

Detection of time-delayed interactions in biosignals using symbolic coupling traces

N. WESSEL^{1(a)}, A. SUHRBIER^{2(a)}, M. RIEDL¹, N. MARWAN³, H. MALBERG², G. BRETTHAUER², T. PENZEL⁴ and J. KURTHS^{1,3}

¹ *Department of Physics, Humboldt-Universität zu Berlin - Berlin, Germany, EU*

² *Institute for Applied Computer Science, Forschungszentrum Karlsruhe GmbH - Karlsruhe, Germany, EU*

³ *Potsdam Institute for Climate Impact Research - Potsdam, Germany, EU*

⁴ *Department of Sleep Medicine, Charité Universitätsmedizin Berlin - Berlin, Germany, EU*

received 12 February 2009; accepted in final form 16 June 2009

published online 17 July 2009

PACS 05.45.-a – Nonlinear dynamics and chaos

PACS 05.45.Tp – Time series analysis

PACS 87.19.Hh – Cardiac dynamics

Abstract – Directional coupling analysis of bivariate time series is an important subject of current research. In this letter, a method based on symbolic dynamics for the detection of time-delayed coupling is presented. The symbolic coupling traces, defined as the symmetric and diametric traces of the bivariate word distribution, allow for the quantification of coupling and are compared with established methods like mutual information and cross recurrence analysis. The symbolic coupling traces method is applied to model systems and cardiological data which demonstrate its advantages especially for nonstationary data.

Copyright © EPLA, 2009

Biological systems usually consist of several subsystems which are interrelated by feedbacks with time delay. To reveal such time-delayed coupling directions from biosignals is a basic task in understanding such systems [1–8]. Data recorded from these systems reflect biological activities of living beings and are characterized on the one hand by real biological information, including nonstationarities, nonlinearities and intrinsic noise, and on the other hand by measurement noise. Therefore, the analysis of biosignals, especially the detection of coupling directions is complicated. Different methods, starting from cross correlation via mutual predictability to information-theoretic approaches [9–18] were applied to biosignals. All these methods are able to find directions of interactions. However, due to the nonstationarity and nonlinearity of the biosignals, the conclusions are not homogenous.

Recently new methods based on order pattern analysis appear to circumvent these problems [19–21]. Order patterns result from a coarse graining (symbolization) of the data into two states: increasing or decreasing amplitudes. This symbolic representation of successive amplitudes is not sensitive to nonstationarities. Figure 1 gives one example showing the potentials of order

patterns: the linear cross correlation analysis obviously is not so applicable as the recurrence quantification analysis with order patterns in the bivariate data set of heart rate (BBI: beat-to-beat interval) and systolic blood pressure (SBP), because it detects too many lags which cannot be interpreted physiologically [22]. From the physiology it is expected that SBP drives BBI by a current and/or a delayed intervention which is caused by two opposite directed cardiovascular regulation mechanisms. On the other hand BBI influences the stroke volume of the heart which tunes SBP. Thus, the order pattern approach reveals these lags τ of the time series more clearly (fig. 1), suggesting the idea that the ordinal structure of nonlinear and nonstationary time series is necessary for the analysis of the dynamics. Here we report an extension of bivariate symbolic dynamics [23] that greatly improves the detection of coupling directions in biosignals. Signals from coupled biological processes tend to move in the same direction or in opposing directions. We show that this type of relationship is reliably indicated by the symmetric and diametric bivariate word distributions. Our very intuitive measure is tested on paradigmatic models and applied to cardiovascular data, especially to bivariate time series consisting of the beat-to-beat systolic blood pressure and heart rate variability values. Revealing

^(a)Both authors equally contributed to this work.

