

Robust methodology for multi-subject RSN analysis – theory and practice...

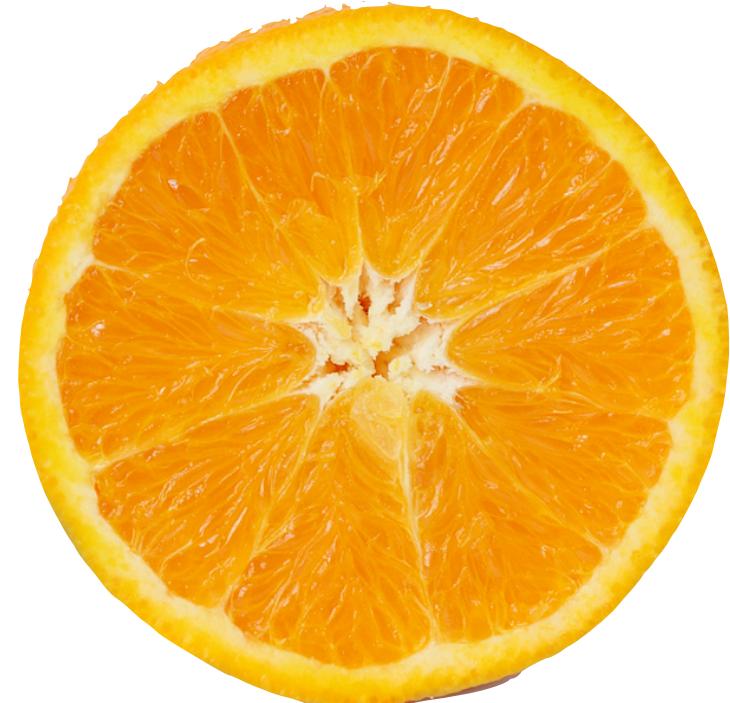
Christian F. Beckmann
beckmann@fmrib.ox.ac.uk

Stephen M. Smith
steve@fmrib.ox.ac.uk

*...or how to compare
RSNs at the population
level*



?
=

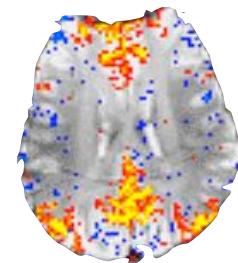
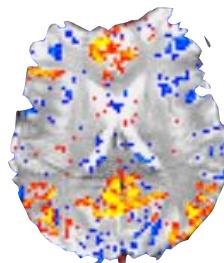


Functional Connectivity & ‘Resting-State Networks’

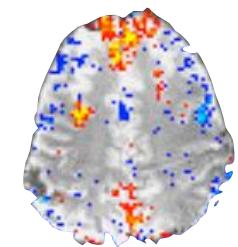
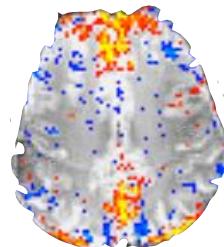
- *temporal correlations between spatially remote neurophysiological events*



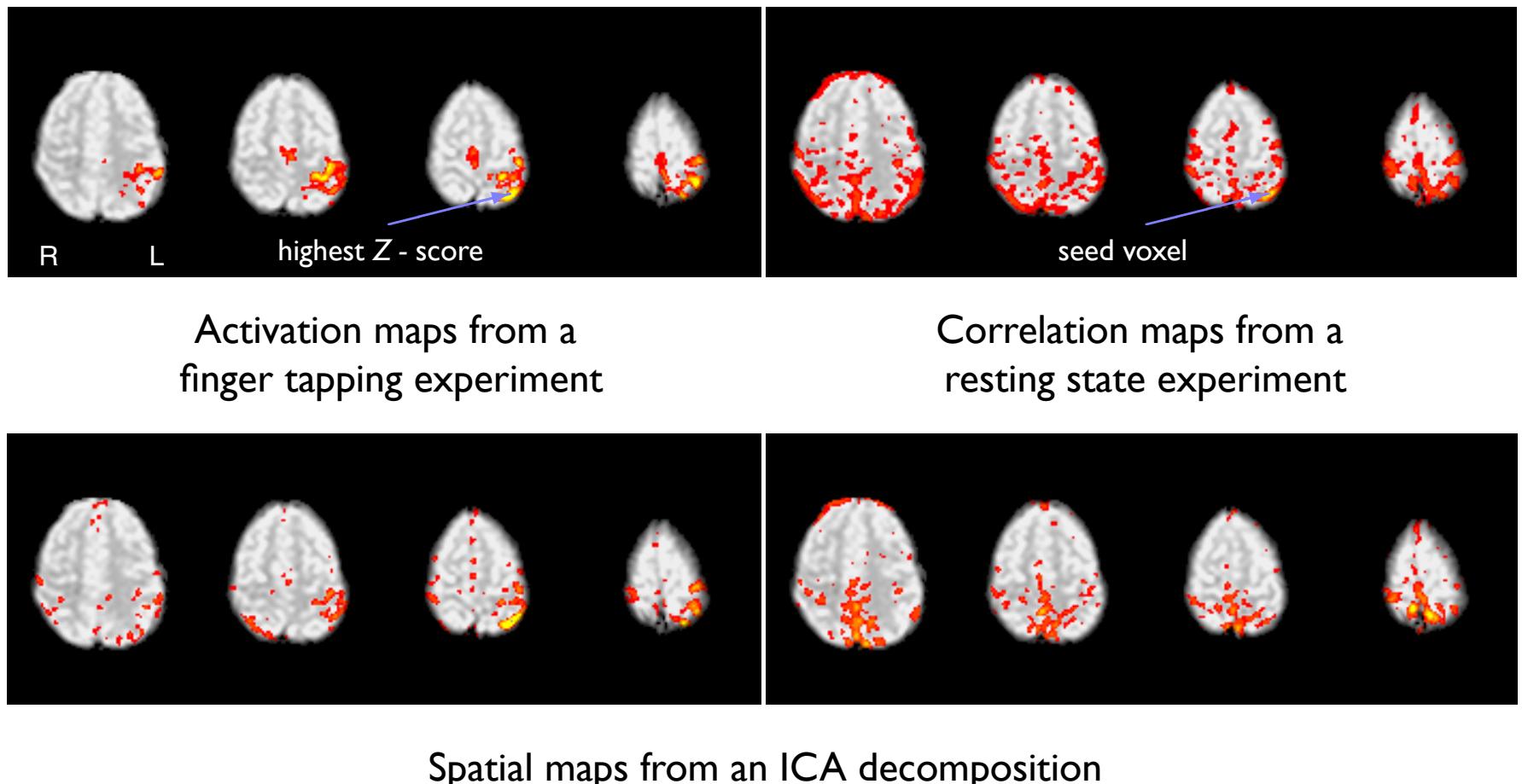
Friston et al. (1993)
J Cereb Blood Flow Metab



- Spatial patterns resembling activation maps
- found in fMRI data (BOLD & ASL) obtained under stimulation and in rest

