

Estimating the Number of Independent Components for Functional Magnetic Resonance Imaging Data

Yi-Ou Li,^{1*} Tülay Adalı,¹ and Vince D. Calhoun^{2,3}

¹Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, Maryland 21250

²Olin Neuropsychiatry Research Center, Institute of Living, Hartford, Connecticut 06106

³Yale University School of Medicine, New Haven, Connecticut 06520

Abstract: Multivariate analysis methods such as independent component analysis (ICA) have been applied to the analysis of functional magnetic resonance imaging (fMRI) data to study brain function. Because of the high dimensionality and high noise level of the fMRI data, order selection, i.e., estimation of the number of informative components, is critical to reduce over/underfitting in such methods. Dependence among fMRI data samples in the spatial and temporal domain limits the usefulness of the practical formulations of information-theoretic criteria (ITC) for order selection, since they are based on likelihood of independent and identically distributed (i.i.d.) data samples. To address this issue, we propose a subsampling scheme to obtain a set of effectively i.i.d. samples from the dependent data samples and apply the ITC formulas to the effectively i.i.d. sample set for order selection. We apply the proposed method on the simulated data and show that it significantly improves the accuracy of order selection from dependent data. We also perform order selection on fMRI data from a visuomotor task and show that the proposed method alleviates the over-estimation on the number of brain sources due to the intrinsic smoothness and the smooth preprocessing of fMRI data. We use the software package ICASSO (Himberg et al. [2004]: Neuroimage 22:1214–1222) to analyze the independent component (IC) estimates at different orders and show that, when ICA is performed at overestimated orders, the stability of the IC estimates decreases and the estimation of task related brain activations show degradation. *Hum Brain Mapp* 28:1251–1266, 2007. ©2007 Wiley-Liss, Inc.

Key words: ICA; fMRI; order selection; sample dependence; entropy rate; subsampling

INTRODUCTION

Functional magnetic resonance imaging (fMRI) data have been analyzed successfully by multivariate methods such as independent component analysis (ICA) to explore brain function (Biswal and Ulmer, 1999; Calhoun et al., 2001; McKeown et al., 1998). Because of the high dimensionality and high noise level of fMRI data, applying ICA on the full spatial or temporal dimension is likely to overfit the data and thus degrades the ICA estimation (Sarela and Vigario, 2003). Therefore, the number of informative components is often assumed to be less than the spatial or temporal dimension of the fMRI data. A lower dimensional subspace

Contract grant sponsor: NIH; Contract grant number: 1 R01 EB 000840-02.

*Correspondence to: Yi-Ou Li, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250. E-mail: liyiou1@umbc.edu

Received for publication 20 February 2006; Revised 4 August 2006; Accepted 7 September 2006

DOI: 10.1002/hbm.20359

Published online 1 February 2007 in Wiley InterScience (www.interscience.wiley.com).

©2007 Wiley-Liss, Inc.

containing the informative sources thus needs to be identified prior to ICA and this step has important implications in the final results of ICA as discussed by Beckmann and Smith (2004), Calhoun et al. (2001), and McKeown (2000).

ICA can be performed to either estimate the independent spatial maps (spatial ICA) or the independent time courses (temporal ICA) of the fMRI data (Calhoun et al., 2003). In this work, the more commonly used spatial ICA approach is adopted, with the generative model:

$$\mathbf{X} = \sum_{k=1}^M \mathbf{a}_k \mathbf{s}_k^T + \mathbf{n} \quad (1)$$

where \mathbf{s}_k is an $N \times 1$ vector containing the voxel values, i.e., the intensity of activations at each voxel, for the k th independent spatial map, \mathbf{a}_k is a $T \times 1$ vector expressing the temporal dynamics (time course) of the k th independent brain activation, M is the number of informative brain sources (a set of spatial maps and their time courses) contained in the data, i.e., the order of the fMRI data, and \mathbf{n} is a $T \times N$ matrix of Gaussian noise, which can be explicitly incorporated into the estimation, as in the noisy ICA model, or neglected as in the noiseless ICA model used in this work. It is worth noting that *within each independent spatial map*, the voxel values are *not* independent in general, which is the issue we address in this work in the context of order selection.

Among different approaches for model order selection, the information-theoretic criteria (ITC) have proven to be particularly attractive for many signal processing applications. Since ITC do not require the specification of an empirical threshold for order selection, they fit naturally into the framework of exploratory data analysis methods such as ICA. A commonly used ITC for order selection, Akaike's information criterion (AIC), is developed based on the minimization of the Kullback–Leibler divergence between the true model and the fitted model (Akaike, 1973). AIC is extended by Cavanaugh as the Kullback–Leibler information criterion (KIC) (Cavanaugh, 1999) using a symmetric Kullback–Leibler divergence between the true and fitted models. The minimum description length (MDL) criterion and the Bayesian information criterion (BIC) are developed based on the minimum code length (Rissanen, 1978) and the Bayes solution to model order selection (Schwartz, 1978), respectively.

The practical formulations of ITC are developed by Wax and Kailath (1985) in the context of detecting the number of signals in noise where both the signals and the noise are modeled by multidimensional complex stationary Gaussian random processes. Wax and Kailath's formulations are directly applicable to the multivariate order selection problem and have been used in estimating the number of latent sources in blind source separation (Karhunen et al., 1997) as well as in ICA of fMRI data (Calhoun et al., 2001; Beckman and Smith, 2004).

Wax and Kailath's order selection formulations are derived based on the i.i.d. sample assumption. When the for-

mulations are applied in the spatial ICA model for order selection on fMRI data, it is assumed that the voxel values of the spatial maps are i.i.d. In fact, however, there is inherent spatial smoothness due to the point spread function of the scanner. Furthermore, smoothing is a common preprocessing step used to suppress the high frequency noise in the fMRI data and to minimize the impact of spatial variability among subjects. Both factors contribute to dependence among the samples in fMRI volume data, thus, weaken the i.i.d. sample assumption. When the dependent samples are taken as if they were i.i.d. samples for order selection, as we show in the next section, Wax and Kailath's formulations tend to over-estimate the order. Overestimation on the dimension of the signal subspace in fMRI data could result in splitting of the informative sources in ICA estimation (Beckman and Smith, 2004), making the interpretation of the ICA results difficult and thus limiting its utility.

To address the voxel-wise sample dependence, we model the fMRI volume data by a three-dimensional (3D) finite-order moving average (MA) process in the spatial domain. Since the local spatial dependence in the fMRI data is due to the MRI spatial point spread function as well as spatial correlation induced by the hemodynamic sources being measured, its effective span does not extend beyond few voxels (Calhoun, 2002; Menon and Goodyear, 1999; Parkes et al., 2005). The spatial distribution of the fMRI signal is often modeled as a Gaussian random field (Friston et al., 1996; Worsley and Friston, 1995), which is a specific type of MA model. A moving average process is the output of a linear system, in this case, a smoothing filter, having an i.i.d. Gaussian process as the input. Assume that the linear system is shift invariant, the resulting moving average process is a stationary Gaussian random process. Based on this model, we propose a subsampling scheme on the fMRI volume data to identify an *effectively i.i.d.* sample set, i.e., determine a grid of locations at which the dependence among the samples is small enough such that it can be easily ignored. Specifically, an information-theoretic concept, entropy rate, is used to measure the sample dependence in the stationary Gaussian process to control the subsampling scheme. By comparing the entropy rate of the subsampled data with that of an i.i.d. Gaussian process, we infer the grid of locations on which the data samples can be considered to be effectively i.i.d.

Once the effectively i.i.d. samples are identified, Wax and Kailath's order selection formulations can be applied on the identified sample set without violating the underlying i.i.d. sample assumption.

In the next section, we discuss details of ITC for order selection and develop the entropy rate matching principle applied to identify the effectively i.i.d. sample set. In Experiments section, we show experimental results of the proposed scheme on order selection of both the simulated data and the fMRI data from a visuomotor test. Furthermore, we study the impact of order selection on the subsequent ICA estimation. We conclude the work with a discussion of the results.

METHODS

Information-Theoretic Criteria for Order Selection

The formulas for AIC, KIC, and MDL criterion all have similar structures:

$$E_{\text{AIC}}(k) = -2L(\mathbf{x}|\Theta_k) + 2G(\Theta_k) \quad (2)$$

$$E_{\text{KIC}}(k) = -2L(\mathbf{x}|\Theta_k) + 3G(\Theta_k) \quad (3)$$

$$E_{\text{MDL}}(k) = -L(\mathbf{x}|\Theta_k) + \frac{1}{2}G(\Theta_k) \log N \quad (4)$$

where $L(\mathbf{x}|\Theta_k)$ is the maximum log-likelihood of the observations, \mathbf{x} , based on the model parameter set Θ_k of the k th order and $G(\Theta_k)$ is the penalty for model complexity given by the total number of free parameters in Θ_k . For MDL, the penalty term is scaled by $\log N$, where N is the sample size.

The original Wax and Kailath's order selection formulations are derived for complex-valued data, the adaptation to the real-valued case is straightforward. The maximum log-likelihood is given by

$$L(\mathbf{x}|\Theta_k) = \frac{N}{2} \log \left(\frac{\prod_{i=k+1}^T \lambda_i^{1/(T-k)}}{\frac{1}{T-k} \sum_{i=k+1}^T \lambda_i} \right)^{(T-k)} \quad (5)$$

where T is the original dimension of the multivariate data, k is the candidate order, N is the sample size, and λ_i s are the eigenvalues of the sample covariance matrix of the multivariate observations. The number of free parameters in Θ_k for the real-valued data is given by

$$G(\Theta_k) = 1 + Tk - \frac{1}{2}k(k-1).$$

When dependent samples are used, the actual number of i.i.d. sample is less than N , as a result, the likelihood term given by Eq. (5) improperly dominates the ITC criteria (for MDL, even though N is included in the penalty, since its effect is through the logarithm of N , it has a slower growing rate compared to the N as a scaling factor in the likelihood), resulting in an over-estimation of the order. Figure 1 shows the spectra of the negative log-likelihood of eleven spatially unsmoothed and smoothed fMRI data sets. It is observed that the log-likelihood of the smoothed fMRI data is "inflated" due to the increased sample dependence. Details of the data sets are discussed in the Experiments section.

