Resting-State fMRI:
ICA and Dual Regression

FSL Course 2024

19 June, Osaka, Japan

Resting state fMRI and ICA

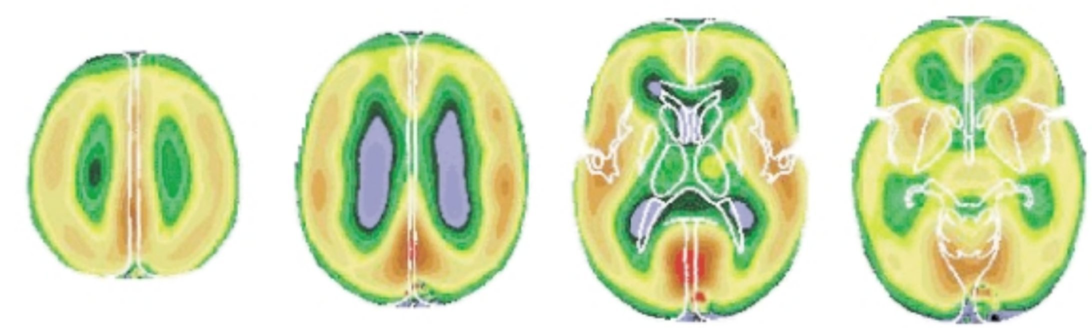
- Introduction to resting state
- Independent Component Analysis
- Single-subject ICA
- Multi-subject ICA
- Dual regression

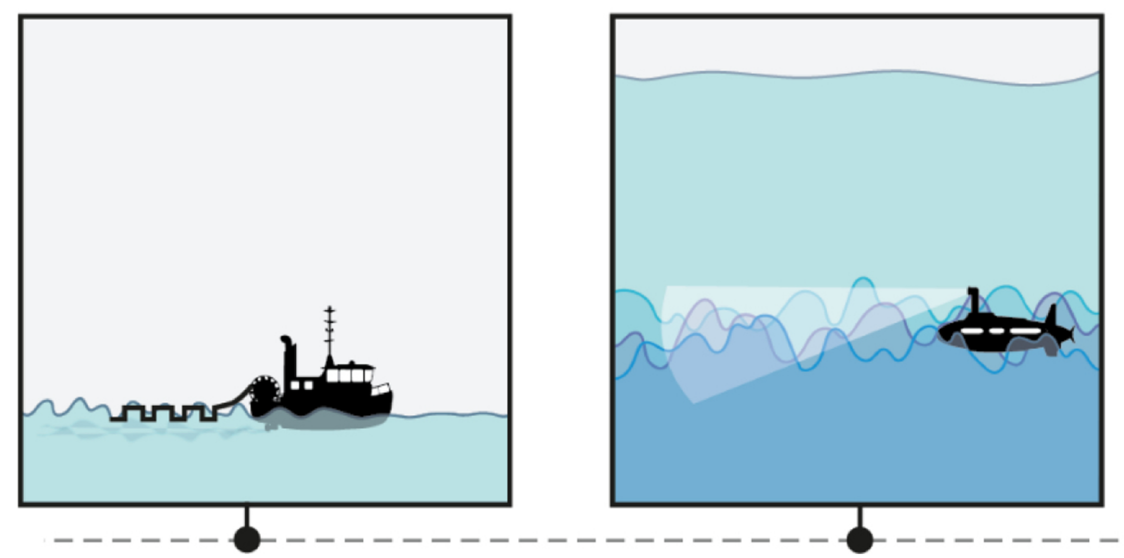
Why resting state fMRI?

Energy consumption in the brain

- Brain < 2% body weight but consumes ~20% of total energy
- Estimated 60-80% of this energy used to support communication between cells
- fMRI provides a window to brain activity
- Task-evoked activity accounts for ~1%

Oxygen consumption



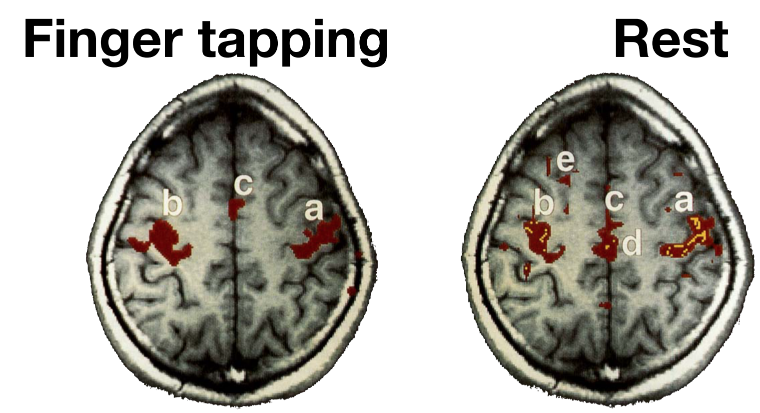


Finn 2021

Resting state -> intrinsic functional brain organisation

Why study the brain at rest?

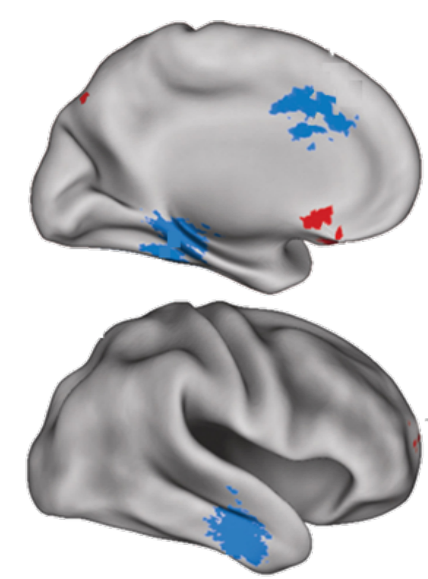
- Understand the inherent functional organisation of the brain



Biswal et al (1995)

- Clinical/ cognitive biomarker

- Pragmatic benefits: can be done in any population, with relatively little setup and expertise required



Sheline et al (2010)

- Localisation vs connectivity

Overview of Resting State Analysis

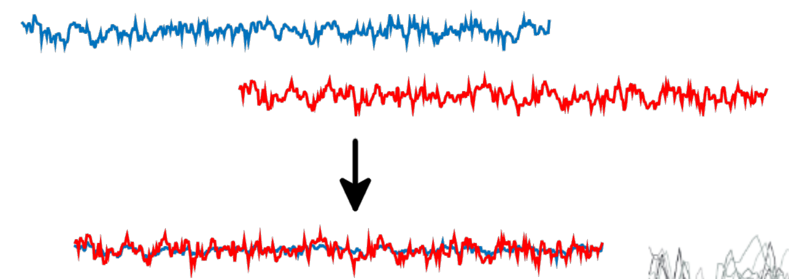
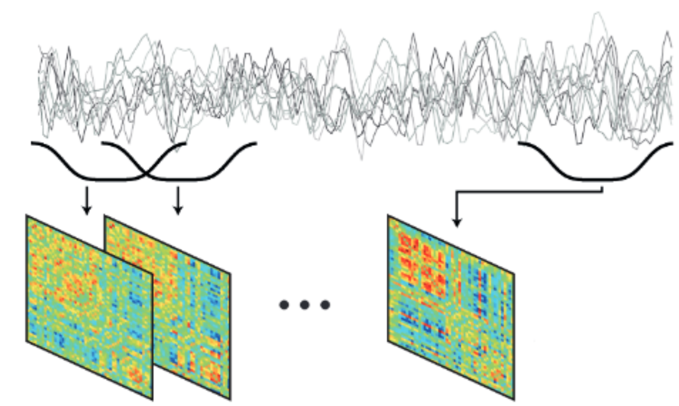
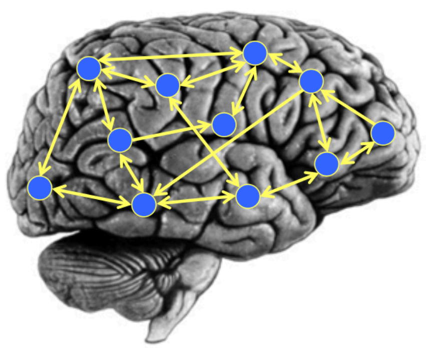
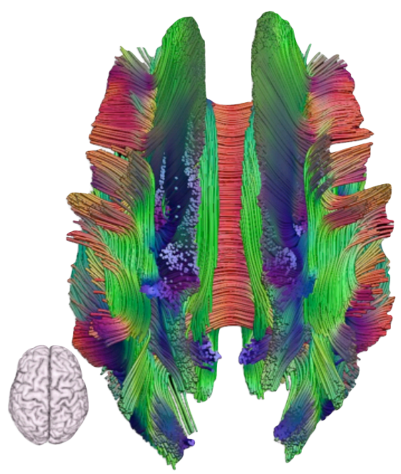
Principles of resting state analysis

- Many different methods available for analysis
- All have one assumption in common:
  - The definition of functional connectivity is based on a statistical dependency between timeseries
- Differences between methods lie in the way these similarities are estimated and/or represented

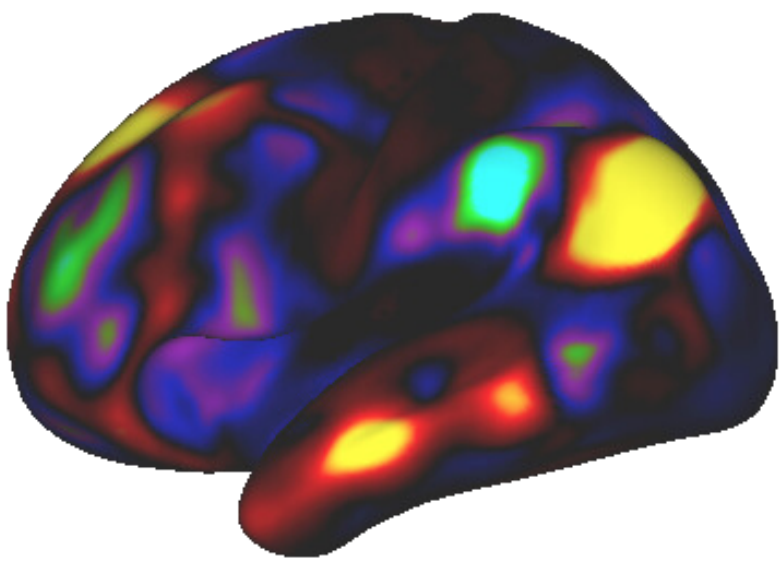
If two brain regions show similarities in their BOLD timeseries, they are functionally connected

Types of connectivity

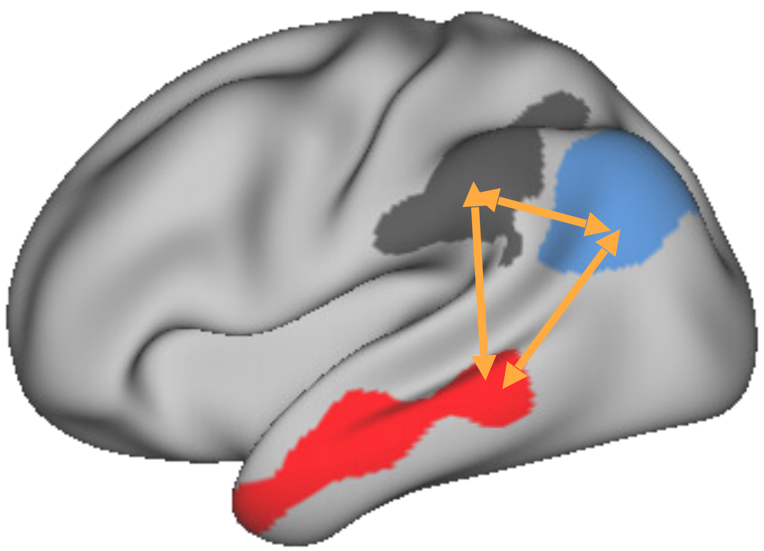
- Functional connectivity
    - Statistical dependency
- Dynamic connectivity
    - Changes in functional connectivity over time
- Effective connectivity
    - Directional influence
- Anatomical (structural) connectivity
    - Presence of a white matter tract

Two broad categories of resting state methods



Voxel-based methods



Node-based methods

Overview of resting state methods

Voxel-based methods

- Seed-based correlation analysis
  - SCA
- Independent component analysis
  - ICA
- Amplitude of low frequency fluctuations
- Regional homogeneity

Node-based methods

- Network modelling analysis
  - FSLNets
- Graph theory analysis
  - Such as degree, hub, path length
- Dynamic causal modelling
- Non-stationary methods

Overview of resting state methods

Voxel-based methods

- Seed-based correlation analysis
    - SCA
- Independent component analysis
    - ICA -> this afternoon
- Amplitude of low frequency fluctuations
- Regional homogeneity

Node-based methods

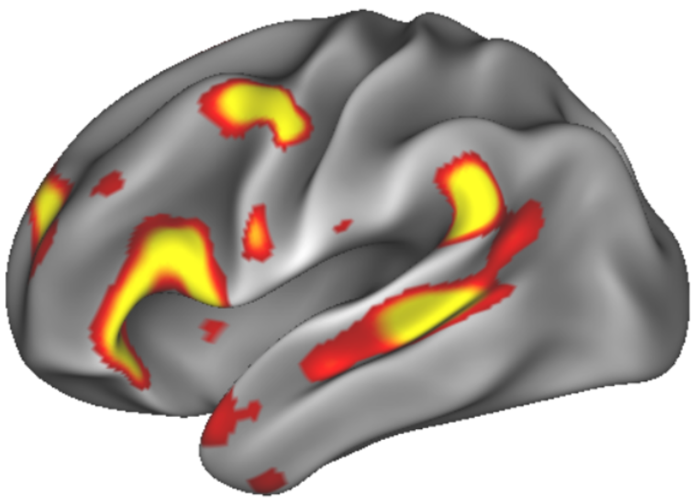
- Network modelling analysis
    - FSLNets -> tomorrow morning
- Graph theory analysis
    - Such as degree, hub, path length
- Dynamic causal modelling
- Non-stationary methods

Resting state fMRI and ICA

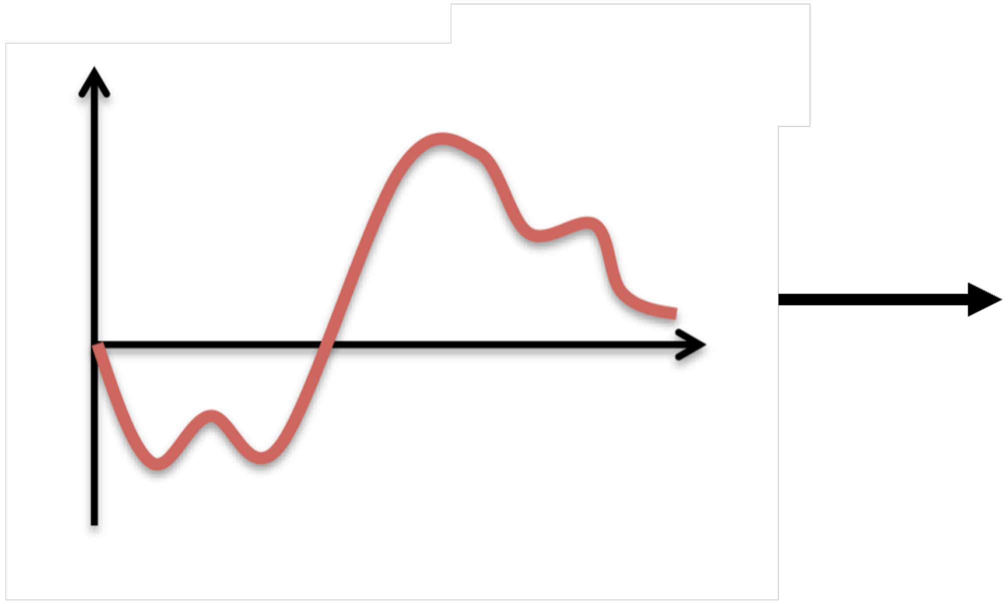
- Introduction to resting state
- Independent Component Analysis
- Single-subject ICA
- Multi-subject ICA
- Dual regression

Resting state functional components

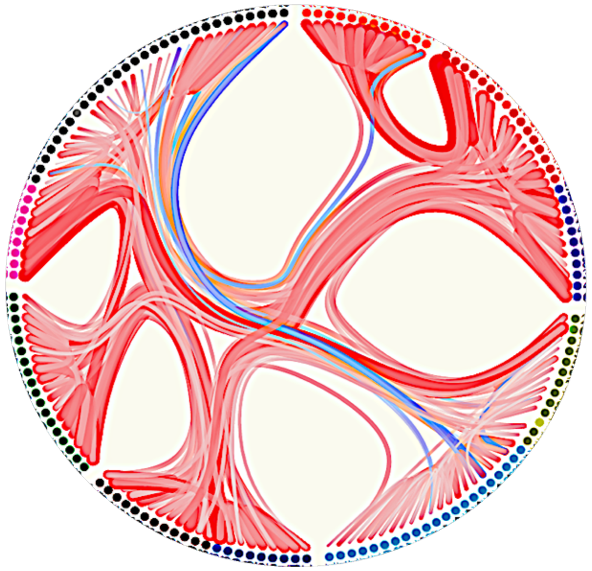
Spatial Maps or Configuration



Time Courses



Network Modelling

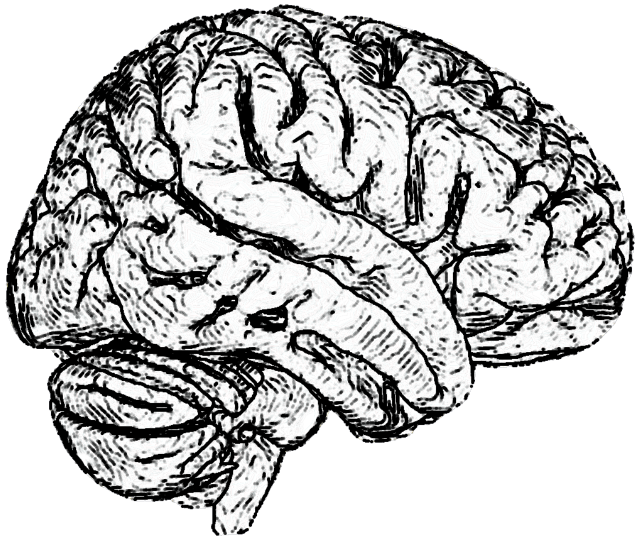


- Each component -> a set of brain regions that work in synchrony
- Timecourses can then be used for network modelling

MELODIC: ICA tool in FSL

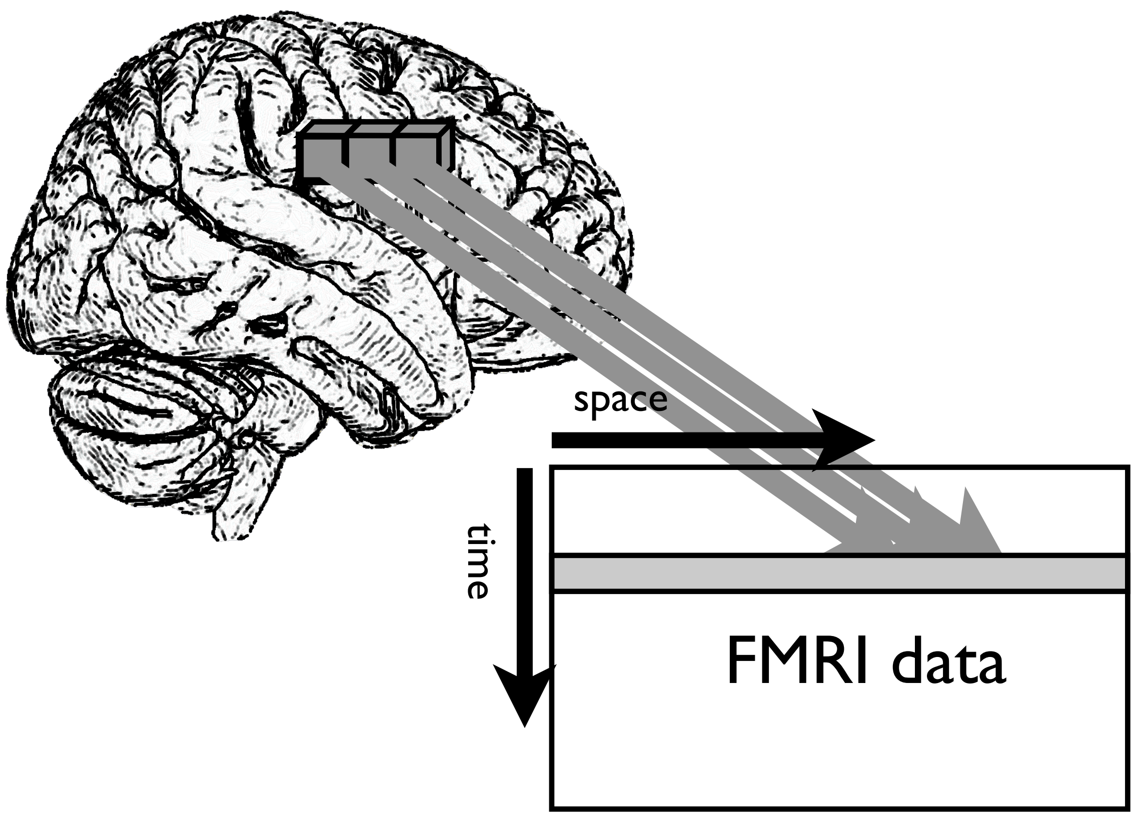
Melodic

multivariate linear decomposition:



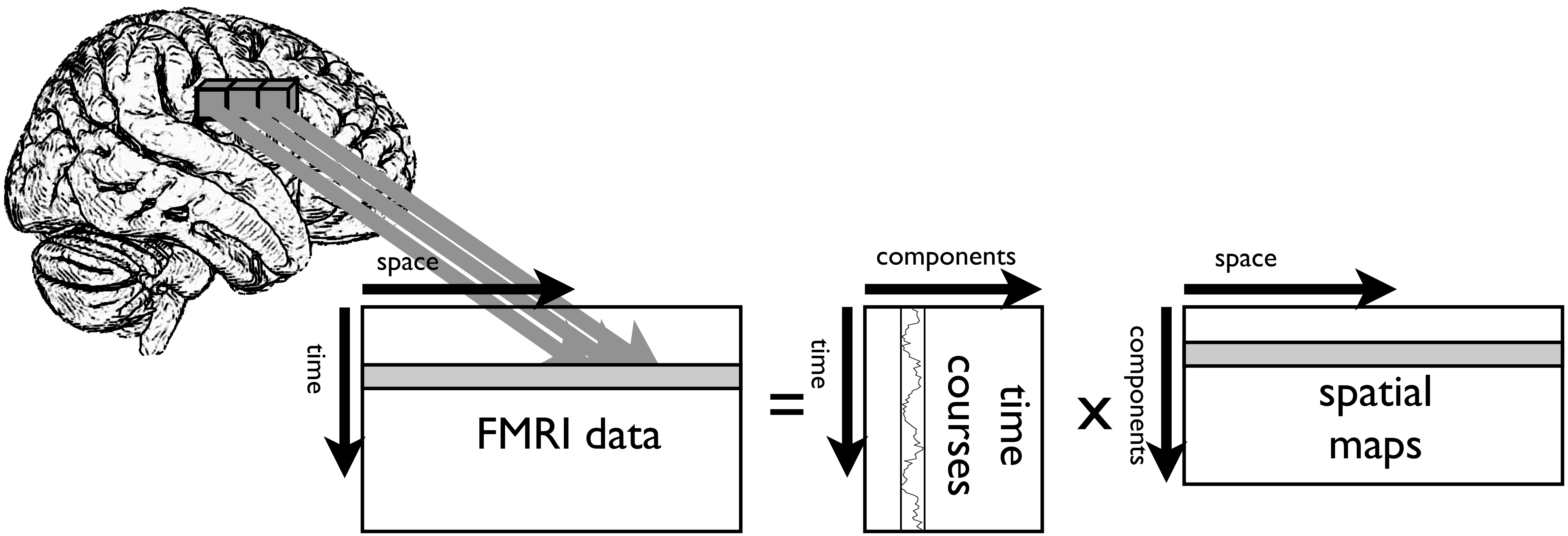
Melodic

multivariate linear decomposition:



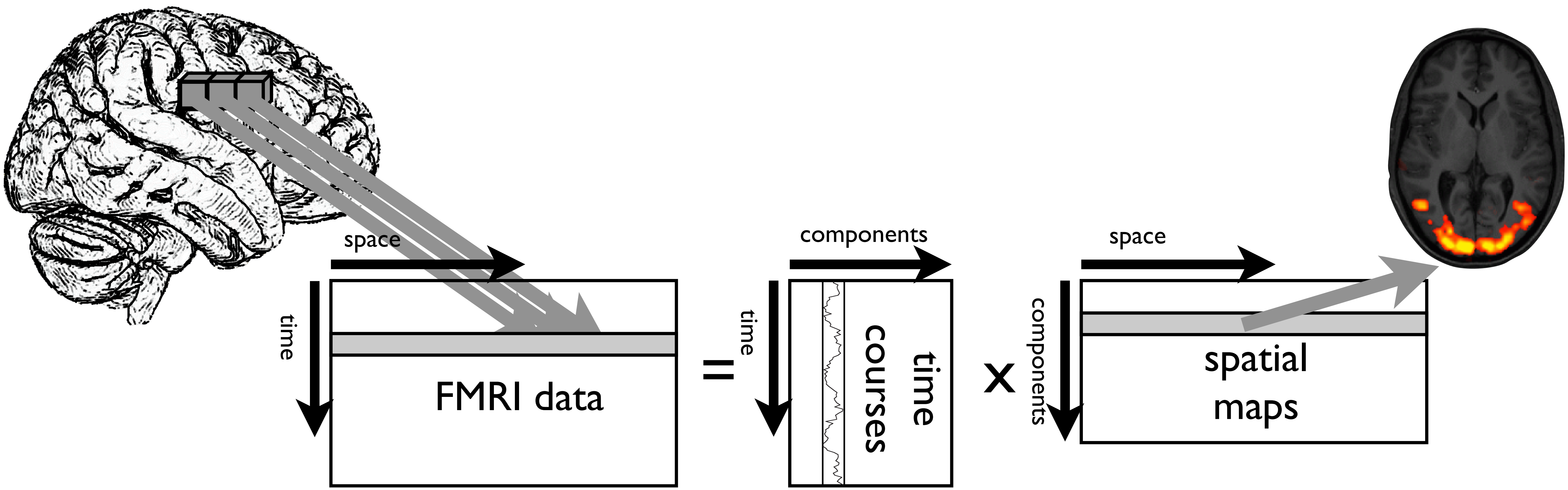
Melodic

multivariate linear decomposition:



