Quantification of intra- and inter-cluster relations in nonstationary and noisy data

Christian Rummel*
Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, 62209 Cuernavaca, Mexico
(Received 13 July 2007; revised manuscript received 19 September 2007; published 17 January 2008)

The interrelation between data channels of multivariate data sets may lead to cluster formation. Revealing the cluster structure can give important information about the underlying systems' properties. Here we investigate the features of a recent genuinely multivariate cluster detection algorithm that is suitable for time-resolved and unsupervised application to nonstationary and noisy time series. Using numerical test systems it is discussed under which conditions intra- and inter-cluster relations can be disentangled and quantified. In addition different types of errors occurring when channels are automatically attributed to clusters are investigated quantitatively. Finally, the algorithm is applied to nonstationary model time series and its time-dependent performance is compared to other cluster detection algorithms.

DOI: 10.1103/PhysRevE.77.016708 PACS number(s): 05.10.−a, 89.75.Fb, 05.45.Tp

I. INTRODUCTION

Many complex systems stemming from different fields of science like physics, engineering, geology, meteorology, sociology, and medicine are assessed by measuring multichannel data. Another prominent example is the dynamics of financial data of stock markets. In all these fields important problems are the revealing of structures in the data, information reduction or the characterization of collective dynamics in complex systems whose interactions are not always known. To this end identification and classification of clusters (groups) within the $M \gg 1$ data channels that are related to each other in terms of any kind of interrelation measure is desired. A reliable, computationally simple and parameter free approach would find a wide range of applications in all the fields mentioned.

An overview over classic clustering algorithms is given in [1,2]. In general they can be divided into partitional and hierarchical methods. Another important division can be made between hard and fuzzy algorithms. In addition to the classic approaches in the past two decades methods of statistical physics have increasingly been used for cluster detection. Most of them postulate or derive a *cost function*, which is minimized in high-dimensional spaces by different techniques [3–9]. An alternative method bases on the size of the components of the largest *eigenvectors* of the equal-time cross-correlation matrix [10–12] or synchronization matrix [13,14]. As the algorithm proposed in the present paper uses eigenvectors as a basic ingredient, too, the foundations, success, and problems of this approach will be discussed in more detail in the subsequent section. An important problem of most of the methods mentioned so far is that they either suffer from an elevated computational effort or require the definition of some kind of threshold parameter (occasionally fixed by the visual inspection of some statistics). In addition for application to nonstationary data a continuous adjustment of the parameters may become necessary.

Very recently, an approach to detection of synchronization clusters in multivariate time series was developed that is based on coarse graining of Markov chains [15] and seems to be suitable for unsupervised application. In the present paper, we investigate the properties of another recently introduced method for automated partitional clustering of nonstationary time series that does not suffer from the aforementioned deficiencies [16]. Both number and size of the clusters are determined in a self-contained manner, i.e., without setting any artificial parameters. The algorithm requires an optimization procedure in a relatively low dimensional space only which makes it computationally fast. Consequently, it is a good candidate for future unsupervised application to long-term recordings of possibly nonstationary data, such as, e.g., human electroencephalograms (EEG).

The paper is organized as follows. In the subsequent section we discuss cluster algorithms based on the largest eigenvectors of interrelation matrices. Thereafter, we present in Sec. III an algorithm that still uses eigenvalues and eigenvectors but does not suffer from their typical deficiencies. In Sec. IV it is discussed how the intra- and inter-cluster relations in the analyzed data can be quantified. The performance of the algorithm is explored in Sec. V. Finally, application is made to model data with time-dependent interrelation structure in Sec. VI before discussing our results in Sec. VII.

II. CLUSTERS AND EIGENVECTORS

A central quantity used for cluster detection in [6–16] are normalized $M \times M$ interrelation or similarity matrices $\mathbf{C}$ whose elements satisfy the properties

$$-1 \leq C_{ij} \leq 1, \tag{1}$$

$$C_{ii} \equiv 1, \tag{2}$$

$$C_{ij} = C_{ji}. \tag{3}$$

Here $C_{ij}=0$ describes absent and $|C_{ij}|=1$ maximal mutual interrelation of data channels "i" and "j" with respect to an interrelation measure which we do not specify further at present. The property (2) of the diagonal elements reflects that each channel is perfectly related to itself and the symmetry property (3) implies real eigenvalues that will be or-

*crummel@web.de

dered according to $\lambda_l \geq \lambda_{l-1}$ in the sequel. Examples for interrelation measures that satisfy these demands are the linear equal-time cross-correlation or Pearson's coefficient [17,18] or nonlinear measures like mean phase coherence [19,20], partial phase synchronization [28] and (normalized) mutual information [21,22]. For the latter recently an algorithm for hierarchical clustering has been proposed that makes use of the grouping property of that measure [23]. Note that many other valuable measures for the interrelation of two time series like Granger causality [24,25], directed transfer function [26], and the very similar partial directed coherence [27], which can be used to identify a directedness of the mutual influence, do not satisfy Eq. (3). We do not consider such measures in the present publication. As the clustering algorithm to be investigated does not depend on the details of the used interrelation measure the discussion will be kept general, specifying to linear cross-correlation first in Sec. V.

In the sequel we define clusters on the basis of interrelation matrices $\mathbf{C}$ as closed groups of channels that are mutually interrelated by matrix elements C_{ij} whose absolute value is on average significantly larger than the average connection from this group to the remaining channels. In the idealized situation of total absence of connections from K clusters to the rest such matrices can in principle be written in block diagonal form by introducing a useful labeling of the channels. Then the cluster structure becomes immediately visible: $\mathbf{C}=\mathrm{diag}(\mathbf{C}_{11}\ldots\mathbf{C}_{KK})$. As relabeling of channels does not affect the structure of the network we start our discussion from the following block model without loss of generality:

$$\mathbf{C}=\begin{pmatrix} \mathbf{C}_{11} & \mathbf{d}_{12} & \cdots & \mathbf{d}_{1K} & \mathbf{0} \\ \mathbf{d}_{12} & \mathbf{C}_{22} & \cdots & \mathbf{d}_{2K} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{d}_{1K} & \mathbf{d}_{2K} & \cdots & \mathbf{C}_{KK} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{4}$$

Here the quadratic matrices $\mathbf{C}_{kk}$ $(k=1,\ldots,K)$ have dimension $m_k \geq 2$ which fulfill $\Sigma_k m_k \leq M$. All diagonal elements fulfill Eq. (2) and its off-diagonal elements have identical values $0 \leq c_k \leq 1$. The $K(K-1)/2$ matrices $\mathbf{d}_{kl}$ have identical elements d_{kl}. In the Appendix some properties of this simplified but generic model are discussed in detail.

Let us mention that different points of view are possible here. One possibility is to interpret the matrices $\mathbf{C}_{kk}$ as the clusters with *intra-cluster* relations modeled by the c_k. The matrices $\mathbf{d}_{kl}$ model possible *inter-cluster* relations. In the limit $c_k \gg |d_{kl}|$ this interpretation in terms of distinct but weakly connected clusters is surely a natural one. We will stick to this point of view throughout the paper. However, in the opposite limit $c_k \gtrsim |d_{kl}|$ it might be more appropriate to have in mind a single cluster with substructure.

Using model (4) cluster detection algorithms can try to find block patterns in interrelation matrices that are constructed from the data under investigation. Indeed, in [8] an (appropriately filtered) correlation matrix constructed from stock return time series is brought into an approximately block diagonal form with $|c_k| \gg |d_{kl}|$. To this end the channel labels "i" are relabeled in such a way that a cost function is minimized using a computationally expensive "traveling salesman" algorithm. However, there is the deficiency that for any time series of finite length T the genuine system-specific correlations described by the elements C_{ij} of correlation matrices are contaminated by noise and random correlations of the order $\sim 1/\sqrt{T}$. A step toward the solution of this problem consists in matrix diagonalization; see e.g. [11,29,30]. It has been found that the genuine system-specific correlations one is interested in manifest themselves via nonrandom level repulsion [12,31] and an increased rigidity of the spacing distributions [32] between subgroups of eigenvalues of the correlation matrix.

In the present paper we exploit properties of the eigenvalues λ_l and eigenvectors $\mathbf{v}_l$ of interrelation matrices $\mathbf{C}$ in order to gain information about the cluster structure of the data. It can easily be shown for the model (4) with $d_{kl}=0$ that for systems containing K clusters with m_k contributing channels each, the rule is that for every cluster one eigenvalue is increased with respect to the uncorrelated situation and m_k-1 eigenvalues are decreased; see the Appendix. The unchanged eigenvalues $\lambda_l \equiv 1$ will be addressed as the "bulk" in the sequel. In addition in this simplified situation the cluster structure can directly be deduced from the eigenvectors of the K largest eigenvalues (henceforth abbreviated as "largest EV"). In general, in the sense of principal component analysis (PCA) these eigenvectors point to the directions of maximal variance; see e.g. [33].

Already in one of the first applications of methods taken from random matrix theory to analysis of financial time series it was suggested to investigate the components of eigenvectors having large inverse participation ratios in order to identify industry sectors [30]. Indeed, using correlation matrices constructed from stock returns, in [10,11] financial market sectors could be identified by looking for large components a_{il}^2 of the K' largest EV $\{\mathbf{v}_l\}$. In [12] this approach was refined by introducing an additional criterion that is meant to suppress wrong attributions due to incidentally large a_{il}^2. However, in many practical situations a clear separation of the clusters is difficult to obtain due to the presence of random and inter-cluster relations, which provoke a mixing of the eigenvector components. For example, in Fig. 1 of [10] some industry sectors are represented by symmetric and antisymmetric linear combinations of eigenvectors corresponding to large eigenvalues of a correlation matrix constructed from stock return time series. Similarly in Fig. 1 of [8] the mixed structure of the second to fifth largest EV is shown.

