Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB)
Department of Clinical Neurology
University of Oxford

Modelling with Independent Components

Christian F. Beckmann
beckmann@fmrib.ox.ac.uk

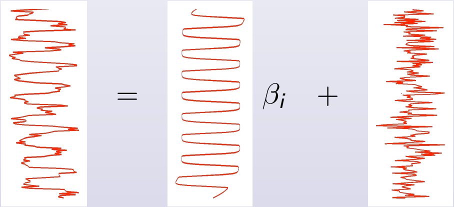
IPAM Mathematics in Brain Imaging - 22/07/04

Outline

1. Exploratory Data Analysis
   - Principles of EDA
   - Principal Component Analysis
   - Independent Component Analysis
2. Probabilistic Independent Component Analysis for FMRI
   - Estimating the model order
   - Estimating Independent Components
   - Statistical Inference on IC maps
   - Full PICA model
3. Application of (P)ICA to FMRI data
   - Investigating the BOLD response
   - Artefact detection
   - Estimating 'difficult' activation pattern
   - Investigation into resting-state networks
4. Tensor-PICA

Review of the GLM

- model each measured time-series as a linear combination of signal and noise: $\mathbf{x}_i = \mathbf{Y}\beta_i + \eta_i$



- If the design matrix does not model all signal, we get wrong inferences!

Classical vs. Exploratory Data Analysis

Classical Data Analysis (e.g. GLM)

- "How well does the model fit the data"
- Problem → Data → Model → Analysis → Conclusions
- results depend on the model
- test specific hypothesis

Exploratory Data Analysis (e.g. ICA)

- "Is there anything interesting in the data"
- Problem → Data → Analysis → Model → Conclusions
- results depend on the data
- can give 'surprising' results

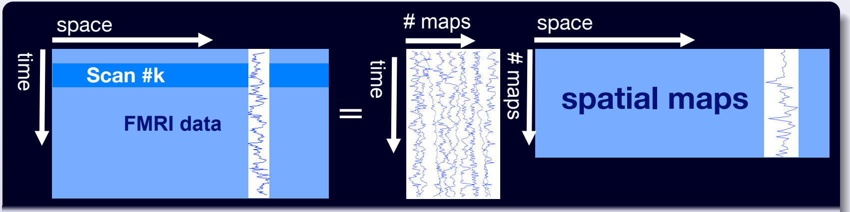
Exploratory Data Analysis techniques

- try to explain / represent the data
  - by calculating quantities which summerise data
  - by extracting underlying (hidden) variables that are 'interesting'

- differ in what is considered to be interesting
  - signals which explain (co-)variances *(PCA, FDA, FA)*
  - signals which have large (co-)variances with e.g. a design matrix *(PLS, CVA)*
  - signals which are clustered in space/time *(clustering)*
  - signals which are statistically independent / maximally non-Gaussian *(ICA)*

Exploratory Data Analysis techniques

- often are *multivariate*
- often provide a multivariate *linear* decomposition



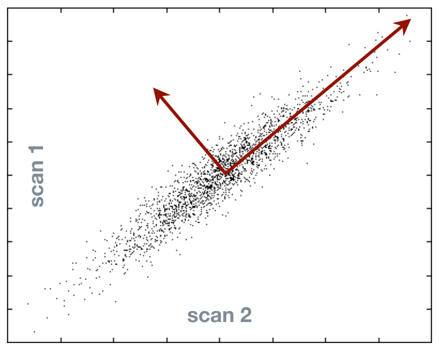
$$\mathbf{X} = \sum_r^R \mathbf{a}_r \otimes \mathbf{b}_r + \boldsymbol{\eta}$$

Data are represented as a 2D matrix and decomposed into factor matrices **A** and **B**, representing the characteristics of R processes in time and across space

Principal Component Analysis (PCA)

- finds new variables which are linear combinations of the observed data along axis of maximum variation

Example

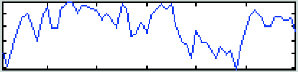


$$\mathbf{w}_1 = \arg\max_{||\mathbf{w}||=1} \mathrm{E}\{(\mathbf{w}^t\mathbf{x})^2\},$$

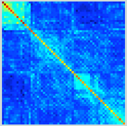
$$\mathbf{w}_k = \arg\max_{||\mathbf{w}||=1} \mathrm{E}\{(\mathbf{w}^t(\mathbf{x} - \textstyle\sum_{i<k} \mathbf{w}_i^t\mathbf{x}\mathbf{w}_i))^2\}$$

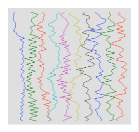
Principal Component Analysis for FMRI

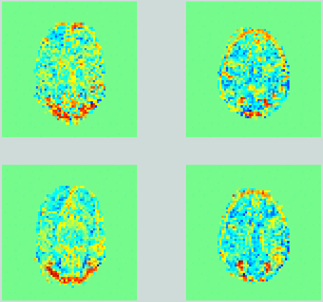
- calculate the data covariance matrix
- calculate the full set of Eigenvectors
- calculate the Eigenimages by projecting the data onto the Eigenvectors



Example





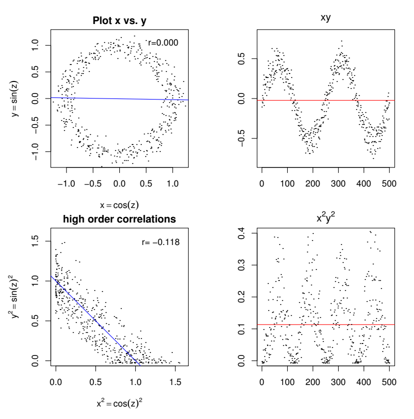


Statistical independence and correlation

- de-correlated signals can still be dependent
- higher-order statistics (beyond mean and variance) can reveal those dependencies

Stone, *Trends Cog. Sci.*, 6(2):59–64 (2002)

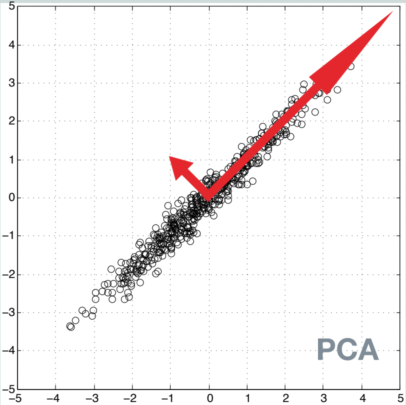
Example



PCA versus ICA

- Principal Component Analysis (PCA) finds directions of maximal variance in Gaussian data (uses second-order statistics)
- Independent Component Analysis (ICA) finds directions of maximal independence in non-Gaussian data (higher-order statistics)

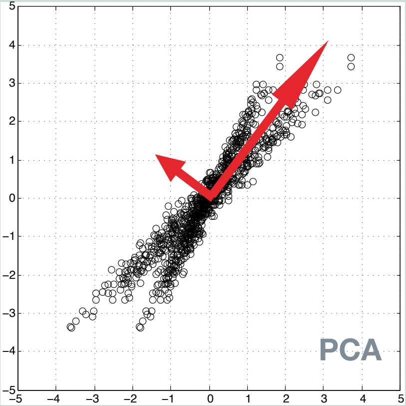
Gaussian data



PCA versus ICA

- Principal Component Analysis (PCA) finds directions of maximal variance in Gaussian data (uses second-order statistics)
- Independent Component Analysis (ICA) finds directions of maximal independence in non-Gaussian data (higher-order statistics)