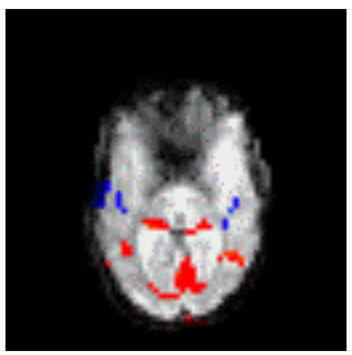
Melodic

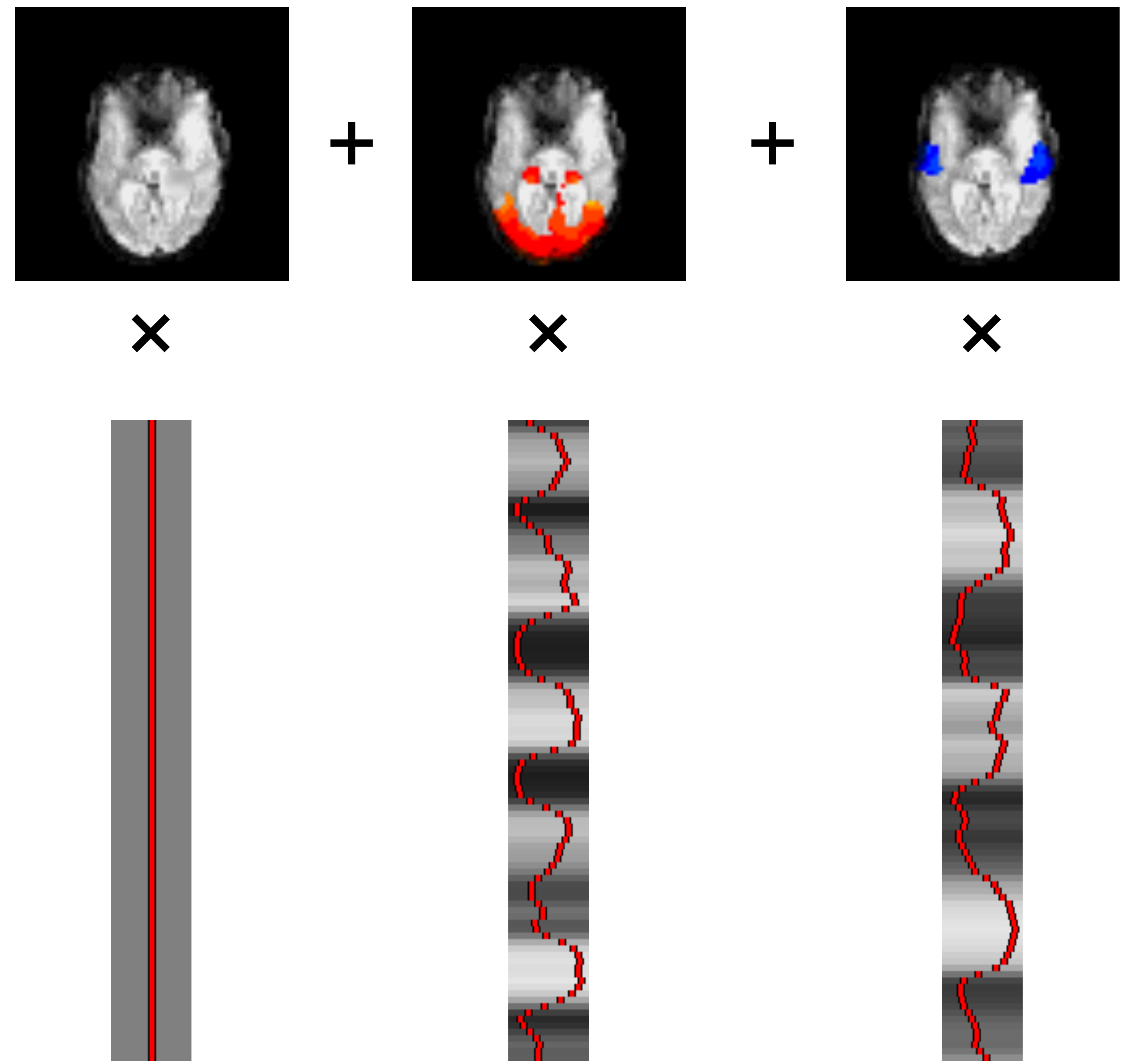
multivariate linear decomposition:



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

What are components?





- express observed data as linear combination of spatio-temporal processes.
- Components can be liberally recombined to reconstruct the original data.

Some characteristics of MELODIC

- Multivariate voxel-based approach

- Exploratory “model-free” method to find interesting structure in the data

- Gives “spatially independent” components

- “Avoiding overfitting” through automatic model order selection

- “Thresholding" to remove background from main signal

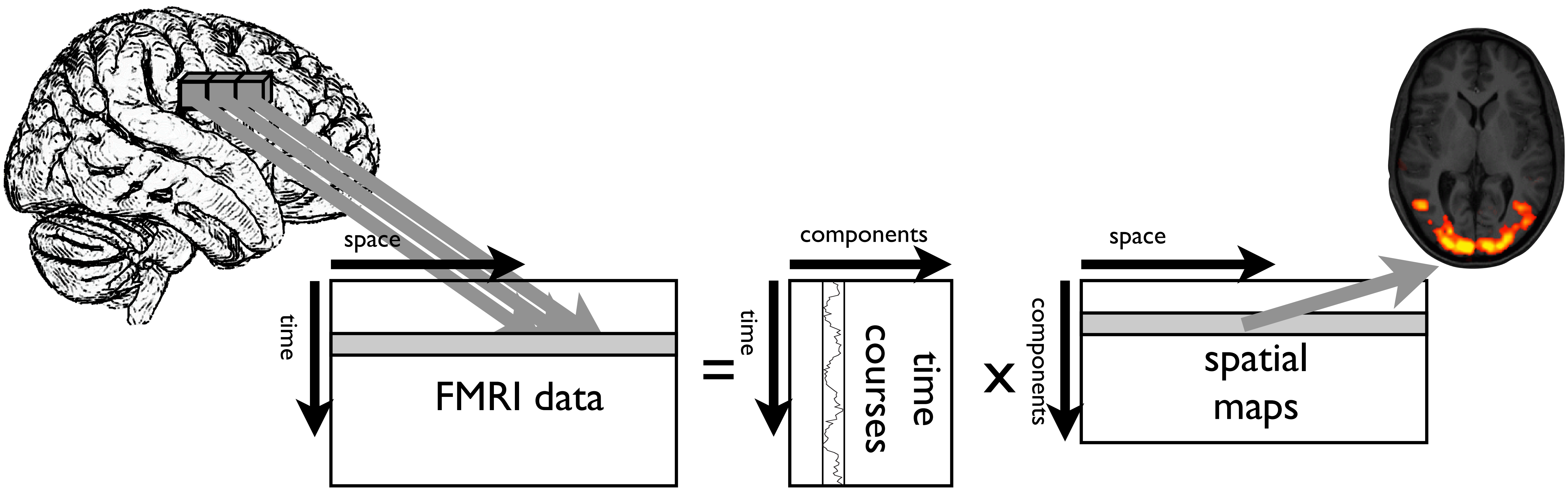
Some characteristics of MELODIC

- Multivariate voxel-based approach
- Exploratory “model-free” method to find interesting structure in the data
- Gives “spatially independent” components
- “Avoiding overfitting” through automatic model order selection
- “Thresholding" to remove background from main signal

ICA vs GLM:
Exploratory vs Confirmatory

Melodic

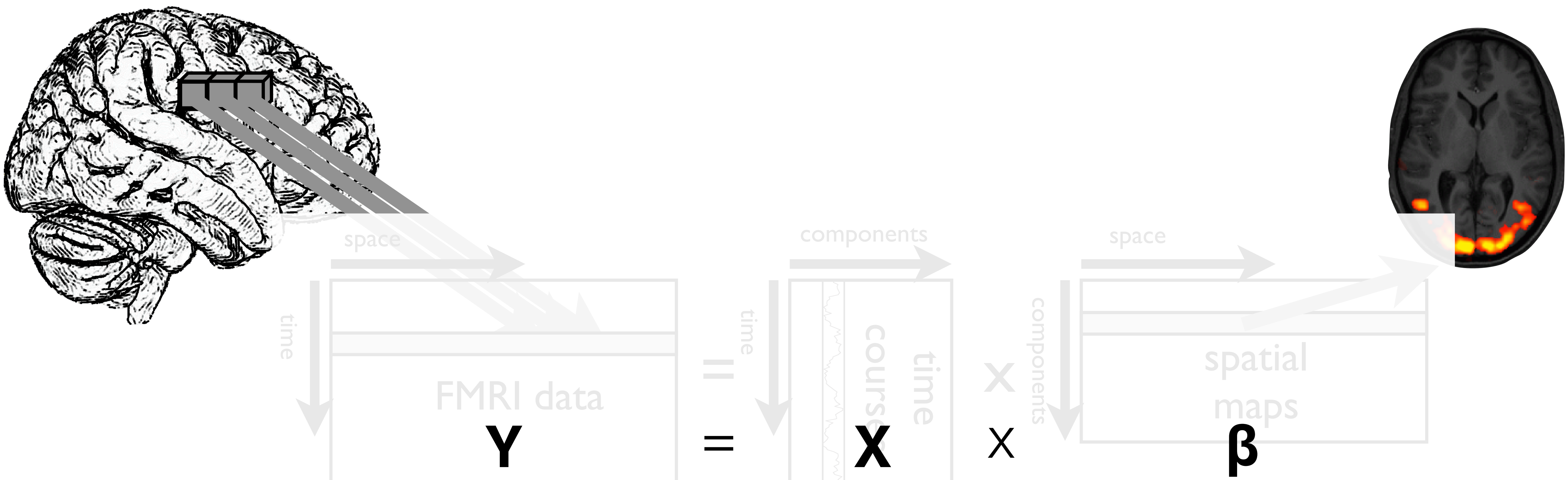
multivariate linear decomposition:



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

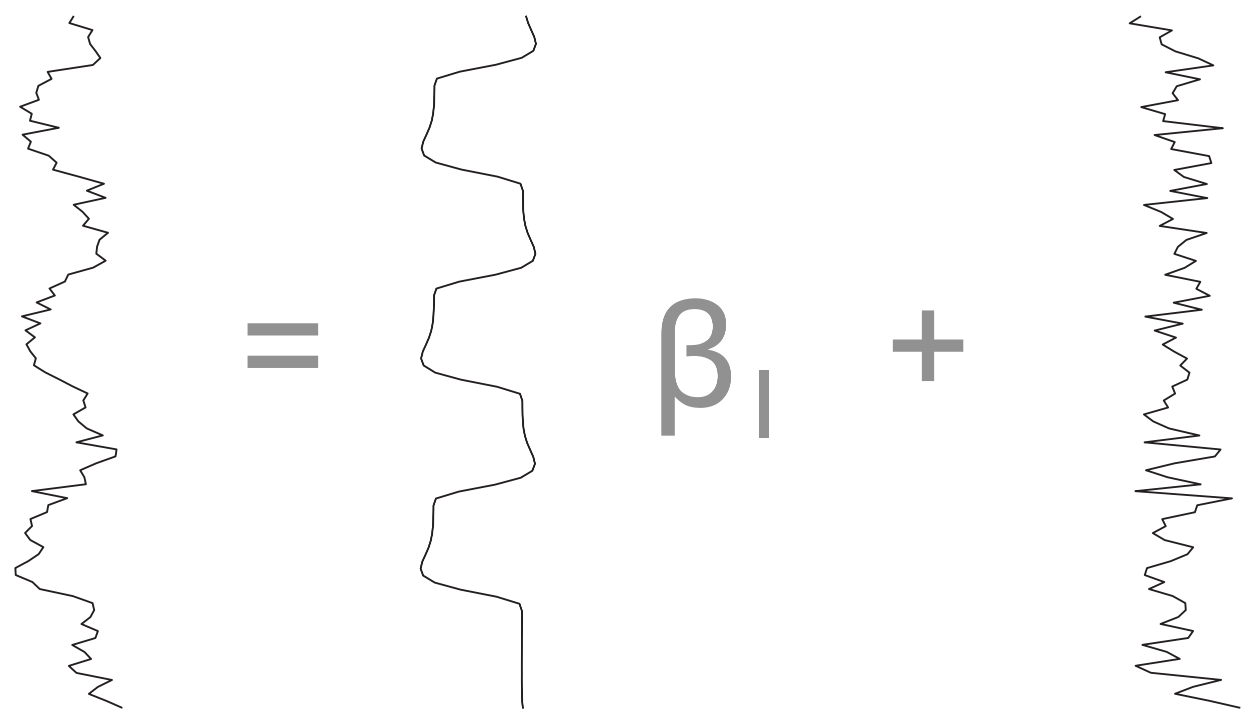
Melodic

multivariate linear decomposition:



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

Model-based (GLM) analysis



- Model each measured time-series as a linear combination of signal and noise
- We know the expected response -> use that to define the design matrix

Resting state = Model-free?

Resting state timeseries are unconstrained -> no design matrix

Task: Confirmatory vs Rest: Exploratory

Confirmatory

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

Problem → Data →

$\boxed{\text{Model} \rightarrow \text{Analysis}}$

→ Results

- results depend on the model

Exploratory

- "Is there anything interesting in the data?"

Problem → Data →

$\boxed{\text{Analysis} \rightarrow \text{Model}}$

→ Results

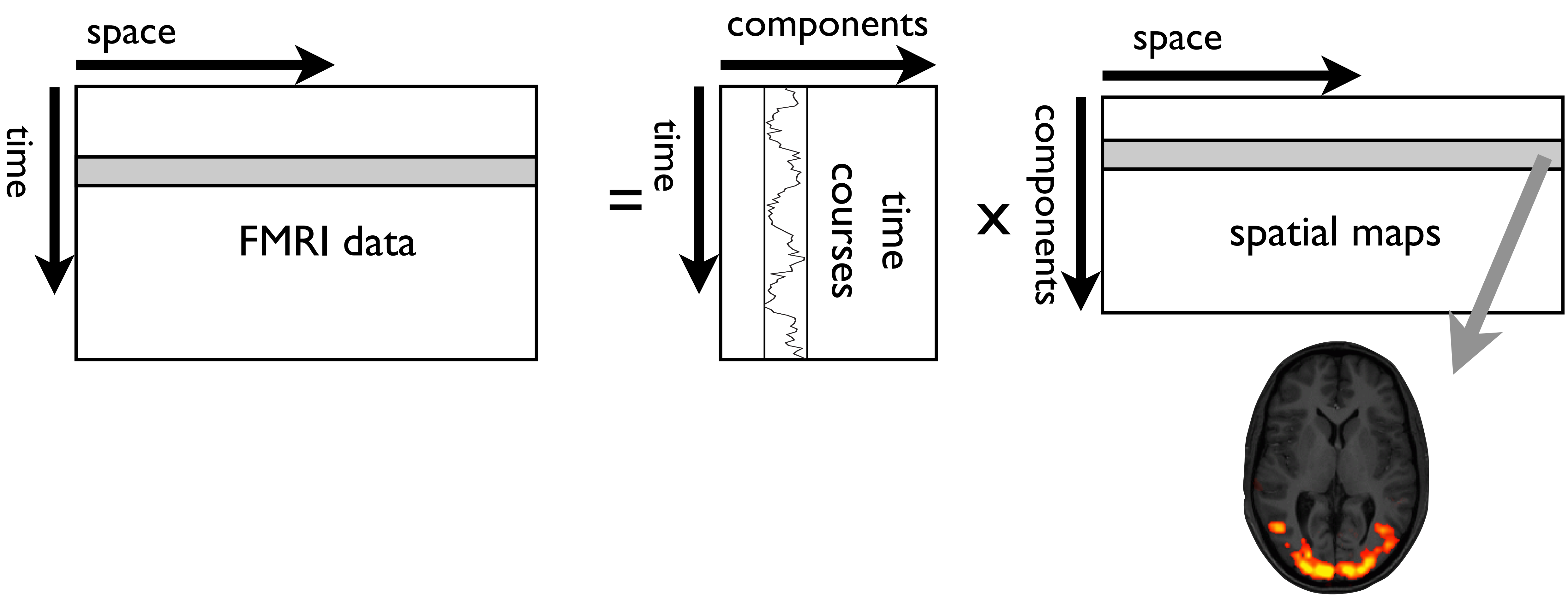
- can give unexpected results

Some characteristics of MELODIC

- Multivariate voxel-based approach
- Exploratory “model-free” method to find interesting structure in the data
- Gives “spatially independent” components
- “Avoiding overfitting” through automatic model order selection
- “Thresholding" to remove background from main signal

Spatially independent components

Spatial ICA for FMRI



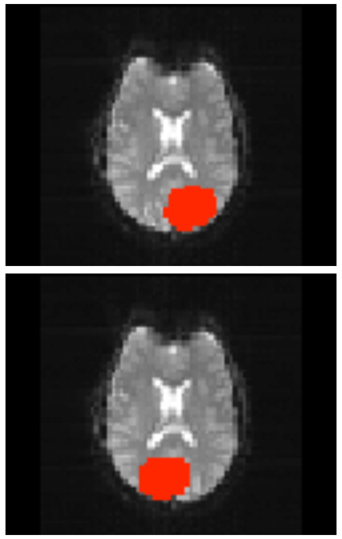
– data is decomposed into a set of spatially independent maps and a set of time-courses

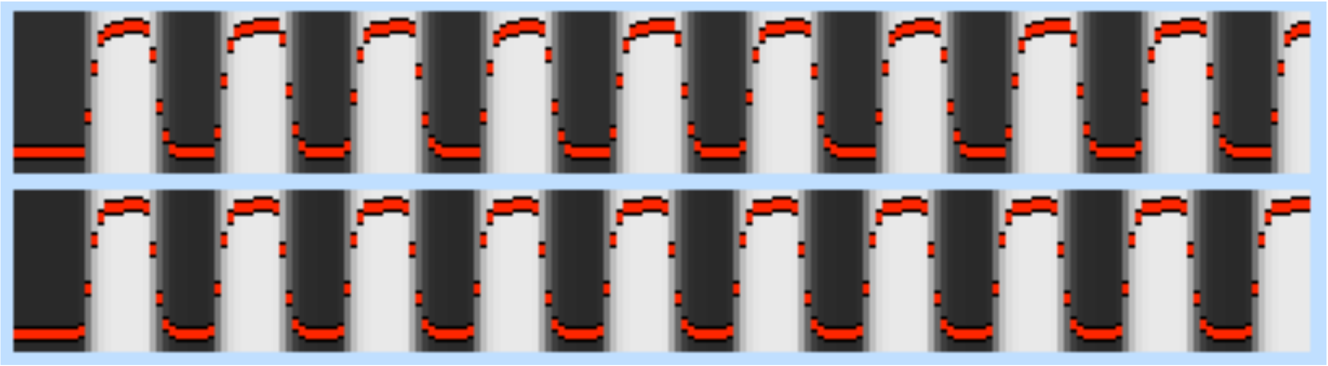
McKeown et al. HBM 1998

PCA vs. ICA ?

Simulated Data

(2 components, slightly different timecourses)

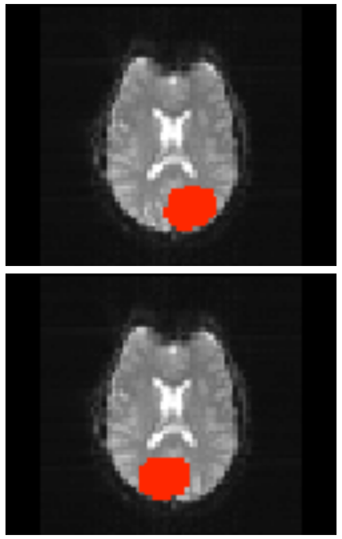


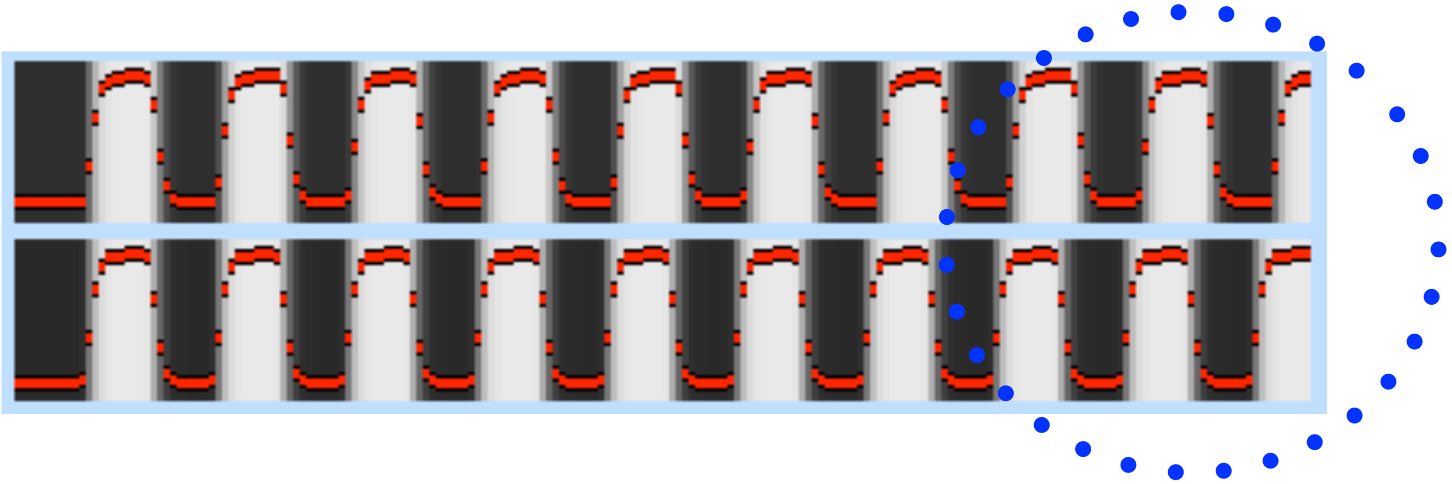


PCA vs. ICA ?

Simulated Data

(2 components, slightly different timecourses)

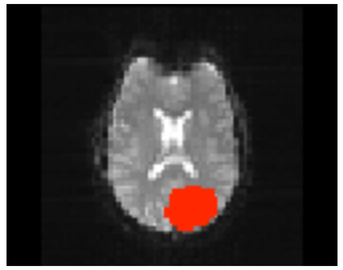


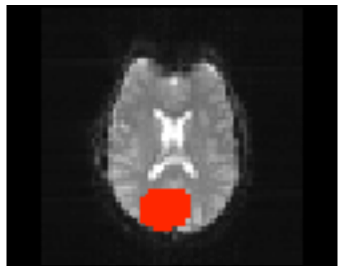


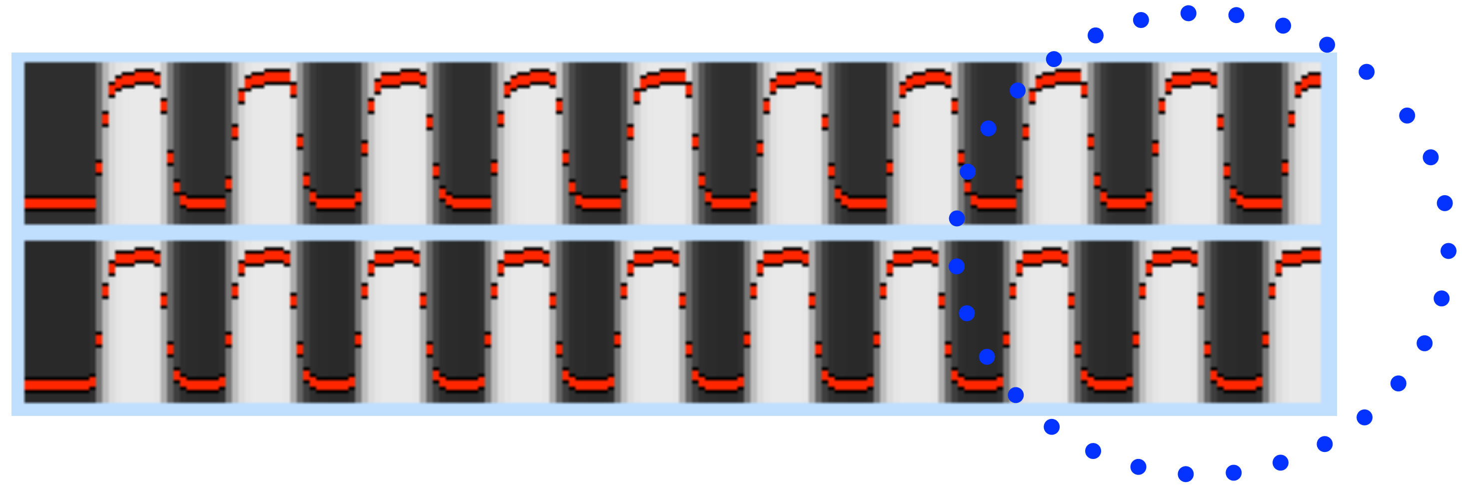
PCA vs. ICA ?

