Neurocomputing 72 (2009) 1042–1055

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Estimating intrinsic dimensionality of fMRI dataset incorporating an AR(1) noise model with cubic spline interpolation

Xiaoping Xie [a,*], Zhitong Cao [a], Xuchu Weng [b], Dan Jin [a]

[a] Physics Department, Zhejiang University, Hangzhou 310027, China
[b] Laboratory for Higher Brain Function, Institute of Psychology, The Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 4 August 2007
Received in revised form
22 March 2008
Accepted 2 April 2008
Communicated by T. Heskes
Available online 26 April 2008

Keywords:
Dimensionality estimation
Autoregressive noise model
Cubic spline interpolation
Functional magnetic resonance imaging
Dimensionality reduction

ABSTRACT

Estimating the true dimensionality of the data to determine what is essential in the data is an important but a difficult problem in fMRI dataset. In this paper, cubic spline interpolation is introduced to detect the number of essential components in fMRI dataset. By constructing proper interpolation variable, more reasonable estimation of the coefficient of an autoregressive noise model of order 1 can be made. Simulation data and real fMRI dataset of resting-state in human brains are used to compare the performance of the new method incorporating an autoregressive noise model of order 1 with cubic spline interpolation (AR1CSI) with that of the method based only on an autoregressive noise model of order 1 (AR1). The results show the AR1CSI method leads to more accurate estimate of the model order at many circumstances, as illustrated in simulated datasets and real fMRI datasets of resting-state human brain.

1. Introduction

Functional magnetic resonance imaging (fMRI) has emerged as a useful and noninvasive technique for studying the function of the brain. A typical fMRI dataset consists of a hundred or more 3D images that represent an indirect (metabolic) measure of neural activity related to brain process. While the spatial and temporal resolution allow for an effective exploration of brain metabolism, the presence of noise and various confounds (physiological rhythms) makes signal extraction challenging.

Although the input dimensionality of fMRI dataset may be quite high (e.g., 600 images for 64 voxels $\times$ 64 voxels $\times$ 5 slices), the meaningful structure of these data, characterized by the intrinsic dimensionality, has many fewer independent degrees of freedom. Dimensionality reduction, namely finding meaningful low-dimensional structures hidden in their high-dimensional observations, is an inevitable problem confronted regularly in dealing with large volumes of high-dimensional data [16,28,31].

A possible way to solve this problem is to consider that the fMRI dataset is a mixture of several patterns corrupted by noise. Detection of such patterns is the goal of principal components analysis (PCA) [18], independent components analysis (ICA) [2,9,24], clustering analysis [11,13], and more generally of multivariate analyze.

Recently, dimensionality estimation and dimensionality reduction have generated a great deal of interest. Many methods are presented to estimate the intrinsic dimensionality (ID) of data and to reduce the dimensionality of the data. Traditionally, the methods to estimate the intrinsic dimensionality can be classified two kinds of approaches, i.e., local methods and global methods [7]. Local (or topological) methods try to estimate the topological dimension of the data manifold, such as the vectors embedded method [12], the near neighbor method [23,32], topology representing network method [20], and locally linear embedding method [28], etc. Global methods try to estimate the ID of a dataset, unfolding the whole dataset in the d-dimensional space. Unlike local methods that use only the information contained in the neighborhood of each data sample, global methods make use of the whole dataset. Projection techniques are widely used global methods [7]. These methods search for the best subspace to project the data by minimizing the projection error, such as PCA, multidimensional scaling methods [26,27], a measured local metric information method to learn the underlying geometry of a dataset [31], and a deep autoencode networks method [16]. As global methods, fractal-based techniques have also been successfully applied to estimate the attractor dimension of the underlying dynamic system generating time series [7,15]. Unlike other global methods, they can provide a non-integer value for ID estimation [18]. Since fractals are generally characterized by

* Corresponding author. Tel.: +86 57187953227.
E-mail addresses: iap@zju.edu.cn, xiexptt@163.com (X. Xie).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.04.003

a non-integer dimensionality, these methods are called fractal. In nonlinear dynamics many definitions of fractal dimensions have been proposed. The Box-Counting and the Correlation dimensions [8,30] are the most popular.

In addition, there are also several methods based on information-theoretic criteria to estimate the dimension from the eigenspectrum, such as the maximum likelihood extension method [1], the Minimum Description Length method [25,33], a Laplace approximation method [22], Bayesian PCA [3,4], and the Bayesian Information Criterion method [29]. Recently, the Bayesian Information Criterion method has been applied to functional neuroimages acquired during a visual activation study [17].

It is well known that fMRI time series typically are demonstrated complex and locally variable autocorrelation structure, even when the data have been acquired with the subject "at rest" [10,21]. There is preliminary evidence that fMRI noise often has long memory in time, or $1/f$ spectral properties, meaning it is positively autocorrelated and there is disproportionate power in the spectrum at low frequencies [21].

All these methods, including information-theoretic criteria method, perform virtually identical when the noise model is purely white. However, using simulated data, it can be shown [10] that none of these methods provide very accurate dimension estimation when the correlation of the noise cannot be neglected. Cordes and Nandy [10] regard the presence of autocorrelations in the noise is the main reason why the above methods fail, because autocorrelations decrease the slope (in absolute value) of the likelihood function.

In order to deal with this problem, an empirical method based on an autoregressive noise model of order 1 (AR1) is introduced to the problem of detecting robustly the number of components in fMRI data [10]. However, from our research, it is found by use of simulations that AR1 method can provide underestimation to the ID at many circumstances, especially as the number of voxels in fMRI data is small.

In this paper, by constructing proper interpolation variable, a method incorporating the autoregressive noise model of order 1 with cubic spline interpolation (AR1CSI) is introduced. In order to compare AR1CSI method with AR1 method and a fractal-based (FB) intrinsic dimension estimating method [8], simulated data with a few signals contaminated by correlative Gaussian noise, and with different number voxels, temporal size of the data, numbers of signal, signal-to-noise ratio and the noise model, were used to estimate the model order. At the same time, the IDs of real resting-state fMRI dataset in human brain were also estimated using AR1 method, FB method and AR1CSI method, respectively. The results analyzing the simulated data and real fMRI dataset prove that AR1CSI method seems to work better for ID estimation and leads to more accurate ID estimation than the AR1 method and FB method, no matter what the voxel number, the temporal size of the data, the signal number and the signal-to-noise ratio are. The ID underestimation of AR1 method can be improved by AR1CSI method.

2. Material and methods

2.1. AR noise model of fMRI data and noisy mixing model

It is proved [21] that the noise in fMRI time series is not uncorrelated but has an autoregressive structure, even when the data have been acquired when the subject is "at rest". One of the most successful modeling strategies for fMRI noise $G=(g_1, g_2, \ldots, g_T)$, T being the number of time points, has been the adoption of autoregressive, linear time invariant models [14,21] of the form

$$g_t=\sum_{i=1}^{q}\phi_i g_{t-i}+\varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma^2), \tag{1}$$

where q is the order of the autoregressive AR(q) process and $t=1, 2, \ldots, T$ denotes time. However, AR models will require many parameters to account for long-range autocorrelated processes. Though the variability of autocorrelation among voxels suggests that it might be suitable to adapt the order of AR process to each individual time series, which can be automated using model selection criteria such as the Bayesian information criterion, this is not always done in practice [21].

Without doubt, for a voxel time series which has a strong correlation structure, the higher the order of AR model is, the better the effect will be. For water phantom data and also for preprocessed real data, when having corrected motion artifacts and detrended signal drifts, it is proved [10] that modeling the noise by an AR(1) process seems to work well and leads to consistent estimates for real and simulated data. The AR(1) model can be described as

$$g_t=\phi g_{t-1}+\varepsilon_t, \tag{2}$$

where ϕ is the AR(1) coefficient and ε_t is a random variable with Gaussian distribution.

The fMRI dataset is a mixture of several patterns corrupted by noise and can be represented by a noisy linear mixing model. The noisy linear mixing model [10] for a fMRI dataset with mean removed can be described as follows:

$$z_i=As_i+G_i, \quad i=1,2,\ldots,N, \tag{3}$$

where z_i and s_i denote the observation and the signal at voxel i with T time points, respectively. The signals are composed of p components with zero mean at voxel i. The mixing matrix A is a $T\times p$ dimensional matrix. The vector G_i is the multivariate correlated Gaussian noise with zero mean. The number of voxels, i.e. N, is satisfied $T/N\to\gamma>0$. Though p, the number of the signal components, is usually less than T, the observation $\boldsymbol{Z}=[z_1, \ldots, z_N]$ is always a matrix of full rank T because the observation is composed of not only signal, but also noise, and thus the dimension of the full fMRI dataset is always larger than the number of biological components [10].

Slightly different from Ref. [10], the signal-to-noise ratio (SNR) is defined in terms of the eigenvalues of covariance matrix for a dataset, i.e.

$$\mathrm{SNR}=\left(\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{j=p+1}^{T}\lambda_j}\right)^{1/2}, \tag{4}$$

where λ_i label the first p eigenvalues of the covariance matrix, which corresponds to signal space of the dataset, and λ_j is the rest $T-p$ eigenvalues of the covariance matrix, which is regarded as noise space.

2.2. ID estimating by AR1CSI method

2.2.1. ID estimating by AR1 method

It is shown [10] that with the conditions of the autoregressive coefficients $\phi\in[0, 0.3]$ and typical fMRI parameters ($N=20{,}000$, $T=160$), $\lambda(k)$, the kth eigenvalue of the sample eigenvalue spectrum for AR(1) Gaussian noise, can be fit excellently by an exponential function $\lambda(k)=a\,\mathrm{e}^{-bk}$, excluding very small k ($k<10$) and very large k ($k>T-10$) where they deviate from the exponential behavior. Similarly, coefficients $a(\phi)$ and $b(\phi)$ can be also accurately parameterized by an exponential function following $a(\phi)=a_1\,\mathrm{e}^{-b_1\phi}$ and $b(\phi)=a_2\,\mathrm{e}^{-b_2\phi}$, respectively. Supposing that the tail eigenvalues of the sample covariance matrix for fMRI dataset have no contributions to the signals and only serve as the noise subspace, then they can be used to estimate the AR(1) coefficient of the corresponding fMRI data.

To estimate the ID of the data, the AR1 method is carried out by following steps [10]. First, in order to get the parameterization $b(\phi) = a_2\,e^{-b_2\phi}$, $a(\phi)$ and $b(\phi)$ are determined by the eigenvalues of simulated data with pure AR(1) Gaussian noise and the proper γ (γ can easily be determined from real data). Second, the tail eigenvalue spectrum of the real data is fitted to $\lambda(k) = a_\gamma\,e^{-b_\gamma k}$ except for the very last 10 eigenvalues, and thus the coefficients, a_γ and b_γ, can be obtained. Then, the AR(1) coefficient ϕ_γ can be estimated by the equality $b(\phi) = a_2\,e^{-b_2\phi}$. Third, the tail eigenvalues of covariance matrix for the real data are always smaller by a shift Δ than the corresponding eigenvalues for simulated pure AR(1) noise. This is because real data contain not only signal but also noise, and each time series is normalized by its variance, however, the corresponding simulated pure AR(1) noise data only contain noise, and the variance of each time series is also normalized to one. Thus, the noise eigenvalue spectrum generated by estimated ϕ_γ can be properly adjusted by the shift Δ. (The shift Δ can be determined by the tail eigenvalue spectra of both real data and simulated pure AR(1) noise.) Fourth, the adjusted noise eigenvalue spectrum for the estimated AR(1) coefficient is then compared to the eigenvalue spectrum of the real data. The number of eigenvalues that are larger than the corresponding adjusted simulated noise eigenvalues and that are before the first the intersection of two eigenvalue spectra then defines the dimension of the signal subspace. It is easy to get the ID estimation by counting the number of eigenvalues that are larger than the corresponding adjusted simulated noise eigenvalue.

