An introduction to Multiway Methods for Multi-Subject fMRI experiment.

FMRIB Technical Report

Didier G. Leibovici and Christian Beckmann

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK
`[didier]or[beckamnn]@fmrib.ox.ac.uk`

Abstract

Model free exploratory methods available for functional MRI analysis of one subject such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are investigated for generalisations when analysing more than one subject. Introducing the subject dimension makes a three-way data: brain, time, subject, and multiway methods are proposed either to optimise the variance (PCA) or the Negentropy (ICA). Results of folded methods and purely multiway ones are described on a 12 subjects verbal study.

1 Introduction

Principal Component Analysis (PCA)[6] and Independent Component Analysis (ICA)[7] are established exploratory methods for single subject analysis in Functional MR imaging. The spatio-temporal decomposition of the *data matrix* (up to a given level r) in both cases can be written:

$$X = \sum_{i}^{r} B_i \otimes T_i + \varepsilon \tag{1}$$

with different statistical properties according to the chosen method. Inferences for *location* on the brain with components B_i [1][2] and *stimulus influence* with associated time course component T_i (e.g. correlation with the paradigm) describe the functional activation. For multi-subject analysis generalisations of these methods to 3-way arrays are needed to get a decomposition of the *data tensor* of order 3 (space, time, and subject) in a form:

$$Y = \sum_{i}^{r} B_i \otimes T_i \otimes S_i + \varepsilon \tag{2}$$

Testing the subject component would allow a *population inference* and can be viewed as a spatio-temporal omnibus test. For the variance criterion (PCA) the PTA-k method [3, 4] offers a decomposition like (2) and potentials to describe multi-subject fMRI data will be illustrated. On the way to optimal rank one decomposition using ICA criterion, analyses of a three-way data seen as a stacked two-way data offers a first step forward: the Single-ICA and Multiple-ICA methods. With an analogy of a new interpretation of the algorithm used for PTA-k method [3] we will also present a *rank-one version* of the Single-ICA. Other 3-way ICA methods can be derived fixing the independence criterion in one dimension only, or on two dimensions (*space* and *time*) with a similar algorithm. These latter will be illustrated with a rank one version of the Multiple-ICA.

2 Variance criterion

For the variance criterion (PCA) the PTA-k method [3, 4] offers decompositions like (2). The optimisation scheme was derived from searching for *singular values*:

$$\sigma_1 = \max_{\substack{\|\psi\|_E=1 \\ \|\varphi\|_F=1 \\ \|\phi\|_G=1}} Y..(\psi \otimes \varphi \otimes \phi)$$

Table 1: PTA-3modes listing of the decomposition up to the third $k$modes and associated solutions.

```
 ++++ PTA-  3 modes ++++
              data=  J12.gm16   576 100 12
   just slice 16
               ------Percent Rebuilt---- 23.34021%
               ------Percent Rebuilt from Selected ---- 17.01323%
                -no- --Sing Val-- --ssX-- --local Pct-- --Global Pct--
vs111                1      191.376  599772        6.1064          6.1064
576 vs111 100 12     3       85.062   70189       10.3088          1.2064
100 vs111 576 12     6       97.382   81266       11.6694          1.5812
100 vs111 576 12     7       87.138   81266        9.3434          1.2660
12 vs111 576 100     9       87.616   80815        9.4988          1.2799
vs222               11      133.299  440752        4.0314          2.9625
vs333               21       97.991  370027        2.5950          1.6010
12 vs333 576 100    29       77.825   42284       14.3237          1.0098

 ++++               ++++
 Shown are selected over  21  PT  with var> 1% total
```

$$= \quad Y..(B_1 \otimes T_1 \otimes S_1). \tag{3}$$

where E, F, G are the Hilbert spaces of finite dimensions (embedding of components in functional Hilbert spaces is also possible), the operation .. is the contraction of the the tensors considered, here equivalent to inner product in the tensor space. To search for the second singular values and Principal Tensor associated an orthogonality constraint is added [4].

An illustration of the output of the method using the `PTAk` package [4] follows. The table (1) gives the beginning of the decomposition (pruned). The first Principal Tensor shown on figure (2) is well correlated (the time component) to the paradigm of the experiment. Each component can be tested (with appropriate tests) separately to confirm BOLD activation at a population level, *i.e.* the correlation is significant showing some activation, related globally to the population (with no apparent differences due to sex), and high values[1] locates activation on the left fronto-temporal of the brain (right on the image). The second Principal Tensor (figure 2) shows a motion artifact of one particular subject.

