PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

rsta.royalsocietypublishing.org

Research

Cite this article: Ávila GMR, Gapelyuk A, Marwan N, Walther T, Stepan H, Kurths J, Wessel N. 2013 Classification of cardiovascular time series based on different coupling structures using recurrence networks analysis. Phil Trans R Soc A 20110623. http://dx.doi.org/10.1098/rsta.2011.0623

One contribution of 13 to a Theme Issue 'Assessing causality in brain dynamics and cardiovascular control'.

Subject Areas:
complexity, medical physics, computational physics

Keywords:
time-series analysis, cardiac dynamics, networks and genealogical trees, hemodynamics, blood flow in cardiovascular system, coupling analysis

Author for correspondence:
Niels Wessel
e-mail: niels.wessel@physik.hu-berlin.de

Royal Society Publishing
Informing the science of the future

Classification of cardiovascular time series based on different coupling structures using recurrence networks analysis

Gonzalo Marcelo Ramírez Ávila[1,2,3], Andrej Gapelyuk[1], Norbert Marwan[2], Thomas Walther[4], Holger Stepan[5], Jürgen Kurths[1,2,6] and Niels Wessel[1]

[1]Institut für Physik, Humboldt-Universität zu Berlin, Germany
[2]Potsdam Institut für Klimafolgenforschung, Germany
[3]Instituto de Investigaciones Físicas Universidad Mayor de San Andrés, La Paz, Bolivia
[4]Centre for Cardiovascular and Metabolic Research, Hull York Medical School, University of Hull, UK
[5]Department of Obstetrics and Gynecology, University of Leipzig, Germany
[6]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, UK

We analyse cardiovascular time series with the aim of performing early prediction of preeclampsia (PE), a pregnancy-specific disorder causing maternal and foetal morbidity and mortality. The analysis is made using a novel approach, namely the ε-recurrence networks applied to a phase space constructed by means of the time series of the variabilities of the heart rate and the blood pressure (systolic and diastolic). All the possible coupling structures among these variables are considered for the analysis. Network measures such as average path length, mean coreness, global clustering coefficient and scale-local transitivity dimension are computed and constitute the parameters for the subsequent quadratic discriminant analysis. This allows us to predict PE with a sensitivity of 91.7 per cent and a specificity of 68.1 per cent, thus validating the use of this method for classifying healthy and preeclamptic patients.

1. Introduction

Q2

Detection of cardiovascular (CV) disorders has been considerably improved due to both technological advances and new methods of time-series analysis. Nevertheless, there are still difficulties that cannot be explained by standard data analysis. Nonlinear data analysis and modelling methods of CV physics allow to improve clinical diagnostics and give the opportunity for a better understanding of CV regulation. One of the most important aspects of these methods is that they focus on non-invasive recordings of biosignals. A review of these methods is outlined in the earlier studies [1–5]. Among the biosignals that CV physics deals with are the heart rate variability (HRV), the variabilities of systolic blood pressure (SBPV) and diastolic blood pressure (DBPV) and the baroreflex sensitivity (BRS). A detailed description of these biosignals and their use can be found in the earlier studies [6–10] and underlying models using these measures have been developed [11–13].

Preeclampsia (PE) is a major hypertensive disorder in pregnant women also characterized by proteinuria for which the pathophysiology remains unclear and constitutes a serious risk for both the mother and the foetus. PE affects healthy nulliparous women in a range between 2 and 7 per cent worldwide [14]. Several strategies are used in order to predict PE, among them we can mention some biochemical markers, such as fms-like tyrosine kinase 1 (sFlt-1), placental growth factor (PlGF), soluble endoglin [15,16], maternal autoantibody, the angiotensin II type I receptor agonistic autoantibody (AT1-AA) [17], the urinary biomarkers [18], ultrasonographic markers [19], non-invasive CV markers [20] or the combination of some of those [21–24].

Time-series analysis plays an essential role in the understanding of various real systems regardless of their nature. The techniques used in time-series analysis evolve systematically supported by new insights in statistics, mathematics and physics. Numerous works are devoted to nonlinear time-series methods [25,26] and further developments arise continuously [27–32] allowing the improvement of the analysis and interpretation, and also providing a deeper understanding of the phenomena dealt with. Recurrence methods have recently become a useful tool in order to study time series and acquired importance because they do not need long time series to identify transitions in dynamical systems [28] and their use may be applied to a wide diversity of systems and phenomena as for instance climate ones [33]. In particular, Recurrence Quantification Analysis (RQA) has been successfully used in the study of CV signals [34,35]. Recently, the recurrence concept has been extended to networks and novel time-series analysis methods arose [36]. In this work, we apply the approach of ε-recurrence networks to analyse non-invasive CV indicators with the aim of developing a classification method to identify healthy subjects (control) from patients who develop PE. The method used in this article differs from the methods dealing with time-varying dynamics [37] in the fact that in a recurrence network, information about the temporal ordering of observations is not explicitly regarded, and we focus on the topological features of the resulting graphs that might reflect dynamically invariant properties associated with the specific dynamical system. The quantification of these topological features enables us to classify patients who develop the pathology from the control individuals.

The paper is organized as follows: in §2, we give an overview of the used complex network-based model; we also explain the clinical aspects (patients and their CV measures), the data processing and statistics. In §3, we present the most important results for the classification. Finally, in §4, we discuss the results giving the conclusion and perspectives of this work.

2. Material and methods

(a) Recurrence networks

Since the foundation of graph theory in 1736 by Leonhard Euler who solved the Königsberg bridge problem, the use of graphs composed of vertices and edges, has witnessed an astonishing evolution towards the complex networks theory in which the principal elements are nodes and links both exhibiting dynamic aspects in the sense that each node might have an intrinsic

behaviour governed by nonlinear dynamical equations and even the links can evolve due to the motion of the nodes [38]. In the last two decades, complex networks theory became an important research topic and its application to numerous complex systems of very different nature comprising physical [39], chemical [40], biological [41], economical [42], social [43] and climatic [33], among others, gives evidence of its practical importance. Details of complex network theory can be found in the study of Albert & Barabasi [44] and Boccaletti *et al*. [45].