The basis of the AR1 method is that $a(\phi)$ and $b(\phi)$ can be accurately parameterized by an exponential function. Nevertheless, we found that this is not always the case by simulation. On one hand, simulation results show that neither $a(\phi)$ against ϕ nor $b(\phi)$ against ϕ can always be accurately parameterized by an exponential function, especially for small voxels number N. As shown in Fig. 1, when $N = 1700$, $\log(a)$ against ϕ is nonlinear for $\phi \in [0, 0.30]$ and $\log(b)$ against ϕ is the same. This means neither $a(\phi)$ against ϕ nor $b(\phi)$ against ϕ can be accurately fit by exponential function under this circumstance.

On the other hand, the key of the AR1 method is that the AR(1) noise spectrum generated by estimated ϕ need to be adjusted so that it coincides with the tail eigenvalues of the real data [10], i.e., the slope of AR(1) noise spectrum also coincides with the slope of tail eigenvalues for the real data. However, for a exponential function $\lambda(k) = a\,e^{-bk}$, the slope is determined by its derivation

$$\frac{d\lambda}{dk} = -ab\,e^{-bk}, \tag{5}$$

i.e., the slope is determined by not only b, but also a. As shown in Fig. 2, though their b is the same, there is significant difference of the slope between two exponent functions because of their different parameter a. Therefore, we believe that it is not enough to estimate coefficient ϕ by using only b (ignoring the effect of a) in AR1 method. In the later of this paper, we will prove that it can result in underestimating ID of the data.

2.2.2. Cubic spline interpolation

Interpolation is used to estimate the value of a function between known data points without knowing the actual function. Cubic spline interpolation [6] is a useful technique to interpolate between known data points due to its stable and

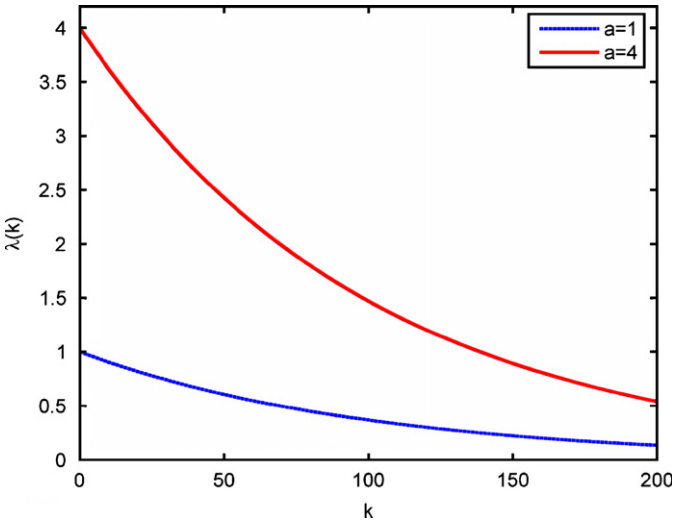

Fig. 2. Two exponent functions ($\lambda(k) = a\,e^{-bk}$) have the same coefficient b, and the different coefficient a. The coefficient $b = 0.01$. The dot–dash line corresponds to $a = 1$. The solid line represents $a = 4$. There is significant difference of the slope between these two exponent functions.

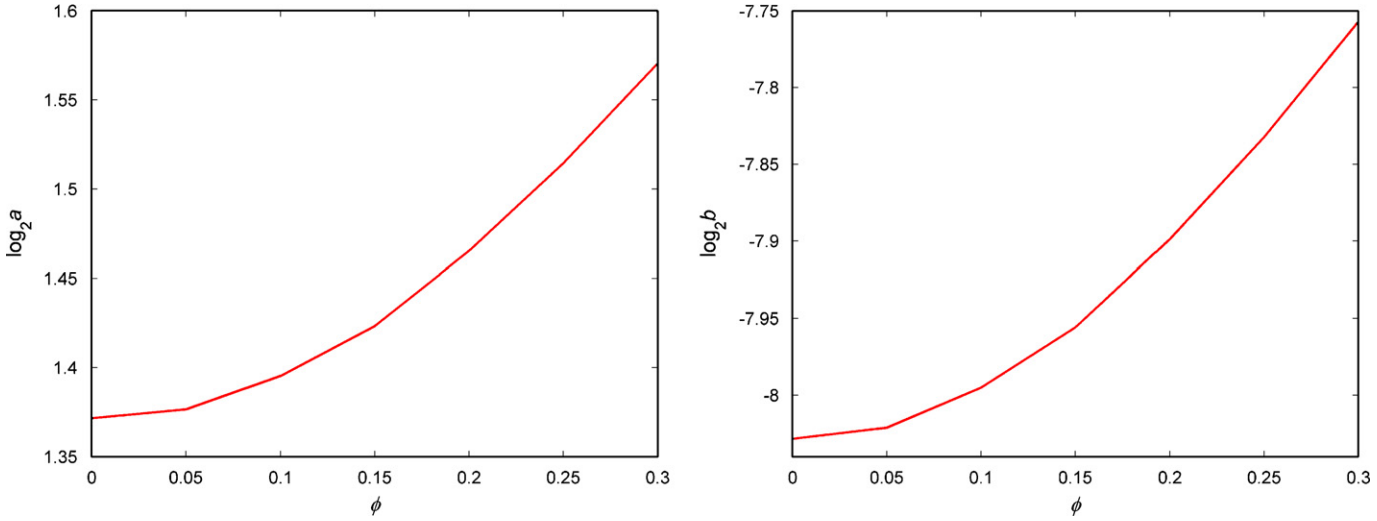

Fig. 1. Logarithm of a typical coefficients $a(\phi)$ and $b(\phi)$ for correlated Gaussian noise corresponding to an AR(1) model (600 time dimension, 1700 samples). ϕ is the AR(1) coefficient. The left plot is logarithm of coefficient a, and the right plot is logarithm of coefficient b. The curves are not linear means that both $a(\phi)$ and $b(\phi)$ are not accurately fit to exponent functions.

smooth characteristics. They are generally well behaved and continuous up to the second order derivative at the data points.

One of the most useful and well-known classes of functions, which map a set of real numbers into itself, is the class of algebraic polynomials. Polynomials are used for approximation because they can be evaluated, differentiated, and integrated easily and in finitely many steps using just the basic arithmetic operations of addition, subtraction and multiplication [19]. A polynomial of order n is a function of the form

$$p(x) = a_1 + a_2x + \cdots + a_nx^{n-1}, \tag{6}$$

and $\mathbf{P_n}$ denotes the set or linear space of all polynomials of order n. Let $\tau_1, \tau_2, \ldots, \tau_n$ be a sequence of n distinct points. For an arbitrary given function g, the expressions $g[\tau_i, \tau_{i+1}, \ldots, \tau_{i+k}]$ are called divided differences of g [6]. The zeroth divided difference with respect to τ_i, denoted $g[\tau_i]$, is simply the value of g at τ_i: $g[\tau_i] = g(\tau_i)$. The remaining divided differences are defined inductively. The first divided difference of g with respect to τ_i, τ_{i+1} is denoted $g[\tau_i, \tau_{i+1}]$ and is defined as

$$g[\tau_i, \tau_{i+1}] = \frac{g[\tau_{i+1}] - g[\tau_i]}{\tau_{i+1} - \tau_i}. \tag{7}$$

Similarly, after the $(k-1)$st divided differences, $g[\tau_i, \tau_{i+1}, \ldots, \tau_{i+k-1}]$ and $g[\tau_{i+1}, \tau_{i+2}, \ldots, \tau_{i+k}]$, have been determined, the kth divided difference relative to $\tau_i, \tau_{i+1}, \tau_{i+2}, \ldots, \tau_{i+k}$ is given by

$$g[\tau_i, \tau_{i+1}, \ldots, \tau_{i+k-1}, \tau_{i+k}] = \frac{g[\tau_{i+1}, \ldots, \tau_{i+k-1}, \tau_{i+k}] - g[\tau_i, \tau_{i+1}, \ldots, \tau_{i+k-1}]}{\tau_{i+k} - \tau_i}. \tag{8}$$

With this notation, the nth Lagrange polynomial $p_n(x)$, which agrees with the function g at the distinct number $\tau_1, \tau_2, \ldots, \tau_n$, can be written as

$$p_n(x) = \sum_{k=1}^{n} g[\tau_1, \ldots, \tau_k](x - \tau_1) \cdots (x - \tau_{k-1}). \tag{9}$$

This equation is also known as Newton's interpolation polynomial [19].

There are many methods to realize interpolation. Piecewise-polynomial approximation [6] is an alterative approach, which is to divide the interval into a collection of subintervals and construct a different approximating polynomial on each subinterval.

Given the data $g(\tau_1), g(\tau_2), \ldots, g(\tau_n)$ with $a = \tau_1 < \tau_2 < \cdots < \tau_n = b$, which consists of joining a set of data points $\{(\tau_1, g(\tau_1)), (\tau_2, g(\tau_2)), \ldots, (\tau_n, g(\tau_n))\}$, a piecewise cubic interpolant f to g is constructed as follows [5]. On each interval $[\tau_i, \tau_{i+1}]$, f agrees with some polynomial p_i of order 4,

$$f(x) = p_i(x) \quad \text{for } \tau_i \leqslant x \leqslant \tau_{i+1} \quad \text{for some } p_i \in \mathbf{P}_4, \quad i = 1, 2, \ldots, n-1. \tag{10}$$

The ith polynomial piece p_i is made to satisfy the conditions:

$$p_i(\tau_i) = g(\tau_i), \quad p_i(\tau_{i+1}) = g(\tau_{i+1}), \quad p_i'(\tau_i) = s_i, \quad p_i'(\tau_{i+1}) = s_{i+1}, \quad i = 1, 2, \ldots, n-1. \tag{11}$$

Here, $s_1, s_2, \ldots, s_n$ are free parameters. The resulting piecewise cubic function f agrees with g at $\tau_1, \tau_2, \ldots, \tau_n$, and is in $C^{(1)}[a, b]$, i.e., is continuous and has a continuous first derivative on $[a, b]$, regardless of how we choose the free "slopes" s_i.

In order to compute the coefficients of ith polynomial piece p_i, its Newton form are used

$$p_i(x) = p_i(\tau_i) + (x - \tau_i)p_i[\tau_i, \tau_i] + (x - \tau_i)^2 p_i[\tau_i, \tau_i, \tau_{i+1}] + (x - \tau_i)^2(x - \tau_{i+1})p_i[\tau_i, \tau_i, \tau_{i+1}, \tau_{i+1}]. \tag{12}$$

Its coefficients are determined from a divided difference table (Table 1) [5] for p_i based on the conditions data for p_i.

Table 1
The divided differences for p_i based on the conditions data for p_i

	$p_i[\ ]$	$p_i[\ ,\ ]$	$p_i[\ ,\ ,\ ]$	$p_i[\ ,\ ,\ ,\ ]$
τ_i	$g(\tau_i)$			
		s_i		
τ_i	$g(\tau_i)$		$(g[\tau_i, \tau_{i+1}] - s_i)/\Delta\tau_i$	
		$g[\tau_i, \tau_{i+1}]$		$(s_{i+1} + s_i - 2g[\tau_i, \tau_{i+1}])/(\Delta\tau_i)^2$
τ_{i+1}	$g(\tau_{i+1})$		$(s_{i+1} - g[\tau_i, \tau_{i+1}])/\Delta\tau_i$	
		s_{i+1}		
τ_{i+1}	$g(\tau_{i+1})$			

Note: $p_i[\ ]$ denotes the zeroth divided differences. $p_i[\ ,\ ]$ denotes the first divided differences. $p_i[\ ,\ ,\ ]$ denotes the second divided differences. $p_i[\ ,\ ,\ ,\ ]$ denotes the third divided differences. $\Delta\tau_i = \tau_{i+1} - \tau_i$.