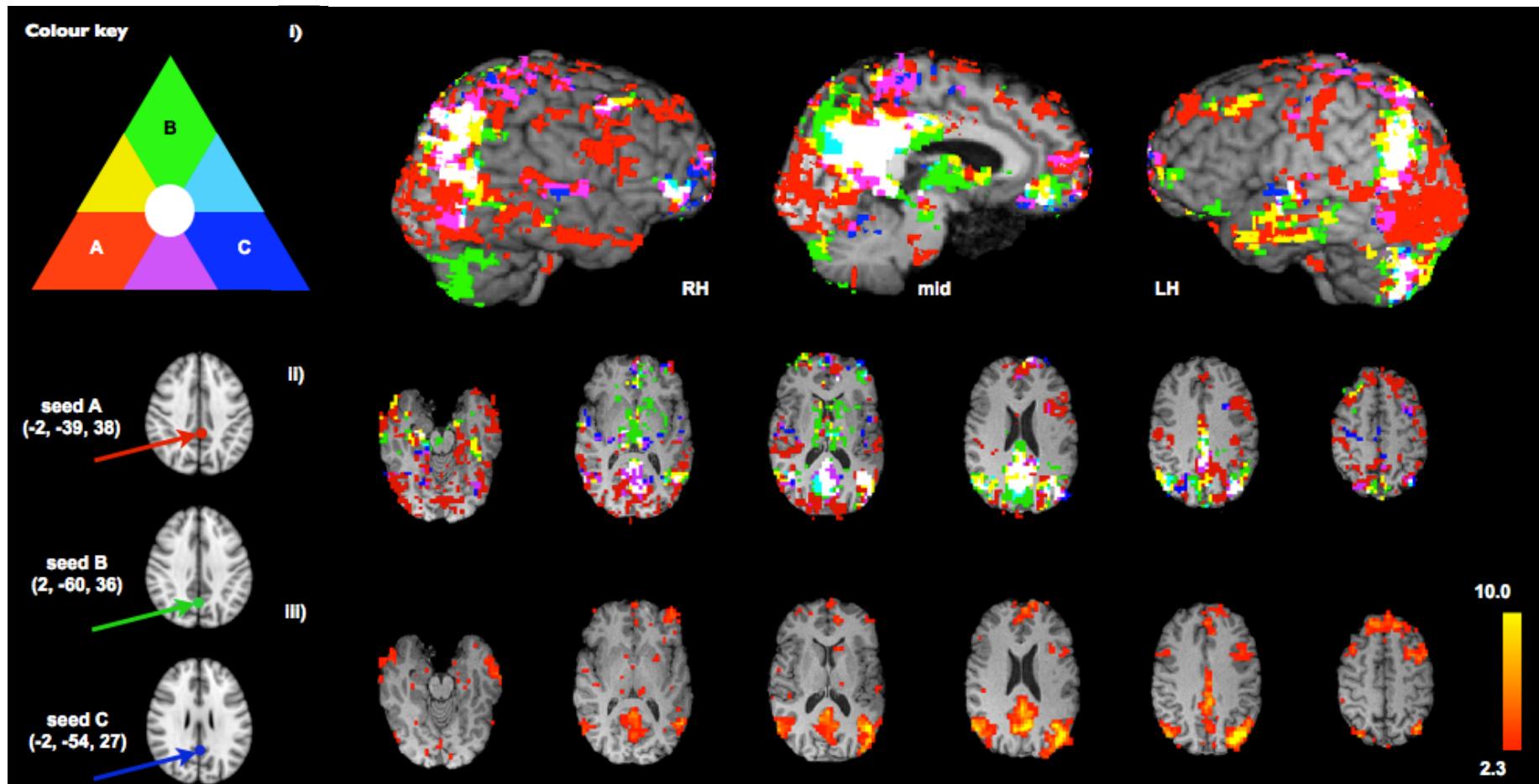


Seed-selection



Beckmann et al (2005)
Philos Trans R Soc Lond, B, Biol Sci

Seed-selection bias



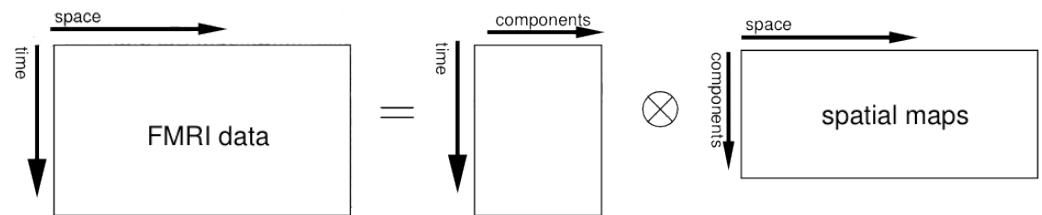
Different ICA models

Single-Session ICA

each ICA component comprises:



spatial map & timecourse



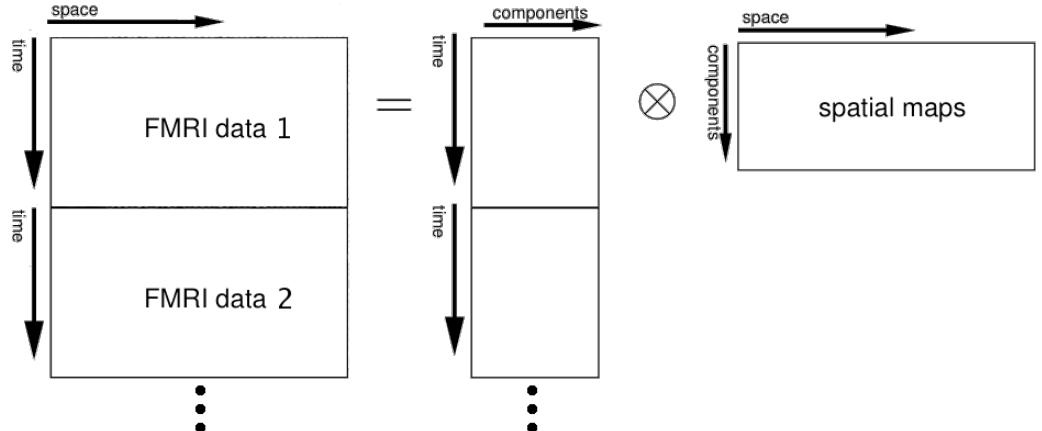
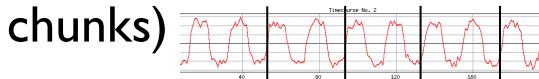
Multi-Session or Multi-Subject ICA: Concatenation approach

each ICA component comprises:



spatial map & timecourse

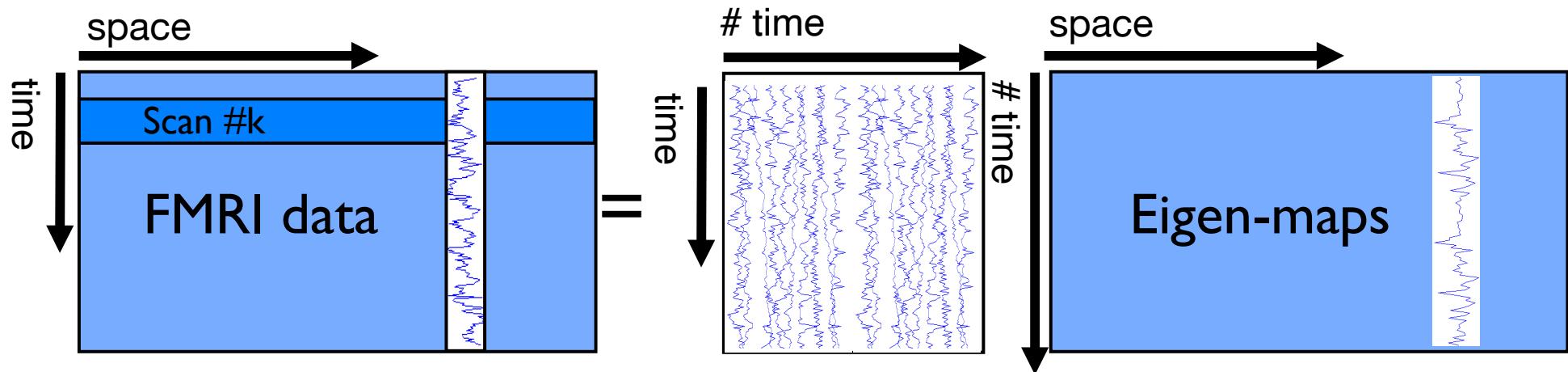
(that can be split up into subject-specific
chunks)



ICA & dimensionality reduction

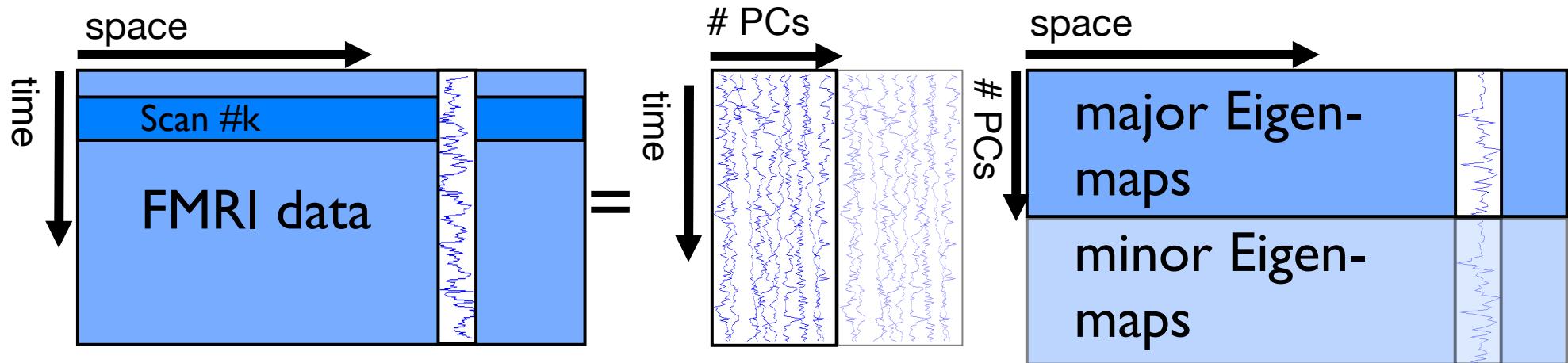
- FMRI data are large...
- ...particularly at the group level
- common to reduce dimensionality by means of a *Principal Component Analysis*