One way to address the above problem is to identify a set of i.i.d. samples within the original data and use the i.i.d. samples for order selection. Hence, a feasible statistical model of the sample space is needed for the i.i.d. sample identification. We propose to model the fMRI volume data as a finite-order MA process, i.e., a stationary Gaussian random process.

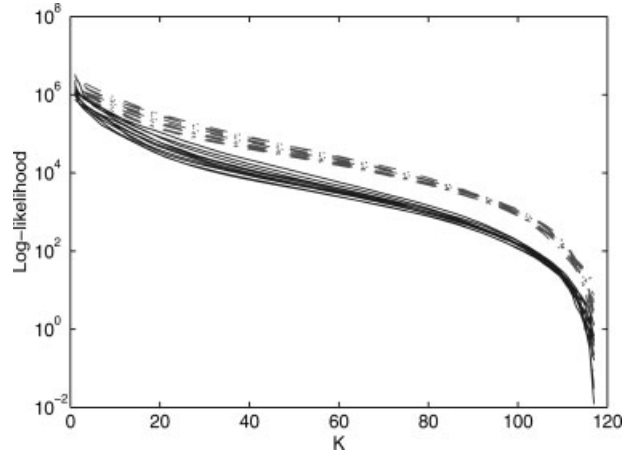


Figure 1.

Comparison of the negative log-likelihood between eleven unsmoothed (solid) and smoothed (dash-dot) fMRI data sets for different candidate order K .

Stationary Gaussian Process and Its Properties

Let $x[n]$, $n = 1, 2, \dots, N$ be a stationary Gaussian random sequence. Without loss of generality, we assume that $x[n]$ has zero mean and unit variance, i.e., $E\{x[n]\} = 0$ and $E\{x^2[n]\} = 1$. The entropy rate of $x[n]$, a measure of the amount of information carried by each sample of a Gaussian random sequence, is given by (Papoulis, 1991)

$$h_x = \ln \sqrt{2\pi e} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln s(\omega) d\omega \quad (6)$$

where $s(\omega)$ is the power spectral density function of the Gaussian random sequence $x[n]$. It is a differential entropy measure due to the continuous nature of the Gaussian distribution.

The entropy rate can be used to measure the autocorrelation of a Gaussian random process. For example, when an i.i.d. Gaussian process is smoothed, its entropy rate decreases. This is intuitively meaningful, since the smoothing operation also filters out part of the information carried by the data. We generate a two-dimensional (2D) i.i.d. Gaussian process and filter it by 2D Gaussian smoothing kernels with full-width half-maximum (FWHM) values 1×1 , 2×2 , and 3×3 voxel(s). The estimated entropy rate of the original process is 1.41, the estimated entropy rates of the three smoothed processes are, respectively, 1.40, 0.67, and -0.17 . In this example, the entropy rate measures the average smoothness in the sample space, or, in other words, the average sample dependence. High sample dependence results in low entropy rate of the process.

In our development, the following observation plays the key role:

The entropy rate of a stationary Gaussian random sequence with unit variance is upper bounded by $\ln \sqrt{2\pi e}$ and the upper

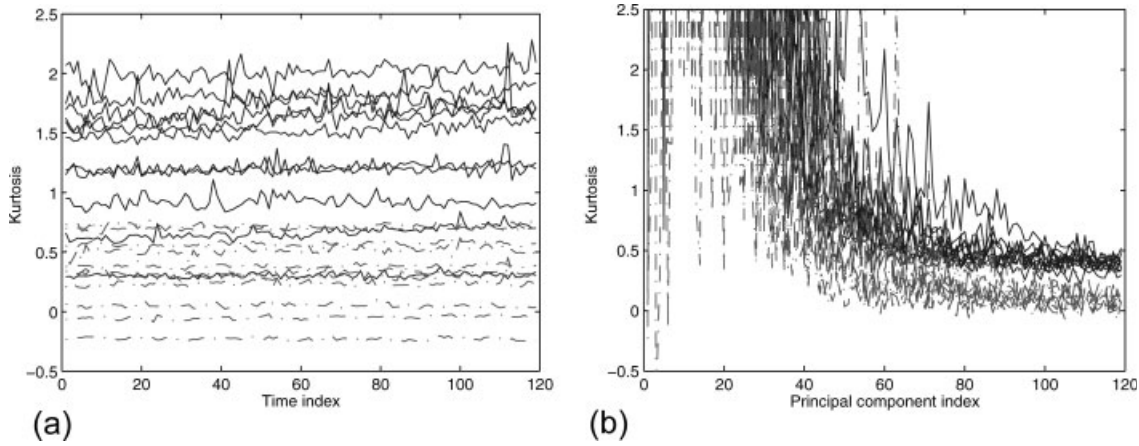


Figure 2.

Kurtosis distribution of the spatially unsmoothed (solid) and smoothed (dash-dot) fMRI volume data (Panel a) and the kurtosis distribution of the corresponding principal components (Panel b), for the eleven fMRI data sets.

bound is achieved if and only if the sequence is a white Gaussian sequence, i.e., all the samples of the sequence are i.i.d.

The above argument can be verified by using the log inequality $\ln(x) \leq x - 1$, where $x > 0$, such that

$$\int_{-\pi}^{\pi} \ln s(\omega) d\omega \leq \int_{-\pi}^{\pi} s(\omega) d\omega - \int_{-\pi}^{\pi} 1 d\omega = 0. \quad (7)$$

The last equality in Eq. (7) is due to the fact that $x[n]$ has unit variance, i.e.

$$\sigma_x^2 = r[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} s(\omega) d\omega = 1.$$

We have equality in Eq. (7) if and only if $s(\omega) \equiv 1$, i.e., when $x[n]$ is a white Gaussian sequence.

As an extension of the argument, the entropy rate of any stationary random sequence with a continuous distribution is upper bounded by $\ln \sqrt{2\pi e}$. This is true because among all continuous probability distributions with equal variance, Gaussian distribution achieves the maximum entropy.

The second observation that leads us to the sampling scheme proposed in the next section can be stated as:

Assume that a stationary Gaussian random sequence $x[n]$ has an autocorrelation function of finite length, i.e., $r[m] = E\{x[n]x[n+m]\} = 0$ for $|m| \geq L$, the subsampled sequence $x_s[n] = x[Ln]$ is a white Gaussian random sequence.

To observe this, let $r_s[m]$ be the autocorrelation function of the subsampled sequence $x_s[n]$

$$r_s[m] = E\{x_s[n]x_s[n+m]\} = E\{x[Ln]x[Ln+Lm]\} = r[Lm]$$

since $x[n]$ is stationary. Because $r[m]$ is of finite length, we have $r_s[m] = r[Lm] = \delta[m]$ where $\delta[m]$ is the Kronecker delta function that assumes value “1” at $m = 0$ and “0”

otherwise. Therefore, the subsampled sequence $x_s[n]$ is a white Gaussian sequence.

In our work, the spatial dependence in the fMRI data is modeled by a finite length correlation in the sample space, i.e., a finite order moving average process, which is reasonable because the major factors contributing to sample dependence in the fMRI data are localized (Grinvald et al., 1994; Menon and Goodyear, 1999; Parkes et al., 2005).

Entropy Rate Matching Principle for Identification of the Effectively i.i.d. Samples

Entropy rate matching principle and its applicability

The properties discussed in last section indicate that (i) the entropy rate upper bound can be used to identify an i.i.d. Gaussian sequence; and (ii) an i.i.d. Gaussian sequence can be obtained by subsampling a finite order moving average sequence. Therefore, we propose the following entropy rate matching principle:

If the estimated entropy rate of a subsampled Gaussian sequence reaches the upper bound of the entropy rate, $\ln \sqrt{2\pi e}$, the subsampled sequence is an i.i.d. sequence.

In general, the fMRI volume data acquired at each time point are not Gaussian distributed. We calculate the normalized kurtosis values of the fMRI volume data at each time point from the eleven spatially unsmoothed and smoothed fMRI data sets. We observe that the kurtosis values are typically distributed around a positive value for all the unsmoothed data sets and most of the smoothed data sets, as shown in Figure 2a. The theoretical value of the normalized kurtosis for a Gaussian distribution is zero. When the kurtosis measure is taken on the principal components calculated from the fMRI data, a group of the least significant principal components are observed to have kurtosis

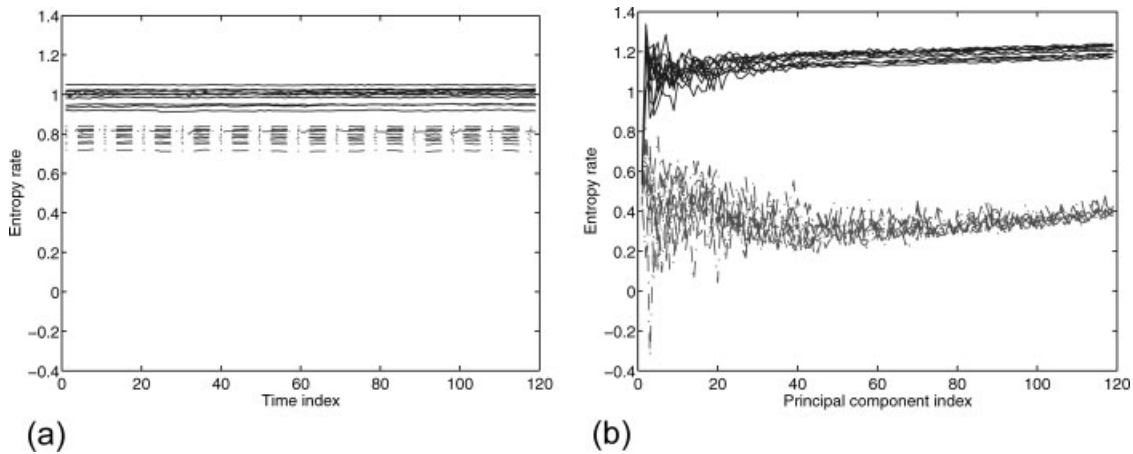


Figure 3.

