Tensorial Extensions of Independent Component Analysis for Multi-Subject FMRI Analysis

FMRIB Technical Report TR04CB1
(A related paper has been accepted for publication in NeuroImage)

Christian F. Beckmann and Stephen M. Smith

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
Corresponding author is Christian F. Beckmann: beckmann@fmrib.ox.ac.uk

Abstract

We discuss model-free analysis of multi-subject or multi-session FMRI data by extending the single-session Probabilistic Independent Component Analysis model (PICA; [2]) to higher dimensions. This results in a three-way decomposition which represents the different signals and artefacts present in the data, in terms of their temporal, spatial and subject-dependent variations. The technique is derived from and compared with Parallel Factor Analysis (PARAFAC; [11]). Using simulated data as well as data from multi-session and multi-subject FMRI studies we demonstrate that the tensor-PICA approach is able to efficiently and accurately extract signals of interest in the spatial, temporal and subject/session domain. The final decompositions improve upon PARAFAC results in terms of greater accuracy, reduced interference between the different estimated sources (reduced cross-talk), robustness (against deviations of the data from modelling assumptions and against overfitting) and computational speed. On real FMRI 'activation' data, the tensor-PICA approach is able to extract plausible activation maps, time courses and session/subject modes as well as provide a rich description of additional processes of interest such as image artefacts or secondary activation patterns. The resulting data decomposition gives simple and useful representations of multi-subject/multi-session FMRI data that can aid the interpretation and optimisation of group FMRI studies beyond what can be achieved using model-based analysis techniques.

Keywords: Tensor decomposition; Independent Component Analysis; PARAFAC; functional magnetic resonance imaging;

1 Introduction

Exploratory data analysis techniques like Principal Component Analysis (PCA) or Independent Component Analysis (ICA) are becoming increasingly popular for the analysis of data from functional imaging experiments, mainly for their potential to account for unknown yet structured spatio-temporal processes in neuroimaging data [21, 26, 2].

Current PCA/ICA methodology typically represents the original 4-D data from a single FMRI experiment as a two-dimensional (time×space) data matrix $\boldsymbol{X}$, which is decomposed into a sum of R outer products of individual *factors*:

$$\boldsymbol{X} = \sum_{r}^{R} \boldsymbol{a}_r \otimes \boldsymbol{b}_r + \boldsymbol{E}.$$

In this spatio-temporal decomposition, the entire data set is represented by different spatial processes, encoded as vectors $\boldsymbol{b}_r$, and associated temporal dynamics, encoded as vectors $\boldsymbol{a}_r$ and confounded by typically Gaussian noise $\boldsymbol{E}$. The exact relationship within and between the sets of vectors differs according to the chosen analysis methodology, e.g. a PCA decomposition enforces orthogonality within the set of time courses and within the set of spatial maps, while an ICA decomposition relaxes the requirements on the time courses and instead places stronger restrictions (statistical independence) on the set of spatial maps. In either case, however, the underlying algorithmic concepts are rooted within the matrix algebraic framework and factorise a single 2-dimensional data matrix into time courses and spatial maps.

Typical neuroimaging studies, however, involve the generation of data from multiple subjects, potentially over a set of different sessions. In [1] we demonstrated how these data generation scenarios fit into a hierarchical multi-level general linear modelling (GLM) framework, where - at every level - results from lower level GLM analysis are combined into a new set of higher-level estimates. While this approach would also be applicable to exploratory techniques like PCA or ICA[1], it requires resorting to model-based analysis at the higher level. In the case of exploratory data analysis techniques there is a potential loss in not modelling and infering on the entire set of data at once: when data is analysed separately there is little scope in mutually conditioning analysis results. This is less of an issue in the model-based (GLM) FMRI analysis, for it is being used to test very specific, *a-priori* hypotheses about the structure of the data. In the case of exploratory techniques, however, analysing multiple data all at once can improve the ability to extract spatio-temporal modes of interest and aid subsequent interpretation.

Current approaches to multi-subject group FMRI ICA decompositions involve concatenating the data either in space [18, 27] or time [6], and then apply a standard two-dimensional ICA decomposition approach. This results either in time courses which are common across subjects together with subject-specific spatial maps or vice versa. Therefore, either in space or time, multiple factor estimates represent the way signal is contained in the individual subjects. In these cases the final representation at the group level differs from standard group FMRI GLM analysis, where experimenters typically are interested in finding *single* spatial maps which, together with *single* temporal modes, jointly describe an individual source process in space and time across the set of subjects[2]. This sparse representation of signal is attractive particularly for its conceptual simplicity. Concatenating data in space or time across subjects prior to the analysis, however, effectively ignores the existence of modes of variation beyond time and space (like within-group variability) and potentially sacrifices accurate data modelling for algorithmic simplicity. Within the current group ICA methodologies, this is partly addressed by performing some heuristic meta-analysis after estimation of the individual modes, e.g. by calculating the mean temporal response or back-projection in order to obtain individual spatial maps which then can be averaged [6]. During the estimation stage itself, however, the multi-dimensional structure of the data is not reflected in the analysis.

In this work we are going to discuss alternative approaches based on generalising the standard bilinear (two-way) exploratory analysis methodology to higher dimensions. To this end, we are going to introduce an iterated rank-1 tensor ICA approach that will decompose a three-way data set into a set of independent spatial maps together with associated time courses *and* estimated subject modes. In analogy to the two-dimensional case, where ICA is often introduced as an extension to PCA, we will derive the tensor ICA technique as an extension to three-way PCA generalisations.

One such possible generalisation is known as *Parallel Factor Analysis* (PARAFAC[3] [10, 12]), where a three-way

[1] i.e. by feeding the results of the decomposition into the model-based higher-level analysis

[2] restricting ourselves to a simple single mean-group analysis for the moment.

[3] Also known as 'canonical decomposition' (CANDECOMP [8])

array is represented by a *tri-linear* combination of 3 outer products:

$$\boldsymbol{X} = \sum_{r}^{R} \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r + \boldsymbol{E}.$$

Techniques like PARAFAC have gained popularity in some scientific disciplines, e.g. in chemometrics they have been used extensively to decompose fluorescence spectroscopy data and in neuroimaging have recently been applied to EEG data to give a space-time-frequency decomposition [23]. For FMRI data, these vectors might represent variation in time, across space and between subjects[4] and jointly form a data tensor of order 3.

In order to obtain a practical data analysis technique, the generative model needs to be augmented with a suitable cost function. Standard PARAFAC analysis generalises PCA and treats the decomposition as a sum-of-squares minimisation problem. Unlike PCA, however, the PARAFAC decomposition does not require orthogonality between any of the vectors in the representation [10]. We will review the PARAFAC approach to the decomposition of three-way arrays and, using experiments on artificial FMRI data, will demonstrate that final PARAFAC estimates can exhibit significant amounts of cross-talk between estimated factors, especially in the spatial domain, with negative impact on the interpretability of results.

In the bilinear case previous research [2] has already demonstrated that the ICA approach has beneficial properties compared to a purely variance based representation as used in PCA. In particular, observed cross-talk between estimated sources was much reduced. Here, we extend this approach to a probabilistic tensor ICA model, where the generative model assumes three-way data in the presence of noise. Instead of optimising for minimum residual sum-of-squares error, we propose to optimise for maximum neg-entropy (non-Gaussianity) of estimated spatial modes.

We begin with a technical description of PARAFAC and a development of the tensor ICA method. The two techniques will then be compared on a set of artificial data sets and on a real FMRI group study.

2 Parallel Factor Analysis (PARAFAC)

The three-way PARAFAC technique is characterised by the following generative model:

$$x_{ijk} = \sum_{r}^{R} a_{ir} b_{jr} c_{kr} + \epsilon_{ijk} \tag{1}$$

$(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K)$ with an associated sum-of-squares loss:

$$\min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \sum_{ijk} \left\| x_{ijk} - \sum_{r}^{R} a_{ir} b_{jr} c_{kr} \right\|^2 . \tag{2}$$

Here, $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R), \boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_R)$ and $\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R)$ denote the $I \times R$, $J \times R$ and $K \times R$ matrices containing the R different factor loadings in the temporal, spatial and subject domain as column vectors. Within this model, any solution to equation 1 is a maximum likelihood solution under the assumptions of Gaussian noise.

The tri-linear model can alternatively be written in matrix notation, giving an expression for the individual 2-D subsets of $\boldsymbol{X}$ [5]:

$$\boldsymbol{X}_{i..} = \boldsymbol{B}\mathrm{diag}(\boldsymbol{a}_i)\boldsymbol{C}^t + \boldsymbol{E}_{i..} \quad i = 1, \ldots, I \tag{3}$$

$$\boldsymbol{X}_{.j.} = \boldsymbol{C}\mathrm{diag}(\boldsymbol{b}_j)\boldsymbol{A}^t + \boldsymbol{E}_{.j.} \quad j = 1, \ldots, J \tag{4}$$

$$\boldsymbol{X}_{..k} = \boldsymbol{A}\mathrm{diag}(\boldsymbol{c}_k)\boldsymbol{B}^t + \boldsymbol{E}_{..k} \quad k = 1, \ldots, K, \tag{5}$$

where $\mathrm{diag}(\boldsymbol{a}_i)$ denotes a $R \times R$ diagonal matrix where the diagonal elements are taken from the elements in row i of $\boldsymbol{A}$ (similarly for $\mathrm{diag}(\boldsymbol{b}_i)$ and $\mathrm{diag}(\boldsymbol{c}_i)$). This gives rise to a set of coupled sum-of-square loss functions. Based on these, a simple way of estimating the factor matrices is to use an iterative *Alternating Least Squares* (ALS) approach, iterating between the least-squares estimates for one of $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ separately while keeping the other

[4]Note that, similar to the two-dimensional case, we can freely pass scalar factors between estimates and also introduce permutations. Absolute amplitude in any of the factors is only meaningful when fixing all other factors to e.g. unit standard deviation or unit range.

two matrices fixed at their most recent estimate:

$$\begin{aligned}
\widehat{\boldsymbol{A}} &= \left(\sum_k \boldsymbol{X}_{..k}\boldsymbol{B}\mathrm{diag}(\boldsymbol{c}_k)\right)\left((\boldsymbol{B}^t\boldsymbol{B})\circ(\boldsymbol{C}^t\boldsymbol{C})\right)^{-1} \\
\widehat{\boldsymbol{B}} &= \left(\sum_i \boldsymbol{X}_{i..}\boldsymbol{C}\mathrm{diag}(\hat{\boldsymbol{a}}_i)\right)\left((\boldsymbol{C}^t\boldsymbol{C})\circ(\widehat{\boldsymbol{A}}^t\widehat{\boldsymbol{A}})\right)^{-1} \\
\widehat{\boldsymbol{C}} &= \left(\sum_j \boldsymbol{X}_{.j.}\widehat{\boldsymbol{A}}\mathrm{diag}(\hat{\boldsymbol{b}}_j)\right)\left((\widehat{\boldsymbol{A}}^t\widehat{\boldsymbol{A}})\circ(\widehat{\boldsymbol{B}}^t\widehat{\boldsymbol{B}})\right)^{-1},
\end{aligned} \tag{6}$$

where $\circ$ denotes the direct (or element-wise) product. The ALS algorithm iteratively calculates OLS estimates for the three factor matrices. Directly fitting these so as to minimise the sum-of-squares error provides a simple way of jointly estimating the factor loadings that describe processes in the temporal, spatial and subject domain without requiring orthogonality between factor loadings in any one of the domains: the multi-way PARAFAC model, unlike PCA, does not suffer from rotational indeterminacy, i.e. a rotation of estimated factors has impact on the overall fit [11, 12]. The ALS algorithm , however, can suffer from slow convergence, in particular, when a set of column vectors in one of the factor matrices is (close to being) collinear. Also, it is sensitive to specifying the correct number of factors R (i.e. the number of columns in $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$). In order to address these issues, [7] have proposed to extend the standard PARAFAC loss function to include a diagonalisation error, such that

$$\begin{aligned}
\mathcal{L}(\boldsymbol{A}) &= \sum_k \left|\left|\boldsymbol{X}_{..k} - \boldsymbol{A}\mathrm{diag}(\boldsymbol{c}_k)\boldsymbol{B}^t\right|\right|_F^2 \\
&\quad + \sum_i \left|\left|\boldsymbol{B}^\dagger\boldsymbol{X}_{i..}(\boldsymbol{C}^\dagger)^t - \mathrm{diag}(\boldsymbol{a}_i)\right|\right|_F^2,
\end{aligned} \tag{7}$$

(similarly for $\mathcal{L}(\boldsymbol{B})$ and $\mathcal{L}(\boldsymbol{C})$). Here, $||\boldsymbol{V}||_F^2 = \mathrm{tr}\,(\boldsymbol{V}^t\boldsymbol{V})$ denotes the Fröbenius norm and $\boldsymbol{B}^\dagger$ denotes the pseudo-inverse of $\boldsymbol{B}$. The first term corresponds to the sum-of-square loss function while the second term penalises the I different $R \times R$ projection matrices. A modified ALS algorithm can be derived by iterating solutions for

$$\frac{\partial\mathcal{L}(\boldsymbol{V})}{\partial\boldsymbol{V}} = 0$$

with $\boldsymbol{V} = \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. The ordinary least-squares solutions then becomes [7]:

$$\widehat{\boldsymbol{A}} = \left(\sum_k \boldsymbol{X}_{..k}\boldsymbol{B}\mathrm{diag}(\boldsymbol{c}_k) + \boldsymbol{P}\right)\left(\boldsymbol{I} + (\boldsymbol{B}^t\boldsymbol{B})\circ(\boldsymbol{C}^t\boldsymbol{C})\right)^{-1},$$

where $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_I]^t$ and where $\boldsymbol{p}_i$ are column vectors formed by the elements on the main diagonal of the $R \times R$ matrix $\boldsymbol{B}^\dagger\boldsymbol{X}_{i..}(\boldsymbol{C}^\dagger)^t$ (similar for $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{C}}$). This modified ALS algorithm has been used for all later PARAFAC calculation.

