Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 89–96

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

Time Series Analysis using Embedding Dimension on Heart Rate Variability

Ronakben Bhavsar[a,*], Neil Davey[a], Na Helian[a], Yi Sun[a], Tony Steffert[b], David Mayor[a]

[a]*University of Hertfordshire, Hatfield, AL10 9AB, UK*
[b]*The Open University, Milton Keynes, MK7 6AA, UK*

Abstract

Heart Rate Variability (HRV) is the measurement sequence with one or more visible variables of an underlying dynamic system, whose state changes with time. In practice, it is difficult to know what variables determine the actual dynamic system. In this research, Embedding Dimension (ED) is used to find out the nature of the underlying dynamical system. False Nearest Neighbour (FNN) method of estimating ED has been adapted for analysing and predicting variables responsible for HRV time series. It shows that the ED can provide the evidence of dynamic variables which contribute to the HRV time series. Also, the embedding of the HRV time series into a four-dimensional space produced the smallest number of FNN. This result strongly suggests that the Autonomic Nervous System that drives the heart is a two features dynamic system: sympathetic and parasympathetic nervous system.

Keywords: Time series analysis; HRV; Embedding Dimension; False Nearest Neighbours; Parasympathetic; Sympathetic; Linear Regression.

1. Introduction and Related Work

Heart Rate Variability (HRV) can be measured using Electrocardiography (ECG). ECG records the electrical activities of the heart, where each beat of the heart is initiated by an electric signal from the heart muscle (Vague). HRV is the estimation of neurocardiac function that reflects heart-brain interactions and autonomic nervous system dynamics [12]. The measurement of HRV is a valuable investigative tool in clinical cardiology as it gives a fundamental method to evaluate the physiological state of the heart directly. Many neurological and psychological investigations have used HRV to assess the effects of stress, emotion, and work on the autonomic nervous system [11]. The heart rate and rhythm are mainly under the control of the Autonomic Nervous System (ANS) and is the part of the Peripheral

* Corresponding author. Tel.: +44-757-436-9688.
E-mail address: r.bhavsar2@herts.ac.uk

10.1016/j.procs.2018.11.015

Nervous System (PNS) that acts as a control system that functions mostly below the level of consciousness to control physical functions. ANS contains two primary role on components: Sympathetic and Parasympathetic Nervous system. Both the sympathetic and parasympathetic nervous systems innervate the heart. The parasympathetic nervous system functions in regulating heart rate through the vagus nerve, with increased vagal activity producing a slowing of heart rate. The sympathetic nervous system has an excitatory influence on heart rate and contractility, and it serves as the final common pathway for controlling the smooth muscle tone of the blood vessels [15].

Time series, such as HRV, is the measurements sequence of one or more visible variables of an underlying dynamic system, whose state changes with time. These time series will be the results of the interaction of many underlying variables. For example, a stock market is affected by many interacting factors, such as economic data, exchange rates and so on. In practice, it is difficult to know what variables determine the actual dynamic system. It is shown by [16], if only one scalar value can be measured from an active system, then by windowing a sufficient number of consecutive values, the nature of the original multivariable dynamic system can be recaptured. In fact, [16] also mentioned if the original dynamic system had a dimension of N, then an embedding of size 2N will be fully regained the original system. The size of this window is called the Embedding Dimensions (ED) [1]. ED estimation has successfully used in neural network approaches for time series prediction [7]. They concluded that optimal performance could be achieved using the correct ED. Furthermore, ED has been adopted by [18], for recurrence plot generation from the reconstructed phase space to represent many real application scenarios when not all variables to describe a system were available. The number of independent variables sufficient for modeling the hair cell response has been estimated utilizing ED approach [6]. Moreover, ED has been considered in a multilayer perceptron neural network to measure hyperchaotic Rssler system state variables [2].

The ED plays a vital role in nonlinear time series analysis[5], as discussed earlier. With its extensive use of finding the nature of an underlying dynamical system, ED is used in this work for HRV time series analysis. The False Nearest Neighbours (FNN) method of estimating ED has been adapted for analysing and predicting variables responsible for the HRV time series. HRV time series taken from participants over a fixed period.