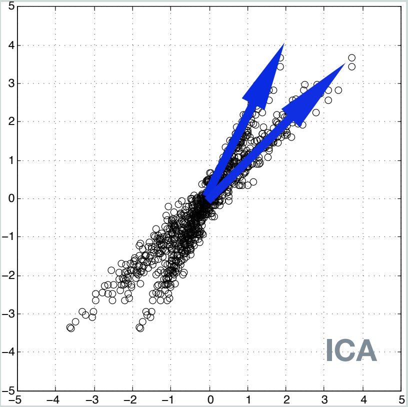
non-Gaussian data



PCA versus ICA

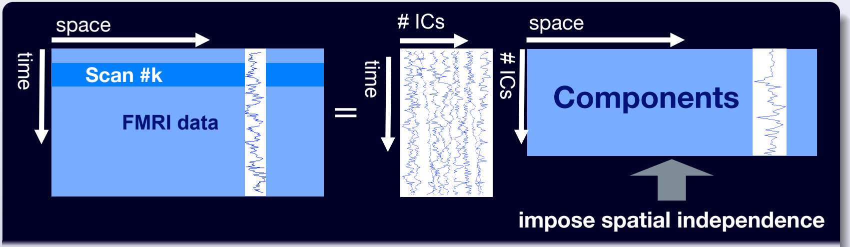
- Principal Component Analysis (PCA) finds directions of maximal variance in Gaussian data (uses second-order statistics)
- Independent Component Analysis (ICA) finds directions of maximal independence in non-Gaussian data (higher-order statistics)

non-Gaussian data



Spatial ICA for FMRI

- the data is represented as a 2D matrix and decomposed into a set of *spatially independent component maps* and a set of associated time-courses

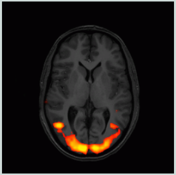


Noise-free generative model

McKeown *et.al.*, *Human Brain Mapping*, 6(5):368–372 (1998)

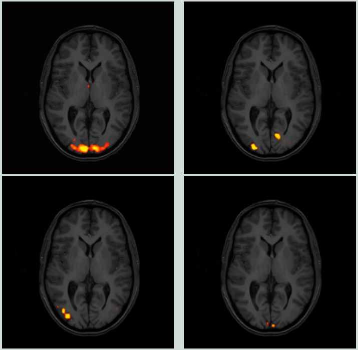
The 'Overfitting Problem'

Example: visual stimulation, b/w reversing checkerboard (8Hz)



GLM results
(using FEAT)

- caused by fitting a noise free model to noisy data
- in the absence of a noise model, everything is significant!

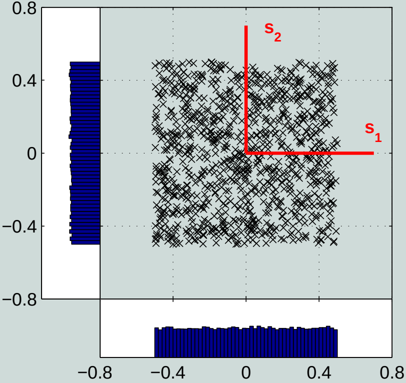


std. ICA results
(all maps with $r > 0.3$ temp. corr. between time-course and design)

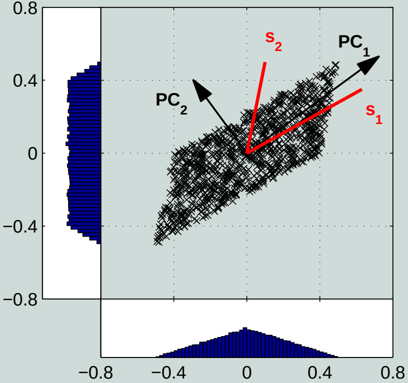
Probabilistic ICA

- statistical 'latent variables' model: we observe linear mixtures of hidden sources
$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i + \boldsymbol{\eta}_i$$
- If $\boldsymbol{\eta}_i \sim \mathcal{N}[0, \sigma^2 \boldsymbol{\Sigma}_i]$ we can use voxel-wise pre-whitening (e.g. Woolrich *et.al*, *NeuroImage*, 14(6):1370–1386 (2001))
- If $\boldsymbol{\eta}_i \sim \mathcal{N}[0, \sigma^2 \mathbf{I}]$ then $\mathbf{R_X} \rightarrow \mathbf{A}\mathbf{A}^t + \sigma^2 \mathbf{I}$ as $N \rightarrow \infty$, i.e. for isotropic Gaussian noise the eigenspectrum is raised by σ^2
- we can estimate the *model order* from the Eigenspectrum of the data covariance matrix $\mathbf{R_X}$
- but $\mathbf{R_X} = \mathbf{R_{XQ}}$ for any $\mathbf{Q}$ with $\mathbf{Q}\mathbf{Q}^t = \mathbf{I}$, i.e. $\mathbf{R_X}$ is *rotational invariant*

Rotational invariance: the geometry of PCA and ICA

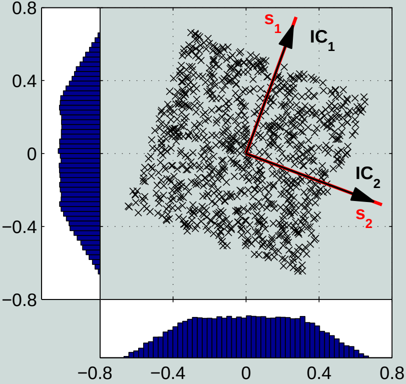


2 independent, uniformly distributed sources s_1 and s_2

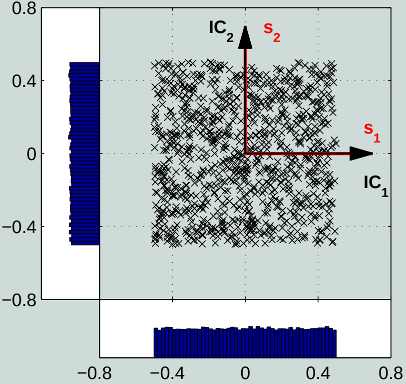


linear mixtures of sources with principal directions

Rotational invariance: the geometry of PCA and ICA



PCA solution (projection on to PC_1 and PC_2)



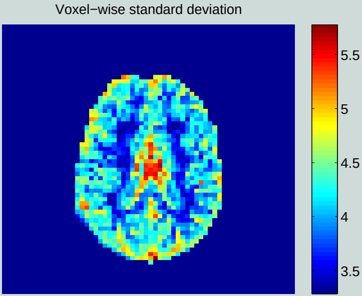
ICA solution (rotation of PCA solution)

Variance-normalisation

- need to normalise by the voxel-wise variance
- this amounts to modelling the spatial covariance matrix as as diagonal:

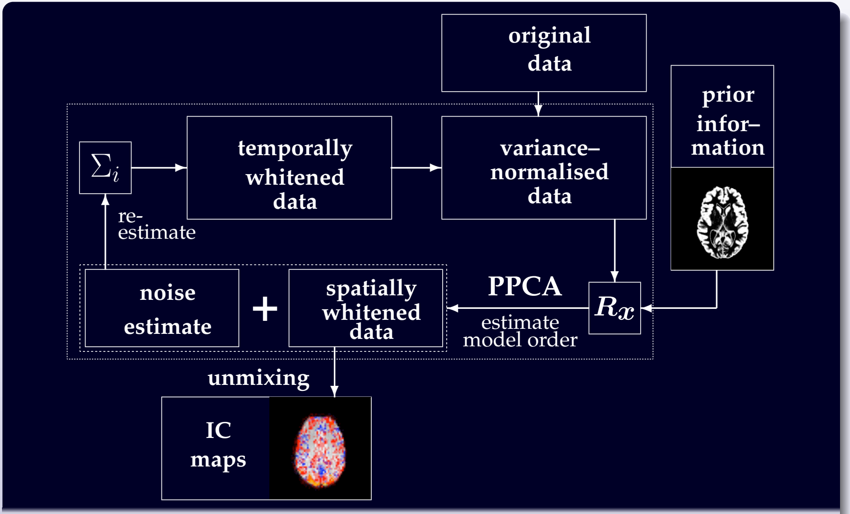
$$\mathbf{V}^{-1/2} = \mathrm{diag}(\sigma_1, \dots, \sigma_N)$$

Example: FMRI resting state data



Estimated voxel-wise noise standard deviation (log-scale)

Probabilistic ICA (I)



