SAMPLE DEPENDENCE CORRECTION FOR ORDER SELECTION IN FMRI ANALYSIS

Yi-Ou Li[1], Tülay Adalı[1], and Vince D. Calhoun[2,3]

[1]Department of CSEE, University of Maryland Baltimore County, Baltimore, MD 21250
[2] Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT 06106
[3] School of Medicine, Yale University, New Haven, CT 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 the brain function. The selection of the proper number of signals of interest is an important step in the analysis to reduce the risk of over/underfitting. The inherent sample dependence in the spatial or temporal dimension of the fMRI data violates the assumption of independent and identically distributed (i.i.d.) samples and limits the usefulness of the practical formulations of information-theoretic order selection criteria. We propose a novel method using an entropy rate matching principle to mitigate the effects of such sample dependence in order selection. We perform order selection experiments on the simulated fMRI data and show that the incorporation of the proposed method significantly improves the accuracy of the order selection by different criteria. We also use the proposed method to estimate the number of latent sources in fMRI data acquired from multiple subjects performing a visuomotor paradigm. We show that the proposed method improves the order selection by alleviating the over-estimation due to the intrinsic smoothness and the effect of smooth preprocessing on the fMRI data.

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) data have been analyzed successfully by multivariate methods such as independent component analysis (ICA) to explore the brain function, see, *e.g.*, [1],[2],[3]. Although ICA can be applied on the full spatial or temporal dimension of the fMRI data, to avoid overfitting due to high noise level in the fMRI data and for computational efficiency, the number of informative components to be estimated is often assumed to be less than the spatial or temporal dimension of the fMRI data. A lower dimensional subspace containing the sources of interest thus needs to be identified before further analysis and this step has important implications in the final results of the ICA analysis as discussed in [3],[4],[5]. For example, overestimation of the number of components to be estimated leads to splitting of the components making the interpretation difficult and limiting the utility of ICA analysis [4]. On the other hand, underestimation of the dimension may lead to loss of relevant information.

In a typical ICA analysis of fMRI data, the following generative model is assumed

$$\mathbf{X} = \sum_{k=1}^{M} \mathbf{a}_k \mathbf{s}_k^T + \mathbf{E} \tag{1}$$

where $\mathbf{X}$ is the $p \times N$ fMRI data set containing N voxels at p time points, $\mathbf{s}_k$ is an $N \times 1$ vector consisting of the spatial map of the kth brain source, $\mathbf{a}_k$ is a $p \times 1$ vector representing the time course of the kth brain source, $\mathbf{E}$ is a p-variate Gaussian noise with N, M is the number of brain sources contained in the data, *i.e.*, the order of the fMRI data.

Among different approaches for order selection, information-theoretic criteria are found to be particularly attractive for many signal processing applications. One of the major advantages of the approach is the automation of the order selection process so that no empirical threshold value needs to be specified. A commonly used order selection criterion, Akaike's Information Criterion (AIC) is developed based on the minimization of the Kullback-Leibler divergence between the true model and the fitted model [6]. AIC is extended by Cavanaugh as the KIC criterion [7], based on the Kullback-Leibler symmetric divergence of the true and fitted models. The Minimum Description Length (MDL) criterion, also known as the Bayesian Information Criterion (BIC), is developed by Rissanen [8] and Schwartz [9] respectively based on the minimum code length and the Bayesian model. The formulas of AIC, KIC, and MDL criteria have the following common structure:

$$-\mathcal{L}(\mathbf{x}|\mathbf{\Theta}_k) + \mathcal{G}(\mathbf{\Theta}_k)$$

where $\mathcal{L}(\mathbf{x}|\mathbf{\Theta}_k)$ is the maximum log-likelihood of the observed data, $\mathbf{x}$, based on the model parameter set $\mathbf{\Theta}_k$ of the kth candidate model, $\mathcal{G}(\mathbf{\Theta}_k)$ is the model penalty term that is directly proportional to the total number of free parameters in $\mathbf{\Theta}_k$. For MDL, the penalty term is also scaled by $\log N$ where N is the sample size.

