Data-driven estimates of the number of clusters in multivariate time series

Christian Rummel,[1,2,*] Markus Müller,[2,3] and Kaspar Schindler[1]
[1]*Department of Neurology, University Hospital and University of Bern, 3010 Bern, Switzerland*
[2]*Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, 62209 Cuernavaca, Mexico*
[3]*Max-Planck-Institute for the Physics of Complex Systems, 01187 Dresden, Germany*
(Received 25 April 2008; revised manuscript received 13 October 2008; published 19 December 2008)

An important problem in unsupervised data clustering is how to determine the number of clusters. Here we investigate how this can be achieved in an automated way by using interrelation matrices of multivariate time series. Two nonparametric and purely data driven algorithms are expounded and compared. The first exploits the eigenvalue spectra of surrogate data, while the second employs the eigenvector components of the interrelation matrix. Compared to the first algorithm, the second approach is computationally faster and not limited to linear interrelation measures.

DOI: 10.1103/PhysRevE.78.066703 PACS number(s): 05.10.−a, 89.75.Fb, 05.45.Tp, 87.19.L−

I. INTRODUCTION

The general problem of finding clusters in multivariate data, i.e., decomposing the data set into subsets that share a common characteristic, has a long history and yielded a variety of different solutions, see, e.g., [1–3]. Despite these efforts in time series analysis, a reliable, computationally simple and parameter-free approach is still lacking and might be helpful for a wide range of applications in diverse fields such as physics, finance, social sciences, physiology, and medicine. In all these fields multivariate time series are recorded. To better understand the underlying processes one may identify groups within the $M \gg 1$ signal channels that are related with respect to some bivariate interrelation measure.

Recently, in [4] a clustering algorithm based on "affinity propagation" has been proposed and in [5] coarse-graining of Markov chains has been applied. Within the variety of approaches to data clustering, these methods are of particular interest because they treat two fundamental problems of clustering on the same footing:

(a) estimation of the number K of clusters and

(b) attribution of the objects or data channels to these clusters.

This feature distinguishes the methods [4,5] from many clustering approaches that either start from prior knowledge or from assumptions about K or they require optimizing K in terms of *a posteriori* evaluation of the "goodness of fit."

Here we extend our previous work [6,7] in order to determine K in the context of another recent approach to data clustering. This approach consists of analyzing the components of the eigenvectors corresponding to the largest eigenvalues (in the sequel abbreviated as "large eigenvectors") of equal-time correlation matrices [8–10] or of synchronization matrices [11,12]. Using appropriate channel labeling, interrelation clusters are delineated in the matrix $\mathbf{C}$ of bivariate interrelation coefficients C_{ij} $(i,j=1,\ldots,M)$ between pairs of data channels. These clusters appear as blocks with on average larger interrelation coefficients within the blocks than between them. In this case, it is straightforward to show that for systems containing K clusters with $m_k > 1$ $(k=1,\ldots,K)$ contributing channels (and sufficiently small intercluster relations), for each cluster one eigenvalue of $\mathbf{C}$ is increased with respect to the uncorrelated situation and $m_k - 1$ eigenvalues are decreased ("level repulsion"). Unclustered channels (note that different to parts of the literature we do not address channels without partners as clusters of size $m_k = 1$ here) are not affected by the level repulsion and consequently represent the center of the eigenvalue spectrum, which is termed "bulk." Consequently, it can be expected that the number K of clusters and the total number of clustered channels $\sum_{k=1}^{K} m_k$ can be estimated by counting the numbers of "large" and of "small" eigenvalues in an appropriate way. Note that the described repulsion scheme of eigenvalues is independent of channel labeling, while the block structure of the interrelation matrix becomes visible only for appropriate channel numberings.

The presence of interrelation clusters is also reflected in the structure of the eigenvectors. Using the signal channels for representing the M-dimensional phase space ("channel basis"), the eigenvectors corresponding to the repelled eigenvalues at both edges of the eigenvalue spectrum have dominant entries exclusively for those components corresponding to the correlated data channels. In other words, these eigenvectors are dominantly aligned to the subspace spanned by the channels that contribute to one of the clusters ("cluster subspace"). The bulk eigenvectors, on the other hand, have nonzero components mainly in the orthogonal subspace. While the identification of channels that contribute to any of the clusters can be achieved on the basis of the largest eigenvector, their attribution to a particular interrelation cluster is a more delicate problem. This is due to mixing of the largest eigenvectors by any finite intercluster relations and by noise such that the clusters can no longer be deduced directly from their components.

To approach this problem, in [6,7], the concept of cluster participation vectors (CPV) was introduced. They consist of those orthogonal linear combinations of largest eigenvectors that maximize a certain distance measure. Each CPV has dominant components for those data channels only that contribute to the same interrelation cluster. Consequently, differ-

*crummel@web.de

ent to the largest eigenvectors themselves, the CPV can be used to group channels to clusters, even if considerable intercluster relations are present. Like the majority of clustering algorithms, these recent developments concentrate on problem (b), whereas for (a), only heuristic ideas based on repulsion of the largest few eigenvalues from the bulk are discussed briefly in [6,7]. Some of those ideas are limited *a priori* to very special situations like linear cross-correlation as interrelation measure or signals with white power spectra.

To overcome the limitations of our previous work, we here introduce two nonparametric and data driven solutions for problem (a). Both of them identify the numbers K^{L} of "large" and K^{S} of "small" eigenvalues that are affected by the level repulsion without requiring a (prejudiced) *ad hoc* setting of thresholds. Consequently, apart from choosing a significance level $0<\alpha<1$ for statistical tests, the methods are parameter-free. The first solution (Sec. III) employs eigenvalue spectra of ensembles of surrogate time series generated independently for each of the data channels. For linear cross-correlation clusters and arbitrary power spectra, this approach yields reliable results, but is computationally slow. The second approach (Sec. IV) makes use of the eigenvector components of the underlying interrelation matrix **C**, regardless of the eigenvalue spectrum. Focusing on appropriately defined distances between normalized eigenvectors, the method overcomes the restriction of the surrogate-based method to linear interrelation measures. In addition, it is computationally much faster.

The paper is organized as follows: In Sec. II, we present a model used subsequently as a test framework to evaluate the performance of the algorithms. This model offers large flexibility to set up different linear correlation patterns between data channels. Additionally, linear univariate properties of the signals can be adjusted to experimental data. In this sense, the employed data sets reflect important features of the real world data one is interested in, while at the same time a variety of correlation patterns can be simulated. Thereafter, the algorithms are described separately in Secs. III and IV. The performance of both approaches is compared in Sec. V by testing under which conditions the number of interrelation clusters is determined correctly. Finally, we apply the methods exemplarily to the electroencephalogram (EEG) of an epileptic patient in Sec. VI.

II. ARTIFICIAL LINEAR COUPLING OF REAL WORLD TIME SERIES

In Refs. [6,7] time series sampled from Gaussian white noise were used as a model system for performance tests. Cross-correlation clusters were constructed in a controlled way by mixing joint and individual white noise components. For this model some results are known analytically for the uncorrelated case. The amount of random correlations (nonzero values of the correlation coefficient that are solely due to the finite length T of the time series but do not reflect genuine interrelations) decreases approximately as $\sim 1/\sqrt{T}$. Furthermore, the density $\rho(\lambda)$ of the eigenvalues of the equal-time cross-correlation matrix can be written in a closed expression in the limit of an infinite M and T such that the ratio $Q=T/M=\mathrm{const}>1$ [13]. As shown in [9] numerical data obtained for the same Q values are in excellent agreement with the analytical result. Therefore one may use deviations from the analytical formula of the level density for an eigenvalue-based estimation of the cluster number [7,8]. However, finite autocorrelation times are typical for many kinds of real world data. They may drastically alter the widths of the distribution of the elements C_{ij} of the cross-correlation matrix and consequently they change also the shape of the level density and the width of the eigenvalue spectrum [14]. This effect becomes more pronounced as the power of slow Fourier frequencies increases. For this reason, we extend the mixing model of [6,7] such that power spectra of real world time series are adopted, while linear cross-correlations remain adjustable.

