



Published in final edited form as:

J Signal Process Syst. 2012 May 1; 67(2): 117–128. doi:10.1007/s11265-010-0509-2.

Order Selection of the Linear Mixing Model for Complex-valued FMRI Data

Wei Xiong[†] [Student Member, IEEE], Yi-Ou Li[†] [Student Member, IEEE], Nicolle Correa[†] [Student Member, IEEE], Tülay Adalı[†] [Fellow, IEEE], and Vince D. Calhoun[‡] [Senior Member, IEEE]

[†]University of Maryland Baltimore County, Baltimore, MD 21250, USA

[‡]The Mind Research Network and University of New Mexico, Albuquerque, NM 87131, USA

Abstract

Functional magnetic resonance imaging (fMRI) data are originally acquired as complex-valued images, which motivates the use of complex-valued data analysis methods. Due to the high dimension and high noise level of fMRI data, order selection and dimension reduction are important procedures for multivariate analysis methods such as independent component analysis (ICA). In this work, we develop a complex-valued order selection method to estimate the dimension of signal subspace using information-theoretic criteria. To correct the effect of sample dependence to information-theoretic criteria, we develop a general entropy rate measure for complex Gaussian random process to calibrate the independent and identically distributed (i.i.d.) sampling scheme in the complex domain. We show the effectiveness of the approach for order selection on both simulated and actual fMRI data. A comparison between the results of order selection and ICA on real-valued and complex-valued fMRI data demonstrates that a fully complex analysis extracts more information about brain activation.

Index Terms

Order selection; complex-valued fMRI; linear mixing model; i.i.d. sampling; entropy rate

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a non-invasive and powerful brain imaging technique that has been utilized since the early 1990s [1] to study human brain function. The MRI signal is intrinsically complex-valued due to the fact that the signal contrast is originally acquired in a spatial frequency space, *i.e.*, the k -space, and image reconstruction is achieved through an inverse Fourier transform. Conventionally, only the magnitude of fMRI data is used in analysis and the phase information is discarded. Studies show that phase of fMRI signal contains useful information for inference on blood oxygenation during functional activation [2] and the orientation of large blood vessels [3] as well as different tissue types [4]. In a recent proposed fMRI technology [5], the phase change is utilized to estimate the brain activation.

To utilize phase information from complex-valued fMRI data, the general linear model (GLM) based analysis has been extended to the complex domain [6], [7]. Although widely used, the GLM is only able to infer brain activations from *a priori* response time sequences. On the other hand, data-driven methods such as independent component analysis (ICA)

provide a more flexible alternative to estimate multiple brain activation sources without a specified response model [8]. A group of commonly used complex ICA approaches, such as maximum likelihood, maximization of non-Gaussianity and nonlinear decorrelations [9], [10], have been effectively applied in the analysis of complex-valued fMRI data [11], [12], and shown advantages in estimation of brain activation [13], [14].

Given the abundance of information in complex-valued fMRI data, several challenges exist for applying the complex ICA approach. First, fMRI data has low signal to contrast ratio, typically about 0 dB for a robust task paradigm, which indicates that a noise model has to be incorporated into the analysis. Second, fMRI data has high spatial and temporal dimension, *e.g.*, 10,000–100,000 spatial voxels by 100–1000 time points. Direct application of a multivariate analysis method, such as ICA, on high dimensional datasets is liable to overfitting. Therefore, order selection and dimension reduction are typically incorporated as steps before ICA on fMRI data. Information-theoretic criteria, such as Akaike's information criterion (AIC) [15], Draper's information criterion (DIC) [16] and the minimum description length (MDL) criterion [17] are reasonable candidates of the order selection criteria, for that the optimal model orders in these criteria are automatically selected based on the trade-off between the maximum likelihood of the model and the penalty on model complexity [18], [19]. In addition, other criteria are developed based on, *e.g.*, maximizing a Laplace approximation to the posterior distribution of the model evidence [20], and the l_2 norm of residual error [21].

One important assumption in estimating information-theoretic criteria scores for a selected model is that the data samples are independent and identically distributed (i.i.d.). However, the fMRI data samples have both spatial and temporal correlation. In the spatial domain, the localization of brain function causes the brain activation pattern to be spatially smooth and clustered [22], [23]. The dynamic cerebral perfusion and the point spread function of the imaging system also introduce sample dependence on adjacent voxels in the image. In the temporal domain, the hemodynamic response function introduces dependence on fMRI time sequences. The fMRI sample dependence is localized and can be measured by entropy rate when the spatial or temporal fMRI data are modeled as a random process. Using the theoretical upper bound of entropy rate for Gaussian random process as the benchmark, a sampling scheme is proposed in [19], to remove localized sample dependence in real-valued fMRI data and identify a subset of effectively i.i.d. samples to correctly calculate information-theoretic criteria for selecting model order.

In [14], information-theoretic criteria for order selection on complex-valued model is developed and preliminary results are presented on complex-valued fMRI data. However, the i.i.d. sampling scheme in the complex domain is based on a limiting assumption that the real and imaginary parts of the fMRI samples are uncorrelated. A general form of entropy rate for complex random Gaussian process which takes into account the correlation between real and imaginary parts of complex-valued processes is given in [24]. We use the general entropy rate formula to calibrate the i.i.d. sampling scheme and hence to correctly calculate the information-theoretic criteria scores for order selection. We compare order selection results on real-valued (magnitude-only) and complex-valued fMRI dataset to observe that the additional information provided by the complex-valued data causes information-theoretic criteria to select a model with higher order. We also study real-valued and complex-valued ICA estimation results to identify common components estimated by both methods, as well as additional components obtained by complex ICA only. The common and additional components suggest that a fully complex analysis may be more efficient in utilizing the fMRI data and hence provides a promising way of investigating brain function.

In the next section, we introduce the order selection with information-theoretic criteria and i.i.d. sampling on complex-valued fMRI data. We present experimental results of the order selection scheme using i.i.d. sampling, on simulated and actual fMRI data in Section III, and provide a discussion of the results in Section IV. In the Appendix, we give the entropy rate formula for a complex-valued Gaussian random process.

II. METHODS

A. Linear mixing model of complex-valued fMRI data

We assume independence of spatial brain activations (spatial ICA) in fMRI data, and write the complex ICA model as:

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

Here, $\mathbf{s}_k, \mathbf{a}_k \in \mathbb{C}^N$, represent, respectively, the activation intensity of each voxel for the k th spatial map and the k th corresponding time course, M is the number of informative spatial map sources, N is the number of voxels in each spatial map source, T is the number of time points in the time course, and \mathbf{N} is the $T \times N$ matrix of Gaussian noise.

With different assumptions on \mathbf{s}_k and \mathbf{a}_k , the linear model can also be fitted to other multivariate analysis methods, such as principal component analysis (PCA) and ICA. For this linear model, order number M is often assumed to be less than the temporal dimension T , and it needs to be identified prior to further analysis such as PCA and ICA. If we choose M to be too small, some meaningful components may be omitted in the subsequent analysis. On the other hand, if we choose M to be too large, a meaningful component would split into several ones and also, a number of components would represent the noise in the data, which all lead to insensitivity and instability of ICA analysis [19].

