Normalized linear variance decay dimension density and its application of dynamical complexity detection in physiological (fMRI) time series

Xiaoping Xie\textsuperscript{a}, Xiaohu Zhao\textsuperscript{b}, Youtong Fang\textsuperscript{c}, Zhitong Cao\textsuperscript{a}, Guoguang He\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} Department of Physics, Zhejiang University, Hangzhou 310027, People’s Republic of China
\textsuperscript{b} Department of Radiology, Tongji Hospital, Tongji University, Shanghai 200065, People’s Republic of China
\textsuperscript{c} Department of Electrical Engineering, Zhejiang University, Hangzhou 310027, People’s Republic of China

\textbf{A B S T R A C T}

The upper and lower bounds of the linear variance decay (LVD) dimension density are analytically deduced using multivariate series with uncorrelated and perfectly correlated component series. Then, the normalized LVD dimension density ($\delta_{\text{normLVD}}$) is introduced. In order to measure the complexity of a scalar series with $\delta_{\text{normLVD}}$, a pseudo-multivariate series was constructed from the scalar time series using time-delay embedding. Thus, $\delta_{\text{normLVD}}$ is used to characterize the complexity of the pseudo-multivariate series. The results from the model systems and fMRI data of anxiety subjects reveal that this method can be used to analyze short and noisy time series.

\begin{center}
\textbullet\textsuperscript{1} Corresponding author. Tel.: +86 57187953227.
E-mail address: gghe@zju.edu.cn (G. He).
\end{center}

1. Introduction

Since the breakthrough of nonlinear dynamics around the end of the seventies of the 20th century, the analysis of time series in terms of nonlinear dynamics has been applied tentatively to a wide variety of experimental and observational data in physics, engineering, biomedicine, psychology, economics, meteorology, and other fields. A number of nonlinear techniques, such as dimensions \cite{1-5}, entropies \cite{6-8}, Lyapunov exponents \cite{9-11}, and Recurrence quantification analysis \cite{12-23}, have been utilized to measure the complexity of real-world time series, but in order to obtain statistically significant results, long data sets are generally required \cite{1-3}. As they are applied to short and noisy time series \cite{12} in physiology, psychology, and biomedicine, some difficulties will be encountered, especially in voxel-based analysis of functional magnetic resonance imaging (fMRI) data sets, which typically consist of a hundred or more 3D images \cite{18}.

An ingenious method of resolving this problem is to obtain large enough data sets with the aid of spatial information \cite{19-22}, i.e., with spatiotemporal data or multivariate series. Several nonlinear methods originally used in univariate series have been generalized to analyze multivariate data, including the scaling of fractal dimensions \cite{23}, dimension densities \cite{24-26}, and the Lyapunov spectrum \cite{27-29}. In addition, there are also several methods based on the decomposition of multivariate series. A multivariate series can be statistically decomposed by either purely linear methods such as Principal Component Analysis (PCA) \cite{2}, Karhunen–Loève decomposition (KLD) \cite{30,31} and wavelet decomposition \cite{32}, or nonlinear approaches such as Nonlinear Principal Component Analysis (NLPCA) \cite{33}, Locally Linear Embedding (LLE) \cite{34}, and Independent Component Analysis (ICA) \cite{35}. However, many difficulties may be encountered in extending these methods to short and noisy scalar time series analysis.

Recently, linear variance decay (LVD) dimension density $\delta_{\text{LVD}}$ \cite{36}, which characterizes the scaling of the component variances of multivariate data sets, has been introduced to measure the complexity of interrelationships among the components of short and noisy multivariate data sets. In this work, the upper and lower bounds of $\delta_{\text{LVD}}$ are analytically deduced using multivariate series for two extreme cases (uncorrelated and perfectly correlated component series). Thus, the normalized LVD dimension density $\delta_{\text{normLVD}}$ is introduced by means of the upper and lower bounds of the LVD dimension density. In order to employ $\delta_{\text{normLVD}}$ to analyze a short and noisy scalar time series and detect the dynamical complexity of the scalar time series, a pseudo-multivariate series was constructed from the univariate time series with time-delay embedding. $\delta_{\text{normLVD}}$ for multivariate data sets was extended to measure the complex-
ity of the constructed pseudo-multivariate series. This procedure is equivalent to the dimension estimates based on singular system analysis [36]. This is equivalent to measuring the complexity of the original univariate time series. Thus, $\delta_{\text{normLVD}}$ is used to detect dynamical complexity of the scalar time series generated by the logistic map and the Lorenz system. Furthermore, $\delta_{\text{normLVD}}$ is also used for voxel-based analysis of short-term fMRI series. The results reveal that $\delta_{\text{normLVD}}$ can be used to characterize the dynamical complexity of short and noisy univariate time series.

2. Theory

2.1. The LVD dimension density for multivariate series

In the case of observations with $n$ simultaneously measured variables and $m$ time points, the $(m \times n)$-dimensional data matrix $A$, which is rescaled to zero means and unit variances for any component time series, can be obtained. Without loss of generality, suppose $m \geq n$. Thus $N = \min(m, n) = n$. According to matrix theory, any arbitrary rectangular matrix $A$ of size $m \times n$ can be decomposed in the form of Eq. (1).

$$A = U\Sigma V^T,$$

where $\Sigma$ is a diagonal matrix of the same dimensions as matrix $A$. Its $N$ nonnegative diagonal elements are called singular values and denoted by $\sigma_1, \sigma_2, \ldots, \sigma_N$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$. They are the square roots of the eigenvalues of $AA^T$ or equivalently $A^T A$. $U$ and $V$ are orthogonal matrices of sizes $m \times m$ and $n \times n$ respectively. Representation in the form of Eq. (1) is called singular value decomposition (SVD) [37,38] of matrix $A$.

In order to characterize the complexity of the multivariate series, the Karhunen–Loève decomposition (KLD) dimension [30,31,36], a positive integer $m_{\text{KLD}}$, has been introduced using these singular values of matrix $A$.

$$D_{\text{KLD}}(f) = \min \left\{ p : \left( \sum_{i=1}^{p} \sigma_i^2 / \sum_{i=1}^{N} \sigma_i^2 \right) \geq f \right\},$$

with $0 \leq f \leq 1$. $D_{\text{KLD}}$ represents the minimum number of KLD modes needed to capture some specified fraction $f$ of the total variance $\sum_{i=1}^{N} \sigma_i^2$. $D_{\text{KLD}}$ has been suggested as a fractal dimension to measure the complexity of spatiotemporal data [30].

Simulations of spatiotemporally chaotic systems show that $D_{\text{KLD}}$ exhibits a linear scaling with the system size $n$ for any fraction $f$ [30,36]. This finding justifies the normalization of the KLD dimension and thus the KLD dimension density $\delta_{\text{KLD}} = D_{\text{KLD}}/N$ [30,31,36], which has a value within the unit interval, is introduced. Since the KLD dimension density is a step function of $f$, it is in some cases a very coarse one [36].

In order to detect the complexity of multivariate series more effectively, the LVD dimension density $\delta_{\text{LVD}}$ [36], which characterizes the exponential decay rate of the remaining variances of multivariate data sets, is introduced in the form of Eq. (3).

$$\delta_{\text{LVD}}(f) = -\frac{\int_{\log_{10}(1-f)}^{0} \frac{1}{\log_{10}(1-f)} \delta_{\text{KLD}}(x) x 10^x \, dx}{\int_{\log_{10}(1-f)}^{0} x^2 10^x \, dx},$$

where $x = \log_{10}(1-\phi)$ and $\phi \in [0, f]$.

2.2. The normalized LVD dimension density

Obviously, the denominator of Eq. (3) can be denoted by

$$\int_{\log_{10}(1-f)}^{0} x^2 10^x \, dx = \frac{1}{(\ln 10)^3} \left\{ 2 - \left[ \ln^2(1-f) - 2 \ln(1-f) + 2 \right](1-f) \right\}.$$ (4)

It is known that any $(m \times n)$-dimensional data matrix $A$ can be used to define an $(n \times n)$-dimensional symmetric and positive semi-definite matrix $S = A^T A$. Matrix $S$ can be completely described by its nonnegative eigenvalues $\sigma_i^2$ ($i = 1, \ldots, N$) and their corresponding eigenvectors [36].

The eigenvalue distribution of matrix $S$ is directly related to the degree of correlations between the components of the multivariate data set [39]. If $n$ components of the multivariate data set are all uncorrelated or orthogonal (with zero means) and each component time series has the same variance (denoted by $\sigma^2$), the non-diagonal elements of matrix $S$ tend to zero. Because each component time series of the multivariate data set is perfectly correlated with itself, all diagonal elements of matrix $S$ are identical and equal to $\sigma^2$. Thus, the eigenvalues of matrix $S$ are also identical, i.e. $\lambda_i = \sigma^2 \forall i$.

In the uncorrelated extreme (i.e. Case I), for a specified fraction $0 < \phi < 1$, there exists an integer $i$ ($1 \leq i \leq N$), such that

$$\frac{(i-1)}{N} < \phi \leq \frac{i}{N}.$$ (5)

Then, $\delta_{\text{KLD}}$ in this case can be obtained easily by

$$\delta_{\text{KLD}}(\phi) = \frac{i}{N}.$$ (6)

Similarly, for a specified fraction $0 < f < 1$, there exists also an integer $p$ ($1 \leq p \leq N$), which satisfies

$$\frac{(p-1)}{N} < f \leq \frac{p}{N}.$$ (7)

Thus, we have the numerator of Eq. (3)

$$\int_{\log_{10}(1-f)}^{0} \delta_{\text{KLD}}(x) x 10^x \, dx = \sum_{i=1}^{p-1} \int_{\log_{10}(\frac{N-i}{N})}^{\log_{10}(\frac{i}{N})} \frac{x 10^x \, dx}{\log_{10}(1-f)} + \int_{\log_{10}(\frac{p}{N})}^{\log_{10}(1-f)} \frac{x 10^x \, dx}{\log_{10}(1-f)}$$

$$= \frac{1}{N(\ln 10)^2} \left\{ \sum_{i=1}^{p-1} \left[ \ln \left( \frac{N-i}{N} \right) - 1 \right] \frac{N-i}{N} - p \left[ \ln(1-f) - 1 \right] (1-f) - 1 \right\}.$$ (8)

Finally, we can obtain the $\delta_{\text{LVD1}}$ for Case I.

$$\delta_{\text{LVD1}}(f) = -\frac{\ln 10}{N} \times \sum_{i=1}^{p-1} \left[ \ln \left( \frac{N-i}{N} \right) - 1 \right] \frac{N-i}{N} - p \left[ \ln(1-f) - 1 \right] (1-f) - 1$$

$$\times \frac{(1-f) \ln(1-f) [2 - (1-f)] + 2 f}{(1-f) \ln(1-f) [2 - (1-f)] + 2 f}.$$ (9)

On the contrary, if all $n$ component time series of the multivariate data set are perfectly correlated (i.e. Case II), all elements of matrix $S$ tend to be identical. Only one eigenvalue of matrix $S$ is nonzero, i.e. $\lambda_{\text{max}} = N\sigma^2$, and the rest of the eigenvalues are zero.
Thus, in Case II, the $\delta_{\text{KLD}}$ can be obtained for a specified fraction $0 < \phi < 1$:
\[
\delta_{\text{KLD}}(\phi) = \frac{1}{N}. \tag{10}
\]

Then, the LVD dimension density, which is denoted by $\delta_{\text{LVD2}}$ for Case II, can be derived easily and shown as Eq. (11).
\[
\delta_{\text{LVD2}}(f) = \frac{\ln 10}{N} \cdot \frac{\ln(1 - f) - f \ln(1 - f) + f}{(1 - f)\ln(1 - f)[2 - \ln(1 - f)] + 2f}. \tag{11}
\]

Because the component time series of the real multivariate data set are neither perfectly correlated nor uncorrelated, their eigenvalue spectrum is given neither by only one nonzero eigenvalue nor by completely identical eigenvalues. Suppose that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_N^2 > 0$ are $N$ eigenvalues of matrix $S$ of a real multivariate data set, where $\sum_{i=1}^{N} \sigma_i^2 = N\sigma^2$. Then, there is an integer $j$, such that $\sigma_j^2 \geq \sigma_2^2 \geq \sigma_{j+1}^2$. Obviously, for an arbitrary fraction $f$, the KLD dimension $D_{\text{KLD}}(f)$ of the real multivariate data set satisfies $D_{\text{KLD}}(f) \geq D_{\text{KLD2}}(f) = 1$, where $D_{\text{KLD2}}(f)$ is the KLD dimension for Case II.

On the other hand, for Case I, the KLD dimension $D_{\text{LVD1}}(f)$ can be estimated by $D_{\text{LVD1}}(f) = p = \min\{q: q/N \geq f\}$. If $p < j$, then $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_p^2 > \sigma_j^2$. Thus, $\sum_{i=1}^{p} \sigma_i^2 = \sigma_1^2 > \sigma_j^2$. If $p > j$, then $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_{p+1}^2 \geq \sigma_j^2$. Therefore, we have also $\sum_{i=1}^{p+1} \sigma_i^2 \geq \sum_{i=1}^{j} \sigma_i^2 = N\sigma^2 - \sum_{i=p+1}^{N} \sigma_i^2 = \sigma_j^2$. Then, we have $D_{\text{LVD1}}(f) \leq D_{\text{LVD2}}(f) = p$. Consequently, we obtain $\delta_{\text{LVD2}}(f) = 1/N \leq \delta_{\text{LVD}}(f) \leq \delta_{\text{LVD1}}(f) = p/N$.

It follows from Eq. (3) that
\[
\delta_{\text{LVD1}}(f) - \delta_{\text{LVD}}(f) = \frac{\int_{0}^{1} \frac{f(x_1)}{x_1^2} x_{10}^4 \ dx}{\int_{0}^{1} \frac{f(x_1)}{x_1^2} x_{10}^4 \ dx} \geq 0 \tag{12}
\]
and
\[
\delta_{\text{LVD}}(f) - \delta_{\text{LVD2}}(f) = \frac{\int_{0}^{1} \frac{f(x_1)}{x_1^2} x_{10}^4 \ dx}{\int_{0}^{1} \frac{f(x_1)}{x_1^2} x_{10}^4 \ dx} \geq 0. \tag{13}
\]

Finally, we have $\delta_{\text{LVD2}}(f) \leq \delta_{\text{LVD}}(f) \leq \delta_{\text{LVD1}}(f)$. This shows that $\delta_{\text{LVD1}}$ and $\delta_{\text{LVD2}}$ are the upper and lower bounds of the LVD dimension density respectively.

Thus, we can introduce the normalized LVD dimension density $\delta_{\text{normLVD}}$ for fixed $f$, $m$ and $n$
\[
\delta_{\text{normLVD}}(f) = \frac{\delta_{\text{LVD}}(f) - \delta_{\text{LVD2}}(f)}{\delta_{\text{LVD1}}(f) - \delta_{\text{LVD2}}(f)}. \tag{14}
\]

Obviously, $\delta_{\text{normLVD}}$ has a value within the interval $[0, 1]$.

2.3. The normalized LVD dimension density for univariate series

Multivariate systems can be assumed to represent either different observables of a single attractor measured simultaneously or independent realizations of the same observable in a given system [19]. In order to employ $\delta_{\text{normLVD}}$ to measure dynamical complexity of scalar time series, it is necessary to construct a pseudo-multivariate series. Here, we use time-delay embedding to construct the pseudo-multivariate series.

Let $Y = \{y_i; i = 1, 2, \ldots, T\}$ denote a scalar time series with zero mean, where $T$ represents the number of time points. A pseudo-multivariate series can be constructed from $Y$ using time-delay embedding. Thus, an $(m \times n)$-dimensional lag matrix $A$, which is a representation of the pseudo-multivariate series, can be constructed as follows:
\[
A = \begin{bmatrix}
\delta_{a11} & \delta_{a12} & \cdots & \delta_{a1n} \\
\delta_{a21} & \delta_{a22} & \cdots & \delta_{a2n} \\
\vdots & \vdots & \ddots & \vdots \\
\delta_{an1} & \delta_{an2} & \cdots & \delta_{ann}
\end{bmatrix}
= \begin{bmatrix}
y_1 & y_1 + \tau & \cdots & y_1 + (n - 1)\tau \\
y_2 & y_2 + \tau & \cdots & y_2 + (n - 1)\tau \\
\vdots & \vdots & \ddots & \vdots \\
y_m & y_m + \tau & \cdots & y_m + (n - 1)\tau
\end{bmatrix}, \tag{15}
\]
where $\tau$ is the time delay and $m + (n - 1)\tau = T$. In fact, this kind of data embedding technique [40] is widely used in the reconstruction of dynamical information from a time series.

It is known that any structure in a time series can be expressed in the form of correlations between the individual sampled values [1]. Thus, if matrix $A$ in Eq. (1) is replaced by the lag matrix in Eq. (15), i.e., the constructed pseudo-multivariate series, the normalized LVD dimension density can be extended to measure the complexity of the constructed pseudo-multivariate series. This offers the possibility to characterize the complexity of the original univariate data.

3. Results and discussions

3.1. Limiting cases

In order to numerically estimate the upper and lower bounds of the LVD dimension density, the first step was to construct the multivariate series corresponding to Case I or Case II. For the Case I, we created a number of orthonormal random data sets (i.e., $A$) with zero means, where $m = 100, 200, 300$ and $n = 32, 48, 64$ ($m > n$), respectively. Then we estimated the $\delta_{\text{LUD}}$ of the reconstructed matrix $A$ and compared them with the theoretical values calculated by Eq. (9). Fig. 1(a–c) shows the results of the comparison. As shown in Fig. 1, it can be observed that $\delta_{\text{LUD}}$ depends on the parameter $f$, which is trivial and has been previously observed and discussed for the multivariate case [36]. Within the range of fraction $f \in [0.1, 0.9999]$, differences (denoted by $\Delta_{\text{LUD}}$) between the theoretical and estimated $\delta_{\text{LUD}}$ are all less than 0.01. Furthermore, the larger the fraction $f$ is, the smaller the difference $\Delta_{\text{LUD}}$ is, which indicates that the larger the parameter $f$, the higher the accuracy of the estimated $\delta_{\text{LUD}}$. The parameter $n$ has a small influence on the results and the parameter $m$ has almost no influence on the results.

In order to construct the multivariate series corresponding to Case II, we created a number of random multivariate data sets with identical component time series (normal distributions and zero means). Then, we estimated their $\delta_{\text{LUD}}$ and compared them with the theoretical value calculated by Eq. (11). Fig. 1(d–f) show the results. It is obvious that Case II has much smaller $\delta_{\text{LUD}}$ than Case I. The rest of the results are similar to those in Case I except that the estimated $\delta_{\text{LUD}}$ is in inverse proportion to the parameter $N$ in Case II (e.g., for $f = 0.9999$, the estimated $\delta_{\text{LUD}}$ with $n = 32$, 48 and 64 are 0.036131, 0.024087 and 0.018070, respectively), which directly follows from Eq. (11).

For the multivariate data set, a typical example of Case I is the orthonormal time series with zero means. Case II corresponds to a coupled network with full synchronization. For the univariate time series, Case I corresponds to a stochastic process with zero auto-correlation at all considered lags (i.e., essentially white noise). An example of Case II is the periodic motion of period 2.
3.2. Example 1: Logistic map

Let us now study one of the simplest dynamical systems, the logistic map [13].

\[ x_{i+1} = f(x_i) = \mu x_i (1 - x_i), \]

(16)

where \( \mu \) is the control parameter. After possible initial transients have been discarded by removing the first 5000 iterations, the time series \( x(\mu) \) of the logistic map were generated by starting from \( x_0 = 0.65 \) and for different values of \( \mu \) between 3.4 and 4.0 with a step width of \( 10^{-4} \).

Fig. 2(a) shows the bifurcation diagram of the logistic map. Fig. 2(b) shows the estimated \( \delta_{\text{normLVD}} \) of the scalar time series generated from the logistic map. The parameters were \( T = 500, n = 48, \tau = 1 \) and \( f = 0.9999 \). The estimated \( \delta_{\text{normLVD}} \) with \( \tau = 3 \) is shown in Fig. 2(d). Fig. 2(c) shows the Lyapunov exponent of the logistic map. It is clear that \( \delta_{\text{normLVD}} \) and the Lyapunov exponent have a very similar profile in the region with \( 3.57 < \mu < 4 \). The above results show that both \( \delta_{\text{normLVD}} \) and the Lyapunov exponent can describe the dynamical complexity of the time series very well. Furthermore, a time delay has a small influence on the results.

In order to compare \( \delta_{\text{normLVD}} \) with other measures of complexity, the length \( l_{\text{max}} \) of the longest diagonal line found in the recurrence plot [17] was also estimated. The \( l_{\text{max}} \) measure can be used as an estimator for \( K_2 \) entropy and for the lower limit of the sum of the positive Lyapunov exponents [17]. The parameters chosen for estimating \( l_{\text{max}} \) were radius of neighborhoods \( \varepsilon = 0.3\sigma \), embedding dimension \( d = 3 \), and time delay \( \tau = 1 \). The estimated \( l_{\text{max}} \) is plotted in Fig. 2(e). \( l_{\text{max}} \) has also the similar profile to \( \delta_{\text{normLVD}} \) and the Lyapunov exponent. Thus, these measures show similar behavior regarding the control parameter \( \mu \) and can describe the dynamical complexity of the time series very well.

The \( \delta_{\text{normLVD}} \) method is also effective in the presence of noise. We superimposed Gaussian white noise to the time series of the logistic map with the signal to noise ratio \( \text{SNR} = 20 \) and \( \text{SNR} = 5 \), then estimated \( \delta_{\text{normLVD}} \) and \( l_{\text{max}} \), respectively. The parameters for \( \delta_{\text{normLVD}} \) were \( T = 500, n = 48, \tau = 1 \) and \( f = 0.97 \). The parameters for \( l_{\text{max}} \) were \( d = 3, \varepsilon = 0.3\sigma \) and \( \tau = 1 \). The results of \( \delta_{\text{normLVD}} \) and \( l_{\text{max}} \) are shown in Fig. 3. We can see from Fig. 3 that \( \delta_{\text{normLVD}} \) and \( l_{\text{max}} \) were radius of neighborhoods \( \varepsilon = 0.3\sigma \), embedding dimension \( d = 3 \), and time delay \( \tau = 1 \).
can also capture all the bifurcation points except that the noise causes a small increase of dimension value in the periodic region (3.4 < \mu < 3.57). In the range of 3.57 < \mu < 4, \delta_{normLVD} remains almost the same for SNR > 5. This demonstrates that small noise only induces a small change of \delta_{normLVD}. However, \lambda_{max} is much more easily influenced by the noise than \delta_{normLVD}.

3.3. Example 2: Lorenz system

The 3D Lorenz system [41], which serves as a low-dimensional approximate model for atmospheric convection processes, is described by three coupled nonlinear ordinary differential equations

\[
\begin{align*}
\frac{dx}{dt} &= -\sigma(x - y), \\
\frac{dy}{dt} &= r x - y - x z, \\
\frac{dz}{dt} &= -b z + x y,
\end{align*}
\]

(17)

where \(x\) is proportional to the intensity of the convection motion, \(y\) is proportional to the horizontal temperature variation, \(z\) is proportional to the vertical temperature variation, and \(\sigma, r, b\) are positive constant parameters.

After the transients had been excluded, the time series of the Lorenz system were generated by starting from \(x_0 = 1, y_0 = 1, z_0 = 1\). The parameters \(\sigma = 10, r = 28, b = 8/3\) have been kept fixed, while \(r\) is increased from 10 to 200 with a step size of 0.1.

Fig. 4(a) shows the largest Lyapunov exponent, denoted by \(\lambda_{max}\), estimated using the Wolf method [42] based on the dynamical equations of the Lorenz system. Fig. 4(b) shows the estimated \(\delta_{normLVD}\) of scalar time series (x component) generated from the Lorenz system. The parameters were \(T = 500, n = 48, \tau = 1\) and \(f = 0.9999\). The estimated \(\lambda_{max}\) (x component too) with \(\epsilon = 0.3\sigma, d = 3, \tau = 1\) is plotted in Fig. 4(c). Obviously, the results observed in the Lorenz system are analogous with those in the logistic map. It should be noted that if only time series are available instead of dynamical equations, it would never be possible to estimate the Lyapunov exponent numerically from only \(T = 500\) points with reasonable confidence.

3.4. Voxel-based fMRI data

The dysfunction in anxiety disease has been widely observed in clinical studies. Though the underlying neural mechanism of anxiety disease is not yet well understood, it has been attributed theoretically to a neuro-degenerative disorder. Recently, researches on the disease by means of functional magnetic resonance imaging (fMRI) have made noteworthy progress [21,22]. More and more methods have been introduced to analyze fMRI data in order to probe the underlying neural mechanism of certain diseases.

In neuroimaging studies, fMRI time-series is considered from a signal processing perspective with particular focus on optimal experimental design and efficiency and hence can be viewed as a linear admixture of signal and noise [43]. Signal corresponds to neurally mediated hemodynamic changes responding to changes in experimental factors. These hemodynamic changes can be modeled as a [non]linear convolution of some underlying neuronal process by a hemodynamic response function (HRF) [44]. In block designs, the design is completely specified by the times of stimulus presentation or trials. Thus, there exist deterministic components in the Blood Oxygenation Level Dependent (BOLD) signal which is modeled by neuronal causes that are expressed via HRF [43].

Ten right-handed anxiety subjects (including 7 males, mean age of 42 years) participated in this study. They had not received any psychopharmacological treatment or cognitive therapy during the two weeks prior to this study. All patients underwent noninvasive fMRI while listening actively to emotionally neutral words in Chinese alternating with no words, i.e. silence, as the control condition and to threat-related words alternating with emotionally
neutral words as the experimental condition [45]. Each word was presented in pseudo-random order in each 16s block of 12 words of the same type. Eight alternating blocks of emotionally neutral words and no (or threat-related) words were presented for about 256s.

All fMRI data were obtained with a Marconi 1.5T EDGE ECLIPSE scanner equipped with a prototype fast gradient system for echo-planner imaging. A T2* weighted, gradient recalled echo-planner imaging sequence was obtained for functional images (Slice thickness/gap = 6/1 mm, TE = 40 ms, TR = 2500 ms, Flip angle = 90°, Fov = 24 cm, Matrix size: 64 × 64). After discarding initial scans (to allow for magnetic saturation effects), each fMRI data set is comprised of 120 vol images.

Preprocessing of imaging data was carried out by means of SPM2 software (http://www.fil.ion.ucl.ac.uk/spm, K.J. Friston et al.). The preprocessing included slice timing, realignment, spatial normalization and spatial smoothing using a Gaussian kernel. A mask containing only brain voxels was generated using SPM2 software. The preprocessing included slice timing, realignment, spatial normalization and spatial smoothing using a Gaussian kernel. The linear drifts of all voxel time series were removed by a least squares method.

Voxel-based analysis of fMRI data was carried out using δnormLVD with n = 48, T = 120, τ = 1 and f = 0.9. Fig. 5(a) shows δnormLVD for eight of those voxels with significant differences of δnormLVD between the experimental condition and the control condition through the t test of two samples (p < 0.001). Fig. 5(b) shows the Lmax for the same eight voxels with d = 3, ε = 0.45σ and τ = 1. Obviously, the control condition and the experimental condition cannot be identified by their Lmax through the t test of two samples (p > 0.05) except voxel 4 (p = 0.015). One cause of failing to identify the control condition and the experimental condition with Lmax may be that Lmax is susceptible to noise, whereas the fMRI dataset is a mixture of several patterns corrupted by noise [18,22]. Another cause may be that the embedding dimension d = 3 and time delay τ = 1 chosen for the estimation of Lmax are too poor to describe sensitively the differences among various physiological or pathological states.

As shown in Fig. 5(a), δnormLVD values obtained for the control condition are smaller than those of the experimental condition for these voxels. The results show that the complexity of time series under the control condition is lower than that during the experimental conditions in these brain regions. Thus, the dynamical behaviors of the control condition are closer to a regular signal, which is a feature of the nervous stimulating signal. This implies neutral words alternating with no words (control) more strongly activate the brain region than do threatening words alternation with neutral words (experimental). The explanation for these results in terms of pathophysiological mechanisms lies beyond the scope of this Letter and is open to further investigation(s).

4. Conclusions

Whatever the unknown underlying dynamics may be, the first problem an observer is confronted with is to capture and measure useful information from observational time series. The LVD dimension density is defined to measure the complexity of interrelationships among the component series of short and noisy multivariate data sets. Comparisons between theoretical and simulated results show that in conditions Case I and Case II, the fraction f has some effect on the estimated δnormLVD. Furthermore, the larger the fraction f, the higher the accuracy of estimated δnormLVD. Thus, the normalized LVD dimension density is introduced using the upper and lower bounds of the LVD dimension density.

With time-delay embedding, the normalized LVD dimension density was used to measure the complexity of pseudo-multivariate series constructed from a scalar time series, which is proved to be equivalent with measuring the complexity of the original scalar time series. The simulated results of the logistic map and the Lorenz system indicate that the δnormLVD method can be used to measure dynamical complexity of underlying dynamics very well. Furthermore, the δnormLVD method is also effective in the presence of noise.

The data sets obtained from clinical and physiological studies are usually short and noisy. The normalized LVD dimension density was also used for voxel-based analysis of the short-term fMRI series as a complexity measure. The results of fMRI series with 120 time points reveal that δnormLVD can be used to detect the difference of complexity among various physiological and pathological states sensitively and has great potential for application in clinical studies.

The δnormLVD method has the advantage of simplicity for very short data sets. This method enables us to analyze very short and noisy time series effectively, and can be applied to real-world time series analysis.

Acknowledgement

This work was supported by the Foundation of the Zhejiang Educational Committee of China (Grant No. Y200909251).
References