Principal Component Analysis



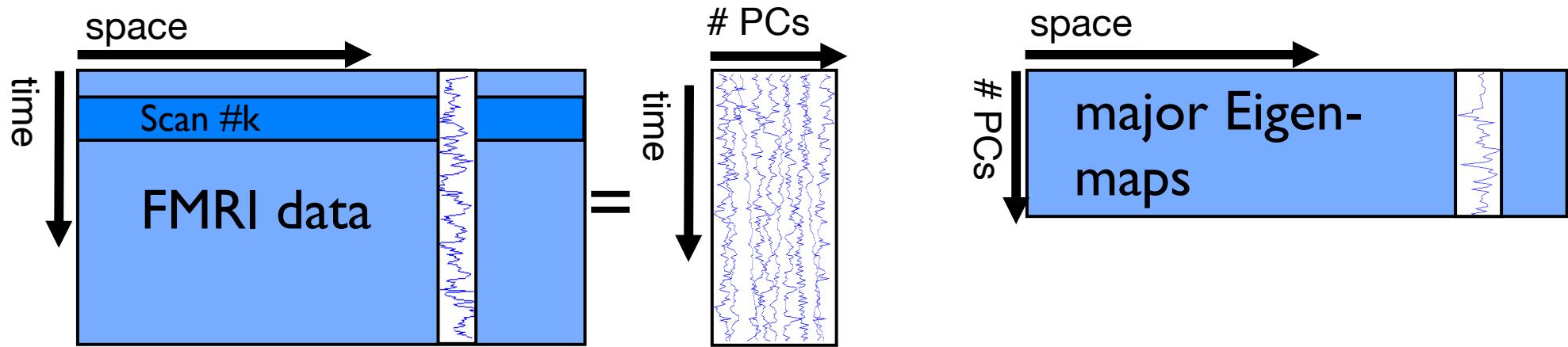
- PCA: data is decomposed into a set of uncorrelated spatial maps and uncorrelated time courses

Principal Component Analysis



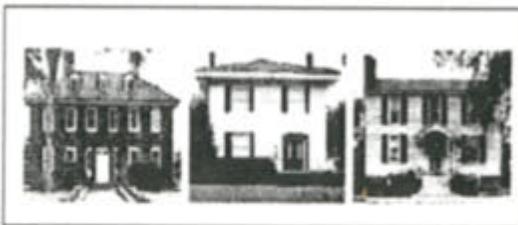
- PCA: data is decomposed into a set of uncorrelated spatial maps and uncorrelated time courses such that a *maximal amount of variance* is retained

Principal Component Analysis

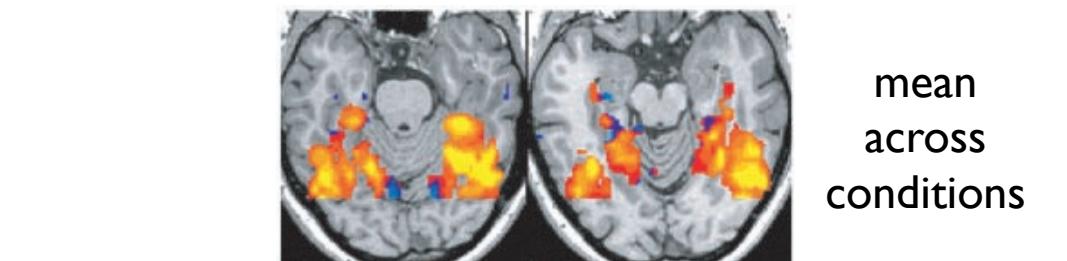
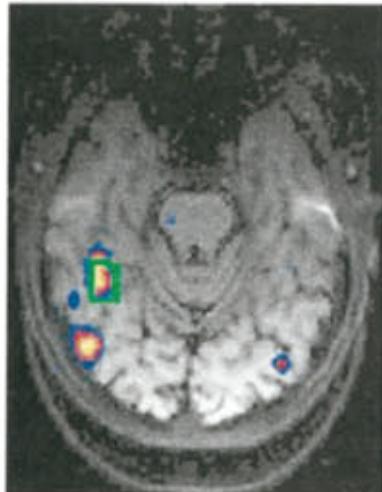


- optimal wrt sum-of-squares error
- minor PCA subspace get's ignored!
- lossy process

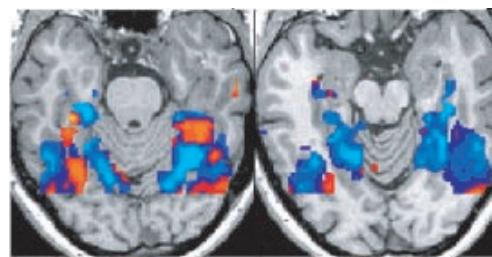
Detour: local vs global brain function



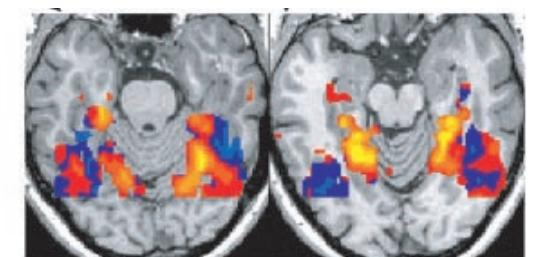
Faces > Houses



mean
across
conditions



Faces



Houses



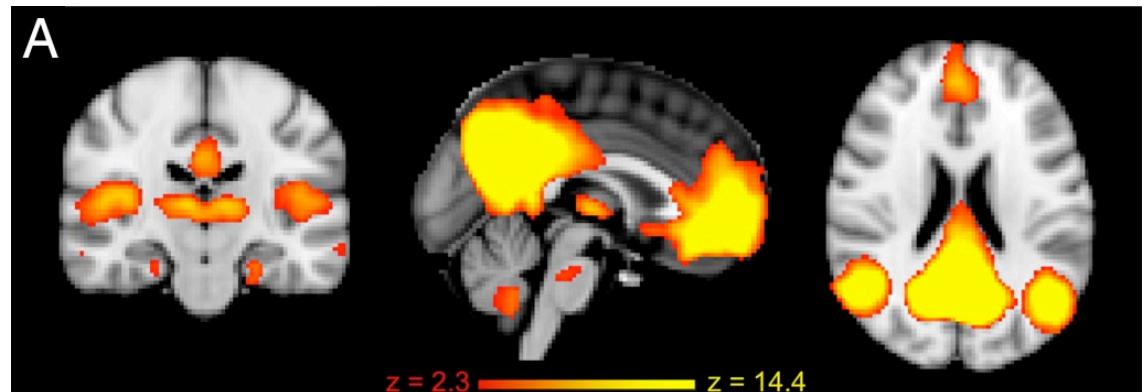
Kanwisher et al. (1997)
J NeuroSci



Haxby et al. (2001)
Science

Detour: primary vs. differential contrasts

- Even when primary effects are large...
(express loads of variance)
- ... *differential* effects can be very small
(express little variance)
- PCA suitable for the former, less suitable for the latter!



PCA bias

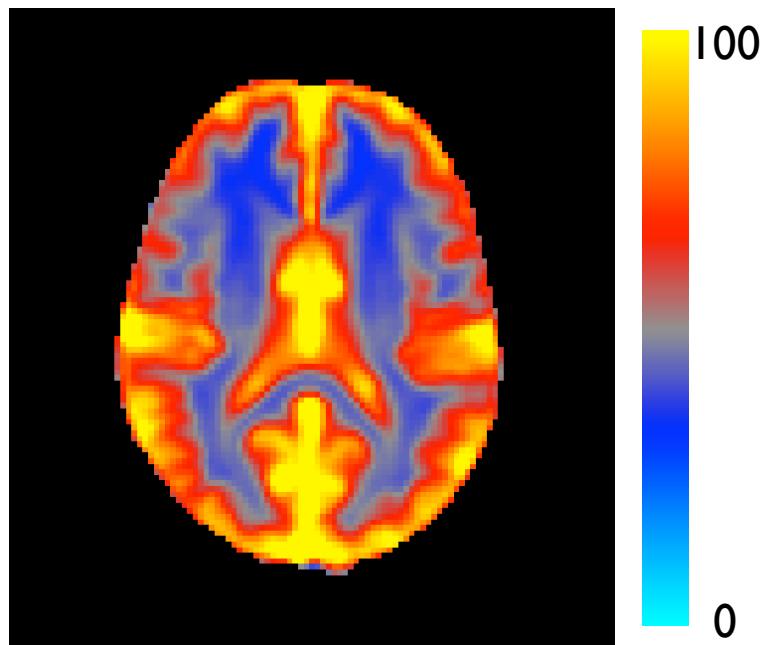
- within-subject bias:
 - *tissue-type bias & variance normalisation*
- between-subject bias:
 - *between major & minor subspace*
 - *within the major subspace*