Simulated Data

(2 components, slightly different timecourses)

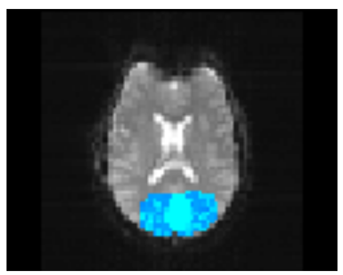


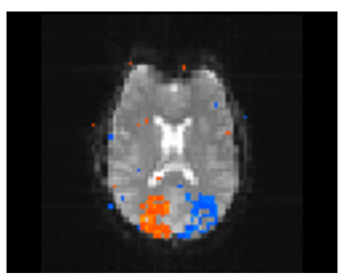


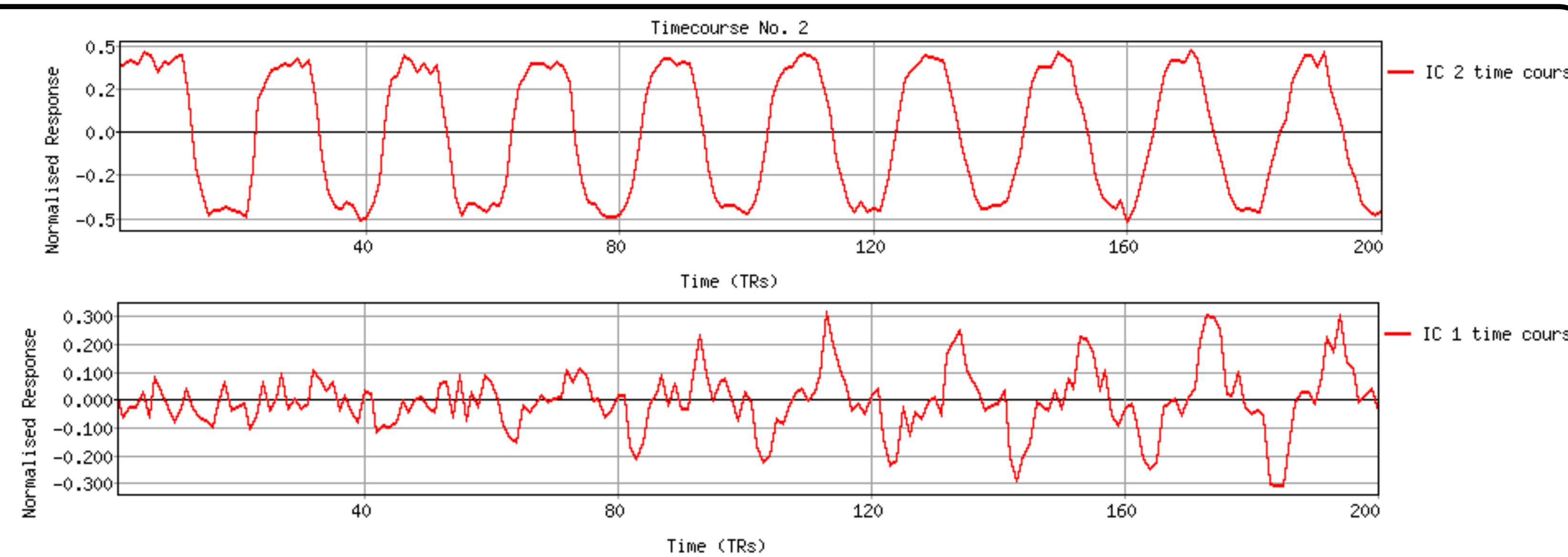


PCA

- Timecourses orthogonal
- Spatial maps and timecourses "wrong"

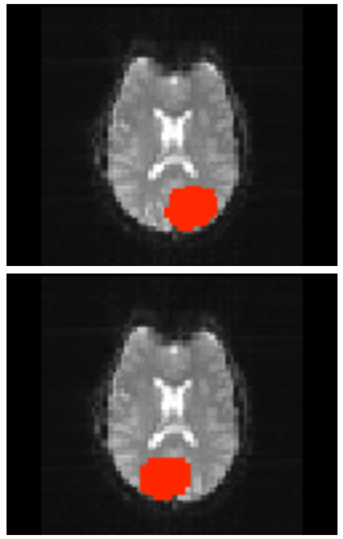


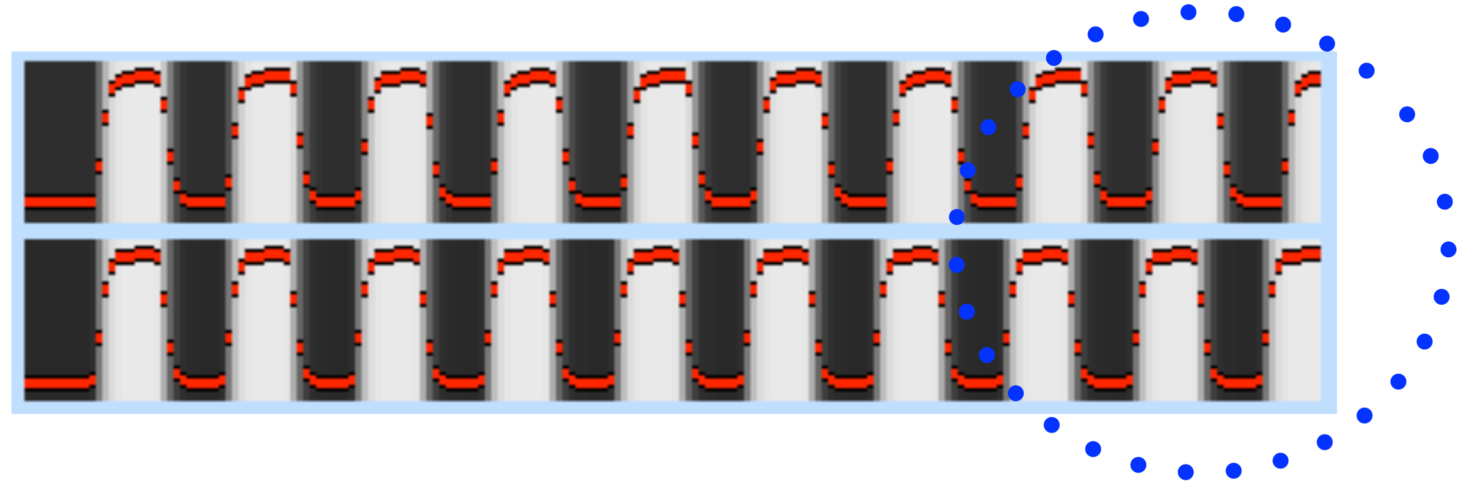




Simulated Data

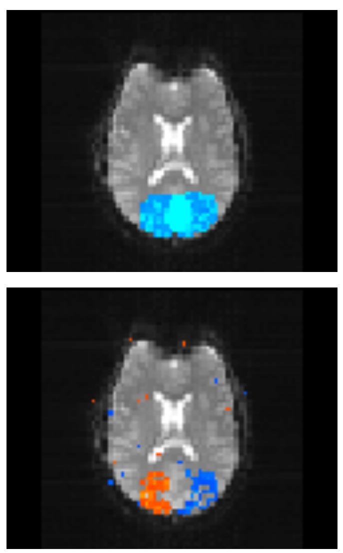
(2 components, slightly different timecourses)

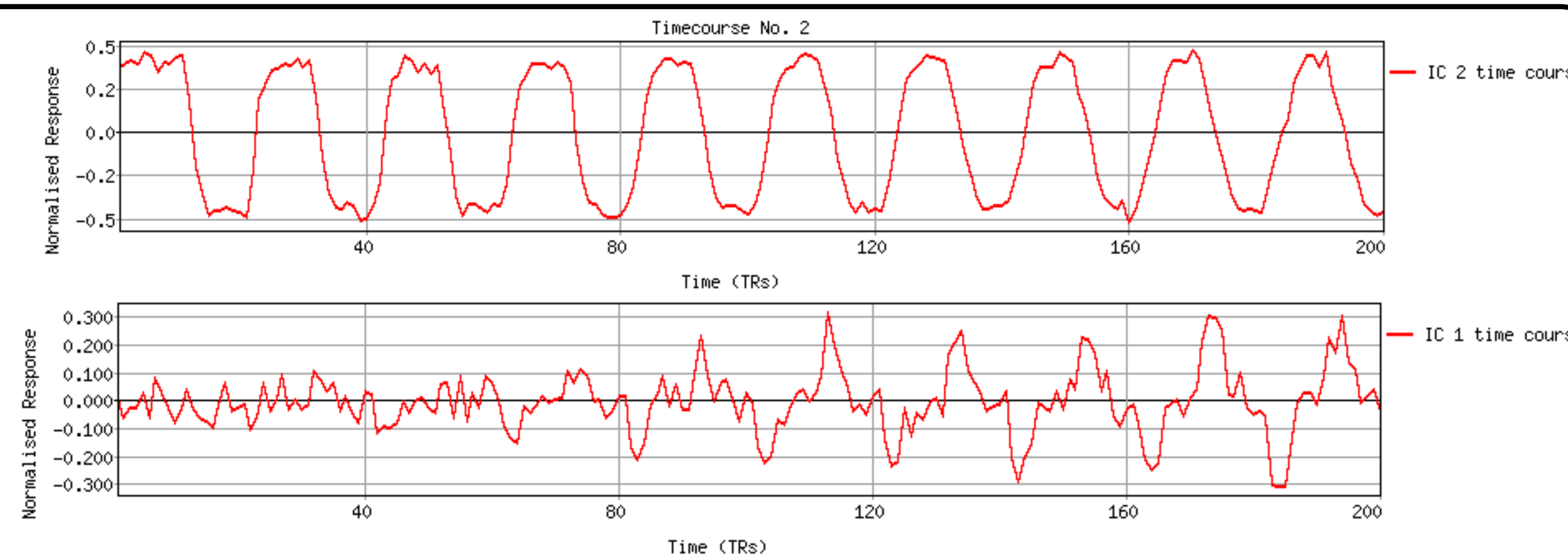




PCA

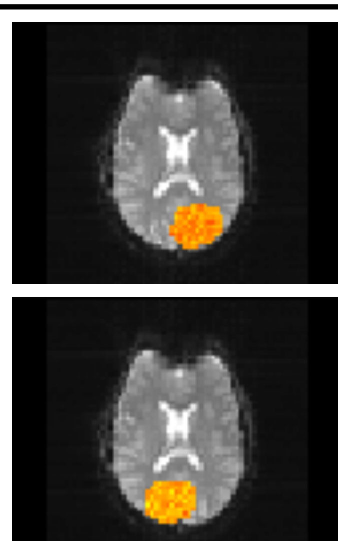
- Timecourses orthogonal
- Spatial maps and timecourses "wrong"

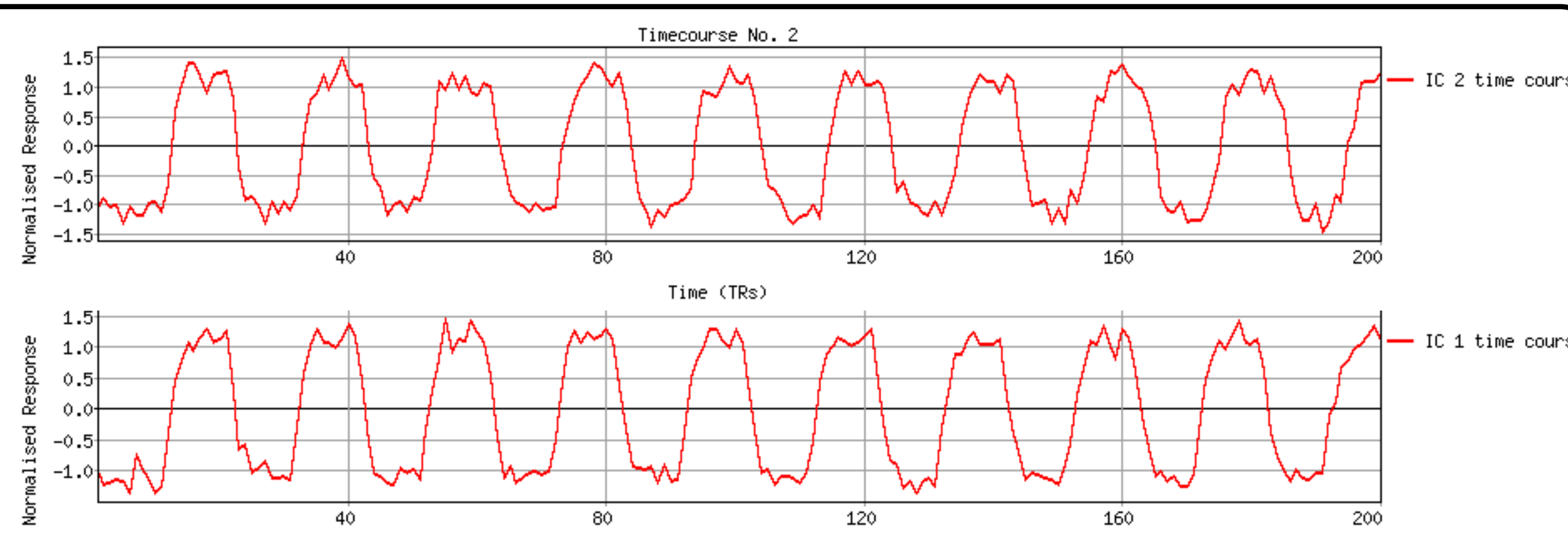




ICA

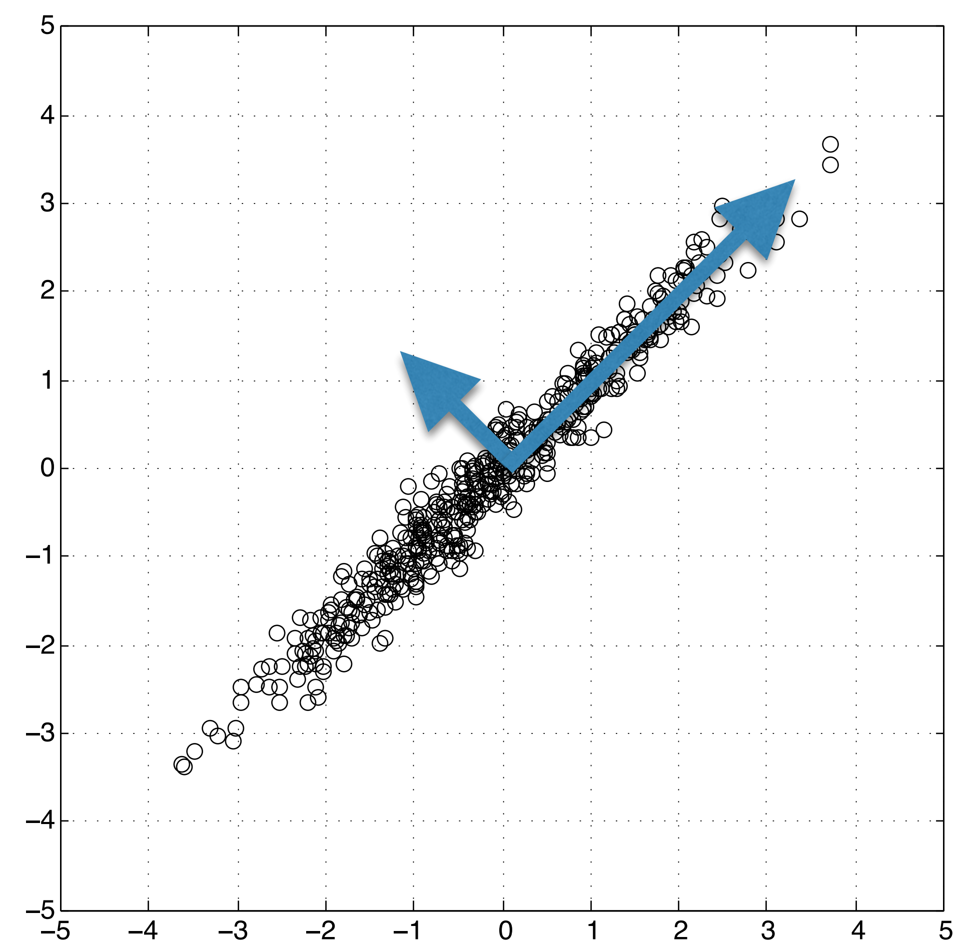
- Timecourses non-co-linear
- Spatial maps and timecourses "right"





PCA vs. ICA

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

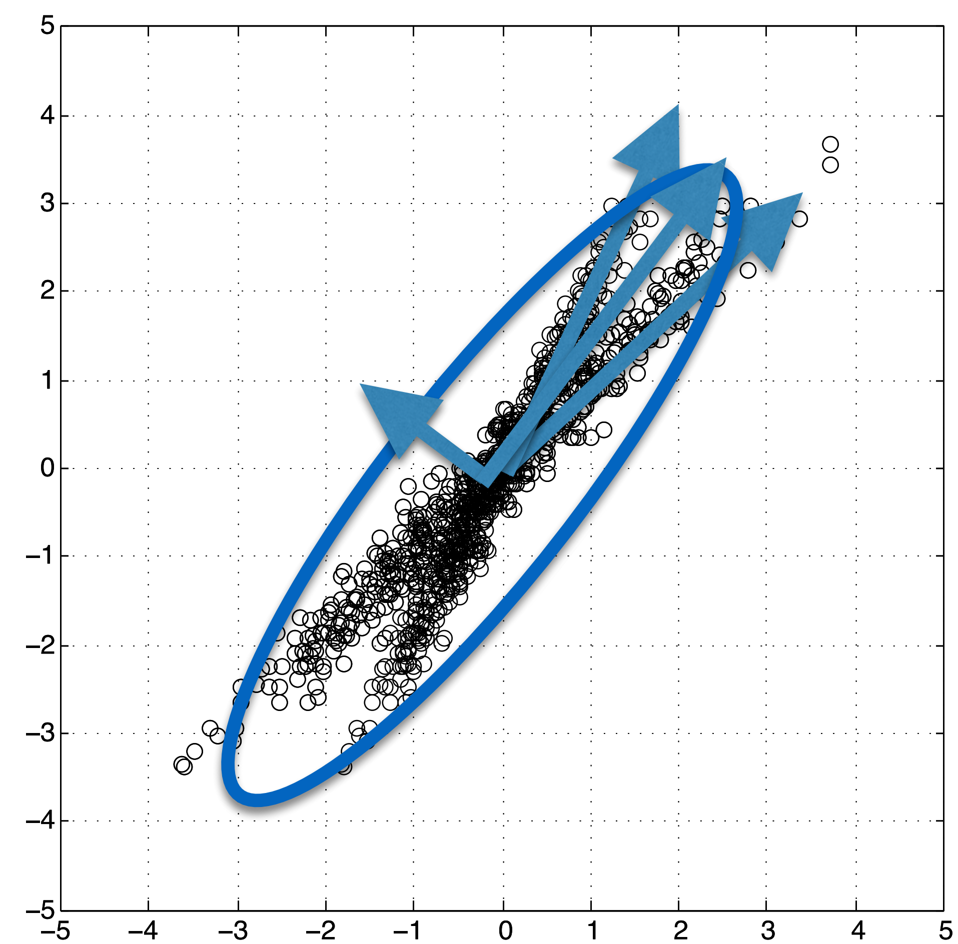


Gaussian data

PCA vs. ICA

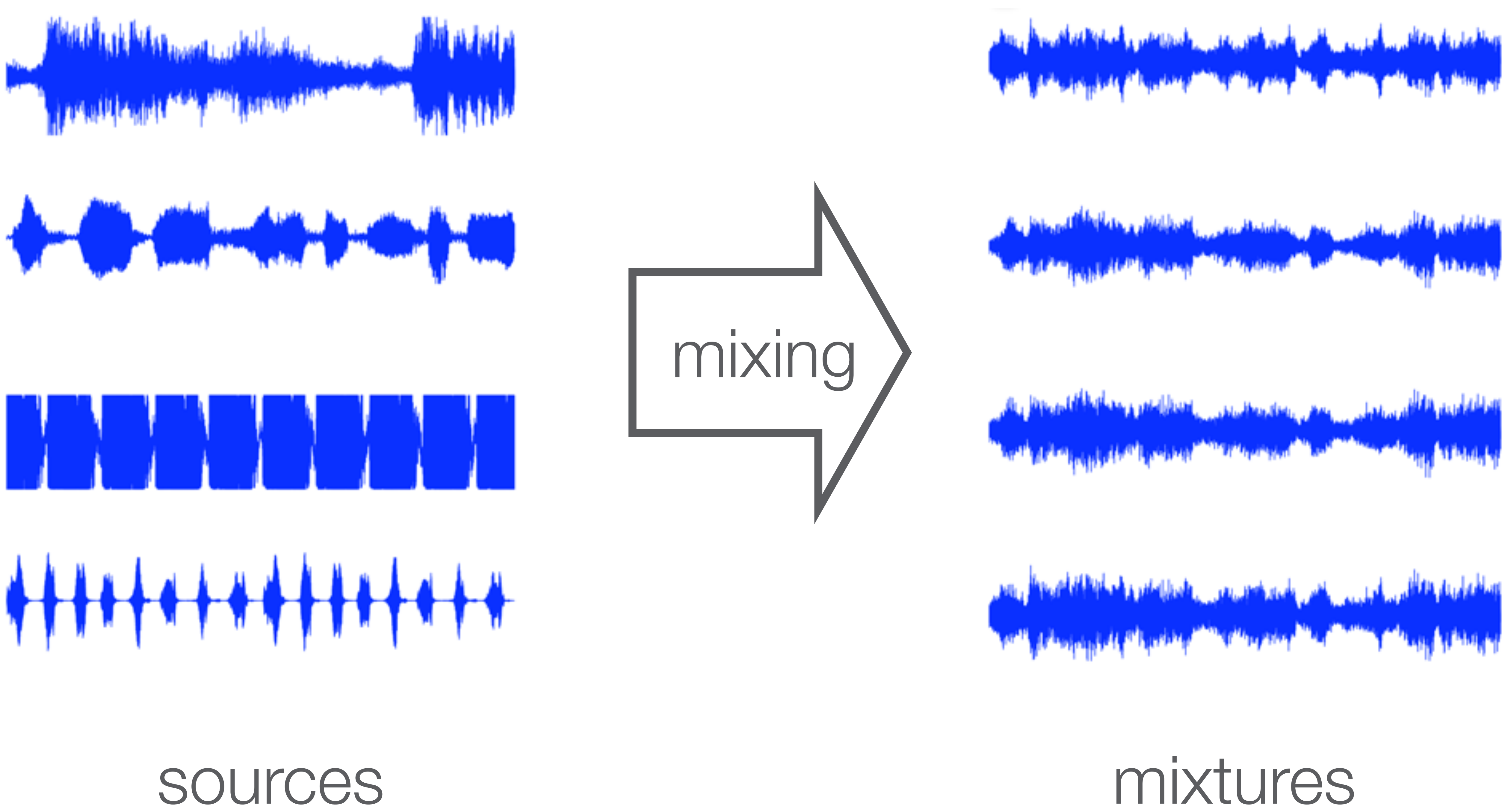
- PCA finds projections of maximum amount of variance in Gaussian data (uses 2nd order statistics only)

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



non-Gaussian data

Non-Gaussianity



Non-Gaussianity

non-Gaussian

Gaussian

ICA estimation

- Random mixing results in more Gaussian-shaped PDFs (Central Limit Theorem)
- ICA turns this around:

  if we estimate components with less Gaussian distributions 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: a) maximises independence between components and b) maximises non-Gaussianity of components
- For (a) need a contrast (objective/cost) function to drive the unmixing which measures statistical independence and for (b) need an optimisation technique:
  - gradient descent & kurtosis or cumulants (Jade)
  - gradient descent & maximum entropy (Infomax)
  - fixed point iteration & neg-entropy (FastICA)

Some characteristics of MELODIC

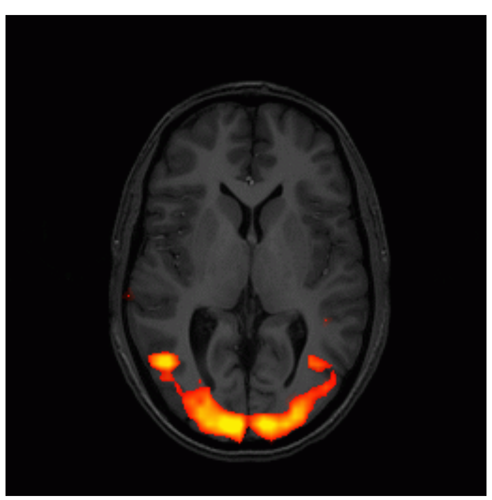
- Multivariate voxel-based approach
- Exploratory “model-free” method to find interesting structure in the data
- Gives “spatially independent” components
- “Avoiding overfitting” through automatic model order selection
- “Thresholding" to remove background from main signal

Managing overfitting: Automatic model order selection

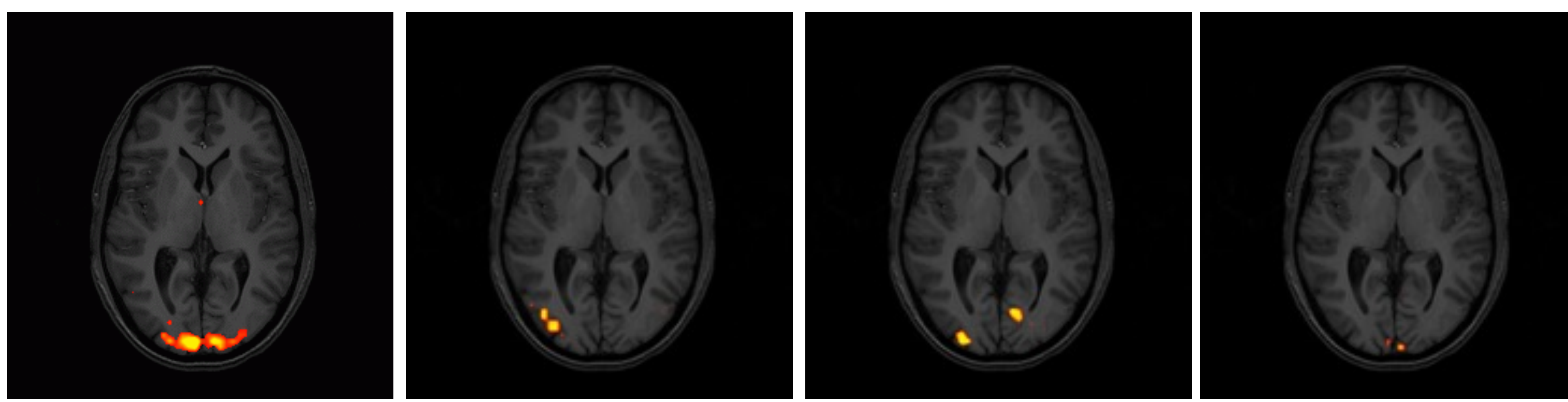
The ‘overfitting’ problem

Fitting a noise-free model to noisy observations:

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



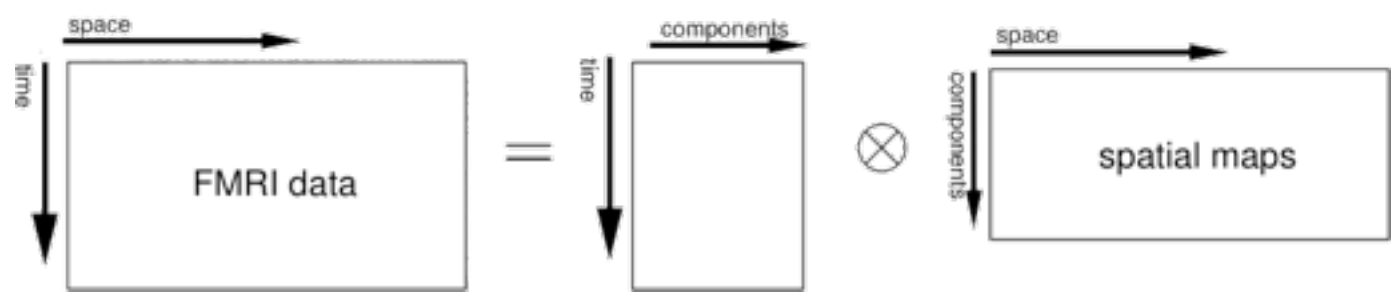
GLM analysis



standard ICA (unconstrained)

Probabilistic ICA model in MELODIC

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



+ noise

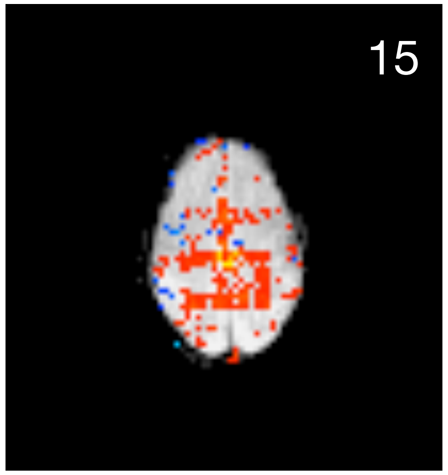
$$\mathbf{Y} = \mathbf{X} \quad \boldsymbol{\beta} + \mathbf{E}$$

Issues:

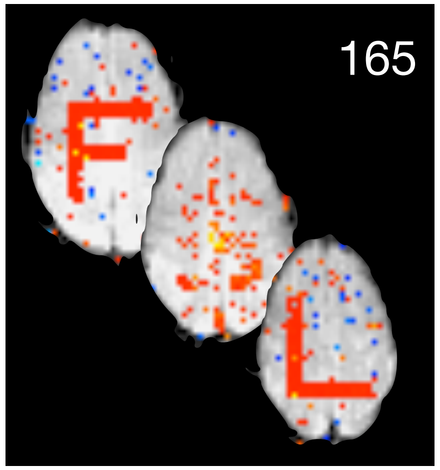
- Model Order Selection: how many components?

