Technical report: Multivariate signal analysis by recurrence plots

Tamara Tosic, Axel Hutt

Technical report: Multivariate signal analysis by recurrence plots

Tamara Tošić and Axel Hutt
INRIA Nancy Grand-Est, team NEUROSYS, Villers-lés-Nancy, France

Abstract

This paper investigates methods for analysing dynamics of biological systems characterised by multimodal signals of different dimensionality, such as polysomnography (sleep) signals or population tracking. We study statistical properties of low dimensional representations of multimodal signals, which we want to use as features for detection of intrinsic signal state changes. We assume that distinctive state transition features remain preserved in low dimensional signal representations of dynamical system signals. We first build recurrence plots for well-behaved biological systems (predator-prey model and Lorenz system which models a forward osmosys). Then, we provide recurrence plot analysis for the real dataset (visual stimulation of ferrets) and we show that preliminary statistical analysis of ferret datasets has stable characteristics. This motivates us to further pursue study on sleep signal statistics that we plan to use for sleep stage detection.

Index Terms

Dynamical systems, Recurrence plots, Detection, Multimodal systems, Wavelets, Synchrosqueezing method.

I. Introduction

Understanding dynamics of non-linear phenomena such as biophysiological signals in humans or population dynamics nowadays became one of the most studied research problems. Appropriate tools for representation and tracking of nonlinear patterns is very important, because nonlinear phenomena occur very often in nature. Nonlinear signals are commonly found in numerous research fields, such as meteorology, social sciences and economy, to name a few.

Analysis and processing of multimodal sensor signals faces numerous challenges, such as reducing a signal dimensionality while preserving the information significant for the detection task or building methods robust to acquisition or sensor synchronisation noise. More information on state-of-the art nonlinear dynamical analysis of multimodal systems is available in [1] and references within.

The aim of this project is to jointly analyse different physiological signals for detecting transitions between system state changes. One of the most challenging signals are polysomnography or sleep signals. Polysomnography signals are captured in the partner hospital CHU Nancy (Service de Neurologie). A standardised method for recording sleep signals, known as a polysomnogram (PSG), captures a vast amount of heterogeneous data. In particular, PSG monitors a set of biophysiological changes that arise during sleep and captures brain activity (EEG - electroencephalogram), heart activity (ECG - electrocardiogram), eye movements (EOG - electrooculogram) and other body function signals. More details on PSG sleep study protocol and measures used for sleep scoring is given in [2] and references within. In the literature, most of the works on sleep signal processing consider a classification of sleep into different stages. Depending on the purpose, classical classifiers (nearest neighbour, neural networks, one vs. all, etc.) use features recommended by the sleep scoring protocol.

In this work, we study recurrence plots given several common nonlinear biological models by tracing the dynamical transitions of the signals. Commonly, multimodal signals have different dimensions and their correlation models are hard to grasp. To efficiently deal with these problems, we propose to build signal features based on fundamental properties of dynamical systems and to analyse them jointly. In particular, we focus on analysing statistical properties of recurrence plots.

Recurrence is one of fundamental properties of dynamical systems which characterises the behaviour of the system in the phase space [3]. It represents an instance in time when the signal trajectory returns to a neigborhood of a location it has already visited previously. Deterministic dynamical systems are described by its trajectory. A signal $\mathbf{x}_i \in \mathbb{R}^n$ is sampled at times $t = i\Delta t$, $i \in \{1, 2, \ldots, N\}$, where Δt is the sampling rate and N is the total number of samples. Then, the recurrence plot is defined as $N \times N$ matrix $\mathbf{R}$ whose elements $r_{i,j}$ take nonzero values if the trajectories of their elements lay within the open ball $\mathcal{B}(\epsilon)$ of radius ϵ:

$$r_{i,j} = \begin{cases} 1, & \text{if } d(\mathbf{x}_i, \mathbf{x}_j) < \epsilon, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $d(\cdot, \cdot)$ is an arbitrary distance function and $i, j \in \{1, \ldots, N\}$. In other words, recurrence plots represent two-dimensional outputs of a nonlinear transform that have more efficient analysis and storage. However, we can not guarantee that some of the significant features are not discarded by such thresholding. To minimize the potential thresholding error for a given distance function, we compute optimal ϵ that maximises a symbolic entropy, as in [4]. Under the assumption that signal states have uniform distribution for a given the recurrence plot, the algorithm builds disjunct and transitive symbolic matrices for each ϵ and computes its entropy as

$$H(\epsilon) = -\frac{1}{S_{k,\epsilon}} \sum_{k=1}^{S_{k,\epsilon}} p_k log(p_k),$$

where $S_{k,\epsilon}$ is the number of states and p_k is the occurrence probability for a state k. Then the chosen value

$$\epsilon^* = \arg\max_{\epsilon} H(\epsilon)$$

is the one that maximizes the entropy of the system.

To increase the accuracy of detection result for sleep stage changes, we propose to appropriately integrate the values of recurrence plots of multimodal signal readings that are captured by multiple sensors. We propose a signal detection method based on solving the optimisation problem with correlation-based constraints, that assumes that the recurrence plots preserve the information required for the signal detection.

Since the pioneering work [3] which have introduced the recurrence plots (RP) as the visualisation tool for describing the states of dynamical systems, most of the research efforts were focused on providing appropriate measures that can quantify recurrence plots [5], [6] and describe the underlying phenomena. For the signal analysis we use a set of measures studied in [7], namely, the correlation dimension (CD), approximate entropy (ApEn), largest Lyapunov exponent (λ_{max}), Hurst exponent (H), fractal dimension (FrD), Katz algorithm (K) and recurrence plot (RP), as the comparison features.

The rest of this paper is organised as follows. In Sections II-V we study the structure and properties of recurrence plots for ferret dataset, Lotka Volterra, Lorenz attractor model and auditory potentials in both, time and frequency domains, and analyse the obtained patterns. Section VI describes statistical approaches to recurrence plot analysis. In Section VII we briefly discuss the future works.

II. Ferret data - time domain recurrence plot analysis

We compute and analyse recurrence plots for ferret dataset[1] in time and frequency domain. We use two frequency domain representations, in particular, Discrete Wavelet transform (WT) and Synchrosqueezing transform (SST) to obtain better insights about recurrences of signal components and noise influence.

A. Time domain recurrence plot analysis

In Fig. 1 we provide time domain recurrence plots for several values of parameter ϵ. Entropy values for the same dataset for a range of ϵ values are provided in Fig. 2.

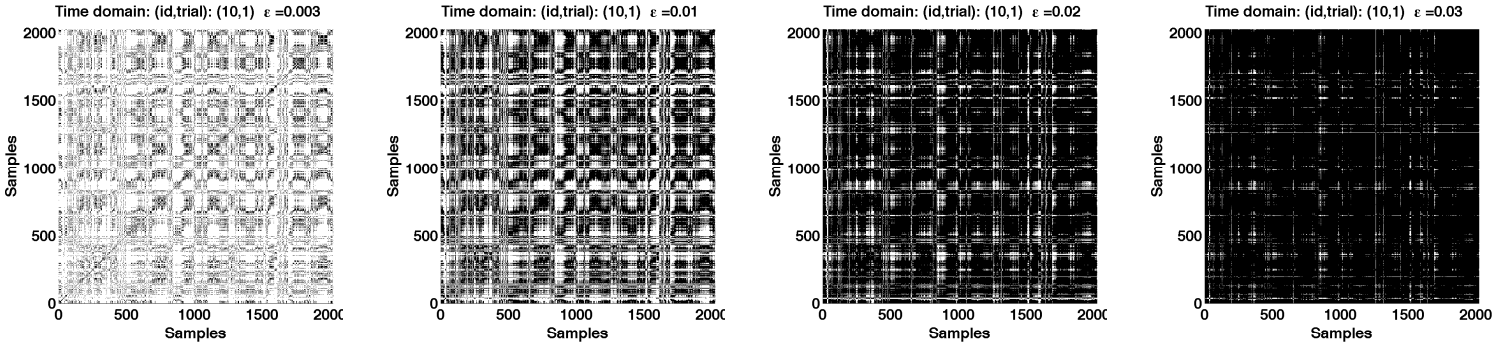


Fig. 1. Ferret measurements (time domain): recurrence plots obtained for the set of threshold values $\epsilon \in \{0.003, 0.01, 0.02, 0.03\}$.

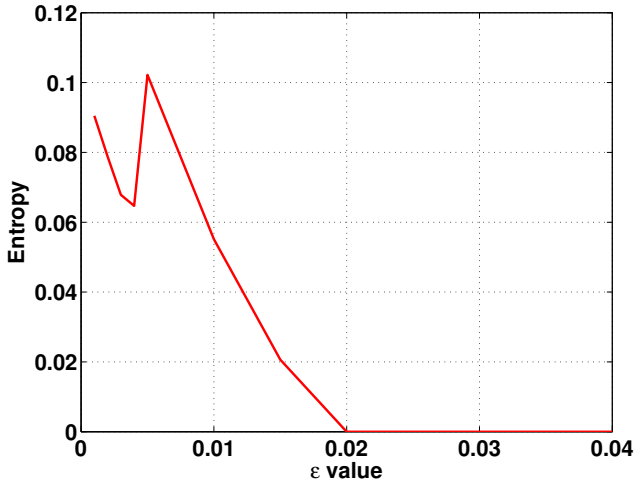


Fig. 2. Entropy values for a ferret dataset (time domain): optimal parameter value is $\epsilon = 0.005$.

B. Frequency domain recurrence plot analysis

1) Discrete wavelet analysis: In Fig. 3 we provide frequency domain recurrence plots (discrete wavelet transform analysis) for several values of parameter ϵ. Entropy values for the same dataset for a range of ϵ values are provided in Fig. 4.

2) Synchrosqueezing analysis: Several illustrations of recurrence plots for synchrosqueezing transform are provided in Fig. 5 and Fig. 6.

Fig. 7 shows the recurrence plot for the optimal values of the parameter ϵ, which are given in Table II.

[1]Ferret recordings used in this work are provided by Kristin K. Sellers, Davis V. Bennet and F. Frohlich, University of North Carolina at Chappel Hill.

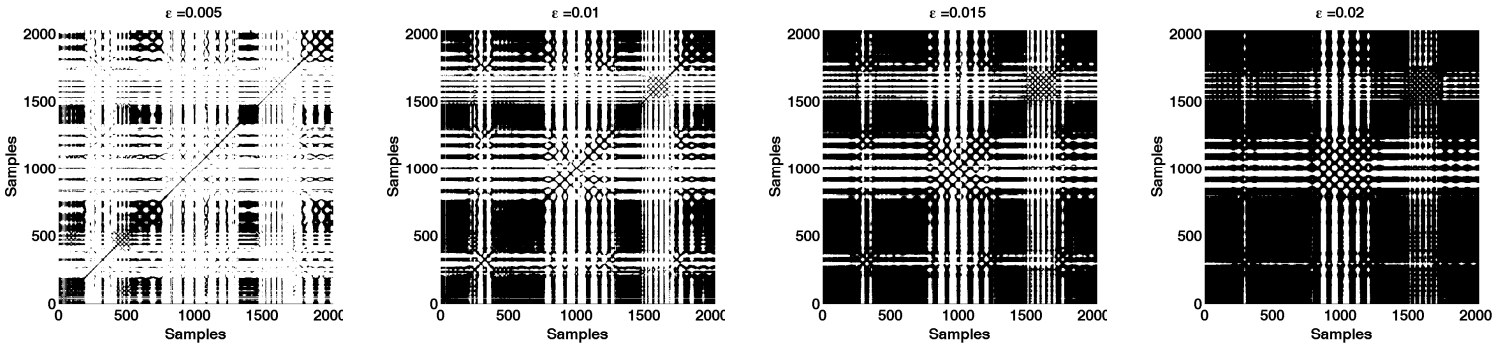


Fig. 3. Ferret measurements (time domain): recurrence plots obtained for the set of threshold values $\epsilon \in \{0.005, 0.01, 0.015, 0.02\}$.

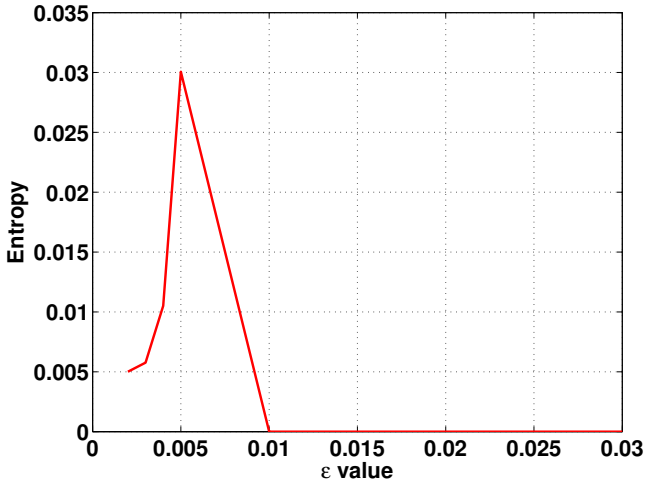


Fig. 4. Entropy values for a ferret dataset (frequency domain, discrete wavelet analysis): optimal parameter value is $\epsilon =$.

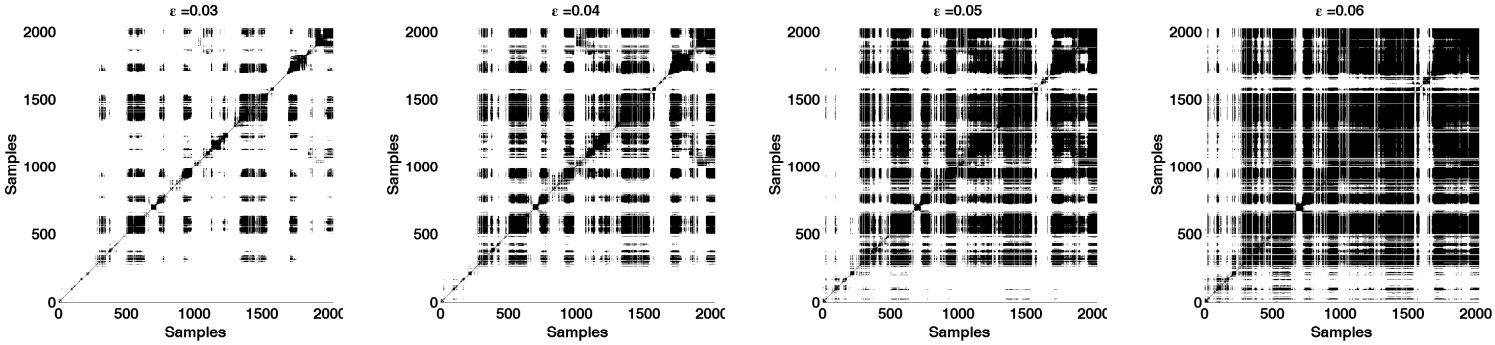


Fig. 5. Ferret measurements (time domain): recurrence plots obtained for the set of threshold values $\epsilon \in \{0.03, 0.04, 0.05, 0.06\}$.

dataset	optimal ϵ values
TD	0.005
FD: DWT	0.005
FD: SST	0.04

TABLE I

FERRET DATA: VALUES OF PARAMETERS ϵ THAT MAXIMISE THE RELATIVE ENTROPY FOR TIME DOMAIN (TD) AND FREQUENCY DOMAIN (FD) ANALYSIS. RECURRENCE PLOTS OBTAINED GIVEN THE DISCRETE WAVELET (DWT) AND SYNCHROSQUEEZING TRANSFORM (SST) ARE SEPARATELY ANALYZED.

