

## ASSESSMENT OF MULTIDIMENSIONAL FUNCTIONAL NEUROIMAGING DATA MODEL BY STATISTICAL RESAMPLING

RADU MUTIHAC

*Department of Electricity and Biophysics, University of Bucharest, 077125 Bucharest, Romania  
E-mail: mutihac@gmail.com*

(Received July 7, 2009)

*Abstract.* Artificially generated functional magnetic resonance imaging (fMRI) data drawn from a block-based visual stimulation paradigm were analyzed by the stochastic neuromorphic extended BS Infomax algorithm [1] implementing spatial independent component analysis (ICA) [2]. Variance estimate based on bootstrap resampling [3] was employed as model selection criterion and reliability assessment of ICA decomposition of neuroimaging data.

*Key words:* statistical analysis, independent component analysis (ICA), functional magnetic resonance imaging (fMRI).

### 1. INTRODUCTION

The current techniques in functional neuroimaging data analysis are loosely dichotomized into *hypothesis-driven* methods, like the general linear model (GLM) [4], and *data-driven* model-free methods, such as independent component analysis (ICA). These two approaches are complementary rather than competitive, and mirror the exploratory and confirmatory aspects of scientific investigation. The analytical hypothesis-driven techniques require *a priori knowledge* and/or *specific assumptions* about the time courses of processes generating the measured signals. In contrast, no prior assumptions are necessary for the data-driven methods on the time courses of activation of any components, or whether a given component is due to a specific psychophysiological activity, or is related to machine noise and/or other artifacts.

ICA is a data-driven model-free multivariate exploratory analysis based on the covariance paradigm and formulated as a generative linear latent variables model. ICA estimates distributions of independent and non-Gaussian multivariate random variables, which can be observed as linearly mixtures only. In contrast to confirmatory (inferential) approaches, exploratory data analysis (EDA) like ICA reveals task-related, transiently task-related, and function-related activity without

reference to any experimental protocol, including unanticipated activations [5]. ICA can effectively remove artifacts and separate sources of the brain signals on the basis of minimal statistical suppositions on their underlying distributions. Fourier and wavelet transforms, which contain weak implicit models of temporal noise, and whitening, which is driven by an explicit noise model, were employed to resampling and setting the appropriate thresholds [6].

## 2. METHODS

Linear transforms are often envisaged to find an adequate representation of multivariate data due to their computational and conceptual simplicity. ICA has emerged as a useful linear transform extension of the principal component analysis (PCA) and developed in the context of blind source separation (BSS) [7]. PCA and ICA have major applications in exploratory data analysis (EDA), such as signal characterization, optimal feature extraction, and data compression, as well as providing bases of subspace classifiers in pattern recognition. First, they are aiming at building generative models that are likely to have produced the observed data. In a neuromorphic approach, the model parameters are treated as the network weights trained in an unsupervised learning scheme [8]. And secondly, they perform information preservation and redundancy reduction. Redundancy in the sensory input contains structural information about the environment. Completely nonredundant stimuli are indistinguishable from noise and the percept of structure is driven by the dependencies.

### 2.1. ICA MODEL

ICA is formulated as a generative linear latent variables model. ICA showed up as a generalization of PCA that separates the high-order dependencies in the input, in addition to second-order dependencies used up by PCA [9]. ICA does not constraint the axes to be orthogonal as PCA (Fig. 1 left), rather it attempts to place them in the directions of maximal statistical dependencies in the data (Fig. 1 right). Each estimate in ICA attempts to encode a portion of the dependencies in the input, so that the dependencies are removed from between the output components. As such, ICA, contrarily to other statistical methods searching for underlying factors in multivariate data, looks for components that are simultaneously *independent* and *non-Gaussian*. The Gaussian is the most unstructured and non-committal distribution given its mean and variance. It follows that signal mixtures are rather uninteresting and the only compelling items ought to be non-Gaussian (assuming that Gaussian distributions arise from mixing only!). This perspective motivates why ICA is so pertinent to biological time series and attractive to data analysis in

imaging neuroscience. Therefore, the ICA problem is solved on the basis of optimizing certain measures of “interestingness” [10] like the departure from Gaussianity, which leads to a numerical optimization problem.