Incorporating prior information

- use regularised PCA (or FDA) to regularise time courses (Ramsay & Silverman, *Functional Data Analysis* (1997))
- signal+noise sub-space is determined from data cov. matrix:

$$\begin{aligned}\mathbf{R_x} &= \sum_i w_i(\mathbf{x}_i - \langle\mathbf{x}\rangle)(\mathbf{x}_i - \langle\mathbf{x}\rangle)^t \quad (\text{typically } w_i = \tfrac{1}{N}) \\ &\propto \sum_{ij} w_i w_j m_{ij}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^t \\ &\quad + \sum_{ij} w_i w_j (1 - m_{ij})(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^t,\end{aligned}$$

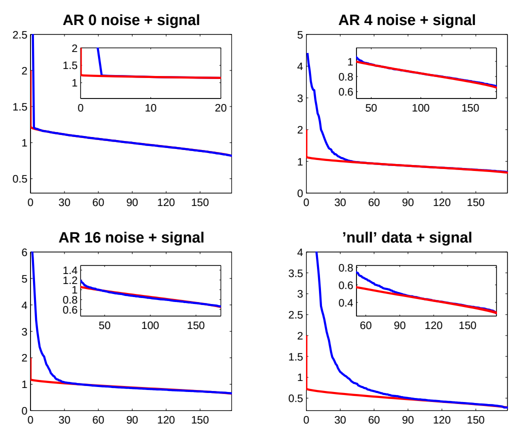
- the matrix $\mathbf{M} = (m_{ij}); m_{ij} \in [0, 1]$ defines a weighted graph of N nodes: can be used to spatially regularise PPCA

Model order selection (Probabilistic PCA)

- the *sample covariance matrix* has a Wishart distribution and we can calculate the empirical distribution function for the eigenvalues

Everson & Roberts, *IEEE TSP*, 48(7):2083–2091 (2000)

Example: 2 signals in noise

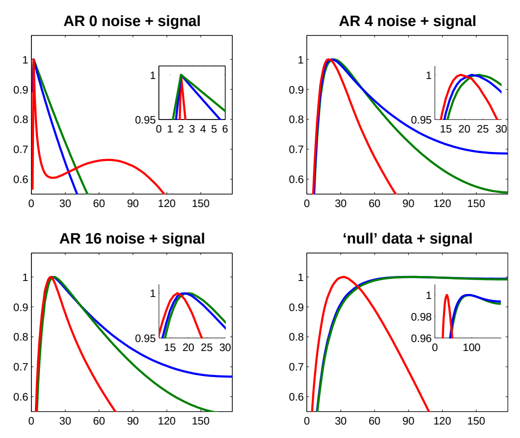


Model order selection (Probabilistic PCA)

- use a probabilistic PCA model and calculate (approximate) the Bayesian evidence for the model order

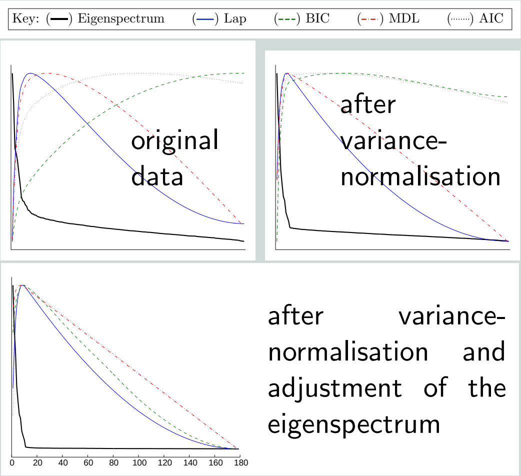
Minka, *TR 514 MIT Media Lab* (2000)

Example: 2 signals in noise



- different Bayesian estimators at different points in the processing chain
- different estimators give similar results
- *Laplace approximation of the Bayesian evidence* is most robust

Example: 10 signals in Gaussian noise

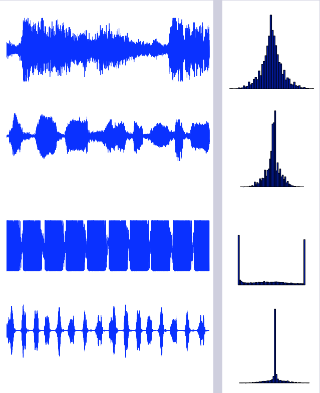


Component estimation

- estimate an 'unmixing matrix' $\mathbf{W} = \mathbf{A}^{\dagger}$ such that the statistical dependency between the estimated sources $\widehat{\mathbf{s}}_i = \mathbf{W}\mathbf{x}_i$ is minimised
- use (i) a *contrast function* and (ii) an *optimisation technique*:
  - kurtosis or cumulants & gradient descent (Jade)
  - maximum entropy & gradient descent (Infomax)
  - neg-entropy & fixed-point iteration (FastICA)

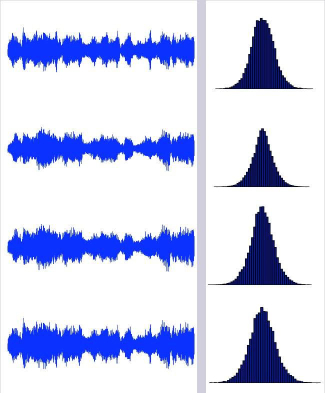
non-Gaussianity is interesting

original sounds



mixtures are more Gaussian

mixtures

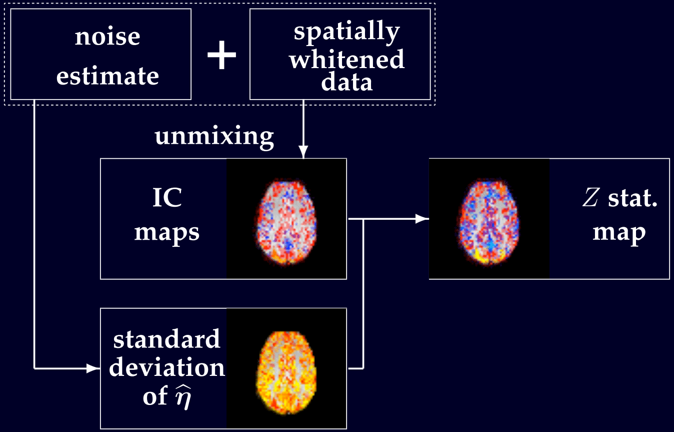


Component estimation

- random mixing results in more Gaussian shaped pdfs (*Central Limit Theorem*)
- if an 'unmixing matrix' produces non Gaussian signals, this is unlikely to be a random result
- use neg-entropy as a measure of non-Gaussianity:
  $$\mathcal{J}(\mathbf{s}) = \mathcal{H}(\mathbf{s}_{\text{gauss}}) - \mathcal{H}(\mathbf{s})$$
- allows for the identification of exactly those source processes which violate standard GLM assumptions
- can use fast approximations:
  $$\mathcal{J}(\mathbf{s}) \simeq \sum_i^p \kappa_i(\mathrm{E}\{g_i(\mathbf{s})\} - \mathrm{E}\{g_i(\mathbf{s}_{\text{gauss}})\})$$

Hyvärinen & Oja, *Neural Computation*, 9(7):1483–1492 (1997)

Probabilistic ICA (II)