B. Information-theoretic criteria

Information-theoretic criteria are commonly used for order selection in many signal processing problems. Since they do not require the specification of an empirical threshold to decide the optimal model order, they fit naturally into the framework of exploratory data analysis methods such as PCA and ICA. There are a number of information-theoretic criteria commonly used for order selection, such as, AIC [15], Kullback-Leibler information criterion (KIC) [25], DIC [16], and MDL [17] or the Bayesian information criterion [26]. The formulas for AIC, KIC, DIC and MDL criteria assume similar forms:

$$\begin{aligned} \mathcal{E}_{AIC}(k) &= -2\mathcal{L}(\mathbf{x}|\Theta_k) + 2\mathcal{G}(\Theta_k) \\ \mathcal{E}_{KIC}(k) &= -2\mathcal{L}(\mathbf{x}|\Theta_k) + 3\mathcal{G}(\Theta_k) \\ \mathcal{E}_{DIC}(k) &= -\mathcal{L}(\mathbf{x}|\Theta_k) + \frac{1}{2}\mathcal{G}(\Theta_k) \log\left(\frac{N}{2\pi}\right) \\ \mathcal{E}_{MDL}(k) &= -\mathcal{L}(\mathbf{x}|\Theta_k) + \frac{1}{2}\mathcal{G}(\Theta_k) \log N \end{aligned}$$

where $\mathcal{L}(\mathbf{x}|\Theta_k)$ is the maximum log-likelihood of i.i.d. observations \mathbf{x} based on the model parameter set Θ_k , and $\mathcal{G}(\Theta_k)$ is the penalty for model complexity given by the total number of free parameters in Θ_k . For DIC and MDL, the penalty term is scaled by terms related to the sample size N . The order number is determined as the value for which the information-theoretic criteria are minimized,

$$M = \arg \min_k \mathcal{E}(k).$$

In [27], Wax and Kailath provide a practical form of the maximum log-likelihood for complex-valued data, based on multivariate Gaussian data model, as

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

where T is the original dimension of 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 for complex-valued data is given by

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

All these order selection formulations are based on the assumption of i.i.d. samples. When dependent samples are used, the actual number of i.i.d. samples is less than N , and the likelihood term given by (2) improperly dominates the information-theoretic criteria, resulting in an over-estimation of the order. Since samples of fMRI data have spatial correlation, induced at scanning and preprocessing procedure, directly using information-theoretic criteria often leads inflated order numbers. In Section III, we show the over-estimation of order number on simulated and actual fMRI data.

C. IID sampling in the complex domain

Commonly there is sample dependence among the fMRI data samples, which violates the i.i.d. assumption of information-theoretic criteria. In fMRI data, spatially the samples are not i.i.d. due to the point spread function of the scanner as well as the use of spatial smoothing as a preprocessing step. However, the dependence among the data is typically localized, *i.e.*, lies within a few adjacent samples. Hence, an i.i.d. sampling scheme previously applied to real-valued data [19] can be extended to identify an effectively i.i.d. sample set in the complex domain.

First, we model the data as a complex-valued finite-order moving average (MA) process, *i.e.*, an second-order stationary Gaussian random process, which is the output of a linear system with an i.i.d. complex Gaussian input. The second-order statistics of the finite-order MA sequence $z(n)$ has finite length, *i.e.*, $R(m) = E\{z(n+m)z^*(n)\} = 0$ and $\tilde{R}(m) = E\{z(n+m)z(n)\} = 0$ for $|m| > L$, where L is a small positive integer. The sampled sequence $z_s(n) = z(Ln)$ is a doubly white Gaussian random sequence, *i.e.*, after normalization, $R_s(m) = \delta(m)$ and $\tilde{R}_s(m) = c\delta(m)$, where $c \in \mathbb{C}$, $|c| \leq 1$. Here c is a measure for the degree of noncircularity and $c = 0$ in circular case.

The entropy rate can be used to measure the sample dependence, and it reaches the upper bound when the samples are i.i.d. The entropy rate of a complex second-order stationary Gaussian random process is given by

$$h_c = \log(\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log [S(\omega)S(-\omega) - |\tilde{S}(\omega)|^2] d\omega$$

where $S(\omega)$ is power spectrum function and $\tilde{S}(\omega)$ is pseudo power spectrum function, which is introduced in the Appendix. Thus, the theoretical upperbound value for the entropy rate to the normalized complex Gaussian random sequence $z(n)$, *i.e.*, $R(0) = 1$ and $\tilde{R}(0) = c$, where $c \in \mathbb{C}$, $|c| \leq 1$, is given by

$$h_c(Z) \leq \log(\pi e) + \frac{1}{2} \log(1 - |c|^2).$$

Since subsampling removes localized sample dependence, by progressively decreasing the sampling rate, the entropy rate of the spatial process increases. A grid of locations on which the data samples are considered to be effectively independent is determined when the entropy rate reaches its upperbound. Therefore, an effective i.i.d. sample set is obtained on this grid of spatial locations at which the dependence among the samples is small enough to be ignored. Since the sampling procedure decreases the number of samples for estimation, an eigenspectrum adjustment scheme [20] is used to mitigate the finite sample effect.

To show the effect of i.i.d. sampling, we plot the entropy rate of the sampled data with different sampling depths on sixteen sets of unsmoothed and smoothed complex-valued fMRI data sets in Fig. 1. The standard deviation is stacked on the mean value in each bar plot. The difference between unsmoothed and smooth data is that there is no spatial smoothing applied to the unsmoothed data, which is a typical preprocessing step on fMRI data to decrease the effect of noise. Details of the utilized data sets are described in Section III. We observe that as the sampling rate decreases, the entropy rate increases and converges to its upperbound. Thus the sampling depth is determined when all the samples are independent. Since the smoothed data have greater sample dependence compared to the unsmoothed data, lower subsampling rate is required to achieve convergence.

III. EXPERIMENTS

A. Simulated data

We generate eight complex-valued spatial maps to simulate the fMRI sources and corresponding time-courses, the magnitudes of which are similar to the ones used in [28], as in Fig. 2. In an fMRI experiment, the phase difference induced by the task activation is typically less than $\pi/9$ [2], [13]. Therefore, we keep the phase of each pixel uniformly distributed in the range $[-\pi/18, \pi/18]$. The phase of each complex-valued time point is generated proportional to its magnitude, but is again restricted to a small range [7], which in our case is $[-\pi/18, \pi/18]$. The spatial sources are rearranged into one-dimensional vectors and mixed by the corresponding time-courses as in (1). Complex-valued Gaussian noise is added to the data set with a specified contrast to noise ratio (CNR), calculated as the ratio of the standard deviation of the mixed data set without noise to the standard deviation of the Gaussian noise. The mixture data are spatially smoothed, separately for the real and imaginary parts, by a Gaussian kernel with a full-width at half maximum (FWHM) of 2 pixels.

The mixtures of eight sources, with noise levels of CNR = -3, 0, 3 and 6 dB, are created and the complex-valued order selection with i.i.d. sampling scheme is applied to these mixtures. To study the effect of sample dependence on the estimated order, we also apply order selection without sampling to simulated data with CNR = 3 dB. The criteria used in the experiment are AIC, KIC, DIC and MDL. Fig. 3 and Fig. 4 show the results of 10 Monte Carlo simulations where a different noise realization is used for each run. The standard deviation is also shown on the bar plot.

Without i.i.d. sampling, the order number is significantly over-estimated as observed in Fig. 3, since the samples are spatially correlated. As shown in Fig. 4, the criteria based on the effective i.i.d. samples yield accurate estimates (8 sources) when the CNR is higher than 0 dB. Fig. 5 shows the DIC criterion value corresponding to each number of sources on the

data with different CNR values, using effective i.i.d. samples. It is observed that as the CNR value increases, the DIC curve becomes more discriminative, and hence, the minimum value in the DIC curve is more effectively identified, which is corresponding to the correct order number $M=8$. For the other criteria, it is also observed that the robustness of the order selection increases as the data are more informative. The CNR of actual fMRI data is typically in the range $[0, 3]$ dB, and the complex-valued order selection scheme is thus effective in this CNR range, as shown in Fig. 4.

B. fMRI data

1) Data acquisition and preprocessing—All fMRI experiments are performed at the Mind Research Network on a 3T Siemens TRIO TIM system with a 12-channel radio frequency coil. The fMRI experiment uses a standard Siemens gradient-echo echo-planar imaging sequence modified so that it stores real and imaginary data separately. We use a Field-of-View = 240 mm, Slice Thickness = 3.5 mm, Slice Gap = 1 mm, Number of slices = 32, Matrix size = 64×64 , TE = 29 ms, and TR = 2 s. The fMRI experiment uses a block design with periods of 30 s OFF and 30 s ON. Sixteen healthy subjects, all of whom provide informed consent, participate in the experiment. The subjects tap their fingers during the ON period and rest during the OFF period. There are six and a half cycles, starting with OFF and ending with the OFF period. We collect 15 whole head fMRI images during each ‘ON’ or ‘OFF’ period. The total experiment time is 6.5 minutes.

