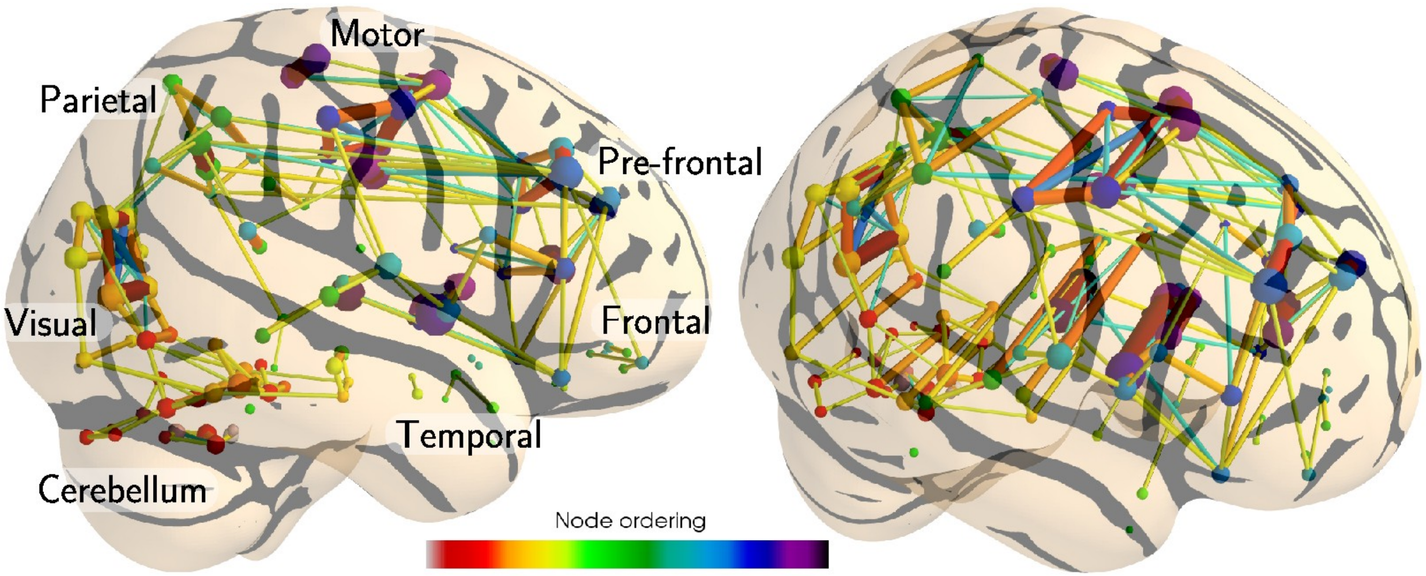
Statistical methods for functional brain connectivity mapping

Bertrand Thirion

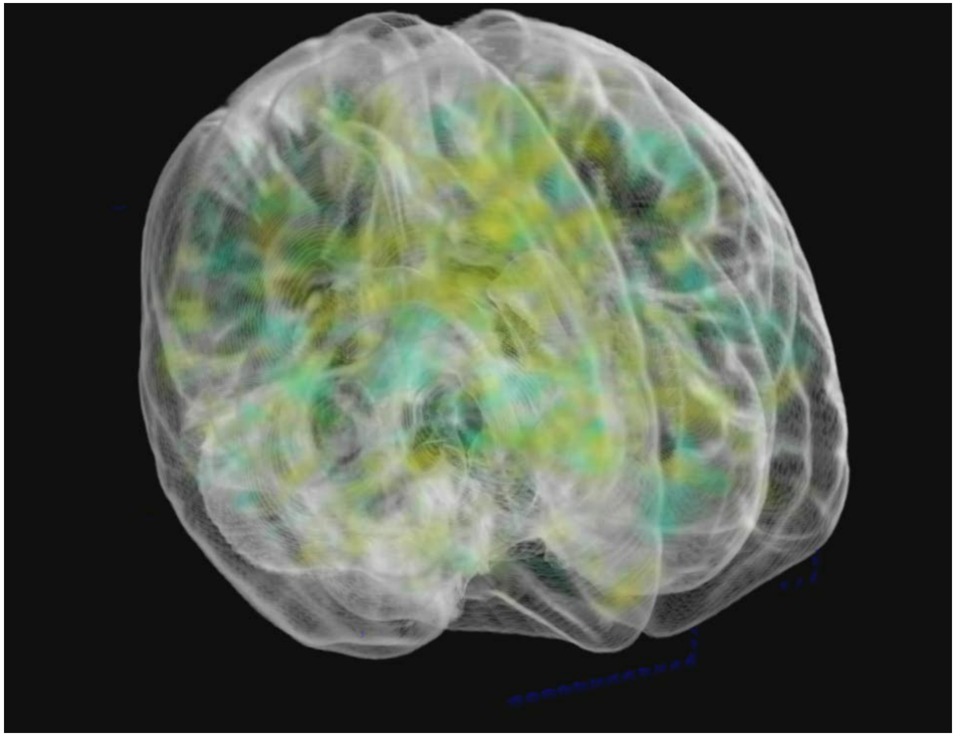
Inria, CEA, Université Paris-Saclay
bertrand.thirion@inria.fr

Outline

- Introduction to functional connectivity analysis
- Defining regions from fMRI data
- Statistical inference on covariance models



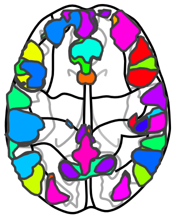
Functional Connectivity



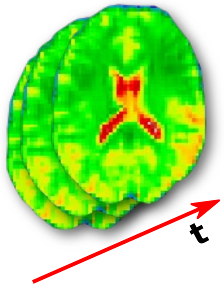
- **Resting-state experiments**: neural signals / absence of task
  - Informative on brain organization: **segregation** & **integration**
  - In task-fMRI, task covariates fit 2-5% of brain signals; remaining *intrinsic* activity not understood
- Neural signals **mixed** with external signal: perceptual & motor activity, biological rhythms etc.

The standard analysis pipeline

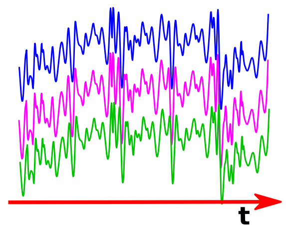
Atlas
p ROIs



r-fMRI

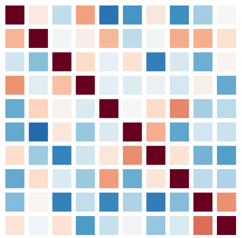


Timeseries extraction

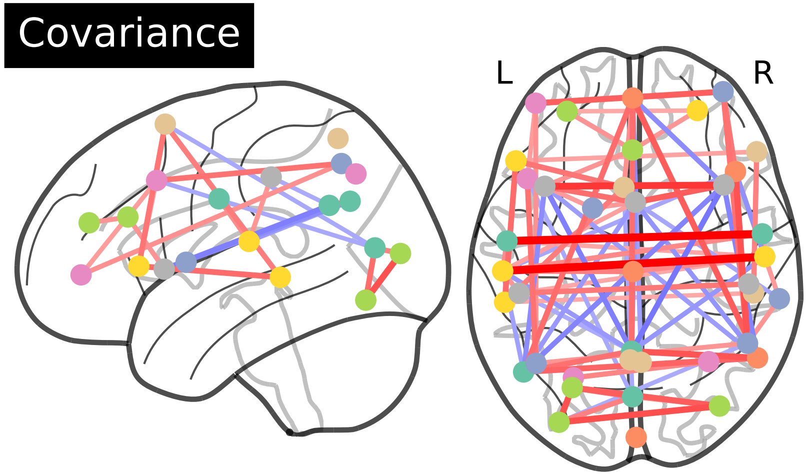


p timeseries

Empirical covariance

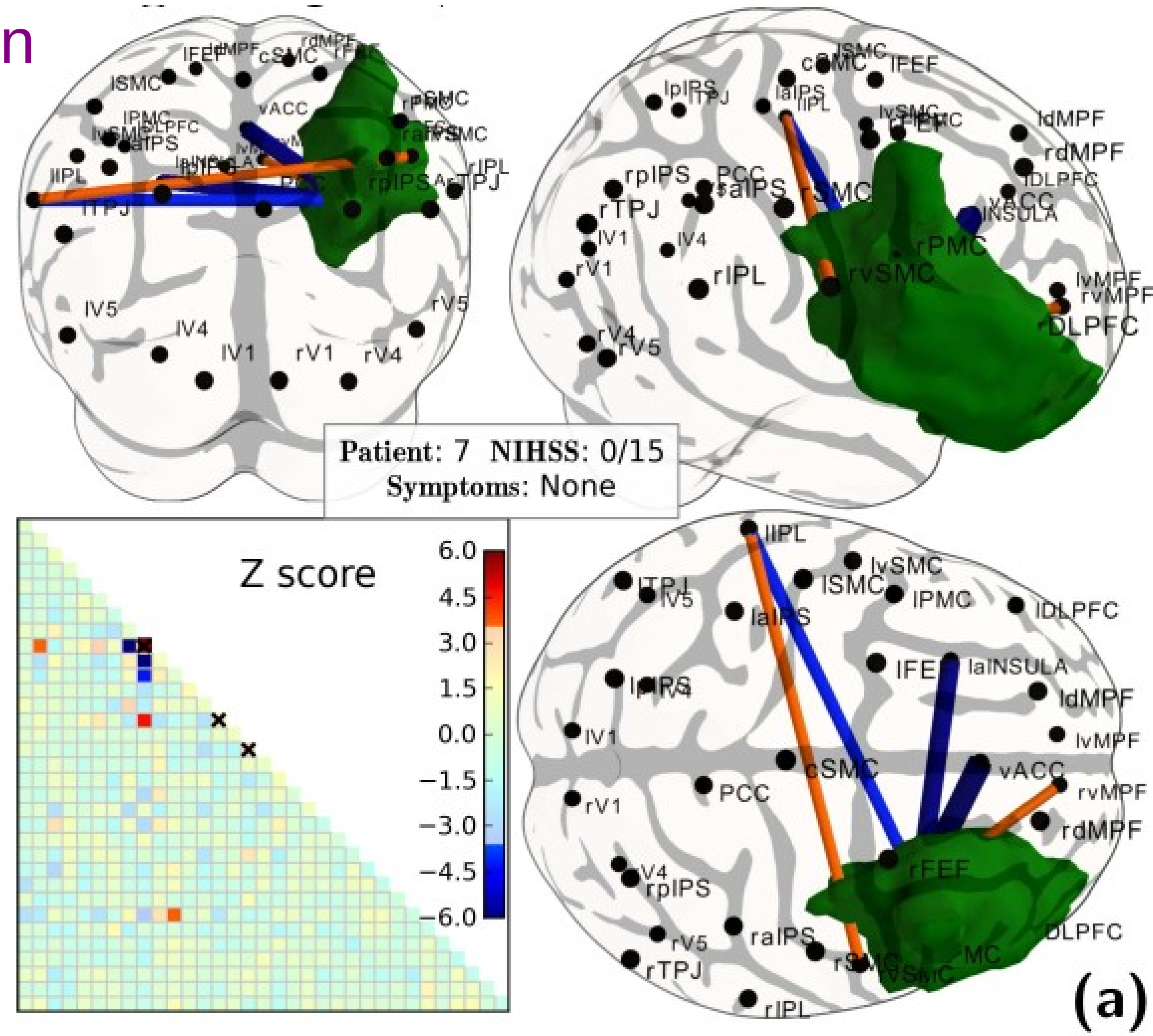


$\mathbf{S}$



Functional connectivity: objectives

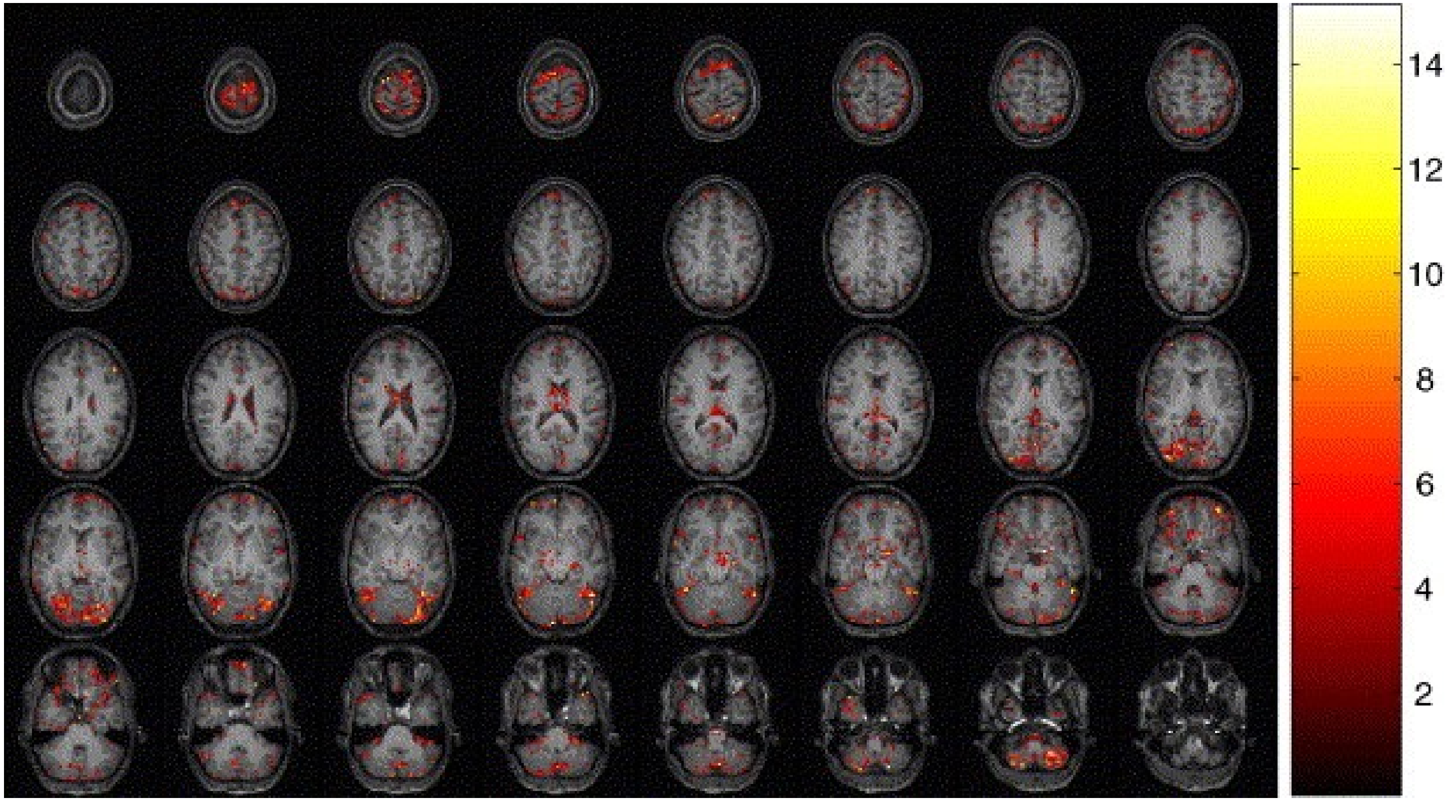
- Understand brain organization
    - Compare with anatomical connectivity
    - Compare with activation data
- marker for several brain diseases, and recovery processes
- Understand some brain diseases
- Interplay of ongoing with evoked activity



Signal cleaning

- Removing spurious physiology based signal

[Lund et al., NeuroImage 2006]

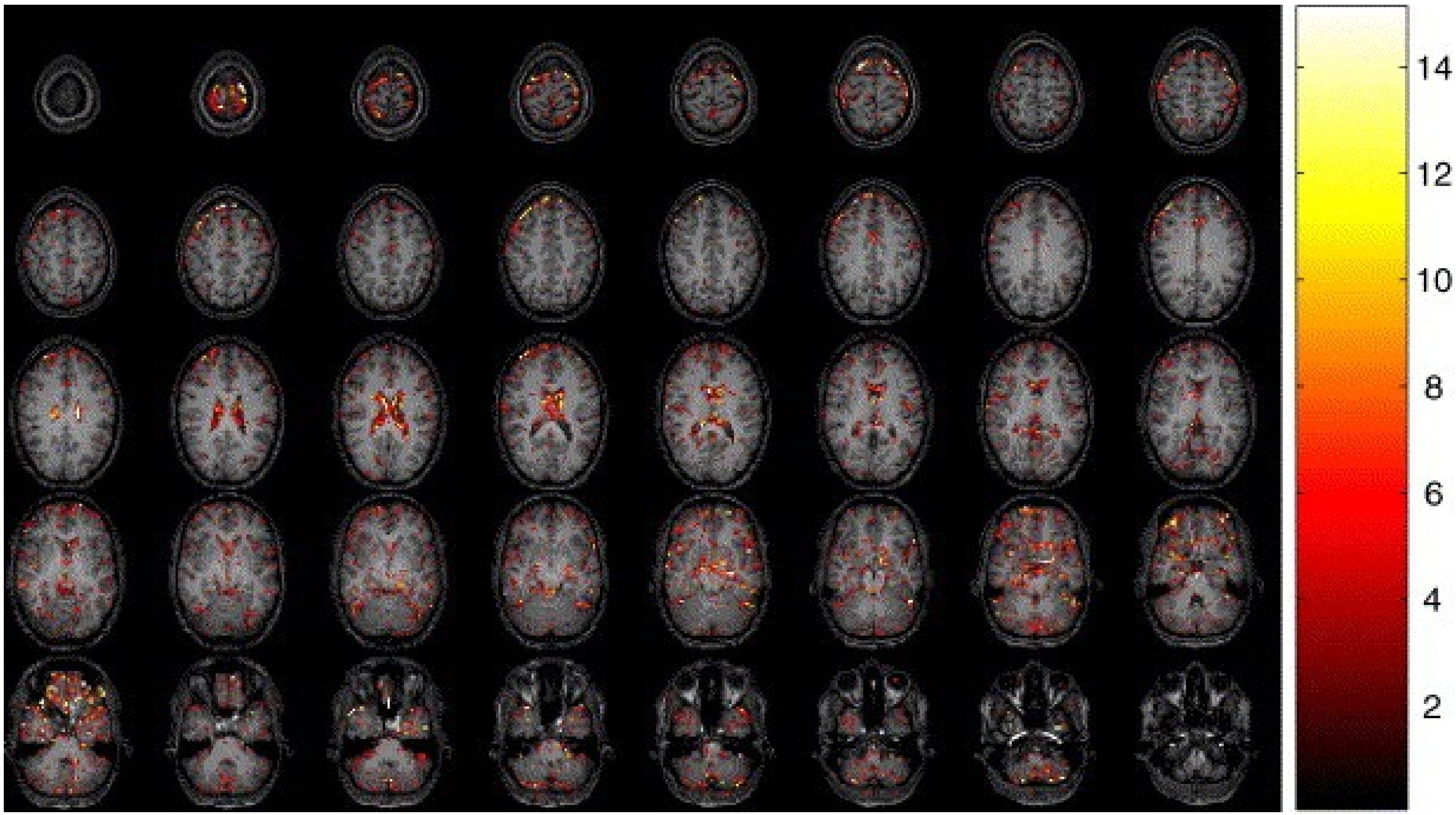


Residual movement effects.

