Quantification of cross correlations in complex spatiotemporal systems

Markus Müller,[1,*] Katja Wegner,[2] Ursula Kummer,[2] and Gerold Baier[1]
[1]*Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, 62210 Cuernavaca, Morelos, Mexico*
[2]*Bioinformatics and Computational Biochemistry, EML Research, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany*
(Received 16 August 2005; revised manuscript received 16 December 2005; published 6 April 2006)

We propose a design of the equal time correlation matrix suitable for the analysis of multivariate time series with ill-defined phases. We present the cross-correlation analysis of model data sets taken from coupled stochastic oscillators and compare the concept with the results obtained from a conventional correlation matrix analysis. We show that the concept provides a higher sensitivity combined with a better statistical significance when quantifying weak cross correlations.

DOI: 10.1103/PhysRevE.73.046106 PACS number(s): 02.50.Sk, 05.45.Tp, 05.10.−a, 89.75.−k

I. INTRODUCTION

During the last decade, analysis of the equal time correlation matrix has become a prominent tool to detect and classify correlations in highly irregular multivariate data sets [1–3]. In [4] it was shown that genuine correlations are imprinted in the spectrum of the eigenvalues and eigenvectors of the correlation matrix via level repulsion at the edges of the spectrum. By this mechanism a separation of the relevant information from statistical fluctuations can be achieved. While the central part of the spectrum is governed by random correlations, analysis of the largest and smallest eigenvalues and their corresponding eigenvectors yields detailed information about the correlation structure of the data set. The results proved particularly valuable for the detection of weak correlations in data sets with strong noise contamination. However, as the standard cross-correlation coefficient essentially compares the phases of oscillations it remains unclear whether the correlation matrix formalism can be applied successfully to data with ill-defined phases.

In the present paper we suggest a design of the correlation matrix tailored to the special requirements of stochastic time series with ill-defined phases like those obtained in various biological systems. Their common characteristic is the occurrence of sudden outbursts of activity from a more or less uniform baseline. To account for this feature we calculate the correlation coefficients from amplitude-weighted conditional probabilities of the bursts. Comparing the results with those obtained from the conventionally constructed correlation matrix we provide evidence that the design yields higher sensitivity and statistically more significant information about spatial correlations.

Data of spatiotemporal bursting are widespread in excitable biological and biochemical systems. Classical examples are provided by groups of electrically bursting neurons in substructures of brains like the thalamus, e.g., Ref. [5], and bursting beta cells in the mammalian pancreas [6]. In addition, recently a great variety of new imaging techniques allow to record spatiotemporal dynamics of both excitable and nonexcitable biochemical systems. A prominent representative is the fluorescence-based measurement of cytosolic calcium in living tissue, e.g., the liver, where irregular patterns associated with bursting in single cells were observed [7,8]. For the latter, a role of deterministic chaos in the generation of the irregularity has been proposed [9] but more recently detailed studies have emphasized stochastic fluctuations due to the small particle numbers involved [10]. As the quantification of correlations is important for the interpretation of the corresponding spatiotemporal patterns in both excitable and nonexcitable systems, we elaborated a method of analysis for such irregular multivariate time series.

II. CORRELATION MATRICES

The conventional equal time correlation matrix is constructed over a stationary time window t_k $(k=1,\ldots,T)$ of a multivariate dataset $Z_i(t_k)$ $(i=1,\ldots,M)$ where M is the number of time series. T, the number of data points in the time window, should be considerably larger than M. To provide a well-defined scale for all correlation coefficients, the data points of the chosen time window T are normalized,

$$\tilde{Z}_i(t_k) = \frac{Z_i(t_k) - \langle Z_i \rangle}{\sigma_i}, \tag{1}$$

where $\langle Z_i \rangle$ and σ_i are the average amplitude and the standard deviation, respectively, both calculated for the time window of length T. The equal time correlation matrix $\mathbf{C}$ [11] can then be constructed as

$$C_{ij} = \frac{1}{T}\sum_k \tilde{Z}_i(t_k)\tilde{Z}_j(t_k). \tag{2}$$

Due to the normalization (1) all entries of matrix (2) vary between plus and minus one depending in whether correlations or anticorrelations are present. All diagonal elements are equal to 1 because each signal is perfectly autocorrelated (zero lag). In consequence, the trace of the correlation matrix is given by $\mathrm{Tr}(\hat{C}) = \Sigma_{i=1}^{M} C_{ii} = \Sigma_{i=1}^{M} \lambda_i$ where λ_i $(i=1,\ldots,M)$ denote the eigenvalues of (2). By this construction, the measured correlations exclusively reflect relationships between the shapes and phases of the signals.