The basic idea of time-series analysis based on complex network techniques lies on the fact that a time series might be transformed into a complex network from which we can extract the adjacency matrix allowing us to obtain local and global network properties. Roughly speaking, there are three classes of transforming methods: visibility graphs, proximity networks and transition networks [46]. The latter two are based on the concept of recurrence which has been stated firstly by Poincaré [47], and reemerges with the introduction of recurrence plots [48] and its further developments and refinements such as RQA [28]. The concept of recurrence applied to a single trajectory of the dynamical system allows us to obtain the recurrence matrix whose elements are given by $R_{i,j} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|)$, where $\Theta(\cdot)$ represents the Heaviside function, $\|\cdot\|$ is a suitable norm, and ε is a threshold distance that should be chosen adequately according to the characteristics of the embedded attractor into the phase space. We interpret the recurrence matrix $\mathbf{R}$ as the adjacency matrix of an unweighted and undirected complex network, commonly called the ε-recurrence network which is associated with a given time series. Possible self-loops must be avoided in this network, thus a Kronecker delta must be subtracted from the recurrence matrix to obtain a zero-diagonal and as a consequence, the elements of the adjacency matrix for an ε-recurrence network are:

$$A_{i,j}(\varepsilon) = R_{i,j}(\varepsilon) - \delta_{i,j}, \tag{2.1}$$

where the ε-dependence (i.e. the level of coarse-graining of phase space involved in this procedure) is considered explicitly [49]. There is not a universal criterion for choosing ε but the choice must be made avoiding too small values that lead to a situation in which there are not enough recurrence points, or too large values implying that every vertex is connected with many other vertices irrespective of their actual mutual proximity in phase space [31]. Having reconstructed the adjacency matrix $\mathbf{A}$ from a time series, we can apply appropriate networks characteristics to analyse and obtain information of the underlying system. In this work, we focus our interest in four global network measures.

(i) Average path length ($\mathcal{L}$)

Is the mean value of the shortest geodetic path lengths $l_{i,j}$ considering all pair of vertices (i, j). Thus,

$$\mathcal{L} = \langle l_{i,j} \rangle = \frac{2}{N(N-1)} \sum_{i<j} l_{i,j}, \tag{2.2}$$

being N the number of nodes in the network (i.e. the length of the time series [31]).

(ii) Mean coreness ($\mathcal{C}_2$)

The notion of core was introduced in social networks related to the concept of cohesiveness [50]. The concept of node coreness indicates the significance of a node and its popularity in the network. The k-core of a graph is the maximal subgraph in which each vertex has at least degree (in the subgraph) k. The coreness of a vertex is the highest order of a k-core containing the vertex and is calculated using the Batagelj's algorithm [51]. The mean coreness is obtained by averaging the corenesses of all the vertices.

(iii) Global clustering coefficient $\mathcal{C}$

As the coreness, the clustering coefficient might be defined locally for each vertex as the ratio of triangles, including vertex i and the number of triples centred on vertex i where triple refers to a

pair (j,k) of vertices that are both linked with i, but not necessarily mutually linked. In terms of the adjacency matrix elements,

$$\mathcal{C}_i = \frac{\sum_{j,k=1}^{N} A_{j,k}A_{i,j}A_{i,k}}{k_i(k_i-1)}, \tag{2.3}$$

where k_i is the connectivity (degree) of the vertex i. The global clustering coefficient is obtained from the averaging of $\mathcal{C}_i$. Thus, $\mathcal{C} = \sum_{1=1}^{N} \mathcal{C}_i/N$ [45].

(iv) Scale local transitivity dimension $\mathrm{D}_{\mathcal{T}}$

Transitivity $\mathcal{T}$ is defined as the ratio of the number of triangles in the network times three and the number of linked triples of vertices [45]. The latter can be written as

$$\mathcal{T} = \frac{\sum_{i,j,k=1}^{N} A_{j,k}A_{i,j}A_{i,k}}{\sum_{i,j,k=1}^{N} A_{i,j}A_{i,k}(1-\delta_{j,k})}. \tag{2.4}$$

The scale local transitivity dimension is defined as $\mathrm{D}_{\mathcal{T}} = \log \mathcal{T}/\log(\frac{3}{4})$ [49].

Note that all these measures depend on ε and have a global character. Some of these measures served as well to characterize the network topology such as in the case of small-world networks that have high clustering coefficient and short characteristic path length [52]. On the contrary, random networks follow the relationships: $\mathcal{C} \sim 1/N$ and $\mathcal{L} \sim \ln N$ [53], being N the number of nodes.

(b) Clinical aspects

We consider for this study 96 patients with abnormal uterine perfusion, followed by means of Doppler sonography in the second trimester, between the 18th and the 26th week of gestation (WOG) of pregnancy, at the Department of Obstetrics and Gynecology of the University of Leipzig. Immediately after the Doppler examination, non-invasive continuous blood pressure monitoring was conducted via finger cuff during 30 min. The continuous blood pressure curves were used to extract the time series of beat-to-beat intervals, systolic and diastolic blood pressures allowing us to obtain the CV indicators (HRV, SBPV and DBPV). The length of the dataset per variable is roughly 1700. At the time of examination, the women were healthy, normotensive, without clinical signs of cervical incompetence and on no medication. After the 30th WOG, 24 patients developed PE. Further details on the methodology can be found in the study of Malberg *et al*. [20].

(c) Data processing and statistics

In order to avoid artefacts such as double recognition of beats, the original recurrence rate (RR) time series were filtered using a preprocessing algorithm which first removes obvious recognition errors; then applies an adaptive percent filter, and finally, an adapting controlling filter [1]. An example of these filtered signals that constitute the CV indicators is shown in figure 1. A preliminary inspection of figure 1 allows us to realize that there are no significant differences in the median values of the CV indicators between both groups.

With the aim of using a recurrence network approach, we consider the three CV indicators and some possible coupling structures. An estimation of the coupling structure of CV indicators has been performed using nonlinear additive autoregressive models with external input following the idea of Granger causality [54]. This coupling analysis shows that HRV, DBPV and SBPV respond to respiration; SBPV respond to DBPV and the latter to HRV. In our case, we do not consider respiration; thus, the coupling structure might be represented as in figure 2*a* where according to the coupling scheme, there is a delay between the HRV, the DBPV and the SBPV. For simplicity, we write down the coupling structure as (HRV(t), DBPV($t+1$), SBPV($t+2$)) or simply $H(t)D(t+1)$