Significant (FDR<0.05) effect of residual movement effects.

dominant near the **edges of the brain.**

Respiratory...

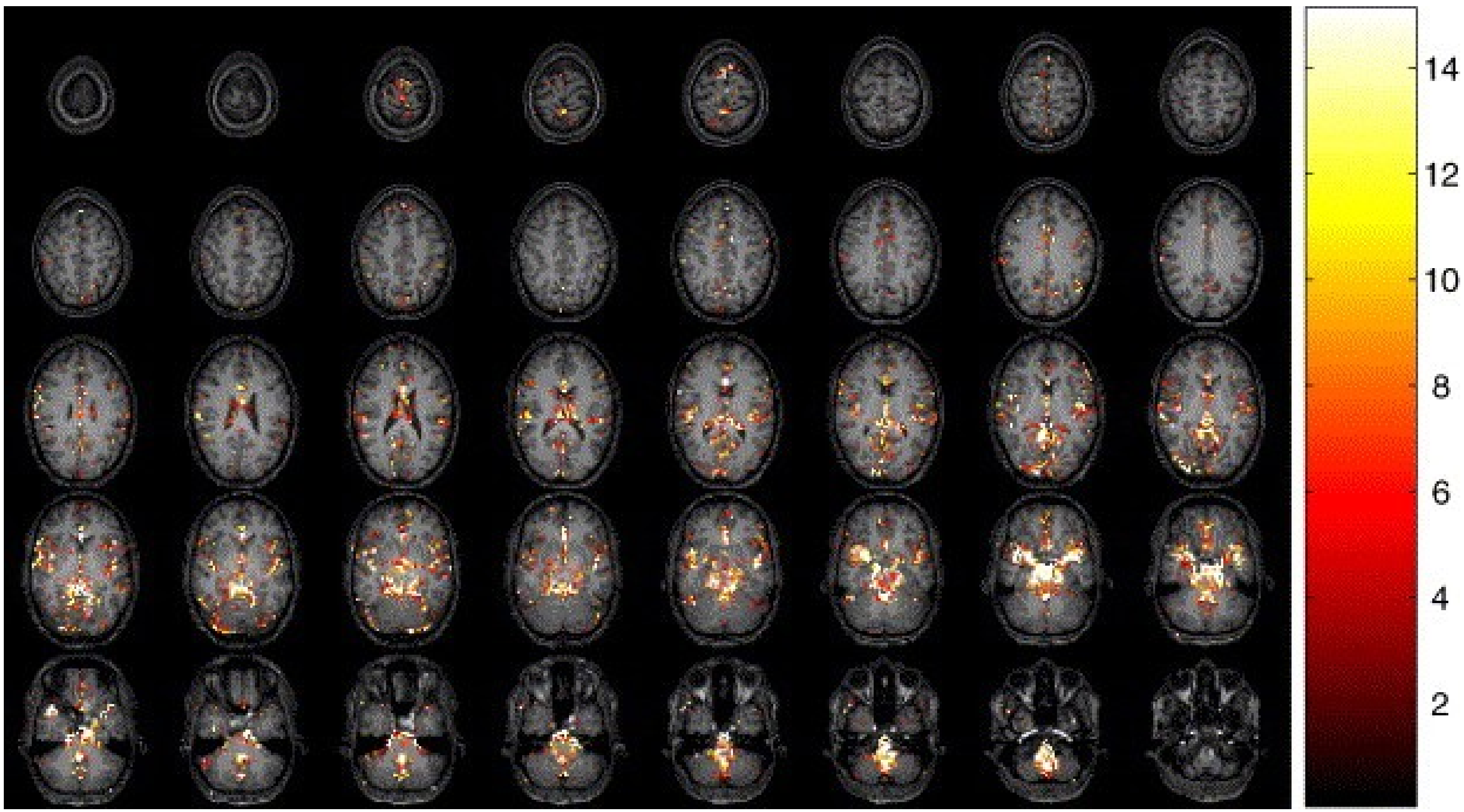


[Lund et al., NeuroImage 2006]

Respiratory-induced noise. significant (FDR<0.05) effect of respiratory oscillations.

Dominant near the **edges of the brain** as well as near in the **larger veins** and in the **ventricles**.

Cardiac ...



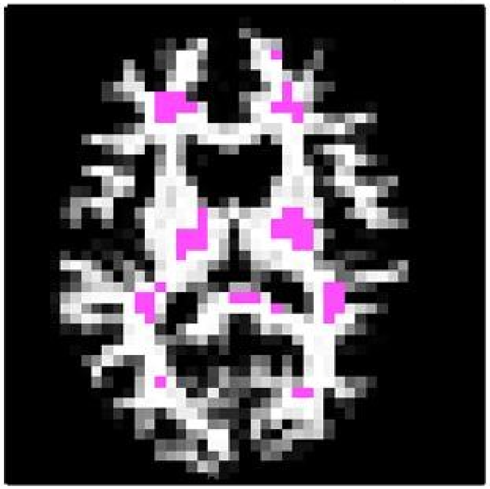
Significant (fwer<0.05) effect of aliased cardiac oscillations.

Dominant near **larger vessels (e.g. medial cerebral artery and Circle of Willis)**.

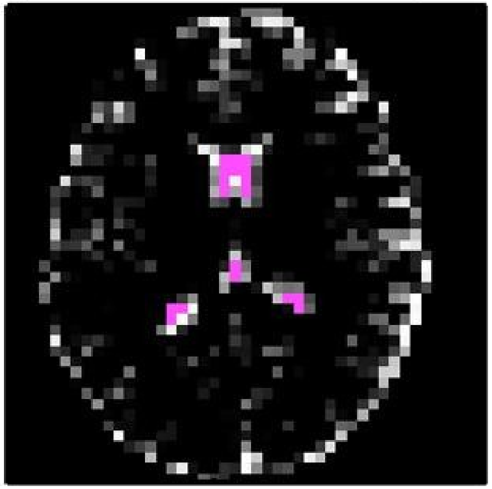
Signal cleaning in practical settings

1. Band-pass filtering: [0.01, 0.1 Hz]

a) White Matter Partial Volume Mask, Subject 1



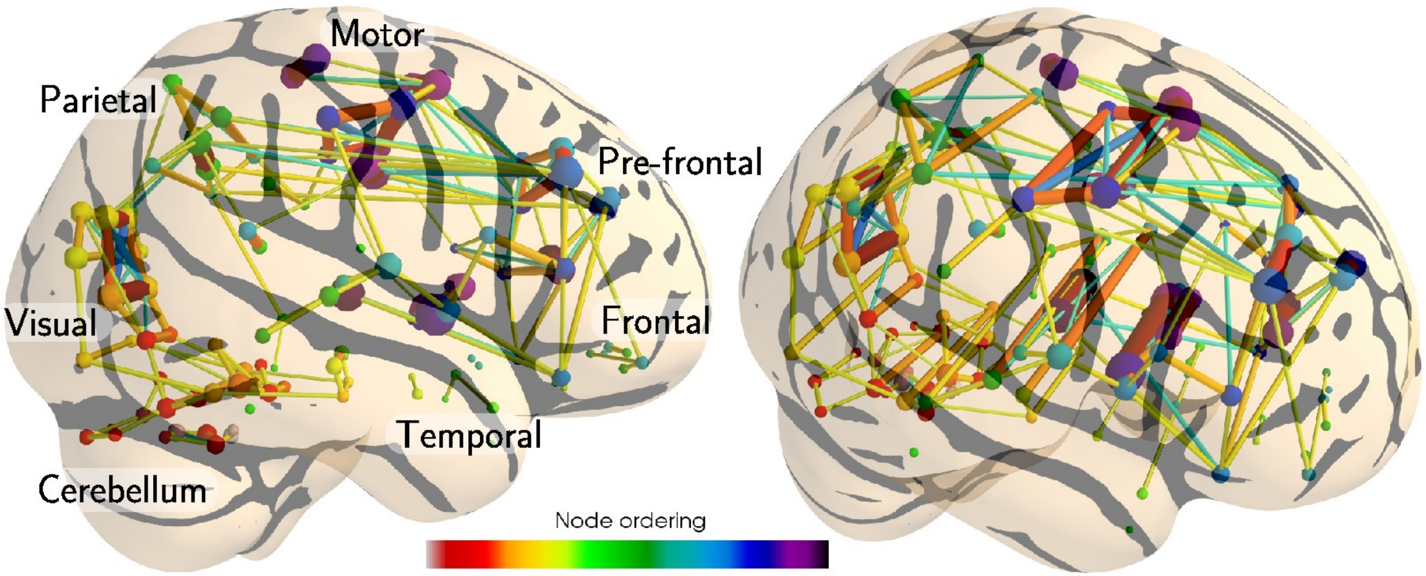
b) CSF Partial Volume Mask, Subject 1



2. Confound removal
[Behzadi et al. Nimg 2007]
PCA of the signals in the voxels with highest variance
or
remove signal components from WM, ventricles

Outline

- Objectives & data preparation
- Multivariate decompositions of brain images
- Statistical inference on covariance models



Learn brain atlases from fMRI data:

- Capture the structure of interest in **multivariate decompositions**

Y : (n, p) matrix
n = # time samples
p = # voxels

$$Y = UV^T + E$$

data — temporal components — spatial components — noise

PCA:	$\mathbf{U}_k \perp \mathbf{U}_l$ and $\mathbf{V}_k \perp \mathbf{V}_l$ if $k \neq l$
Clustering:	$\mathbf{V}_k \in \{0,1\}\, \forall k$ and $\mathbf{V}.\mathbf{e}_K = \mathbf{e}_p$
ICA:	$\mathbf{V}_k$ independent of $\mathbf{V}_l$ if $k \neq l$
sparse PCA:	$\mathbf{V}_k \perp \mathbf{V}_l$ if $k \neq l$, $\|\mathbf{V}_k\|_1$ small $\forall k$

Choice of K: model selection problem

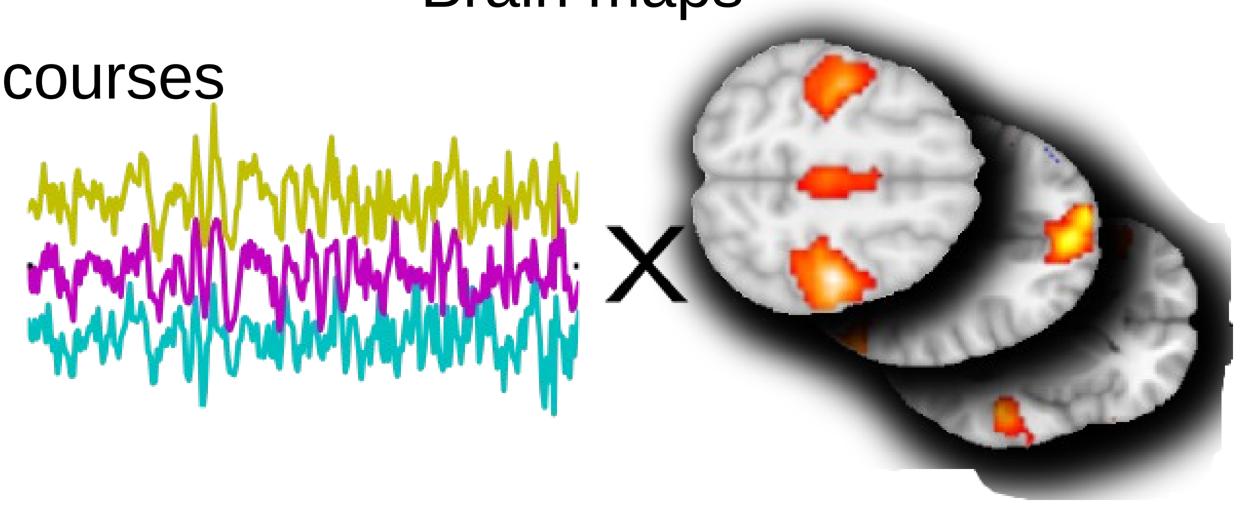
Learn brain atlases from fMRI data

Analogy with the GLM

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^T + \varepsilon$$
$$= \mathbf{X}\beta + \varepsilon$$

Brain maps

Time courses



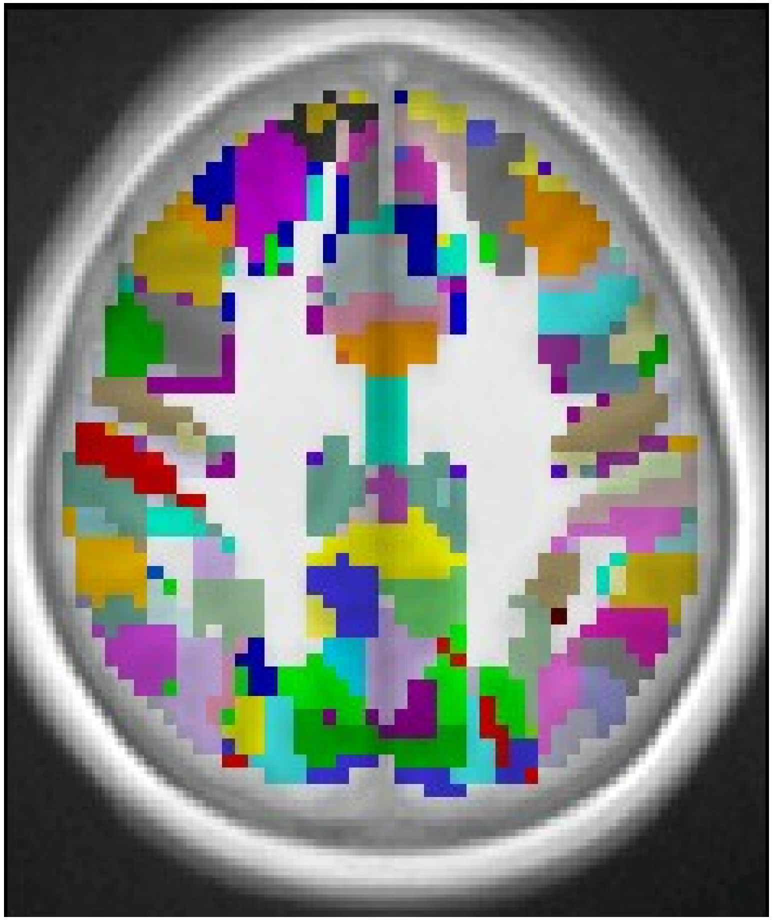
Main difference: “design matrix” unknown; to be estimated from data
→ “mixing matrix”

Learning brain atlases: Estimation techniques

- Clustering: k-means, Ward
- PCA: centering + singular value decomposition
- ICA: minimize marginal entropy of the reconstructed sources
- Sparse PCA: alternated optimization
  - Update of U: least squares
  - Update of V: K lasso problems $\min_{\mathbf{V},\mathbf{U}:\|\mathbf{U}_k\|\leq 1,\forall k} \|\mathbf{Y}-\mathbf{U}\mathbf{V}^T\|^2+\lambda\|\mathbf{V}\|_1$

Online algorithms for L_1 penalty in |Mairal et al. JMLR 2010]