PCA bias

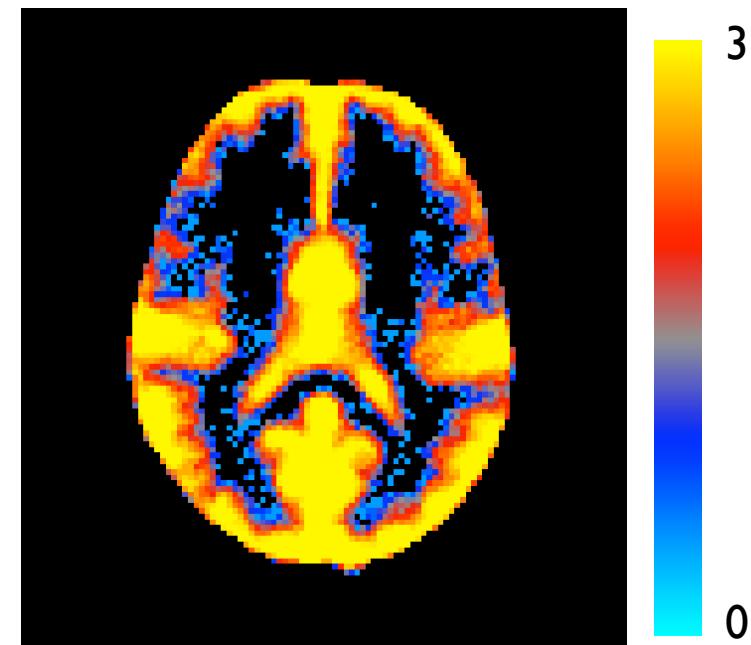
- within-subject bias:
 - *tissue-type bias & variance normalisation*
- between-subject bias:
 - *between major & minor subspace*
 - *within the major subspace*

PCA tissue-type bias

voxel-wise temporal
standard deviations



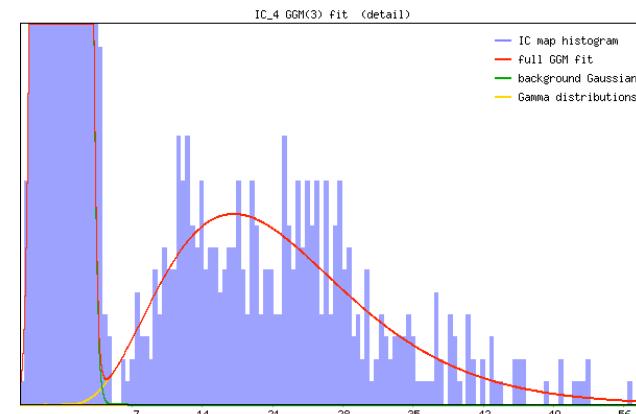
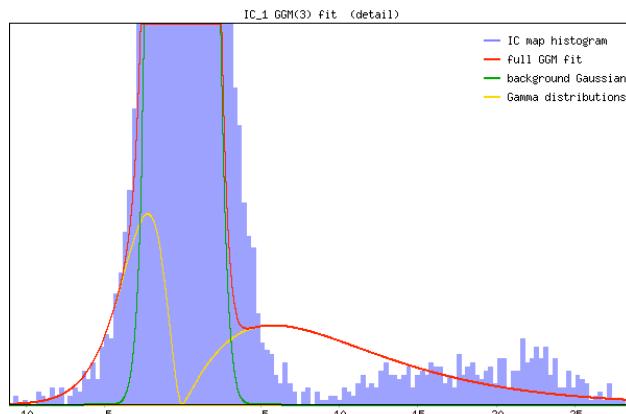
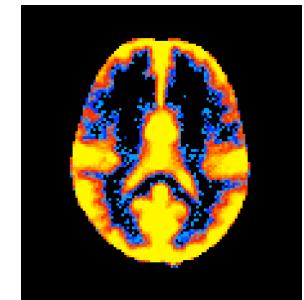
(log) - frequency of voxels
appearing in major PCA space



- Without normalisation PCA is **biased** towards tissue exhibiting strong temporal variation

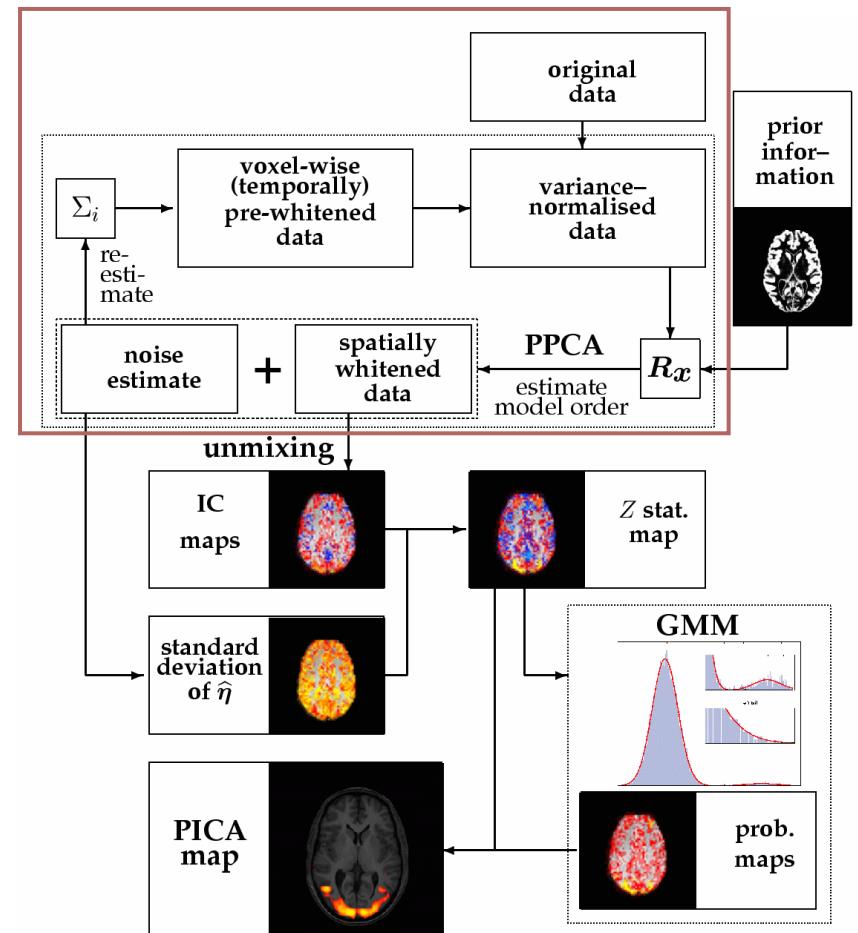
PCA tissue-type bias

- *Inhomogeneous sensitivity* to signal
(particularly in WM / GM relative to CSF)
- *Thresholding* more difficult (MOG vs Gaussian noise model)
- *Interaction* with most ICA cost function



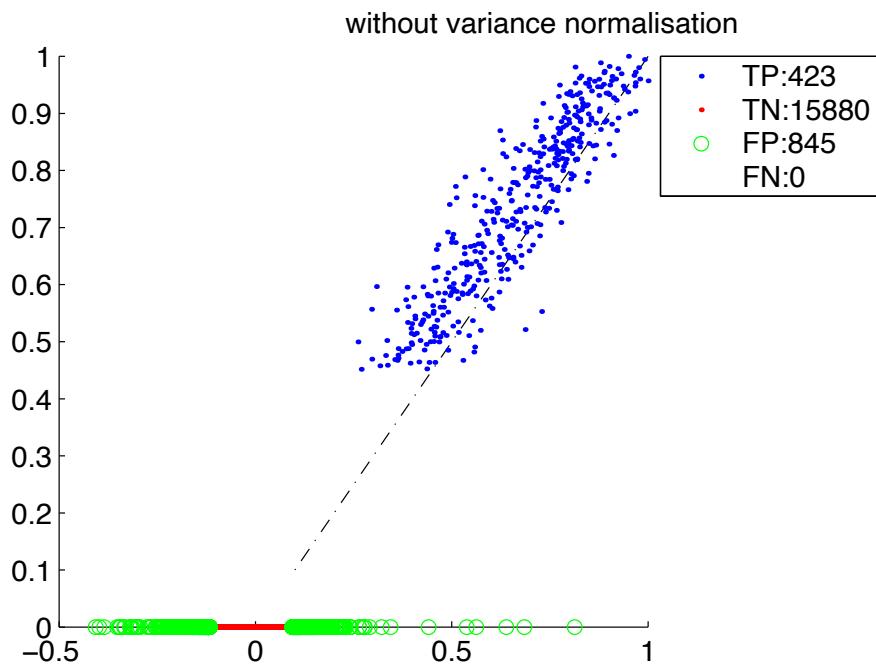
Variance-normalisation

- normalise every voxel's time series wrt to it's noise std-deviation (inputs become Z-stats)