It is interesting to note that the ALS approach to three-way PARAFAC does provide a unique decomposition, provided the data has appropriate 'system variation' [11, 12], i.e. when $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are of full rank and there are proportional changes in the relative contribution from one factor to another in all three domains so that no two factors in any domain are collinear. In FMRI, however, we might expect the individual vectors in subject space to exhibit a significant amount of collinearity between some of them, e.g. in the case of two spatially different physiological signals, we might expect the relative contribution of the individual subjects to be very similar, so that two columns in $\boldsymbol{C}$ are (close to being) collinear. The effects of collinearity of some of the factors on the ability to extract the latent structure of the data will be evaluated in section 4.

3 Tensor PICA

The set of equations 3–5 can alternatively be expressed as simple matrix products, e.g. the set of equations 5 can be expressed as:

$$\boldsymbol{X}_{IK\times J} = (\boldsymbol{C} \,|\otimes|\, \boldsymbol{A})\boldsymbol{B}^t + \widetilde{\boldsymbol{E}}. \tag{8}$$

Here, $\boldsymbol{X}_{IK\times J}$ denotes the $IK \times J$ data matrix formed by concatenating all K different data sets in the temporal domain and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ are matrices containing the R individual temporal, spatial and session/subject factors in their columns. The first factor $(\boldsymbol{C} \,|\otimes|\, \boldsymbol{A})$ denotes the *Khatri-Rao* product of $\boldsymbol{A}$ and $\boldsymbol{C}$, i.e. a $IK \times R$ matrix formed by K copies of $\boldsymbol{A}$ stacked and column-wise scaled by diagonal matrices formed from rows of $\boldsymbol{C}$ such that $(\boldsymbol{C} \,|\otimes|\, \boldsymbol{A}) = ((\boldsymbol{A}\text{diag}(\boldsymbol{c}_1))^t, \ldots, (\boldsymbol{A}\text{diag}(\boldsymbol{c}_K))^t)^t$ [5].

From equation 8, the matrix of spatial factors, $\boldsymbol{B}$, has least-squares estimates of

$$\widehat{\boldsymbol{B}}^t = (\boldsymbol{C} \,|\otimes|\, \boldsymbol{A})^\dagger \boldsymbol{X}_{IK\times J}. \tag{9}$$

Similarly, $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{C}}$ can be estimated from equations 3 and 4. Thus, ALS estimates for each of the three matrices can be calculated via a linear projection of the three-way data, reshaped to 3 different 2-D matrices.

The PARAFAC model and the ALS algorithm for estimation treat all three domains equally and do not utilise any domain-specific information. Section 4 demonstrates how this can lead to PARAFAC results which are difficult to interpret, mainly due to significant cross-talk between estimated spatial maps.

In order to address this, we formulate a tensor-PICA model which incorporates the assumption of maximally non-Gaussian distributions of estimated spatial maps, $\boldsymbol{B}$: equation 8 is identical to a standard (2-D) factor analysis or noisy ICA model [2], where the matrix $(\boldsymbol{C} \,|\otimes|\, \boldsymbol{A})$ denotes the 'mixing' matrix and $\boldsymbol{B}^t$ contains the set of spatial maps as row vectors. Unlike the single subject (2-D) PICA model, however, the mixing matrix now has a special block structure which can be used to identify the factor matrices $\boldsymbol{A}$ and $\boldsymbol{C}$. Given the first matrix factor in equation 8, it is easy to recover the two underlying matrices $\boldsymbol{A}$ and $\boldsymbol{C}$: each of the R columns in $(\boldsymbol{C} \,|\otimes|\, \boldsymbol{A})$ is formed by K scaled repetitions of a single column from $\boldsymbol{A}$, i.e. when reshaped into a $I \times K$ matrix is of rank 1. Thus, we can transform each column r into a $I \times K$ matrix and calculate its (single) non-zero left Eigenvector of length I, together with a set of K factor loadings (projections of the matrix onto the left Eigenvector), using a Singular Value Decomposition (SVD) and use these to re-constitute a column of the underlying factor matrices $\boldsymbol{A}$ and $\boldsymbol{C}$. This needs to be repeated for each of the R columns separately and the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are proportional to the R different Eigenvectors and factor loadings respectively, i.e. the values obtained by projecting the matrix of $I \times K$ matrix of time courses onto the Eigenvector of the SVD.

This gives the following algorithm for a rank-1 tensor PICA decomposition of three-way data $\boldsymbol{X}$:

i) perform an iteration step for the decomposition of the full data

$$\boldsymbol{X}_{IK\times J} = \boldsymbol{M}^c \boldsymbol{B}^t + \widetilde{\boldsymbol{E}}_1 \tag{10}$$

using the 2-D PICA approach for the decomposition into a compound mixing matrix $\boldsymbol{M}^c$ of dimension $IK \times R$ and a $R \times J$ matrix containing the associated spatial maps $\boldsymbol{B}^t$.

ii) decompose the estimated mixing matrix $\boldsymbol{M}^c$ such that

$$\boldsymbol{M}^c = (\boldsymbol{C} \,|\otimes|\, \boldsymbol{A}) + \widetilde{\boldsymbol{E}}_2 \tag{11}$$

via a column-wise rank-1 Eigenvalue decomposition: for each column r of $\boldsymbol{M}^c$, form the $I \times K$ matrix $\widetilde{\boldsymbol{M}}_r = (\boldsymbol{M}^c_{r,1}, \ldots, \boldsymbol{M}^c_{r,K})$. Under the model, this matrix contains K scaled repetitions of a single temporal factor $\boldsymbol{A}_r$ which can be found by calculating its least-squares rank-1 approximation using SVD. Along with the temporal factor $\boldsymbol{A}_r$ (left Eigenvectors), the SVD provides the individual scalings (right Eigenvectors) which define the corresponding vector in $\boldsymbol{C}_r$. This needs to be repeated for each column of $\boldsymbol{M}^c$ and provides estimates for $\boldsymbol{A}$ and $\boldsymbol{C}$.

iii) iterate decomposition of $\boldsymbol{X}_{IK\times J}$ and $\boldsymbol{M}^c$ until convergence[5].

Note that, like PARAFAC, the rank-1 tensor PICA decomposition estimates factor matrices for the generative model of equation 1. The estimated matrices, however, provide a different structural representation of the three-way data $\boldsymbol{X}$. Note also, that the singular value decomposition of each matrix $\widetilde{\boldsymbol{M}}_r$ not only provides the left and right Eigenvectors which form the relevant columns in $\boldsymbol{A}_r$ and $\boldsymbol{C}_r$ but also gives a set of Eigenvalues. The ratio of the largest Eigenvalue and the sum of all Eigenvalues can be used to assess the quality of the rank-1 approximation: if the matrix $\widetilde{\boldsymbol{M}}_r$ is not well approximated by the outer product of the left and right Eigenvectors the corresponding ratio will be low, i.e. only represent a small amount of variability in $\widetilde{\boldsymbol{M}}_r$.

[5] e.g. when $||\boldsymbol{A}^{\text{new}} - \boldsymbol{A}^{\text{old}}||_F + ||\boldsymbol{B}^{\text{new}} - \boldsymbol{B}^{\text{old}}||_F + ||\boldsymbol{C}^{\text{new}} - \boldsymbol{C}^{\text{old}}||_F < \epsilon$.

3.1 Relation to mixed-effects GLMs

Within the general linear model, the data from an individual subject k at time i and voxel location j is expressed as:

$$x_{ijk} = \sum_{r}^{R} \bar{\boldsymbol{A}}_{ir}^{(k)} \bar{\beta}_{jr}^{(k)} + \bar{\epsilon}_{ij}^{(k)}. \tag{12}$$

Here, $\bar{\boldsymbol{A}}^{(k)}$ denotes the (potentially subject-specific) lower-level GLM design matrix containing R regressors, $\bar{\beta}^{(k)}$ denotes the subject specific lower-level linear model parameters at voxel location j and $\bar{\epsilon}_{ij}^{(k)}$ is the subject-specific (fixed-effects) error. Typically, the subject specific linear model parameters are then related to group parameters in a second linear model[6]:

$$\bar{\boldsymbol{\beta}}_{jr}^{(k)} = \bar{\boldsymbol{C}}_{kr} \bar{\boldsymbol{B}}_{jr} + \bar{\eta}_{jr}, \tag{13}$$

where now $\bar{\boldsymbol{C}}_{kr}$ denotes the k^{th} entry of the group-level design matrix for lower-level regressor r (which, in the case of mean group activation studies, is a vector of ones for each r), $\bar{\boldsymbol{B}}_{jr}$ is the group effect size at voxel location j and where $\bar{\eta}$ is the random-effects variance contribution of the lower-level regressor r at voxel location j. Within this model, the estimated random-effects variance contribution is given by the variance of the estimated parameter estimates around their mean. The mixed-effects two-level GLM relates the group level parameters of interest to the original data as

$$x_{ijk} = \sum_{r}^{R} \bar{\boldsymbol{A}}_{ir}^{(k)} \bar{\boldsymbol{C}}_{kr} \bar{\boldsymbol{B}}_{jr} + \bar{\nu}_{ijk}, \tag{14}$$

where $\bar{\nu}$ is the combined error term with associated mixed-effects variance. Similar to the generative model of equation 1, the two-level mixed-effects GLM expresses the data via a tri-linear decomposition.

In order to compare the variance terms, assume that $\bar{\boldsymbol{A}}^{(k)} = \boldsymbol{A}, \bar{\boldsymbol{B}} = \boldsymbol{B}$ and $\bar{\boldsymbol{C}} = \boldsymbol{C}$. Then the mixed-effects error $\bar{\nu}$ corresponds to the error term $\widetilde{E}$ in equation 8. In contrast to the PARAFAC/tensor ICA model, however, equation 14 uses *a-priori* specified design matrices $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{C}}$. When fixing the GLM group-level design $\bar{\boldsymbol{C}}$ to a column vector of ones in order to calculate mean group activation size, the resulting group-level error, $\bar{\eta}$, has associated voxel-wise random-effects variance. That is, the multi-level GLM permits different voxel locations to have different random-effects variance. Even in the case where two voxels show similarly significant amplitude modulation to the same regressor, r, their random-effects variance contribution is allowed to be different. As such, the multi-level GLM random-effects variance is the variance of the individual subjects' responses around the expected population mean response *at a given voxel location*, i.e. the random-effects variance is averaged over time but not over space. By comparison, $\boldsymbol{C}_r$ in equation 8 represents the amplitude of signal modulation of the entire spatio-temporal process defined by $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$ for different processes, r, and subjects k, independent of voxel location j. As such, the variance of a single vector $\boldsymbol{C}_r$ signifies the variance of the individual subjects' responses around the expected population mean response for the entire spatio-temporal process described by $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$. Within the standard multi-level GLM this quantity is not readily available but can be approximated by the spatially averaged random-effects variance weighted by the normalised voxel-wise parameter estimates within post-thresholded clusters [25].

3.2 Data pre-processing for tensor-PICA

As in the 2-D case, the data will be voxel-wise de-trended (using Gaussian-weighted least squares straight line fitting; [19]) and de-meaned separately for each data set k before the tensor-PICA decomposition. In order to compare voxel locations between subjects/sessions, the individual data sets need to be co-registered into a common space, typically defined by a high-resolution template image. We do not, however, necessarily need to re-sample to the higher resolution and can keep the data at the lower EPI resolution in order to reduce computational load. After transformation into a common space, the data is temporally normalised by the estimated voxel-wise noise covariances $\boldsymbol{V}_k^{-1/2} = \mathrm{diag}(\sigma_{1,k}, \ldots, \sigma_{J,k})$ using the iterative approximation of the noise covariance matrix from a standard 2-PICA decomposition. This will normalise the voxel-wise variance both within a set of voxels from a single subject/session and between subjects/sessions. The voxel-wise noise variances need to be estimated from the residuals of an initial PPCA decomposition. This, however, cannot simply be done by calculating the individual data covariance matrices $\boldsymbol{R}_{..k} \propto \boldsymbol{X}_{..k}\boldsymbol{X}_{..k}^{t}$.