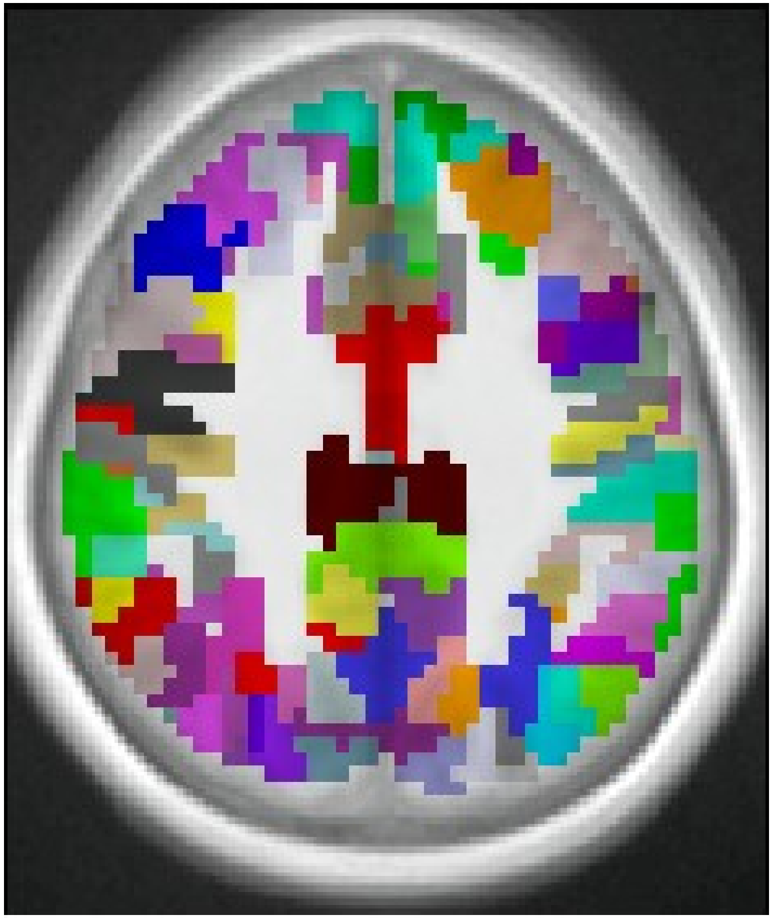
Clustering: k-means



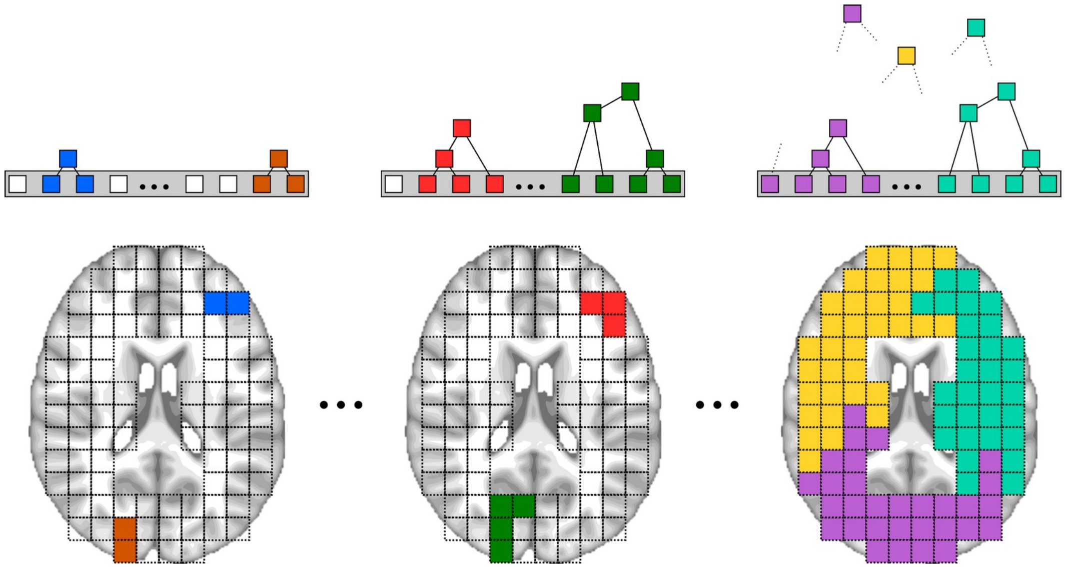
- Works well for small k
- No priors
- Need prior data smoothing

[Yeo et al. J Neurophysiol 2011]

Clustering: Ward



- Works well for large k
- Fast
- Includes spatial constraints



[Michel et al. PR 2011]

Independent Components Analysis (ICA)

Generative model $\mathbf{Y} = \mathbf{U}\mathbf{V}$

(k,p) (k, k) (k,p)

analysis $\hat{\mathbf{V}} = \mathbf{W}\mathbf{Y}$

Classical (noise-free) ICA setting

k = rank of decomposition

First step $\text{Dimension reduction / denoising}$

$$\mathbf{Y} = \mathbf{U}\mathbf{V} + \mathbf{E}$$

(n, p) (n, k) (k,p) (n, p)

$$\text{SVD of } \mathbf{Y} : \ \mathbf{Y} = \mathbf{A}\mathbf{\Lambda}\mathbf{D}$$

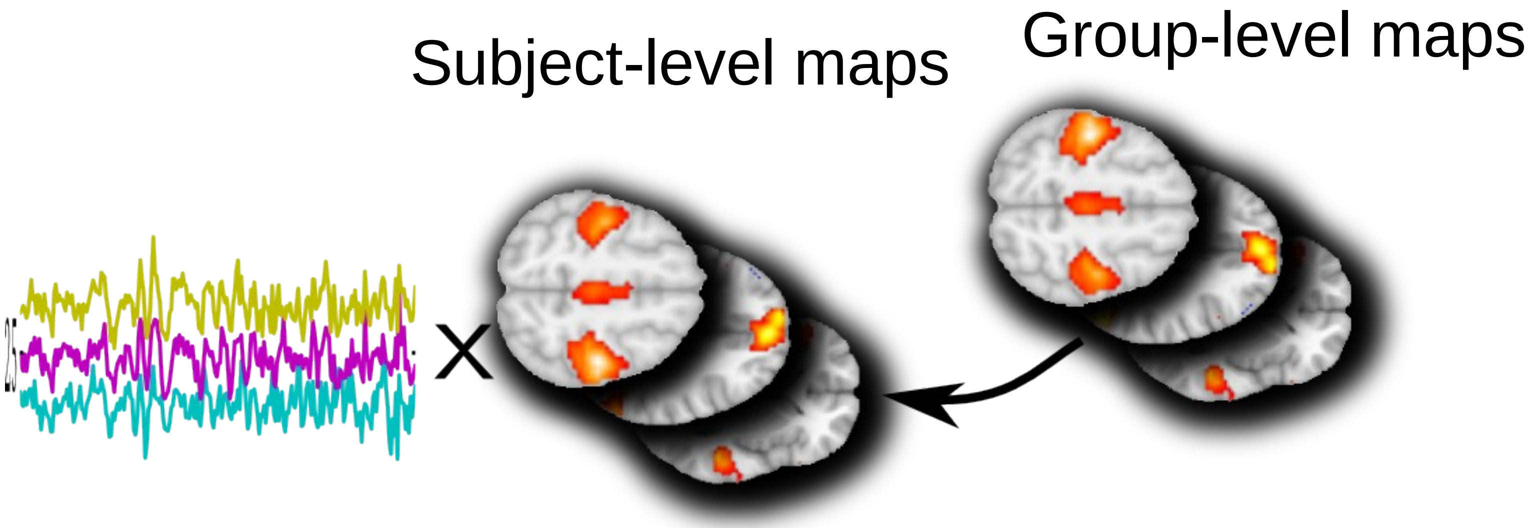
(n, p) (n, p) (p,p) (p, p)

denoising $\mathbf{Y} = \mathbf{A}\mathbf{\Lambda}[:, :k]\mathbf{D}[:k] + \mathbf{A}\mathbf{\Lambda}[:, k:]\mathbf{D}[k:]$

whitening: set singular values to 1 $\tilde{\mathbf{Y}} \leftarrow \mathbf{A}\mathbf{D}[:k]$

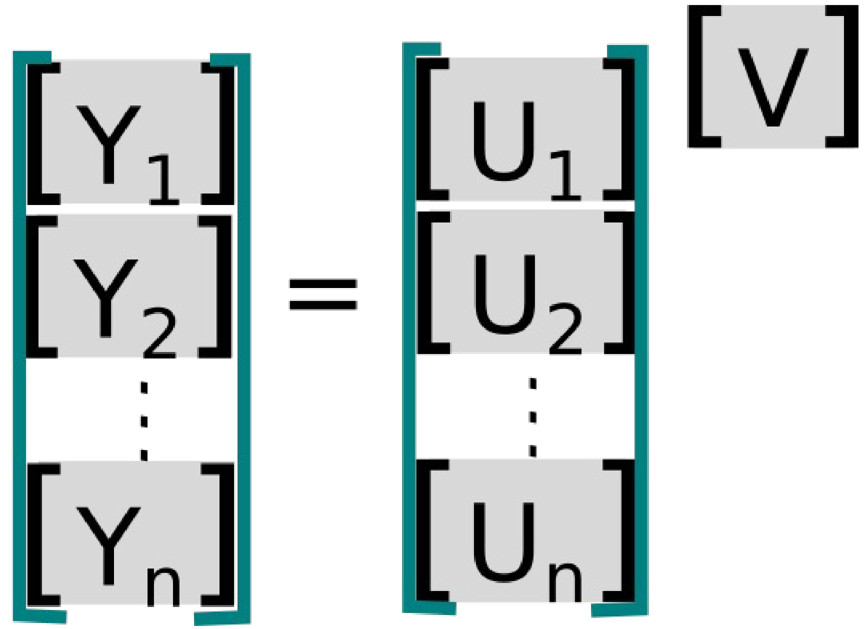
Brain atlases: multi-subject setting

Goal: consistent region segmentation across individuals



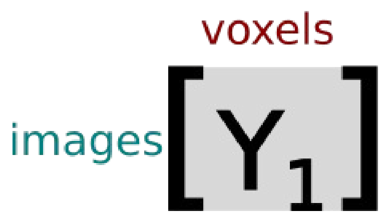
- Concatenated ICA [Beckmann et al. 2005]
- generalized Canonical correlation analysis [Varoquaux et al. Nimg2010]
- Multi-subject dictionary learning [Varoquaux et al. IPMI 2011]

computational strategies

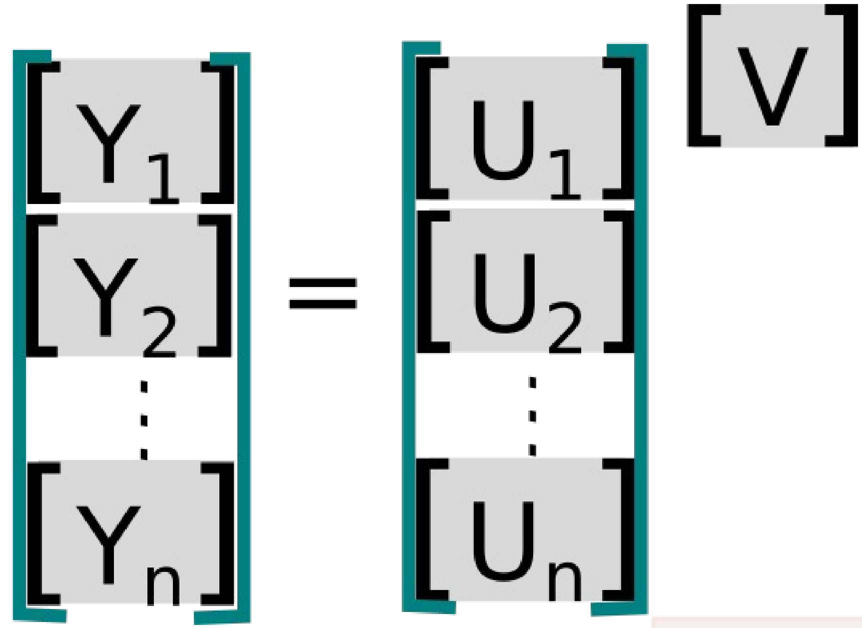


Concatenation approach

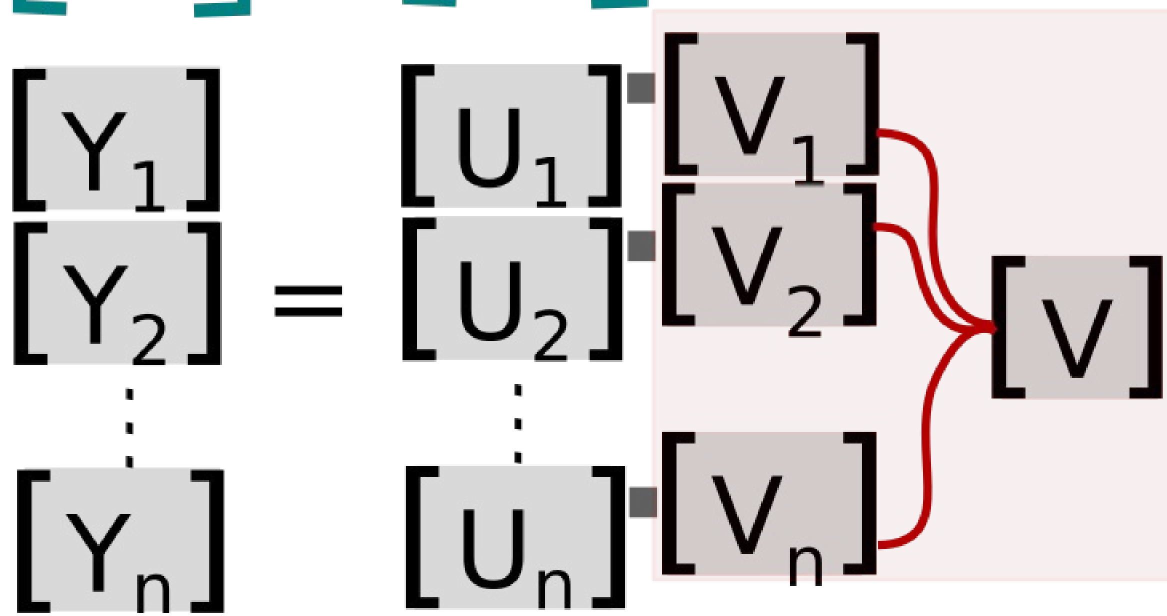
computational strategies



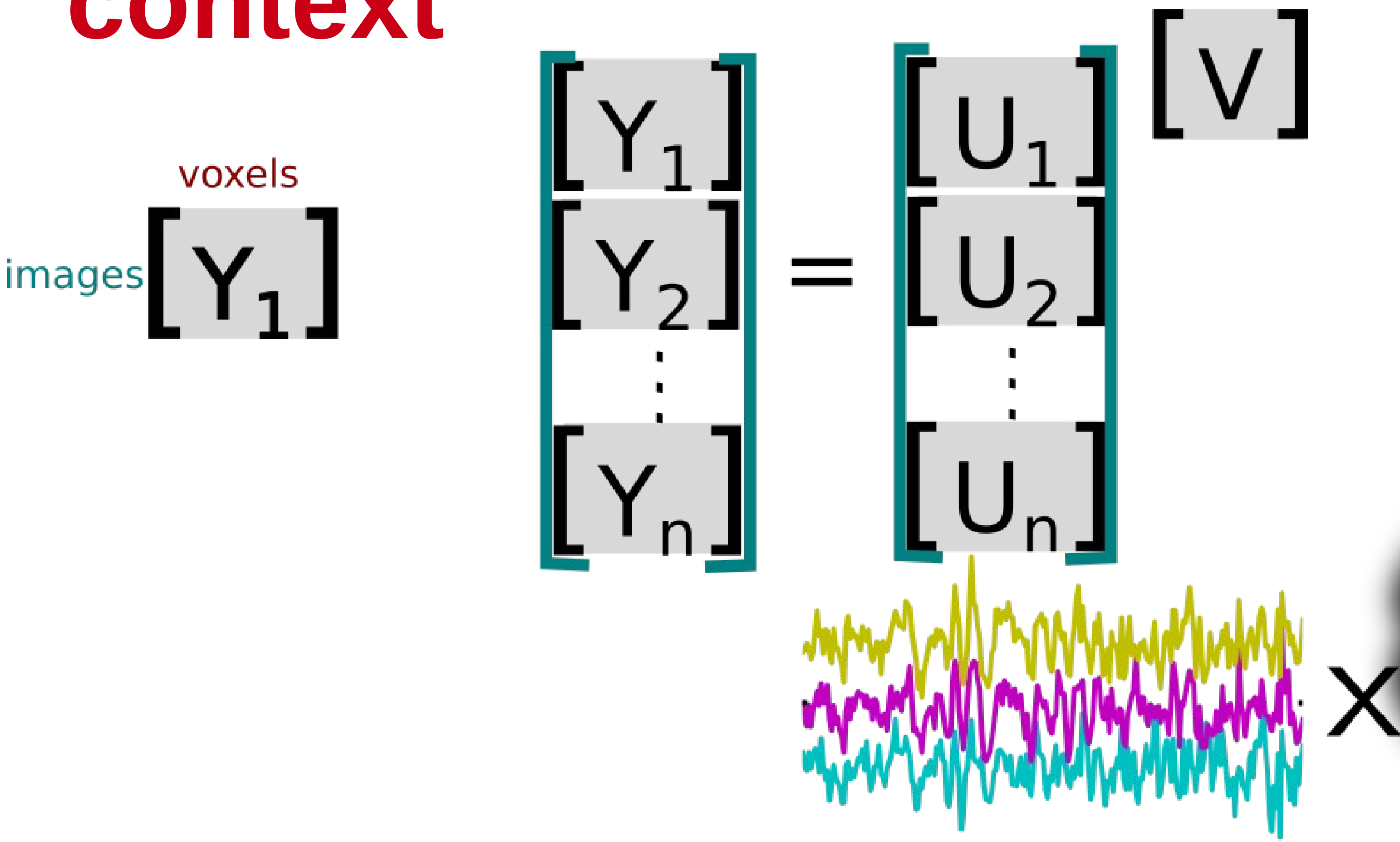
Concatenation approach

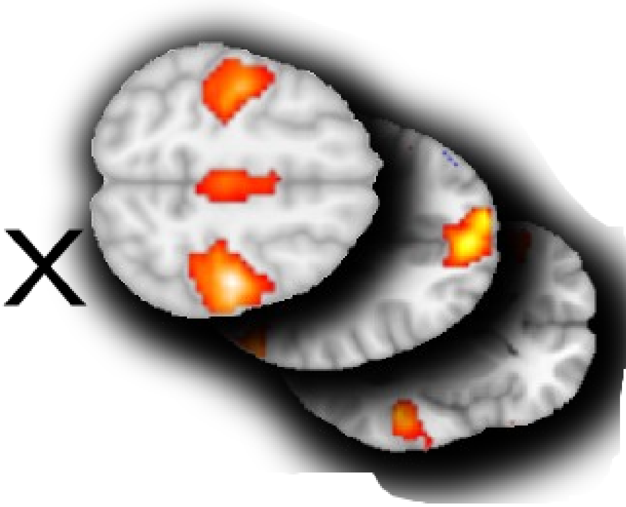


Latent factor approach



Dictionary Learning in a data-rich context





$$\min_{\mathbf{V},\mathbf{U}:\|\mathbf{U}_k\|\leq 1,\forall k} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\|^2 + \lambda\|\mathbf{V}\|_1$$