Model Order Selection

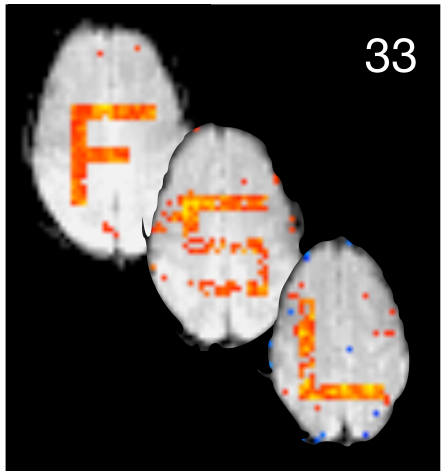
‘How many components’?



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

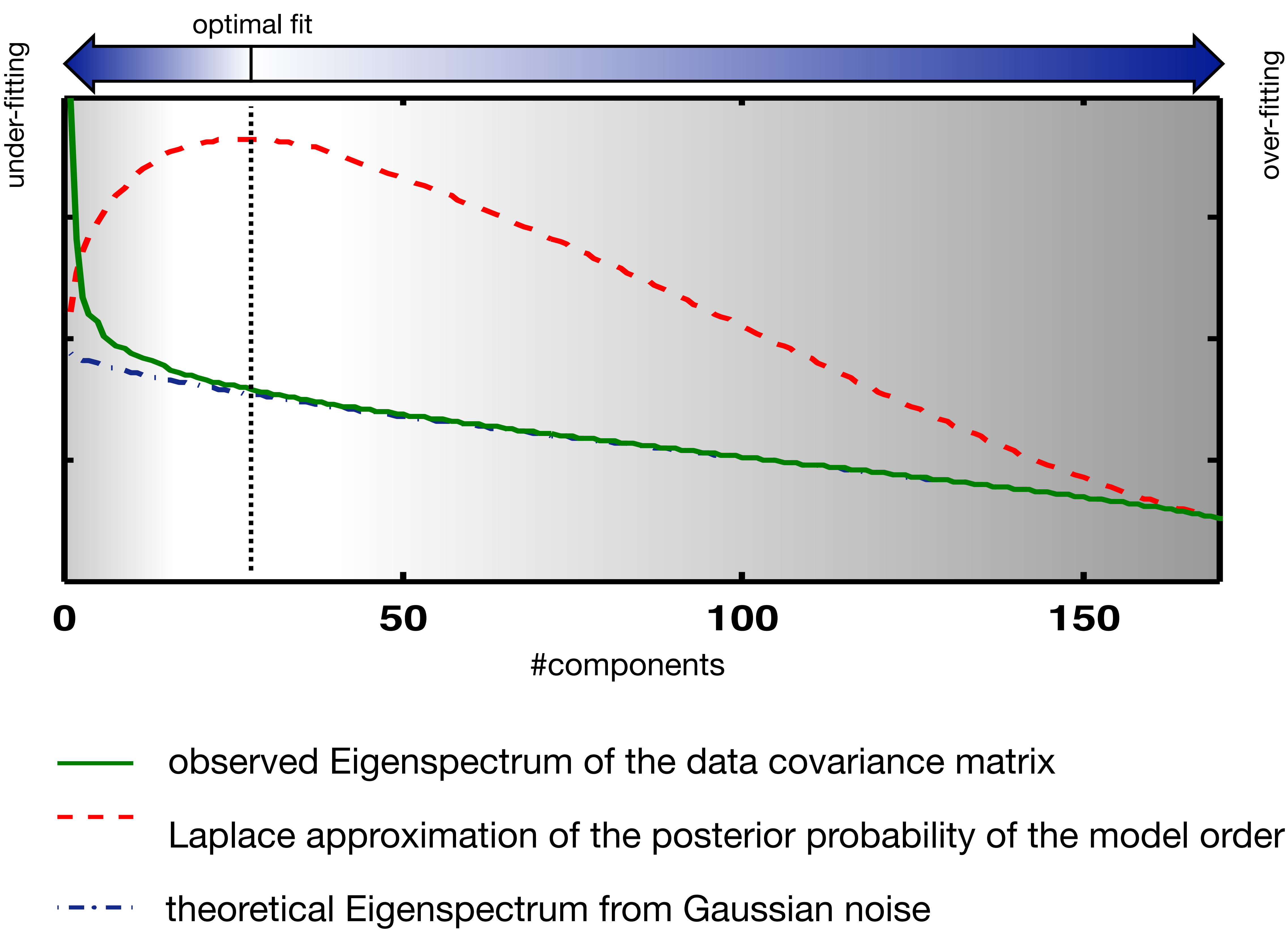


over-fitting: the inclusion of too many components leads to fragmentation of signal across multiple component maps, reducing the ability to identify the signals of interest



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

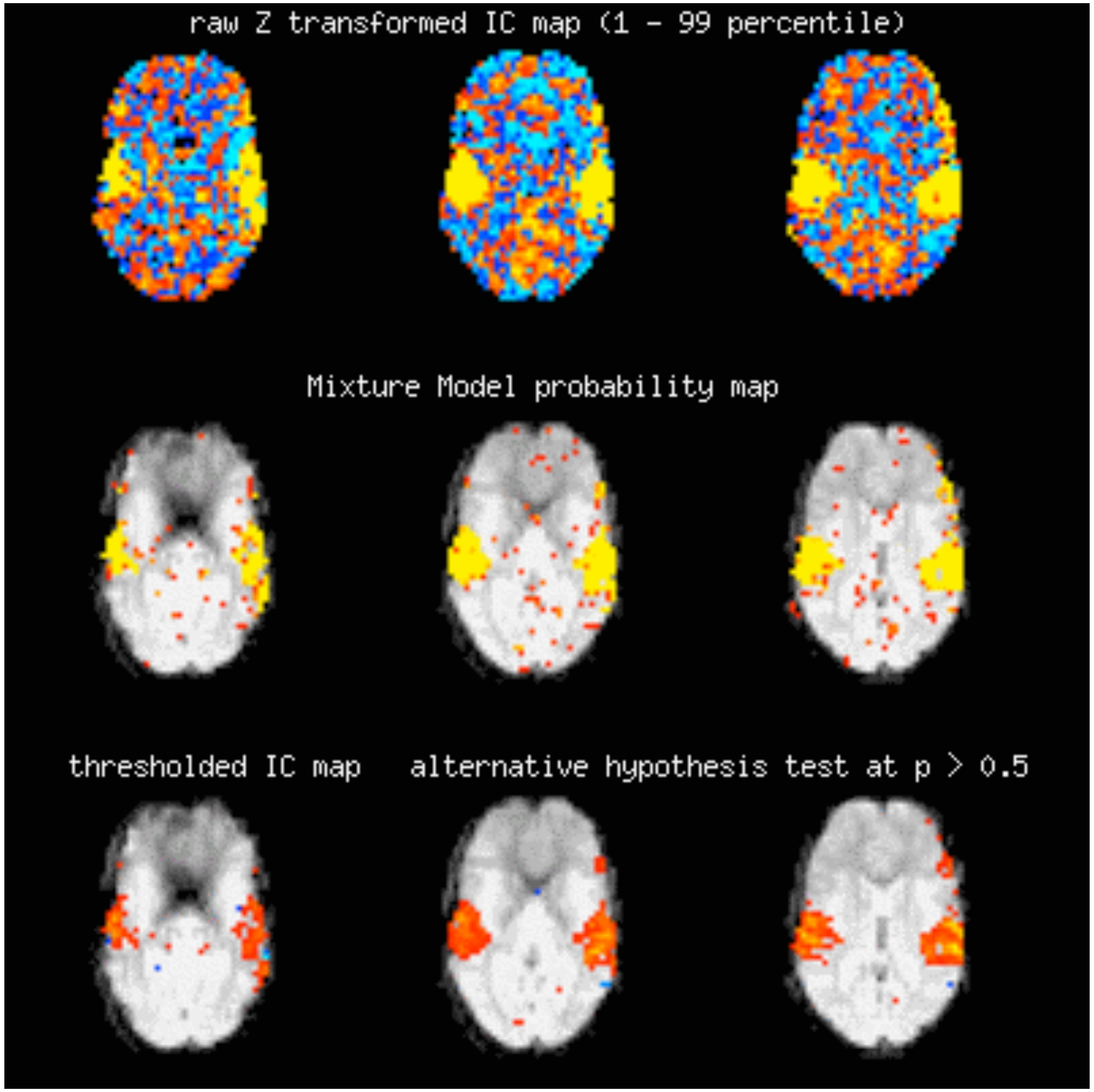


Some characteristics of MELODIC

- Multivariate voxel-based approach
- Exploratory “model-free” method to find interesting structure in the data
- Gives “spatially independent” components
- “Avoiding overfitting” through automatic model order selection
- “Thresholding" to remove background from main signal

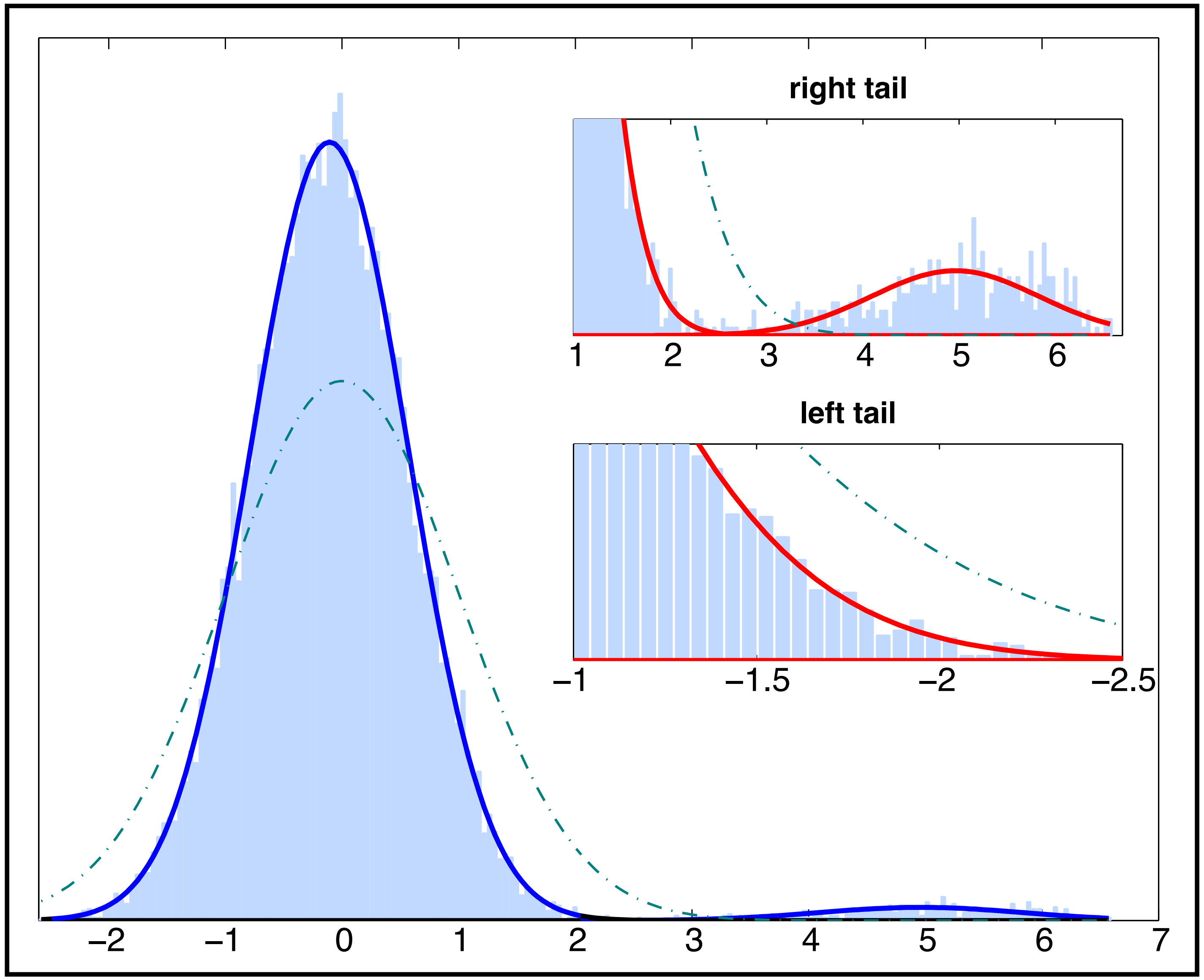
Thresholding

Thresholding

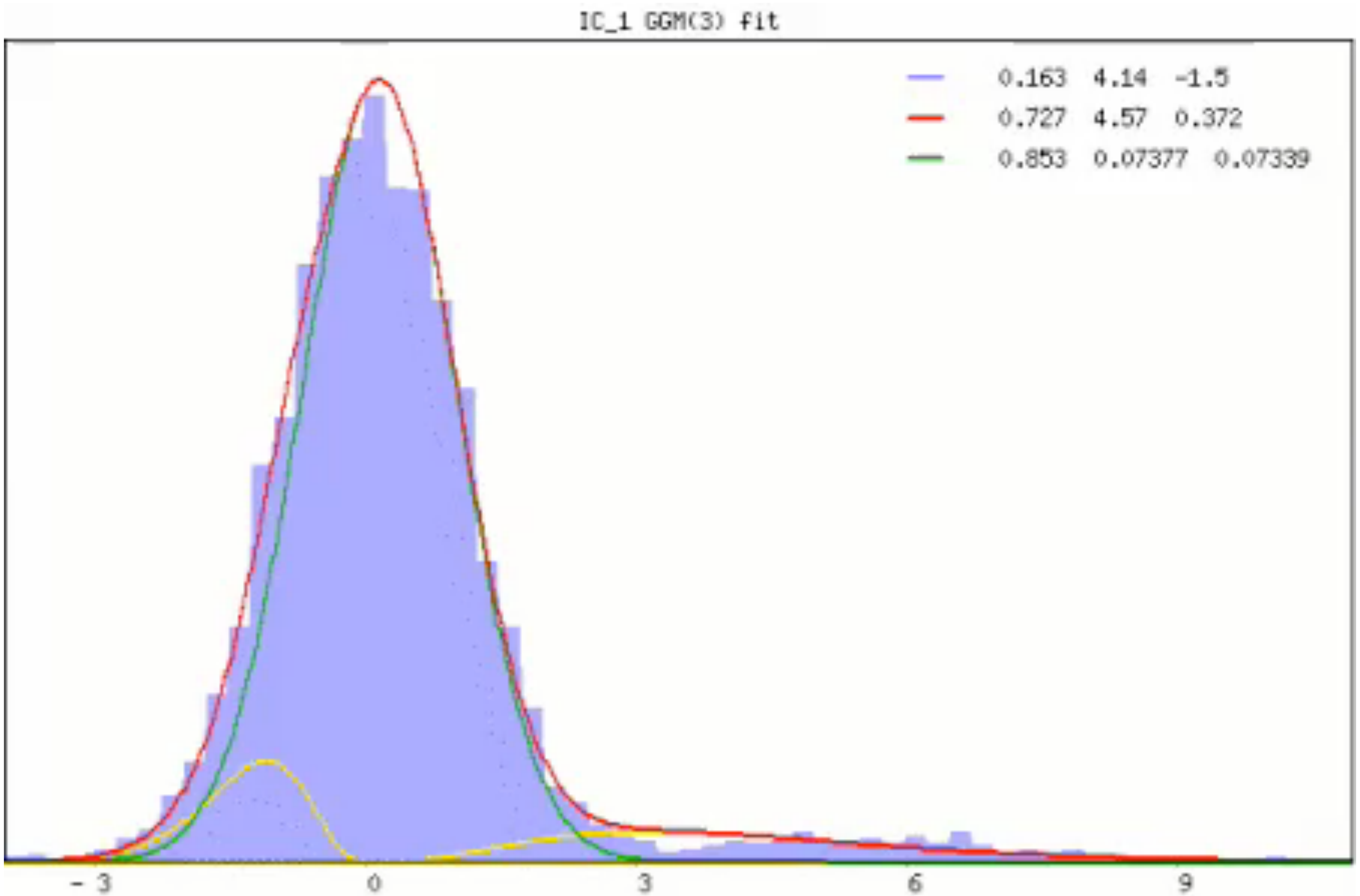


Thresholding

- 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)

Summary of part 1

- Resting state allows us to study the intrinsic organisation of the brain
- It focuses on connectivity and estimates functional components (resting state networks, RSN)
- Each RSN is characterised with a spatial map and a time course
- ICA can be used to characterise these RSNs
- MELODIC is FSL's ICA tool
  - Model-free, spatial independence, non-Gaussianity, model order selection, thresholding via mixture modelling

Resting-State fMRI:

ICA and Dual Regression

Rezvan Farahibozorg

FSL Course 2024

19 June, Osaka, Japan

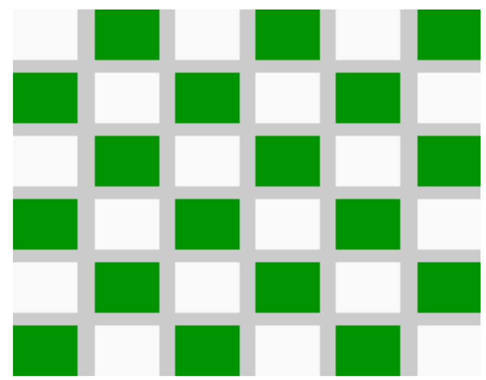
Resting state fMRI and ICA

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

The goal of single subject ICA is artefact detection

FMRI inferential path

Experiment



Physiology

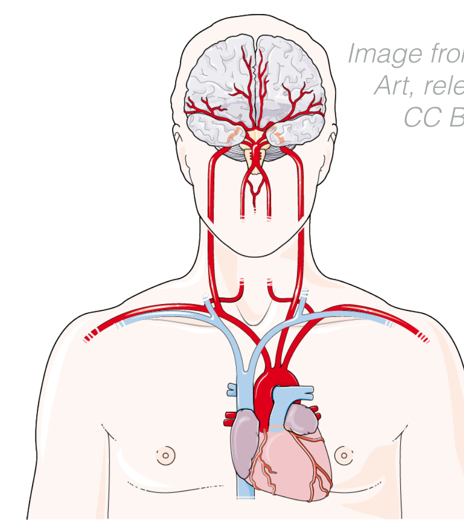
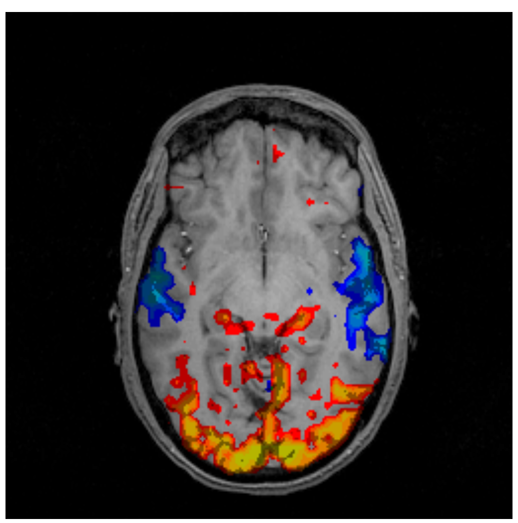


Image from Servier Medical Art, released under the CC BY 3.0 license

Interpretation of final results



Analysis

MR Physics

Image from mos.ru, released under the CC BY 4.0 license

Variability in FMRI

Experiment

suboptimal event timing,
inefficient design, etc.

Physiology

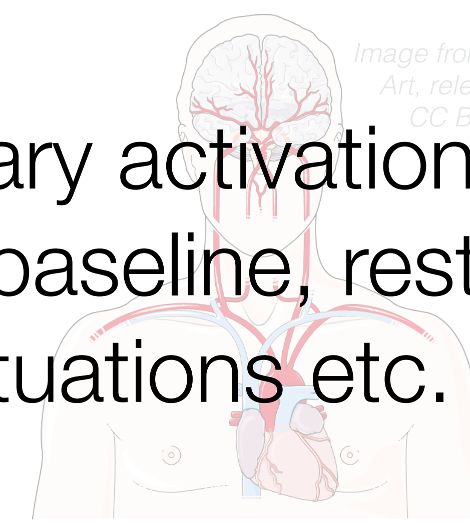
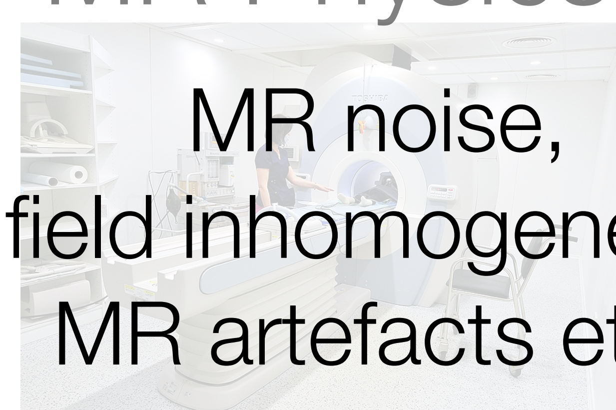


Image from Servier Medical Art, released under the CC BY 3.0 license

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

MR Physics



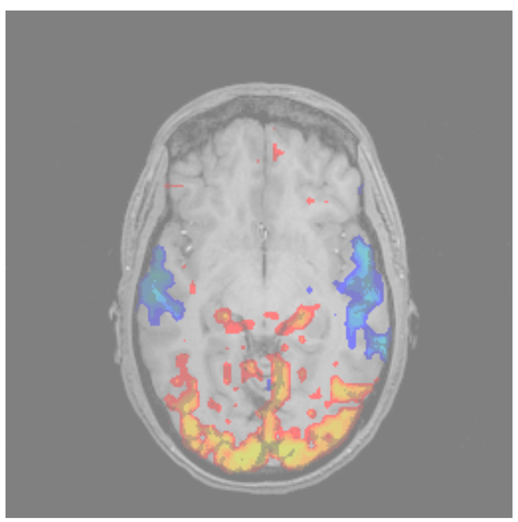
MR noise,
field inhomogeneity,
MR artefacts etc.