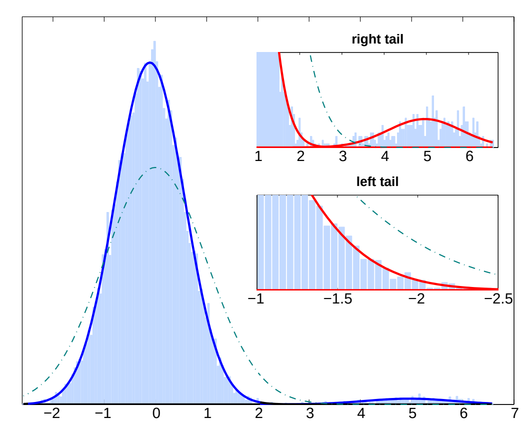


form voxel-wise Z-statistics using the estimated standard deviation of the noise

Thresholding IC maps

- estimated maps have been optimised to violate the noise model
- null-hypothesis test is invalid
- thresholding based on Z-transforming across the spatial domain gives wrong false-positives rate

example histogram and fit to single Gaussian



Thresholding IC maps

- under the model:

$$\widehat{\mathbf{S}}_{\text{ML}} = \widehat{\mathbf{A}}^{\dagger}\mathbf{X} = \widehat{\mathbf{A}}^{\dagger}\mathbf{A}\mathbf{S} + \widehat{\mathbf{A}}^{\dagger}\mathbf{E},$$

  i.e. the *estimated* spatial maps contain a linear projection of the noise

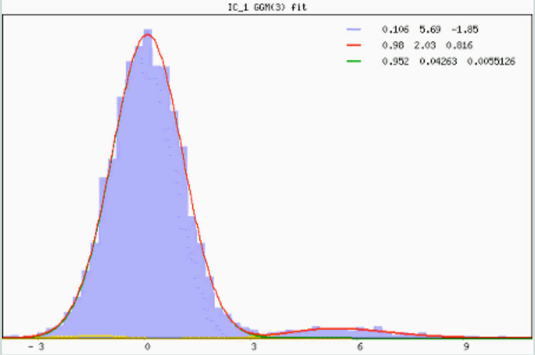
- the distribution of the estimated spatial maps is a mixture distribution
- use Gaussian / Gamma mixture model for each spatial map $\mathbf{s}_r$:

$$\begin{aligned} p(\mathbf{s}_r|\boldsymbol{\theta}) &= \pi_{r,1}\mathcal{N}[\mathbf{s}_r; \mu_{r,1}, \sigma^2_{r,1}] \\ &+ \pi_{r,2}\mathcal{G}^{+}[\mathbf{s}_r - \mu_{r,1}; \mu_{r,2}, \sigma_{r,2}] \\ &+ \pi_{r,3}\mathcal{G}^{-}[-\mathbf{s}_r + \mu_{r,1}; \mu_{r,3}, \sigma_{r,3}] \end{aligned}$$

Thresholding IC maps

- fit using Expectation Maximisation (EM)
- different ways of thresholding: posterior probabilities, NHT, FDR
- no multiple-comparison problem

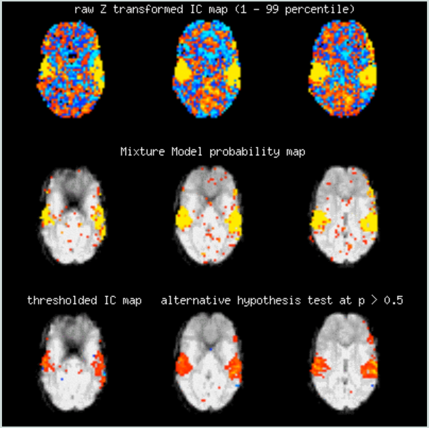
Example



Thresholding IC maps

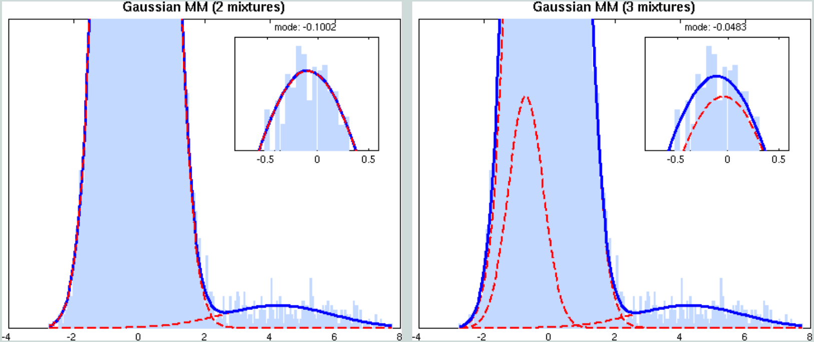
- fit using Expectation Maximisation (EM)
- different ways of thresholding: posterior probabilities, NHT, FDR
- no multiple-comparison problem

Example



Why Gaussian/Gamma mixtures?

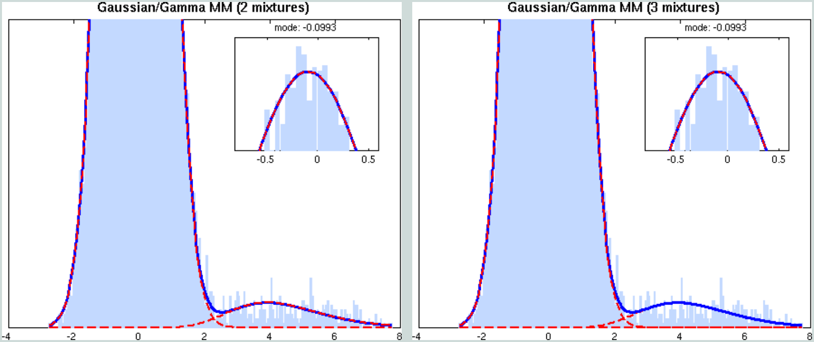
Example: Gaussian MM fit (2,3 mixtures)



- Gaussian densities for 'non-background' classes are suboptimal: 'non-background' densities are typically not symmetric.

Why Gaussian/Gamma mixtures?

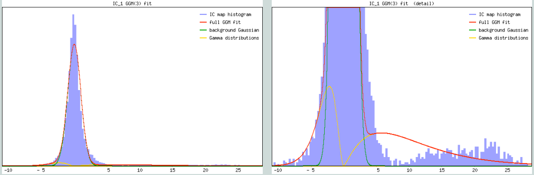
Exampe: Gaussian/Gamma MM fit (2,3 mixtures)

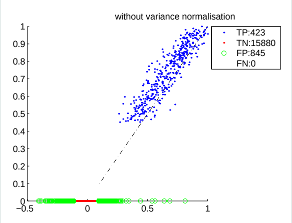


- Gaussian/Gamma MM is more robust wrt specification of the right number of mixtures

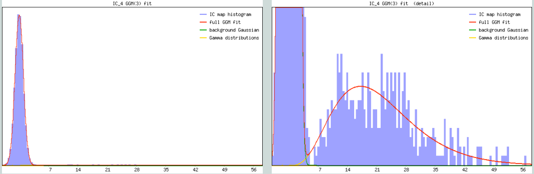
The effect of variance-normalisation

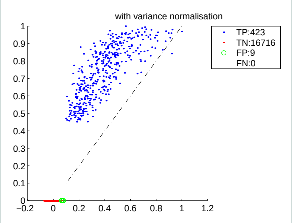
IC histogram (without variance-norm.)

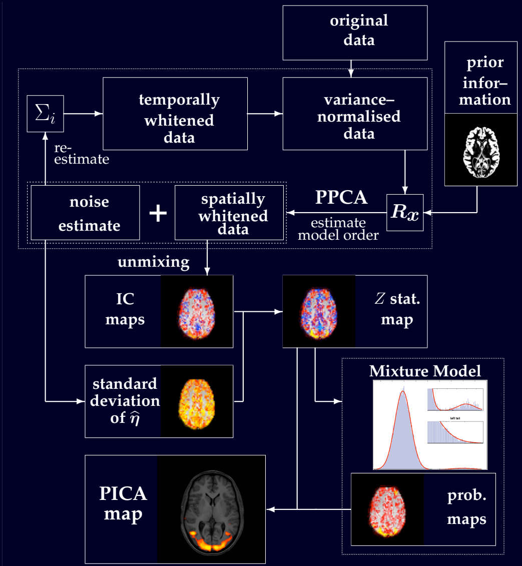




IC histogram (with variance-norm.)