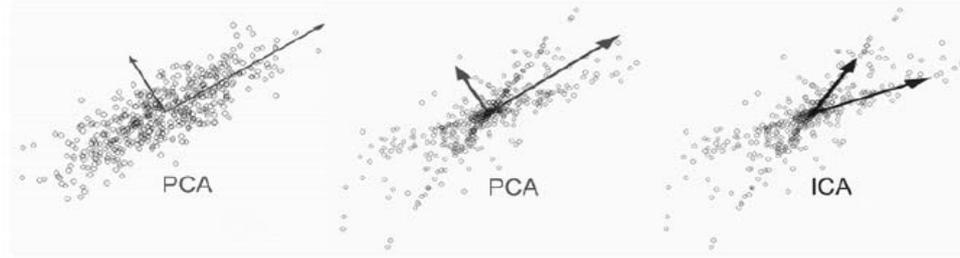


Fig. 1 – Decomposition of bivariate data: PCA of Gaussian distributed data (left), PCA of non-Gaussian distributed data (mid), and ICA of non-Gaussian distributed data (right).

The ICA linear stationary model assumes that  $\mathbf{x}(t)$ ,  $\mathbf{n}(t) \in \mathbb{R}^N$  and  $\mathbf{s}(t) \in \mathbb{R}^M$  are three random (column) vectors for any sample index  $t = 1, 2, \dots, T$  with zero mean and finite covariance, with the components of  $\mathbf{s}(t)$  being statistically independent and one Gaussian at most, and  $\mathbf{A}$  a rectangular constant full column rank  $N \times M$  matrix with at least as many rows as columns ( $N \geq M$ ):

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{i=1}^M s_i(t) \mathbf{a}_i + \mathbf{n}(t). \quad (1)$$

Mixing is supposed to be instantaneous, so there is no time delay between the (latent) source variable  $s_i(t)$  mixing into an observable (data) variable  $x_j(t)$ . The noise  $\mathbf{n}(t)$  is assumed not to be correlated with the data. Then the ICA problem can be formulated as follows: given  $T$  realizations of  $\mathbf{x}(t)$ , estimate both the mixing matrix  $\mathbf{A}$  and the corresponding realizations of  $\mathbf{s}(t)$  (Fig. 2).

There are nevertheless several limitations to solving the basic ICA model. If no suppositions are made about the noise, it cannot be introduced in the model. Next, the size of  $\mathbf{s}(t)$ , usually unknown, should not be greater than the size of acquired data  $\mathbf{x}(t)$ , otherwise the problem becomes under-determined. If the size of  $\mathbf{x}(t)$  is greater than the size of  $\mathbf{s}(t)$ , the problem is over-determined and the extra data can be used for reducing the noise. [11]. In any case, each source signal  $s_i(t)$   $i = 1, 2, \dots, M$  is assumed here to be a stationary zero-mean stochastic process and only one of them is allowed to have a Gaussian distribution. As such,

the source separation consists in updating a  $N \times M$  unmixing matrix  $\mathbf{B}(t)$ , without resorting to any information about the spatial mixing matrix  $\mathbf{A}$ , so that the vector  $\mathbf{y}(t) = \mathbf{B}(t) \mathbf{x}(t)$  becomes an estimate  $\mathbf{y}(t) = \hat{\mathbf{s}}(t)$  of the original independent source signals  $\mathbf{s}(t)$ . It is expected that after a finite number of iterations, the separating matrix  $\mathbf{B}(t)$  to converge to a fixed value  $\mathbf{B}(t) \rightarrow \mathbf{B}$ . Then the estimates of the ICA basis functions (vectors)  $\{\mathbf{a}_i\}$  are the columns of the pseudo-inverse  $\hat{\mathbf{A}} = \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1}$ , which amounts to  $\hat{\mathbf{A}} = \mathbf{B}^{-1}$  if  $\mathbf{B}$  is a square matrix ( $M = N$ ) and the estimated ICs are given by:

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) = \mathbf{B}\mathbf{x}(t). \quad (2)$$