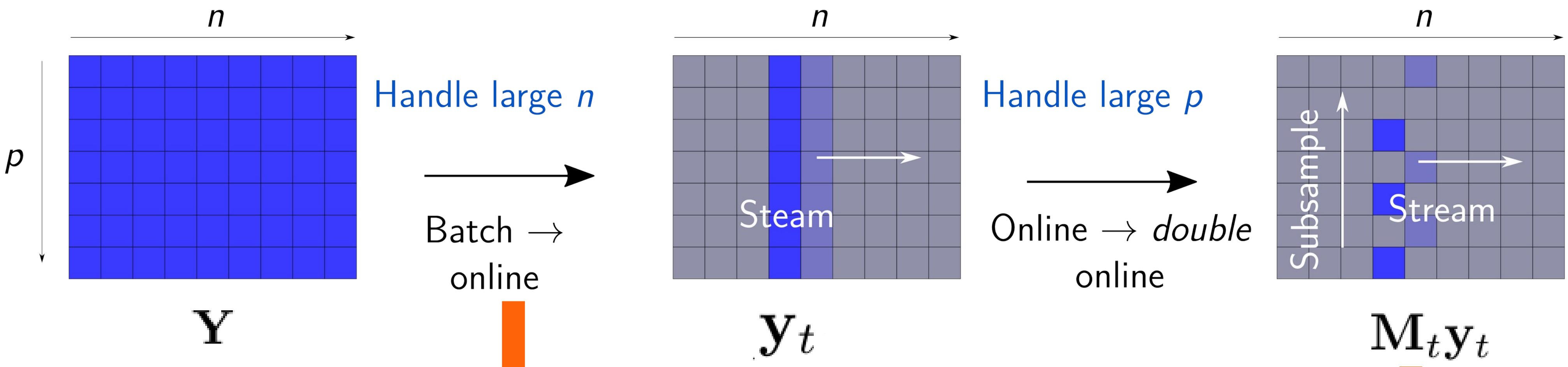
Concatenation mode for multi-subject → long “time courses”

Upscaling the model

- Human Connectome project $n=4.10^6$, $p=2.10^5$, 4**TB** of data
- Spatially-concatenated DL [varoquaux IPMI 2013] does not scale well:
  - Requires loading the data [Mensch et al. ISBI 2016]
- Solution
  - Work on batches of images **and** voxels
    - Online method in both samples and feature dimensions

Stochastic gradient approaches

https://github.com/arthurmensch/modl



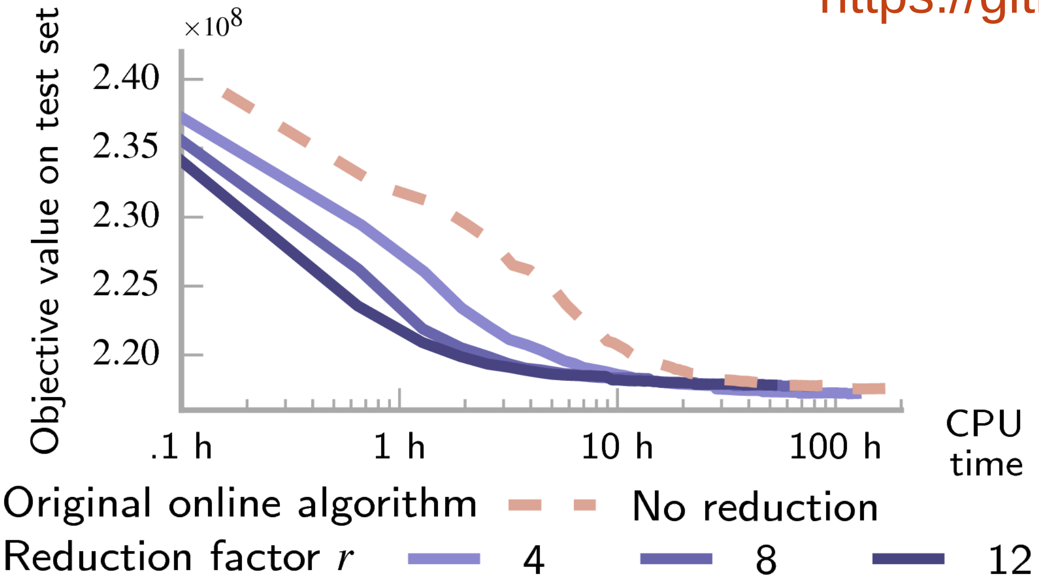
$$\mathbf{u}_t(\mathbf{V}) = \mathrm{argmin}_{\mathbf{u}\in\mathbb{R}^{n\times k}} \|\mathbf{y}_t - \mathbf{u}\mathbf{V}\|_F^2 + \lambda\Omega(\mathbf{u})$$

$$\mathbf{V}_t = \mathrm{argmin}_{\mathbf{V}\in\mathcal{C}} \sum_{i=1}^{t} \|\mathbf{y}_i - \mathbf{u}_i\mathbf{V}\|_F^2$$

$$\mathbf{u}_t(\mathbf{V}) = \mathrm{argmin}_{\mathbf{u}\in\mathbb{R}^{n\times k}} \|\mathbf{M}_t\left(\mathbf{y}_t - \mathbf{u}\mathbf{V}\right)\|_F^2 + \lambda\frac{s}{p}\Omega(\mathbf{u})$$

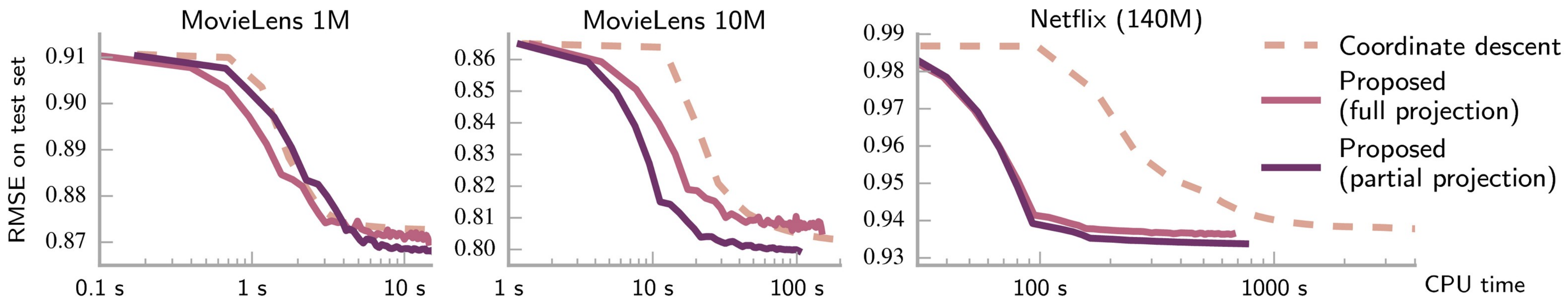
Stochastic gradient approaches

https://github.com/arthurmensch/modl

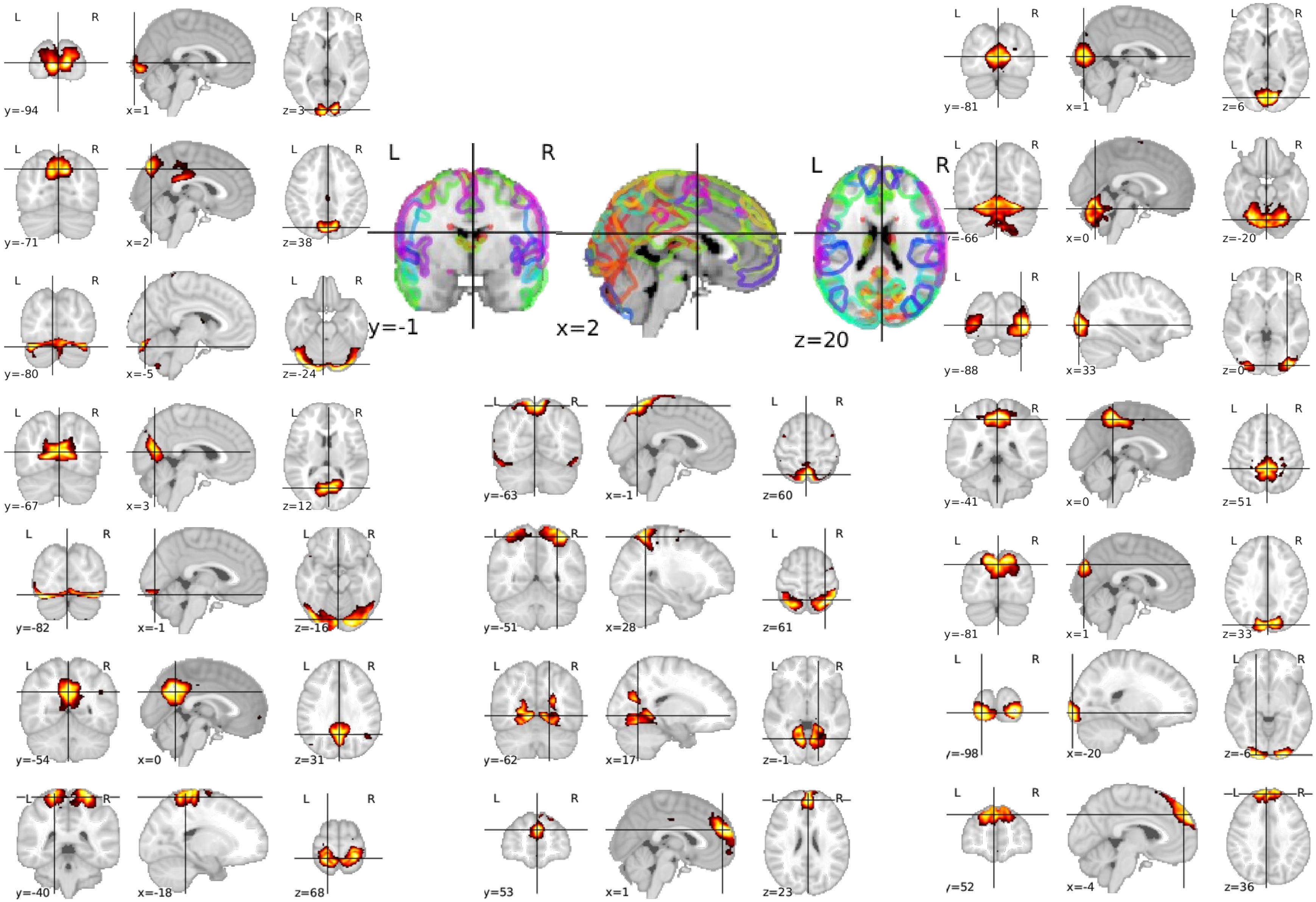


10-fold gain in CPU time
without loss in accuracy

Can be used for recommender systems



Brain atlases



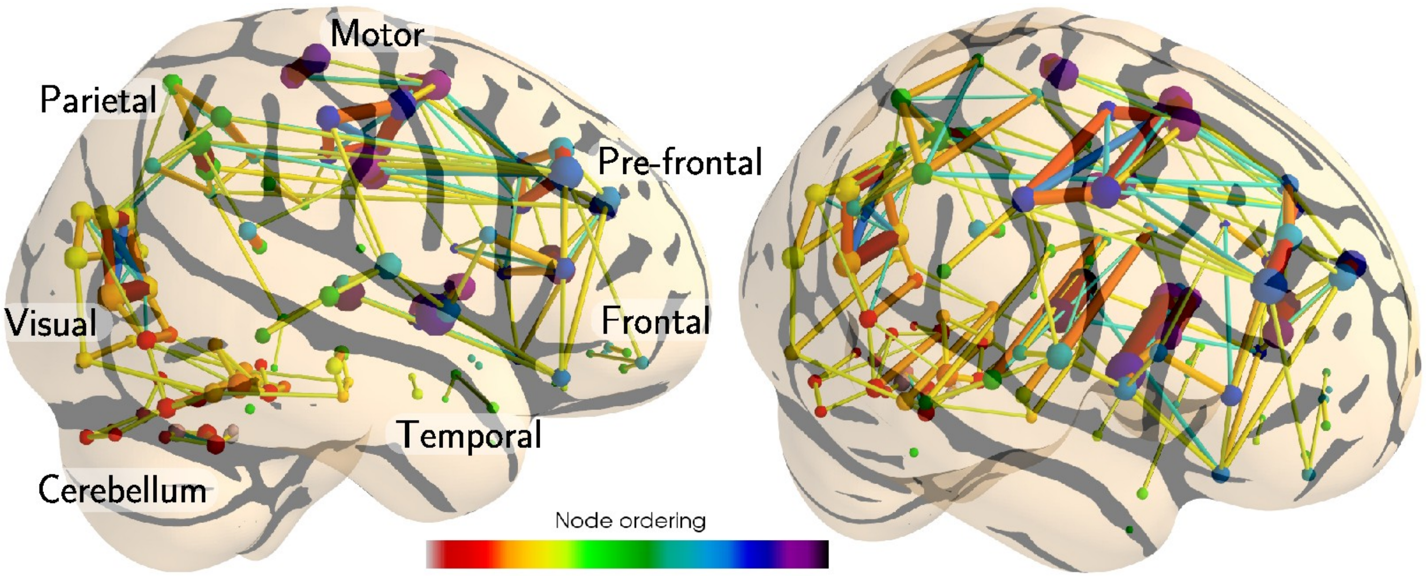
[Mensch et al. ICML 2016, IEEE TSP 2018]

Brain atlases from fMRI data

- Relax clustering to spatio-temporal decompositions
- Model selection is hard
- Sparse-/smooth- PCA algorithms perform well
  - Fine-tuning to individual data is expensive
  - Sophisticated penalties (eg TV+l1) expensive
- Development of online methods to scale to large data [Mairal et al. JMLR 2010, Mensch et al. ICML 2016, Mensch et al IEEE TSP 2018]

Outline

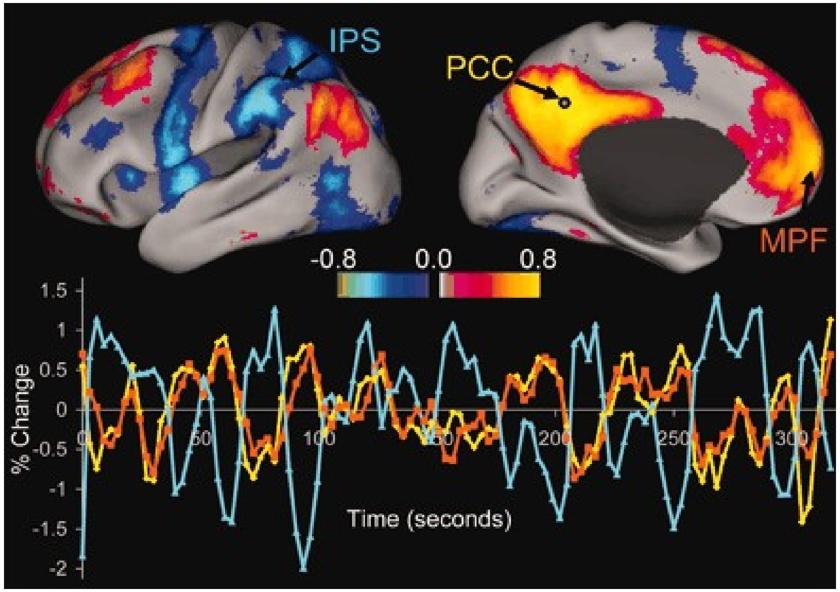
- Objectives & techniques
- Multivariate decompositions and ICA
- Statistical inference on covariance models
  - Covariance estimation
  - Statistical modeling and inference



The covariance structure of fMRI signals: From descriptive statistics to models

Problem setting:

- R regions
- Functional connectivity = statistical dependence of the signals cross regions
  - Integration of brain systems,
  - Population comparison,
  - Normal brain function/ diseases



[Fox et al PNAS 2005]

Main issues

- Relatively short time series ($<10^3$ samples in R regions)
- noise (physiological signals)
- between-subject variability