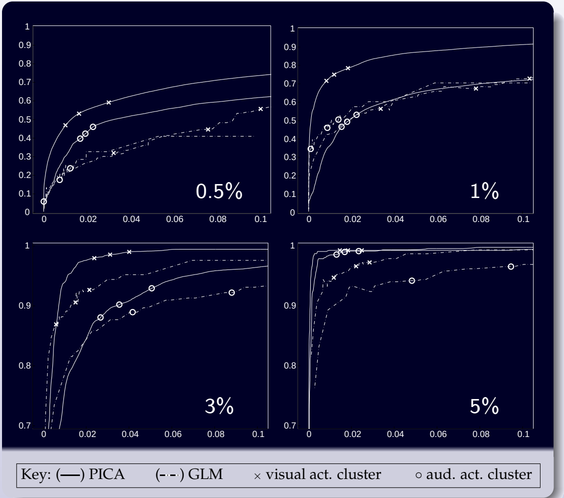
full PICA model

implemented as Melodic, part of FMRIB's Software Library (FSL)

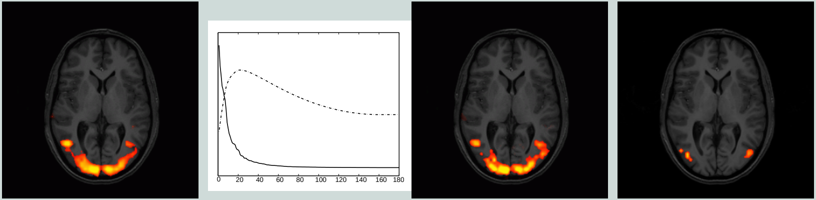
Beckmann & Smith, *IEEE TMI*, 23(2):137–152 (2004)

Receiver-Operator Characteristics

- simulated FMRI data
- PICA vs. GLM at different ‘activation‘ levels and different thresholds
- plot of true-positives rate vs. false-positives rate

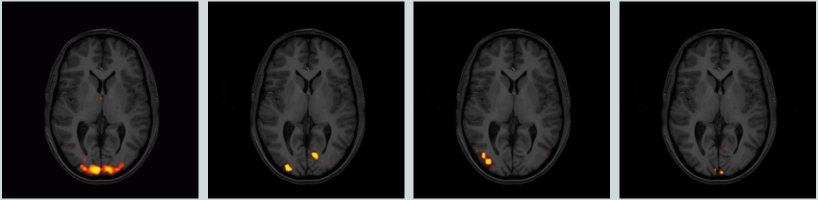


Example: visual stimulation, b/w reversing checkerboard (8Hz)



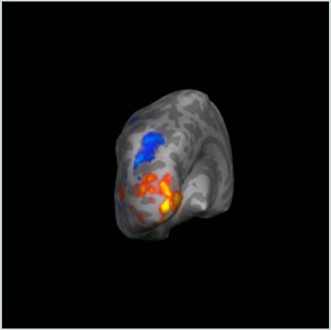
GLM results | dim.-est. | PICA results (constrained)

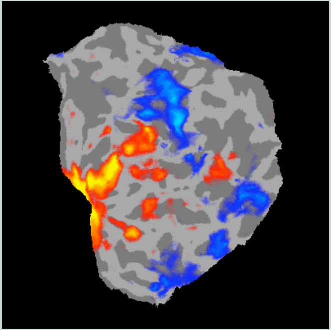
standard ICA (unconstrained)



Example: visual stimulation, b/w reversing checkerboard (8Hz)

- PICA maps show primary visual cortex and V3 (MT)





Can we still estimate spatially correlated signals?

- spatial correlation between 2 sources s_1 and s_2:

$$\rho(s_1, s_2) = \frac{s_1^t s_2}{N\sqrt{\mathrm{Var}(s_1)}\sqrt{\mathrm{Var}(s_2)}}$$

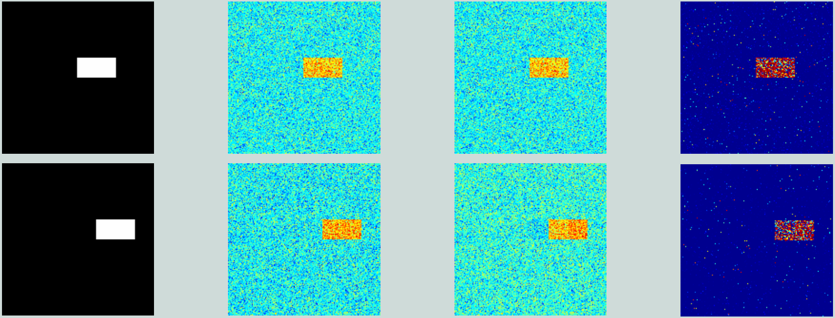
- in the presence of noise:

$$\rho(s_1 + \eta_1, s_2 + \eta_2) = \frac{s_1^t s_2}{N\sqrt{\mathrm{Var}(s_1) + \sigma_1^2}\sqrt{\mathrm{Var}(s_2) + \sigma_2^2}},$$

i.e. for *sparse* signals in noise, imposing orthogonality (de-correlating estimated signals) is not necessarily restrictive

Example: 2 correlated sources

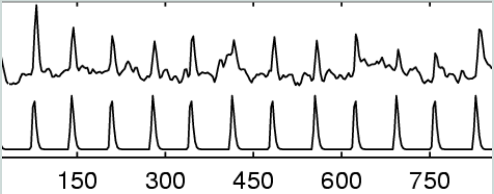
original sources	sources with noise	de-correlated sources	thresholded sources



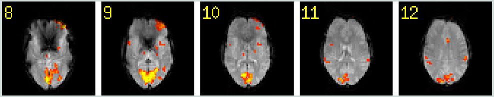
$\rho = 0.5$	$\rho < 0.1$	$\rho = 0$	$\rho \approx 0.5$

Investigating the temporal characteristics of the BOLD response

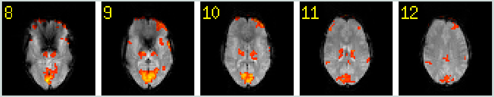
- pain study: 14 short bursts of painful heat
- estimated (top) and expected (bottom) temporal response to stimulation



- GLM result using canonical model



- GLM result using estimated model

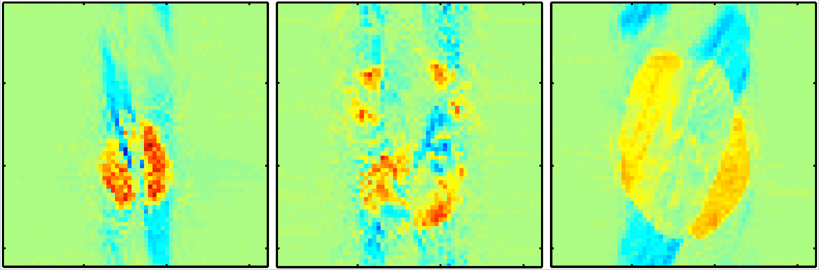


Wise & Tracey, 2000

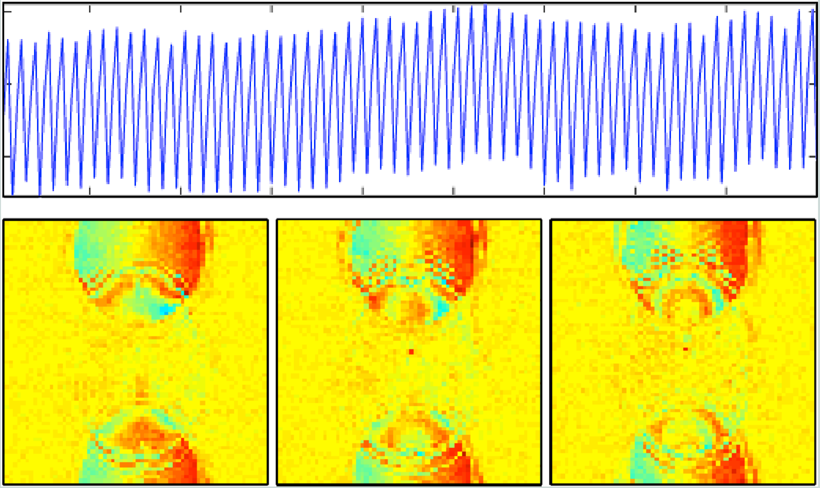
Detecting artefacts in FMRI data

- FMRI data contain a variety of source processes
- Artefactual sources typically have unknown spatial and temporal extent and cannot easily be modelled accurately
- exploratory techniques do not require a priori knowledge of time-courses and/or spatial maps

