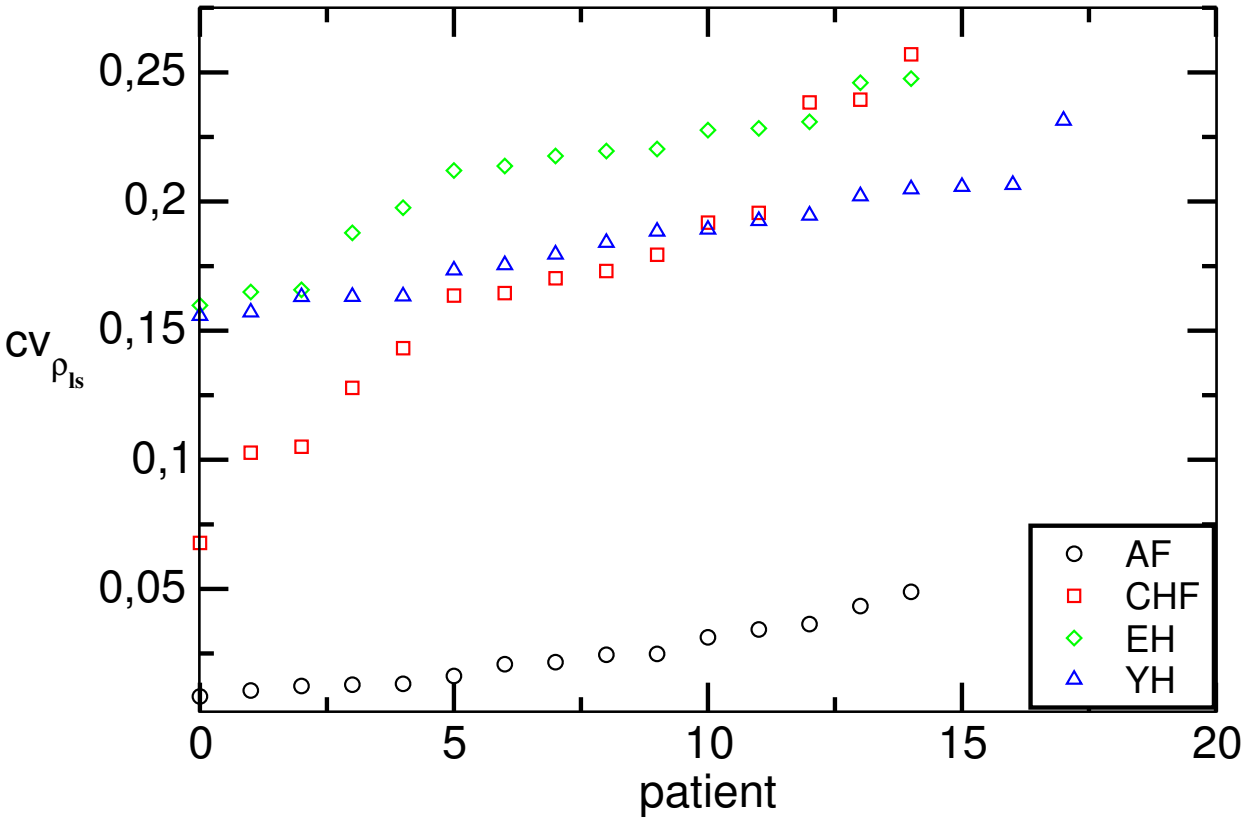

FIG. 5: Comparison of the coefficient of variation $cv_{\rho_{\rm ls}}$ of patients with atrial fibrillation (AF), with congestive heart failure (CHF) and elderly healthy persons (EH).

avoided.

In order to investigate the physiological correlates for LASDID we perform a correlation analysis. Pearson correlation coefficients between different HRV parameters and $\mu_{\rho_{\rm ls}}$, $\sigma_{\rho_{\rm ls}}$ and $cv_{\rho_{\rm ls}}$ are given in table III. Mean heart rate (inversely related to MeanNN) as well as sdNN, the standard deviation of the time series, does not correlate with $\mu_{\rho_{\rm ls}}$ and $cv_{\rho_{\rm ls}}$. For rmssd, the root mean square of successive differences, however, we see a significant relation to $\mu_{\rho_{\rm ls}}$, 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 $\mu_{\rho_{\rm ls}}$, demonstrating that the Mayer waves having the strongest influence for estimating LASDID. Interstingly, $cv_{\rho_{\rm ls}}$ did not show any significant relation to HRV parameters.

