Resting state fMRI and ICA

- Introduction to resting state
- Independent Component Analysis
- Single-subject ICA
- Multi-subject ICA
- Dual regression

Resting state methods

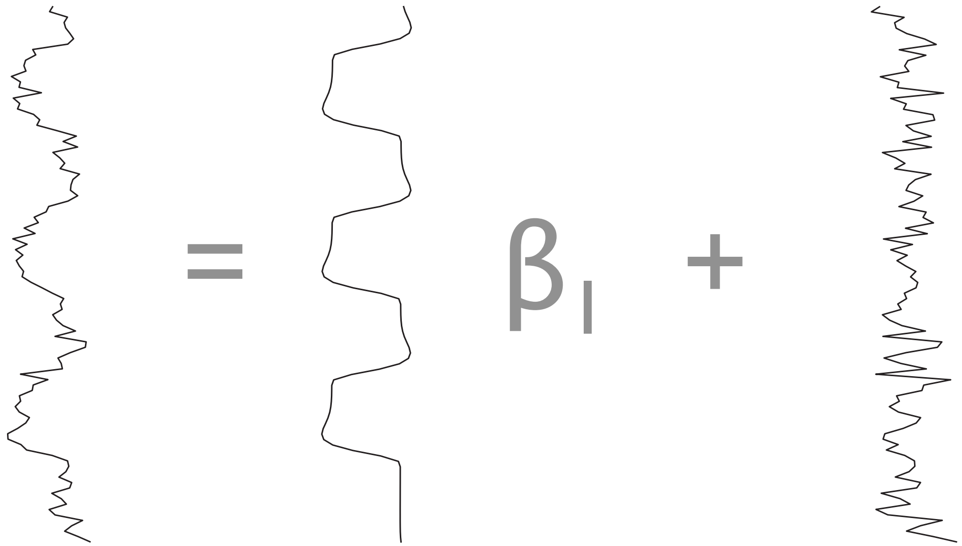
ICA

- Multivariate voxel-based approach
- Finds interesting structure in the data
- Exploratory “model-free” method
- Spatial approach

Network modelling

- Node-based approach (first need to parcellate the brain into functional regions)
- Map connections between specific brain regions (connectomics)
- Temporal approach

Model-based (GLM) analysis



- model each measured time-series as a linear combination of signal and noise
- If the design matrix does not capture every signal, we typically get wrong inferences!

Data Analysis

Confirmatory

- "How well does my model fit to the data?"

Problem → Data →

Model → Analysis

→ Results

- results depend on the model

Exploratory

- "Is there anything interesting in the data?"

Problem → Data →

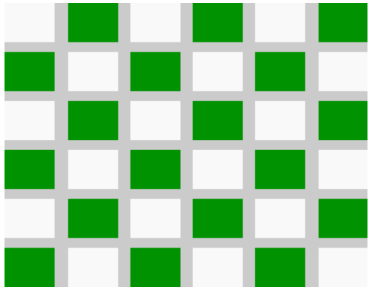
Analysis → Model

→ Results

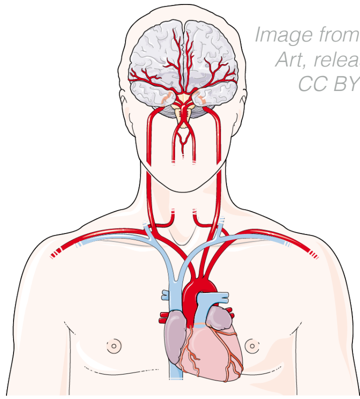
- can give unexpected results

FMRI inferential path

Experiment



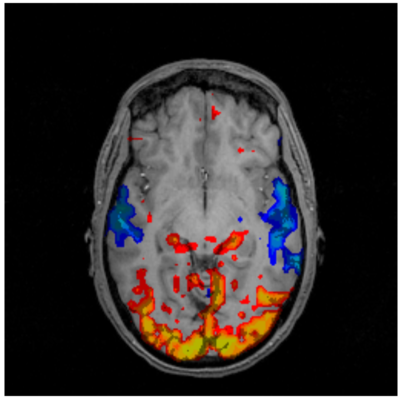
Physiology



MR Physics

Analysis

Interpretation of final results

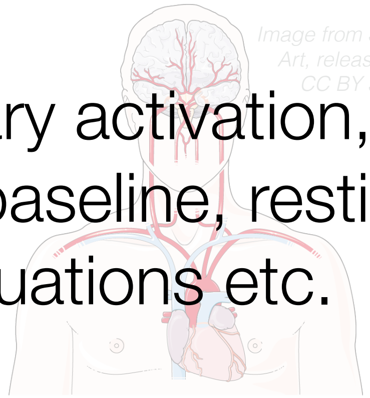


Variability in FMRI

Experiment

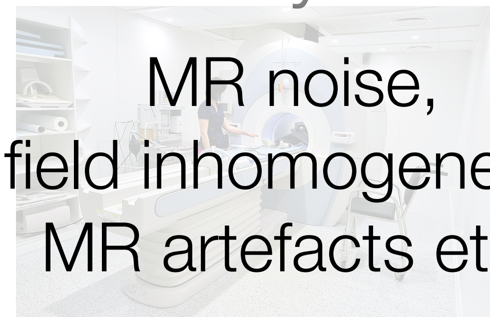
suboptimal event timing, inefficient design, etc.

Physiology



secondary activation, ill-defined baseline, resting-fluctuations etc.

MR Physics

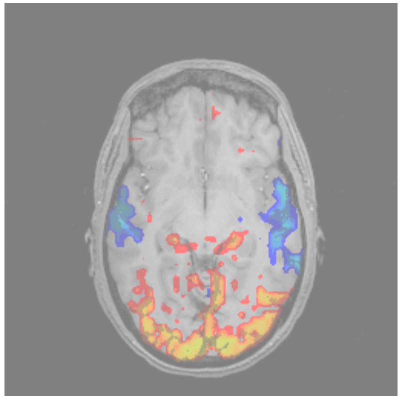


MR noise, field inhomogeneity, MR artefacts etc.

Analysis

filtering & sampling artefacts, design misspecification, stats & thresholding issues etc.

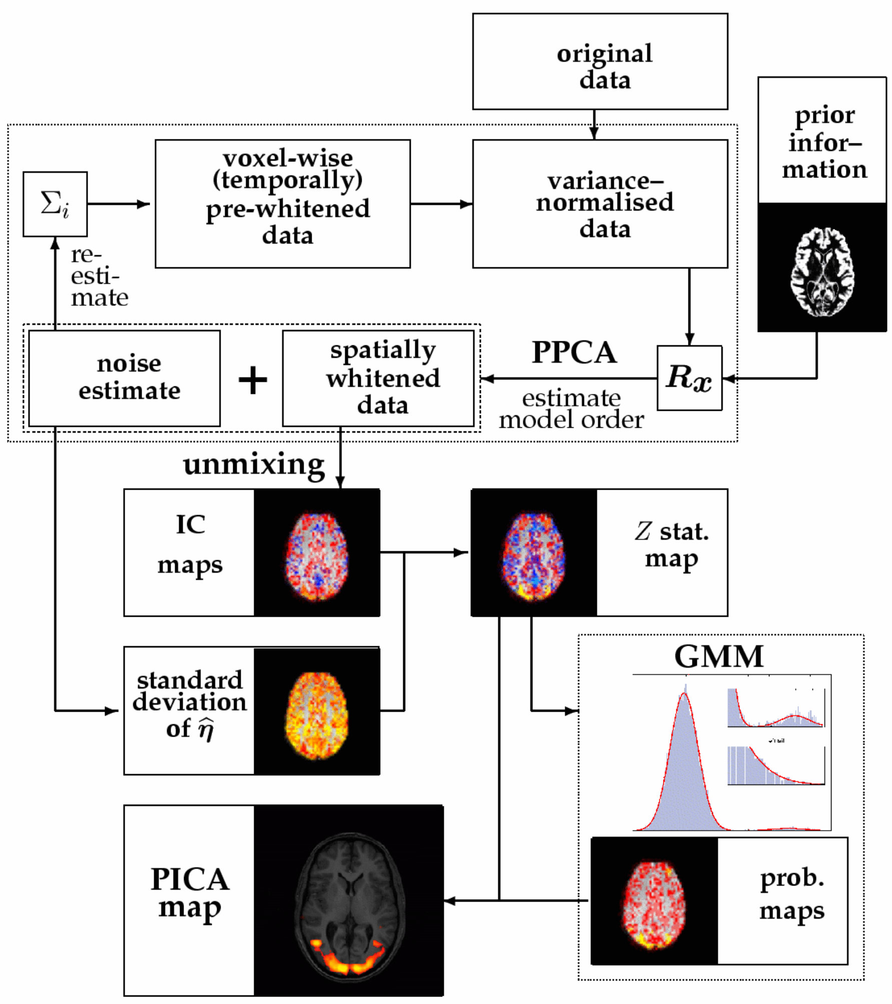
Interpretation of final results



Model-free?

There is no explicit time-series model of assumed ‘activity’

Model-free?



$$Y^i = S^i A^i + E^i, \quad \text{where} \quad E^i_{\cdot j} \sim \mathcal{N}(0, \sigma^2_Y \boldsymbol{I})$$

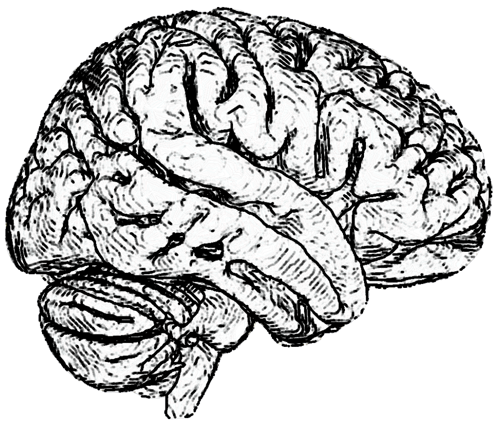
There is an underlying mathematical (generative) model

Decomposition techniques

- try to 'explain' / represent the data
  - by calculating quantities that summarise the data
  - by extracting underlying 'hidden' features that are 'interesting'
- differ in what is considered 'interesting'
  - are localised in time and/or space (Clustering)
  - explain observed data variance (PCA, FDA, FA)
  - are maximally independent (ICA)

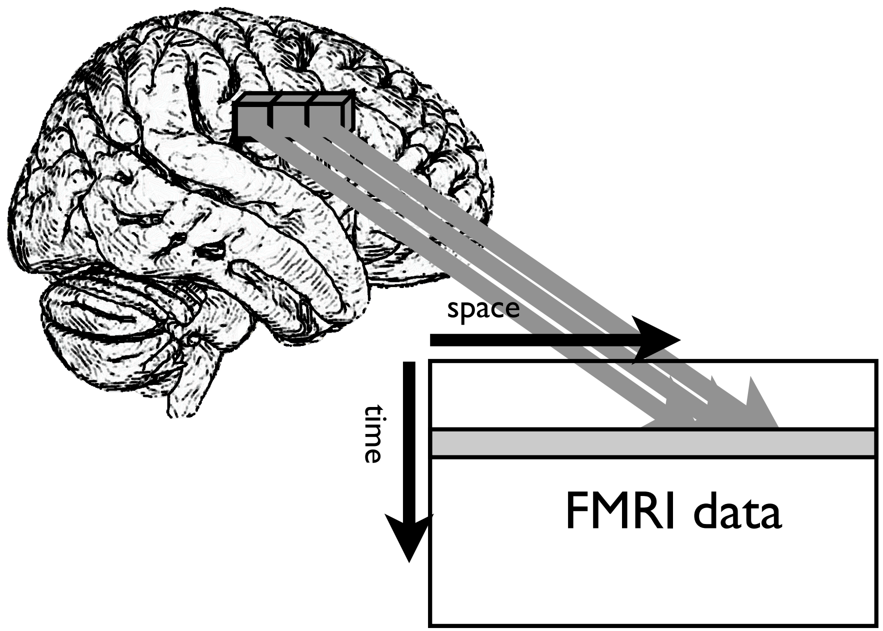
Melodic

multivariate linear decomposition:



