LETTER

Reconstructing multi-mode networks from multivariate time series

View the article online for updates and enhancements.

Reconstructing multi-mode networks from multivariate time series

Zhong-Ke Gao[1], Yu-Xuan Yang[1], Wei-Dong Dang[1], Qing Cai[1], Zhen Wang[2], Norbert Marwan[3],
Stefano Boccaletti[4] and Jürgen Kurths[3]

[1] *School of Electrical and Information Engineering, Tianjin University - Tianjin 300072, China*
[2] *Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University - Xi'an 710072, China*
[3] *Potsdam Institute for Climate Impact Research - Telegraphenberg A31, 14473 Potsdam, Germany*
[4] *Institute of Complex Systems - Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy*

Abstract – Unveiling the dynamics hidden in multivariate time series is a task of the utmost importance in a broad variety of areas in physics. We here propose a method that leads to the construction of a novel functional network, a multi-mode weighted graph combined with an empirical mode decomposition, and to the realization of multi-information fusion of multivariate time series. The method is illustrated in a couple of successful applications (a multi-phase flow and an epileptic electro-encephalogram), which demonstrate its powerfulness in revealing the dynamical behaviors underlying the transitions of different flow patterns, and enabling to differentiate brain states of seizure and non-seizure.

Introduction. – Characterizing dynamical processes in complex systems from observed time series is a vitally important, yet intractable, issue in a broad range of research areas. Various solutions sprung up (such as fractal analysis [1], power spectra [2], recurrence plot [3], de-trended cross-correlation analysis [4] and decomposed transfer entropy [5]), but the increase of system complexity makes it largely difficult to depict the dynamical behavior from time series, and conventional methods encountered resistance to produce reliable results.

Regarding each system component as a node and determining the edges in terms of units' interactions pave a way for mapping from a system to a graph [6–11]. Thereafter, complex network theory can be used to investigate the qualities of the system. Several effective methodologies have been proposed to map a univariate/multivariate time series into a complex graph [12–21] and have been applied to various areas like climate dynamics [22,23], epilepsy diagnosis [24], rainfall prediction [25], thermo-acoustic instability detection [26] and two-phase flow analysis [27–31].

In this letter, we put forward a novel powerful functional network, namely a multi-mode weighted network, by combining multi-mode conversion with recurrence networks to focus on the analysis of high-dimensional time series. The key lies in delineating pair-wise system element interrelationships by searching for the times when both of their trajectories recur simultaneously, *i.e.*, the occurrence of joint recurrences. In particular, we first perform a multivariate empirical mode decomposition (MEMD) [32] on each high-dimensional time series to obtain several multivariate intrinsic mode functions (IMFs) corresponding to different modes. Next, for each targeted IMFs (at each targeted mode), we infer a weighted recurrence network to form a multi-mode weighted graph. The method is then applied to measured data from fluid dynamics and neuroscience, which validate its effectiveness. Finally, comparisons with traditional functional networks will also be discussed.

Methodology. – We start by introducing the MEMD method. The past decades have witnessed a rapid development of traditional time-frequency analysis methods, from Fourier transform to wavelet approaches. Such traditional methods are affected by intrinsic limitations: standard Fourier methods project data onto fixed functions and are then unable to properly analyze short and intermittent real-world data, while the employment of integral transforms in typical time-frequency analysis harms the analytic signal representation resulting from the blurring of the notion of time. These and other limitations triggered the appearance of the empirical mode decomposition (EMD) method, which holds the advantages of