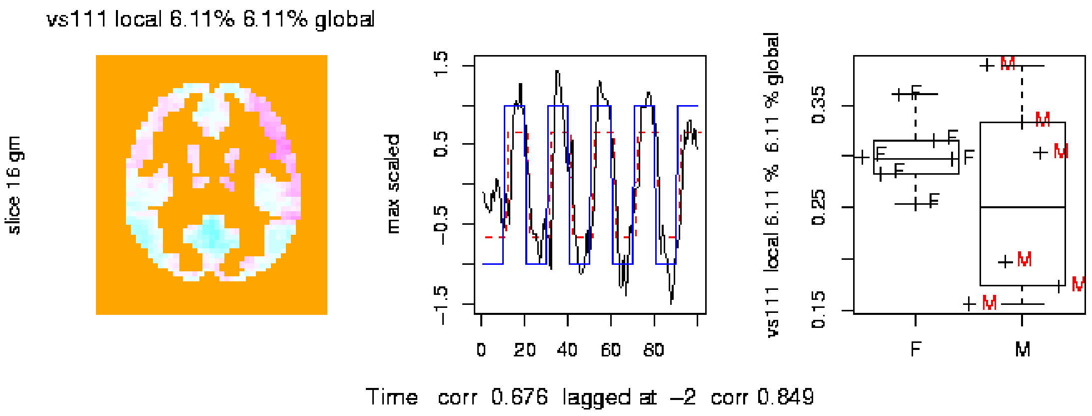

Figure 1: PTA3-modes of the detrended time series and subject scaled of a 12 subjects fMRI data: activation

In fact the criterion of singular values (3) can be written in variance criterion form reminiscent of the optimisation for eigenvalues of a "covariance" matrix:

$$\begin{aligned}
\sigma_1^2 &= \max_{\substack{\|\varphi\|_F=1 \\ \|\phi\|_G=1}} E([Y'..(\varphi \otimes \phi)]2) \\
&= \max_{\substack{\|\varphi\|_F=1 \\ \|\phi\|_G=1}} E([Y' \otimes Y'])..([\varphi \otimes \phi] \otimes [\varphi \otimes \phi]) \\
&= \max_{\|\varphi\|_F=1} E([Y'..\phi_1] \otimes [Y'..\phi_1])..(\varphi \otimes \varphi)
\end{aligned}$$

[1]Note one should test the image for local maximum to confirm inferentially the spatial description of activation

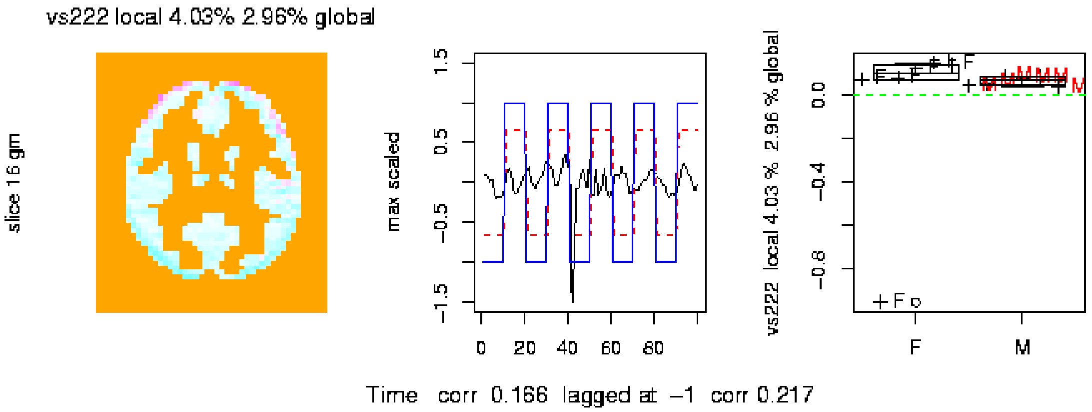

Figure 2: PTA3-modes of the detrended time series and subject scaled (whole sequence masked brain)of a 12 subjects fMRI data: motion artefact

$$
\begin{aligned}
&= \max_{\|\phi\|_G=1} E([Y'..\varphi_1]\otimes[Y'..\varphi_1])..(\phi\otimes\phi) \\
&= E(Y'\otimes Y')..(\varphi_1\otimes\phi_1) \qquad (4)
\end{aligned}
$$

where here Y' is a random matrix in $F\otimes G$ from which N observations (here voxels) were collected in Y. This random version is interesting in giving some differnets ways of implementing an algorithm and in describing the PTA-3modes method in a statistical framework rather than in algebra. By analogy generalisation in term of a different optimisation index based on distribution of the component rather than only on the variance of can be derived.

