Statistical learning and optimization for functional MRI data mining

Alexandre Gramfort
alexandre.gramfort@telecom-paristech.fr

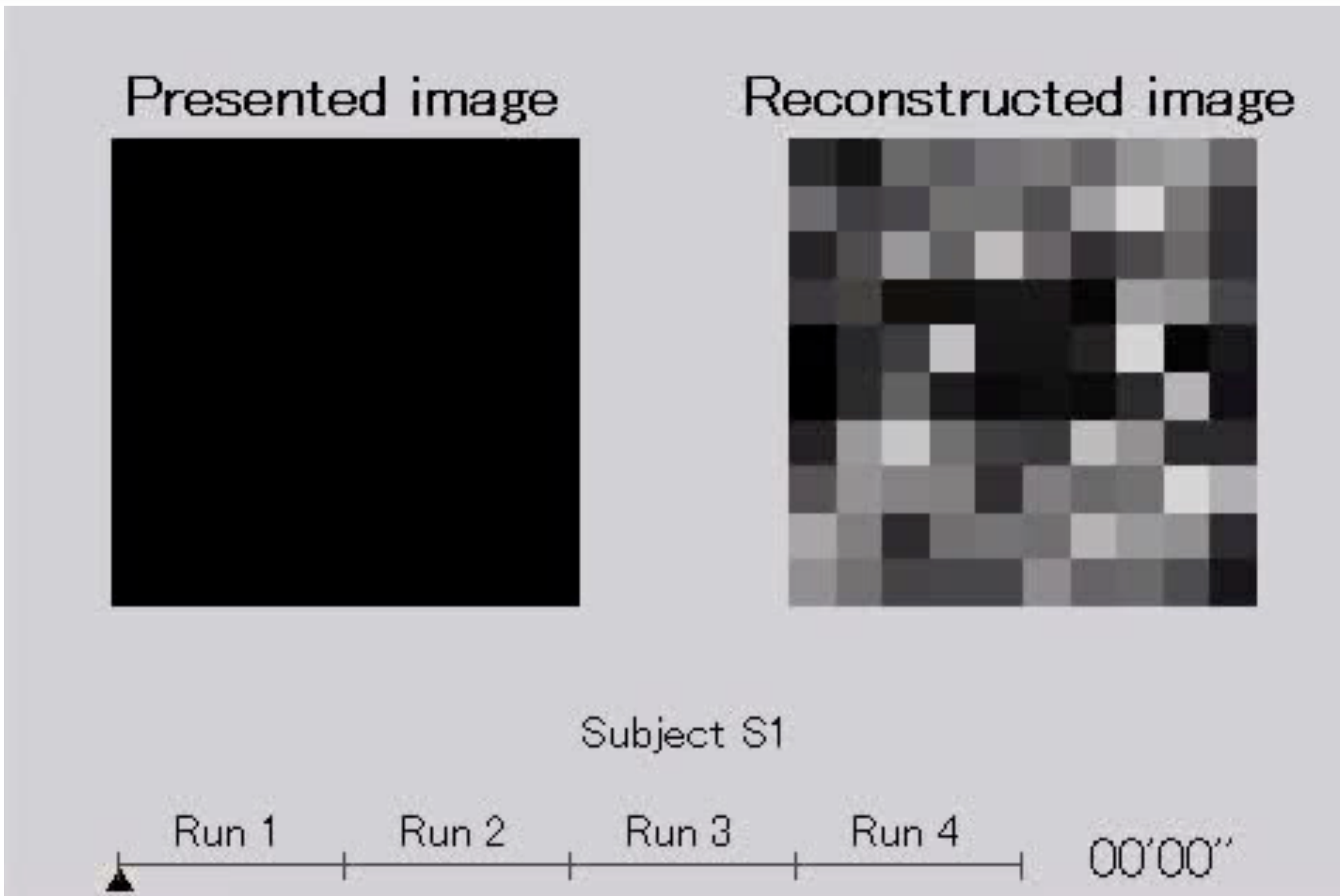
Assistant Professor
LTCI, Télécom ParisTech, Université Paris-Saclay

Macaron Workshop - INRIA Grenoble - 2017

Neuron
Article

Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders

Yoichi Miyawaki,[1,2,6] Hajime Uchida,[2,3,6] Okito Yamashita,[2] Masa-aki Sato,[2] Yusuke Morito,[4,5] Hiroki C. Tanabe,[4,5] Norihiro Sadato,[4,5] and Yukiyasu Kamitani[2,3,*]



http://www.youtube.com/watch?v=h1Gu1YSoDaY

Current Biology 21, 1641–1646, October 11, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.08.031

Report

Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies

Shinji Nishimoto,[1] An T. Vu,[2] Thomas Naselaris,[1] Yuval Benjamini,[3] Bin Yu,[3] and Jack L. Gallant[1,2,4,*]

mental processes. It has therefore been assumed that fMRI data would not be useful for modeling brain activity evoked

Presented clip

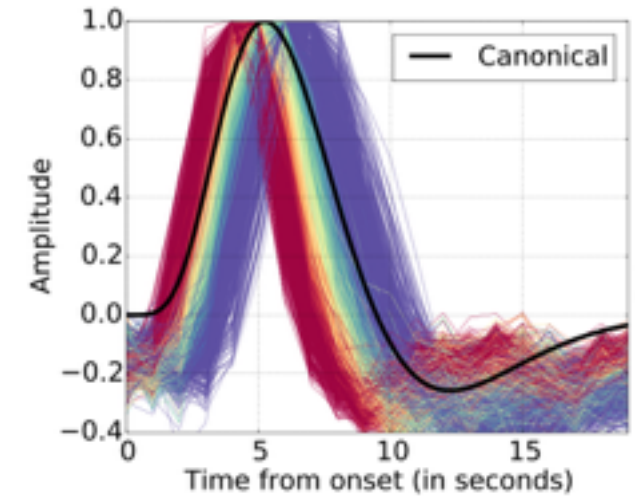
Clip reconstructed from brain activity

http://www.youtube.com/watch?v=nsjDnYxJ0bo

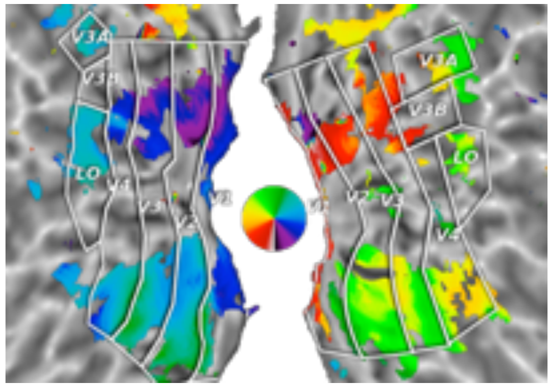
Outline

- Background

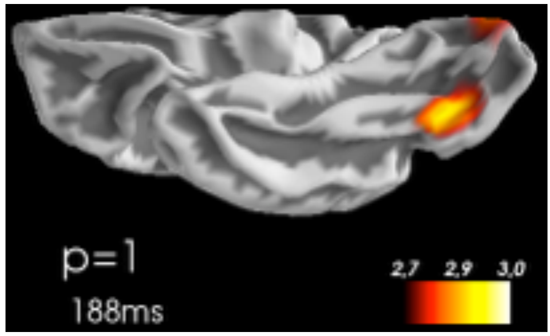
- Estimating the hemodynamic response function [Pedregosa et al. Neuroimage 2015]



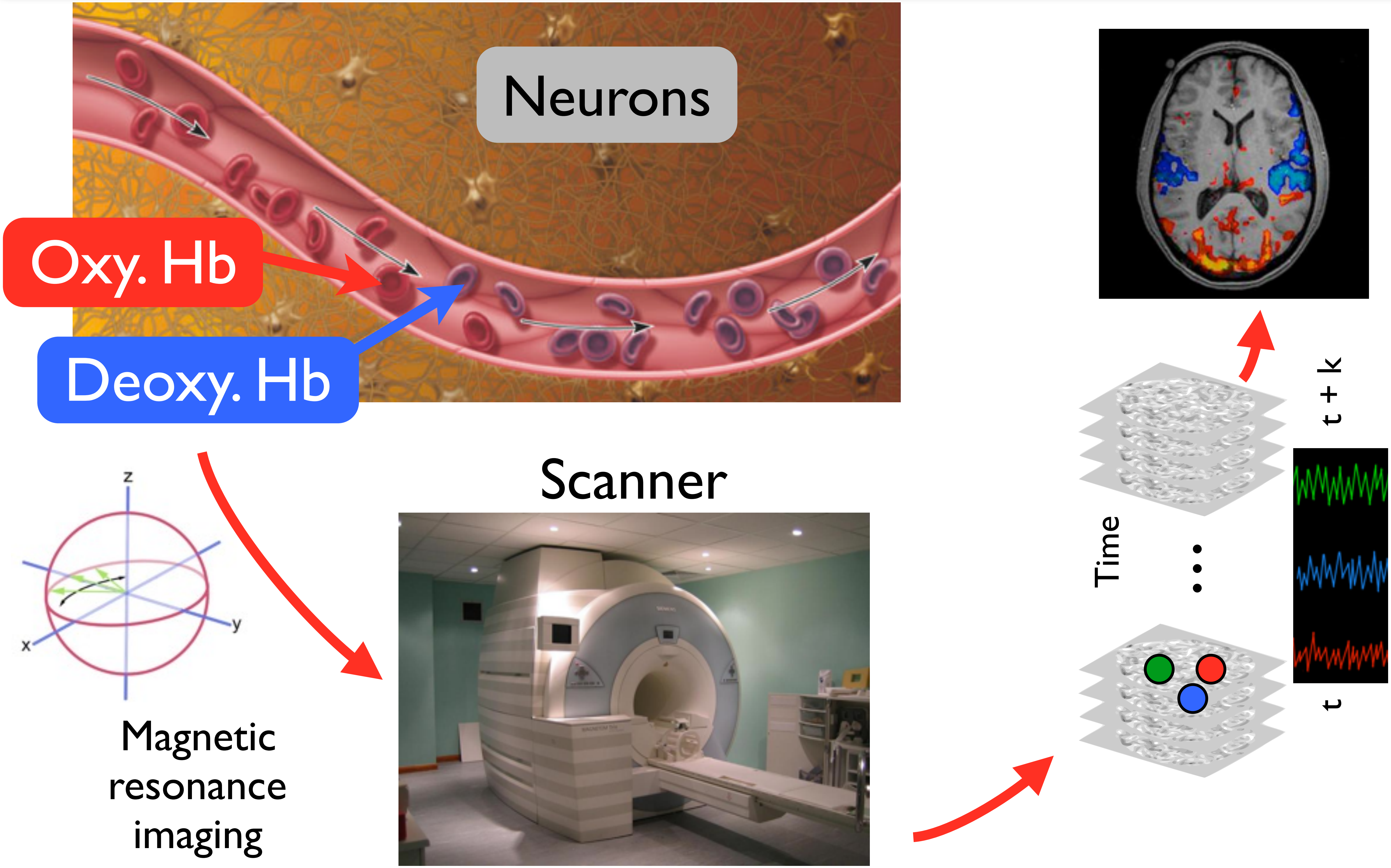
- Mapping the visual pathways with computational models and fMRI [Eickenberg et al. Neuroimage 2016]

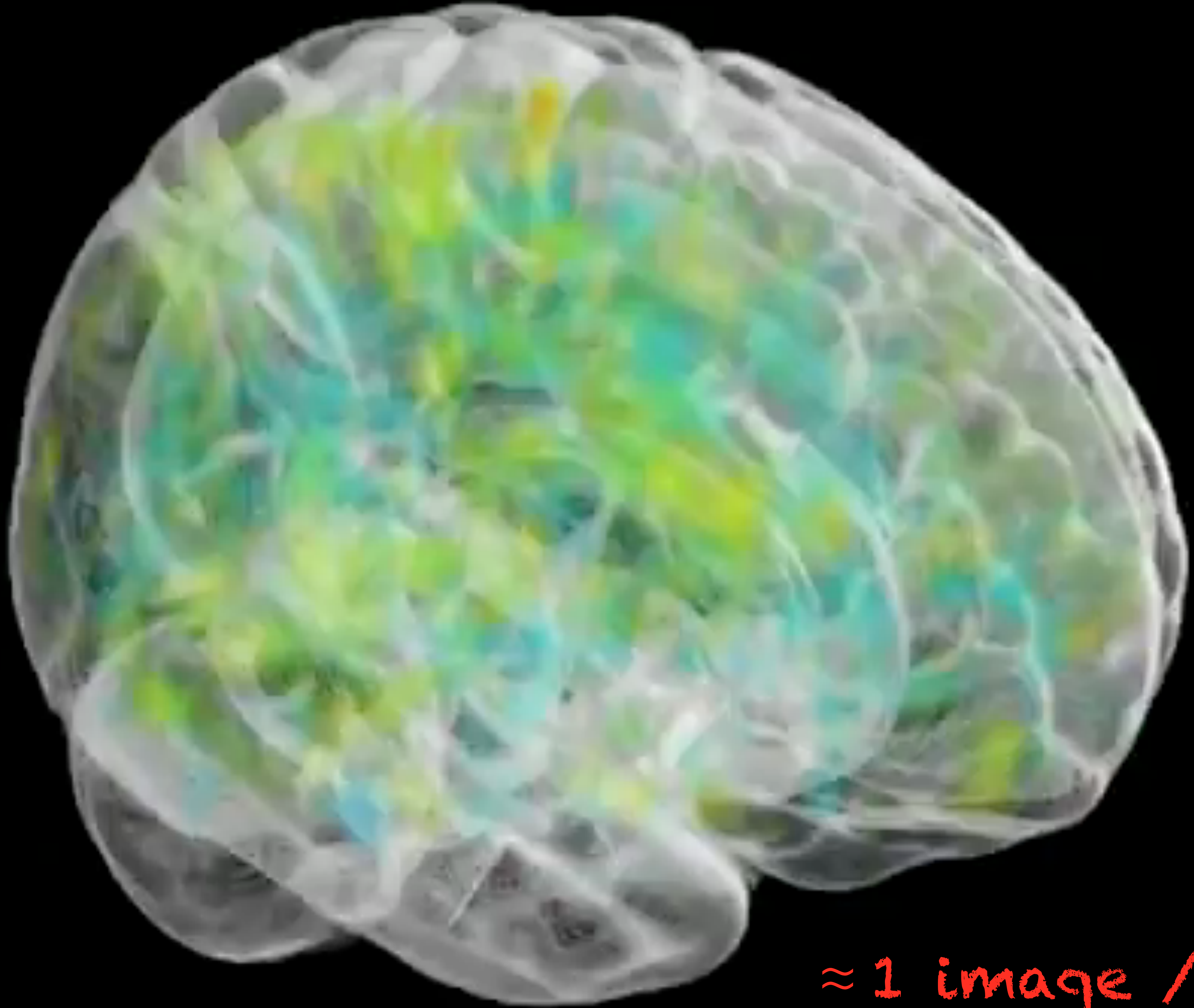


- Optimal transport barycenter for group studies [Gramfort et al. IPMI 2015]



Functional MRI



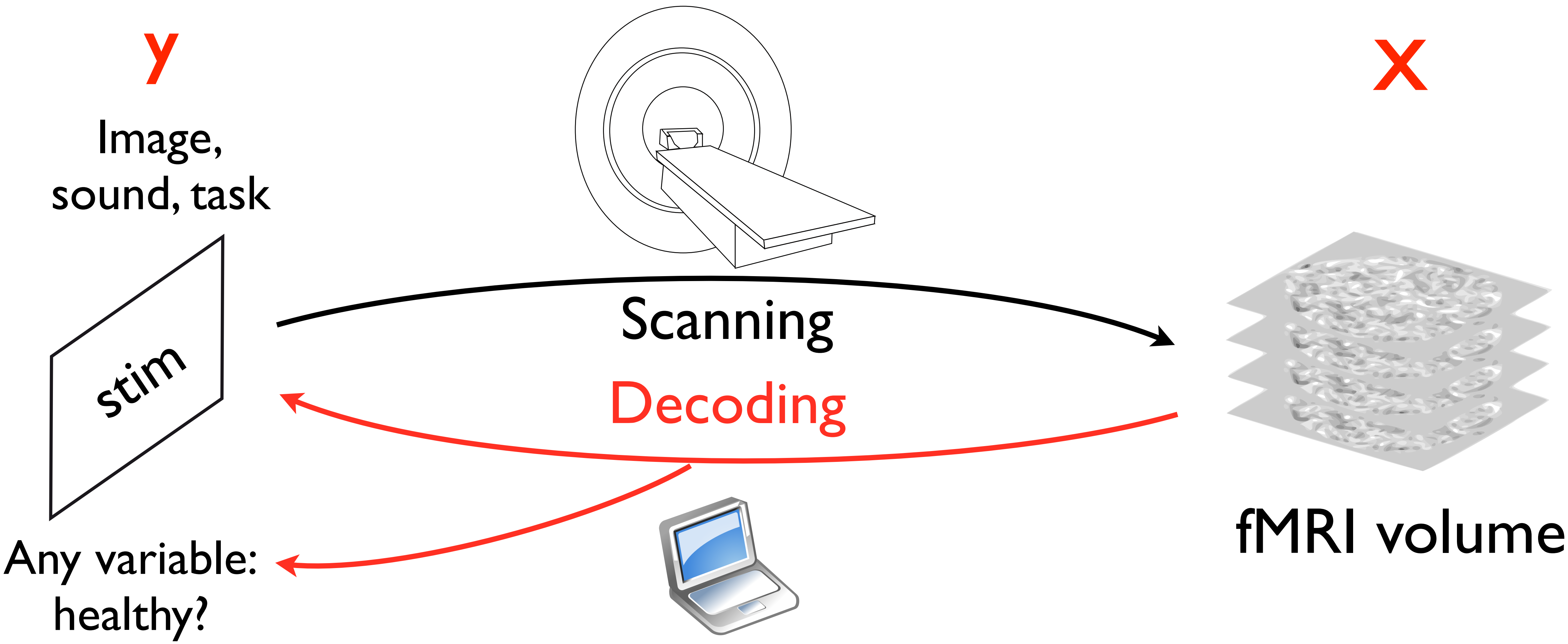


≈ 1 image / 2s

http://www.youtube.com/watch?v=uhCF-zlk0jY

courtesy of Gael Varoquaux

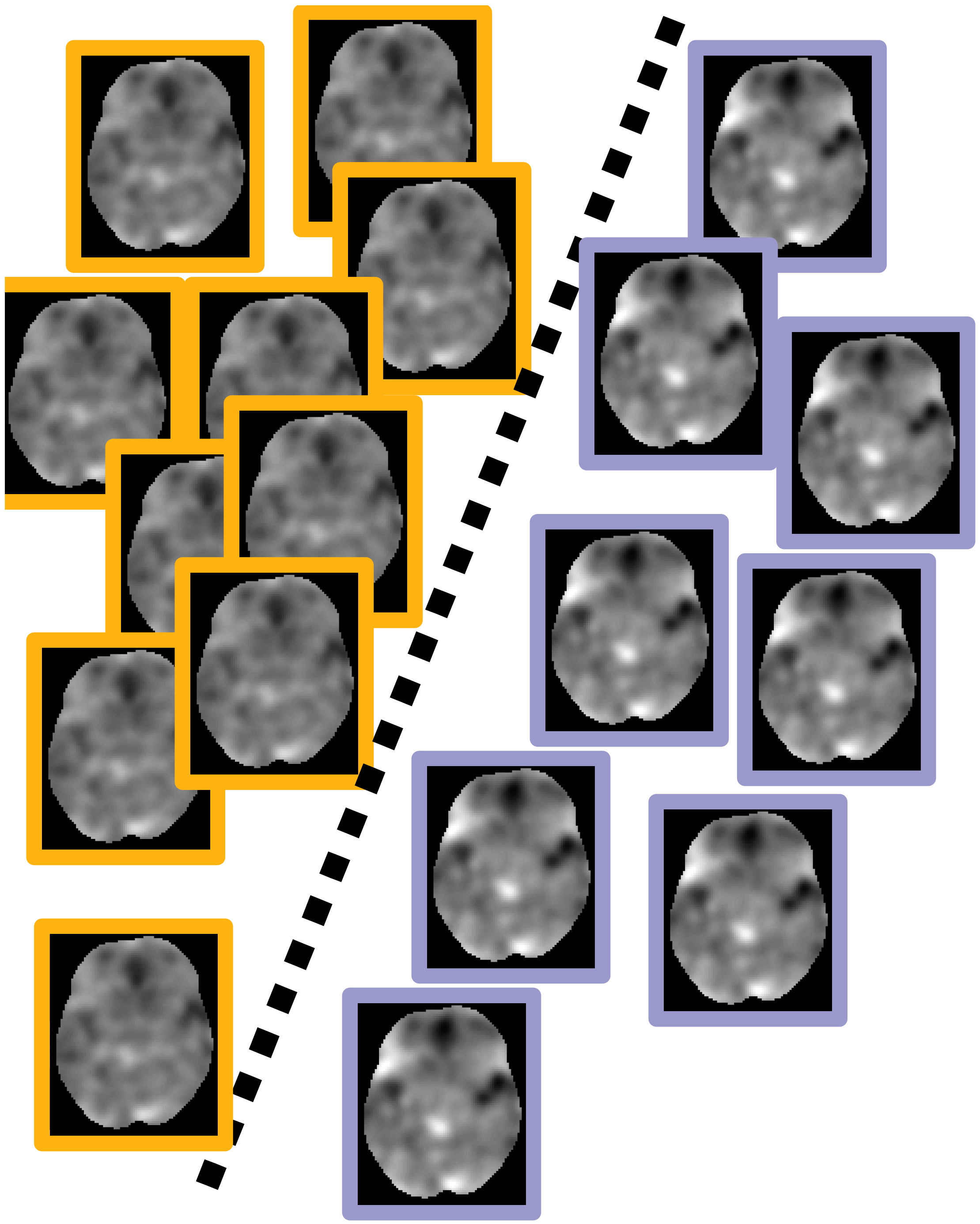
fMRI supervised learning (decoding)