Entropy rate distribution of the spatially unsmoothed (solid) and smoothed (dash-dot) fMRI volume data (Panel **a**) and the entropy rate distribution of the corresponding principal components (Panel **b**), for the eleven fMRI data sets.

values close to zero as shown in Figure 2b. The normality of the least significant components can also be examined by a statistical test such as the Jarque–Bera test (Judge et al., 1988) or goodness-of-fit to a normal distribution (Conover, 1980).

To study the variation of sample dependence of the fMRI volume data, we directly calculate the entropy rate of the fMRI volume data at each time point by slightly violating the Gaussian assumption. It is observed that entropy rate is invariant with respect to time, as being observed in Figure 3a. This is also true for a large portion of the least significant principal components obtained by PCA from the original volume data as being observed in Figure 3b. For the most significant principal components, the entropy rates assume large variation, a result from the mismatch with the Gaussian distribution assumption. Based on the observation that the entropy rate of the fMRI volume data is stationary in the temporal domain, the effectively i.i.d. samples are inferred from the least significant principal components since their distributions better match the Gaussian assumption.

Therefore, we use PCA to obtain a set of least significant components of the fMRI data and estimate the effectively i.i.d. sample set from the least significant components with the subsampling scheme controlled by the entropy rate matching principle. Since the principal components share the same sample space with the original volume data, once the grid of effectively i.i.d. sample set is identified in the volume, the original fMRI data can be subsampled on the grid and the resulting subsampled data can be considered to be effectively i.i.d.

Procedure for the identification of effectively i.i.d. samples

We first subsample the selected least significant principal components by the smallest subsampling depth $\Delta = 2$, i.e.,

keep every-other sample. Since there is less dependence in the subsampled sequence, the entropy rate of the subsampled sequence increases. We progressively increase the depth of subsampling till the estimated entropy rate of the subsampled sequence reaches its upper bound. At this point, the resulting sequence is closest to a white Gaussian sequence and the resulting subsampling depth defines the grid in the sample space on which the samples are deemed to be effectively i.i.d. Correspondingly, $N_e = N/\Delta$ is the effective number of i.i.d. samples. If subsampling is performed on the 3D fMRI volume data, $N_e = N/\Delta^3$, where N is the total number of in-brain voxels. To improve the estimation, we estimate Δ from a set of least significant principal components and take the average values of those.

Calculation of the entropy rate

To estimate the entropy rate of a Gaussian sequence numerically, summation is used to approximate the integral in Eq. (6), i.e.,

$$h_x \approx \ln \sqrt{2\pi e} + \frac{1}{4\pi} \sum_k \ln \hat{s}(\omega_k) \Delta\omega$$

where $\Delta\omega$ is given by $\Delta\omega = 2\pi/\Sigma\hat{s}(\omega_k)$ since the sequence has unit variance. The power spectral density estimate $\hat{s}(\omega)$ is obtained by taking discrete Fourier transform of the estimated autocorrelation sequence smoothed by a Parzen window (Wei, 1989), i.e.,

$$\hat{s}(\omega_k) = \sum_{m=-N+1}^{N-1} \hat{r}[m]W[m]e^{-j2\pi\frac{m\omega_k}{N}}$$

where ω_k is the discrete frequency index and $W[m]$ is the windowing sequence.

The autocorrelation sequence is estimated by

$$\hat{r}[m] = \frac{1}{N - |m|} \sum_{n=0}^{N-|m|-1} x[n + |m|]x[n].$$

Since the fMRI volume data are 3D, all the computations above are extended into their 3D forms.

Order selection with effectively i.i.d. samples

A set of effectively i.i.d. samples of the original fMRI data is obtained by subsampling the data at the grid resulting from the proposed subsampling scheme. Since the subsampling causes a decrease in the amount of samples for estimating the eigenvalues of the covariance matrix in Eq. (5), an eigenspectrum adjustment (Beckmann and Smith, 2004) is used to mitigate the finite sample effect. It is also worth noting that the eigenspectrum adjustment does not address the effect of sample dependence in order selection. As we show in the simulations, incorporating such an adjustment on the original dependent data can not avoid the over-estimation on the order due to sample dependence. However, this adjustment plays an important role in correcting the finite sample effect that becomes significant due to the subsampling.

The proposed procedure for order selection is summarized in the following pseudo code:

```

X1 = PCA(X);
X2 = X1(kurtosis(X1) < κth);
{κth is the kurtosis threshold of Gaussianity}
[M, N] = sizeof(X2);
for i = 1 to M do
    Δi = 1;
    while h(X2(i)) ≤ ln √(2πε) do
        X2(i) = subsampling(X2(i), Δi);
        Δi = Δi + 1;
    end while
end for
Δ̄ = 1/M ∑i Δi;
X3 = subsampling(X, Δ̄);
{λi}i=1,2,3,...,T = eigenvalue(covariance(X3));
Ne = N/Δ̄3;
{λi}i=1,2,3,...,T = eigenspectrum-adjustment({λi}i=1,2,3,...,T, Ne);
E(k) = Ne/2 log(∏i=k+1T λi1/(T-k) / ∑i=k+1T λi) + β(Tk - 1/2 k(k-1));
{β is the weight of the penalty term specific to the
ITC used}
K = argmin{E(k)};

```

EXPERIMENTS

We perform two sets of experiments: (i) order selection on simulated data and (ii) order selection on fMRI data acquired from subjects performing a visuomotor task. We compare the order selection results based on the original data with the results based on the effectively i.i.d. samples obtained by the proposed subsampling scheme. For both the two approaches, eigenspectrum adjustment (Beckmann and Smith, 2004) is used to compensate the finite sample effect. We also compare the estimated independent components in the subspace of the different selected orders to study the impact of order selection on the ICA estimation.

Order Selection on Simulated Data

We generate eight sources and the associated time courses similar to the ones used by Correa et al. (2005) to create the simulated fMRI data according to the model

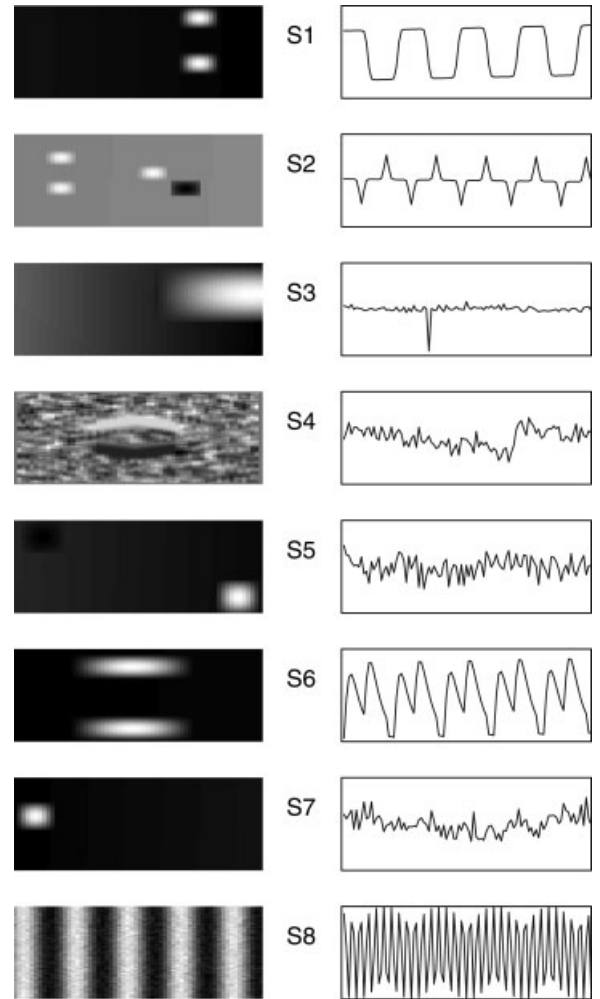


Figure 4.

Eight simulated sources and their time courses.