To compare LASDID to correlation dimensions of HRV data calculated by others, we also estimated $\rho_{\rm ls}$ for higher embedding dimensions m up to $m = 200$ with $80,000$

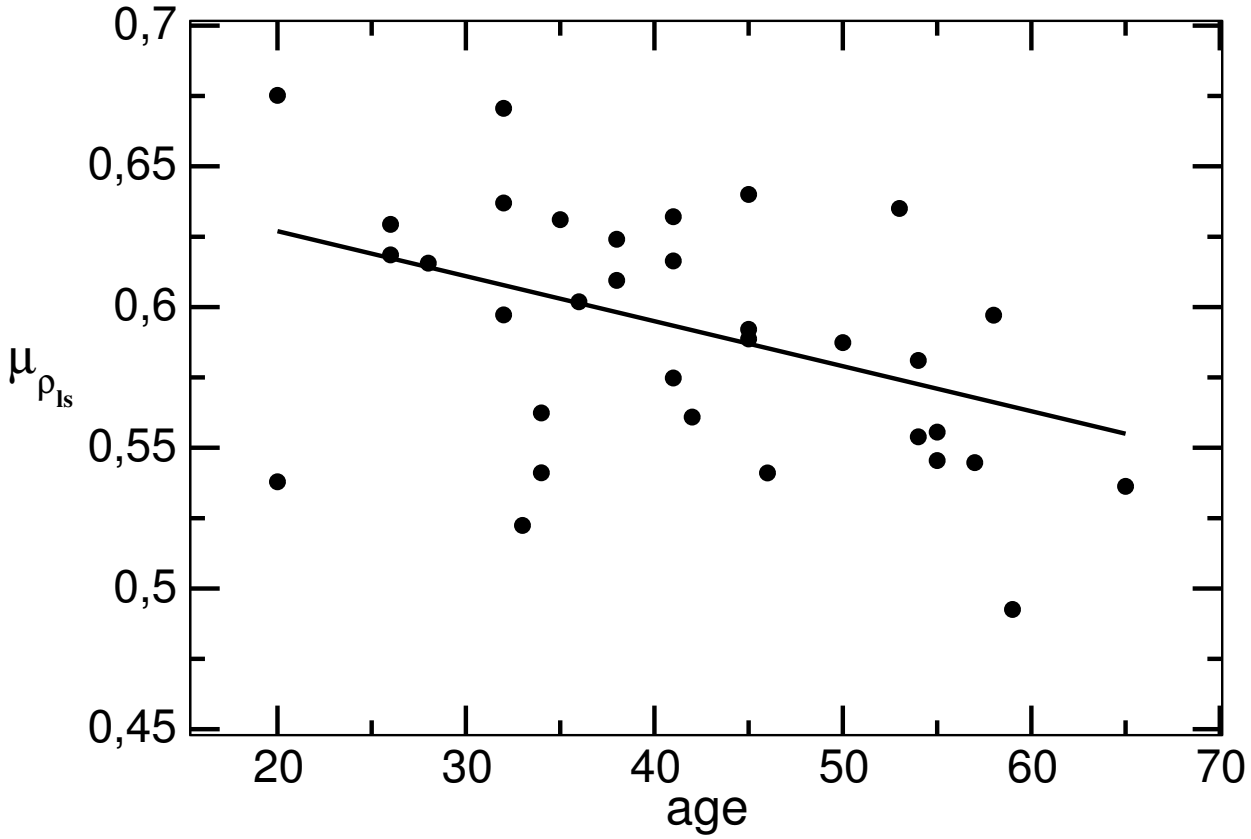

FIG. 6: The mean values $\mu_{\rho_{\rm ls}}$ of the large scale dimension density time series for healthy persons are decreasing with the age of the persons. This Correlation is significant: $r = -0.45$ and $p < 0.01$.