For illustration, we give in Fig. 1(a) the example of $N_{\mathrm{ens}}=100$ numerical realizations of a variant of the block model (4) where $K=3$ groups of size $m_1=7$, $m_2=5$, and $m_3=3$ are correlated within a total of $M=20$ channels. Clusters "1" and "2" have positive inter-cluster relation, whereas clusters "2" and "3" are antirelated. Different to the restriction to identical matrix elements c_k and d_{kl} made below Eq. (4) the off diagonal matrix elements of the blocks $\mathbf{C}_{kk}$ and $\mathbf{d}_{kl}$ are drawn independently from Gaussian distributions with center $c_k^{(0)}=c^{(0)}=0.6$, $d_{12}^{(0)}=-d_{13}^{(0)}=-d_{23}^{(0)}=d^{(0)}=0.3$ and (relatively large) widths $\delta c=\delta d=0.15$. In addition the matrices $\mathbf{0}$ of Eq. (4) are replaced by random numbers that are drawn from a Gaussian

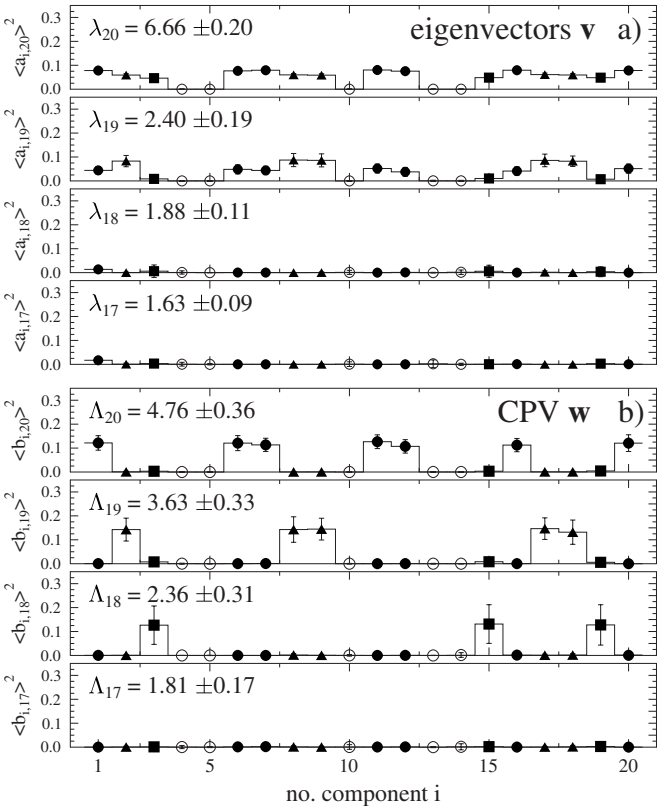

FIG. 1. Comparison of the squared average components of the $K'=4$ largest EV $\mathbf{v}_{M+1-k}$ of the interrelation matrix with those of their linear combinations $\mathbf{w}_{M+1-k}$, $k=1,\ldots,K'$ for $N_{\mathrm{ens}}=100$ independent realizations. The error bars are often of the size of the symbols. The channels that contribute to the $K=3$ clusters are marked with full symbols (●, ▲, ■) and the uncorrelated ones with open circles (○).

distribution with zero mean and width $\delta=0.15$. Whenever a matrix element left the range $[-1,1]$ it was set to the maximal or minimal value. In Fig. 1(a) almost all channels that correspond to one of the clusters contribute to the largest two eigenvectors $\mathbf{v}_{20}$ and $\mathbf{v}_{19}$ with a comparable strength a_{ik}^2. In the context of M financial time series the contribution to $\mathbf{v}_M$ has been called the "market wide effect." Contributions to the lower eigenvectors $\mathbf{v}_{18}$ and $\mathbf{v}_{17}$ average out entirely. As one can see, in the presence of random and inter-cluster relations using the largest few eigenvectors a clear distinction is only possible between the total set of clustered (full symbols) and the unclustered channels (open circles) whose squared components are often orders of magnitude smaller.

In [13,14] *participation indices* (PI) $p_{il}=\lambda_l a_{il}^2$ were introduced as a step towards the solution of the problems that arise when channels are attributed automatically to clusters on the basis of the largest EV. The PI quantify the "absolute involvement" of channel "i" in cluster "l." Its "strength" is given by the eigenvalue λ_l and the "relative involvement" of the channels by the squared components a_{il}^2 of the corresponding eigenvector. Every channel "i" is attributed to the cluster "k" with the maximal PI: $p_{ik}=\max_l p_{il}$. Finally only clusters corresponding to an (often small) number K' of largest eigenvalues $\lambda_k>1$ are accepted. The concept of PI works well in situations with a pronounced cluster structure and a strong repulsion of the largest eigenvalues from the "bulk" that is affected by random correlations only. Furthermore, it provides a conceptionally simple algorithm which is applicable in automated manner and real time. However, it tends to find "pseudoclusters" [14] from channels having incidentally large components in eigenvectors that correspond to smaller eigenvalues. In addition it was found in [14] that the PI are easily misled in situations of equal cluster size and large inter-cluster relations.

III. CPV ALGORITHM

A. Cluster participation vectors

The basic idea of our approach [16] is to find a set of orthonormal linear combinations $\{\mathbf{w}_l\}$—called *cluster participation vectors* (CPV) in the sequel—of the $K'<M$ largest EV $\{\mathbf{v}_l\}$ that have dominant entries exclusively for those components which correspond to the cluster "l." For matrices of the form (4) with $\mathbf{d}_{kl}=\mathbf{0}$ the CPV are identical to the largest EV: $\mathbf{w}_{M+1-k}\equiv\mathbf{v}_{M+1-k}^{(0)}(k=1,\ldots,K')$; see the Appendix. If this condition is violated the $\{\mathbf{w}_l\}$ are of course *not* eigenvectors of the interrelation matrix $\mathbf{C}$ anymore. Rather suitable vectors can be found by maximizing the *distance measure*

$$D_{ln}=D(\mathbf{w}_l,\mathbf{w}_n)=\sum_{i=1}^{M}|b_{il}^2-b_{in}^2|, \tag{5}$$

where the b_{il} are the components of vector $\mathbf{w}_l$. Note, that D_{ln} satisfies all requirements of a metric (symmetry, positive semidefiniteness, triangle inequality). D_{ln} is equal to zero for vectors where all components are identical to each other up to a sign and assumes its maximum value $D_{ln}=2$ for normalized vectors $\mathbf{w}_l$ and $\mathbf{w}_n$ without common components.

It is important to note that we do not use Eq. (5) for distance clustering [1,2] in the sequel, but to minimize common components between linear combinations of the largest EV. In general the CPV can be obtained by a rotation of the K' largest EV $\{\mathbf{v}_l\}$ about the generalized Euler angles (GEA) α_κ, $\kappa=1\ldots K'(K'-1)/2$ such that the sum

$$\Delta=\sum_{l=M-K'+1}^{M}\sum_{n>l}D(\mathbf{w}_l,\mathbf{w}_n) \tag{6}$$

is maximized. An explicit recipe for the construction of the rotation matrix needed in K' dimensions (in Sec. III C it is discussed how one can fix this dimension) is given in [34]. Due to the symmetry of the problem under reflection of the axes it suffices to use angles in the interval $0\le\alpha_\kappa\le\pi/2$ in our case. In order to obtain the global maximum of $\Delta(\alpha_\kappa)$ in the $K'(K'-1)/2$-dimensional space of the α_κ we apply the great deluge algorithm (GDA) of [35] which works as follows: Arbitrary points in the $K'(K'-1)/2$-dimensional landscape are tested for their value of Eq. (6). All points are accepted that satisfy the only restriction that Δ is larger than a "water level" $W\ge W_{\min}=\min(\Delta)=0$. If the point is in the allowed region W is increased by the "rainfall" rate δW and

TABLE I. Mutual distances $D(\mathbf{v}_l,\mathbf{v}_n)$ (above diagonal) and $D(\mathbf{w}_l,\mathbf{w}_n)$ (below diagonal) between the largest EV and the CPV for the situation shown in Fig. 1.

	$l=20$	$l=19$	$l=18$	$l=17$
$n=20$	0	0.69 ± 0.18	1.25 ± 0.18	1.30 ± 0.16
$n=19$	1.65 ± 0.14	0	1.38 ± 0.17	1.38 ± 0.16
$n=18$	1.68 ± 0.13	1.61 ± 0.13	0	1.26 ± 0.22
$n=17$	1.78 ± 0.12	1.68 ± 0.15	1.66 ± 0.15	0

the procedure is repeated. Otherwise, in total N trials are tested independently, leading to a break of the process if no points $\Delta\geq W$ are found anymore. This procedure is computationally fast and known to get stuck in local maxima of multidimensional landscapes only rarely due to the fact that random moves into "wrong" (downward) directions are allowed. In our implementation, we use $\delta W=0.001$ and $N=1000$.

An example for the usefulness of the construction of the CPV $\{\mathbf{w}_l\}$ is illustrated in Fig. 1(b) for the same situation as in (a): $K'=4$ eigenvectors were taken into account. Each of the CPV $\mathbf{w}_{20}\ldots\mathbf{w}_{18}$ corresponds to exactly one of the $K=3$ clusters and different to the eigenvectors themselves a clear distinction between clusters and uninvolved channels is possible. As opposed to that the structure of $\mathbf{w}_{17}$ cannot be distinguished from $\mathbf{v}_{17}$. This fact can be used to deduce the number K of clusters that are actually formed in the data. Table I shows that the distance (6) has increased on average about 39% by the rotation (from 7.26 for the $\{\mathbf{v}_l\}$ to 10.06 for the $\{\mathbf{w}_l\}$). Note that for $N_{\mathrm{ens}}=100$ calculations of the mutual distance (5) between two independently sampled normalized random vectors in $M=20$ dimensions with Gaussian distributed components we obtain $D_{ln}=1.11\pm0.08$. Consequently, the distance $D(\mathbf{v}_{20},\mathbf{v}_{19})$ between the two largest EV lies significantly below and all distances $D(\mathbf{w}_l,\mathbf{w}_n)$ between the CPV significantly above this value, see entries below the diagonal in Table I. In contrast most of the remaining distances $D(\mathbf{v}_{l<20},\mathbf{v}_{n<19})$ have a considerable overlap with the result D_{ln} for random vectors, see entries above the diagonal in Table I.

The squared components b_{il}^2 of the CPV give information about the involvement of the channels "i" in the cluster "l." In order to demonstrate this we show in Fig. 2 the largest EV and CPV of a single equal-time correlation matrix that was constructed from $M=32$ time series which were sampled from Gaussian noise. In order to have a well defined cluster structure the first $m_1=13$ and the next $m_2=9$ time series were coupled via an admixture of a common noise component; see Sec. V A for details of the test system. Thereby the last 7 and 5 channels are coupled to the cluster with only 40 and 60% of the common noise component, respectively. Two observations can be made: First, an interpretation that bases on the size of the squared eigenvector components a_{il}^2 could easily be lead to the wrong conclusion that four clusters were present. This is not the case for the squared components b_{il}^2 of the CPV. Here the channels that contribute to the two clusters are separated from the non-contributing ones by orders of magnitude and can be distinguished clearly. Second,

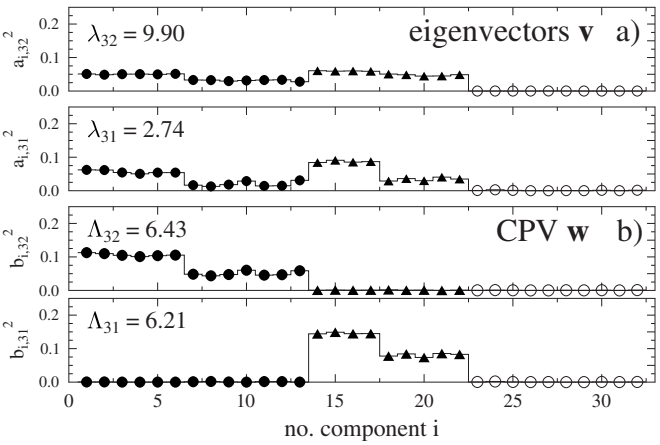

FIG. 2. Comparison of the squared components of the $K'=2$ largest EV $\mathbf{v}_{M+1-k}$ of the C matrix with those of their linear combinations $\mathbf{w}_{M+1-k}$, $k=1,2$. The channels that contribute to the $K=2$ clusters are marked with full symbols (●, ▲, ■) and the uncorrelated ones with open circles (○).

the relative involvement of the channels in the clusters as given by the size of the common noise component (see Sec. V A) is reflected correctly by the b_{il}^2.