For cubic spline interpolation, the free slopes $s_2, \ldots, s_{n-1}$ are determined from the condition that f should be twice continuously differentiable [5,6], i.e., so that f has also a continuous curvature. This gives the conditions that, for $i = 2, \ldots, n-1$,

$$p_{i-1}''(\tau_i) = p_i''(\tau_i), \tag{13}$$

supposing that the two remaining free-parameters s_1 and s_n have been chosen somehow, we now have a tridiagonal linear system of $n-2$ equations for the $n-2$ unknowns $s_2, \ldots, s_{n-1}$ which is strictly row diagonally dominant. Hence, the system has exactly one solution.

The free-parameters s_1 and s_n can be obtained by boundary condition, including free or natural boundary and clamped boundary, etc. [6]. If nothing is known about end point derivatives, then the "not-a-knot" condition [5] should be tried. Here, we use the "not-a-knot" boundary condition, which s_1 and s_n are selected so that $p_1 = p_2$ and $p_{n-2} = p_{n-1}$. This requires that f''' be continuous across τ_2 and τ_{n-1}.

Before estimating the autoregressive coefficient ϕ using cubic spline interpolation method, a very important thing we should do is to construct interpolation variable x. To make the eigenvalues of AR(1) noise coinciding with the tail eigenvalues of the data after shifted, x should be identical with the slope of the tail eigenvalues (i.e. the derivation of the tail eigenvalues shown in formula (5)). However, in order to let interpolation work properly, x must be monotonically increasing or decreasing. Though both a and b are monotonically increasing with ϕ, $ab\,e^{-bk}$ can break down the monotonic property and make the cubic spline interpolation fail. For keeping the monotonic property of x, we consider to replace $ab\,e^{-bk}$ with $ab\,e^{bk}$ as variable x in this paper.

2.2.3. ID estimating by AR1CSI method

The algorithm of estimating the ID by AR1CSI method carries out as the following:

- *Step* 1. Construct $(ab\,e^{bk} = f(\phi), \phi)$ by pure noise simulation data.
  The $a(\phi)$ and $b(\phi)$ are constructed by the eigenvalue spectrum of the covariance matrix for the simulation data of pure Gaussian noise with AR(1) covariance structure. The simulation data has the same T and N ($\gamma \to T/N$) as the corresponding fMRI data. Thus some pairs of discrete $a(\phi)$ against ϕ and $b(\phi)$ against ϕ are obtained. Therefore, a set of data points $(ab\,e^{bk} = f(\phi), \phi)$ can be constructed from $a(\phi)$ and $b(\phi)$.
- *Step* 2. Determine the proper ϕ_γ corresponding to fMRI data.
  In order to determine the proper ϕ_γ corresponding to fMRI data, the eigenvalue spectrum of real fMRI data (all voxel time

series were detrended) is calculated. The tail eigenvalues of real fMRI data are fitted to $\lambda(k) = a_\gamma \mathrm{e}^{-b_\gamma k}$ except for the very last 10 eigenvalues (due to the same reason as in AR1 method), and thus the coefficients, namely a_γ and b_γ, can be obtained. Then, unlike the AR1 method that uses only the coefficient b_γ to estimate the coefficient ϕ_γ, the proper ϕ_γ corresponding to fMRI data, which is also the AR(1) coefficient, can be calculated by use of cubic spline interpolation method based on the equality $ab\,\mathrm{e}^{bk} = f(\phi)$.

- *Step* 3. Adjust the eigenvalue spectrum of the simulated AR(1) noise data.
  With the proper coefficient ϕ_γ estimated from fMRI data in Step 2, the simulated pure AR(1) noise data are generated again using the same T and N as the corresponding fMRI data. Due to the same reason as in AR1 method, the tail eigenvalues for spectrum of real fMRI data are also a shift Δ smaller than the corresponding eigenvalues of the simulated pure AR(1) noise. The shift Δ is also determined from the tail spectrum, and thus the noise eigenvalue spectrum can be properly adjusted by the shift Δ. Then, the adjusted eigenvalue spectrum of the simulated pure AR(1) noise data is compared to the eigenvalue spectrum of corresponding fMRI data.
- *Step* 4. Estimate the intrinsic dimensionality.
  After adjusting the eigenvalue spectrum of the simulated pure AR(1) noise data, the number of eigenvalues of real fMRI data that are larger than the corresponding adjusted eigenvalues of simulated pure AR(1) noise and that are before the first intersection of two eigenvalue spectra defines the dimension of the signal space as AR1 method does [10]. We can get the estimation of the intrinsic dimensionality of the real fMRI data by counting the number of those eigenvalues that are larger than the corresponding adjusted eigenvalues of simulated pure AR(1) noise from $k = 1$ to the first intersection between the eigenvalue spectrum of the real fMRI data and the adjusted eigenvalue spectrum of the simulated pure AR(1) noise data.

2.2.4. Effectiveness analysis of AR1CSI method

Suppose that the real dataset comprises p signal components and its tail eigenvalues of the covariance matrix can be fit by an exponent function $\lambda_0(k) = a\,\mathrm{e}^{-bk}$, and $k = p{+}1, p{+}2, \ldots, T$; the AR(1) coefficient ϕ_γ can then be determined by using AR1 (using only b) and AR1CSI (using $ab\,\mathrm{e}^{bk}$) method, respectively. The corresponding simulated AR(1) noise datasets can be generated by using the estimated ϕ_γ. If the estimation is accurate enough, the eigenvalues for simulated AR(1) noise data by using the ϕ_γ estimated by AR1 method can be fit by an exponent function $\lambda_2(k) = a''\,\mathrm{e}^{-bk}$, which has the same b as the real dataset. At the same time, the eigenvalues for simulated AR(1) noise data by using the ϕ_γ estimated by AR1CSI method can also be fit by an exponent function $\lambda_1(k) = a'\,\mathrm{e}^{-b'k}$, which holds $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$. For $\lambda_1(k)$ and $\lambda_2(k)$, the domain of definition is $k = 1, 2, \ldots, T$. In order to make comparison easy, $\lambda_0(k)$ is also extended to $k = 1, 2, \ldots, T$ following the exponent function $\lambda_0(k) = a\,\mathrm{e}^{-bk}$.

Because of normalization, the eigenvalues of the covariance matrices form the pattern

$$T = \sum_{k=1}^{T} a''\,\mathrm{e}^{-bk} = \sum_{k=1}^{T} a'\,\mathrm{e}^{-b'k} > \sum_{k=1}^{T} a\,\mathrm{e}^{-bk}. \tag{14}$$

Lemma. If formula (14) can been held by three exponent functions $\lambda_0(k) = a\,\mathrm{e}^{-bk}$, $\lambda_1(k) = a'\,\mathrm{e}^{-b'k}$ and $\lambda_2(k) = a''\,\mathrm{e}^{-bk}$ with $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$, $k = 1, 2, \ldots, T$, then we have

(1) $b' < b$ and $a < a' < a''$;
(2) $ab\,\mathrm{e}^{-bi} < a'b'\,\mathrm{e}^{-b'i}$ and $ab\,\mathrm{e}^{-bi} < a''b\,\mathrm{e}^{-bi}$, $\forall i \in [1, T]$.

Proof. (1) Suppose $b' \geqslant b$, then we have

$\frac{1}{b'\,\mathrm{e}^{b'(k+i)}} \leqslant \frac{1}{b\,\mathrm{e}^{b(k+i)}}$. Since $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$, thus

$$\sum_{i=1}^{T} a'\,\mathrm{e}^{-b'i} = \sum_{i=1}^{T} \frac{ab\,\mathrm{e}^{bk}}{a'b'\,\mathrm{e}^{b'k}} a'\,\mathrm{e}^{-b'i} = \sum_{i=1}^{T} \frac{ab\,\mathrm{e}^{bk}}{b'\,\mathrm{e}^{b'k}}\,\mathrm{e}^{-b'i} \leqslant \sum_{i=1}^{T} \frac{ab\,\mathrm{e}^{bk}}{b\,\mathrm{e}^{bk}}\,\mathrm{e}^{-bi} = \sum_{i=1}^{T} a\,\mathrm{e}^{-bi}.$$

This contradicts our hypothesis, so $b' < b$.

Since $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$, then $a < a'$.

If $a' \geqslant a''$, then $a''\,\mathrm{e}^{-bi} \leqslant a'\,\mathrm{e}^{-bi}$. Since $b' < b$, we have $a''\,\mathrm{e}^{-bi} \leqslant a'\,\mathrm{e}^{-bi} < a'\,\mathrm{e}^{-b'i}$. Thus,

$$\sum_{k=1}^{T} a''\,\mathrm{e}^{-bk} < \sum_{k=1}^{T} a'\,\mathrm{e}^{-b'k}.$$

This also contradicts with our hypothesis, so $a' < a''$. Since $a < a'$, thus $a < a' < a''$.

(2) Since $b' < b$ and $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$, thus $ab\,\mathrm{e}^{-bi} = ab\,\mathrm{e}^{bk}\mathrm{e}^{-bk}\mathrm{e}^{-bi} = a'b'\,\mathrm{e}^{b'k}\mathrm{e}^{-bk}\mathrm{e}^{-bi} < a'b'\,\mathrm{e}^{b'k}\mathrm{e}^{-b'k}\mathrm{e}^{-b'i} = a'b'\,\mathrm{e}^{-b'i}$.

Due to $a < a''$, $ab\mathrm{e}^{-bi} < a''b\,\mathrm{e}^{-bi}$. □

Since the real dataset comprises p signal components, its tail eigenvalues of the covariance matrix is composed of the last $T-p$ eigenvalues.

Theorem 1. If formula (14) can been held by the exponent functions $\lambda_0(k) = a\,\mathrm{e}^{-bk}$, $\lambda_1(k) = a'\,\mathrm{e}^{-b'k}$ with $a'b'\,\mathrm{e}^{b'k} = ab\,\mathrm{e}^{bk}$, $k = 1, 2\ldots, T$, as well as

$$\Delta = \frac{1}{T-p} \sum_{i=p+1}^{T} (a'\,\mathrm{e}^{-b'i} - a\,\mathrm{e}^{-bi}). \tag{15}$$

Then, the inequality

$$a'\,\mathrm{e}^{-b'(p+1)} - \Delta > a\,\mathrm{e}^{-b(p+1)} \tag{16}$$

can hold.

Proof. Constructing a function $\beta(x) = a'\,\mathrm{e}^{-b'x} - a\,\mathrm{e}^{-bx}$ with $(0 < x \leqslant T)$, then its derivation $\beta'(x) = ab\,\mathrm{e}^{-bx} - a'b'\,\mathrm{e}^{-b'x}$. From Lemma, we have $\beta'(x) < 0$ within $(0 < x \leqslant T)$. Thus, $\beta(x)$ is a function of monotonous decrement at the scope of $x \in (0, T]$. Therefore, for $\beta(k) = a'\,\mathrm{e}^{-b'k} - a\,\mathrm{e}^{-bk}$ and $k = 1, 2, \ldots, T$, the mean of the last $T-p$ terms, which is also the shift Δ used to adjust the pure AR(1) noise spectrum, is less than the term $\beta(p{+}1)$. Thus, we have the inequality (17).

$$\Delta = \frac{1}{T-p} \sum_{i=p+1}^{T} (a'\,\mathrm{e}^{-b'i} - a\,\mathrm{e}^{-bi}) < a'\,\mathrm{e}^{-b'(p+1)} - a\,\mathrm{e}^{-b(p+1)}. \tag{17}$$

So, $a'\,\mathrm{e}^{-b'(p+1)} - \Delta > a\,\mathrm{e}^{-b(p+1)}$. □

Theorem 1 means that after adjusted, the eigenvalue $\lambda_1(p{+}1) - \Delta$ of covariance matrix for simulated pure AR(1) noise, which generated using ϕ_γ estimated by AR1CSI, is still larger than $\lambda_0(p{+}1)$ of real data.