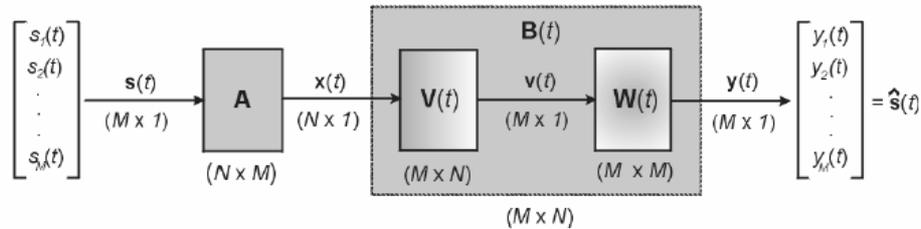


Fig. 2 – Overall processing of the (unknown) signal sources  $\mathbf{s}(t)$  in the basic noiseless ICA model: linear mixing by the (unknown) constant matrix  $\mathbf{A}$ , and separating by the (estimated) matrix  $\mathbf{B}(t)$ .

In practice, applying the ICA model amounts to determining an *estimation principle* for measuring the departure from Gaussianity of the estimated ICs and to selecting an adequate *algorithm* to run the computations required by the estimation principle. Nevertheless, ICA decomposition of data is unique up to sign, scaling, and permutation. Nonquadratic functions are generally involved by the estimation methods that require computationally demanding numerical algorithms. The current algorithms for ICA can loosely be classified in two categories [12]. One category contains *adaptive algorithms* generally based on stochastic gradient methods and implemented in neural networks (NNs) [13]. Adaptive algorithms may also be based either on optimization of cumulant-based contrast functions [14], or on “estimating equations” involving nonlinear distortions of the output [15]. The neural adaptive algorithms exhibit slow convergence and their convergence heavily depends on the correct choice of the learning rate parameters [16]. The second category relies on *batch computation* optimizing some relevant criterion functions. Generally, they imply complex matrix or tensorial operations. Neuromorphic block technique algorithms based on *2nd-* and *4th-*order cumulants [17], as well as (quasi)-likelihood approaches [18] were also proposed.

## 2.2. RESAMPLING METHODS

Resampling is the process of artificially generating null-data sets with a statistical dependence structure similar to an original data set [6]. Resampling methods produce surrogate data sets that may be used to predict the probability of an event occurring. By generating a large number of surrogate fMRI time series and evaluating the test statistic ( $t$  or  $F$ ) for each, the probability of observing a test statistic that exceeds a certain threshold can be estimated. Specifically, the task in resampling fMRI time series is to estimate the probability that the employed test statistic exceeds a certain threshold given the absence of activity in the voxel or in the region of interest. In other words, the threshold that produces a desired  $P$ -value is to be found.

The main difficulty in resampling fMRI time series stems from the spatiotemporal character of fMRI data. Since fMRI time series are serially correlated [19], a plain randomization of the samples in the time series is not preserving the temporal dependence structure leading to biased threshold estimates. As such, fMRI time series need transforming to a domain where data samples become approximately uncorrelated and permutation preserves the statistical structure, then back projected by the inverse transform in the original domain [6]. Likewise, the inherent spatial correlation structure is preserved by applying the same permutation to every time series in the fMRI data set [20]. An alternative approach consists in multidimensional spatio-temporal resampling [21].

Three different transforms have so far been proposed for resampling: whitening, Fourier, and wavelet transforms. A *whitening* (sphering) transform renders the time series uncorrelated. Prior to applying a whitening transform, a model for the serial correlation structure has to be adopted, such as the AR(1) noise model [22] or ARMA model [23]. The *Fourier resampling* method is based on Wiener-Khinchin theorem, which relates the autocorrelation function of a signal to the magnitude of its Fourier transform. This procedure reduces to the Fourier transform (FT), keeping the magnitude of each frequency component but permuting the phase components, and then applying the inverse FT [20]. Fourier resampling is nevertheless not appropriate for fMRI analysis due to (i) several degrees of freedom it has to mimic the autocorrelation function in the original time series, and (ii) because the test statistic depends on the magnitude of the FT of the time series. Wavelet methods approach the analysis of statistical fields by estimating the time series at various resolutions (scales) among the random fluctuations. The wavelet-based statistical analysis provides (i) multiresolution decomposition suitable for scale-invariant process analysis, (ii) sparse representation of typical brain activation maps, (iii) optimally whitening of data producing approximately decorrelated wavelet coefficients, (iv) Karhunen-Loève expansions for  $1/f$ -like processes, which is the case in fMRI, (v) good estimators for the noise process parameters, and (vi) adaptiveness to nonstationary and local