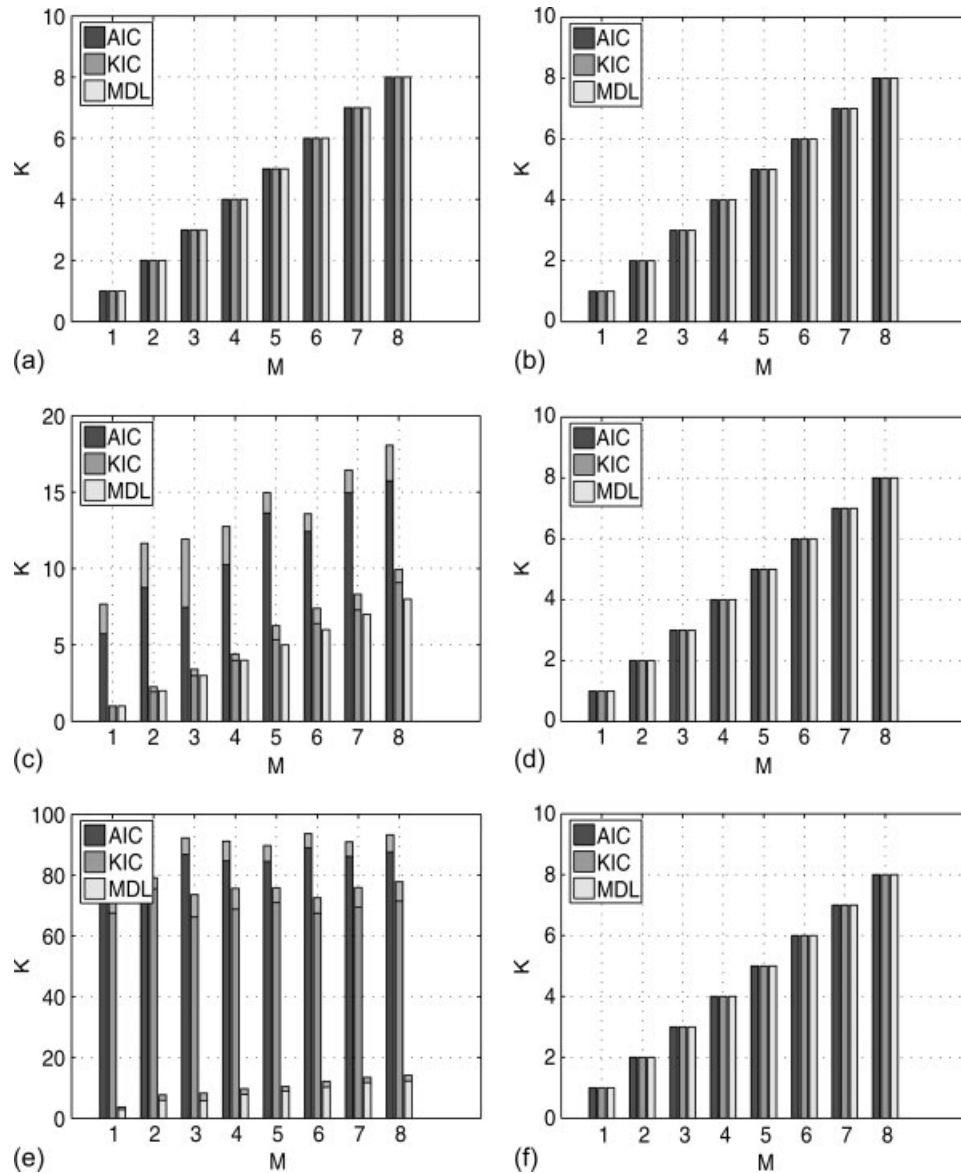


Figure 5.

Order selection on simulated data (CNR = 1 and $T = 100$) at three different smoothness levels: (a) No smoothing, using original data; (b) No smoothing, using effectively i.i.d. samples; (c)

FWHM = 2 voxels, using original data; (d) FWHM = 2 voxels, using effectively i.i.d. samples; (e) FWHM = 3 voxels, using original data; (f) FWHM = 3 voxels, using effectively i.i.d. samples.

described in Eq. (1). Each simulated spatial source is a 60×60 pixels 2D image, each time course is a 100 point waveform simulating the temporal dynamics of the corresponding spatial sources. The spatial sources are rearranged into 1D vectors and mixed by the time courses according to the generative model as Eq. (1), resulting in a 100×3600 spatio-temporal data set. Zero mean Gaussian noise is added to the data set on each of the 100 channels independently with the contrast to noise ratio (CNR) value of 1 (0 dB), typical for the fMRI scan from a robust task paradigm. The sources and the time courses are shown in Figure 4.

Figure 5 shows the order selection results on the simulated data spatially smoothed by the Gaussian kernel with the FWHM of three different sizes. Both order selection based on the original data samples and the effectively i.i.d. samples are implemented. In each panel of Figure 5, “ M ” indicates the actual number of sources used to generate the data and “ K ” is the estimated number of sources by order selection. All the results are based on the average of 20 Monte Carlo simulations with $M = 1, 2, \dots, 8$ randomly selected sources. The standard deviation is stacked on the mean value in each bar plot.

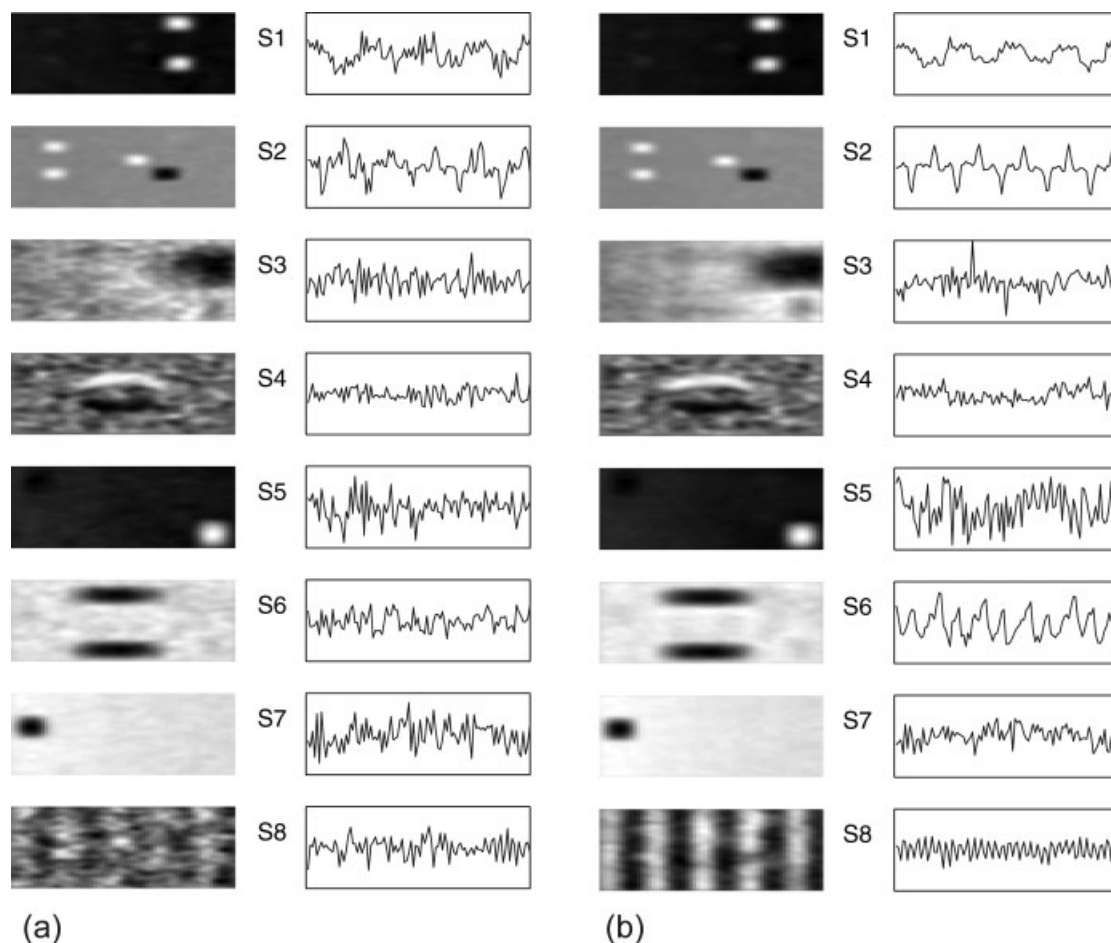


Figure 6.

Estimated sources and time courses by ICA from the data set with $\text{CNR} = 1$ and smoothed by the Gaussian kernel with $\text{FWHM} = 3$ voxels: (a) ICA after dimension reduction to $K = 13$; (b) ICA after dimension reduction to $K = 8$.

From the results, we observe that when the original data samples are i.i.d., the order selection results based on the original data are correct (Fig. 5a). However, for the smoothed data, the order selection criteria based on the original data samples over-estimate the number of sources. Although all the criteria over-estimate the true order due to the sample dependence, over-estimation by MDL is less severe, which can be directly explained by its heavier penalty term compared to that of AIC and KIC [see Eqs. (2)–(4)]. However, MDL still can not select the true order accurately without the correction for sample dependence, as observed in Figure 5e. When order selection is performed on the effectively i.i.d. samples, the effect of sample dependence is removed and the results are accurate and stable as shown in Figure 5b,d,f.

To demonstrate the impact of order selection on ICA estimation, we apply the Infomax algorithm (Bell and Sejnowski, 1995) to the smoothed data ($\text{FWHM} = 3$ voxels) with $\text{CNR} = 1$, $T = 100$, and $M = 8$. Dimension of the data is reduced by

PCA according to the order selected by the MDL criterion based on the original data samples ($K = 13$) and the one based on effectively i.i.d. samples ($K = 8$). Dimension reduction is achieved by performing PCA and keeping a set of the most significant components, i.e., the principal components with the largest variances. Figure 6a,b shows, respectively, the estimated sources and time courses by Infomax after the dimension is reduced to $K = 13$ and $K = 8$. Because of the scaling ambiguity of ICA, both the estimated spatial sources and estimated time courses are normalized for a uniform representation. The correlation between each true and estimated spatial source is calculated and presented in Table I.

By comparing the two cases, degradation on the ICA estimation is observed for sources S3, S4, and S8 when the dimension of the signal subspace is over-estimated. Besides the degradation of the ICA estimation, performing ICA in the over-estimated dimension introduces unnecessary computation load and longer convergence time for iterative ICA algorithms.