*Corresponding author. Electronic address: muellerm@servm.fc.uaem.mx

 ©2006 The American Physical Society

For cases however, where the time series lack smooth continuous oscillations, a method that quantifies phase-shape-correlations like the correlation matrix formalism [4] may not be the optimal tool of analysis. For instance, systems like the biochemical oscillators described below produce time series that are composed of a nonoscillatory baseline interrupted by sudden bursts of irregular shape. In this case the main question might be whether the times of occurrence and the amplitudes of the bursts in different signals are correlated. To answer this question we modify the design of the correlation matrix.

Let P_i be the probability to find a finite amplitude of a burst at time t of time series $Z_i(t)$, and let P_{ij} denote the conditional probability to find a nonzero burst amplitude simultaneously in time series Z_i and Z_j. Hence, if both time series are uncorrelated P_{ij} factorize $P_{ij}=P_iP_j$ on the average.

To take the magnitude of the bursts into account we define a matrix **A**, the elements of which are the amplitude weighted conditional probabilities,

$$A_{ij} = \frac{1}{T_a}\sum_{k=1}^{T} \frac{Z_i(t_k)Z_j(t_k)}{\sqrt{\langle Z_i^2\rangle\langle Z_j^2\rangle}}. \tag{3}$$

where $T_a = T^{\frac{P_i+P_j}{2}}$ is the arithmetic average over the number of data points where the signals $Z(t_k)$ are nonzero and $\langle Z^2\rangle$ is the average power of time series Z. Thus, $A_{ij}=A_{ji}$, as in the case of the conventional correlation matrix (2). The diagonal elements A_{ii} are equal to one and hence the trace of matrix **A** is equal to the number of time series M: $\mathrm{Tr}(\hat{A})=\sum_{i=1}^{M}A_{ii}=\sum_{i=1}^{M}\Lambda_i=M$ where Λ_i $(i=1,\ldots,M)$ denote the eigenvalues of matrix **A**.

If time series i and j are noncorrelated the value $A_{ij}=A_iA_j$ is equal to the amount of random coincidences of finite amplitudes of the signals with

$$A_i = \frac{1}{\sqrt{TP_i}}\sum_{k=1}^{T} \frac{Z_i(t_k)}{\sqrt{\langle Z_i^2\rangle}}. \tag{4}$$

The correlation coefficients of both C_{ij} and A_{ij} appear to be dominated by random entries and further analysis is required to extract quantitative information about genuine spatiotemporal correlations pattern. The main difference to the conventional correlation matrix (2) is that for data sets showing bursts with positive amplitudes and a zero baseline all nondiagonal elements of matrix **A** are positive by definition, $A_{ij}>0$.

III. TEST SYSTEM

In this paper we use a simple model of linearly coupled bursting oscillators to generate artificial multivariate time series.

An individual oscillator is represented by a system of three ordinary differential equations which describes the temporal evolution of three concentrations. The model is described in Ref. [9] and consists of the following equations:

$$\frac{dX(t)}{dt} = k_1 + k_2X - k_3\frac{XY}{X+k_4} - k_5\frac{XZ}{X+k_6},$$

$$\frac{dY(t)}{dt} = k_7X - k_8\frac{Y}{Y+k_9},$$

$$\frac{dZ(t)}{dt} = k_{10}X - k_{11}\frac{Z}{Z+k_{12}} = F_Z(X,Z). \tag{5}$$

Originally the model was derived as a simplification of a more realistic model for the bursting calcium dynamics in liver cells [9], but it is used here to construct a representative prototype of a bursting reaction-diffusion system.

To obtain a spatiotemporal system 10 of these cells were coupled diffusively in the Z variable by adding the following coupling term:

$$\frac{dZ_i}{dt} = F_Z(X,Z) - D(2Z_i - Z_{i-1} - Z_{i+1}), \tag{6}$$

where D is the coupling constant and Z_i denotes a concentration in the ith cell. The first and tenth cell are connected by periodic boundary conditions to yield a ring of identical oscillators.