Challenge: Predict a behavioral variable from the fMRI data

Objective: Predict y given X or learn a function f : X -> y

Classification example with fMRI



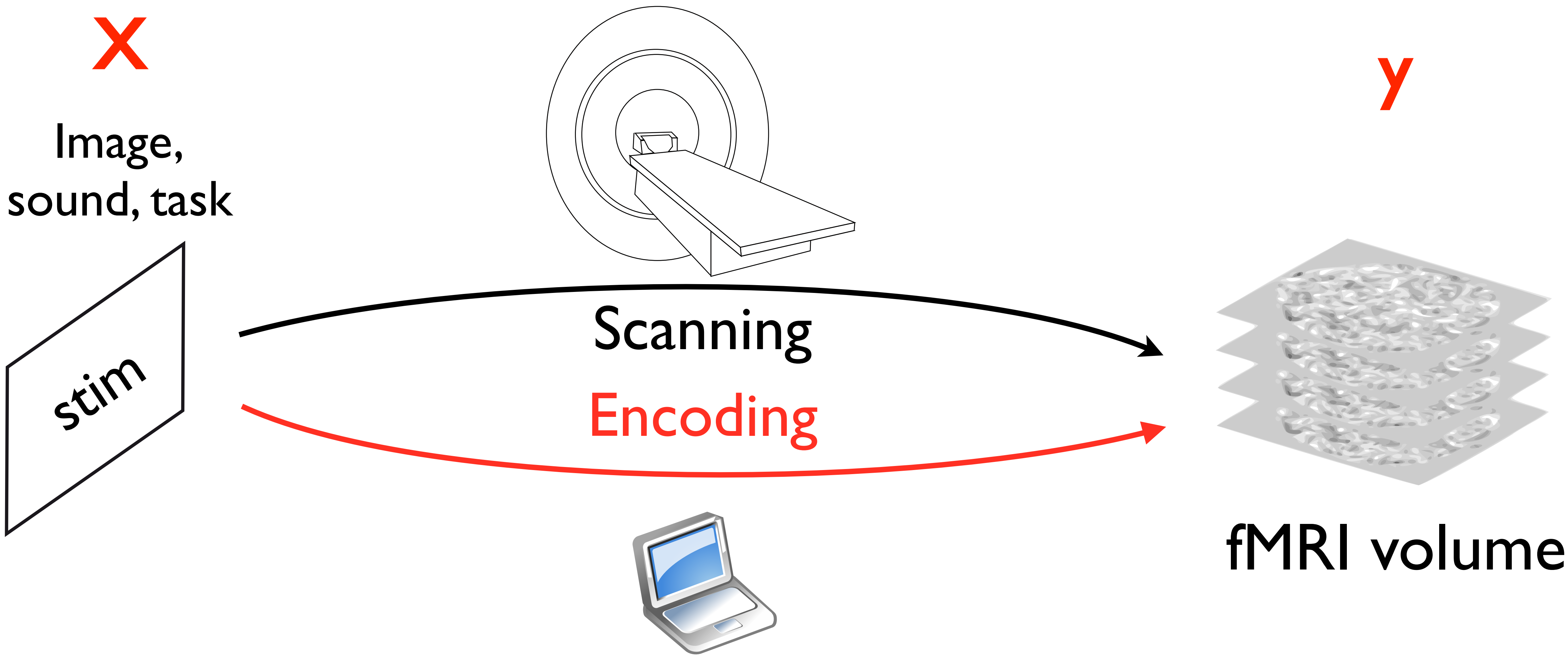
The **objective** is to be able to **predict** or given an fMRI activation map

Patient	vs.	Controls
Faces	vs.	Houses
...	vs.	...
1	vs.	-1

ie. $y = \{-1, 1\}$

objective: Predict $y = \{-1, 1\}$ given $x \in \mathbb{R}^p$

fMRI supervised learning (Encoding)



Challenge: Predict the BOLD response from the stimuli descriptors

Objective: **Predict y given X or learn a function f : X -> y**

[Thirion et al. 06, Kay et al. 08, Naselaris et al. 11, Nishimoto et al. 2011, Schoenmakers et al. 13 ...]

Learning the hemodynamic response function (HRF) for encoding and decoding models

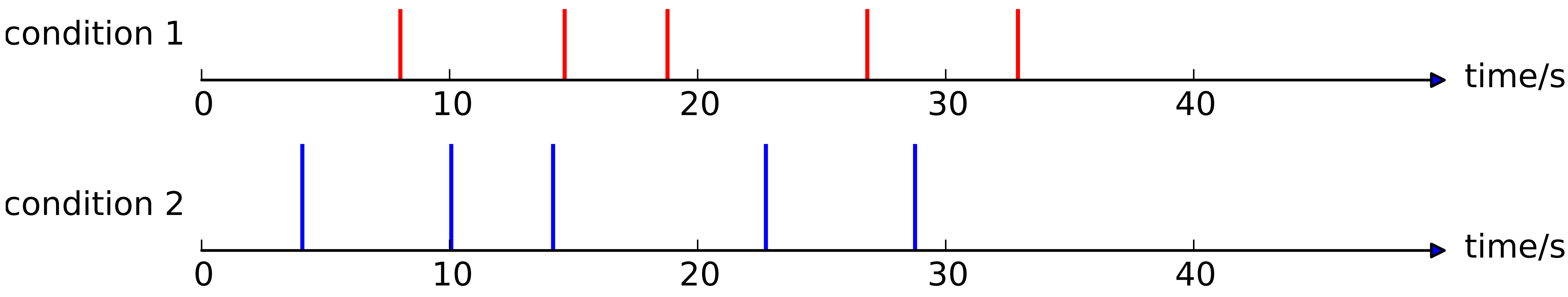
thanks to Fabian Pedregosa Michael Eickenberg

Data-driven HRF estimation for encoding and decoding models, Fabian Pedregosa, Michael Eickenberg, Philippe Ciuciu, Bertrand Thirion and Alexandre Gramfort, Neuroimage 2015

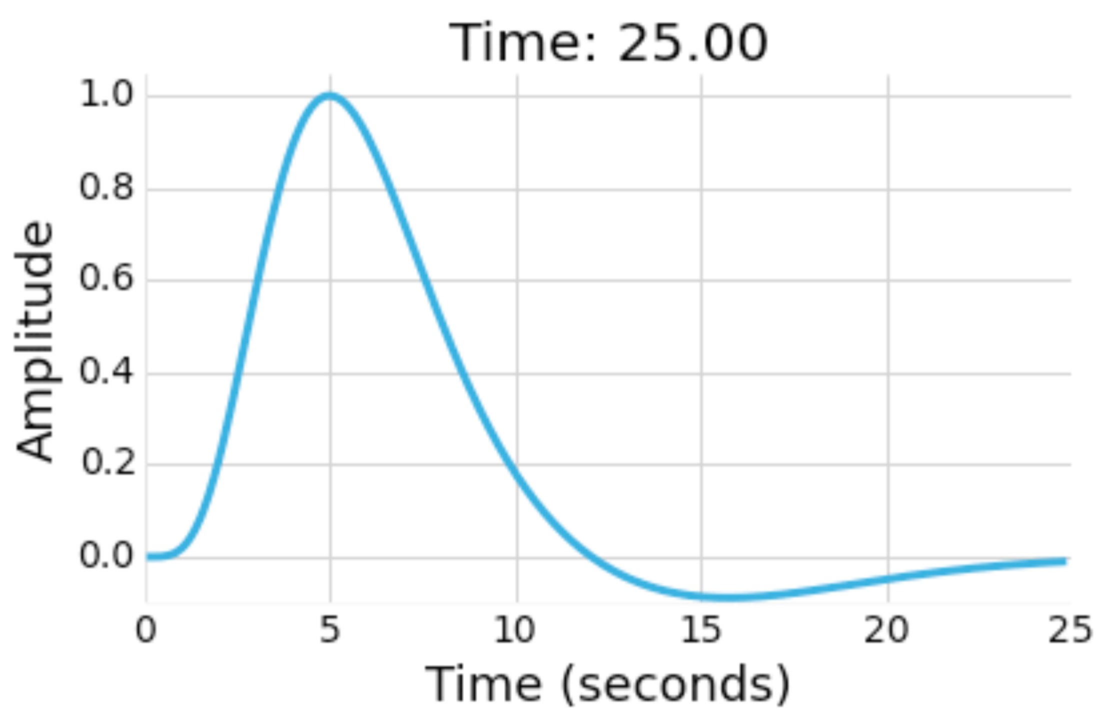
PDF: https://hal.inria.fr/hal-00952554/en

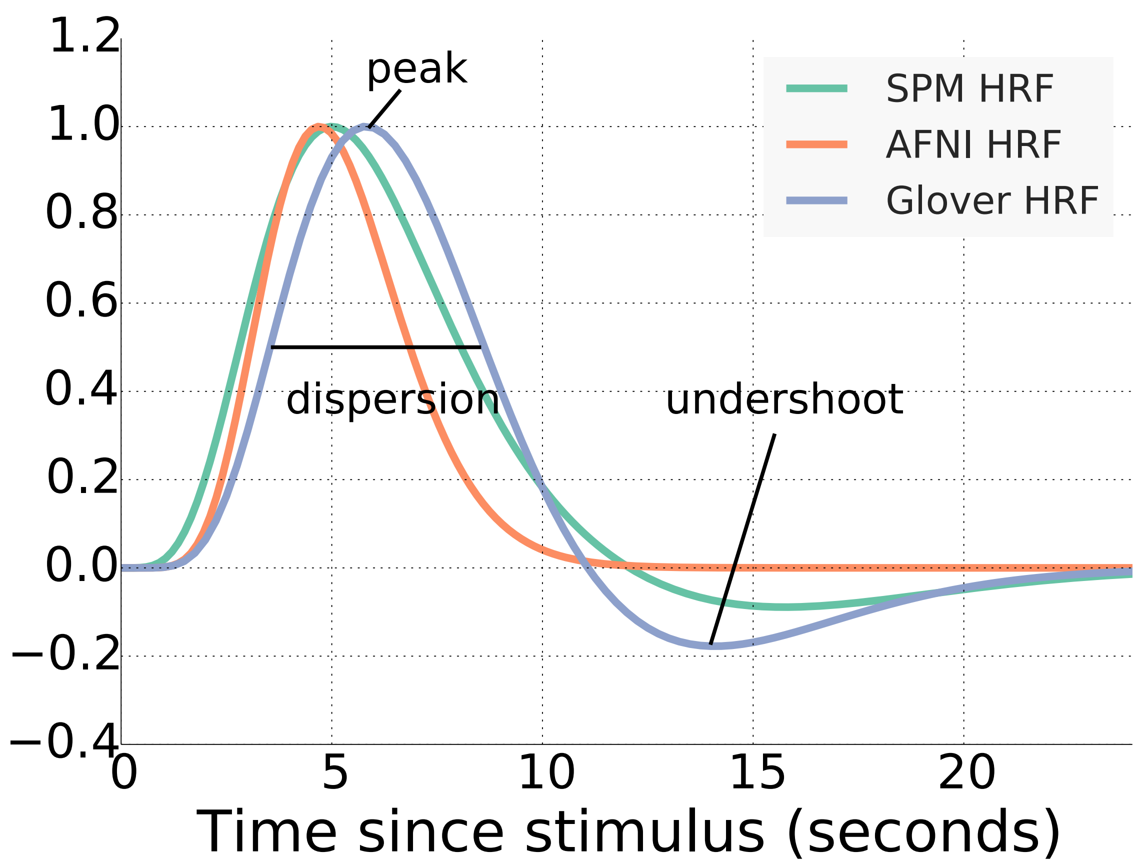
Code: https://pypi.python.org/pypi/hrf_estimation

fMRI paradigm and HRF

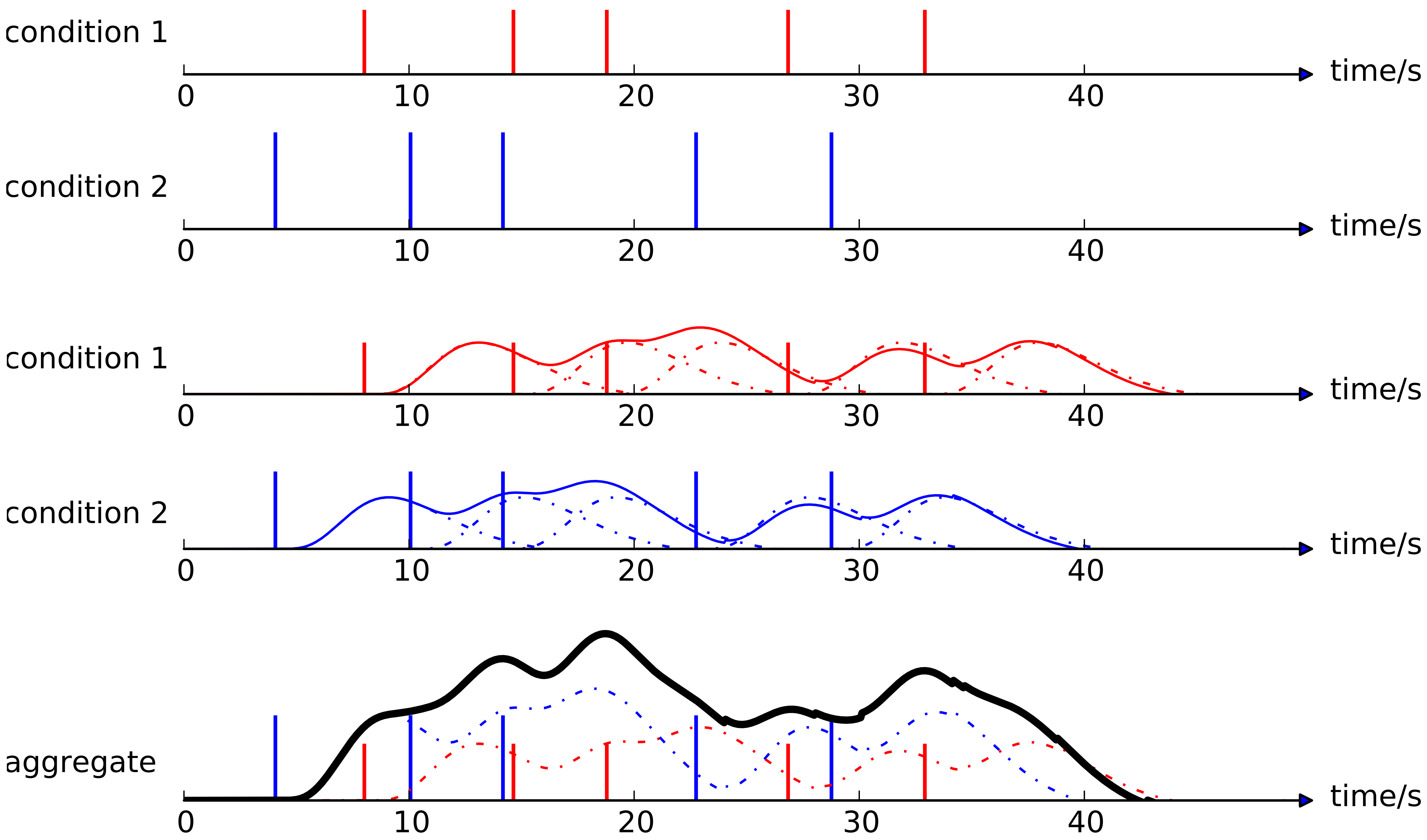


HRF: Hemodynamic response function



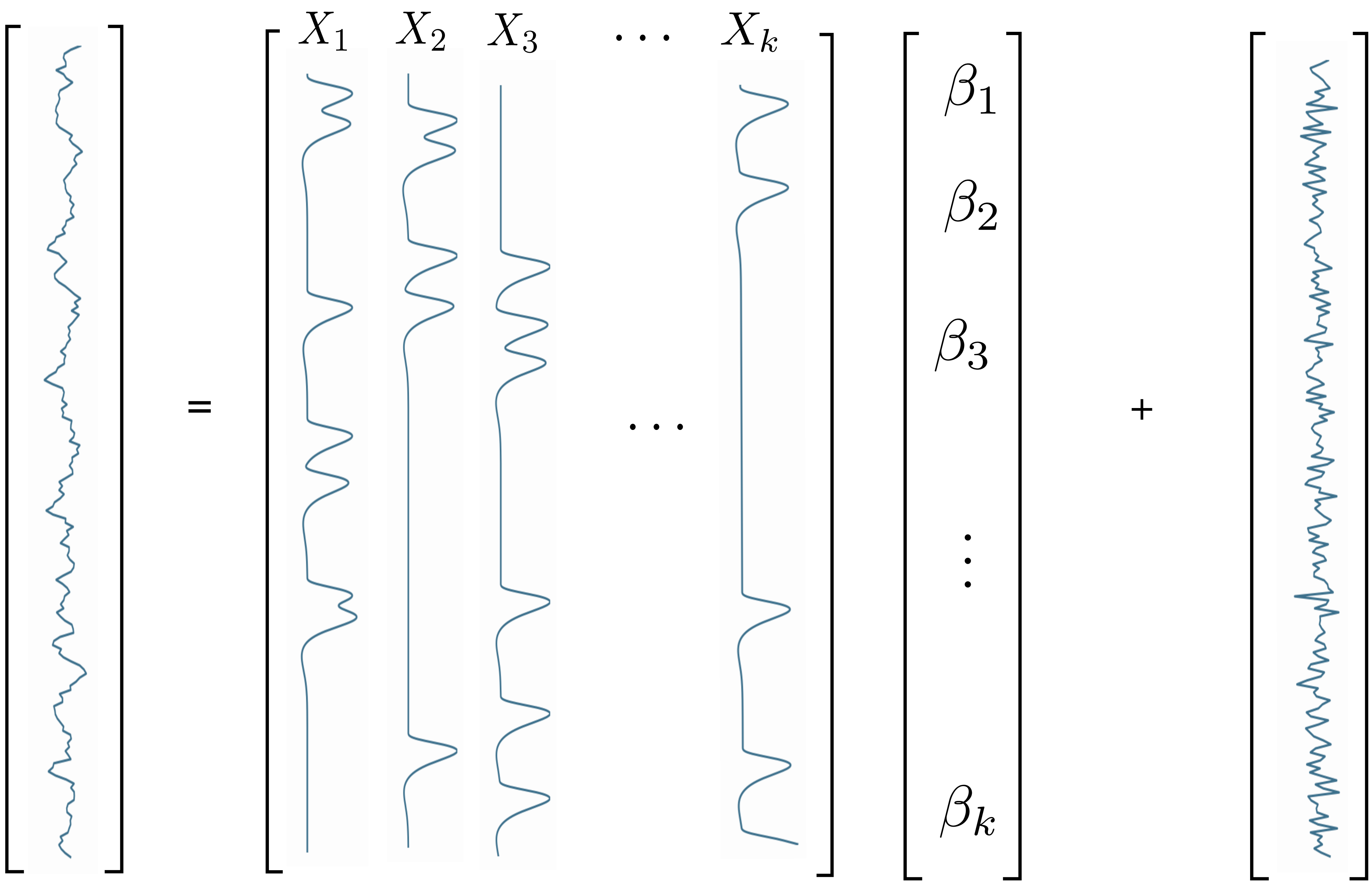


fMRI paradigm and HRF



General Linear Model (GLM)

y = Observed BOLD $\mathbf{X}$ = Design Matrix $\boldsymbol{\beta}$ = Activation coefficients $\boldsymbol{\varepsilon}$ = Noise