Wax and Kailath [10] derive the maximum log-likelihood and the model penalty for the AIC and MDL criteria in a context of detecting the number of signals in noise where both the signals and the noise are modeled by Gaussian random processes. The maximum log-likelihood in their formulation is given by

$$\mathcal{L}(\mathbf{x}|\mathbf{\Theta}_k) = N \log \left(\frac{\prod_{i=k+1}^{p} l_i^{1/(p-k)}}{\frac{1}{p-k} \sum_{i=k+1}^{p} l_i} \right)^{(p-k)} \tag{2}$$

where p is the original dimension of the data, k is the candidate order, N is the number of data samples, and l_i's are the eigenvalues of the sample covariance matrix of the observation vector $\mathbf{x}$. The number of free parameters in $\mathbf{\Theta}_k$ is given by

$$\mathcal{G}(\mathbf{\Theta}_k) = k(2p - k) + 1.$$

Wax and Kailath's formulation thus is directly applicable to the multivariate order selection problem and it is used in estimating the number of latent sources in blind source separation [11] as well as in ICA for fMRI data [3], [4].

One important assumption used in the derivation of Wax and Kailath's order selection formula is that the samples of the observation vector $\mathbf{x}$ are i.i.d. Specifically, for the fMRI data modeled by Eq. (1), the samples from each voxel are assumed to be i.i.d. In fact, however, there is inherent spatial smoothness due to the point spread function of the scanner. Furthermore, smoothing is a commonly used preprocessing step to suppress the noise in the fMRI data. Both of these factors contribute to the dependence among the fMRI data samples, thus, weaken the i.i.d sample assumption. When the samples are dependent, the actual number of i.i.d. samples is less than N. Therefore, the likelihood term shown by Eq. (2) improperly dominates the order selection criteria because the penalty term does not increase as much by the inflation of N. As a result, the Wax and Kalaith's formulation is liable to over-estimate the order.

To address the sample dependence introduced by spatial smoothness, we model the fMRI data at each time point by a finite order moving average sequence in the spatial domain. Based on this model, we propose a subsampling scheme on the fMRI volume data to estimate the effective number of i.i.d samples, N_e. The estimated N_e is therefore used instead of the original sample size N in the likelihood computation given by Eq. (2) for the order selection criteria. Specifically, we use an information-theoretic concept, entropy rate, to measure the sample dependence. By comparing the entropy rate of the subsampled fMRI data with that of an i.i.d. sequence, we infer the effective number of i.i.d. samples in the original fMRI data.

In the next section, we develop the entropy rate matching principle and outline the implementation of the principle to estimate N_e. In section 3, we show experimental results by the proposed method on order selection of the simulated and true fMRI data. We conclude the work with a discussion in the last section.

2. ENTROPY RATE MATCHING

2.1. Stationary Gaussian process and its properties

Let $x[n]$, $n = 1, 2, ..., N$ be a stationary Gaussian random sequence, and $r[m]$ be the autocorrelation function of $x[n]$, the Fourier transform of the autocorrelation function, $s(\omega) = \sum_{m=-\infty}^{\infty} r[m]e^{-j\omega m}$, is the power spectral density function of the random sequence $x[n]$. Without loss of generality, we assume that $x[n]$ has zero mean and unit variance, i.e., $E\{x[n]\} = 0$ and $E\{x^2[n]\} = 1$. This is assumed in sequel unless otherwise mentioned. As a measure of the amount of information carried by each sample of the random sequence, the entropy rate of $x[n]$ is given by [12]

$$h_x = \ln \sqrt{2\pi e} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln s(\omega)d\omega. \qquad (3)$$

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 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 $\ln(x) \leq x - 1$, $(x > 0)$, such that

$$\int_{-\pi}^{\pi} \ln s(\omega)d\omega \leq \int_{-\pi}^{\pi} s(\omega)d\omega - \int_{-\pi}^{\pi} 1d\omega = 0 \qquad (4)$$

The last equality in (4) 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 (4) if and only if $s(\omega) \equiv 1$, i.e., when $x[n]$ is a white Gaussian sequence.

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., $r[m] = 0$ for $|m| \geq \lambda$, the subsampled sequence $x_s[n] = x[\lambda n]$ is a white Gaussian random sequence.

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

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

because $x[n]$ is stationary. Hence, we have $r_s[m] = r[\lambda m] = \delta[m]$ 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.

2.2. Entropy rate matching principle and its implementation

2.2.1. Entropy rate matching principle and its applicability

The properties discussed in Section 2.1 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 entropy rate upper bound, $\ln \sqrt{2\pi e}$, the subsampled sequence is an i.i.d. sequence.