In order to account for low particle numbers in the living cell, Eq. (5) together with Eq. (6) was stochastically simulated using the Gillespie algorithm [12] in the Copasi simulation software [13]. This stochastic simulation introduces intrinsic, particle number dependent noise to the dynamics of the deterministic model. In Fig. 1 we show representative time series of two Z-variables of our test system. There are occasional bursts occurring in only one of the signals (i.e., at 270–290 and 480–490 time units), bursts in both signals which occur almost simultaneously with approximately the same amplitudes (i.e., at 50–70 time units), and bursts which approximately coincide in time but not in amplitude (i.e., at 370–380 time units).

Figure 1(b) shows a space-time plot of 10 uncoupled units $(D=0)$ and Fig. 1(c) a space-time plot of 10 weakly coupled units $(D=0.1)$ of the test system. All time series are characterized by comparatively long segments of silence $[Z_i(t)=0$, coded in black] which are aperiodically interrupted by short bursts of irregular shape (grey coded). No apparent relationships are visible, neither between the bursts within one signal nor between the activity of neighboring cells.

The question arises, whether there are mathematical tools that are sensitive enough to detect even weak correlations in data of the type shown in Fig. 1 on an objective scale and hence distinguish them from random correlations and noise. For instance, one might expect that the coupling in Fig. 1(c) induces some kind of correlations when compared to the uncoupled case of Fig. 1(b). Simultaneously one is always interested in the statistical significance of the obtained results, an important aspect especially if only short data segments are available, a typical situation with experimental recordings in biochemistry.

IV. RESULTS

At first we determine the average cross-correlation coefficient c and its standard deviation σ_c of the nondiagonal elements of matrix **C**. Then we construct an $M\times M$ matrix

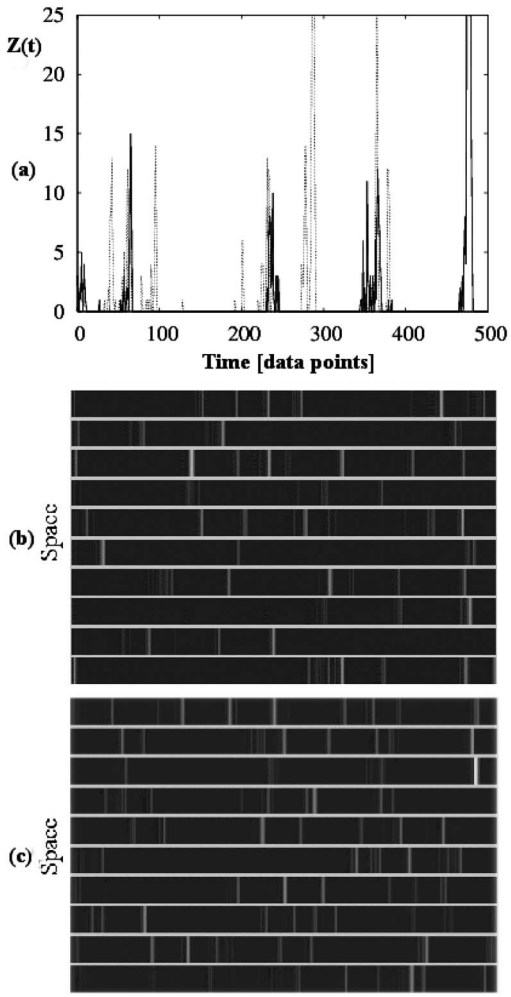

FIG. 1. Stochastically simulate time series of the test system Eq. (5) with Eq. (6). (Parameters, $k_1=0.212$, $k_2=3.2$, $k_3=4.88$, $k_4=1.18$, $k_5=1.52$, $k_6=0.19$, $k_7=1.24$, $k_8=32.24$, $k_9=29.09$, $k_{10}=13.58$, $k_{11}=153$, $k_{12}=0.16$, $D=0.1$, volume set to unity.) (a) Concentration-time plot of variables Z_4 (dashed line) and Z_5 (dotted line) with $D=0.1$ [cf. Fig. 1(c)], (b) space-time plot of the 10 units with $D=0$, (c) space-time plot of the 10 units with $D=0.1$.

with diagonal elements equal to unity and all nondiagonal entries equal to c,

$$\mathbf{C}_{\mathrm{av}} = \begin{pmatrix} 1 & & \\ c & 1 & c \\ & & 1 \end{pmatrix} \tag{7}$$

and two matrices with all nondiagonal entries based on σ_c which provide a measure for the statistical error,

$$\mathbf{C}_{\pm} = \begin{pmatrix} 1 & & \\ c \pm \frac{\sigma_c}{2} & 1 & c \pm \frac{\sigma_c}{2} \\ & & 1 \end{pmatrix}. \tag{8}$$