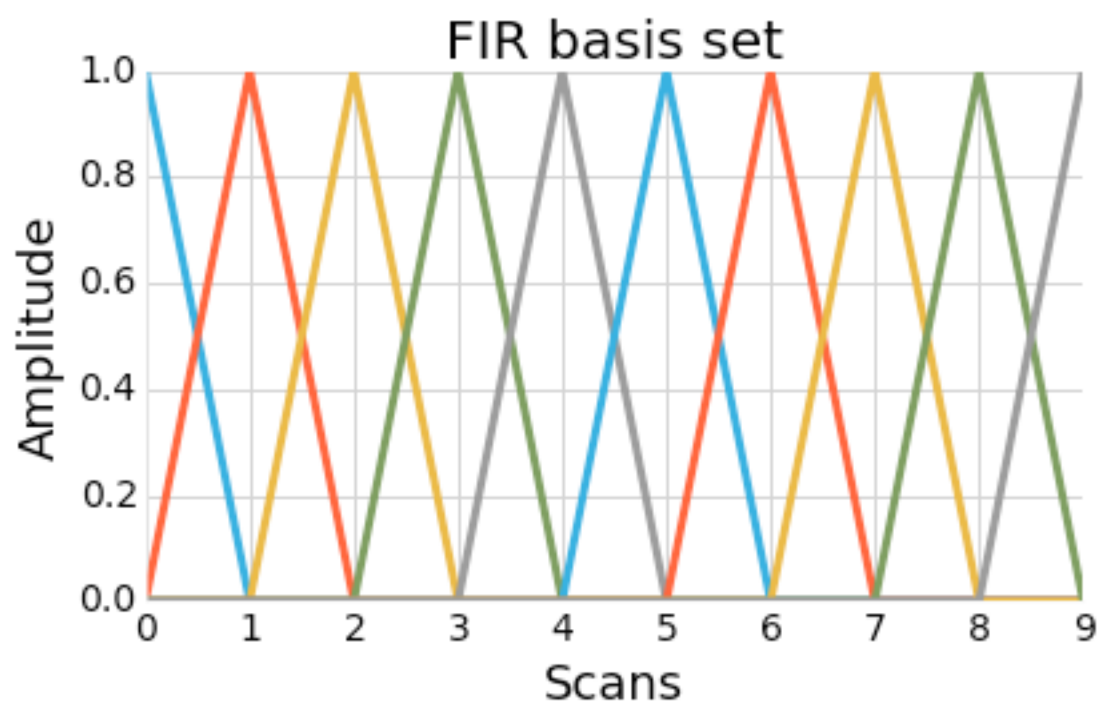
Basis constrained HRF

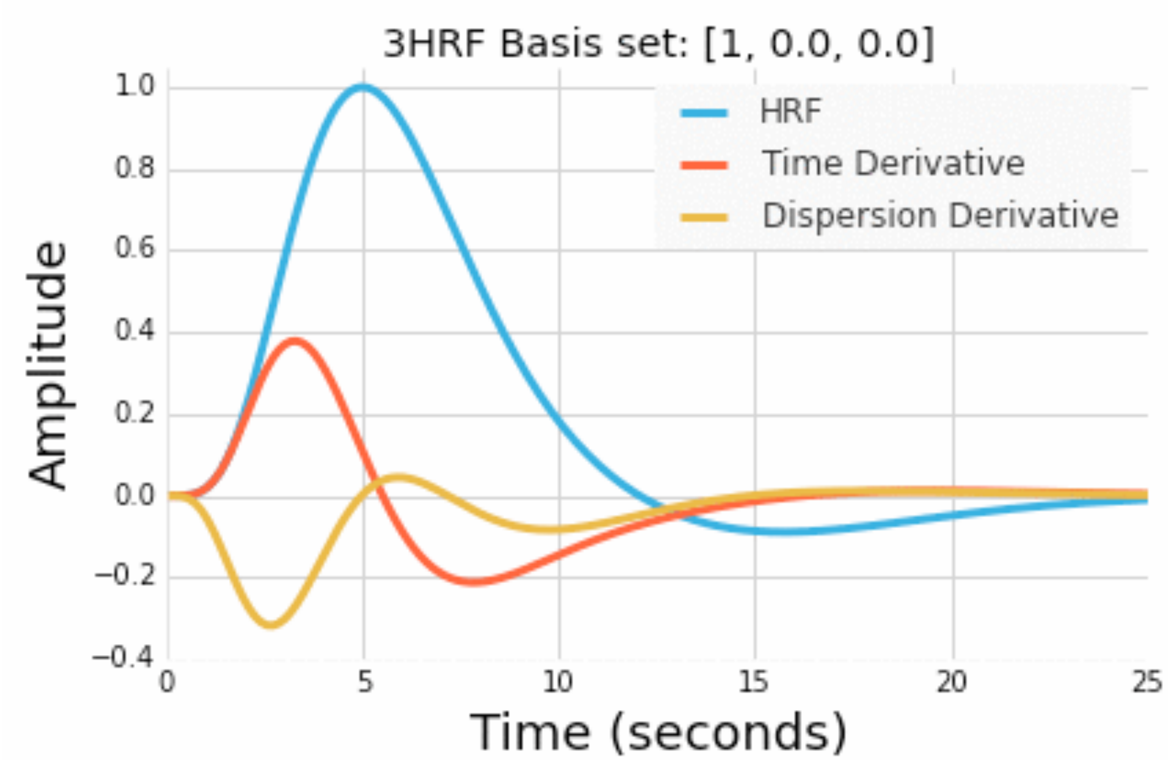
Hemodynamic response function (HRF) is known to vary substantially across subjects, brain regions and age.

D. Handwerker et al., "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses.," Neuroimage 2004.

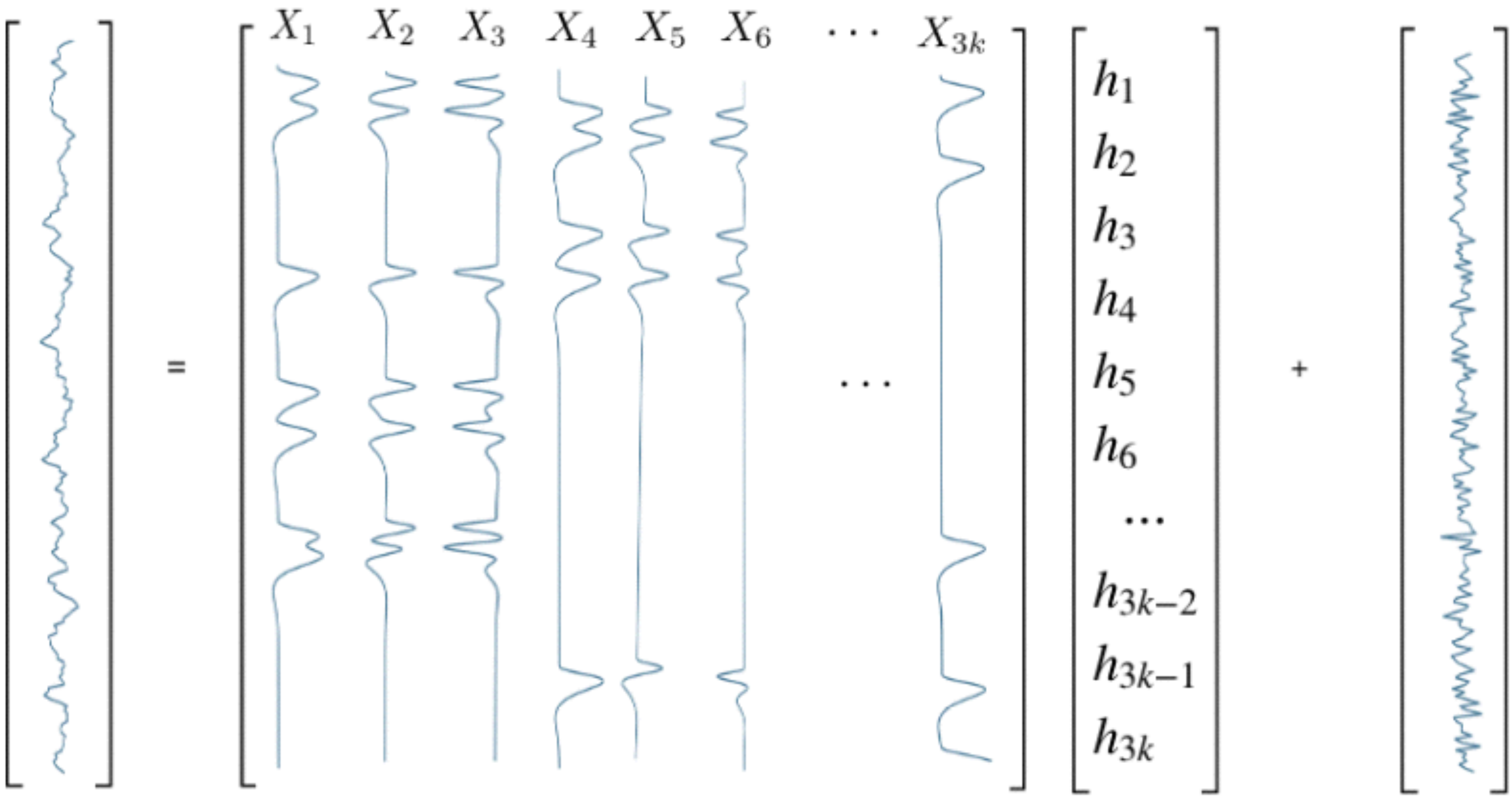
S. Badillo et al., "Group-level impacts of within- and between-subject hemodynamic variability in fMRI," Neuroimage 2013.

Two basis-constrained models of the HRF: FIR and 3HRF





Rank1-GLM


=
X_1 X_2 X_3 X_4 X_5 X_6 $\cdots$ X_{3k}
$\cdots$
h_1
h_2
h_3
h_4
h_5
h_6
$\cdots$
h_{3k-2}
h_{3k-1}
h_{3k}
+

Rank1-GLM

From 1 HRF per condition

$$\begin{pmatrix} h_1 & h_4 & & h_{3k-2} \\ h_2 & h_5 & \vdots & h_{3k-1} \\ h_3 & h_6 & & h_{3k} \end{pmatrix}$$

From 1 HRF shared between all conditions

$$\begin{pmatrix} h_1 & h_1 & & h_1 \\ h_2 & h_2 & \vdots & h_2 \\ h_3 & h_3 & & h_3 \end{pmatrix}$$

Rank1-GLM

Assuming 1 HRF shared between all conditions and a different amplitude/scale per condition this leads to:

$$\begin{pmatrix} \beta_1 h_1 & \beta_2 h_1 & & \beta_k h_1 \\ \beta_1 h_2 & \beta_2 h_2 & \vdots & \beta_k h_2 \\ \beta_1 h_3 & \beta_2 h_3 & & \beta_k h_3 \end{pmatrix} = \mathbf{h}\boldsymbol{\beta}^T$$

Rank1-GLM

$$\begin{pmatrix} \beta_1 h_1 & \beta_2 h_1 & & \beta_k h_1 \\ \beta_1 h_2 & \beta_2 h_2 & \vdots & \beta_k h_2 \\ \beta_1 h_3 & \beta_2 h_3 & & \beta_k h_3 \end{pmatrix} = \mathbf{h}\boldsymbol{\beta}^T$$

$$\text{argmin}_{\mathbf{h},\ \boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\text{vec}(\mathbf{h}\boldsymbol{\beta}^T)\|^2$$

$$\text{subject to } \|\mathbf{h}\|_\infty = 1 \text{ and } \langle \mathbf{h}, \mathbf{h}_{\text{ref}} \rangle > 0$$

$$\implies \text{ solved locally using quasi-Newton methods}$$

Challenge: This optimization problem is not big yet it needs to be done tens of thousands of time (typically 30,000 to 50,000 times for each voxel)

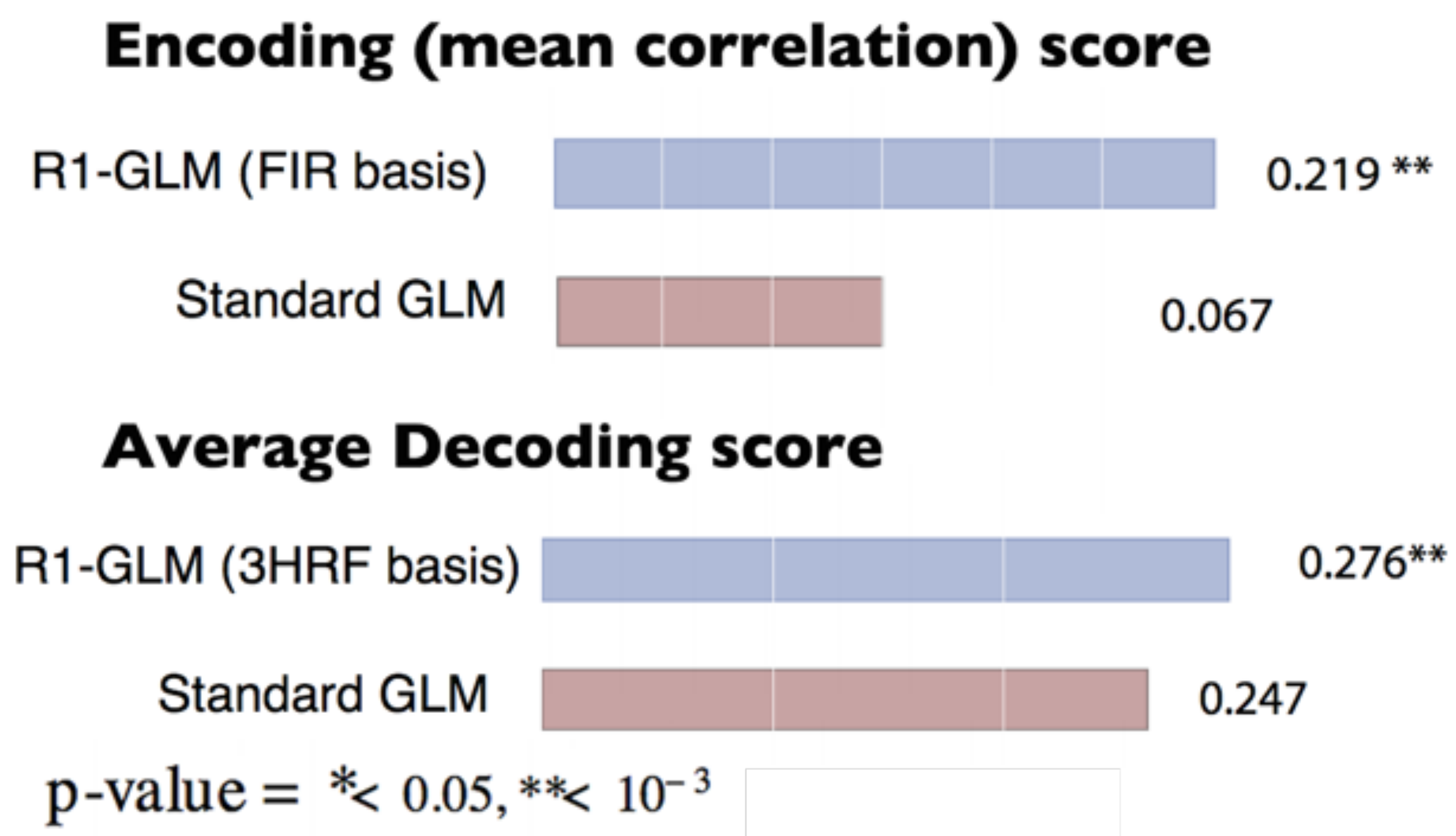
Remark: Worked better than alternated optimization or 1st order methods

Results

Cross-validation score in two different datasets

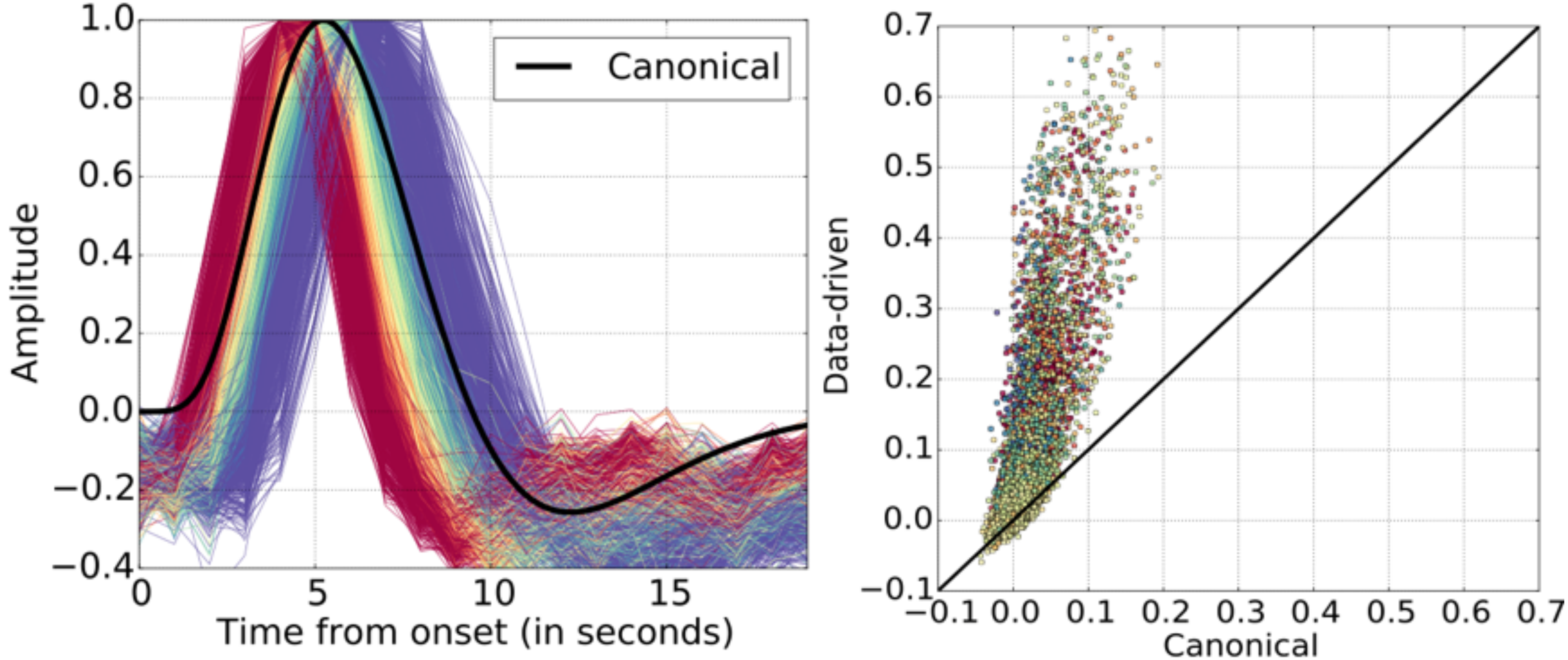
S. Tom et al., "The neural basis of loss aversion in decision-making under risk," Science 2007.