B. Cluster participation coefficients

Using the notation $\langle\cdot|\cdot\rangle$ for the scalar product the relation

$$\sum_{k=M-K'+1}^{M}|\langle\mathbf{v}_k|\mathbf{w}_l\rangle|^2=\sum_{l=M-K'+1}^{M}|\langle\mathbf{v}_k|\mathbf{w}_l\rangle|^2=1 \tag{7}$$

is satisfied by construction for the K' largest EV and the corresponding CPV. In addition to the CPV we introduce *cluster participation coefficients* (CPC) as defined by

$$\Lambda_l=\sum_k|\langle\mathbf{v}_k|\mathbf{w}_l\rangle|^2\lambda_k. \tag{8}$$

The CPC Λ_l reduce to the eigenvalues λ_l in the limit of vanishing inter-cluster relations (where $\mathbf{w}_l\equiv\mathbf{v}_l$) and share the following properties: $\lambda_{M+1-K'}\leq\Lambda_l\leq\lambda_M$ and $\Sigma_l\Lambda_l=\Sigma_l\lambda_l=M$. In the following the CPC will be used to assign an ascending order to the CPV.

Figure 3 illustrates the dependence of the eigenvalues λ_k and CPC Λ_k on the relative strength $d^{(0)}/c^{(0)}$ of the inter-cluster relations for the same model as in Fig. 1. Shown is a band of one standard error deviation from the average. The most important observation is that the CPC stay almost constant with weakly increasing fluctuations up to $d^{(0)}/c^{(0)}\approx0.7$ and can be nicely described by zero order perturbation theory (A1) to Eq. (4): $\Lambda_k\approx\lambda_k^{(0)}$. In contrast the repulsion of the two largest eigenvalues (the three clusters merge into a common one) for increasing intercluster relation $d^{(0)}$ can be described well by second order perturbation theory (A5) for $d^{(0)}/c^{(0)}\lesssim0.3$.

Figures 1 and 3 together make it possible to understand, why the PI algorithm of [13,14] fails in situations with strong inter-cluster relations: If the largest eigenvalue λ_M is increased strongly as shown in Fig. 3 an estimate of the cluster

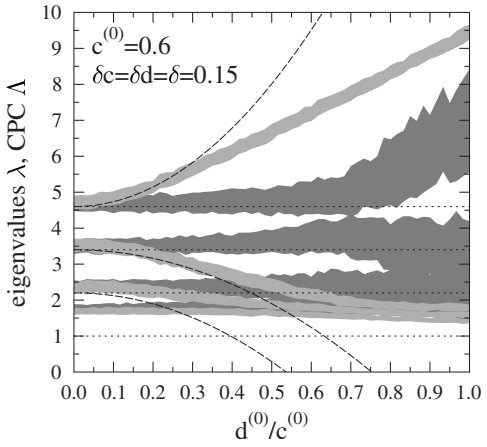

FIG. 3. Dependence of the four largest eigenvalues λ_k (light gray) and CPC Λ_k (dark gray) on the relative inter-cluster relation. $N_{\mathrm{ens}}=100$ independent realizations of the same block-matrix model as in Fig. 1 with $K=3$ clusters were sampled using $K'=4$. Dotted and dashed lines indicate the approximation to the eigenvalues in zero and second order perturbation theory, respectively; see the Appendix.

"M" 's "strength" by λ_M leads to an overestimation whereas at the same time the "strength" of the clusters "$M-1$" to "$M-K+1$" is underestimated. In addition the distribution of the components a_{iM}^2 becomes flat and consequently many channels "i" are erroneously attributed to the cluster "M"; see Fig. 1(a).

C. Definition of K'

Still it has not been clarified how the number K' of largest EV used for the maximization described in Sec. III A has to be chosen in situations where the number K of clusters is unknown beforehand. In the optimal situation K' can be deduced from the number of "large" eigenvalues that are clearly separated from the "bulk" via a gap in the eigenvalue spectrum. In the presence of purely random correlations of size δ the "bulk" is centered at $\lambda^{(0)}=1$ and has a width $\sim\delta$. From Eq. (A1) it becomes obvious that the condition for a clear separation between "large" eigenvalues and "bulk" eigenvalues is $\min_k(m_k-1)c_k \gg \delta$ if inter-cluster relations can be neglected.

In Fig. 3 this condition is not fulfilled, mainly due to the large width of random fluctuations and the small number of $m_3=3$ participants in cluster "3." In such situations comparison with the spectrum of reference data without strong interrelations can be made use of. For time series that can be decomposed into (approximately) stationary segments it is conceivable to use univariate surrogate time series; see e.g. [36], that are independently produced for all M data channels. Such a procedure by definition destroys all linear and nonlinear relations between the channels and K' can be fixed by the number of eigenvalues that are larger than the largest eigenvalue of the interrelation matrix of the surrogates: $\lambda_l > \lambda_M^{\mathrm{surr}}$. For noisy time series this is very similar to using the upper boundary λ_+ of the level density of Wishart ensembles [17,37] for the same purpose. Under the condition $T,M \gg 1$ but $T/M=\mathrm{const}>1$ the eigenvalue density of the correlation matrix of time series that are sampled independently from Gaussian white noise is nonzero only in the range $\lambda_- \leq \lambda \leq \lambda_+$ with $\lambda_\pm=(1\pm\sqrt{M/T})^2$ [38]. In [10] the condition $\lambda_l > \lambda_+$ is used for selection of the largest EV for direct cluster identification.

For time-dependent applications of the algorithm to nonstationary data also an appropriately chosen reference interval where only weak interrelations are suspected can be used to define the number K' of "large" eigenvalues dynamically. As an example it can be defined as the number of eigenvalues that climb over a threshold which is not trespassed in, say, 90 or 95% of the reference interval.

For the correlation matrix it has been found recently that spectral fluctuations as displayed by the distribution $P(s_l)$ of *individually unfolded* nearest neighbor spacings $s_l=\lambda_{l+1}-\lambda_l$ indicate even more sensitively in which part of the eigenvalue spectrum genuine correlations are induced via nonrandom level repulsions than the level density itself [32]. Investigation of possible exploitation of these effects for our purposes is underway.

Independently of how K' is fixed, a nonoptimal choice of $K'>K$ does not affect the results dramatically (however increases the computational effort as the dimension of the space of the GEA α_κ increases). Rather the algorithm focuses automatically on the K largest EV that correspond to the K clusters as illustrated in Fig. 3: Different to $\Lambda_{20}\ldots\Lambda_{18}$ and within their respective error bars Λ_{17} and λ_{17} behave very much alike. Similarly in Fig. 1 the structure of the CPV $\mathbf{w}_{17}$ and the corresponding eigenvector $\mathbf{v}_{17}$ cannot be distinguished whereas $\mathbf{w}_{20}\ldots\mathbf{w}_{18}$ nicely indicate the channels that contribute to the clusters.

D. Automatic attribution of channels to clusters

As we are aiming at a time resolved and unsupervised application of the algorithm to interrelation matrices constructed from possibly non-stationary time series the channels have to be attributed to clusters automatically, once the CPV are found. For separation of large from small components of $\mathbf{w}_l$ it is suitable to first arrange the b_{il}^2 in increasing order by rank ordering. For the components of the CPV $\mathbf{w}_{32}$ of Fig. 2 this is shown in Fig. 4(a). Two measures are candidates for a good separation algorithm: The largest ratio y and the largest difference x between neighboring rank ordered b_{il}^2. Either option alone fails in certain situations as can be seen from panels (b) and (c) of Fig. 4. Using the difference criterion x it is not clear whether the cluster should close after the 6th or the 13th largest b_{il}^2 [cf. also 4(a)]. The criterion based on the ratio y indicates the 13th largest value correctly. However, an even larger ratio appears between the two smallest b_{il}^2, again leading to an ambiguity. Effects of these kind become even more pronounced as the noise contamination of the matrix elements C_{ij} increases, a situation that has to be taken into consideration in real world applications.

However, the product $x\cdot y$ of both measures develops a pronounced and unique peak at the correct position; see Fig. 4(d). Consequently it allows us to deduce a robust threshold

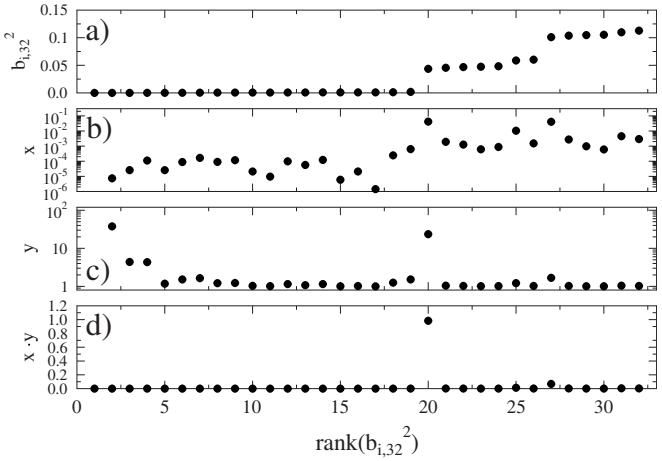

FIG. 4. Different candidates of measures that are suitable for automatic attribution of channels to clusters. (a) Squared components b_{iM}^2 of Fig. 2 put into increasing order. (b) Difference x between adjacent values. (c) Ratio y between adjacent values. (d) Product $x \cdot y$ of difference and ratio between adjacent values. Note the logarithmic scale in (b) and (c).

which is defined in a self-contained manner and can therefore be used for automated attribution of channels to clusters. In rare cases this recipe assigns a channel "i" to several clusters "l." To establish uniqueness in such situations the cluster with the largest b_{il}^2 is selected.