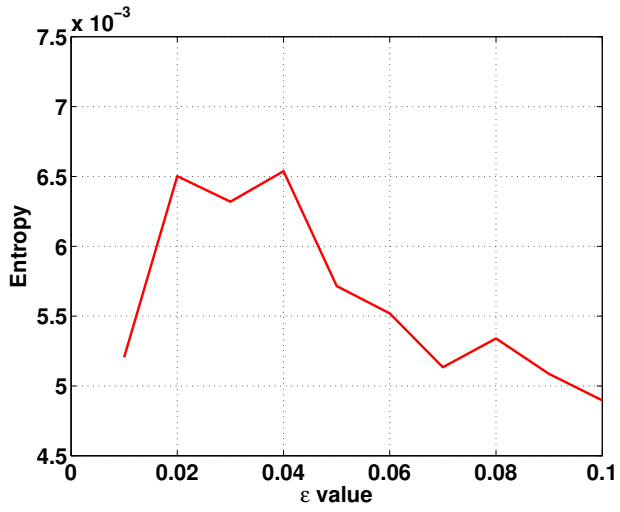


Fig. 6. Entropy values for a ferret dataset (frequency domain, synchrosqueezing analysis) provides optimal parameter $\epsilon = 0.04$.

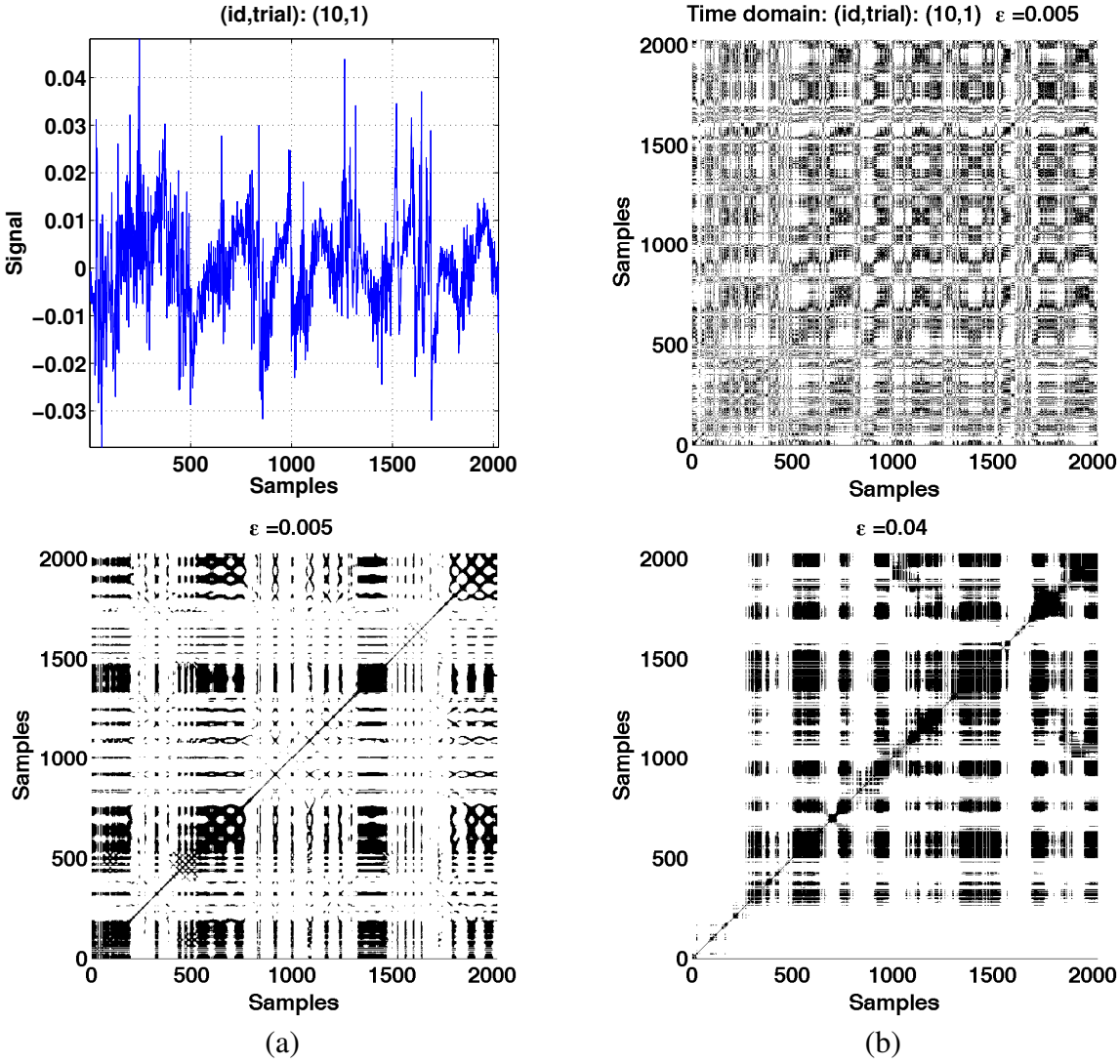


Fig. 7. Visual stimuli measurements at ferrets and recurrence plots for optimal values of parameter ϵ in time and frequency domains. Top row: (a) input signal; (b) time domain recurrence plot obtained for optimal $\epsilon = 0.005$. Bottom row: (a) recurrence plot obtained from the discrete wavelet transform of the original signal for optimal $\epsilon = 0.005$. (b) recurrence plot obtained from the synchrosqueezing transform of the original signal for optimal $\epsilon = 0.04$

III. LOTKA-VOLTERRA MODEL

We first analyse the Lotka-Volterra competitor model (often denoted as *predator-prey* model) with three attractor states. Simulation dataset is a sum of three noisy sinusoids, where pendulum-like oscillations model transitions between the attractor states. The sinusoid frequencies are chosen such that exactly one sinusoid per time instance has a dominant amplitude. We generate two datasets with the same set of parameters and different random noise component, as illustrated in Fig. 8.

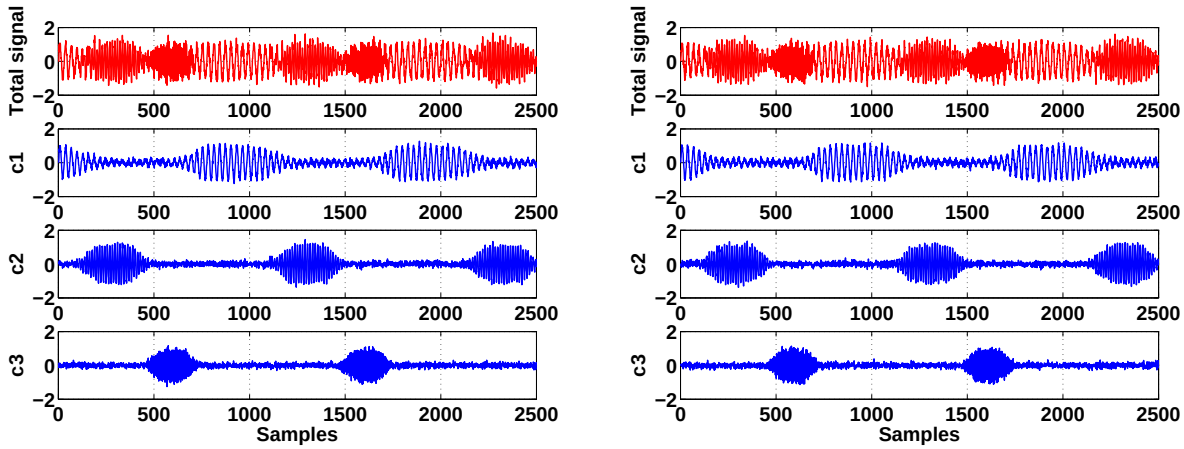


Fig. 8. Lotka-Volterra model: (left) dataset model_1 and (right) model_2. Figures show the dataset (top) and its three components. Channel values are renormalized to the interval $[-0.5, 0.5]$ for easier comparison of recurrence plots for different components.

A. Time domain analysis of Lotka-Volterra model

We compute the value of the parameter ϵ based on the maximum entropy criteria given in [4]. Fig. 9 illustrates entropy values for different threshold parameters.

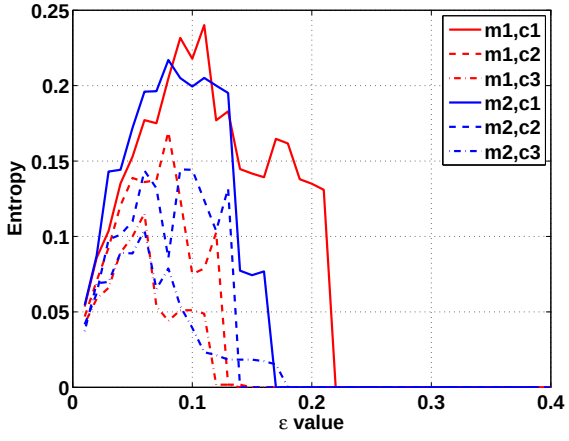


Fig. 9. Lotka-Volterra model (time domain): relative entropy values for range of ϵ values. Full and dashed curves of the same colour mark same signal components, where both considered signal models {m1,m2} consist of three components {c1,c2,c3}.

We summarise maximal entropy values for two datasets in Table II. Recurrence plots corresponding to these ϵ values are illustrated in Fig. 10.

We repeat same analysis for the clean signal in time and in frequency domain. Recurrence plots obtained for the optimal values of threshold ϵ are given in Table III.

dataset	ϵ values		
	c_1	c_2	c_3
model_1	0.11	0.08	0.06
model_2	0.08	0.09	0.06
clean_model	0.05	0.02	0.03

TABLE II

LOTKA-VOLTERRA MODEL (TIME DOMAIN): VALUES OF THE PARAMETER ϵ THAT MAXIMISES THE RELATIVE ENTROPY, FOR ALL SIGNAL MODELS AND THEIR COMPONENTS.

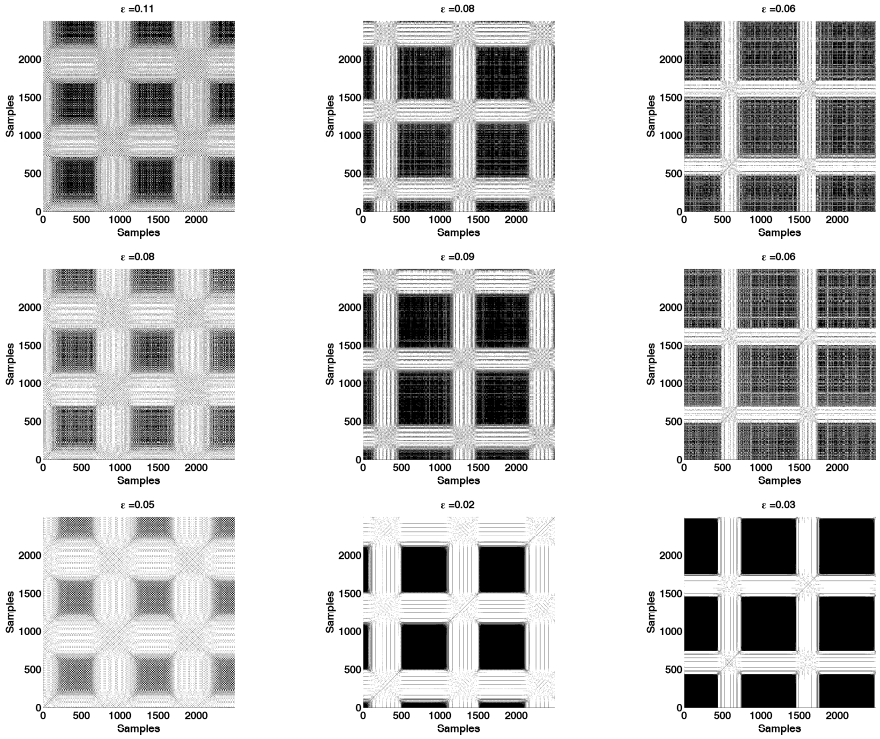


Fig. 10. Lotka-Volterra model (time domain): recurrence plots obtained for the optimal set of threshold values ϵ: (first row) three components of dataset model_1, (second row) components of model_2 and (third row) the clean model signal components.

B. Frequency domain analysis of Lotka-Volterra model

We now analyse recurrence plots in the frequency domain. To obtain signal scalograms (power spectrum of the signal over different scales), we use the continuous wavelet transform of Gabor (Morlet) wavelets that are known to be closely related to human visual and auditory systems [8]. Scalograms for two signal models (model_1, model_2) and for a non-noisy signal model and their three components are illustrated in Fig. 12.

In the frequency domain, we analyze recurrence plots obtained from scalograms of each of the signal components. Recurrence plots are obtained by using values of power spectrum vector at particular instance instead of vectors of sensor measurements directly. The optimal parameter ϵ is calculated for all the components, as illustrated in Fig. 13, and the recurrence plots for the optimal values of the threshold ϵ are given in Fig. 14. We note that more informative recurrence plots are obtained in case when the search space is limited to the interval $[0, 0.25]$, as illustrated in Fig. 15. This is the consequence of the increasing tendency of the number of possible sequences in the domain $[0, \epsilon_{all-one}]$, where $\epsilon_{all-one}$ is the maximal value of the distance function that provides *all-one* recurrence atria.