K. N. Kay et al., "Identifying natural images from human brain activity.," Nature 2008.



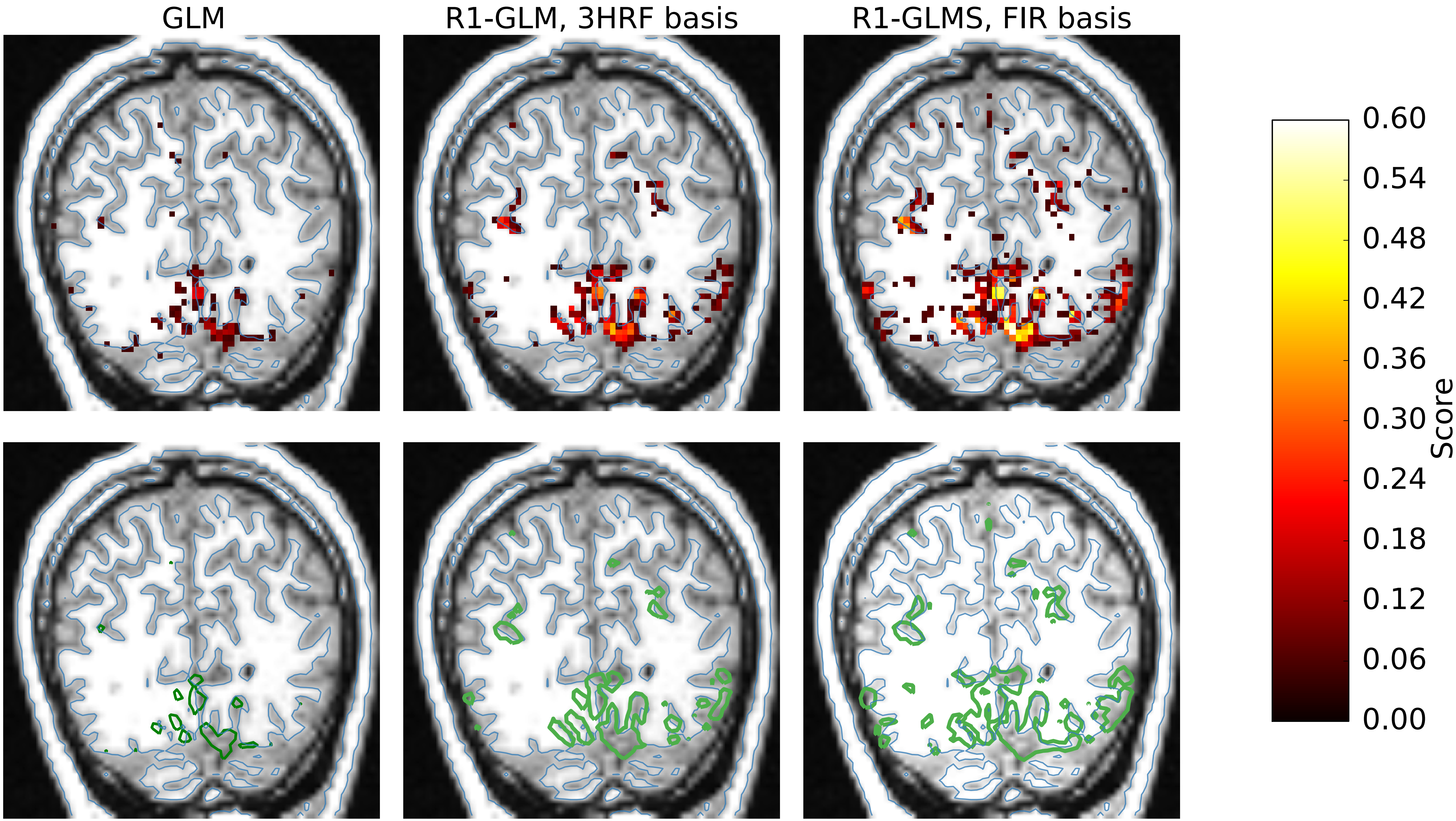
Results

Measure: voxel-wise encoding score. Correlation with the BOLD at each voxel on left-out data.

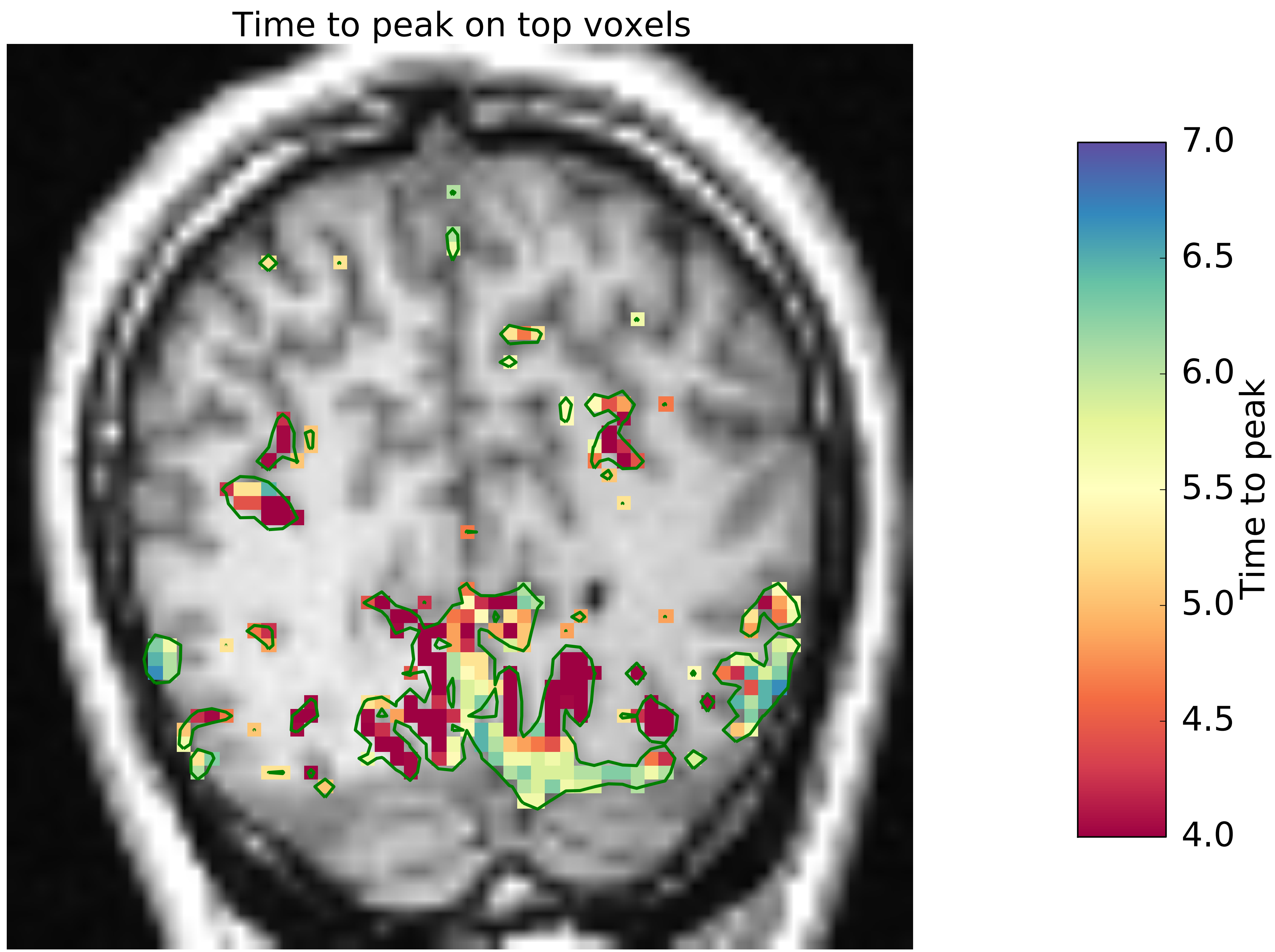


R1-GLM (FIR basis) improves voxel-wise encoding score on more than 98% of the voxels.

Results


GLM
R1-GLM, 3HRF basis
R1-GLMS, FIR basis
0.60
0.54
0.48
0.42
0.36
0.30
0.24
0.18
0.12
0.06
0.00
Score

Results



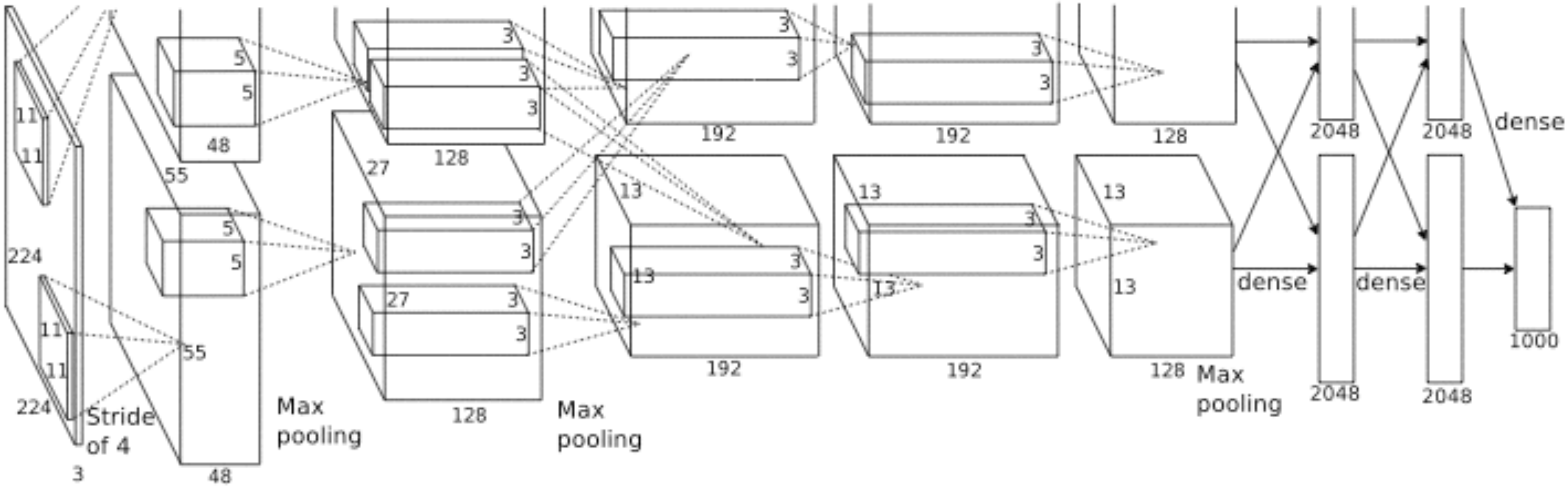
Convolutional Networks Map the Architecture of the Human Visual System

work of Michael Eickenberg

joint work with Bertrand Thirion and Gaël Varoquaux

"Seeing it all: Convolutional network layers map the function of the human visual system"
Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, Bertrand Thirion (submitted)

Convolutional Nets for Computer Vision

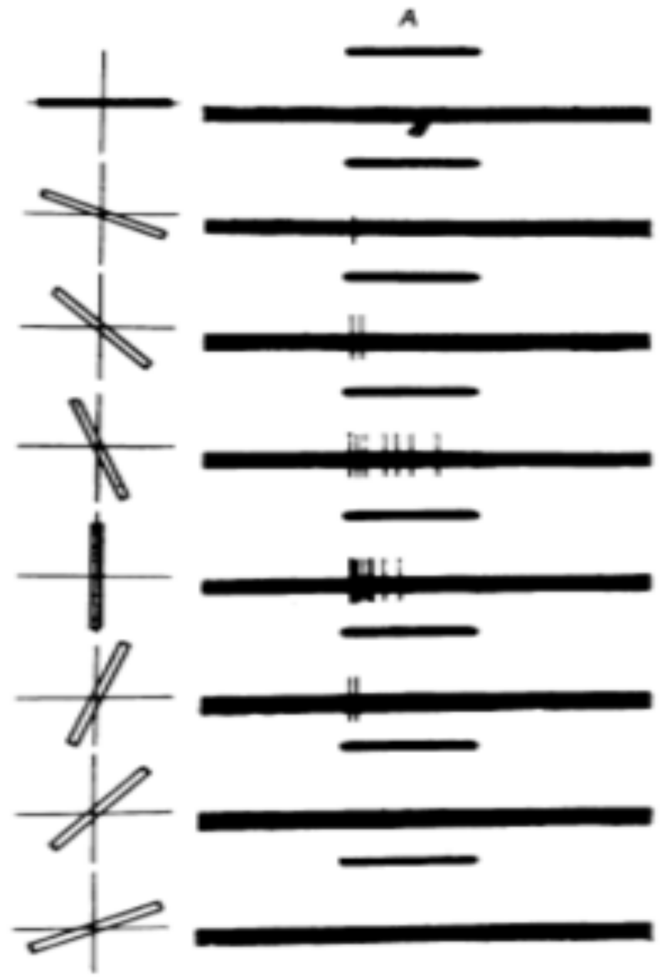


[Krizhevski et al, 2012]

Relating biological and computer vision

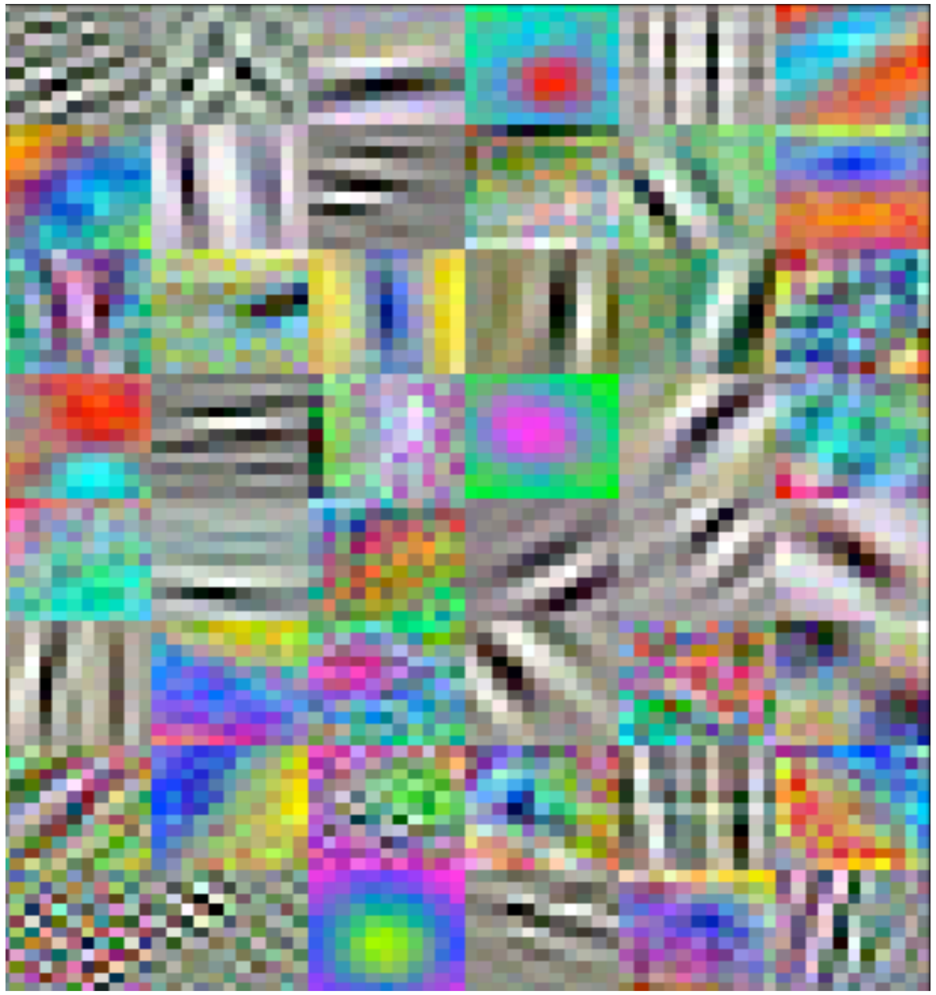
Low
Level

Cat V1
orientation selectivity



[Hubel & Wiesel, 1959]

ConvNet Layer 1



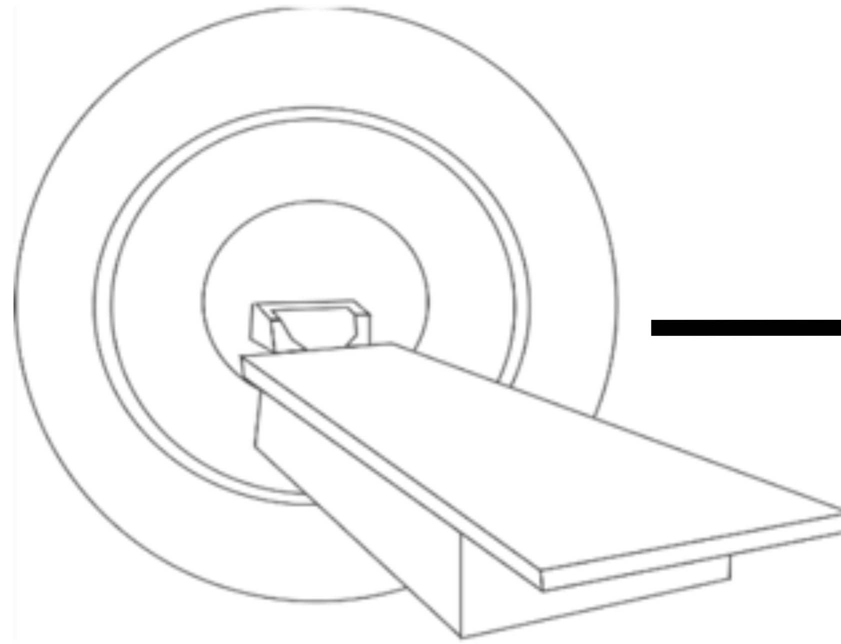
[Sermanet 2013]

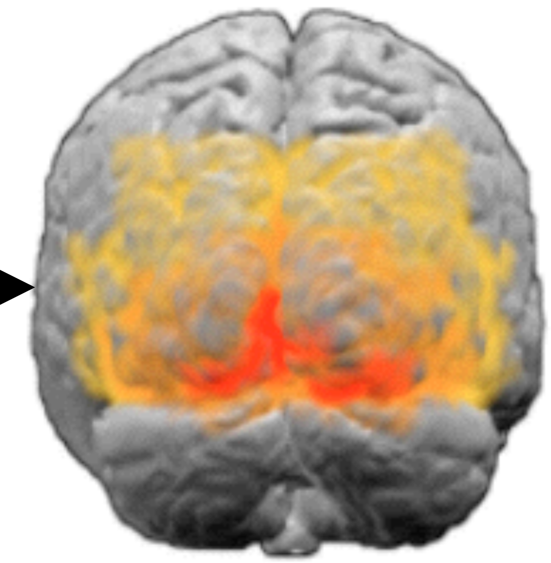
- V1 functionality comprises edge detection
- Convolutional nets learn edge detectors, color boundary detectors and blob detectors

Can we use computer vision models and a large fMRI data to better understand human vision?