	$\mu_{\rho_{\rm ls}}$	$\sigma_{\rho_{\rm ls}}$	$cv_{\rho_{\rm ls}}$
meanNN	0.053^{+}	0.110^{+}	0.107^{+}
sdNN	0.227^{+}	0.232^{+}	0.152^{+}
rmssd	$0.509^{\diamond}$	0.276^{+}	0.059^{+}
pNN50	$0.510^{\diamond}$	0.267^{+}	0.046^{+}
LF/HF	-0.607^{*}	$-0.453^{\diamond}$	-0.226^{+}
LFn	-0.735^{*}	$-0.461^{\diamond}$	-0.163^{+}
fwshannon	-0.659^{*}	-0.261^{+}	0.037^{+}
polvar10	$-0.553^{\diamond}$	-0.353^{∇}	-0.145^{+}

TABLE III: Correlations coefficients r (p-value) between large scale dimension densities and heart rate variability parameters (* $p < 0.001$, $^{\diamond}$ $p < 0.01$, $^{\nabla}$ $p < 0.05$, $^{+}$ not significant).

heart beats. For the EH group this value is decreasing from $\rho_{\mathrm{ls}} = 0.35$ for $m = 5$ to values between $\rho_{\mathrm{ls}} = 0.018$ to $\rho_{\mathrm{ls}} = 0.056$ for $m = 200$. This corresponds to large-scale dimensions $m\rho_{\mathrm{ls}}$ between 4 and 11 for $m = 200$. This is in accordance with results of the correlation dimension calculated by Carvajal et al. ($D_2 = 7.5 - 10.8$) [31], Babloyantz and Destexhe ($D_2 = 5.5 - 6.3$) [32], Kanters et al. ($D_2 = 9.6 - 10.2$) [33], Govindan et al. ($D_2 = 2.8 - 5.8$) [34] and Guzzetti et al. ($D_2 = 4 - 7$) [35]. The maximal embedding dimension used by them was about $m = 20$. We could calculate with embedding dimensions up to $m = 200$ because LASDID needs less data-points than the Grassberger-Procaccia algorithm. But we did not find an upper limit for $m\rho_{\mathrm{ls}}$, even for $m = 200$ we get increasing results.

CONCLUSIONS

In this paper, we showed that our new method of LASDID can be used to analyse very short, instationary and unfiltered data.

Firstly, we presented a way of discriminating the 46 simulated and physiological HRV time series from the 2002 Computers in Cardiology challenge [36] using only one parameter. Next, we demonstrated its potentials for risk stratification of cardiac diseases. Patients with atrial fibrillation showed averaged large-scale dimension densities near to one and can be completely discriminated from the other groups. A dimension density near one means, that atrial fibrillation leads to a broad range of random heart beats. The comparison of the results of LASDID for filtered and unfiltered data showed, that for this group filtering is senseless, because too many heart beats are exclude due to their randomness. For the CHF group filtering also sometimes destroys important HRV information, as shown for the patients with ventricular premature beats. In addition, we showed that the group of the elderly healthy subjects is statistically different in $\mu_{\rho_{\mathrm{ls}}}$ 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 $\mu_{\rho_{\mathrm{ls}}}$ decreases too (see YH vs. EH in Tab.I and fig. 6). In the CHF group $\mu_{\rho_{\mathrm{ls}}}$ 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 $cv_{\rho_{\mathrm{ls}}}$ the same phenomena can be detected: Patients with AF persisting over 24h do not show circadian complexity changes, the young healthy group is inbetween the CHF and the elderly healthy group.

Compared to other methods of analysing RR-intervals of HRV data we only needed one parameter to separate the simulated and physiological HRV time series from the 2002 Computers in Cardiology challenge. In a previous paper [25] we used three different parameter for this, we quantified 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 chosen too rigid - one can easily detect this time series as an artificial one from its frequency spectrum. The averaged LASDID $\mu_{\rho_{\mathrm{ls}}}$, 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 $cv_{\rho_{\mathrm{ls}}}$, 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.

Finally, looking at the correlation of LASDID to standard HRV parameters and finding no statistical significant relation for $cv_{\rho_{\mathrm{ls}}}$ demonstrates the independence of our approach and the fact, that we do not need to filter the data is a big advantage for stratification of cardiac diseases.

Corinna Raab has been funded by the Ministry of Science, Research and Culture

of the Federative State Brandenburg. Jürgen Kurths and Niels Wessel acknowledge financial support by the EU Network of Excellence NoE 005137 BioSim as well as by the by the Deutsche Forschungsgemeinschaft (KU 837/23-1, KU 837/20-1).

---

* corinna@agnld.uni-potsdam.de

[1] C. Alberte and D. P. Zipes, J Cardiovasc Electrophysiol **14**, 87 (2003).

[2] H. V. Barron and M. D. Lesh, J Am Coll Cardiol **27(5)**, 1053 (1996).

[3] R. J. Damiano Jr, J Card Surg **7(1)**, 36 (1992).

