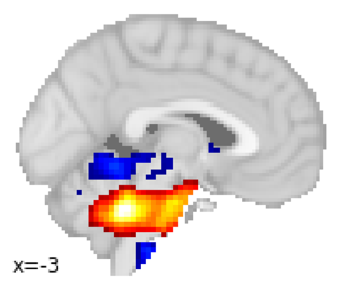
Compressed Online Dictionary Learning for Fast Resting-State fMRI Decomposition #4103

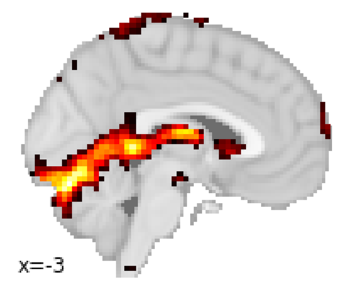
Arthur Mensch(1) Julien Mairal(2) Gaël Varoquaux(1) Bertrand Thirion(1)

(1)Parietal team, Inria, CEA, Neurospin, Paris-Saclay University. Gif-sur-Yvette, France
(2)Thoth team, Inria. Grenoble, France

Low-rank sparse decomposition methods for rfMRI



ICA + thresholding



Dictionary Learning

- **Dictionary learning** is a good alternative to ICA
- Efficient online algorithm
- Samples are *time-series* for each voxels

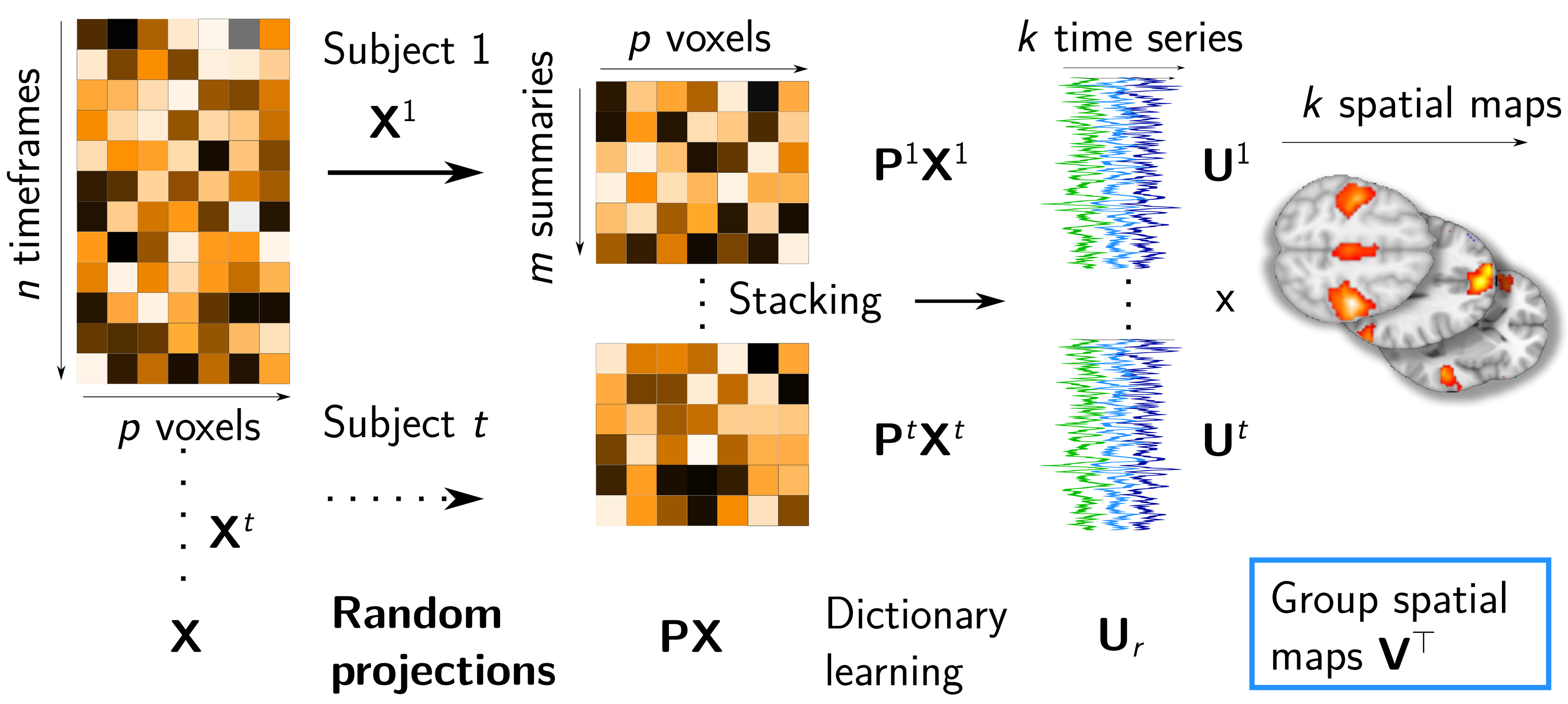
Dictionary learning for resting-state fMRI