In this manner, one eliminates the fluctuations within the matrices but takes into account the statistical error in an explicit manner. It is straightforward to prove by induction that

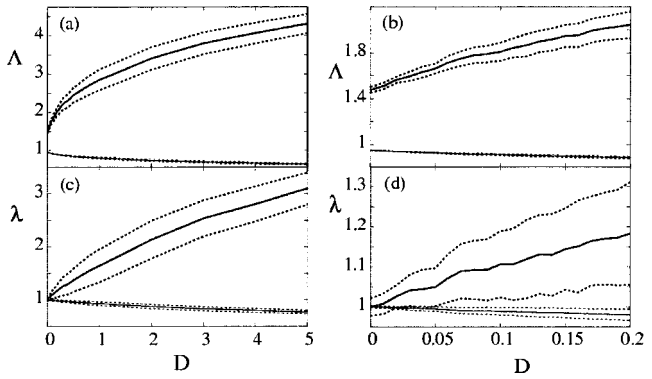

FIG. 2. Comparison of the conventional correlation matrix and the amplitude weighted correlation matrix. One matrix for **A** and **C** is constructed over a data segment of 150 000 sampling points. Eigenvalues of $\mathbf{C}_{\mathrm{av}}$ (solid line), $\mathbf{C}_{\pm}$ (dashed line), and $\mathbf{A}_{\mathrm{av}}$ (solid line), $\mathbf{A}_{\pm}$ (dashed line) are drawn as a function of coupling constant D. (a) Eigenvalues of $\mathbf{A}_{\mathrm{av}}$ and $\mathbf{A}_{\pm}$, (b) amplification of (a) for the range $D=0$ to 0.2, (c) eigenvalues of $\mathbf{C}_{\mathrm{av}}$ and $\mathbf{C}_{\pm}$, (d) amplification of (c) for the range $D=0$ to 0.2.

the characteristic polynomial of such matrices is given by

$$P_M(\lambda) = (1-\lambda-B)^{(M-1)}[1-\lambda+(M-1)B], \tag{9}$$

where B is the value of the nondiagonal elements of (7) and (8), respectively. These matrices have one large eigenvalue $\lambda_l=1+(M-1)B$ and $(M-1)$ degenerate small ones $\lambda_s=1-B$. Applying the same procedure to (3) yields matrices $\mathbf{A}_{\mathrm{av}}$ and $\mathbf{A}_{\pm}$ with one large eigenvalue $\Lambda_l=1+(M-1)B$ and $(M-1)$ degenerate small eigenvalues $\Lambda_s=1-B$, respectively. We would like to remark, that the difference of the eigenvalues from the matrices $\mathbf{A}_{\pm}|\lambda_{s+}-\lambda_{s-}|$ or equivalently $|\Lambda_{s+}-\Lambda_{s-}|$ is essentially smaller than the difference of the smallest and second largest eigenvalues of the full matrices **C** or **A**, respectively. The average over the nondiagonal elements (which is equivalent to a kind of spatial average) reduces the amount of statistical fluctuations caused by the random correlations.

Figure 2 shows the eigenvalues of matrices $\mathbf{C}_{\mathrm{av}}$ and $\mathbf{A}_{\mathrm{av}}$ and their statistical error (eigenvalues of $\mathbf{C}_{\pm}$ and $\mathbf{A}_{\pm}$) as a function of the coupling strength D calculated for a time segment of $T=150\,000$ sampling points in the test system.

Due to the normalization (1) the eigenvalues λ_l and λ_s of matrix $\mathbf{C}_{\mathrm{av}}$ are equal to 1 for $D=0$ because $C_{ij}=0$ on the average. With increasing coupling the degeneration is lifted and the difference of the eigenvalues $\Delta_c(D)=\lambda_l(D)-\lambda_s(D)$ increases continuously from zero to some maximum value. In contrast, $\Delta_a(D)=\Lambda_l(D)-\Lambda_s(D)$ of the eigenvalues of matrix $\mathbf{A}_{\mathrm{av}}$ is always finite, even for $D=0$. In this case, the average nondiagonal elements of the matrix **A** are given by the product of the average of the A_i of Eq. (4), i.e., $\langle A_{ij}\rangle = \langle A_i\rangle\langle A_j\rangle > 0$. However, as in the case of $\Delta_c(D)$, the difference of the eigenvalues increases with increasing coupling and hence a correlation measure can be defined based on the differences of the eigenvalues for both cases:

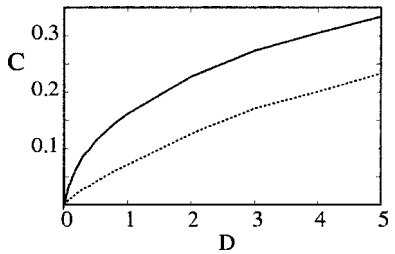

FIG. 3. Correlation measure C calculated from matrix **A** (solid line) and matrix **C** (dashed line) as a function of coupling constant D. One matrix for **A** and **C** is constructed over a data segment of 150 000 sampling points.

$$C(D) = \frac{\Delta(D) - \Delta(D=0)}{M - \Delta(D=0)}, \tag{10}$$

where $\Delta(D)$ denotes $\Delta_a(D)$ and $\Delta_c(D)$, respectively. Theoretically, $C(D)$ takes values between zero (no genuine correlations present) and one (perfectly correlated state, i.e., identical time series). However, due to the stochastic nature of the chosen data $C(D)$ practically never reaches its possible maximum value, not even for very large values of D.

Figure 3 shows a comparison of the correlation measure $C(D)$ calculated for the matrices **C** and **A**, respectively. In both cases, $C(D)$ increases continuously from 0 to some maximum value. Nevertheless, except for $D=0$, $C(D)$ always assumes higher values when calculated from matrix **A** indicating that the correlation matrix based on amplitude weighted conditional probabilities is more sensitive to the detection of correlations than the conventional correlation matrix. This is especially true for the case of small couplings where $C(D)$ calculated for matrix **A** increases more rapidly than in the case of matrix **C**. At $D=0.2$ the correlation measure $C(D)$ calculated for matrix **A** is about three times the value for the case of matrix **C** pointing to a higher sensitivity of the concept. For large couplings the slopes of both curves approach the same value, as should be expected.

In order to estimate the statistical significance of $C(D)$ we define a function

$$R(D) = \frac{2[\Delta(D) - \Delta(D=0)]}{\Sigma(D)} \tag{11}$$

where $\Sigma(D)=[\sigma_l(D)+\sigma_s(D)]$ is the sum of the error bars of the large and the small eigenvalue, i.e., $\sigma_l(D)=|\lambda_{l+}-\lambda_{l-}|$ and $\sigma_s(D)=|\lambda_{s+}-\lambda_{s-}|$.

Thus, the statistical significance of the separation of eigenvalues and hence the correlation measure (10) can be defined as

$$S(D) = \exp[-R(D)]. \tag{12}$$

If the statistical error $\Sigma(D)/2$ is smaller than the average separation of the eigenvalues $\Delta(D)-\Delta(D=0)$, the value of $S(D)$ will be smaller than $1/e$, otherwise it takes values between zero and $1/e$. In Fig. 4 $S(D)$ is shown as a function of coupling strength D for correlation matrices (2) and (3). The results demonstrate the superior performance of the correlation measure calculated from matrix **A**. At $D=0.2$ the value

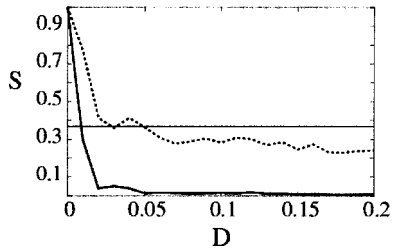

FIG. 4. Statistical significance S of the correlation measure as a function of the coupling D for the case of matrix **A** (solid line) and **C** (dashed line). For comparison the value e^{-1} is drawn as a horizontal solid line. The matrices are constructed over a data segment of 150 000 sampling points.

of $S(D)$ is about two orders of magnitude smaller in the case of matrix **A** than for matrix **C**. In this way, $S(D)$ provides an objective criterion for the statistical significance of the detected cross correlations.

The results presented so far were obtained for comparatively long stationary time series, which may not be a realistic assumption for some experiments. Therefore, in Fig. 5 we present the results for $C(D)$ and $S(D)$ in a case where instead of $T=150\,000$ only $T=1000$ data points (which corresponds to approximately 20 bursts) are taken into account for the construction of the matrices **C** and **A**, respectively. In contrary to the case of the long sequences, the curves are obtained by averaging over 100 trials. Again, as seen particularly from measure $S(D)$, the performance of the method using matrix **A** is significantly better than the approach based on matrix **C**. Already at $D\approx 0.05$ the value of $S(D)$ for matrix **A** crosses the critical value $1/e$, whereas for matrix **C** the same is true at $D\approx 0.45$. The results in Fig. 5 demonstrate the superior performance of the method based on matrix **A** especially in the case of short, weakly correlated time series.