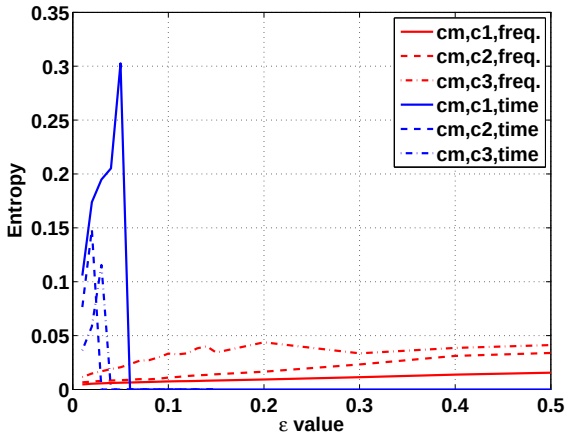


Fig. 11. Lotka-Volterra model (time and frequency domain): relative entropy values for range of ϵ values. Full and dashed curves of the same colour mark the frequency or time domain signal components of the clean signal model (cm): {c1,c2,c3} denote signal components and {freq.,time} corresponding domains.

clean_model dataset	ϵ values		
time domain	0.05	0.02	0.03
frequency domain	0.5	0.5	0.2

TABLE III

LOTKA-VOLTERRA MODEL (TIME AND FREQUENCY DOMAIN): VALUES OF ϵ PARAMETER THAT GIVE THE MAXIMAL RELATIVE ENTROPY RESULTS FOR EACH OF THREE SIGNAL COMPONENTS, FOR TWO TYPES OF RECURRENCE PLOTS - OBTAINED FROM THE TIME SEQUENCE AND THE FREQUENCY SCALOGRAMS.

C. Total Lotka-Volterra signal analysis with partially overlapped spectrum: analysis in time and frequency domains

Recurrence plots for total, fully normalized overlap_model dataset with partially overlapped time domain are illustrated in Fig. 16. We conclude that due to the noise, patterns are very hard to spot in this scenario. Therefore, under the assumption that the signal components are known, we proceed by analysing recurrence plots for different signal components, as given in Fig. 17.

We now perform the analysis in the frequency domain. Spectrogram of total signal and its components (from left to right, respectively) are illustrated in Fig. 18. We do not optimize the values of the parameter ϵ for this case since we do not observe monotonic properties for dependence of the relative entropy on the parameter ϵ.

Frequency components of recurrence plots (normalized signal) with partially overlapped spectrum are illustrated in Fig. 19. Though the recurrence patterns are more visible than in the time domain analysis, some artefacts remain (full signal given in left column of Fig. 19) and clean distinction of the state change detection time is not straight forward. More conclusive plots are given in *per-component* signal analysis. We note however that good recurrence plot for a signal component with dominant energy has similar ϵ value as the total signal recurrence plot, therefore, it is noise-sensitive. Therefore, if components are known, component-wise normalization in a frequency domain is important signal preprocessing step.

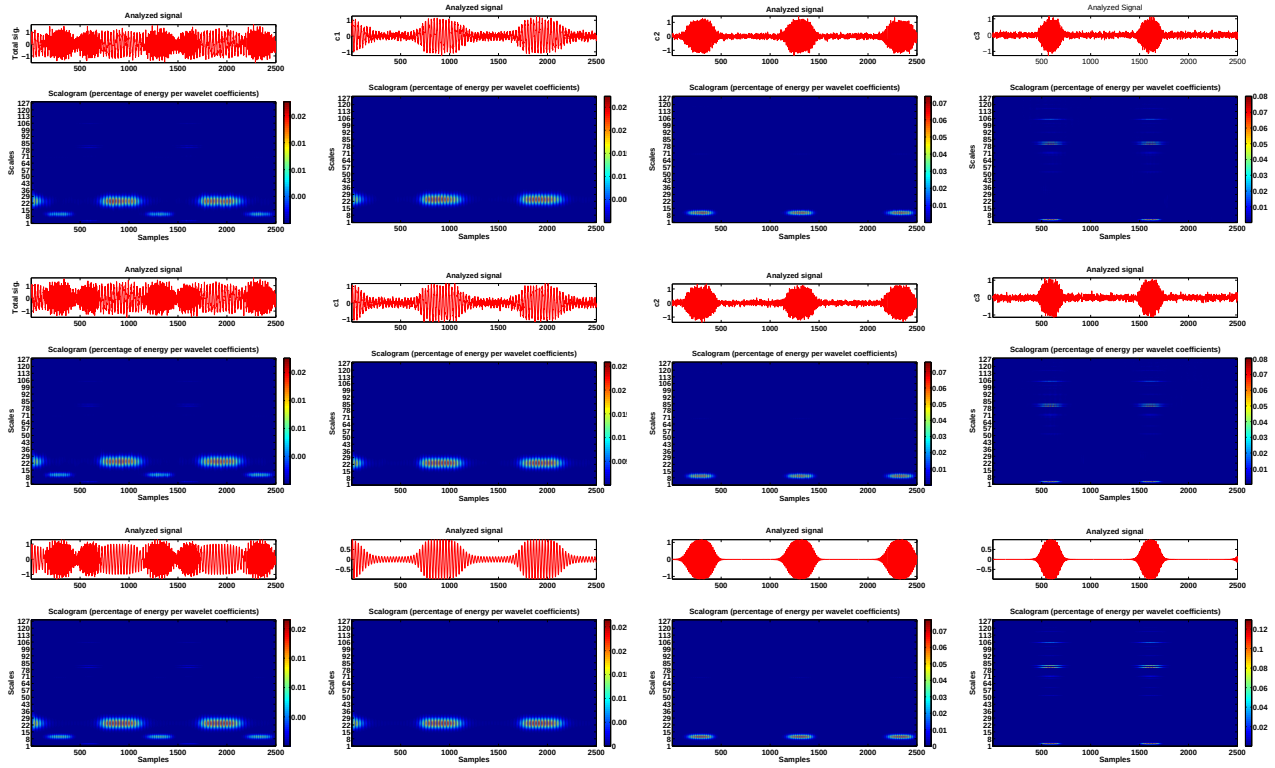


Fig. 12. Lotka-Volterra model (frequency domain): Scalograms of the signals and its three components are given row-wise: (top row) model_1, (middle row) model_2 and (bottom row) noiseless model clean_model. Columns represent (first column) full signal spectrograms and its three signal components $\{c_1, c_2, c_3\}$, respectively.

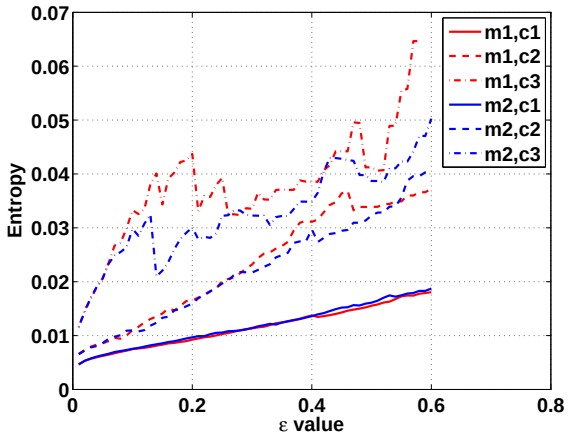


Fig. 13. Lotka-Volterra model (frequency domain): Dependence of the relative entropy from the parameter ϵ: full and dashed curves of the same colour mark the components of the same signal model: {m1,m2} denote two signal models and {c1,c2,c3} their components.

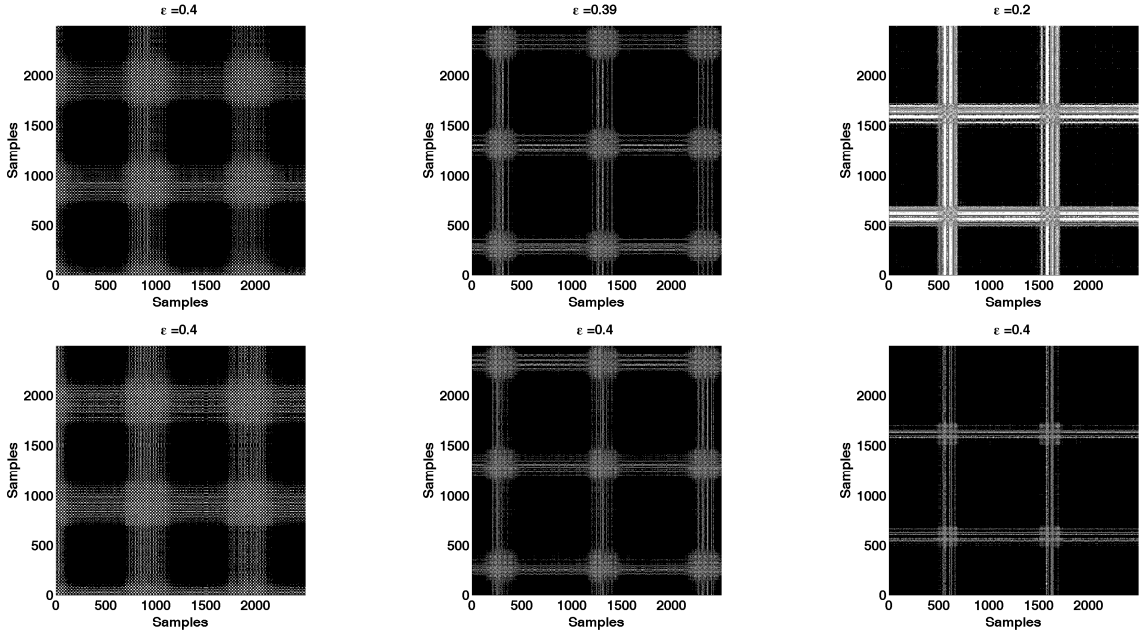


Fig. 14. Lotka-Volterra model (frequency domain): Scalogram-based recurrence plots for optimal values ϵ values, given the interval $[0, 0.4]$: (first row) three components of dataset model_1 and (second row) components of model_2.

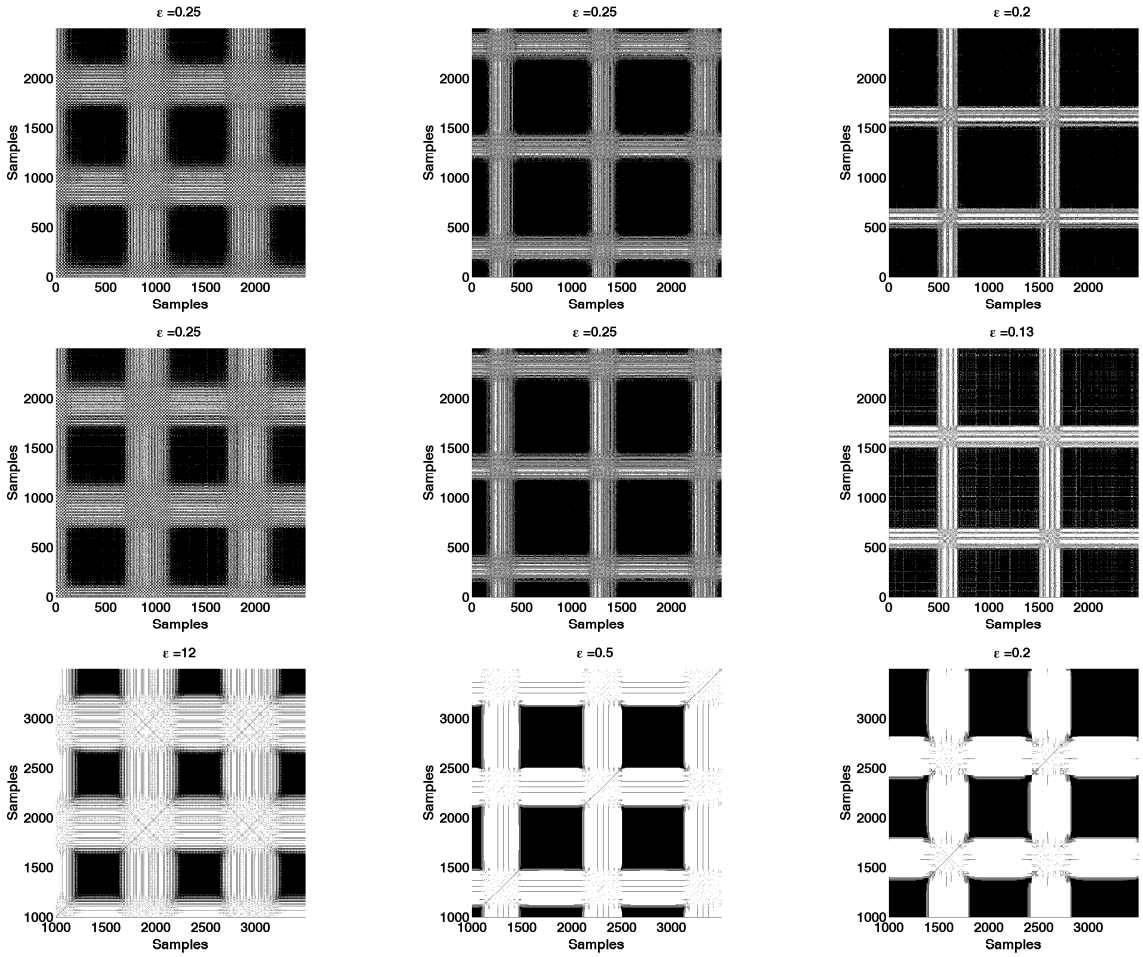


Fig. 15. Lotka-Volterra model (frequency domain): Scalogram-based recurrence plots for optimal ϵ values, given the interval $[0, 0.25]$: (first row) three components of dataset model_1 and (second row) components of model_2. Note that in (third row) that represents the components of clear_model the values of $\epsilon = 12$.