Covariance structure of fMRI signals

- Alternative: Generative model
  - fMRI time series $Y_1,..,Y_R$ as n samples from centered R-dimensional Gaussian variable
  - Infer the covariance/precision of this variable

$$\hat{\mathbf{K}} = \mathrm{argmin}_{\mathbf{K}\succ 0} \mathrm{tr}\left(\hat{\mathbf{\Sigma}}\mathbf{K}\right) - \mathrm{logdet}(\mathbf{K})$$

  - Where $\hat{\mathbf{\Sigma}}$ is the empirical covariance (correlation) matrix

$(\hat{\mathbf{\Sigma}}_{i,j}), (i,j) \in [1,p]^2$ sample (Pearson) correlations

$\left(\frac{-\mathbf{K}_{i,j}}{\sqrt{\mathbf{K}_{i,i}\mathbf{K}_{j,j}}}\right), (i,j) \in [1,p]^2$ partial correlations

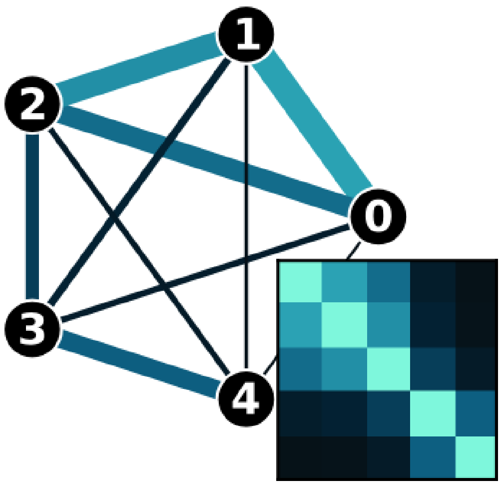
[Marrelec et al. 2006, Fransson et al. 2008]

The covariance structure of fMRI signals

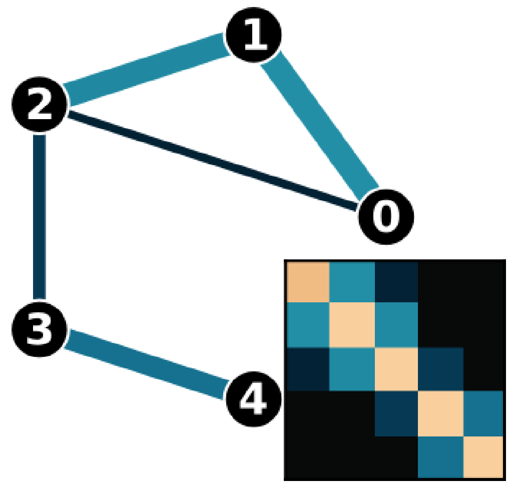
$$\hat{\mathbf{K}} = \mathrm{argmin}_{\mathbf{K}\succ 0}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}\mathbf{K}\right) - \mathrm{logdet}(\mathbf{K})$$

Empirical correlations

$\hat{\mathbf{\Sigma}}$

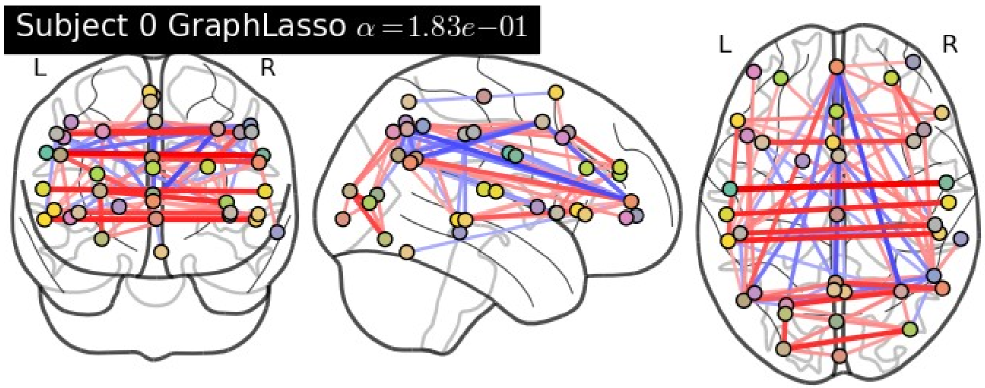


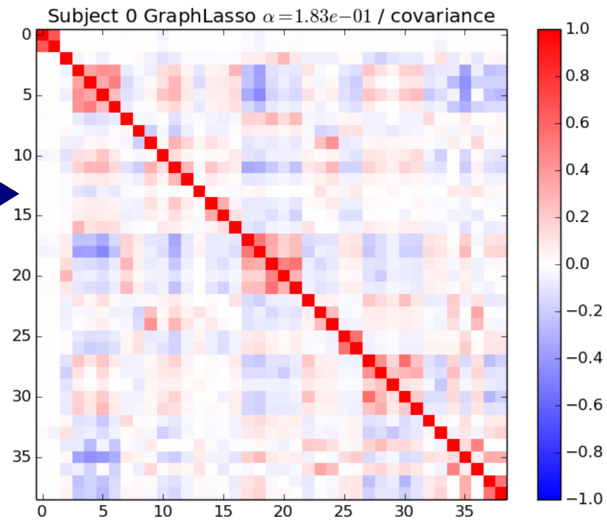
Underlying interactions
(=partial correlations) $\mathbf{K}$



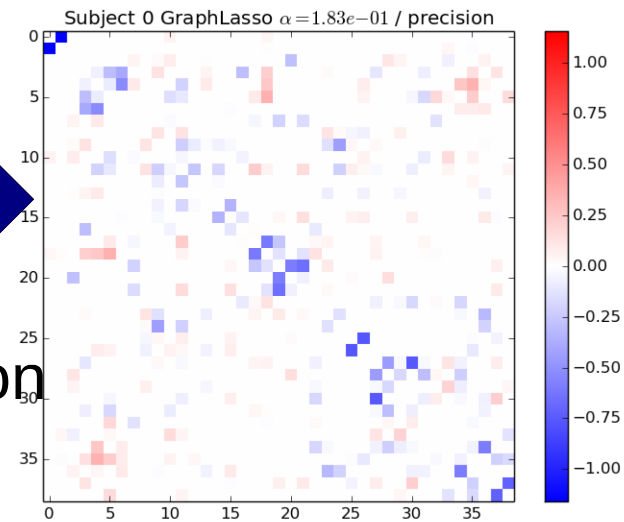
[Fransson et al. 2008]
Partial correlations highlight the backbone of brain networks

The covariance structure of fMRI signals





Covariance matrix



Precision matrix

The Ledoit Wolf estimator

Shrink empirical covariance $\min_{\rho_1,\rho_2} \mathbb{E}\left(\|\boldsymbol{\Sigma}^\star - \boldsymbol{\Sigma}\|^2\right)$ s.t. $\boldsymbol{\Sigma}^\star = \rho_1 \mathbb{I} + \rho_2 \hat{\boldsymbol{\Sigma}}$

Solution [Ledoit & Wolf, 2004] $\rho_1 = \frac{b^2}{d^2} Tr(\hat{\boldsymbol{\Sigma}}) \quad \rho_2 = 1 - \frac{b^2}{d^2}$

$$d^2 = \|\hat{\boldsymbol{\Sigma}} - Tr(\hat{\boldsymbol{\Sigma}})\mathbb{I}\|^2 \qquad b^2 = \frac{1}{n^2}\sum_{k=1}^{n} \|(\mathbf{X}_k \mathbf{X}_k)^T - \hat{\boldsymbol{\Sigma}}\|^2$$

CF L_2-shrinkage:

$$\hat{\mathbf{K}} = \mathrm{argmin}_{\mathbf{K}\succ 0} \mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}\mathbf{K}\right) - \mathrm{logdet}(\mathbf{K}) + \lambda \mathrm{tr}(\mathbf{K})$$

$$\hat{\mathbf{K}} = \left(\hat{\boldsymbol{\Sigma}} + \lambda \mathbb{I}\right)^{-1}$$

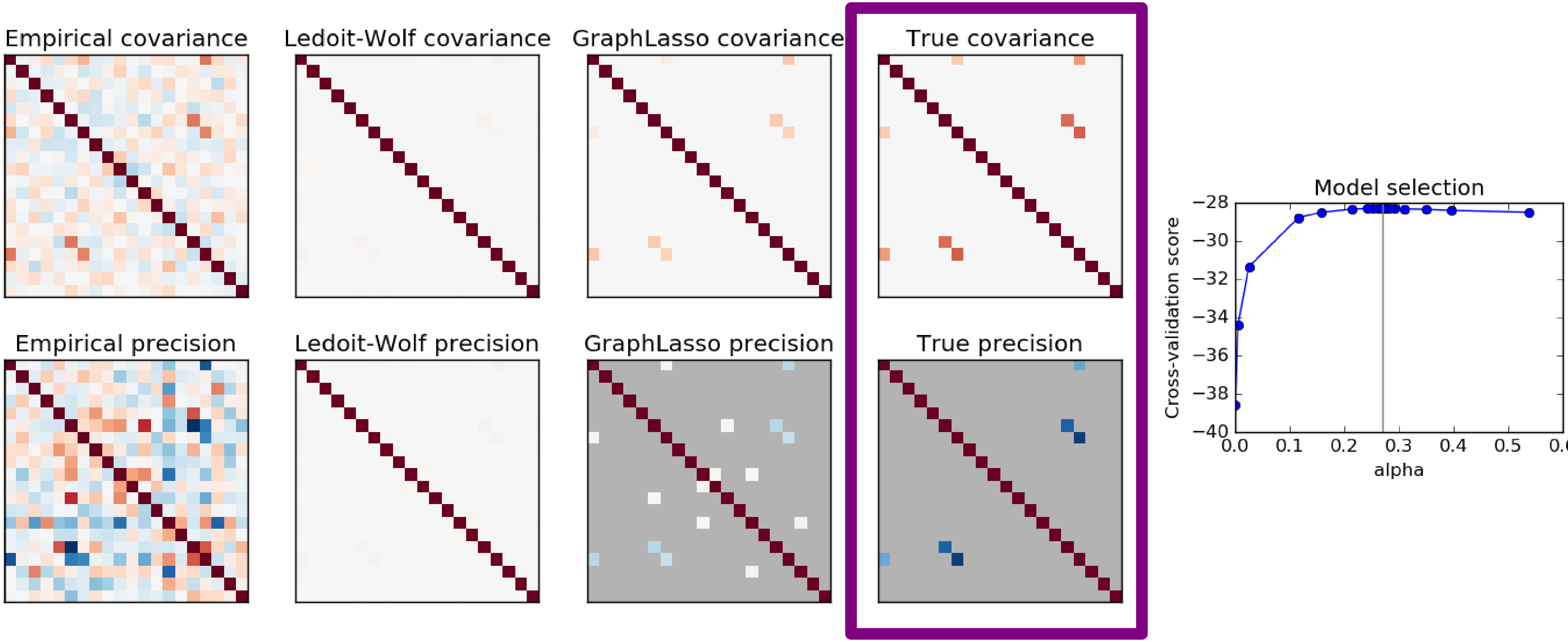
The covariance of fMRI signals: Covariance selection

- **enforce sparsity** instead of shrinkage
  - Few brain regions anatomically connected
  - Benefits to covariance estimation: covariance selection [Dempster at al. 1972]
  - 0's of K <=> independence conditional to the other nodes [Lauritzen, 1996]
  - L1 shrinkage on off-diagonal precision coefficients → convex estimation [Banejee et al. ICML 2006] [Friedmann et al. Biostat 2008]

$$\hat{\mathbf{K}}_{\ell_1} = \mathrm{argmin}_{\mathbf{K}\succ 0} \mathrm{tr}\left(\mathbf{K}\,\hat{\mathbf{\Sigma}}_{\text{sample}}\right) - \log\det\mathbf{K} + \lambda\|\mathbf{K}\|_1$$

Main issue: set λ

Covariance estimators

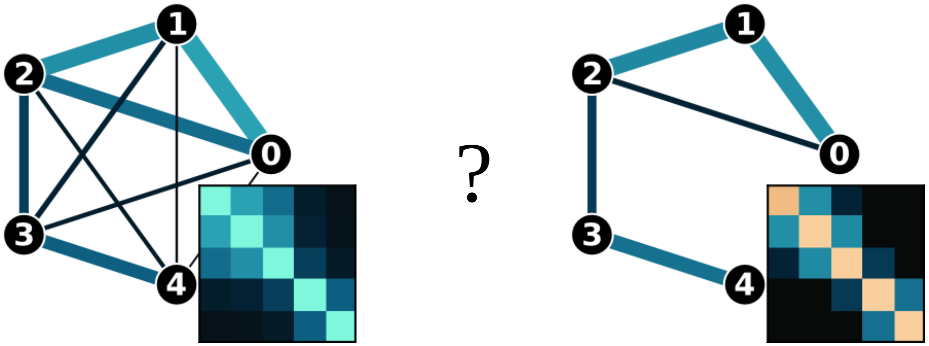


Ground truth

http://scikit-learn.org/stable/auto_examples/covariance/plot_sparse_cov.html#example-covariance-plot-sparse-cov-py

Statistical testing on graphical models

- State of the art: t-test [Fair et al. PNAS 2007] or discriminative models [Richiardi et al. Neuroimage 2011] on pairwise correlations
- Known issues
  - a test on correlations is not the optimal test
    - correlations are **not** observed **independently**
  - **diluted** effects: a local perturbation has widespread effects
  - Non-independent tests → **cannot localize** the true effects



Statistical testing on graphical models

- Use of affine invariant metric on SPD matrices

$$\mathrm{dist}_g(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \|\mathrm{Log}\left(\mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-\frac{1}{2}}\right)\|_2$$

- Geometric mean

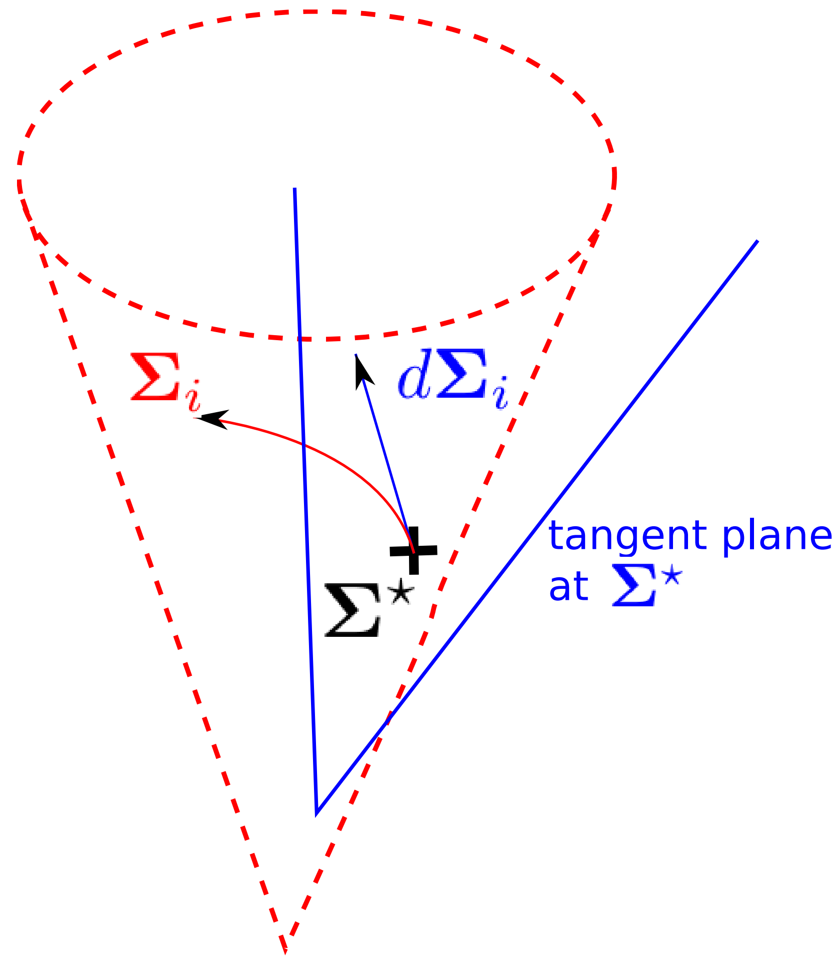
$$\mathbf{\Sigma}_\star = \mathrm{argmin}_{\mathbf{\Sigma}} \sum_{i=1}^{n} \mathrm{dist}_g(\mathbf{\Sigma}, \mathbf{\Sigma}_i)^2$$

- Geometric deviations

$$\mathrm{Let}\ \Phi_{\mathbf{\Sigma}} : \mathbf{S} \rightarrow \mathrm{Log}\left(\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{S}\mathbf{\Sigma}^{-\frac{1}{2}}\right)$$

$d\Sigma_i = \Phi_{\mathbf{\Sigma}_\star}(\mathbf{\Sigma}_i)$

“tangent” representation



[Varoquaux et al. MICCAI 2010]

Statistical testing on graphical models

Geometric distance on SPD matrices

$$\mathrm{dist}_g(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \|\mathrm{Log}\left(\mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-\frac{1}{2}}\right)\|_2$$

Geometric matrix average

$$\mathbf{\Sigma}_\star = \mathrm{argmin}_{\mathbf{\Sigma}} \sum_{i=1}^{n} \mathrm{dist}_g(\mathbf{\Sigma}, \mathbf{\Sigma}_i)^2$$