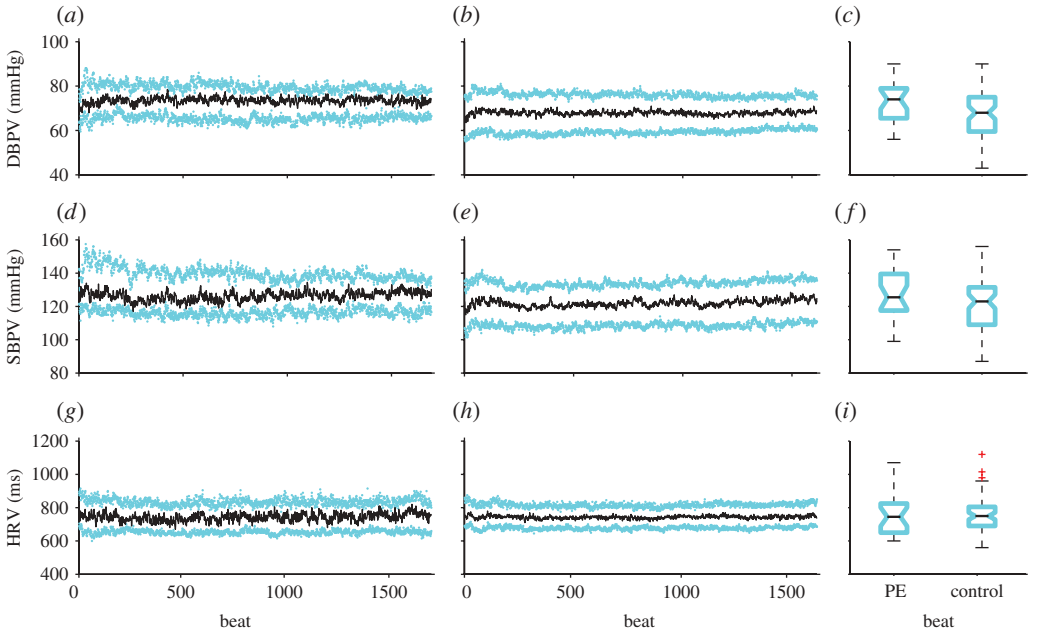

Figure 1. Representative time series of the CV indicators (*a*,*b*) DBPV, (*d*,*e*) SBPV and (*g*,*h*) HRV using the median and the corresponding Q_1 and Q_3 (25% and 75% quartiles, respectively) for both groups: PE (first column) and control (second column). Box-and-whisker plots considering a whisker of 1.5 times the interquartile range for (*c*) DBPV, (*f*) SBPV and (*i*) HRV. (Online version in colour.)

$S(t+2) \equiv 012$. Here, an arrow indicates a causal relation, it is why, when there are two arrows, it indicates that the 'first source' of arrows ($H(t)$) plays the role of the 'main driver' so that, the other variables do not influence it (consequently, we adopt the 0 for that, i.e. no arrows converge on it); the 'second source' of arrows ($D(t)$) plays the role of the 'secondary driver', thus, it is driven by H but it drives to S (consequently, we adopt the 1 (label $t+1$) for that); finally, S is driven by D and consequently by H, i.e. there are two time steps separating S from the 'first source' H, and for this reason, we adopt the 2 (label $t+2$) for that.

We sought to predict whether or not a patient develops PE using the CV indicators embedded in a phase space determined by the structure of coupling. We consider a minimalist assumption in which the structure of coupling between HRV, DBPV and SBPV is equal in each subject of a group and that this structure does not change during the measurement. In this study, we set out to test all the possible structures of coupling shown in figure 2 and a wide range of the threshold ε going from 0.01σ to 0.99σ being σ the standard deviation of the underlying process in the embedded phase space. From a simple CV time series corresponding to each patient, we construct a complex network for each possible structure of coupling and each value of ε. Then, we compute the four network measures: $(\mathcal{C}, \mathcal{L}, \mathcal{C}\prime, \mathrm{D}_{\mathcal{T}})$, and with these new measures we perform an analysis to classify the groups of patients. For that purpose, we firstly verify whether or not these new parameters are significant by means of a Mann–Whitney U-test and considering a significance level of 5 per cent; being the null hypothesis that data in the vectors corresponding to control and preeclamptic patients are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians.

The methodology to obtain the results might be summarized in the following steps: (i) preprocessing the raw data and obtaining the values of HRV, DBPV and SBPV. (ii) Identifying the interaction among HRV, DBPV and SBPV (coupling structure). (iii) Obtaining the recurrence matrix for HDS(t). (iv) Obtaining the recurrence network. (v) Use the structure of the coupling and the threshold ε to construct the parameter space. (vi) Perform the classification analysis from the parameter space and the statistical issues.