Image from mos.ru, released under the CC BY 4.0 license

Analysis

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

Interpretation of final results

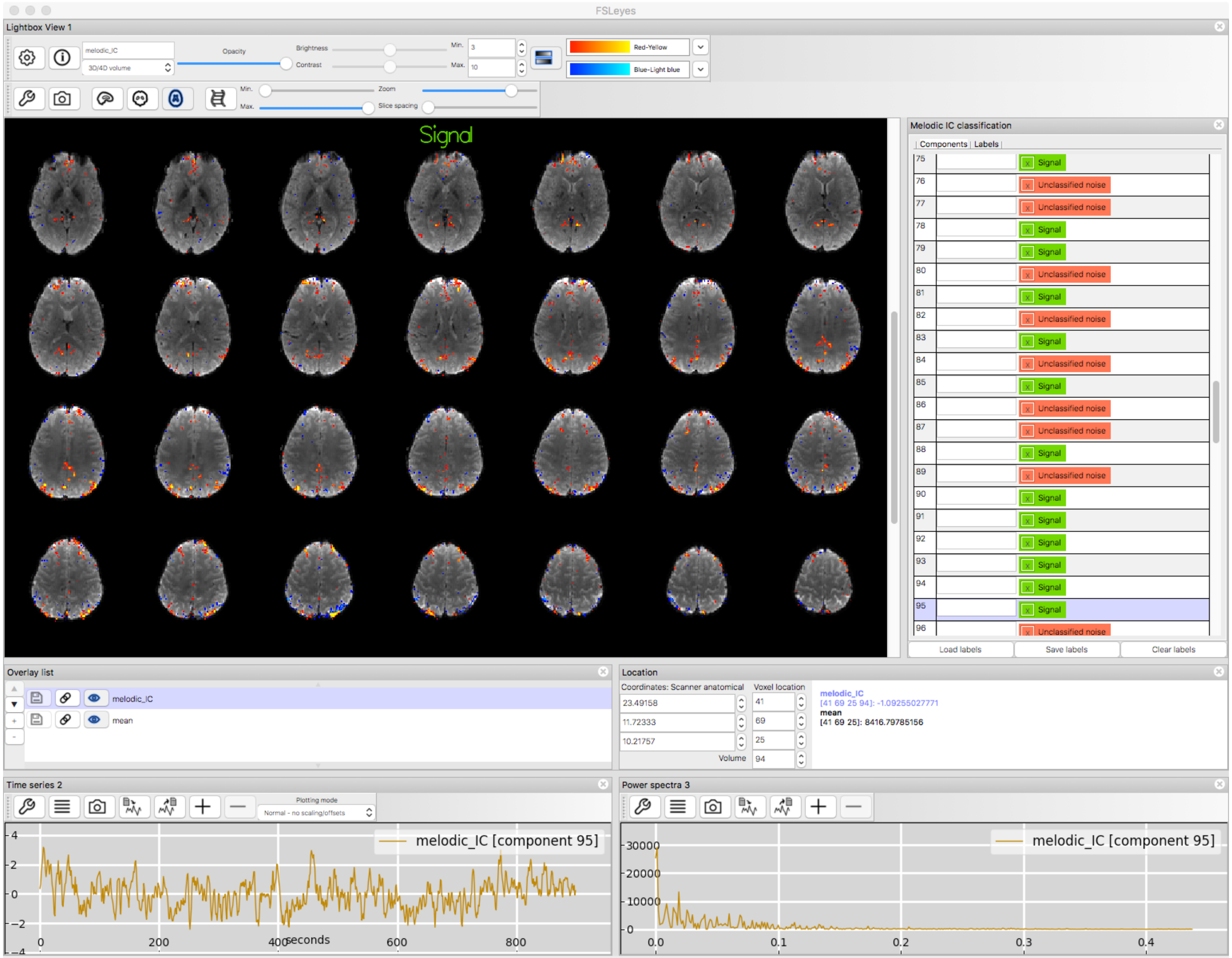


Artefact detection

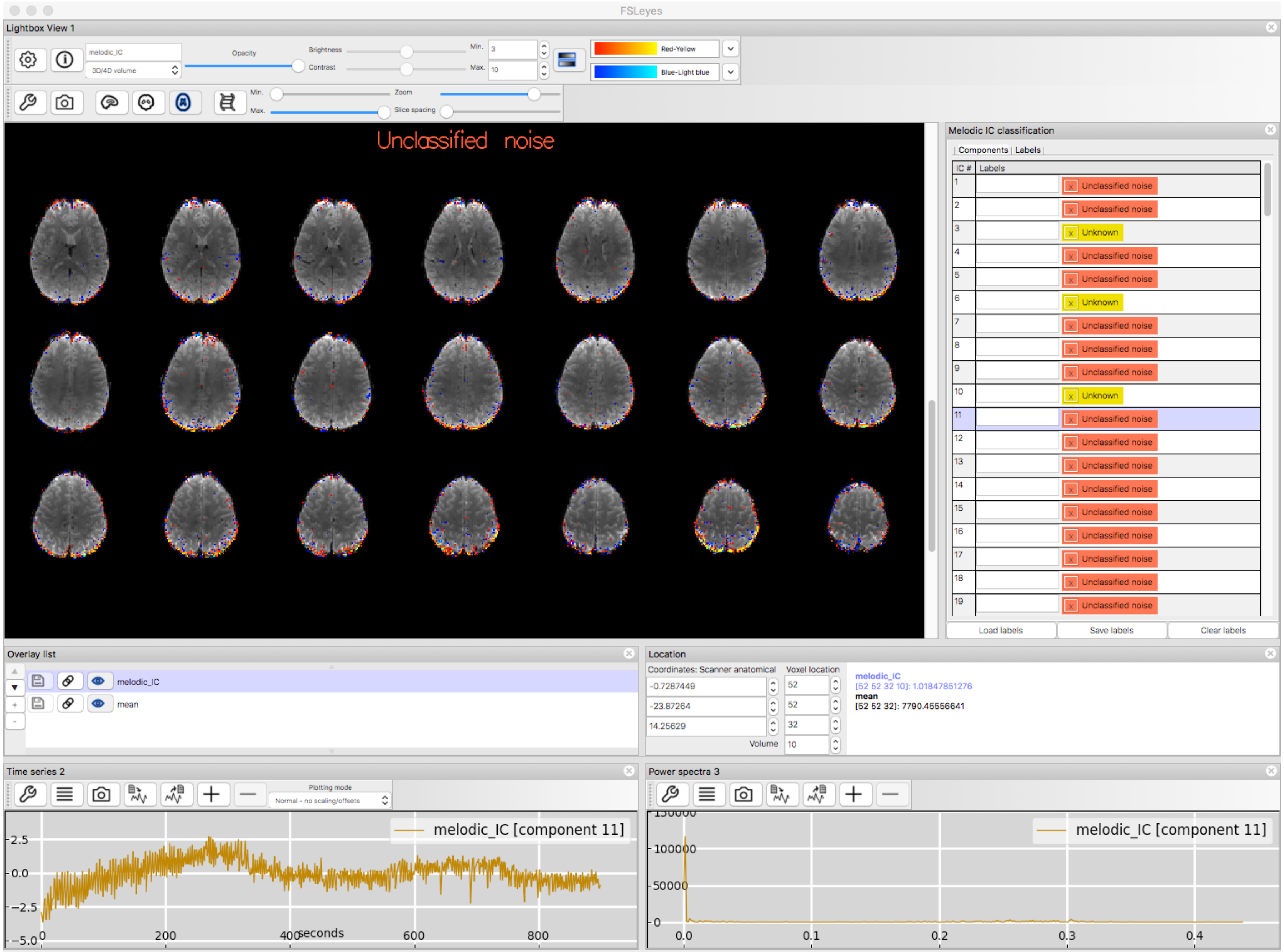
- Resting state FMRI data contain a variety of source processes
- Some of these sources are interesting signals and others are artefacts such as motion, cardiac pulsation, respiration
- Artifactual sources typically have unknown spatial and temporal extent and cannot easily be modelled accurately
- Good news: they are spatiotemporally distinct from true signals in the brain
- Therefore, ICA which is an exploratory tool which requires minimal knowledge of spatiotemporal characteristics of components can be used to identify artefacts, a.k.a, **Structured noise**.
- This is done at **single-subject level.**

Manual labelling of ICA components as signal vs artefact

FSLeyes Melodic Mode

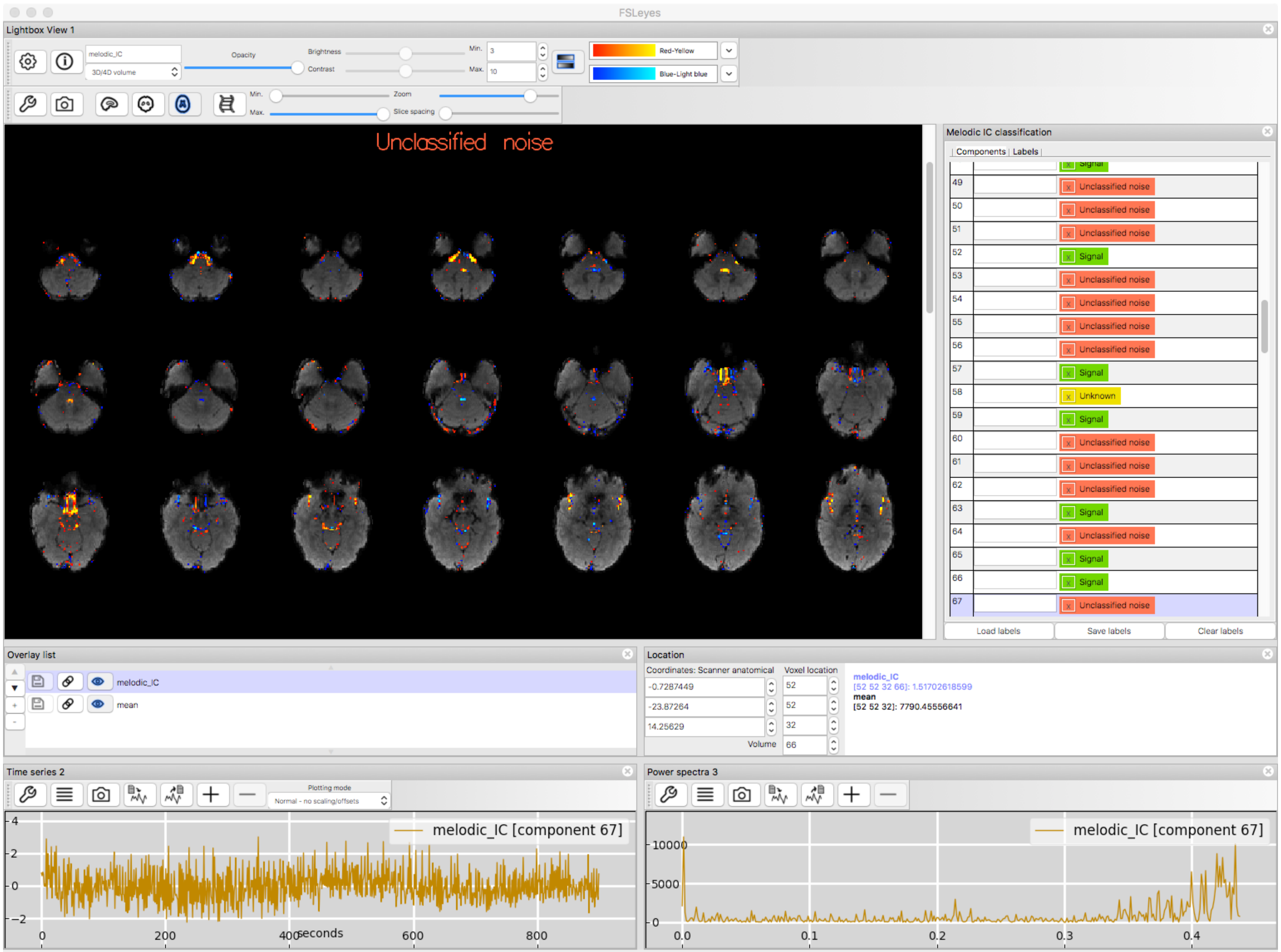


motion



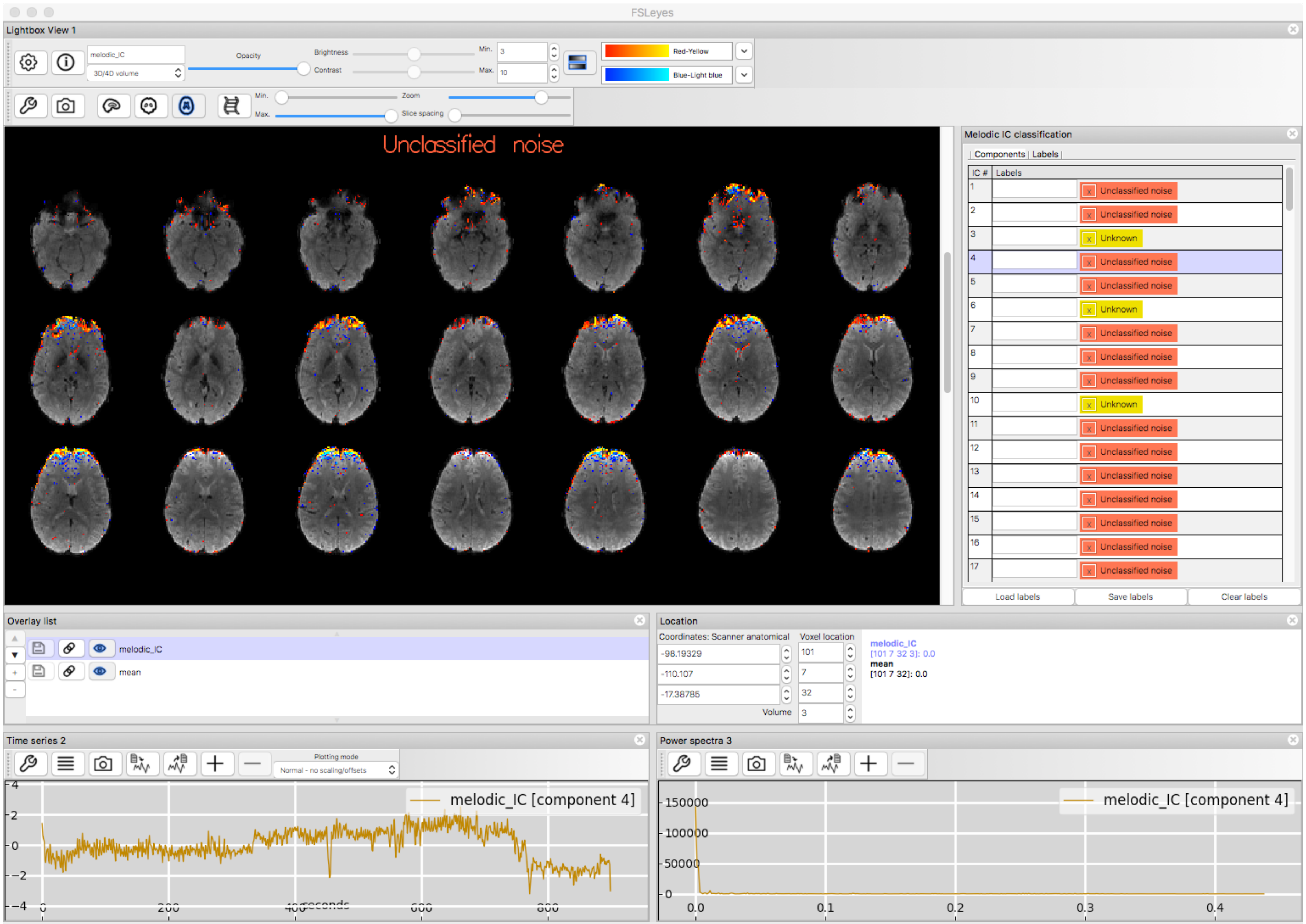
FSLeyes
Lightbox View 1
melodic_IC
3D/4D volume
Opacity
Brightness
Contrast
Min.
Max.
3
10
Red-Yellow
Blue-Light blue
Zoom
Slice spacing
Unclassified noise
Melodic IC classification
Components
Labels
IC #
1 Unclassified noise
2 Unclassified noise
3 Unknown
4 Unclassified noise
5 Unclassified noise
6 Unknown
7 Unclassified noise
8 Unclassified noise
9 Unclassified noise
10 Unknown
11 Unclassified noise
12 Unclassified noise
13 Unclassified noise
14 Unclassified noise
15 Unclassified noise
16 Unclassified noise
17 Unclassified noise
18 Unclassified noise
19 Unclassified noise
Load labels
Save labels
Clear labels
Overlay list
melodic_IC
mean
Location
Coordinates: Scanner anatomical
Voxel location
-0.7287449
-23.87264
14.25629
52
52
32
Volume
10
melodic_IC
[52 52 32 10]: 1.01847851276
mean
[52 52 32]: 7790.45556641
Time series 2
Plotting mode
Normal - no scaling/offsets
melodic_IC [component 11]
2.5
0.0
−2.5
−5.0
0
200
400
600
800
seconds
Power spectra 3
melodic_IC [component 11]
150000
100000
50000
0
0.0
0.1
0.2
0.3
0.4

cardiac

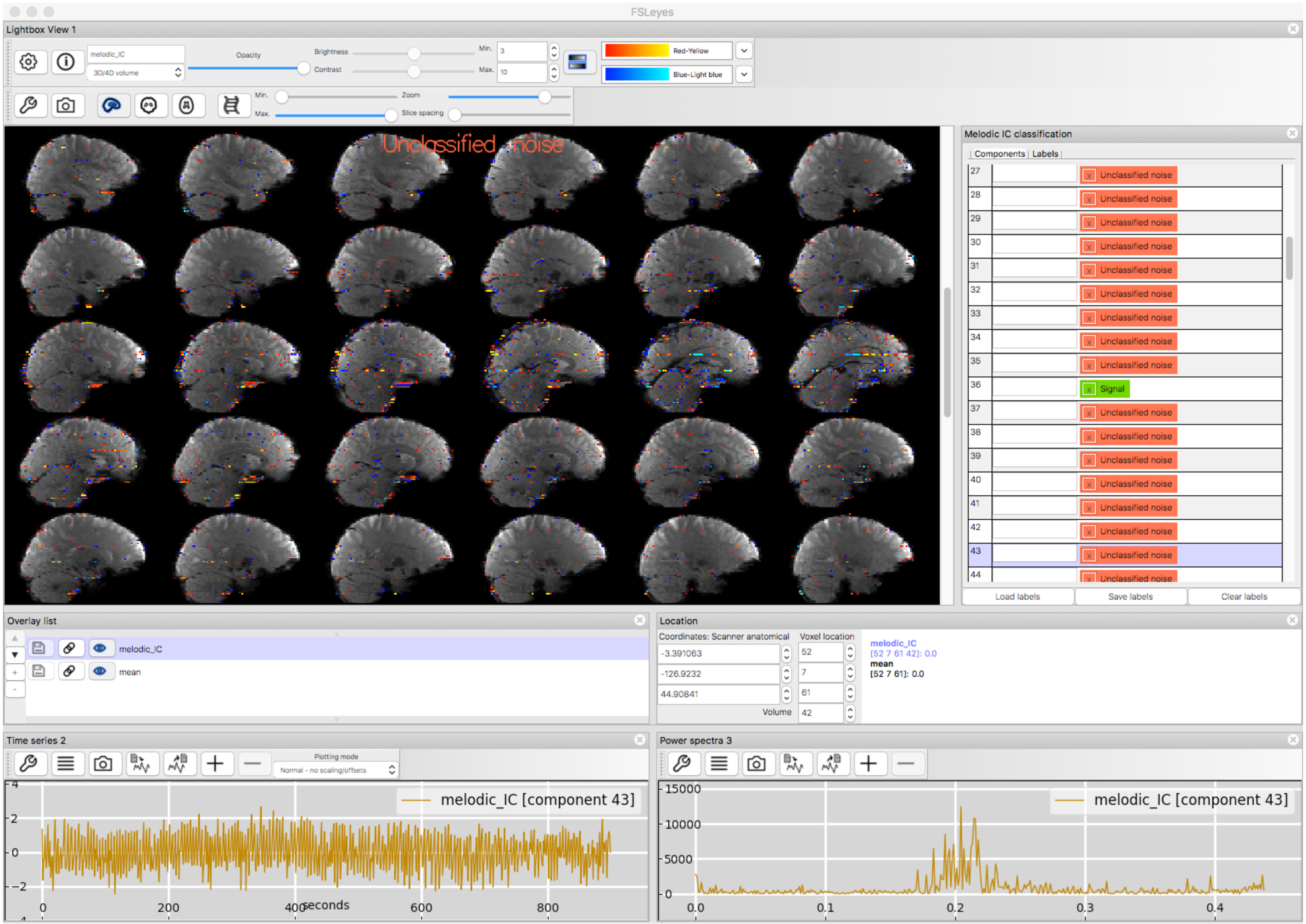


FSLeyes
Lightbox View 1
Unclassified noise
Melodic IC classification
Components | Labels
49 Unclassified noise
50 Unclassified noise
51 Unclassified noise
52 Signal
53 Unclassified noise
54 Unclassified noise
55 Unclassified noise
56 Unclassified noise
57 Signal
58 Unknown
59 Signal
60 Unclassified noise
61 Unclassified noise
62 Unclassified noise
63 Signal
64 Unclassified noise
65 Signal
66 Signal
67 Unclassified noise
Load labels
Save labels
Clear labels
Overlay list
melodic_IC
mean
Location
melodic_IC
[52 52 32 66]: 1.51702618599
mean
[52 52 32]: 7790.45556641
Time series 2
melodic_IC [component 67]
seconds
Power spectra 3
melodic_IC [component 67]

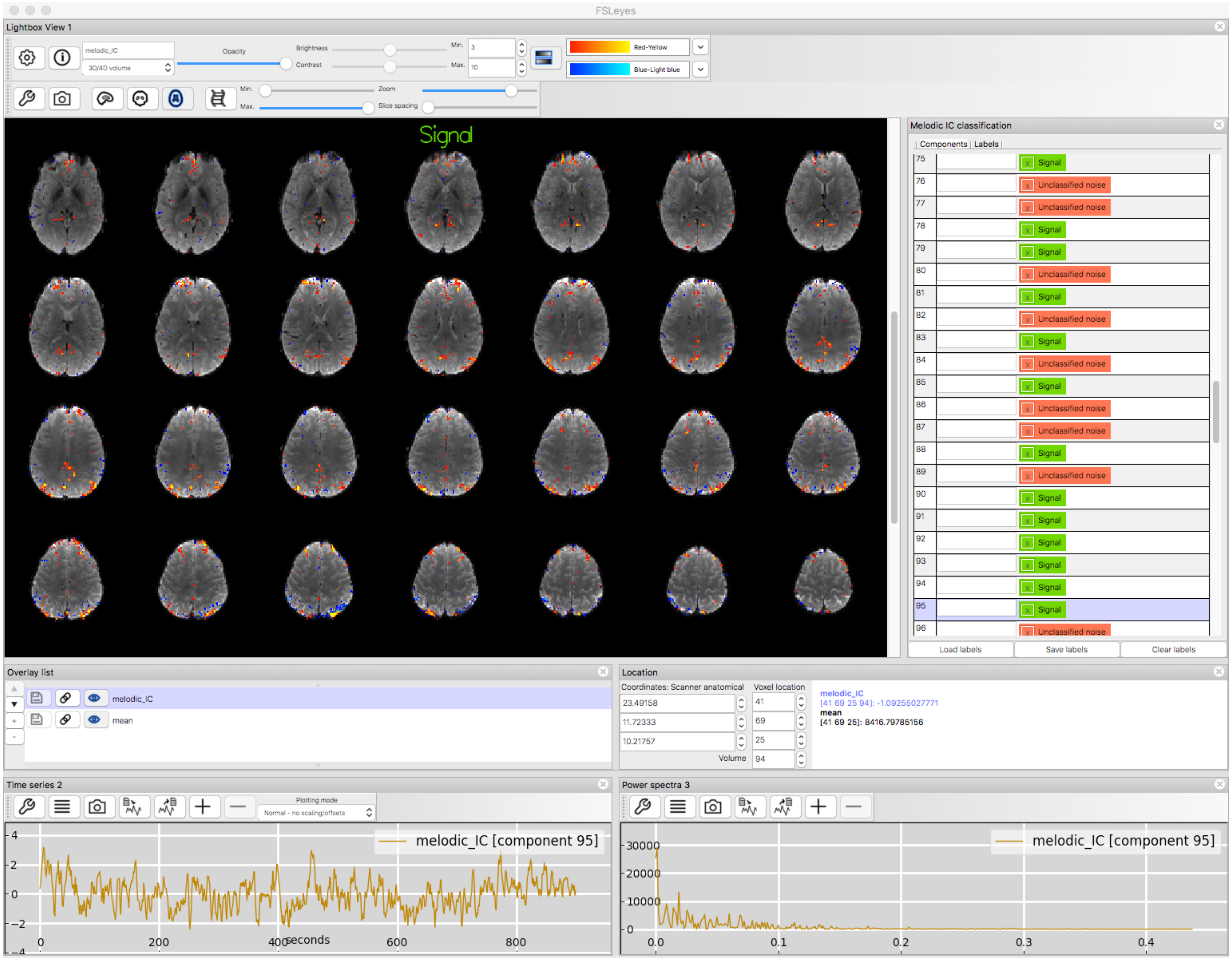
susceptibility motion



multiband

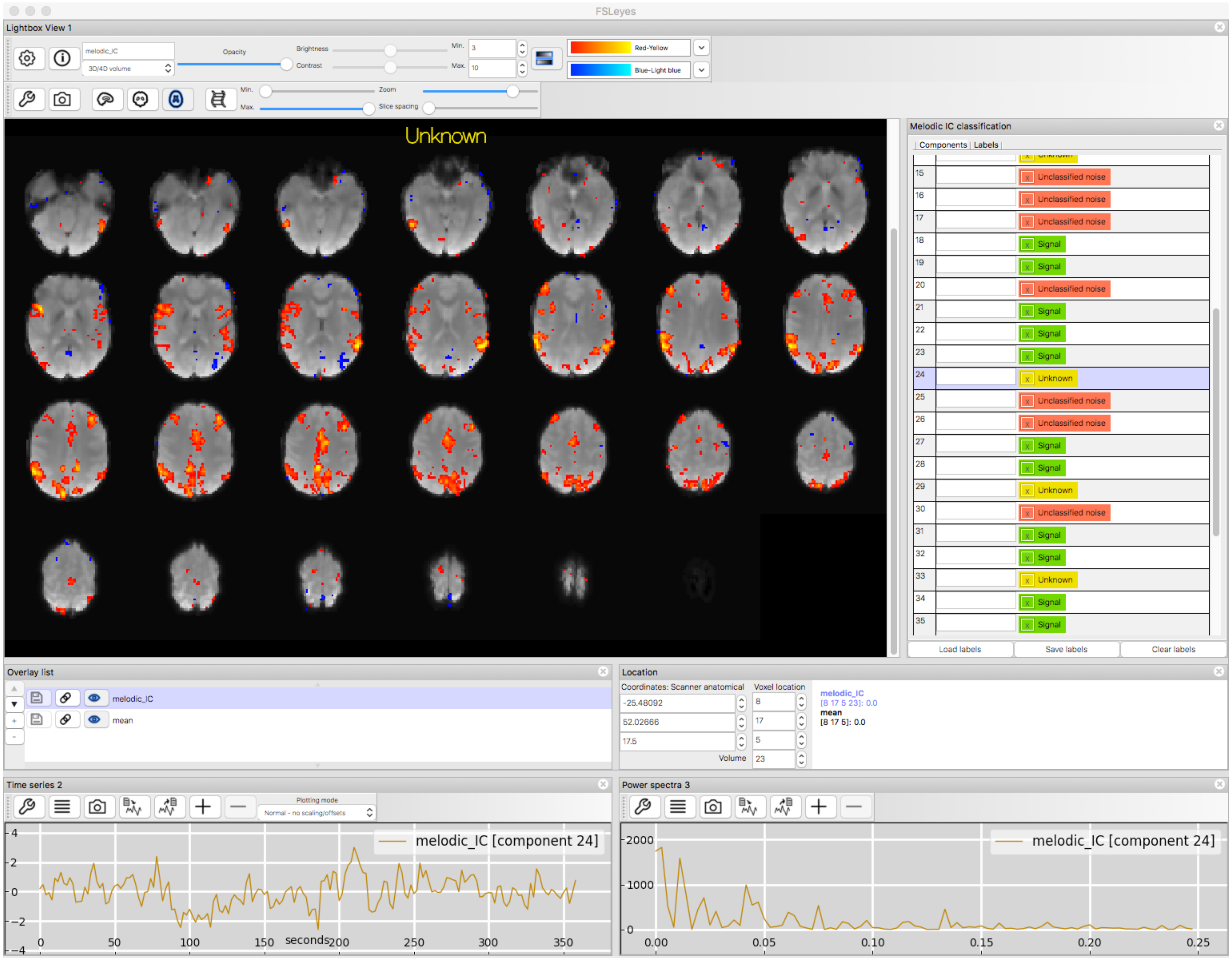

FSLeyes
Lightbox View 1
Unclassified noise
Melodic IC classification
Components | Labels
Signal
Unclassified noise
Load labels
Save labels
Clear labels
Overlay list
melodic_IC
mean
Location
Coordinates: Scanner anatomical
Voxel location
Volume
Time series 2
Plotting mode
melodic_IC [component 43]
seconds
Power spectra 3
melodic_IC [component 43]