2. Dataset Information

The HRV time series used in this work are of the data where participants undergoing acupuncture in the experimental settings. Three different methods of acupuncture have been used: Electro-Acupuncture(EA), Transcutaneous Electro-Acupuncture (TEA) and Manual Acupuncture (MA). EA is a method of inserting needles at specific points on the body. The needles have then been connected to a device that generates continuous electric pulses. These devices are used to adjust the frequency and intensity of the delivered impulse [3]. TEA is a safe, standardized acupuncture technique in which there is no needle insertion. It involves applying cutaneous electrical stimulation by placing electrodes at classical Chinese acupoints [14]. The electrodes (patches) are attached to the participant's skin when the unit is switched on; mild electrical current travels through the electrodes wires into the body. MA is an acupuncture method similar to EA, in which needles have inserted at specific points on the body. These needles are then twisted by or otherwise manipulated by the acupuncturist instead of passing electric pulses through the needle [17].

2.1. Dataset 1

This dataset consists of HRV data of 7 participants, derived using TEA method of acupuncture. HRV monitoring was carried out in nine 5-minute slots: three baseline slots and six acupuncture stimulation slots. The stimulation parameters (e.g., body location) are kept constant within each intervention but varies between interventions. Each participant visited twice, during which the TEA stimulation of either 2.5Hz or 10Hz is applied (randomised order used) at six different body locations (Slot 3 to 8) with eyes closed. The baseline measurements are slots 1, 2, and 9.

1. No stimulation with Eyes Closed.
2. No stimulation with Eyes Open.
3. TEA stimulation on Left Hand and below Left Knee.
4. TEA stimulation on Right Hand and below Right Knee.
5. TEA stimulation on 3. and 4. together.

6. TEA stimulation on Upper Body (Left and Right Hands).
7. TEA stimulation on Lower Body (below Left and Right Knees).
8. TEA stimulation on 6. and 7. Together.
9. No stimulation with Eyes Open.

2.2. Dataset 2

This dataset consists of HRV data of 12 participants, derived using both EA and MA method of acupuncture. All participants attended for four visits, during which stimulation performed at a different location (in randomized order): Right (Below Right Knee and Right Hand), Left (below Left Knee and Left Hand), Upper Body (Right and Left Hands) and Lower Body (below Left and Right Knees). HRV monitoring was carried out in eight 5-minute sequential slots with stimulation at a single location: EA stimulation of 2.5Hz, 10Hz, 20Hz and 80Hz is applied (Slot 3 to 6), MA stimulation applied in two slots (Slot 2 and Slot 7), and baseline measurements are slots 1 and slot 8.

1. No stimulation with Eyes Open 1
2. MA Stimulation 1
3. EA Stimulation at 2.5Hz
4. EA Stimulation at 10Hz
5. EA Stimulation at 20Hz
6. EA Stimulation at 80Hz
7. MA Stimulation 2
8. No stimulation with Eyes Open 2

3. Embedding Dimension

ED is used to find out the nature of an underlying dynamical system. The method FNN is used to determine how many dimensions are sufficient to embed a particular time series [8]. The FNN is designed to determine how many features are enough to integrate a specific time series [8]. The basic idea behind FNN is that points in a state space should be close to each other because their dynamical state is similar, not because they have been projected close to each other as an artefact of constructing the embedding with a dimension which is too low. In an embedding of dimension D, each point is established as a vector.

The FNN algorithm can be summarized as follows:

1. Find the nearest neighbour for each point in an embedding of dimension D;
2. Find the percentage of those nearest neighbours which do not remain the nearest neighbour within embedding of dimension D+1, such points turns as false nearest neighbours;
3. Increase the embedding dimension until the number of false nearest neighbour is sufficiently small.

3.1. An Example of Embedding Dimension Calculation

In order to find the correct embedding dimension, n, an incremental search, from $n = 1$, is performed. A set of time lagged vectors x_n , for a given n, is formed. The nearest neighbour relation within the set of x_n's is then computed. When the correct value of n has been reached, the addition of an extra dimension to the embedding should not cause these nearest neighbours to spring apart. Any pair whose additional separation is of a high relative size is deemed FNN. Specifically, if x_n has nearest neighbour $\widetilde{x}_n$, then the relative additional separation when the embedding dimension is incremented is given by [1]:

$$FNN = \left| \frac{d(x_n, \widetilde{x}_n) - d(x_{n+1}, \widetilde{x}_{n+1})}{d(x_n, \widetilde{x}_n)} \right|, \tag{1}$$