Theorem 2. If formula (14) can been held by the exponent functions $\lambda_0(k) = a\,\mathrm{e}^{-bk}$, $\lambda_2(k) = a''b\,\mathrm{e}^{-bk}$, $k = 1, 2, \ldots, T$, as well as

$$\Delta'' = \frac{1}{T-p} \sum_{i=p+1}^{T} (a''\,\mathrm{e}^{-bi} - a\,\mathrm{e}^{-bi}). \tag{18}$$

Then, the inequality

$$a''\,\mathrm{e}^{-b(p+1)} - \Delta'' > a\,\mathrm{e}^{-b(p+1)} \tag{19}$$

can hold.

The proof of Theorem 2 is similar to the proof of Theorem 1. Theorem 2 implies that after adjusted, the eigenvalue $\lambda_2(p{+}1) - \Delta''$

of covariance matrix for simulated pure AR(1) noise, which generated using ϕ_γ estimated by AR1, is also larger than $\lambda_0(p+1)$ of real data.

For simplifying the problem, assume that AR1 method has the same shift Δ as AR1CSI method, i.e. $\Delta \approx \Delta''$. Because of normalization, the sum of all eigenvalues of simulated pure AR(1) noise is T, i.e.

$$T = \sum_{k=1}^{T} a'' e^{-bk} = a'' e^{-b} \frac{1 - e^{-bT}}{1 - e^{-b}}. \quad (20)$$

If b is very small and T is large enough, Eq. (20) can be simplified to $T \approx a'' b^{-1}$. Therefore, the eigenvalue $\lambda_2(p+1)$ of AR1 can be simplified to

$$\lambda_2(p+1) = a'' e^{-(p+1)b} \approx Tb\, e^{-(p+1)b}. \quad (21)$$

Construct the function $\rho(x) = x e^{-(p+1)x}$ for $x > 0$. Its derivation $\rho'(x) = e^{-(p+1)x} - x(p+1) e^{-(p+1)x}$. As $(p+1)x < 1$, the $\rho'(x) > 0$ and thus $\rho(x)$ is a function of monotonous increment. Due to $b' < b$, we can obtain

$$a'' e^{-(p+1)b} > a' e^{-(p+1)b'}. \quad (22)$$

For fMRI data, the condition $(p+1)b < 1$ is satisfied easily. Combining inequalities (16) and (22), it is clear that after adjusted, both the eigenvalue $\lambda_2(p+1) - \Delta$ and $\lambda_1(p+1) - \Delta$ of covariance matrix for simulated pure AR(1) noise are larger than $\lambda_0(p+1)$ of the real data. Furthermore, the eigenvalue $\lambda_1(p+1) - \Delta$ of AR1CSI method is located between the eigenvalue $\lambda_0(p+1)$ of the real dataset and $\lambda_2(p+1) - \Delta$ of AR1 method. This means that the ID estimated by AR1CSI is closer to the true number of signal components than that estimated by AR1.

2.3. Data acquisition and procedure

2.3.1. Simulated data

To evaluate the performance of AR1CSI method, simulated data with a few signals contaminated by Gaussian noise with autocorrelation structure were generated to estimate the model order by AR1 method, FB method and AR1CSI method, respectively. With fixed autoregressive coefficient $\phi \in [0, 0.5]$, the simulated data were constructed from a few orthogonal signals, mixed by an orthogonal mixtures matrix and added AR(1) Gaussian noise. All simulated data were composed of 20,000 or 2000 voxels. The number of time frames was from 150 to 600. Prior to analysis, each voxel time series was normalized by its variance. Fig. 3 shows the corresponding eigenvalue spectrum of the covariance matrix for simulated data with $p = 50$, $\phi = 0.2$ and SNR = 2. The pure AR(1) noise eigenvalues, which coefficients (ϕ_γ) were estimated by AR1 and AR1CSI method, respectively, are also plotted. The left plot includes the noise eigenvalue spectrum generated from the estimated AR(1) coefficient before adjusted and the right plot includes the noise eigenvalue spectrum after adjusted.

As shown in Fig. 3, due to sufficient SNR used in the simulation, the slope difference is significant between the large eigenvalues corresponding to the signals and the small eigenvalues regarded as the noise region. Thus a natural break forms between the large and small eigenvalues. The eigenvalues corresponding to the signals are indicated by this natural break [10] and their number can be count through this break. On the other hand, variance normalization brings about the result that the tail eigenvalues of spectrum for data become smaller than the corresponding eigenvalues for the pure AR(1) noise, and thus the eigenvalue spectrum of the pure AR(1) noise data need to be adjusted. After adjusting the noise eigenvalues, the intersection of the data eigenvalue spectrum with the pure AR(1) noise eigenvalue spectrum will provide the correct number of signals [10].

2.3.2. fMRI dataset

Gradient-echo echo planar imaging (EPI) data were acquired from nine healthy volunteers in resting-state with closing eyes, stopping thinking if any idea comes up. The data was obtained on a 1.5 T PHILIPS MEDICAL SYSTEMS Gyroscan NT (TR = 700 ms, flip angle = 70° and FOV = 23 cm, with five transection slices covering the visual cortex and other five transection slices covering the motor cortex, 5 mm slice thickness, matrix size: 64×64). After discarding initial scans (to allow for magnetic saturation effects) each time series is comprised of 600 scan images.

The data were preprocessed with using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software/spm2) software. The time series were slice-time corrected, corrected differences in image

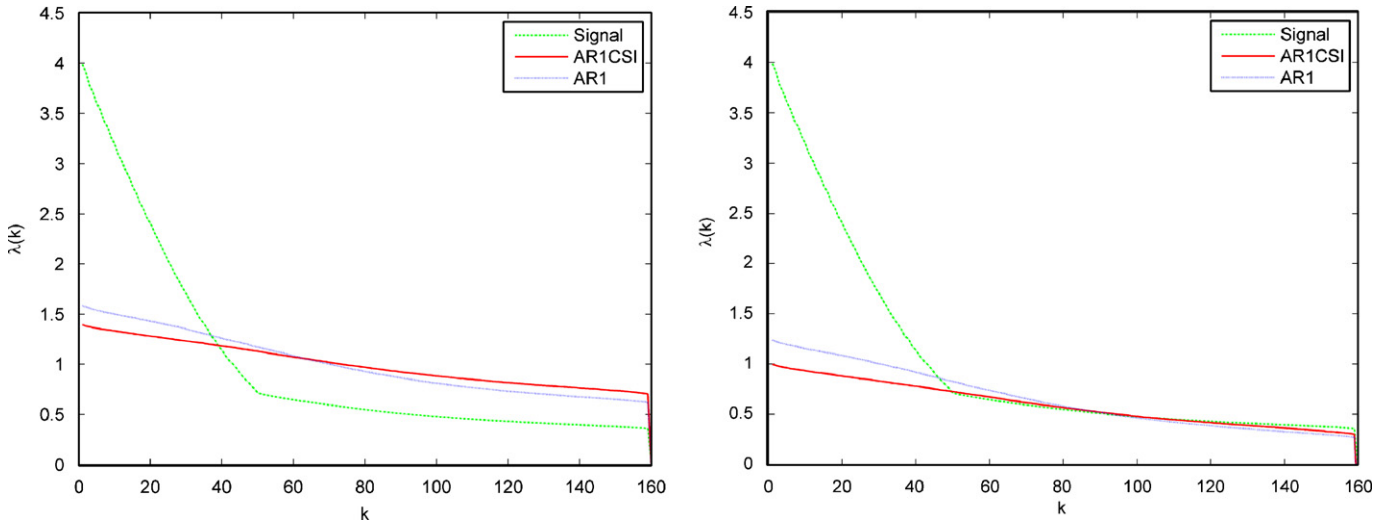

Fig. 3. Eigenvalues for simulated data with correlated noise corresponding to AR(1) mode. The dot–dash line represents 50 signals corrupted by noise with a SNR = 2. The broken line corresponds to the pure noise spectrum which coefficient ϕ is estimated by AR1 method. The solid line corresponds to the pure noise spectrum which coefficient ϕ is estimated by AR1CSI method. All time series were normalized by their variances. The right plot is the tail-adjusted eigenvalue spectrum of pure AR(1) noise, and the left plot is the eigenvalue spectrum of pure AR(1) noise but not tail-adjusted.

acquisition time between slices, and realigned, corrected for movement-related effects. A mask containing only brain voxels was generated by threshold. In order to retain as much information as possible, no further preprocess was done.

3. Results and discussion

Fig. 4 shows the block diagram of the procedure for estimating the ID of the fMRI data.

3.1. For simulated data

To compare the results of the ID estimation at various circumstances, we applied FB, AR1 and AR1CSI method respectively to the simulated data respectively. The results are shown in Figs. 5–14.

3.1.1. Simulated data as a function of the autoregressive coefficient ϕ (Figs. 5 and 6)

Fig. 5 shows the comparison of the estimated IDs as a function of the autoregressive coefficient ϕ between AR1 and AR1CSI method. Fig. 6 shows the comparison between FB and AR1CSI method. For 20,000 voxels, AR1CSI method almost produces the correct estimate of the number of signals. FB method significantly overestimates the model order. As ϕ close to zero, both AR1 and AR1CSI method slightly underestimate the model order. With increasing noise correlation, AR1 always underestimate the model order, whereas AR1CSI method retains its stability and produces an accurate estimate of the model order except for $\phi<0.08$. For 2000 voxels, though both AR1 and AR1CSI method produce underestimate the model order, the dimension estimated by AR1CSI method is closer to the correct number of signals than that by AR1 method. FB method overestimates the model order at small ϕ, and underestimates the model order at large ϕ.

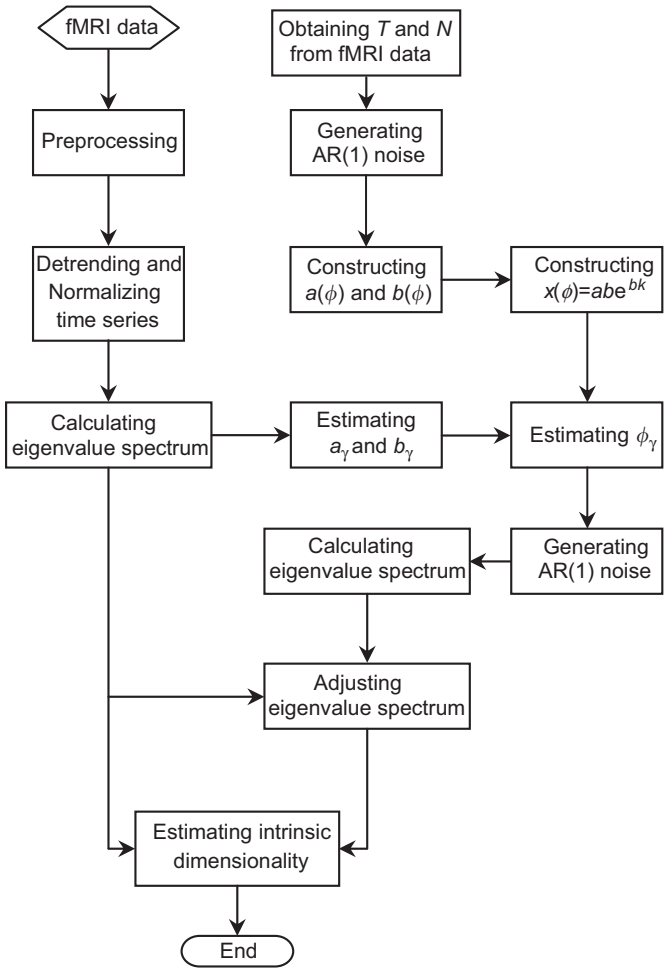

Fig. 4. The block diagram of the procedure for estimating the ID of the fMRI data.