3 Negentropy criterion

With PCA one obtains a decomposition like 1 with variance criterion and the constraints on the components is no correlation. Generalisations of these criterion and constraints has been a continuous endeavour and the literature is abondant about this subject. Recently the concept of Independent Components Analysis has been widely used in fMRI analysis. Initially the purpose was to look for Independent Components at a higher order than two (no correlation). The motivation was to recover non-gaussian latent variables (called sources) who generated the data by linear mixing of them, and supposed to be mutually independent. The link between this two concepts can be understood as follows.
A measure of mutual independence [5] of the p variables y_j is the Kullback-Liebler divergence between the joint density and the product of the marginal densities :

$$
\begin{aligned}
MI(y) = KL(f_y||\prod_{j=1}^{p} f_{y_j}) &= E_y(\ln(\frac{f_y(u)}{\prod_{j=1}^{p} f_{y_j}(u_j)}) \qquad (5)\\
&= \int f_y(u)\ln(\frac{f_y(u)}{\prod_{j=1}^{p} f_{y_j}(u_j)}du \qquad (6)
\end{aligned}
$$

and is often termed mutual information. Negentropy itself can be defined as the Kullback-Liebler divergence between the density (single or vector variable) and the normal equivalent *i.e.* with the same mean and variance, or difference of the corresponding entropies [2]:

$$N(y) = h(\phi_y) - h(f_y) = -E_\phi(\ln(\phi_y(u)) + E_f(\ln(f_y(u)) = KL(f_y||\phi_y) \qquad (7)$$

The Negentropy can be related to the mutual information:

$$MI(y) = N(y) - \sum_{j=1}^{p} N(y_j) + \frac{1}{2}\ln(\frac{\prod_j V_{jj}}{det(V)}) \qquad (8)$$

[2]The difference is always positive *i.e.* the entropy of a normal distribution $h(\phi_y) = -E(\ln(\phi_y)) = 1/2(p + p\ln(2\pi) + \ln(det(V)))$ has the largest entropy among p-dimensional distributions. The proof of the equivalent definitions comes from $E_{\phi_y}(\ln(\phi_y)) = E_{f_y}(\ln(\phi_y))$ because $\ln(\phi_y)$ is a polynomial of degree two and ϕ_y and f_y have same first and second moments [5]

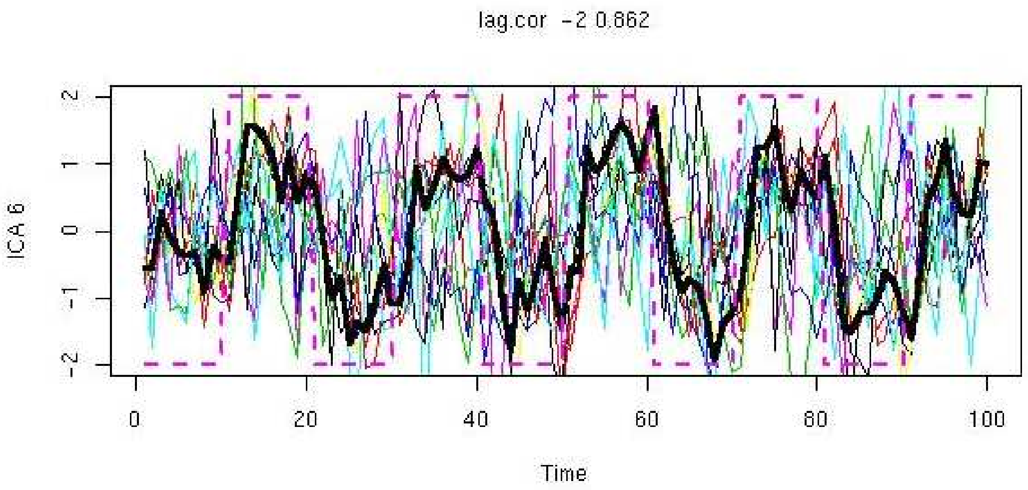

Figure 3: Single-ICA of the detrended time series and subject scaled of a 12 subjects fMRI data: unmixing matrix split by subject with best mean time-course lagged correlation to paradigm; the mean time-course is in black.

then as the Negentropy is invariant by invertible transformation minimising the mutual information is equivalent to maximising negentropies of components which then are the least Gaussian variables obtained by linear transformation of the original vector variable. This assertion is in fact true is the third term is null as it is not invariant by invertible transformation. This term which would correspond to a PCA criterion is null when the data is spherised. One can see the analogy with Projection-Pursuit where projections of the data who are the least Gaussian are looked for often using also the negentropy as non-Gaussianity measure. The difference is that in Projection-Pursuit [8] the projections are not necessarily fixed to unidimensional ones.