Approach

[Kay et al, 2008]





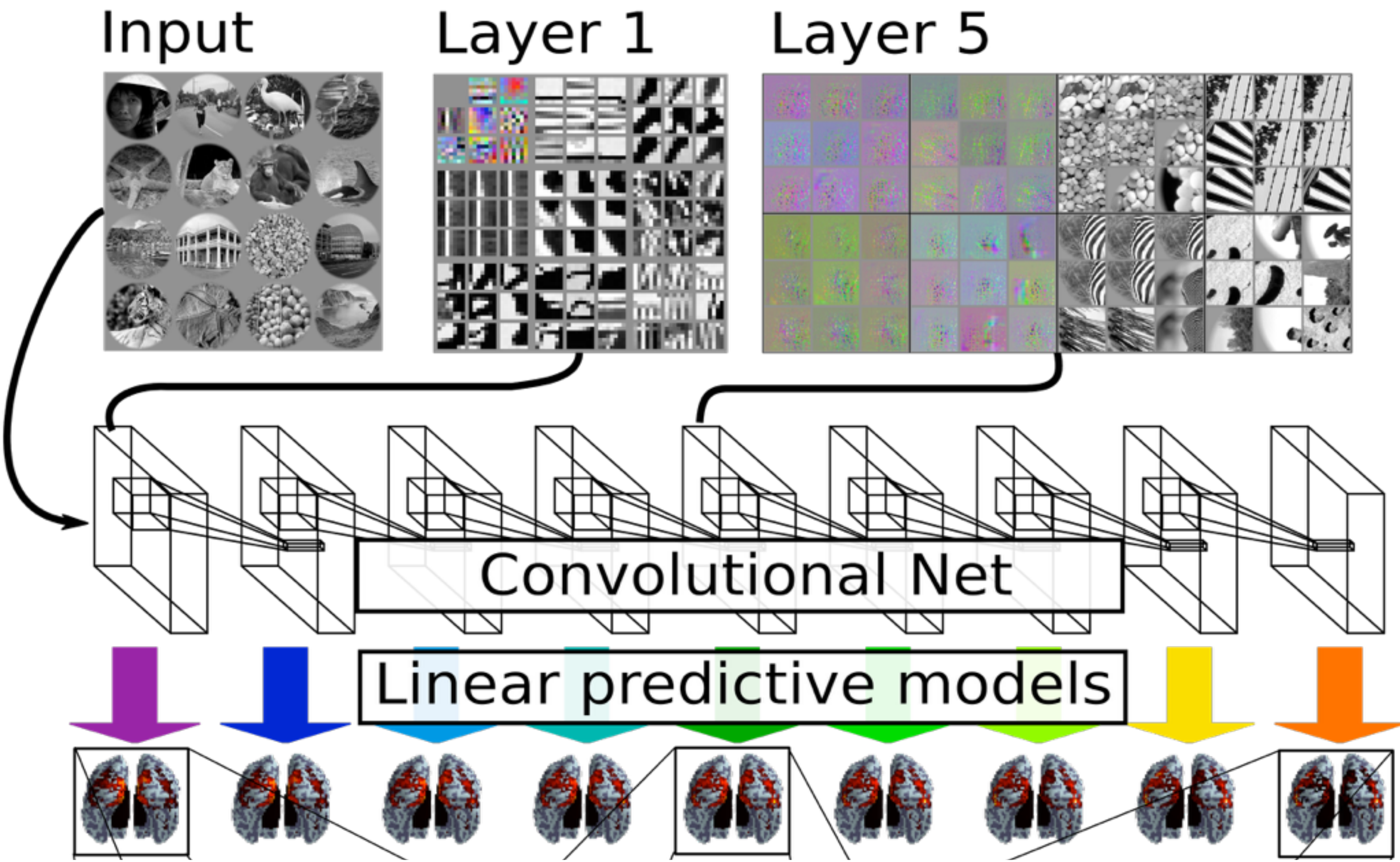
Nonlinear Feature Extraction
Via
Convolutional Net Layers

Voxel-Wise Prediction
Using **Linear** Model
(Ridge Regression)

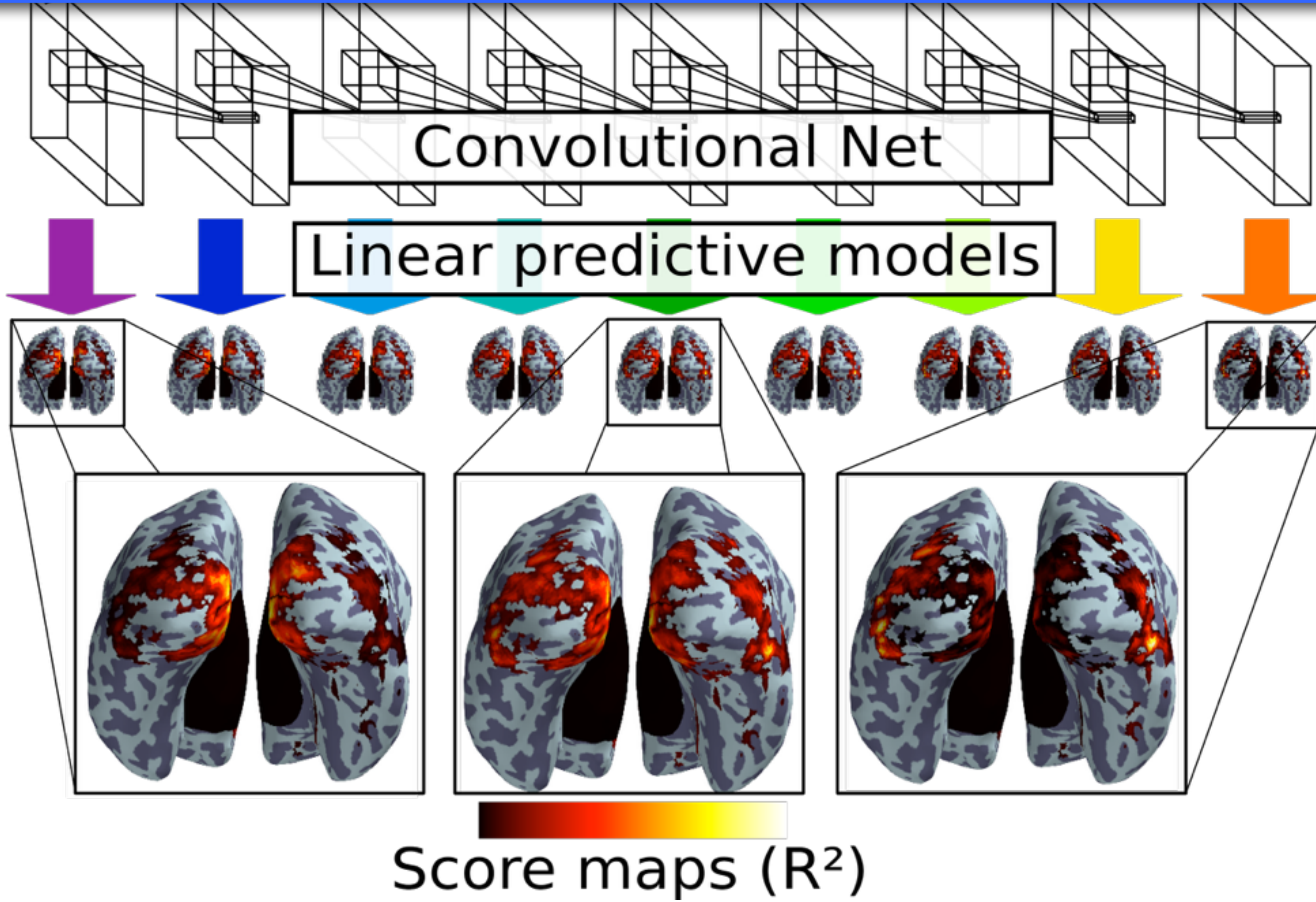
Forward Model Setup:

- Encoding model [Naselaris et al., 2011]
- Make sure complexity resides in feature extraction