3.1.2. Simulated data as a function of the SNR (Figs. 7 and 8)

Fig. 7 shows the comparison of the estimated IDs as a function of the SNR between AR1 and AR1CSI method. Fig. 8 shows the comparison between FB and AR1CSI method. For the number of voxels is 20,000, varying the SNR for data with $\phi = 0.2$ shows that though AR1 provide an underestimate dimensionality even for data with relatively high SNR whereas AR1CSI method provides a correct dimensionality estimation and a stable estimation. FB method produces a significantly overestimates of the model order. Furthermore, the larger the SNR is, the more serious the overestimate of the FB method is. For 2000 voxels, both AR1 and AR1CSI method underestimate the model order in most SNR. AR1CSI method also produces closer dimensionality estimation to the correct number of signals than AR1 does. FB method overestimates the model order at large SNR and underestimates the model order at small SNR.

3.1.3. Simulated data as a function of the temporal size of the data (Figs. 9 and 10)

Fig. 9 shows the comparison of the estimated IDs as a function of the temporal size of the data between AR1 and AR1CSI method. Increasing the temporal size of the data has some effect on the model order estimate using AR1 and AR1CSI method. For any temporal size of the data, not only 20,000 voxels but also 2000 voxels, AR1CSI method can also produces almost correct and stable dimensionality estimate. But for AR1 method, it is shown again that underestimate dimensionality is obtained at 2000 voxels. When the number of voxels is 20,000, AR1CSI method still provides better results than AR1 method except $T>450$, which provides almost the same results as AR1CSI method. Fig. 10 shows the comparison between FB and AR1CSI method. For 20,000 voxels, FB method always overestimates the model order. However, for 2000 voxels, FB always underestimates the number of signals.

3.1.4. Simulated data as a function of the number of signals (Table 2, Fig. 11)

Table 2 shows the comparison of the estimated IDs as a function of the number of signals between AR1 and AR1CSI method. Fig. 11 shows the comparison between FB and AR1CSI method. For 20,000 voxels, increasing the number of signals shows that AR1CSI method is very accurate for the number of signals from 20 to 70 whereas AR1 method underestimates the dimensionality of the data. FB method still overestimates significantly the model order. For 2000 voxels, though both AR1CSI method and AR1 method underestimate the ID of the data at most circumstance, AR1CSI method produces a better results than AR1 method again. FB method overestimates the number of signals at small number of signals and underestimates the model order at large number of signals.

3.1.5. Simulated data as a function of the number of voxels (Fig. 12)

Fig. 12 shows the comparison of the estimated IDs as a function of the number of voxels. The left plot is the comparison between

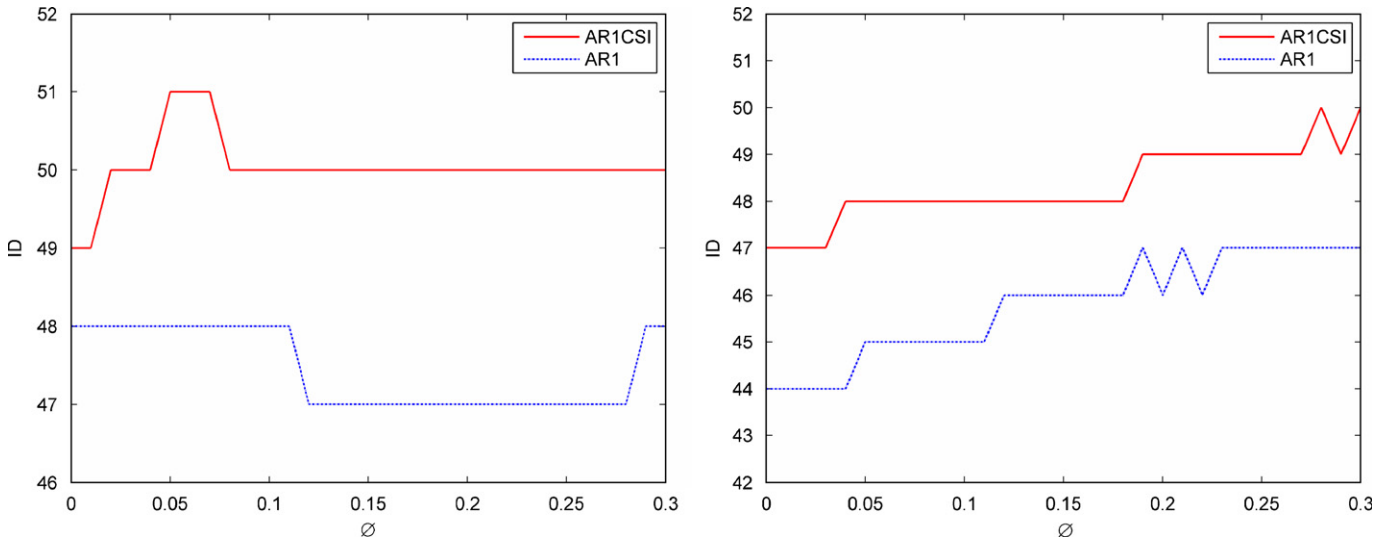

Fig. 5. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the AR(1) coefficient ϕ. $\phi \in [0, 0.3]$, $T = 160$ and SNR $= 2$. The dot–dash line is the dimension estimated by AR1 method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

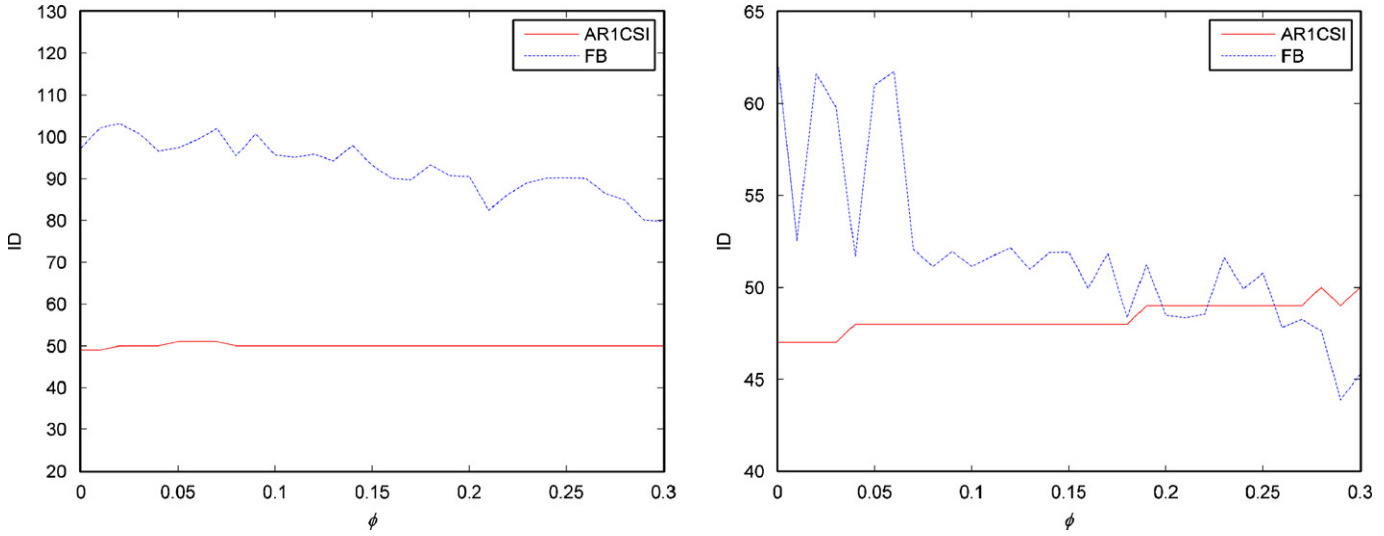

Fig. 6. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the AR(1) coefficient ϕ. $\phi \in [0, 0.3]$. $T = 160$ and SNR $= 2$. The dot–dash line is the dimension estimated by FB method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

AR1 and AR1CSI method. The right plot is the comparison between FB and AR1CSI method. Increasing the number of voxels shows that AR1CSI method is very accurate whereas AR1 method underestimates the dimensionality of the data. When the number of voxels is more than 4000, AR1CSI method produces the correct estimate of the number of signals. Even for a large number of voxels, AR1 method cannot produce a correct result yet and gives a lower estimation than the correct dimensionality. FB method underestimates the model order at small number of voxels, whereas overestimates the number of signals at large number of voxels. Furthermore, the larger the number of voxels is, the more serious the overestimate of the FB method is.

To show the effectiveness of AR1CSI method for a more complicated noise model, simulated data with a few signals corrupted by an ARMA(4,4) noise, which follows a $A(q)y(t) = C(q)e(t)$ process, were generated. Here, $e(t)$ stands for white noise following Gaussian distribution, $A(q)$ denote the AR coefficients and $C(q)$ label the MA coefficients, and q is the sign of the shift operator. The average ARMA(4,4) coefficients were obtained by use of a typical resting-state data, where $A = [1\ -0.007\ 0.025\ 0.111\ -0.073]$ and $C = [1\ 0.061\ 0.030\ 0.089\ -0.044]$ [10]. With ARMA(4,4) noise model and these coefficients, 20,000 and 2000 noisy time series were generated and 50 orthogonal signals were added. FB, AR1 and AR1CSI method were then applied to the artificial data

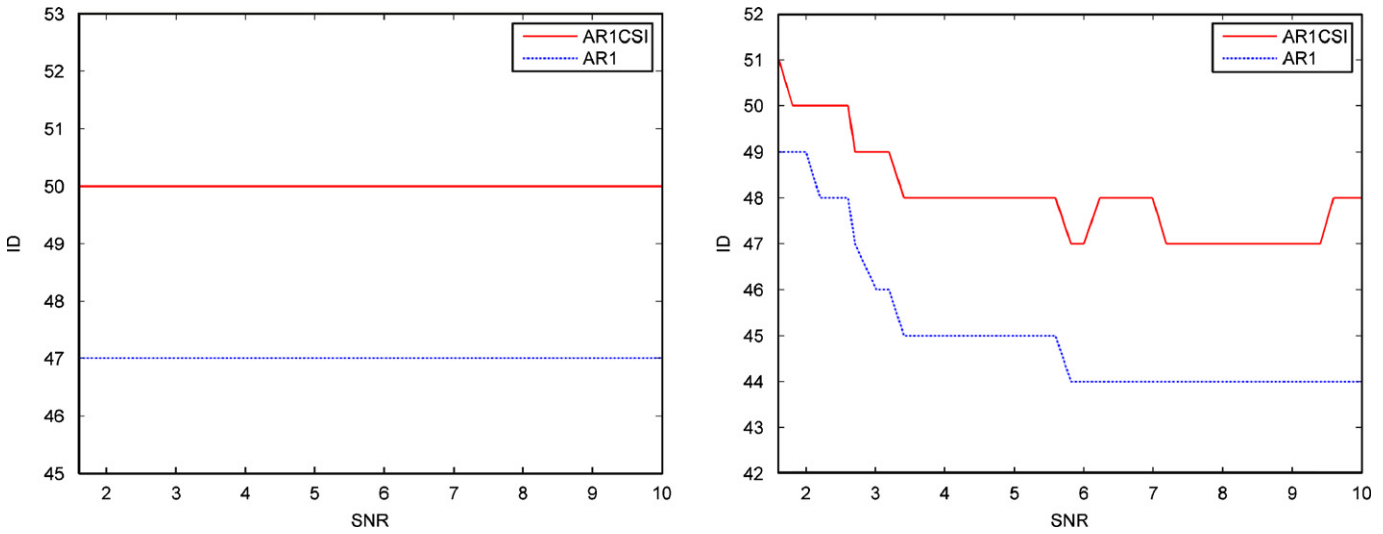

Fig. 7. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the SNR. $T=160$ and $\phi=0.2$. The dot–dash line is the dimension estimated by AR1 method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