[6] In its general form the multi-level GLM allows to combine *contrasts* of lower-level parameter estimates. Such designs, however, can be re-formulated as higher-level GLMs operating on simple parameter estimates; see [1] for details.

Within the tensor-PICA framework, the temporal modes (contained in $\boldsymbol{A}$) are assumed to describe the temporal characteristics of a given process r for all data sets k. We will therefore estimate the initial temporal Eigenbasis from $\boldsymbol{R} = \frac{1}{K}\sum_k \boldsymbol{R}_{..k}$, i.e. by the mean data covariance matrix. This corresponds to a PPCA analysis of $\boldsymbol{X}_{I\times JK}$, i.e. the original data reshaped into a (# time point) by (# voxels) $\times$ (# subjects/sessions) matrix. We use the Laplace approximation to the model order [22] to infer on the number of source processes, R. The projection of the data sets onto the matrix $\boldsymbol{U}_R$ (formed by the first R common Eigenvectors of $\boldsymbol{R}$) reduces the dimensionality of the data in the temporal domain in order to avoid overfitting. This projection is identical for all K different data sets, given that $\boldsymbol{U}_R$ is estimated from the mean sample covariance matrix $\boldsymbol{R}$. Therefore, we can recover the original time-courses by projection onto $\boldsymbol{U}_R$: when the original data $\boldsymbol{X}_{IK\times J}$ is transformed into a new set of data $\widetilde{\boldsymbol{X}}_{RK\times J}$ by projecting each $\boldsymbol{X}_{..k}$ onto $\boldsymbol{U}_R$, the original data can be recovered from

$$\boldsymbol{X}_{IK\times J} = (\mathbf{I}_R|\otimes|\boldsymbol{U}_R)\widetilde{\boldsymbol{X}}_{RK\times J},$$

where $\mathbf{I}_R$ denotes the identity matrix of rank R. If the new data $\widetilde{\boldsymbol{X}}$ is decomposed such that $\widetilde{\boldsymbol{X}}_{RK\times J} = \left(\boldsymbol{C}|\otimes|\widetilde{\boldsymbol{A}}\right)\boldsymbol{B}^t + \widetilde{E}$, then $\boldsymbol{X}_{IK\times J} = (\boldsymbol{C}|\otimes|\boldsymbol{A})\,\boldsymbol{B}^t + \widetilde{E}$, where $\boldsymbol{A} = \boldsymbol{U}_R\widetilde{\boldsymbol{A}}$. This approach is different from e.g. [6], where an individual data set k is projected onto a set of Eigenvectors of the data covariance matrix $\boldsymbol{R}_{..k}$. As a consequence, each data set in [6] has a different signal+noise subspace compared to the other data sets.

Similar to the 2-D PICA model, the set of pre-processing steps is iterated in order to obtain estimates for the voxel-wise noise variance $\boldsymbol{V}$, the PPCA Eigenbasis $\boldsymbol{U}$ and the model order R before decomposing the reduced data $\widetilde{\boldsymbol{X}}_{RK\times J}$ into the factor matrices $\widetilde{\boldsymbol{A}}, \boldsymbol{B}$ and $\boldsymbol{C}$ (see [2] for details).

3.3 Group-level inference

In equation 9, the estimated spatial maps are given by the projection of the original data $\boldsymbol{X}$ transformed into its 2-D representation $\boldsymbol{X}_{IK\times J}$ and projected onto the estimated 'unmixing' matrix $(\boldsymbol{C}|\otimes|\boldsymbol{A})^\dagger$. To generate statistic values we transform spatial maps $\boldsymbol{B}$ into voxel-wise Z-scores by dividing the estimated spatial maps by the residual mixed-effects variance and model the histogram of Z-statistics values using the Gaussian/Gamma mixture model approach [3, 29]. The fitted mixture model can then be used to threshold spatial maps using the voxel-wise posterior probability of 'activation' or the expected false positive rate over the brain or over the voxels classified as 'non-background noise' (FDR).

4 Experimental Methods

We illustrate the approach of rank-1-tensor-PICA and compare it with standard PARAFAC on a set of artificial and on real FMRI group data.

4.1 Simulated data

We acquired whole brain volumes ($64 \times 64 \times 21, 4 \times 4 \times 6$mm) of FMRI data on a Varian 3T system (TR=3sec; TE=30ms) under resting condition. The data were corrected for subject motion using MCFLIRT [14], temporally high-pass filtered (Gaussian-weighted least-squares straight line fitting, with sigma=20.0s) [19] and masked for non-brain voxels using BET [24]. The pre-processed data was used to estimate background noise parameters (voxel-wise means and std. deviations) which were used to generate 3 artificial data sets with Gaussian noise characteristics. Artificial signal was linearly added to the Gaussian background noise data using spatial maps and time courses depicted in figure 1. The time courses correspond to the stimulus trains from a simple block design, a single-event (fixed inter-stimulus interval) design and a single-event (random inter-stimulus interval) convolved with a canonical hæmodynamic response function (Gamma variate with 3s standard-deviation and 6s lag).

Five different data sets (196 time points $\times$ 2800 voxels $\times$ 3 subjects each) were generated as example FMRI studies with different signal characteristics:

(A) Each subject's data $\boldsymbol{X}_{..k}$ contains all three spatial maps shown in figure 1. Each spatial map has a different associated time course: time course 1 modulates spatial map 1, time course 2 modulates spatial map 2 and time course 3 modulates spatial map 3. This defines three spatio-temporal processes which are introduced at different strengths into the individual subjects' data. The 'activation'levels were set to (3,4,5),(2,3,4) and (2,2,3) times the mean noise standard deviation for subjects 1-3. The complete 3-way data conforms to the generative model of equation 1 with $R = 3$ source processes in the data.

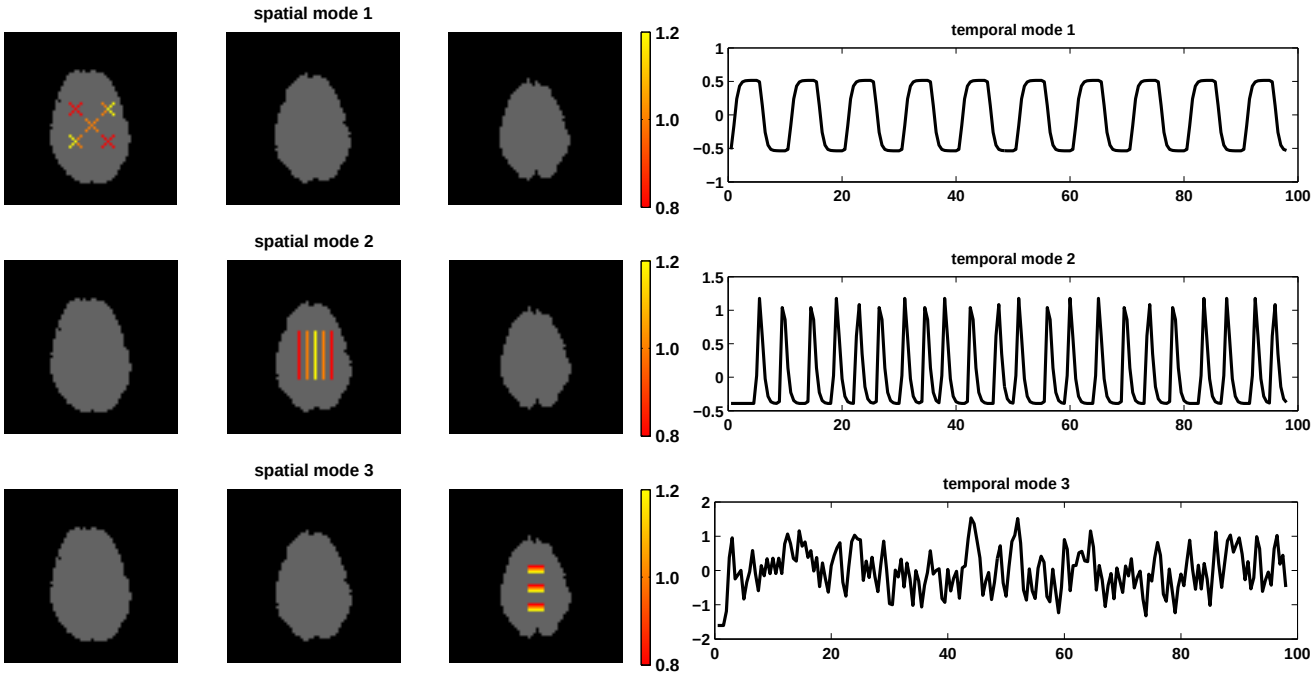


Figure 1: Artificial spatial maps and time-courses used for the generation of artificial group data.

(B) Each subject contains spatial map 1 modulated by time-course 1. In addition, subject 2 contains spatial map 2 modulated by time course 2, while subject 3 contains spatial map 3 modulated by time course 3. This data set is a special case of data set (A) with strength set to (3,0,0),(2,3,0) and (2,0,3). The data still conforms to the generative model of equation 1 and is used to demonstrate the performance of PARAFAC and tensor-PICA on data where the matric $\mathbf{C}$ is sparse, i.e. for data which contains subject-specific source processes in addition to a common source process.

(C) Like data set (A), but with the individual convolution parameters for the generation of the signal time-courses differing between subjects in mean lag and standard deviation used for the Gamma HRF ($\sigma = 3, 3.5$ and 4 seconds, mean lag of $4, 5$ and 6 seconds). This induces small differences in the temporal signal characteristics between subjects. This data set is used to test for robustness against small deviations from the model assumptions in the temporal domain (e.g. small differences between subjects in the BOLD response to the same set of stimuli). Note that this data set still conforms to the tri-linear model, as these different time courses together with the spatial maps can be interpreted as separate source processes (i.e. with $\boldsymbol{A}$ containing 9 time courses with sets of 3 time courses being close to collinear and with $\boldsymbol{B}$ containing 9 spatial maps where sets of 3 are identical). The data does not, however, conform to the tensor-PICA model, as the spatial maps are not statistically independent.

(D) Subject 1 does not contain any 'activation' signal. Subjects 2 and 3 contain 'activation' signal in the area defined by spatial map 2, modulated by the simple block-design (time course 1). Subject 3 also contains extra 'activation' signal in the area defined by spatial map 3, modulated again by time course 1. In addition, all three subjects contain 'nuisance' signals (spatial map 1 modulated by a different time course in each subject). This data simulates cases where FMRI data is confounded by e.g. resting-state networks which are spatially consistent but differ in the temporal characteristics of the resting-state BOLD signal. The data conforms to the tri-linear model when viewed as a set of 5 spatial maps with 5 associated time courses.

(E) Each data set contains all three spatial maps, but modulated by a different time course, i.e. the association between the spatial maps 1-3 and time courses 1-3 changes between subjects. The data conforms to the tri-linear model when viewed as a set of 9 spatial maps and 9 associated time courses. However, like data set (C) some of the spatial maps are identical and thus not statistically independent.

4.2 Multi-Session FMRI data

The data were originally used in [20] to study session variability in repeated FMRI experiments: a healthy 23-year old right handed male was scanned under a visual, cognitive and motor paradigm in 33 separate sessions over a period of 2 months. The data used here consists of the first 10 sessions under the motor paradigm: a block design with 24.6s on/off periods and right index finger tapping at 1.5 Hz during the 'active' condition. Data (78 volumes) was collected on a Siemens Vision (2T) with TR=4.1s, 48*64*64 (3mm isotropic voxels). In addition, a single T1 weighted structural image was taken at 1x1x1.5mm resolution.

4.2.1 Data pre-processing

The data were individually corrected for head-motion using MCFLIRT. Mean-based intensity normalisation of all volumes by the same factor was applied (i.e. grand-mean scaling so that each of the 10 sessions had the same mean intensity value when averaged over 78 volumes and all brain voxels), followed by high-pass temporal filtering (see above) was performed. The individual data sets were registered into the space of the high resolution T1 image using FLIRT [15]. In order to decrease computational load, the T1 high resolution image was segmented into different tissue types using FMRIB's Automated Segmentation Tool (FAST) [30]. This provided maximum a-posteriori estimates for voxel-wise grey matter probability. Voxels with $p > 0.2$ (N=47168; $\sim 18\%$ of all intra-cranial voxels) were included in the tensor analysis so that $\boldsymbol{X}$ is a 3-way array of dimension $78 \times 47168 \times 10$. Based on the estimated sample covariance matrix of the 78×471680 matrix $\boldsymbol{X}_{I \times JK}$, the Laplace approximation to the model estimated a 19-dimensional signal sub-space. The data for each session was projected onto the space spanned by the first 19 Eigenvectors and spatially normalised by the voxel-wise variance estimate from the residuals of the projection.