- Set of t fMRI resting-state sequences $\mathbf{X}^0, \dots, \mathbf{X}^{t-1}$
- Find k sparse spatial components capturing the data variance
- Dictionary learning formulation

$$\min_{\substack{\mathbf{V}\in\mathbb{R}^{p\times k}\\ \mathbf{U}\in\mathbb{R}^{nt\times k}, \|\mathbf{U}_j\|_2\leq 1}} \|\mathbf{X} - \mathbf{U}\,\mathbf{V}^\top\|_F^2 + \lambda\,\|\mathbf{V}\|_1$$

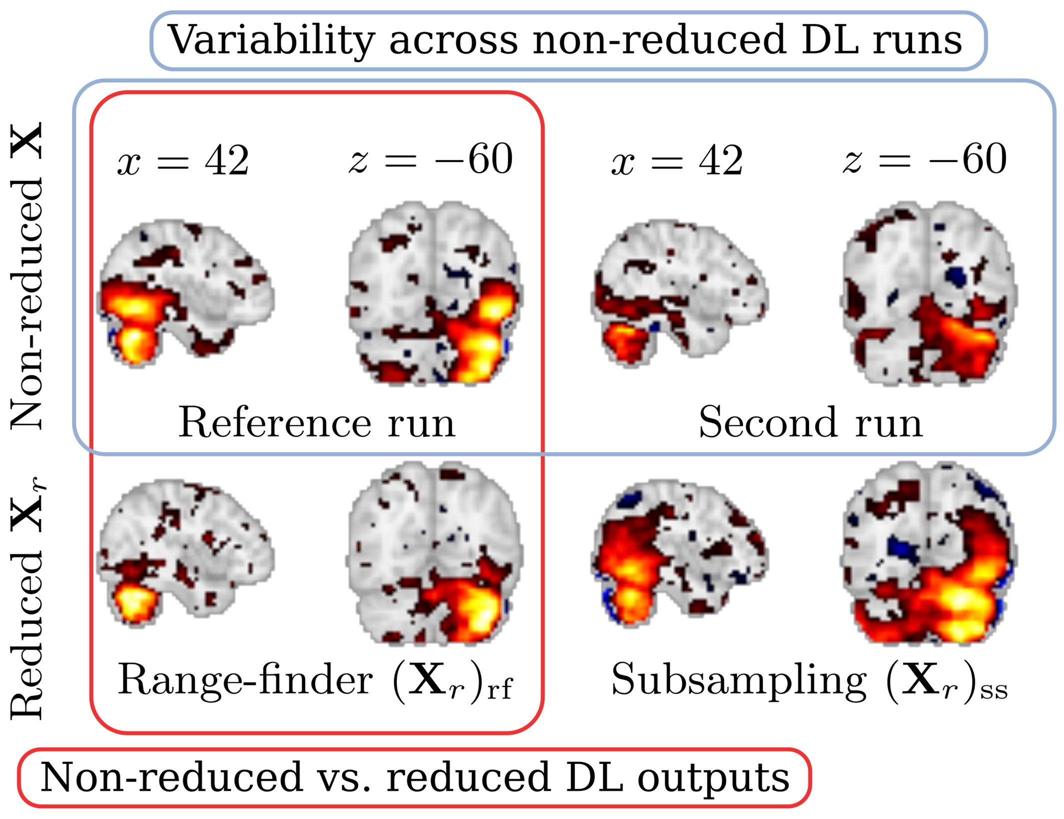
How to apply dictionary learning on large datasets ?