features [24]. Wavelet resampling amounts to wavelet transforming the time series, permuting the wavelet coefficients within each level of detail, and then applying the inverse wavelet transform [22]. The implicit noise model is that the noise spectrum is flat within each frequency subband. It comes out that in terms of the assumptions on the noise structure, wavelet resampling is positioned between the model-driven whitening resampling and non-parametric Fourier resampling [6].

All unsupervised procedures applied to arbitrary data sets come up with solutions within their model class irrespective of the applicability of the used model to data. For instance, the ICA model yields an estimate of the mixing matrix  $\mathbf{A}$ , even if the observed data  $\mathbf{x}$  contain no structure at all. In unsupervised training, a data set  $\mathbf{x} = \{x_1, x_2, \dots, x_T\}$  is given and a set of parameters  $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_p\}$  is learned that presumably characterizes the generating law of data. All parameters are dependent on the given data set:  $\theta_i = \theta_i \{x_1, x_2, \dots, x_T\}$   $i = 1, 2, \dots, P$ . Usually, a random variable  $\mathbf{X}$  is considered distributed to a stochastic process  $F$ , whereas  $\mathbf{x}$  is regarded as a realization of  $\mathbf{X}$ . Resampling generates surrogate data sets that eventually allow approximating the variance of each parameter (estimator) by a repeated learning of that parameter. In a bootstrap resampling approach, from a block of  $N$  data points, a sample is obtained by randomly drawing  $N$  elements from the original data set. It can be shown that in the case of independent and identically-distributed (i.i.d.) random variables, the bootstrap estimators of the distributions of common statistics are consistent [25]. The estimated parameters are denoted by  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{x}) = \{\hat{\theta}_1(\mathbf{x}), \hat{\theta}_2(\mathbf{x}), \dots, \hat{\theta}_p(\mathbf{x})\}$ , so that a scalar parameter  $\theta_i$  is estimated with an estimator  $\hat{\theta}_i(\mathbf{x})$  whose stability is assessed by the *root mean square error*:

$$\sigma_i = \sqrt{E_F \left\{ \left[ \theta_i - \hat{\theta}_i(\mathbf{X}) \right]^2 \right\}}, \quad i = 1, 2, \dots, P, \quad (3)$$

where  $E_F \{ \cdot \}$  stands for the expectation for  $F$ . Denote by  $\hat{F}$  the empirical distribution of data  $\mathbf{x}$ , so that a random variable from  $\hat{F}$  take the values  $x_t$  ( $t = 1, 2, \dots, T$ ) with equal probabilities  $1/T$ . Hence  $B$  surrogate data sets  $\mathbf{x}^{*b} = (x_1^{*b}, x_2^{*b}, \dots, x_T^{*b})$  are generated with  $b = 1, 2, \dots, B$  by taking  $T$  i.i.d. random variable  $x_1^{*b}, x_2^{*b}, \dots, x_T^{*b}$ , from the empirical distribution  $\hat{F}$ . For each bootstrap sample  $\mathbf{x}^{*b}$ , the estimator  $\hat{\theta}_i^{*b}(\mathbf{x}) = \hat{\theta}_i(\mathbf{x}^{*b})$  is calculated that results in  $B$  bootstrap replications  $\hat{\theta}_i^{*b}$ ,  $b = 1, 2, \dots, B$  of each  $\hat{\theta}_i$   $i = 1, 2, \dots, T$ . In this context, the robustness of estimation against small resampling changes of data is given by the bootstrap estimate of error [3]:

$$\hat{\sigma}_i(B) = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\theta}_i - \hat{\theta}_i^{*b})^2}, \quad i=1,2,\dots,P. \quad (4)$$

Accordingly, the bootstrap estimator (4) can be employed to rank different algorithms, or to choose hyperparameters and select among the solutions for one algorithm only.