Data are preprocessed using the SPM5 software package¹. Data are motion corrected using INRIalign – a motion correction algorithm unbiased by local signal changes [29]. Real and imaginary parts of the complex data are spatially smoothed with a $10 \times 10 \times 10$ mm FWHM Gaussian kernel respectively, and spatially normalized into the standard Montreal Neurological Institute space. Motion correction and spatial normalization parameters are computed from the magnitude data. The magnitude of the smoothed complex-valued data is taken as real-valued fMRI data. To study the effect of sample dependence on order selection, the data without spatial smoothing is used in the experiments as the “unsmoothed” fMRI data set in contrast to the “smoothed” fMRI data which are fully preprocessed.

2) Order selection—We apply the order selection scheme with and without i.i.d. sampling to smoothed and unsmoothed fMRI data. To compare the order number of real-valued and complex-valued fMRI data, a real-valued order selection scheme in [19] is applied to the magnitude-only fMRI data. Fig. 6 and Fig. 7 show the results based on sixteen subjects. The standard deviation across different subjects is stacked on the mean value in each bar plot.

In Fig. 6, we can see that the order numbers are close to original temporal dimension without i.i.d. sampling scheme for both unsmoothed and smoothed fMRI data, due to the intrinsic spatial correlation in the fMRI data. For order selection results based on effectively i.i.d. samples, the estimated order number for smoothed fMRI data is lower than that of the unsmoothed data, since smoothing leads to a certain degree of signal loss in the high frequency range. Unlike the case of simulated data where the true order is known and thus can be used to justify the order selection results, the order numbers on the fMRI data can not be directly verified. However, the impact of order selection manifests itself, for example, by ICA estimation results at different selected orders. The stability of the components from multiple Monte Carlo ICA trials is a relevant index closely linked to the order number, and is studied in [19], [30]. As observed in [14], the range of values indicated by KIC, DIC and

¹<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>

MDL on i.i.d. samples are appropriate for performing complex ICA of fMRI data and lead to stable estimation results.

In Fig. 7, it is observed that the order number estimated for complex-valued fMRI data is higher than that of real-valued data, which indicates that complex-valued fMRI data contain more information than real-valued data. Fig. 8 shows the DIC criterion value corresponding to each number of sources on one set of complex-value fMRI data and the corresponding real-valued data, using effective i.i.d. samples. The minimum value in the DIC curve of complex-valued data is more effectively identified, and thus order selection on complex-valued fMRI data is more robust than that of the real-valued data, which also can be observed from other criteria curves (not shown here as they are similar). The complex-valued fMRI data are thus expected to be more informative, as also further investigated by ICA results on fMRI data presented next. For subsequent ICA analysis, we use the order estimated by DIC, *i.e.*, 20 for real-valued data and 25 for complex-valued data.

3) ICA analysis on complex-valued fMRI data—In our experiments, a group ICA analysis [18] is applied on fMRI data sets, which improves statistical power by incorporating the inferences across a group of subjects compared with ICA on a single subject. Although the group ICA method in [18] is based on the analysis of real-valued data, it can be extended to complex-valued data straightforwardly, since PCA and ICA are applicable to both real-valued and complex-valued data. For performing ICA, we use nonlinear decorrelations with the nonlinear function $\text{atanh}(\cdot)$ [9], [10], implemented in group ICA [18]. In ICA analysis, we find that the motion artifact has much influence to the results for complex-valued fMRI data, since the values of the boundary area, especially phase values, are deteriorated by even slight brain movement in experiments [31]. Therefore, we use an eroded mask to remove the background and boundary area, while in the real-valued case, only the background area outside the brain boundary is removed.

The time-courses and spatial maps are reconstructed after ICA. For the real-valued fMRI data, the average of resulting spatial maps across the subjects are converted to Z -scores, while for the complex-valued data, there are two sets of average spatial maps, *i.e.*, magnitude and phase. The magnitude spatial maps are thresholded at $|Z| > 1.2$, and the phase spatial maps are thresholded at $|Z| > 1$. It is observed that the typical components for finger tapping task [32] exist in both real-valued and complex-valued results of our experiments, such as, the task-related component in the primary motor cortex, the default mode network component and components in auditory and visual cortex.

Furthermore, order number suggested by the information-theoretic criteria for the complex-valued data is higher than that of the real-valued data. The interesting question is whether this indicates the existence of new components of interest not observed in the real-valued data. As an example, in Fig. 9, we show a component that is consistently estimated as a result of performing ICA on complex fMRI data. The active area in both the magnitude and phase spatial maps indicates parietal cortex, part of parietal lobe, which is related to visual and motor function.

IV. DISCUSSION

In this paper, we study order selection for multivariate analysis on complex-valued fMRI data. By modeling the complex data as a finite-order MA process, we extend the i.i.d. sampling scheme proposed in [19] into complex-valued sample space, using entropy rate derived for the complex domain. It is shown both by simulated and actual data that the scheme improves the performance of order selection with information-theoretic criteria when the samples are correlated. The order selection scheme with i.i.d. sampling on

complex-valued fMRI data is important for the performance of data-driven analysis approaches such as ICA. The developed method can be utilized in other signal processing scenarios where a lower dimensional informative subspace needs to be identified from high dimensional noisy observations.

We study the issue of model order selection in complex-valued fMRI data analysis in this work. Compared with magnitude-only fMRI data analysis, the results from order selection and ICA on complex-valued fMRI data suggest that a more comprehensive approach, adding the phase information, provides additional benefits. The fully complex analysis of fMRI data is a promising direction where several issues are of interest, such as study of order selection criteria other than information-theoretic criteria for complex fMRI data analysis, comparison of complex ICA algorithms for brain source estimation, preprocessing strategies for complex-valued fMRI data, and investigation of the phase signal variation on estimation of brain function. In Section III, we show a brain activation pattern consistently estimated by complex ICA on selected model order but not obtained in the real-valued analysis on the same dataset. The component shows spatially localized activation in part of the parietal lobe, suggesting that, part of it is involved in the performed finger tapping task. Although preliminary, the result motivates the application of a fully complex data-driven method on fMRI data, either as an alternative to conventional analysis, or, as a novel tool to investigate brain function in more complicated cognitive tasks.

Acknowledgments

This work is supported by the NSF grants NSF-CCF 0635129 and NSF-IIS 0612076.

APPENDIX

We present the entropy rate of a complex-valued second-order stationary Gaussian random process using a widely linear model following an approach similar to the one given in [33].

Given a second-order stationary and zero mean random process Z_k , the *covariance function* is defined by $R(m) = E\{Z_{k+m}Z_k^*\}$ and the *pseudo covariance function* [34], also called the *relation function* [35], as $\tilde{R}(m) = E\{Z_{k+m}Z_k\}$. Without loss of generality, the random processes and vectors discussed in this paper are assumed to be zero mean. A random process is called second-order stationary if it is wide sense stationary and its pseudo covariance function only depends on the index difference. The Fourier transform of the covariance function yields the *power spectrum* (or *spectral density*) function $S(\omega)$. Similarly, we define the Fourier transform of the pseudo covariance function as the *pseudo power spectrum function* $\tilde{S}(\omega)$.

Entropy rate is a measure of average information in a random sequence, which can be written for a complex random process Z_k as

$$h_c(Z) = \lim_{n \rightarrow \infty} \frac{H(Z_1, Z_2, \dots, Z_n)}{n} \quad (3)$$

when the limit exists. As in the real case, $H(Z_1, Z_2, \dots, Z_n) \leq \sum_{k=1}^n H(Z_k)$, with equality if and only if the random variables Z_k are independent. Therefore the entropy rate can be used to measure the sample dependence and it reaches the upper bound when all samples of the process are independent.

The widely linear filter is introduced in [36], and any second-order stationary complex signal can be modeled as the output of a widely linear system driven by a circular white

noise, which cannot be achieved by a strictly linear system [35]. Given the input and output vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, a widely linear system is expressed as

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{x}^*$$

where \mathbf{F} and \mathbf{G} are complex-valued impulse responses in matrix form. The system function of a widely linear system is the pair of functions $[F(\omega), G(\omega)]$.