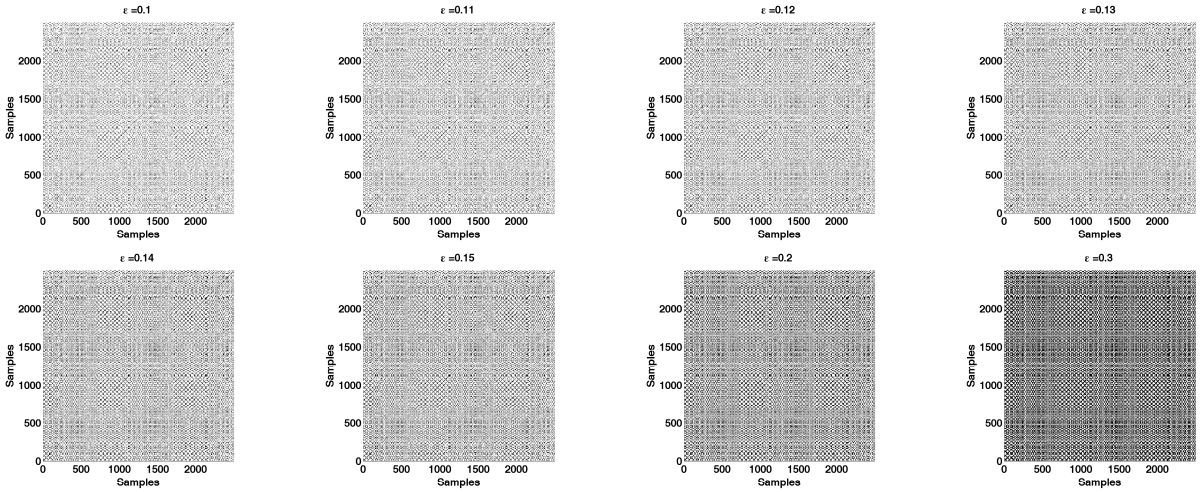


Fig. 16. Lotka-Volterra model with partially overlapped spectrum components (time domain): Recurrence plots of the fully normalised signal overlap_model dataset. (first row) $\epsilon \in \{0.01, 0.11, 0.12, 0.13\}$, (second row) $\epsilon \in \{0.14, 0.15, .2, 0.3\}$.

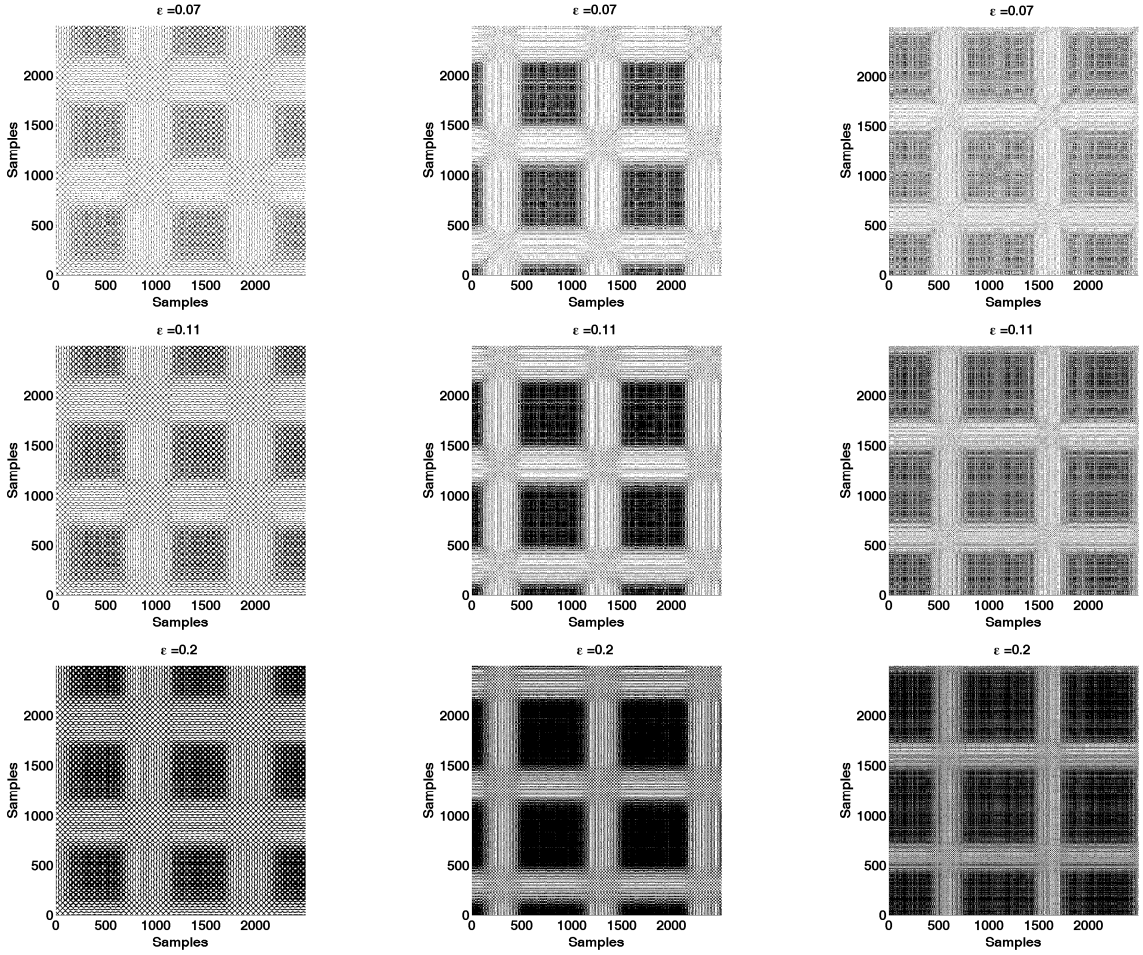


Fig. 17. Lotka-Volterra model with partially overlapped spectrum components (time domain): Recurrence plots for components of normalised overlap_model dataset. (first row) $\epsilon = 0.07$, (second row) $\epsilon = 0.11$; (third row) $\epsilon = 0.2$. Columns represent recurrence plots for three signal components.

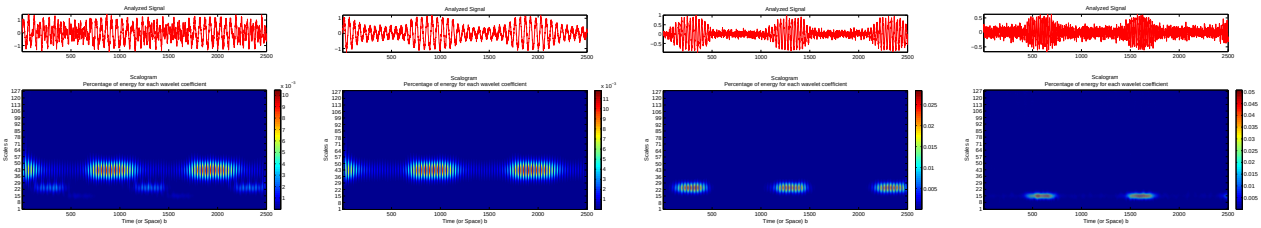


Fig. 18. Lotka-Volterra model with partially overlapped spectrum components: Spectrogram illustrations for total signal (left) and its three signal components.

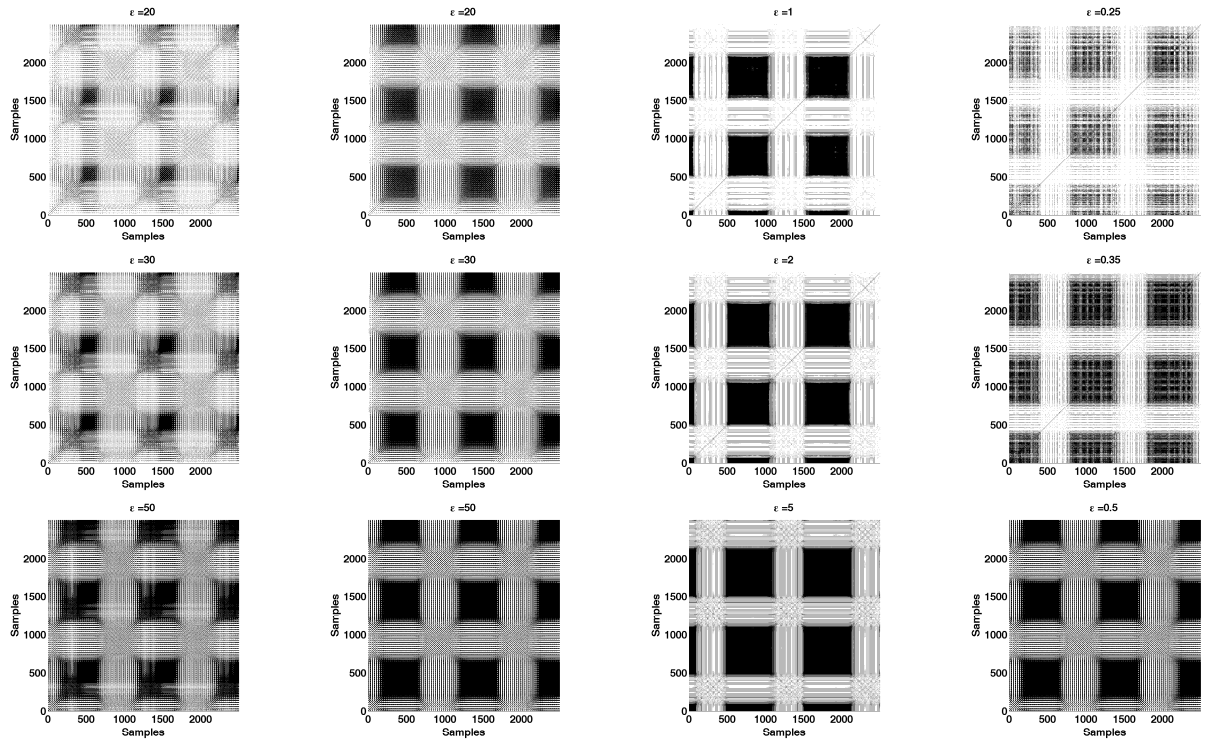


Fig. 19. Lotka-Volterra model with partially overlapped spectrum components (frequency domain): Recurrence plots total signal (left) and its three components are given row-wise for several ϵ values. (first row) $\epsilon \in \{20, 20, 1, 0.25\}$, (second row) $\epsilon \in \{25, 25, 1.5, 0.3\}$; (third row) $\epsilon \in \{30, 30, 2, 0.35\}$; (fourth row) $\epsilon \in \{35, 35, 2.5, 0.4\}$, (fifth row) $\epsilon \in \{40, 40, 3, 0.45\}$.

IV. Lorenz attractor model

Input Lorenz attractor datasets are illustrated in Fig. 20. We analyse a Lorenz attractor model dataset

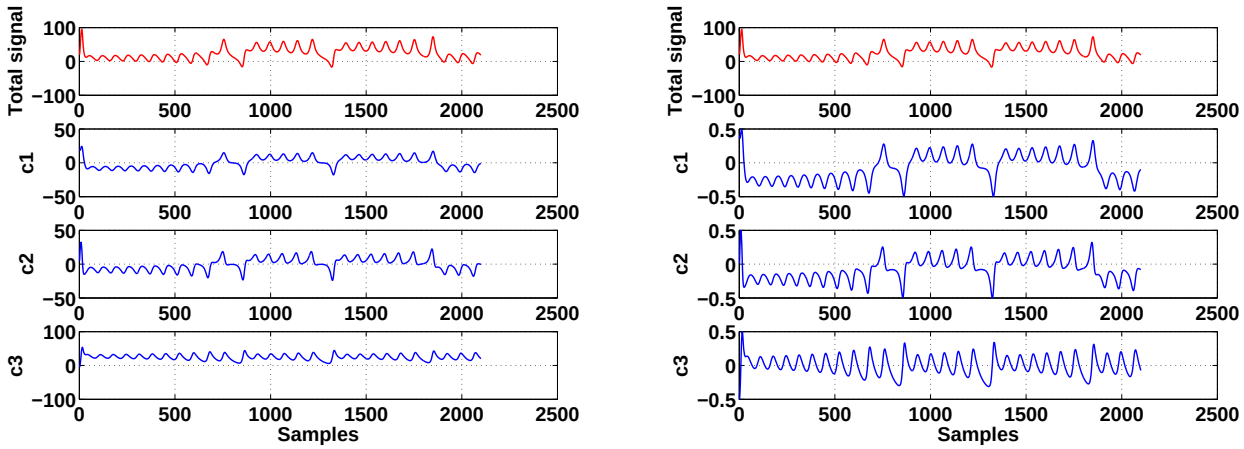


Fig. 20. Lorenz attractor model: Signals and their components: (left) Lorenz_non-normalized_model dataset; (right) normalized version Lorenz_normalized_model.

and obtain recurrence plots for the full signal and its components in time and frequency domains.

A. Time domain analysis

In Fig.21 we illustrate the recurrence plot of the full time-domain signal composed out of the three components and each of the components.

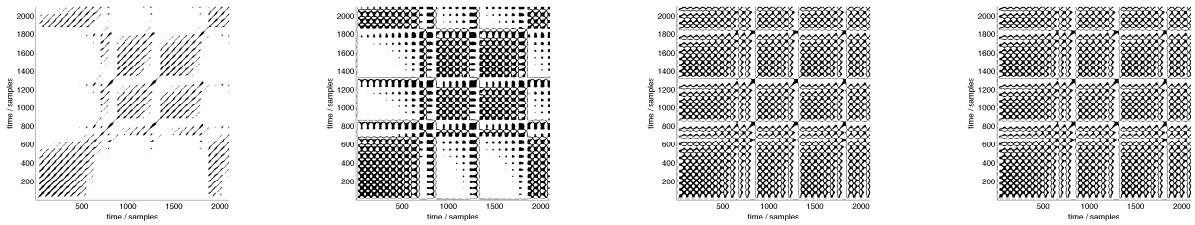


Fig. 21. Lorenz attractor model (time domain): Recurrence plots given for the optimal parameter ϵ Lorenz_normalized_model dataset; (left) total signal and its three signal components, respectively.

Next, we illustrate evolution of several Lorenz signal recurrence plots for each of the components in Fig. 23.

We next plot the sequence of recurrence plots that correspond to the normalized signal (single normalization, without knowledge of particular components), illustrated in Fig. 23.

To observe the influence of normalisation on recurrence plots, we also analyse the scenario with double normalization (the signal components are normalized to the range $[-0.5, 0.5]$ prior to their summation and second normalisation to the interval $[-0.5, 0.5]$). Results of the double normalization are illustrated in Fig. 24.

B. Frequency domain analysis

Similarly, we perform experiments in the frequency domain. Spectrogram of the full signal and its components are illustrated in Fig. 25.

Recurrence plots obtained from the spectrograms of the normalized signal are illustrated in Fig. 26.