IV. ESTIMATES FOR THE INTERRELATION STRENGTH

A. Intra-cluster relation

Under the assumption that the studied interrelation matrix $\mathbf{C}$ has a clear cluster structure in the sense that intra-cluster relations are stronger than inter-cluster relations and noise contamination is weak, the situation can be modeled by block matrices. In the Appendix such a simplified model with constant intra- and inter-block elements is treated in perturbation theory. In cases where the studied matrix $\mathbf{C}$ can indeed be approximated by a block matrix these results can be used to estimate the intra-cluster and inter-cluster relations.

The most direct estimate for the intra-cluster relations can be deduced from the average absolute off-diagonal matrix element $\langle |C_{i_k<j_k}| \rangle$ inside the block $\mathbf{C}_{kk}$. Here the indices i_k and j_k are restricted to the channels that (according to the used algorithm) actually contribute to cluster "k." Exploiting the observation $\Lambda_k \approx \lambda_k^{(0)}$ shown in Fig. 3 and Eq. (A1) the total intra-cluster relation C_k in cluster "k" can in addition be estimated from the CPC via

$$C_k \approx \frac{\Lambda_k - 1}{m_k - 1}. \tag{9}$$

In order to check the reliability of both estimates we again employed a variant of the three-cluster block-model of the type (4) with a finite width $\delta c = \delta d = \delta = 0.1$ of the distribution of the matrix elements. In Fig. 5 we show an example with $M=32$ channels and cluster sizes $m_1=8$, $m_2=11$, $m_3=5$ and average intra-cluster relation $c_1^{(0)}=0.7$, $c_2^{(0)}=0.4$, and $c_3^{(0)}=0.8$. All inter-cluster relations were set to the same average value $d_{12}^{(0)}=-d_{13}^{(0)}=-d_{23}^{(0)}=d^{(0)}$ that was restricted to the range $0 \le d^{(0)} \le \min c_k^{(0)} = c_2^{(0)}$ in order to assure a clearly defined cluster structure. $N_{\mathrm{ens}}=100$ independent realizations of the matrix were sampled and $K'=3$ was kept fixed.

Within the given statistics both estimates agree with the values $c_k^{(0)}$. It is remarkable that for not too large $d^{(0)}/c_2^{(0)} \lesssim 0.6$ the errors of the estimate (9) are considerably smaller than those of $\langle |C_{i_k<j_k}| \rangle$ which are of the order of δc. This can be understood by calling the fact that the CPC depend on the largest eigenvalues and eigenvectors only. For moderate inter-cluster relations the latter are known to be less contaminated by purely random interrelations than the elements of the matrix $\mathbf{C}$; see e.g. [11] for the example of the correlation matrix. It remains to make a remark on the origin of the mavericks in (d) and (f). For increasing inter-cluster relations the errors of the estimate for the CPC grows as can be seen from Fig. 3. Therefore it may happen occasionally that the algorithm interchanges the cluster-labels "2" and "3." Indeed

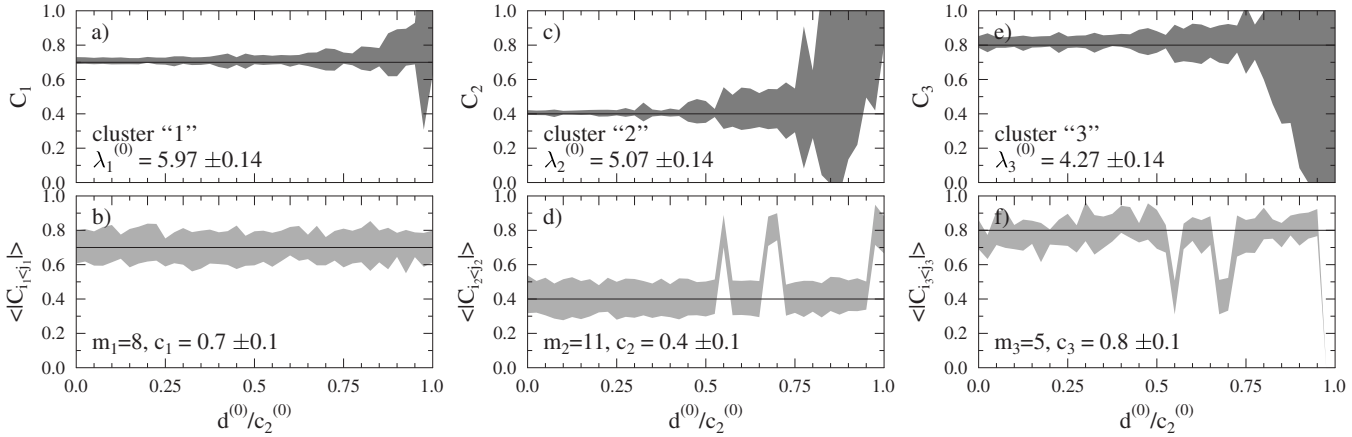

FIG. 5. Comparison of different estimates for the intra-cluster relations of the three clusters of a model system (see text). Shown is a band of two standard deviations width around the average over $N_{\mathrm{ens}}=100$ independent realizations. Top: C_k of Eq. (9); bottom: average absolute matrix element $\langle |C_{i_k<j_k}| \rangle$ between the clustered channels.

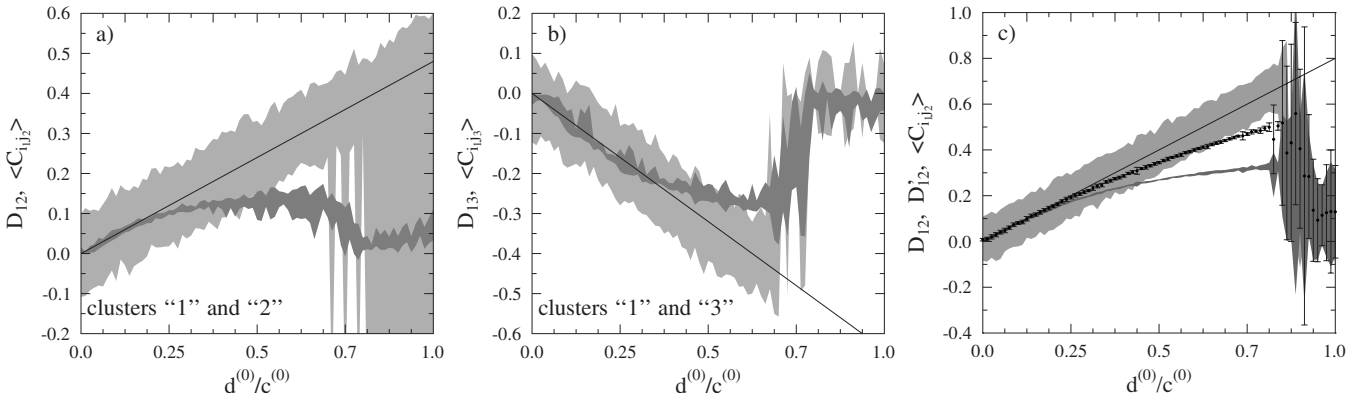

FIG. 6. (a) and (b) Comparison of the estimates D_{kl} of Eq. (10) (dark gray) and $\langle C_{i_k j_l}\rangle$ (light gray) for the inter-cluster relations between the three clusters of the model (see text). As fully drawn lines also the center of the distribution of the matrix elements $d_{kl}^{(0)}$ is shown. The figure for D_{23} (not shown) looks qualitatively like the one for D_{13}. (c) Comparison of the estimates for the inter-cluster relations of a two-cluster system. The estimates given by D_{kl} of Eq. (10) and the average inter-cluster matrix element are again shown as dark and light gray bands, respectively. In contrast the estimate (11) is shown with its error bars.

the switching of the average matrix elements $\langle |C_{i_k<j_k}| \rangle$ takes place between the values 0.4 and 0.8.

B. Inter-cluster relation

Under the condition that the CPC are nondegenerate and the inter-cluster relations moderate their strength can be estimated from the overlap of the eigenvectors and the CPV. To this end we remind the observation $\Lambda_k \approx \lambda_k^{(0)}$ of Fig. 3 and use Eq. (A4) as well as $\mathbf{v}_k^{(0)} = \mathbf{w}_k$ to obtain the estimate

$$D_{kl} \approx \langle \mathbf{w}_l | \mathbf{v}_k \rangle \frac{\Lambda_k - \Lambda_l}{\sqrt{m_k m_l}} \quad \text{for} \quad k \neq l \tag{10}$$

for the interrelation between clusters "k" and "l." Again we used the three-cluster block-model introduced above to explore the reliability of this estimate. Now clusters of size $m_1=7$, $m_2=5$, and $m_3=3$ are formed within a total of $M=20$ channels. The parameters are set to the following values: $c_k^{(0)}=c^{(0)}=0.8$, $d_{12}^{(0)}=0.6d^{(0)}$, $d_{13}^{(0)}=-0.8d^{(0)}$, $d_{23}^{(0)}=-d^{(0)}$, and $d^{(0)}$ is varied in the range $[0, c^{(0)}]$. The width of the distribution of the matrix elements is again $\delta c=\delta d=\delta=0.1$.

In Figs. 6(a) and 6(b) the results are shown for $N_{\text{ens}}=100$ independent realizations of the matrix $\mathbf{C}$ and $K'=3$. For relative inter-cluster relations $d^{(0)}/c^{(0)} \lesssim 0.7$ the average of the inter-cluster matrix elements $\langle C_{i_k j_l}\rangle$ reproduces $d_{kl}^{(0)}$ very well, though with large error bars of size δd. Beyond this value the clusters can no longer be distinguished by the algorithm. As compared to the average matrix element the estimate D_{kl} of Eq. (10) supplies a much better approximation to $d_{kl}^{(0)}$ for small $d^{(0)}/c^{(0)} \lesssim 0.3$. However, for larger $d^{(0)}/c^{(0)}$ the absolute value of $|d_{kl}^{(0)}|$ is systematically underestimated.

For $K'=2$, a case that, e.g., appears frequently in applications to electroencephalograms, the absolute value of the inter-cluster relation d_{12} can in addition be estimated from the eigenvalue and the CPC alone via Eq. (A5):

$$|D_{12}''| \approx \sqrt{\frac{|\lambda_M - \Lambda_M|(\Lambda_M - \Lambda_{M-1})}{m_1 m_2}} \approx \sqrt{\frac{|\lambda_{M-1} - \Lambda_{M-1}|(\Lambda_M - \Lambda_{M-1})}{m_1 m_2}}. \tag{11}$$

The sign of the interrelation is not accessible. To compare the quality of this estimate to the earlier one we employed a two-cluster block-model of type (4) with $M=32$ channels. Here we show the case where the cluster parameters are chosen as $m_1=13$, $m_2=9$, $c_k^{(0)}=c^{(0)}=0.8$, and $\delta c=\delta d=\delta=0.1$. Figure 6(c) shows that for $N_{\text{ens}}=100$ repetitions the result of Eq. (11) has very small error bars and delivers reliable estimates up to larger values $d^{(0)}/c^{(0)}$ than (10). Unfortunately, for $K'>2$ a similar estimate based on the CPC alone is no longer possible: The arising system of K' linear equations for $K'(K'-1)/2$ coupling parameters d_{kl}^2 is underdetermined (for $K'=3$ the resulting matrix is singular).