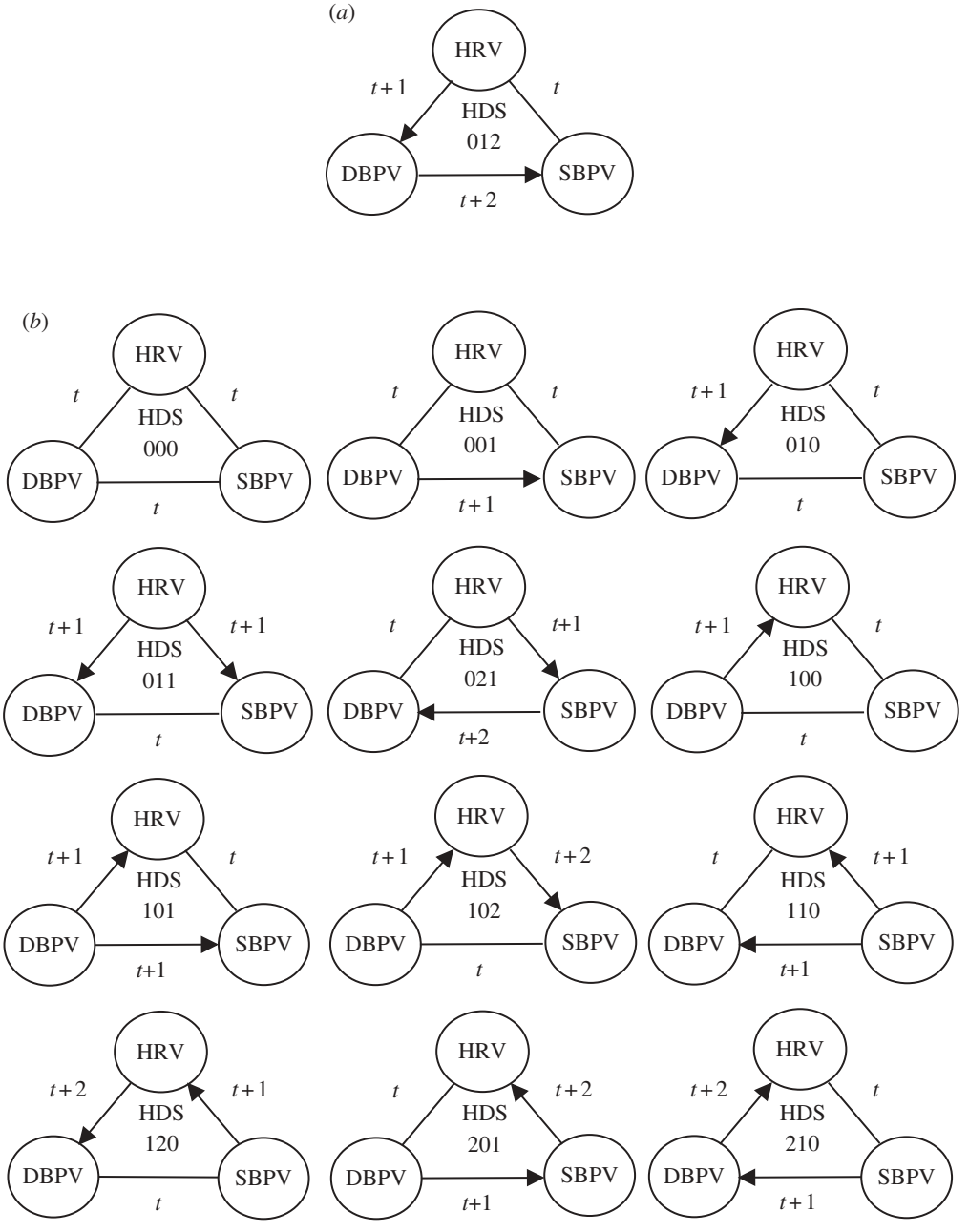

Figure 2. (*a*) Coupling structure considering that HRV drives the DBPV and this in turn the SBPV (directed arrows from HRV to DBPV, and from DBPV to SBPV). Note that when the variables are linked only for a line, it means that these are coupled but without any delay. This might be written schematically as $H(t)D(t+1)S(t+2) \equiv 012$; the latter number can change according to the delay among the sequential variables HRV, DBPV and SBPV, represented as *HDS*. (*b*) All the other possibilities of coupling structures. The numbers corresponding to the sequence *HDS* take into account the possible delay among the variables. We only consider one time-step delays between two consecutive variables. For instance, 000, indicates that there is no delay among the three variables and the corresponding coupling structure is $(H(t), D(t), S(t))$. On the contrary, 210, indicates that *S* drives *D* and this turn in *H*, thus, the coupling structure is $(H(t+2), D(t+1), S(t))$. Note that in general, $\mathrm{HDS}(t) = (H(t+\tau_1), D(t+\tau_2), S(t+\tau_3))$, where τ_i represent the delay of each of the variables that satisfies $\sum_{i=1}^{3} \tau_i \leq 3$ and at least one of the values must be zero.

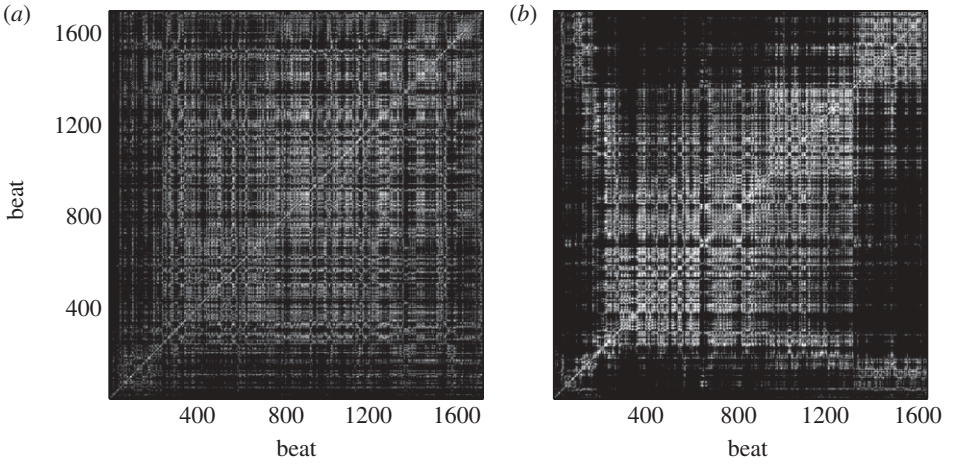

Figure 3. Recurrence plots obtained using the time series of the medians for both groups of patients (*a*) PE and (*b*) control.

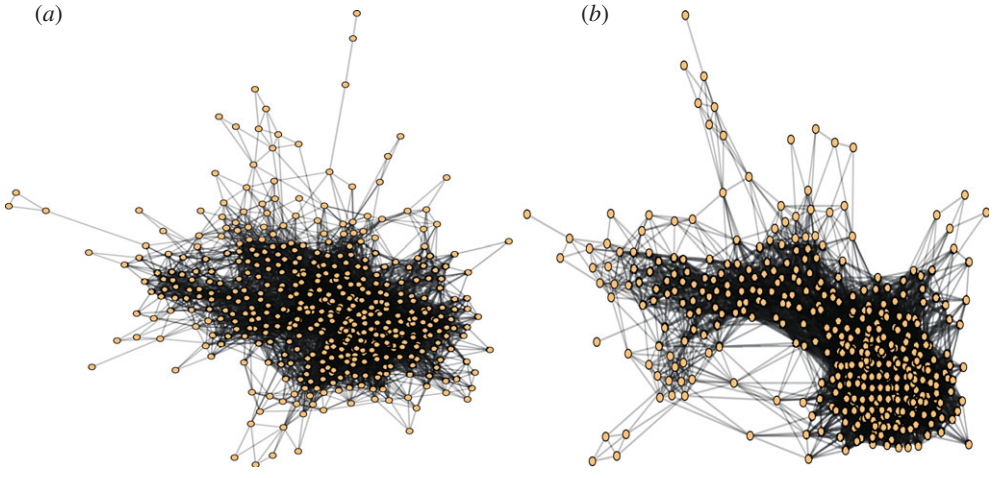

Figure 4. Visualization of the networks obtained by using a layout algorithm that considers an optimization procedure by physical principles (software Network Workbench [55]); in this case, for a better visualization, we take only 500 nodes into consideration. The networks visualizations were constructed with the medians of the time series for both groups of patients (*a*) PE and (*b*) control. (Online version in colour.)

3. Results

As the approach is based on recurrence complex networks, firstly we obtain the matrices **R** and **A**, and we expect that visualization of structures related to these matrices could also give some information about the differences between the PE and the control group. A visualization of the recurrence plots and the associated networks, obtained using the medians of the time series are shown in figures 3 and 4, respectively. We note slight differences between the recurrence plots corresponding to PE and control group (figure 3). The network representations are made using a layout algorithm that uses basic principles of physics to iteratively determine an optimal layout, i.e. a given mass and an electric charge is associated with each node, and each edge is represented as a spring. For this second construction, we have used only 500 nodes in order to better visualize the differences between the PE and control networks (figure 4*a*,*b*, respectively). Nonetheless, this visualization inspection is just a first checkup that cannot replace the quantification of the network measures.