We start from a set of multivariate real world time series with arbitrary correlation structure $X_i(t)$ $(i=1,\ldots,M$ and $t=1,\ldots,T)$ and normalize the channels to zero mean and unit variance as defined over T data points by

$$X_i(t) \to \tilde{X}_i(t) = \frac{X_i(t) - \langle X_i \rangle}{\sigma_i}. \tag{1}$$

Next, for every channel "i" of the normalized real world time series $\tilde{X}_i(t)$, an independent surrogate [15,16] is generated $\tilde{X}_i \to \tilde{X}_i^{\mathrm{surr}}$. By definition they conserve univariate linear properties like the power spectrum and consequently the autocorrelations of each data channel, whereas phase relations and thus linear as well as nonlinear relationships between the data channels are destroyed. We used an implementation of the iterative amplitude adjusted Fourier transform (IAAFT) algorithm [15,16] independently for each data channel. To introduce an adjustable correlation pattern within $1 \leqslant K \leqslant M/2$ subgroups of size $2 \leqslant m_k \leqslant M$ $(k=1,\ldots,K;\ 2 \leqslant \Sigma_k m_k \leqslant M)$, the average signals of all channels involved in cluster "k" and cluster pair "kk'" are calculated at each time step t:

$$\xi_k(t) = \frac{1}{m_k} \sum_{i \in k} \tilde{X}_i^{\mathrm{surr}}(t), \tag{2}$$

$$\zeta_{(kk')}(t) = \frac{1}{m_k + m_{k'}} \sum_{i \in k \cup k'} \tilde{X}_i^{\mathrm{surr}}(t). \tag{3}$$

The degree of correlation between subgroups of artificial time series $Y_i(t)$ is controlled via mixing the normalized time series $\tilde{\xi}_k(t)$ (representing the common intracluster component) and $\tilde{\zeta}_{(kk')}(t)$ (representing the common intercluster component) to the channels' individual components $\tilde{X}_i^{\mathrm{surr}}(t)$:

$$Y_i(t) = \sum_k \rho_{ik} \tilde{\xi}_k(t) + \sum_{(kk')} \sigma_{i(kk')} \tilde{\zeta}_{(kk')}(t) + \tau_i \tilde{X}_i^{\mathrm{surr}}(t). \tag{4}$$

Here the channels' independence

$$\tau_i = 1 - \sum_k \rho_{ik} - \sum_{(kk')} \sigma_{i(kk')} \tag{5}$$

must fall into the interval $[0,1]$ for all channels "i" in order to be able to vary between completely correlated and com-

pletely uncorrelated situations. In-existence of clusters ($K=0$) corresponds to $\tau_i \equiv 1$. If the independence $\tau_i \equiv \tau$ of all channels is equal, $1-\tau$ is a measure for the total correlation in the system. The strength of the linear "intracluster correlations" is controlled by the coupling parameters ρ_{ik}, which have nonzero values only if channel "i" belongs to cluster "k." Similarly, the coupling parameters $\sigma_{i(kk')}$ control the strength of the linear "intercluster correlations" and are finite only if channel "i" belongs to one of the clusters k or $k' \neq k$.

Note that besides noise and random correlations, the presence of intercluster correlations complicates searching for the cluster number as well as attributing the channels to the clusters. In addition, finite intercluster correlations even add some arbitrariness to the genuine concept of clusters. For sizable intercluster correlations, it is not obvious whether separate clusters or a common cluster with substructure is the "correct" interpretation. For these reasons, we set in the following the conditions $\sigma_{i(kk')} \leqslant \max(\rho_{ik}, \rho_{ik'})$ to ensure that intracluster correlations are stronger than intercluster correlations.

Due to mixing [Eq. (4)] the linear univariate properties of the input channels X_i and the output channels Y_i cannot be identical. Rather, the model supplies multivariate time series Y_i that offer a compromise between (a) easily adjustable linear correlation patterns and (b) power spectra and autocorrelations that are "realistic" enough to serve as a model for the class of real world time series X_i of interest. As our final application will be cluster detection in time series of human electroencephalograms (EEG), we illustrate this statement with the example of a seizure and artifact-free EEG epoch (2 days after electrode implantation and 2 days before the first seizure) of $T=1024$ data points length, selected from a continuous intracranial recording of an epileptic patient (see Sec. VI for further details concerning the clinical data set). The results shown in this and the subsequent sections do not depend sensitively on the specific choice of the data segment. Figures 1(a) and 1(b) show the power spectrum (normalized to total power) and autocorrelation functions for two channels X_1 and X_2 of the original data. In Figs. 1(c) and 1(d) the corresponding quantities are shown for two strongly coupled channels Y_1 and Y_2. Due to the mixing procedure (4) these quantities are more similar to each other for the output channels Y_i than for the input data X_i. For weaker coupling, this effect diminishes. The important point here, however, is that *qualitatively* characteristic features of the input data like the decay of the power spectrum (approximately $1/f$) and of the autocorrelation function at small delays Δt are well captured by the correlated output data.

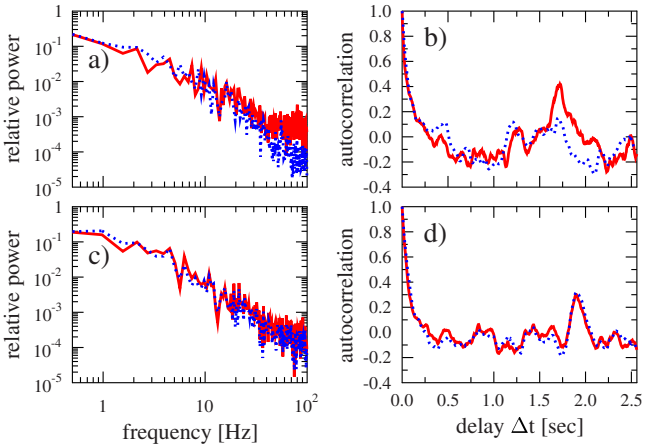

FIG. 1. (Color online) Comparison of relative power (normalized to the total power, left) and autocorrelation functions (right) of two original EEG channels (top) and two artificially correlated ($\rho=0.9$, $\sigma=0.0$, $\tau=0.1$) data channels (bottom). The fully drawn red and dotted blue lines refer to data channels "1" and "2" respectively.

III. SURROGATE-BASED ALGORITHM

In order to determine the number K of linear cross-correlation clusters, it was suggested in [7] to compare the eigenvalue spectrum λ_l ($l=1,\ldots,M$) of the equal-time cross-correlation matrix [17,18]

$$\mathbf{C} = \frac{1}{T}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^t \tag{6}$$

constructed from the measured data with the distribution of eigenvalues λ_l^n ($n=1,\ldots,N_{\mathrm{surr}}$) of an ensemble of surrogates produced from the same data. In Eq. (6), $\tilde{\mathbf{Y}}$ is the $M \times T$ data matrix with elements $\tilde{Y}_{it}=\tilde{Y}_i(t)$, normalized as in Eq. (1). In the sequel, it is assumed that the eigenvalues are ordered according to their size: $\lambda_{l+1} \geqslant \lambda_l$. Based on the eigenvalue repulsion scheme described in the Introduction, a cluster will be defined for each eigenvalue λ_k at the upper end of the spectrum that is significantly larger than the corresponding surrogate eigenvalues λ_k^n. Thus we begin with the largest eigenvalue λ_M and count for how many large eigenvalues the null hypothesis H_k that λ_k stems from the distribution of the $\{\lambda_k^n\}$ can be rejected (without interruption). This number gives an estimate $K_{\mathrm{surr}}^{\mathrm{L}}$ of the number of clusters in the data. For time series recorded from neurons *in vitro*, a similar surrogate-based idea was used recently in [19].

A few technical details are in order here: As each cluster consists of at least two channels, the maximal possible cluster number is $K_{\max}=M/2$. To formalize the comparison of the eigenvalue spectra in a nonparametric way, we rank order for each of the $K_{\max}$ largest eigenvalues ($k=M-K_{\max}+1,\ldots,M$) the joint distribution $\Lambda_k=\{\lambda_k,\lambda_k^n\}$ of the original and the surrogate eigenvalues. The probability that the original eigenvalue λ_k has at least rank $N_{\mathrm{surr}}+1-\nu_k$ (where $0 \leqslant \nu_k \leqslant N_{\mathrm{surr}}$) in Λ_k if it is actually consistent with the surrogates' distribution can be calculated. Given N_{surr} sets of surrogates, it is $p_k=(\nu_k+1)/(N_{\mathrm{surr}}+1)$. To detect significant deviations in the eigenvalue spectrum on an overall significance level $0<\alpha<1$, the individual significance levels α_k of the multiple comparisons must be reduced. We use the Holm-Bonferroni correction [20] and reject H_k if $p_k<\alpha_k=\alpha/[K_{\max}-\mathrm{rank}(p_k)+1]$. Similar to the more conservative Bonferroni correction where $\alpha_k \equiv \alpha/K_{\max}$ the size N_{surr} of the surrogate ensemble is fixed by α and $K_{\max}$ in the following way: $N_{\mathrm{surr}}=K_{\max}/\alpha-1$. The maximally needed number of surrogates is thus given by $N_{\mathrm{surr}}=M/(2\alpha)-1$. As this number can be large, the workload can be reduced considerably for practical applications by confining the search to a reasonably chosen $1<K_{\max}<M/2$ *a priori*. A flowchart of this algorithm is given in Fig. 2.