Hierarchical dimension reduction

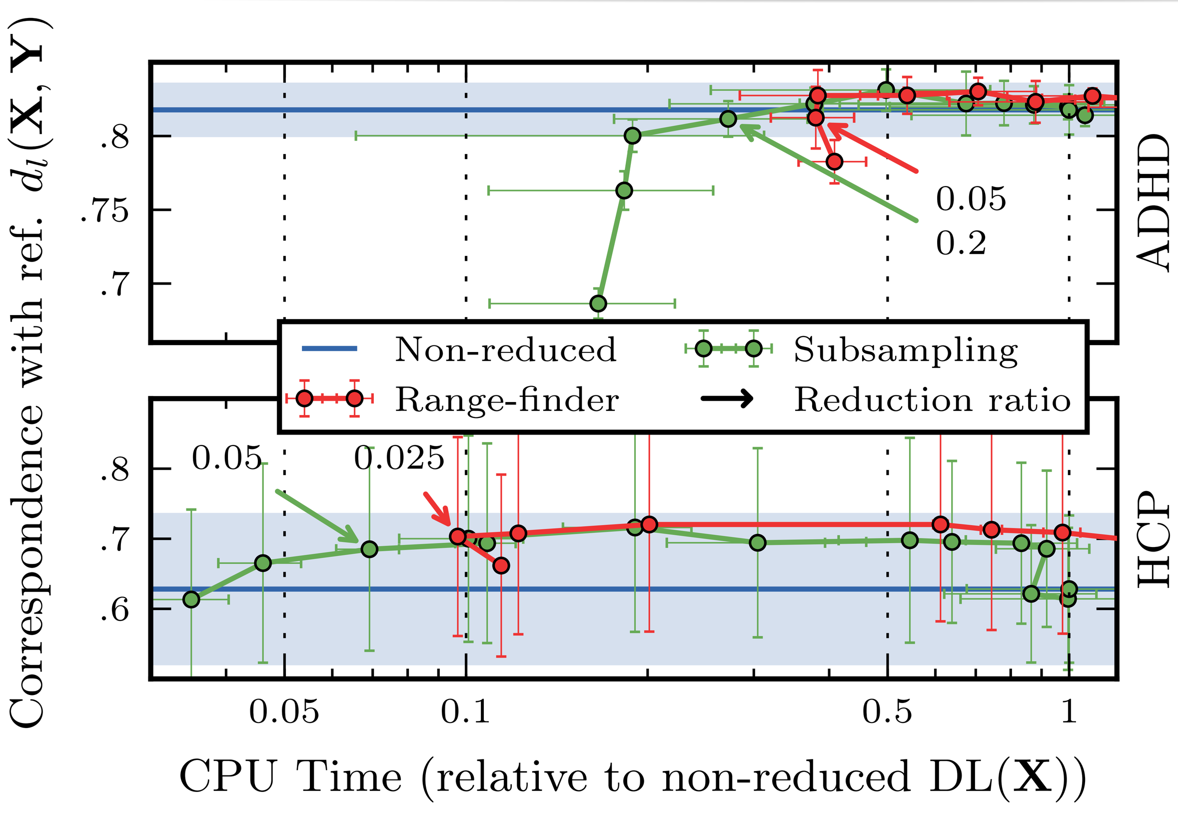


Spatial decomposition quality

- Reduce $n \to m$ by dimension reduction
- Per subject reduction:
  $\mathbf{X}^s_r \in \mathbb{R}^{m\times p} = \mathbf{P}^s\mathbf{X}^s$
- Concatenate
  $\mathbf{X} \in \mathbb{R}^{mt\times p} = \mathbf{PX}$
- **Qualitative validation**
  Quality of spatial maps
  → comparable to simple stacked dictionary learning



Measuring speed-up



- **Validation:** Correlation between obtained maps and simple dictionary learning output
- Repeated runs (stochastic algorithm) to capture most *obtainable* maps