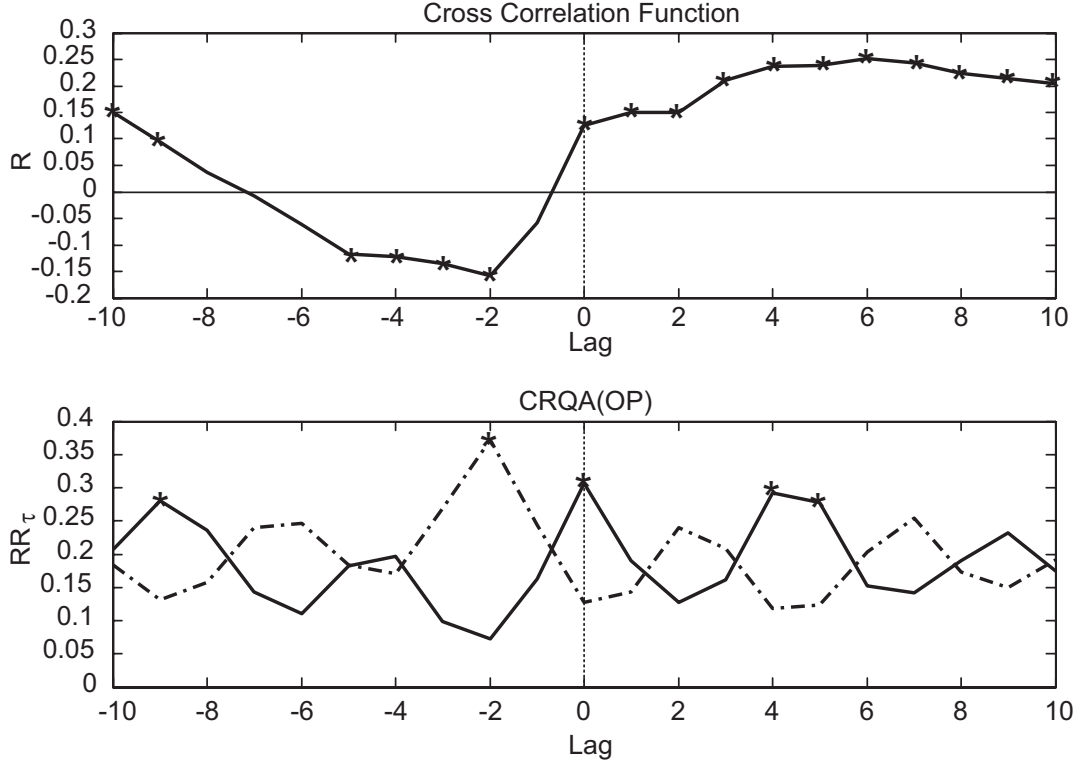


Fig. 1: Linear cross correlation function (R) as well as recurrence quantification analysis (recurrence rate RR_τ) for the detection of coupling directions of real data (see fig. 2). Top: R between heart rate (BBI) and systolic blood pressure (SBP) data. This analysis reveals interrelations between both time series for almost any lag (marked with asterisks) and is, therefore, not very specific. Bottom: cross recurrence quantification analysis based on order patterns (CRQA(OP)) applied to BBI and SBP data [19,20] reveals the most significant positive interrelation between both time series for lag $\tau = 0$ and negative interrelation for lag $\tau = -2$. Solid line: SBP \rightarrow BBI (positive linkage RR_+ in [24]); dotted line: BBI \rightarrow SBP (negative linkage RR_-). The order patterns are constructed using the dimension $m = 3$ and delay $\vartheta = 3$.

the coupling relations for the latter data enables us to quantify the short term regulation of the cardiovascular system, and thus to quantify the risk of cardiovascular disorders. In the following we develop the theory and give examples of *symbolic coupling traces* (SCT) which are also based on the analysis of structural patterns but easier to interpret and less computationally intensive.

To introduce the SCT method, we consider a dynamic system represented by two paired one-dimensional time series $x(t)$ and $y(t)$. They are first transformed into two symbol sequences $s_x(t)$ and $s_y(t)$ via the transformation rule

$$s_z(t) = \begin{cases} 1, & z(t) \leq z(t + \vartheta), \\ 0, & z(t) > z(t + \vartheta). \end{cases} \quad (1)$$

Next, we construct series of words $w_x(t)$ and $w_y(t)$ containing l successive symbols from the time series $s_x(t)$ and $s_y(t)$, respectively. Hence, $d = 2^l$ different patterns are possible. Using the notion of nonlinear systems theory and the concept of embedding, the time series $x(t)$ and $y(t)$ can be interpreted as realisations of a dynamical system. Then, the original phase space is reconstructed by time delay embedding $\vec{x} = (x_t, x_{t+\vartheta}, \dots, x_{t+(m-1)\vartheta})$ for sufficiently high dimension m [25]. Depending on the strength of

noise, the structure of the reconstructed trajectory could be very complicated. Therefore, we make it simpler by symbolisation. The encoding of the trajectory yields a decomposition of the phase space into 2^l areas where $l = m - 1$. Afterwards, the bivariate word distribution (BWD) $(p_{ij})_{i=1,\dots,d,j=1,\dots,d}$ is estimated [23]