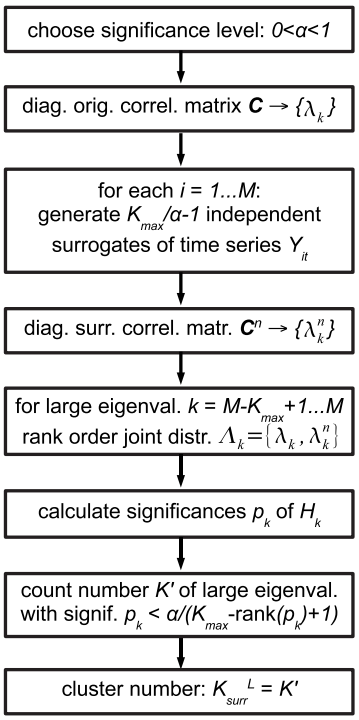

FIG. 2. Flowchart of the surrogate-based algorithm, see text for details.

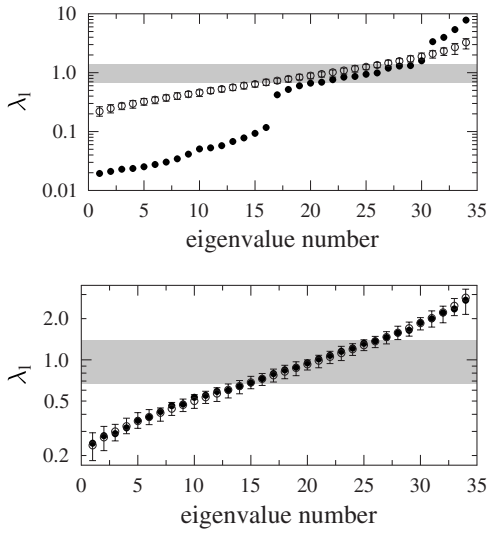

FIG. 3. Illustrative examples for the eigenvalue spectrum of the correlation matrix of model data (full symbols) and the distribution of $N_{\mathrm{surr}}=339$ surrogates, for which the average and extremal values are shown (open symbols, the errors bars span from the maximal to the minimal λ_l^n). Top: $K=4$ clusters of $m=5$ channels each ($\rho=0.44$, $\sigma=0.12$, $\tau=0.2$). Bottom: purely randomly correlated case $K=0$. In both panels the gray shaded region indicates the range of nonzero eigenvalue density for multivariate time series produced from uncorrelated white noise and the same parameters M and T, see text.

To illustrate this method, we show in Fig. 3(a) an example of the eigenvalue spectra for four strongly correlated clusters, each of them containing five data channels, and some moderate intercluster correlations. The total number of data channels is $M=34$ and the cross-correlation matrix (6) is constructed over $T=1024$ sample points. In the original eigenvalue spectrum, increased gaps $\lambda_{l+1}-\lambda_l$ are found between the two largest eigenvalues λ_{34} and λ_{33} (better visible on the linear scale) as well as between λ_{31} and λ_{30}. The first gap suggests a single cluster, whereas the second one implies four clusters. Indeed, the gap between the two largest eigenvalues is caused by the presence of intercluster correlations. As discussed above, several clusters with sizable intercluster correlations could also be interpreted as a single cluster with a certain finer structure. Hence if the first noticeable gap was used for defining the cluster number, this single cluster picture is implicitly applied. In Fig. 3(a), the largest *four* eigenvalues of the original data are clearly larger than the maximal values of the $N_{\mathrm{surr}}=339$ surrogates (open symbols) corresponding to $K_{\max}=M/2$ and a significance level of $\alpha=0.05$. Following the scheme described above, the cluster number $K=4$ in the data can be correctly deduced regardless of the finite intercluster correlations. Note the additional pronounced gap between λ_{17} and λ_{16} that clearly distinguishes a group of $K^{\mathrm{S}}_{\mathrm{surr}}=16$ small eigenvalues from the bulk. Thus, for this example, also the total number of clustered channels can be estimated correctly as $\Sigma_{k=1}^{K} m_k = K^{\mathrm{L}}_{\mathrm{surr}} + K^{\mathrm{S}}_{\mathrm{surr}} = 20$.

For data without any genuine correlations, the situation is shown in Fig. 3(b). Here all eigenvalues are compatible with those derived from the surrogates within the statistical error. Note that in Fig. 3 the surrogate eigenvalue spectra are considerably broader than expected for uncorrelated white noise time series (gray shaded region $0.669 \leqslant \lambda_l \leqslant 1.397$, compare to [13]). This pronounced deviation is solely due to the different power spectra of our EEG based time series (4) and demonstrates why a simple comparison with white noise model data is not sufficient.

From these observations it becomes evident that in principle the cluster number K can be deduced by comparing the original and surrogate eigenvalues. A similar result has been obtained recently from a parametric testing procedure in [19]. However, there are two important deficiencies of this approach. First, the generation of IAAFT surrogate ensembles of size $N_{\mathrm{surr}} \sim K_{\max}/\alpha$ is computationally expensive. In [15] it was shown that in order to reach accuracy δ for the power spectra, roughly $1/\delta$ iterations are necessary. The computational complexity of the surrogate-based algorithm is accordingly given by the generation of $K_{\max}/\alpha$ univariate surrogates of IAAFT type for each of the M channels with accuracy δ using the fast Fourier transform (complexity $T \log_2 T$ [21]) and amounts to $(K_{\max}/\alpha) M (1/\delta)(T \log_2 T)$. Only if one needs to calculate surrogate eigenvalues anyway, e.g., in order to estimate the genuine cross-correlation strength [14] in the multivariate data set, the *additional* workload of the surrogate-based algorithm for estimating K is negligibly small.

Second, the surrogate-based method can be used for linear interrelation measures such as cross correlation only. Within the error tolerated during generation of the surrogates, *linear* univariate properties like the autocorrelation function are the same as for the original data. Genuine cross correlations, however, are destroyed and the surrogates may serve to test the null hypothesis of their complete absence, even though

the correlation coefficient C_{ij} is influenced by the autocorrelation. On the other hand, *nonlinear* univariate properties of the signals are destroyed by the IAAFT surrogates together with linear and nonlinear interrelations. Otherwise, this technique could not be used for nonlinearity testing [16]. This implies that in cases where nonlinear univariate properties might have some impact on nonlinear bivariate dependencies, IAAFT surrogates are no longer suitable to probe the null hypothesis of zero nonlinear interrelations. An alternative might consist of the computationally even more expensive constrained randomization of time series [22] tailored to conserve exactly those univariate nonlinear properties of interest.

IV. EIGENVECTOR-BASED ALGORITHM

Different to the previous section, in the following we do not need to restrict ourselves to any specific interrelation measure between two of the M data channels. We use linear cross correlation only for illustration. All we need in general is that the matrix $\mathbf{C}$ constructed from bivariate interrelation coefficients fulfills three requirements. (i) The measure must be normalizable, $-1 \leqslant C_{ij} \leqslant 1$. (ii) Each channel must be perfectly interrelated with itself, $C_{ii}=1$. (iii) The measure must be symmetric, $C_{ji}=C_{ij}$, in order to obtain real eigenvalues λ_l and eigenvectors $\mathbf{v}_l$.