TABLE I. Correlation between the true sources and the sources estimated by ICA after dimension reduction to $K = 13$ and $K = 8$

	Source							
	S1	S2	S3	S4	S5	S6	S7	S8
$K = 13^a$	0.97 ± 0.01	0.97 ± 0.01	0.81 ± 0.08	0.71 ± 0.12	0.97 ± 0.01	0.94 ± 0.20	0.98 ± 0.01	0.70 ± 0.01
$K = 8^b$	0.97 ± 0.01	0.97 ± 0.01	0.93 ± 0.01	0.86 ± 0.01	0.97 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.92 ± 0.01

^a $K = 13$ is the order selected by MDL criterion based on the original data.

^b $K = 8$ is the order selection result by MDL based on the effectively i.i.d. samples.

Order Selection on fMRI Data From Visuomotor Task

Participants and Experimental Paradigm

Eleven right-handed participants with normal vision—five females, six males, average age 30 years—participated in the study. Subjects performed a visuomotor task involving two identical but spatially offset, periodic, visual stimulus, shifted by 20 s from one another. The visual stimuli were projected via an LCD projector onto a rear-projection screen subtending $\sim 25^\circ$ of visual field, visible via a mirror attached to the MRI head coil. The stimuli consisted of an 8 Hz reversing checkerboard pattern presented for 15 s in the right visual hemifield, followed by 5 s of an asterisk fixation, followed by 15 s of checkerboard presented to the left visual hemifield, followed by 20 s of asterisk fixation. The 55-s set of events was repeated four times for a total of 220 s. The motor stimuli consisted of participants touching their thumb to each of their four fingers sequentially, back and forth, at a self-paced rate using the hand on the same side on which the visual stimulus is presented.

Imaging parameters

Scans were acquired at the Olin Neuropsychiatry Research Center at the Institute of the Living on a Siemens Allegra 3T dedicated head scanner equipped with a 40 mT/m gradients and a standard quadrature head coil. The functional scans were acquired using gradient-echo echo-planar imaging with the following parameters: repeat time (TR) = 1.50 s, echo time (TE) = 27 ms, field of view = 24 cm, acquisition matrix = 64×64 , flip angle = 60° , slice thickness = 4 mm, gap = 1 mm, 28 slices, ascending acquisition. Six “dummy” scans were performed at the beginning to allow for longitudinal equilibrium, after which the paradigm was automatically triggered to start by the scanner.

Preprocessing

Data were processed using the MATLAB Toolbox for statistical parametric mapping (SPM; <http://www.fil.ion.ucl.ac.uk/spm2>). Images were realigned using INRIalign—a motion correction algorithm unbiased by the local signal changes (Freire et al., 2001; Freire and Mangin, 2001). Data were spatially normalized into the standard Montreal

Neurological Institute space (Friston et al., 1995). The data (originally acquired at $3.75 \times 3.75 \times 4 \text{ mm}^3$) were slightly resampled to $3 \times 3 \times 5 \text{ mm}^3$, resulting in $53 \times 63 \times 28$ voxels. The data is spatially smoothed with an $8 \times 8 \times 8 \text{ mm}^3$ FWHM Gaussian kernel, resulting in the smoothed fMRI data set. To study the effect of sample dependence on order selection, the data obtained after motion correction and spatial normalization but *not* spatial smoothing is also used in the experiments as the “unsmoothed” fMRI data set in contrast to the “smoothed” fMRI data, which are fully preprocessed.

Order selection

Order selection is performed on the fMRI data for each subject using the practical formulations of AIC, KIC, and MDL criteria, based on the original fMRI data samples and the effectively i.i.d. samples obtained by the proposed subsampling scheme.

Figure 7 shows the entropy rate increase during the subsampling on the eleven fMRI data sets. It can be observed that as the subsampling depth increases, the entropy rate approaches the theoretical upper bound. Since the smoothed data have more sample dependence, greater subsampling depth is required to achieve the entropy rate upper bound.

Figure 8a shows the order selection results based on the original fMRI data, while Figure 8b shows the results based on the effectively i.i.d. samples. The standard deviation across different subjects is stacked on the mean value in each bar plot.

Orders selected based on the original fMRI data are high for both the unsmoothed and the smoothed fMRI data, given the typically observed components of significance from ICA estimation. For the smoothed fMRI data, the estimated orders are even higher than the unsmoothed data and close to the original temporal dimension. This is pathological since smoothing—a filtering operation—can not lead to an increase in the number of components, i.e., an increase in the information contained in the data.

For order selection results based on the effectively i.i.d. samples, the estimated orders are close for the unsmoothed and the smoothed fMRI data. In this case, the selected order of the smoothed fMRI data is slightly lower than that of the unsmoothed fMRI data due to a loss of the data variability by smoothing.

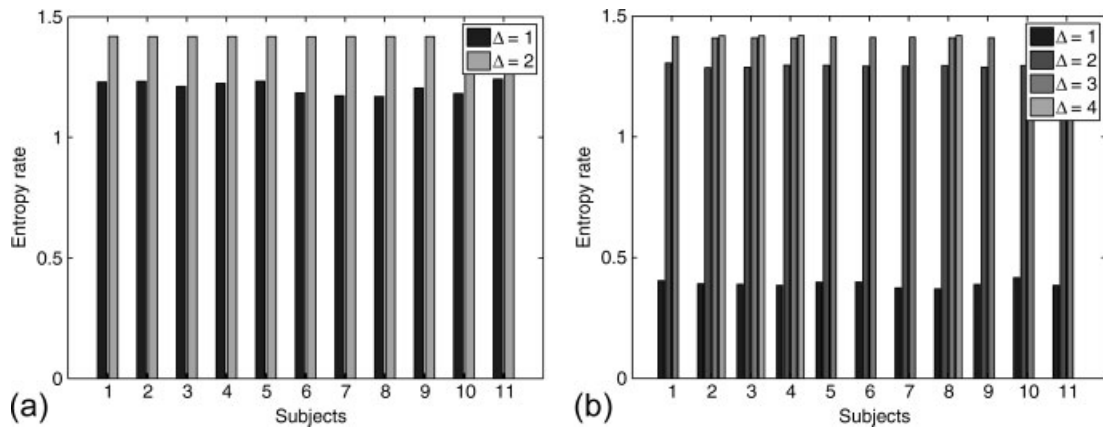


Figure 7.

The increase on entropy rate of the subsampled data during the estimation of effectively i.i.d. samples: (a) Entropy rate increase of the unsmoothed fMRI data sets; (b) Entropy rate increase of the smoothed data sets.

ICA estimation

Unlike the case of the simulated data where the true order is known and thus can be used to justify the order selection results, the order selection on the fMRI data can not be directly verified. However, the impact of order selection manifests itself on the ICA estimation at different selected orders. For example, the stability of the IC estimates from multiple Monte Carlo ICA trials is a relevant index closely linked to the order selection. The stability on the ICA estimation is studied by, e.g., Himberg et al. (2004), Meinecke et al. (2002), and Ylipaavalniemi and Vigario (2004). In this work, the Infomax algorithm is applied to dimension reduced data at different orders for multiple Monte Carlo trials and the ICA estimation results at each order are analyzed with the ICASSO software package (Himberg et al., 2004).

In ICASSO, absolute correlation is used as the similarity measure among the IC estimates and group-average agglomeration strategy is used to identify the cluster of IC estimates attributing to the same underlying independent source. The reliable IC estimates are obtained by retrieving the centrotypic of each cluster, i.e., the one of the estimates that is most similar to other estimates in the cluster. ICASSO also provides quantitative evaluations on the compactness of the clusters of IC estimates, which is used in this work to validate order selection with respect to the stability of ICA estimation. A compactness index close to unity indicates that the estimation is stable and consistent, i.e., similar components are estimated at each run of the ICA algorithm.

Figure 9 shows the compactness indices of the clusters of IC estimates for the eleven data sets. The three panels correspond to the results from the subspace of different dimen-

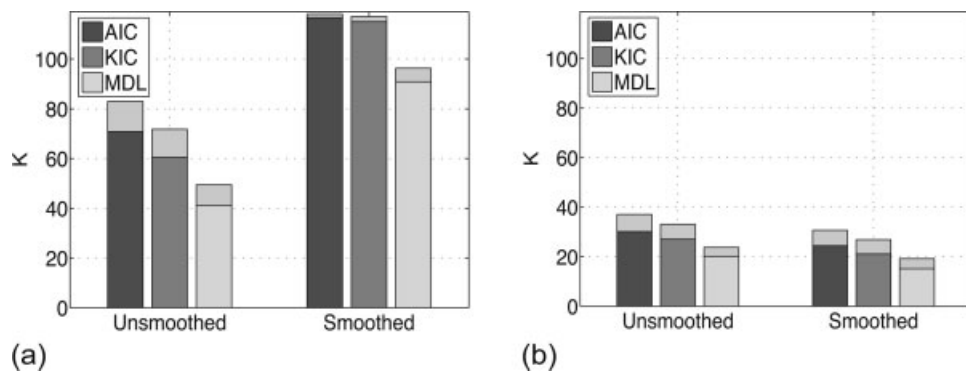


Figure 8.

Order selection results on fMRI data of eleven subjects performing a visuomotor task. For the smoothed fMRI data, $8 \times 8 \times 8 \text{ mm}^3$ FWHM Gaussian smoothing kernel is used. (a) Order selection from original fMRI data; (b) Order selection from effectively i.i.d. fMRI data samples.