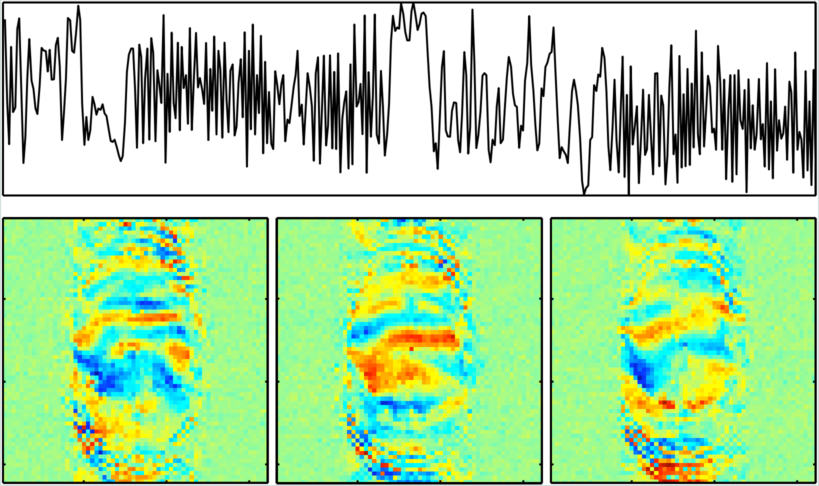
slice-dropout (scanner instability)



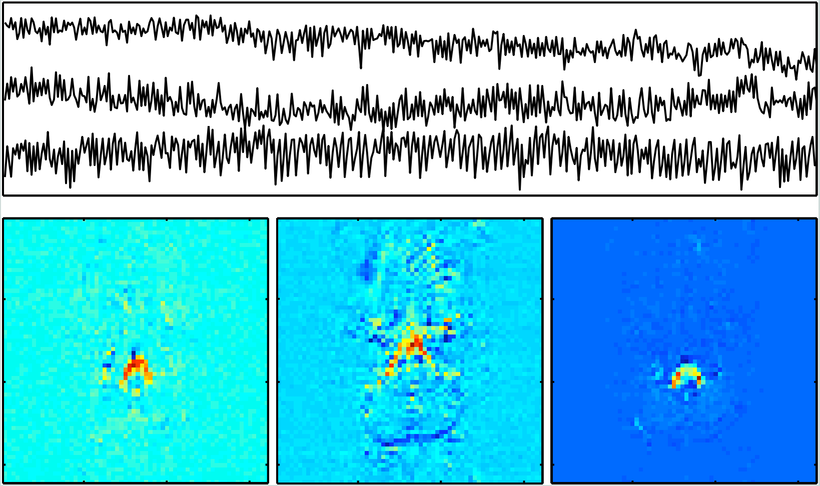
EPI ghost (phantom)



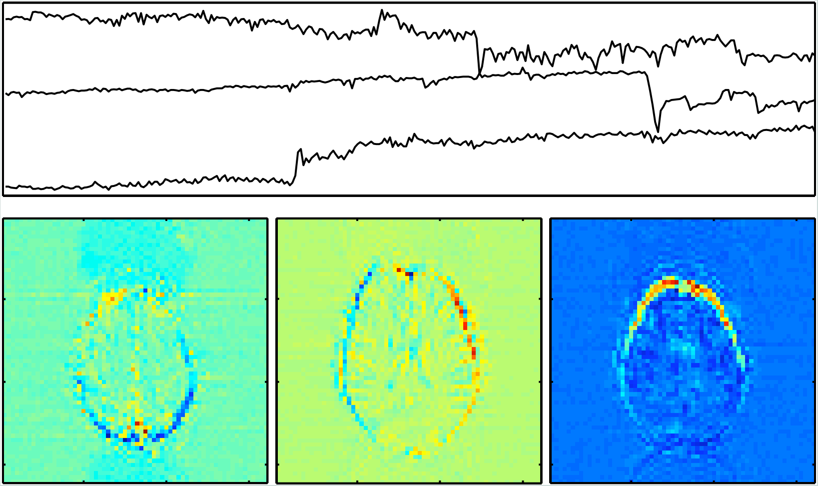
EPI ghost (in human subject with head motion)



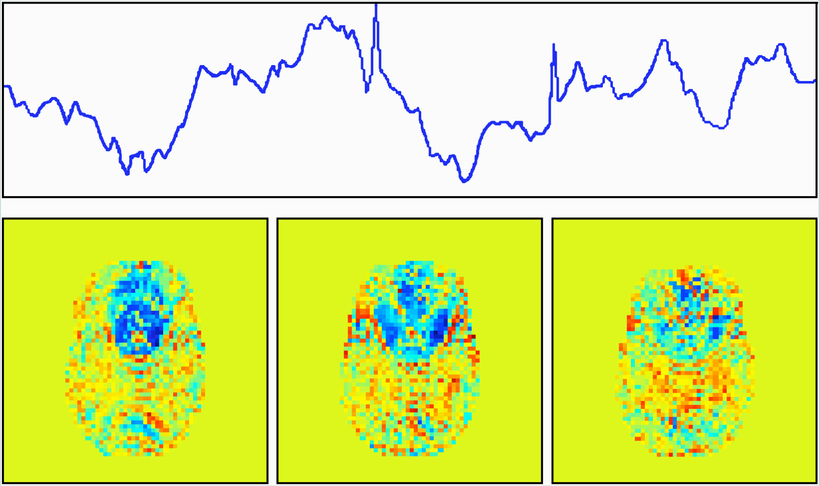
high-frequency noise (mainly in ventricles)



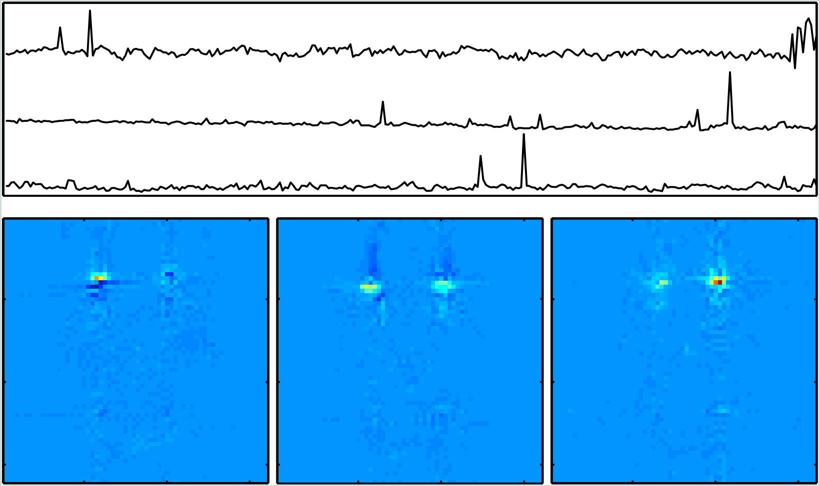
head motion



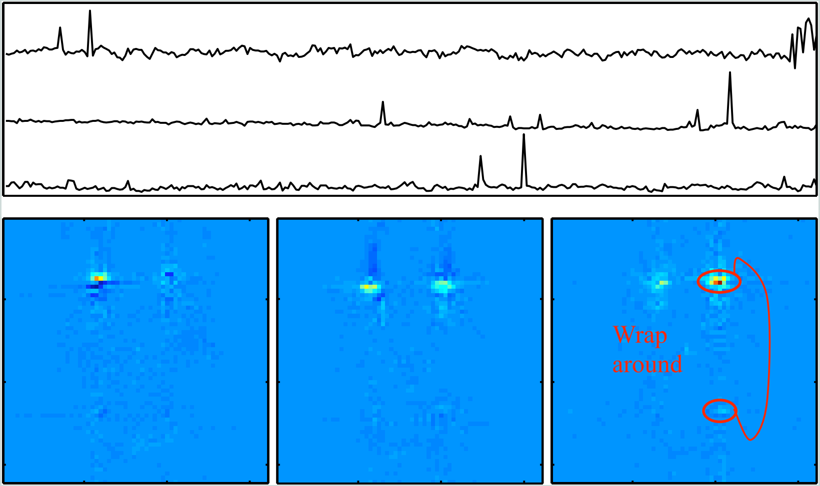
B_0 field inhomogeneity



eye-related artefacts (eyeblink, eyball motion ?)