4.3 Multi-Subject FMRI data

Five healthy right-handed subjects perfomed 30s blocks of a visually cued reaction time task involving left index finger movement, left hand sequential finger movement and left hand random finger movement (see [16] for details).

For each subject, 122 axial echo-planar volumes (21x6mm slices, TE=30ms, TR=3s, 64*64*21 voxels at 4x4x6mm) were acquired on a 3T Varian/Siemens MRI system at the Oxford Centre for Functional Magnetic Imaging of the Brain together with a T1-weighted anatomical image at 1x1x1.5mm resolution.

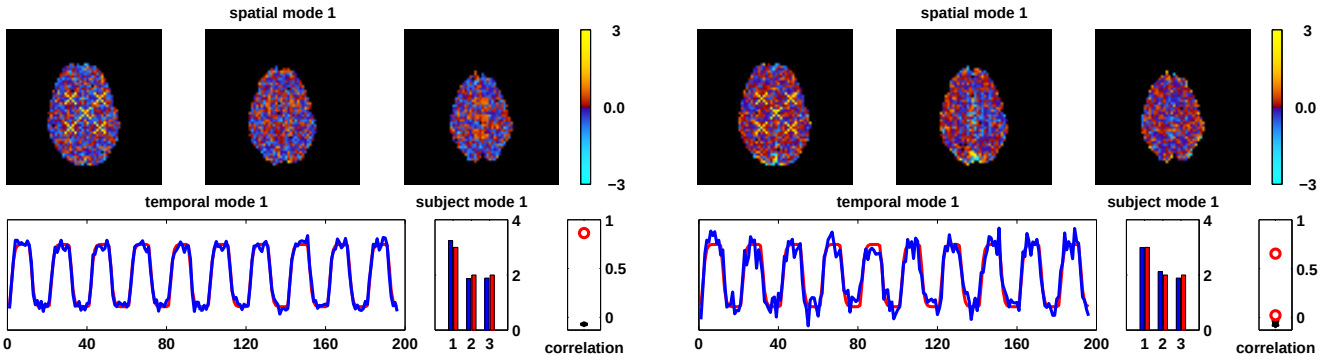
4.3.1 Data pre-processing

The data were individually corrected for head-motion using MCFLIRT and spatially smoothed using a Gaussian kernel of FWHM 5mm. Mean-based intensity normalisation of all volumes by the same factor was applied, followed by high-pass temporal filtering (see above) was performed. The individual data sets were registered into MNI space using FLIRT [15] while keeping the data at the functional resoultion in order to decrease computational load. The final 3-way data $\boldsymbol{X}$ was of size $122 \times 12839 \times 5$. Based on the estimated sample covariance matrix of the matrix $\boldsymbol{X}_{I \times JK}$, the Laplace approximation to the model estimated a 12-dimensional signal sub-space. The data for each session was projected onto the space spanned by the first 12 Eigenvectors and spatially normalised by the voxel-wise variance estimate from the residuals of the projection.

5 Results

5.1 Simulated data

In this section we compare the results from PARAFAC and tensor-PICA on the different artificial data sets (A)-(E). For each method, we show unthresholded spatial maps, together with both estimated time-courses and the 'true' (data sets (A) and (B)) or best rank-1 approximation to the set of true (data sets (C) - (E)) time-courses. The accuracy in the subject domain is shown via bar plots with both the estimated (left bars) and true (right bars) weights for the three subjects. In each plot, we show the results of the estimated source with highest spatial correlation with the true sources. The boxplots signify how strong the correlation is relative to the correlation between all other processes and the true spatial maps. Multiple large correlation coefficients signify substantial cross-talk between estimated spatial maps.



(i) tensor-PICA (ii) PARAFAC

Figure 2: Tensor-PICA and PARAFAC decomposition results for data set (A) and the first spatial map. Both for tensor-PICA and PARAFAC, $R = 13$ maps were estimated, based on the Laplace approximation to the Bayesian model order. The spatial maps are normalised to unit standard deviation. The estimated time courses (blue) are shown together with the 'true' signal time courses (red), both are scaled to mean 0 and unit standard deviation. The barplots show the accuracy of the estimation in the subject domain for each of the three subjects (blue:estimated, red:true), while the boxplot shows the correlation of the 'true' spatial map with each of the R estimated maps as an indicator of cross-talk between estimated maps of interest.

5.1.1 Data set (A)

Data set (A) conforms to the assumptions of the model in equation 1 and figure 2 shows the estimate of the first spatio-temporal process (spatial map 1 and time course 1 at different 'strengths' 3,2,2 for the three subjects) for tensor-PICA and PARAFAC where $R = 13$ was used based on the Laplace approximation to the model order. Within the spatial, temporal and subject domain, both techniques identify the artificial signals well. For both techniques the boxplots clearly shows only a single process having high spatial correlation with the 'true' spatial map 1. In the case of PARAFAC, however, the maximum correlation is reduced, possibly an effect of suboptimal convergence. The estimation of the PARAFAC solution has used almost 15 times the number of floating point operations compared to tensor-PICA. Estimates for the two other source processes are qualitatively similar to what is shown in figure 2.

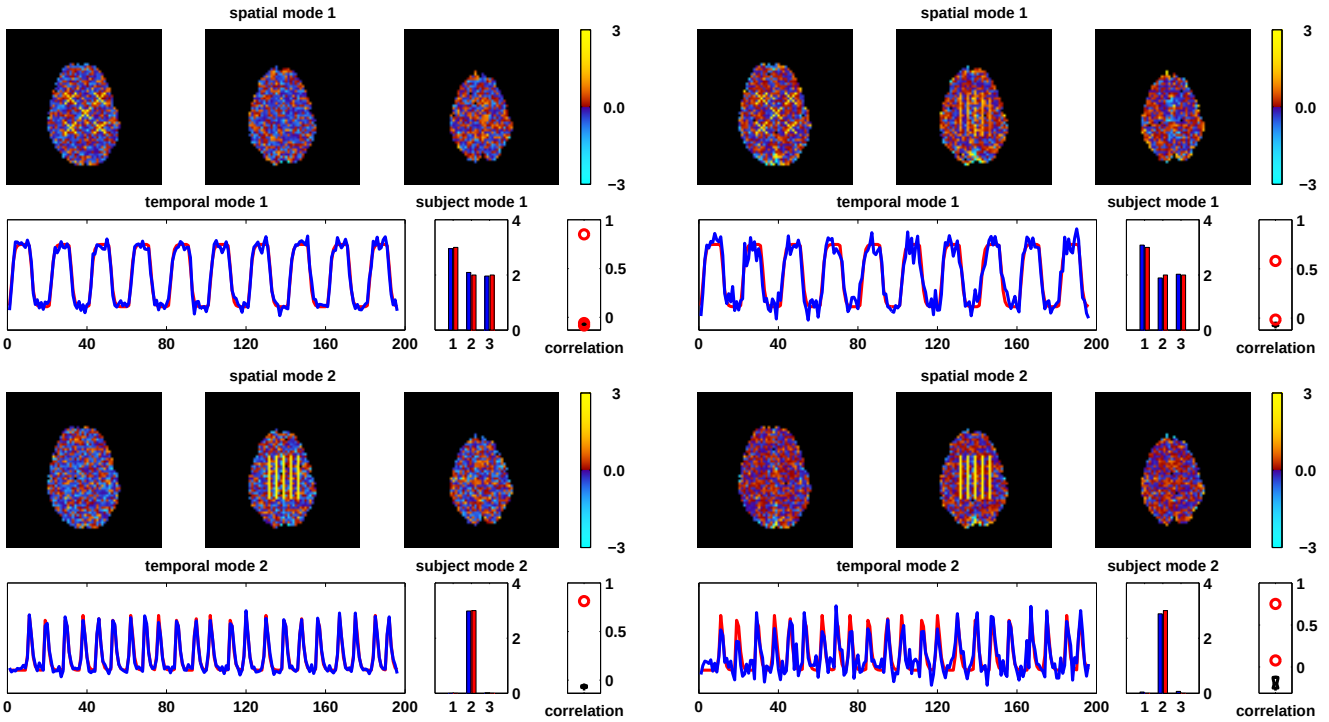
5.1.2 Data set (B)

Data set (B) differs from data set (A) only by having some processes which are subject specific, i.e. they have zero amplitude modulation in some of the subjects' data. Figure 3 shows the results for the 'common' source process (top) and one of the secondary source process (bottom) which only exists in subject 2. The model order was estimated to be $R = 13$. Both techniques still estimate the source processes, with the PARAFAC estimate of the common source process showing some cross-talk with spatial map 2. The PARAFAC estimation in this case involved a 23-fold increase in number of floating point operations.

5.1.3 Data set (C)

In data set (C), different parameters for the HRF convolutions resulted in small differences in the time courses. As a consequence, each of the three spatial maps has a slightly different associated time-course in each subject. As such, there is no longer a single 'true' associated time course across subjects. The induced variation in the temporal domain, however, is small (time courses have $r > 0.75$ temporal correlation) and the different time courses are well approximated by their dominant Eigenvector, i.e. by the best rank-1 approximation. The data permits two different representations: firstly, the signal content can either be approximated as a linear combination of 3 processes (where in the temporal domain the rank-1 approximation to the 3 slightly different time courses is used), or can fully be expressed as a linear combination of 9 processes with large co-linearity in $\boldsymbol{A}$ and 3 multiple versions for each of the true spatial sources in $\boldsymbol{B}$.

Figure 4 shows the estimated set of source processes for tensor-PICA and PARAFAC. The tensor-PICA decomposition, due to the independence assumption in the spatial domain, represents the data via a set of 3 source



(i) tensor-PICA (ii) PARAFAC

Figure 3: Tensor-PICA and PARAFAC decomposition results for data set (B) with $R = 13$: the first process is contained in all three subjects while the second process only appears in subject 2.

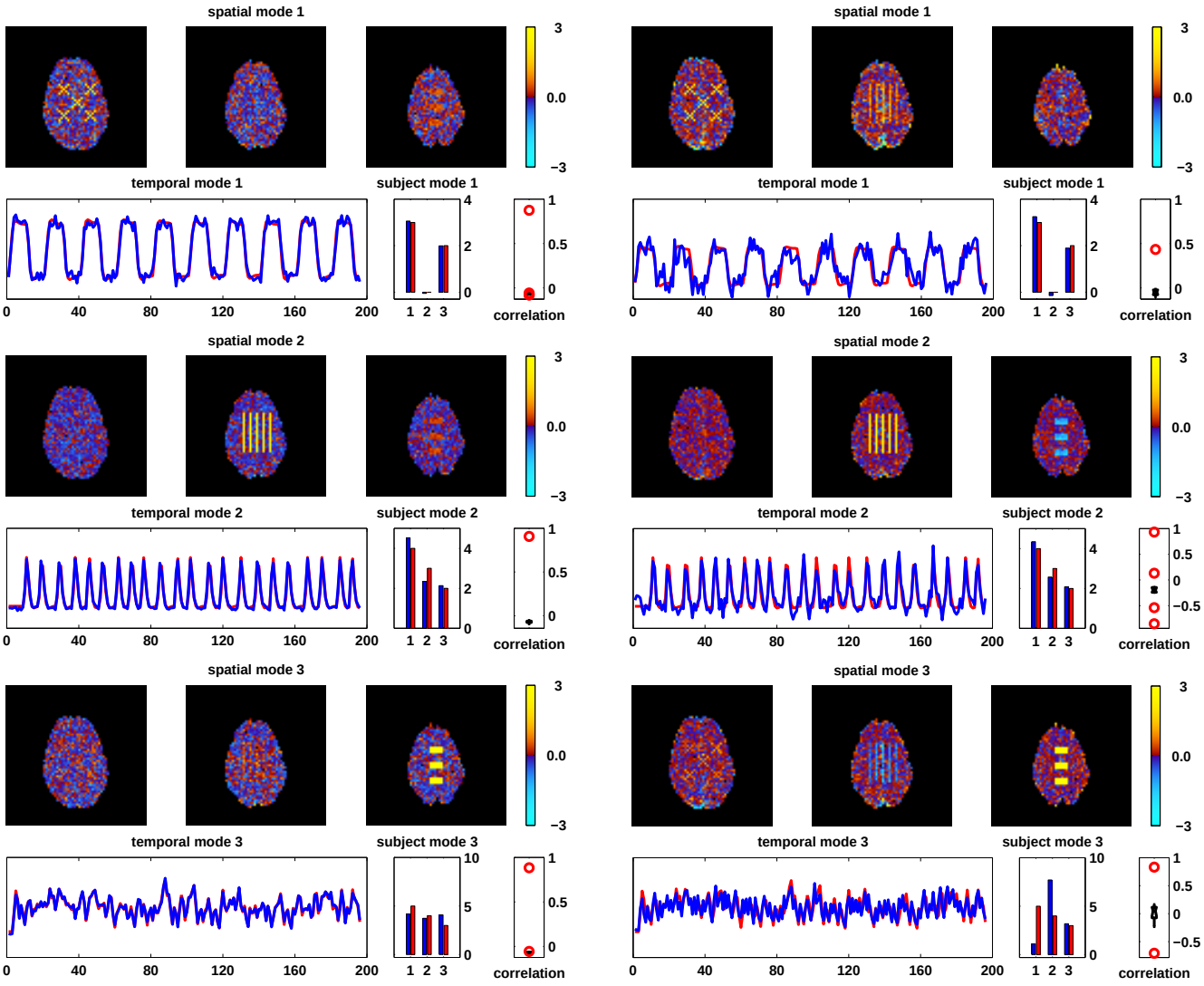
processes. Compared to the tensor-PICA results, the PARAFAC spatial estimates exhibit some cross-talk, e.g. the first spatial map is visibly confounded by map 2 and map 3. Also, in the temporal- and subject domains, PARAFAC finds less accurate estimates of the true source processes. The two approaches differ most significantly in the way in which true spatial maps correlate with each of the $R = 14$ estimated maps: while the spatial tensor-PICA decomposition always results in only one source which correlates strongly with the true spatial map, the PARAFAC decomposition shows that, especially for sources 2 and 3, multiple PARAFAC estimates correlate with the true maps. As such, PARAFAC does not represent the signal of interest via 3 different source processes but equally does not find the representation by 9 sources: almost all of the estimated correlated maps show significant amount of cross-talk. Convergence in this case is particularly slow, with 47 times the number of floating point operations compared to tensor-PICA.