adaptivity, enhanced accuracy and integrity due to the fact that its bases are data-driven rather than fixed. However, channel-wise EMD analysis may generate IMFs with different numbers, and same-index IMFs are not necessarily containing equal scales. The latter problem led to the emergence of MEMD [32]. MEMD is capable of containing mode-aligned intrinsic joint rotational modes, which is achieved by analyzing the data in a p-dimensional domain, and performing uniform sampling for the projection vector directions. Therefore, the p channels of data have the same number of mode-aligned IMFs, including the same rotational modes. In addition, MEMD allows the occurrence of mixing modes, which makes it applicable for coherent multivariate time-frequency analysis. The MEMD procedure is as follows: given a p-dimensional time series $x(t)$, 1) we first obtain a uniform sampling of V points on a p-dimensional sphere to form direction angles θ_v, where $v = 1, 2, \ldots, V$, which are used to generate projection vectors s_{θ_v} (for this a low-discrepancy Hammersley sequence namely quasi-Monte-Carlo sampling is performed) [33]); 2) we project the p-dimensional time series $x(t)$ along the direction vector s_{θ_v}, where $v = 1, 2, \ldots, V$, so that the set of projections $\{q_{\theta_v}(t)\}_{v=1}^{V}$ can be achieved; 3) we find the time instants $\{t_{\theta_v}^i\}_{v=1}^{V}$ at which the set of the projected signals $\{q_{\theta_v}(t)\}_{v=1}^{V}$ reach the local extremum, and then interpolates $[t_{\theta_v}^i, x(t_{\theta_v}^i)]$ via cubic splines to acquire the multivariate envelope curves $\{e_{\theta_v}(t)\}_{v=1}^{V}$; 4) we subtract the mean of the V multidimensional envelopes $\mathrm{m}(t)$ from $x(t)$ to gain the detail $\mathrm{d}(t)$, where $\mathrm{m}(t) = \frac{1}{V} \sum_{v=1}^{V} e_{\theta_v}(t)$; 5) if $\mathrm{d}(t)$ satisfies the stopping criteria described in [32] for a multivariate IMFs, then $x(t) = x(t) - \mathrm{d}(t)$ otherwise $x(t) = \mathrm{d}(t)$. With the new $x(t)$, the procedure goes back to step 1). Eventually, a multivariate time series is converted into several multivariate IMFs corresponding to different frequency bands. The quasi-dyadic filterbank property of MEMD and the ability in producing mode-aligned IMFs greatly facilitate us to pick out the targeted frequency bands and focus on the analysis of those targeted modes of highest significance.

The weighted recurrence network at each targeted multivariate IMFs is then constructed. In order to capture synchronization, here the joint recurrence plot (JRP) [34] technique is considered. The construction of the weighted recurrence network can be described as follows. For a multi-channel signal $\{x_{k,l}\}_{l=1}^{L}, k = 1, 2, \ldots, p$ which contains p sub-signals of equal length L, the phase-space reconstruction is first conducted channel-wise (by choosing a suitable embedded dimension m and time delay τ by the False Nearest Neighbors algorithm [35] and C-C method [36], respectively) as $\vec{x}_k(t) = (x_{k,t}, x_{k,t+\tau}, \ldots, x_{k,t+(m-1)\tau}), t = 1, 2, \ldots, N$, where N is the number of vector points in the reconstructed trajectory. For any generated trajectory $\vec{x}_m(i)$, a $N \times N$ recurrence plot (RP) can be obtained

$$RP_{i,j}^{\vec{x}_m}(\varepsilon^{\vec{x}_m}) = \Theta(\varepsilon^{\vec{x}_m} - \|\vec{x}_m(i) - \vec{x}_m(j)\|),$$
$$i = 1, \ldots, N; \quad j = 1, \ldots, N, \quad (1)$$

where ε is a pre-chosen threshold resulting from fixing the recurrence rate (a measure of the density of recurrence points in the recurrence plot (RP)) to 0.1, and $\| \cdot \|$ is the maximum norm. Thus, for a multi-channel signal $\{x_{k,l}\}_{l=1}^{L}, k = 1, 2, \ldots, p$ containing a p sub-signal, the number of obtained recurrence plots is p. For pair-wise RPs, the joint recurrence plot (JRP) is

$$JRP_{i,j}^{\vec{x}_m, \vec{x}_n}(\varepsilon^{\vec{x}_m}, \varepsilon^{\vec{x}_n}) = R_{i,j}^{\vec{x}_m}(\varepsilon^{\vec{x}_m}) R_{i,j}^{\vec{x}_n}(\varepsilon^{\vec{x}_n}). \quad (2)$$