Proposition 1: The entropy rate of output $y(n)$ of a widely linear system $[F(\omega), G(\omega)]$, where $F(\omega)$ and $G(\omega)$ are minimum phase, is given by

$$h_c(Y) = h_c(X) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \{ [|F(e^{j\omega})|^2 - |G(e^{j\omega})|^2] [|F(e^{-j\omega})|^2 - |G(e^{-j\omega})|^2] \} d\omega$$

where $h_c(X)$ is the entropy rate of input $x(n)$.

Theorem 1: If $x(n)$ is a complex second-order stationary Gaussian random process with power spectrum function $S(\omega)$ and pseudo power spectrum function $\tilde{S}(\omega)$, its entropy rate h_c is given by

$$h_c = \log(\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log [S(\omega)S(-\omega) - |\tilde{S}(\omega)|^2] d\omega.$$

The proof of *Proposition 1* and *Theorem 1* are given in [24].

For a second-order circular process, we have $\tilde{S}(\omega) = 0$, thus yielding the entropy rate of a second-order circular Gaussian random process as

$$h_{\text{circ}} = \log(\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log [S(\omega)S(-\omega)] d\omega.$$

For the general case in *Theorem 1*, $|\tilde{S}(\omega)|^2 \geq 0$. Hence, for the second-order circular and noncircular Gaussian random sequences with the same covariance function $R(m)$, we have

$$h_{\text{noncirc}} \leq h_{\text{circ}},$$

which can be also verified using the result for complex entropy [34], [37] and the definition of entropy rate given in (3).

REFERENCES

1. Ogawa S, Tank S, Menon R, Ellermann J, Kim S, Merkle H, Ugurbil K. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci.* 1992; vol. 89:5851–5955.
2. Hoogenrad FG, Reichenbach JR, Haacke EM, Lai S, Kuppusamy K, Sprenger M. In vivo measurement of changes in venous blood-oxygenation with high resolution functional MRI at .95 Tesla by measuring changes in susceptibility and velocity. *Magn. Reson. Med.* 1998; vol. 39:97–107.
3. Menon RS. Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magn. Reson. Med.* 2002; vol. 47:1–9. [PubMed: 11754436]

4. Rauscher A, Sedlacik J, Barth HMM, Reichenbach JR. Magnetic susceptibility-weighted MR phase imaging of the human brain. *Am. J. Neuroradiol.* 2005; vol. 26:736–742. [PubMed: 15814914]
5. Lee J, Shahram M, Schwartzman A, Pauly JM. Complex data analysis in high-resolution SSFP fMRI. *Mag. Res. Med.* 2007; vol. 57:905–917.
6. Nan FY, Nowak RD. Generalized likelihood ratio detection for fMRI using complex data. *IEEE Trans. Med. Imaging.* 1999; vol. 18:320–329. [PubMed: 10385289]
7. Rowe DB. Modeling both the magnitude and phase of complex-valued fMRI data. *Neuroimage.* 2005; vol. 25:1310–1324. [PubMed: 15850748]
8. McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent components. *Hum. Brain Mapp.* 1998; vol. 6:160–188. [PubMed: 9673671]
9. Adalı, T.; Kim, T.; Calhoun, V. Proc. ICASSP. Vol. vol. 5. Canada: Montreal; 2004. Independent component analysis by complex nonlinearities; p. 525-528.
10. Adalı T, Li H, Novey M, Cardoso JF. Complex ICA using nonlinear functions. *IEEE Trans. Signal Process.* 2008; vol. 56:4536–4544.
11. Calhoun, VD.; Adalı, T.; Li, Y-O. Independent component analysis of complex-valued functional MRI data by complex nonlinearities. Proc. ISBI; Arlington, VA. 2004.
12. Adalı T, Calhoun VD. Complex ICA of brain imaging data. *IEEE Signal Processing Mag.* 2007; vol. 24:136–139.
13. Calhoun VD, Adalı T, van Zijl PCM, Pekar JJ. Independent component analysis of fMRI data in the complex domain. *Magn. Reson. Med.* 2002; vol. 48:180–192. [PubMed: 12111945]
14. Xiong, W.; Li, Y-O.; Li, H.; Adalı, T.; Calhoun, VD. On ICA of complex-valued fMRI: advantages and order selection. Proc. ICASSP; Las Vegas, NV. 2008. p. 529-532.
15. Akaike H. A new look at statistical model identification. *IEEE Trans. Autom. Control.* 1974; vol. 19:716–723.
16. Draper D. Assessment and propagation of model uncertainty. *J. Roy. Stat. Soc.* 1995; vol. 57:45–97.
17. Rissanen J. Modeling by the shortest data description. *Automatica.* 1978; vol. 14:465–471.
18. Calhoun VD, Adalı T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 2001; vol. 14:140–151. [PubMed: 11559959]
19. Li Y-O, Adalı T, Calhoun VD. Estimating the number of independent components for fMRI data. *Hum. Brain Mapp.* 2007; vol. 28:1251–1266. [PubMed: 17274023]
20. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imag.* 2004; vol. 23:137–152.
21. McKeown MJ. Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. *Neuroimage.* 2000; vol. 11:24–35. [PubMed: 10686114]
22. Phillips CG, Zeki S, Barlow HB. Localization of function in the cerebral cortex, past, present and future. *Brain.* 1984; vol. 107:327–361. [PubMed: 6421455]
23. Pascual-Marqui RD, Michel CM, Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 1994; vol. 18:49–65. [PubMed: 7876038]
24. Xiong W, Adalı T, Li Y-O, Li H, Calhoun VD. On entropy rate for the complex domain. *IEEE Trans. Signal Process.* 2009 submitted to.
25. Cavanaugh JE. A large-sample model selection criterion based on Kullback's symmetric divergence. *Stat. Probab. Lett.* 1999; vol. 44:333–344.
26. Schwartz G. Estimating the dimension of a model. *Ann. Stat.* 1978; vol. 6:461–464.
27. Wax M, Kailath T. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Process.* 1985; vol. 33:387–392.
28. Correa, N.; Li, Y-O.; Adalı, T.; Calhoun, VD. Comparison of blind source separation algorithms for fMRI using a new Matlab toolbox: GIFT. Proc. ICASSP; Philadelphia, PA. 2005. p. 401-404.
29. Freire L, Mangin JF. What is the best similarity measure for motion correction in fMRI time series. *IEEE Trans. Med. Imag.* 2001; vol. 21:470–484.

30. Hyvarinen A, Himberg J, Esposito F. Validating the independent components of neuroimaging time-series via clustering and visualization. *Neuroimage*. 2004; vol. 22:1214–1222. [PubMed: 15219593]
31. McKeown MJ, Hansen LK, Sejnowski TJ. Independent component analysis of functional MRI: what is signal and what is noise? *Curr. Opin. Neurobiol.* 2003; vol. 13:620–629. [PubMed: 14630228]
32. Moritz CH, Houghton VM, Cordes D, Quigley M, Meyerand ME. Whole-brain functional MR imaging activation from a finger-tapping task examined with independent component analysis. *Am. J. Neuroradiol.* 2000; vol. 21:1629–1635. [PubMed: 11039341]
33. Papoulis A. Maximum entropy and spectral estimation: a review. *IEEE Trans. Acoust. Speech, Signal Process.* 1982; vol. 29:1176–1186.
34. Neeser FD, Massey JL. Proper complex random processes with applications to information theory. *IEEE Trans. Inf. Theory.* 1993; vol. 39:1293–1302.
35. Picinbono B, Bondon P. Second-order statistics of complex signals. *IEEE Trans. Signal Process.* 1997; vol. 45:411–420.
36. Picinbono B, Chevalier P. Widely linear systems for estimation. *IEEE Trans. Signal Process.* 1995; vol. 43:2030–2033.
37. Schreier PJ. Bounds on the degree of impropriety of complex random vectors. *IEEE Signal Process. Lett.* 2008; vol. 15:190–193.