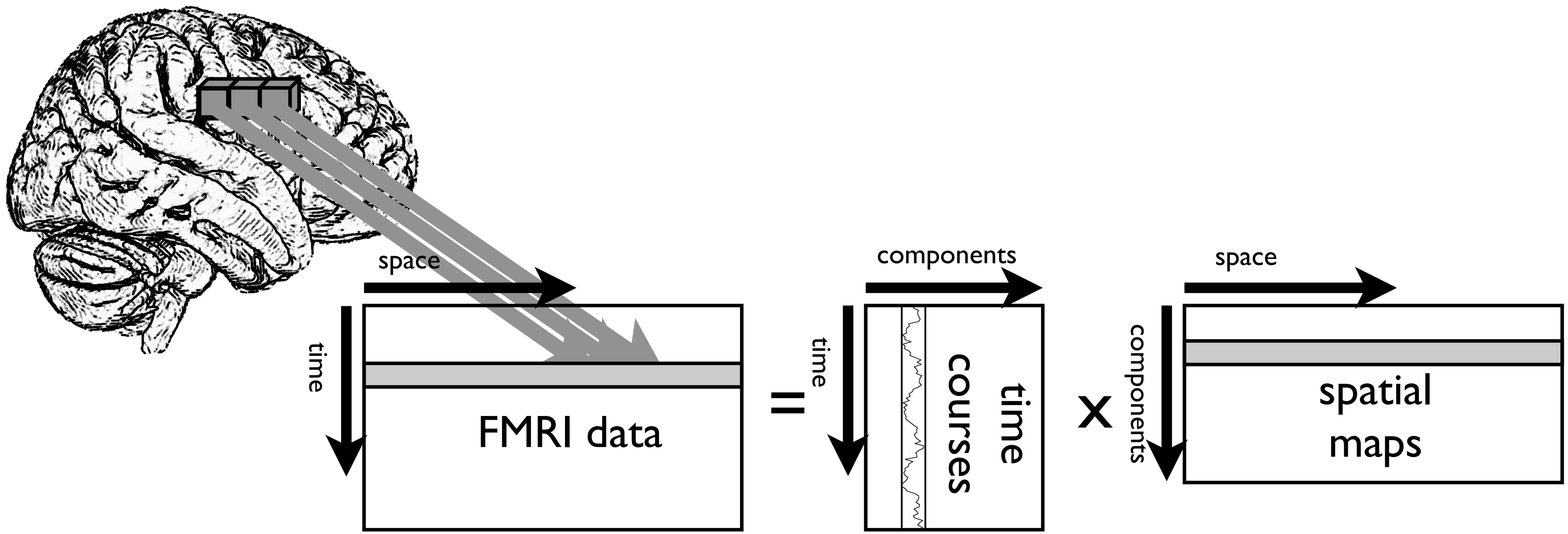
Melodic

multivariate linear decomposition:



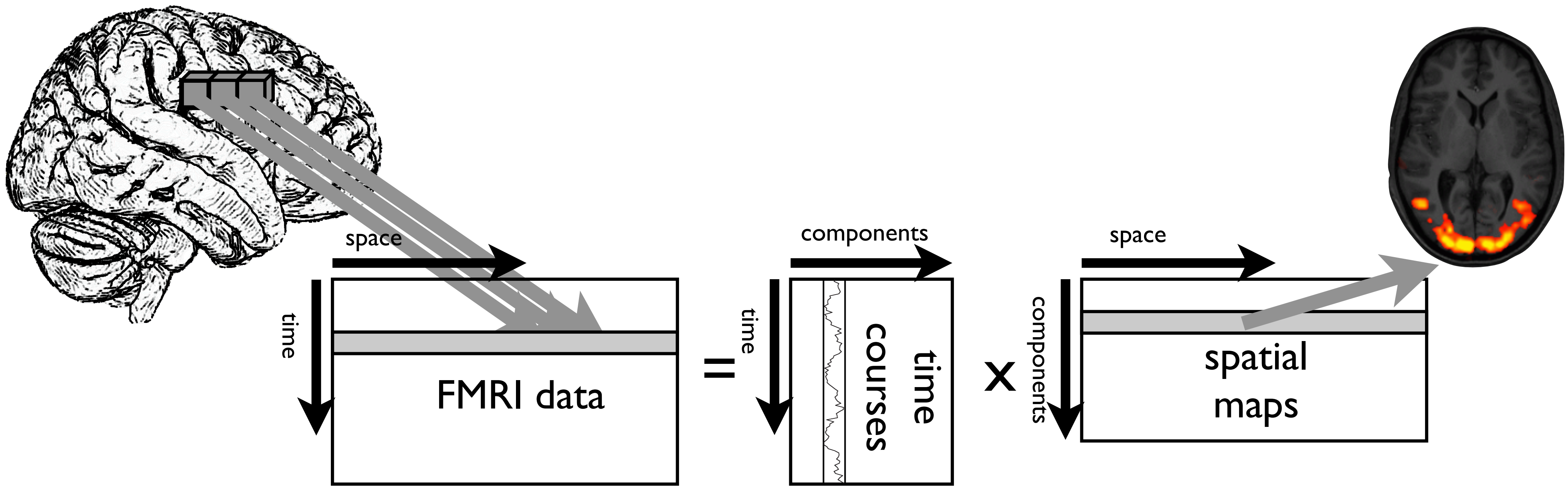
Melodic

multivariate linear decomposition:



Melodic

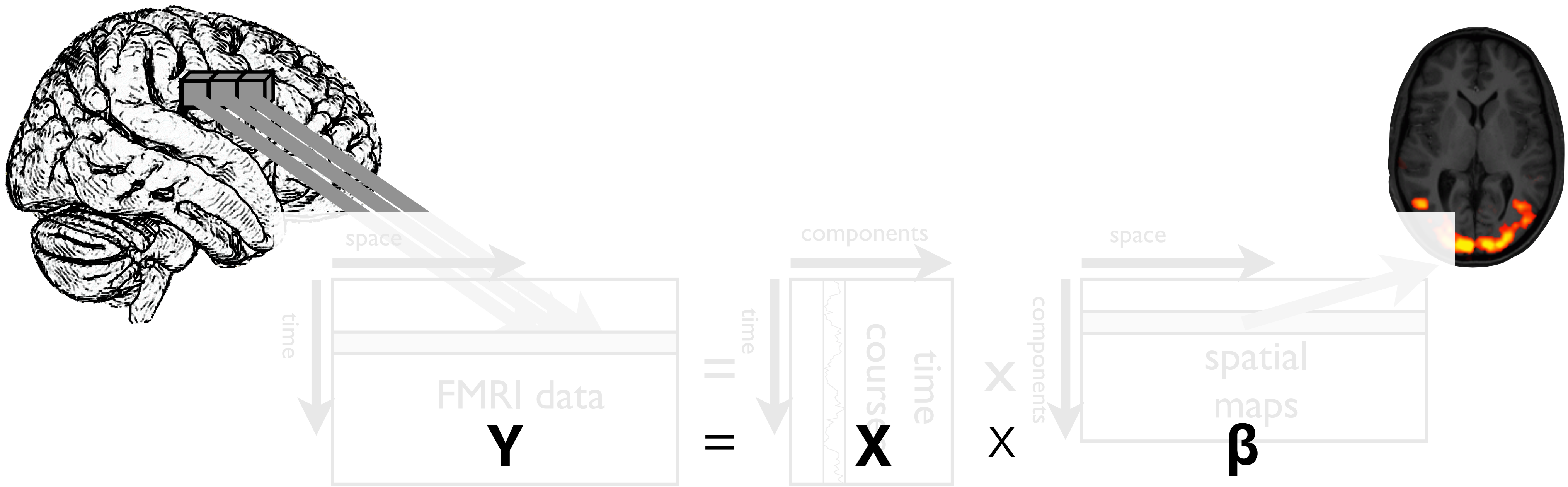
multivariate linear decomposition:



Data is represented as a 2D matrix and decomposed into components

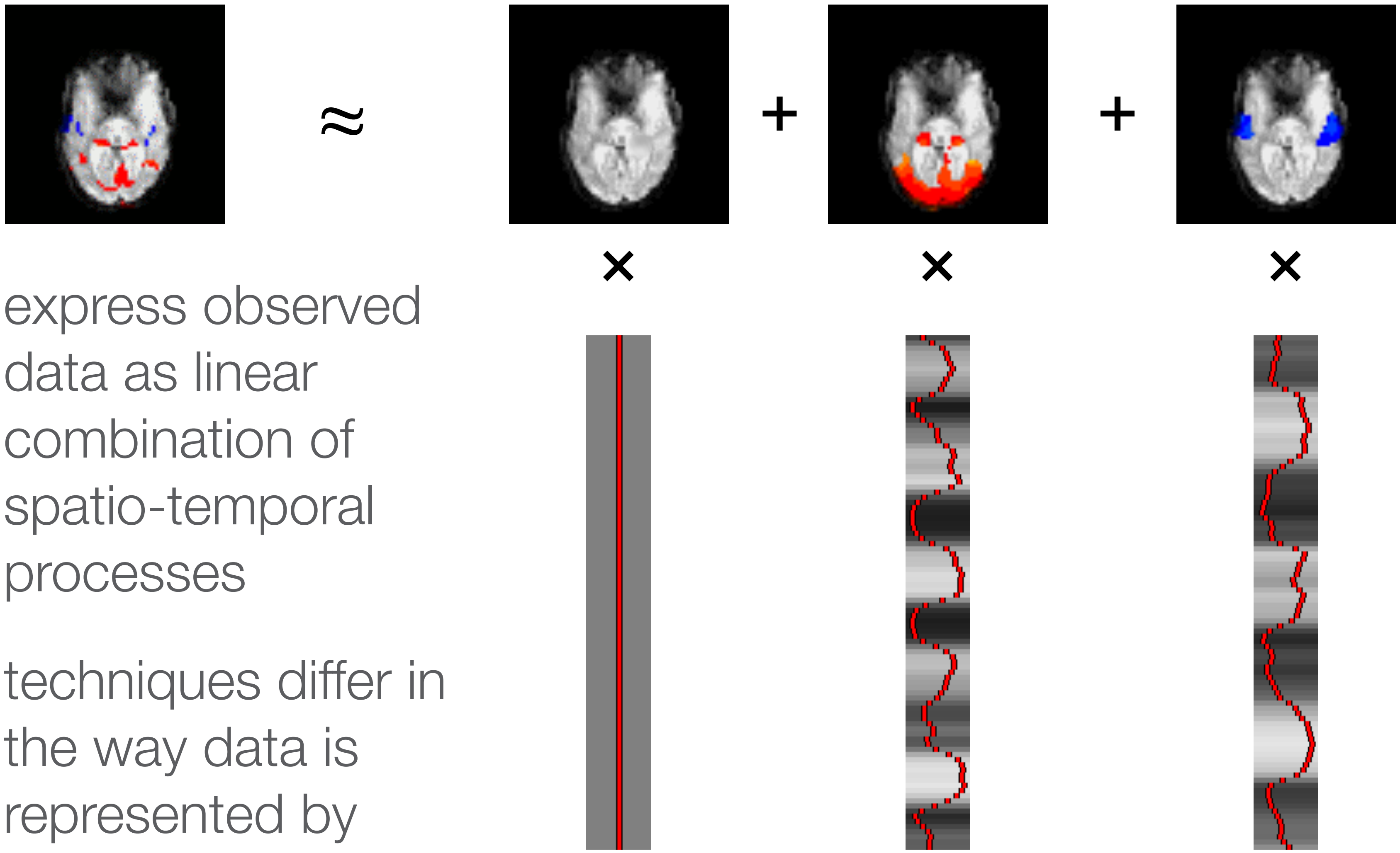
Melodic

multivariate linear decomposition:



Data is represented as a 2D matrix and decomposed into components

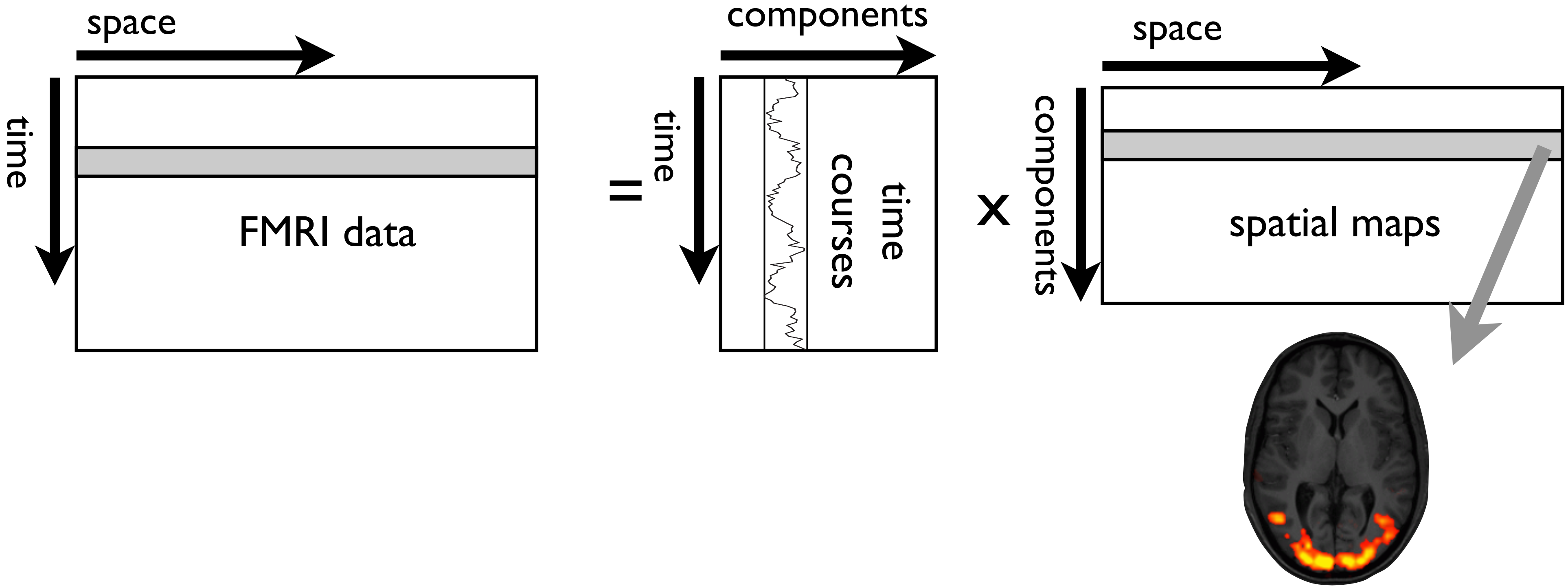
What are components?



- express observed data as linear combination of spatio-temporal processes

- techniques differ in the way data is represented by components

Spatial ICA for FMRI



– data is decomposed into a set of spatially independent maps and a set of time-courses

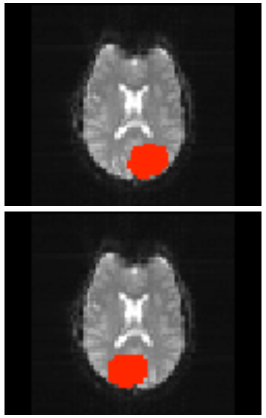
McKeown et al.
HBM 1998

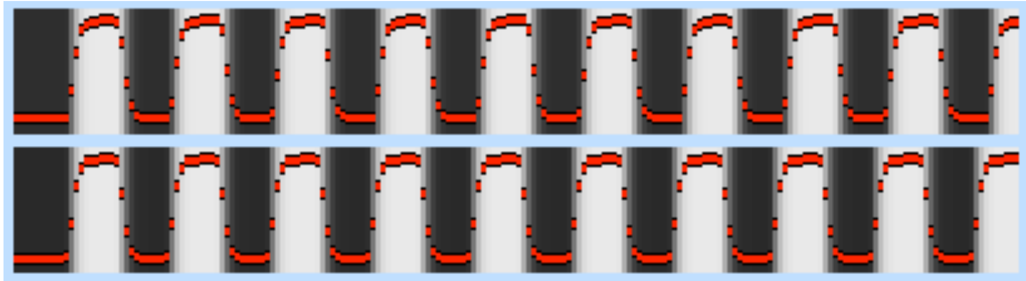
Independence

PCA vs. ICA ?

Simulated Data

(2 components, slightly different timecourses)

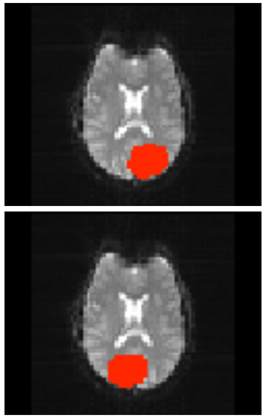


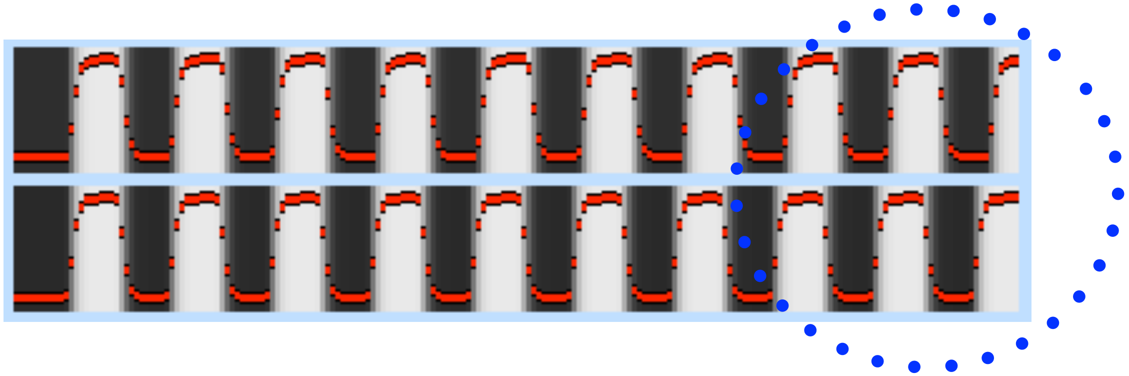


PCA vs. ICA ?

Simulated Data

(2 components, slightly different timecourses)

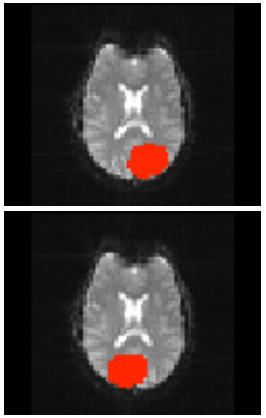


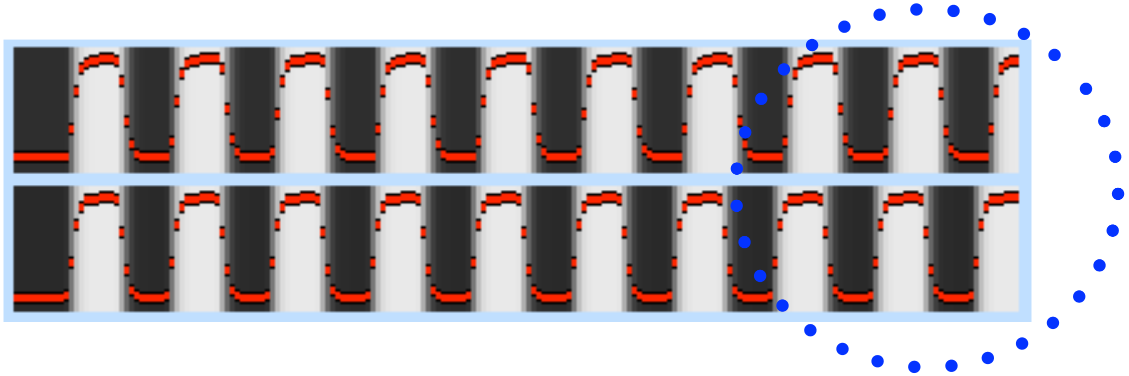


PCA vs. ICA ?

Simulated Data

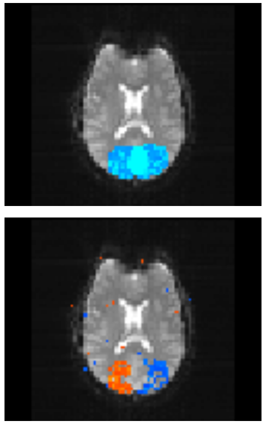
(2 components, slightly different timecourses)

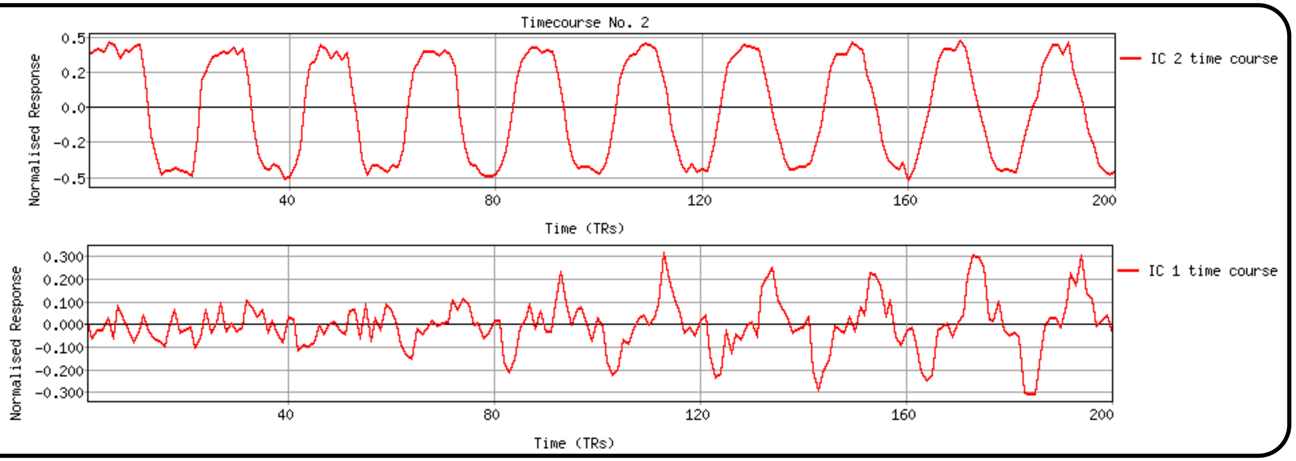




PCA

- Timecourses orthogonal
- Spatial maps and timecourses "wrong"