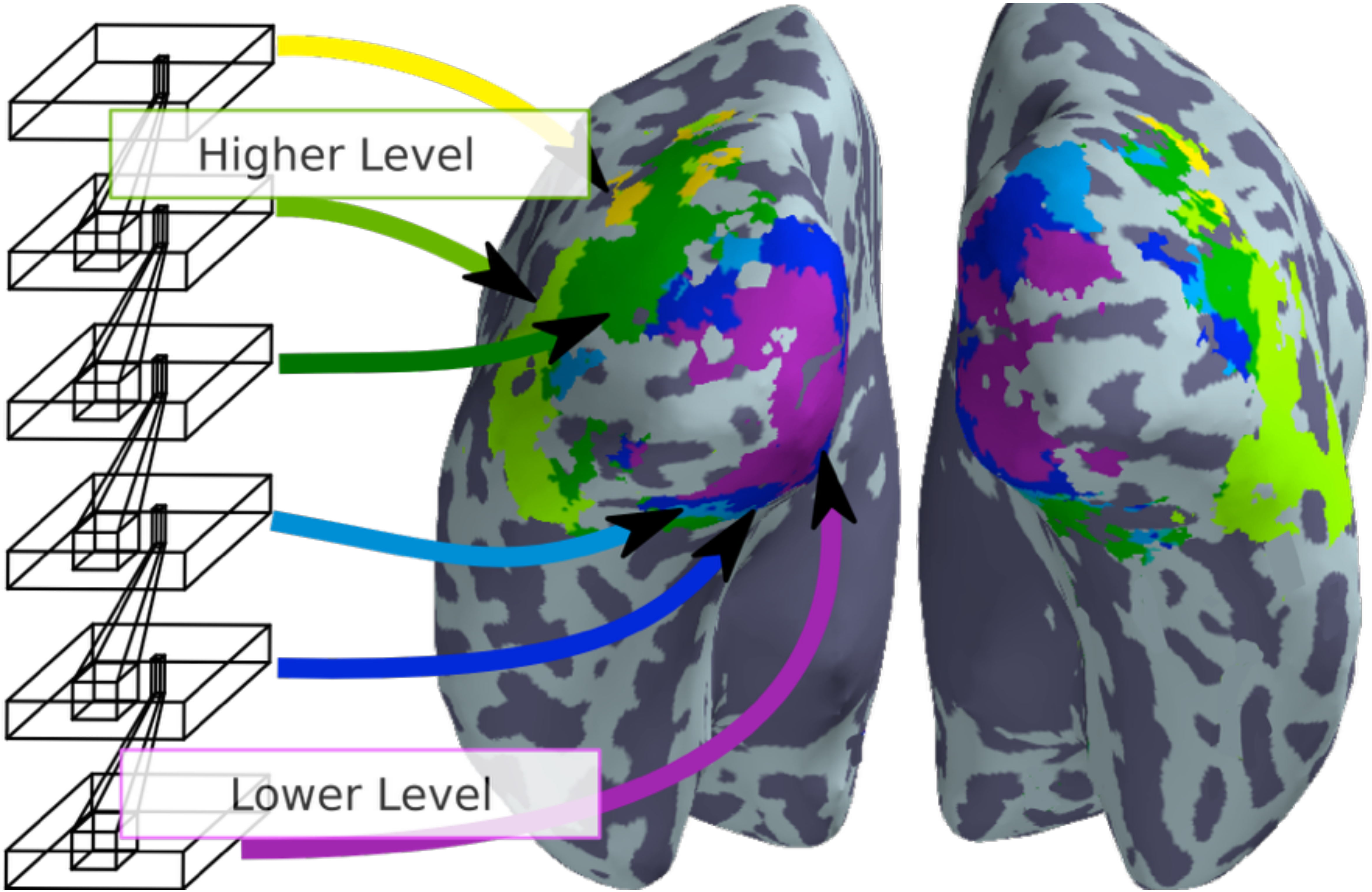
Convolutional Net Forward Models



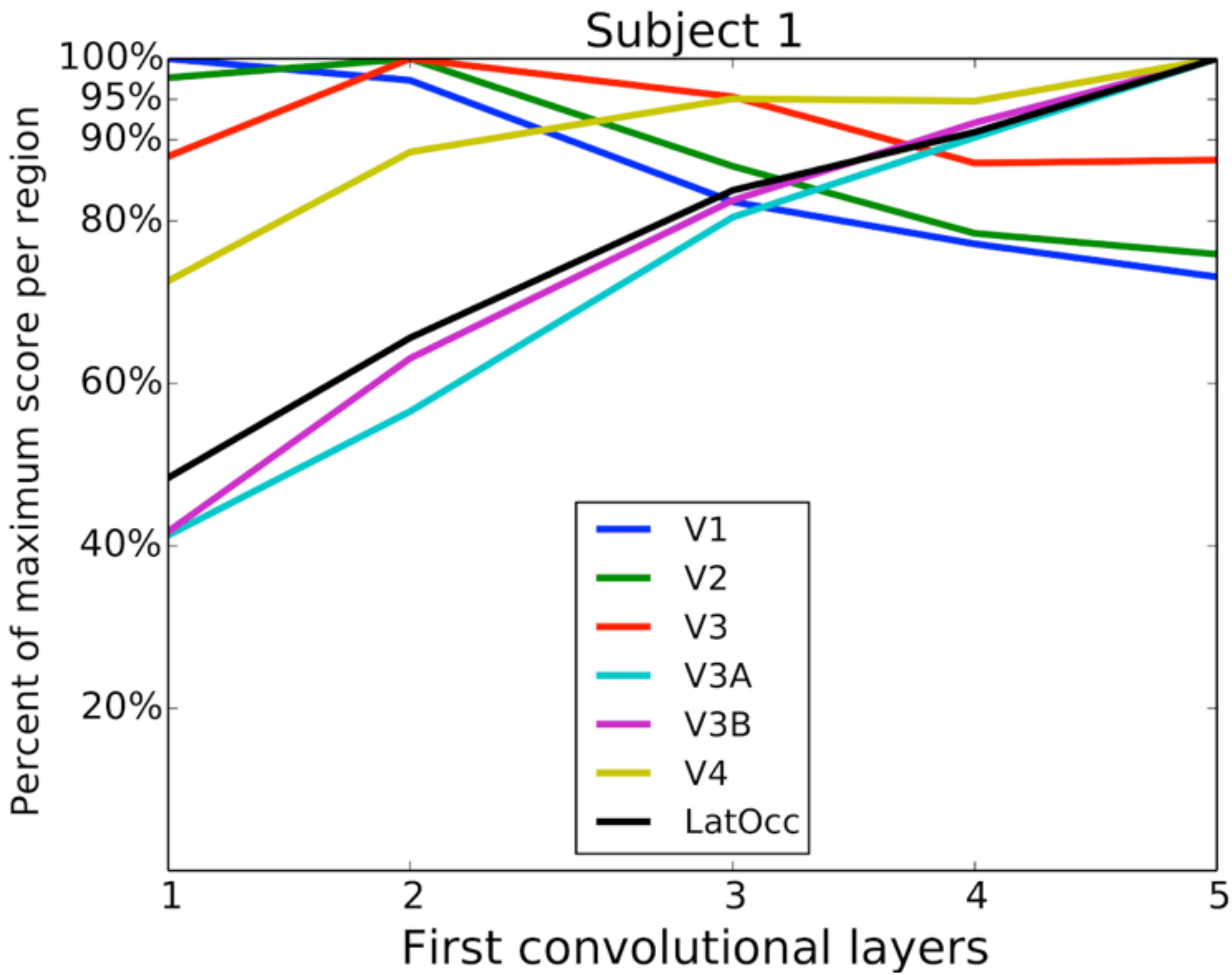
Convolutional Net Forward Models



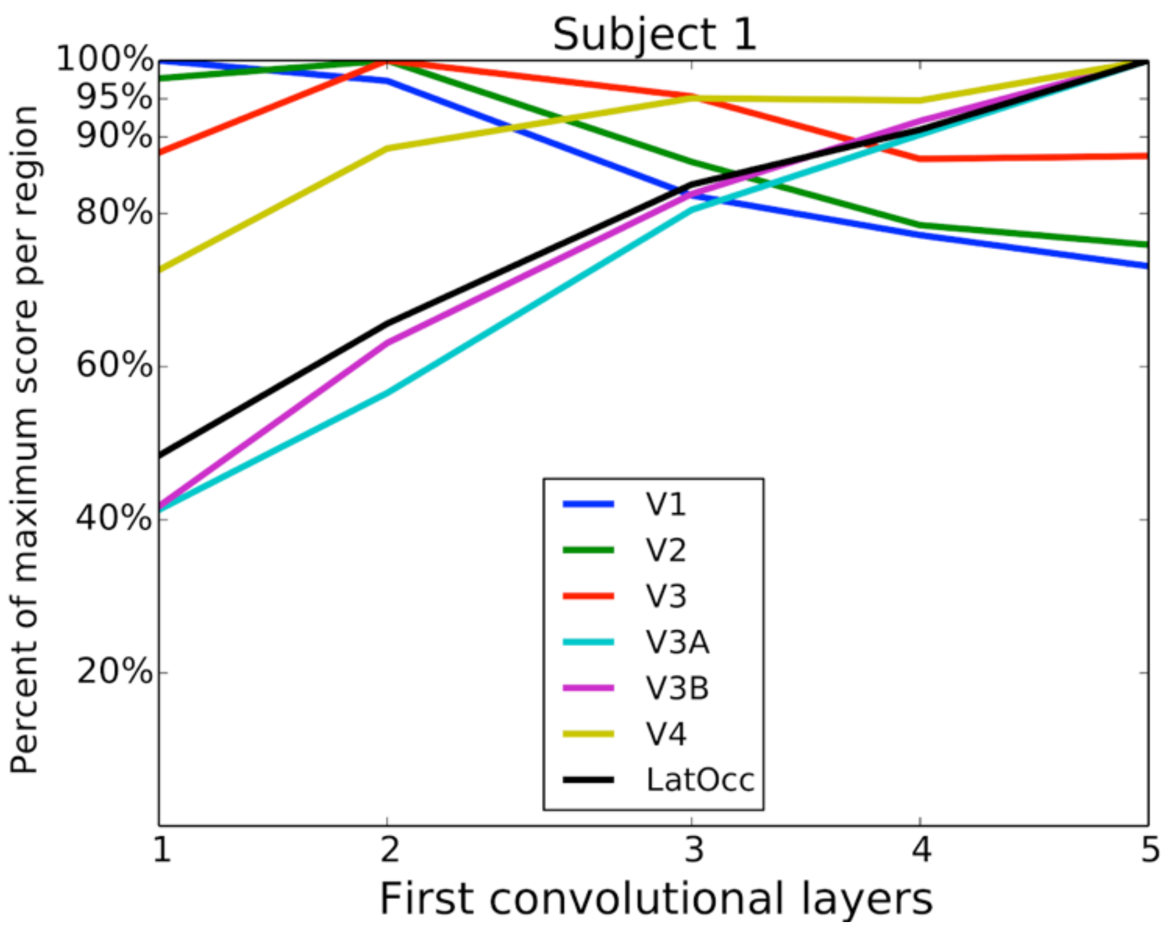
Best Predicting Layers per Voxel

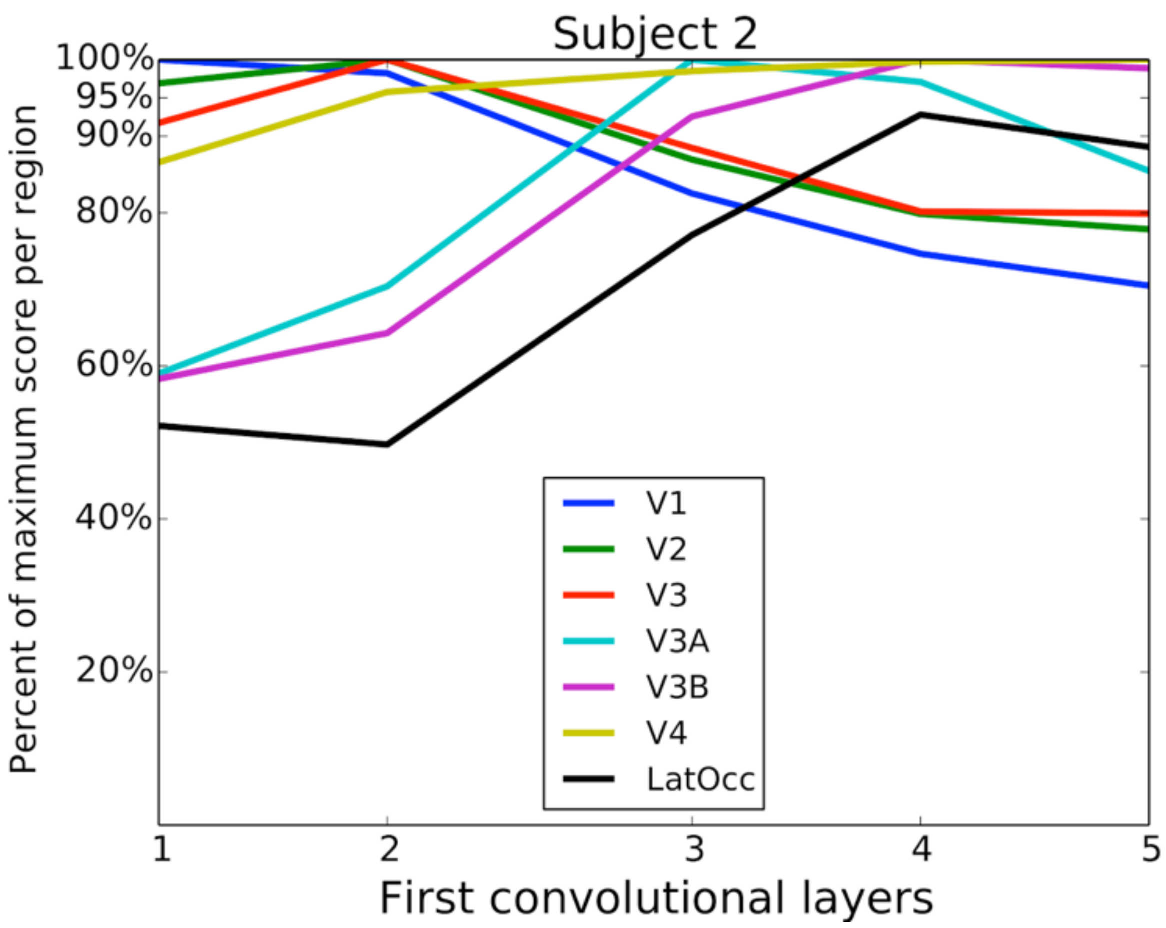


Score Fingerprints per Region of Interest

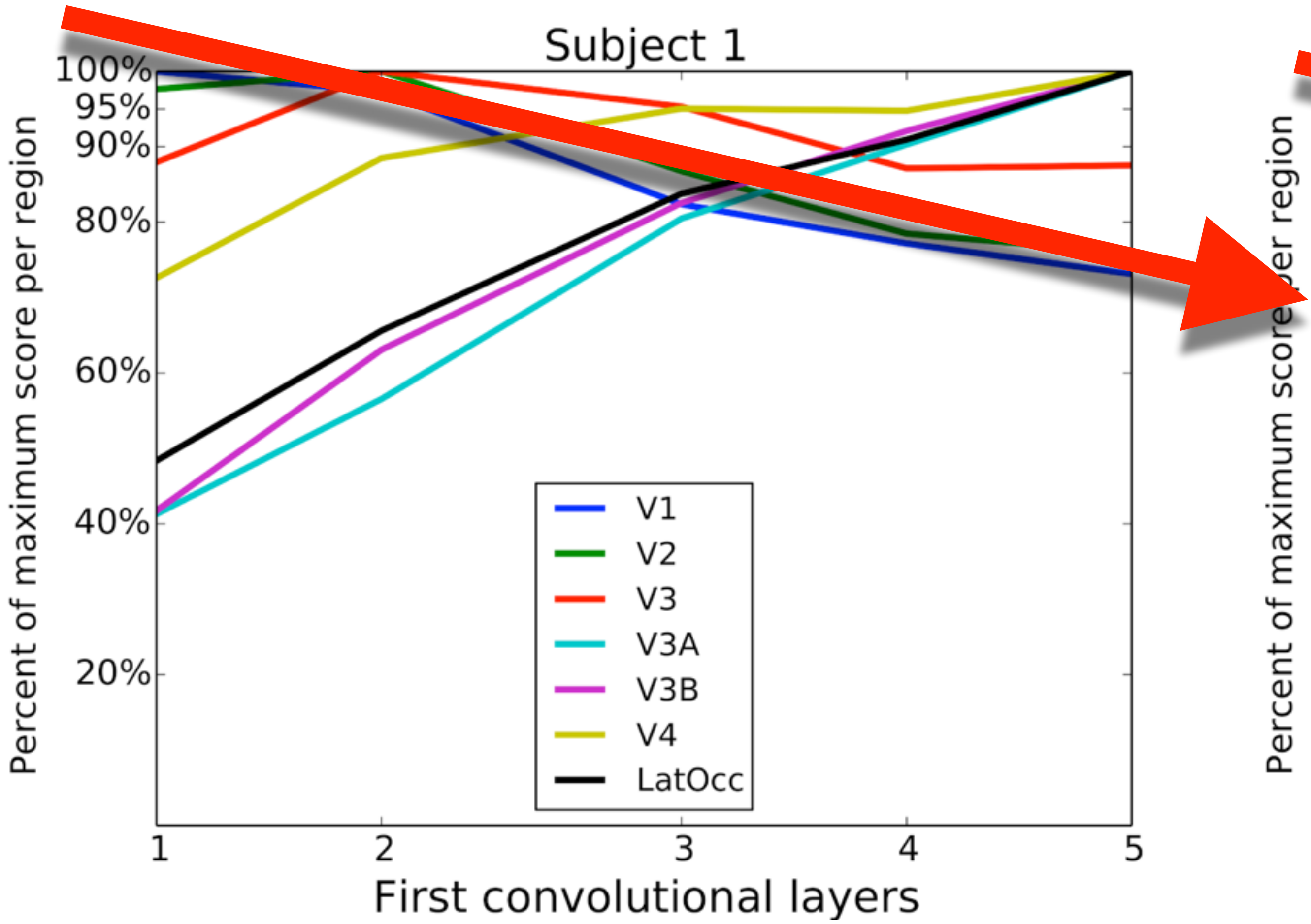


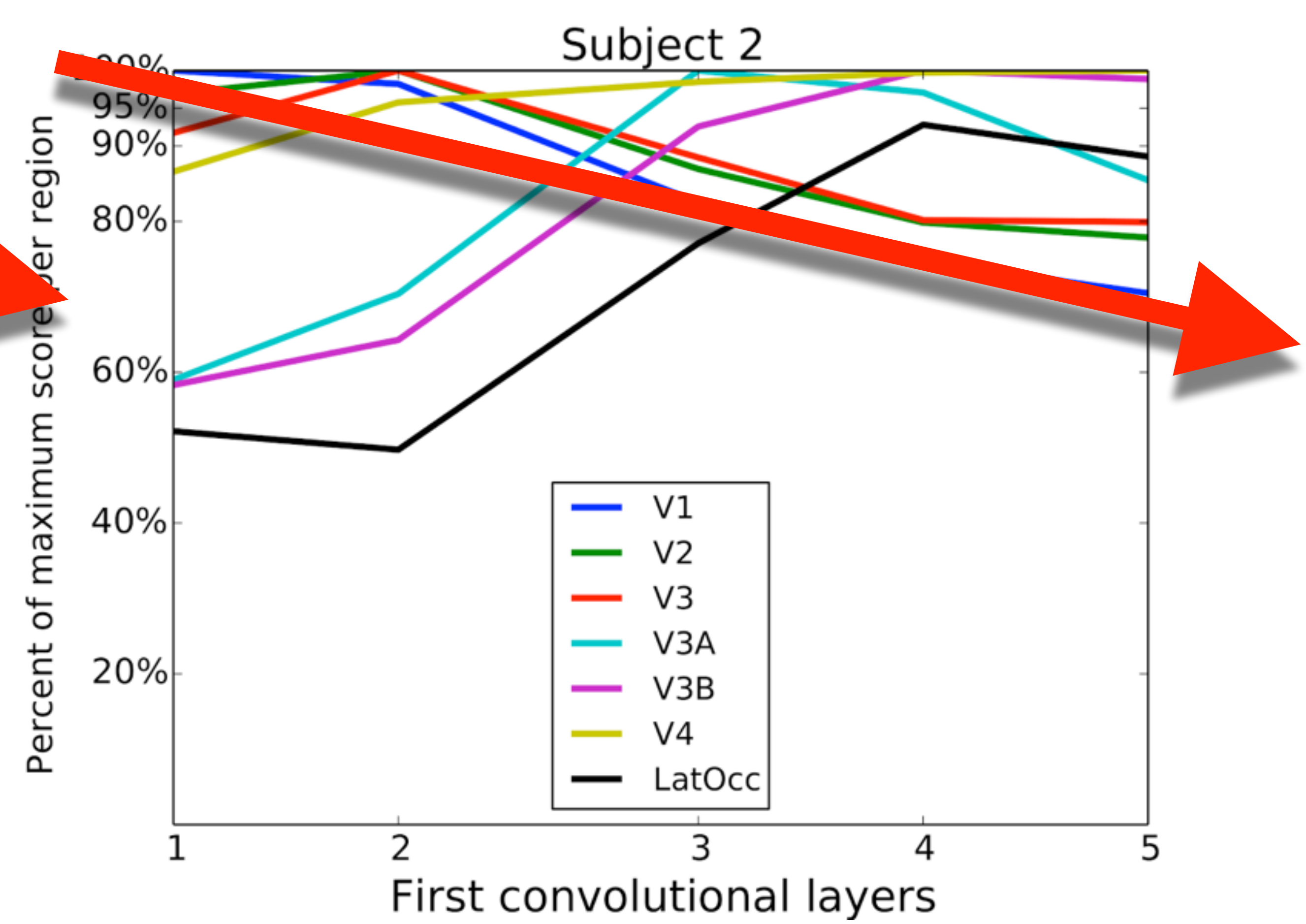
Score Fingerprints per Region of Interest