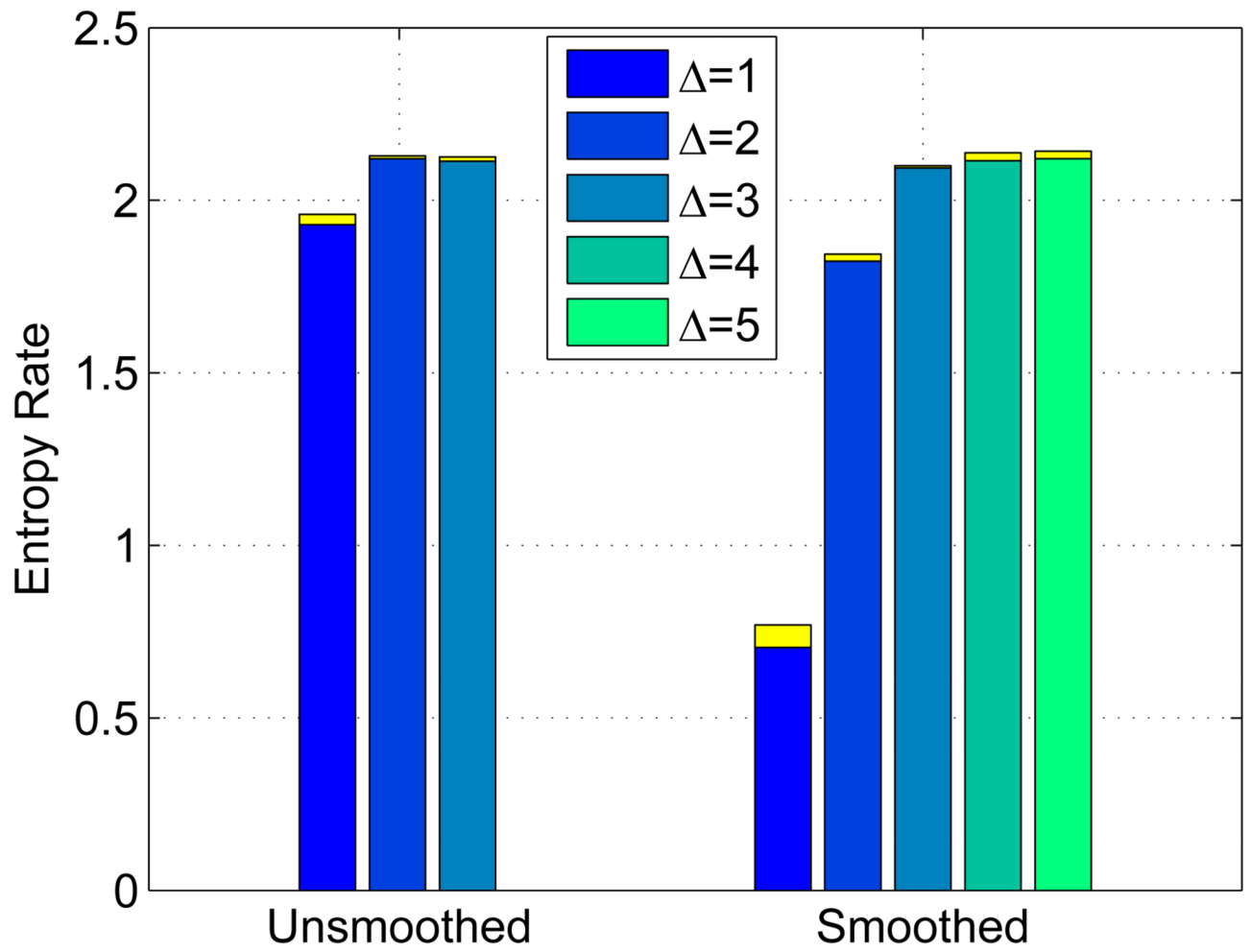


Fig. 1. The entropy rate of the sampled data with different sampling depths Δ on unsmoothed and smoothed fMRI data sets