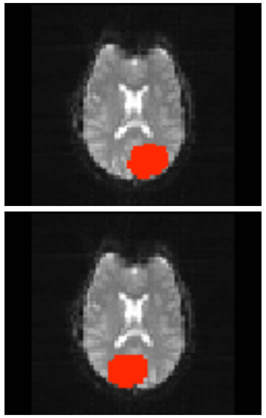


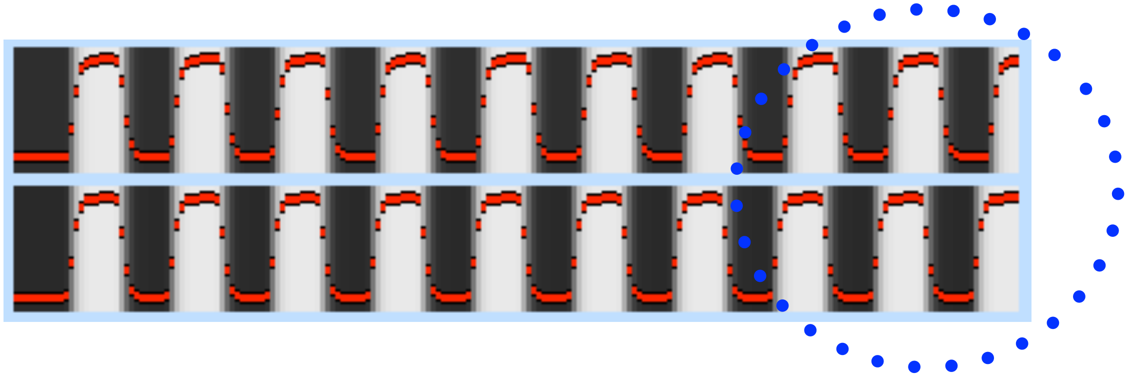


PCA vs. ICA ?

Simulated Data

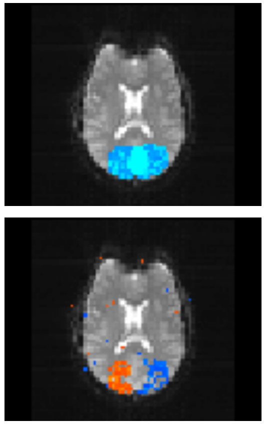
(2 components, slightly different timecourses)

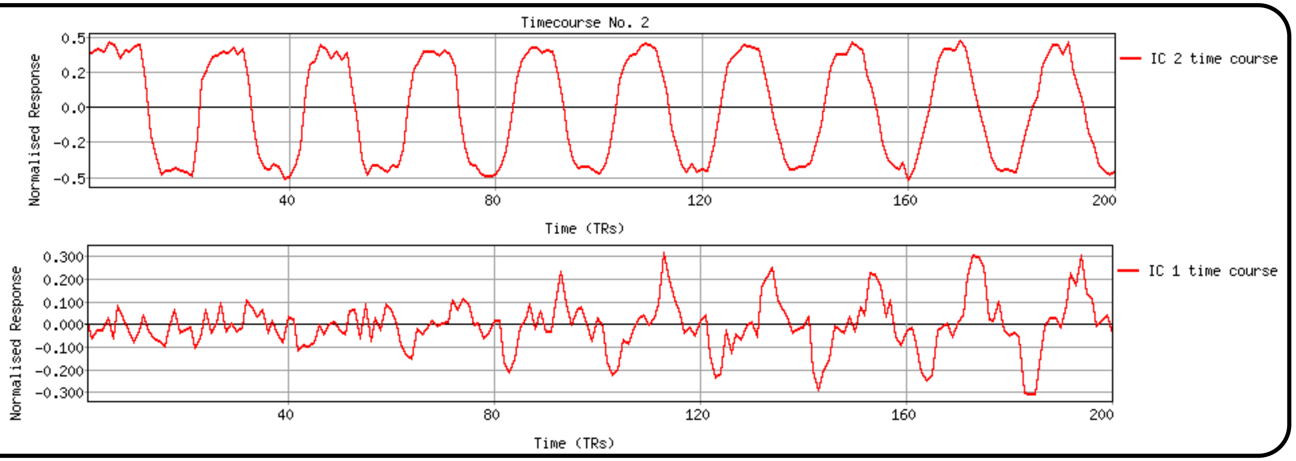




PCA

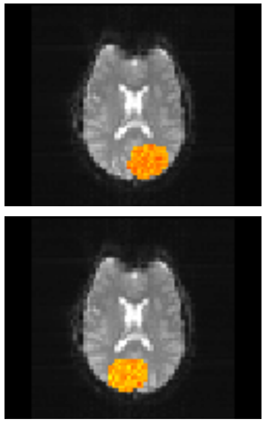
- Timecourses orthogonal
- Spatial maps and timecourses “wrong”

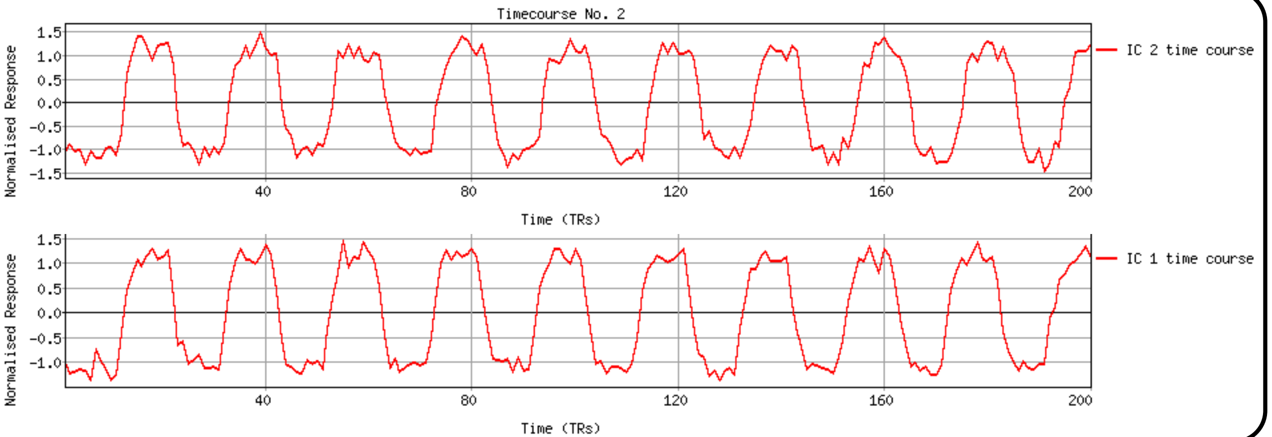




ICA

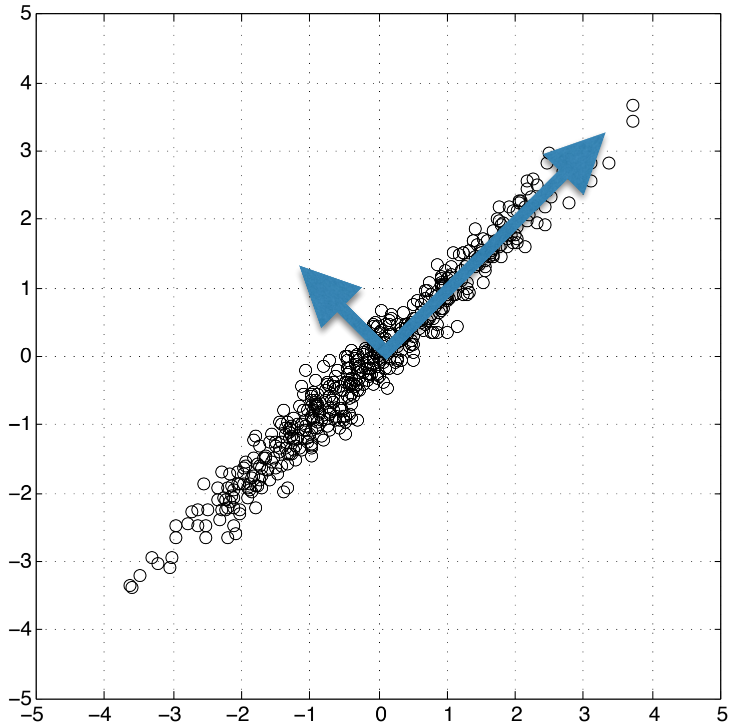
- Timecourses non-co-linear
- Spatial maps and timecourses “right”





PCA vs. ICA

- PCA finds projections of maximum amount of variance in Gaussian data (uses 2nd order statistics only)

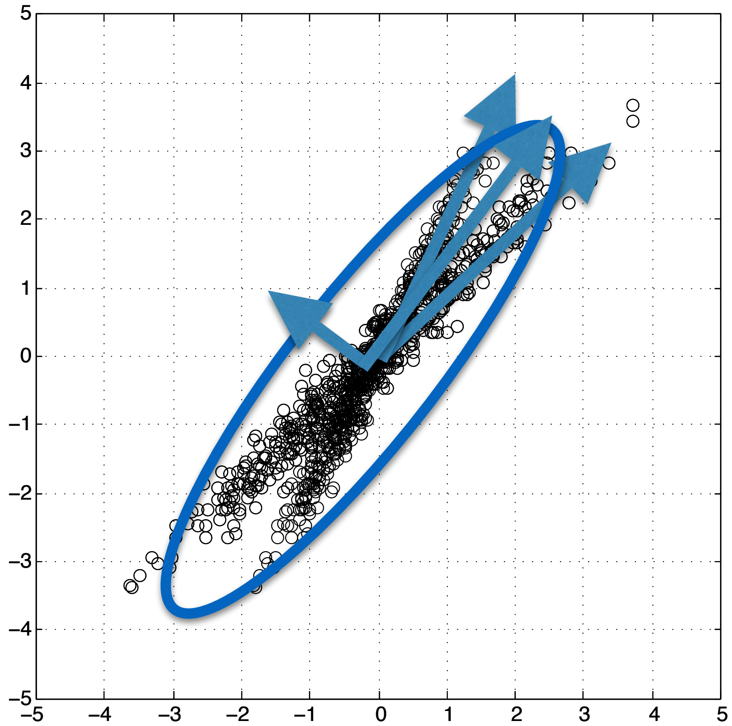


Gaussian data

PCA vs. ICA

- PCA finds projections of maximum amount of variance in Gaussian data (uses 2nd order statistics only)

- Independent Component Analysis (ICA) finds projections of maximal independence in non-Gaussian data (using higher-order statistics)