A first step towards identifying the number of interrelation clusters of a multivariate data set from the eigenvectors of $\mathbf{C}$ consists of separating the subspace of interrelated data channels ("cluster subspace") from the subspace of independent signals ("bulk subspace"). A criterion for such a discrimination can be derived directly from the structure of the eigenvectors of $\mathbf{C}$. As described in the Introduction, the eigenvectors of the repelled eigenvalues (i.e., those at both edges of the spectrum) have dominant components in the channels corresponding to correlated signals, while the remaining eigenvectors are essentially orientated into the orthogonal bulk subspace. In order to achieve the desired separation, we compute the following *distance measure* for all pairs of normalized eigenvectors:

$$D_{ll'} = \sum_{i=1}^{M} |a_{il}^2 - a_{il'}^2|. \tag{7}$$

This quantity was introduced in [6,7] to construct the CPV. Here, the a_{il} of Eq. (7) are the components of the eigenvectors $\mathbf{v}_l$ and $\mathbf{v}_{l'}$ in the channel basis. It is easily checked that Eq. (7) satisfies all requirements of a *metric* (symmetry, positive semidefiniteness, and the triangle inequality). $D_{ll'}$ is equal to zero for vectors where all components are identical to each other up to a sign and assumes its maximum value $D_{ll'}=2$ when the normalized vectors $\mathbf{v}_l$ and $\mathbf{v}_{l'}$ have no common components. Note the similarity to the "taxicab metric" or "Manhattan distance" $D_{\mathrm{taxi}}=\Sigma_{i=1}^{M}|a_{il}-a_{il'}|$ [23].

The measure (7) is designed such that vectors belonging to orthogonal subspaces generically have larger distances than vectors that are taken from the same subspace. Hence distances between eigenvectors corresponding to eigenvalues that show a genuine repulsion either at the lower or upper edge of the spectrum on the one hand, or eigenvectors corresponding to bulk eigenvalues on the other ("within subspace distances"), show on the average smaller distances than those between these groups ("intersubspace distances"). Furthermore, in cases where the bulk subspace is not empty, it is possible to distinguish within the correlated subspace between "small" and "large" eigenvectors, provided the eigenvalues (and corresponding eigenvectors) are labeled in rank order.

Thus if nonrandom correlations are present, the matrix $\mathbf{D}$ shows a characteristic pattern where regions of predominantly smaller and larger distances alternate. The former appear between eigenvectors corresponding to the small (region S), the large (region L), or the bulk (region B) eigenvalues as well as between combinations of one small and one large eigenvalue (region SL). These regions are interrupted by regions of larger distances between bulk eigenvectors and eigenvectors corresponding to either the large (region BL) or the small eigenvalues (region SB). An example of this characteristic pattern is illustrated in Fig. 4(a) for the same data as in Fig. 3(a). The matrix elements $D_{ll'}$ displayed above the diagonal show a clear block structure that can be exploited to distinguish the subspaces. Below the diagonal, the separation into regions of smaller within subspace distances (white) and larger intersubspace distances (black) is symbolized accordingly. The number $K_{\mathrm{ev}}^{\mathrm{L}}$ of eigenvectors belonging to region L can be used to define the number of correlation clusters. A separation of regions L and S is possible whenever region B is nonempty, i.e., for the eigenvector based algorithm at least one data channel must be unclustered. In the present example, the cluster number can be defined correctly as $K_{\mathrm{ev}}^{\mathrm{L}}=K=4$ and the total number of clustered data channels is given by $\Sigma_{k=1}^{K} m_k = K_{\mathrm{ev}}^{\mathrm{L}} + K_{\mathrm{ev}}^{\mathrm{S}} = 20$ where $K_{\mathrm{ev}}^{\mathrm{S}}=16$ is the number of eigenvectors belonging to region S.

The optimal partitioning of the matrix $\mathbf{D}$ can be achieved automatically by studying the distributions of the within and intersubspace distances $D_{ll'}$ [white and black regions below the diagonal in Fig. 4(a)] under independent variation of $K_{\mathrm{ev}}^{\mathrm{L}}$ and $K_{\mathrm{ev}}^{\mathrm{S}}$. As an objective criterion for the goodness of the partitioning, we utilize nonparametric Kolmogorov-Smirnov (KS) tests [21,24] to compare the unbinned partial distributions of the within and the intersubspace distances to their joint distribution (distribution of all $D_{ll'}$ independent of any supposed partitioning). The KS test quantifies the dissimilarity between two distributions and gives a significance p_{KS}, which measures the probability that the dissimilarity is incidentally equal or larger than the obtained value, although the distributions are actually consistent [21,24]. Varying $K_{\mathrm{ev}}^{\mathrm{L}}$ and $K_{\mathrm{ev}}^{\mathrm{S}}$, we pick the partitioning that has the least chance p_{KS} to come about by incidence. In Fig. 4(b) the partial distributions of this optimal partitioning are displayed together with the joint one. For the shown example it is obvious that the null hypothesis, that both partial distributions are consistent with the joint one, can be rejected on an extremely small significance level α.

Some technical details of the procedure are in order here. As in the surrogate based algorithm of the previous section, the range where $K_{\mathrm{ev}}^{\mathrm{L}}$ must be searched is limited by the number of channels M: $K_{\mathrm{ev}}^{\mathrm{L}}=1,\ldots,M/2$. To form a cluster, each

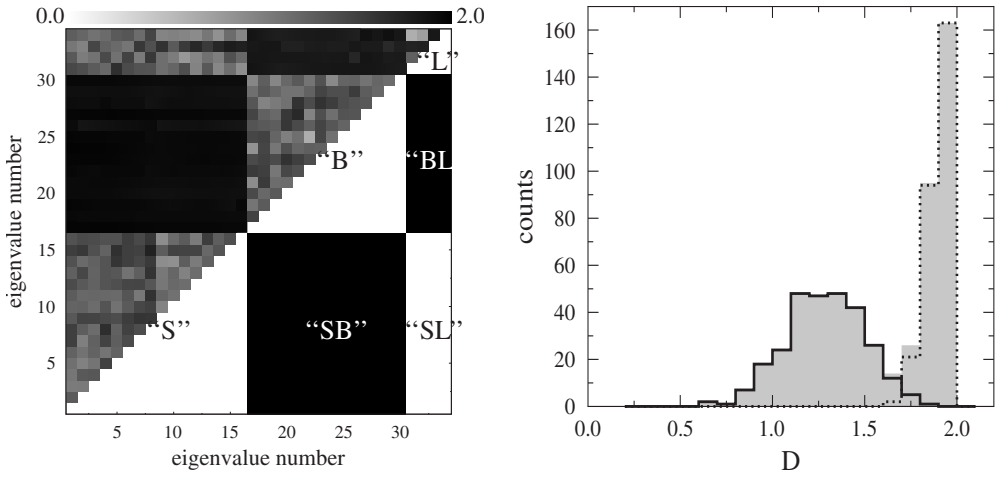

FIG. 4. Left: Distance matrix for the same example as in Fig. 3(a) ($K=4$, $m=5$, $\rho=0.44$, $\sigma=0.12$, $\tau=0.2$). Whereas above the diagonal the original matrix elements are shown, the idealized block structure of within and intersubspace distances to be revealed by the algorithm is symbolized as white and black regions below the diagonal. Right: Histograms of the distributions of the within (fully drawn) and intersubspace distances (dotted). The significances of a KS test for consistency of the partial with the joint distribution of all off-diagonal matrix elements (gray shaded) are $p_{\mathrm{KS}}^{\mathrm{within}}=6.0\times10^{-40}$ and $p_{\mathrm{KS}}^{\mathrm{inter}}=3.9\times10^{-40}$, respectively.

large eigenvector must have at least one small partner and hence $K_{\mathrm{ev}}^{\mathrm{S}}\geqslant K_{\mathrm{ev}}^{\mathrm{L}}$. In addition, in order to be able to distinguish between regions L and S, at least one eigenvector must fall into the bulk subspace B. Thus $K_{\mathrm{ev}}^{\mathrm{S}}$ is limited to the range $K_{\mathrm{ev}}^{\mathrm{S}}=K_{\mathrm{ev}}^{\mathrm{L}},\ldots,M-1-K_{\mathrm{ev}}^{\mathrm{L}}$. The $\sim M^2/4$ elements of the space spanned by $K_{\mathrm{ev}}^{\mathrm{L}}$ and $K_{\mathrm{ev}}^{\mathrm{S}}$ can be scanned systematically for the best partitioning.