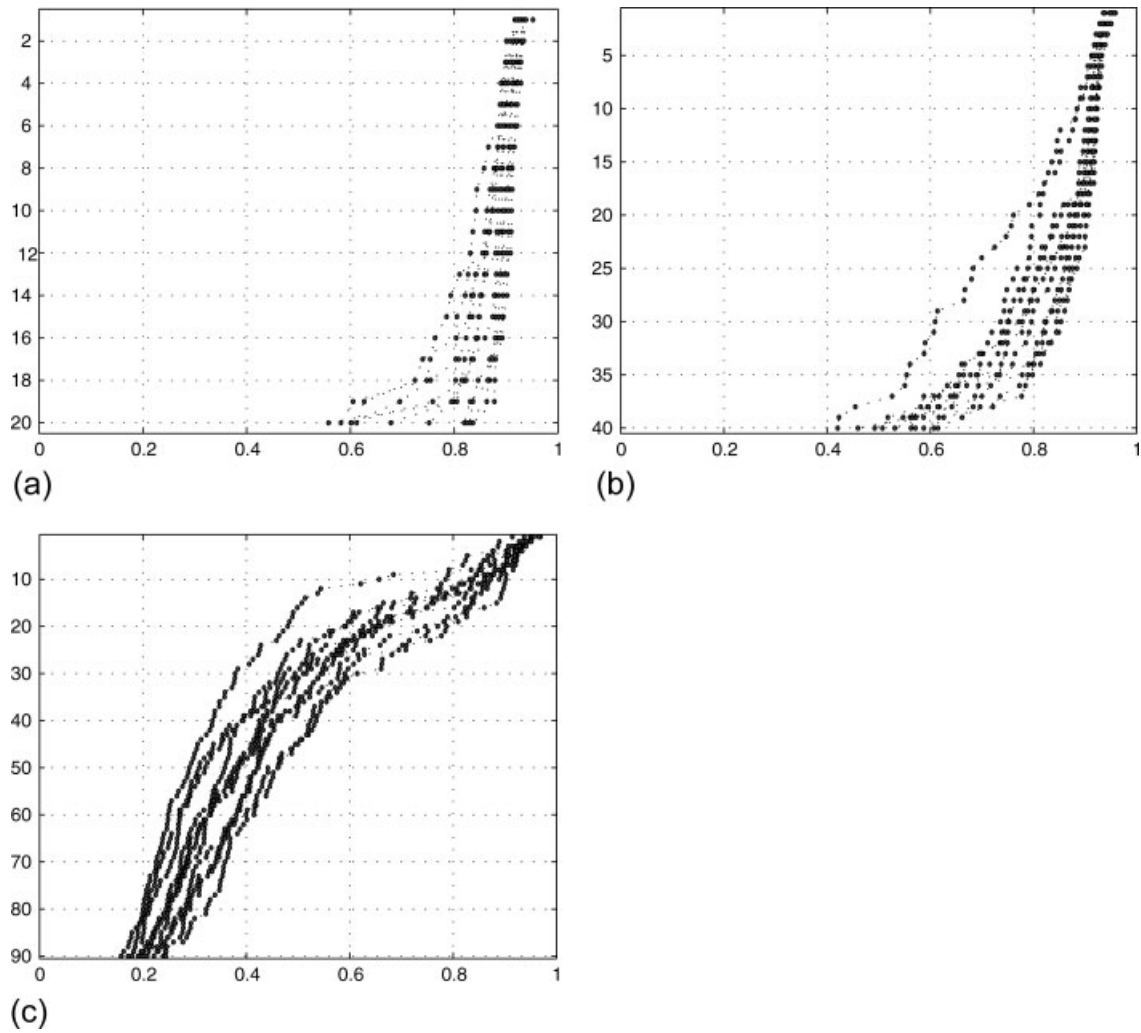


Figure 9.

The stability index of the components estimated in the subspace of different dimensions for eleven subjects: (a) $K = 20$, (b) $K = 40$, and (c) $K = 90$.

sions indicated by K . The compactness index of each cluster is calculated as the average intra-cluster similarity subtracted by the average extra-cluster similarity (Himberg et al., 2004). For the cases of $K = 20$, the resulting stability indices of most of the clusters range from 0.6 to 0.9. For $K = 40$, the stability indices decrease below 0.6 for some of the least stable clusters between cluster 30 and 40. For the case of $K = 90$, the stability indices of most of the data sets decrease below 0.6 after the first 20–30 clusters.

In line with the stability indices, Figure 10 shows the 2D projections of the clustered IC estimates of one subject given by ICASSO. Each IC cluster is prescribed by a convex hull with the black dots representing the individual IC estimates. The background color of the convex hull indicates the average intra-cluster similarity (darker background indicates higher similarity). By comparing the three panels

in Figure 10, it is observed that the clusters of IC estimates are mostly compact and well separated when ICA is performed in the subspace of $K = 20$ dimensions. As the dimension of the subspace increases to $K = 40$, most clusters become less compact and some of the clusters run into each other. When the dimension is increased to $K = 90$, a large portion of the IC estimates distributed in the central part of the graph does not have significant correlation with each other. As a result, the clustering for the ICs in this case is pathological. The pathological clustering of the IC estimates at high orders indicates that ICA estimation becomes less stable in these dimensions. Same pattern is observed on the cluster plots of data sets from all the eleven subjects.

We retrieve each of the reliable IC estimates from ICASSO to obtain the independent brain activations and the associated time courses. The task-related ICs are selected accord-

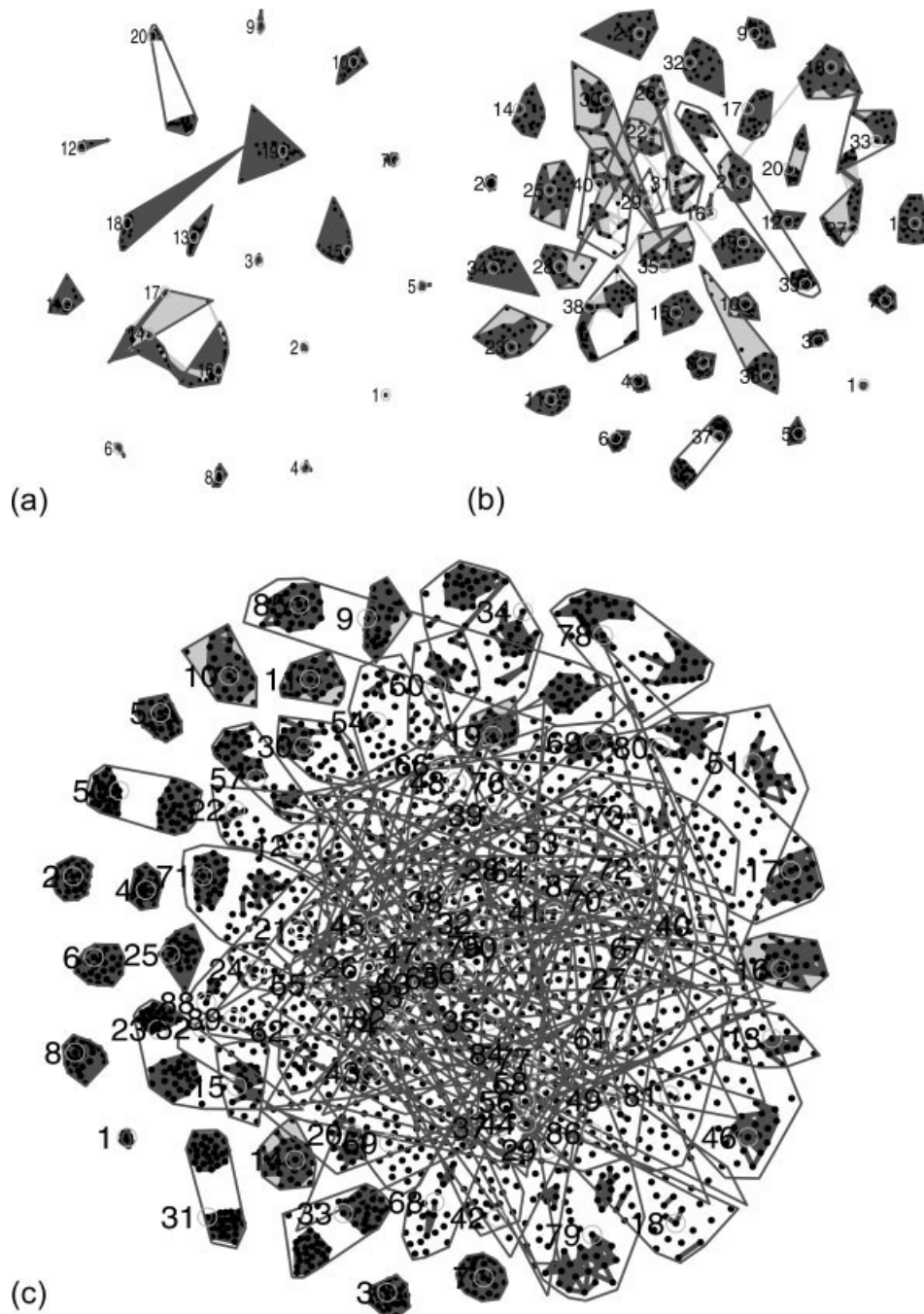


Figure 10.