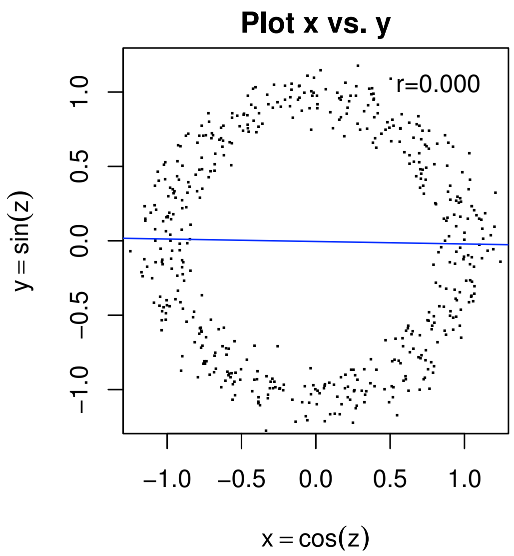


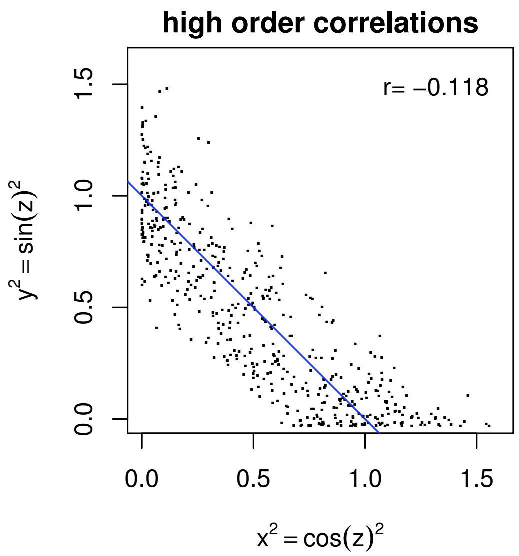
non-Gaussian data

Correlation vs. independence

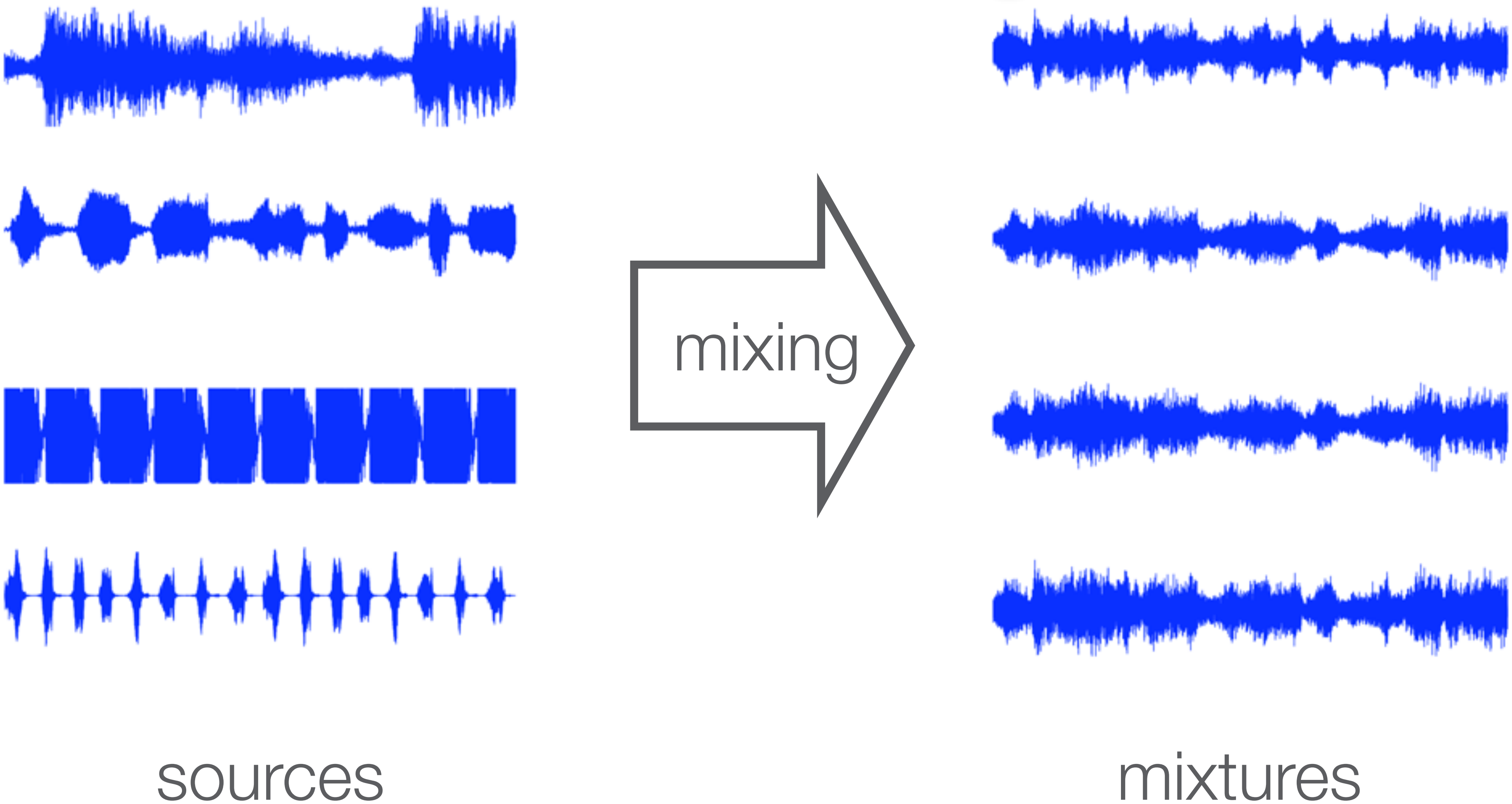
- de-correlated signals can still be dependent
- higher-order statistics (beyond mean and variance) can reveal these dependencies

Stone et al. 2002

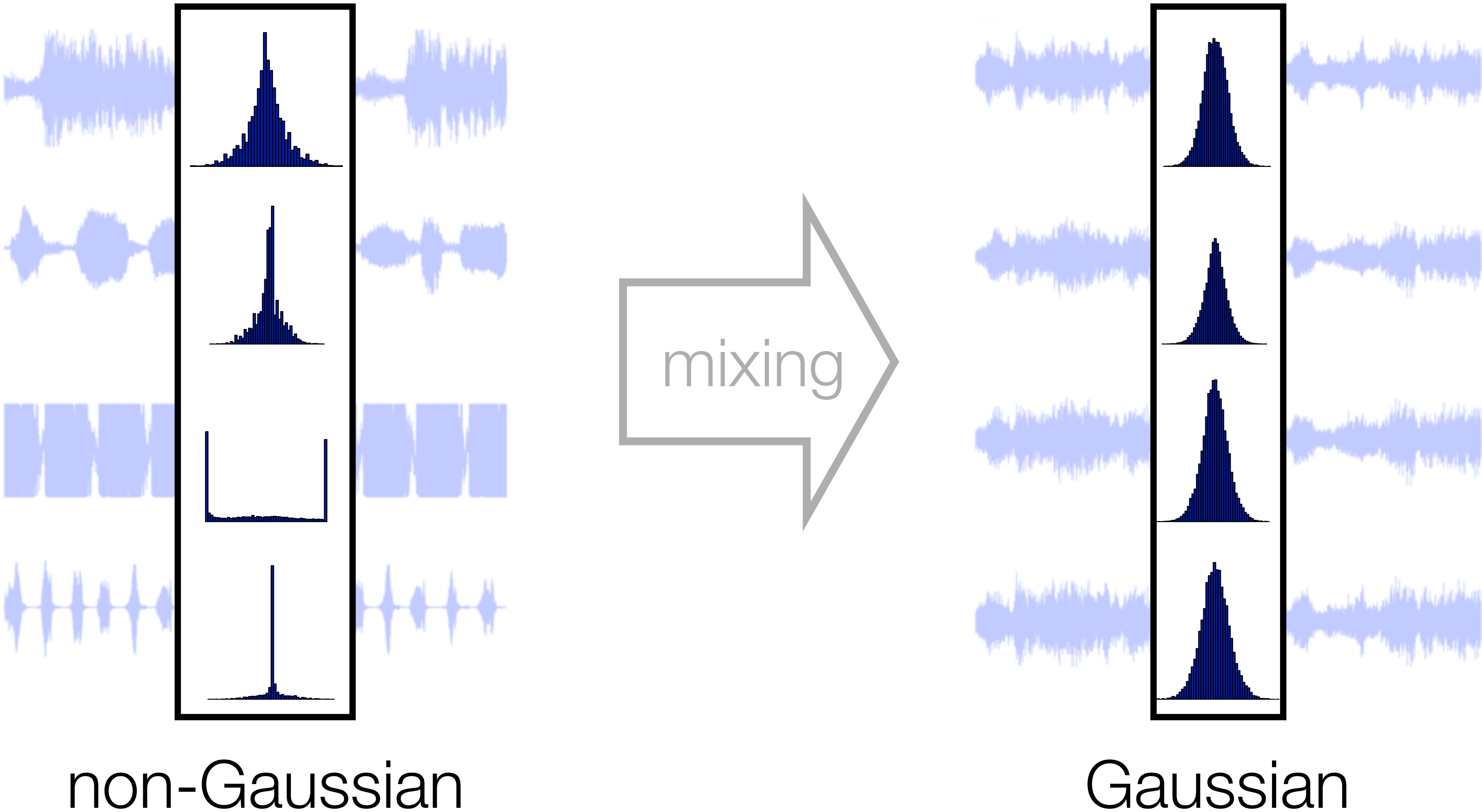




Non-Gaussianity



Non-Gaussianity



ICA estimation

- Random mixing results in more Gaussian-shaped PDFs (Central Limit Theorem)
- conversely:

  if mixing matrix produces less Gaussian-shaped PDFs this is unlikely to be a random result

  ➡ measure non-Gaussianity

- can use neg-entropy as a measure of non-Gaussianity

Hyvärinen & Oja 1997

ICA estimation

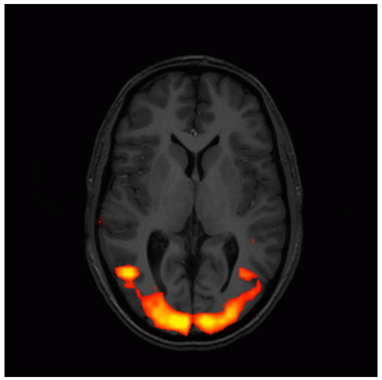
- need to find an unmixing matrix such that the dependency between estimated sources is minimised
- need (i) a contrast (objective/cost) function to drive the unmixing which measures statistical independence and (ii) an optimisation technique:
- kurtosis or cumulants & gradient descent (Jade)
- maximum entropy & gradient descent (Infomax)
- neg-entropy & fixed point iteration (FastICA)

Overfitting & thresholding

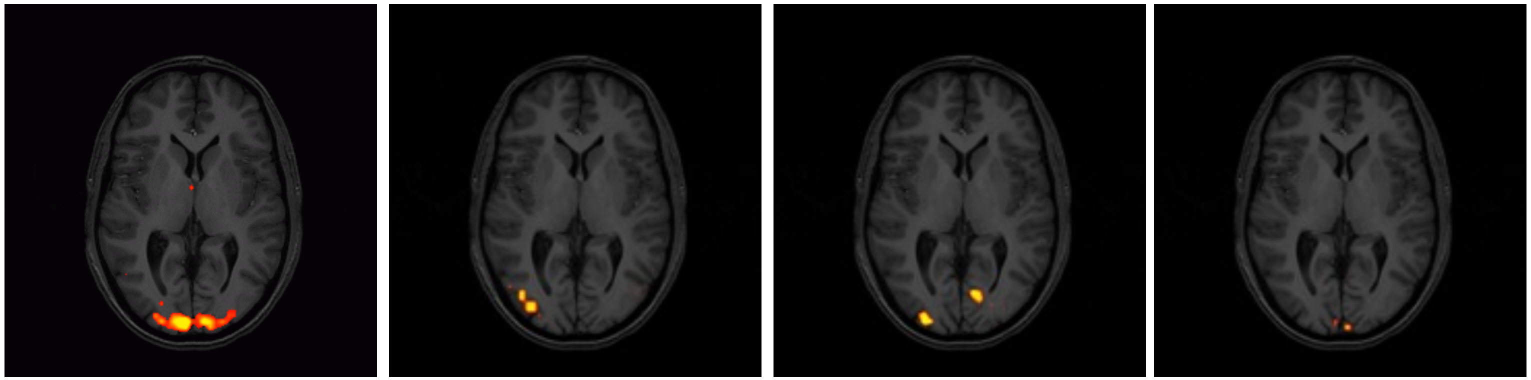
The ‘overfitting’ problem

fitting a noise-free model to noisy observations:

- no control over signal vs. noise (non-interpretable results)
- statistical significance testing not possible



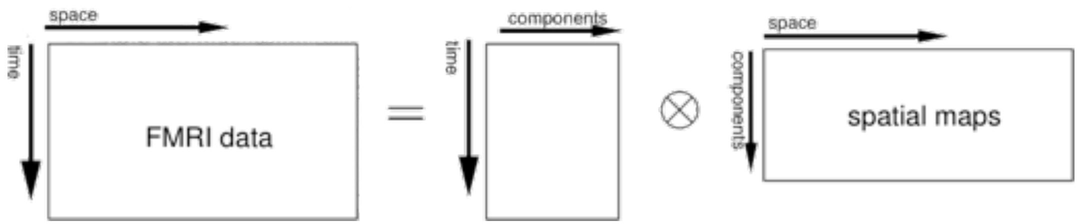
GLM analysis



standard ICA (unconstrained)

Probabilistic ICA model

statistical "latent variables" model: we observe linear mixtures of hidden sources in the presence of Gaussian noise



+ noise

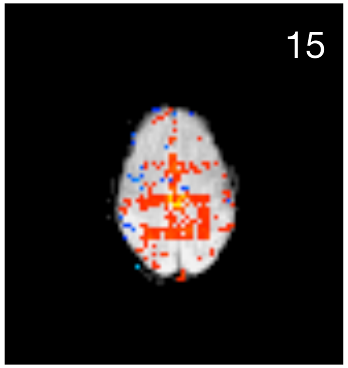
$$\mathbf{Y} = \mathbf{X} \quad \boldsymbol{\beta} + \mathbf{E}$$

Issues:

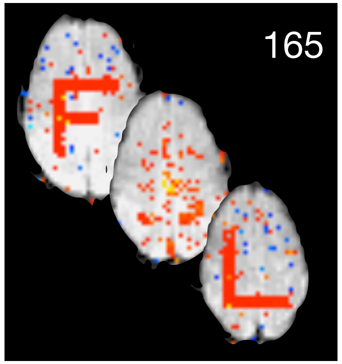
- Model Order Selection: how many components?
- Inference: how to threshold ICs?

Model Order Selection

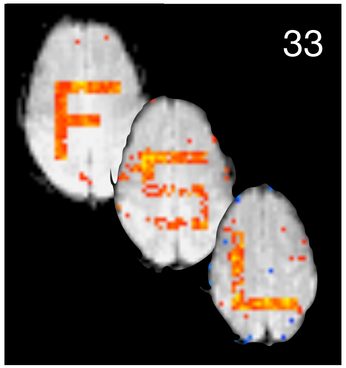
‘How many components’?



under-fitting: the amount of explained data variance is insufficient to obtain good estimates of the signals

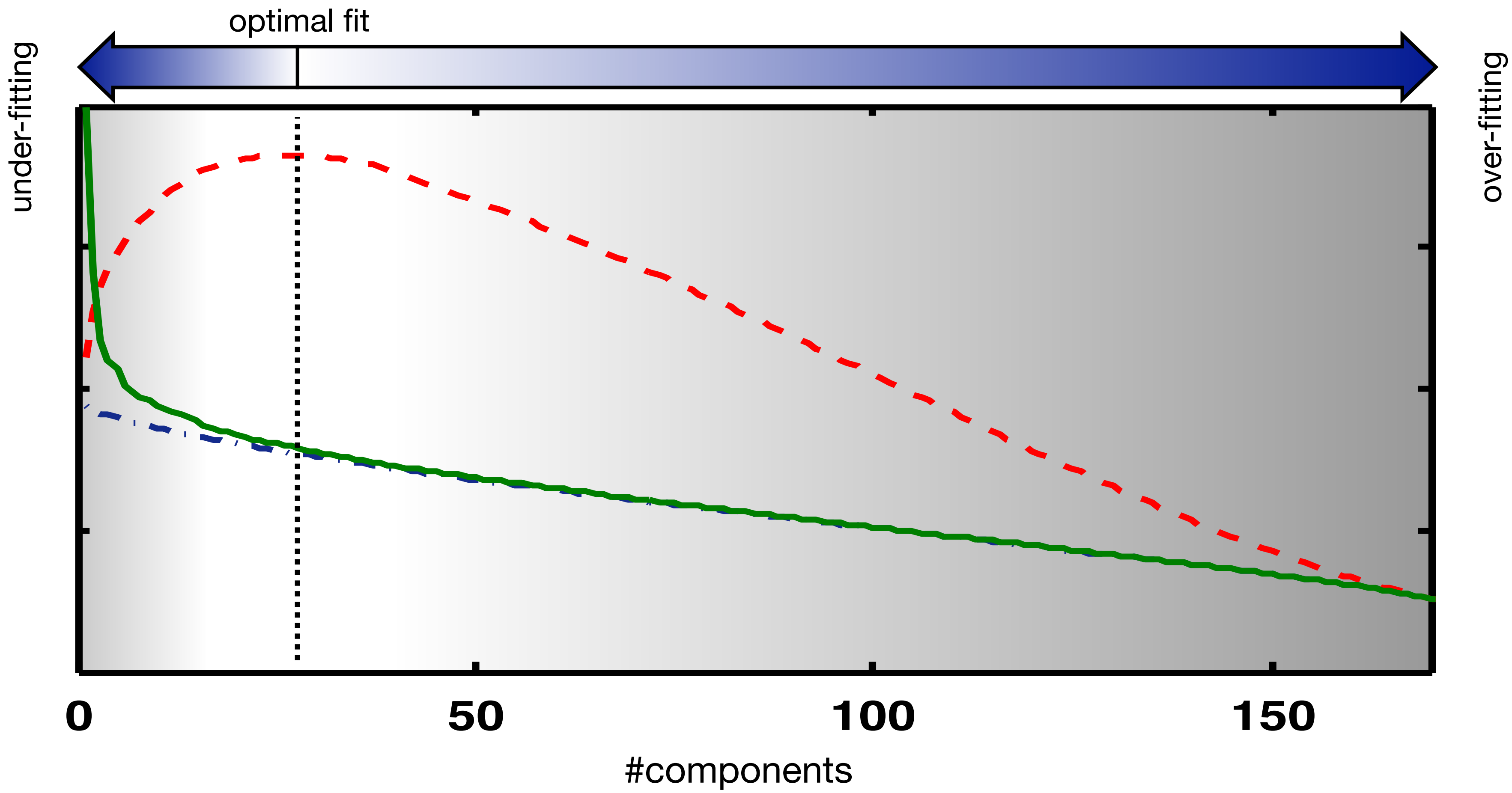


over-fitting: the inclusion of too many components leads to fragmentation of signal across multiple component maps, reducing the ability to identify the signals of interest



optimal fitting: the amount of explained data variance is sufficient to obtain good estimates of the signals while preventing further splits into spurious components