When this value exceeds an absolute value, then x_n and $\widetilde{x}_n$ are denoted as FNN. Where, x is the time series, n is the index for x, and d is the euclidean distance. In roder to calculate nearness of neighbours, Euclidean Distance is used [10]. For example a time series is a sequence of values $x_n(t)$, where x is the time series, n is the index for x, and t represent time. Theoretically, x may be a value which varies continuously with t. An ED of 2 forms vectors $(x_0, x_1), (x_1, x_2)$ and so on. An ED of 3 forms the vectors (x_0, x_1, x_2) and so on. Since this is numeric vectors, the distance apart of any pair of these vectors could be calculated. So for each vector in a given embedding, the nearest neighbour can be found. However, some of these nearest neighbour may be false neighbour, in a sense that they are not nearest neighbour in the embedding with one extra dimension [7]. A geometric explanation of the concept that is at the core of the FNN technique is as shown in Fig.1.

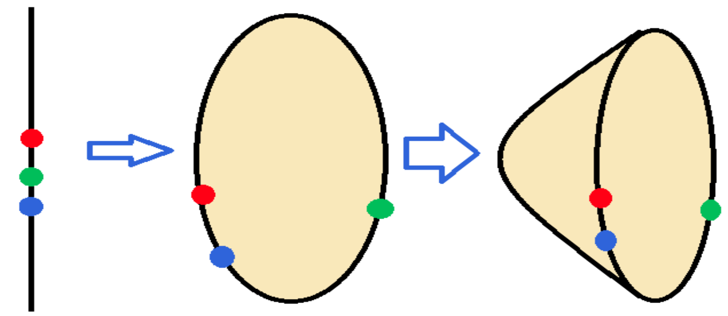

Figure 1. Geometric explanation of the FNN Algorithm [8]: (a) In one-dimensional, Red and Green Point are nearest neighbour, (b) In two-dimensional Red and Green point are not nearest neighbour (i.e. false nearest neighbour), but Red and Blue are nearest neighbour, and (c) In three-dimensional Red and Blue are still nearest neighbour, so they are real nearest neighbour.

The line at the bottom represents a dimensional state space X_1(Red Point), X_2 (Green Point), and X_3 (Blue Point)) and the nearest neighbour of the X_1 (Red Point) is the X_2 (Green Point). Next, the time series embedded into two-dimensional state space, represented by the oval in the middle of the picture. The X_1 (Red Point) and X_2 (Green point) are no longer near to each other. So, the X_2(Green point) is labelled as a false nearest neighbour because it was only near to the X_1 (Red Point) due to the projection of the time series onto the line.

Next, the nearest neighbour for each point in the two-dimensional state space found. Now the nearest neighbour to the X_1 (Red Point) is the X_3 (Blue Point). The time series is now embedded into a three-dimensional state space as represented in the rotated parabola at the top of the picture. The X_1 (Red Point) and the X_3 (Blue Point) are still near to each other, and so the X_3 (Blue Point) is not a false nearest neighbour. This process continues until either there are no further malicious nearest neighbour or the data set becomes so sparse in a high dimensional space that no points can be considered to be near neighbours, to begin with. The resulting percentage of FNN for each ED is then plotted against the corresponding ED to create FNN plot.

3.2. An Example of Lorenz Attractor

The well known Lorenz Attractor as shown in Fig.2. has three underlying cross-coupled variables. However, the attractor itself is almost two dimensional. The minimum dimension of the attractor for a Lorenz data set as shown in Fig.3. is 4 or 5, which suggests that the actual underlying dimension system has features of around 2 [7].