In what follows, the joint recurrence rate (JRR) [37] is used to quantify the density of recurrence points in each JRP, *i.e.*, $JRR(\vec{x}_m, \vec{x}_n) = \frac{1}{N^2} \sum_{i,j=1}^{N} JRP_{i,j}^{\vec{x}_m, \vec{x}_n}$. Then, we characterize the synchronization of pair-wise time series by $S(\vec{x}_m, \vec{x}_n) = \frac{JRR(\vec{x}_m, \vec{x}_n)}{RR}$, where RR represents the recurrence rate of each recurrence plot, which is 0.1. Thus, for a multi-channel signal $\{x_{k,l}\}_{l=1}^{L}, k = 1, 2, \ldots, p$, we eventually obtain a synchronization matrix $S(\vec{x}_m, \vec{x}_n)$ of size $p \times p$. By means of these calculations, we infer a weighted recurrence network by regarding each sub-signal of the multivariate IMFs as a node and deeming the value of the synchronization index $S(\vec{x}_m, \vec{x}_n)$ as the weight of the link between nodes m and n. Thus, the novel functional network (namely the multi-mode weighted network) is formed, and consists of several weighted networks calculated at each targeted mode. The schematic diagram of our method is described in fig. 1.

The application in a two-phase flow system. – Validation of the performance and effectiveness of the method is provided with two applications in real systems. The first is an oil-water two-phase flow, which is frequently encountered in many industrial fields. Particularly, the combination of different total flow velocities and water cut K_w (identified as different flow conditions) gives rise to the formation of three typical flow patterns, *i.e.*, oil-in-water slug flow (D OS/W), oil-in-water bubble flow (D O/W) and very fine dispersed oil-in-water bubble flow (VFD O/W) [38]. Recognition of the transition between flow patterns, and characterization of the underlying dynamical behaviors are extremely important, especially for the enhancement of oil recovery. Our method identifies the flow patterns in the complicated oil-water two-phase flow system and detect the underlying dynamical behaviors during the transitions. We use tap-water and No. 3 white oil, with a density of $856 \, \mathrm{kg/m^3}$ and a viscosity of $11.984 \, \mathrm{mPa \cdot s}$, to conduct a small-diameter vertical oil-water two-phase flow experiment, where a high-speed cycle motivation conductance sensor is designed to record the conductivity of the oil-water mixture resulting from the evolution of different flow patterns. A series of 48-channel time series is then obtained. MEMD conversion is first performed on each 48-channel time series. According to the energy distribution of each mode, we select as targeted modes those whose energy have only little difference with the original signal and whose frequency distribution is in accordance with the typical frequency bands of the three flow patterns. For each generated weighted
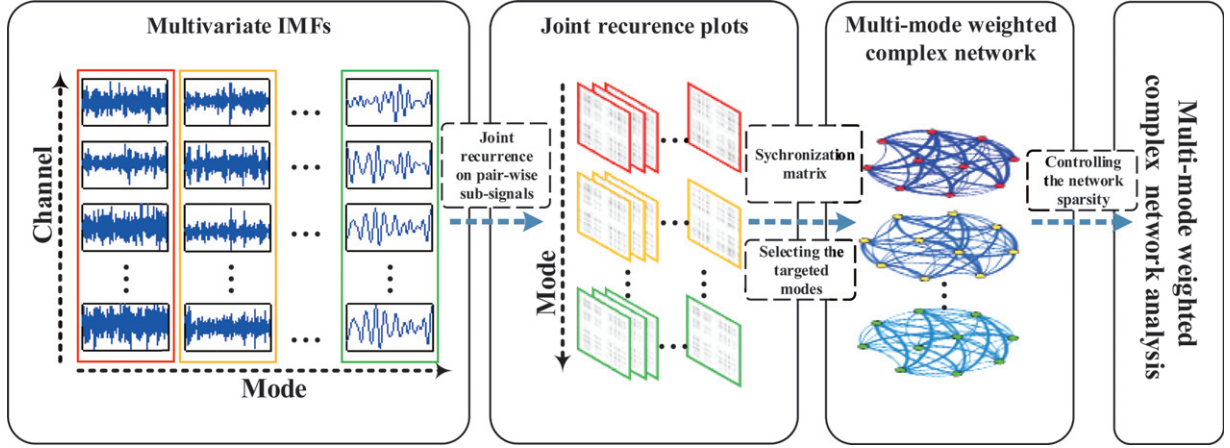
Fig. 1: (Color online) The schematic diagram of the method leading to a multi-mode weighted network representation of the original data set.
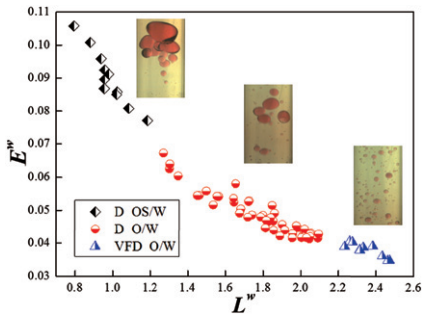