4 Single and Multiple ICA

For a three-way data array Y being a sample of size N (number of voxels of the brain image) of a random matrix Y' of dimensions $T \times S$ (T= number of Time points , S=number of Subjects), different two-ways ICA can be investigated. It is possible to look for an Independent Components in the *spatial* dimension associated to $time - course \times subject$ unmixing matrices. We call it Single-ICA by analogy with single subject ICA. Note this is done with classical ICA on a data matrix, the unmixing matrices are in fact the unmixing vectors of length the product $T.S$.

$$Y = \sum_{i}^{r} B_i \otimes T_{S_i} + E \tag{9}$$

Intuitively one supposes the unmixing vector as “different” from subject to subject but the sources should be similar i.e. we hope the sub-components (unmixing for each subjects) to be similar. This should be considered as a random subject model but here is considered only as a fixed subject effect as in fact we may want to look at the variation *a posteriori*. At this stage in a similar way of nested PCA one could look at the SVD decomposition of the unmixing vector as bilinear forms ($time - course \times subject$) and probably retain the best low rank approximation of the unmixing matrix.
One gets very similar results as with `PTAk` for the correlation with the paradigm and spatial representation. For the subject representation on figure (4) is slightly different even if the interpretation of it remains the same. The variances within group seem similar not like for `PTAk` where the Female group had a smaller variance.
It is also possible to look for *combined subject-brain* Independent Components associated to the same *time-course* mixing vector:

$$Y = \sum_{i}^{r} B_{S_i} \otimes T_i + E \tag{10}$$

It is called Multiple-ICA as the “traditional” spatial (brain) component is multiple. One would hope the component to reflect a “narrow” distribution over subjects of this traditional spatial component. This last model correspond more to the intuitive consideration that the subjects responses to an fMRI experiment are

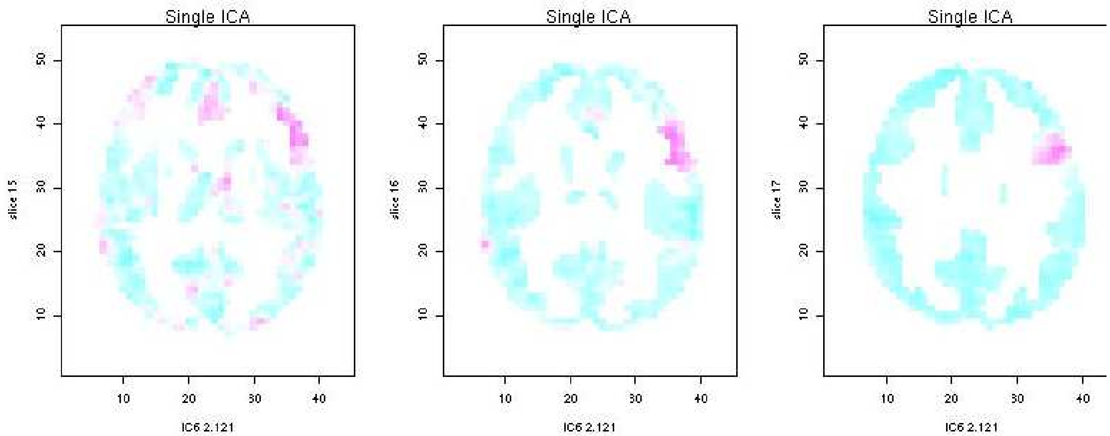

Figure 4: Single-ICA of the detrended time series and subject scaled of a 12 subjects fMRI data: brain component associated to the best correlated average time-course to paradigm; 2.121 is the Jones & Sibson approximation of the Negentropy for this component.