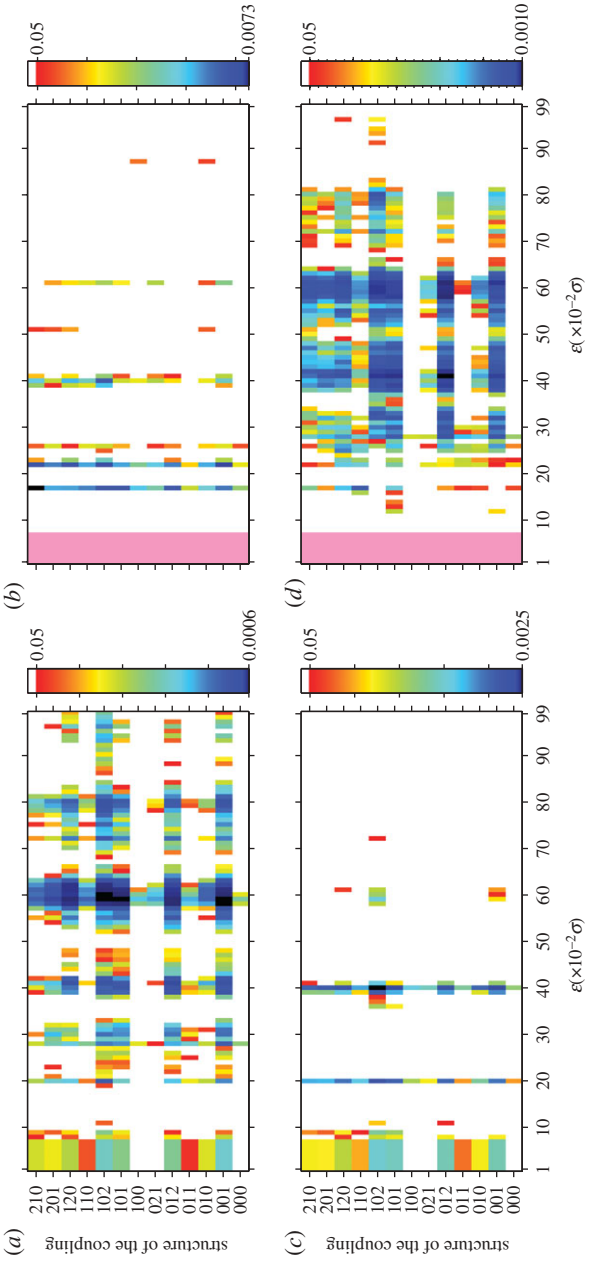

Figure 5. Phase plane structure of the coupling versus ε showing the significance level p computed by means of a Mann–Whitney U-test for establishing differences between the control and PE groups and using the network measures (*a*) $\mathcal{C}$, (*b*) $\mathcal{L}$, (*c*) $\mathcal{C}_\mathcal{I}$ and (*d*) $\mathrm{D}_\mathcal{T}$. The colour code indicates the p-values. Note that some special pixels are used such as white ($p \geq 0.05$; H_0 cannot be rejected), black (minimum p-value) and pink (the strip comprised in the interval $[1, 7] \times 10^{-2}\sigma$ and covering all the cases of the coupling structure, where it is not possible to compute the p-value, i.e. it is undetermined).

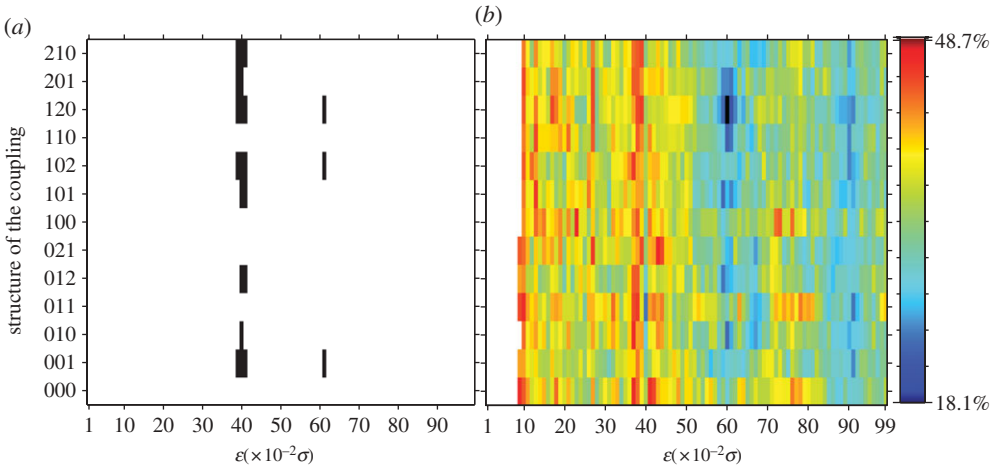

Figure 6. Same phase planes as in figure 5 showing the situations (*a*) in which the four considered network measures satisfy simultaneously the condition $p < 0.05$ (black pixels); and (*b*) the misclassification errors (colour code) in the classification of control and PE groups after a quadratic discriminant analysis for the four network measures. Here, the black pixel indicates the minimum value of the error, whereas white pixels indicate that the discriminant analysis cannot be performed and it is related to the fact that for these cases, at least one of the network measures has an undetermined *p*-value.

The results for each network measure are represented in the phase plane, embedding (structure of coupling) versus ε as shown in figure 5. The colour code indicates the *p*-values of the statistical test when the null hypothesis H_0 of equal medians at 5 per cent significance level is rejected. The white pixels denote that there is no difference between both groups ($p \geq 0.05$), and pink ones, the impossibility to compute *p*. On the contrary, the black pixels represent the minimum *p*-value among all the possibilities on the phase plane.

According to figure 5, the significant values for each network measure occur only for some coupling structures and thresholds ε. Figure 6*a* shows the same plane as in figure 5 but considering the cases in which all the four network measures are simultaneously significant, i.e. $p < 0.05$ (black pixels). The inspection of figure 6*a* shows that there are 22 situations in which the four network measures satisfy simultaneously the statistical significance test, and we further restrict the analysis to these selected cases which do not necessary correspond to the lower *p*-values. Now, considering these four measures as the parameters for the classification of control and PE groups, we perform a quadratic discriminant analysis for all the possible structures of the coupling and ε (figure 6*b*). At a first sight, we identify the minimum *p*-value which corresponds to the coupling structure of 120 and $\varepsilon = 0.60\sigma$. Nevertheless, this situation does not correspond to the selected ones in figure 6*a* and consequently it is discarded. Table 1 shows the statistical measures of the performance of a binary classification test for the 22 selected cases. Such measures are misclassification error (percentage of observations that are misclassified), sensitivity (proportion of true positives that are correctly identified by the test), specificity (proportion of true negatives correctly identified by the test) [56], positive predictive value (PPV), i.e. the proportion of patients with positive test results who are correctly diagnosed and negative predictive value (NPV), i.e. the proportion of patients with negative test results who are correctly diagnosed [57]. From table 1, we select the situation corresponding to a coupling structure 120 and $\varepsilon = 0.61\sigma$ (italic fonts) whose misclassification error is 20.1 per cent giving consequently the best values for the classification results, i.e. a sensitivity of 91.7 per cent, a specificity of 68.1 per cent, a PPV of 48.9 per cent and an NPV of 96.1 per cent. We stress again the fact that the *p*-values for the four considered network measures are statistically significant (table 2). It is worthwhile to mention that performing 10-fold cross-validation, the results obtained using this technique give for sensitivity and specificity 60.0 per cent and 69.7 per cent, respectively.