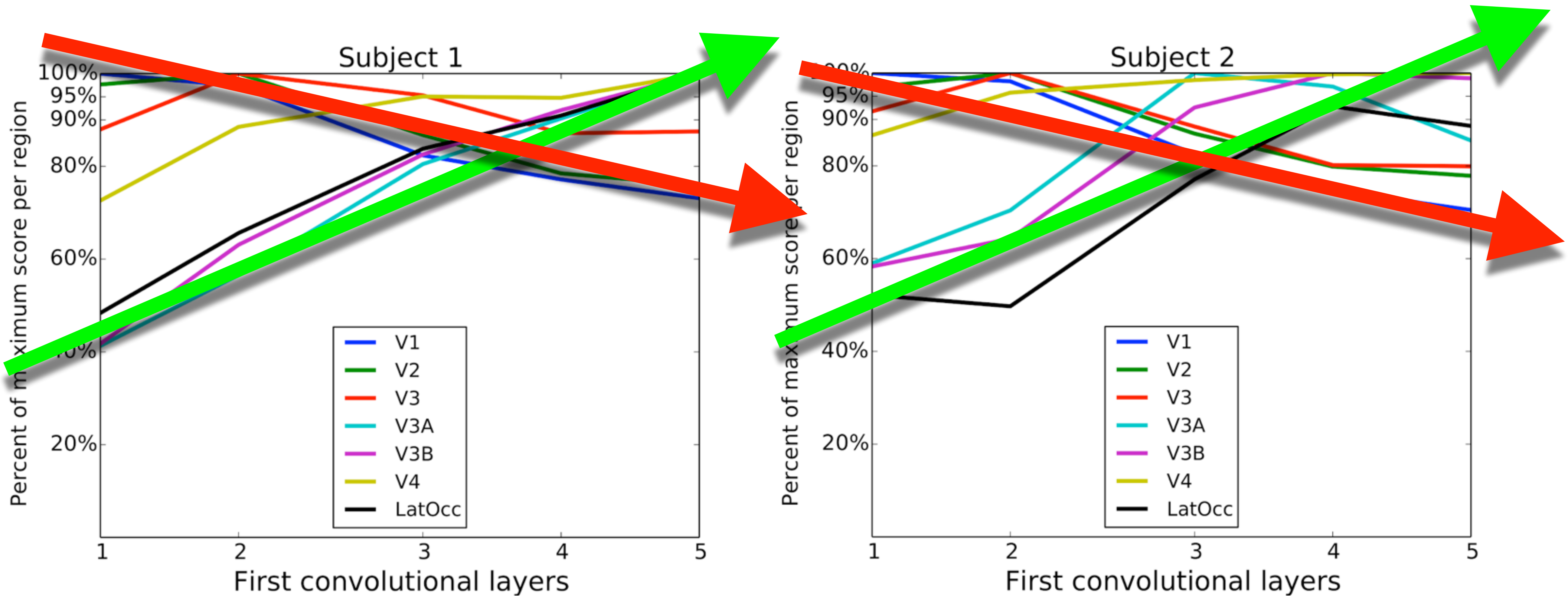


Score Fingerprints per Region of Interest



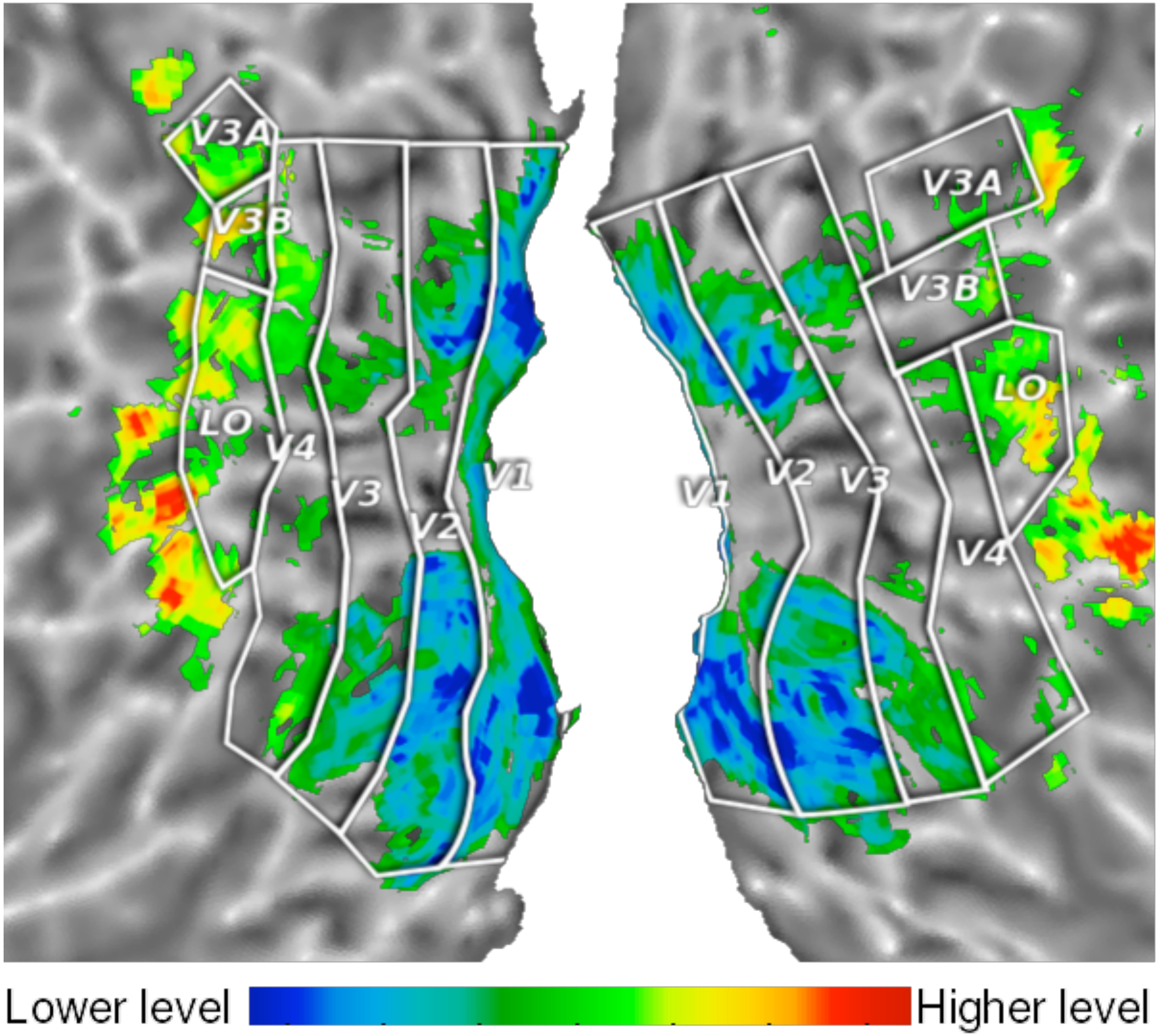


Score Fingerprints per Region of Interest

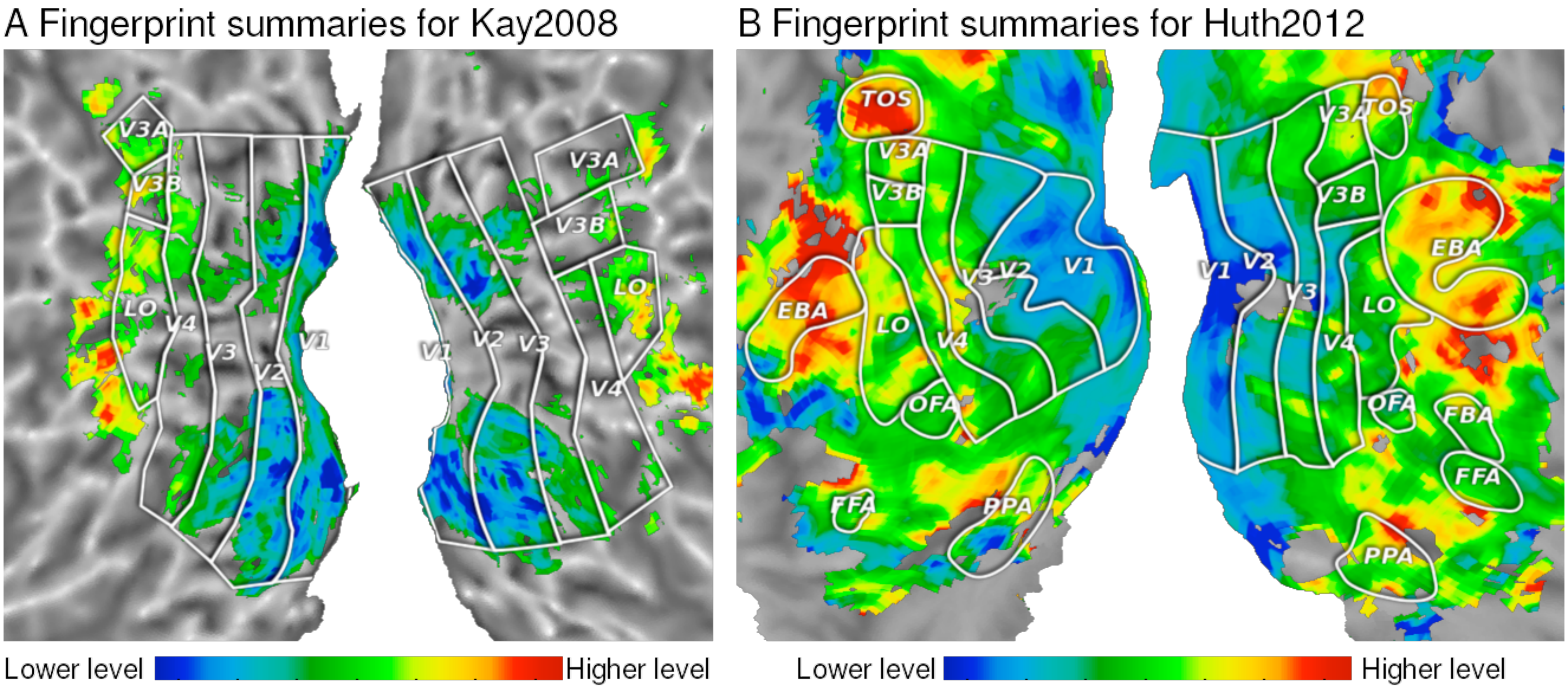


Fingerprints summary statistic

A Fingerprint summaries for Kay2008



Fingerprints summary statistic



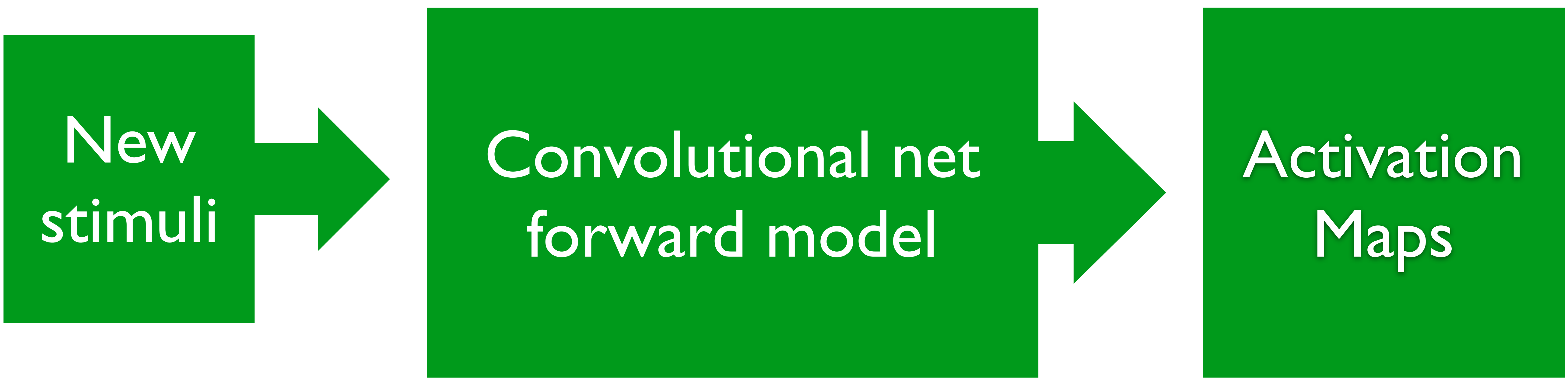
Photos

Videos

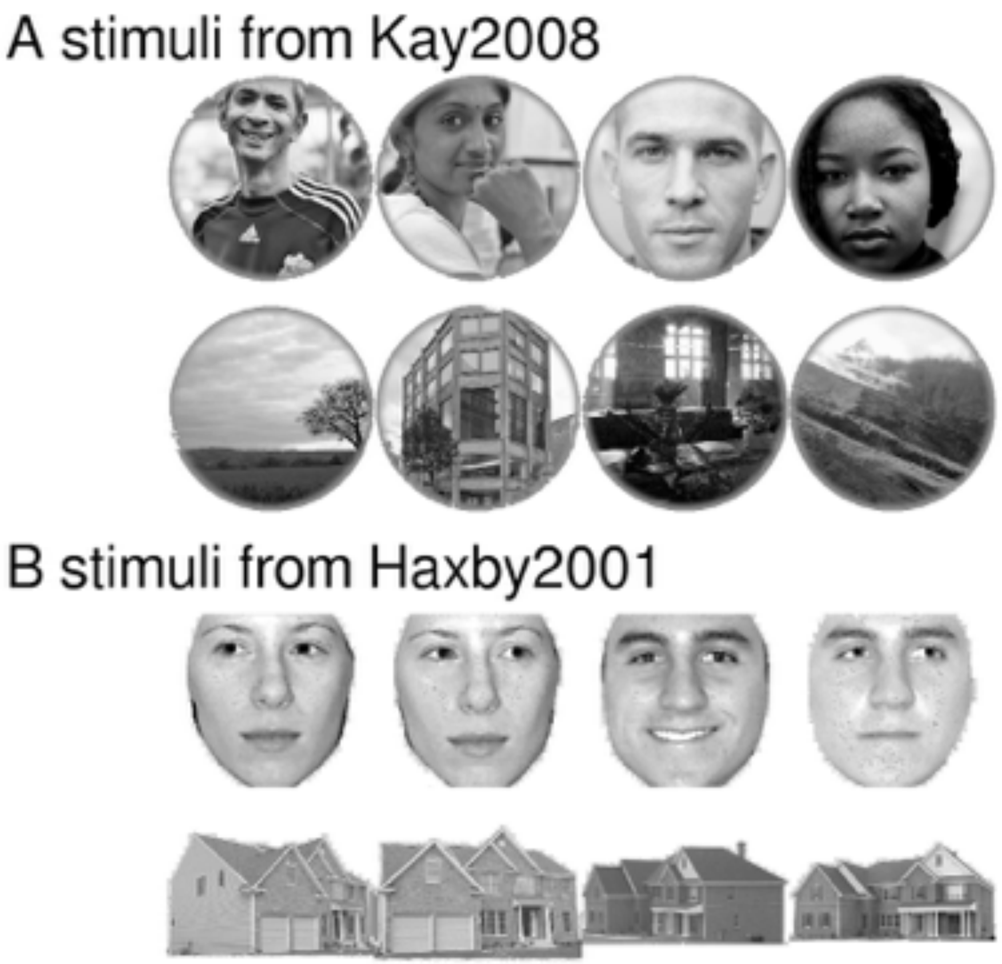
Synthesizing Brain activation maps

If our model is strong enough, we can use it to reproduce known experiments

Generate BOLD response, do GLM analysis



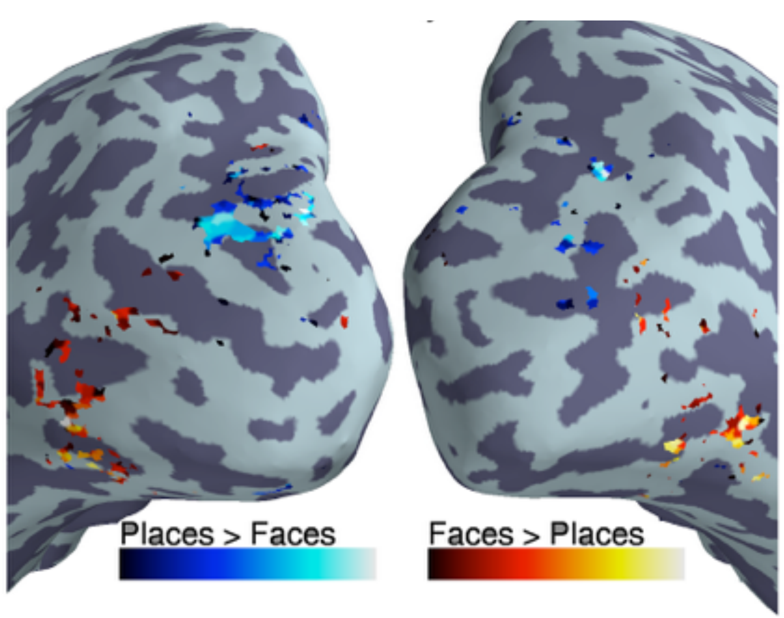
High-level Validation: Faces / Places



Convolutional Net Forward Model

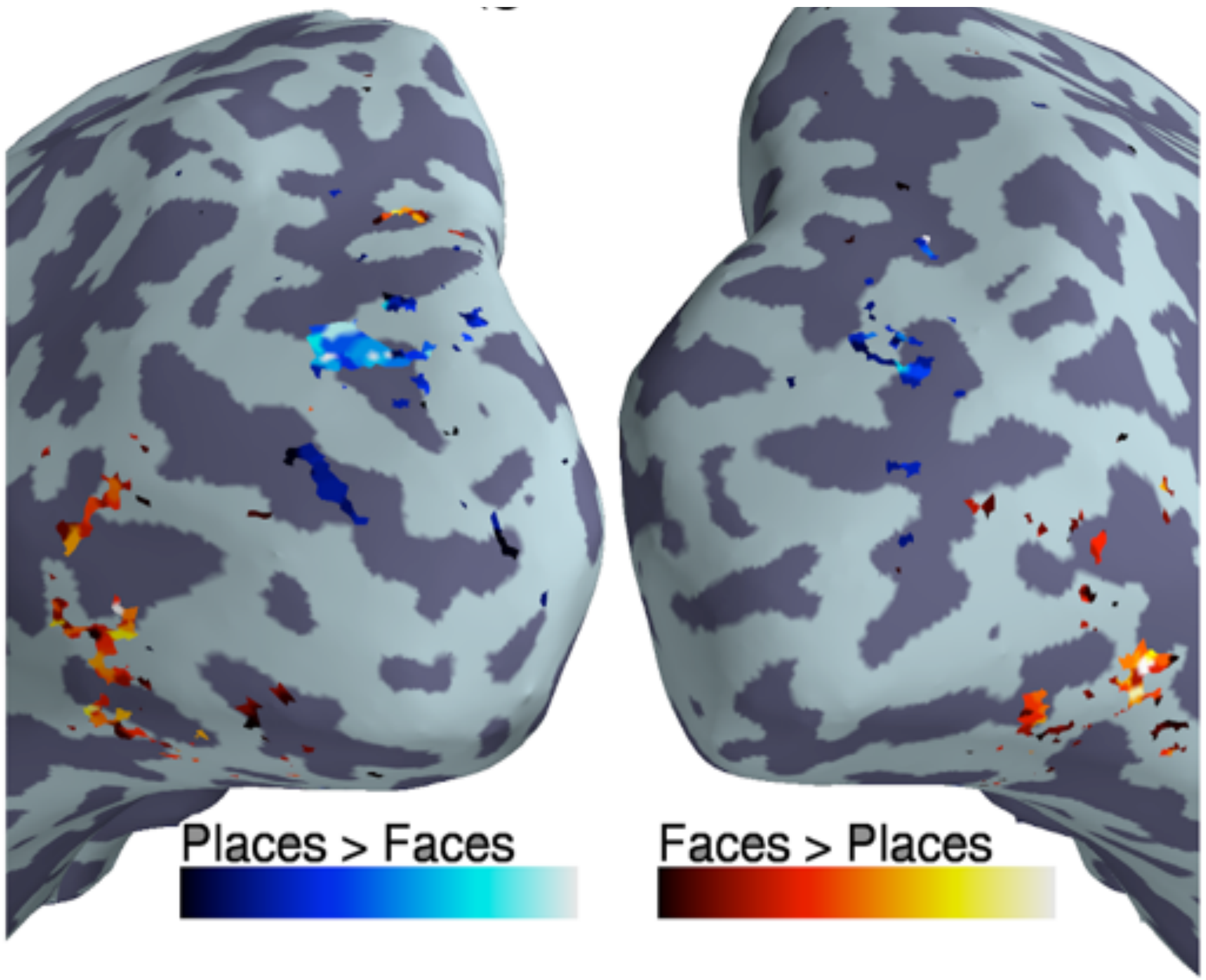
Activation Maps

GLM Contrast Maps



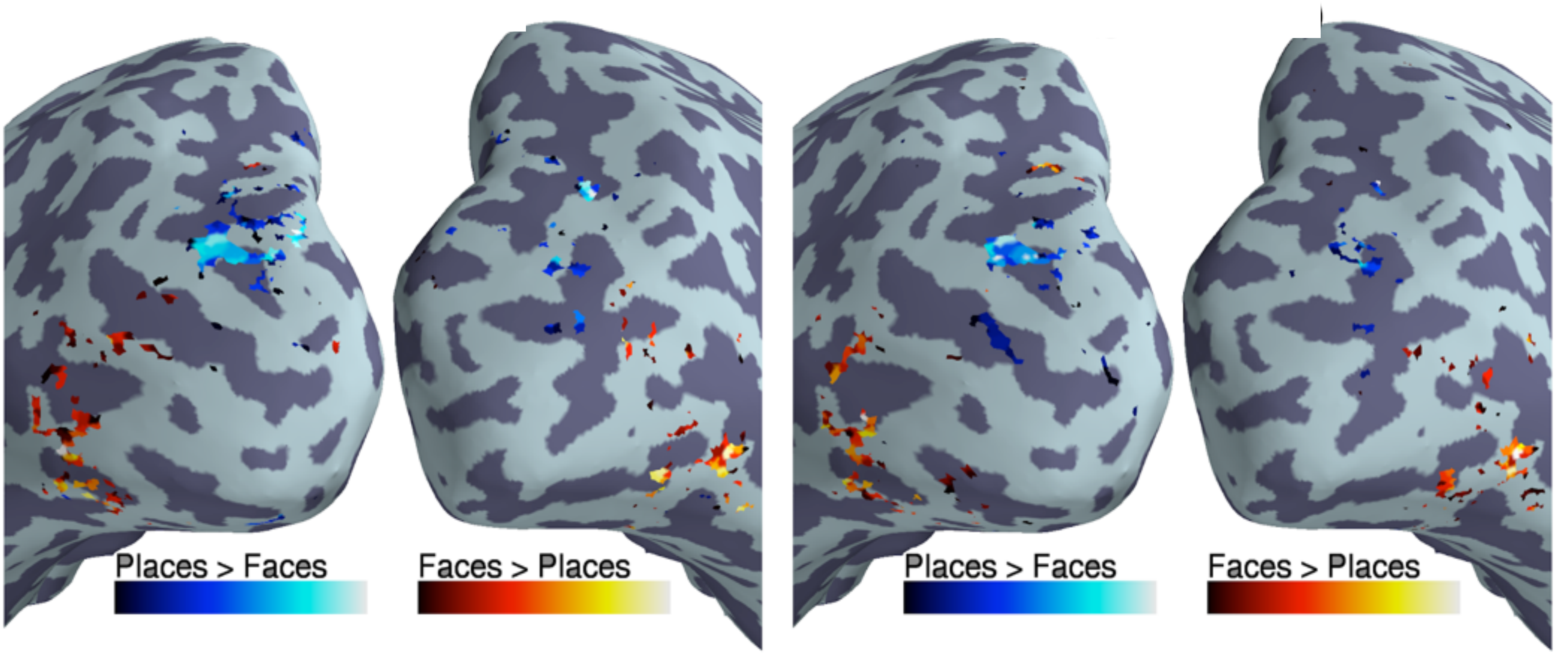
Faces vs Places: Ground Truth

Stimuli from [Kay 2008]
Close-up faces and scenes



Contrast of
stimuli from [Kay 2008]
Close-up faces and scenes

Faces vs Places

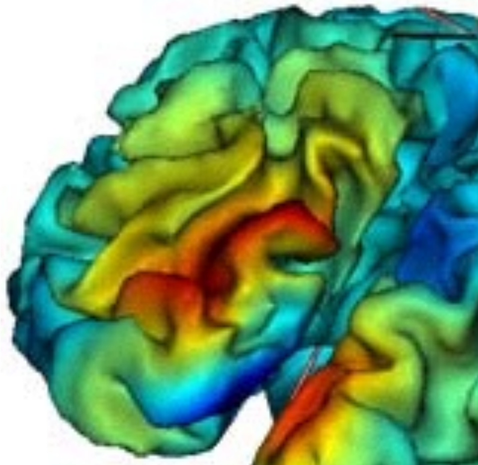


Simulation on
[Kay 2008] Left out stimuli

BOLD ground truth

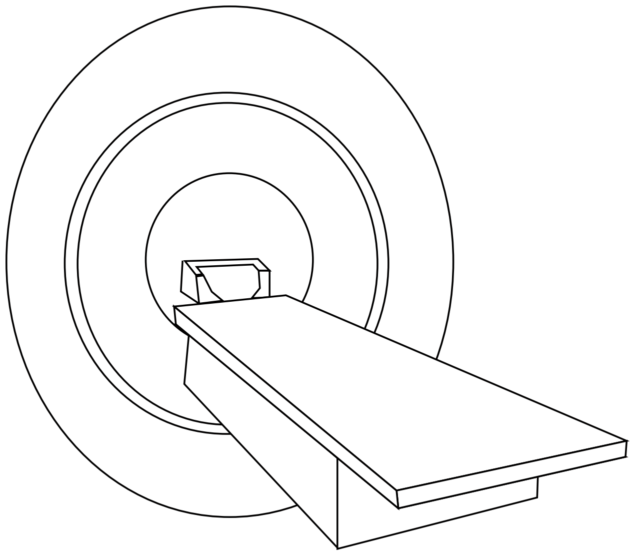
Fast Optimal Transport Averaging of Neuroimaging Data

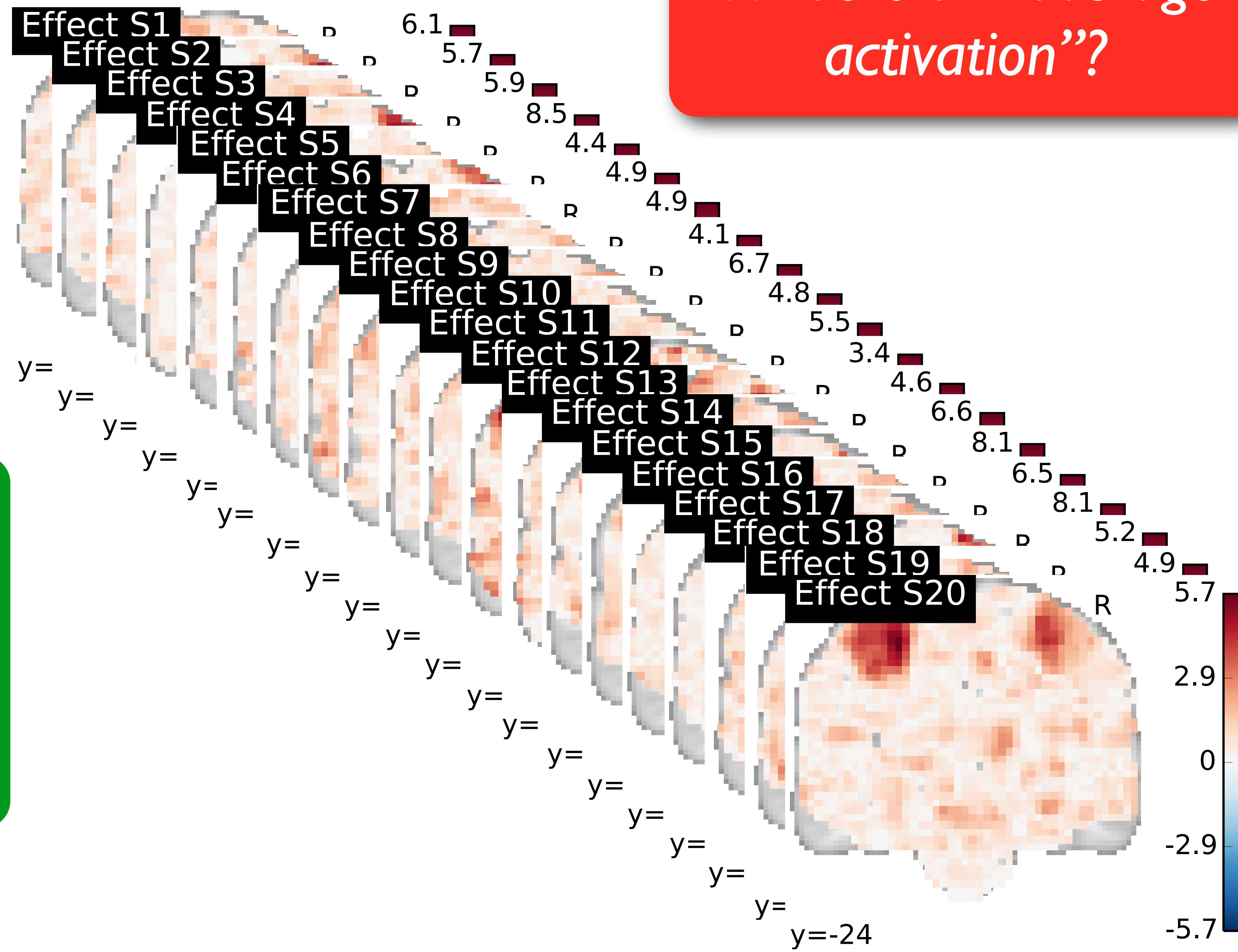
Joint work with: Gabriel Peyré Marco Cuturi