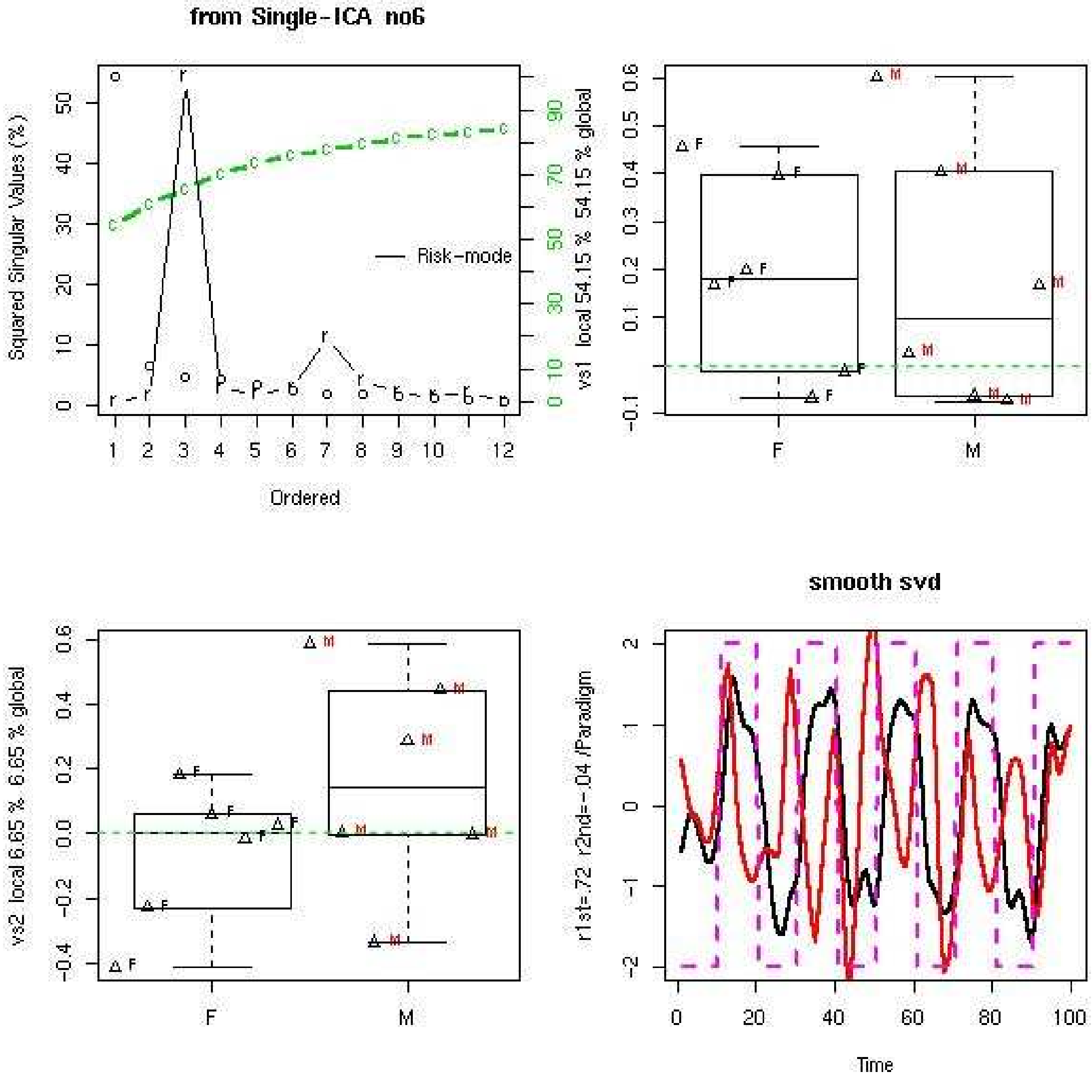

Figure 5: Single-ICA smooth SVD [4] of the unmixing matrix (fig.4) of ICA6; the first component well correlated with the paradigm is associated with no apparent difference for sex; on the scree plot c is for cumulated and r an estimation of a risk in choice dimension.