In many situations not only the detection of correlations is desired but also the the spatial distribution of dependencies is

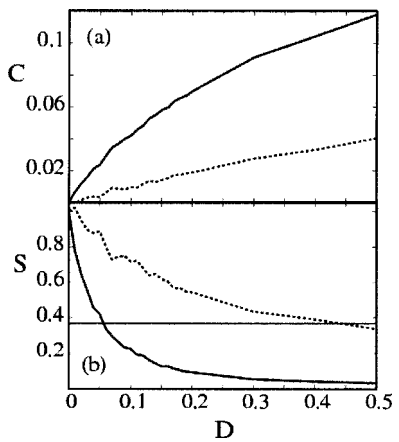

FIG. 5. Comparison of the correlation measure C (a) and the statistical significance S (b) of the correlation matrices **A** (solid line) and **C** (dashed line), respectively, as a function of the coupling strength D. The matrices are constructed over a data segment of 1000 sampling points. The results are obtained by averaging over 100 trials.

of interest. In Ref. [4] it was shown how to extract information about the spatial properties of the correlation structure by analyzing the eigenvectors of the correlation matrix. In general, if K time series of an M-dimensional multivariate data set are correlated, K eigenstates of the correlation matrix repel. This repulsion occurs exclusively between states at the edges of the spectrum, where the number of increasing states at the upper end and decreasing states at the lower end of the eigenvalue spectrum is determined by the specific structure of the correlation pattern [4]. In order to check the performance of the correlation matrix (3) in this respect, we perform an analysis of a data set retrieved from the test system when only two of the 10 units are coupled with $D>0$. The coupling between the remaining oscillators is set to zero. In this situation, the eigenvector $\vec{v}_{10}$ corresponding to the largest eigenvalue Λ_{10} of matrix $\mathbf{A}$ repels with the eigenvector $\vec{v}_1$ corresponding to Λ_1, independently of which pair of the $M=10$ signals is correlated.

In order to identify a subset of $K<M$ correlated signals of an M-dimensional multivariate data set one must provide an interpretation of the basis in which the matrix $\mathbf{A}$ is written. Each column or row of matrix $\mathbf{A}$ (and matrix $\mathbf{C}$, respectively) represents the correlations between a particular signal $Z_i(t)$ and all others. Hence, each basis vector represents the mutual correlations measured in a particular time series. Therefore, each of the components a_{ij} $j=1,\ldots,M$ of an eigenstate $\vec{v}_i$ can be assigned to a specific time series $Z_j(t)$ (the "channel basis" [3,4]).

The repulsion of states at the edges of the spectrum of matrix $\mathbf{A}$ leads to a mixing of their components in a similar manner as in interference processes of quantum states [14,15]. In our case, the eigenvectors involved in this repulsion process collect significant contributions of those components which belong to the correlated subspace [4], i.e., it is expected that those components a_{ij} of $\vec{v}_1$ and $\vec{v}_{10}$ which belong to the two correlated signals should be larger than the others. The magnitude of those a_{ij} depend on the amount of correlations and hence on the strength of the coupling between the units. Therefore, by investigating the structure of the eigenvectors it is possible to determine which of the time series are correlated.

In order to extract information about the spatial structure of the correlation pattern, it is not convenient to analyze the eigenvectors of the matrices $\mathbf{A}_{\mathrm{av}}$ and $\mathbf{A}_{\pm}$, because they reflect a spatial average of the correlation. Instead one must investigate the properties of matrix $\mathbf{A}$ which do contain specific information about the spatial distribution of the correlations.

Figure 6 shows the results for the eigenvector components $|a_{ij}|^2$ of matrix $\mathbf{A}$ when only unit 1 and 10 of the test system are coupled with $D=0.1$. The results are obtained by averaging over 500 matrices constructed from data segments of length $T=15\,000$. The 95% confidence level is drawn with dashed lines (note that the error given by the dashed lines corresponds to a range of about four times the standard deviation assuming normal distribution for $|a_{ij}|^2$). The figure shows a comparatively large value for the first and last components of the eigenvectors $\vec{v}_{10}$ and $\vec{v}_1$ corresponding to the largest and smallest eigenvalue. The components of all other eigenvectors are almost uniformly distributed. As an example, the components of $\vec{v}_9$ are drawn for comparison.