5.1.4 Data set (D)

This data set simulates a scenario where subjects differ in spatial extent of signal: subject 1 does not contain any 'activation' signal, while subjects 2 contains 'activation' signal in spatial map 2 and subject 3 contains 'activation' signal in spatial maps 2 and 3. For both subjects 2 and 3, signal is temporally modulated by time course 1. All three subjects contain 'nuisance' signal in spatial area 1, but modulated each time by a different time course (time course $2, 3$ and $2+3$). This simulates spatially consistent but temporally inconsistent effects like resting state networks [4]

Estimated sources are shown in figure 5. The PARAFAC decomposition no longer reflects the spatial or temporal extent of signal well. Similar to data sets (A)-(C), the tensor-PICA decomposition identifies 3 source processes which are strongly correlated with the true spatial maps 1-3. The estimated spatio-temporal decomposition closely matches the way that data was generated. In the case of the main nuisance effect (contained inside spatial map 1), the tensor-PICA decomposition approximates the best rank-1 decomposition of the different time-courses involved as the time-course which best summarises the 3 different temporal signals associated with this single map.

Note that each estimated time course in a tensor-PICA decomposition is calculated from an SVD of a single column of $\boldsymbol{M}^c$, reshaped into a $I \times K$ matrix. This will not only provide the single time course which best represents



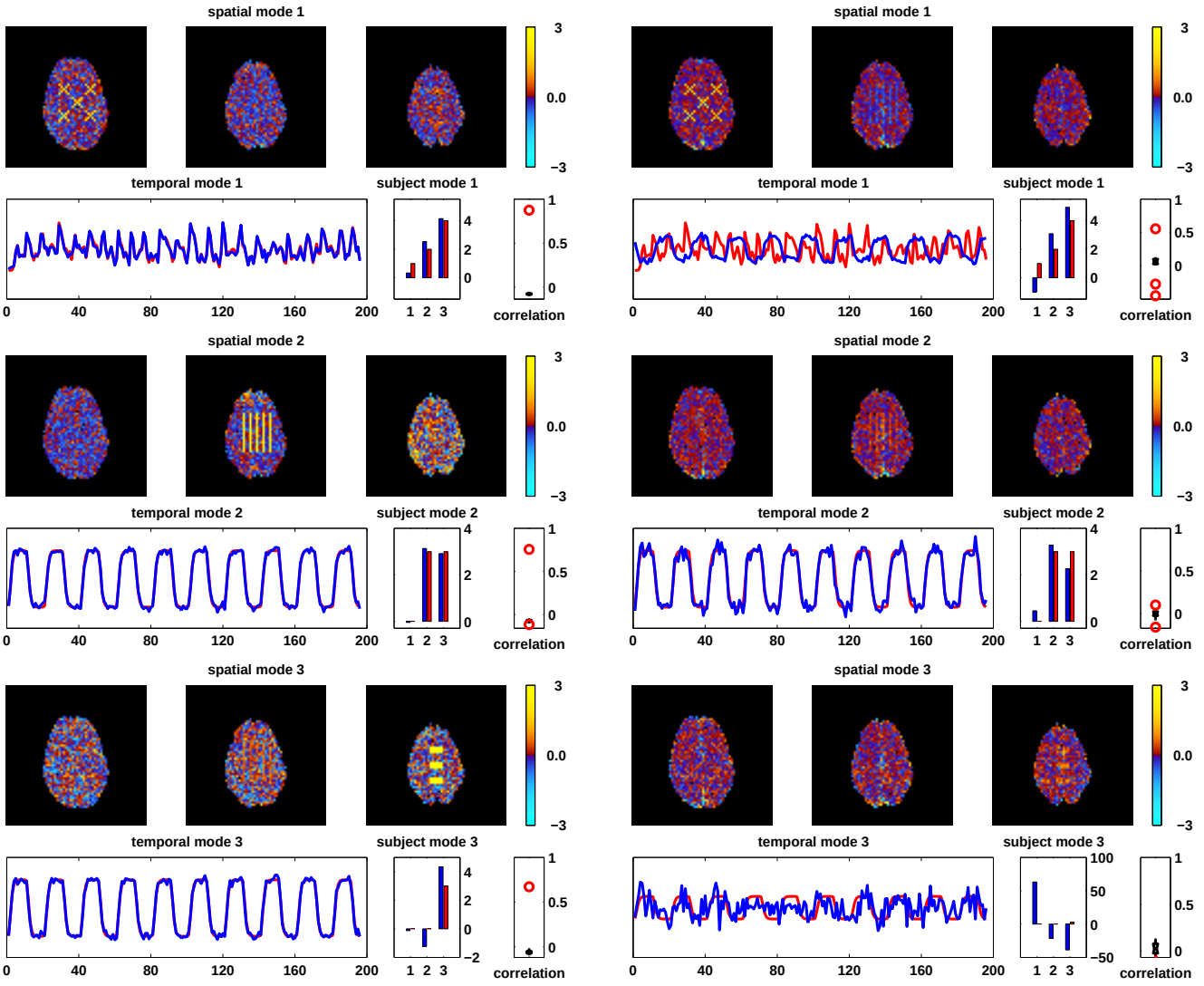
(i) tensor-PICA (ii) PARAFAC

Figure 4: Tensor-PICA and PARAFAC decomposition results for data set (C) with $R = 14$. Each of the 'true' spatial maps no longer has a single 'true' signal time course associated and estimated time courses (blue) are shown together with the dominant Eigenvector of all possible time courses (red).

the K individual time courses for each column in $\boldsymbol{M}^c$, but also provides information about the amount of variance that this individual time course explains. For this data, the time courses for estimated source processes 2 and 3 represent 99.27% and 99.57% of the total variance contained in the relevant columns of $\boldsymbol{M}^c$. By comparison, the time course for source process 1 only represents 54.07% of the variance in the temporal domain. This indicates that the rank-1 approximation of the time-courses associated with spatial map 1 is not very descriptive of the temporal characteristics in each of the subjects.

5.1.5 Data set (E)

In data set (E), all signals are temporally and spatially inconsistent between subjects. While each subject contains each of the spatial maps and time courses in figure 1, the association between true spatial maps and true time courses differs in each subject. Figure 6 shows the results for one of the source signals. Despite the fact that the signal content in this data set does conform to the tri-linear model, the high degree of collinearity in this representation prevents PARAFAC from clearly identifying the source process in the spatial or temporal domain. The tensor-PICA results contain exactly one source with high spatial correlation for each of the three 'true' spatial



(i) tensor-PICA (ii) PARAFAC

Figure 5: Tensor-PICA and PARAFAC decomposition results for data set (D) and all three spatial maps.

maps (only one of the maps is shown, the results for spatial map 1 and 2 are qualitatively similar). The estimated subject-factor differs significantly from the 'true' relative activation strengths. Also, the associated time course only explains 46.26% of the final amount of variance among the set of 3 time-courses, indicating that a rank-1 approximation might not be sufficient to capture the temporal dynamics.

5.1.6 Accuracy and dimensionality

The accuracy of estimation, both for PARAFAC and tensor-PICA, depends on the number of processes, R, estimated with each method. All results presented above have used a value of R as estimated via the Laplace approximation to the model order for the Eigenspectrum of the data covariance matrix $\boldsymbol{R}_{I\times JK}$. Figure 7 compares the accuracy of estimation for both PARAFAC (P) and tensor-PICA (T) on all 5 data sets (A)-(E) for different values of $R = 3, 10$ and 40 in the spatial, temporal and subject domain.

Circles denote the source process with highest absolute correlation with one of the three true spatial maps while dots show the correlation of the remaining sources. In almost all cases, the source process with highest spatial correlation also has largest temporal correlation with the associated true time course (or to the best rank-1 approximation)[7]. For data sets (A)-(C), i.e. when signals conform to the generative model of equation 1, the

[7] In the subject domain, the correlations with all other processes is not shown, as there are only 3 subjects in this study

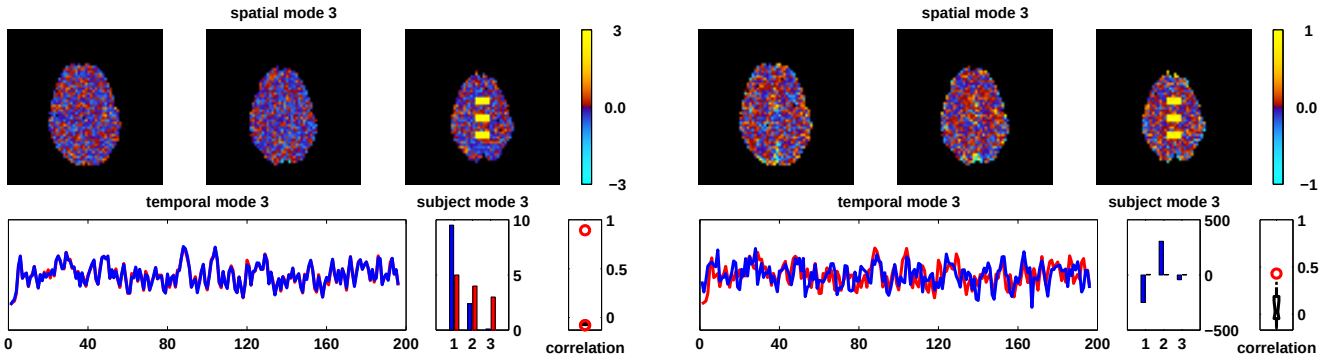


Figure 6: Tensor-PICA and PARAFAC decomposition results for data set (E) and estimated spatial map 3.