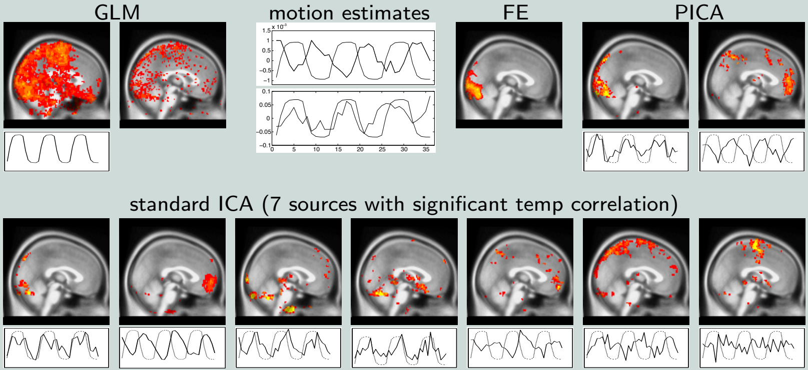


wrap-around in FOV due to interaction with the EPI ghost



- Data from McGonigle *et.al.*, *NeuroImage*, 11:708–734 (2000)
- 33 sessions under visual stimulation - some data was discarded

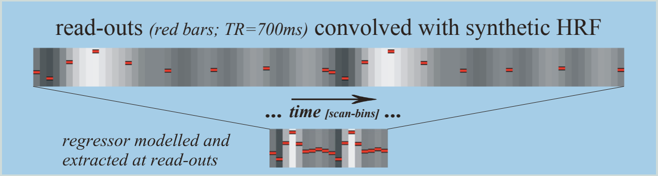
stimulus-correlated motion



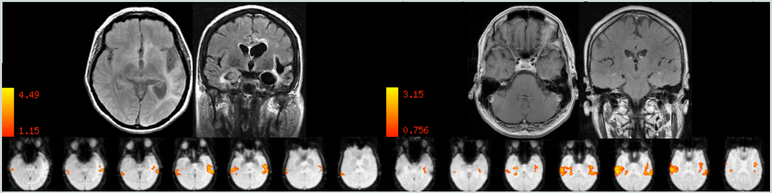
Example: 'Scanning for the scanner'

utilise the effective fluctuation of the EPI sequence noise to scan for residual auditory responses in patients Bartsch *et.al.*, HBM (2004)

MODIFIED EPI GRADIENT-TRAIN WITH READ-OUT OMISSIONS & EXPECTED AUDITORY BOLD SIGNAL MODULATIONS:



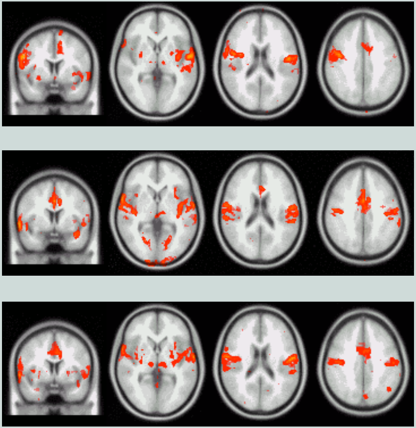
Herpes-simplex-encephalitis postoperative deafness



PICA on resting data

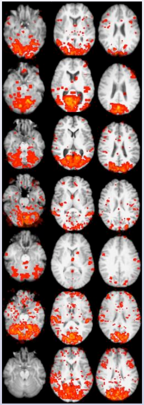
- perform ICA on null data and compare spatial maps between subjects/scans
- ICA maps depict spatially localised and temporally coherent signal changes that are confounding effects for the GLM

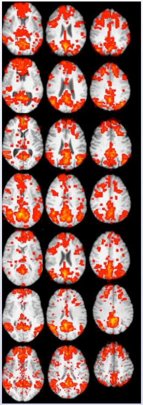
Example: 1 subject, 3 sessions



RSN classification (7 normals): 4 consistent maps

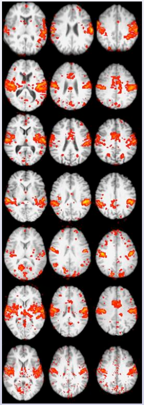
visual cortex
medial occipital

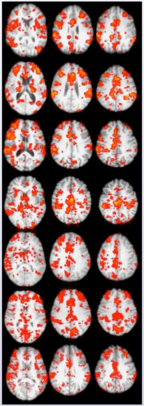




visual, lateral
occipital, medial
parietal

primary and secondary
sensory, anterior
insula, pain





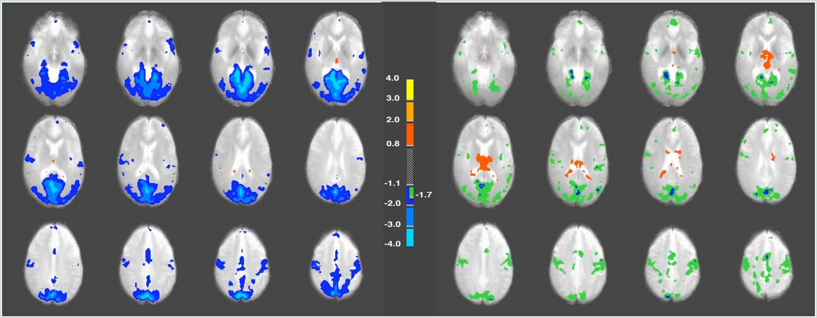
posterior parietal,
prefrontal: attention,
working memory

DeLuca *et.al.*, ISMRM (2004)

Simultaneous EEG/FMRI

- record single bipolar EEG channel recording during FMRI
- estimate subject specific alpha power und use for GLM

PICA vs. GLM



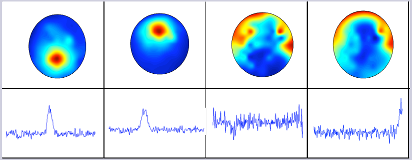
Goldman & Cohen, HBM (2003)

Simultaneous EEG/FMRI

task

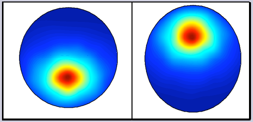
ERP data

temporal ICA



FMRI data

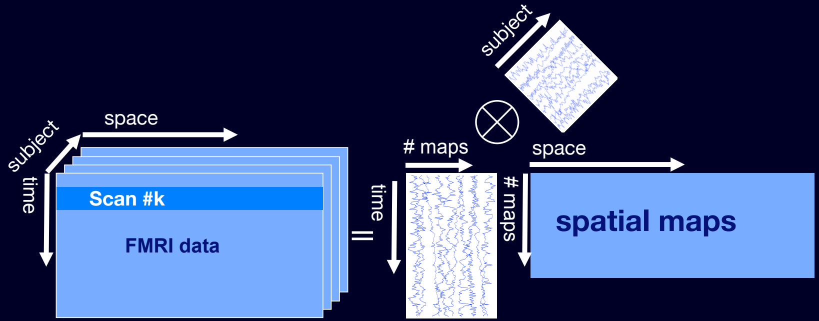
spatial ICA and project maps to scalp (need forward model only)



match ERP and FMRI sources using the scalp spatial maps

Loftus *et.al.*, HBM (2003)

Tensor-PICA: multi-way generalisation of PICA



$$x_{ijk} = \sum_r^R \mathbf{a}_{ir} \times \mathbf{b}_{jr} \times \mathbf{c}_{kr} + \eta_{ijk}$$

Data are represented as a 3D array and decomposed into factor matrices **A**, **B** and **C**.