In order to optimize the partitioning, we search the minimum of the sum $P=p_{\mathrm{KS}}^{\mathrm{within}}+p_{\mathrm{KS}}^{\mathrm{inter}}$, where $p_{\mathrm{KS}}^{\mathrm{within}}$ gives the significance of the dissimilarity between the distribution of the within subspace distances [white region in Fig. 4(a)] and the joint distribution of all off-diagonal matrix elements $D_{ll'}$ [gray shaded in Fig. 4(b)]. $p_{\mathrm{KS}}^{\mathrm{inter}}$ is the same for intersubspace distances [black region in Fig. 4(a)]. Note that different to only quantifying the dissimilarity of two distributions, the significances p_{KS} in addition take into account the effect of different sample sizes induced by varying $K_{\mathrm{ev}}^{\mathrm{L}}$ and $K_{\mathrm{ev}}^{\mathrm{S}}$

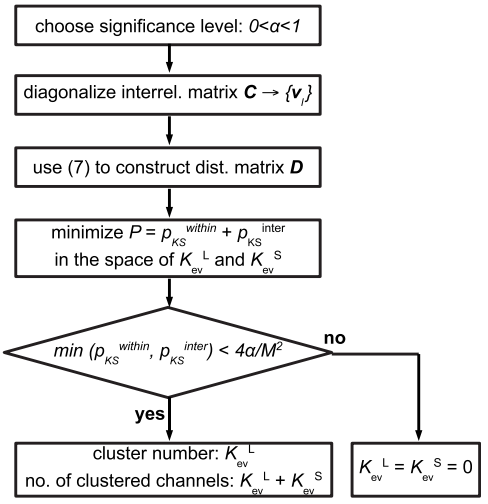

FIG. 5. Flowchart of the eigenvector-based algorithm, see text for details.

[21,24]. Finally, the partitioning that minimizes P is accepted only if at least one of the partial distributions is incompatible with the joint one on a chosen significance level α. As the systematic scan of the space spanned by $K_{\mathrm{ev}}^{\mathrm{L}}$ and $K_{\mathrm{ev}}^{\mathrm{S}}$ may lead to incidentally small p_{KS} in every trial, again a Bonferroni correction for the number of trials $\sim M^2/4$ must be carried out. Thus the partitioning is accepted only if $\min(p_{\mathrm{KS}}^{\mathrm{within}},p_{\mathrm{KS}}^{\mathrm{inter}})<4\alpha/M^2$. If this requirement is not fulfilled, we reject the partitioning as probably incidental and set $K_{\mathrm{ev}}^{\mathrm{L}}=K_{\mathrm{ev}}^{\mathrm{S}}=0$.

This eigenvector-based algorithm for estimating the cluster number $K_{\mathrm{ev}}^{\mathrm{L}}$ and the total number of clustered channels $\Sigma_{k=1}^{K}m_k=K_{\mathrm{ev}}^{\mathrm{L}}+K_{\mathrm{ev}}^{\mathrm{S}}$ is summarized in the flowchart in Fig. 5. Different to the surrogate-based algorithm of Sec. III, application of this recipe to nonlinear interrelation measures is possible. A second advantage is that neither the chosen level of significance α influences the eigenvector-based algorithm's additional workload given by M^3 (construction of the distance matrix **D**) nor are iterations necessary. Therefore, as compared to IAAFT surrogates, the time complexity $\mathcal{T}$ of the eigenvector based algorithm is

$$\frac{\mathcal{T}_{\mathrm{surr}}}{\mathcal{T}_{\mathrm{ev}}}\sim\frac{K_{\max}}{\alpha}\frac{1}{\delta}\frac{T\log_2 T}{M^2} \tag{8}$$

times smaller, i.e., typically *several orders of magnitude*. In this respect, online estimation of the cluster number in continuously recorded multivariate signals is easily feasible, which is in practice a serious problem for the algorithm based on IAAFT surrogates due to huge computation times.

V. PERFORMANCE TEST

In the following, the performance of the two algorithms is assessed in terms of correct estimates of the cluster number K. Using the model of Sec. II the performance is quantified by the ratio $R=N_K/N_{\mathrm{ens}}$ of the number N_K of correct detec-

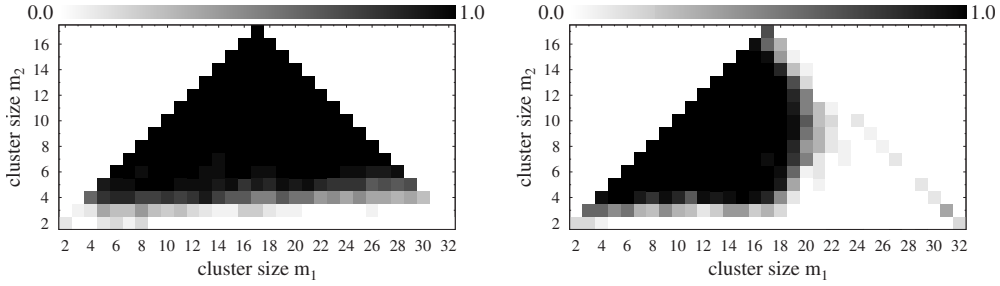

FIG. 6. Dependence of the performance measure R on the cluster sizes m_1 and m_2 in a two cluster situation with parameters $M=34$, $T=512$, $\rho_1=\rho_2=0.7$, $\sigma=0$, $\tau=0.3$, and $N_{\mathrm{ens}}=20$. Left: Surrogate-based algorithm. Right: Eigenvector-based algorithm. The significance level is chosen equally as $\alpha=0.05$ in both cases.

tions $K^{\mathrm{L}} \equiv K$ and the size N_{ens} of the used ensemble of independent realizations of the data matrix (4) produced under identical conditions. Not surprisingly, we found that both algorithms perform better for larger T, i.e., the longer the analyzed time series are (data not shown). This can be easily explained by the suppression of random correlations with increasing T. For our model data derived from interictal EEG, we found that both algorithms yield good results for $T/M \gtrsim 10$; though this finding may not apply for data with considerably different power spectra. In addition, we confirmed the intuitive expectation that reliability of both algorithms decreases with increasing signal independence τ_i (data not shown), i.e., the smaller the correlation of the clusters, the larger the error rate.

In the following, we investigate situations that are unfavorable for at least one of the algorithms. In Fig. 6 the performances are compared for a two cluster situation $K=2$ with $\rho_1=\rho_2=0.7$ and $\sigma=0$ as a function of the individual cluster sizes m_1 and m_2. We set $M=34$ and $T=512$ in this and the subsequent figures and choose a significance level $\alpha=0.05$ requiring the generation of $N_{\mathrm{surr}}=339$ surrogates. Both algorithms perform poorly if one of the two clusters is very small, $m_1, m_2 \lesssim 3$. For small $m_1 \approx m_2$, the eigenvector-based algorithm slightly outperforms the surrogate-based one. In this case, repulsion of the two largest eigenvalues due to cluster formation is too weak to be detected reliably by the surrogate-based algorithm. In the case of small bulk sizes $m_1+m_2 \approx M$ the situation is opposite as the eigenvector-based algorithm has difficulties in partitioning the distance matrix $\mathbf{D}$. A more serious weakness of the eigenvector-based algorithm is revealed in situations with large asymmetries $m_1 \gg m_2$ of the cluster sizes. For $\rho_1=\rho_2$ in these cases mainly a single cluster is found $K_{\mathrm{ev}}^{\mathrm{L}}=1$. Using the CPV algorithm [6,7] only the channels contributing to the larger group are detected, while the smaller cluster is missed. We found that in these situations the performance of the eigenvector-based algorithm improves considerably if the asymmetry $m_1 > m_2$ is balanced by stronger coupling strengths $\rho_1 < \rho_2$ in the smaller cluster. The surrogate-based algorithm is almost unaffected by asymmetric cluster sizes.

To detect performance differences not due to asymmetries and to simultaneously keep the huge parameter space of possible cluster patterns manageable, we constrain the analysis to the special case of equal cluster sizes $m_k \equiv m$ in the sequel. In addition, all intra- and inter-cluster coupling strengths of Eq. (4) are set to the same constant: $\rho_{ik} \equiv \rho$ and $\sigma_{i(kk')} \equiv \sigma$ (for all data channels "i" of a cluster and $k,k'=1,\ldots,K$). For these situations we investigated the dependence of the algorithms' performances on the ratio $r_i = \Sigma_{(kk')}\sigma_{i(kk')}/\Sigma_k \rho_{ik}$ between the total inter- and intra-cluster coupling strengths affecting channel "i" (data not shown). It turned out that different to allocating the channels to clusters [6,7] the performance for estimating the cluster number K was rather insensitive to the specific composition of inter- and intra-cluster couplings. Therefore we also fix all r_i as $r_i \equiv r = (K-1)\sigma/\rho = 0.5$ in the sequel.