$$p_{ij} = P(w_x(t) = W_i, w_y(t) = W_j). \quad (2)$$

p_{ij} is the joint probability that the words W_i and W_j occur at the same time t in the word sequences $w_x(t)$ and $w_y(t)$, respectively. In the univariate symbolic dynamics approach only one word sequence is considered. A schematic illustration of the encoding procedure and the BWD-calculation is depicted in fig. 2. To measure the delay-time probability matrix $\Pi(\tau) = (p_{ij}(\tau))$ that the word W_i occurs in w_x at time t and W_j occurs in w_y at time $(t + \tau)$, we introduce

$$p_{ij}(\tau) = P(w_x(t) = W_i, w_y(t + \tau) = W_j), \quad (3)$$

based on eq. (2). With the symbol transformation of eq. (1) we take a loss of amplitude information $x(t)$ and $y(t)$, however, in time series with moderate noise and nonstationarities this information can be unreliable. Through

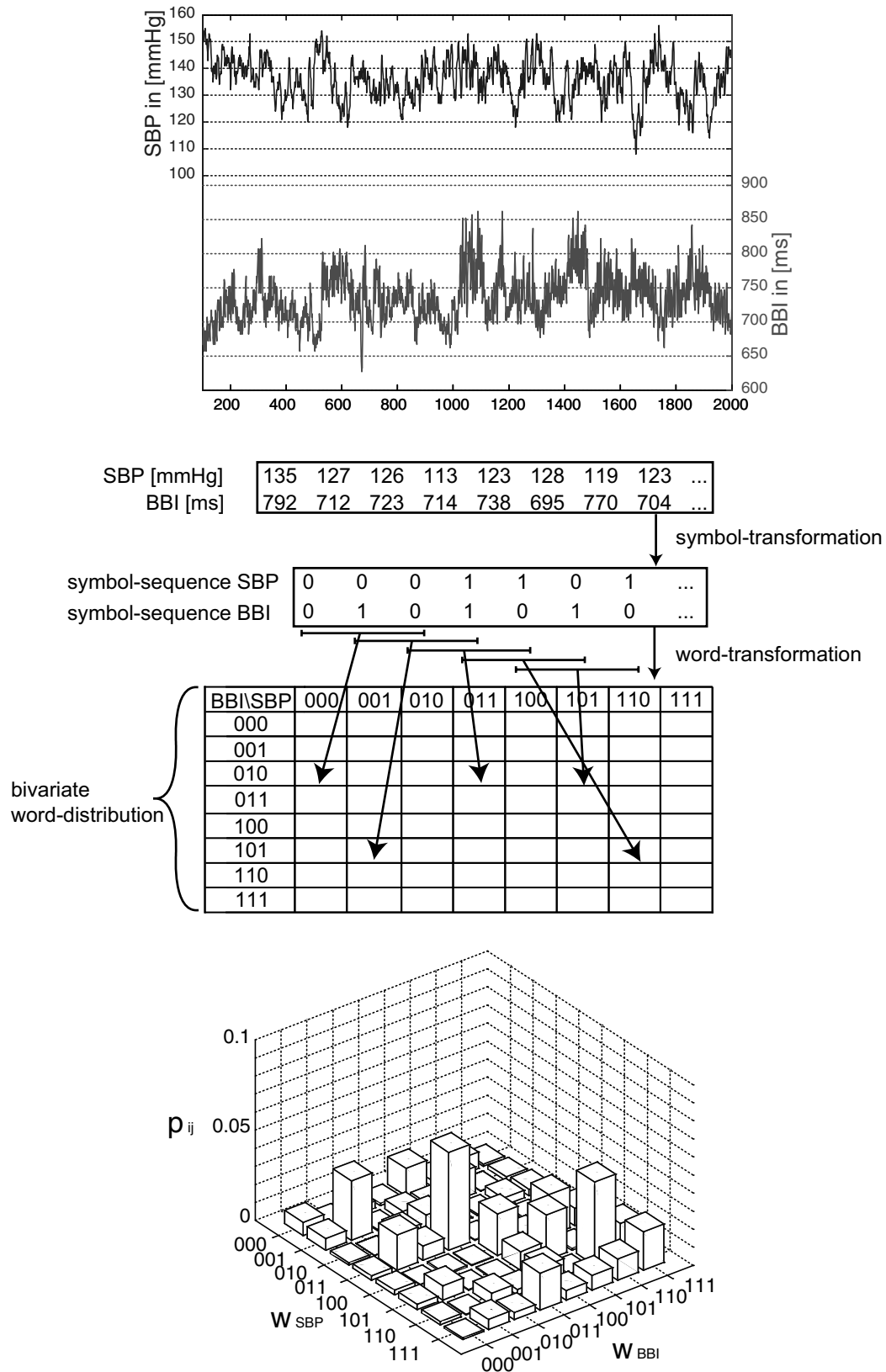


Fig. 2: Scheme for calculating the bivariate word distribution. Starting from two time series (*e.g.* SPB and BBI upper panel), a two-dimensional symbol sequence (middle panels) is calculated by a symbol transformation ($\vartheta = 1$ and $l = 3$) which leads then to the bivariate word distribution (lower panel) as the basis of parameter calculation.

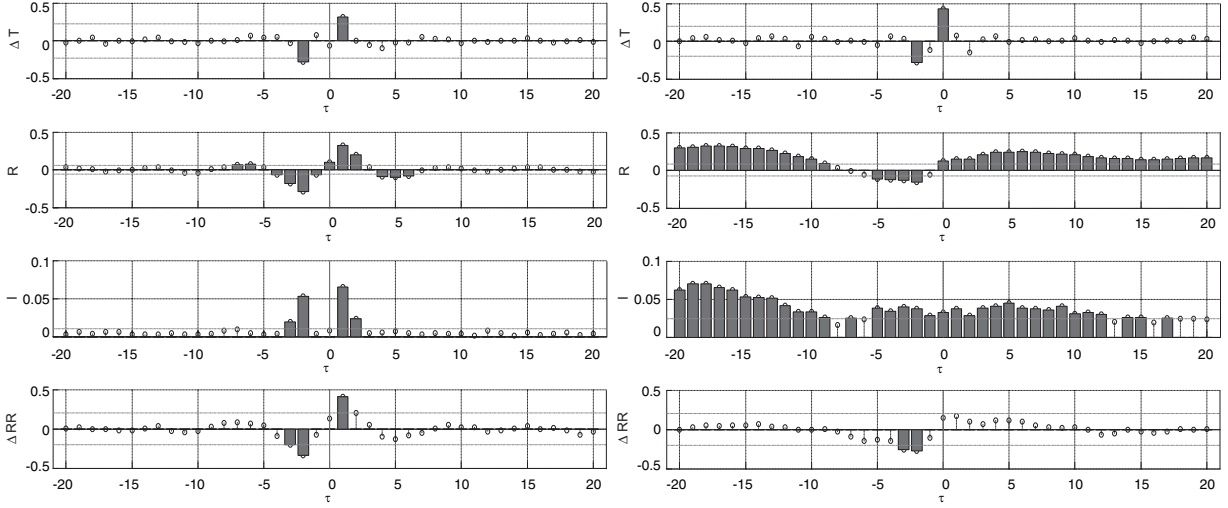