[4] L. Glass, Nature **410**, 277 (2001).

[5] C. Schäfer, M. G. Rosenblum, J. Kurths, and H.-H. Abel, Nature **392**, 239 (1998).

[6] K. B. Marvel, Nature **411**, 252 (2001).

[7] A. L. Goldberger, D. R. Rigney, J. Mietus, E. M. Antman, and S. Greenwald, Experientia **44**, 983 (1988).

[8] L. Glass and D. Kaplan, Med. Prog. Technol. **19**, 115 (1993).

[9] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Review of Modern Physics **65**, 1331 (1993).

[10] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[11] R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths, and H. Scheingraber, Chaos, Solitons & Fractals **4**, 133 (1994).

[12] P. E. Rapp, C. J. Cellucci, K. E. Korslund, T. A. Watanabe, and M. A. Jimenez-Montano, Physical Review E **64**, 16209 (2001).

[13] J. Kurths, A. Voss, A. Witt, P. Saparin, H. J. Kleiner, and N. Wessel, Chaos **5**, 88 (1995).

[14] A. Voss, J. Kurths, H. J. Kleiner, A. Witt, N. Wessel, P. Saparin, K. J. Osterziel, R. Schurath, and R. Dietz, Cardiovasc Res **31**, 419 (1996).

[15] N. Wessel, A. Voss, J. Kurths, A. Schirdewan, K. Hnatkova, and M. Malik, Med Biol Eng Comput **38**, 680 (2000).

[16] U. Schwarz, A. O. Benz, J. Kurths, and A. Witt, Astron. Astrophys. **277**, 215 (1993).

[17] A. Witt, J. Kurths, F. Krause, and K. Fischer, Geoph. Astroph. Fluid Dyn. **77**, 79 (1994).

[18] A. Hempelmann and J. Kurths, Astron. Astrophys. **232**, 356 (1990).

[19] J. Skinner, M. Molnar, and C. Tomberg, Integr Physiol Behav Sci. **29**, 217 (1994).

[20] C. Raab and J. Kurths, Physical Review E **64** (2001).

[21] P. Grassberger and I. Procaccia, Physical Review Letters **50**, 346 (1983).

[22] L. A. Smith, Physics Letters A **133**, 283 (1988).

[23] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, Circulation **101**, E215 (2000).

[24] N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, and A. Voss, Phys. Rev. E **61**, 733 (2000).

[25] N. Wessel, H. Malberg, U. Mayerfeldt, A. Schirdewan, and J. Kurths, Computers in Cardiology **29**, 133 (2002).

[26] N. Wessel, A. Voss, H. Malberg, C. Ziehmann, H. Voss, A. Schirdewan, U. Meyerfeldt, and J. Kurths, Herzschr. Elektrophys. **11**, 159 (2000).

[27] N. Wessel, A. Schirdewan, and J. Kurths, Phys Rev Lett **91**, 119801 (2003).

[28] A. Goldberger, L. Amaral, J. Hausdorff, P. Ivanov, C.-K. Peng, and H. Stanley, Journal of American College of Cardiology **24**, 1700 (1994).

[29] N. Iyengar, C. Peng, R. Morin, A. Goldberger, and L. Lipsitz, Am J Physiol **271**, 1078 (1996).

[30] Y. Yoshikawa and Y. Yasuda, Bulletin of Toyohashi Sozo College **7**, 63 (2003).

[31] R. Carvajal, N. Wessel, M. Vallverdu, P. Caminal, and A. Voss, Computer Methods and Programs in medicine **78**, 133 (2005).

[32] A. Babloyantz and A. Destexhe, Biol. Cybern. **58**, 203 (1988).

[33] J. K. Kanters, N. H. Holstein-Rathlou, and E. Agner, J. Cardivasc. Electrophysiol. **5**, 591 (1994).

[34] R. Govindan, K. Narayanan, and M. Gopinathan, Chaos **8**, 495 (1998).

[35] S. Guzzetti, M. G. Signorini, C. Cogliati, S. Mezzetti, A. Porta, S. Cerutti, and A. Malliani, Cardiovasc. Res. **31**, 441 (1996).

[36] http://www.physionet.org/challenge/2002/, Computers in Cardiology **29** (2002).

[] An implementation of this adaptive filtering procedure, which considers the instantaneous variability, is available from tocsy.agnld.uni-potsdam.de ¡http://tocsy.agnld.uni-potsdam.de¿.