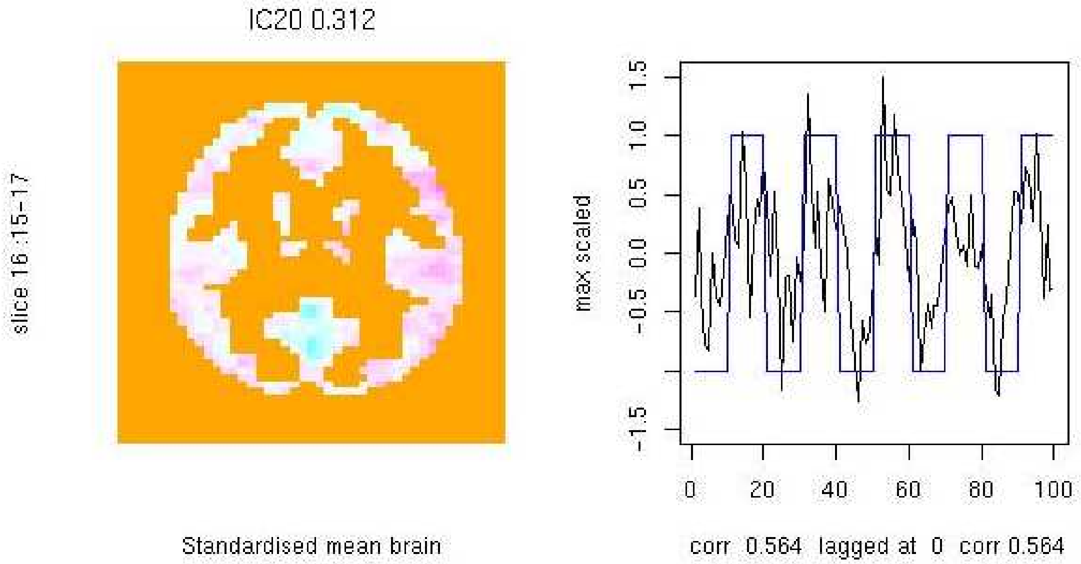

Figure 6: Multiple-ICA standardised mean over subjects of ICA20 and unmixing time-course.

independent identically distributed and we are looking for a random variation of activation pattern hopefully i.e. the random variation between subjects is accessible at each voxel. Nonetheless this last method is biased towards non-Gaussian subject samples, i.e. as well as looking for spatial distribution the least Gaussian the method will try to separate the subjects and with small samples of subjects this may destroy any hope of getting similar spatial distributions.

5 Towards rank-1-ICA

The last point of the previous section illustrates that in fMRI we may be interested in Single-ICA (unmixing dimension is time-subject) or in a Multiple-ICA if the multiple component is the spatio-temporal dimension, the subject dimension being the unmixing dimension. For these two ways of approaching ICA for multi-subject experiments a rank-one version of the ICA is certainly more interesting. With the Single-ICA model we would look for a rank-one unmixing matrix ($time \times Subject$)and with the Multiple-ICA model we will look for a rank-one component (a *spatio-temporal* component).
The first idea of getting a rank-one component is in nested way of performing analysis, to do it once the ICA is finished and in summarising the unmixing matrix by it best rank-one approximation by an SVD. Note that doing the same thing for the Multiple-ICA approach (*i.e.* best approximation of the ICA component) one would not get necessary an optimised negentropy! Note also that for the Single-ICA the two steps are optimised separately but that usually does not mean that the rank-one unmixing is achieving maximum negentropy. This is what we naturally did on figure 4 or when looking at the average time-course a a pot-hoc way of describing the result.
In order to achieve a better algorithm one have to rewrite the optimisation incorporating the rank-one "constraint". For Single-ICA that look really like a constraint as one wants the unmixing matrix to be of rank one. With the Multiple-ICA approach it is different as one want the component to be of rank one. This last one is more complicated, so we will focuses on the first one which was also the most sensible approach for fMRI data. A simple approach for rank-one-Single-ICA is to use a penalised SVD as describe in [4] and force "smoothing" on the principal components during optimisation as to be issued from a rebuild from a one component ICA model. This was suggested by the equivalence (modulo some orthogonality constraints) between penalised (smoothing as best rank-one SVD) SVD and the PTA-3modes. The problem with this approach is that the "smoothing" may violate too much the conditions for least squares optimisation under smoothing constraint making the algorithm fails to converge [11]. Another rank-one-Single-ICA can be derived incorporating in the ICA algorithm a constraint on the unmixing to be of rank-one. This is done after every updating unmixing vector then alternating uptdating the Newton algorithm for best direction and best least squares rank-one approximation of the unmixing vector put as a matrix $time \times subject$. Figure 5 shows one set of components resulting from this approach. The problem here is that the unmixing vector length is huge comparatively to