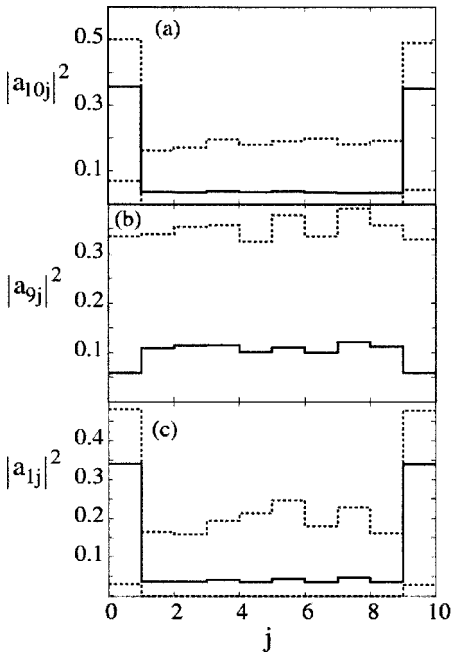

FIG. 6. Eigenvector components from $\mathbf{A}$ calculated for 500 matrices over a data segment of 15 000 sampling points per matrix when only two from 10 units of the test system are coupled with $D=0.1$. (a) Components of the eigenvector $\vec{v}_{10}$ corresponding to the largest eigenvalue Λ_{10}, (b) components of $\vec{v}_9$, and (c) components of $\vec{v}_1$. Solid line, average value over 500 trials, dashed lines, 95% confidence levels.

These results fit precisely to the interpretation given above. The repelling states mix almost exclusively within the coupled subspace of the M-dimensional configuration space. Hence, the corresponding eigenstates show large values of those $|a_{ij}|^2$ which correspond to the coupled units. Consequently, the spatial structure of the coupling between the M oscillators of the test system (and hence the induced correlations) can be seen directly via the structure of the repelling eigenstates.

In the given example the quality of the results can be improved by averaging over the $\vec{v}_1$ and $\vec{v}_{10}$. In Fig. 7 we show the results of the average over the first and last com-

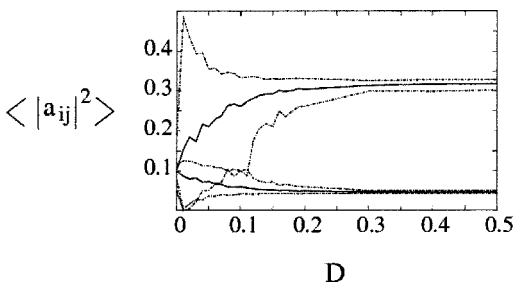

FIG. 7. Average over the first and tenth component of the eigenvectors corresponding to the largest and smallest eigenvalue of matrix $\mathbf{A}$ as a function of the coupling strength D, calculated over a data segment of 15 000 sampling points when only two from 10 units of the test system are coupled. Solid line, average over the first and tenth component; dashed line, average over the components from 2 to 9, dashed-dotted line, 95% confidence levels.

ponent as well as the average over the remaining components as a function of the coupling strength. For each value of D an ensemble of 500 matrices are calculated over a data segment of 15 000 sampling points. The 95% confidence level for each of the averages is drawn with dashed lines.

For increasing coupling the average value of the components 1 and 10 increases almost monotonically whereas the average of components from 2 to 9 decreases. Simultaneously the statistical significance of the obtained results increases with stronger coupling, i.e., the width of the confidence interval gets smaller. Hence, the stronger the coupling between the oscillating units, the stronger is the mixing of the eigenstates within the correlated subspace and consequently the larger is the magnitude of the components belonging to the time series of the correlated oscillators.

Therefore, if the coupling is strong enough the method allows to identify the coupled units via the structure of the largest and smallest eigenvectors. In general, for successful identification it is true that the weaker the coupling the longer is the required time segment for the construction of the correlation matrix. This is especially true for time series with a large noise component as in the present case.