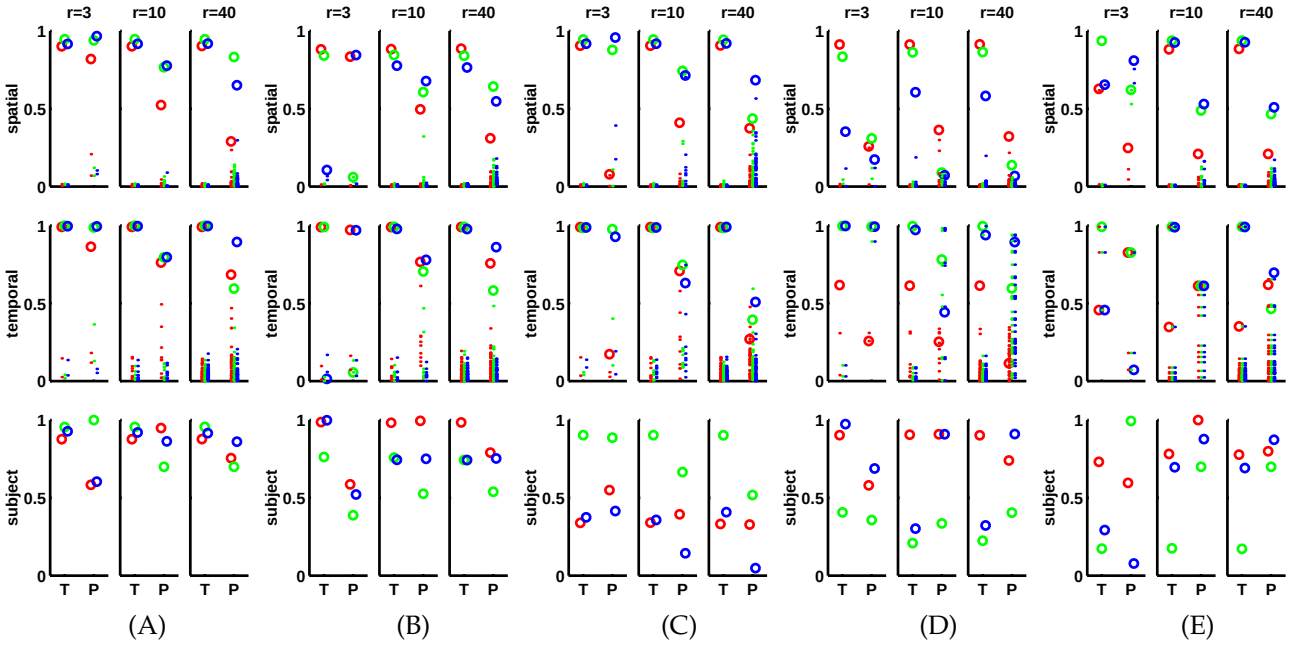


Figure 7: Accuracy of signal estimation for PARAFAC and tensor-PICA on the artificial FMRI group data (A)-(E). Plots show the correlation between 'true' (or best rank-1) modes and estimated modes in the spatial, temporal and subject domain (top to bottom rows respectively) for PARAFAC (P) and tensor-PICA (T). For each method and data set, the analysis was performed for $R = 3, 10$ and 40. Different colours show the estimation accuracy for the three spatial maps shown in figure 1.

correlations in the spatial and temporal domain between true sources and estimates from tensor-PICA are very high and always clearly identify a single process (i.e. for each 'true' spatial map, one of the estimated spatial maps has high spatial correlations while at the same time all other estimated spatial modes have low spatial correlation). Furthermore, the estimation is relatively robust when estimating a different number of sources. This is of prime importance, since the exact number of source processes is not known *a-priori* and the Laplace approximation is not expected to always give very accurate results (see [2] for a detailed discussion). The PARAFAC estimates, by comparison, exhibit a stronger dependence on the number of estimated sources R. As the number of estimated sources increases, a larger number of source processes show 'spurious' correlations with the true spatial maps. In the case of data sets (D) and (E), the PARAFAC results are significantly worse compared to the tensor-PICA results and do not identify the source processes in any domain. These simulations suggests that tensor-PICA is less sensitive to the model order as well as deviations of the signal content in the data from the generative three-way

model.

5.2 Multi-session FMRI data

For the multi-session FMRI data, we compare PARAFAC and tensor ICA results to GLM mixed-effects group analysis maps as generated by FLAME (FMRIB's Local Analysis of Mixed Effect; [28]). For comparison, spatial maps generated by any of the three techniques were thresholded using the Gaussian/Gamma mixture-modelling approach described in [2] at a posterior probability level of $p > 0.5$, i.e. at the intensity level where the probability of 'activation' as modelled by the Gamma densities exceeds the probability under the background noise Gaussian density. Individual mixture model fits are given below each of the thresholded maps, together with the relevant time course and (in the case of PARAFAC and tensor-PICA) the estimate of the relative 'activation' strength per session.

The results for GLM are shown in figure 8(i) (all maps are shown in neurological convention, i.e. left hemisphere is displayed on the left). Similar to the thresholded Z-stat maps presented in the original paper by [20], the super-thresholded clusters coincide with areas typically involved in motor processing: bilateral premotor, contra-lateral primary motor and sensory areas, SMA, bilateral secondary somatosensory and the ipsilateral anterior lobe of the cerebellum. Based on the Gaussian/Gamma mixture model fit, significantly negative group-level Z-scores are found in ipsilateral primary motor areas, bilateral intra-parietal sulcus and occipital parietal cortex (blue). Below the GLM map is the first-level GLM regressor together with its power spectrum. Also shown is the normalised (to unit standard deviation) set of first level parameter estimates, weighted by the group-level Z-scores and averaged within post-threshold group-level activation clusters. At the group level, the averaged and weighted set of first-level estimates expresses the change in effect 'strength' between different sessions similar to what is estimated explicitly as part of the PARAFAC and tensor-PICA decomposition as the third mode, $\boldsymbol{C}$ (see [25] for examples of the usefulness of this quantity in the context of model-based FMRI group analysis).

The main PARAFAC map[8] in figure 8(ii) similarly shows super-thresholded clusters in premotor and motor areas, but shows fewer voxels in secondary somatosensory areas and does not identify an ipsilateral cluster in the cerebellar cortex. The power spectrum of the associated time course has highest power at the fundamental frequency of the design (6.5 cycles) but also large power at the first harmonic and some higher frequencies.

By comparison, the primary tensor-PICA map (figure 9(i)) shows much larger correlation with the GLM map than the main PARAFAC map. The spatial map from tensor-PICA shows areas similar to the GLM mixed-effects map, with the tensor-PICA map more prominently showing clusters in bilateral secondary somatosensory (S2) areas. Additionally, cingulate motor and ipsilateral primary motor areas have survived thresholding. Among the 19 estimated sources, this process has not only the highest spatial correlation with the GLM map, but also the highest temporal correlation with the GLM design and highest mean effect size.The rank-1 approximation explains 78.3% of the variation between the temporal responses for each of the sessions. Similar to the PARAFAC results, there is some correspondence between the normalised[9] estimated session response (bottom) and the weighted averaged GLM first-level parameter estimates (figure 8(i), bottom).

The "negative" activation in the GLM map (e.g. ipsilateral motor areas) no longer shows up in this map but is contained within a separate tensor-PICA map[10] shown in figure 9(ii). The most strongly de-activated areas include the ipsilateral primary motor areas and somatosensory areas, possibly de-activating 'non-hand' motor areas as shown previously for the somatosensory system [9]. The plot of the normalised response size over sessions does show that this de-activation is consistent over sessions. The amount of explained variance in the rank-1 approximation , however, is reduced to $\sim 46\%$, suggesting that, unlike primary activation, the de-activation is less consistent in the temporal characteristics between sessions.

Some parts of the de-activation as identified in the GLM analysis (blue in figure 8(i)), however, do not appear in the primary de-activation map shown in figure 9(ii). Instead, a third tensor-PICA map (with correlation of $\rho = 0.27$ to the GLM design) shows de-activation in the superior occipital lobule, an area commonly involved in stereo vision (see figure 10). Unlike the de-activation depicted in figure 9(ii), only a few of the 10 sessions show a significantly non-zero effect size: the boxplot shows sessions 8 and 10 as 'outliers', possibly due to visual fixation.

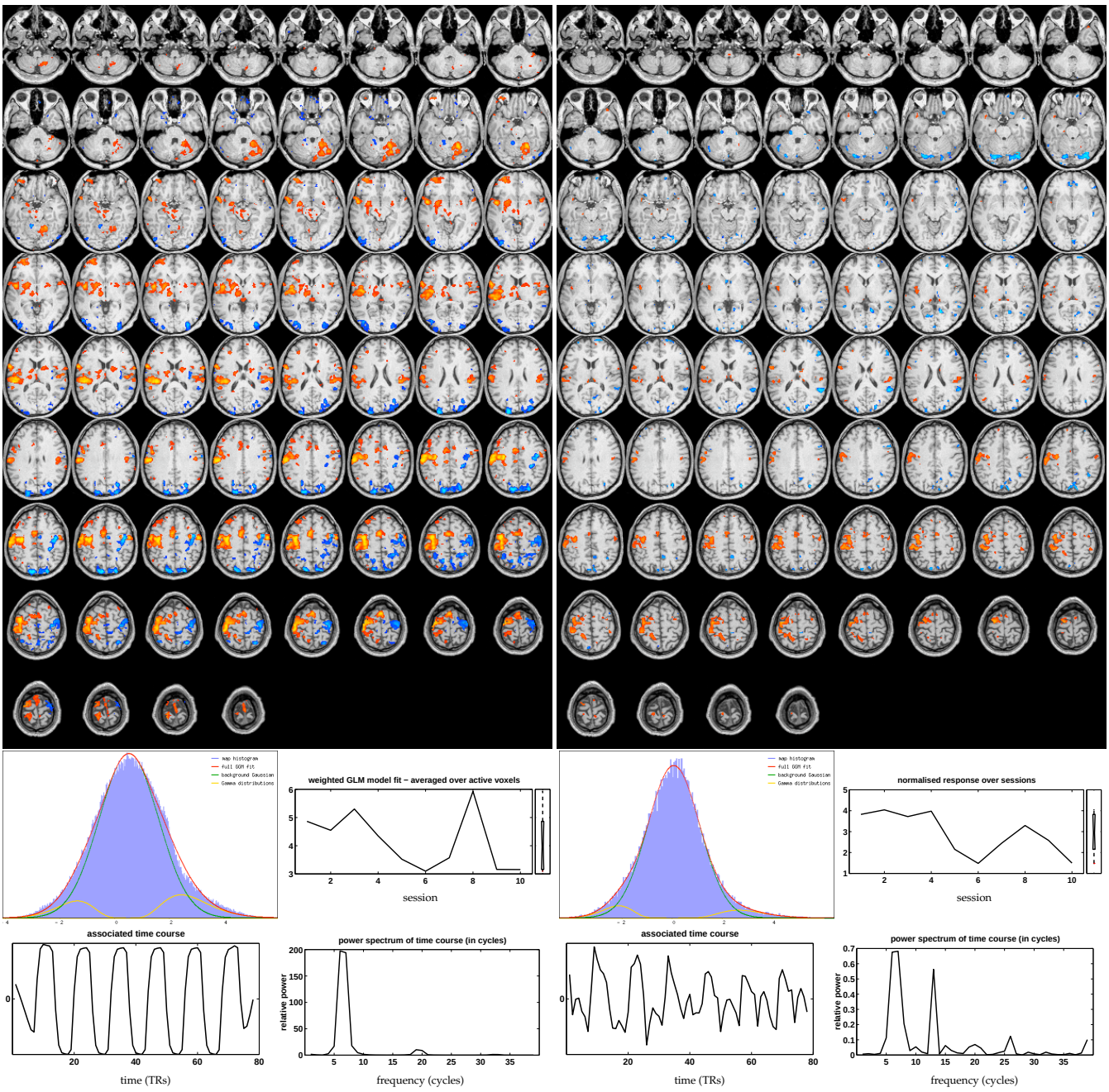
Similar to the case of the artificial data, figure 11 demonstrates that the tensor-PICA results show a much clearer identification of a single "activation" map as well as reduced cross-talk between estimated maps.

Additional 'interesting' maps from the tensor-PICA decomposition are shown in figure 12: the first map (i) depicts the spatial extent of an image artefact showing signal fluctuations possibly due to RF signal aliased into the

[8] Identified as the map with highest spatial correlation with the GLM map out of 19 estimated sources; see figure 11(i). Also, the associated time course has highest correlation with the GLM regressor.

[9] Estimated response size between sessions is normalised to unit standard deviation, thus showing relative response size only.