Reminding $\langle \mathbf{v}_k^{(0)} | \delta \mathbf{v}_k^{(2)} \rangle \approx \langle \mathbf{w}_k | \mathbf{v}_k \rangle - 1$ also formula (A8) can be used to determine the absolute value of d_{12} uniquely if $K'=2$:

$$|D_{12}''| \approx \sqrt{2(1 - |\langle \mathbf{w}_M | \mathbf{v}_M \rangle|)\frac{(\Lambda_M - \Lambda_{M-1})^2}{m_1 m_2}} \approx \sqrt{2(1 - |\langle \mathbf{w}_{M-1} | \mathbf{v}_{M-1} \rangle|)\frac{(\Lambda_M - \Lambda_{M-1})^2}{m_1 m_2}}. \tag{12}$$

For larger K' the d_{kl}^2 also remain underdetermined here. We have investigated the utility of this formula and found that it leads to estimates very similar to those of Eq. (10). Comparing Eq. (A4) with Eq. (A8) in the special case $K'=2$ we find $(\langle \mathbf{v}_M^{(0)} | \delta \mathbf{v}_{M-1}^{(1)} \rangle)^2 = -2\langle \mathbf{v}_M^{(0)} | \delta \mathbf{v}_M^{(2)} \rangle$ and a similar relation for $\mathbf{v}_{M-1}^{(0)}$. This close relation explains why the estimates (10) and (12) lead to very similar results.

V. STATISTICAL EVALUATION OF THE PERFORMANCE FOR CORRELATION CLUSTERS

The simple block model (4) exploited up to now can be used as an idealization of interrelation matrices **C** that are constructed from any measure that satisfies the conditions (1) to (3). As the CPV algorithm itself is independent of the details of the underlying interrelation measure we specify to one of the simplest and computationally easiest ones. In order to evaluate the applicability of the CPV algorithm for the analysis of mutual relations of time series we use the linear equal-time cross-correlation matrix [17,18] from now on:

$$\mathbf{C} = \frac{1}{T}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^t. \tag{13}$$

Here the $M \times T$ data matrix $\tilde{X}_{it} = \tilde{X}_i(t)$ contains the M signal channels measured on T time steps after normalization to zero mean and unit variance as defined over T data points:

$$\tilde{X}_i(t) = \frac{X_i(t) - \langle X_i \rangle}{\sigma_i}. \tag{14}$$

A. Time series based test framework for cluster detection algorithms

In order to be able to properly evaluate and compare the performance of the CPV algorithm for detection of clusters in multivariate time series we set up a time series based test framework. Note that a more direct sampling of C-matrices as done up to now bears the danger of producing "unphysical" matrices. For instance the unit matrix in M dimensions (corresponding to completely uncorrelated channels) is contained in the model (4) but no set of corresponding time series can be found for any finite T. More severely, the positive definiteness of the correlation matrix (13) can easily be violated by sampling C-matrices instead of time series.

For such reasons we prefer to use an ensemble of size N_{ens} of multivariate noise time series. The degree of correlations between subgroups of the M time series of length T is controlled via common noise components of adjustable amplitude; cf. the model of [39]. To model a system that contains K clusters the following prescription is used [16]:

$$X_{it} = \left(1 - \sum_k \rho_{ik} - \sum_{(kk')} \sigma_{i(kk')}\right)\eta_{it} + \sum_k \rho_{ik}\xi_{kt} + \sum_{(kk')} \sigma_{i(kk')}\zeta_{(kk')t}. \tag{15}$$

The strength of the "intra-cluster correlations" is controlled via the parameters ρ_{ik} which have a nonzero value only if channel "i" belongs to cluster "k." Similarly the parameters $\sigma_{i(kk')}$ control the strength of the "inter-cluster correlations" and are finite only if channel "i" belongs to one of the clusters k or $k' \neq k$. Finally, η_{it}, ξ_{kt}, and $\zeta_{(kk')t}$ denote the individual and common noise components that are drawn independently from a Gaussian distribution with zero mean and unit variance on every time step. In order to have a well defined cluster structure in the sense that intra-cluster correlations must be stronger than inter-cluster correlations the conditions $\sigma_{i(kk')} \leq \max(\rho_{ik}, \rho_{ik'})$ must be satisfied. Moreover, in order to be able to switch between completely correlated and completely uncorrelated channels the value of the bracket in Eq. (15) must fall inside the interval $[0,1]$ for all channels "i."

Using Eq. (15) and the fact that η_{it}, ξ_{kt}, and $\zeta_{(kk')t}$ are mutually uncorrelated the average covariance of time series X_{it} and X_{jt} can easily be derived as

$$\langle X_{it}X_{jt}\rangle = \delta_{ij}\left(1 - \sum_k \rho_{ik} - \sum_{(kk')} \sigma_{i(kk')}\right)^2 + \sum_k \rho_{ik}\rho_{jk} + \sum_{(kk')} \sigma_{i(kk')}\sigma_{j(kk')}. \tag{16}$$

From this the average elements of the C-matrix can be calculated via $C_{ij} = \langle X_{it}X_{jt}\rangle / \sqrt{\langle X_{it}^2\rangle\langle X_{jt}^2\rangle}$.

For a general situation of K clusters the *attribution errors* may be sorted into the following four categories: (Ik) channels that contribute to a given cluster "k" are not detected (K subcategories), (IIkk') channels that belong to cluster "k" are attributed to cluster "$k' \neq k$" [$K(K-1)$ subcategories], (IIIk) uncorrelated channels are attributed to the cluster "k" (K subcategories), and (IV) false clusters ("pseudoclusters" [14]) are found within the uncorrelated channels. The attribution errors of either category can be measured by the average ratio of misattributions over the N_{ens} realizations of Eq. (15) with the same parameters.

B. Evaluation

For a systematic evaluation of the different categories of attribution errors we concentrate on the generic problem of distinction of $K=2$ clusters with finite inter-cluster correlations. The case $K>2$ has also been studied, however, the huge parameter space makes a comprehensive but lean presentation of the results impossible. An example of the performance of the algorithm for $K=4$ is given in [16] where also comparison with PI and the standard k-means algorithm [1,2] is made. To this end the average of the attribution errors of categories I to IV is shown there, rather than a separate discussion of the categories.

In Fig. 7 we show the attribution errors for $N_{\text{ens}}=100$ independent realizations of Eq. (15) for a two-cluster situation with $T=1024$, $M=20$, $m_1=10$, $m_2=5$, and $\rho_1=\rho_2=\rho$. With the exception of (**f**) (where $K'=3$ is kept fixed) the value of K' is defined automatically by the number of eigenvalues that become larger than the upper boundary λ_+ of the eigenvalue density of the Wishart ensemble; cf. Sec. III C. It has been tested before that for only randomly correlated time series all eigenvalues λ_l nicely fall into the interval $[\lambda_-, \lambda_+]$.

Comparison of Figs. 7(a) and 7(b) reveals a small asymmetry of the channels that are not correctly attributed to the clusters "1" and "2." Due to the different cluster size $m_1 = 2m_2$ the region where errors are committed is slightly larger for the smaller cluster. In general errors of category I are committed if the total correlation is too small (i.e., $\rho, \sigma \ll 1$). In this situation the cluster structure is simply not pronounced enough to be detected by the CPV algorithm. In

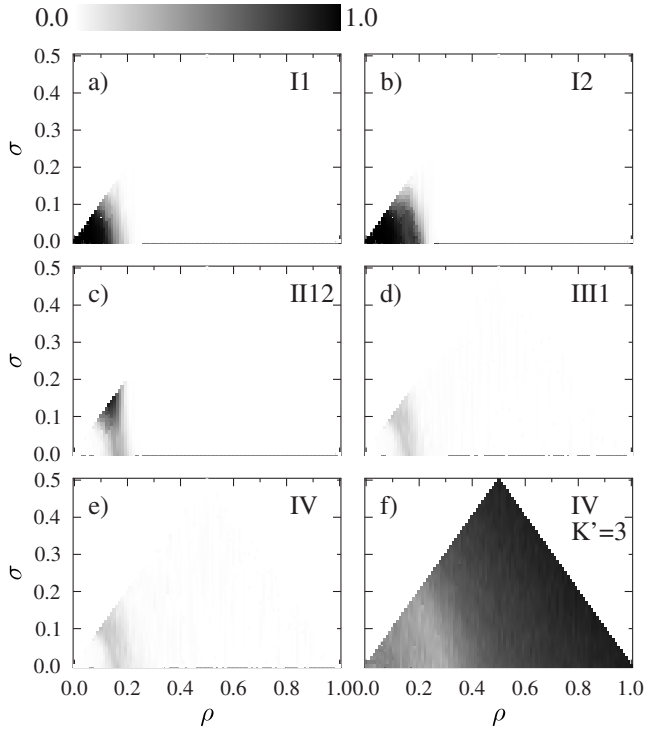

FIG. 7. Statistical evaluation of the categories I to IV of the attribution errors for a two-cluster situation with equal intra-cluster correlations $\rho_1=\rho_2=\rho$ and inter-cluster correlation σ. The triangle of the allowed region of the parameters ρ and σ is visible in (f).

[16] it is shown that other algorithms have similar performance difficulties in this region. Figure 7(c) shows the errors stemming from channels that belong to cluster "1" but are erroneously attributed to cluster "2." Examination of the structure of the corresponding CPV $\mathbf{w}_{20}$ and $\mathbf{w}_{19}$ reveals that in these cases the algorithm is unable to distinguish between the two clusters and finds a common one. As expected this kind of problems occurs for relatively small intra-cluster couplings and inter-cluster couplings of comparable amplitude. Note that the CPV algorithm does not encounter problems, when dealing with strong inter-cluster correlations in general. Figure 7(d) shows that in the same region occasionally unclustered channels are erroneously attributed to cluster "1." The errors of categories II21 and III2 are not shown in the figure as they are almost identical to those of II12 and III1. Finally, Figs. 7(e) and 7(f) show to what extent "pseudoclusters" are found by the CPV algorithm. If K' is defined automatically by the number of eigenvalues $\lambda_k > \lambda_+$ (e) errors of category IV are committed in the same region as those of II and III. However, a huge amount of "pseudoclusters" are found within the unclustered channels if $K'=3$ is kept fixed. Comparison with Fig. 1(b) shows that this shortcoming is not caused by a failure of the maximization of the function (6). The structure of the corresponding low lying CPV is very similar to the eigenvectors. Rather, what causes "pseudoclusters" is the fact that the automatic attribution algorithm is forced to attribute channels to a cluster "3" that simply does not exist.