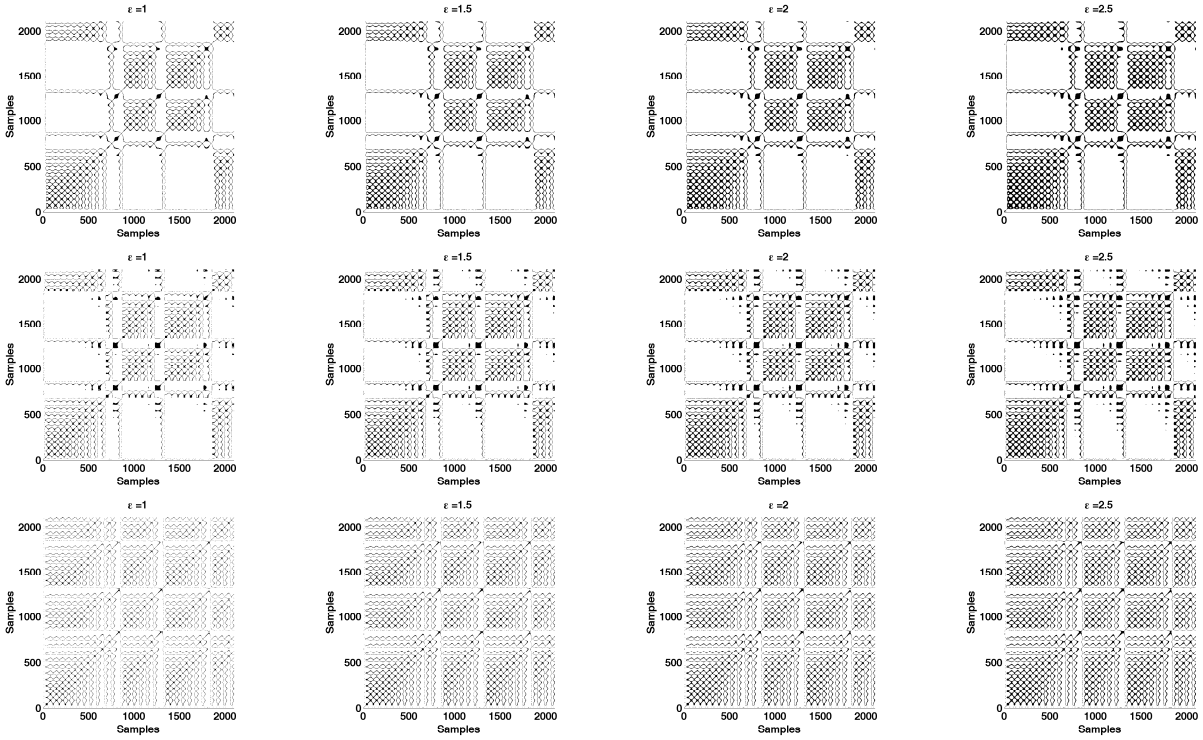


Fig. 22. Lorenz attractor model (time domain): Recurrence plots for a fully normalized signal Lorenz_normalized_model dataset. (first column) $\epsilon = 1$; (second column) $\epsilon = 1.5$; (third column) $\epsilon = 2$; (fourth column) $\epsilon = 2.5$. Rows represent recurrence plots for three signal components.

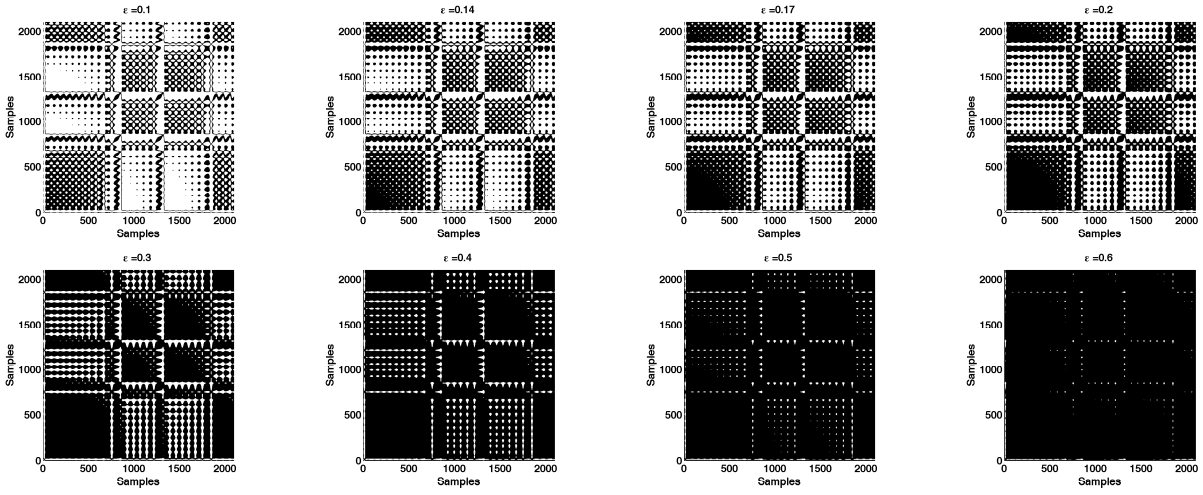


Fig. 23. Lorenz attractor model (time domain): Recurrence plots for normalized (total) signal dataset. (first row) $\epsilon \in \{0.1, 0.14, 0.17, 0.2\}$, (second row) $\epsilon \in \{0.3, 0.4, 0.5, 0.6\}$.

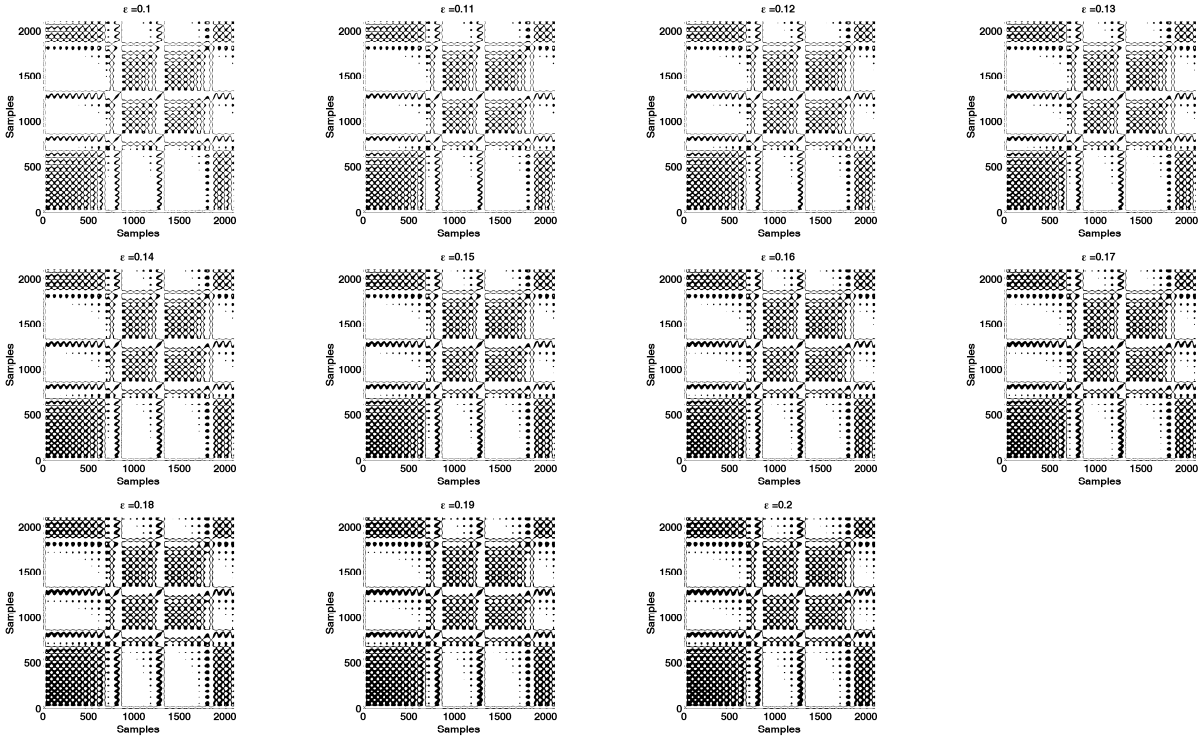


Fig. 24. Lorenz attractor model (time domain): Recurrence plots for normalized signal dataset (double normalisation: per components and per signal). (first row) $\epsilon \in \{0.1, 0.11, 0.12\}$, (second row) $\epsilon \in \{0.13, 0.14, 0.15\}$, (third row) $\epsilon \in \{0.16, 0.17, 0.18\}$, (forth row) $\epsilon \in \{0.19, 0.2\}$.

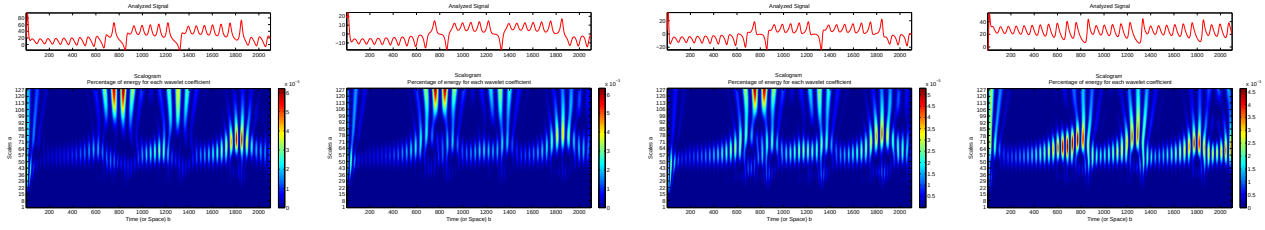


Fig. 25. Lorenz attractor model (frequency domain): Spectrogram of the Lorenz attractor for a Lorenz_normalized_model dataset with overlapped spectrum components.

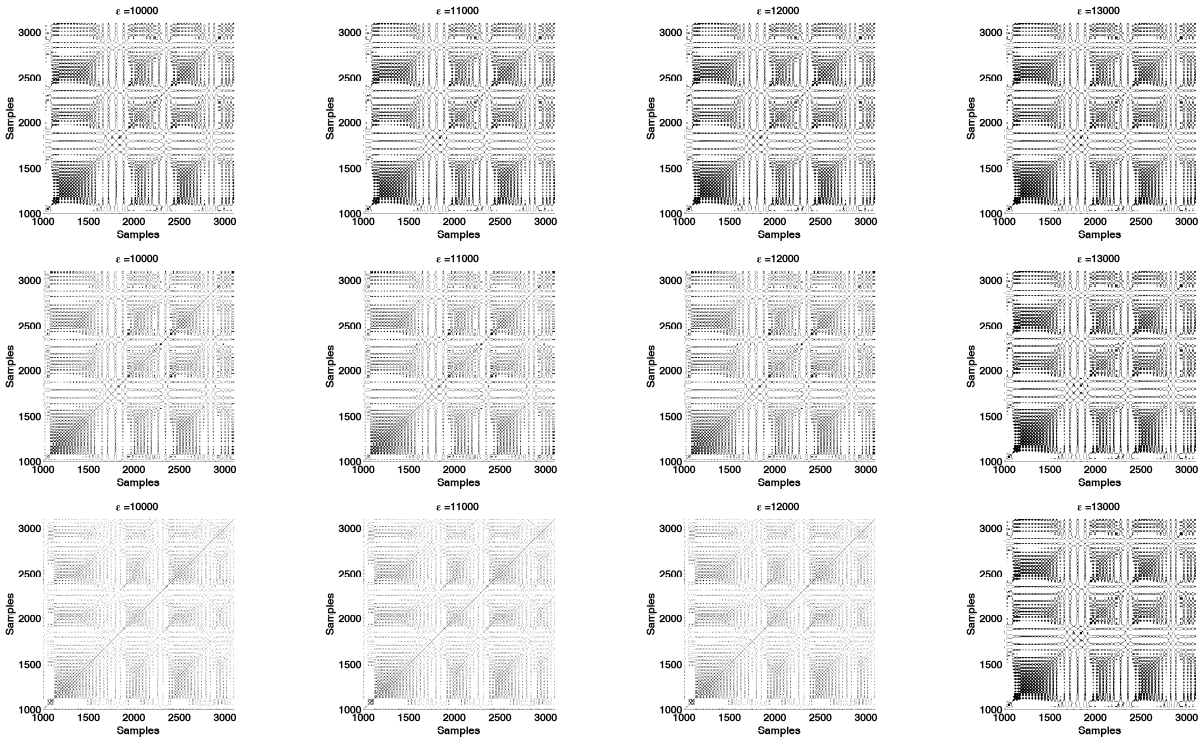
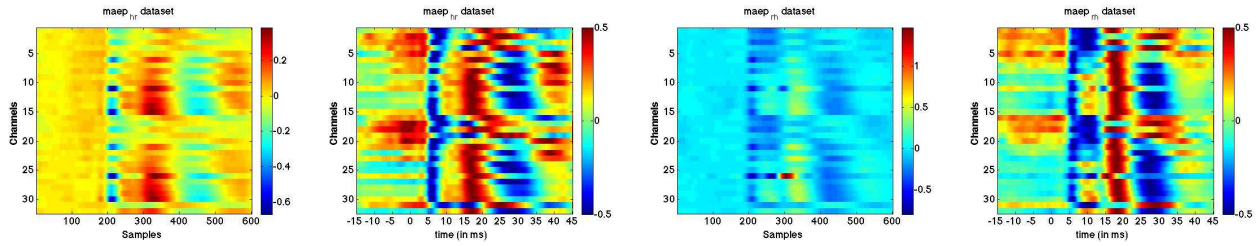


Fig. 26. Lorenz attractor model (frequency domain) with overlapped spectrum components: Recurrence plots obtained from spectrograms, (first column) $\epsilon = 10 \times 10^3$, (second column) $\epsilon = 11 \times 10^3$; (third column) $\epsilon = 12 \times 10^3$; (fourth column) $\epsilon = 13 \times 10^3$. Rows represent recurrence plots for three signal components.

V. AUDITORY EVOKED POTENTIALS



(a) (b) (c) (d)

Fig. 27. Auditory evoked potential signal (32 electrodes): (a) maep_hr and (c) maep_rh datasets. Channel values are renormalized to the values in the interval $[-0.5, 0.5]$ with purpose to observe channel synchronisation and credibility of their measurements. Normalized datasets are illustrated in figures (b) maep_hr and (d) maep_rh.