It is worth noting that the proposed entropy rate measure is improper for non-Gaussian processes. It is also observed that the original fMRI volume data are not Gaussian in general (results not shown). Assuming that the spatial sample dependence of the fMRI data is invariant with respect to time, N_e can be inferred from fMRI volume data at any time point or a linear combination of such data set. According to the model of Eq. (1), principal component analysis (PCA) produces such combinations of fMRI volume data that are (approximately) Gaussian – as a set of the least significant components i.e., the principal components with the least variances. Therefore, we use PCA to obtain a set of least significant components of the fMRI data and estimate the effective number of i.i.d. samples from the least significant components by the entropy rate matching principle.

2.2.2. Estimation of N_e by the subsampling scheme

We first subsample the selected least significant principal components by the smallest possible subsampling depth $i = 2$, i.e., drop all but every two samples. As the samples being kept after subsampling have less dependence, the entropy rate of the subsampled sequence increases. We increase the depth of subsampling till the estimated entropy rate of the subsampled sequence reaches the entropy rate upper bound. At this point, we consider the resulting sequence as a white Gaussian sequence and its length, $N_e = N/i$, as the effective number of i.i.d. samples contained in the original sequence. To improve the estimation, we estimate N_e from each selected least significant principal component individually and take the average as the estimated N_e of the data.

2.2.3. Calculation of entropy rate

To estimate the entropy rate of a Gaussian sequence of finite length, summation is used to approximate the integral in Eq. (3), i.e.,

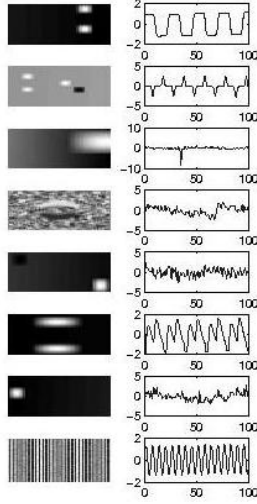$$h_x \approx \ln \sqrt{2\pi e} + \frac{1}{4\pi} \sum_k \ln \hat{s}(\omega_k)\Delta\omega$$

Fig. 1. Simulated fMRI sources and their time courses

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

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

3. EXPERIMENTAL RESULTS

3.1. Order estimation on simulated fMRI data

We use the eight simulated fMRI sources and their associated time courses similar to the ones used in [13] to create the mixed spatial-temporal data according to the generative model in Eq. (1). The sources and the time courses are shown in Figure 1. White Gaussian noise is added to the mixtures at three different contrast to noise ratio (CNR) levels. The typical CNR of fMRI data is approximately 1 (0dB) for a robust paradigm. Order estimation experiments are performed on the unsmoothed data as well as the data spatially smoothed by a 3×3 pixel full-width-half-maximum (FWHM) Gaussian kernel.

Figure 2 shows the order selection results by AIC, KIC, and MDL without incorporating N_e. Figure 3 shows the results with the estimated N_e values incorporated into the order selection formulas of the three criteria. In each figure, 'M' indicates the true number of sources mixed in the data and 'K' is the estimated number of sources. All the results are based on an average of ten Monte Carlo simulations using different source and noise realizations. The standard deviation is stacked on the mean value in each bar plot.

For the unsmoothed data, both the regular and the proposed order selection methods give correct estimation since the data samples are i.i.d. For the smoothed data, all the criteria significantly overestimate the number of sources without incorporating N_e (Figures 2b, 2d, and 2f). When N_e is incorporated, however, the order estimation is correct in most of the cases (Figures 3b, 3d, and 3f).
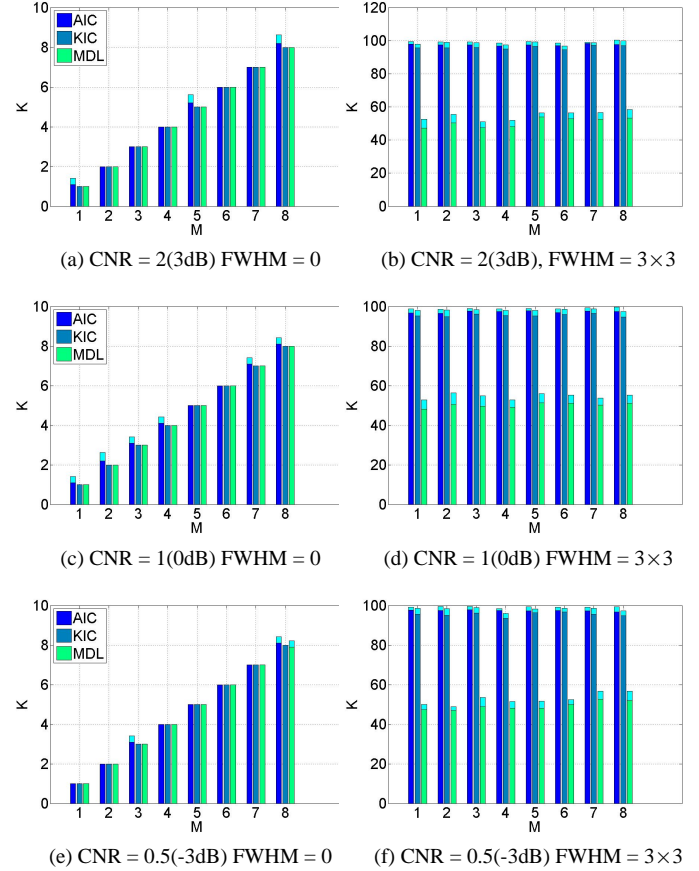