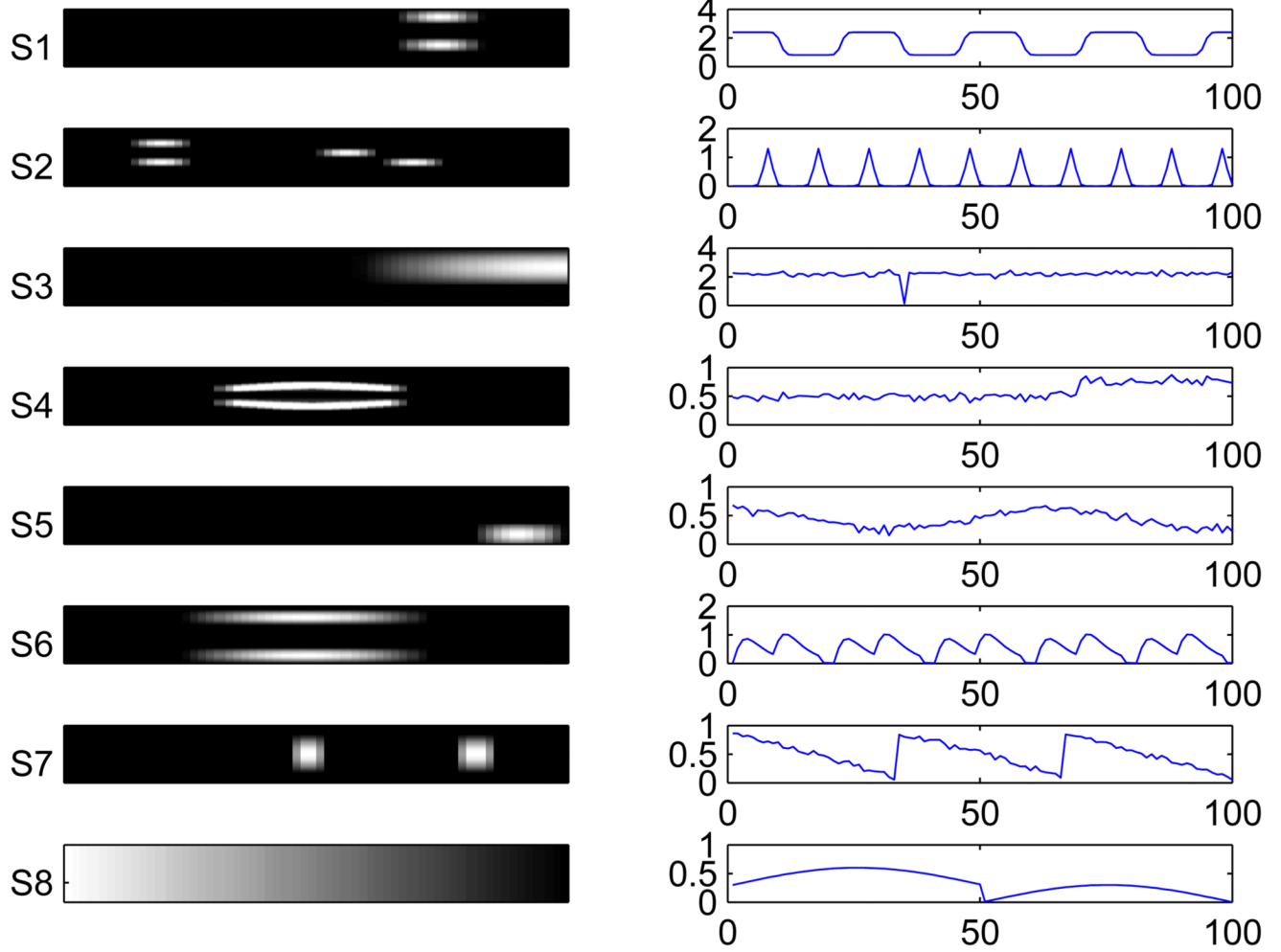


Fig. 2.
Magnitude of eight simulated sources and magnitude of their time-courses

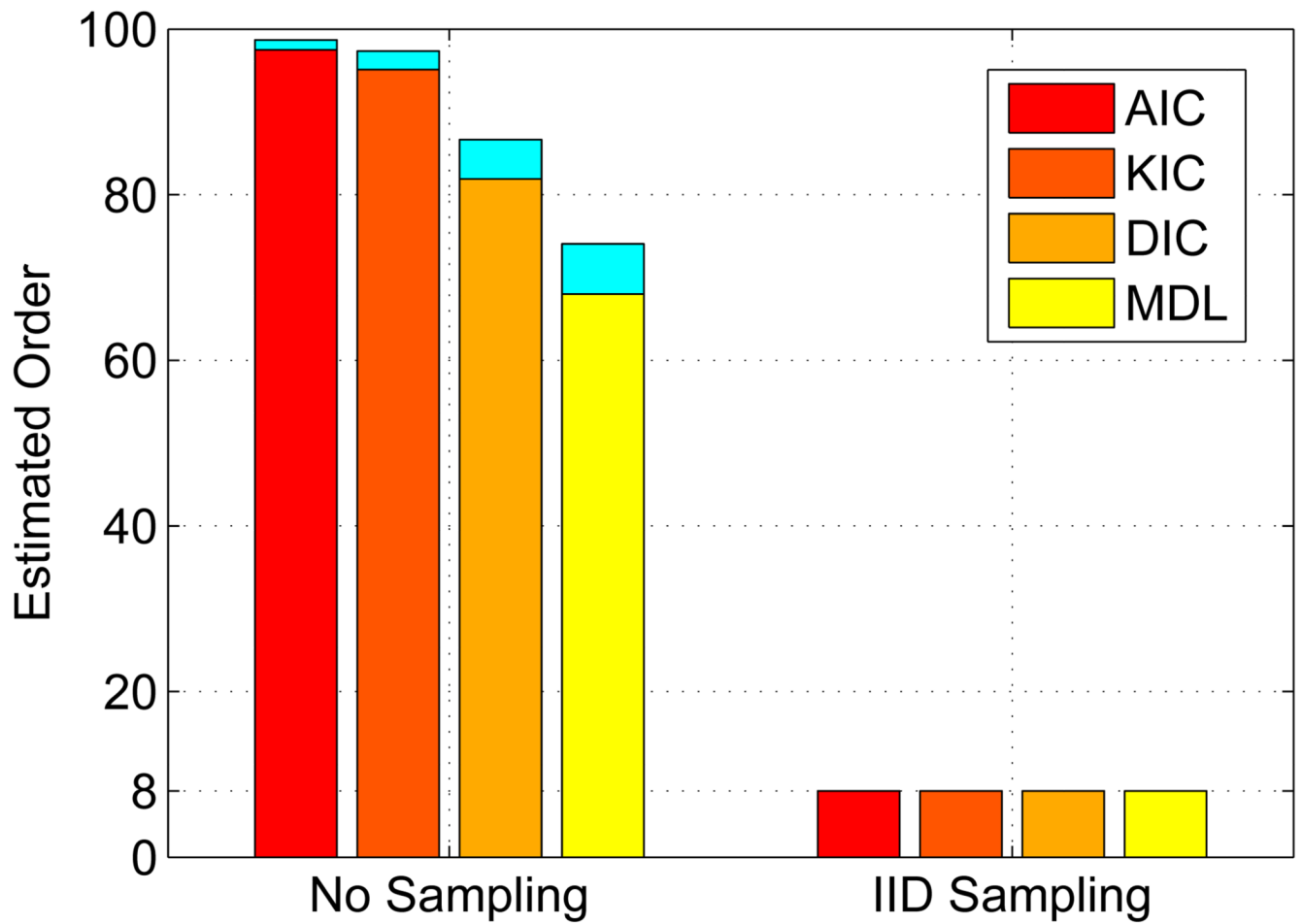


Fig. 3. Order selection with and without i.i.d. sampling on simulated data (CNR = 3 dB), where the true order is eight

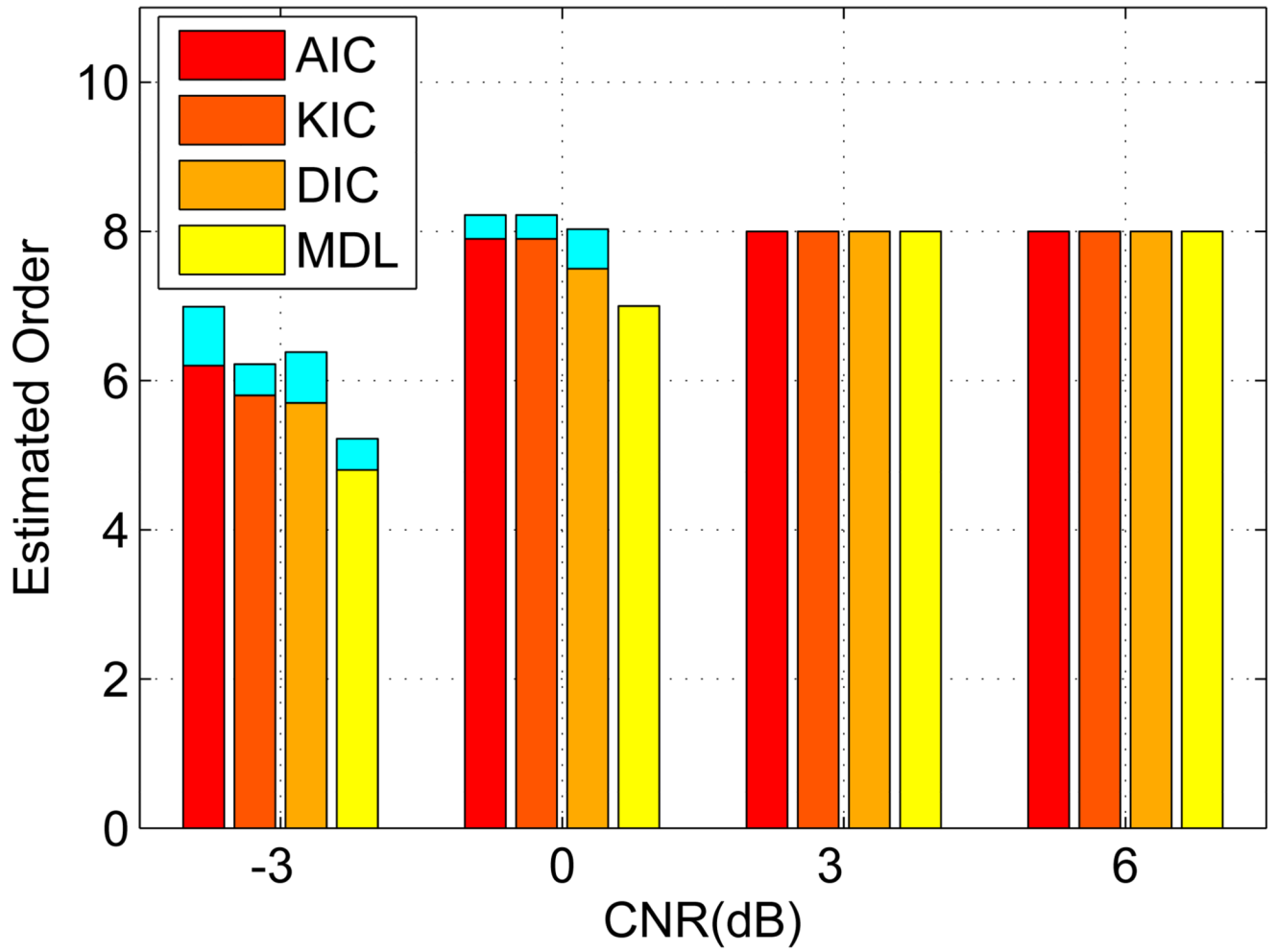


Fig. 4. Order selection on simulated data with different CNR values: -3 , 0 , 3 and 6 dB, using effectively i.i.d. samples, where the true order is eight