Fig. 2: (Color online) The combined graph of L^w and E^w (see text for definition). Insets are photos of the corresponding flow patterns.

Fig. 3: (Color online) The evolutional process of the two metrics during the transition among different patterns, while the water cut K_w is 86% and 97%, respectively: (a) E^w; (b) L^w; (c) E; (b) L (see text for definitions of all quantities).

recurrence network, only those edges whose weight is the highest (in the range 10%–35%) among all the links are retained, aiming at inspecting the topology of relatively sparse networks [39]. In addition, the integrals (over the sparsity range) of two network metrics are considered for two metrics: the weighted global efficiency which is defined as $E^w = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} (d_{ij}^w)^{-1}}{n-1}$, and the weighted characteristic path length which can be calculated as $L^w = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^w}{n-1}$ (both indicating the functional integration of networks). Note that, d_{ij}^w is the shortest weighted path length between nodes i and j and is decided by $d_{ij}^w = \sum_{a_{uv} \in g_i \overset{w}{\leftrightarrow} j} f(w_{uv})$, where $g_{i \overset{w}{\leftrightarrow} j}$ is the shortest weighted between i and j and f is a map from weight to length [40].

The final values of L^w and E^w of each flow condition are obtained by averaging the integrated E^w and L^w from each generated recurrence network. The combined distribution graph of these two metrics is reported in fig. 2, together with images of the different flow conditions. In addition, fig. 3(a) and (b) presents some detailed results of these two metrics, aiming at portraying the evolution of different flow patterns. Notice that sparse networks are here weighted, while traditional functional
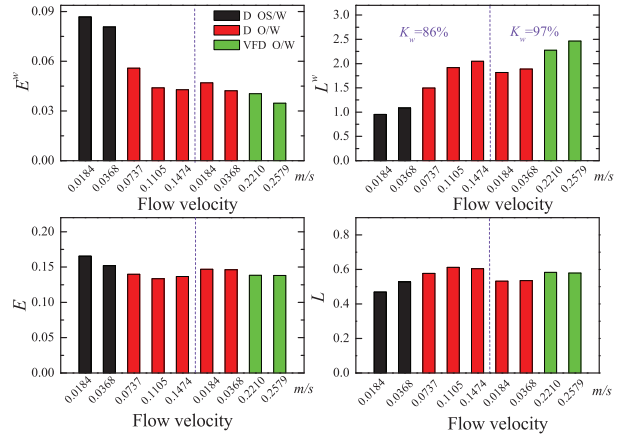
networks are commonly un-weighted. As a result, we can compare our weighted sparse networks with multi-mode un-weighted networks and calculating the integrated un-weighted global efficiency and characteristic path length (denoted as E and L, respectively) for the same flow conditions as in fig. 3(a) and (b) (the detailed results are shown in fig. 3(c) and (d)). The definitions of E and L are similar to that of E^w and L^w, but d_{ij}^w should be substituted for d_{ij} which represents the shortest path length (distance) between nodes i and j. Furthermore, the Pearson correlation coefficient (a widely used metric for functional networks) is here also used on the original data to infer a traditional functional network for the flow conditions described in fig. 3, and again the integrated E^w and L^w are calculated (the results are reported in fig. 4).