signal

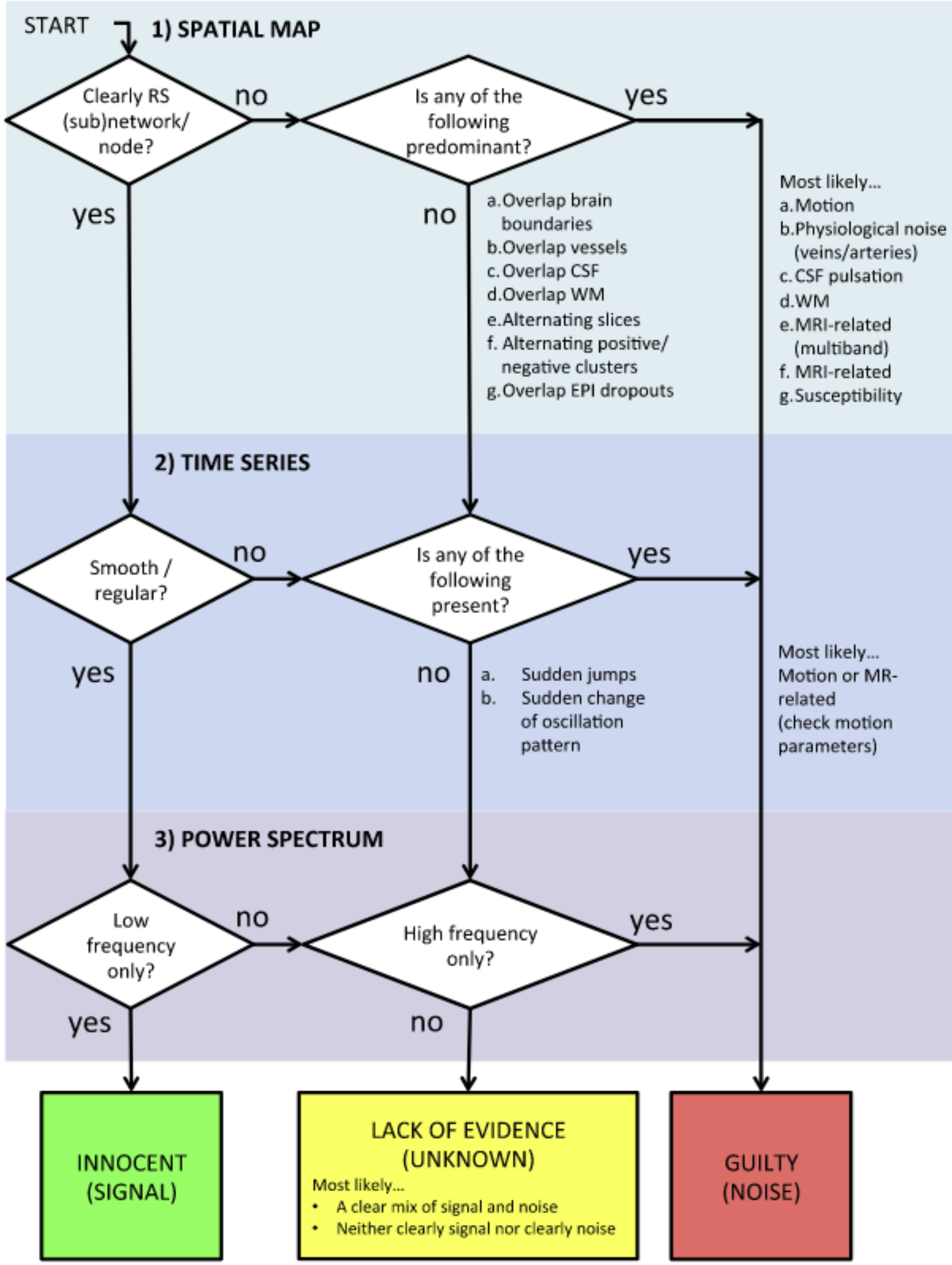


FSLeyes
Lightbox View 1
Signal
Melodic IC classification
Components
Labels
Load labels
Save labels
Clear labels
Overlay list
melodic_IC
mean
Location
Time series 2
Power spectra 3
melodic_IC [component 95]
seconds

effects of scan parameters



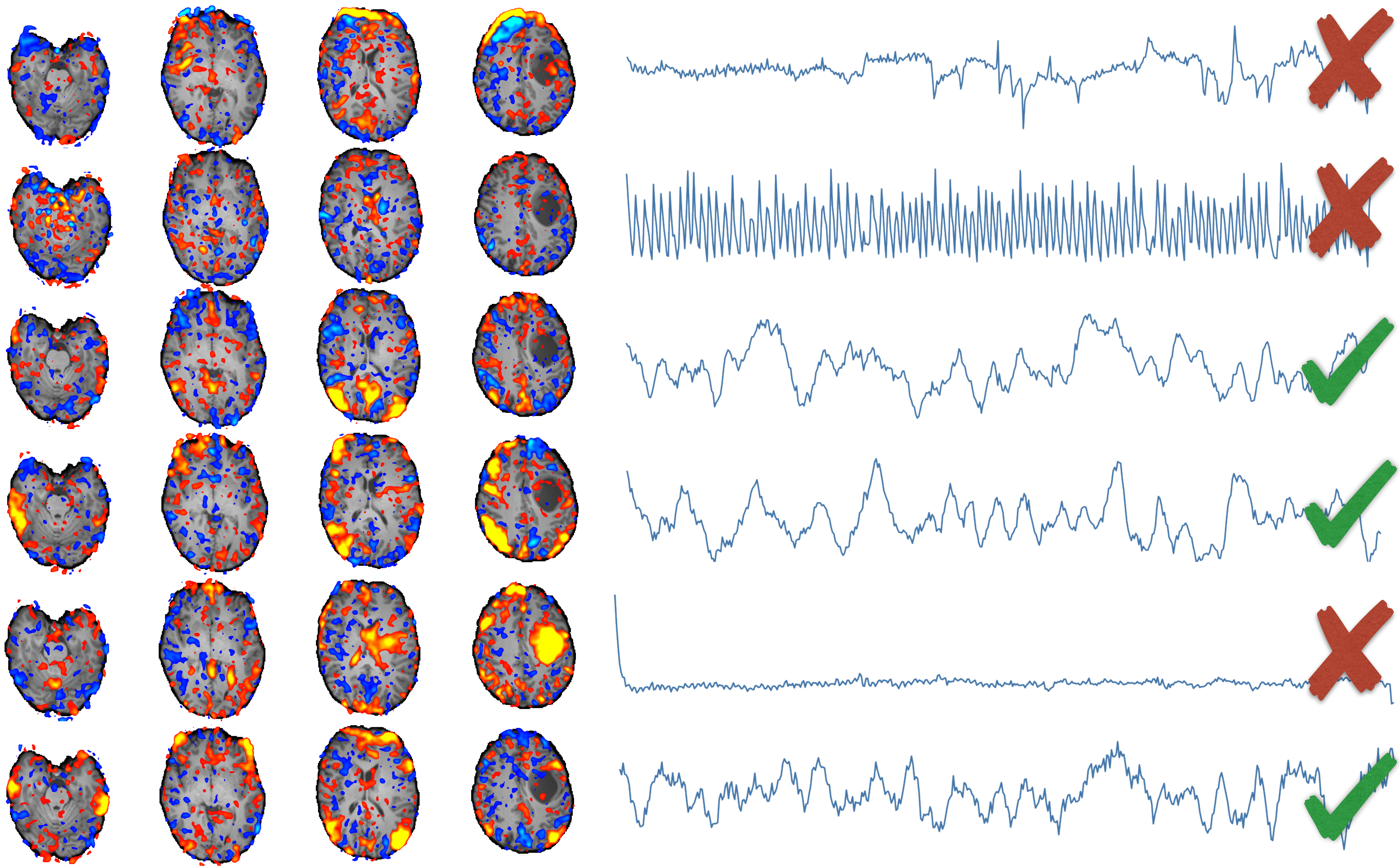
manual classification



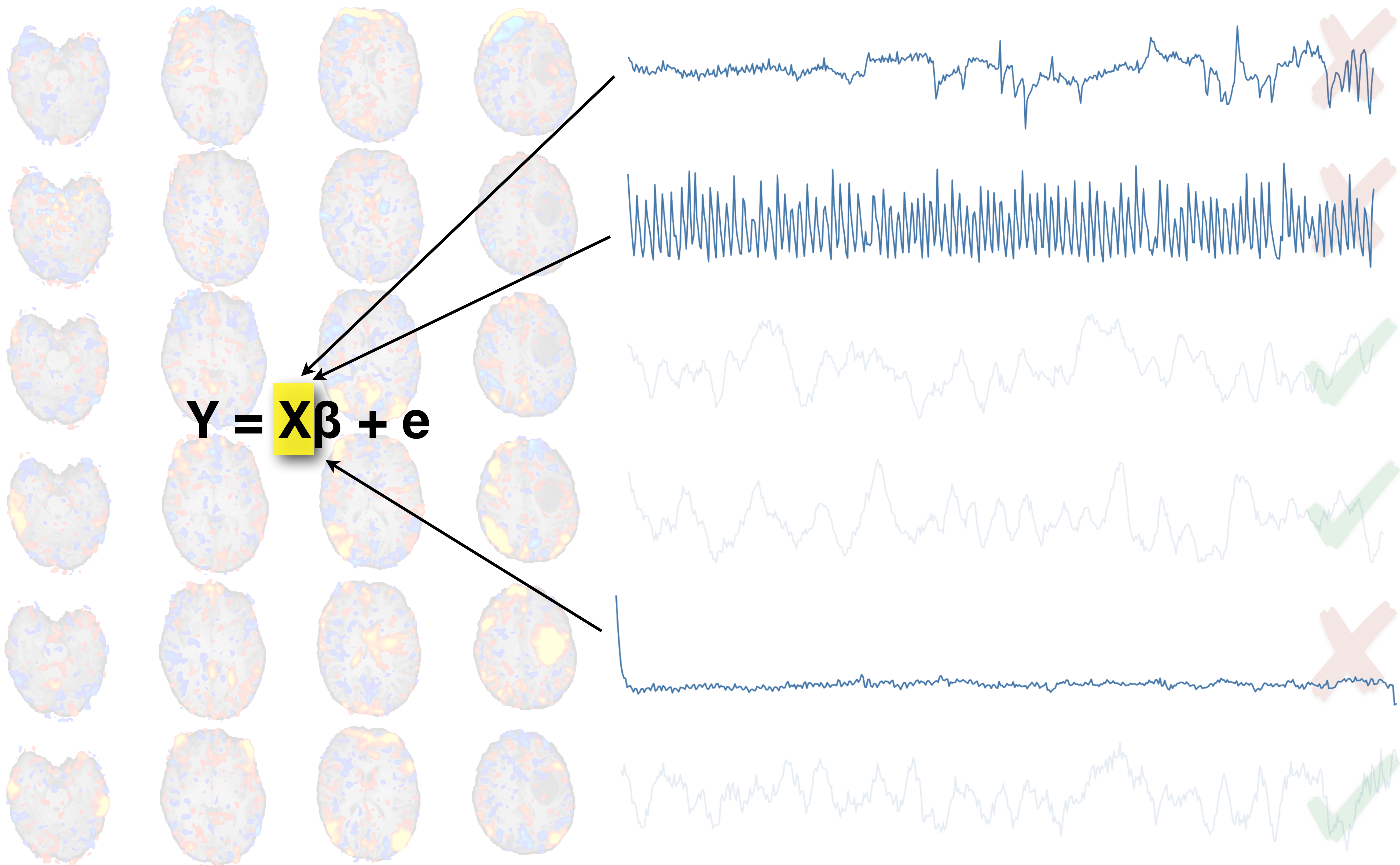
Griffanti et al (2016).
https://doi.org/10.1016/j.neuroimage.2016.12.036

Removing artefacts

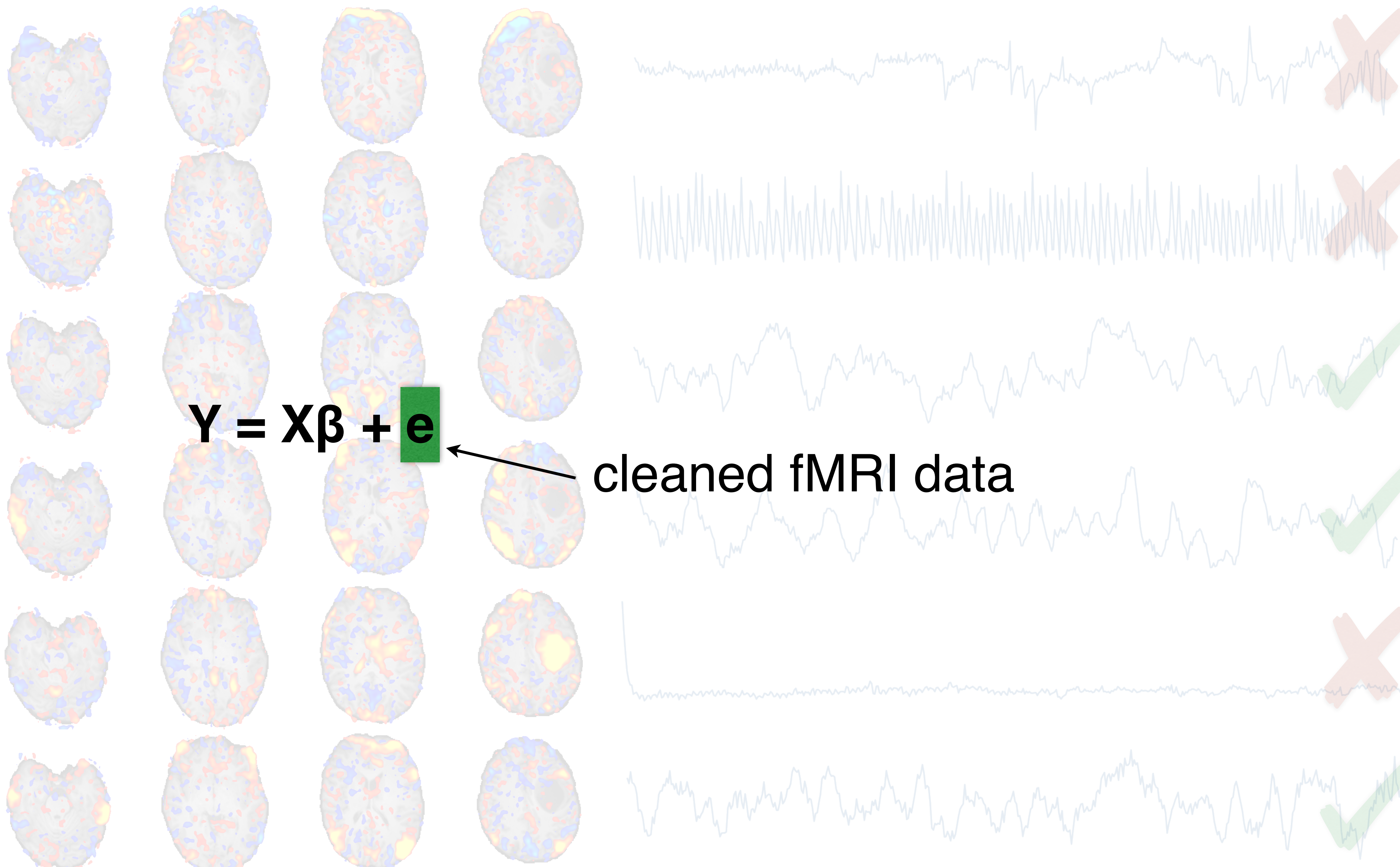
ICA-based denoising



ICA-based denoising



ICA-based denoising

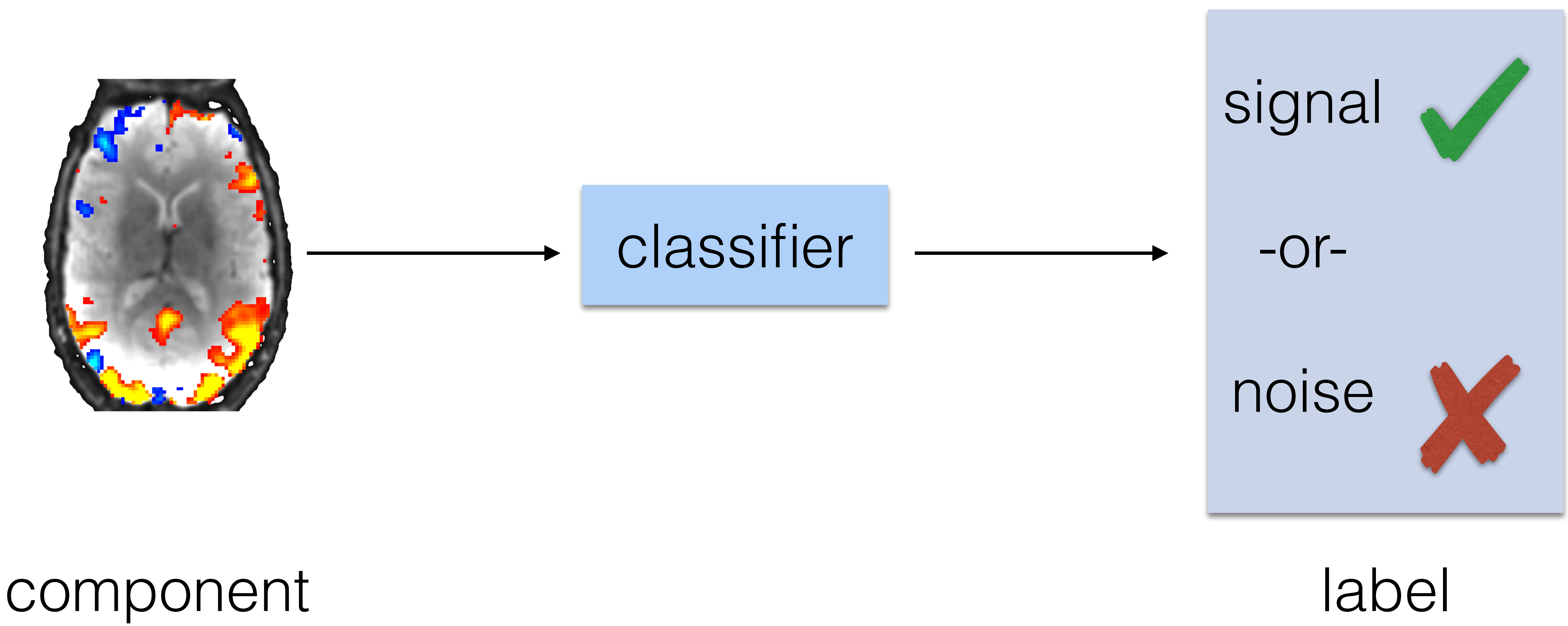


$$Y = X\beta + \boxed{e}$$

cleaned fMRI data

Semi-automatic artefact detection

Semi-automatic classification



semi-automatic classification

- FIX (https://fsl.fmrib.ox.ac.uk/fsl/docs/#/resting_state/fix)
  - Classifier with many features
  - Requires manually labelled training data
  - 99% accuracy on high-quality data
- The latest version is reimplemented in Python and installed as part of FSL, unlike previous MATLAB/R version that needed separate installation.

semi-automatic classification

- FIX (https://fsl.fmrib.ox.ac.uk/fsl/docs/#/resting_state/fix)
  - Classifier with many features
  - Requires manually labelled training data
  - 99% accuracy on high-quality data
- ICA-AROMA (github.com/rhr-pruim/ICA-AROMA)
  - Simple classifier with only 4 features
  - No training data required
  - Mainly designed for motion artefacts

Resting state fMRI and ICA

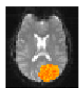
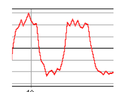
- Introduction to resting state
- Independent Component Analysis
- Single-subject ICA
- Multi-subject ICA
- Dual regression

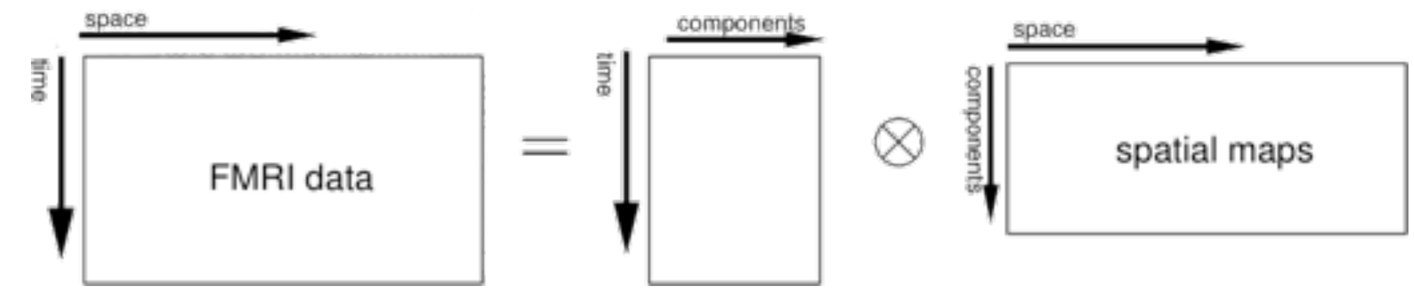
The goal of multi subject ICA is to characterise Resting State Networks (RSNs)

Different ICA models

Single-Session ICA

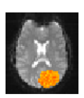
each ICA component comprises:

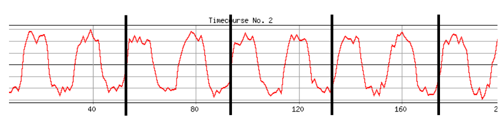
 spatial map & timecourse 

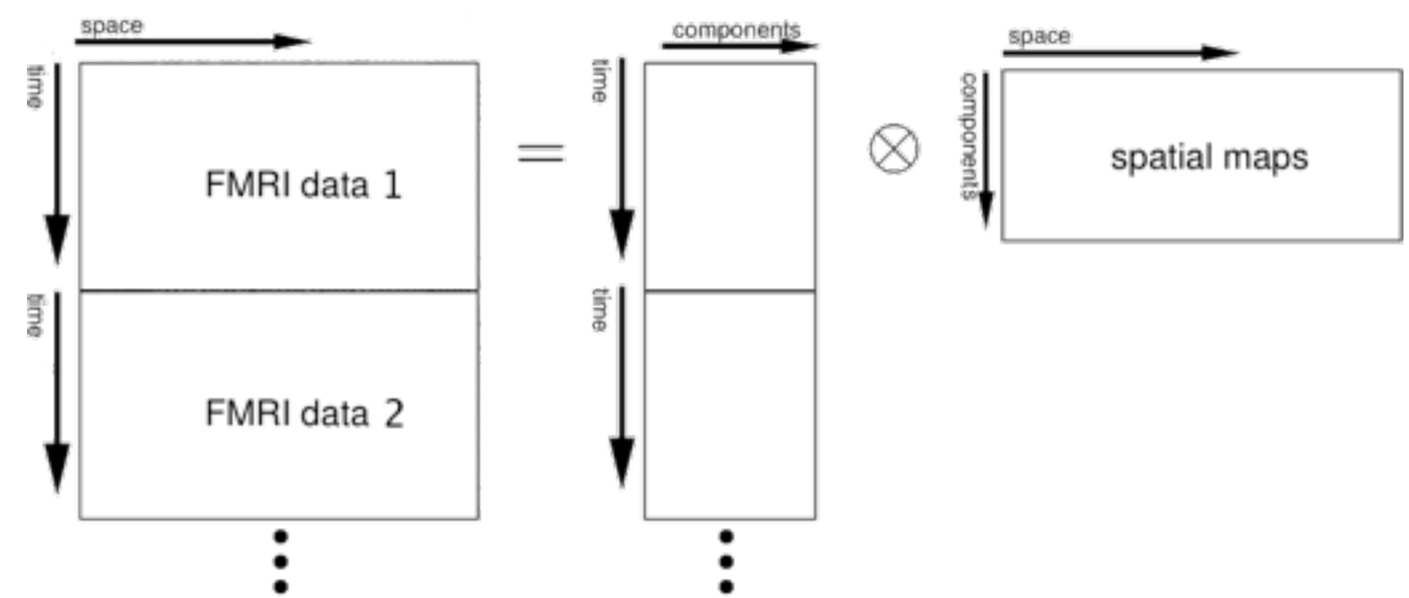


Multi-Session or Multi-Subject ICA: **Concatenation approach**

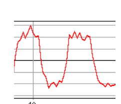
each ICA component comprises:

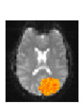
 spatial map & timecourse

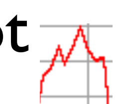
(that can be split up into subject-specific chunks) 