Auditory brain stem response (ABR) and auditory evoked potentials of middle latency (MAEP) were captured for three subjects. The EEG signal was recorded by 32 electrodes placed in the 10-20 extended system, where Cs was a reference electrode.

Preprocessing: Low pass filtering ($f_{cut-off} = 2kHz$), sampling with the rate $f_s = 10kHz$, number of samples $n_s = 600$ with sampling interval $\delta t = 60ms$, sampling interval $[-15, 45]$ ms ($t = 0$ ms marks the sound click). Signal is obtained by averaging over $n_t = 10000$ single trials. Signals and their normalised values are illustrated in Fig. 27.

The original datasets are given in Fig. 27 (a), (c), where image rows represent single electrode values over a time interval. We observe that the range of values for datasets differs. We normalise signals from every channel such that all the channel signals take the values from the interval $[-0.5, 0.5]$, as illustrated in Fig. 27 (b) and (d). In both datasets, the set of electrodes (the channels $\{2, 4, 17, 19, 31\}$) show different behaviour to other electrodes. Therefore, we analyze their influence on recurrence plot for full and reduced number of channels per dataset.

Several recurrence plots for different ϵ values for the maep_hr dataset are illustrated in Fig. 28, while the similar plots for the dataset maep_rh are given in Fig. 29. The results for normalised maep_hr and maep_rh datasets are given in Figs. 30 and 31, respectively.

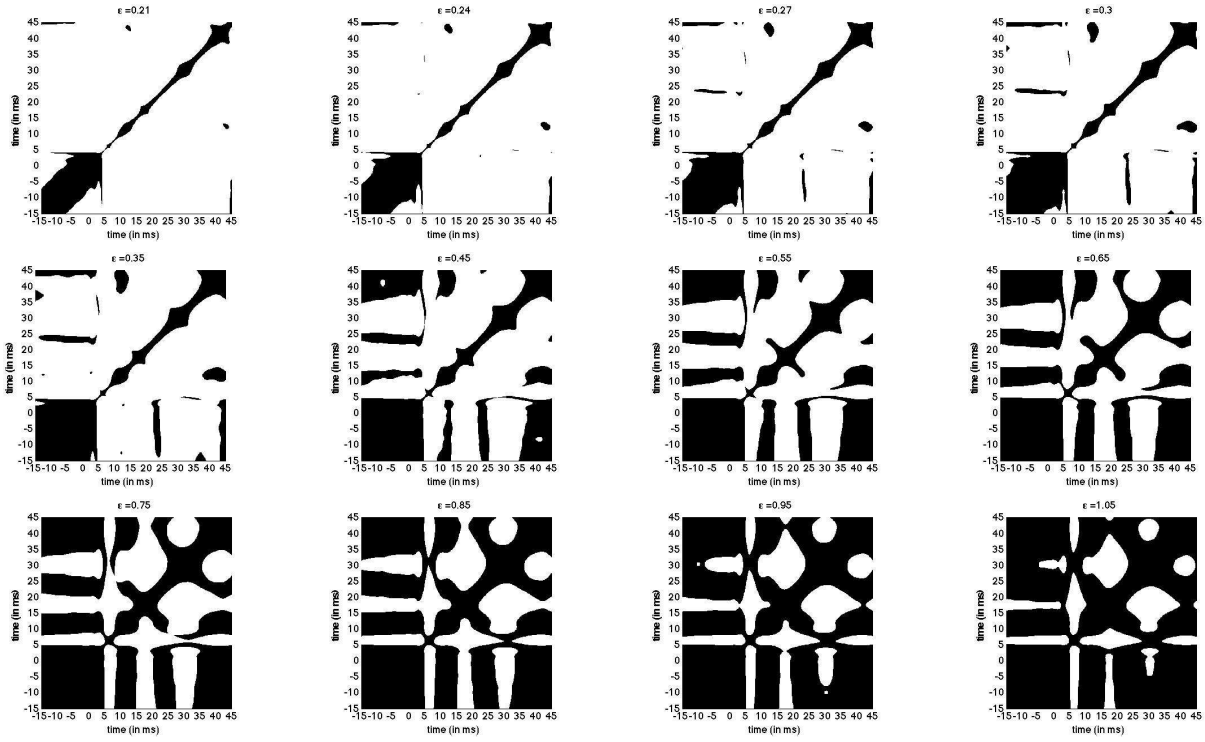
Fig. 32 (left) gives the optimal epsilon value $\epsilon_0 = 0.11$ and the corresponding recurrence plot.

A. Analysis for the surrogate dataset to the original maep_hr dataset: Iterative Amplitude Adjusted FT

The surrogate dataset is formed as described in Section 4.3 of the review paper [9], as described in short.

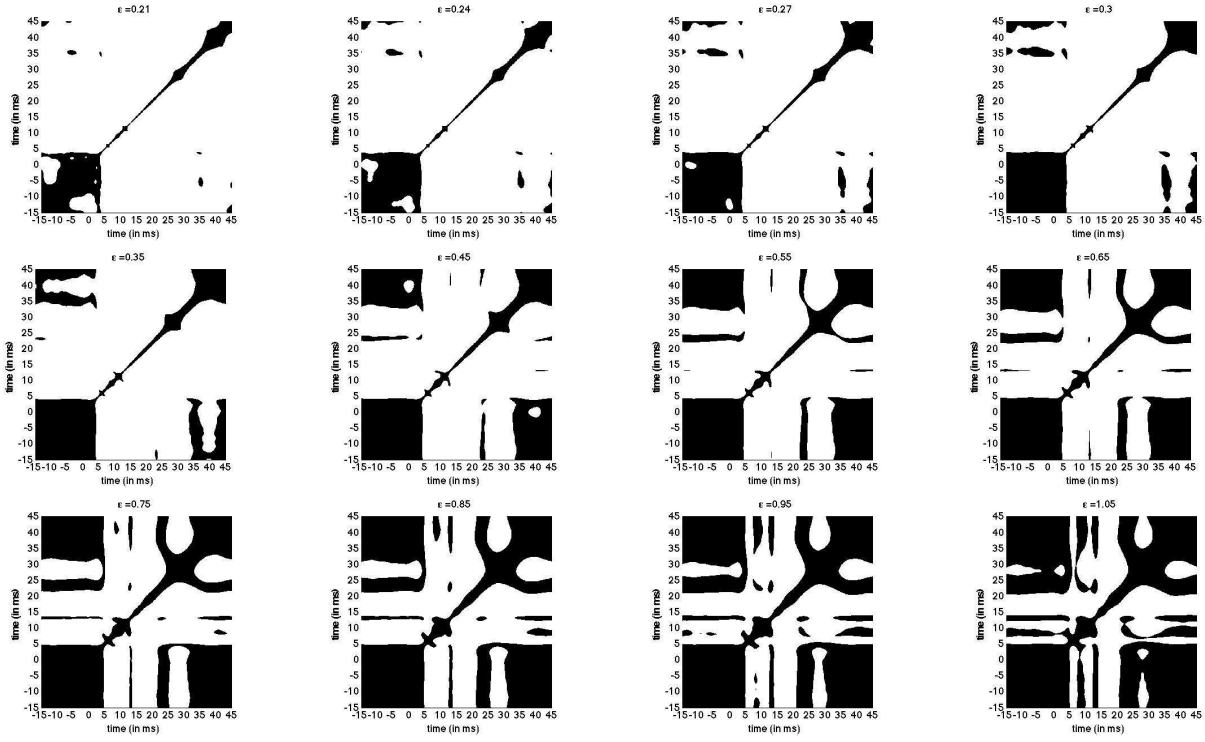
- $\{\bar{r}_n^{(0)}\}$ shuffle of the data or IAAFT (Iterate Amplitude Adjusted FT)
- compute FT of $\{\bar{r}_n^{(i)}\}$: $\{\bar{R}_k^{(i)}\} = \frac{1}{\sqrt{N}} \bar{r}_n e^{\frac{i2\pi kn}{N}}$
- compute the inverse FT, where the phase is $e^{i\psi_k^{(i)}} = \frac{\bar{R}_k^{(i)}}{|\bar{R}_k^{(i)}|}$
  $s_n^{(i)} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{i\psi_k^{(i)}} |S_k| e^{-\frac{2\pi kn}{N}}$
- rank ordering used to compute new $r_n^{(i+1)} = C_{rank}(s_n^{(i)})$

The surrogate maep_hr dataset is illustrated in Fig. 33 and several recurrence plots in Figs. 34 show that the same statistics signal does not show the recurrence state.



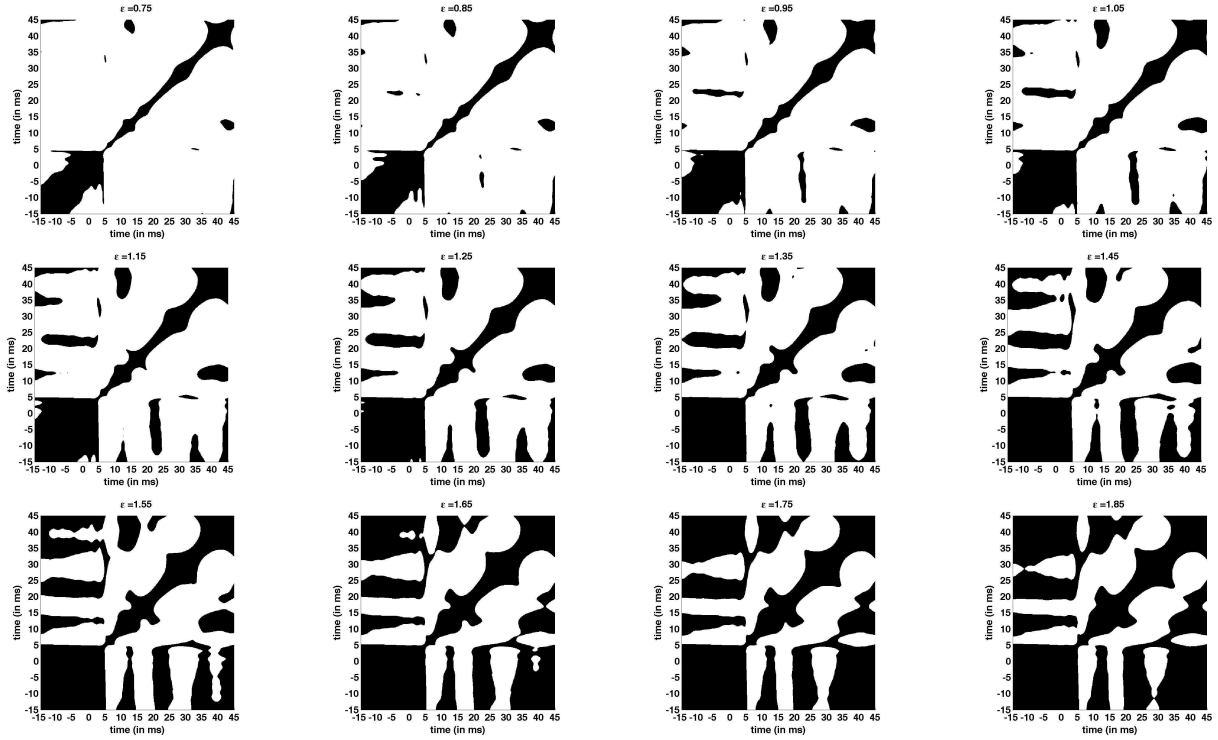
(a) (b) (c) (d)

Fig. 28. Auditory evoked potential signal: Recurrence plots obtained for different parameters ϵ for dataset **maep_hr**: (first row) (a) $\epsilon = 0.21$, (b) $\epsilon = 0.24$, (c) $\epsilon = 0.27$, (d) $\epsilon = 0.30$; (second row) (a) $\epsilon = 0.35$, (b) $\epsilon = 0.45$, (c) $\epsilon = 0.55$, (d) $\epsilon = 0.65$; (third row) (a) $\epsilon = 0.75$, (b) $\epsilon = 0.85$, (c) $\epsilon = 0.95$, (d) $\epsilon = 1.05$.



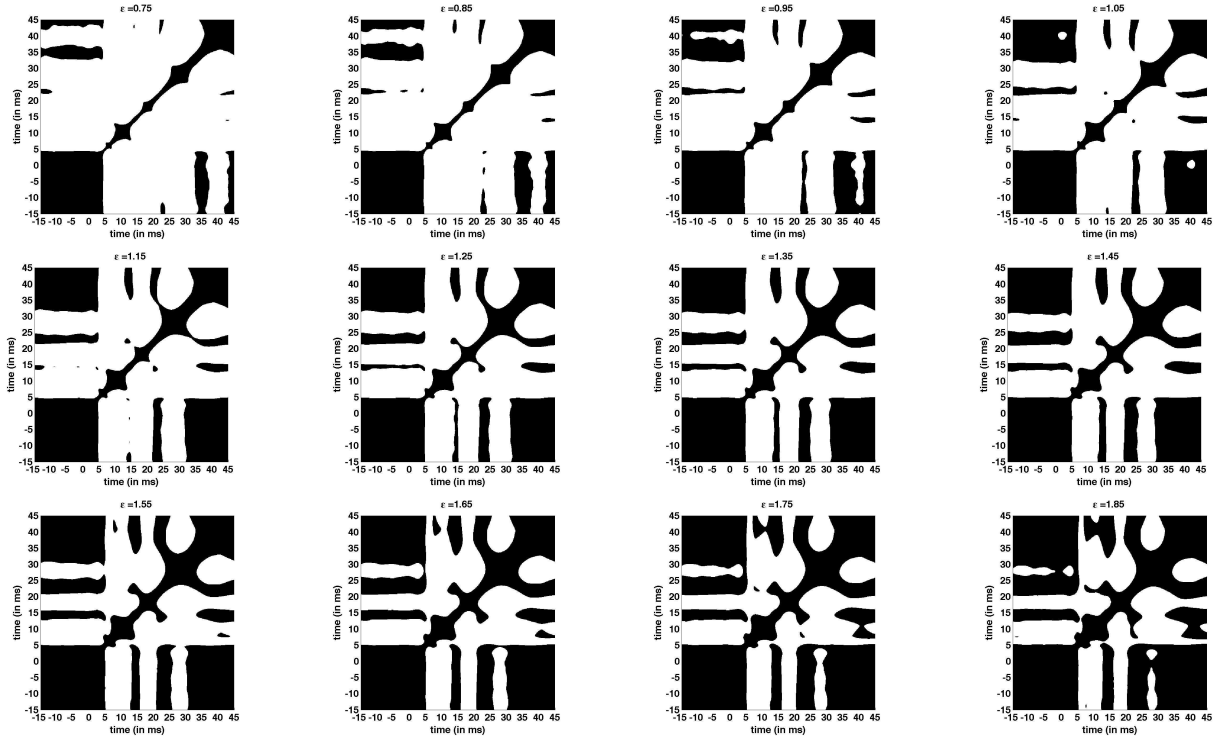
(a) (b) (c) (d)

Fig. 29. Auditory evoked potential signal: Recurrence plots obtained for different parameters ϵ for dataset **maep_rh**: (first row) (a) $\epsilon = 0.21$, (b) $\epsilon = 0.24$, (c) $\epsilon = 0.27$, (d) $\epsilon = 0.30$; (second row) (a) $\epsilon = 0.35$, (b) $\epsilon = 0.45$, (c) $\epsilon = 0.55$, (d) $\epsilon = 0.65$; (third row) (a) $\epsilon = 0.75$, (b) $\epsilon = 0.85$, (c) $\epsilon = 0.95$, (d) $\epsilon = 1.05$.