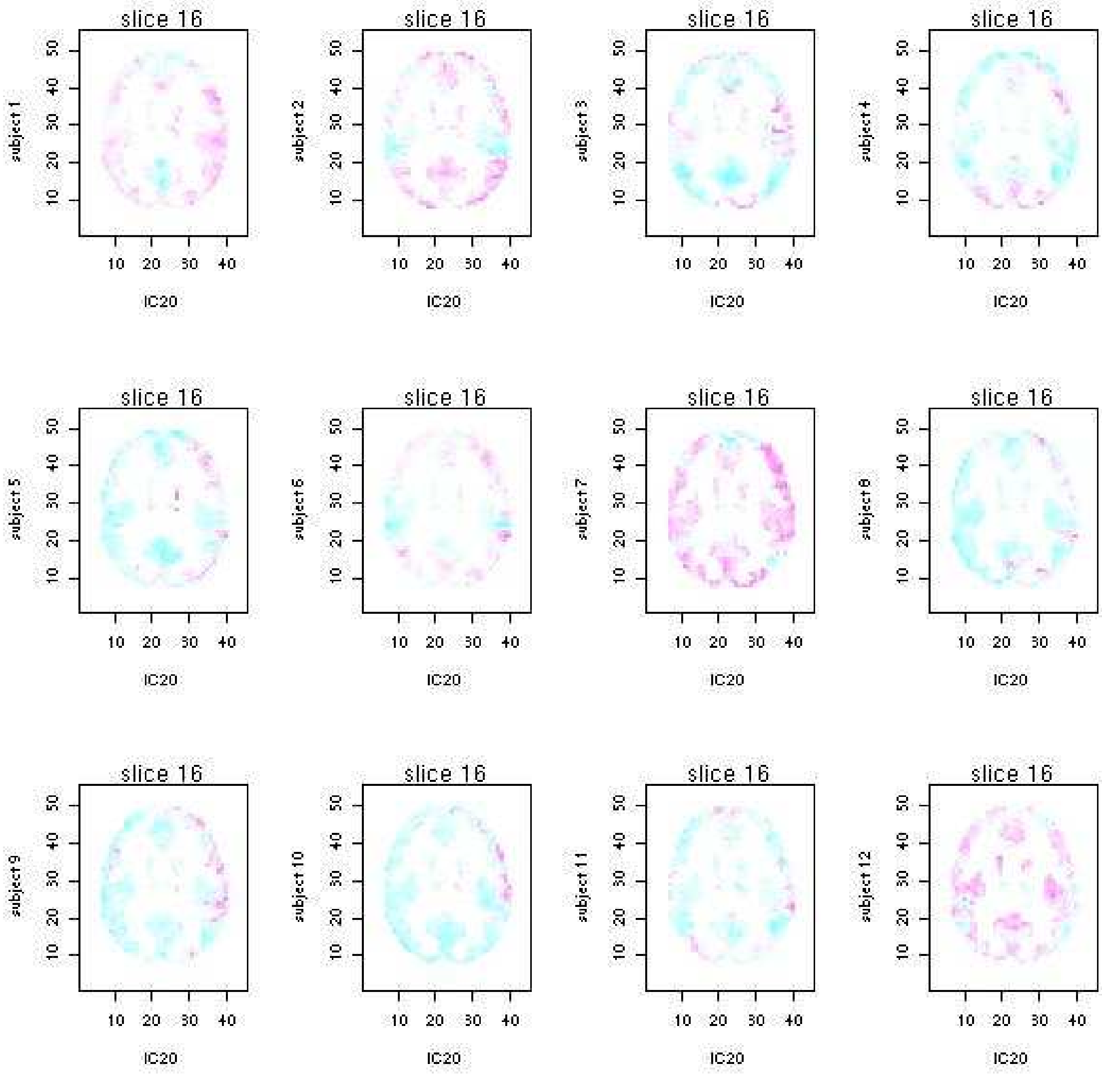

Figure 7: Multiple-ICA ICA20; the scale is different for each subject.