PARAFAC

- as a symmetric least-square problem this is known as *PARAFAC (Parallel Factor Analysis)* and can be solved using Alternating Least Squares (ALS), i.e. by iterating least-squares solutions for

$$\begin{aligned}
\mathbf{X}_{i..} &= \mathbf{B}\mathrm{diag}(\mathbf{a}_i)\mathbf{C}^t + E_{i..} \quad \forall i \\
\mathbf{X}_{.j.} &= \mathbf{C}\mathrm{diag}(\mathbf{b}_j)\mathbf{A}^t + E_{.j.} \quad \forall j \\
\mathbf{X}_{..k} &= \mathbf{A}\mathrm{diag}(\mathbf{c}_k)\mathbf{B}^t + E_{..k} \quad \forall k
\end{aligned}$$

- requires *system variation* (no co-linearity in $\mathbf{A}, \mathbf{B}$ or $\mathbf{C}$)
- treats all modes the same

Tensor-PICA: estimation

- rewrite:

$$\mathbf{X}_{IK \times J} = (\mathbf{C}|\otimes|\mathbf{A})\mathbf{B}^t + \mathbf{E}$$

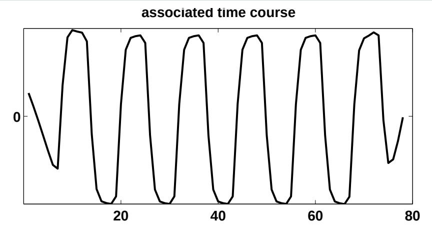
- can be treated as a 2-stage estimation problem:
  1. PICA estimation of $\mathbf{B}$ from $\mathbf{X}_{IK \times J}$ by estimating $\mathbf{M}$ as the mixing matrix
  2. rank-1 Eigen-decomposition of each column $\mathbf{M}^{(r)}$, reshaped into a $I \times K$ matrix, in order to find the underlying factor matrices such that $\mathbf{M} = (\mathbf{C}|\otimes|\mathbf{A})$
- Jointly estimates R modes which describe signal characteristics in the temporal, spatial and subject/session domain.

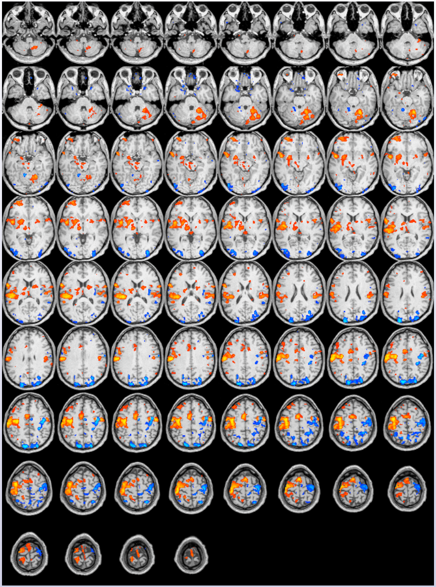
Example

10 sessions under motor paradigm (right index finger tapping)

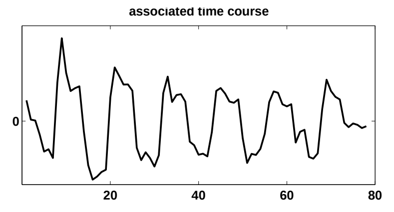
McGonigle *et.al*, NeuroImage 11:708–735, 2000

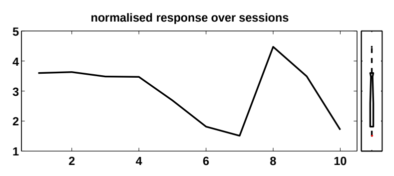
Group-level mixed-effects results



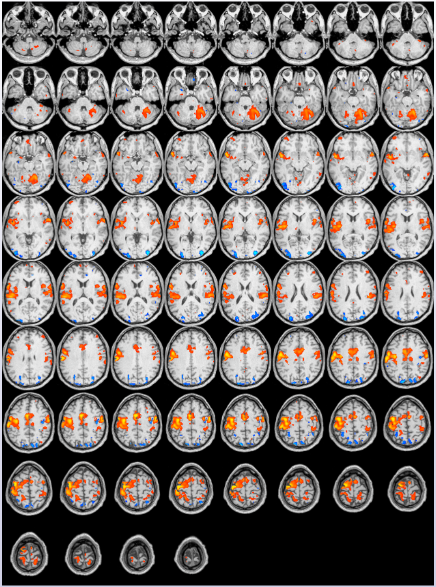


Tensor PICA: primary activation

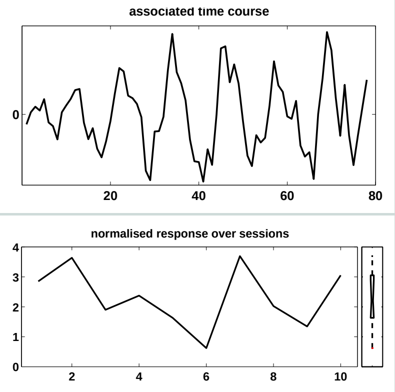




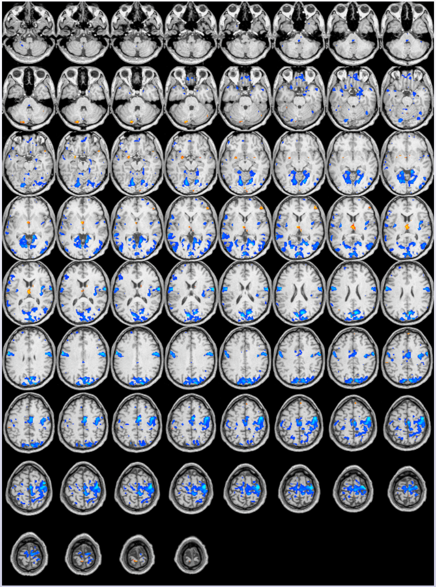
contra-lateral primary motor/sensory; SMA; bi-lateral secondary somatosensory; anterior lobe of cerebellum



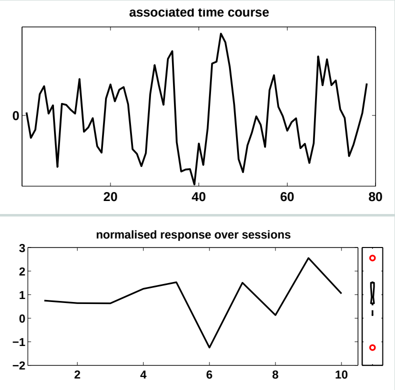
Tensor PICA: primary 'de-activation'



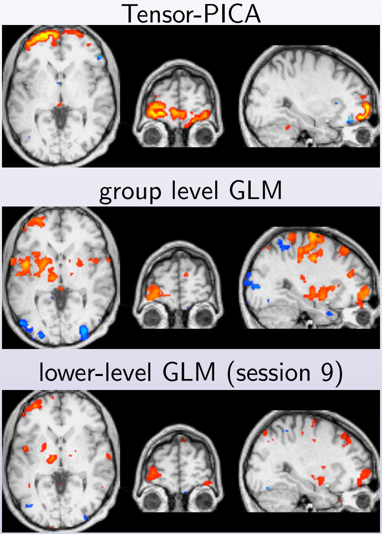
ipsi-lateral primary motor/somatosensory



Tensor PICA: artefacts



stimulus-correlated motion
(strong in 2 sessions)



Conclusions

- exploring your data is important in order to get a better understanding
- don't just look at post-thresholded stats images!
- model-free analysis is complementary to GLM - make use of it
- PCA/ICA techniques are easy to use - results are often less easy to interpret, though
- probabilistic ICA can produce plausible activation maps and associated time-courses

Acknowledgements

MELODIC software
available as part of FSL
http://www.fmrib.ox.ac.uk/fsl

- Caitlinn Loftus (HFI Melbourne)
- Andreas Bartsch (BJMU Würzburg)
- Robin Goldmann (Columbia Univ.) & Mark Cohen (UCLA)
- Dave McGonigle (CNRS Paris)
- Pedro Valdés-Sosa (CNC Havana)
- Stephen Smith, Mark Jenkinson, Mark Woolrich, Marilena DeLuca (FMRIB)
- GlaxoSmithKline