Model Order Selection

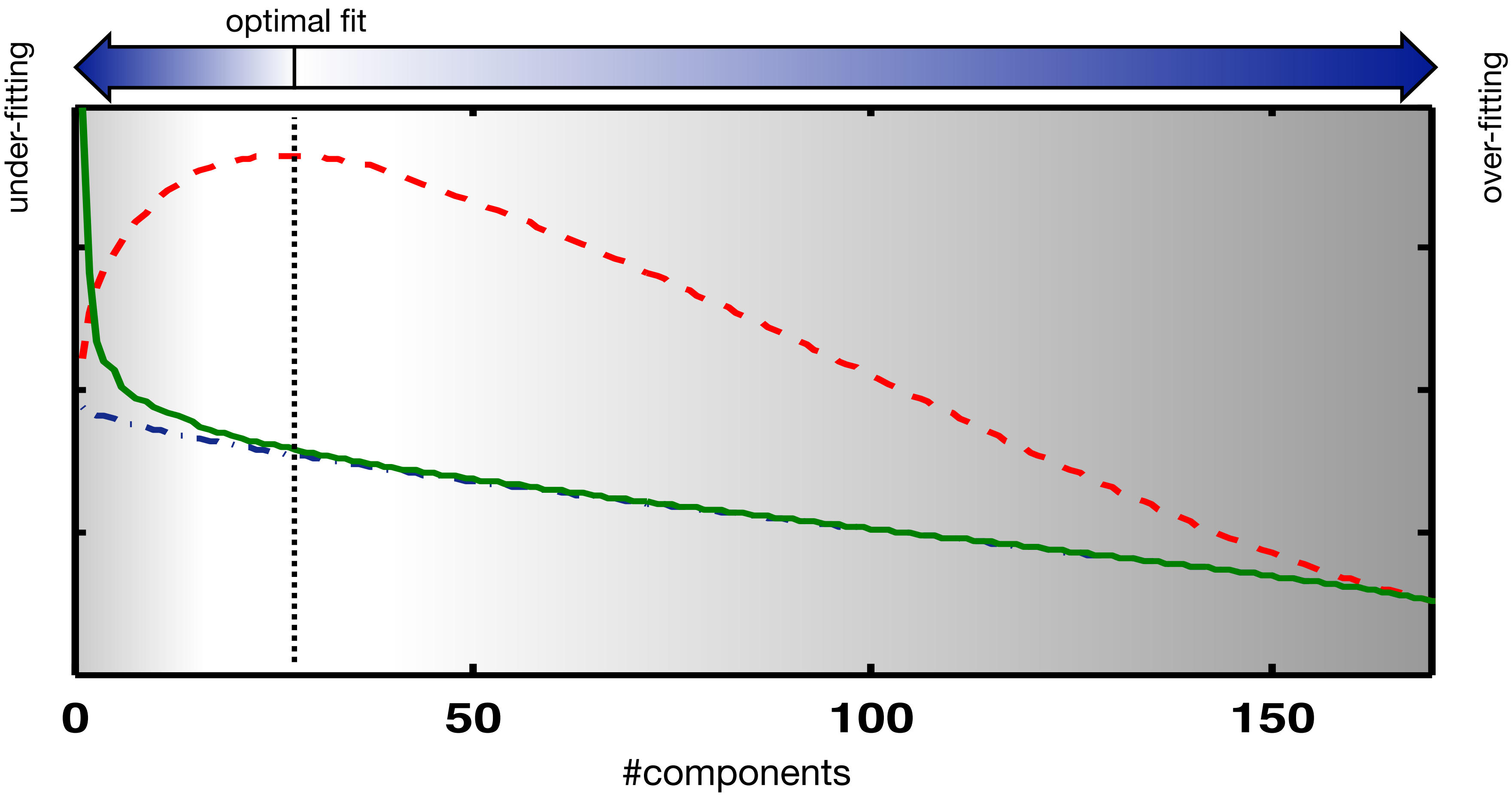


observed Eigenspectrum of the data covariance matrix

Laplace approximation of the posterior probability of the model order

theoretical Eigenspectrum from Gaussian noise

Model Order Selection

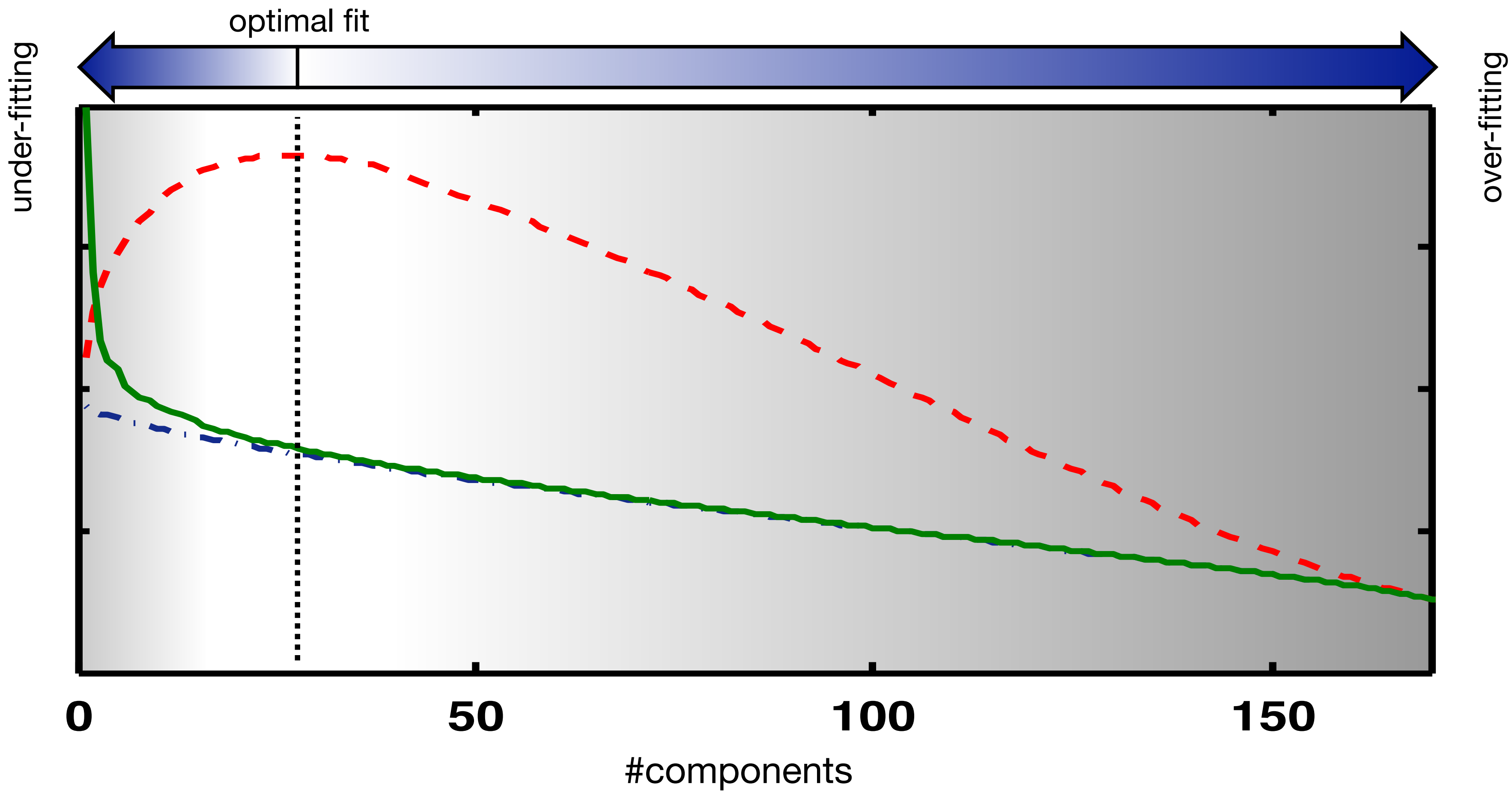


observed Eigenspectrum of the data covariance matrix

Laplace approximation of the posterior probability of the model order

theoretical Eigenspectrum from Gaussian noise

Model Order Selection

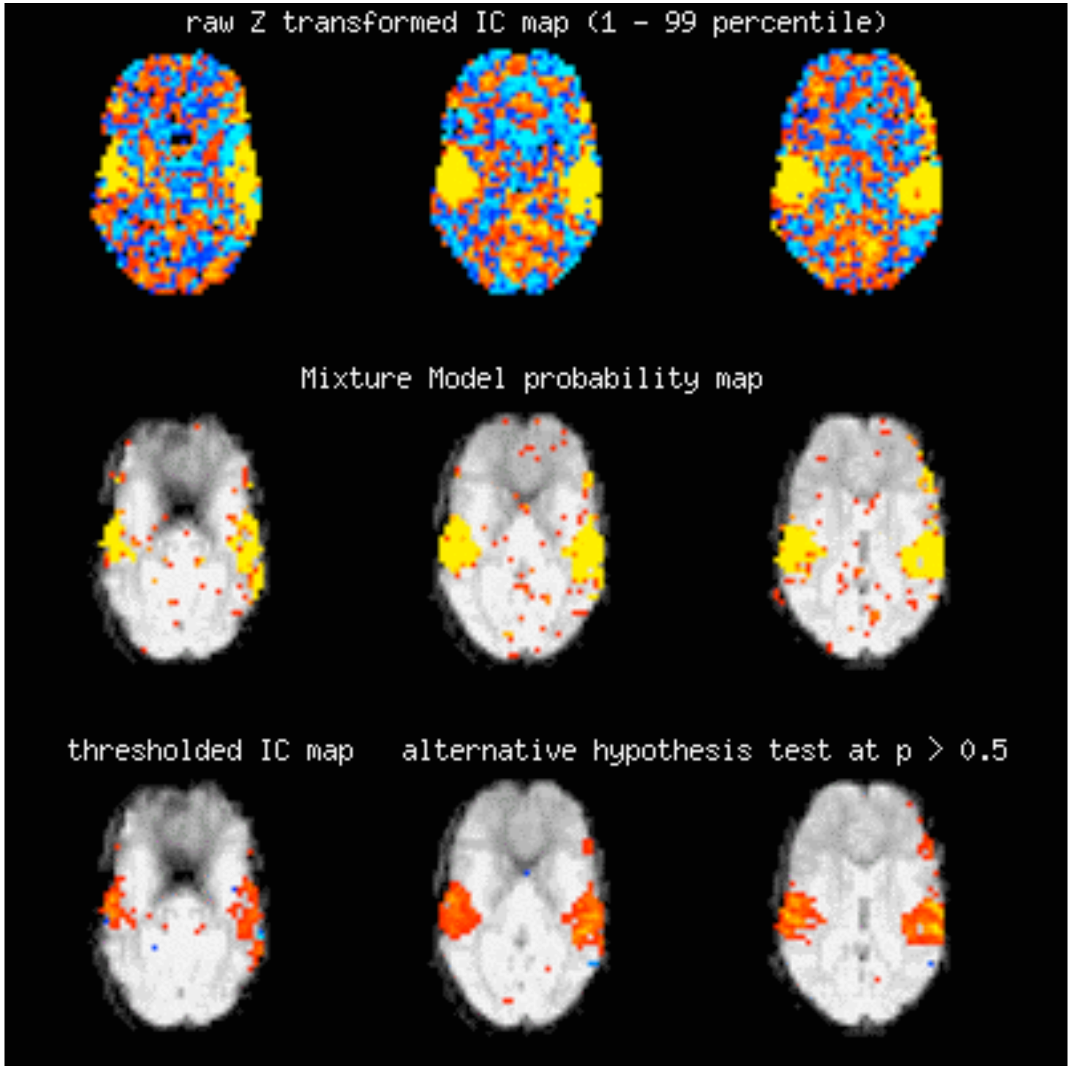


observed Eigenspectrum of the data covariance matrix

Laplace approximation of the posterior probability of the model order

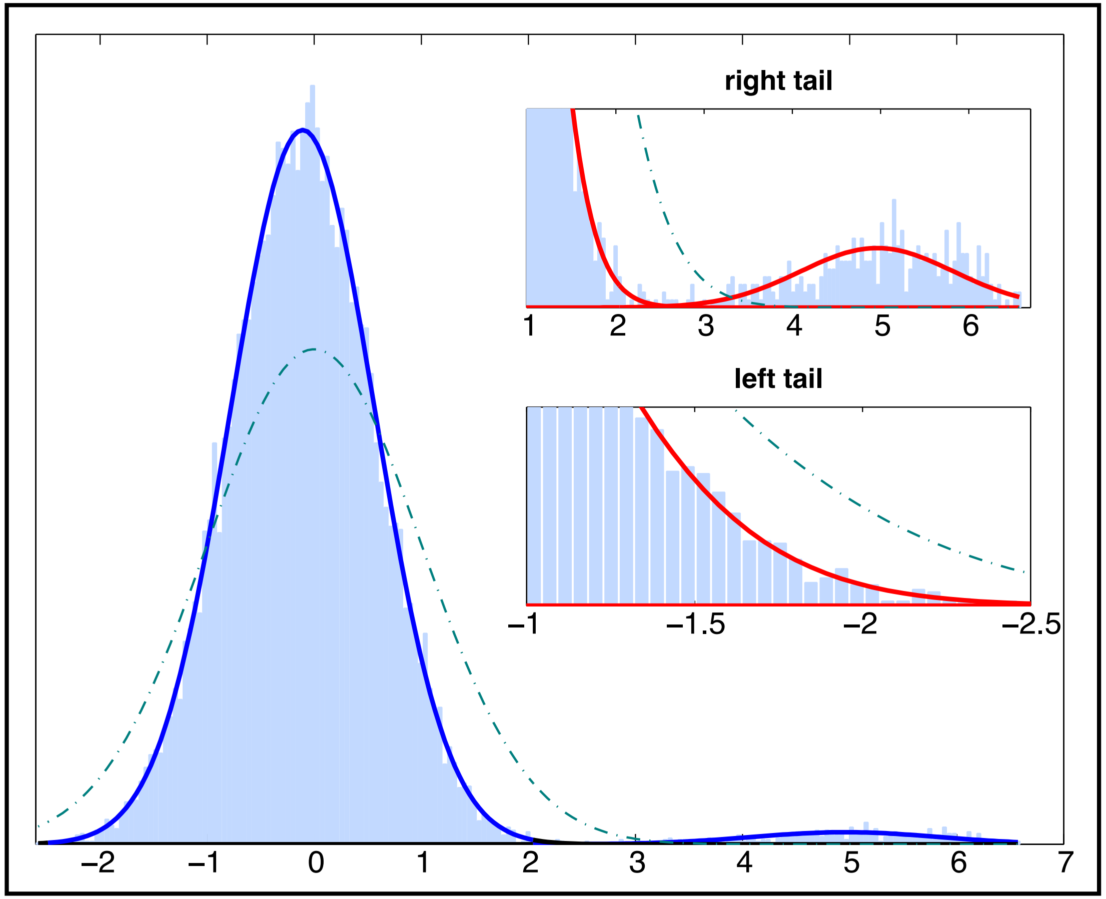
theoretical Eigenspectrum from Gaussian noise