### 2.3. MULTIDIMENSIONAL ICA MODEL

In the case of independent multidimensional subspaces [26], not all of the  $M$  latent sources assumed mutually statistically independent, rather they can be  $K < M$  higher dimensional independent subspaces such as:

$$P(S_1, S_2, \dots, S_M) = \prod_{k=1}^K P_k(S_{i_{k-1}}, \dots, S_{i_k}), \quad (5)$$

where the set of indices is  $1 = i_0 < i_1 < \dots < i_K = M + 1$ , and  $P(\cdot)$  denotes the joint probability density (pdf) function of the whole data set. Each independent subspace  $k$  is characterized by a pdf  $P_k(\cdot)$  that cannot be further factorized into a product of marginal pdf's. Since standard ICA comes out with an unordered set of  $1D$  source signal subspaces, a natural distance measure between two sources is the angle between their respective subspaces. Consider an orthonormal basis  $\{\mathbf{e}_i\}_i$  of the data space and let  $\mathfrak{A}$  and  $\mathfrak{B}$  be two  $1D$  subspaces and  $\mathbf{u} \in \mathfrak{A}$   $\mathbf{v} \in \mathfrak{B}$ , then the distance between  $\mathfrak{A}$  and  $\mathfrak{B}$  with respect to  $\{\mathbf{e}_i\}_i$  may be defined like:

$$d_{\{\mathbf{e}_i\}}(\mathbf{u}, \mathbf{v}) = \arccos \frac{\sum_i (\mathbf{u}\mathbf{e}_i) \cdot (\mathbf{v}\mathbf{e}_i)}{\sqrt{\sum_i (\mathbf{u}\mathbf{e}_i)^2 \cdot \sum_i (\mathbf{v}\mathbf{e}_i)^2}}. \quad (6)$$

The mixing/unmixing process can be expressed in terms of changing the coordinates from orthonormal basis  $\{\mathbf{e}_i\}_i$  to the canonical basis of the estimated latent sources  $\{\mathbf{f}_j\}_j$ , which is the basis of the  $M$  latent sources on a bootstrap sample:

$$\mathbf{s} = \sum_i \mathbf{e}_i s_i = \sum_j \mathbf{f}_j x_j, \quad \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, \quad i, j = 1, 2, \dots, M. \quad (7)$$

The linear transform between the two coordinate systems in the case of ICA reads like  $\mathbf{f}_j = \sum_{i=1}^M \mathbf{e}_i A_{ij}^{-1}$ ,  $j=1,2,\dots,M$ ; denoting the ICA estimated sources by

$\hat{\mathbf{s}}_i = \sum_{j=1}^M \hat{A}_{ij}^{-1} x_j$ ,  $j=1,2,\dots,M$  and  $\hat{\mathbf{s}} = \sum_{i=1}^M \hat{\mathbf{e}}_i \hat{\mathbf{s}}_i$ , then the estimated orthogonal basis is:  $\hat{\mathbf{e}}_i = \sum_{k,j} \mathbf{e}_k \hat{A}_{kj}^{-1} \hat{A}_{ji}$ ,  $i=1,2,\dots,M$  and the estimation error for the  $i$ -th component is:

$$E_i = d_{\{\mathbf{e}_i\}}(\mathbf{e}_i, \hat{\mathbf{e}}_i) = \arccos \frac{(A^{-1} \hat{A})_{ij}}{\sqrt{\sum_k (A^{-1} \hat{A})_{ki}^2}}. \quad (8)$$

In this context, the bootstrap estimator for the root mean square error is:

$$\hat{\sigma}_i = \sqrt{\frac{1}{B} \sum_{b=1}^B d_{\{\mathbf{e}_i\}}^2(\hat{\mathbf{e}}_i - \hat{\mathbf{e}}_i^{*b})^2}, \quad i=1,2,\dots,P, \quad (9)$$

where  $\{\hat{\mathbf{e}}_i^{*b}\}_i$  is the basis estimated from the  $b$ -th surrogate data set. Whitening implies that the separation matrix  $\mathbf{B}$  reduces to the rotation matrix  $\mathbf{W}$ . The separating matrices  $\hat{\mathbf{R}}^{*b}$  resulted from the surrogate data sets are approximately representing small rotations. Finding higher dimensional source signal subspaces is carried out by decomposing each rotation into  $N(N-1)/2$  elementary rotations within all  $2D$  planes spanned by the coordinate axes [3]:

$$\hat{\boldsymbol{\alpha}}^{*b} = \ln(\hat{\mathbf{R}}^{*b}), \quad (10)$$

where each component  $\alpha_{ij}$  of  $\hat{\boldsymbol{\alpha}}^{*b}$  is the angle of a rotation in the  $(ij)$ -plane. The variance of the  $\hat{\boldsymbol{\alpha}}^{*b}$ -matrix component-wise gives the *separability matrix* [3] that measures the instability of ICA projection w.r.t. a rotation in the  $(ij)$ -plane:

$$\hat{S}_{ij} = \sqrt{\frac{1}{B} \sum_{b=1}^B (\alpha_{ij}^{*b})^2}, \quad i, j=1,2,\dots,M. \quad (11)$$

The sICA algorithm separates the independent subspaces such that the separability matrix has a block structure that groups together  $ID$  ICA projections that cannot be reliably separated from each other but form a multidimensional stable subspace. The *uncertainty* of an ICA estimated  $K$ -dimensional subspace  $\mathfrak{S}$  spanned by the basis vectors  $\{\hat{\mathbf{e}}_i\}$ ,  $i \in I(\mathfrak{S}) = \{i_1, i_2, \dots, i_K\}$  can also be defined in terms of the separability matrix:

$$U(\mathfrak{S}) = \max \hat{S}_{ij}, \quad i \in I(\mathfrak{S}), \quad j \notin I(\mathfrak{S}). \quad (12)$$

Uncertainty is inversely related with the separation power of the ICA algorithm and directly related with the expected error.

### 3. RESULTS AND DISCUSSION

One healthy subject was selected for single-shot gradient-echo planar imaging (EPI) blood oxygenation level dependent fMRI scanning at 1.5 T, while running a simple block-type visual task during 12 identical sessions of 216 s each (Fig. 3). Acquisition and reconstruction matrices were both of dimensions  $64 \times 64 \times 64$  with voxel size of  $3.8 \times 3.8 \times 3.75 \text{ mm}^3$ . Sessions consisted of 72 volumes (scans) acquired at repetition time  $TR = 3 \text{ s}$ .

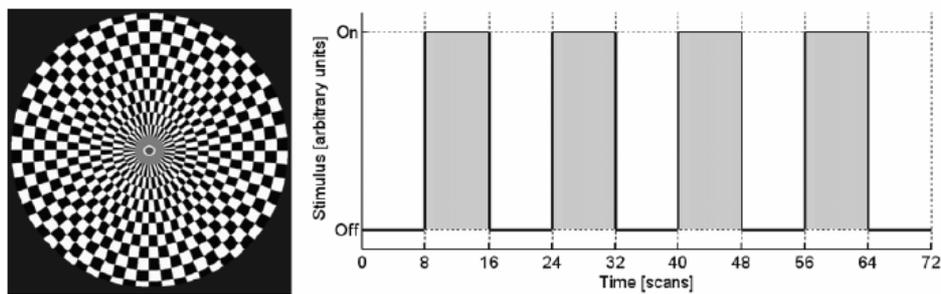


Fig. 3 – A flashing checkerboard (left) was presented in blocks of 24 s followed by 24 s of fixation (right), starting with fixation.

Acquired data underwent some preprocessing in SPM5 [4], and were subsequently plugged onto BS extended infomax spatial ICA algorithm [1]. Maximizing the information transfer in a nonlinear artificial NN minimizes the mutual information (MI) among the outputs when optimization is done over both the synaptic weights *and* the nonlinear transfer function.