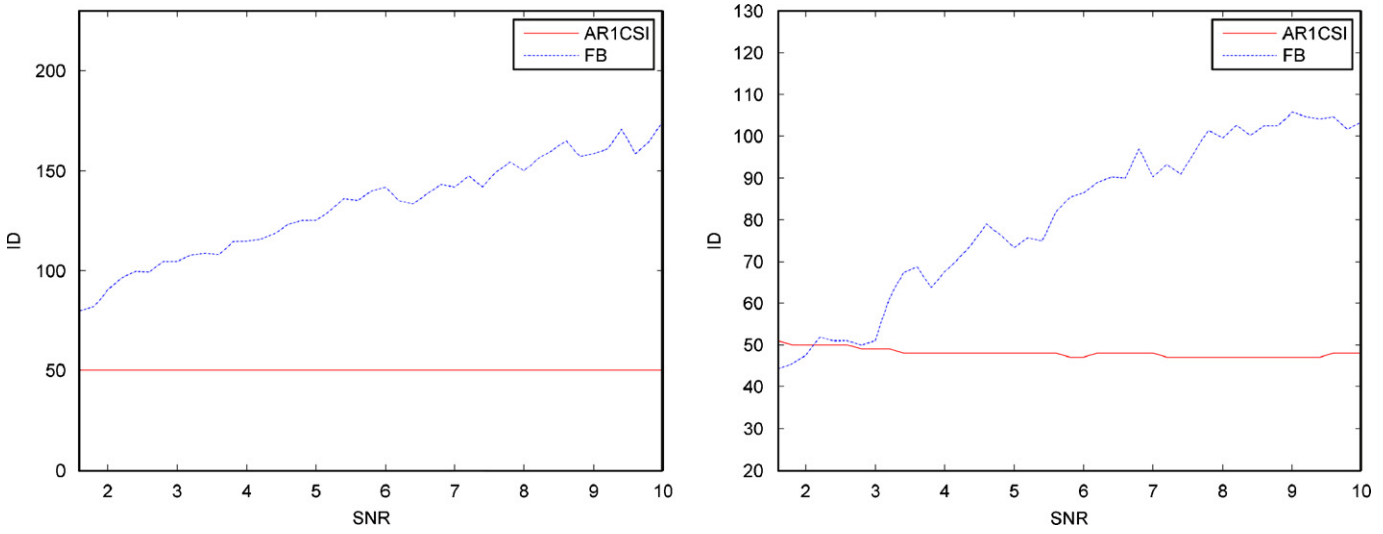

Fig. 8. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the SNR. $T=160$ and $\phi=0.2$. The dot–dash line is the dimension estimated by FB method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

respectively and the IDs were estimated as a function of the temporal data size.

3.1.6. ARMA(4,4) model of the noise (Figs. 13 and 14)

Fig. 13 shows the comparison of the estimated IDs as a function of the temporal size of the data between AR1 and AR1CSI method. Though AR1 method slight underestimate the dimensionality of the data at 20,000 voxels, it has a significant underestimate of the dimensionality at 2000 voxels whereas AR1CSI method almost produces stable and accurate estimates at 20,000 voxels and slight underestimate at 2000 voxels. Fig. 14 shows the comparison between FB and AR1CSI method. As in AR(1) noise simulated data, FB method always overestimates the number of signals for 20,000 voxels and underestimates the model order for 2000 voxels.

3.2. For resting-state fMRI dataset

For real fMRI dataset, before we get the ID estimation, we may do not know where the tail eigenvalues are. Therefore, there is an important problem how to identify the tail eigenvalues of the dataset. As discussion before, if p components are correspondence to signals, the rest of $T-p$ eigenvalues in covariance matrix will be regard as noise parts, i.e., these $T-p$ eigenvalues comprise tail eigenvalues. Thus, we can estimate the ID using iteration until the first number of tail eigenvalue is ID+1. During iteration, it is important that they should be damped to avoid numerical oscillations that arise in some circumstances. The first number of tail eigenvalue is set to the damping factor μ times its value from the previous iteration plus $(1-\mu)$ times its prescribed updated value, where the damping factor μ is between 0 and 1. Here, we used the damping factor of $\mu=0.5$.

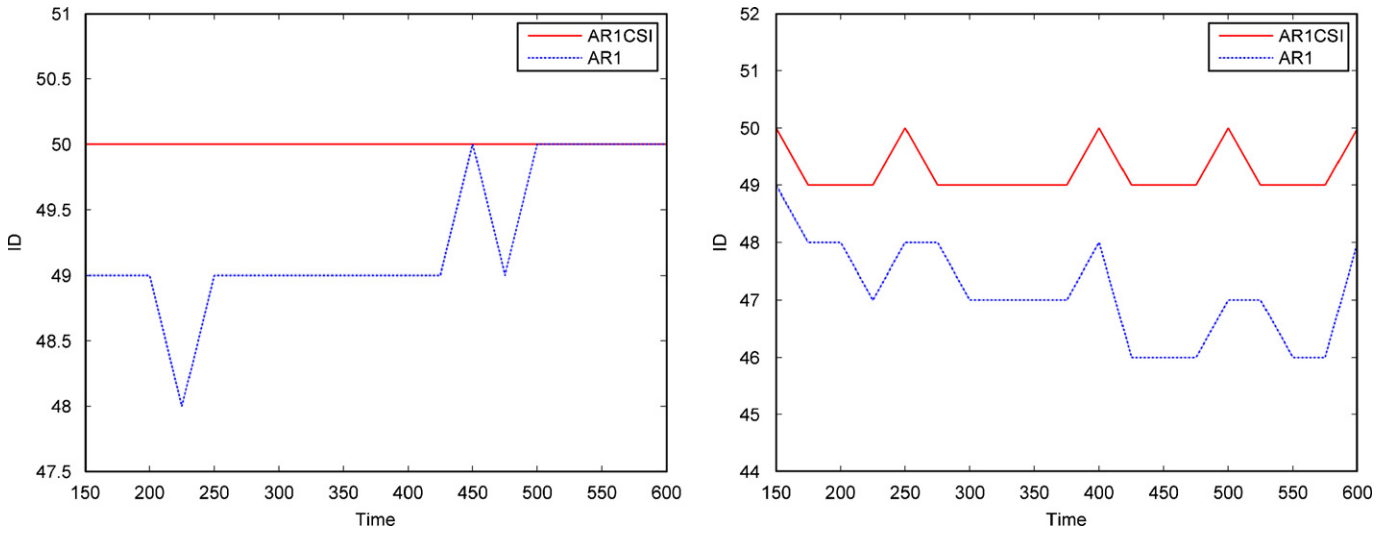

Fig. 9. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the temporal size of the data. $\phi = 0.2$ and SNR = 2. The dot–dash line is the dimension estimated by AR1 method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

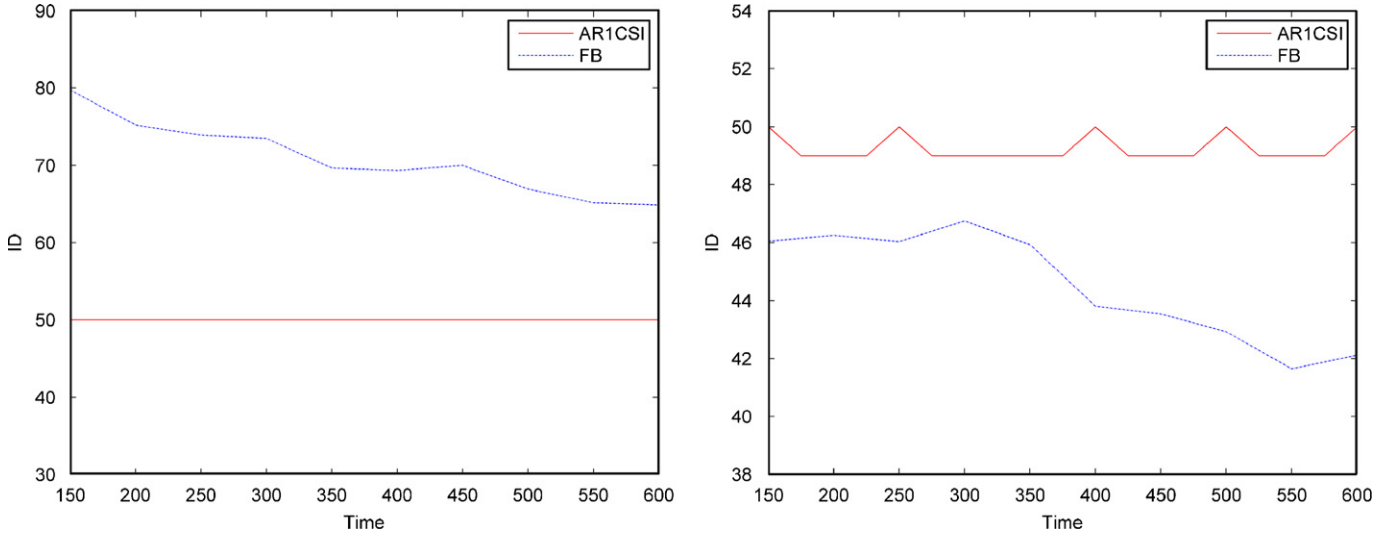

Fig. 10. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the temporal size of the data. $\phi = 0.2$ and SNR = 2. The dot–dash line is the dimension estimated by FB method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

Real resting-state fMRI dataset of nine human brains were investigated where the results were presented for the dataset using truncating from the 150 time frames in the front part of dataset to the full dataset with a temporal size of 600.

A plot of the ID means among all subjects varying with temporal size of the datasets is shown in Fig. 15. The IDs were estimated by FB method, AR1 method and AR1CSI method, respectively, where the AR(1) coefficient was determined from fitting the tail spectra of the datasets. Fig. 15 shows the ID estimated by AR1 method is lower than by AR1CSI method. On the other hand, the ID of the dataset increases with the temporal size of the datasets. For the temporal size of the datasets $T<350$, the variation of the ID is rapidly. For the temporal size of the datasets $T>350$, the slow rate of curvature of the ID is observed. For the ID estimated by AR1CSI method, the shifts are more than 10 units with the temporal size of the datasets T from 150 to 350. For the temporal size of the datasets T from 350 to 600, however, it appears to be shifted by no more than 6 units. It is identified with the recent report that in real resting-state fMRI dataset, although the true dimension may increase when more data are acquired in time, the dimensionality does not grow about linearly with the number of time frames used in the data [10]. Moreover, Fig. 15 shows that the IDs estimated by FB method are much larger than the IDs estimated by AR1 and AR1CSI method.

Table 3 shows the IDs, estimated by AR1 and AR1SCI method respectively, of real full fMRI datasets with the temporal size of 600, which were acquired from nine subjects in resting-state.