In fig. 3(a) and (b) the distribution of the two measures is significantly different in the transition among the three typical oil-water flow patterns. While the total flow
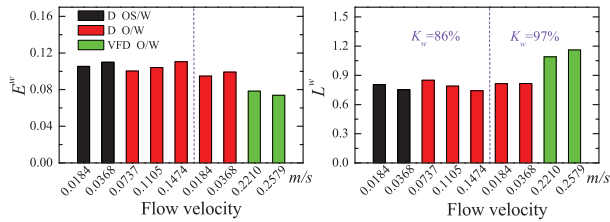
Fig. 4: (Color online) The evolutional process of the two metrics, calculated using a traditional (Pearson correlation coefficient based) functional network, during the transition among different flow patterns while the water cut K_w is 86% and 97%, respectively: (a) E^w; (b) L^w.

Fig. 5: (Color online) The p-value for each t-test: (a) C^w; (b) E_{loc}^w (see text for definitions of both quantities).

velocity is low and the oil cut is relatively high, the oil-in-water slug flow appears as an aggregation of cup-shaped oil slugs and small oil droplets moving in a water continuum quasi-periodically. The increase of the flow velocity and the turbulent kinetic energy destroys the joint movement conditions, and makes it impossible to form oil slugs which separate into relatively uniformly distributed small oil bubbles. At that moment, the movement of oil bubbles becomes stochastic, resulting in a gradual weakening of the synchronization of the system and further leading to the increase of L^w and the decrease of E^w. Figure 3(a) and (b) provides the concrete evolutional process of the two metrics, related to the transition of flow patterns, and the water cut K_w is 86% and 97%, respectively. When the oil cut is pretty low but the turbulent kinetic energy is huge, the giant crushing force triggers oil bubbles scattering into the whole tube and transforming them into tiny droplets, indicating that the very fine dispersed oil-in-water bubble flow shows up. Thus, the synchronization of the system nearly dies away and L^w possesses a growth trend while E^w changes in the opposite trend simultaneously. On the other side, as shown in fig. 3(c) and (d), the reconstruction of un-weighted network is infeasible to identify the three different flow patterns, let alone characterize the complicated evolution process. In addition, fig. 4(a) and (b) shows that the traditional approach, based on Pearson correlation analysis, is also unable to uncover the flow behavior in the transitions between different flow conditions. All these findings render our multi-mode weighted complex network extraordinarily effective for identifying flow patterns, and for characterizing the dynamical behaviors in the evolution of different flow patterns.