4. Experiments and Results

4.1. HRV Analysis using ED

For analysing the HRV time series using the ED, for each participant in each 5-minute slot, the number of FNN has been calculated as the ED has increased. MATLAB code is used to gather ED performance of HRV [13]. Once, the

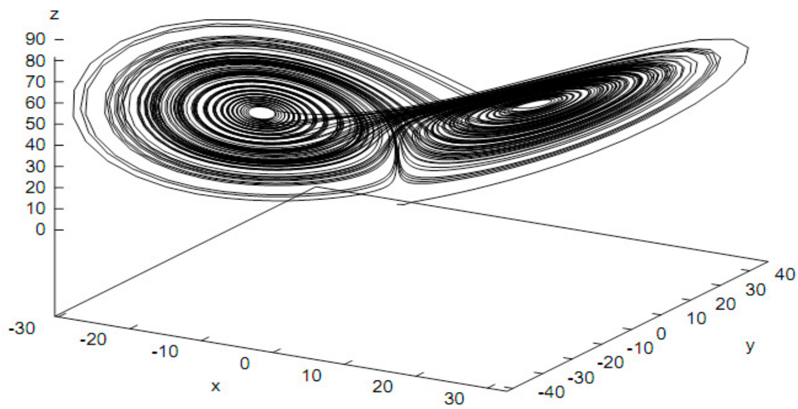

Figure 2. A visualisation of the Lorenz attractor in 3-dimensional phase space x(t), y(t),z(t)[7].

Embedding Dimension	*Percentage of False Nearest Neighbours*
	Clean Data
2	77%
3	3.3%
4	0.3%
5	0.3%

Figure 3. The percentage of false nearest neighbours in the Lorenz data set [7].

percentage of FNN for the ED is gathered for each 5-minute slot, the result is plotted for each participant, as shown in Fig.4. A comparison is made on the changes of various stimulus locations for each participant (i.e., between each slot's result), and among the ED result of different participants. The first notable result for all participants is that the optimal ED is about 4. It is also noteworthy that this optimal ED is independent of stimulus location.

The ED results of 4 participants, two from each dataset (Participant 1 and Participant 2 from Dataset 1, and Participant 3 and Participant 4 from Dataset 2) as shown in Fig.4. For all the other participants from both datasets, results are similar to as shown in Fig.4. It is important to note that, some health-related problems found for few participants. For example, Participant 1 from dataset 1 has Thyroid insufficiency and menstrual irregularity, and participant 4 from dataset 2 have Asthma.

In the Fig.4., the X axes (Horizontally) represent ED from 1 to 10, and Y axes (Vertically) represent the percentage of FNN for the ED 1-10. The 8-9 different colors (Curves) in the graph correspond to slots containing baseline and acupuncture stimulation locations for an individual participant. There are two important findings from these results:

1. The first notable result for all participants is that the optimal ED is about 4. It is also notable that this optimal ED is independent of the specific stimulus location. This result suggests that an ED of 4 or 5 is most appropriate for HRV data for all slots and all participants.
2. In Fig.4., right-hand side figures (Participant 2 and Participant 4) shows increasing numbers of false nearest neighbours as ED increases above its optimal value of 4, whereas left-hand side figures (Participant 1 and Participant 3) does not display this. An increase in the number of false nearest neighbours with increasing ED is normally suggestive of noise in the data [1].

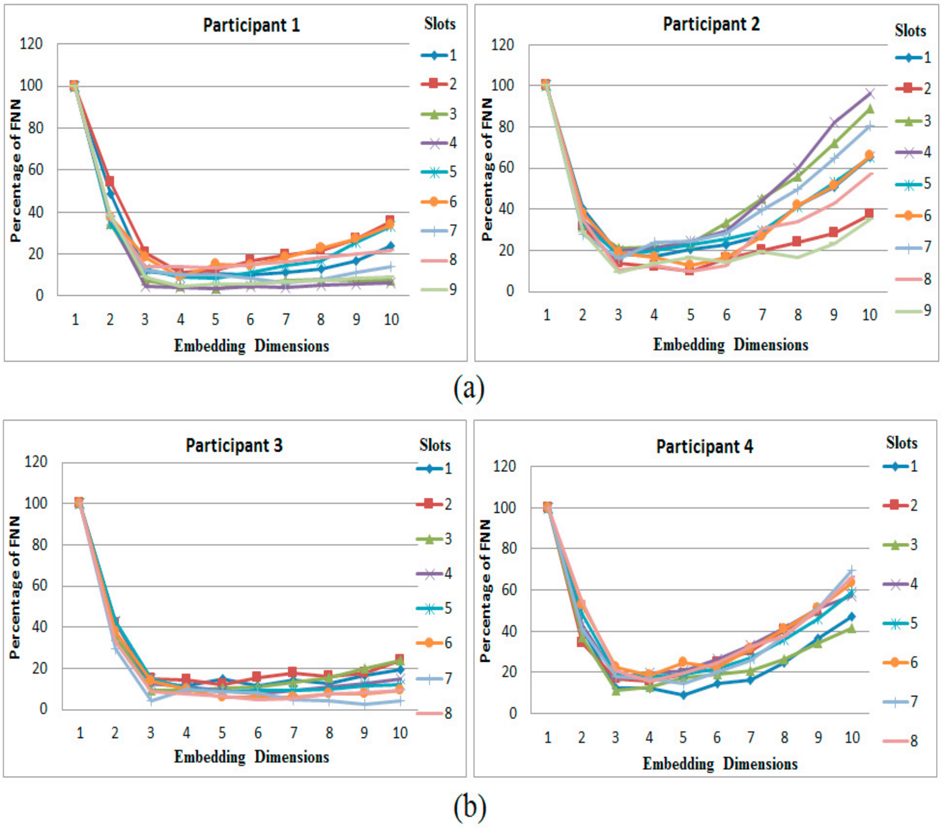

Figure 4. Embedding Dimension Result: (a) Two participant's (Participant 1 and 2) ED result from Dataset 1, and (b) Two participant's (Participant 3 and 4) ED result from Dataset 2.

4.2. HRV Prediction using Linear Regression

The standard Linear Regression is utilized to produce a linear predictor for our embedded data. It was suggested by [9], that the performance of a time series predictor is affected by the size of a window, in which the time series embedded. Therefore, Linear Regression is used to predict the window size for HRV series to achieve the best result. For each window, data is split into a training set of 250 vectors and a test set of 106 vectors. The size of the embedding varies between 2 and 6. Errors are calculated as relative error [4], and results for five different window size has shown in Fig.5. It is clear that the best regressor has four inputs, and changing this number either way harms the performance.

The linear prediction of the HRV suggests that window size of 4 will be enough to fit the HRV time series data. Also, this reflects the ED result 4 is the minimum ED for the HRV Data as shown in Fig.4.

5. Discussion and Conclusion

Our result indicating that the HRV has an estimated ED of 4 suggests that the underlying dynamic system has 2 features; based on [16], if the original dynamic system had a dimension of N, then an embedding of size 2N will be fully regained the original system. This result is interesting because HRV is driven by two underlying systems, the sympathetic and parasympathetic neural pathways; HRV is a marker of sympathetic and parasympathetic influences on the modulation of heart rate [19], and this reflects in the ED result. The effect of the sympathetic pathway is to increase heart rate and blood pressure (Fight or Flight response), whereas the parasympathetic path acts to decrease heart rate and blood pressure (Rest and Digest response). Therefore, the main finding here is that in all circumstances an Embedding of the HRV time series into a four-dimensional space produced the smallest number of false nearest neighbours.

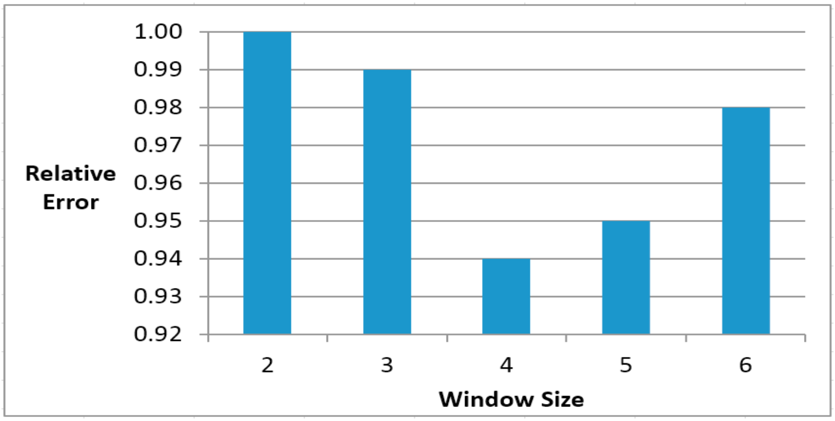

Figure 5. Relative error of the predictor for varying window size for HRV Time series.

This finding strongly suggests that the Autonomic Nervous System that drives the heart is a two-dimensional dynamic system.

From the participant's questionnaire, a variety of subjective responses to the acupuncture stimulation found. However, this did not appear to have much effect on the HRV time series, which robustly kept its two-dimensional dynamic system.