(a) CNR = 2(3dB) FWHM = 0

(b) CNR = 2(3dB), FWHM = 3×3

(c) CNR = 1(0dB) FWHM = 0

(d) CNR = 1(0dB) FWHM = 3×3

(e) CNR = 0.5(-3dB) FWHM = 0

(f) CNR = 0.5(-3dB) FWHM = 3×3

Fig. 2. Order selection without i.i.d. sample estimation

3.2. Order estimation on true fMRI data from visuomotor test

We perform order selection on the unsmoothed and smoothed fMRI data from twelve subjects taking a visuomotor test. See [14] for the detailed experimental settings. Data were processed using the MATLAB Toolbox for Statistical Parametric Mapping (SPM)[1]. Images were realigned using INRIalign - a motion correction algorithm unbiased by the local signal changes [15], [16]. Data were spatially normalized into the standard Montreal Neurological Institute space [17], spatially smoothed with a $8\times8\times8$mm FWHM Gaussian kernel. The data (originally acquired at $3.75\times3.75\times4$ mm) were slightly resampled to $3\times3\times5$ mm, resulting in $53\times63\times28$ voxels.

Figure 4a shows the order selection results without incorporating N_e while Figure 4b shows the results using the estimated N_e values. The standard deviation across different subjects is stacked on the mean value in each bar plot.

When N_e is not used in the formulas, most order selection criteria give high order estimate, *i.e.*, suggest a very large number of sources to be estimated by ICA. For the smoothed data, the estimated orders are higher than that of the unsmoothed data (Figure 4a).

When N_e is incorporated into the order selection criteria, the estimated orders tend to be more reasonable. Also the estimated order of the smoothed data are slightly lower than that of the unsmoothed
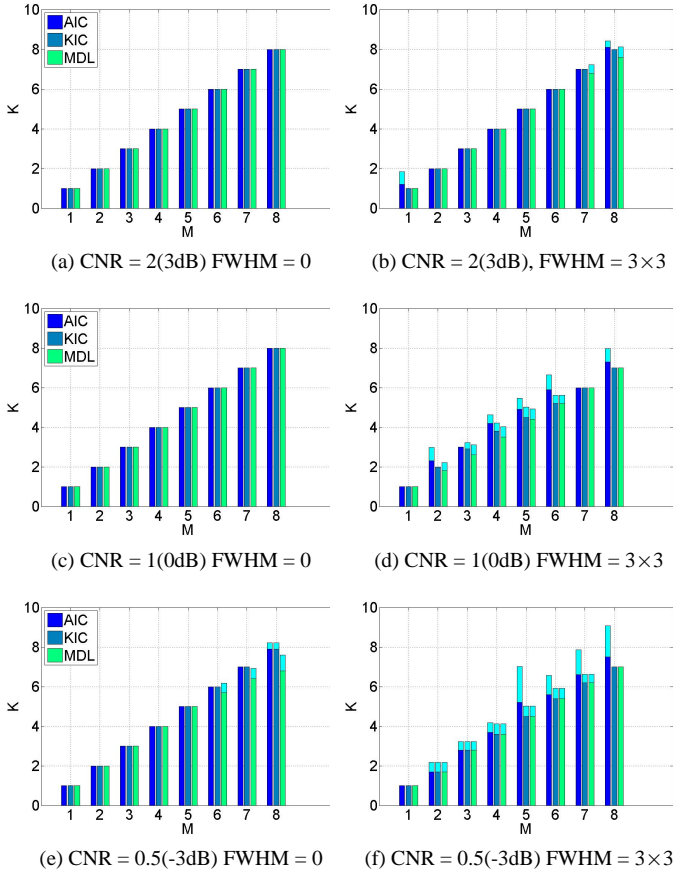
---