Fig. 3: Comparison of the calculated SCT parameter $\Delta T = T - \bar{T}$, cross correlation R , the mutual information I , and the recurrence rate difference $\Delta RR = RR_+ - RR_-$ (based on order pattern) for a simulation (left: $x_i = ax_{i-1} + by_{i-\tau_1} + \epsilon_i$ and $y_i = cy_{i-1} + dx_{i-\tau_2} + \nu_i$, two coupling terms coupling $y_{i-\tau_1}$ and $x_{i-\tau_2}$, $\epsilon = N(0, 0.1)$, $\nu = N(0, 0.1)$, $a = 0.3$, $b = 0.7$, $c = 0.3$, $d = 0.7$) and experimental data of a healthy volunteer (right) at different lags τ . Significant parameters are drawn as boxes. A simulation with symmetric coupling at $\tau = 1$ as well as diametric coupling at $\tau = -2$ is indicated on the left. The exact detection of the lags by ΔT is obvious. This is also true if nonstationarities, such as additive trends or time-dependent variance are present. R and I as well as ΔRR help to draw conclusions about the lags, but do not show them clearly. On the right, the most significant lags in the real data are revealed by ΔT at $\tau = 0$ and $\tau = -2$, *i.e.* BBIs correspond diametrically ($\tau = -2$) and symmetrically ($\tau = 0$) with SBP. The parameters R , I and ΔRR do not show these lags as clearly as ΔT , respectively show false significant lags.

symbolisation, word transformation and symmetric bivariate selection of the diagonals we can exclude random effects and include significant coupling information only.