The application in an epileptic electro-encephalogram. – As a paroxysmal neurological disorder, epilepsy occurs with recurrent and abrupt seizures closely linked to neural synchronization [41]. Most of the studies on epilepsy which use complex networks analysis [42–44] aim at differentiating epilepsy patients from healthy controls. However, the identification of pre-seizure and seizure periods (*i.e.*, the seizure prediction task) can further improve the precaution and even the treatment of epilepsy [45,46]. We here apply our method to the identification of pre-seizure
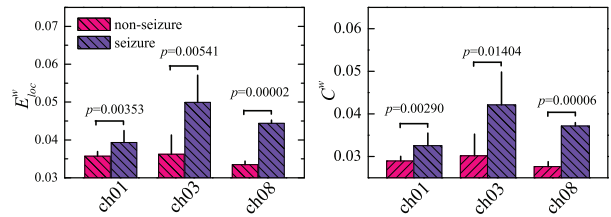
and seizure in the scalp electroencephalogram (EEG) recordings [47] of three pediatric patients (labeled as ch01, ch03, ch08), who feature 7, 7 and 5 seizure events respectively. For each one of them, we select a 20-s-long non-seizure recording immediately prior to the seizure onset (the pre-seizure recording), and a 20-s-long seizure recording. After the MEMD procedure, we choose mode 2 to 8 as our targeted modes (corresponding to the 1–48 Hz broad frequency band), and construct the multi-mode weighted complex networks. Thereafter, we calculate the integrated weighted local efficiency and clustering coefficient (representing the efficiency of local communication and the clustered connectivity in networks, respectively) of each weighted recurrence network and then average the values of all of them as the final results. Besides, the weighted local efficiency is defined as $E_{loc}^w = \frac{1}{2n} \sum_{i \in N} \frac{\sum_{j,h \in N, j \neq i}(w_{ij}w_{ih}[d_{jh}^w(N_i)]^{-1})^{1/3}}{k_i(k_i - 1)}$, where $d_{jh}^w(N_i)$ is the length of the shortest path between nodes j and h that contains only neighbors of node i, and k_i is the degree of node i [40]. And the weighted clustering coefficient is defined as $C^w = \frac{1}{n} \sum_{i \in N} \frac{2t_i^w}{k_i(k_i - 1)}$, where $t_i^w = \frac{1}{2} \sum_{j,h \in N}(w_{ij}w_{ih}w_{jh})^{1/3}$ is the geometric mean of triangles around node i and w_{ij}, w_{ih} and w_{jh} represent the weight between node i and j, i and h, j and h, respectively [40].

A t-test is performed between non-seizure and seizure results for all events of each patient. The p-value of each t-test is shown in fig. 5, where each p-value is much less than 0.05, reflecting the excellent performance of the two metrics on discriminating between non-seizure and seizure periods. C^w presents a growth trend, namely, the clustering of synchronization among different brain regions experiences a sharp change due to the onset of seizure. In addition, the increase of the integrated E_{loc}^w implies that during the seizure period a well-pronounced local synchronization behavior appears and compared to the non-seizure period, its efficiency among different brain regions is stronger. Once again, all these findings validate that our method possesses the capacity of exploring the dynamical properties underlying seizure processes, and is able to provide a more accurate seizure prediction.

Conclusions. – In summary, our method allows realizing multi-information fusion. Successful utilization in vertical oil-water two-phase flow experiments demonstrates

that the reconstructed network enables to distinguish between typical flow patterns, and further detects the inherent dynamical property during the transition among the different flow patterns. Furthermore, the results also illustrate that our technique performs much better than un-weighted networks and traditional approaches based on Pearson correlation analysis. On the other hand, a strong performance in patient-specific classification of non-seizure and seizure periods paves the way for the utilization of our method in the prediction task of seizure onset. Our method is however not limited to the above applications, but can make contributions in all cases in which addressing multi-information fusion challenges is needed, which is generally encountered in many data-intensive research areas like smart power grids and cities, as well as intelligent transport and health care.