The 2D clustered representation of the estimated components in the subspace of different dimensions: (a) $K = 20$, (b) $K = 40$, and (c) $K = 90$. Clustering is based on the similarity

between the components, the 2D projection is based on the Euclidean distance as a metric of the dissimilarity between the components.

ing to the correlation between the spatial map and the anatomical template of the visual and motor cortices, and the correlation between the time course and the task paradigm. The anatomical templates for the visual and motor cortices include (Correa, 2005):

- Broadman's areas (BAs) 1–3: somatosensory areas
- Broadman's area (BA) 4: primary motor area
- BA 6: secondary motor area
- BA 17: primary visual area
- BAs 18, 19: secondary visual areas

TABLE II. ICA estimation of the left task-related (LTR) and right task-related (RTR) activations of eleven subjects at three different orders^a

Subject	20		40		90	
	LTR	RTR	LTR	RTR	LTR	RTR
1	0.58	0.42	0.61	0.38	0.61	0.34
2	0.77	0.59	0.61	0.77	0.61	0.77
3	0.48	0.61	0.48	0.58	(s)0.46/0.17 ^b	(s)0.62/0.10 ^b
4	0.69	0.26	(s)0.77/0.22 ^b	0.27	0.81	0.26
5	0.71	0.59	0.74	0.55	0.72	0.50
6	0.72	0.69	0.71	0.71	0.71	(s)0.67/0.41 ^b
7	0.50	0.44	0.44	0.38	0.46	0.40
8	0.38	0.69	0.44	0.76	0.42	0.76
9	0.61	0.62	0.67	0.64	0.62	0.44
10	0.59	0.55	0.53	(s)0.59/0.52 ^b	(s)0.46/0.55 ^b	(s)0.64/0.19 ^b
11	0.28	0.69	0.64	0.22	0.66	0.38

The values 20, 40, and 90 represent the K values.

^aThe coefficient of determination (R^2) between the estimated time course and the corresponding task paradigm is tabulated.

^b“(s)” followed by two R^2 values “ i/ii ” indicates a splitting case, where “ i ” and “ ii ” are the R^2 values between each of the splitting time course and the task paradigm.

The right and left hemispheres containing the above regions are chosen for the right and left task-related templates, respectively.

Table II shows the coefficient of determination (R^2) of the estimated task-related time courses in representing the corresponding task paradigm by a simple linear regression $Y = aX + b$, where Y is the task paradigm and X is the estimated time course. R^2 measures the fraction of variability in the task paradigm that can be explained by the variability in the estimated time courses. In a simple linear regression, R^2 also equals to the square of the correlation coefficient between Y and X .

The splitting components are identified as those components whose spatial maps (resp., time courses) assume significant correlation with the same anatomical template (resp., task paradigm). It is observed that, for $K = 40$ and 90 , the task-related ICs in Subjects 3, 4, 6, and 10 are splitting. For the cases where we observe splitting, the respective R^2 values between each of the splitting time course and the task paradigm are calculated. Furthermore, the correlation between the two splitting time courses in each case is calculated and found to range from 0.45 to 0.92, indicating that the splitting time courses represent similar temporal processes.

As an example, Figure 11 shows the two splitting right task-related (RTR) components of Subject 10 at $K = 90$, while Figure 12 shows the integrated RTR component estimated at $K = 20$. The activation maps are converted to Z-score and thresholded with $Z \geq 1$. In this example, the task-related visual activations are estimated in two ICs at $K = 90$, as marked by the arrows on the activation maps in Figure 11.

DISCUSSION

We propose an i.i.d. sampling scheme to improve the order selection performance of different ITCs. The impact of order selection to the estimation of the brain activations is investigated on data sets from eleven subjects performing

a visuomotor task. The ICASSO software package is used for the quantitative evaluation on the stability of the ICA estimation at different selected orders. As an exploratory data analysis method, ICA is expected to find statistically significant independent sources from the noisy fMRI data. The independent sources come from different types of non-Gaussian distributions, including artifacts contributing to the fMRI data with significant variances. The selection of the useful sources from the ICA estimation result is an ongoing research topic closely related to the objective of the cognitive experiment. In this work, we study the order selection and compare the ICA estimation on the task-related components at different orders for each individual subject. ICA estimation on a group of subjects can be addressed, e.g., by performing ICA on the aggregated fMRI data (Calhoun et al., 2004), or clustering on the IC estimates from each individual subject (Esposito et al., 2005). The experimental results in this work give a preliminary justification that performing ICA in an unnecessarily high dimensional subspace decreases the stability of ICA estimation and hence could degrade the integrity of the ICA representation on brain activations. The dimension of the signal subspace of fMRI data can be decided with information criteria so that the reliability of the ICA estimation is guaranteed at a reasonable level.

Two major issues for research in order selection are the study of the effects of finite sample size and the compensation of the effect of dependencies in the sample space. In our current work, we address the latter issue and propose a scheme to identify the set of effectively i.i.d. samples from the dependent data by subsampling and entropy rate matching. The i.i.d. sample set is used to improve data order selection. Hence, the proposed scheme we show can facilitate the subsequent analysis procedures such as ICA. Incorporation of a dependent data model and derivation of the likelihood, and hence the information-theoretic criterion for this case is

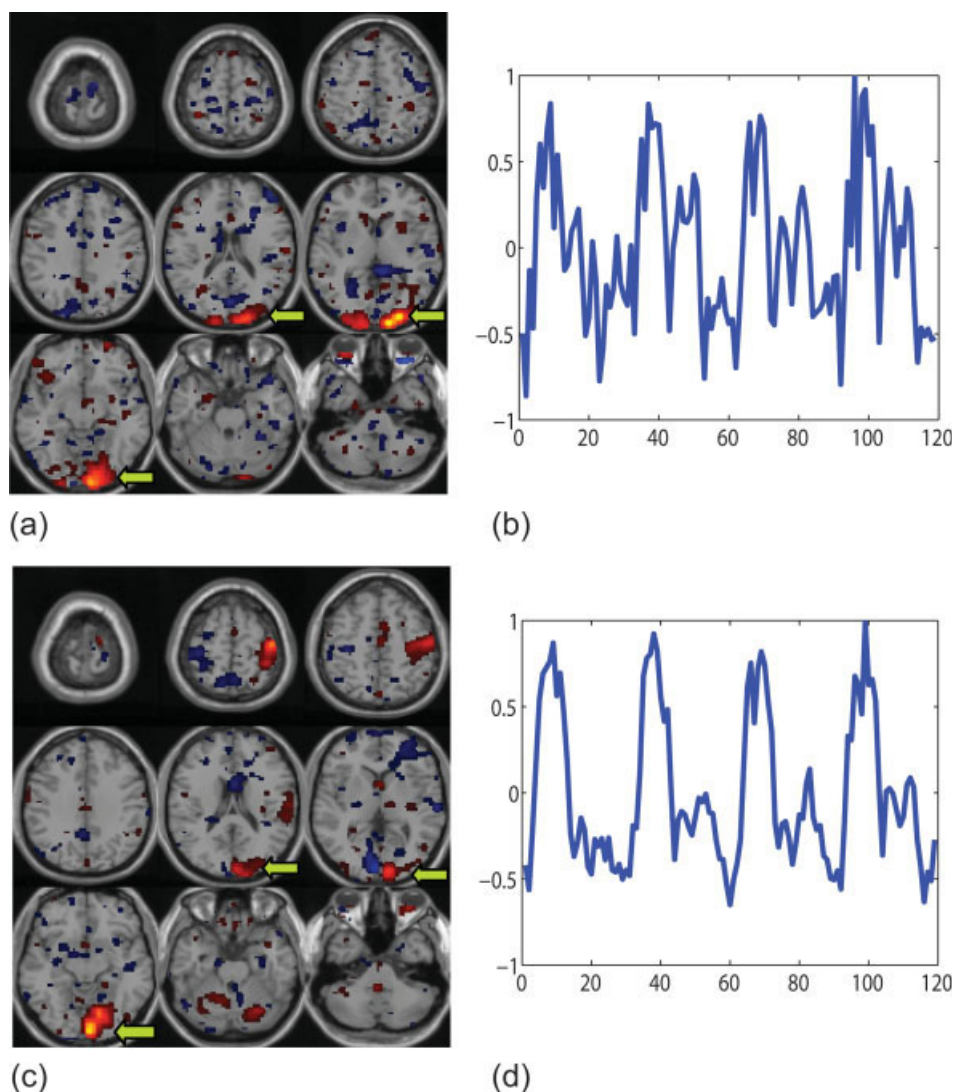


Figure 11.

Split of the RTR activation estimated from the smoothed fMRI data of Subject 10 in the subspace of dimension $K = 90$: (a) The estimated spatial map containing part of the RTR activations; (b) The estimated time course of the activations in (a); (c) The esti-

estimated spatial map containing another part of the RTR activations; (d) The estimated time course of the spatial map in (c). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

another approach to address the problem of dependencies in the sample space. Although this is a rigorous approach, it is not likely to lead to an easily tractable solution. We model the spatial data as a moving average process to justify that subsampling can remove the sample dependence. Another approach to address the sample dependence especially for image processing is to represent the data in the transformed domain by certain basis functions such as linear scale space and wavelets. These models are more involved with machine vision tasks and could possibly be used in the interpretation of the estimated fMRI activation maps.

The proposed method to address the sample dependence effect in order selection is motivated by the characteristics

of the fMRI data. Because of the localization and connection of the functional organization of the brain (Pascual-Marqui et al., 1994; Phillips et al., 1984), the latent sources in the fMRI data are typically smoother than noise in the spatial domain. Therefore, spatial smoothing reduces noise variance more significantly than it does the variance of the informative spatial maps. As a result, the dynamic range of the eigenvalues, i.e., the difference between the source variance and the noise variance, increases after the smoothing operation. Therefore, the likelihood term given in Eq. (5) decreases for the smoothed data at each order because of the greater difference between the geometric mean and the arithmetic mean of the least significant $T-K$ eigenvalues.