[1]http://www.fil.ion.ucl.ac.uk/spm2

(a) CNR = 2(3dB) FWHM = 0

(b) CNR = 2(3dB), FWHM = 3×3

(c) CNR = 1(0dB) FWHM = 0

(d) CNR = 1(0dB) FWHM = 3×3

(e) CNR = 0.5(-3dB) FWHM = 0

(f) CNR = 0.5(-3dB) FWHM = 3×3

Fig. 3. Order selection with i.i.d. sample estimation

(a) Without i.i.d. sample estimation

(b) With i.i.d. sample estimation

Fig. 4. Estimated number of sources in fMRI data from a visuomotor test

data, which is intuitively satisfying because the smoothing operation may reduce the information contained in the data, hence, a smaller number of signals of interest are preserved (Figure 4b).

4. DISCUSSION

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 practical scheme to improve order selection when the samples are not i.i.d. Hence, the proposed scheme we show can facilitate the subsequent analysis procedures such as ICA.

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. The utility of our approach is further confirmed since the proposed scheme achieves reasonable improvement on order selection as shown in the simulation results and the experimental results for the fMRI data are reasonable and consistent with what is expected.
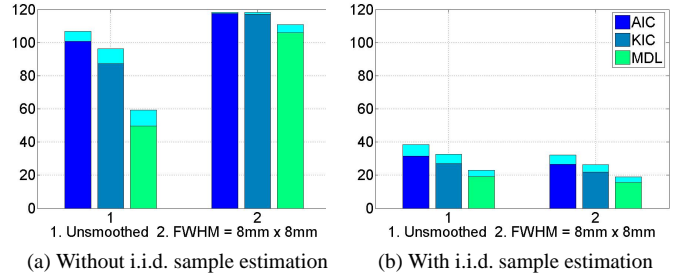
5. REFERENCES

[1] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent components," *Human Brian Mapping*, vol. 6, pp. 160–188, 1998.

[2] B. B. Biswal and J. L. Ulmer, "Blind source separation of multiple signal sources of fMRI data sets using independent component analysis," *J. Comput. Assist. Tomogr*, vol. 23, pp. 265–271, 1999.

[3] V. D. Calhoun, T. Adalı, J. J. Pekar, and G. D. Pearlson, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, pp. 140–151, 2001.

[4] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. on Medical Imaging*, vol. 23, pp. 137–152, 2004.

[5] M. J. McKeown, "Detection of consistently task-related activations in fMRI data with hybrid independent component analysis," *NeuroImage*, vol. 11, pp. 24–35, 2000.

[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inform. Theory, suppl. Problems of Control and Inform. Theory*, Philadelphia, PA.

[7] J. E. Cavanaugh, "A large-sample model selection criterion based on Kullback's symmetric divergence," *Statistics and Probability Letters*, vol. 44, pp. 333–344, 1999.

[8] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[9] G. Schwartz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, 1978.

[10] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 33, pp. 387–392, 1985.

[11] J. Karhunen, A. Cichocki, W. Kasprzak, and P. Pajunen, "On neural blind separation with noise suppression and redundancy reduction," *Int. Journal of Neural Systems*, vol. 8, pp. 219–237, 1997.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes 3rd Ed.*, McGraw-Hill, Inc., 1991.

[13] N. Correa, T. Adalı, Y.-O. Li, and V. D. Calhoun, "Comparison of blind source separation algorithms for fMRI using a new Matlab toolbox: GIFT," in *Proc. ICASSP 2005*, Philadelphia, PA.

[14] V. D. Calhoun, T. Adalı, J. J. Pekar, and G. D. Pearlson, "Latency (in)sensitive ICA: Group independent component analysis of fMRI data in the temporal frequency domain," *NeuroImage*, vol. 20, pp. 1661–1669, 2003.

[15] L. Freire, A Roche, and J. F. Mangin, "What is the best similarity measure for motion correction in fMRI time series?," *IEEE Trans. Med. Imaging*, vol. 21, pp. 470–484, 2001.

[16] L. Freire and J. F. Mangin, "Motion correction algorithms may create spurious activations in the absence of subject motion," *NeuroImage*, vol. 14, pp. 709–722, 2001.

[17] K. Friston, J. Ashburner, C. D. Frith, J. P. Poline, J. D. Heather, and R. S. Frackowiak, "Spatial registration and normalization of images," *Human Brain Mapping*, vol. 2, pp. 165–189, 1995.