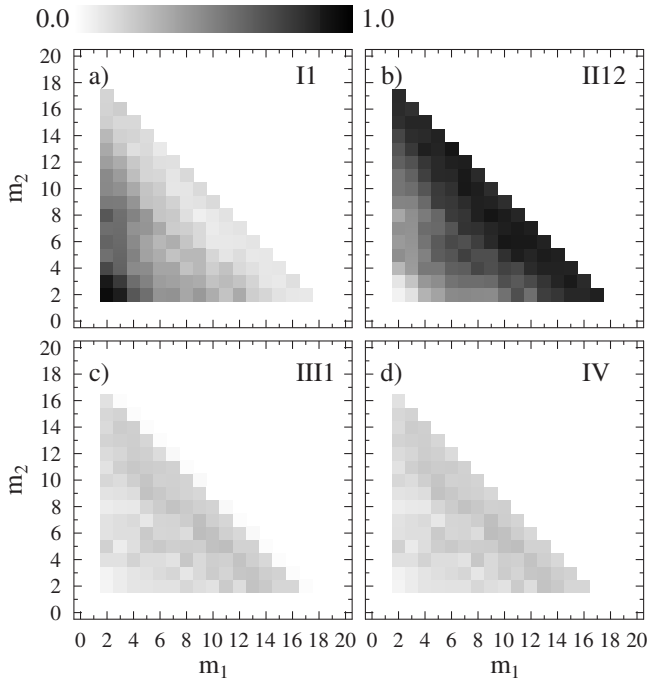

FIG. 8. Statistical evaluation of the categories I to IV of the attribution errors for a two-cluster situation with equal intra-cluster correlations $\rho_1=\rho_2=0.15$ and inter-cluster correlation $\sigma=0.14$ as a function of the cluster sizes.

C. Problems of the CPV algorithm

Despite the overall satisfactory performance of the algorithm we focus on the problematic situations in the following in order to investigate in more detail under which conditions the algorithm tends to fail. In Fig. 8 we show the dependence of the attribution errors on the two cluster sizes m_1 and m_2 for $N_{\mathrm{ens}}=100$ independent realizations of Eq. (15). The parameters were set to $T=1024$, $M=20$, $\rho_1=\rho_2=0.15$, and $\sigma=0.14$. As the inter-cluster correlations are almost as strong as the (weak) intra-cluster correlations these parameters are very unfavorable for any cluster detection algorithm; cf. Fig. 7(c) for the CPV. Again K' is defined automatically via $\lambda_k > \lambda_+$. Attribution errors of category I1 are mainly committed for small cluster sizes m_1 and m_2. In contrast for $\sigma \lesssim \rho \approx 0.15$ error category II12 becomes worst for $m_1+m_2 \approx M$. Under these unfavorable conditions the algorithm again fails to separate the clusters. Compared to categories I1 and II12 again III1 and IV turn out of minor importance. We have checked that these kinds of attribution errors in addition diminish rapidly as the shaded areas of Figs. 7(d) and 7(e) are left. Therefore, we concentrate on categories I1 and II12 in the following.

In Fig. 9 the attribution errors of category I1 (channels that belong to cluster "1" are not found) and II12 (channels that belong to cluster "1" are erroneously attributed to cluster "2") are shown for $M=20$, $T=1024$, $N_{\mathrm{ens}}=100$ and $\rho_1=\rho_2=0.3$, $\sigma=0.2$. Under these more favorable conditions attribution errors of categories I1 and II12 are committed only if one of the two clusters is small $1 < m_1, m_2 \lesssim 3$. Errors of type III and IV are no longer committed at all and therefore not shown.

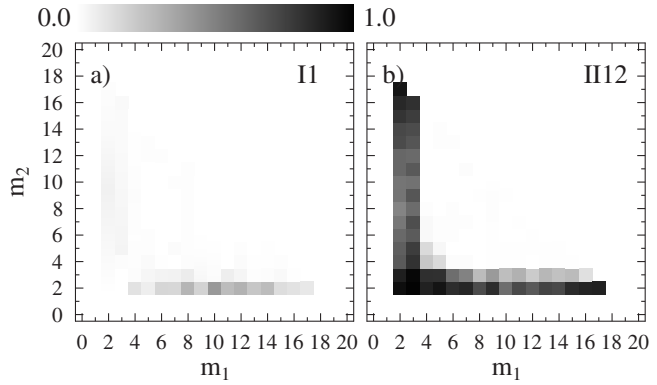

FIG. 9. Statistical evaluation of the categories I1 and II1 of the attribution errors for a two-cluster situation with equal intra-cluster correlations $\rho_1=\rho_2=0.3$ and inter-cluster correlation $\sigma=0.2$ as a function of the cluster sizes. Shown are average attribution errors. Note that the apparent asymmetry in (a) disappears if the error bars of this quantity are taken into account.

To investigate the behavior of the error category II if one of the clusters is very small, we finally show the result of $N_{\mathrm{ens}}=100$ realizations with $M=40$, $\rho_1=\rho_2=\rho$, and $T=1024$ in Fig. 10. If one of the clusters is large ($m_2=17$) the small one ($m_1=2$) cannot be distinguished and the corresponding channels are erroneously attributed to the larger one if $\sigma\approx\rho$; see (a) and (b). Misattributions in the opposite direction are extremely rare. When the smaller cluster grows, e.g., $m_1=5$, misattributions are committed in both directions but occur in the region $\sigma\approx\rho\approx0.2$ only.

VI. APPLICATION TO TIME-DEPENDENT MODEL DATA

The examples shown in Sec. V consist of averages over an ensemble of size N_{ens} that has been produced under stationary conditions. Such a situation cannot always be realized in real world applications where time series are often intrinsically nonstationary. For instance in [14,16] first applications of cluster detection algorithms to electroencephalographic recordings containing epileptic seizures are made. If cluster detection algorithms are to be applied to such data using a running analysis window of length T reliable results must be delivered on every time step. We show in the following by studying nonstationary model time series that the CPV algorithm is capable of dealing with these conditions, too.

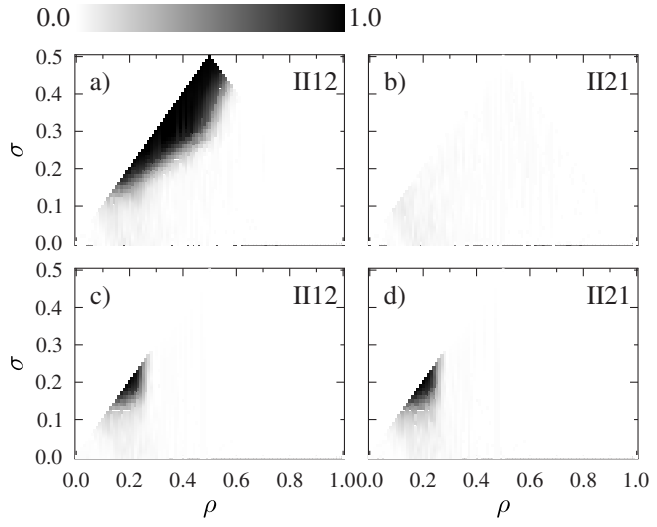

FIG. 10. Statistical evaluation of the category II of the attribution errors for a two-cluster situation with equal intra-cluster correlations $\rho_1=\rho_2=\rho$ and one small cluster. Top: $m_1=2$, $m_2=17$; bottom: $m_1=5$, $m_2=17$.

TABLE II. Time dependence of the intra-cluster couplings of the test model.

Segment	A Channels 8–11	B Channels 12–17	C Channels 18–20
$0\le t<5$	$\rho_A=0$	$\rho_B=0$	$\rho_C=0$
$5\le t<10$	$\rho_A=0$	linear	linear
$10\le t<15$	$\rho_A=0$	$\rho_B=0.3$	$\rho_C=0.15$
$15\le t<20$	linear	linear	linear
$20\le t<25$	$\rho_A=0.4$	$\rho_B=0.4$	$\rho_C=0$
$25\le t<30$	linear	linear	$\rho_C=0$
$30\le t\le35$	$\rho_A=0$	$\rho_B=0$	$\rho_C=0$

For test purposes we used $M=20$ time series produced from the model (15) with time dependent cluster structure. A time unit was arbitrarily set to 256 data points and time series of 35 units length were sampled with changing cluster structure. The parameters were chosen such that the advantage of the CPV algorithm over the competing methods of PI [13,14] and the components of large eigenvectors alone [10,11] becomes most pronounced; cf. [16]. In the first and last five time units no correlation clusters are present in the test system: $\rho_k=\sigma_{kk'}=0$ for $k,k'=A,B,C$. In the segment $t=10-15$ two clusters are formed by the channels 12–17 and 18–20. The amplitudes of the intra- and inter-cluster correlations ρ_k and $\sigma_{kk'}$ are given in Table II and Table III, respectively. Later on in $t=20-25$ the cluster "C" has been replaced by cluster "A" which is formed by the channels 8–11 and the intra-cluster correlations of cluster "B" have increased slightly. In the remaining segments all parameters ρ_k and $\sigma_{kk'}$ of Eq. (15) are shifted linearly between the values of the adjacent segments.

Figure 11 shows the cluster structure revealed by the CPV algorithm for different window lengths T. The number K' is

TABLE III. Time dependence of the inter-cluster couplings of the test model.