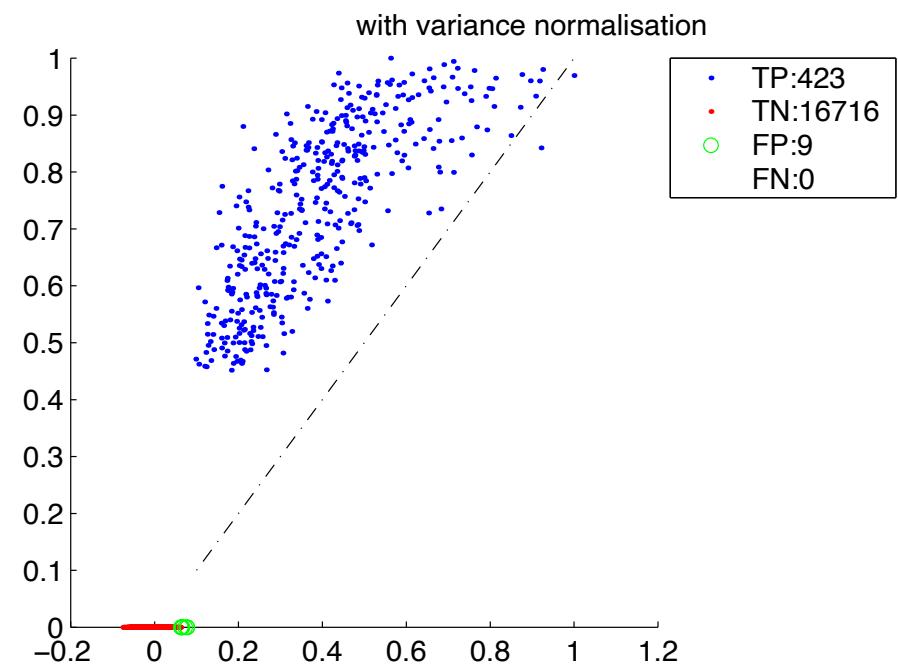


Example

- Simulated signal in noise



ICA without variance
normalisation



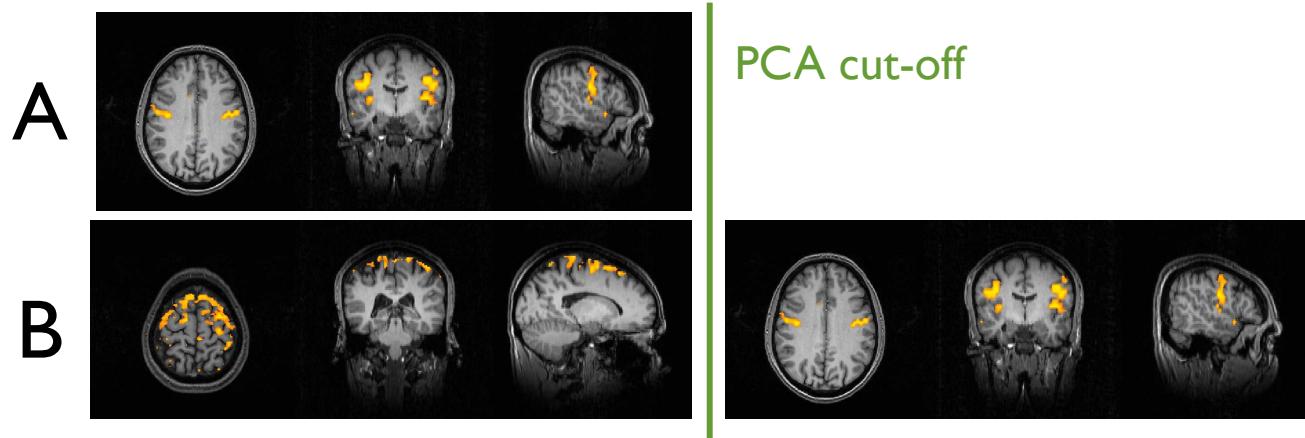
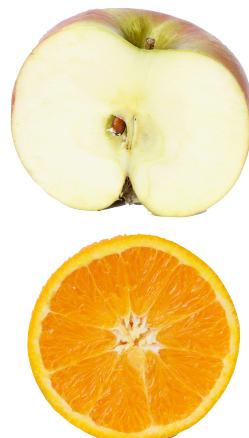
ICA with variance
normalisation

PCA bias

- within-subject bias:
 - *tissue-type bias & variance normalisation*
- between-subject bias:
 - *between major & minor subspace*
 - *within the major subspace*

between major & minor subspace

- 2 subjects, only a single (identical) RSN
- one of the subjects has strong motion



- Inverting PCA will find significant difference wrt RSN

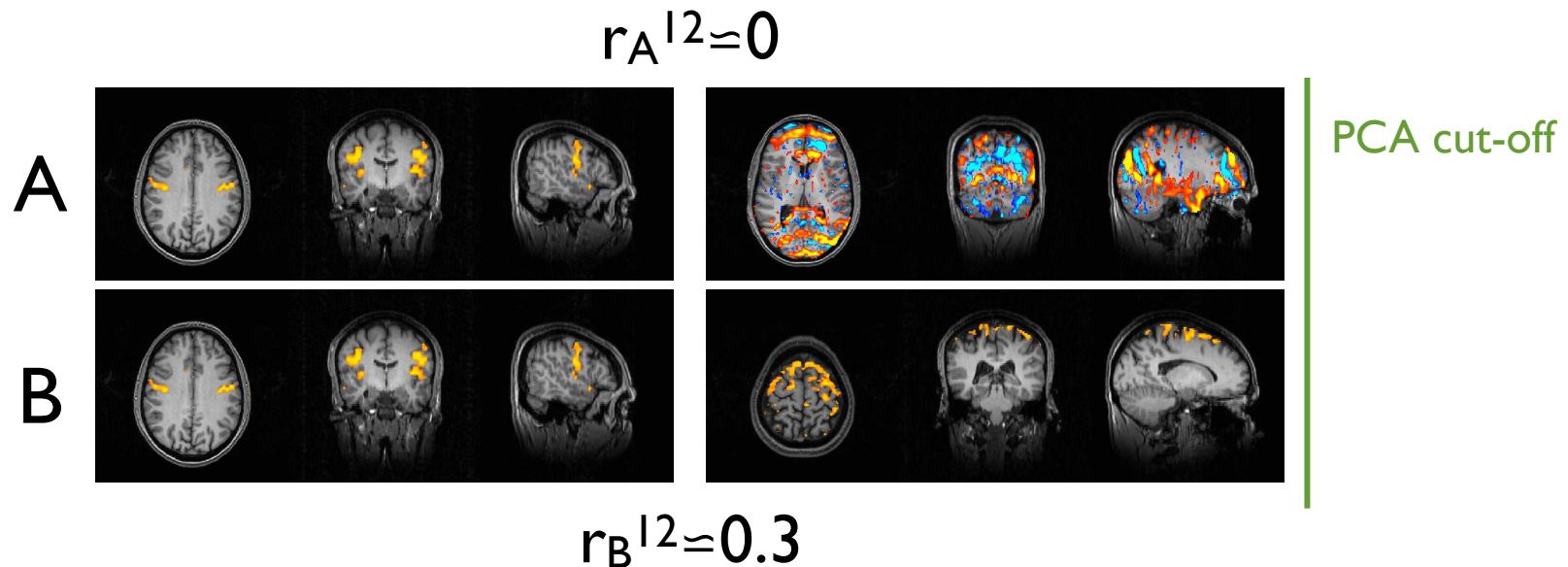
False-positive detection!