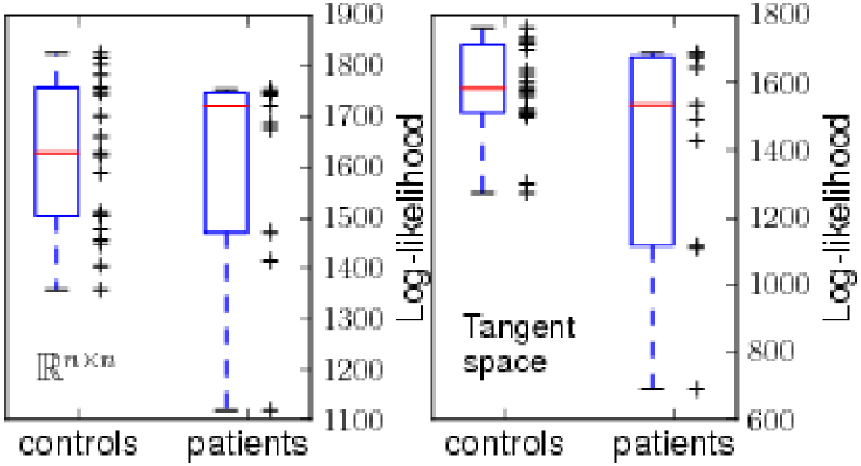
Geometric deviation

$$\text{Let } \Phi_{\mathbf{\Sigma}} : \mathbf{S} \to \mathrm{Log}\left(\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{S}\mathbf{\Sigma}^{-\frac{1}{2}}\right)$$

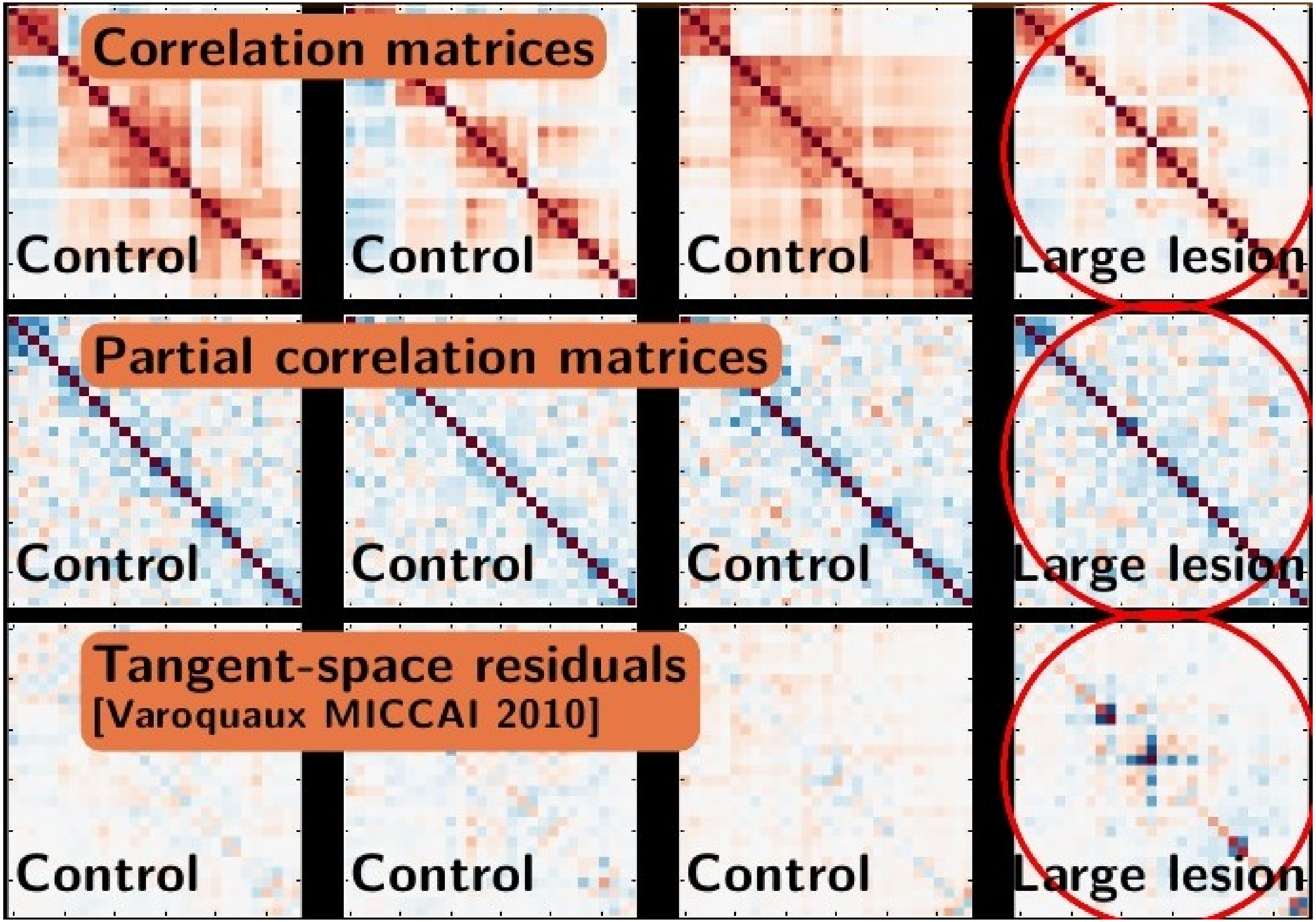
$$\Phi_{\mathbf{\Sigma}_\star}(\mathbf{\Sigma}_i)$$

[Varoquaux et al. MICCAI 2010]

Ensuing test discriminates patients from controls

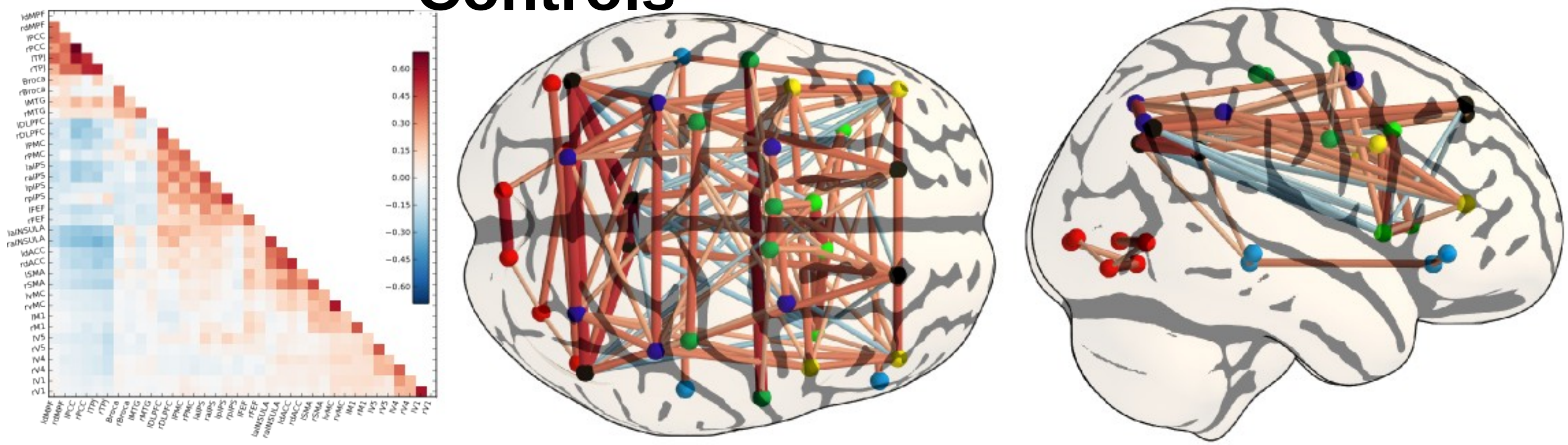


Statistical testing on graphical models

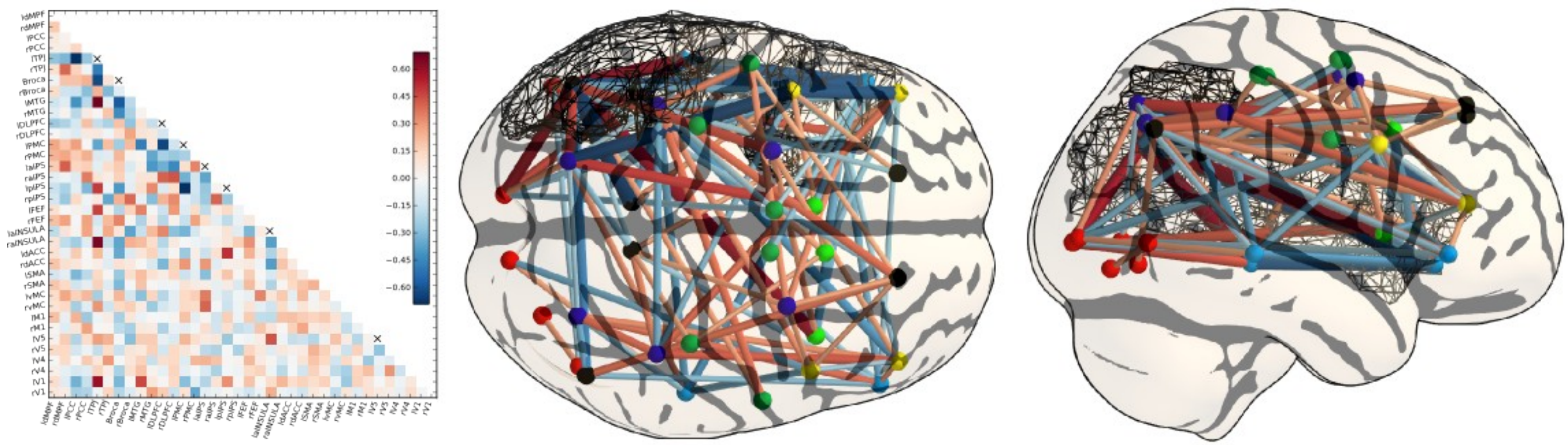


Statistical testing on graphical models

Controls



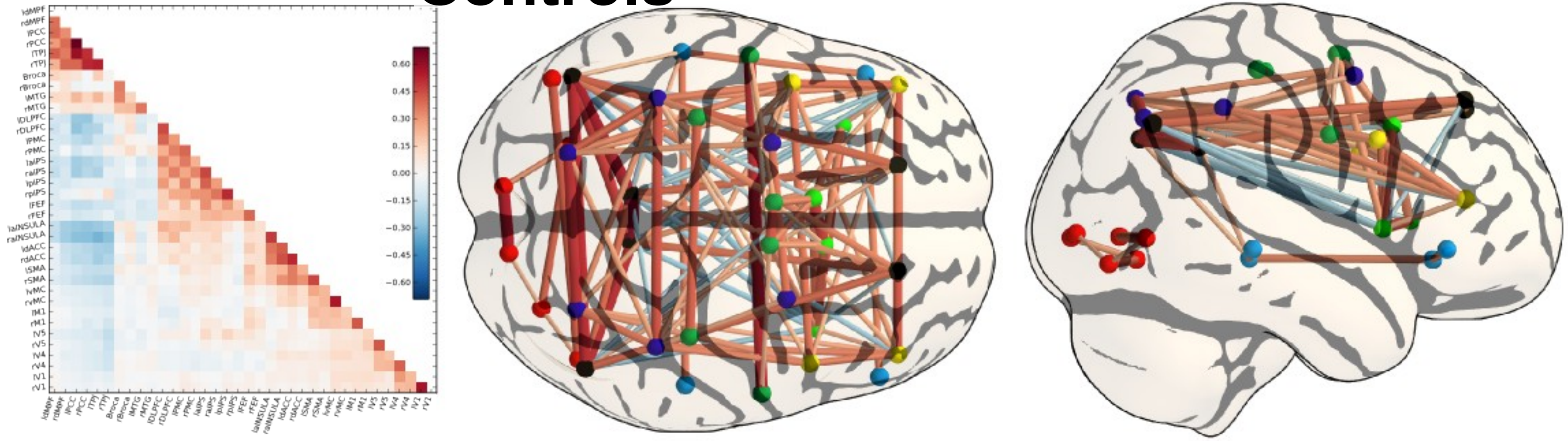
Perturbated Correlations after a stroke



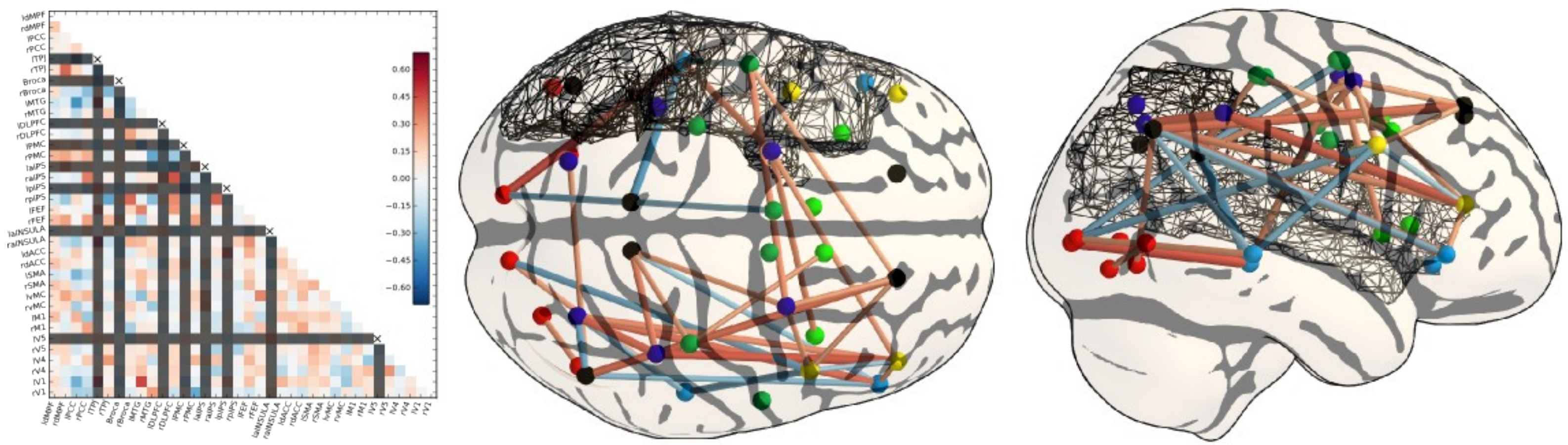
Data by F. Baronnet and A. Kleinschmidt. See [Varoquaux et al MICCAI 2010]

Statistical testing on graphical models

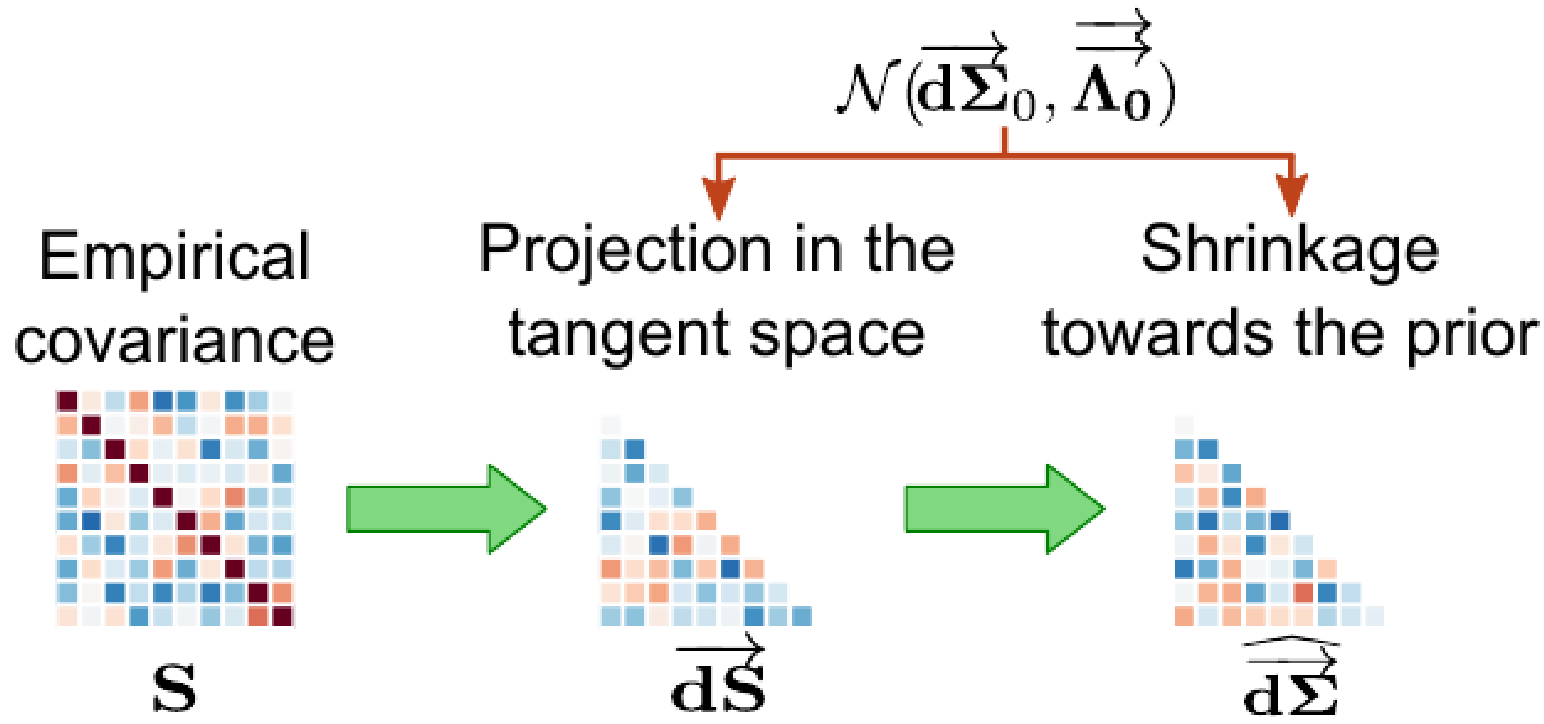
Controls



Perturbated Correlations after a stroke



Population model covariance



- Idea: regularize estimation in tangent representation

  → shrinkage toward population distribution

- Low-variance, biased estimator

Population model covariance

Tangent model $\mathbf{d\Sigma} = \mathrm{logm}(\mathbf{\Sigma}_0^{-\frac{1}{2}} \mathbf{\Sigma}\, \mathbf{\Sigma}_0^{-\frac{1}{2}})$