Segment	$A-B$	$A-C$	$B-C$
$0\le t<5$	$\sigma_{AB}=0$	$\sigma_{AC}=0$	$\sigma_{BC}=0$
$5\le t<10$	linear	linear	linear
$10\le t<15$	$\sigma_{AB}=0.1$	$\sigma_{AC}=0.1$	$\sigma_{BC}=0.25$
$15\le t<20$	linear	linear	linear
$20\le t<25$	$\sigma_{AB}=0.38$	$\sigma_{AC}=0.15$	$\sigma_{BC}=0.15$
$25\le t<30$	linear	linear	linear
$30\le t\le35$	$\sigma_{AB}=0$	$\sigma_{AC}=0$	$\sigma_{BC}=0$

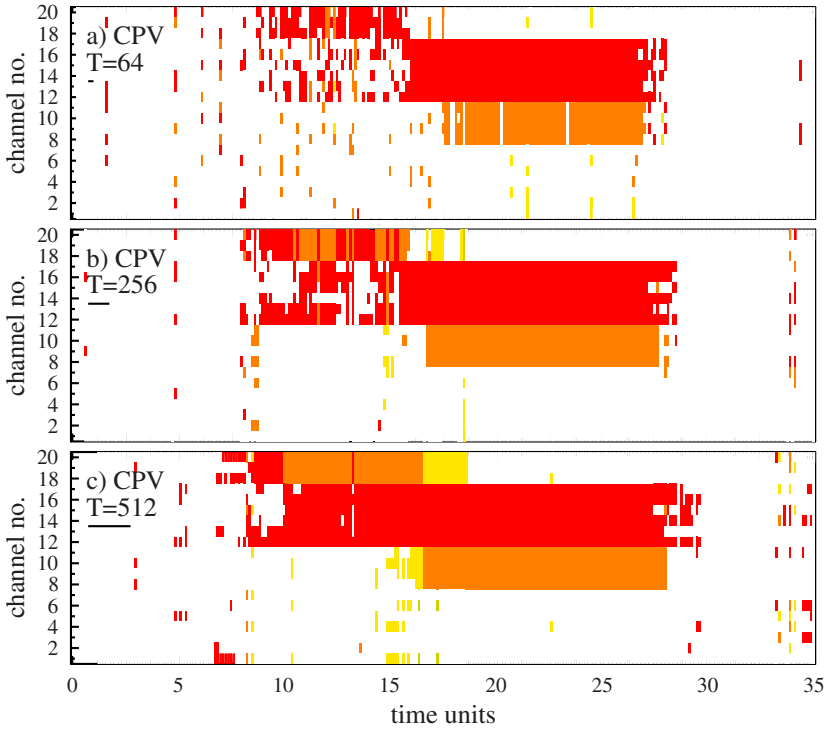

FIG. 11. (Color online) Cluster structure of the same model system analyzed with the CPV algorithm and three different window lengths T. The bar below the specification of the window length T shows which part of the time series is used at a time. All channels marked by the same color (gray value) belong to the same cluster ($\mathbf{w}_{20}$: dark gray; $\mathbf{w}_{19}$: gray; $\mathbf{w}_{18}$: light gray).

defined dynamically as explained in Sec. III C: $0 \leq t \leq 5$ is chosen as a reference interval for which the upper boundary of the 90% significance interval of the eigenvalues λ_l is calculated. Channels that contribute to different clusters are marked by different colors (gray values). In principle the involvement of a channel in a cluster as given by b_{ik}^2 could be coded by the color depth. For better graphical clarity here a channel was given full color whenever $b_{ik}^2 > 1/M = 0.05$ and left white otherwise. The attribution errors committed by the CPV algorithm decrease with increasing window length T. In $t = 10-15$ the inter-cluster correlations σ_{BC} are too strong and the intra-cluster correlations ρ_B and ρ_C too weak for the two clusters to be disentangled for $T = 64$. In addition not all channels of cluster "B" are found. In the situation given in $t = 20-25$ this is already possible for $T = 64$. Increasing T the dynamically changing cluster structure is revealed perfectly well by the CPV algorithm. Of course this has to be bought with a loss of time resolution.

In Fig. 12 we display the time resolved performance of three alternative approaches, cf. [16], for the same situation as in Fig. 11(c). The dynamical definition of K' makes it possible to compare all algorithms on the same footing. First, the method for automated attribution of channels to clusters is used on the basis of the largest EV instead of the CPV. This can be seen as an unsupervised version of the approach used in [10,11]. Second, we compare to the PI algorithm (including the trimming procedure of [13,14] but without the additional concept of bit strings that was introduced in [14] to filter out the most frequent cluster patterns). Third, we use the standard k-means algorithm for hard partitional clustering of data [1,2] with equal-time cross-correlation as proximity measure. As this algorithm is known to depend on the initial choice of the cluster centroids we checked the performance of the three possibilities beforehand: (i) Heuristic linear combination of K' disjoint groups of time series "i" on the basis of large C-matrix elements, (ii) use of K' randomly chosen time series, and (iii) use of K' randomly sampled time series. We used the best performing heuristic initial centroids (i) in our comparison. While for EV and PI the strength of the involvement of a channel "i" in a cluster "k" is estimated from the eigenvector components a_{ik}^2 in the case of the k-means algorithm we use the absolute value of the cross-correlation of the channels with the cluster centroids for the same purpose. Again only channels with involvement $> 1/M = 0.05$ are shown.

In the segment $t = 10-15$ a separation of the clusters is still unstable using the largest EV alone [Fig. 12(a)]. Here mainly attribution errors of category II are committed. For the same situation the PI are unable to distinguish the two clusters even in the segment $t = 20-25$ [Fig. 12(b)]. From Fig. 12(c) it becomes clear that in the segment $t = 10-15$ our implementation of the k-means algorithm does not perform convincingly either. Channel "14" of cluster "B" is constantly mixed into cluster "C" (error cat. II). We have checked that the deficiencies of the alternative algorithms cannot be cured by further increasing T.

In Fig. 12 strong inter-cluster couplings were chosen in order to put emphasis on performance differences between the CPV on the one hand and EV and PI on the other. Now we investigate the dynamical reliability of our estimates for the intra- and inter-cluster correlations. In order to be able to exploit second order perturbation theory the ratio of the average inter- and intra-cluster matrix elements $d^{(0)}/c^{(0)}$ must be confined to small values $|d^{(0)}/c^{(0)}| \lesssim 0.25$; cf. Fig. 6. Therefore the values of the intra-cluster couplings of Table II were reduced by a factor 0.8 and the inter-cluster couplings of Table III by a factor 0.4. In Fig. 13 the temporal evolution of the estimates (9) and (11) (black symbols) is shown for the strongest two clusters in comparison to the average of the involved C-matrix elements (gray symbols with error bars)

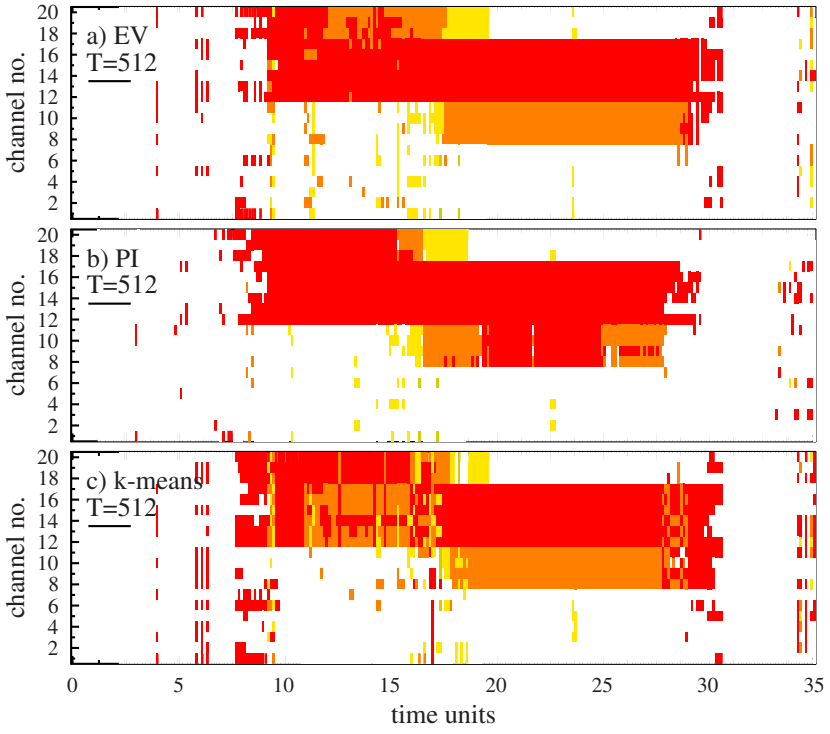

FIG. 12. (Color online) Cluster structure of the same model system as in Fig. 11(c) analyzed with three alternative algorithms and $T=512$. All channels marked by the same color (gray value) belong to the same cluster ($\mathbf{w}_{20}$: dark gray; $\mathbf{w}_{19}$: gray; $\mathbf{w}_{18}$: light gray).

and as estimated directly from the coupling strengths via Eq. (16) (fully drawn lines). If the inter-cluster coupling is not too weak the estimates C_k (where $k=A,B$) of Eq. (9) are compatible with $\langle|C_{i_k j_k}|\rangle$ and close to the value $c^{(0)}$calculated directly from Eq. (16). From (c) it becomes obvious that the estimate (11) for the inter-cluster correlations is closer to $d^{(0)}$ than $\langle C_{i_k j_l}\rangle$. Indeed the ratio $d^{(0)}/c^{(0)}\approx 0.2$ is in the region where second order perturbation theory should give good estimates for the inter-cluster correlations. Situations with larger $d^{(0)}/c^{(0)}$ were also explored. There the obtained result $D_{kl}<D'_{kl}<d^{(0)}$ is consistent with the findings of Fig. 6(c).

VII. SUMMARY AND DISCUSSION

In the present paper the features of a recent cluster detection algorithm based on cluster participation vectors (CPV) [16] have been explored. The algorithm can be used for any bivariate interrelation measure C_{ij} between two channels "i"

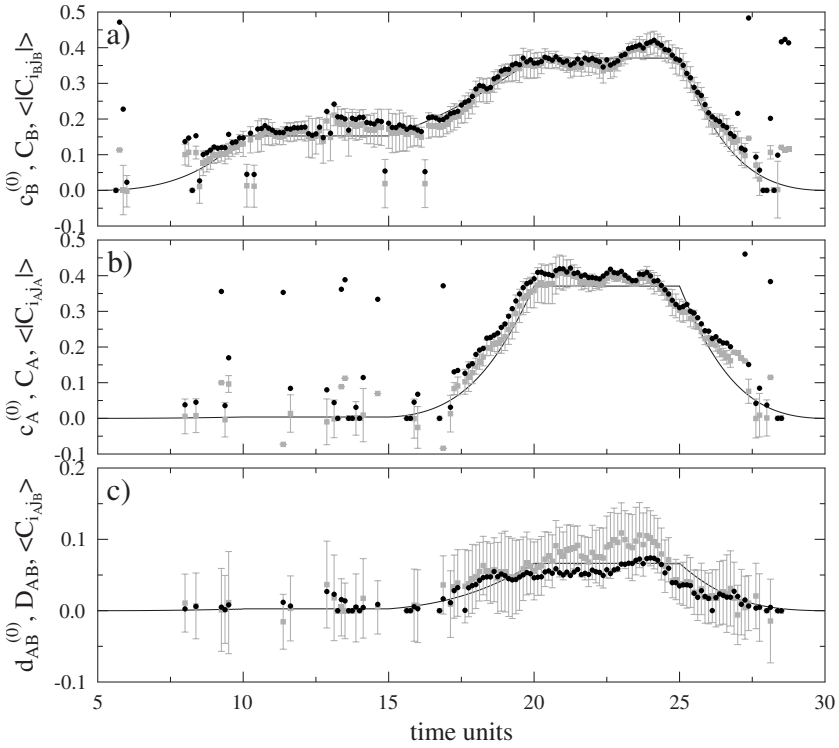

FIG. 13. (a) and (b) Estimates C_k (black symbols) and $\langle|C_{i_k j_k}|\rangle$ (gray symbols with error bars) for the intra-cluster correlations. (c) Estimates D_{kl} (black symbols) and $\langle C_{i_k j_l}\rangle$ (gray symbols with error bars) for the inter-cluster correlations. For comparison in all panels also the time evolution of the average C-matrix elements $c^{(0)}$ and $d^{(0)}$ as calculated from the coupling strengths via Eq. (16) is shown as fully drawn lines.

and "j" of a multivariate time series that satisfies the conditions (1)–(3). Details of the algorithm, which is particularly not restricted to the use of the linear equal-time cross-correlation coefficients, have been discussed in the present paper in more elaborate fashion than possible in [16].