PCA bias

- within-subject bias:
 - *tissue-type bias & variance normalisation*
- between-subject bias:
 - *between major & minor subspace*
 - *within the major subspace*

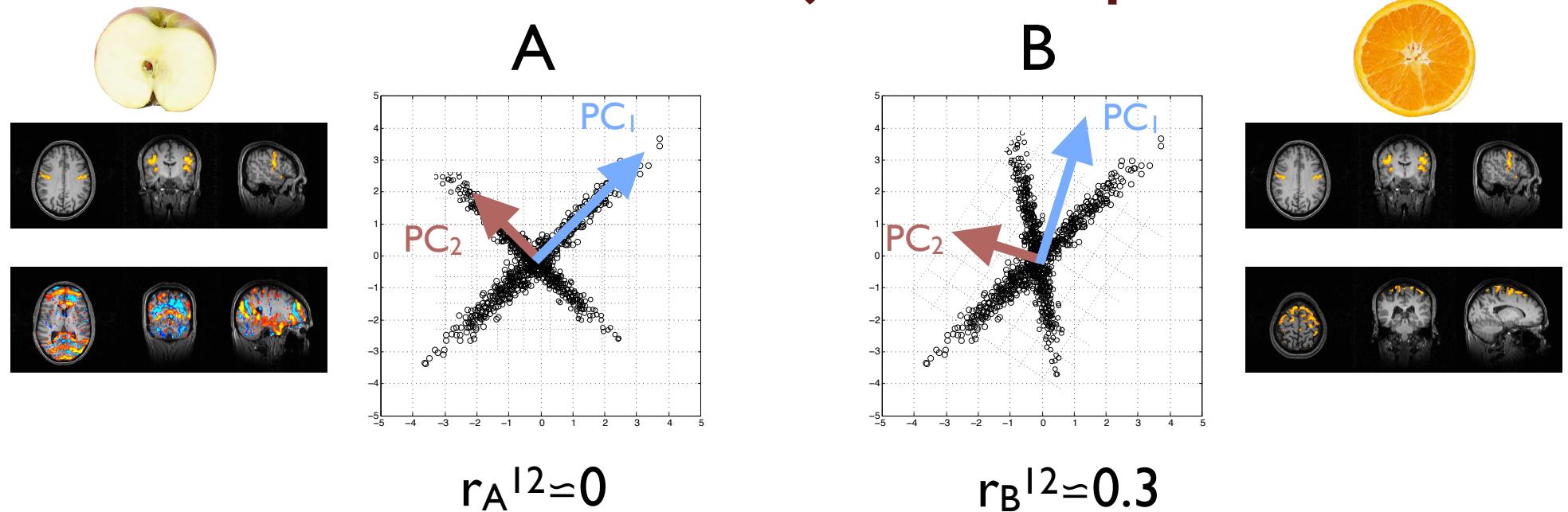
within the major subspace

- 2 subjects, 2 effects, different degree of temporal correlation



- PCA will represent effect I differently in the two subjects

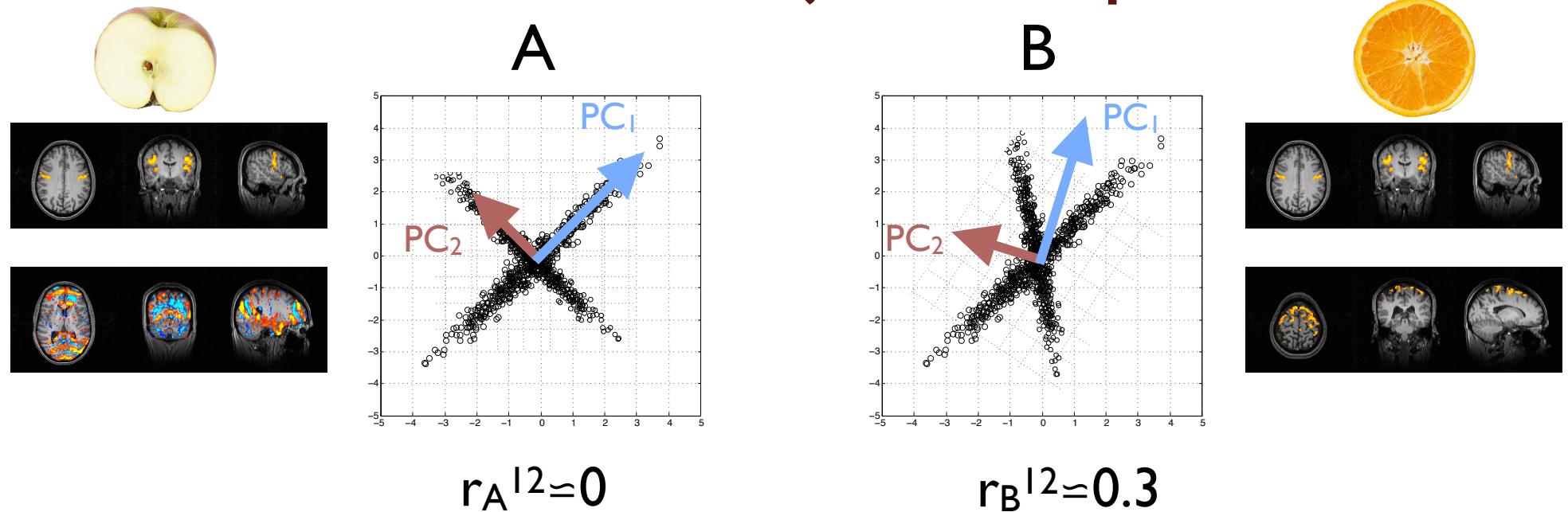
Difference in signal representation within the major subspace



- Inverting PCA will use deflected vectors and over- or under-estimate effects in B

False-positive or false-negative detection!

Difference in signal representation within the major subspace



- Note: in higher-dimensional cases the subspaces for A and B will in general not be the same (don't just differ by rotation)!
- Can only talk about the joint (intersection) space in confidence

How to do multi-subject PCA?

- Ideally want to use the *full* $NT \times NT$ covariance matrix

| | | NxT | |
|--|--|----------|----------|
| | | C_{11} | C_{12} |
| | | C_{21} | C_{22} |
| | | \vdots | \ddots |
| | | C_{n1} | C_{1n} |

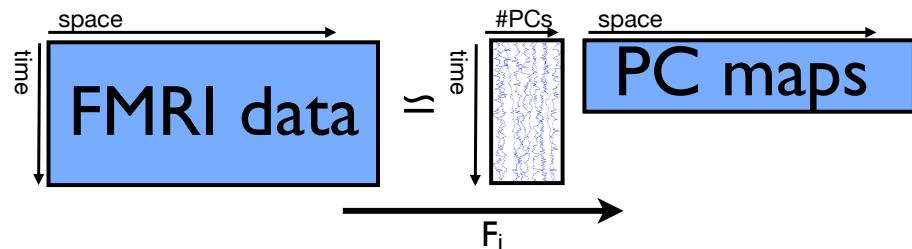
How to do multi-subject PCA?

- Ideally want to use the *full* $NT \times NT$ covariance matrix
- *subject-specific* PCA uses C_{ii} only

| | | NxT | | | |
|-----|----------|----------|----------|-----|----------|
| | | C_{11} | C_{12} | ... | C_{In} |
| NxT | C_{21} | C_{22} | | | |
| | : | | | | |
| | C_{n1} | | | ⋮ | C_{In} |
| | | | | ⋮ | |

Subject-wise PCA

- For every subject this defines a mapping F_i



- F_i is lossy and ignores minor subspace
- worse for multiple subjects,
(remember: we can only infer
on the joint projection space
with confidence!)

| NxT | | | |
|----------|----------|-----|----------|
| C_{11} | C_{12} | ... | C_{1n} |
| C_{21} | C_{22} | | |
| : | | ⋮ | |
| C_{n1} | | | C_{1n} |

How to do multi-subject PCA?

- Ideally want to use the *full* $NT \times NT$ covariance matrix
- *subject-specific* PCA uses C_{ii} only
- $C_{ij} (i \neq j)$ contains the relevant cross-subject information

| | | NxT | | | |
|-----|----------|----------|----------|-----|----------|
| | | C_{11} | C_{12} | ... | C_{In} |
| NxT | C_{21} | C_{22} | | | |
| | : | | | | |
| | C_{n1} | | | | C_{In} |
| | | | | | |

How to do multi-subject PCA?