Vector model $\overrightarrow{d\Sigma} = \mathrm{vec}(\mathbf{d\Sigma}) = \{ \sqrt{2}\, \mathbf{d\Sigma}_{i,j},\ j < i,\ \mathbf{d\Sigma}_{i,i},\ i = 1 \ldots p\}$

prior $\overrightarrow{d\Sigma} \sim \mathcal{N}(\overrightarrow{d\Sigma}_0, \overrightarrow{\overrightarrow{\mathbf{\Lambda}}}_0)$

$$\overrightarrow{\overrightarrow{\mathbf{\Lambda}}}_0 = \frac{1}{N_{\mathrm{train}} - 1} \sum_{i=1}^{N_{\mathrm{train}}} \overrightarrow{dS}_i \otimes \overrightarrow{dS}_i$$

$$\overrightarrow{d\Sigma}_0 = \overrightarrow{0}$$

likelihood $p(\overrightarrow{DS} | \overrightarrow{d\Sigma}) = \mathcal{N}(\overrightarrow{dS}, \overrightarrow{\overrightarrow{\mathbf{\Lambda}}})$

Posterior inference $p(\overrightarrow{d\Sigma} | \overrightarrow{DS}) = \mathcal{N}(\widehat{\overrightarrow{d\Sigma}}, \overrightarrow{\overrightarrow{C}})$

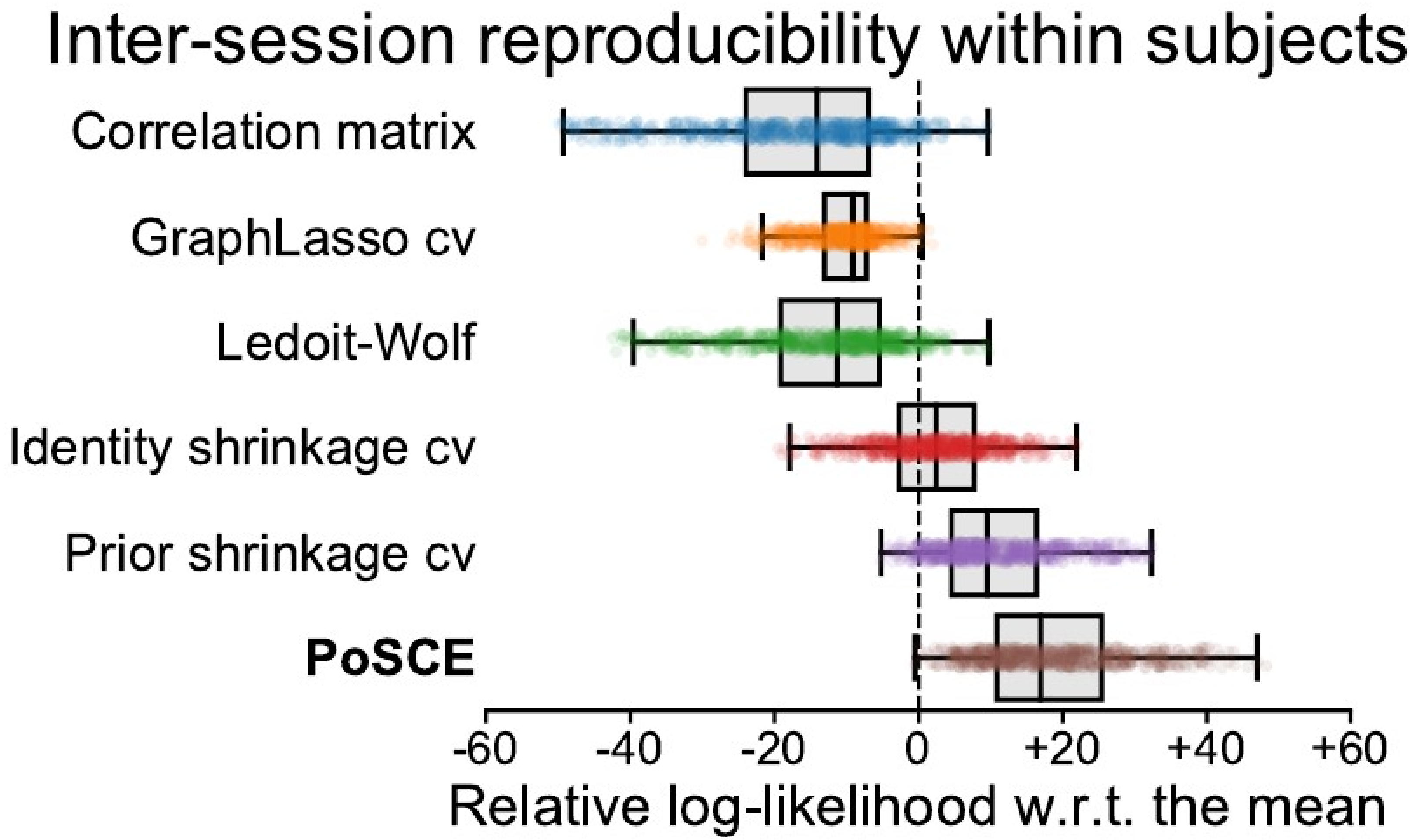
$$\overrightarrow{\overrightarrow{C}}^{-1} = \overrightarrow{\overrightarrow{\mathbf{\Lambda}}}^{-1} + \overrightarrow{\overrightarrow{\mathbf{\Lambda}}}_0^{-1}$$

$$\widehat{\overrightarrow{d\Sigma}} = \left(\overrightarrow{\overrightarrow{\mathbf{\Lambda}}}^{-1} + \overrightarrow{\overrightarrow{\mathbf{\Lambda}}}_0^{-1}\right)^{-1} \overrightarrow{\overrightarrow{\mathbf{\Lambda}}}^{-1} \overrightarrow{dS}$$

Back to covariance matrices $\hat{\mathbf{\Sigma}}_{\mathrm{PoSCE}} = \mathbf{\Sigma}_0^{\frac{1}{2}} \,\mathrm{expm}\left(\widehat{\mathbf{d\Sigma}}\right) \mathbf{\Sigma}_0^{\frac{1}{2}}$

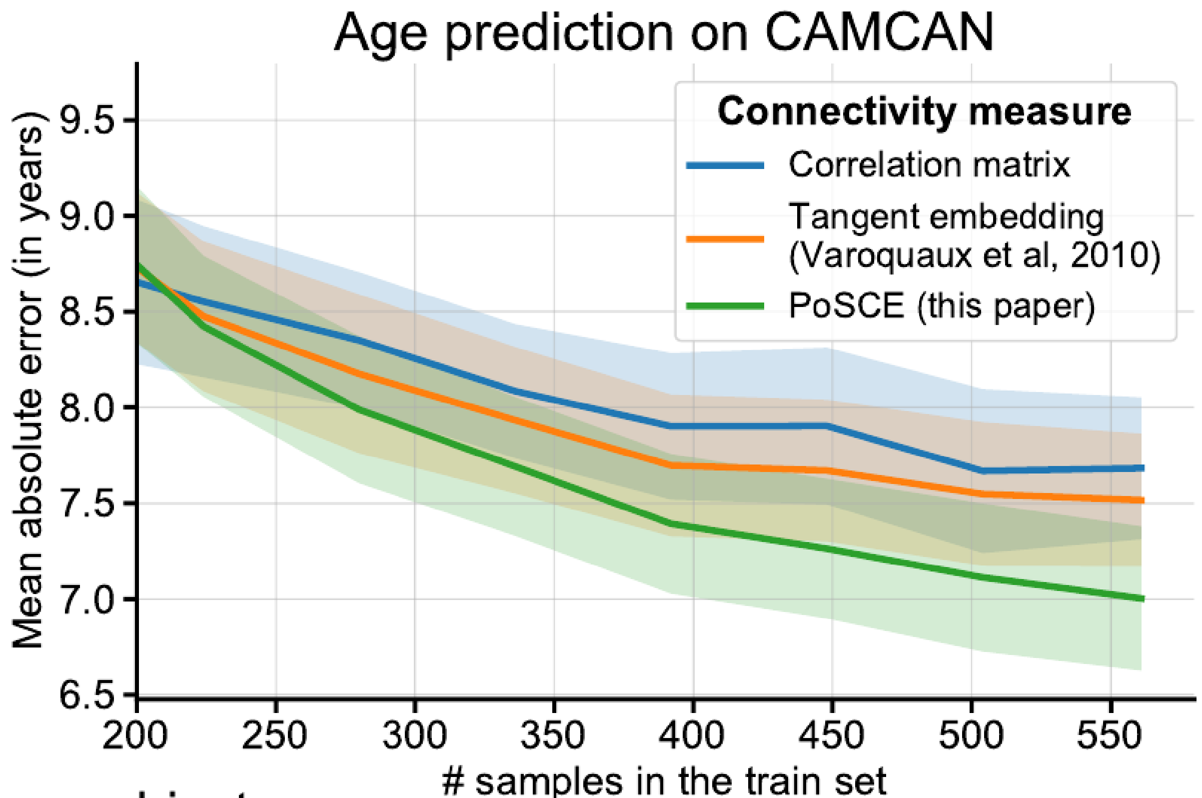
Applications of POSCE

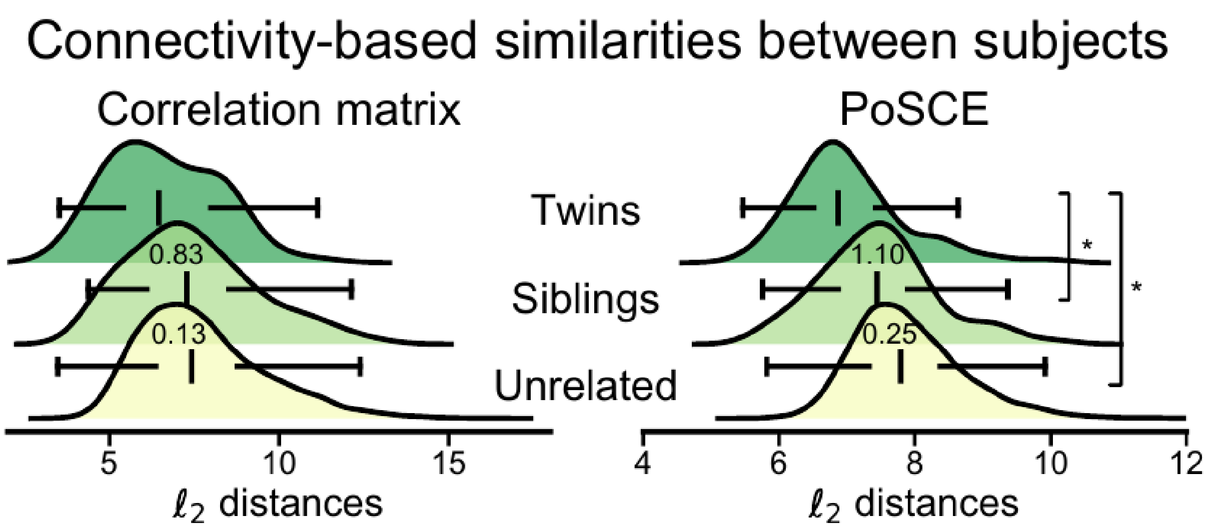
Repeated measures experiment: POSCE better generalizes to data from same subject



Applications of POSCE

Age prediction experiment: regularization improves estimation

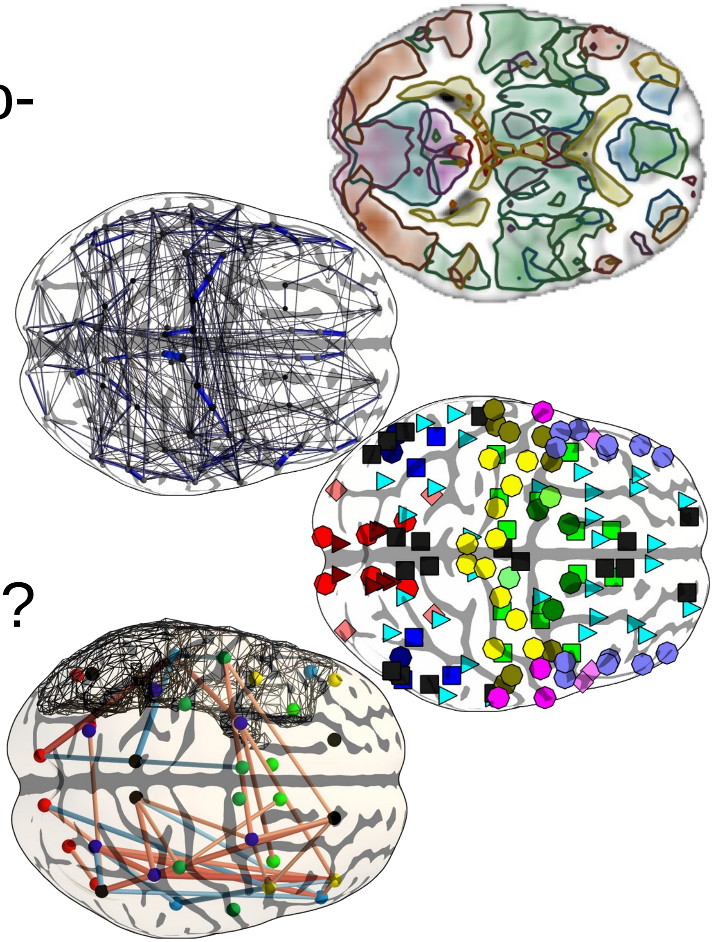




Genetic similarity better represented by POSCE estimator

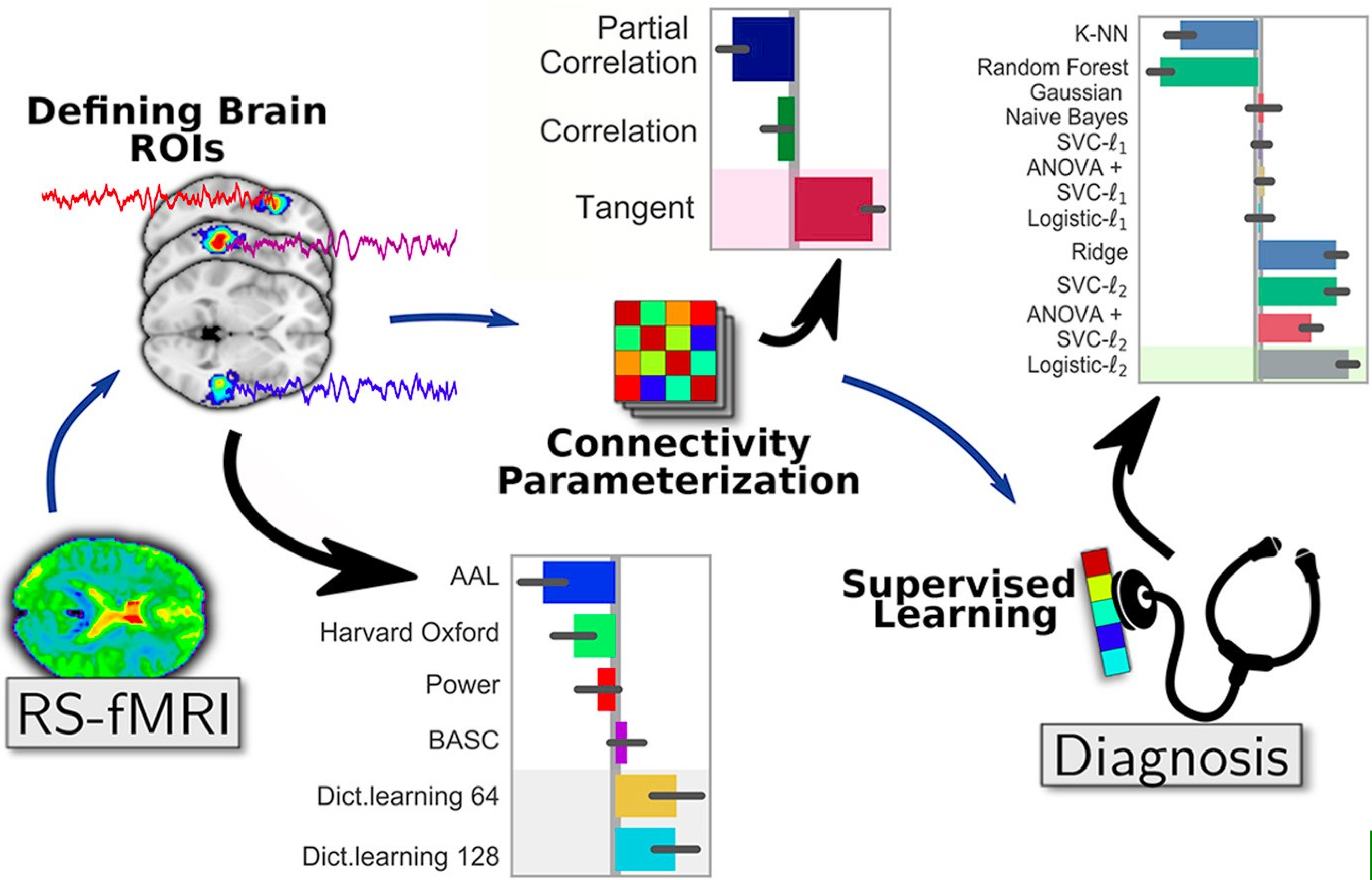
The covariance structure of fMRI signals

- Deriving brain atlases from functional co-activations
- Graph of functional interactions
- Intrinsic structures in brain connectivity
- Individualized diagnostic markers
- Towards a functional-anatomical model ?