It was shown that different to the squared coefficients a_{ik}^2 of the largest eigenvectors of the interrelation matrix $\mathbf{C}$, the components b_{ik}^2 of the CPV give information about the relative involvement of channel "i" in cluster "k" even in the presence of relatively strong inter-cluster relations.

The introduction of cluster participation coefficients (CPC) and the use of second order perturbation theory allows for the estimates (9), (10), and (11) of the intra- and inter-cluster relations. By sampling interrelation matrices the reliability of these estimates was explored systematically, finding that they outperform simple averages over the bivariate interrelations C_{ij} of the involved channels. The uncertainties of Eqs. (9), (10), and (11) are significantly smaller due the fact that they depend on large eigenvalues and the corresponding eigenvectors only, which are known to be less influenced by random correlations and noise.

After these basic studies we have specified to the special case of positive definite equal-time cross-correlation matrices and explored four different categories of attribution errors systematically on the basis of model time series with known cluster structure. In large areas of the parameter space the CPV algorithm reliably separates clusters—even in the presence of considerable inter-cluster correlations where many competing methods fail; cf. [16]. However, it gets confused for small total correlation. This is almost natural; if the cluster structure is too weak the involved channels cannot be detected reliably by any algorithm. For the CPV algorithm this is especially true if at least one of the clusters is small. In such situations the small cluster can only hardly be separated from the larger one.

Using model data with time-dependent cluster structure we found that the CPV algorithm outperforms competing approaches while still delivering reliable estimates for intra- and inter-cluster relations if only their ratio is small enough to allow for second order perturbation theory. In [16] first application of the CPV algorithm to EEG recordings containing primary generalized epileptic seizures of absence type has been presented. More work in this direction is in progress and will be published elsewhere.

ACKNOWLEDGMENTS

The author thanks Markus Müller, Gerold Baier, and Kaspar Schindler for inspiring discussions as well as Carsten Allefeld and Stephan Bialonski for helpful remarks on an early version of the algorithm. The work was supported by Deutsche Forschungsgemeinschaft, Germany (Grant Nos. RU 1401/1-1 and RU 1401/1-2).

Appendix: A simplified model for the K-cluster situation

In this appendix we give some analytical results for the block model (4) for a K-cluster situation. For vanishing inter-cluster dependencies $\mathbf{d}_{kl}=\mathbf{0}$ every block $\mathbf{C}_{kk}$ produces one large eigenvalue

$$\lambda_{k\uparrow}^{(0)} = 1 + (m_k - 1)c_k \tag{A1}$$

and m_k-1 degenerate small ones

$$\lambda_{k\downarrow}^{(0)} = 1 - c_k. \tag{A2}$$

In the m_k-dimensional subspaces of the blocks $\mathbf{C}_{kk}$ the eigenvectors corresponding to the eigenvalue (A1) are completely symmetric: $\mathbf{v}_{k\uparrow}^{(0)} = 1/\sqrt{m_k}(1,\ldots,1)^t$. Any linear combination of the antisymmetric vectors $\mathbf{v}_{k\downarrow}^{(0)} = 1/\sqrt{2}(1,0,\ldots,-1,0,\ldots)^t$ of these subspaces with only two nonzero entries with opposite sign is an eigenvector corresponding to the degenerate eigenvalues (A2). All eigenvalues that do not correspond to one of the blocks $\mathbf{C}_{kk}$ are equal to 1 with canonical basis vectors as eigenvectors. These eigenvalues are often called "bulk" eigenvalues.

For finite but small matrix elements d_{kl} the mixing of this clear situation can be assessed by perturbation theory as known from quantum mechanics. We restrict our discussion to the K largest eigenvalues (A1) and the corresponding eigenvectors in the sequel: $k,l=1,\ldots,K$. The first order correction of the eigenvalues (A1) vanishes. Using the abbreviation $\Delta\lambda_{kl}^{(0)} = \lambda_{M+1-k}^{(0)} - \lambda_{M+1-l}^{(0)}$ the lowest correction of the K largest orthonormal eigenvectors reads

$$\delta\mathbf{v}_{M+1-k}^{(1)} = \sum_{l\neq k} \frac{\sqrt{m_k m_l} d_{kl}}{\Delta\lambda_{kl}^{(0)}} \mathbf{v}_{M+1-l}^{(0)}. \tag{A3}$$

Note that because of $\langle \mathbf{v}_{k\uparrow}^{(0)} | \mathbf{d}_{kl} | \mathbf{v}_{l\downarrow}^{(0)} \rangle = 0$ the antisymmetric eigenvectors of small eigenvalues (A2) do not contribute. The overlap with the unperturbed eigenvectors becomes for $l \neq k$

$$\langle \mathbf{v}_{M+1-l}^{(0)} | \delta\mathbf{v}_{M+1-k}^{(1)} \rangle = \frac{\sqrt{m_k m_l} d_{kl}}{\Delta\lambda_{kl}^{(0)}} \tag{A4}$$

and $\langle \mathbf{v}_{M+1-k}^{(0)} | \delta\mathbf{v}_{M+1-k}^{(1)} \rangle = 0$.

The leading correction to the large eigenvalues (A1) is given by second order perturbation theory:

$$\delta\lambda_{M+1-k}^{(2)} = \sum_{l\neq k} \frac{m_k m_l d_{kl}^2}{\Delta\lambda_{kl}^{(0)}}. \tag{A5}$$

Similarly, the correction to the large eigenvectors in second order perturbation theory gives the leading order contribution to the overlap with its own unperturbed eigenvectors:

$$\delta\mathbf{v}_{M+1-k}^{(2)} = \sum_{l\neq k} \sum_{n\neq k} \frac{m_k \sqrt{m_l m_n} d_{kl} d_{kn}}{\Delta\lambda_{kl}^{(0)} \Delta\lambda_{kn}^{(0)}} \mathbf{v}_{M+1-l}^{(0)} - \frac{1}{2} \sum_{l\neq k} \frac{m_k m_l d_{kl}^2}{(\Delta\lambda_{kl}^{(0)})^2} \mathbf{v}_{M+1-k}^{(0)}. \tag{A6}$$

The first term contributes for $l \neq k$

$$\langle \mathbf{v}^{(0)}_{M+1-l} | \delta \mathbf{v}^{(2)}_{M+1-k} \rangle = \sum_{n \neq k} \frac{m_k \sqrt{m_l m_n} d_{kl} d_{kn}}{\Delta\lambda^{(0)}_{kl} \Delta\lambda^{(0)}_{kn}}, \tag{A7}$$

whereas the last term contributes to the overlap for $k=l$:

$$\langle \mathbf{v}^{(0)}_{M+1-k} | \delta \mathbf{v}^{(2)}_{M+1-k} \rangle = -\frac{1}{2} \sum_{l \neq k} \frac{m_k m_l d^2_{kl}}{(\Delta\lambda^{(0)}_{kl})^2}. \tag{A8}$$

Higher order perturbation theory does not lead to formulas that can be used for estimation of the inter-cluster relations.

[1] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data* (Prentice Hall, Englewood, NJ, 1988).
[2] A. Jain, M. Murty, and P. Flynn, ACM Comput. Surv. **31**, 264 (1999).
[3] K. Rose, E. Gurewitz, and G. C. Fox, Phys. Rev. Lett. **65**, 945 (1990).
[4] M. Blatt, S. Wiseman, and E. Domany, Phys. Rev. Lett. **76**, 3251 (1996).
[5] S. Wiseman, M. Blatt, and E. Domany, Phys. Rev. E **57**, 3767 (1998).
[6] L. Kullmann, J. Kertész, and R. Mantegna, Physica A **287**, 412 (2000).
[7] L. Giada and M. Marsili, Phys. Rev. E **63**, 061101 (2001).
[8] D.-H. Kim and H. Jeong, Phys. Rev. E **72**, 046133 (2005).
[9] S.-W. Son, H. Jeong, and J. D. Noh, Eur. Phys. J. B **50**, 431 (2006).
[10] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Phys. Rev. E **64**, 035106(R) (2001).
[11] V. Plerou, P. Gopikrishnan, B. Rosenow, Luis A. Nunes Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[12] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).
[13] C. Allefeld, M. Müller, and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 3493 (2007).
[14] S. Bialonski and K. Lehnertz, Phys. Rev. E **74**, 051909 (2006).
[15] C. Allefeld and S. Bialonski Phys. Rev. E **76**, 066207 (2007).
[16] C. Rummel, G. Baier, and M. Müller, Europhys. Lett. **80**, 68004 (2007).
[17] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
[18] T. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 2003).
[19] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, Physica D **144**, 358 (2000).
[20] F. Mormann, R. G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. E. Elger, and K. Lehnertz, Phys. Rev. E **67**, 021912 (2003).
[21] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Phys. Rev. E **65**, 041903 (2002).
[22] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[23] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger, Europhys. Lett. **70**, 278 (2005).
[24] C. W. J. Granger, Econometrica **37**, 424 (1969).
[25] W. Hesse, E. Moller, M. Arnold, and B. Schack, J. Neurosci. Methods **124**, 27 (2003).
[26] M. Kaminski and K. Blinowska, Biol. Cybern. **65**, 203 (1991).
[27] K. Sameshima and L. Baccala, J. Neurosci. Methods **94**, 93 (1999).
[28] B. Schelter, M. Winterhalder, R. Dahlhaus, J. Kurths, and J. Timmer, Phys. Rev. Lett. **96**, 208103 (2006).
[29] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[30] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[31] M. Müller, G. Baier, A. Galka, U. Stephani, and H. Muhle, Phys. Rev. E **71**, 046116 (2005).
[32] M. Müller, Y. L. Jiménez, C. Rummel, G. Baier, A. Galka, U. Stephani, and H. Muhle, Phys. Rev. E **74**, 041119 (2006).
[33] I. T. Jolliffe, *Principal Component Analysis* (Springer, Berlin, 1986).
[34] D. K. Hoffman, R. C. Raffenetti, and K. Ruedenberg, J. Math. Phys. **13**, 528 (1972).
[35] G. Dueck, J. Comput. Phys. **104**, 86 (1993).
[36] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).
[37] J. Wishart, Biometrika **20A**, 32 (1928).
[38] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[39] J. D. Noh, Phys. Rev. E **61**, 5981 (2000).