- Ideally want to use the *full* $NT \times NT$ covariance matrix
- *subject-specific* PCA uses C_{ii} only
- $C_{ij} (i \neq j)$ contains the relevant cross-subject information
- *subject-specific* PCA ignores these terms!

| | | NxT | | | |
|-----|----------|----------|----------|-----|----------|
| | | C_{11} | C_{12} | ... | C_{In} |
| NxT | C_{21} | C_{22} | | | |
| | : | | | | |
| | | | | | |
| | C_{n1} | | | | C_{In} |

group-PCA

- assume every subjects' SVD can be expressed in terms of common maps

$$Y^i = U^i D^i V^t$$

- The average data matrix then can be expressed with V as a set of right Eigen-maps,

$$\frac{1}{N} \sum_i Y^i = \widetilde{U} \widetilde{D} V^t$$

- ...provided:
 - $(U^k)^t U^j = 0$ for all k, j (i.e. temporal dynamics are uncorrelated)
 - $\sum_{i \in K} Y^i$ is of full rank for any subset K of $\{1 \dots N\}$

group-PCA

- based on the average data matrix find Eigen-maps

$$\frac{1}{N} \sum_i Y^i = \widetilde{U} \widetilde{D} V^t$$

- solve for $U^i D^i$ s.t.

$$Y^i = U^i D^i V^t$$

- reduce dim. per subject

| | | | |
|---|----------|----------|----------|
| | | | NxT |
| | C_{11} | C_{12} | ... |
| | C_{21} | C_{22} | |
| ⋮ | | | ⋮ |
| | C_{n1} | | C_{In} |

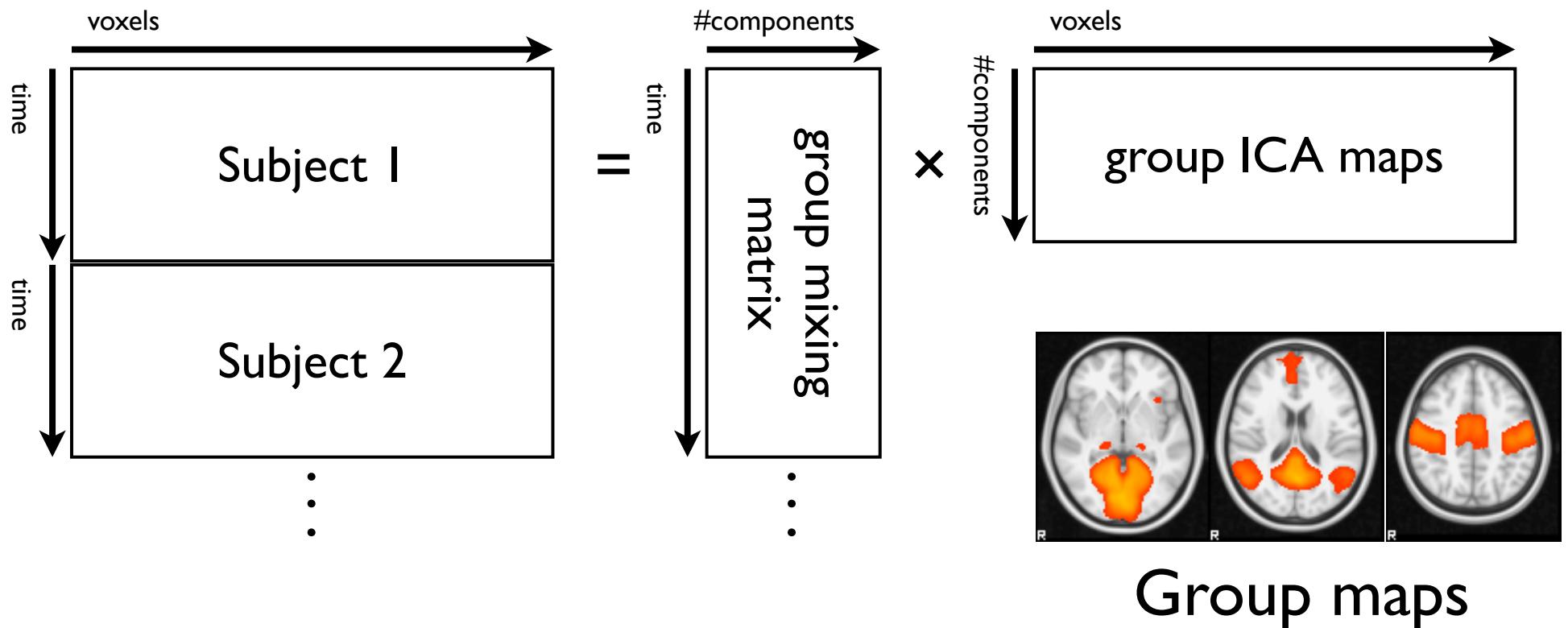


enforces that in *all* subjects the same effects
can be represented and compared



Concat-ICA

- Concatenate all subjects' data temporally
(Actually: group-based PCA reduction on each subject, then concat, then PCA reduction)
- then run ICA to obtain group maps



How to get subject-specific maps?

- could track all (temporal) transformation F_i used to find spatial group maps, then work backwards, e.g. ‘Back-Projection’ (Calhoun et al. 2001, HBM)
- if forward projections are lossy this can’t be recovered in the backward projection!
- instead, need to **avoid PCA bias**
 - ▶ remove all dependence on the initial PCA(s)

Bayesian model

- Model each subject's data as a linear combination of spatial maps and noise

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

- assume all spatial maps are sampled from underlying ‘average’ group maps

$$S^i = S_g + E_g \quad \text{where} \quad E_g \sim \mathcal{N}(0, \sigma_g^2 I)$$

‘Dual regression’ model

Hierarchical Normal-Normal model

$$Y^i \sim \mathcal{N}(S^i A^i, \sigma_Y^2 I)$$

$$S^i \sim \mathcal{N}(S_g, \sigma_g^2 I)$$

S_g known

- can use *Empirical Bayes* for estimation

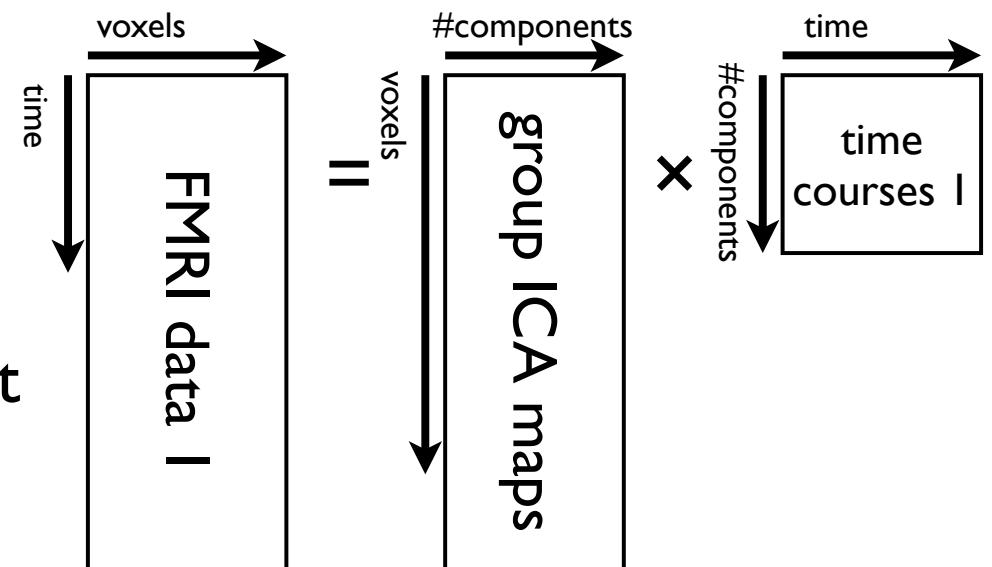
‘Dual regression’

- wrt temporal modes

$$Y^i = S_g A^i + \tilde{E} \quad \text{where} \quad \tilde{E} \sim \mathcal{N}(0, \sigma_g^2 I + \sigma_Y^2 I)$$

$$\widehat{A}^i = S_g^\dagger Y^i$$

- ...so estimates can be obtained using simple spatial regression of group-ICA maps against data



‘Dual regression’

- wrt subject-specific spatial maps

$$p(\mathbf{S}) \propto \frac{1}{\sigma_g^{2N}} \prod \exp\left(-\frac{1}{2\sigma_g^2} (S_g - S^i)^2\right) \\ \times \prod \exp\left(-\frac{1}{2\sigma_Y^2} (S^i A^i - Y^i)^2\right)$$

- combined likelihood function

‘Dual regression’