In some circumstances, the number of FNN increase as the ED became more massive than the optimal value. This increase is suggestive of noise in the data that may have come from the ECG measuring equipment.

It was suggested by [9], that the performance of a time series predictor is affected by the window size, in which the time series embedded. So that the best predictor would be the one that used correct ED. Our experiments, reported here, using a Linear Regression to predict the HRV series confirmed this as a window size of four gave the best result.

References

[1] Henry DI Abarbanel, Reggie Brown, John J Sidorowich, and Lev Sh Tsimring. The analysis of observed chaotic data in physical systems. *Reviews of modern physics*, 65(4):1331, 1993.

[2] Massimo Camplani and Barbara Cannas. The role of the embedding dimension and time delay in time series forecasting. *IFAC Proceedings Volumes*, 42(7):316–320, 2009.

[3] Lynn Casimiro, Les Barnsley, Lucie Brosseau, Sarah Milne, Vivian Welch, Peter Tugwell, and George A Wells. Acupuncture and electroacupuncture for the treatment of rheumatoid arthritis. *The Cochrane Library*, 2005.

[4] Chris Chatfield and Andreas S Weigend. Time series prediction: Forecasting the future and understanding the past: Neil a. gershenfeld and andreas s. weigend, 1994,the future of time series, in: As weigend and na gershenfeld, eds.,(addison-wesley, reading, ma), 1-70., 1994.

[5] Bian Chun-Hua and Ning Xin-Bao. Determining the minimum embedding dimension of nonlinear time series based on prediction method. *Chinese Physics*, 13(5):633, 2004.

[6] Justin Faber and Dolores Bozovic. Chaotic dynamics of inner ear hair cells. *Scientific reports*, 8(1):3366, 2018.

[7] Ray J Frank, Neil Davey, and Stephen P Hunt. Time series prediction and neural networks. *Journal of Intelligent and Robotic Systems*, 31(1-3):91–103, 2001.

[8] Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6):3403, 1992.

[9] Rhee M Kil, Seon Hee Park, and Seunghwan Kim. Optimum window size for time series prediction. In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, volume 4, pages 1421–1424. IEEE, 1997.

[10] Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[11] Marek Malik and A John Camm. Heart rate variability. *Clinical cardiology*, 13(8):570–576, 1990.

[12] Rollin McCraty, Mike Atkinson, Dana Tomasino, and William P Stuppy. Analysis of twenty-four hour heart rate variability in patients with panic disorder. *Biological psychology*, 56(2):131–150, 2001.

[13] Mirwais. Finds minimum Embedding Dimension with false nearest neighbours method. https://uk.mathworks.com/matlabcentral/fileexchange/37239-minimum-embedding-dimension, 2012. [Online; accessed 18-May-2018].

[14] MML Ng, Mason CP Leung, and DMY Poon. The effects of electro-acupuncture and transcutaneous electrical nerve stimulation on patients with painful osteoarthritic knees: a randomized controlled trial with follow-up evaluation. *The Journal of Alternative & Complementary Medicine*, 9(5):641–649, 2003.

[15] Brian F Robinson, Stephen E Epstein, G David Beiser, and Eugene Braunwald. Control of heart rate by the autonomic nervous system: studies in man on the interrelation between baroreceptor mechanisms and exercise. *Circulation Research*, 19(2):400–411, 1966.

[16] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[17] Elizabeth A Tough, Adrian R White, T Michael Cummings, Suzanne H Richards, and John L Campbell. Acupuncture and dry needling in the management of myofascial trigger point pain: a systematic review and meta-analysis of randomised controlled trials. *European Journal of Pain*, 13(1):3–10, 2009.

[18] Dadiyorto Wendi, Norbert Marwan, and Bruno Merz. In search of determinism-sensitive region to avoid artefacts in recurrence plots. *International Journal of Bifurcation and Chaos*, 28(01):1850007, 2018.

[19] Yuru Zhong, Hengliang Wang, Ki Hwan Ju, Kung-Ming Jan, and Ki H Chon. Nonlinear analysis of the separate contributions of autonomic nervous systems to heart rate variability using principal dynamic modes. *IEEE transactions on biomedical engineering*, 51(2):255–262, 2004.