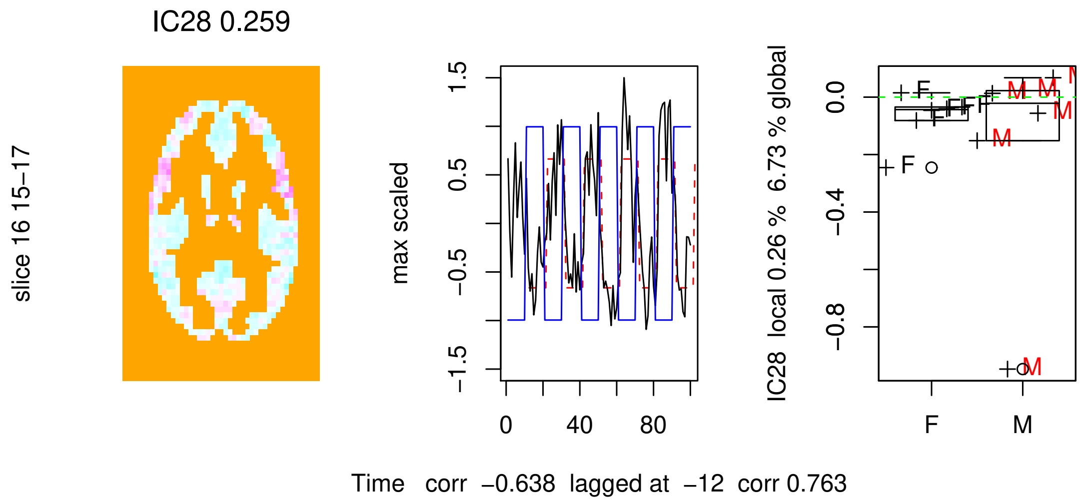

Figure 8: rank one Single-ICA by constraint in an ICA: ICA28 the best correlated time-course

dimension the data is susppected to lies in [2]. A dimension reduction is usually done before ICA and the unmixing vector "projected back" onto the original space. Because of the rank-one contraint here one has either not to reduce the dimension before ICA making difficult to find ICA's in a raisonnable time or to try to incorporate this reduction anyway!
This heuristic looking approach can be in fact derived directly from the rewriting of the optimisation problem as in [10] with a tensor of rank one as argument instead of a vector.

6 Conclusions and Perspectives

Some statistical properties and comparisons of these new multiway methods are still needed to complete our first ideas. The application on a verbal study with 12 subjects gave some insights on their different interests.
It seems rather obvious that when looking for activation with a regular paradigm one certainly expects the variance criterion to give good results especially if the subject dispersion (between groups and /or within groups) is expected to be relatively Gaussian in the direction of interest. In that purpose Single-ICA achieve similar results as in using PTA3. This is because of the first step of reducing the dimensionality (using variance criterion) in te ICA algorithm. Without this first step too many time-courses have to be fit. Probably with a drastic dimension reduction using a variance criterion, the real interest of ICA is less evident and certainly the rank-one approach is an other way of reducing the dimensionality enforcing initial Gaussianity by forcing independence of *time and subject* in the unmixing component. .
Finding complicated group structures and/or expecting non regular paradigm would be better achieved with Negentropy approach, but the Multiple-ICA approach has too many degrees of freedom for that purpose. In reducing the degree of freedom on the independent Components the rank-one Multiple-ICA should give better results than the Multiple-ICA. Nonetheless using special metrics[9,4] with PTAk may provide interesting alternatives and may be a rank-one Single-ICA using these special metrics on the time-subject unmixing would allow to find these
Another 3-way ICA methods can be thought fixing the independence criterion on two dimensions (*space* and *time*). This is the other rank-one ICA approach *i.e.* like a Multiple-ICA but looking for a rank-one independent components.

References

[1] Benali, H et al.(1997). Space-Time Statistical Model for Functional MRI Image Sequences. *Information Processing in Medical Imaging*, 285-298.

[2] Beckmann, C. (2001). Spatio-Temporal accuracy of ICAs for FMRI. *Human Brain Mapping 7th annual meeting*, section Analysis.

[3] Leibovici, D. and Sabatier, R. (1997). A Singular Value Decomposition of a k-ways array for a Principal Component Analysis of multi-way data, the PTA-kmodes. *Linear Algebra and its Applications*, **269**, 307-329.

[4] Leibovici, D.(2001). An SVD-kmodes R-code, the package `PTAk` version 1.1-4. released on `CRAN`. `ftp://cvs.r-project.org/pub/CRAN/src/contrib/PACKAGES.html#PTAk`, fMRI examples on,`http://www.fmrib.ox.ac.uk/~didier/work/tensor.html`

[5] Comon, P.(1994). Independent Component Analysis, a new concept? *SIGNAL Processing*, 36:287-314.

[6] Worsley,K.J Poline, J-B Friston, K.(1997). Characterizing the Response of PET and fMRI Data Using Multivariate Linear Models. *Neuroimage* **6**:305-319.

[7] Beckmann, C.(2001). Independent Component Analysis for fMRI experiments. (*to be submitted*).

[8] Jones, and Sibson, (1987). What is projection pursuit?) *Journal of the Royal Statistical Society series B*, **150**:1-36.

[9] Caussinus, H and Ruiz, A (1990). Interesting Projections of Multidimensional Data by Means of Generalised Principal Components Analysis. *COMPSTAT90, Physica-Verlag, Heidelberg*,121-126.

[10] Hyvrinen, A (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. on Neural Networks*, 10(3):626-634, 1999.

[11] Leibovici, D. (2001). A Penalised Algorithm for SVD and Multi-way Functional Data Analysis Methods. (*in progress*).