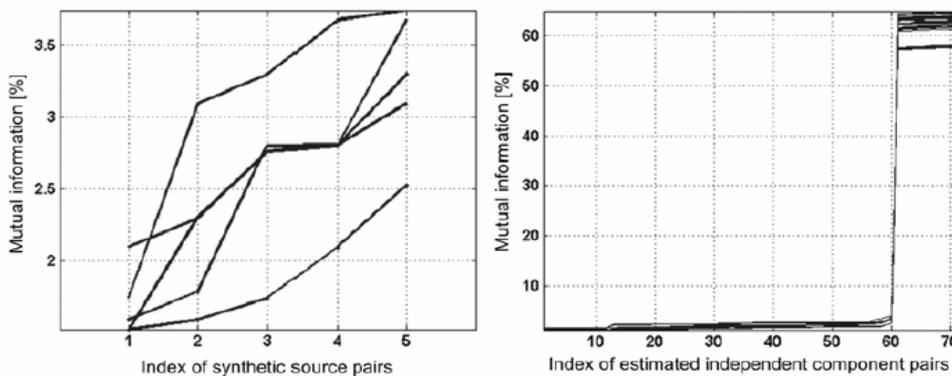


Fig. 4 – Mutual information of the selected centroids (left) and of the estimated components of the artificially generated data (right). The steep flexion at 60 corresponds to the existence of 6 clusters, each comprising 12 components [27].

The first most non-Gaussian 6 estimated ICs were assigned as centroids to artificially generate 6 independent clusters, each one consisting of 12 components obtained by shuffling the time points (*i.e.*, scans) of each centroid but preserving their temporal structure (Fig. 4). Resampling in spectral domain produced degenerated thresholds due to Fourier truncated model used, whereas waveletbased resampling resulted in quite a conservative threshold due to relatively few wavelet coefficients capturing the low frequency paradigm and enforcing poor randomization.

The interpretation of the stability matrix is based on the assumption that the more meaningful an estimated component is, the more frequently it should pop up when analyzing the bootstrap samples, that is, the more stable it should be. Low values of the elements in the stability matrix indicate good separability. The separability matrix corresponding to the estimated components separated by square ICA from a linear mixture of 72 artificially generated time series from 6 independent synthetic sources [27] shows 6 distinct separable clusters (Fig. 5).

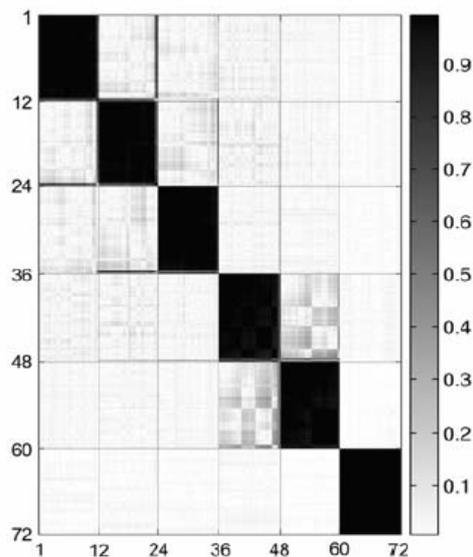


Fig. 5 – Separability matrix.

#### 4. CONCLUSION

Reliability estimation of neuroimaging data model is used to choose an appropriate ICA approach, to boost the separation power of the ICA algorithm, and to detect the estimated ICs that are most likely to have a realistic resampling neurophysiological meaning. The BOLD response in fMRI time series biases the estimation of temporal autocorrelation depending on the experimental paradigm

complexity, which entails biased thresholds. Resampling based on whitening transform has been proved the most robust in the presence of BOLD signal in a block-type fMRI experimental design.

As a general statistical method, the strategy is suited for assessing reliability in several unsupervised learning methods like the independent component [12] and fuzzy clustering analysis [28] of functional neuroimaging data.