(a) (b) (c) (d)

Fig. 30. Auditory evoked potential signal: Recurrence plots obtained for different parameters ϵ for **normalized** values of dataset **maep_hr**: (first row) (a) $\epsilon = 0.75$, (b) $\epsilon = 0.85$, (c) $\epsilon = 0.95$, (d) $\epsilon = 1.05$; (second row) (a) $\epsilon = 1.15$, (b) $\epsilon = 1.25$, (c) $\epsilon = 1.35$, (d) $\epsilon = 1.45$; (third row) (a) $\epsilon = 1.55$, (b) $\epsilon = 1.65$, (c) $\epsilon = 1.75$, (d) $\epsilon = 1.85$.



(a) (b) (c) (d)

Fig. 31. Auditory evoked potential signal: Recurrence plots obtained for different parameters ϵ for **normalized** values of dataset **maep_rh**: (first row) (a) $\epsilon = 0.75$, (b) $\epsilon = 0.85$, (c) $\epsilon = 0.95$, (d) $\epsilon = 1.05$; (second row) (a) $\epsilon = 1.15$, (b) $\epsilon = 1.25$; (c) $\epsilon = 1.35$, (d) $\epsilon = 1.45$; (third row) (a) $\epsilon = 1.55$, (b) $\epsilon = 1.65$, (c) $\epsilon = 1.75$, (d) $\epsilon = 1.85$;

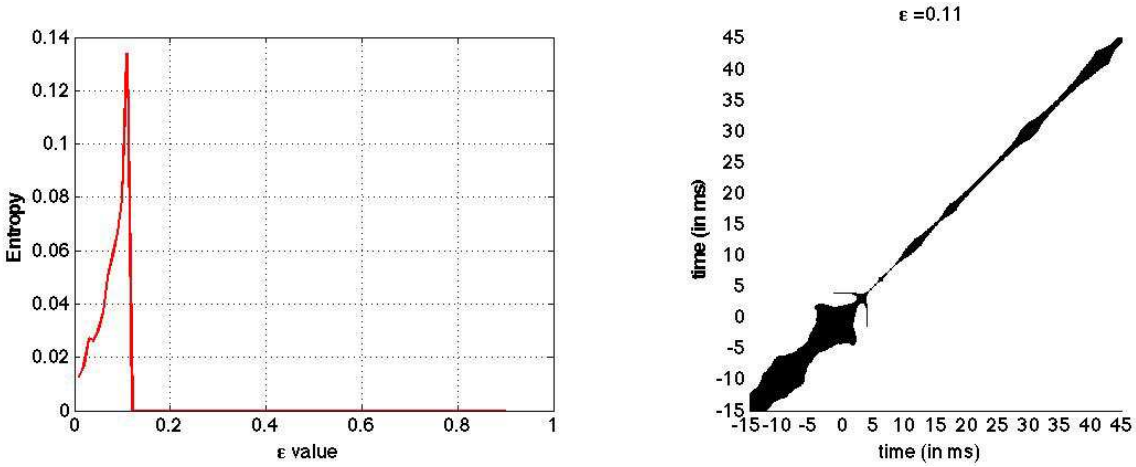


Fig. 32. Illustration of the dependence of the relative entropy from the parameter epsilon (left) and the recurrence plot for optimal $\epsilon = 0.11$ (right). The maximum entropy of this system is given for $\epsilon = 0.1$. This is just a test example performed for the purpose of learning. Note that the data has only one recurrence state and therefore, there is no much sense to apply such a criteria.

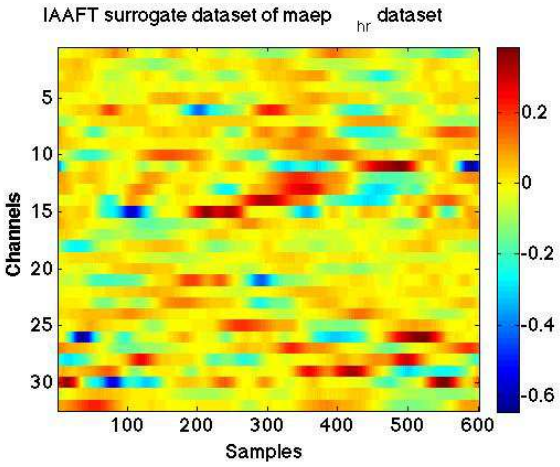


Fig. 33. Surrogate dataset of the original signal maep_hr obtained by IAAFT.

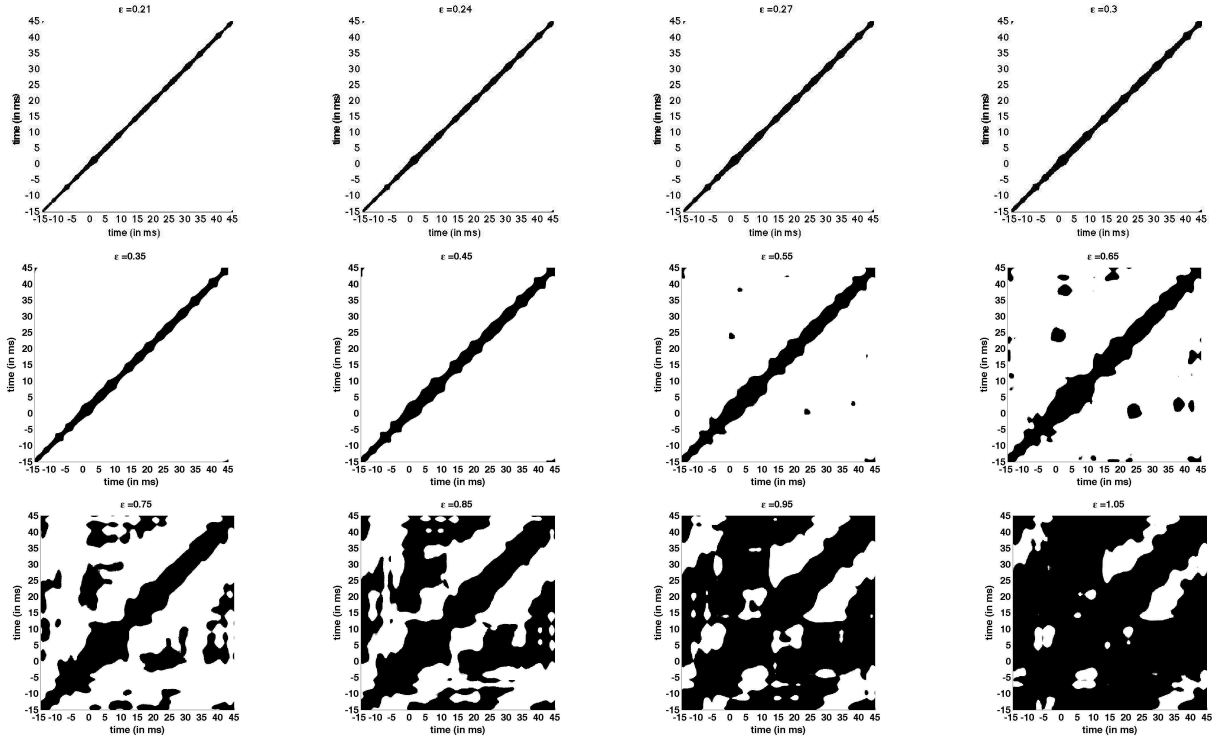


Fig. 34. Surrogate (IAAFT) data of auditory evoked potential signal: Recurrence plots obtained for maep_hr dataset for different parameters ϵ: (first row) (a) $\epsilon = 0.21$, (b) $\epsilon = 0.24$, (c) $\epsilon = 0.27$, (d)$\epsilon = 0.30$; (second row) (a) $\epsilon = 0.35$, (b) $\epsilon = 0.45$, (c)$\epsilon = 0.55$, (d) $\epsilon = 0.65$; (third row) (a) $\epsilon = 0.75$, (b) $\epsilon = 0.85$, (c) $\epsilon = 0.95$, (d) $\epsilon = 1.05$;

VI. Multivariate statistical analysis of recurrence plots

In this section we study statistical properties and trial-to-trial variability of recurrence plots, where signals are captured by locally positioned sensors. Time series in brain studies are highly variable within several trials, even when they are locally captured (e.g., firing patterns of neurons) under well-controlled experimental conditions. On the other hand, many physiological features have characteristic frequency signatures. For example, blinking our eyes is connected with changes in the alpha band (frequencies in the interval $(8-12)Hz$) of EEG recordings. We therefore propose a novel concept for building recurrence plots from time-frequency signal representation, instead of directly from time variate data. Such representation is in general adapted for nonstationary signal analysis. In our case, time-frequency signal representations provide additional flexibility to recurrence plot analysis that is not present in time-domain (for e.g., possibility to weight the importance of some frequency bands) and gives insights into frequency band feature importance.

We study statistical properties of RPs that are obtained from time-frequency signal representations. If the higher order system dynamics remains encoded in two-dimensional signal representations, its structure will have particular statistical properties that differ from those of surrogate datasets, built to have the same energy in time-frequency bands as the original dataset. To perform this analysis, we follow several steps. We first build original and surrogate RPs in time-frequency domain. The surrogate dataset is constructed to have same energy characteristics in time-frequency domain, where the information component encoded is time is randomized. Finally, we use classical multivariate statistical analysis tools (for e.g., t-tests, hypothesis testing, bootstrap methods) to extract and compare the statistics of the original (trial) dataset and surrogates. More details about each of these steps are provided below.

We first build recurrence plots as illustrated in Fig. 35. A set of T time-variate trials $\{x_1, \ldots, x_T\}$ are first represented in a time-frequency domain. We propose to use a synchrosqueezing (SSQ) transform [10], [11]. This transform overcomes the energy smudging problem present in spectrograms computed using short-time Fourier transforms by performing wavelet-based filtering and signal power reassignment to appropriate frequencies. Next, frequency space is split into several frequency bands of interests (see Tab. IV for some examples). At every time instance, we compute mean values for chosen frequency bands and store these values in the final time-frequency representation, which is then used to obtain recurrence plots. Compared to classical recurrence plots built in time domain, frequency band representations allow insights on important signal characteristics and flexibility in analysis of particular frequency bands, with the price of additional computations (synchrosqueezing transform computations and mean filtering).

We note that physiological signals are known to be nonstationary, nonlinear and noisy. Statistical tests in general consider that main properties of underlying signals are not preserved in surrogate sets and underlying signal is often assumed to be stationary. Therefore, it is crucial to make reasonable assumptions about datasets and to perform statistical tests thoughtfully. Additional verifications about signal assumptions are to be done for datasets used in our work. For the completeness of this document, we make a digression to define stationary processes: a stationary process is a stochastic process whose joint probability distribution does not change when shifted in time.

Frequency bands	
$(0.5, 20]$ Hz	(δ, β)
$(0.5, 40]$ Hz	$full$
$[0.5, 4]$ Hz	δ
$(4, 8]$ Hz	θ
$(8, 12]$ Hz	α
$[12, 20)$ Hz	β

TABLE IV

Analyzed frequency bands: note that band definitions vary slightly from author to author.

Next step represents building the surrogate dataset. Classically, surrogates of time variate signals are

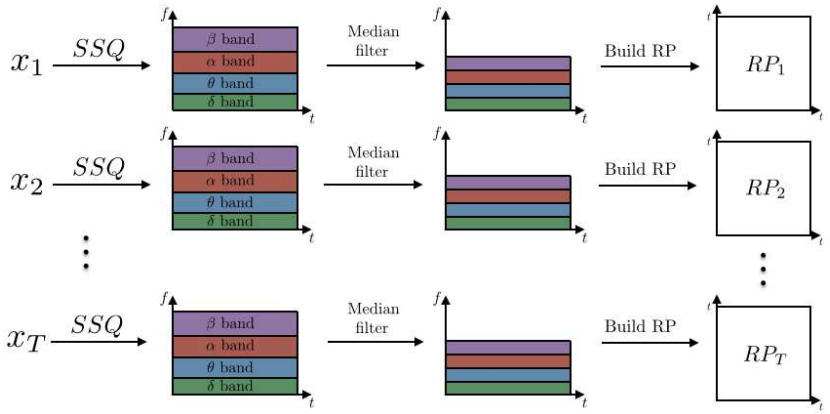


Fig. 35. Building recurrence plots given T time series $\{x_1, \ldots, x_T\}$. Processing blocks represented by arrows: (i) SSQ is a synchrosqueezing transform block used for obtaining time-frequency representations of signals; (ii) Median filter computes mean values for each of the chosen frequency bands, which reduces space dimensionality and computational costs for making RPs; (iii) Recurrence plots are built using distance-based thresholding with optimal parameter value ϵ, Eq. (1).

Measure	Formula
mean $\bar{x}$	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$
st. dev.	$\sigma = \sqrt{\sigma^2}$
Parameters estimated from the data	
variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N-1}(x_i - \bar{x})^2$
st. dev.	$\sigma = \sqrt{\sigma^2}$

TABLE V

CLASSICAL STATISTICAL MEASURES FOR AN N-DIMENSIONAL VECTOR $x \in \mathbf{R}^N$.

built to preserve some of the important features of the original time series, for example the spectrum magnitude, while replacing its spectrum phase by a random sequence of values. The reasoning behind this randomisation is that time domain reshaping destroys non-stationarities (phase keeps information about the spectral components time organisation), which means that local spectral components will vary while the global spectrum remains the same (time-invariant in case of a stationary process). As a consequence, the mean and variance parameters do not change over time. For more details on building time series surrogates are provided in [9] and references within. Other works, such as [12], [13] deal with testing stationarity of time-frequency signal representation using parametric and nonparametric methods to build surrogates.