With these specifications we explore the dependence of the performance measure R on cluster size m and number K. Results are shown in Fig. 7. The empty region above the solid line corresponds to $K > M/m$ and is not accessible under the simplifying assumption of equal cluster sizes $m_k \equiv m$. In the example shown, the independence of channels is fixed at the relatively large value of $\tau=0.4$. We find that both algorithms have problems in detecting small clusters. Again small clusters are missed more frequently by the surrogate-based algorithm resulting in too small estimates $K^{\mathrm{L}} < K$. For smaller τ we obtained better results for both algorithms, while the outcome was qualitatively unchanged. The eigenvector-based algorithm has certain difficulties for small bulk sizes, as can be seen from the areas close to the line $K=M/m$.

To study how strongly the estimated cluster number depends on the chosen significance level α we investigate the unfavorable situation $m=4$ (small cluster sizes) and $\tau=0.4$ (large channels' independence). In Fig. 8 the estimated cluster number K^{L} is shown as a function of K for four different values of α. As indicated by the error bars, the dependence of both algorithms on α is weak. However, as the eigenvector-based algorithm has difficulties in rejecting spurious partitionings if α is too large, $\alpha \leqslant 0.10$ is recommended. The surrogate-based algorithm tends to underestimate K except in the uncorrelated situation ($K=0$), which is revealed securely. On the contrary, for the eigenvector-based algorithm the deviation of K^{L} from K is less pronounced and has smaller standard deviation. The only exceptions are cluster number $K=0$ (for $\alpha > 0.10$) and small bulk sizes ($m_{\mathrm{bulk}}=2$ for $K=8$).

In summary, we found that the performance of the eigenvector-based algorithm is impaired by strongly asymmetric cluster sizes and small bulk sizes. To avoid difficulties in detecting the absence of clusters, we recommend a significance level of at least $\alpha \leqslant 0.10$. The performance of the surrogate-based algorithm is not impaired by these problems.

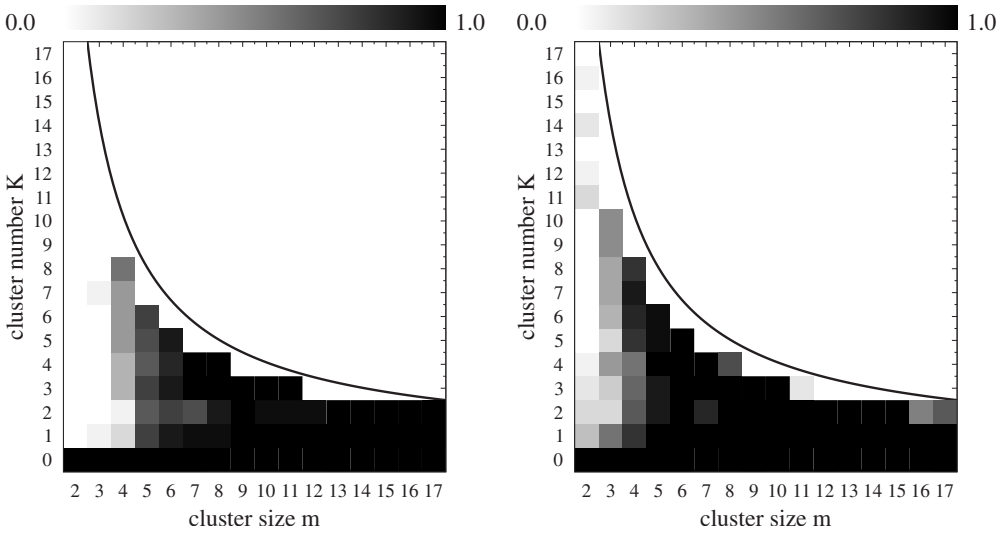

FIG. 7. Dependence of the performance measure R on cluster size m and cluster number K. $M=34$, $T=512$, $\tau=0.4$, $r=0.5$, and $N_{\mathrm{ens}}=20$. Left: Surrogate-based algorithm. Right: Eigenvector-based algorithm. The significance level is chosen equally as $\alpha=0.05$ in both cases. As a fully drawn line the function $K=M/m$ is shown.

However, for small cluster sizes it performs worse than the eigenvector-based algorithm and under certain conditions the underestimation of the cluster number is more pronounced.

VI. APPLICATION TO INTRACRANIALLY RECORDED EEG

Epilepsy is the second most common neurological disease and roughly one-third of the patients are not rendered seizure-free by current therapies. It is plausible that a better understanding of the spatiotemporal evolution of epileptic seizures may ultimately promote development of new tools for seizure control. Importantly, epileptic seizures are not produced by single neurons, but are the result of the emergent activity of large neuronal networks that coalesce into clusters with pathologic collective dynamics [25]. Therefore investigating cluster formation and disintegration in EEG recordings seems to be a reasonable approach to assess seizure dynamics. To demonstrate the usefulness of the methods developed in the present paper in this context, we show in the following their application to EEG recordings. Other interrelation matrix-based clustering approaches to multivariate neurophysiological data have been made in [6,12,19]. Different clustering approaches to EEG data from epilepsy patients were recently presented in [26,27].

Here, we discuss in detail the evolution of the cluster pattern in an intracranially recorded EEG of an 18-year-old male suffering from pharmacoresistant temporal lobe epilepsy. The patient underwent evaluation for the possibility of epilepsy surgery at the Inselspital in Bern, Switzerland. He had signed informed consent that EEG and imaging data might be used for research purposes, and the study protocol had previously been approved by the ethics committee of the University of Bern. Using four implanted strip and two foramen ovale electrodes (40 contacts in total, see the x-ray im-

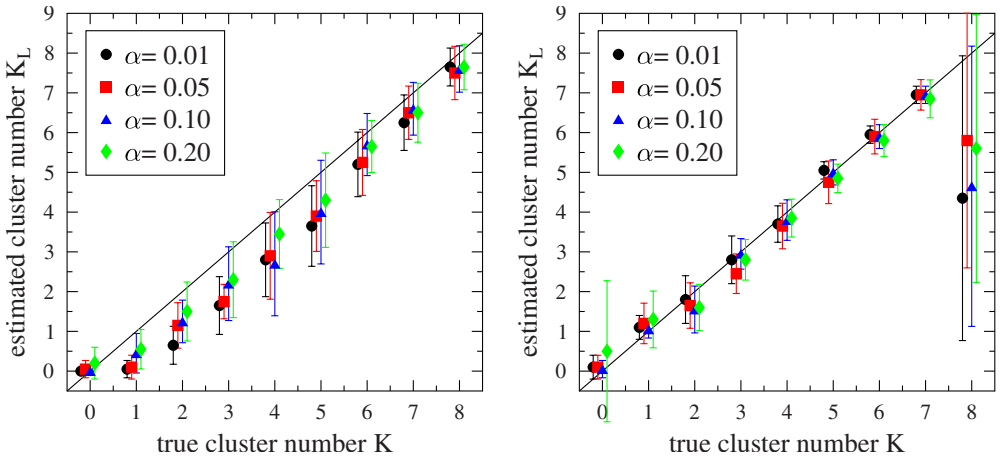

FIG. 8. (Color online) Dependence of the estimated cluster number K^{L} on the true cluster number K for various significance levels α. $M=34$, $T=512$, $m=4$, $\tau=0.4$, $r=0.5$, and $N_{\mathrm{ens}}=20$. Left: Surrogate-based algorithm. Right: Eigenvector-based algorithm. For better visibility the symbols are shifted slightly along the x axis and the line $K^{\mathrm{L}}=K$ is shown for eye guidance.

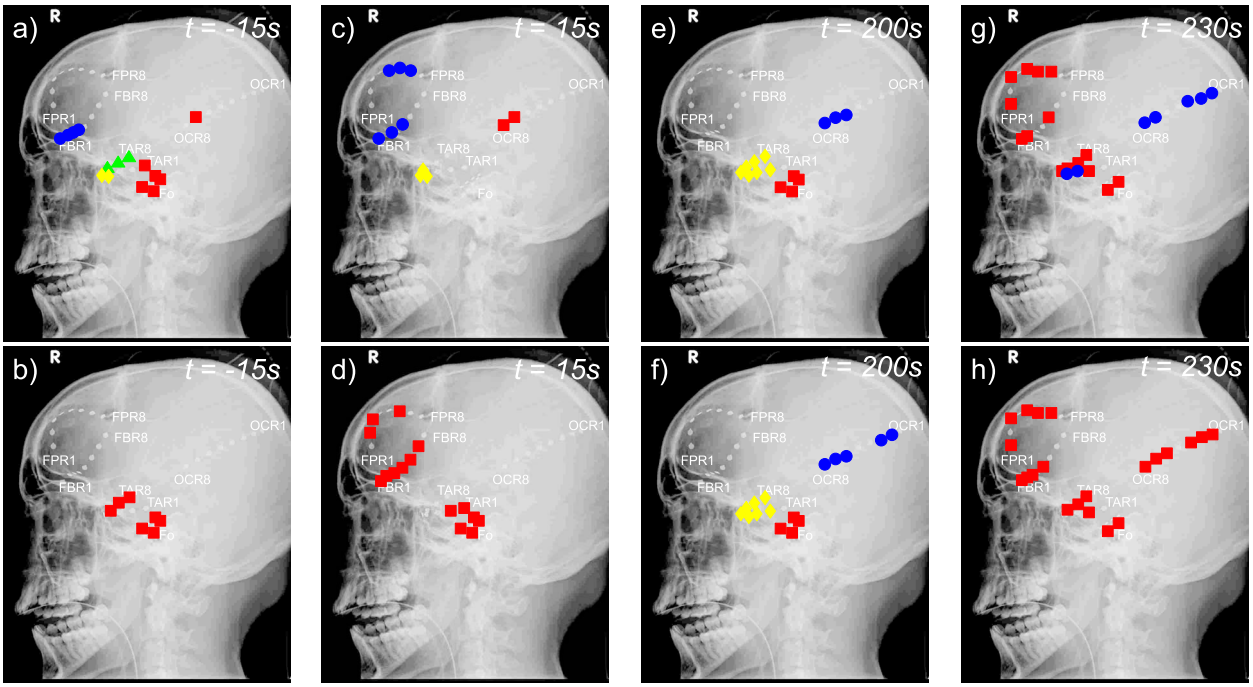

FIG. 9. (Color online) Evolution of the cluster structure during a typical focal onset epileptic seizure of a patient suffering from pharmacoresistant temporal lobe epilepsy. Contacts allocated to the same cluster are marked by the same symbol and color. Top: Cluster number estimated using the surrogate-based algorithm. Bottom: Cluster number estimated using the eigenvector-based algorithm. Shown are snapshots of a pre-ictal [(a) and (b)], an early ictal [(c) and (d)], a late ictal [(e) and (f)], and a post-ictal [(g) and (h)] period. Abbreviations: Fo: foramen ovale electrodes; FPR1–FPR8: right fronto-polar strip electrode; FBR1–FBR8: right fronto-basal strip electrode; TAR1–TAR8: right temporal anterior strip electrode; and OCR1–OCR8: occipital central strip electrode.

aging in Fig. 9) the EEG was recorded directly from the cortical surface (sampled at 200 Hz, bandpass filtered between 0.1 and 70 Hz, analog-to-digital converted with 16-bit resolution). After excluding artifact-contaminated channels (Fo2, Fo5, Fo7, Fo8, FPR4, and FPR8) by visual inspection, $M=34$ EEG signals were analyzed. The average of all channels was used as the reference signal because it had been previously demonstrated that of all the commonly used references, this one distorts the eigenvalue spectrum of correlation matrices the least [28].

An interictal epoch (far away from seizure activity) of the same long-term EEG had already been used as input for the "artificially coupled EEG" in Secs. II–V. Here we show the results for a partial complex seizure that is initiated in the inner part of the tip of the right temporal lobe and impairs consciousness during its evolution. Seizure duration and evolution (concerning symptoms and signs as well as the cluster patterns) of the selected example were representative for all five seizures recorded during the long-term EEG. An analysis window of $T=1024$ corresponding to 5.12 s was shifted over the recording with 200 sample points (1 s) step width in order to analyze eigenvalues and eigenvectors of the cross-correlation matrix (6) in a time resolved way. For the surrogate-based algorithm we limited the maximal cluster number to $K_{\max}=10$ in order to save computation time. We consider this choice to be safe because in the whole data set of 20 min duration comprising this seizure, the obtained cluster number was confined by $K^{\mathrm{L}}_{\mathrm{surr}} \leqslant 7$.

In most parts of the data set the surrogate-based algorithm estimated a larger cluster number K^{L} than the eigenvector-based one. Different to the previously studied model of "artificially coupled EEGs" we can no longer decide strictly which estimate for the cluster number is correct. To assess the plausibility of the results, we therefore employed the CPV algorithm [6,7] to allocate channels to clusters.

In Fig. 9 four snapshots of the cluster patterns are displayed for both algorithms. The snapshots are representative of the corresponding part of seizure evolution, i.e., they do not depend sensitively on the chosen time step. As defined by visual inspection of the EEG by an experienced epileptologist and electroencephalographer (K.S.), the seizure starts at $t=0$ and ends at $t=217$ s. In the examples shown the surrogate-based algorithm indicates up to four clusters, whereas the eigenvector-based one finds maximally three clusters. The best agreement is found in the late ictal period [Figs. 9(e) and 9(f)], which is immediately followed by a discrepancy $K^{\mathrm{L}}_{\mathrm{surr}}=2$ vs $K^{\mathrm{L}}_{\mathrm{ev}}=1$ in the early post-ictal period [Figs. 9(g) and 9(h)]. However, in both cases the cluster number reduces from the late ictal to the post-ictal period as several clusters merge into larger ones also attracting previously uninvolved channels. From a neurophysiological point of view, the latter finding would be consistent with the recently proposed hypothesis that one mechanism for seizure termination is the build up of a spatiotemporally extended refractory zone that blocks the further spread of ictal activity. Such a refractory zone could, for example, be established by a large cluster of neurons firing highly correlated and with slow periodicity becoming simultaneously unexcitable [25,29].

Applying the eigenvector-based algorithm to estimate the cluster number, several seconds before seizure onset [Fig.

9(b)] only one cluster is found. Allocated contacts are foramen ovale electrodes and several contacts of the right temporal anterior strip electrode (TAR), i.e., the electrode that covers the pole of the right temporal lobe. This finding is neurophysiologically plausible because they are closest to the seizure onset zone probed by contact TAR1 and are therefore expected to be entrained by ictal activity early during the seizure or even before seizure activity may be delineated visually. Different from the eigenvector-based algorithm the surrogate-based one detects four clusters [Fig. 9(a)]. Three of them can be seen as a decomposition of the cluster found by the eigenvector-based algorithm. Contacts TAR1 (red square), TAR4 and TAR5 (yellow diamonds), and TAR6 to TAR8 (green triangles) belong to different clusters, though they are located on the same strip electrode. This reveals that spatial closeness does not necessarily imply involvement in the same cluster. The reason for the difference might be that the strip electrode TAR is wrapped around the lateral temporal lobe and the involved contacts record from cortical areas with different functionality. In addition, the surrogate-based algorithm finds a cluster involving contacts of the spatially relatively distant electrode FBR (blue circles), which records from the basal part of the right frontal lobe. This cluster is not found with the eigenvector-based algorithm.

During seizure [Figs. 9(c)–9(f)], typically more contacts become allocated to the correlation clusters. Using the eigenvector-based algorithm, in the early phase of the seizure [Fig. 9(d)], the frontal strip electrodes (FPR and FBR) are involved in the same cluster as the foramen ovale electrodes and the contacts TAR1 and TAR2. Again, the surrogate-based algorithm [Fig. 9(c)] reveals a more detailed result by separating the contacts on the frontal, the temporal, and the occipital electrodes into three distinct clusters. As the clusters on the TAR and the OCR electrodes are much smaller than the frontal one, the fact that these channels are not allocated to clusters by the eigenvector-based algorithm is in line with the findings of Sec. V, compare discussion of Fig. 6(b).

Before seizure termination [Figs. 9(e) and 9(f)] both algorithms find the same cluster number $K^{\mathrm{L}}_{\mathrm{surr}}=K^{\mathrm{L}}_{\mathrm{ev}}=3$ and also the cluster pattern obtained by the CPV algorithm is very similar. One cluster comprises the foramen ovale electrodes (red squares) and a second one the contacts of the temporal anterior electrode (yellow diamonds). Interestingly, the contact TAR1, recording directly from the seizure onset zone is no longer involved in any cluster in this seizure phase. A third cluster comprises the contacts of the occipital central electrode (blue circles) and contains the only difference between both algorithms for the cluster number. It can be attributed to the probabilistic aspect of the maximization procedure of the CPV algorithm, see [6,7].

In the time period early after seizure termination [Figs. 9(g) and 9(h)], most channels are allocated to correlation clusters. Again, TAR1 is not part of one of the clusters. The surrogate-based algorithm [Fig. 9(g)] finds two clusters, the larger one comprising frontal, most temporal anterior, and foramen ovale electrodes (red squares). In the smaller one, contacts of the temporal anterior and the occipital electrodes are involved (blue circles). Different to most of the panels in the upper row, here, the contribution of relatively distant contacts to joint clusters suggests that functional similarities rather than spatial closeness is relevant for involvement in the same cluster.

In contrast to the surrogate-based algorithm the eigenvector-based one [Fig. 9(h)] obtains only one global cluster to which the majority of contacts contribute. To investigate the reason we studied the size of the absolute intra- and inter-cluster elements of the equal-time cross-correlation matrix. The intracluster elements of the cluster symbolized by red squares are 0.45 ± 0.19 and 0.50 ± 0.22 in the one corresponding to blue circles (given are mean and standard deviation). The average matrix elements between these clusters are relatively large, too: 0.28 ± 0.22. This suggests that the surrogate-based algorithm treats this situation as distinct but interrelated clusters, whereas the eigenvector-based one interprets the result as a single cluster with substructure. A similar situation is given for the clusters of the TAR and the foramen ovale electrodes of Fig. 9(a).

As already mentioned, the observed formation of clusters and the subsequent evolution of cluster patterns towards a strongly and globally correlated state after seizure termination via cluster reorganization is consistent with recent findings of [25,29]. More detailed neurophysiological interpretations are subject of ongoing research with a larger number of EEG recordings and go beyond the scope of the present paper.

VII. SUMMARY AND CONCLUSIONS

In the present paper two algorithms for estimating the cluster number K in multichannel data have been investigated. Different to our previous work [6,7] but similar to the algorithms presented in [4,5], the algorithms treat the estimation of K on the same footing as the allocation of data channels to interrelation clusters, which is usually done in a second step. Moreover, the methods go without artificially defining any threshold. Rather, data-driven criteria based on a chosen significance level α for nonparametric statistical tests are used exclusively. This makes both methods candidates for unsupervised data clustering.

Using a model system with power spectra realistic for interictal EEG, in large parts of the scanned parameter space the surrogate-based algorithm presented in Sec. III and the eigenvector-based algorithm of Sec. IV perform comparably well. The occasionally better performance of the surrogate-based algorithm, which appears more robust for strongly asymmetric cluster sizes and small bulk sizes (note that the eigenvector-based algorithm is not applicable if the bulk is empty), is contrasted by two deficiencies, the first one consisting of the long computation time for generating IAAFT surrogate ensembles. The eigenvector-based algorithm is computationally faster by orders of magnitude.

In principle, the generation of surrogate data is a delicate issue, and, depending on the data, the results might not reflect the desired properties, see, e.g., [16]. Therefore we repeated the computations of Sec. V using the much faster method of simple shift surrogates as proposed in [30]. This method conserves power spectra and amplitude distributions

exactly, while destroying cross correlations by independently time shifting the data sets relative to each other and wrapping the extra values around to the beginning of the data set. For the studied data the obtained results differ only marginally from those included in the present paper. Another method of speeding up the surrogate generation process consists of using amplitude adjusted Fourier transform surrogates (AAFT) [31] that omit the costly iteration procedure. However, for any particular class of real world data it must be carefully checked which algorithm for surrogate generation is appropriate. Note that IAAFT surrogates might not always be the optimal solution. For instance, if the experimental data show strong quasiperiodic characteristics, like, e.g., in electrocardiograms or even in EEG of primary generalized seizures with spike and wave activity, it might be more adequate to use the approach proposed in [32].

The second deficiency of the surrogate-based algorithm of Sec. III is its principal restriction to linear interrelation measures. In contrast, application of the eigenvector-based algorithm to symmetric and normalizable nonlinear interrelation measures such as, e.g., mean phase coherence [33,34] or mutual information [35–37] is possible.

In application to equal-time cross-correlation matrices computed from intracranial EEG recordings, the cluster number estimated by both algorithms may differ. However, the cluster patterns revealed by both methods show physiologically plausible characteristics consistent with recent hypotheses about mechanisms for seizure termination [25,29]. In contrast to the eigenvalue-based algorithm the surrogate-based one allows some insight into the clusters' "fine structure."

A problem that our estimation of the cluster number shares with most other approaches to data clustering is that it starts from a model assumption. Obviously, every sharply defined block in the interrelation matrix **C** produces one repelled large eigenvalue and the eigenvector components are confined to the corresponding subspace of the M-dimensional phase space. However, at present our strategies for cluster search invert this reasoning without proving that this is justified: The surrogate-based algorithm defines an interrelation cluster for every large eigenvalue λ_k that is not compatible with the surrogates. Similarly, the eigenvector-based algorithm partitions the distance matrix **D** in an optimal way and defines a cluster for every eigenvector in region L. Although we could show that both approaches perform reasonably well in general, it remains to be investigated which alternative patterns (rings, trees, etc.) are also detected by the methods of the present paper.

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their constructive remarks which helped to improve the paper considerably. The authors thank Gerold Baier, Carsten Allefeld, Stephan Bialonski, and Klaus Lehnertz for fruitful discussions. The work was supported by Deutsche Forschungsgemeinschaft, Germany (Grant No. RU 1401/1-2), CONACyT, Mexico (Project No. 48500), and Novartis Jubiäumsstiftung, Switzerland.

[1] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[2] A. Jain, M. Murty, and P. Flynn, ACM Comput. Surv. **31**, 264 (1999).
[3] J. Kogan, *Introduction to Clustering Large and High-Dimensional Data* (Cambridge University Press, Cambridge, England, 2007).
[4] B. J. Frey and D. Dueck, Science **315**, 972 (2007).
[5] C. Allefeld and S. Bialonski, Phys. Rev. E **76**, 066207 (2007).
[6] C. Rummel, G. Baier, and M. Müller, Europhys. Lett. **80**, 68004 (2007).
[7] C. Rummel, Phys. Rev. E **77**, 016708 (2008).
[8] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley, Phys. Rev. E **64**, 035106(R) (2001).
[9] V. Plerou, P. Gopikrishnan, B. Rosenow, Luis A. Nunes Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[10] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).
[11] C. Allefeld, M. Müller, and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 3493 (2007).
[12] S. Bialonski and K. Lehnertz, Phys. Rev. E **74**, 051909 (2006).
[13] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[14] M. Müller, G. Baier, C. Rummel, and K. Schindler, Europhys. Lett. **84**, 10009 (2008).
[15] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[16] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).
[17] T. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 2003).
[18] D. F. Morrison, *Multivariate Statistical Methods*, 4th ed. (Thomson, Belmont, 2005).
[19] X. Li, D. Cui, P. Jiruska, J. E. Fox, X. Yao, and J. G. R. Jefferys, J. Neurophysiol. **98**, 3341 (2007).
[20] S. Holm, Scand. J. Stat. **6**, 65 (1979).
[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, England 1992).
[22] T. Schreiber, Phys. Rev. Lett. **80**, 2105 (1998).
[23] E. F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry* (Dover, New York, 1986).
[24] S. Siegel, *Non-Parametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, 1956).
[25] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, Brain **130**, 65 (2007).
[26] A. Hedge, D. Erdogmus, D. S. Shiau, J. C. Principe, and C. J. Sackellares, Computational Intelligence Neurosci. **2007**, 83416 (2007).
[27] G. J. Ortega, L. Menendez de la Prida, R. G. Sola, and J. Pastor, Epilepsia **49**, 269 (2008).

[28] C. Rummel, G. Baier, and M. Müller, J. Neurosci. Methods **166**, 138 (2007).
[29] K. Schindler, C. E. Elger, and K. Lehnertz, Clin. Neurophysiol. **118**, 1955 (2007).
[30] T. Netoff and S. Schiff, J. Neurosci. **22**, 7297 (2002).
[31] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer, Physica D **58**, 77 (1992).
[32] M. Small, D. Yu, and R. G. Harrison, Phys. Rev. Lett. **87**, 188101 (2001).
[33] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, Physica D **144**, 358 (2000).
[34] F. Mormann, R. G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. E. Elger, and K. Lehnertz, Phys. Rev. E **67**, 021912 (2003).
[35] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Phys. Rev. E **65**, 041903 (2002).
[36] A. Kraskov, H. Stögbauer, and P. Grassberger, Phys. Rev. E **69**, 066138 (2004).
[37] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger, Europhys. Lett. **70**, 278 (2005).