Though AR1 method is introduced to detect robustly the number of components [10] in fMRI dataset, it may provide

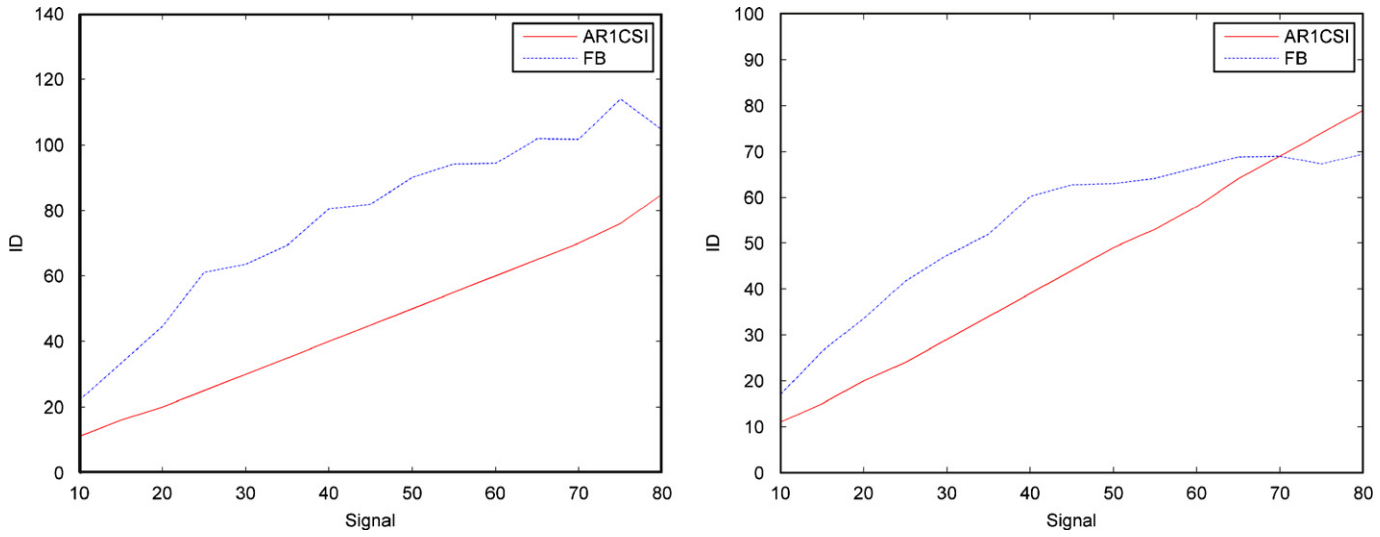

Fig. 11. Estimated dimension for simulated data with a few signals and correlated noise corresponding to AR(1) model as a function of the number of signals. $T = 160$ and $\phi = 0.2$. The dot–dash line is the dimension estimated by FB method. The solid line represents the dimension estimated by AR1CSI method. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

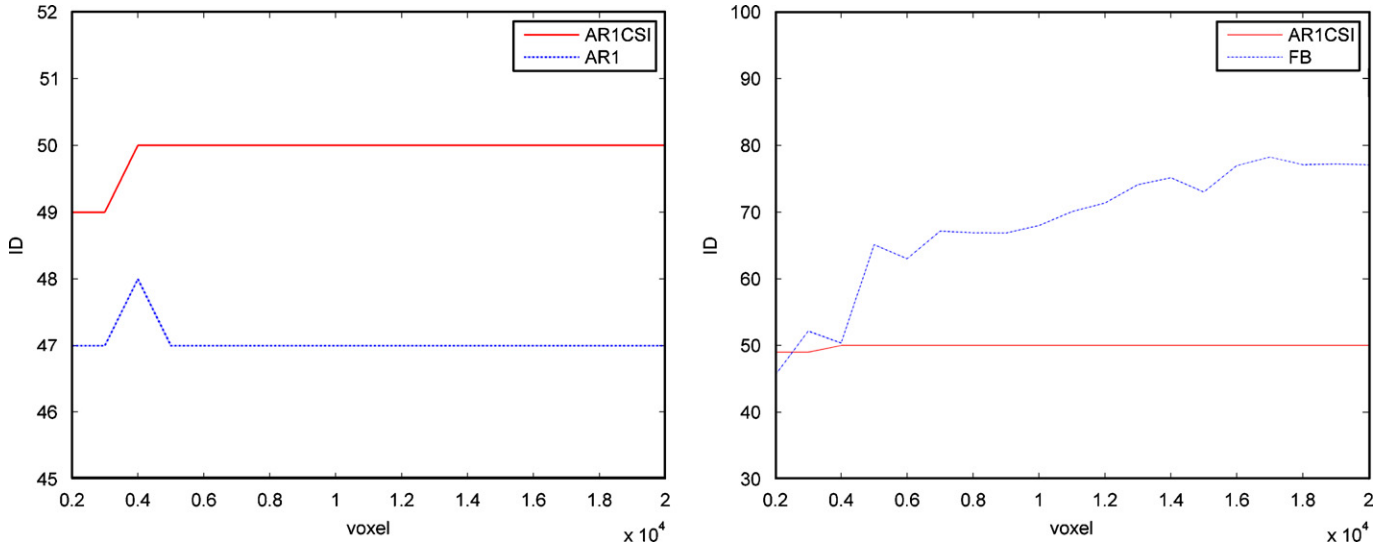

Fig. 12. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to AR(1) model as a function of the number of voxels. $T = 160$, SNR = 2 and $\phi = 0.2$. The solid line represents the dimension estimated by AR1CSI method. In the left plot, the dot–dash line is the dimension estimated by AR1 method. In the right plot, the dot–dash line is the dimension estimated by FB method.

underestimation to the ID at many circumstances, especially at the condition of small number of voxels. Fig. 2 shows that there is difference of the slope between two exponent functions due to their different parameter a, even they have the same parameter b. Furthermore, the larger the difference of parameter a is, the larger the slope difference of two exponent functions is. Because AR1 method estimates coefficient ϕ using only b (overlooking the coefficient a), it will overestimate the slope of eigenvalues for pure AR(1) noise, as shows in Fig. 3. This is the principal cause for AR1 method to underestimate the ID of data. For AR1CSI method, estimation of coefficient ϕ is based on the slope coincidence of the tail eigenvalues between real data and its correspondent pure AR(1) noise, as also shows in Fig. 3. It not only takes b into account, but also uses a to estimate the coefficient ϕ, and thus can estimate the slope of the AR(1) noise eigenvalues more accurately than AR1 method. Hence, it can get the more accurate estimation of ID than AR1 method.

4. Conclusions

In this paper, the proper interpolation variable ($x = ab\,e^{bk}$) is constructed and the method incorporating the autoregressive noise model of order 1 with cubic spline interpolation is introduced to estimate the ID of fMRI data for human brain.

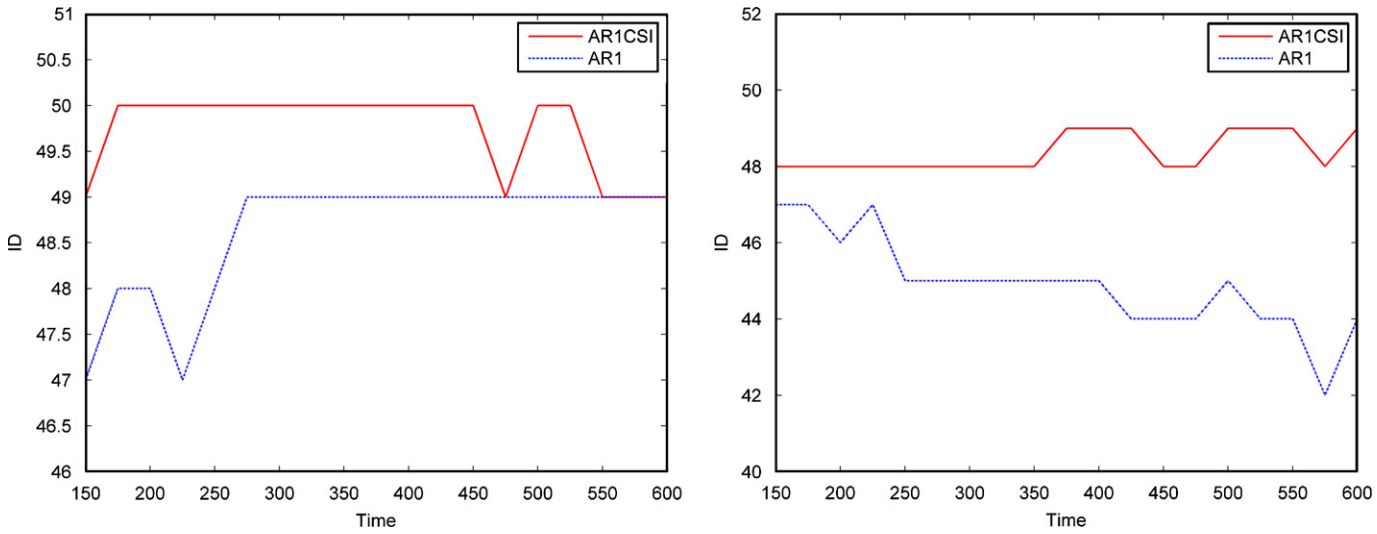

Fig. 13. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to an ARMA(4,4) model as a function of the temporal size of the data. SNR = 2. The dot–dash line is the dimension estimated by AR1 method. The solid line represents the dimension estimated by AR1CSI. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

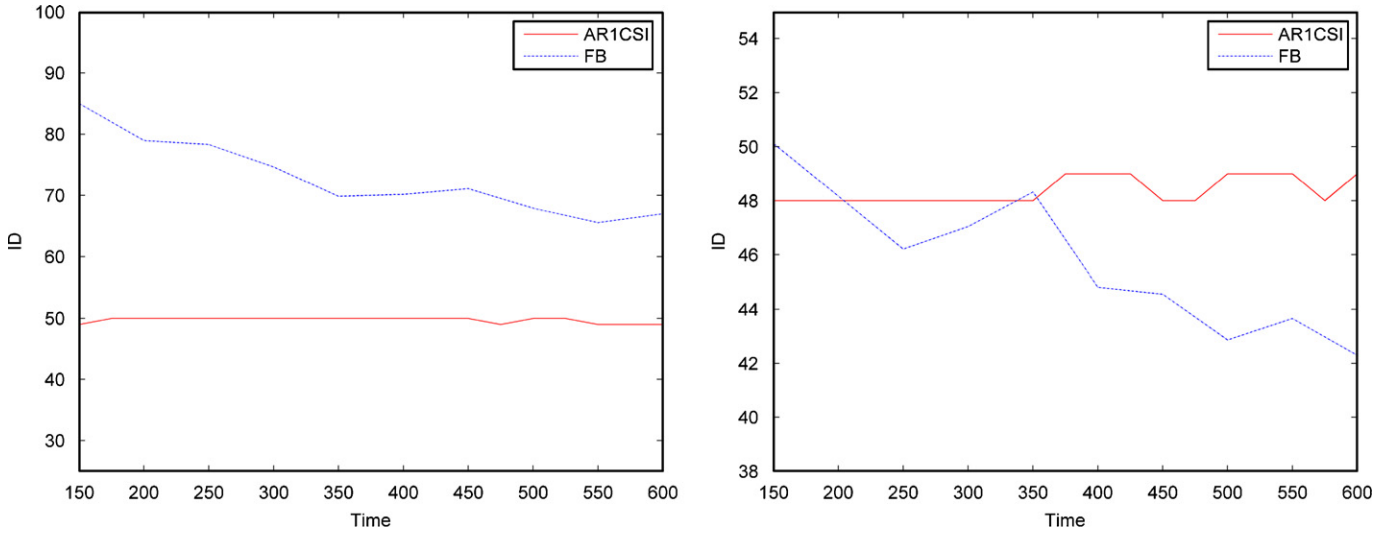

Fig. 14. Estimated dimension for simulated data with 50 signals and correlated noise corresponding to an ARMA(4,4) model as a function of the temporal size of the data. SNR = 2. The dot–dash line is the dimension estimated by FB method. The solid line represents the dimension estimated by AR1CSI. The number of voxels in left plot is 20,000, and the number of voxels in right plot is 2000.