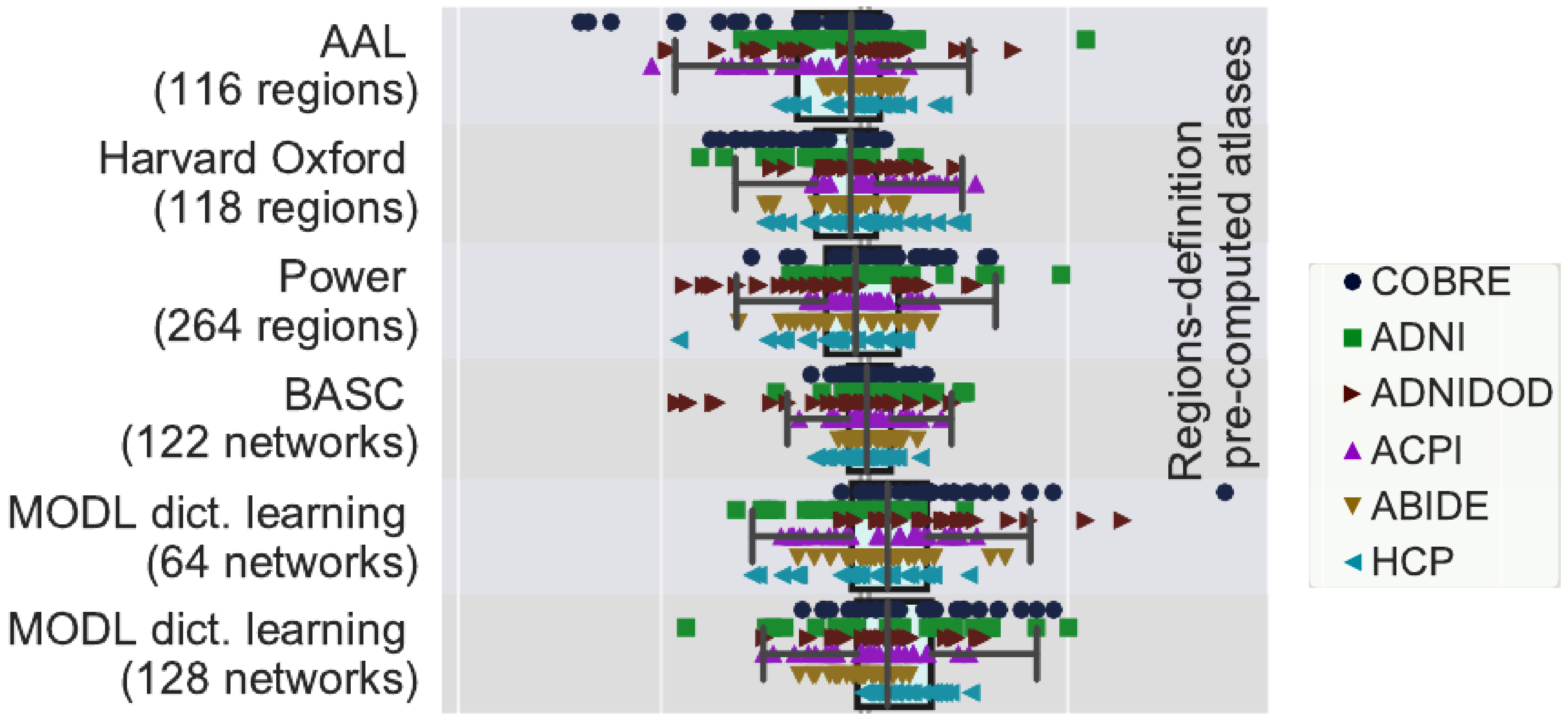


Weak SNR, weak specificity of resting-state signals

What matters most for diagnosis accuracy?



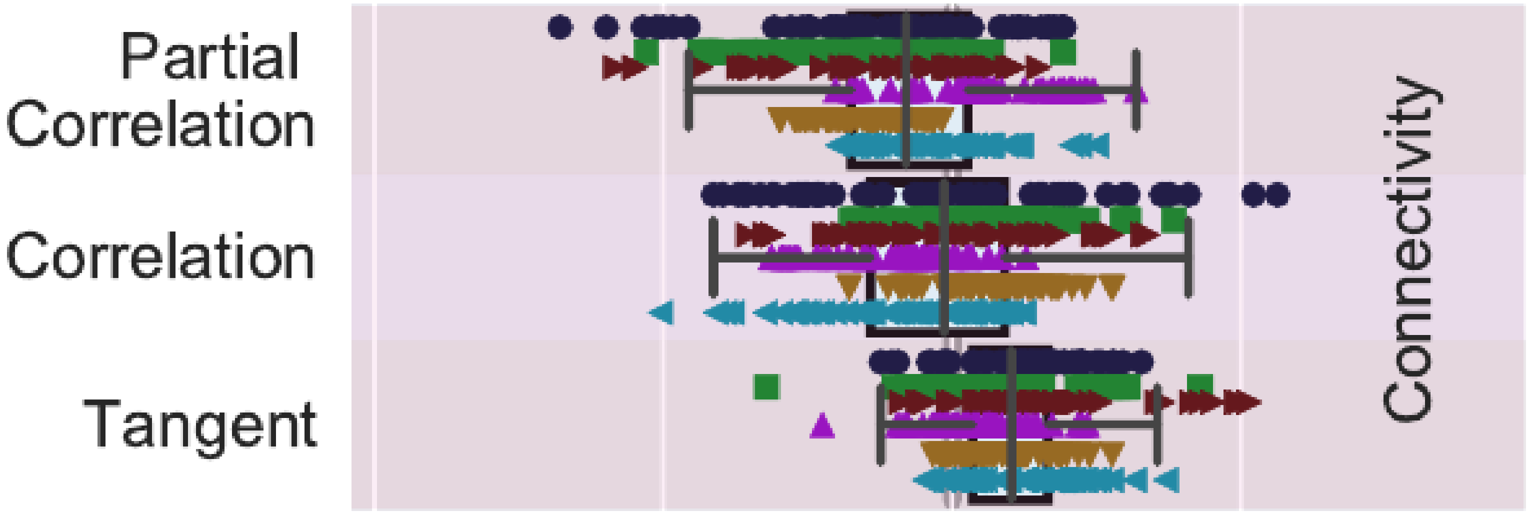
The best brain atlas



Across 6 datasets & cross-validation folds

[Dadi et al. Nimg 2019]

The best connectivity model

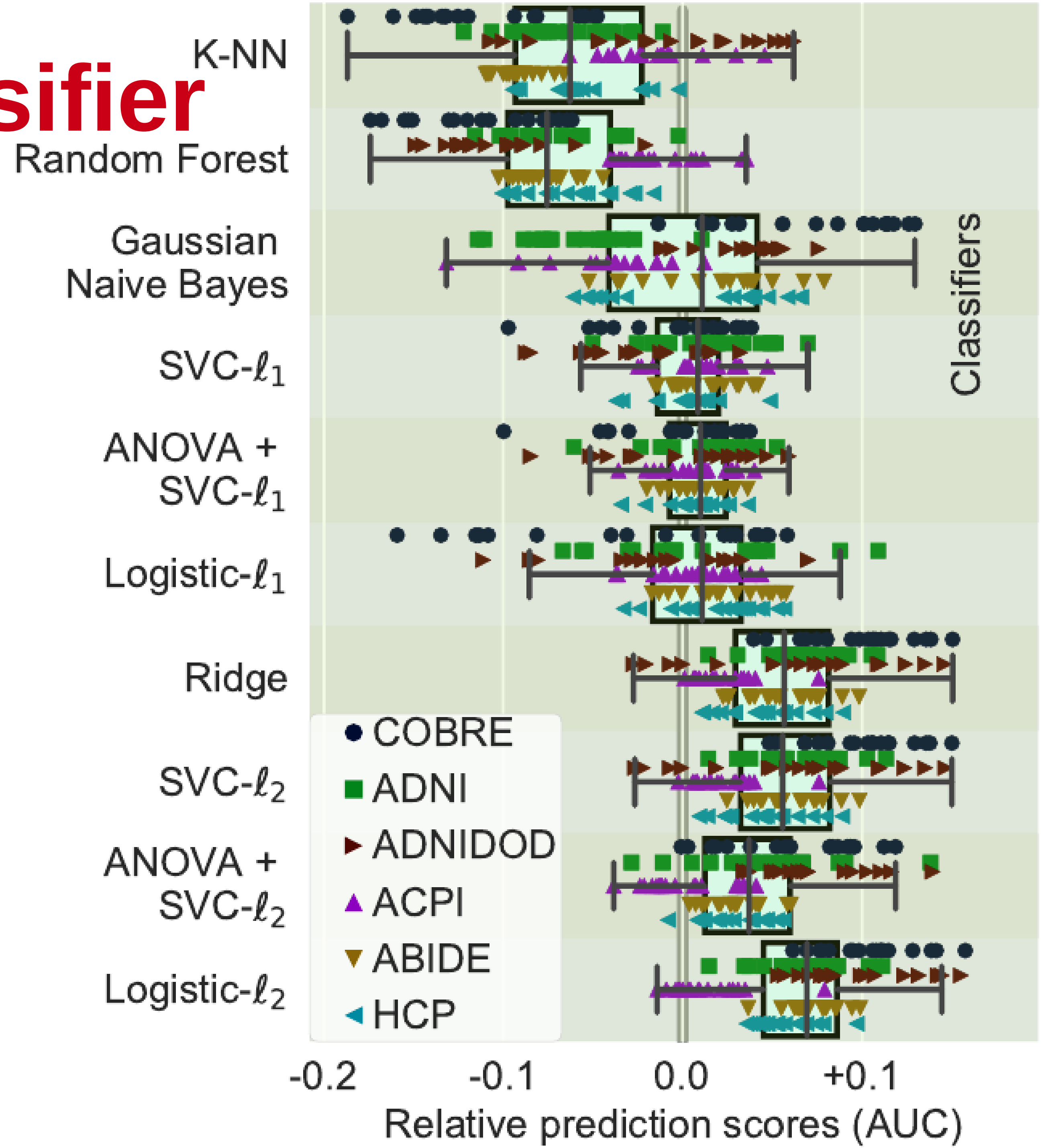


Across 6 datasets & cross-validation folds

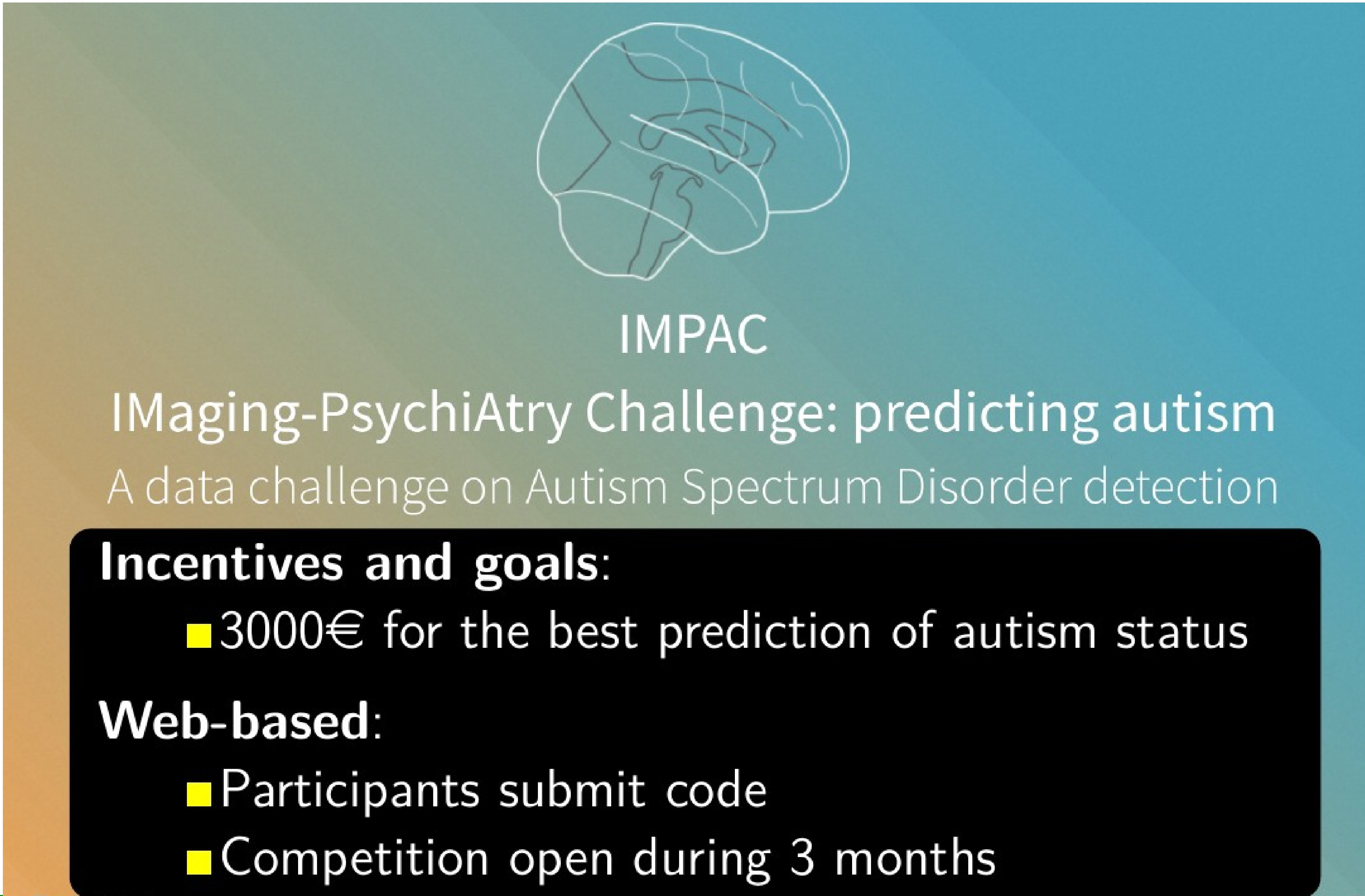
Tangent outperforms alternatives

[Dadi et al. Nimg 2019]

The best classifier

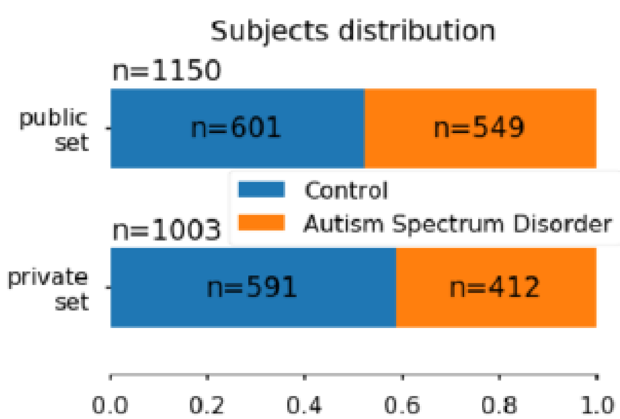


IMPAC

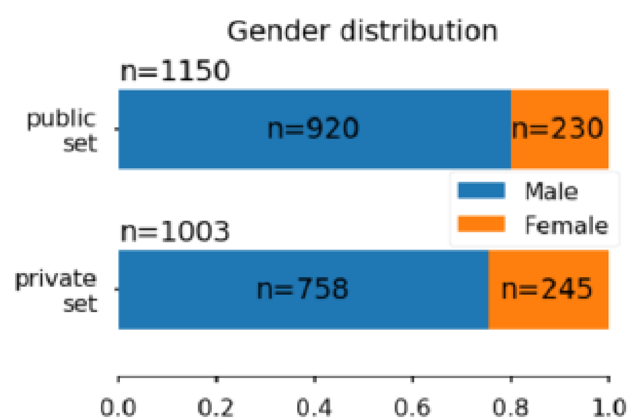

IMPAC
IMaging-PsychiAtry Challenge: predicting autism
A data challenge on Autism Spectrum Disorder detection
Incentives and goals:
3000€ for the best prediction of autism status
Web-based:
Participants submit code
Competition open during 3 months

Blind assessment of biomarkers

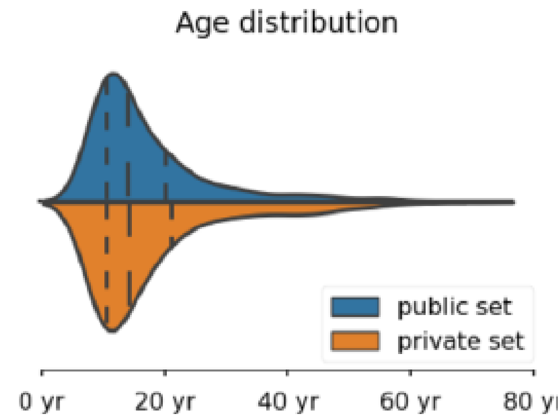
Patient vs Control distribution



Gender distribution



Age distribution



Hidden test set:

Participants never see the private set

Private-set prediction scores are published at the end

Learning curve