We compare statistics of recurrence plots for original time-variate trials for several sensors to understand if recurrence plots preserve the common underlying signal dynamics in statistically significant way. To compare statistics of two signal a set of surrogate signals for each original signal is constructed as illustrated in Fig. 36. Then we compute the statistics of the original and surrogate series and compare it as illustrated in Fig. 37. The original signal is represented by the set of T recurrence plots built from captured trial values. We build S surrogate sets, each consisting of T recurrence plots. To perform, a statistical analysis, we first compute a statistics for the original signal set (for e.g., for each pixel of recurrence plots or for group of locally positioned pixels, as in Tab. V) and build confidence intervals (t-tests) or hypothesis that surrogates should fulfill. We repeat the statistics computation for the surrogate sets and compare it.

We compare pixel-related statistical measures for recurrence plots obtained from the time-frequency plane and their surrogates. We denote the set of the recurrence plots obtained from the original trial data by $\mathcal{T} = \{RP_k\}$, $k \in \{1, \ldots, T\}$ and each of their surrogate sets by $\mathcal{S}^{(k)} = \{S_k^{(l)}\}$, $l \in \{1, \ldots, S\}$. In our

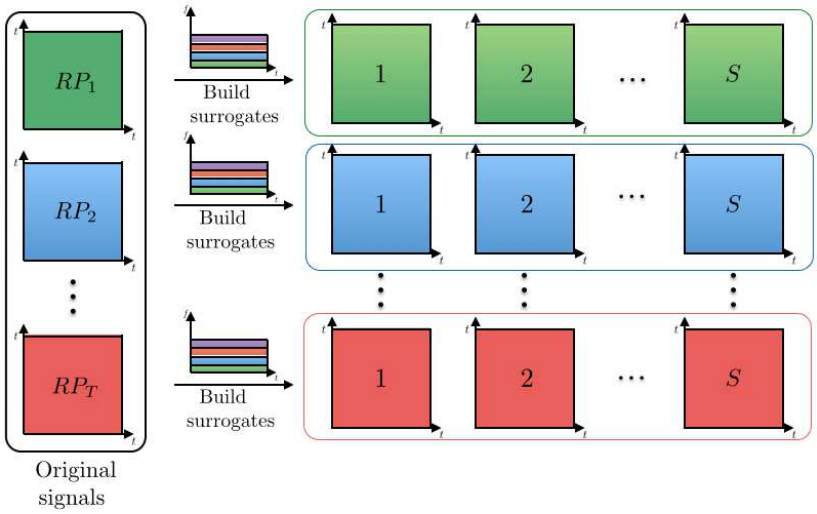


Fig. 36. Surrogate set construction: recurrence plots obtained from the original time-variate signals and surrogate recurrence plots have same energy distribution in time-frequency domain.

simulations, there is in total $T = 10$ trials, where for each of them we build $S = 100$ surrogates. For performing t-tests we compute a set of pixel-wise mean and variance values based on T recurrence plot and obtain $N \times N$ dimensional matrices of values. Then, we compute t-tests which verify if the mean values of S surrogate datasets fall into approximate confidence intervals defined by the original, trial data. The mean and variance values are estimated from the data. In the case that a variable distribution is normal, so will it be its mean. Then, for the mean $\bar{r_{i,j}}$ and variance $s_{r_{ij}}$, the approximate 95% and 99% confidence intervals are given with $(\bar{x} - 1.96 s_x) \leq \mu \leq (\bar{x} + 1.96 s_x)$ and $(\bar{x} - 2.58 s_x) \leq \mu \leq (\bar{x} + 2.58 s_x)$, respectively.

In the case when the variable is not normally distributed, but the number of samples is big enough (for e.g., bigger than 30), according to the central limit theorem we can still use the same confidence intervals as those given above. When this is not the case, we need to use the t-test. The confidence intervals for $T = 10$ trials (degree of freedom is $df = T - 1 = 9$) a range of p-values are given below in Tab. VI that are extracted from the table of Student's t-test for the completeness of this document.

We deal with pixel-based analysis and we visually represent the outcomes of tests in matrix $A = [a_{i,j}]$, whose element values are:

$$a_{i,j} = \begin{cases} 1, & \text{if the mean value falls into the confidence interval,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Mean value is computed as $\bar{x}_{i,j} = \frac{1}{T} \sum_{t=1}^{T} r_{i,j}$, while its variance is equal to $s_{\bar{x}_{i,j}} = \frac{s_{i,j}}{\sqrt{T}}$, $s_{i,j} = \frac{1}{T}(r_{i,j} - \bar{x}_{i,j})^2$.

Preliminary results based on ferret dataset show that surrogate datasets have different statistics than the original data. In Tab. VII we provide a percentage of surrogate recurrence plot pixels that fall into the computed confidence intervals for pixel-wise signal statistics. For example, 59% of mean values for surrogates fall into the $p = 0.2$ confidence interval of the original (if the surrogates had the same structure, this value would be equal to or higher from $1 - p = 80\%$ instead of 60%). This happens for all the explored bands, with exception of α and β bands for the value $p = 0.2$ (marked bold in the Tab. VII).

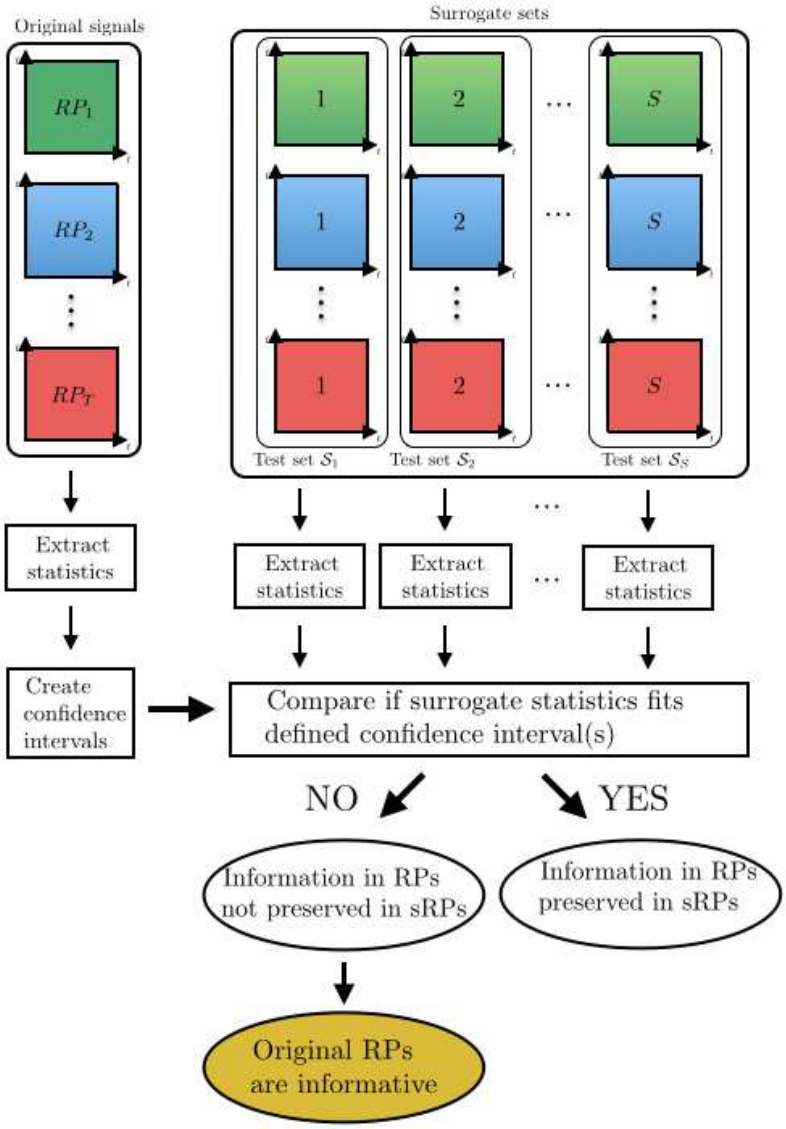


Fig. 37. Statistical analysis algorithm.

	values of p				
$df = 9$	0.2(20%)	0.1(10%)	0.05(5%)	0.02(2%)	0.01(1%)
	1.383	1.8333	2.262	2.821	3.250

TABLE VI

VALUES OF THE CRITICAL PARAMETER t ABOVE WHICH A SPECIFIED PROBABILITY p FAILS ARE GIVEN FOR A DEGREE OF FREEDOM $df = 9$ (NUMBER OF SAMPLES IS 10) AND A RANGE OF p VALUES (TWO-TAILED PROBABILITY).

Frequency bands	Confidence level of the interval				
	$0.2(20\%)$	$0.1(10\%)$	$0.05(5\%)$	$0.02(2\%)$	$0.01(1\%)$
Non-normalized signal					
$[0.5, 20)Hz$	0.5908	0.5927	0.6006	0.6006	0.6022
$[0.5, 40)Hz$	0.5979	0.6001	0.6079	0.6079	0.6097
δ	0.6034	0.6046	0.6104	0.6104	0.6116
θ	0.8325	0.8386	0.8386	0.8386	0.8443
α	0.6944	0.8883	0.8939	0.9040	0.9045
β	0.7315	0.8861	0.8913	0.9005	0.9009
Normalized signal					
$[0.5, 20)Hz$	0.3237	0.3237	0.3321	0.3321	0.3321
$[0.5, 40)Hz$	0.3009	0.3009	0.3085	0.3085	0.3085
δ	0.6194	0.6216	0.6253	0.6253	0.6279
θ	0.7289	0.7318	0.7341	0.7341	0.7362
α	**0.8110**	0.8204	0.8204	0.8205	0.8257
β	**0.8119**	0.8217	0.8217	0.8217	0.8264

TABLE VII

TABLE VALUES REPRESENT PERCENTAGE OF RECURRENCE PLOT POINTS THAT FALL INTO THE GIVEN CONFIDENCE INTERVAL FOR $S = 100$ SURROGATE SETS. ONE SURROGATE SET CONSISTS OF $T = 10$ RECURRENCE PLOTS (ONE SURROGATE FOR EACH OF 10 TRIALS).

VII. Future works

Ongoing project steps include extensive statistical analysis for different signal types and their surrogate datasets, as well as analysis of frequency-band feature signatures. For this purpose we will use hypothesis tests, t-tests and bootstraps, as well as the appropriate corrections where required (for e.g, Bonferroni correction).

Final goal is to provide appropriate measures and algorithms for analysis of human physiological data. A tool for analysis of EEG and MEG systems [14] represents a set of algorithms for building scalp maps and time-frequency images, provides electromagnetic source reconstruction and allows fitting functional integration in neural system. A method known as synchrosqueezing [10], [11] is stable and efficient time-frequency analysis algorithm which well adapts signal space representation to its nonstationarity properties, which makes it very interesting for biomedical signal processing domain. Other approaches study suitable measures for nonlinear analysis of sleep EEG signals [7], [15], [16].

In the following step of this project we will apply the recurrence plot analysis methods on polysomnography data obtained from the partner hospitals, namely, University Hospital of Zurich, Switzerland and CHU Nancy (Service de Neurologie). We will jointly analyse different physiological signals for sleep stage change detection and present its outcome to the Phillips research center in Netherlands.

References

[1] C. Stam, "Nonlinear dynamical analysis of eeg and meg: Review of the emerging field," *Clinical Neurophysiology*, no. 116, pp. 600–616, 2005.

[2] I. C., C. A., and Q. S. F. for the American Academy of Sleep Medicine, "The aasm manual for the scoring of sleep and associated events," *American Academy for Sleep Medicine, Westchester, IL*, 2007.

[3] P. H., "Sur la probleme des trois corps et les équations de la dynamique," *Acta Mathematica*, vol. 13, pp. 1–271, 1890.

[4] P. bei Graben and A. Hutt, "Detecting recurrence domains of dynamical systems by symbolic dynamics," *Physical Revew Letters*, vol. 110, 2013.

[5] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Letters*, vol. 5, pp. 973–977, 1987.

[6] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[7] U. R. Acharya, O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," *Computer Methods and Programs in Biomedicine*, vol. 80, pp. 37–45, 2005.

[8] S. Marčelja, "Mathematical description of the responses of simple cortical cells," *Journal of the Optical Society of America*, vol. 11, pp. 1297–1300, 1980.

[9] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 11, p. 346382, 2000.

[10] S. Meignen, T. Oberlin, and S. McLaughlin, "A new algorithm for multicomponent signals analysis based on synchrosqueezing: With an application to signal sampling and denoising," *IEEE Tras. on Signal Processing*, vol. 60, no. 11, pp. 5787–5798, 2012.

[11] F. Auger, P. Flandrin, Y.-T. Lin, S. McLauhlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing," *IEEE Signal Processing Magazine*, vol. 80, pp. 32–41, 2013.

[12] C. Richard, A. Ferrari, H. Amoud, P. Honeine, P. Flandrin, and P. Borgnat, "Statistical hypothesis testing with time-frequency surrogates to check signal stationarity," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3666–3669, 2010.

[13] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing Stationarity With Surrogates : A Time-Frequency Approach," *IEEE Trans. on Sig. Process.*, vol. 58, no. 7, pp. 3459–3470, 2010.

[14] V. Litvak, J. Mattout, S. Kiebel, C. Phillips, R. Henson, J. Kilner, G. Barnes, R. Oostenveld, J. Daunizeau, G. Flandin, W. Penny, and K. Friston, "EEG and MEG data analysis in SPM8," *Hindawi Publishing Corporation: Computational Intelligence and Neuroscience*, 2011.

[15] T. Kobayashi, K. Misaki, H. Nakagawa, S. Madokoro, H. Ihara, K. Tsuda, Y. Umezawa, J. Murayama, and K. Isaki, "Non-linear analysis of the sleep EEG," *Psychiatry Clin Neurosci*, vol. 53, no. 2, pp. 159–61, 1999.

[16] E. Pereda, A. Gamundi, R. Rial, and J. Gonzáles, "Non-linear behaviour of human EEG: fractal exponent versus correlation dimension in awake and sleep stages," *Neuroscience Letters*, vol. 250, pp. 91–94, 1998.