- wrt subject-specific spatial maps

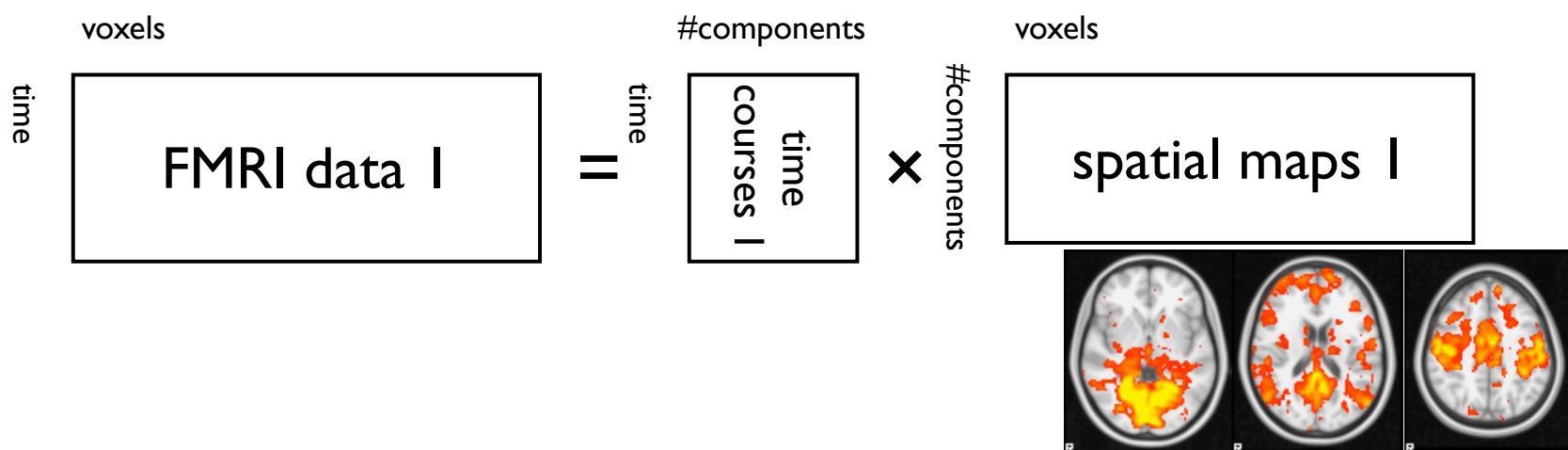
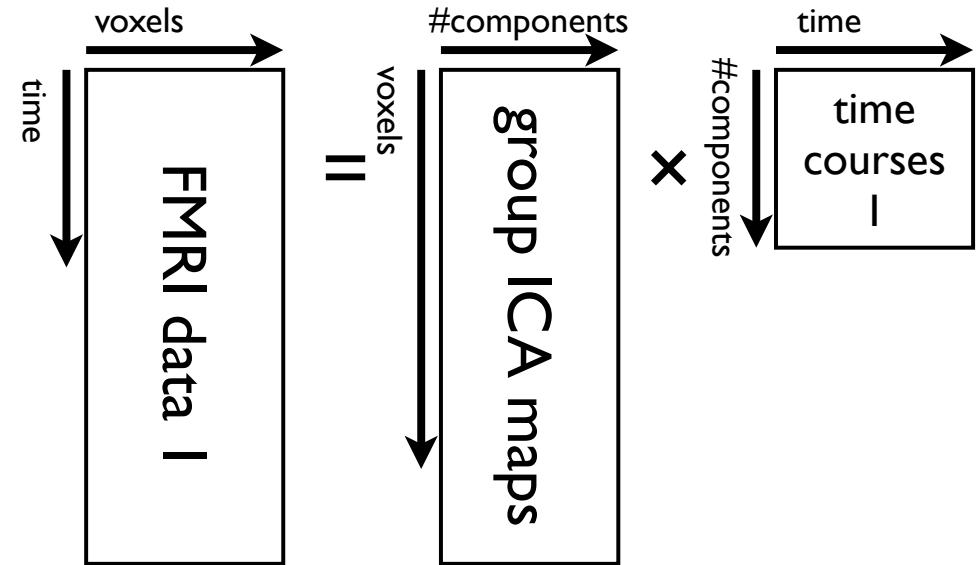
$$S^i \sim \mathcal{N}(\alpha_1 A^{i\dagger} Y^i + \alpha_2 S_g, 1/\beta^2) \quad \text{with}$$

$$\alpha_1 = \frac{1/\sigma_Y^2}{1/\sigma_Y^2 + 1/\sigma_g^2}, \quad \alpha_2 = \frac{1/\sigma_g^2}{1/\sigma_Y^2 + 1/\sigma_g^2}$$
$$\beta^2 = 1/\sigma_Y^2 + 1/\sigma_g^2$$

for $\sigma_g \gg \sigma_Y$: $\widehat{S}^i = \widehat{A}^{i\dagger} Y^i$

‘Dual regression’

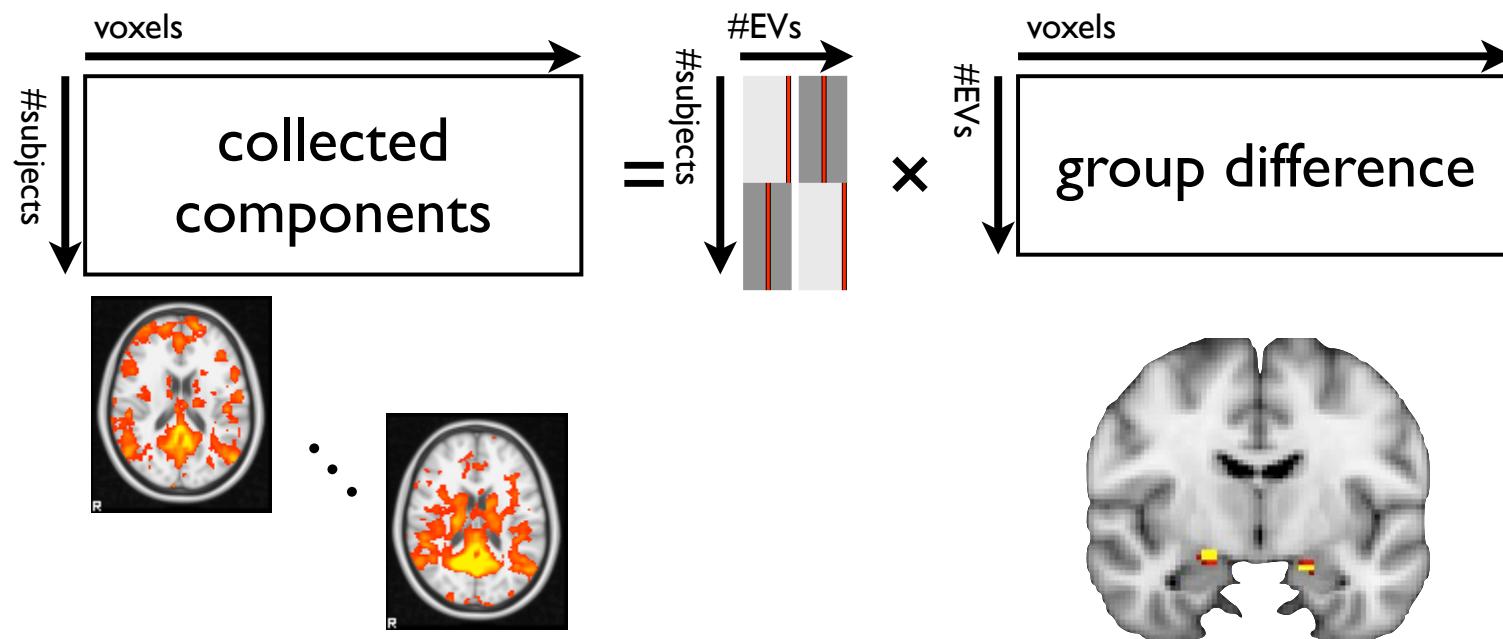
- Regress spatial ICs into each subject’s 4D data to find subject-specific timecourses
- regress these back into the 4D data to find subject-specific spatial ICs associated with the group ICs





‘Dual regression’

- Collect maps and perform voxel-wise non-parametric 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

I. Separate-subject ICAs

E.g. “SOG-ICA” - *BrainVoyager* (*Esposito, NeuroImage, 2005*)

- Separate ICA for each subject
 - can add robustness via multiple runs (“ICASSO”)
- Cluster components across subjects, to achieve robust matching of any given RSN across subjects
- **Good:** Keeps benefits of single-subject ICA - better modelling/ignoring of structured noise in data
- **Bad:** Correspondence problem, in particular different splittings in different subjects caused by even small changes in the data; PCA bias!

2. Group-ICA and Back-Projection

E.g. “GICA” - GIFT (*Calhoun, HBM, 2001*)

- Separate PCA (dimensionality reduction) for each subject
- Concat PCA-output across subjects and do group-ICA
- Back-project (invert) ICA results to get individual subject maps
- **Good:** No correspondence problem
- **Bad:**
 - Lose benefits of single-subject ICA
 - PCA-bias

3. Group-ICA and Dual-Regression

E.g. “*MELODIC+dual-reg*” - FSL (Beckmann, OHBM, 2009)

- Group-average PCA (dimensionality reduction)
 - Project each subject onto reduced group-average-PCA-space
 - Concat PCA-output across subjects and do group-ICA
 - Regress group-ICA maps onto individual subject datasets to get individual subject maps
-
- **Good:** No correspondence problem, no group/PCA bias
 - **Bad:** Lose benefits of single-subject ICA

Summary

- PCA bias prevents comparison of smallish effects
- hard to detect (in SOS evaluations)
- need to be mindful of its implications, particularly when inferring *differences in connectivity* based on maps which show *similarity in connectivity*



Acknowledgements

- Clare Mackay, Nicola Filippini, Mark Woolrich (FMRIB Oxford)
- David Cole (Imperial College London)
- Lisa Nickerson (Boston)
- wikipedia.org