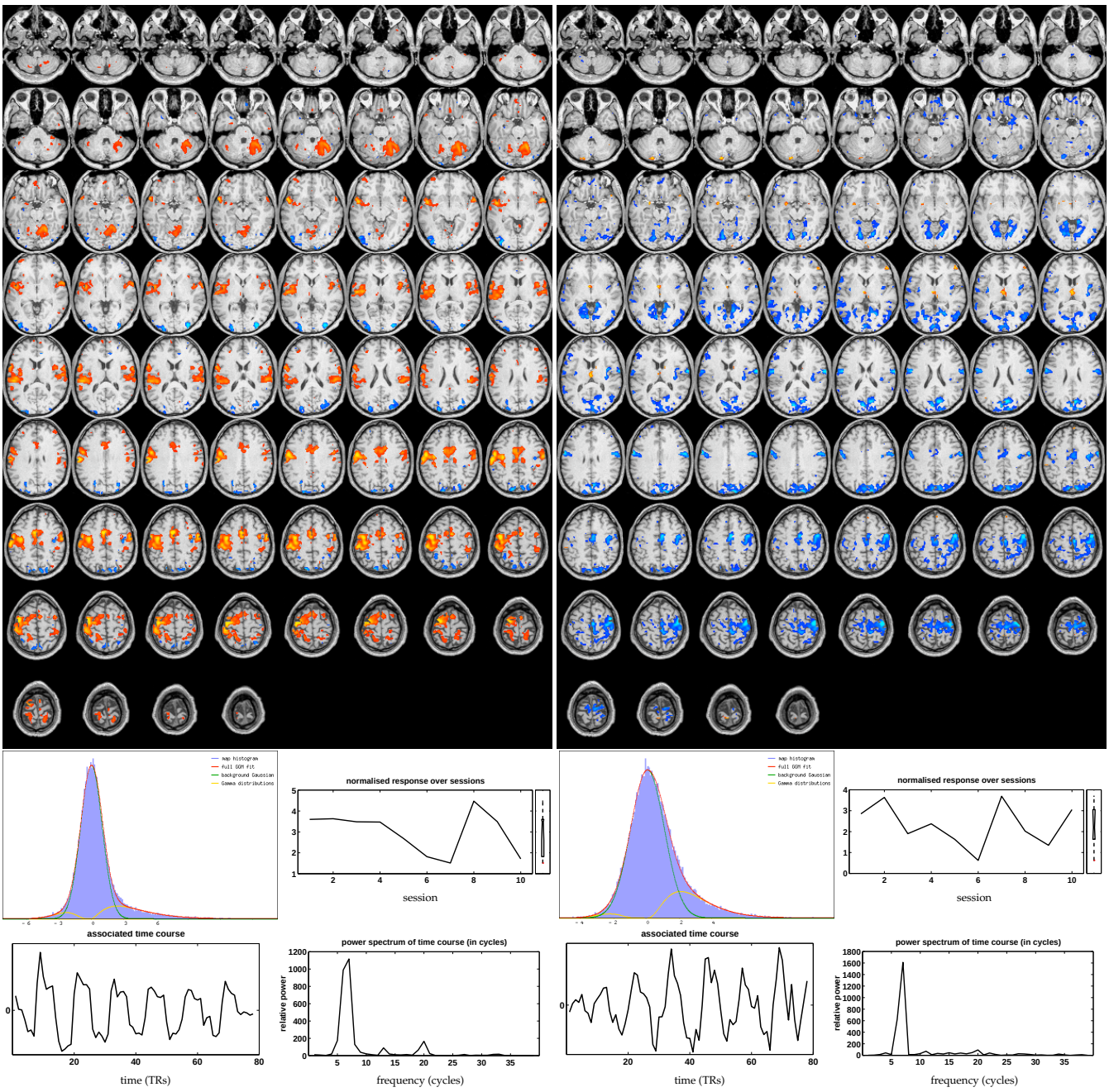
[10] The model is ambiguous with respect to scalar factors and signs. The maps presented here have been scaled manually for comparison.



(i) GLM (ii) PARAFAC

Figure 8: Thresholded FLAME GLM map (i) and PARAFAC map (ii) together with the mixture model fit used for thresholding, associated time courses, power spectra and estimates of relative effect size over subject space.

field-of-view. While the exact origin of these signal components is unknown, they negatively impact on a group GLM analysis as these pattern induce additonal error variance. Figure 12(ii) shows stimulus-correlated residual head motion, most clearly appearing at the frontal lobe intensity boundaries. The presence of this artefact strongly impacts on standard GLM analysis: both the single-level (for session 6) and the group-level GLM estimates for motor activation show false positives around the area where the tensor-PICA map shows the residual motion. Though only a few sessions are estimated to contain this spatio-temporal process, the amplitude modulation induced by the artefacts within these sessions is large enough to be significant even at the group level.



(i) primary tensor-PICA map (ii) secondary tensor-PICA map

Figure 9: Thresholded tensor-PICA maps, mixture model fits, time courses, power spectra of time courses and estimates of relative effect size over subject space: (i) main activation map (map with highest spatial correlation with GLM map and with highest temporal correlation with the GLM first level design); (ii) separate 'de-activation' map identifying ipsilateral primary motor, anterior cerebellum and posterior SMA. Note that in order to facilitate comparison to the de-activation in 8(i), the spatial map and associated time course have been inverted.

5.3 Multi-subject FMRI data

The primary activation map for the multi-subject motor activation study is shown in figure 13. The estimated spatial map shows somatosensory cortex and bilateral primary and secondary motor cortex. Though both left and right motor cortex are shown to activate, the contra-lateral side shows larger amplitude modulation. The associated time course is shown together with the best fit with a 3-level general linear model, where 'activation' during index finger movement (I), sequential finger movement (S) and random finger movement (R) are separate

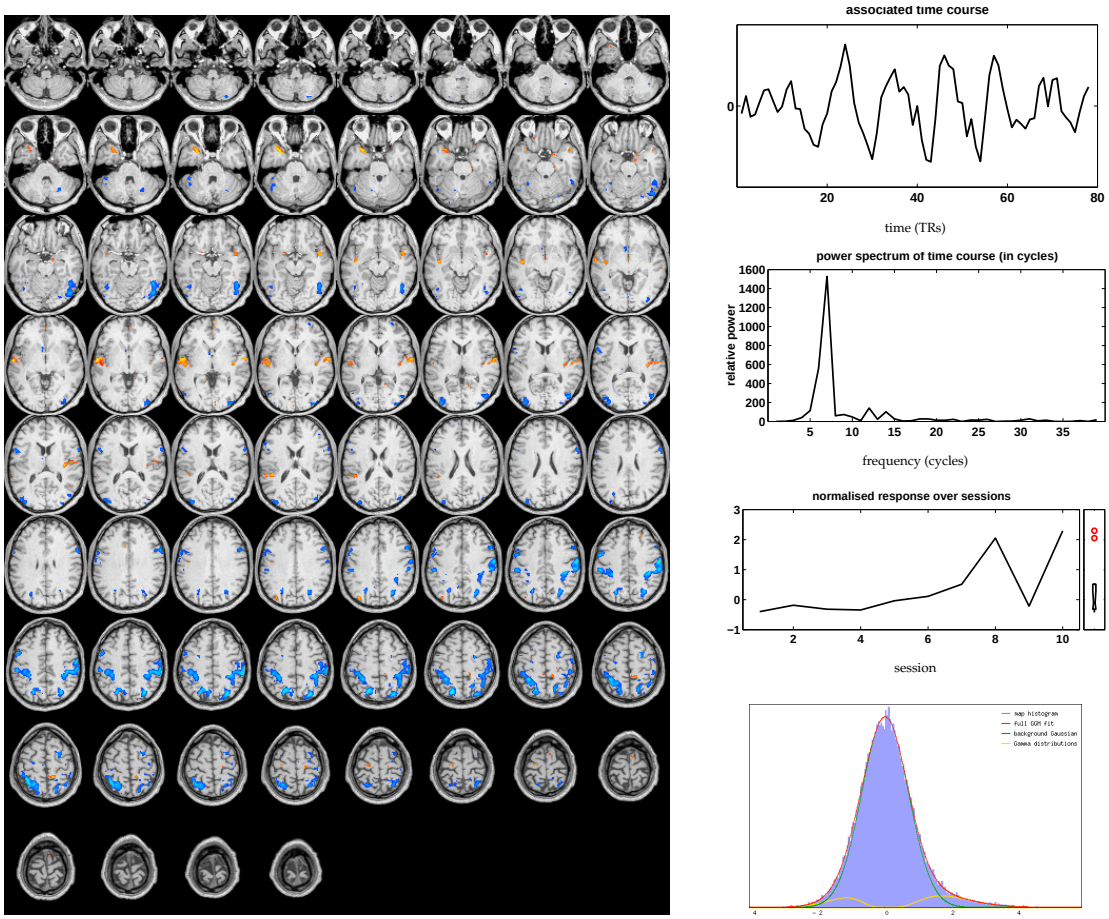


Figure 10: Thresholded tensor-PICA map showing de-activation in the superior occipital lobule and sensory areas.

explanatory variables. The corresponding regression parameters are 0.72 (I), 2.03 (S) and 2.37 (R), suggesting an increase in activation levels I<S<R consistent with results obtained from a GLM analysis of the data (see [16] for details). The final model fit correlates with the estimated data time course at $r > 0.75$.

6 Discussion

We have presented an iterative rank-1 tensor-PICA decomposition for the analysis of single group FMRI data. The method was derived from the three-way PARAFAC model by adding additional maximum non-Gaussianity constraints to the estimated spatial maps. The result of this constraint for estimates in the spatial domain is that the tensor-PICA approach no longer treats all modes of variation as equal. This is an important aspect of the tensor-PICA model, since in FMRI there are substantially different numbers of observations available in the different domains, i.e. a typical FMRI group study involves $10 - 30$ subjects, with $50 - 300$ volumes and 25000 - 45000 intra-cranial voxels (after co-registration into a common space). The tensor-PICA approach, unlike PARAFAC, places stronger statistical constraints on the spatial domain where plentiful data is available.

The approach differs from existing group-ICA methodology [6, 18, 27 and 17, sec. 4] in that it does not simply concatenate the data in space or time in order to perform a single two-dimensional ICA decomposition followed by some meta-analysis to estimate the variation between subjects. Instead, the tensor-PICA approach directly estimates separate modes in the three domains by iterating between estimating a 2-D PICA model on the data concatenated in time and a rank-1 decomposition of the resulting estimate of the mixing matrix $\boldsymbol{M}^c$. Due to the iterative nature, the three-dimensional nature of the data is represented within the estimation stage. This eliminates the need for heuristic post-processing of a set of time-courses (or a set of spatial maps) in order to express the variation across subjects. The technique is fully automated, including the estimation of the model order R.

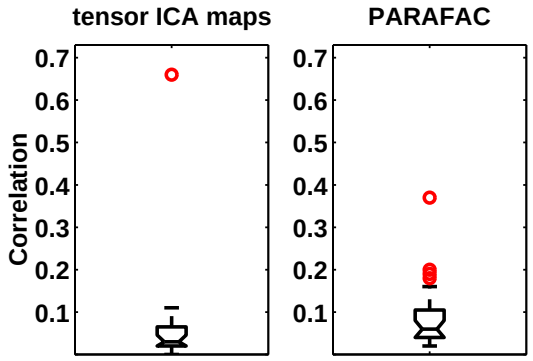


Figure 11: Spatial correlation between GLM map and tensor-PICA resp PARAFAC maps: boxplots of the correlation between the FLAME multi-level GLM map (figure 8 (i)) and tensor-PICA maps (left) and PARAFAC maps (right). Maps with highest correlation are shown in figure 9(i) for tensor-PICA and figure 8(ii) for PARAFAC.

In this paper, we have concentrated on the case of a single group. The methodology can, however, be extended to higher dimensions: under the model of equation 1 where we assume the existence of a single group there is only one single non-zero Eigenvalue for each matrix $\boldsymbol{M}_r^c$ and a rank-1 approximation is appropriate. In practical applications with finite observations and in the presence of noise, the matrices $\boldsymbol{M}_r^c$ will be of full rank. If a sufficient number of observations in the session/subject domain is available, we can apply model order selection techniques and estimate the number of time courses which combine to represent the temporal characteristics of the sources. In the case of 2 groups with similar spatial signal but sufficiently different temporal characteristics, a rank-2 approximation to each matrix $\boldsymbol{M}_r^c$ will then result in a 4-way decomposition of the data. The typically small number of observations in the session/subject domain, however, makes estimation of the model order from the data very difficult. It is possible, however, to impose the number of different time courses that are used to represent the temporal characteristics of all sessions/subjects/groups, e.g. one can use a rank-2 approximation to $\boldsymbol{M}_r^c$ in order to estimate different temporal responses from 2 different subgroups in the population.

For the generative model of equation 1 we have demonstrated, on a set of artificial group data sets, that tensor-PICA can successfully estimate multiple processes in the spatial, temporal and subject/session domain. Compared to a PARAFAC decomposition, the tensor-PICA estimation shows significant improvements in the form of: (i) an increased accuracy for primary 'activation' maps, (ii) reduced cross-talk between the different estimated spatial maps and (iii) an increased robustness against deviation from the model assumptions and against estimating the model order R incorrectly.

All of these improvements are a direct consequence of the optimisation for maximally non-Gaussian spatial source distributions. Typical FMRI 'activation' is sparse in the spatial domain and the estimated linear regression coefficients (spatial maps) will contain only a few 'significantly large' values embedded in random Gaussian distributed 'background noise'. An optimisation for non-Gaussianity of estimated spatial maps optimises for the largest possible separation of the first set of regression coefficients ('activation') from all other remaining coefficients ('background'). Unlike the sum-of-squares error function associated with the PARAFAC model, the error function associated with an optimisation for maximum non-Gaussianity does not improve from 'splitting' components and/or having multiple components which 'explain' the same signals. In particular, an optimisation for jointly maximal non-Gaussian spatial maps implies a minimization of statistical dependence and cross-talk in the spatial domain (see [13] for a clear account of the relation between statistical independence and non-Gaussianity).