Therefore, it is quantified by the BWD-diagonals: i) the trace T of the matrix $\Pi(\tau)$ is defined as

$$T(\tau) = \sum_{i=j} p_{ij}(\tau). \quad (4)$$

It represents the fraction of both time series, which are structurally equivalent to each other at lag τ .

ii) The parameter \bar{T}

$$\bar{T}(\tau) = \sum_{i=1, \dots, d, j=d+1-i} p_{ij}(\tau) \quad (5)$$

describes the fraction of both signals, which are structurally diametric at lag τ (d is the number of different patterns). Both parameters vary from 0 to 1 and comprise the diagonals of the BWD only. The quantification of the complete BWD by means of Shannon-entropy does not reveal the correct lags clearly. The reason for this is, that if all elements of the BWD are included in the parameter, too much information about mixed forms of symmetric and diametric structures are involved. This leads to a blur, hiding the correct lags. Finally, the difference $\Delta T = T - \bar{T}$ of the above parameters is the most appropriate choice which is shown by simulations, for example in fig. 3. Apart from the cross recurrence and SCT parameters, the

classic cross correlation function R [10] and the mutual information I [15] are calculated for comparison.

To study significance limits for ΔT , the parameter is calculated for randomised time series without coupling (for $\tau = 0$). A parameter value in a statistical test is considered as significant (seldom event) if its probability (estimated by $N = 100$ randomized surrogates) is smaller than the chosen significance level. Therefore, the maximum and minimum of ΔT in the group of simulations represent the borders of 1% significance level. In the following we consider only $m = 4$ and $\vartheta = 1$. Higher dimensions are not appropriate, because the word distribution (eq. (2)) cannot be estimated sufficiently well for smaller datasets. The considered lags are limited to $20 \leq \tau \leq 20$ (sampling units) in order to focus on short time-delayed dependencies. To validate the new method, the simplest approach is used: simulations of coupled 2D autoregressive (AR) processes (cf. fig. 3, left panel). The coefficients of the AR models are varied in order to study the influence of varying coupling strengths and of noise. For an example with model-specified lags at $\tau = -2$ and $\tau = 1$ (fig. 3, left panel), all four methods determine the defined lags correctly. In contrast to the SCT, however, the other three methods have additional false detections. The SCT-parameter detected the lags in case of delayed coupling with autocorrelation more clearly than cross correlation, mutual information and recurrence plot based on order pattern (cf. fig. 3, left panel). For higher noise levels this advantage decreases, however, for cardiological

time series autocorrelation with certain coupling have to be expected. Consequently, the SCT parameters are suitable for analysing the coupling of these signals (cf. fig. 3, right panel). For a further validation, we also applied it to nonlinear coupled models, *e.g.* SETARX (self-exciting threshold autoregressive model with external input) systems [26] and found similar results.

A minimal required sampling time is essential for our method. We used the symbolic dynamics method introduced in [27] to verify that the required sampling time for the given time series is fulfilled. This method quantifies the maximum information in a time series when coarse-graining the absolute differences between adjacent sample points. We find that an electrocardiogram sampled with at least 100 Hz and blood pressure recordings with at least 2 mmHg resolution are sufficient. Moreover, it is important to consider the amount of noise in the data when calculating the bivariate symbol sequence according to eq. (1). We investigated the impact of noise on the SCT coupling parameters by simulations. Additive white noise with a maximal signal-to-noise ratio of 10 dB does not significantly influence the coupling results. The SCT method is applied to real cardiological data to analyse the coupling between BBI and SBP values of 20 healthy volunteers (age: 53.0 ± 8.0 years). For all subjects, we measured continuous blood pressure signals (30 min, Portapres Mod. 2, 100 Hz sampling frequency, under standardized supine resting conditions, recorded at the Charité Berlin). Artefacts, such as calibration, motion etc., are removed and interpolated using an adaptive filter technique [28] to avoid corrupted results. The continuous blood pressure curves are used to extract the time series of BBI and SBP. A representative example of the coupling analysis is shown in fig. 3 (right panel). Parameters based on SCT are not influenced by nonstationarities of the time series. The parameter ΔT detects significant lags at $\tau = -2$ and $\tau = 0$ for all subjects. This confirms the prevailing opinion about the cardiovascular short term regulation. The symmetric lag at $\tau = 0$ reflects the mechanically induced arterial pressure fluctuations, whereas the diametric lag at $\tau = -2$ represents the vagal feedback from the BBI to the SBP. In ongoing studies the diagnostic and prognostic value of the new method could be verified with respect to cardiological diseases and the influence of ventilation.

The analysis of model systems and cardiological data with the SCT method demonstrates its advantages especially for physiological data. Nevertheless, for the general assessment of coupling directions in time series, both new and established methods should be used. The amount of nonstationarities, nonlinearities, autocorrelation and noise influences the quality of the results of each method. Coupling in stationary data with strong noise can be well detected via mutual information and cross correlation, whereas in deterministic data cross recurrence should be preferred. The parameters of the SCT method and cross recurrence based on order pattern close the gap in the

coupling analysis of nonstationary time series with strong autocorrelation and moderate noise, where cross correlation, mutual information and other methods are not sufficient to localise the lags exactly. The advantages of the SCT parameters in comparison to cross recurrence are the easy interpretation and the lower computational costs, because only symmetric and diametric words are quantified. Furthermore, SETARX simulations discovered that the reason for the failure of the known methods are instationarities, which are caused by intrinsic nonlinearities or changing conditions. As the time series of heart rate and blood pressure are likely to be of the mentioned kind for which the SCT is appropriate, a higher gain of diagnostic information can be expected. Testing of coupling in terms of time lag and in terms of intensity under well-controlled physiological test conditions can probably give indications for risk factors for cardiovascular diseases. Finally, our approach may be applicable also for multivariate or various other systems, *e.g.* socio-economic systems.

This work was supported by the German Science Foundation (DFG KU837/23-1, DFG PE628/4-1 and BR1303/10-1) as well as by the EU project BRACCIA.

REFERENCES

- [1] IVANOV P. CH. *et al.*, *Europhys. Lett.*, **43** (1998) 363.
- [2] BLASIUS B. *et al.*, *Nature*, **399** (1999) 354.
- [3] GLASS L., *Nature*, **410** (2001) 277.
- [4] MARVEL K., *Nature*, **411** (2001) 251.
- [5] ASHKENAZY Y. *et al.*, *Europhys. Lett.*, **53** (2001) 709.
- [6] LAI Y. C., *Phys. Rev. E*, **66** (2002) 036217.
- [7] MITLON J. G. *et al.*, *EPL*, **83** (2008) 48001.
- [8] CAMMAROTA C. *et al.*, *Phys. Rev. E*, **74** (2006) 042903.
- [9] RUMMEL C. *et al.*, *EPL*, **80** (2007) 68004.
- [10] MULLER M. *et al.*, *EPL*, **84** (2008) 10009.
- [11] GRANGER C. *et al.*, *Econometrica*, **37** (1969) 424.
- [12] DHAMALA M. *et al.*, *Phys. Rev. Lett.*, **100** (2008) 018701.
- [13] FAES L. *et al.*, *Biomed. Tech. (Berlin)*, **51** (2006) 255.
- [14] ROSENBLUM M. *et al.*, *Phys. Rev. E*, **65** (2002) 041909.
- [15] PALUS M. *et al.*, *Phys. Rev. E*, **75** (2007) 056211.
- [16] PORTA A. *et al.*, *Biol. Cybern.*, **86** (2002) 241.
- [17] NOLLO G. *et al.*, *Am. J. Physiol.*, **288** (2005) H1777.
- [18] SCHELTER B. *et al.*, *Phys. Rev. Lett.*, **96** (2006) 208103.
- [19] MARWAN N. *et al.*, *Phys. Rep.*, **438** (2007) 237.
- [20] GROTH A. *et al.*, *Phys. Rev. E*, **72** (2005) 046220.
- [21] MARWAN N., *CRP toolbox for Matlab*, www.recurrence-plot.tk.
- [22] DE BOER R. W. *et al.*, *Am. J. Physiol.*, **253** (1987) 680.
- [23] BAUMERT M. *et al.*, *Med. Biol. Eng. Comput.*, **40** (2002) 241.
- [24] MARWAN N. *et al.*, *Phys. Lett. A*, **302** (2002) 299.
- [25] TAKENS F., *Dynamical Systems and Turbulence*, Vol. **898** (Springer-Verlag) 1981.
- [26] TONG H., *Non-linear Time Series: A Dynamical System Approach* (Oxford Science Publications) 1990.
- [27] MALBERG H. *et al.*, *Somnologie*, **10** (2006) 33.
- [28] WESSEL N. *et al.*, *Phys. Rev. E*, **61** (2000) 733.