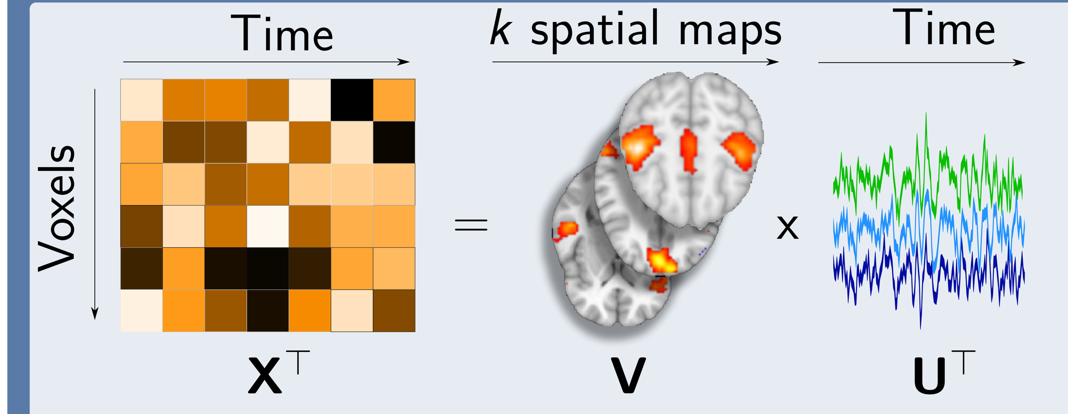
Similar output quality

↑increasing reduction ×40
↓reducing CPU time ×10

In practice: 75 record subset of HCP (150GB)

- **10x speed-up** in time & memory usage
- **Non-significant accuracy loss**
- Runs in two hours
- Easy to grasp, easy to use
- Packaged on `nilearn.github.io` (poster #1876)

Changing paradigm: streaming 3D records



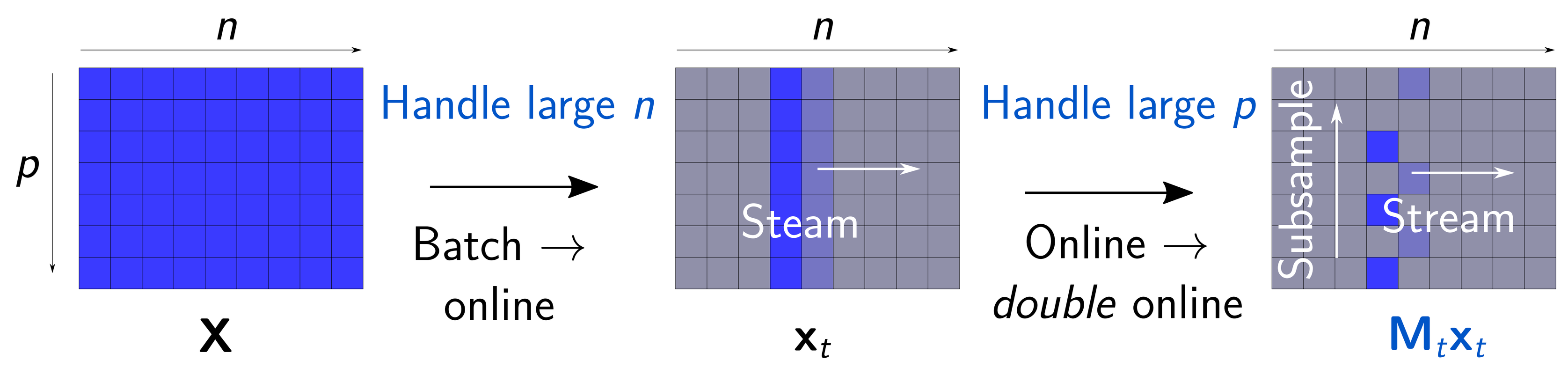
$$\min_{\substack{\mathbf{V}\in\mathbb{R}^{p\times k}\\ \mathbf{U}\in\mathbb{R}^{nt\times k}, \|\mathbf{U}_j\|_1\leq 1}} \|\mathbf{X}^\top - \mathbf{V}\,\mathbf{U}^\top\|_F^2 + \lambda\,\|\mathbf{U}\|_2$$

- **2 TB** dataset (HCP)
- Cannot be loaded in memory
- **Solution**: stream 3D records
- Samples are *spatial maps* for each time/records
- ℓ_1 projection → sparse maps
- ℓ_2 penalty → dense loading