As an additional benefit over PARAFAC, the tensor-PICA decompositions each required significantly less computation (between $1/10$ and $1/100$ times the number of floating point operations) compared to PARAFAC in order to converge to a solution[11]. Again this is a consequence of the fact that the cost function in a Tensor-PICA decomposition is more sensitive to the particular signal characteristics in the spatial domain. As a result, the number of iterations until convergence is much reduced (the number of floating-point operations per iteration is actually greater when using Tensor-PICA as opposed to PARAFAC).

Using real FMRI data we have demonstrated that tensor-PICA can extract plausible spatial maps and time courses. The main activation pattern of the multi-session decomposition identified cortical regions which corres-

[11] E.g. approximately 40min for the tensor-PICA estimation compared to ∼8h for PARAFAC on the real FMRI data on a Compaq Alpha ES40 667MHz Server with Matlab 5.3 (excluding registration of individual session data to the template space) for the results presented in section 5.2.

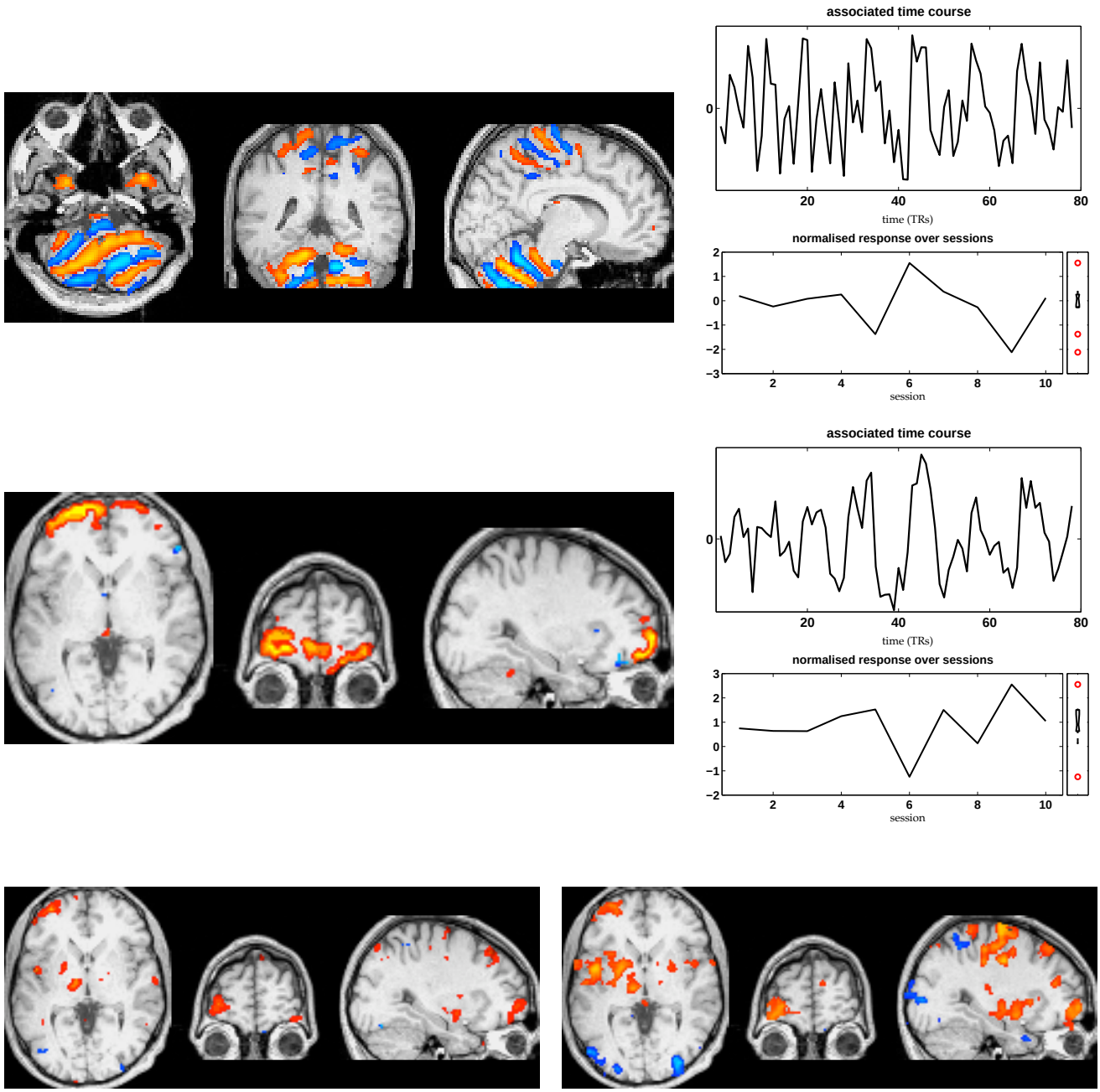


Figure 12: Additional tensor-PICA maps: Image artefact (i), particularly strong in 3 sessions and residual motion artefact (ii). The motion correlates with the stimulus and therefore confounds the individual (single-session) and group GLM maps. (iii) shows the Z-statistics map for session 6, obtained from a standard GLM analysis (left) and the group-level GLM map (right, identical to figure 8(i)).

pond to what has been reported in [20]. Furthermore, the tensor-PICA decomposition gives a rich description of additional processes in the data. For example, the tensor-PICA decomposition separated negative (de-)activation into different plausible spatial maps with associated time-courses and variation across sessions. The final decomposition does suggest that there are at least two distinct processes which contribute to the negative Z-scores: plausible ipsilateral de-activation in the primary motor cortex consistent across sessions and de-activation in the

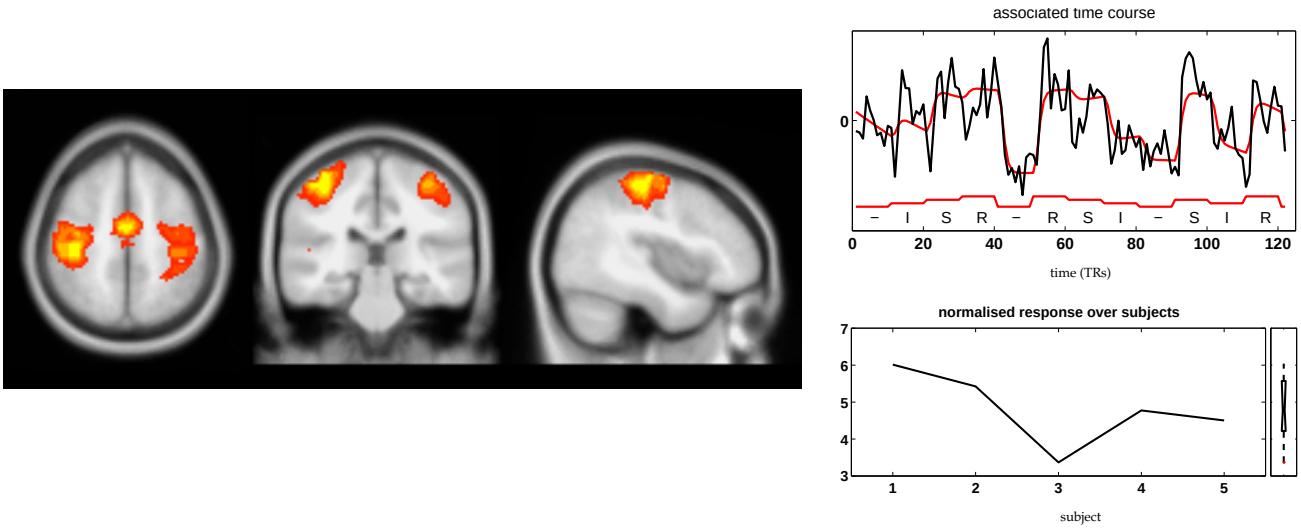


Figure 13: Thresholded tensor-PICA map and associated time course for the multi-subject FMRI data set. Post-thresholded areas include primary and secondary motor areas and SMA. The associated time course is shown together with the best model fit based on linear regression of the estimated temporal mode against the GLM design with 3 exploratory variables (modelling index finger movement (I), sequential finger movement (S) and random finger movement (R) separately).

superior occipital lobule which only appears in a few sessions, possibly due to visual fixation. In addition, the tensor-PICA decomposition identified nuisance effects like artefactual RF signal components or stimulus correlated motion at the group level.

We believe that the tensor-PICA approach can provide simple and useful representations of multi-subject/multi-session FMRI data that can aid the interpretation and optimisation of group FMRI studies.

7 Acknowledgements

The authors wish to thank Dr. D. Leibovici and Prof. P.A. Valdés-Sosa for helpful discussions on multi-way analysis, Dr. D.J. McGonigle for granting access to the FMRI data, Prof. P.M. Matthews and Dr. H. Johansen-Berg for their advice on neuroanatomy and Prof. M.S. Cohen for helpful discussions about MR physics. The authors gratefully acknowledge the financial support from the UK EPSRC and GlaxoSmithKline.

References

[1] Beckmann, C., Jenkinson, M., and Smith, S. (2003a). General multi-level linear modelling for group analysis in FMRI. *NeuroImage*, 20:1052–1063. first two authors contributed equally.

[2] Beckmann, C. and Smith, S. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging*, 23(2):137–152.

[3] Beckmann, C., Woolrich, M., and Smith, S. (2003b). Gaussian / Gamma mixture modelling of ICA/GLM spatial maps. In *Ninth Int. Conf. on Functional Mapping of the Human Brain*.

[4] Biswal, B., DeYoe, E., and Hyde, J. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, 35:107–113.

[5] Bro, R. (1998). *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*. PhD thesis, University of Amsterdam.

[6] Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(140–151).

[7] Cao, Y.-Z., Chen, Z.-P., Mo, C.-Y., Wu, H.-L., and Yu, R.-Q. (2000). A PARAFAC algorithm using penalty diagonalization error (PDE) for three-way data array resolution. *The Analyst*, 125:2303–2310.

[8] Carroll, J. and Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319.

[9] Drevets, W., Burton, H., Videen, T., Snyder, A., Simpson, J. J., and Raichle, M. (1995). Blood flow changes in human somatosensory cortex during anticipated stimulation. *Nature*, 373(6511):198–199.

[10] Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "exploratory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

[11] Harshman, R. and Lundy, M. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In *Research methods for multimode data analysis*, chapter 5, pages 122–215. Praeger, New York.

[12] Harshman, R. and Lundy, M. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis*, 18:39–72.

[13] Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492.

[14] Jenkinson, M., Bannister, P., Brady, J., and Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841.

[15] Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156.

[16] Johansen-Berg, H., Rushworth, M., Bogdanovic, M., Kischka, U., Wimalartna, S., and Matthews, P. (2002). The role of ipsilateral premotor cortex in hand movement after stroke. *Proceedings of the National Academy of Sciences*, pages 14518–14523.

[17] Leibovici, D. and Beckmann, C. (2001). An introduction to multiway methods for multi-subject fMRI experiments. Internal Technical Report TR01DL1, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK.

[18] Lukic, A. S., Wernick, M. N., Hansen, L. K., Anderson, J., and Strother, S. C. (2002). A spatially robust ica algorithm for multiple fMRI data sets. In *IEEE International Symposium on Biomedical Imaging, Proceedings*, pages 839–842.

[19] Marchini, J. and Ripley, B. (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage*, 12(4):366–380.

[20] McGonigle, D., Howseman, A., Athwal, B., Friston, K., Frackowak, R., and Holmes, A. (2000). Variability in fMRI: An examination of intersession differences. *NeuroImage*, 11:708–734.

[21] McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–88.

[22] Minka, T. (2000). Automatic choice of dimensionality for PCA. Technical Report 514, MIT Media Lab.

[23] Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. (2004). Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *NeuroImage*, 22:1035–1045.

[24] Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155.

[25] Smith, S., Beckmann, C., Ramnani, N., Woolrich, M., Bannister, P., Jenkinson, M., Matthews, P., and McGonigle, D. (2004). Variability in FMRI: A re-examination of intersession differences. *Human Brain Mapping*. in press.

[26] Strother, S., Kanno, I., and Rottenberg, D. (1995). Principal component analysis, variance partitioning and "functional connectivity". *J Cereb Blood Flow Metab*, 15(5):353–360.

[27] Svensén, M., Kruggel, F., and Benali, H. (2002). ICA of fMRI Group Study Data. *NeuroImage*, 16:551–563.

[28] Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. (2003). Multi-level linear modelling for FMRI group analysis using Bayesian Inference. *NeuroImage*, 21(4):1732–1747.

[29] Woolrich, M., Behrens, T., Beckmann, C., and Smith, S. (2004). Mixture models with adaptive spatial regularisation for segmentation with an application to FMRI data. *IEEE Trans. on Medical Imaging*. in press.

[30] Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. on Medical Imaging*, 20(1):45–57.