V. CONCLUSIONS

In this paper we presented a design of the equal time correlation matrix based on the probability to find simultaneously bursts of comparable amplitude in two time series of a multivariate data set. We could show via direct comparison that the performance of the formalism in Ref. [4] is superior when the new correlation matrix is used. Especially for weak relationships between the signals the correlation measure calculated from matrix $\mathbf{A}$ is more sensitive with a higher statistical significance than the results obtained with matrix $\mathbf{C}$. This is relevance for the evaluation of nonstationary and comparatively short data sets which do not yield significant quantitative information about spatial correlations when the standard cross-correlation coefficients are evaluated. For strong correlations as well as for continuously oscillating systems the results obtained from both matrices are of comparable quality.

It turned out that by evaluating the difference of the eigenvalues of the matrices $\mathbf{C}_{\mathrm{av}}$ and $\mathbf{A}_{\mathrm{av}}$ the sensitivity for the detection of the presence of correlations can be drastically improved. The error bars given by the eigenvalues of the matrices $\mathbf{C}_{\pm}$ and $\mathbf{A}_{\pm}$ are much smaller than the range between the smallest and second largest eigenvalue of the full matrices $\mathbf{C}$ or $\mathbf{A}$, respectively.

Nevertheless, if for a given system not only the detection of correlations but also the localization of correlated units is of interest, the non-averaged correlation matrix must be considered. Then the spatial correlation structure of the system is imprinted in the structure of the eigenvectors corresponding to the largest and smallest eigenvalues. However, for the identification of the coupled units a better statistics is needed which can be achieved by longer stationary time series.

Furthermore, in many experimental situation the case of zero coupling (noncorrelated case) cannot be realized easily. Hence, information about the value of $\Delta(D=0)$, needed for the calculation of the correlation measure $C(D)$, is not available directly. In this case, the uncorrelated case could be simulated by creating surrogate data from the experimental results. However, the usual random shuffling of partial phases is not appropriate for data like those obtained from the test system. Here, a random permutation and shifting of the bursts seems to be the adequate strategy for the creation of surrogates.

We believe that the method based on matrix $\mathbf{A}$ can be successfully applied to a variety of bursting biochemical systems [7] and bursting media in general [5,6]. Since more and more high-throughput techniques are developed in biology measuring a multitude of different data for many time points, e.g., metabolites or proteins in the cell, and it is often of interest which concentrations correlate, we think that our method will become applicable to more and more systems in the future. In principal, the method should be also valuable for the analysis of microarrays looking for coexpressed genes. However, in this case there is a strong limitation on the number of time points. Therefore, detailed investigations to determine the lower limits for the number of time points needed for the analysis should be done before application.

Another attractive application is the analysis of the highly irregular dynamics of complex excitable systems like networks of firing or bursting neurons. Here, multivariate data can be obtained from recordings of electrode arrays or by application of voltage-dependent dyes. In this case an amplitude threshold should be set in order to discriminate the noisy basal state from the information-containing firing state. Thereby one creates time series consisting of the spikes or bursts separated by intervals of zeros similar to the output of the simulation of the test system. The presented technique based on the matrix $\mathbf{A}$ then provides information about cross correlations between spiking activity of different neurons and the presence of spatiotemporal correlations might be related to information-processing activity of the system.

ACKNOWLEDGMENTS

This work was supported by CONACyT, Mexico (Project No. 40885-F), and the Klaus Tschira Foundation. The authors thank Sven Sahle for helpful discussions.

[1] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[2] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amarel, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[3] V. Plerou, P. Gopikrishnan, B. Rosenow, Luis A. Nunes Amarel T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[4] M. Müller G. Baier, A. Galka, U. Stephani, and H. Muhle, Phys. Rev. E **71**, 046116 (2005).
[5] J. Tsoukatos *et al.*, Exp. Brain Res. **113**, 273 (1997).
[6] P. M. Dean and D. R. Matthews, Nature (London) **219**, 389 (1968).
[7] K. Motoyama *et al.*, Am. J. Physiol. **276**, R575 (1999).
[8] C. J. Dixon *et al.*, Biochem. J. **269**, 499 (1990).
[9] U. Kummer *et al.*, Biophys. J. **79**, 1188 (2000).
[10] U. Kummer, B. Krajnc, J. Pahle, A. K. Green, C. J. Dixon, M. Marhl, Biophys. J., **89**, 1603 (2005)
[11] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
[12] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[13] Copasi Project, www.copasi.org
[14] M. Müller, F.-M. Dittes, W. Iskra, and I. Rotter, Phys. Rev. E **52**, 5961 (1995).
[15] C. Jung, M. Müller, and I. Rotter, Phys. Rev. E **60**, 114 (1999).