Accelerating online dictionary learning algorithm

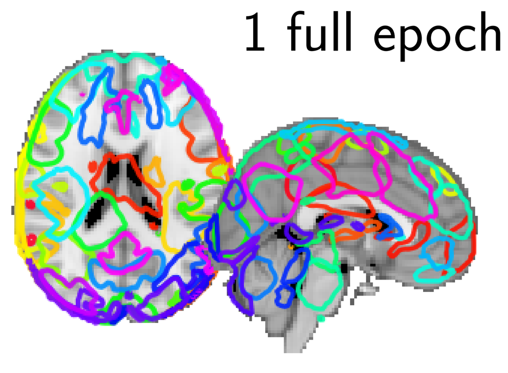
- **Scalable algorithm**: maps can be learned by loading single maps in memory
- But slow : **1 week and a half required to process the full HCP dataset**

We use random subsampling to accelerate online dictionary learning



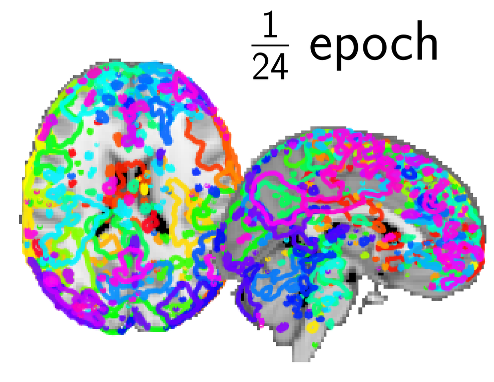
Baseline online algorithm

1 full epoch



235 h run time

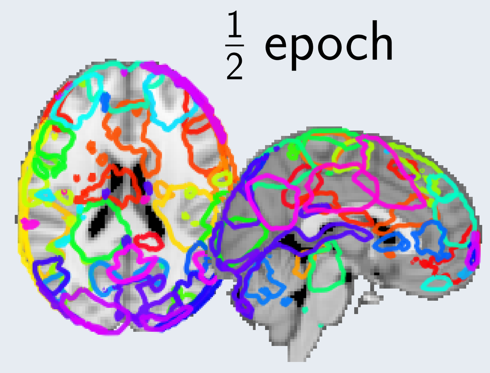
$\frac{1}{24}$ epoch



10 h run time

Reduction $r = 12$

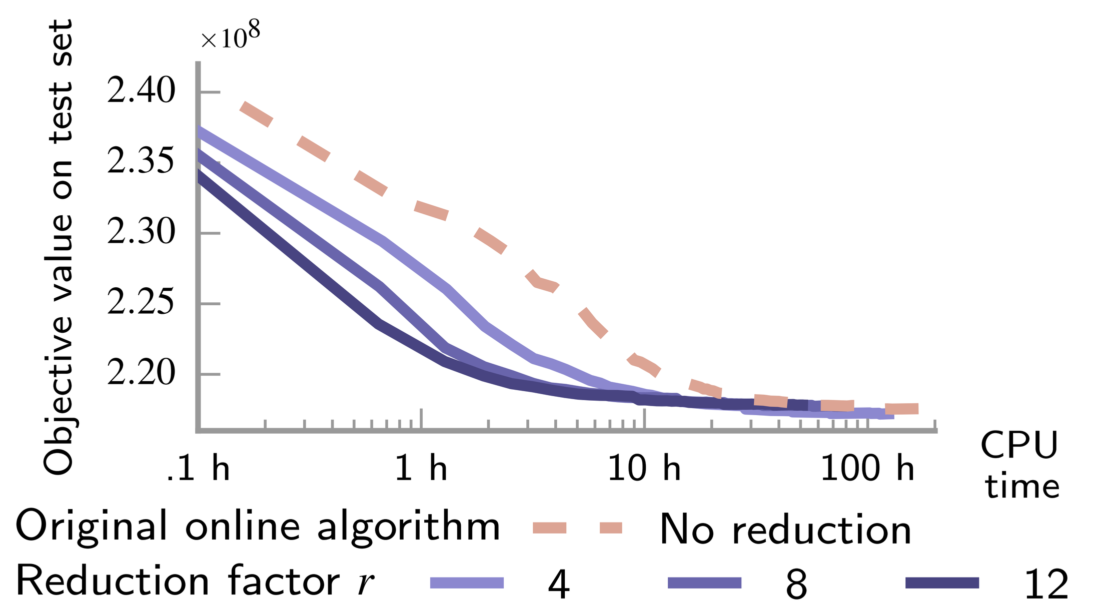
$\frac{1}{2}$ epoch



10 h run time

Well defined spatial maps (noiseless blobs) are obtained 10× faster

Measuring speed-ups



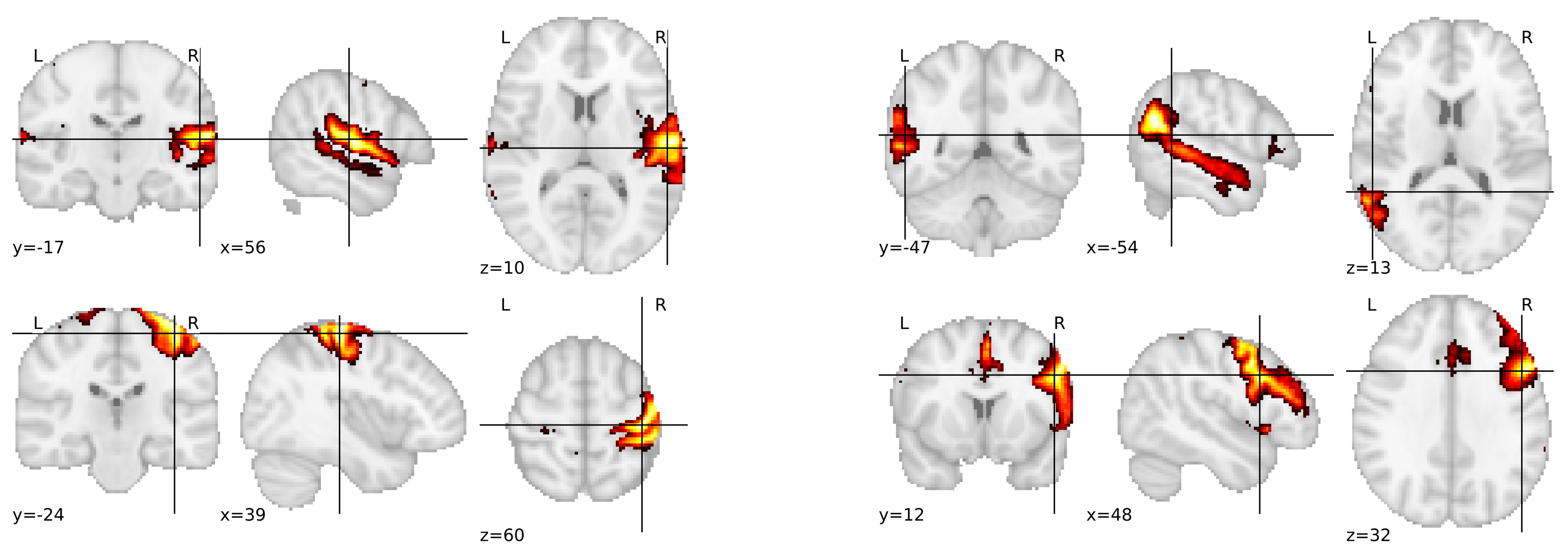
- Explained variance on test set vs. non subsampled algorithm
- Are obtained maps as sparse as original algorithm ?

Results

- Speed-up ~ reduction factor
- Comparable performance with/without subsampling

In practice: 2000 records of HCP (2TB)

- **Full epoch: 20h**. Half is enough
- **4 GB** RAM needed
- Highly detailed maps (see below)
- `github.com/arthurmensch/modl`



[1] A. Mensch, J. Mairal, B. Thirion, and G. Varoquaux. Dictionary Learning for Massive Matrix Factorization. *Proceedings of The 33rd International Conference on Machine Learning*, pages 1737–1746, 2016.

[2] A. Mensch, G. Varoquaux, and B. Thirion. Compressed Online Dictionary Learning for Fast fMRI Decomposition. *International Symposium on Biomedical Imaging*, 2016.