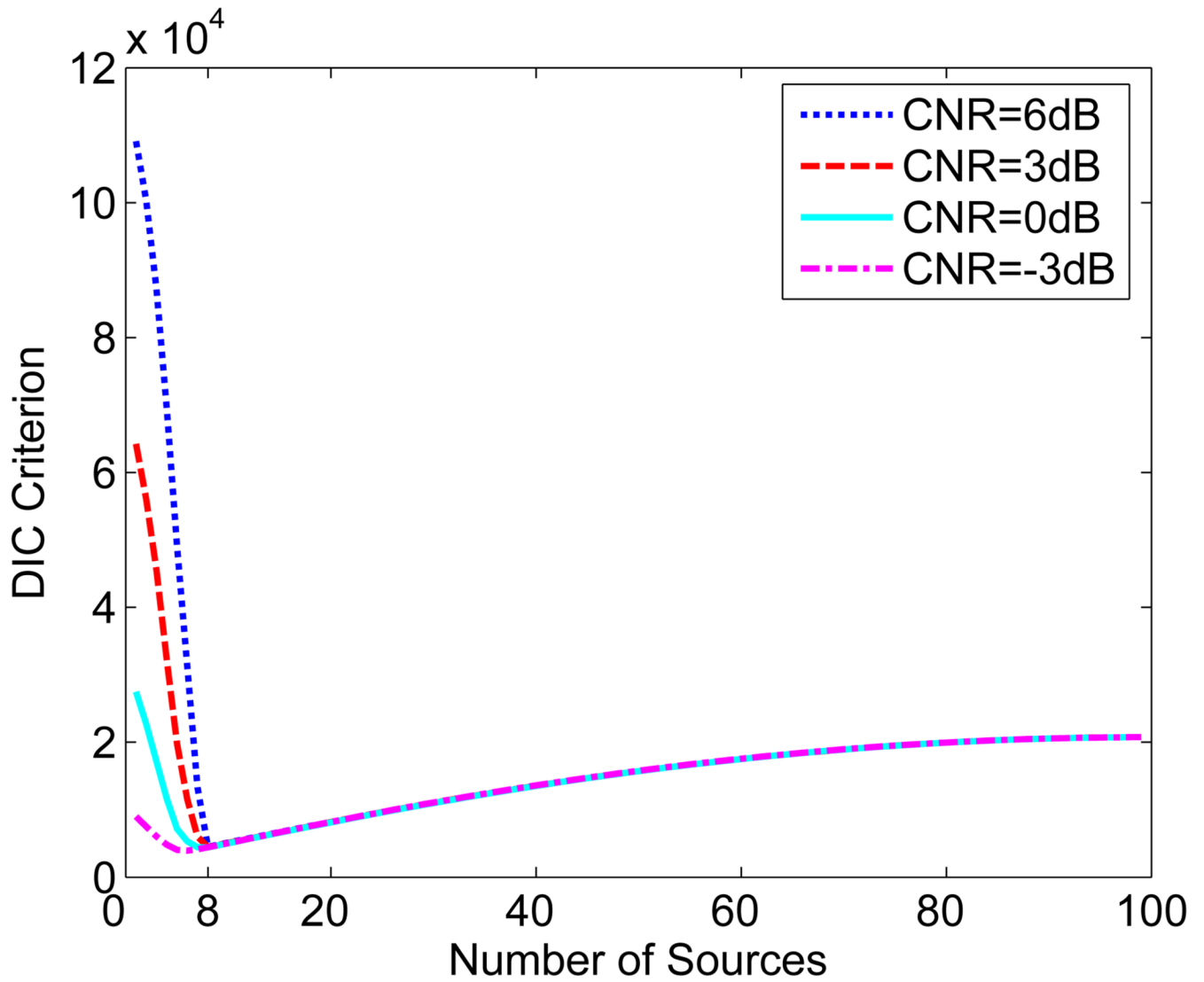


Fig. 5. DIC criterion on simulated data with different CNR values: -3, 0, 3 and 6 dB, using effectively i.i.d. samples

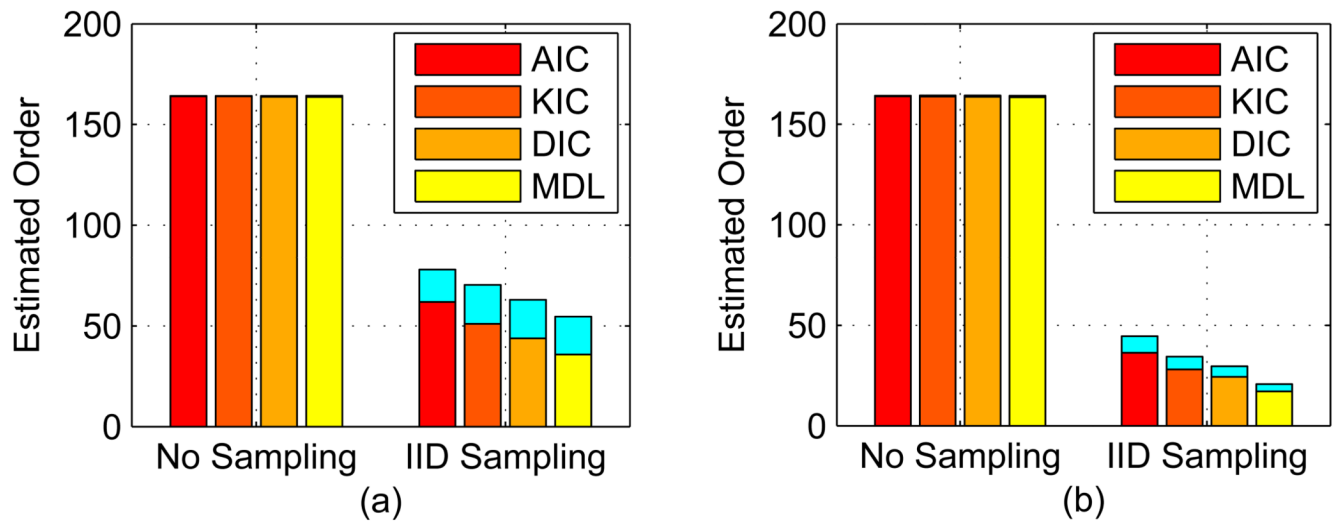


Fig. 6. Order selection with and without i.i.d. sampling, (a) on unsmoothed fMRI data, (b) on smoothed fMRI data