Table 1. Statistical measures of the performance of a binary classification test considering the 22 possible situations in which the four network measures satisfy simultaneously the condition $p < 0.05$.

coupling	ε ($\times\sigma$)	error (%)	sensitivity (%)	specificity (%)	PPV (%)	NPV (%)
001	0.39	40.3	79.2	40.3	30.7	85.3
	0.40	31.2	79.2	58.3	38.78	89.4
	0.41	34.0	89.5	44.4	34.4	91.4
	0.61	29.2	79.2	62.5	41.3	90.0
010	0.40	28.5	83.3	59.7	40.8	91.5
012	0.40	29.2	83.3	58.3	40.0	91.3
	0.41	36.8	87.5	38.9	32.3	90.3
101	0.40	34.0	83.3	48.6	35.1	89.7
	0.41	38.2	87.5	36.1	31.3	89.7
102	0.39	41.0	79.2	38.9	30.2	84.8
	0.40	34.7	83.3	47.2	34.5	89.5
	0.41	40.3	83.3	36.1	30.3	86.7
	0.61	29.2	83.3	58.3	40.0	91.3
120	0.39	45.8	79.2	29.2	27.1	80.8
	0.40	34.0	91.7	40.3	33.9	93.6
	0.41	36.1	91.7	36.1	32.3	92.9
	0.61	*20.1*	*91.7*	*68.1*	*48.9*	*96.1*
201	0.39	42.4	83.3	31.9	29.0	85.2
	0.40	38.2	75.0	48.6	32.7	85.4
210	0.39	46.1	79.2	30.6	27.5	81.5
	0.40	32.6	87.5	47.2	35.6	91.9
	0.41	36.1	87.5	40.3	32.8	90.6

Table 2. *P*-values for the four network measures when using the situation structure of the coupling 120 and $\varepsilon = 0.61\sigma$ chosen from table 1.

network measure	$\mathcal{C}$	$\mathcal{L}$	$\mathcal{C}_l$	$D_{\mathcal{T}}$
p-values	0.0013	0.0386	0.0472	0.0028

Finally, we must point out several aspects in order to justify our methodology compared with other possible approaches. The utilization of ε-recurrence networks to classify using the embedding of the individual time series, i.e. $x(t) = (H(t), H(t-1), H(t-2))$, $y(t) = (S(t), S(t-1), S(t-2))$ or $z(t) = (D(t), D(t-1), D(t-2))$ restricts considerably the possibilities for the analysis; there are few cases in which a set of four network measures satisfies the condition $p < 0.05$ for all the four variables simultaneously. We obtained interesting results using the embedding of the DBPV for small values of ε. On the contrary, for HRV and SBPV, there are not four network measures satisfying simultaneously the condition $p < 0.05$. Despite the 'good results' (sensitivity and specificity, 41.7% and 93.1%, respectively) obtained with the embedding of DBPV, this approach obviates the coupling among the CV indicators. Thus, the conceptual richness is less

than the approach used in this work. We also performed RQA considering the same threshold and coupling structure ($\varepsilon = 0.61\sigma$ and 120) and found that sensitivity and specificity are 91.7 per cent and 45.8 per cent, respectively, for a set of four variables (RR, determinism, longest vertical line and trapping time) satisfying simultaneously the condition $p < 0.05$. The latter confirms that our method is more adequate for this problem of classification.

4. Discussion

Although large improvements concerning the early detection of PE have been made [58], this topic is still one of the most important challenges in obstetrics. Several researches are oriented to cheap and non-invasive methods and the application of new methods and ideas contributes to improve the statistical predictive values. The essential aspect of the approach used in this work lies in its novelty when applying to CV signals, i.e. complex time series that in their raw form are not useful for classification, are transformed into recurrence networks from which we extract several measures that allow a classification with suitable results. In fact, after the choice of an adequate structure of the coupling and the threshold ε, only one complex network is constructed from the three CV indicators for each person and then, we quantify the network features that constitute the parameters for the classification analysis. In summary, our exploratory results show that the used approach constitutes a useful tool to study such a classification problem.

Note that the analysis presented here is in some sense only a first approximation of the recurrence networks approach. Our method takes into account the coupling among the CV indicators (HRV, SBPV and DBPV). The coupling is very important in the dynamics of these quantities and consequently in the causes that could lead to deregulation of the blood pressure. We see for future research several ways of improvements, as explained in the following. In spite of the minimalist assumptions concerning the structure of the coupling, and just one value of ε in order to avoid the ambiguities stated in [59], our results give useful information for the classification and are similar to those obtained in [20], thus validating our method. The consideration of dynamic structures of the coupling (i.e. temporal variations in the coupling structure) could improve our results and also give us a deeper insight in the underlying physiological processes. For that, it is necessary to design an adaptive algorithm taking into account possible transitions in several time windows. This method could be also useful as an alternative to find the adequate structure of coupling and thus minimizing poor performances of the specific coupling structures to describe some casual relations. We are also aware of the major role of respiration [60] and related to the understanding of the PE pathogenesis [54]. Therefore, adequate data processing, including respiration should improve our results. Additionally, in order to improve the accuracy and applicability of the method, further complementary measures of causality could be performed either using Granger causality or convergent cross mapping, a new technique applied successfully to real ecological systems [61]. Finally, the method used in this study offers some additional advantages such as the ability to cope with short time series and shows that it is not mandatory to have observations equally spaced in time.

The used classification method might also be improved considering some other consistent network measures for a better characterization of the phase-space properties or combining with RQA that could give additional insights into the recurrence structure of the underlying dynamical system. The quantification of the motifs [62] consisting of three or four nodes in the underlying networks constitutes another aspect that could be considered in the classification. This quantification seems to be a feasible tool to understand the dynamics of certain systems and deserves a deeper study. The combination of these methods and its adequate interpretation could also help to a better comprehension of the related physiological aspects.

In our tests, we have observed that the discriminant analysis could even be improved by considering shorter time series and other types of coupling structure such as 013, 014, etc. (results not shown here). Nevertheless, we do not report these results because we prefer to keep on the complete time series and the assumptions concerning the coupling structure and consequently, the consistence of the physiological interpretations.

The consideration of other qualitative aspects related to the history of the patients (age, ethnicity and body mass index) [63] and in general predisposing factors such as genetic [64], behavioural [65] or environmental [66] could give additional information to improve the classification analysis combined with the technique used in this paper.

Our study follows the same line as previous works [20] in which the biosignal analysis (in our case, the associated recurrence complex network analysis) constitutes a non-invasive, cheap and universal diagnostic approach whose utilization offers new possibilities both in the understanding of PE pathogenesis and on the envisaging of new therapeutic strategies. Moreover, the method used in this work has similar performance at the level of classification to other ones. The novel aspects considered here are the inclusion of the coupling structure and the fact that few parameters (the network measures) can give us suitable classification results.

Acknowledgements. This work has been financially supported by the German Academic Exchange Service (DAAD), by the Deutsche Forschungsgemeinschaft (KU 837/20-1, KU-837/29-2, KU-837/34-1), by the Federal Ministry of Economics and Technology (FKZ KF2248001FR9) as well as the European projects EU NEST-pathfinder and BRACCIA.

References

1. Wessel N, Voss A, Malberg H, Ziehmann C, Voss HU, Schirdewan A, Meyerfeldt U, Kurths J. 2000 Nonlinear analysis of complex phenomena in cardiological data. *Herzschr. Elektrophys.* **11**, 159–173. (doi:10.1007/s003990070035)
2. Xiao X, Mullen TJ, Mukkamala R. 2005 System identification: a multi-signal approach for probing neural cardiovascular regulation. *Physiol. Meas.* **26**, R41–R71. (doi:10.1088/0967-3334/26/3/R01)
3. Wessel N, Malberg H, Bauernschmitt R, Kurths J. 2007 Nonlinear methods of cardiovascular physics and their clinical applicability. *Int. J. Bifurcation Chaos* **17**, 3325–3371. (doi:10.1142/S0218127407019093)
4. Porta A, Aletti F, Vallais F, Baselli G. 2009 Multimodal signal processing for the analysis of cardiovascular variability. *Phil. Trans. R. Soc. A* **367**, 391–409. (doi:10.1098/rsta.2008.0229)
5. Bravi A, Longtin A, Seely A. 2011 Review and classification of variability analysis techniques with clinical applications. *Biomed. Eng. Online* **10**, 90. (doi:10.1186/1475-925X-10-90)
6. Kurths J, Voss A, Saparin P, Witt A, Kleiner HJ, Wessel N. 1995 Quantitative analysis of heart rate variability. *Chaos* **5**, 88–94. (doi:10.1063/1.166090)
7. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim C, Suri J. 2006 Heart rate variability: a review. *Med. Biol. Eng. Comput.* **44** 1031–1051. (doi:10.1007/s11517-006-0119-0)
8. Elstad M, Toska K, Chon KH, Raeder EA, Cohen RJ. 2001 Respiratory sinus arrhythmia: opposite effects on systolic and mean arterial pressure in supine humans. *J. Physiol.* **536**, 251–259. (doi:10.1111/j.1469-7793.2001.t01-1-00251.x)
9. Karemaker J, Wesseling K. 2008 Variability in cardiovascular control: the baroreflex reconsidered. *Cardiovasc. Eng.* **8**, 23–29. (doi:10.1007/s10558-007-9046-4)
10. Di Rienzo M, Parati G, Radaelli A, Castiglioni P. 2009 Baroreflex contribution to blood pressure and heart rate oscillations: time scales, time-variant characteristics and nonlinearities. *Phil. Trans. R. Soc. A* **367**, 1301–1318. (doi:10.1098/rsta.2008.0274)
11. Cohen MA, Taylor JA. 2002 Short-term cardiovascular oscillations in man: measuring and modelling the physiologies. *J. Physiol.* **542**, 669–683. (doi:10.1113/jphysiol.2002.017483)
12. Eckberg D. 2008 Arterial baroreflexes and cardiovascular modeling. *Cardiovasc. Eng.* **8**, 5–13. (doi:10.1007/s10558-007-9042-8)
13. Riedl M, Suhrbier A, Malberg H, Penzel T, Bretthauer G, Kurths J, Wessel N. 2008 Modeling the cardiovascular system using a nonlinear additive autoregressive model with exogenous input. *Phys. Rev. E* **78**, 011919. (doi:10.1103/PhysRevE.78.011919)
14. Sibai B, Dekker G, Kupferminc M. 2005 Pre-eclampsia. *Lancet* **365**, 785–799. (doi:10.1016/S0140-6736(05)17987-2)

15. Rana S, Karumanchi SA, Levine RJ, Venkatesha S, Rauh-Hain JA, Tamez H, Thadhani R. 2007 Sequential changes in antiangiogenic factors in early pregnancy and risk of developing preeclampsia. *Hypertension* **50**, 137–142. (doi:10.1161/HYPERTENSIONAHA.107.087700)
16. Ohkuchi A, Hirashima C, Matsubara S, Takahashi K, Matsuda Y, Suzuki M. 2011 Threshold of soluble fms-like tyrosine kinase 1/placental growth factor ratio for the imminent onset of preeclampsia. *Hypertension* **58**, 859–866. (doi:10.1161/HYPERTENSIONAHA.111.174417)
17. Siddiqui AH, Irani RA, Blackwell SC, Ramin SM, Kellems RE, Xia Y. 2010 Angiotensin receptor agonistic autoantibody is highly prevalent in preeclampsia. *Hypertension* **55**, 386–393. (doi:10.1161/HYPERTENSIONAHA.109.140061)
18. Carty DM *et al.* 2011 Urinary proteomics for prediction of preeclampsia. *Hypertension* **57**, 561–569. (doi:10.1161/HYPERTENSIONAHA.110.164285)
19. Maulik D, Figueroa R. 2005 Doppler velocimetry for fetal surveillance: Randomized clinical trials and implications for practice. In *Doppler ultrasound in obstetrics and gynecology* (ed. D Maulik), pp. 387–401. Berlin, Germany: Springer.
20. Malberg H, Bauernschmitt R, Voss A, Walther T, Faber R, Stepan H, Wessel N. 2007 Analysis of cardiovascular oscillations: a new approach to the early prediction of pre-eclampsia. *Chaos* **17**, 015113. (doi:10.1063/1.2711660)
21. Walther T, Wessel N, Malberg H, Voss A, Stepan H, Faber R. 2006 A combined technique for predicting pre-eclampsia: concurrent measurement of uterine perfusion and analysis of heart rate and blood pressure variability. *J. Hypertens.* **24** 747–750. (doi:10.1097/01.hjh.0000217858.27864.50)
22. Stepan H, Unversucht A, Wessel N, Faber R. 2007 Predictive value of maternal angiogenic factors in second trimester pregnancies with abnormal uterine perfusion. *Hypertension* **49**, 818–824. (doi:10.1161/01.HYP.0000258404.21552.a3)
23. Stepan H, Geipel A, Schwarz F, Krämer T, Wessel N, Faber R. 2008 Circulatory soluble endoglin and its predictive value for preeclampsia in second-trimester pregnancies with abnormal uterine perfusion. *Am. J. Obstet. Gynecol.* **198**, 175.e1–175.e6. (doi:10.1016/j.ajog.2007.08.052)
24. Giguere Y, Charland M, Bujold E, Bernard N, Grenier S, Rousseau F, Lafond J, Legare F, Forest J-C. 2010 Combining biochemical and ultrasonographic markers in predicting preeclampsia: a systematic review. *Clin. Chem.* **56**, 361–375. (doi:10.1373/clinchem.2009.134080)
25. Schreiber T. 1999 Interdisciplinary application of nonlinear time series methods. *Phys. Rep.* **308**, 1–64. (doi:10.1016/S0370-1573(98)00035-0)
26. Kantz H, Schreiber T. 2003 *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.
27. Maraun D, Kurths J. 2004 Cross wavelet analysis: significance testing and pitfalls. *Nonlin. Proc. Geophys.* **11**, 505–514. (doi:10.5194/npg-11-505-2004)
28. Marwan N, Romano MC, Thiel M, Kurths J. 2007 Recurrence plots for the analysis of complex systems. *Phys. Rep.* **438**, 237–329. (doi:10.1016/j.physrep.2006.11.001)
29. Hastie T, Tibshirani R, Friedman J. 2008 *The elements of statistical learning: data mining, inference, and prediction*. Berlin, Germany: Springer.
30. Brocker J, Engster D, Parlitz U. 2009 Probabilistic evaluation of time series models: a comparison of several approaches. *Chaos* **19**, 043130. (doi:10.1063/1.3271343)
31. Donner RV, Zou Y, Donges JF, Marwan N, Kurths J. 2010 Recurrence networks – a novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**, 033025. (doi:10.1088/1367-2630/12/3/033025)
32. Lacasa L, Toral R. 2010 Description of stochastic and chaotic series using visibility graphs. *Phys. Rev. E* **82**, 036120. (doi:10.1103/PhysRevE.82.036120)
33. Donges JF, Donner RV, Trauth MH, Marwan N, Schellnhuber HJ, Kurths J. 2011 Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proc. Natl Acad. Sci. USA* **108**, 20 422–20 427. (doi:10.1073/pnas.1117052108)
34. Zbilut JP, Thomasson N, Webber CL. 2002 Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Med. Eng. Phys.* **24**, 53–60. (doi:10.1016/s1350-4533(01)00112-6)
35. Marwan N, Wessel N, Meyerfeldt U, Schirdewan A, Kurths J. 2002 Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Phys. Rev. E* **66**, 026702. (doi:10.1103/PhysRevE.66.026702)

36. Marwan N, Donges JF, Zou Y, Donner RV, Kurths J. 2009 Complex network approach for recurrence analysis of time series. *Phys. Lett. A* **373**, 4246–4254. (doi:10.1016/j.physleta.2009.09.042)
37. Shiogai Y, Stefanovska A, McClintock PVE. 1999 Nonlinear dynamics of cardiovascular ageing. *Phys. Rep.* **488**, 51–110. (doi:10.1016/j.physrep.2009.12.003)
38. Fujiwara N, Kurths J, Díaz-Guilera A. 2011 Synchronization in networks of mobile oscillators. *Phys. Rev. E* **83**, 025101. (doi:10.1103/PhysRevE.82.036120)
39. Nicolaides C, Cueto-Felgueroso L, Juanes R. 2010 Anomalous physical transport in complex networks. *Phys. Rev. E* **82**, 055101. (doi:10.1103/PhysRevE.82.055101)
40. Weber S, Porto M. 2006 Multicomponent reaction–diffusion processes on complex networks. *Phys. Rev. E* **74**, 046108. (doi:10.1103/PhysRevE.74.046108)
41. Chang S, Jiao X, Gong X-q, Li C-h, Chen W-z, Wang C-x. 2008 Evolving model of amino acid networks. *Phys. Rev. E* **77**, 061920. (doi:10.1103/PhysRevE.77.061920)
42. Glattfelder JB, Battiston S. 2009 Backbone of complex networks of corporations: the flow of control. *Phys. Rev. E* **80**, 036104. (doi:10.1103/PhysRevE.80.036104)
43. Klemm K, Eguíluz VM, Toral R, San Miguel M. 2003 Nonequilibrium transitions in complex networks: a model of social interaction. *Phys. Rev. E* **67**, 026120. (doi:10.1103/PhysRevE.67.026120)
44. Albert R, Barabasi A-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)
45. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. 2006 Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308. (doi:10.1016/j.physrep.2005.10.009)
46. Donner RV, Small M, Donges JF, Marwan N, Zou Y, Xiang R, Kurths J. 2011 Recurrence-based time series analysis by means of complex network methods. *Int. J. Bifurcation Chaos* **21**, 1019. (doi:10.1142/S0218127411029021)
47. Poincaré H. 1890 Sur le problème des trois corps et les équations de la dynamique. *Acta Math.* **13**, 1–270.
48. Eckmann JP, Kamphorst SO, Ruelle D. 1987 Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973–977. (doi:10.1209/0295-5075/4/9/004)
49. Donner RV, Heitzig J, Donges JF, Zou Y, Marwan N, Kurths J. 2012 The geometry of chaotic dynamics—a complex network perspective. *Eur. Phys. J. B* **84**, 653–672. (doi:10.1140/epjb/e2011-10899-1)
50. Seidman SB. 1983 Network structure and minimum degree. *Soc. Netw.* **5**, 269–287. (doi:10.1016/0378-8733(83)90028-x)
51. Batagelj V, Zaveršnik M. 2002 *An O(m) algorithm for cores decomposition of networks*. University of Ljubljana, Institute of Mathematics, Physics and Mechanics, Department of Mathematics. Q3
52. Watts DJ, Strogatz SH. 2001 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)
53. Amaral LAN, Ottino JM. 2004 Complex networks. *Eur. Phys. J. B* **38**, 147–162. (doi:10.1140/epjb/e2004-00110-5)
54. Riedl M, Suhrbier A, Stepan H, Kurths J, Wessel N. 2010 Short-term couplings of the cardiovascular system in pregnant women suffering from pre-eclampsia. *Phil. Trans. R. Soc. A* **368**, 2237–2250. (doi:10.1098/rsta.2010.0029)
55. Nwb team. 2006. Network Workbench tool. Indiana University, Northeastern University and University of Michigan. See http://nwb.slis.indiana.edu.
56. Altman DG, Bland JM. 1994 Statistics notes: diagnostic tests 1: sensitivity and specificity. *Br. Med. J.* **308**, 1552. (doi:10.1136/bmj.308.6943.1552)
57. Altman DG, Bland JM. 1994 Statistics notes: diagnostic tests 2: predictive values. *Br. Med. J.* **309**, 102. (doi:10.1136/bmj.309.6947.102)
58. Banek CT, Gilbert JS. 2011 Approaching the threshold for predicting preeclampsia. *Hypertension* **58**, 774–775. (doi:10.1161/HYPERTENSIONAHA.111.175992)
59. Donner RV, Zou Y, Donges JF, Marwan N, Kurths J. 2010 Ambiguities in recurrence-based complex network representations of time series. *Phys. Rev. E* **81**, 015101. (doi:10.1103/PhysRevE.81.015101)
60. Porta A, Bassani T, Bari V, Pinna GD, Maestri R, Guzzetti S. 2012 Accounting for respiration is necessary to reliably infer Granger causality from cardiovascular variability series. *IEEE Trans. Biomed. Eng.* **59**, 832–841. (doi:10.1109/tbme.2011.2180379)
61. Sugihara G, May R, Ye H, Hsieh Ch, Deyle E, Fogarty M, Munch S. 2012 Detecting causality in complex ecosystems. *Science* **338**, 496–500. (doi:10.1126/science.1227079)

62. Xu X, Zhang J, Small M. 2008 Superfamily phenomena and motifs of networks induced from time series. *Proc. Natl Acad. Sci. USA* **105**, 19 601–19 605. (doi:10.1073/pnas.0806082105)
63. Kenny LC *et al.* 2010 Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. *Hypertension* **56**, 741–749. (doi:10.1161/HYPERTENSIONAHA.110.157297)
64. Roberts JM, Cooper DW. 2001 Pathogenesis and genetics of pre-eclampsia. *Lancet* **357**, 53–56. (doi:10.1016/S0140-6736(00)03577-7)
65. Wikström A-K, Stephansson O, Cnattingius S. 2010 Tobacco use during pregnancy and preeclampsia risk. *Hypertension* **55**, 1254–1259. (doi:10.1161/HYPERTENSIONAHA.109.147082)
66. Magnus P, Eskild A. 2001 Seasonal variation in the occurrence of pre-eclampsia. *BJOG: Int. J. Obstet. Gynaecol.* **108**, 1116–1119. (doi:10.1111/j.1471-0528.2003.00273.x)