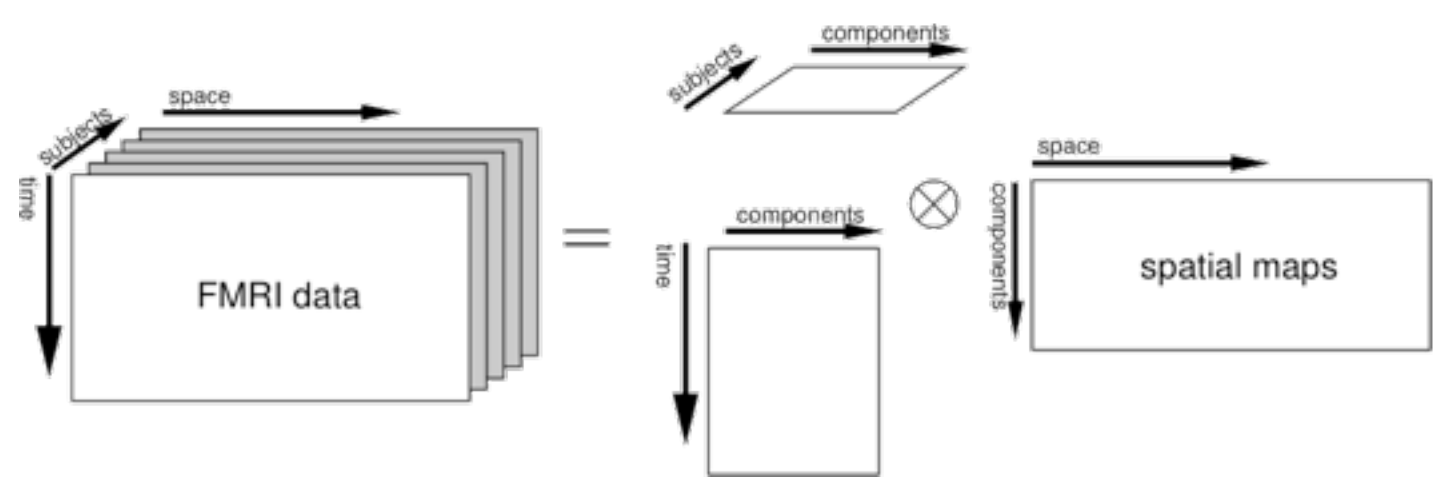


Multi-Session or Multi-Subject ICA: **Tensor-ICA approach**

each ICA component comprises: 

 spatial map, session-long-timecourse

& subject-strength plot 

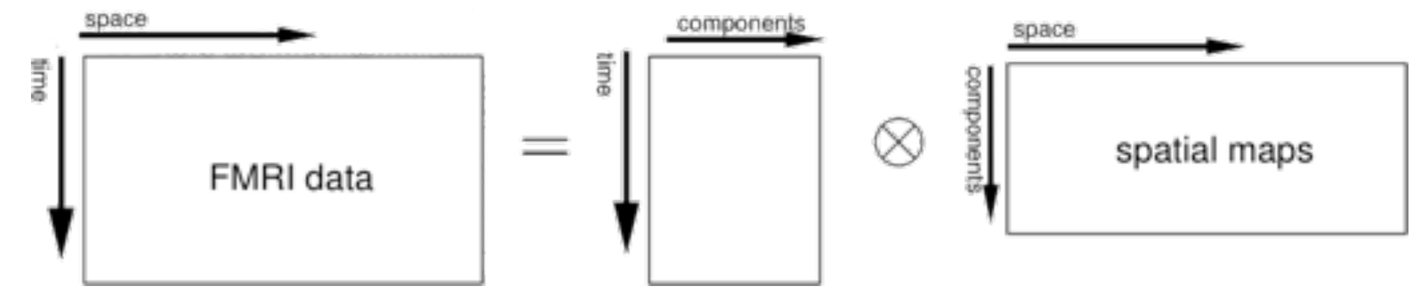


Different ICA models

Single-Session ICA

each ICA component comprises:

spatial map & timecourse

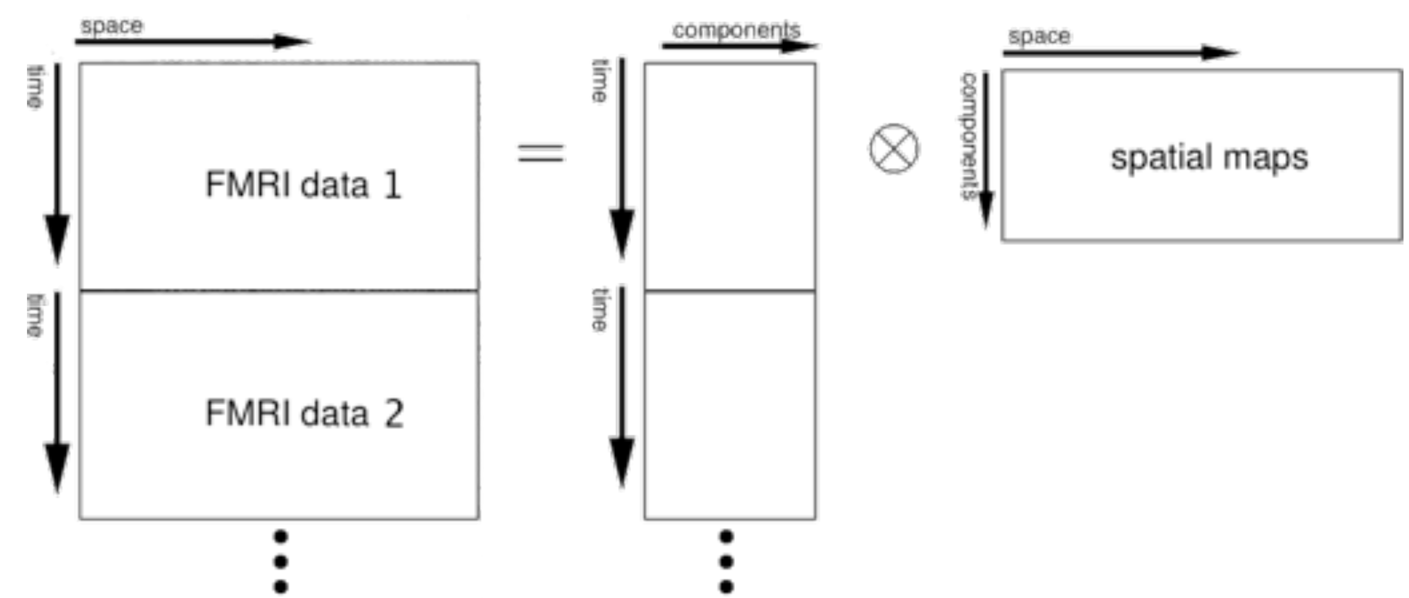


Multi-Session or Multi-Subject ICA: **Concatenation approach**

good when:

each subject has DIFFERENT timeseries

e.g. resting-state FMRI

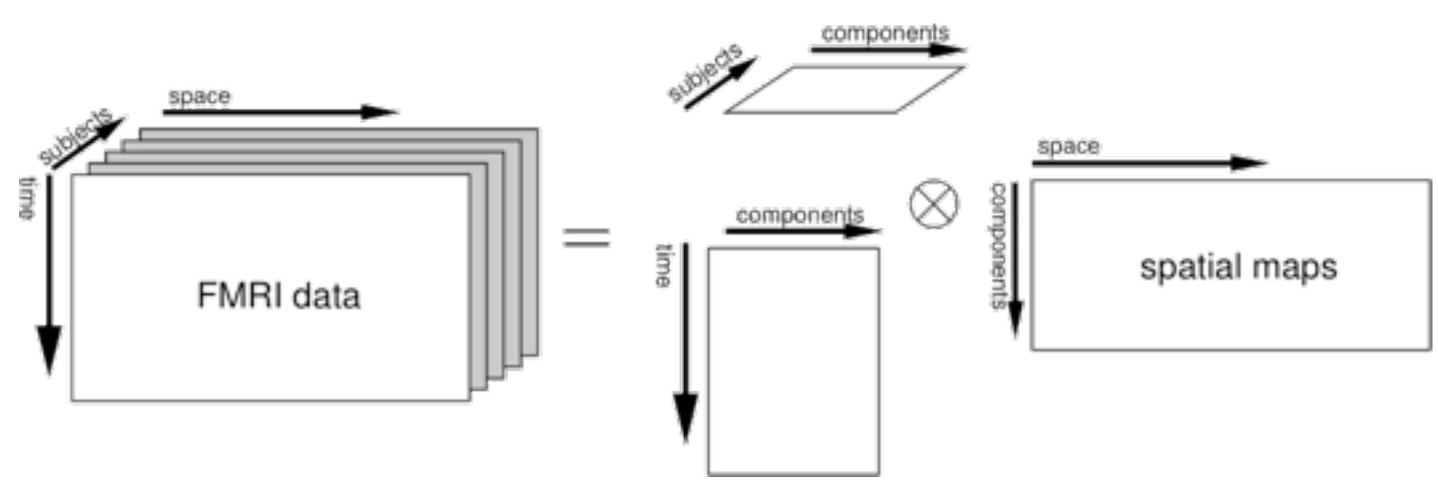


Multi-Session or Multi-Subject ICA: **Tensor-ICA approach**

good when:

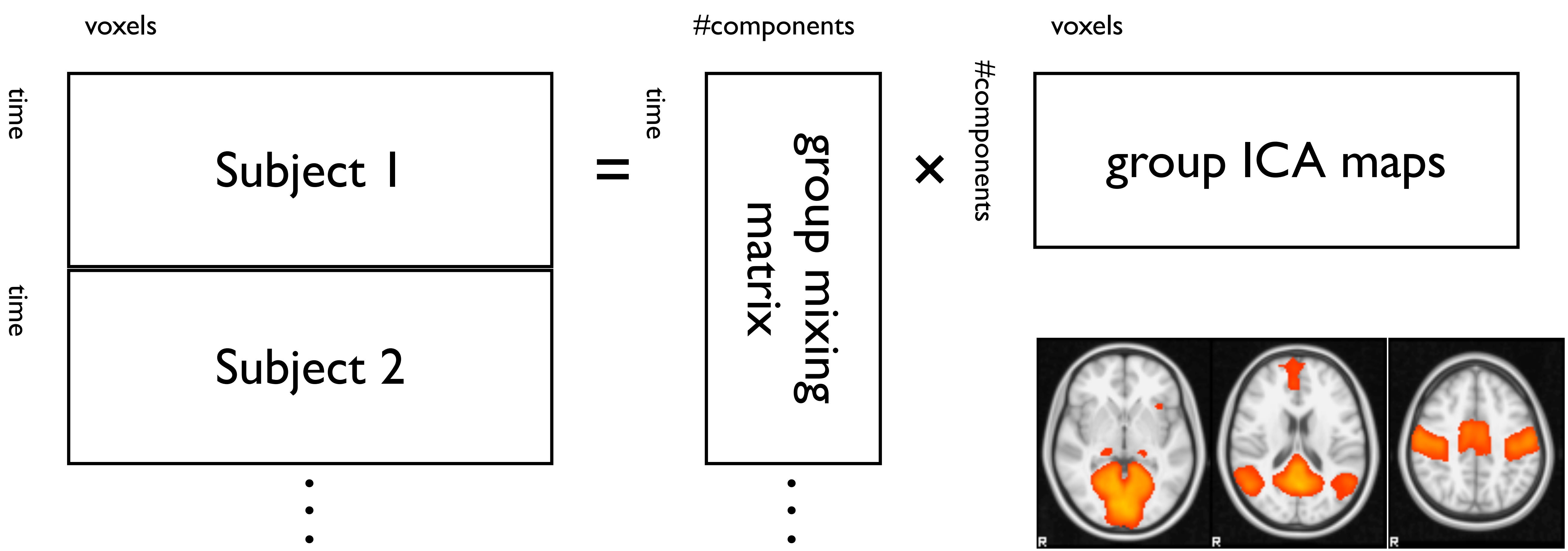
each subject has SAME timeseries

e.g. activation FMRI



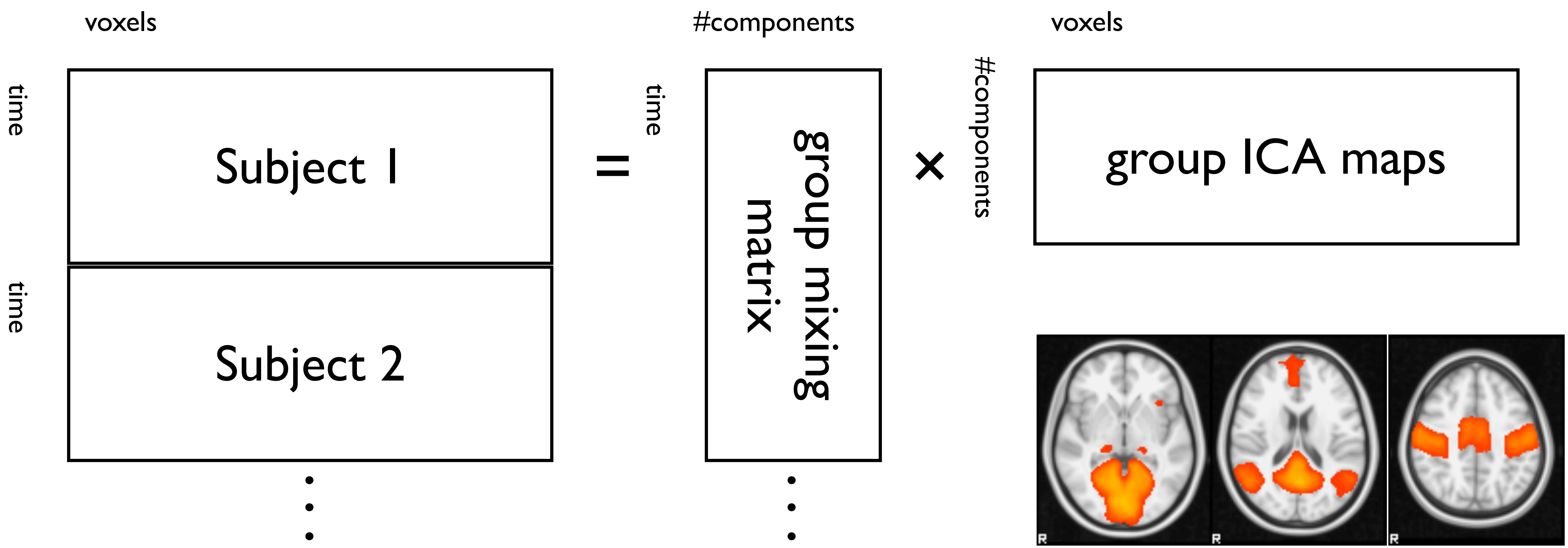
Concatenated ICA

- Concatenate all subjects' data temporally
- Then run ICA
- More appropriate than tensor ICA (for RSNs)



Concatenated ICA

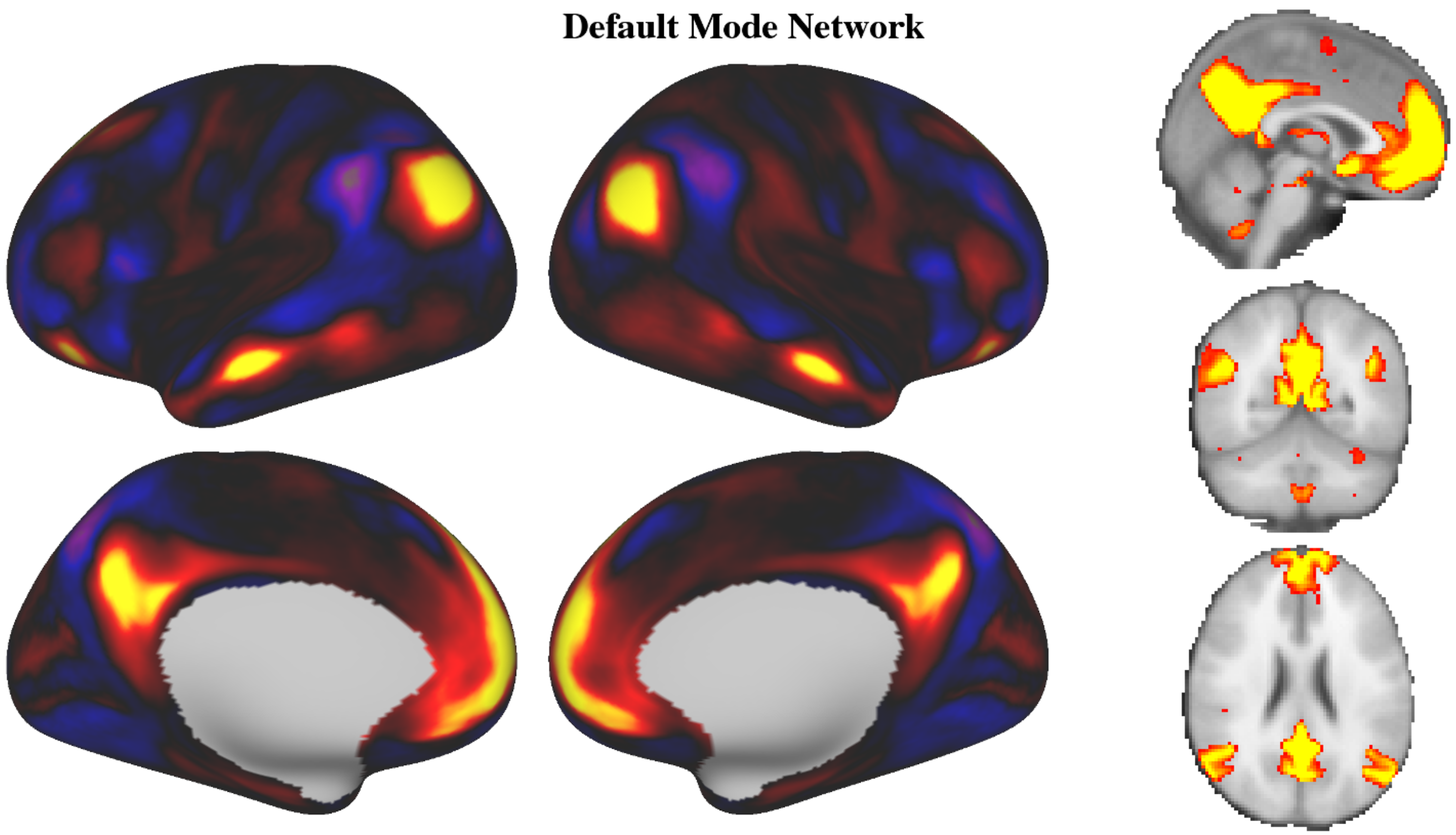
- Data sets must be registered to a common space (anatomical alignment)
- Memory optimisation trick (called MIGP)

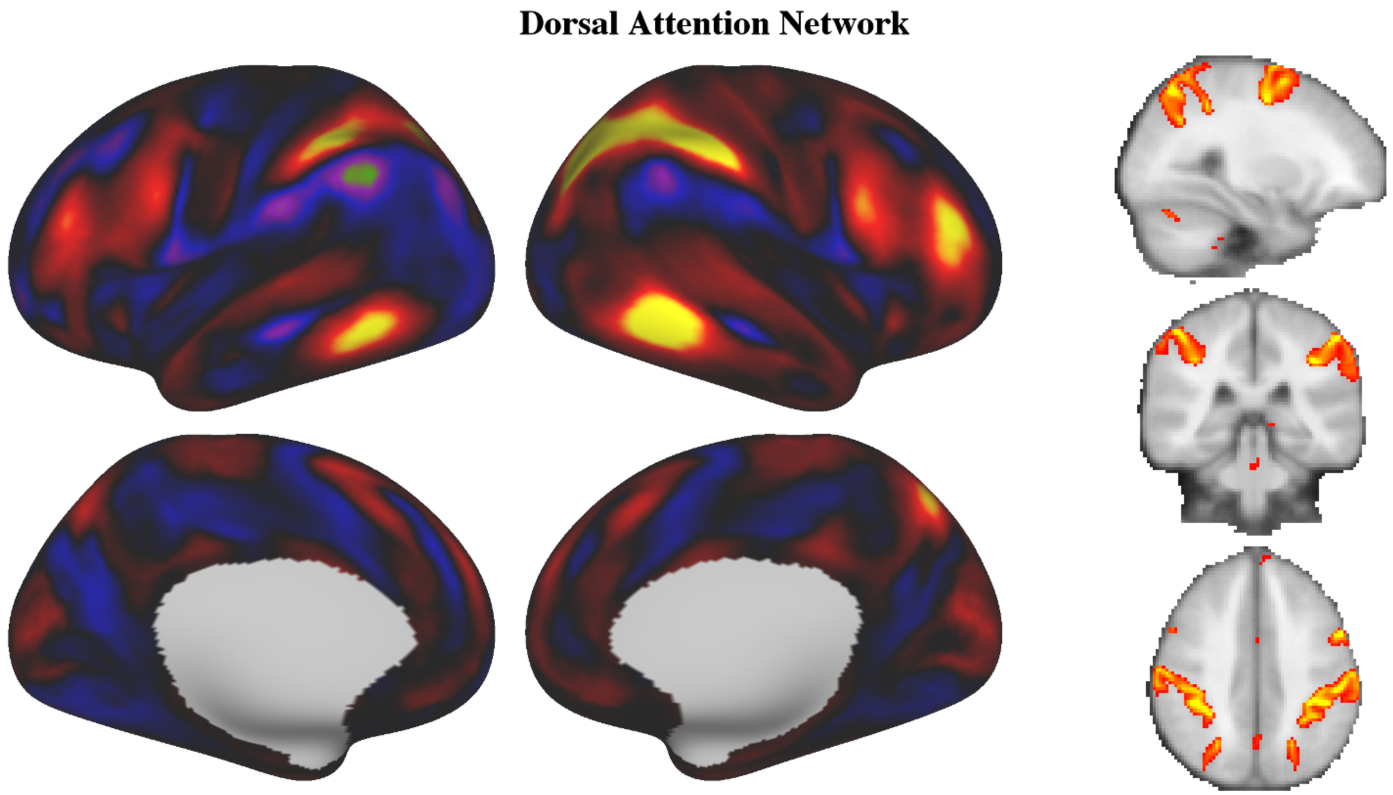


Resting state multi-subject ICA

- Why not just run ICA on each subject separately?
  - Correspondence problem (eg RSNs across subjects)
  - Different splittings sometimes caused by small changes in the data (naughty ICA!)
- Instead - start with a “group-average” ICA
  - But then need to relate group maps back to the individual subjects

Resting state networks

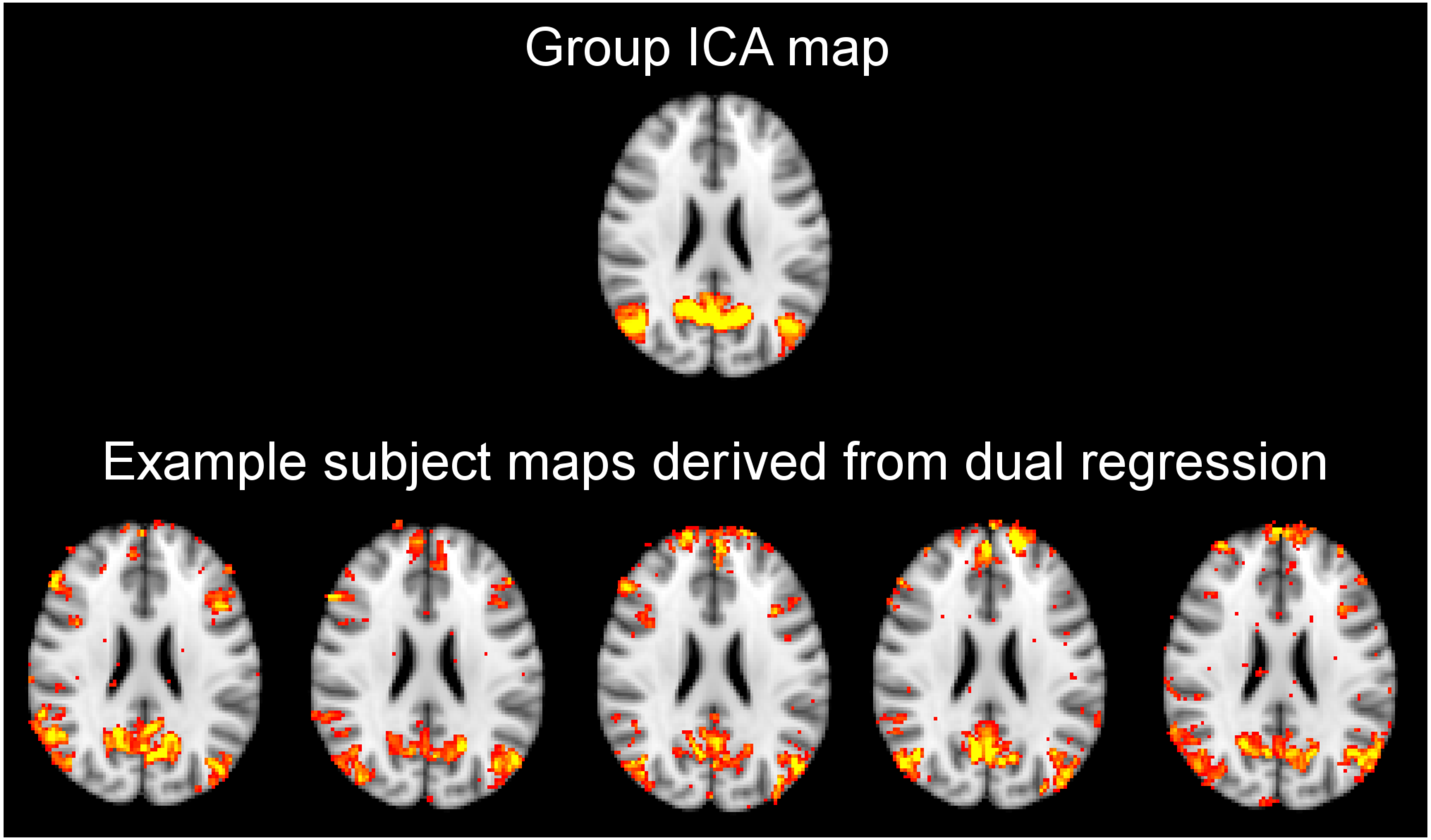




Resting state fMRI and ICA

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

Resting state multi-subject ICA

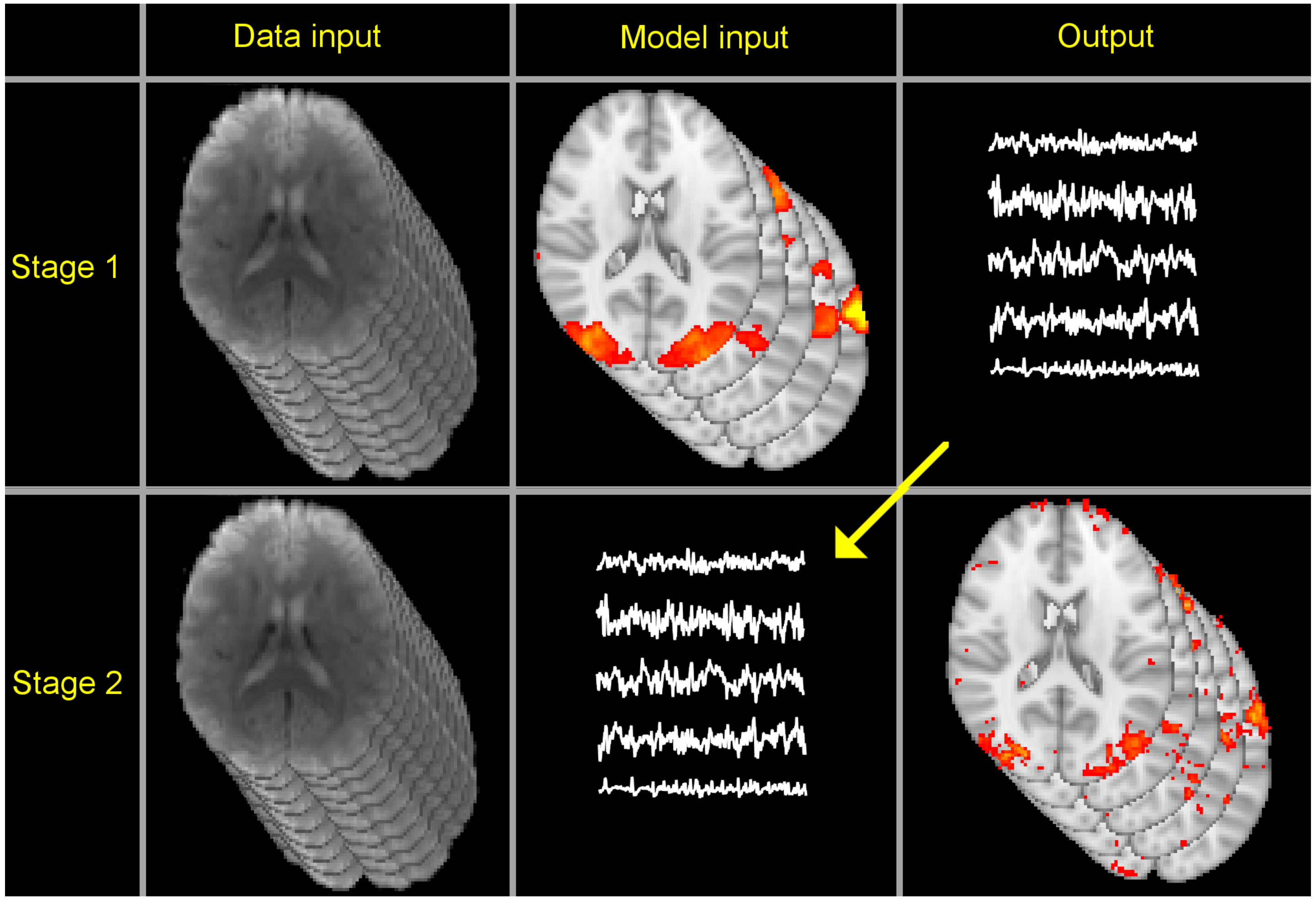


Dual Regression

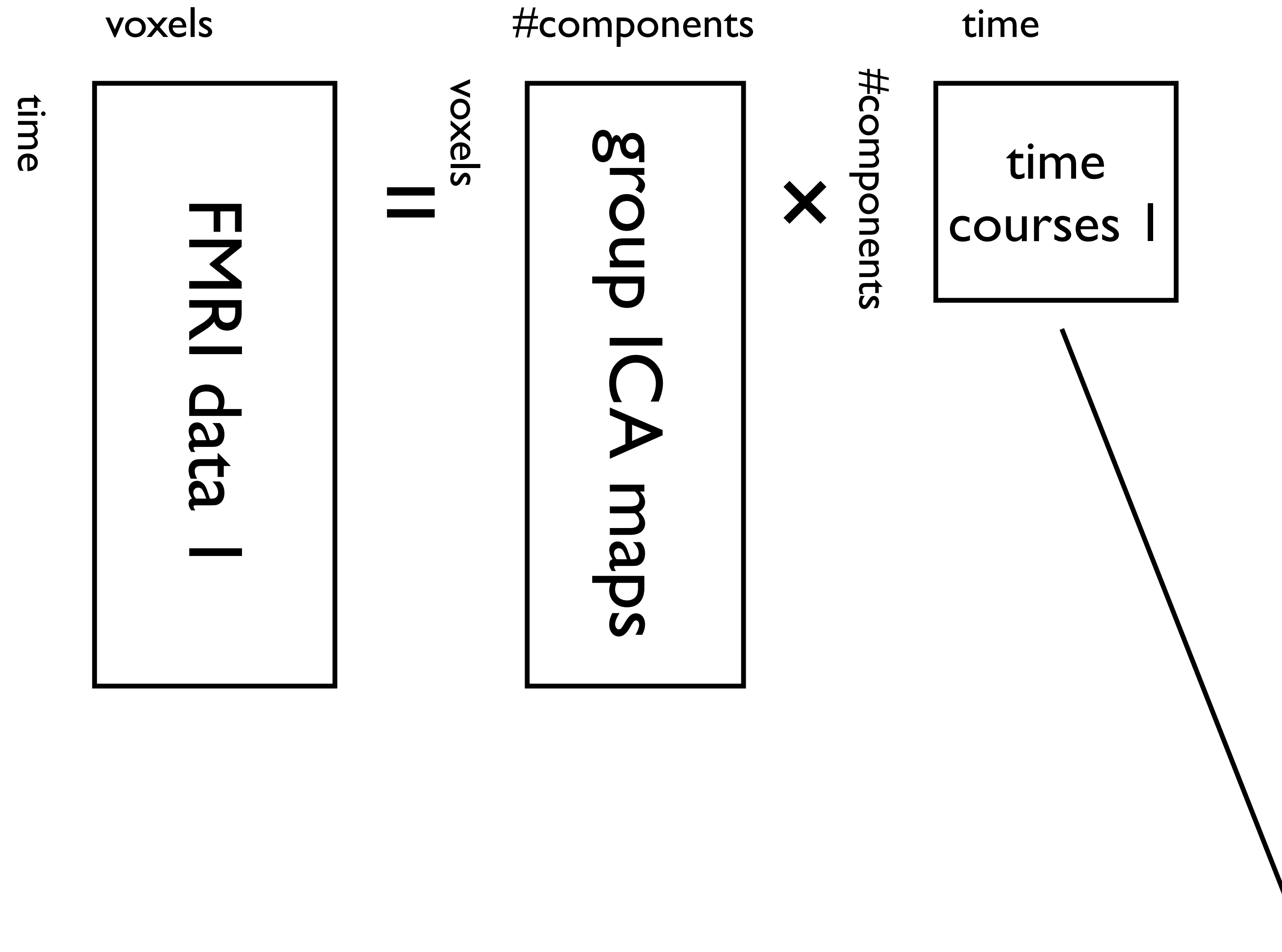
Two steps that both involve multiple regression:

1. Extract subject timeseries

2. Extract subject maps

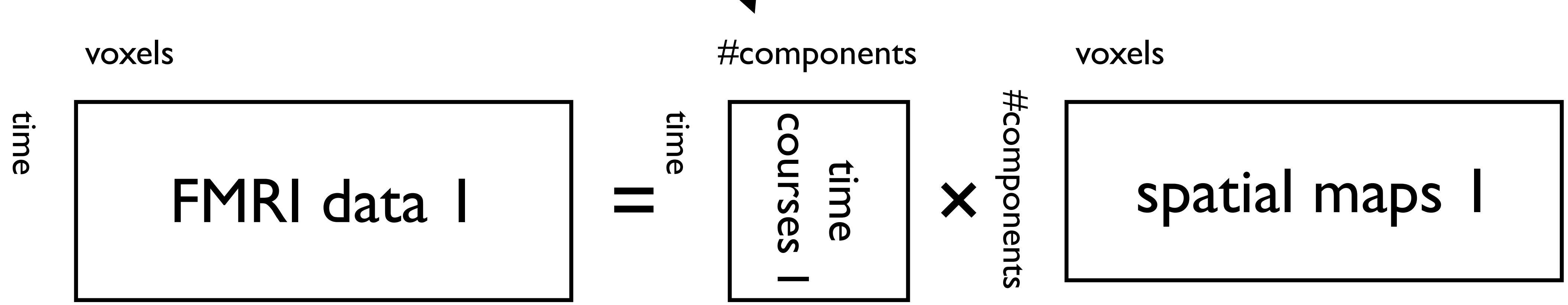


Dual Regression

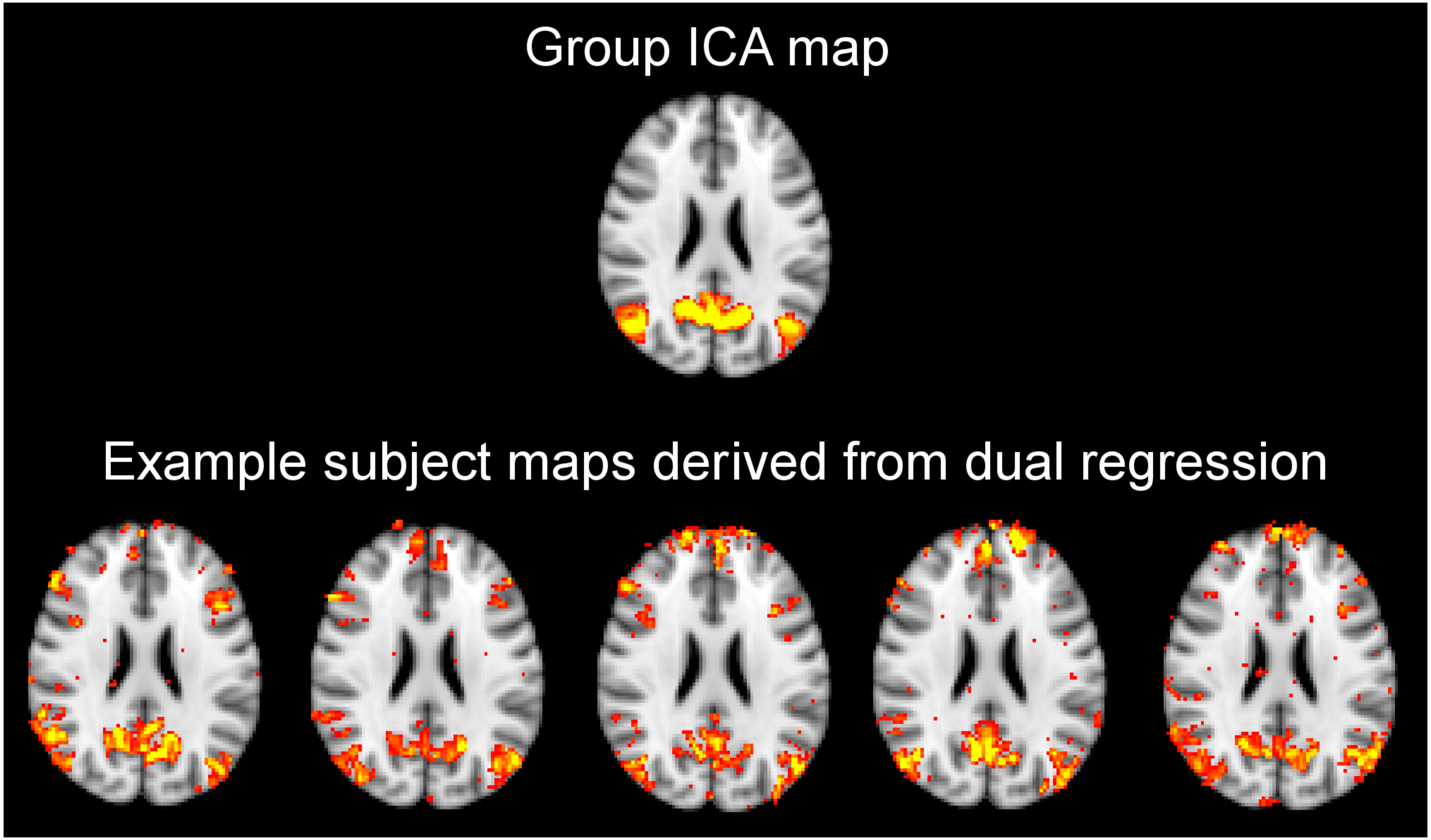


1. Regress group maps into each subject's 4D data to find subject-specific timecourses

2. Regress these timecourses back into the 4D data to find subject-specific spatial maps



Dual Regression



Running dual_regression

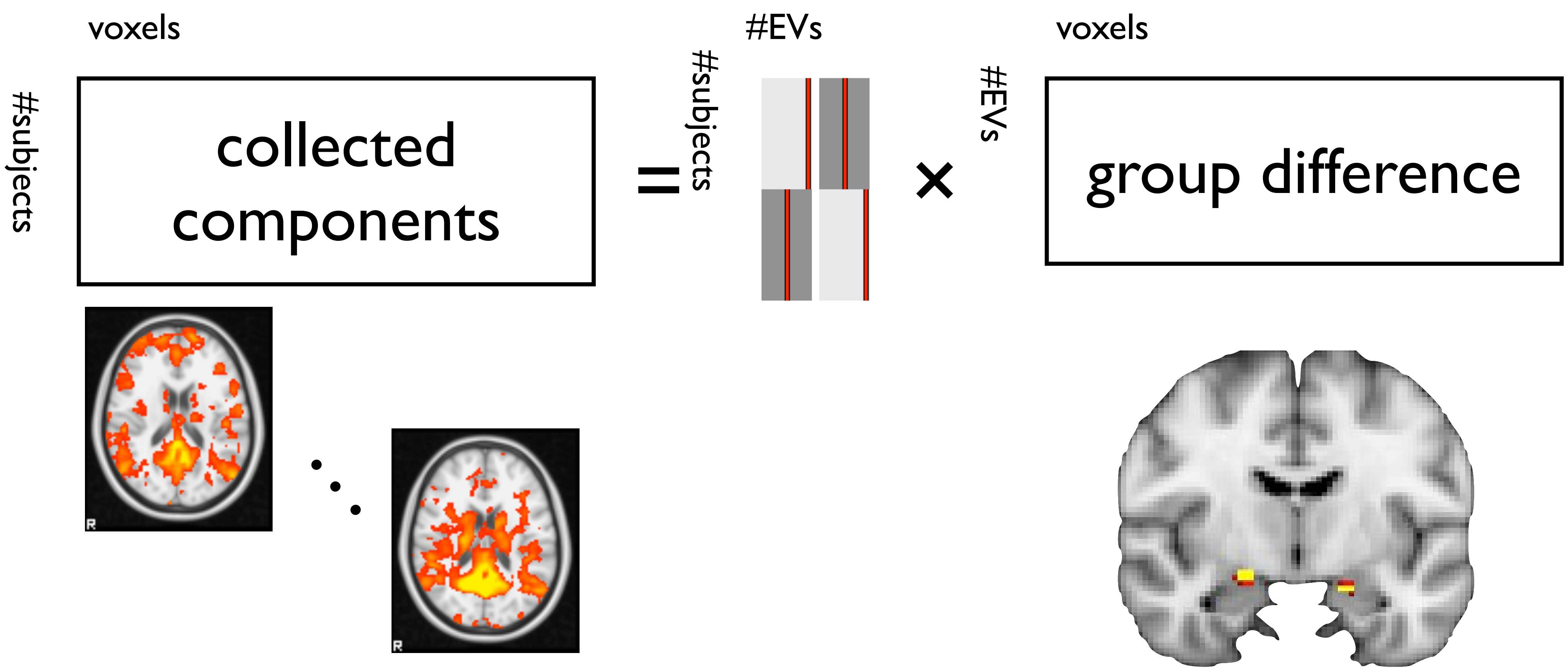
beckmann — bash — bash — 142×23

```
[islay:~] dual_regression.sh

dual_regression v0.5 (beta)

***NOTE*** ORDER OF COMMAND-LINE ARGUMENTS IS DIFFERENT FROM PREVIOUS VERSION

Usage: dual_regression <group_IC_maps> <des_norm> <design.mat> <design.con> <n_perm> <output_directory> <input1> <input2> <input3> .........
e.g.   dual_regression groupICA.gica/groupmelodic.ica/melodic_IC 1 design.mat design.con 500 grot `cat groupICA.gica/.filelist`

<group_IC_maps_4D>            4D image containing spatial IC maps (melodic_IC) from the whole-group ICA analysis
<des_norm>                    0 or 1 (1 is recommended). Whether to variance-normalise the timecourses used as the stage-2 regressors
<design.mat>                  Design matrix for final cross-subject modelling with randomise
<design.con>                  Design contrasts for final cross-subject modelling with randomise
<n_perm>                      Number of permutations for randomise; set to 1 for just raw tstat output, set to 0 to not run randomise at all.
<output_directory>            This directory will be created to hold all output and logfiles
<input1> <input2> ...         List all subjects' preprocessed, standard-space 4D datasets

<design.mat> <design.con>     can be replaced with just
-1                            for group-mean (one-group t-test) modelling.
If you need to add other randomise option then just edit the line after "EDIT HERE" below

[islay:~]
```

- FSL command line tool, combining:
  - DR to create subject-wise estimates (stage 1 + stage 2)
  - Group comparison using randomise (stage 3)

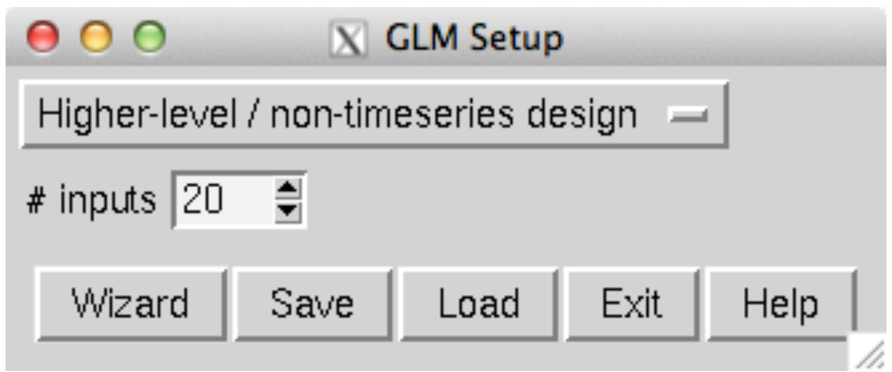
Group comparison

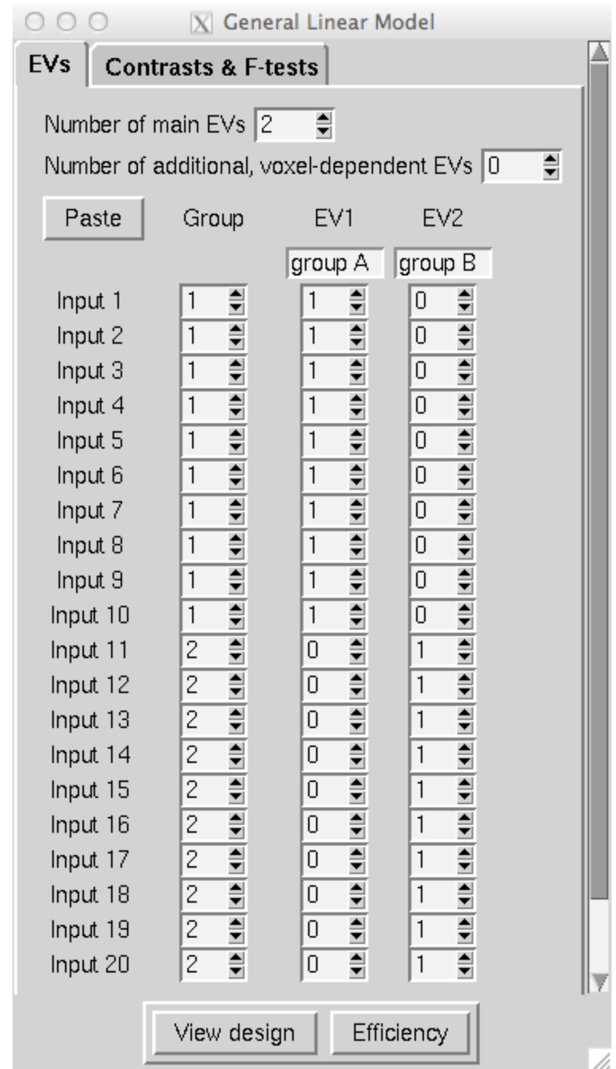
- Collect maps and perform voxel-wise test (e.g. randomisation test on GLM)

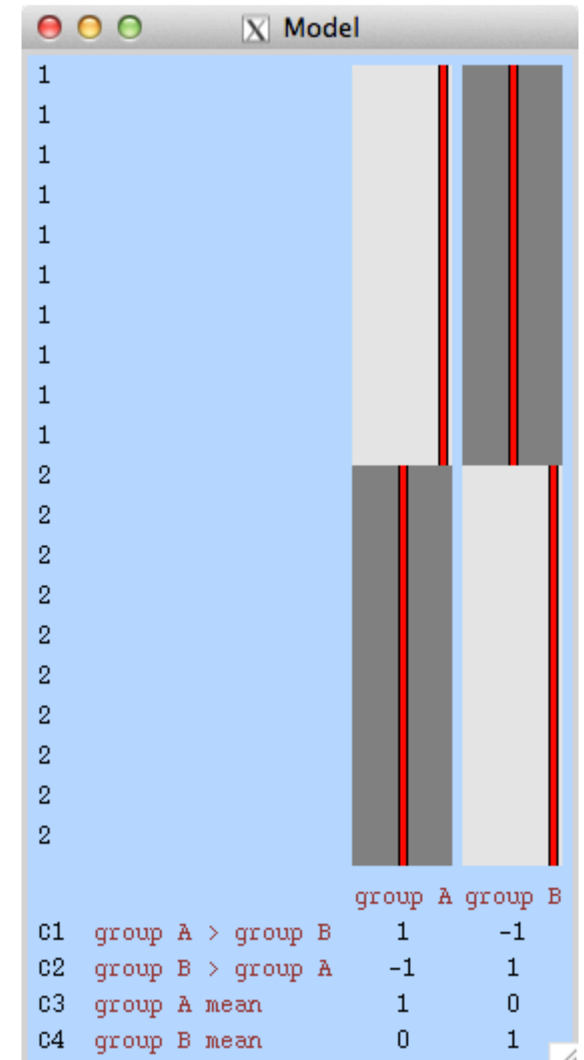


- Can now do voxelwise testing across subjects, separately for each original group ICA map
- Can choose to look at strength-and-shape differences

Group analysis on maps



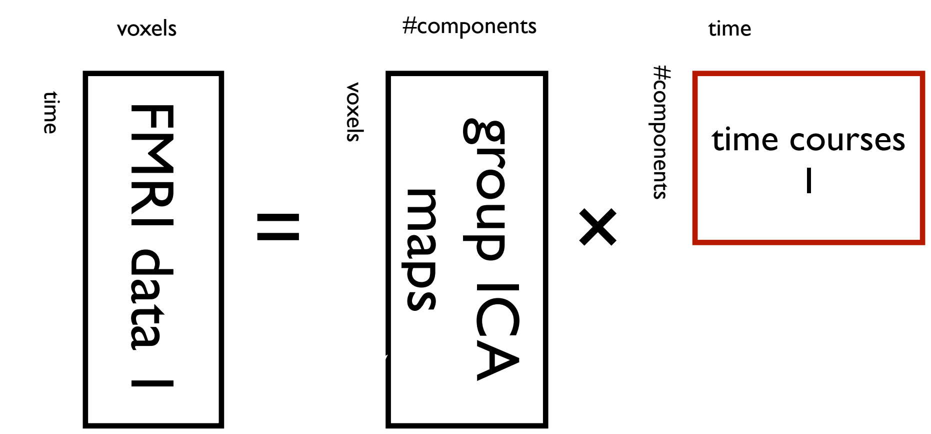




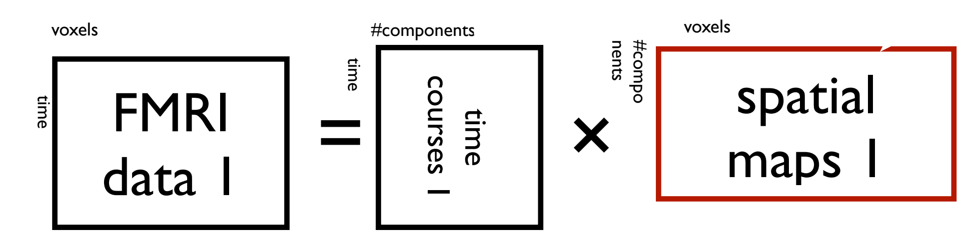
- can use the Glm tool (Glm_gui on mac) to create GLM design and contrast matrices

Dual regression outputs

- dr_stage1_subject[#SUB].txt - the timeseries outputs of stage 1 of the dual-regression.



- dr_stage2_subject[#SUB].nii.gz - the spatial maps outputs of stage 2 of the dual-regression.



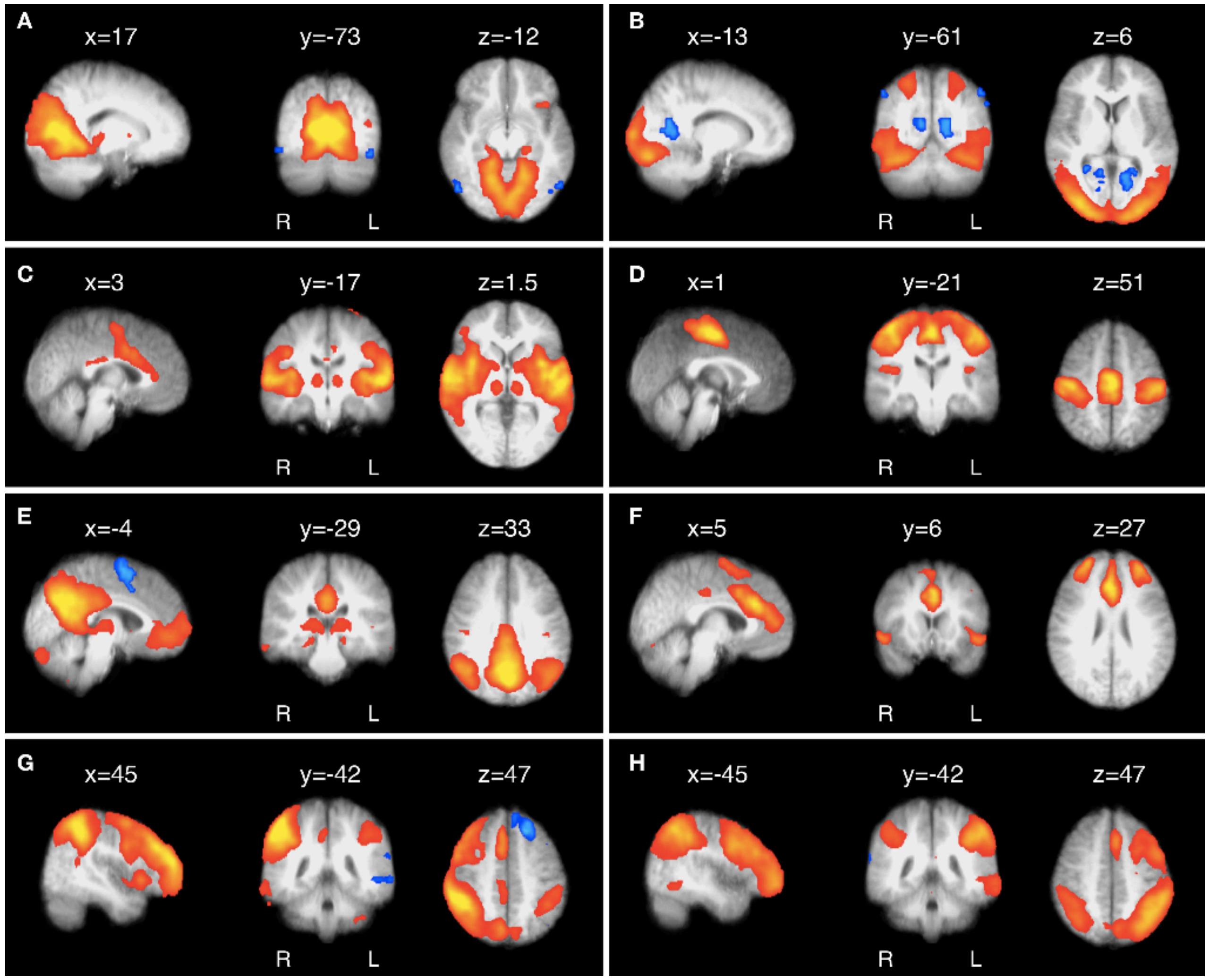
- dr_stage2_ic[#ICA].nii.gz - the re-organised parameter estimate images

- dr_stage3_ic[#ICA]_tstat[#CON].nii.gz - the output from randomise

(corrected for multiple comparisons across voxels but not across #components!!)

Group template maps

- Generate from the data using ICA
  - use all data to get unbiased templates
  - use independent control group
    - will not model signals and artefacts
- use existing template



template RSNs

https://www.fmrib.ox.ac.uk/datasets/royalsoc8/

Summary of part 2

- MELODIC ICA is applied to resting state fMRI for two main purposes: artefact detection, resting state network characterisation
- Artefact detection is done using single subject ICA
  - Artefact components can be classified manually or using FIX/AROMA
  - We then use GLM to regress out the artefacts and clean up the fMRI
- RSN estimation is done using group-level ICA
  - We use concatenation technique
  - This will give consensus RSNs which we then project onto subject fMRI using dual regression
  - Subject specific RSNs obtained from dual regression can be used for statistical comparisons (and input to FSLNets -> we will learn about this tomorrow)

Resting state fMRI and ICA

Available from:

- Oxford University Press
- Amazon