Thresholding

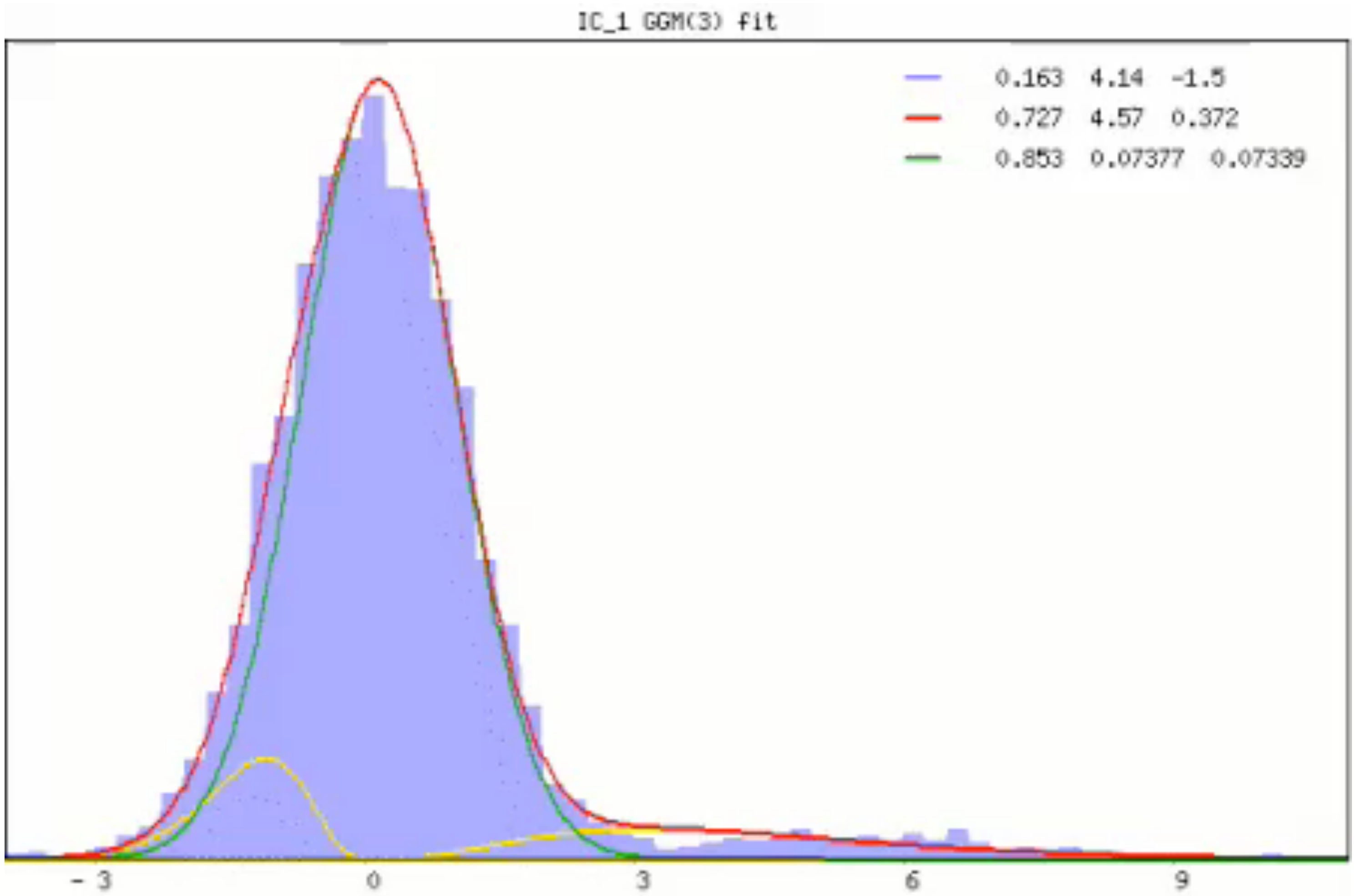


Thresholding

- classical null-hypothesis testing is invalid
- data is assumed to be a linear combination of signals and noise
- the distribution of the estimated spatial maps is a mixture distribution!

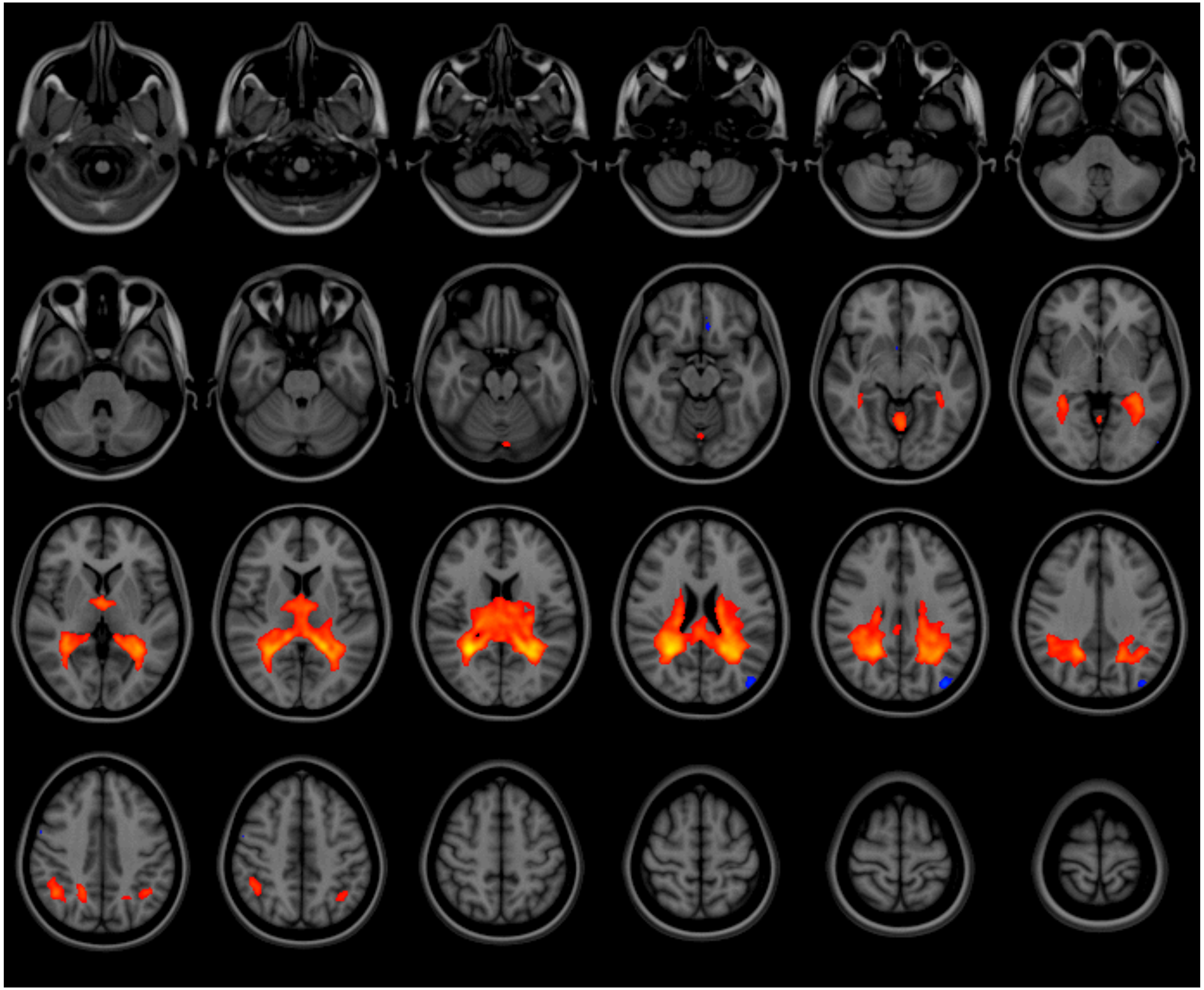


Alternative Hypothesis Test



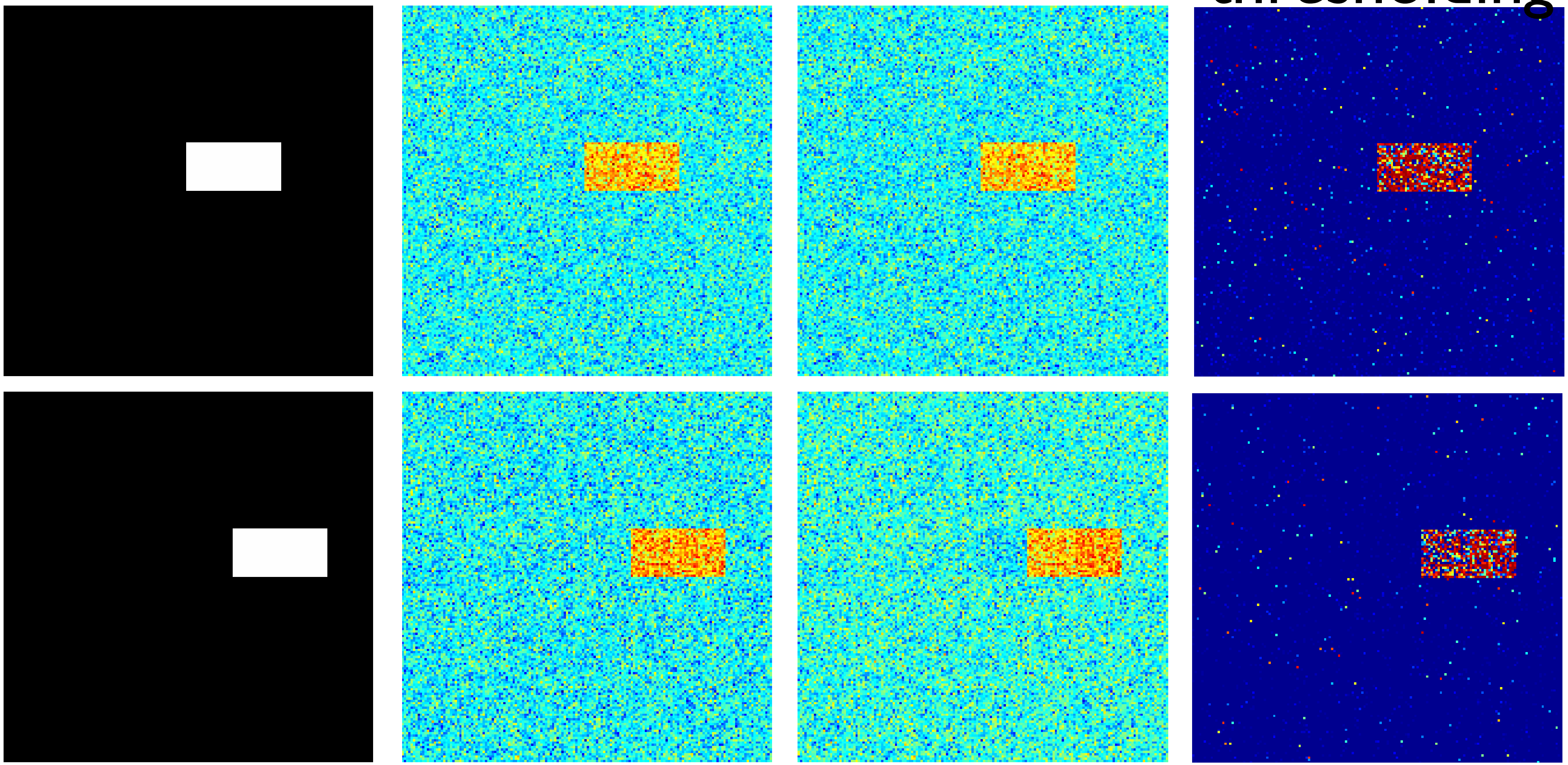
- use Gaussian/Gamma mixture model fitted to the histogram of intensity values (using EM)

What about overlap?



What about overlap?

Sources | Sources + noise | ICA solution | after thresholding



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