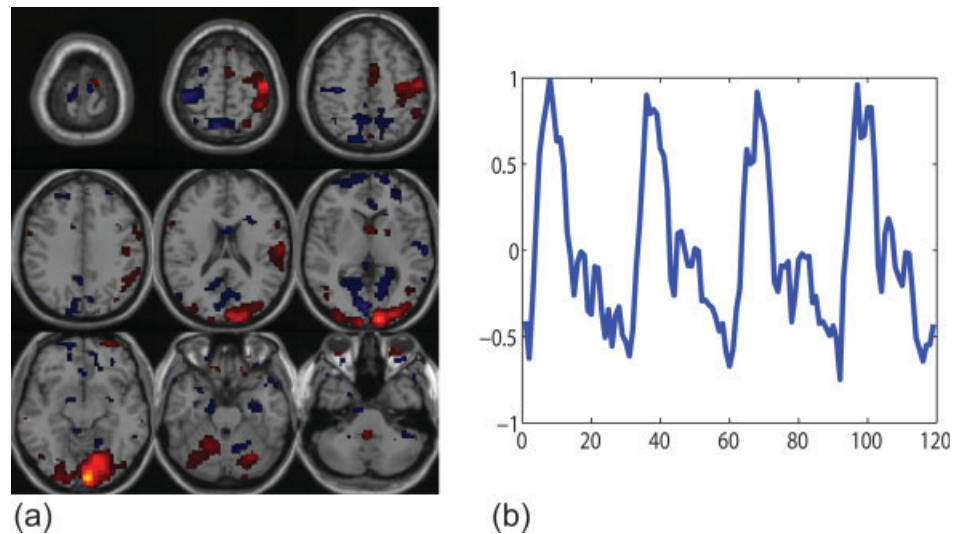


Figure 12.

The integrated RTR activation estimated from the smoothed fMRI data of Subject 10 in the subspace of dimension $K = 20$: (a) The estimated spatial map containing the RTR activations; (b) The estimated time course of the spatial map in (a). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Since the penalty terms are not affected by the smoothing operation, the increase of the negative log-likelihood shifts the minimum of the ITC order selection formulas to higher orders. This is a typical case of over-estimation caused by sample dependence and can be observed in the two groups of bar plots in Figure 8a. When order selection is based on the effectively i.i.d. samples, since the effective sample size N_e of the smoothed data is much less than that of the unsmoothed data, it cancels out the increase of the negative log-likelihood due to the dispersion of the eigenvalues.

It is also important to note that there is temporal dependence in both the fMRI signal and the noise, which have been studied by, e.g., Bullmore et al. (2001) and Friston et al. (2000). The temporal dependence is considered in the estimation of the intrinsic dimensionality of the fMRI data with the noise modeled by a first-order autoregressive process (Cordes and Nandy, 2006), where the dimension is inferred by comparing the eigenspectrum of the fMRI data against the assumed noise spectrum adjusted for the temporal dependence. In this work, we focus on addressing the spatial dependence for order selection in spatial ICA. Specifically, ITC are used for order selection with the adjustment for sample dependence in the spatial domain. In general, the proposed scheme can be applied to address the temporal dependence for order selection in temporal ICA, provided that the short time dependence model is plausible and the temporal length of the fMRI data is long enough for subsampling and entropy rate estimation. The proposed scheme can be easily automated and incorporated into the fMRI analysis procedure between PCA and ICA.

Beckmann and Smith (2004) propose an adjustment on the eigenspectrum of the sample covariance matrix of multidimensional Gaussian noise based on the empirical distribution function of the eigenvalues developed in random matrix theory. The adjustment improves the eigenspectrum estimation for limited amount of data samples and is technically applicable to the ITC for order selection. However,

as we show in the simulations, the eigenspectrum adjustment, which is based on the i.i.d. sampling model, is not capable of correcting the effect of sample dependence in order selection. We incorporate this adjustment in our approach to compensate for the decrease of sample size due to the proposed i.i.d. sampling scheme.

In our scheme, the i.i.d. sample identification is performed on the least significant components of the fMRI volume data to better match the Gaussian process condition for the calculation of the entropy rate. When the test of normality is performed on the least significant principal components, it indicates normality in most cases. However, for the unsmoothed fMRI data, most of the components, though close to passing, fail the test. Upon further inspection, the normalized kurtosis values for these components are found to be close to zero and hence these components, though slightly violating the Gaussian assumption of the entropy rate matching principle, are nonetheless near Gaussian. It is observed that for fMRI data with long temporal records, the least significant components are more close to Gaussian process (results not shown). However, the spatial properties such as the entropy rate assume larger variation across different principal components. Therefore, proper strategy is required for selecting the set of Gaussian components for i.i.d. sample identification.

ACKNOWLEDGMENTS

The authors would also like to thank Srinivas Rachakonda for the test of the order selection program on a variety of fMRI data sets.

REFERENCES

- Akaike H (1973): Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, Budapest, Hungary.
- Beckmann CF, Smith SM (2004): Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* 23:137–152.

- Bell AJ, Sejnowski TJ (1995): An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7: 1004–1034.
- Biswal BB, Ulmer JL (1999): Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *J Compu Assist Tomogr* 23:265–271.
- Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M (2001): Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Hum Brain Mapp* 12:61–78.
- Calhoun VD (2002): Independent component analysis for functional magnetic resonance imaging. Ph.D. Thesis, University of Maryland, Baltimore County, Baltimore, Maryland.
- Calhoun VD, Adali T, Pekar JJ, Pearlson GD (2001): A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 14:140–151.
- Calhoun VD, Adali T, Hansen LK, Larsen J, Pekar JJ (2003): ICA of functional MRI data An overview. In *Proceedings of the ICA2003*, Nara, Japan.
- Calhoun VD, Adali T, Pekar JJ (2004): A method for comparing group fMRI data using independent component analysis: Application to visual, motor, and visuomotor tasks. *Magn Reson Med* 22:1181–1191.
- Cavanaugh JE (1999): A large-sample model selection criterion based on Kullback's symmetric divergence. *Stat Probab Lett* 44: 333–344.
- Conover WJ (1980): *Practical Nonparametric Statistics*. New York: Wiley.
- Cordes D, Nandy RR (2006): Estimation of the intrinsic dimensionality of fMRI data. *Neuroimage* 29:145–154.
- Correa N (2005): Performance of blind source separation algorithms for functional magnetic resonance imaging. M.S. Thesis, University of Maryland Baltimore County, Baltimore, Maryland.
- Correa N, Adali T, Li Y-O, Calhoun VD (2005): Comparison of blind source separation algorithms for fMRI using a new Matlab toolbox: GIFT. In *Proceedings of the ICASSP2005*, Philadelphia, Pennsylvania.
- Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Salle FD (2005): Independent component analysis of fMRI group studies by self-organizing clustering. *Neuroimage* 25:193–205.
- Freire L, Mangin JF (2001): Motion correction algorithms may create spurious activations in the absence of subject motion. *Neuroimage* 14:709–722.
- Freire L, Roche A, Mangin JF (2001): What is the best similarity measure for motion correction in fMRI time series? *IEEE Trans Med Imaging* 21:470–484.
- Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS (1995): Spatial registration and normalization of images. *Hum Brain Mapp* 2:165–189.
- Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996): Detecting activations in PET and fMRI: Levels of inference and power. *Neuroimage* 40:223–235.
- Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J-B (2000): To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage* 12:196–208.
- Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994): Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J Neurosci* 14:2545–2568.
- Himberg J, Hyvarinen A, Esposito F (2004): Validating the independent components of neuroimaging time-series via clustering and visualization. *Neuroimage* 22:1214–1222.
- Judge GG, Hill RC, Lutkepohl H, Griffiths WE, Lee T-C. 1988. *Introduction to the Theory and Practice of Econometrics*. New York: Wiley.
- Karhunen J, Cichocki A, Kasprzak W, Pajunen P (1997): On neural blind separation with noise suppression and redundancy reduction. *Int J Neural Syst* 8:219–237.
- McKeown MJ (2000): Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *Neuroimage* 11:24–35.
- McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ (1998): Analysis of fMRI data by blind separation into independent components. *Hum Brain Mapp* 6:160–188.
- Meinecke F, Ziehe A, Kawanabe M, Muller KR (2002): A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Trans Biomed Eng* 49:1514–1525.
- Menon RS, Goodyear BG (1999): Submillimeter functional localization in human striate cortex using bold contrast at 4 tesla: Implications for the vascular point-spread function. *Magn Reson Med* 41:230–235.
- Papoulis A (1991): *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill. 568p.
- Parkes LM, Schwarzbach JV, Bouts AA, Pullens P, Deckers RH, Kerskens CM, Norris DG (2005): Quantifying the spatial resolution of the gradient echo and spin echo bold response at 3 tesla. *Magn Reson Med* 54:1465–1472.
- Pascual-Marqui RD, Michel CM, Lehmann D (1994): Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int J Psychophysiol* 18:49–65.
- Phillips CG, Zeki S, Barlow HB (1984): Localization of function in the cerebral cortex. Past, present and future. *Brain* 107:327–361.
- Rissanen J (1978): Modeling by the shortest data description. *Automatica* 14:465–471.
- Sarela J, Vigario R (2003): Overlearning in marginal distribution-based ICA: Analysis and solutions. *J Mach Learn Res* 4:1447–1469.
- Schwartz G (1978): Estimating the dimension of a model. *Ann Stat* 6:461–464.
- Wax M, Kailath T (1985): Detection of signals by information theoretic criteria. *IEEE Trans Acoust Speech Signal Process* 33:387–392.
- Wei WWS. 1989. *Time Series Analysis*. Reading, MA: Addison-Wesley.
- Worsley K, Friston KJ (1995): Analysis of fMRI time series revisited-again. *Neuroimage* 6:239–249.
- Ylipaavalniemi J, Vigario R (2004): Analysis of auditory fMRI recordings via ICA: A study on consistency. In *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary.