Large-scale dimension densities for heart rate variability analysis

Corinna Raab,[1,*] Niels Wessel,[1,2] Alexander Schirdewan,[2] and Jürgen Kurths[1]
[1]*Center for Dynamics of Complex Systems, Institute of Physics, University of Potsdam, Potsdam, Germany*
[2]*Franz-Volhard-Clinic, HELIOS Kliniken Berlin, Medical Faculty of the Charité, Humboldt University, Berlin, Germany*
(Received 22 December 2005; published 10 April 2006)

In this work, we reanalyze the heart rate variability (HRV) data from the 2002 Computers in Cardiology (CiC) Challenge using the concept of large-scale dimension densities and additionally apply this technique to data of healthy persons and of patients with cardiac diseases. The large-scale dimension density (LASDID) 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. This way, it is possible to analyze rather short, nonstationary, and unfiltered data, such as HRV. Moreover, this method allows us to analyze 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 the CiC 2002 challenge using only one parameter. In the second part we analyzed 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 ($\rho_{\mathrm{ls}}^{\mu}=0.97\pm0.02$) group from the others and, especially during daytime, the CHF patients show significant differences from 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 ρ_{ls}^{μ} (day, 0.65 ± 0.13; night, 0.66 ± 0.12; n.s.) in contrast to healthy controls (day, 0.54 ± 0.05; night, 0.61 ± 0.05; $p=0.002$). 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.

DOI: 10.1103/PhysRevE.73.041907 PACS number(s): 87.19.Hh, 05.45.−a

I. 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 nonstationary) 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, wavelet analysis, or renormalized entropy [11–13]. They are mostly based on symbolic dynamics and are efficient quantities to characterize measurements of natural systems, such as in cardiology [14–16], cognitive psychology, or astrophysics [17–19]. These methods are often not sufficient for very short data sets. For short data sets the method of point correlations has been introduced [20], 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 based on the method of large-scale dimension densities (LASDID) [21] and apply this methodology to HRV data. The LASDID method allows one to analyze 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 in 24 h.

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 analyze HRV data of atrial fibrillation (AF) patients and congestive heart failure (CHF) patients in comparison to healthy persons [22].

II. METHOD OF LARGE-SCALE DIMENSION DENSITIES

The LASDID [21] 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 analyze rather short and nonstationary data sets.

To calculate the correlation dimension D_2 of a system with the Grassberger-Procaccia algorithm [23] means that the attractor first has to be reconstructed by embedding. The em-

*Electronic address: corinna@agnld.uni-potsdam.de

bedded time series consists of vectors $\{\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_m(t))\}$, 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 which neighboring points are counted for the correlation sum. D_2 is then defined as

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

if this limit exists [23]. 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 distance-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 [23]. 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 [10,21,24]. If the time series is too short, one only gets the part of $D_2(r,m)$ with decreasing values. With the LASDID method we are able to use this part of $D_2(r,m)$ too.

We have recently introduced the large-scale dimension density $\rho_{\mathrm{ls}}(r,m)$ [21] 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:

$$\rho_{\mathrm{ls}}(r,m) = D_2(r,m)/[mD_2(r,1)]. \tag{3}$$

This normalization is the main point of our approach and leads to a surprisingly well-expressed plateau for large scales r yielding an estimate of ρ_{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 one to estimate a reliable value of ρ_{ls} by averaging all values of this region.

The advantage of the LASDID method is that it is possible to estimate it from rather short and nonstationary time series. So we can cut every RR interval time series into M shorter pieces. The minimum length of these pieces will be discussed later. To reduce very large RR intervals which sometimes occur because of measurement errors in the unfiltered data that we use in the second part of the paper, it is necessary to transform the data to a Gaussian distribution. Then, for every one of these 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 ρ_{ls}^{μ} by

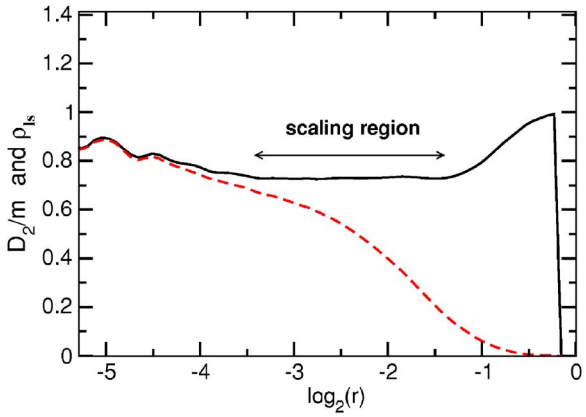

FIG. 1. (Color online) Comparison of LASDID results (solid line) with the Grassberger-Procaccia algorithm (dashed line) calculated for HRV data (see section 'data' and Fig. 2). With LASDID we get a plateau for scales between 1/2 and 1/10 of the attractor diameter, corresponding to $\log_2(r) = -1$ to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

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

the standard deviation $\rho_{\mathrm{ls}}^{\sigma}$ by

$$\rho_{\mathrm{ls}}^{\sigma} = \sqrt{\frac{1}{M-1} \sum_{i=1}^{M} [\rho_{\mathrm{ls}}(t_i) - \rho_{\mathrm{ls}}^{\mu}]^2}, \tag{5}$$

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

$$\rho_{\mathrm{ls}}^{\mathrm{cv}} = \rho_{\mathrm{ls}}^{\sigma} / \rho_{\mathrm{ls}}^{\mu}. \tag{6}$$

As shown in [21] 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, which means that 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 ρ_{ls}^{μ}, $\rho_{\mathrm{ls}}^{\sigma}$, and $\rho_{\mathrm{ls}}^{\mathrm{cv}}$ have to be understood as relative measures. For the calculation of the 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 $m = 4, \ldots, 8$ and delay times $\tau = 1, \ldots, 5$. 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 the Mann-Whitney U test and Pearson correlation coefficients where appropriate. In all tests, the criterion for statistical significance is $p < 0.05$.

III. 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 analyzing in

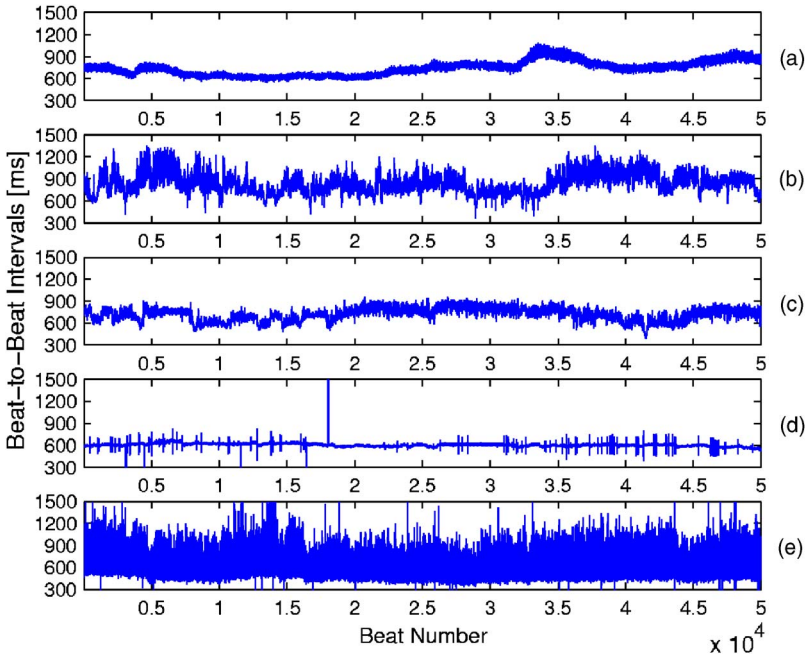

FIG. 2. (Color online) Representative beat-to-beat intervals (RR intervals) from simulations (time series 34 from Computers in Cardiology challenge 2002) (a), from a young and healthy volunteer (b), from an elderly healthy volunteer (c), from a patient with congestive heart failure (CHF) (d), as well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

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, we are far from understanding this system and 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 [25] 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 h 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 was 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 study whether these 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 analyze the 24 h HRV data of 15 patients with atrial fibrillation (15 male, age 67±12), of 15 patients with congestive heart failure (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 h 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 [25]. We calculate LASDID results 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 [14,26,27]. The following HRV parameters are calculated from the filtered time series [28,39]: 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 normalized low-frequency (LFN), the ratio LF to HF, is estimated. Finally, HRV is analyzed by methods of nonlinear dynamics, especially symbolic dynamics [15,29]: FWSHANNON, the Shannon entropy of the word distribution, and POLVAR10, a measure to detect intermittently decreased HRV. Finally, we use LASDID methods to estimate dimensions of HRV data with high embedding dimensions.

IV. RESULTS

A. 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 Figs. 2(a) and 2(b)]. We subdivide every time series into 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 fluctuating values $\rho_{\mathrm{ls}}(t)$ which are analyzed by calculating the mean value ρ_{ls}^{μ} [Eq. (4)], the standard deviation $\rho_{\mathrm{ls}}^{\sigma}$ [Eq. (5)], and the coefficient of variation $\rho_{\mathrm{ls}}^{\mathrm{cv}}$ [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} cannot be calculated reliably. The region of the plateau becomes too 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 informa-

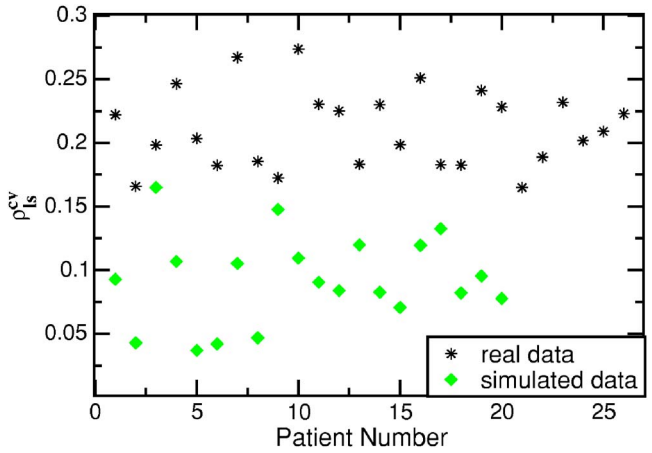

FIG. 3. (Color online) A comparison of the coefficients of variation ρ_{ls}^{cv} [Eq. (6)] of real data and simulated data shows higher values for real data.

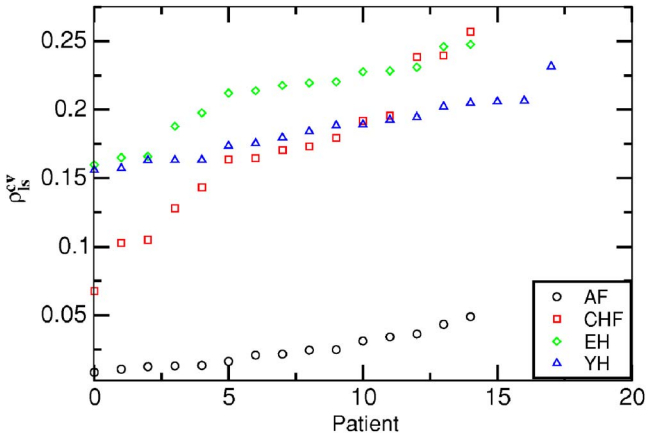

FIG. 5. (Color online) Comparison of the coefficient of variation ρ_{ls}^{cv} of patients with atrial fibrillation (AF), with congestive heart failure (CHF), and elderly healthy persons (EH).

tion 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 ρ_{ls}^{μ} between 0.5 and 0.7, whereas the simulated data range between 0.4 and 0.9; only half of the models generated data which also range between 0.5 and 0.7. Values near one indicate a rather stochastic behavior of the heart rate; values near zero mean deterministic heart beats. Furthermore, real data show stronger fluctuations in the time series of LASDID, i.e., the values of ρ_{ls}^{σ} are higher for real data (ρ_{ls}^{σ} from 0.09 to 0.17 for real data against ρ_{ls}^{σ} 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 ρ_{ls}^{cv}. 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 use the time between 8:00 a.m. and 1:00 p.m. as the day interval and the time from 1:00 a.m. to 6:00 a.m. as the night interval. We use intervals of only 5 h length to minimize the risk of test persons having a nap during the day interval or being awake during the night interval. For real data we find higher values of ρ_{ls}^{μ} for the night for most of the records (day, $\rho_{ls}^{\mu}=0.546\pm0.056$; night, $\rho_{ls}^{\mu}=0.628\pm0.069$). According to the Mann-Whitney U test this difference between day and night is significant (P for day vs night below 0.001). But only a few of the simulated data sets show differences between two different time intervals.

Interestingly, always two data sets of the simulated data have been generated with the same model. These pairs do not differ much in ρ_{ls}^{μ} which enables us to assign the data with lower ρ_{ls}^{μ} to a single model. For data with higher ρ_{ls}^{μ} always two models come into question (see Fig. 4).

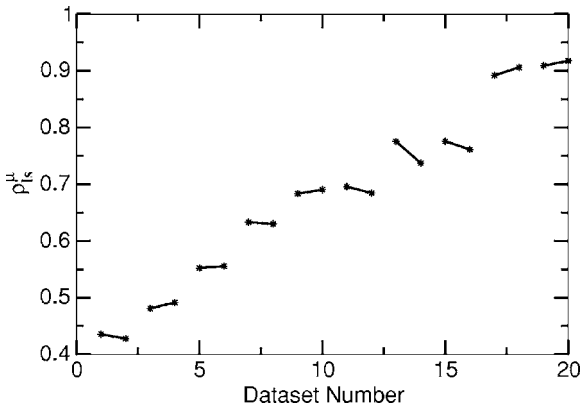

FIG. 4. ρ_{ls}^{μ} of the simulated data. Always two data sets have been generated with the same model and are connected with lines.

B. Risk stratification of cardiac diseases

The second intention of this paper was to demonstrate a possible application of LASDID for risk stratification of cardiac diseases. Therefore, we compare the data of different pathologies and healthy subjects. For patients with atrial fibrillation we find values of ρ_{ls}^{μ} near 1, which indicates almost stochastic heart beats. The coefficient of variation ρ_{ls}^{cv} for these patients is very low (see Fig. 5 and Table I). This means that the AF group separates completely from the others. Elderly patients with congestive heart failure show higher values of ρ_{ls}^{cv}. The highest values we find for elderly

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 ρ_{ls}^{μ} [Eq. (4)], ρ_{ls}^{σ} [Eq. (5)], and ρ_{ls}^{cv} [Eq. (6)].

Group	ρ_{ls}^{μ}	ρ_{ls}^{σ}	ρ_{ls}^{cv}
AF	0.968 ± 0.021	0.023 ± 0.012	0.024 ± 0.013
CHF	0.651 ± 0.125[a]	0.105 ± 0.027[a]	0.168 ± 0.053[a]
EH	0.563 ± 0.042[a,b]	0.120 ± 0.022[a]	0.209 ± 0.028[a,b]
YH	0.606 ± 0.039[a,c]	0.112 ± 0.016[a]	0.185 ± 0.021[a,c]

[a] $p<0.001$ vs AF group.
[b] $p<0.05$ vs CHF group.
[c] $p<0.05$ vs EH group.

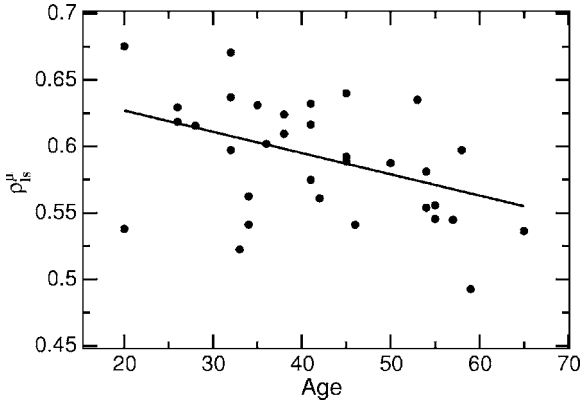

FIG. 6. The mean values ρ_{ls}^{μ} of the large-scale dimension density time series for healthy persons are decreasing with increasing age of the persons. This correlation is significant: $r=-0.45$ and $p<0.01$.

healthy patients (EH) (see Fig. 5 and Table I). This means that for elderly persons low values of ρ_{ls}^{cv} indicate a higher risk of heart disease. For young and healthy persons this value also is low, but not because of heart disease. For the young and healthy heart the number of degrees of freedom is larger than for elderly hearts. The degrees of freedom correspond to ρ_{ls}^{μ}, i.e., ρ_{ls}^{μ} is decreasing for increasing age (see Fig. 6). This result agrees with other studies that have found decreasing dimensionality of heart beats with age, described by Goldberger and co-workers, using detrended fluctuation analysis [30,31] and by Yoshikawa, calculating Lyapunov dimensions [32]. ρ_{ls}^{μ} is inversely proportional to ρ_{ls}^{cv}, which means young persons have lower values of ρ_{ls}^{cv} than elderly persons. On the other hand, CHF patients also have more degrees of freedom than elderly healthy persons because of the disease and CHF patients and young healthy persons cannot be distinguished by ρ_{ls}^{cv} and ρ_{ls}^{μ}, if one analyzes 24 h data only. But the processes of regulation in the young and healthy heart and in the hearts of CHF patients are different. The variability of the healthy heart follows a circadian rhythm which leads to higher values of ρ_{ls}^{μ} for the night for the healthy persons (old and young). For CHF patients this rhythm is disturbed so that the values of ρ_{ls}^{μ} for day and night are the same (see Table II). Table II also shows that there is a significant difference in ρ_{ls}^{μ} for the day between CHF patients and elderly or young healthy persons. For risk stratification of cardiac diseases it is necessary to take the comparison of the different daytimes. Thus, finding no circadian differences in ρ_{ls}^{μ} is also a pathological sign.

For standard analyses it is necessary to exclude artifacts and premature beats from the HRV data to make it for instance possible to estimate spectra reliably. To see, how sensitive the LASDID is to this filtering, we preprocessed [28] the data and calculated the LASDID again. For healthy persons we find almost no differences in ρ_{ls}^{μ}, ρ_{ls}^{σ}, and ρ_{ls}^{cv}. For most of the AF patients ρ_{ls}^{μ} is decreasing and ρ_{ls}^{cv} increasing, respectively. This means that 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 and unfiltered data, but for patient 2, 6, and 15 ρ_{ls}^{μ} is higher and ρ_{ls}^{cv} is lower for the unfiltered data (CHF2, $\rho_{ls}^{\mu}=0.682$ vs 0.596; $\rho_{ls}^{cv}=0.154$ vs 0.128; CHF6, $\rho_{ls}^{\mu}=0.803$ vs 0.720; $\rho_{ls}^{cv}=0.099$ vs 0.067; CHF15, $\rho_{ls}^{\mu}=0.565$ vs 0.543; $\rho_{ls}^{cv}=0.204$ 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 preprocessing 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 3 and 8 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 from 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 preprocessing can be avoided.

In order to investigate the physiological correlates for the LASDID results we perform a correlation analysis. Pearson correlation coefficients between different HRV parameters and ρ_{ls}^{μ}, ρ_{ls}^{σ}, and ρ_{ls}^{cv} are given in Table III. Mean heart rate (inversely related to MeanNN) as well as sdNN, the standard deviation of the time series, do not correlate with ρ_{ls}^{μ} and ρ_{ls}^{cv}. For rmssd, the root mean square of successive differences, however, we see a significant relation to ρ_{ls}^{μ}, i.e., short-term respiratory induced oscillation in HRV plays an important role for LASDID. The highest correlation we find is for the normalized low-frequency band around 0.1 Hz to ρ_{ls}^{μ}, demonstrating that the Mayer waves having the strongest influence for estimating LASDID. Interestingly, ρ_{ls}^{cv} did not show any significant relation to HRV parameters.

TABLE II. Comparison of the day and night values of ρ_{ls}^{μ} for healthy persons (EH and YH) with patients with congestive heart failure (CHF).

Group	ρ_{ls}^{μ} day	ρ_{ls}^{μ} night	p (day vs night)
EH	0.54 ± 0.05[a]	0.61 ± 0.05	0.002
YH	0.57 ± 0.05[a]	0.67 ± 0.05	<0.001
CHF	0.65 ± 0.13	0.66 ± 0.12	n.s.

[a]$p<0.05$ vs CHF group.

TABLE III. Correlation coefficients r (p value) between large-scale dimension densities and heart rate variability parameters.

	ρ_{ls}^{μ}	ρ_{ls}^{σ}	ρ_{ls}^{cv}
meanNN	0.053[a]	0.110[a]	0.107[a]
sdNN	2.227[a]	0.232[a]	0.152[a]
rmssd	0.509[b]	0.276[a]	0.059[a]
pNN50	0.510[b]	0.267[a]	0.046[a]
LF/HF	−0.607[c]	−0.453[b]	−0.226[a]
LFn	−0.735[c]	−0.461[b]	−0.163[a]
fwshannon	−0.659[c]	−0.261[a]	0.037[a]
polvar10	−0.553[c]	−0.353[d]	−0.145[a]

[a]Not significant.
[b]$p<0.01$.
[c]$p<0.001$.
[d]$p<0.05$.

To compare LASDID to correlation dimensions of HRV data calculated by others, we also estimated ρ_{ls} for higher embedding dimensions m up to $m=200$ with 80 000 heart beats. For the EH group this value is decreasing from $\rho_{ls}=0.35$ for $m=5$ to values between $\rho_{ls}=0.018$ and $\rho_{ls}=0.056$ for $m=200$. This corresponds to large-scale dimensions $m\rho_{ls}$ between 4 and 11 for $m=200$. This is in accordance with results of the correlation dimension calculated by Carvajal *et al.* ($D_2=7.5-10.8$) [33], Babloyantz and Destexhe ($D_2=5.5-6.3$) [34], Kanters *et al.* ($D_2=9.6-10.2$) [35], Govindan *et al.* ($D_2=2.8-5.8$) [36], and Guzzetti *et al.* ($D_2=4-7$) [37]. The maximal embedding dimension used by them was about $m=20$. We could calculate with embedding dimensions up to $m=200$ because LASDID needs fewer data points than the Grassberger-Procaccia algorithm. But we did not find an upper limit for $m\rho_{ls}$, even for $m=200$ we get increasing results. As is well known, this could have various reasons, e.g., stochastic influences, very high-dimensionality, etc., which are impossible to identify from real data. But this is not the topic of this paper.

V. CONCLUSIONS

In this paper, we have shown that our method of LASDID can be used to analyze very short, noinstationary, and unfiltered data.

First, we have presented a way of discriminating the 46 simulated and physiological HRV time series from the 2002 Computers in Cardiology challenge [38] using only *one* parameter. Next, we have 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 1 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 excluded 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 have shown that for the daytime the group of the young and elderly healthy subjects is statistically different in ρ_{ls}^{μ} from the congestive heart failure group. Interestingly, considering the complete 24 h data, only the elderly healthy persons and not the young healthy volunteers are statistically different from the CHF group. This is due to the fact that HRV decreases with age; here the number of modes ρ_{ls}^{μ} decreases too (see YH vs EH in Table I and Fig. 6). In the CHF group ρ_{ls}^{μ} is increased compared to elderly healthy subjects. Hence, the number of independent modes increases due to the disease—possible explanations are ventricular ectopy or pulsus alternans. For the circadian variation of ρ_{ls}^{cv} the same phenomena can be detected: patients with AF persisting over 24 h do not show circadian complexity changes, and the young healthy group is in between the CHF and the elderly healthy group.

Compared to other methods of analyzing 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 [27] 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 cutoffs for these parameters, both time series groups could be discriminated completely. The cutoffs 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. 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 ρ_{ls}^{μ}, 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 ρ_{ls}^{cv}, 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 or food intake. In the night, however, you should not have such effects—but there are also significant differences in HRV dependent on the sleep stages. No simulation in this database was able to model all these variability changes.

Finally, looking at the correlation of LASDID to standard HRV parameters and finding no statistical significant relation for ρ_{ls}^{cv} demonstrates the independence of our approach. Moreover, the fact that we do not need to filter the data improves the applicability for clinical risk stratification.

ACKNOWLEDGMENTS

C.R. has been funded by the Ministry of Science, Research and Culture of the Federative State Brandenburg. J.K. and N.W. acknowledge financial support by the EU Network of Excellence Grant No. NoE 005137 BioSim as well as by the Deutsche Forschungsgemeinschaft (Grants No. KU 837/23-1 and No. KU 837/20-1). The authors thank Annette Witt for helpful discussion.

[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**, 1053 (1996).
[3] R. J. Damiano, Jr., J. Card. Surg. **7**, 36 (1992).
[4] L. Glass, Nature (London) **410**, 277 (2001).
[5] K. B. Marvel, Nature (London) **411**, 252 (2001).
[6] C. Schäfer, M. G. Rosenblum, J. Kurths, and H.-H. Abel, Nature (London) **392**, 239 (1998).
[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, Rev. Mod. Phys. **65**, 1331 (1993).
[10] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, U.K., 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. A. Watanabe, and M. A. Jimenez-Montano, Phys. Rev. E **64**, 016209 (2001).
[13] M. Lotric, A. Stefanovska, D. Stajer, and V. Urbancic-Rovan, Physiol. Meas **21**, 441 (2000).
[14] J. Kurths, A. Voss, A. Witt, P. Saparin, H. J. Kleiner, and N. Wessel, Chaos **5**, 88 (1995).
[15] 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).
[16] N. Wessel, A. Voss, J. Kurths, A. Schirdewan, K. Hnatkova, and M. Malik, Med. Biol. Eng. Comput. **38**, 680 (2000).
[17] U. Schwarz, A. O. Benz, J. Kurths, and A. Witt, Astron. Astrophys. **277**, 215 (1993).
[18] A. Witt, J. Kurths, F. Krause, and K. Fischer, Geophys. Astrophys. Fluid Dyn. **77**, 79 (1994).
[19] A. Hempelmann and J. Kurths, Astron. Astrophys. **232**, 356 (1990).
[20] J. Skinner, M. Molnar, and C. Tomberg, Integr Physiol. Behav. Sci. **29**, 217 (1994).
[21] C. Raab and J. Kurths, Phys. Rev. E **64**, 016216 (2001).
[22] C. Raab, N. Wessel, A. Schirdewan, and J. Kurths, Comput. Cardiol. **32**, 985 (2005). With this work we won the Computers in Cardiology 2005 challenge.
[23] P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983).
[24] L. A. Smith, Phys. Lett. A **133**, 283 (1988).
[25] 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).
[26] N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, and A. Voss, Phys. Rev. E **61**, 733 (2000).
[27] N. Wessel, H. Malberg, U. Meyerfeldt, A. Schirdewan, and J. Kurths, Comput. Cardiol. **29**, 133 (2002).
[28] N. Wessel, A. Voss, H. Malberg, C. Ziehmann, H. Voss, A. Schirdewan, U. Meyerfeldt, and J. Kurths, Herzschr. Elektrophys. **11**, 159 (2000).
[29] N. Wessel, A. Schirdewan, and J. Kurths, Phys. Rev. Lett. **91**, 119801 (2003).
[30] N. Iyengar, C. Peng, R. Morin, A. Goldberger, and L. Lipsitz, Am. J. Physiol. **271**, 1078 (1996).
[31] A. Goldberger, L. Amaral, J. Hausdorff, P. Ivanov, C.-K. Peng, and H. Stanley, J. Am. Coll. Cardiol. **24**, 1700 (1994).
[32] Y. Yoshikawa and Y. Yasuda, Bull. Toyohashi Sozo College **7**, 63 (2003).
[33] R. Carvajal, N. Wessel, M. Vallverdu, P. Caminal, and A. Voss, Comput. Methods Programs Biomed. **78**, 133 (2005).
[34] A. Babloyantz and A. Destexhe, Biol. Cybern. **58**, 203 (1988).
[35] J. K. Kanters, N. H. Holstein-Rathlou, and E. Agner, J. Cardiovasc. Electrophysiol. **5**, 591 (1994).
[36] R. Govindan, K. Narayanan, and M. Gopinathan, Chaos **8**, 495 (1998).
[37] S. Guzzetti, M. G. Signorini, C. Cogliati, S. Mezzetti, A. Porta, S. Cerutti, and A. Malliani, Cardiovasc. Res. **31**, 441 (1996).
[38] http://www.physionet.org/challenge/2002/
[39] An implementation of this adaptive filtering procedure, which considers the instantaneous variability, is available from http://tocsy.agnld.uni-potsdam.de