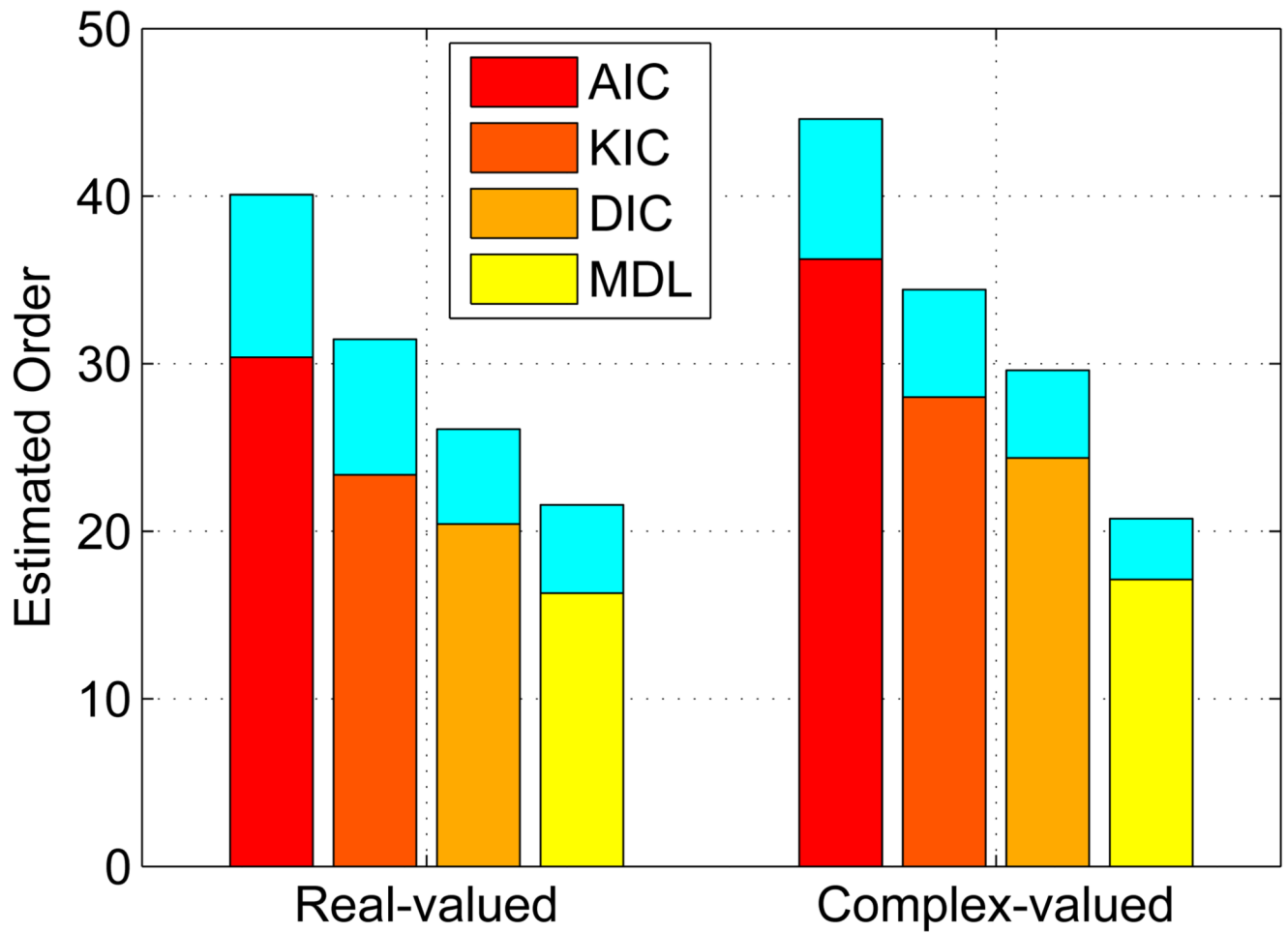


Fig. 7. Order selection on real-valued and complex-valued fMRI data, using effectively i.i.d. samples

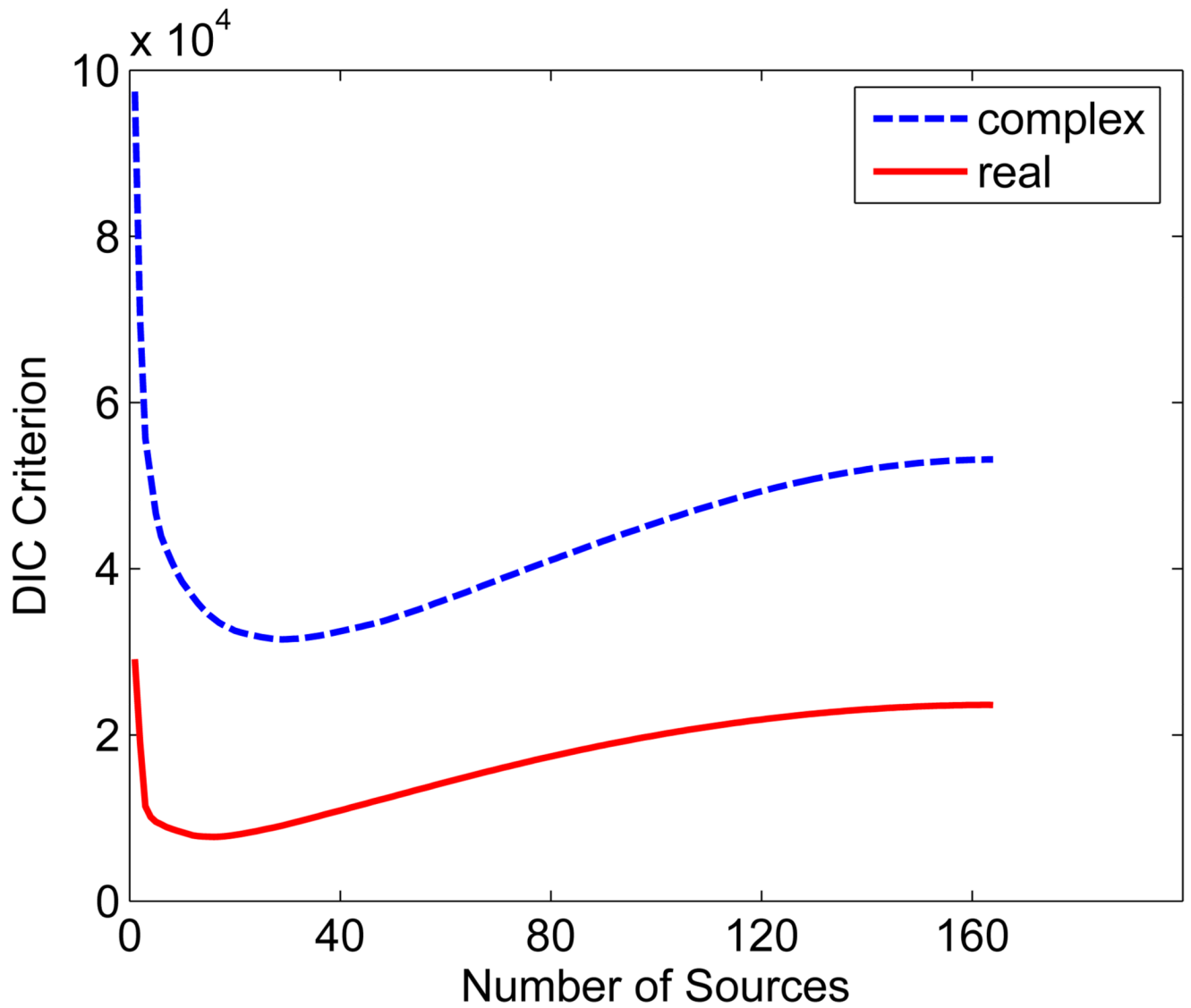


Fig. 8. DIC criterion on real-valued and complex-valued fMRI data, using effectively i.i.d. samples

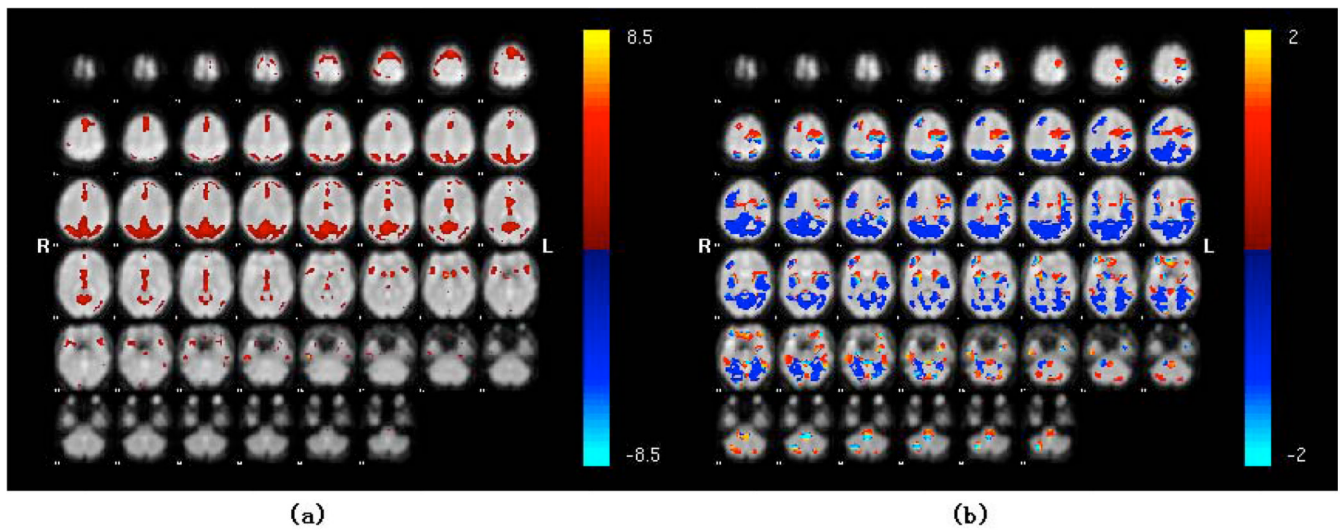


Fig. 9. A component in parietal cortex from group ICA on complex-valued fMRI data, (a) magnitude spatial map, (b) phase spatial map