Table 2
Comparison of ID estimated by AR1 and AR1CSI increasing the number of signals at 20,000 voxels and 2000 voxels

The estimated ID and the difference between estimated ID and the number of signals															
The number of signals	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
AR1(20,000)	11 (1)	15 (0)	19 (−1)	24 (−1)	29 (−1)	33 (−2)	38 (−2)	42 (−3)	47 (−3)	51 (−4)	56 (−4)	62 (−3)	67 (−3)	72 (−3)	78 (−2)
AR1CSI(20,000)	11 (1)	16 (1)	20 (0)	25 (0)	30 (0)	35 (0)	40 (0)	45 (0)	50 (0)	55 (0)	60 (0)	65 (0)	70 (0)	76 (1)	85 (5)
AR1(2000)	10 (0)	15 (0)	19 (−1)	22 (−3)	27 (−3)	31 (−4)	36 (−4)	41 (−4)	46 (−4)	51 (−4)	56 (−4)	62 −3	67 (−3)	74 (−1)	79 (−1)
AR1CSI(2000)	11 (1)	15 (0)	20 (0)	24 (−1)	29 (−1)	34 (−1)	39 (−1)	44 (−1)	49 (−1)	53 (−2)	58 (−2)	64 (−1)	69 (−1)	74 (−1)	79 (−1)

Note: The number in the parenthesis is the difference between estimated ID and the number of signals. $T = 160$, SNR = 2 and $\phi = 0.2$.

Unlike the AR1 method that estimates ϕ using only b (neglecting the effect of coefficient a), the AR1CSI method determines the autoregressive coefficient ϕ using cubic spline interpolation based on the function $ab\,e^{bk} = f(\phi)$. It is proved by simulated data and real fMRI dataset that the AR1CSI method seems to work better and leads to more accurate estimation than the AR1 and FB method, no matter what the voxel number, the temporal size of the data, the signal number and the SNR are. Comparing with AR1 and FB method, the performance for estimating the model order in fMRI dataset can be improved by AR1CSI method.

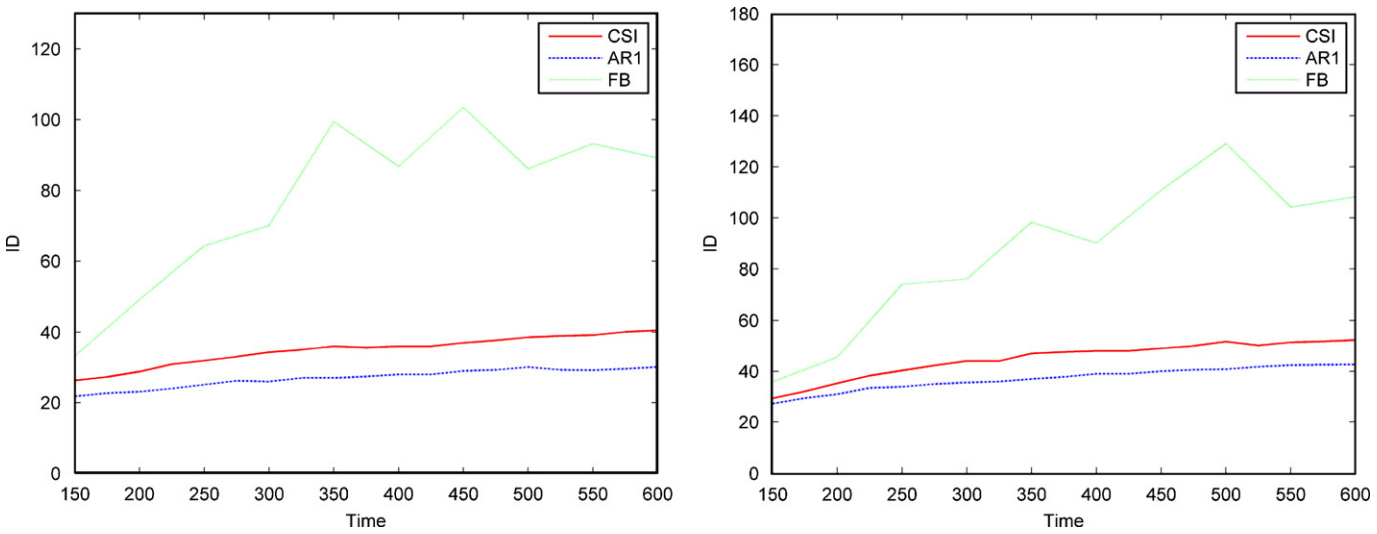

Fig. 15. Estimated dimensionality for real resting-state fMRI dataset as a function of the temporal size of the data. The broken line is the dimensionality estimated by FB method. The dot–dash line is the dimensionality estimated by AR1 method. The solid line represents the dimension estimated by AR1CSI method. The left plot is the mean dimensionality of slices covering the motor cortex among all subjects and the right plot is the mean dimensionality of slices covering the visual cortex among all subjects.

Table 3
The IDs comparison of fMRI estimated by AR1 and AR1CSI

Subject	1	2	3	4	5	6	7	8	9
AR1 (SCMC)	22	23	25	40	30	25	26	40	40
AR1CSI (SCMC)	32	35	40	49	41	38	37	52	41
AR1 (SCVC)	37	31	46	43	52	45	44	43	42
AR1CSI (SCVC)	46	40	57	56	58	53	60	51	50

Note: SCVC means five transection slices covering the visual cortex and SCMC means five transection slices covering the motor cortex.

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China (Grant no. 50577057) and Key Project of Humanities and Social Sciences, Ministry of Education of China (No. 07JZD0029).

References

[1] H. Akaike, Information theory and an extension of maximum likelihood principle, in: Proceedings of the 2nd International Symposium on Information Theory, Akademia Kiado, Budapest, 1973, pp. 267–281.
[2] C.F. Beckmann, S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging, IEEE Trans. Med. Imaging 23 (2004) 137–152.
[3] C.M. Bishop, Bayesian PCA, in: M.S. Kearns, S.A. Solla, D.A. Cohn (Eds.), Advances in Neural Information Processing Systems, vol. 11, 1999, pp. 382–388.
[4] C.M. Bishop, Variational principal components, in: Proceedings of the Ninth International Conference on Artificial Neural Networks, ICANN'99, vol. 1, 1999, pp. 509–514.
[5] C.D. Boor, A Practical Guide to Splines, Springer, New York, 1978.
[6] R.L. Burden, J.D. Faires, Numerical Analysis, Higher Education Press and Thomson Learning Inc., Beijing, 2001.
[7] F. Camastra, Data dimensionality estimation methods: a survey, Pattern Recognit. 36 (2003) 2945.
[8] F. Camastra, A. Vinciarelli, Estimating the intrinsic dimension of data with a fractal-based method, IEEE Trans. Pattern Anal. Mach. Intell. 24 (10) (2002) 1404–1407.
[9] P. Common, Independent component analysis: a new concept?, Signal Process. 24 (1) (1991) 1–10.
[10] D. Cordes, R.R. Nandy, Estimation of the intrinsic dimensionality of fMRI data, NeuroImage 29 (2006) 145–154.
[11] B.S. Everitt, Cluster Analysis, third ed., Edward Arnold, 1993.
[12] K. Fukunaga, D.R. Olsen, An algorithm for finding intrinsic dimensionality of data, IEEE Trans. Comput. 20 (2) (1976) 165–171.
[13] C. Goutte, P. Toft, E. Rostrup, F.A. Nielsen, L.K. Hansen, On clustering fMRI time series, NeuroImage 9 (3) (1999) 298–310.
[14] L. Harrison, W.D. Penny, K. Friston, Multivariate autoregressive modeling of fMRI time series, NeuroImage 19 (2003) 1477–1491.
[15] R. Hegger, H. Kantz, Practical implementation of nonlinear time series methods: the TISEAN package, Chaos 9 (1999) 413–435.
[16] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504–507.
[17] P.d.F.A.R. Højen-Sorenson, O. Winther, L.K. Hansen, Analysis of functional neuroimages using ICA with adaptive binary sources, Neurocomputing 49 (2002) 213–225.
[18] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 1997.
[19] D. Kincaid, W. Cheney, Numerical Analysis: Mathematics of Scientific Computing, third ed., Wadsworth Group and China Machine Press, Beijing, 2003.
[20] T. Martinetz, K. Schulten, Topology representing networks, Neural Networks 3 (1994) 507–522.
[21] V. Maxim, L. Sendur, J. Fadili, J. Suxkling, R. Gould, R. Howard, E. Bullmore, Fractional Gaussian noise, functional MRI and Alzheimer's disease, NeuroImage 25 (2005) 141–158.
[22] T.P. Minka, Automatic choices of dimensionality for PCA, Technical Report 514, MIT, 2000.
[23] K. Pettis, T. Bailey, T. Jain, R. Dubes, An intrinsic dimensionality estimator from near-neighbor information, IEEE Trans. Pattern Anal. Mach. Intell. 1 (1) (1979) 25–37.
[24] J. Reidl, J. Strrke, D.B. Omer, A. Grinvald, H. Spors, Independent component analysis of high-resolution imaging data identifies distinct functional domains, NeuroImaging 34 (2007) 94–108.
[25] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[26] A.K. Rommey, R.N. Shepard, S.B. Nerlove, Multidimensional Scaling. Vol. 1: Theory, Seminar Press, New York, 1992.
[27] A.K. Rommey, R.N. Shepard, S.B. Nerlove, Multidimensional Scaling. Vol. 2: Application, Seminar Press, New York, 1992.
[28] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[29] G. Schwartz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[30] M. Small, Applied Nonlinear Time Series Analysis: Application in Physics, Physiology and Finance, World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.
[31] J.B. Tenenbaum, V.d. Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[32] G.V. Trunk, Statistical estimation of the intrinsic dimensionality of a noisy signal collection, IEEE Trans. Comput. 25 (1976) 161–171.

[33] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust., Speech, Signal Process. 33 (1985) 387–392.

Xiaoping Xie was born in Zhejiang Province, China. He received the B.S. degree in electronics from Hangzhou University, Hangzhou, China, the M.S. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 1986 and 1997, respectively. Now, he is working for the Ph.D. degree in Science from Zhejiang University, Hangzhou, China. His current research interests include fMRI data analysis, image processing and signal processing.

Zhitong Cao received his M.S. degree from Hefei University of Technology, Hefei, China in 1981. He is a professor of the Department of Physics, Zhejiang University, China. His main interests and research activities are in calculation and image of neural information in the field of brain, artifical intelligence including statistical learning theory, chaotic neural network, and pattern identification.

Xuchu Weng received the B.S. degree in Traditional Chinese Medicine from Zhejiang College of Traditional Chinese Medicine, Hangzhou, China, the M.S. degree and Ph.D. in Biological Psychology from Chinese Academy of Science, Beijing, China, in 1985, 1992 and 1995, respectively. Now, he is a professor of Neuroscience and Psychology, Laboratory for Higher Brain Function, Institute of Psychology, Chinese Academy of Sciences.

Dan Jin born in Zhejiang, China, receiving bachelor's degree from Zhejiang University, now studies on neuroinformatics as postgraduate in Applied Physics Institute, Zhejiang University. Major interest is data analysis of fMRI.