## REFERENCES

1. A.J. Bell, T.J. Sejnowski, *An information maximization-approach to blind separation and blind deconvolution*, Neural Comput., **7**, 1129–1159 (1995).
2. P. Comon, *Independent component analysis, A new concept?* Signal Proces., **36**, 287–314 (1994).
3. F. Meinecke, A. Ziehe, M. Kawanabe, K.-R. Müller, *A resampling approach to estimate the stability of one-dimensional or multidimensional independent components*, IEEE Transactions on Biomedical Engineering, **49**, 12, 1514–1525 (2002).
4. K.J. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline, C.D. Frith, R.S.J. Frackowiak, *Statistical parametric maps in functional imaging: A general linear approach*, Hum. Brain Mapp., **2**, 189–210 (1995).
5. L. Mutihac, R. Mutihac, *Mining in chemometrics*, Anal. Chim. Acta, **612**, 1, 1–18 (2008)
6. O. Friman, C.-F. Westin, *Resampling fMRI time series*, NeuroImage, **25**, 859–867 (2005).
7. J.-F. Cardoso, *Blind signal separation: Statistical principles*, Proc. IEEE, **9**, 2009–2025 (1999).
8. R. Mutihac, M.M. Van Hulle, *PCA and ICA neural implementation for source separation*, Proc IEEE on Neural Networks, 2003, Oregon, **1-4**, pp. 769–774.
9. R. Mutihac, M.M. Van Hulle, *Comparison of PCA and ICA for BSS*, Romanian Reports in Physics, **56**, 1, 25–37 (2004).
10. K.J. Friston, *Modes or models: A critique on independent component analysis for fMRI*, Trends in Cognitive Sciences, **2**, 10, 373–375 (1998).
11. J. Karhunen, A. Cichocki, W. Kasprzak, *On neural blind separation with noise suppression and redundancy reduction*, Int. Journal of Neural Systems, **8**, 219–237 (1997).
12. R. Mutihac, R.C. Mutihac, *A comparative study of ICA algorithms for EEG*, Romanian Reports in Physics, **59**, 3, 831–860 (2007).
13. S. Amari, A. Cichocki, H.H. Yang, *A new learning algorithm for blind signal separation*, in Advances in Neural Information Processing, MIT Press, Cambridge, MA, **8**, 757–763 (1995).
14. O. Moreau, O. Macchi, *New self-adaptive algorithms for source separation based on contrast functions*, Proc. IEEE Signal Processing Workshop on Higher Order Statistics, 1993, pp. 215–219.
15. J.-F. Cardoso, A. Belouchrani, B. Laheld, *A new composite criterion for adaptive and iterative blind source separation*, Proc. ICASSP, 1994, pp. 273–276.
16. R. Mutihac, M.M. Van Hulle, *Neural network implementations of independent component analysis*, Proc. IEEE Workshop on Neural Networks for Signal Processing, Martigny, 2002, pp. 505–514.
17. J.-F. Cardoso, A. Souloumiac, *Blind beamforming for nongaussian signals*, IEE Proc.-F., **140**, 362–370 (1993).
18. Pham D.-T., Garrat P., Jutten C., *Separation of a mixture of independent sources through a maximum likelihood approach*, Proc. EUSIPCO, 1992, pp. 771–774.
19. M. Woolrich, B. Ripley, M. Brady, S. Smith, *Temporal autocorrelation in univariate linear modeling of fMRI data*, NeuroImage, **14**, 6, 1370–1386 (2001).

20. A. Laird, B. Rogers, M. Meyerand, *Comparing of Fourier and wavelet resampling methods*, Magn. Reson. Med. **51**, 2, 418–422 (2004).
21. M. Breakspear, M. Brammer, E. Bullmore, P. Das, L. Williams, *Spatiotemporal wavelet resampling for functional neuroimaging data*, Hum. Brain Mapp., **23**, 1, 1–25 (2004).
22. E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, M. Brammer, *Colored noise and computational inference in neurophysiological (fMRI) time series: Resampling methods in time and wavelet domains*, Hum. Brain. Mapp., **12**, 2, 61–78 (2001).
23. J. Locascio, P. Jennings, C. Moore, S. Korkin, *Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging*, Hum. Brain Mapp., **5**, 3, 168–193 (1997).
24. R. Mutihac, *Wavelet-based statistical analysis versus SPM of functional brain imaging*, International Journal of Intelligent Computing in Medical Sciences and Image Processing (IC-MED), **2**, 3, 225–236 (2008).
25. J. Shao, D. Tu, *The Jackknife and Bootstrap*, Springer, New York, 1995.
26. J.-F. Cardoso, *Multidimensional independent component analysis*, Proc. ICASSP, IEEE, 1998.
27. R. Mutihac; *On model complexity estimation in non-square ICA of fMRI data*, Proc. ISMRM Miami Beach, Florida, 2005, p. 1594.
28. R. Mutihac, *ICA of brain imaging data – Validation by resampling and hierarchical clustering*, NeuroImage, **41**, Supp. 1, S161 (2008).