$$* * *$$

REFERENCES

[1] PENG C. K. *et al.*, *Phys. Rev. E*, **49** (1994) 1685.
[2] DAW C. S. *et al.*, *Phys. Rev. Lett.*, **75** (1995) 2308.
[3] MARWAN N. *et al.*, *Phys. Rep.*, **438** (2007) 237.
[4] PODOBNIK B. and STANLEY H. E., *Phys. Rev. Lett.*, **100** (2008) 084102.
[5] RUNGE J. *et al.*, *Phys. Rev. Lett.*, **108** (2012) 258701.
[6] LIU Y. Y., SLOTINEAND J. J. and BARABSI A. L., *Nature*, **473** (2011) 167.
[7] GAO J. X. *et al.*, *Nat. Phys.*, **8** (2012) 40.
[8] ZOU Y. *et al.*, *Phys. Rev. Lett.*, **112** (2014) 114102.
[9] BOCCALETTI S. *et al.*, *Phys. Rep.*, **544** (2014) 1.
[10] GAO J. X., BARZEL B. and BARABSI A. L., *Nature*, **530** (2016) 167.
[11] WANG W. X., LAI Y. C. and GREBOGI C., *Phys. Rep.*, **644** (2016) 1.
[12] ZHANG J. and SMALL M., *Phys. Rev. Lett.*, **96** (2006) 238701.
[13] LACASA L. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008) 4972.
[14] MARWAN N. *et al.*, *Phys. Lett. A*, **373** (2009) 4246.
[15] DONNER R. V. *et al.*, *New J. Phys.*, **12** (2010) 033025.
[16] DONNER R. V. *et al.*, *Int. J. Bifurcat. Chaos*, **21** (2011) 1019.
[17] WANG W. X. *et al.*, *Phys. Rev. Lett.*, **106** (2011) 154101.
[18] HEMPEL S. *et al.*, *Phys. Rev. Lett.*, **107** (2011) 054101.
[19] XIANG R. X. *et al.*, *Chaos*, **22** (2012) 013107.
[20] GAO Z. K., SMALL M. and KURTHS J., *EPL*, **16** (2016) 50001.
[21] GOSAK M. *et al.*, to be published in *Phys. Life Rev.* (2017), `https://doi.org/10.1016/j.plrev.2017.11.003`.
[22] DONGES J. F. *et al.*, *EPL*, **87** (2009) 48007.
[23] DONGES J. F. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **108** (2011) 20422.
[24] CHAVEZ M. *et al.*, *Phys. Rev. Lett.*, **104** (2010) 118701.
[25] BOERS N. *et al.*, *Nat. Commun.*, **5** (2014) 5199.
[26] MURUGESAN M. and SUJITH R. I., *J. Fluid Mech.*, **772** (2015) 225.
[27] GAO Z. K. *et al.*, *EPL*, **109** (2015) 30005.
[28] GAO Z. K. *et al.*, *Sci. Rep.*, **5** (2015) 8222.
[29] MOSDORF R. and GRSKI G., *Int. Commun. Heat Mass Transf.*, **64** (2015) 14.
[30] GAO Z. K. *et al.*, *Int. J. Bifurcat. Chaos*, **27** (2017) 1750059.
[31] GAO Z. K. *et al.*, *Int. J. Bifurcat. Chaos*, **27** (2017) 1750123.
[32] MANDIC D. P. *et al.*, *IEEE Signal Process. Mag.*, **30** (2013) 74.
[33] CUI J. and FREEDEN W., *SIAM J. Sci. Comput.*, **18** (1997) 595.
[34] ROMANO M. C. *et al.*, *Phys. Lett. A*, **330** (2004) 214.
[35] KENNEL M. B., BROWN R. and ABARBANEL H. D. I., *Phys. Rev. A*, **45** (1992) 3403.
[36] KIM H. S., EYKHOLT R. and SALAS J. D., *Physica D*, **127** (1999) 48.
[37] SCHINKEL S. *et al.*, *J. Neurosci. Methods*, **197** (2011) 333.
[38] GAO Z. K. *et al.*, *IEEE Trans. Instrum. Meas.*, **65** (2016) 1690.
[39] ACHARD S. and BULLMORE E., *PLoS Comput. Biol.*, **3** (2007) 174.
[40] RUBINOV M. and SPORNS O., *NeuroImage*, **52** (2010) 1059.
[41] LEHNERTZ K. *et al.*, *J. Neurosci. Methods*, **183** (2009) 42.
[42] HORSTMANN H. T. *et al.*, *Clin. Neurophysiol.*, **121** (2010) 172.
[43] HANEEF Z. *et al.*, *Epilepsy Behav.*, **46** (2015) 227.
[44] GAO Z. K. *et al.*, *Int. J. Neural Syst.*, **27** (2017) 1750005.
[45] NGAMGA E. J. *et al.*, *Phys. Lett. A*, **380** (2016) 1419.
[46] FEI K. L. *et al.*, *Neurocomputing*, **249** (2017) 290.
[47] SHOEB A. *et al.*, *Epilepsy Behav.*, **5** (2004) 483.