[Fast Optimal Transport Averaging of Neuroimaging Data
Alexandre Gramfort, Gabriel Peyré, Marco Cuturi, Proc. IPMI 2015]

The overall goal



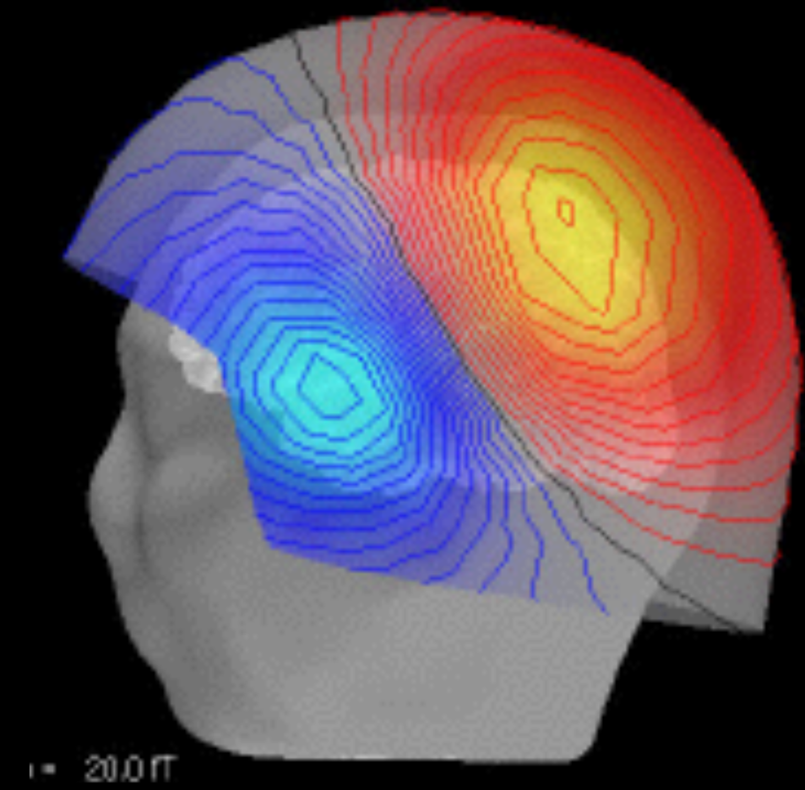


What is an "average activation"?

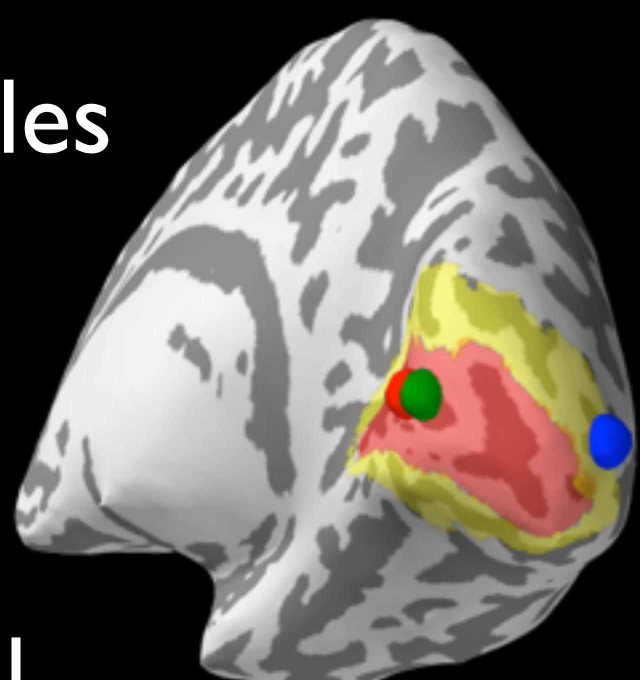
Functional neuroimaging experiment 20 subjects

with Magnetoencephalography (MEG)

From sensors to sources at every ms for each subject



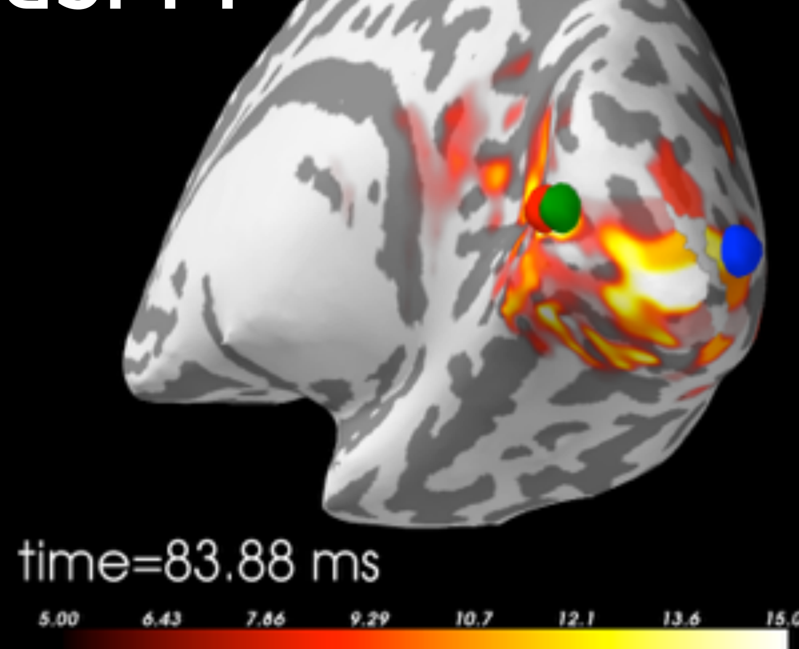
Dipoles



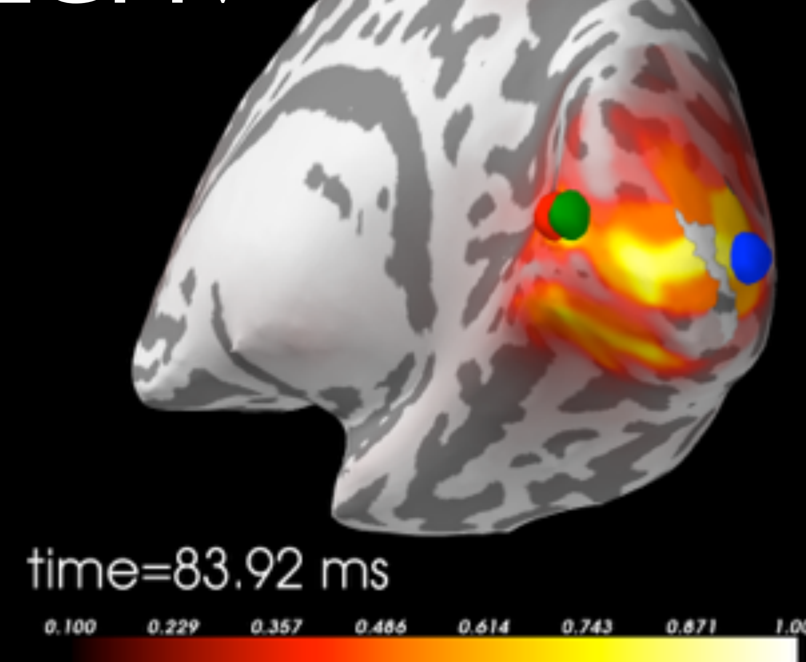
V1

V2d

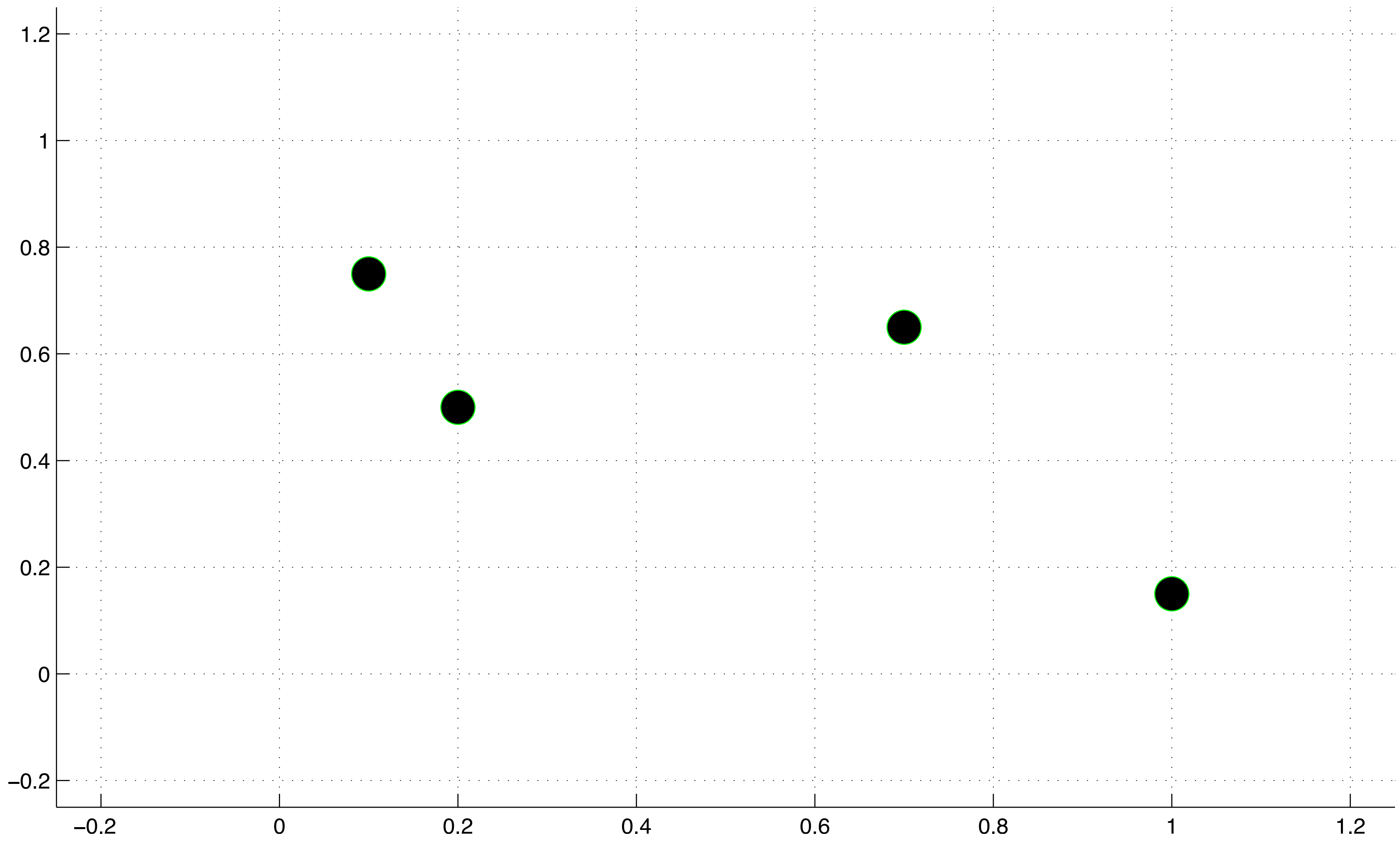
dSPM



LCMV



Motivation

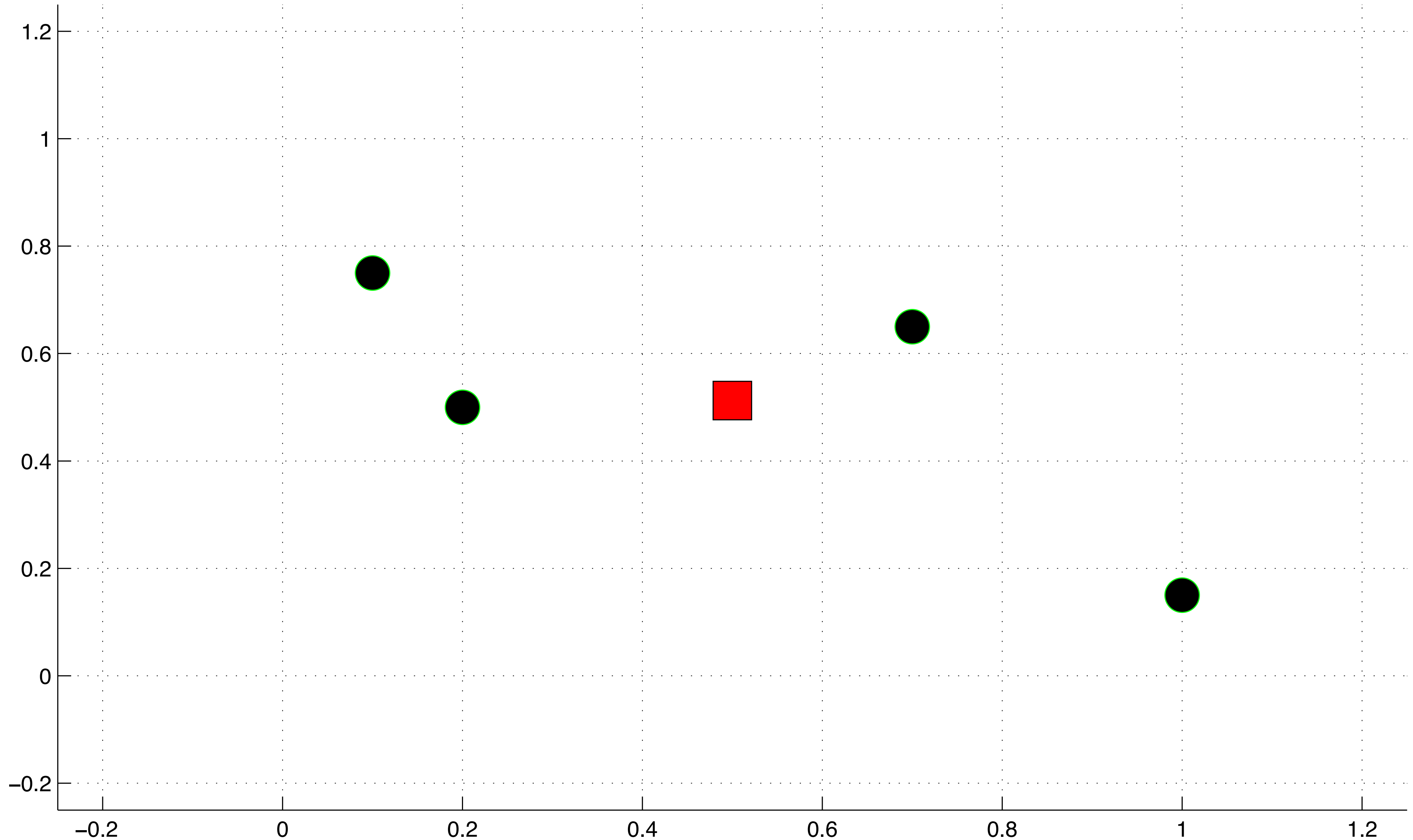


4 points in $\mathbb{R}^2$

$$x_1, x_2, x_3, x_4$$

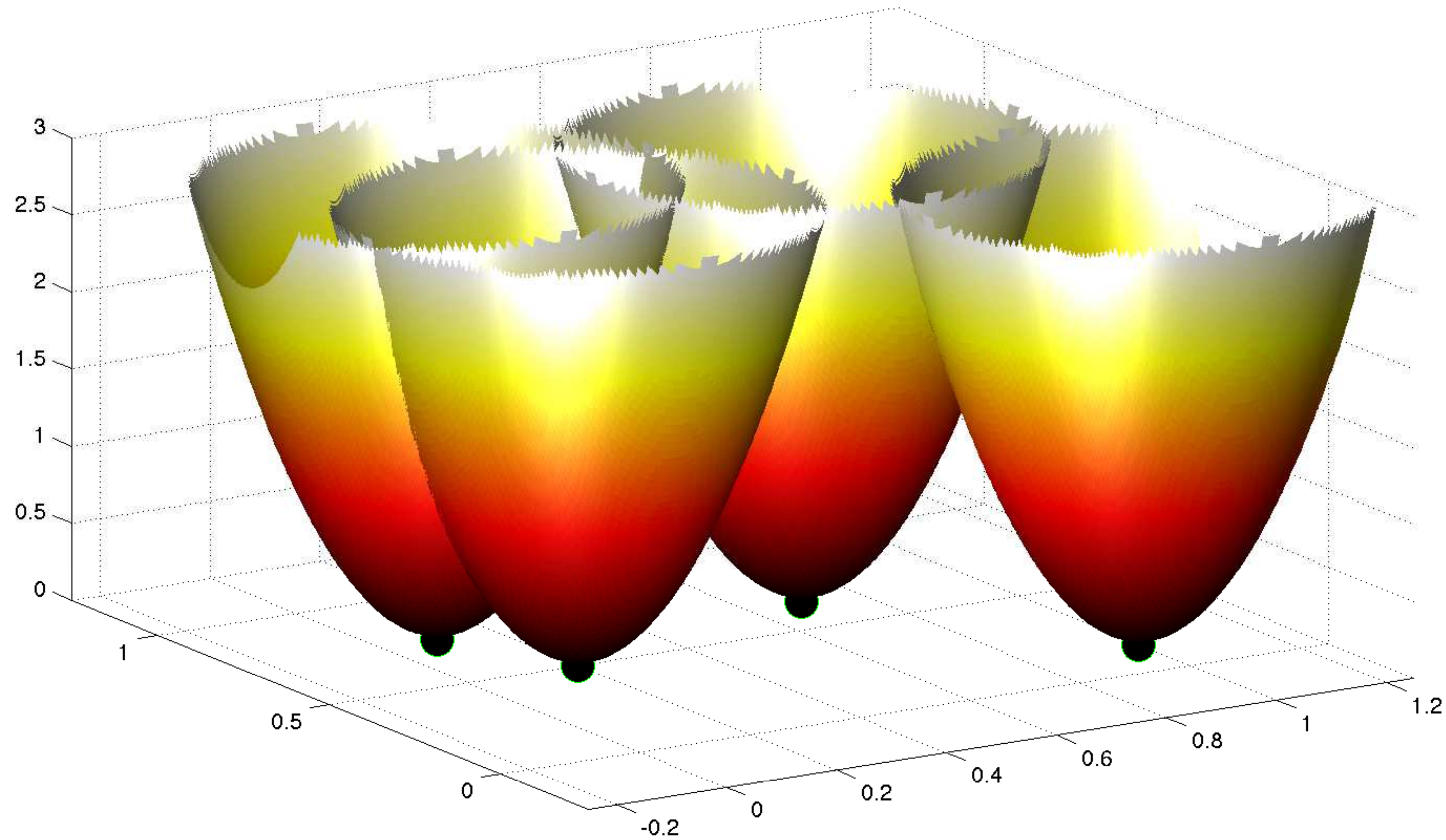
Imagine a 2D flat brain with 4 activations…

Motivation



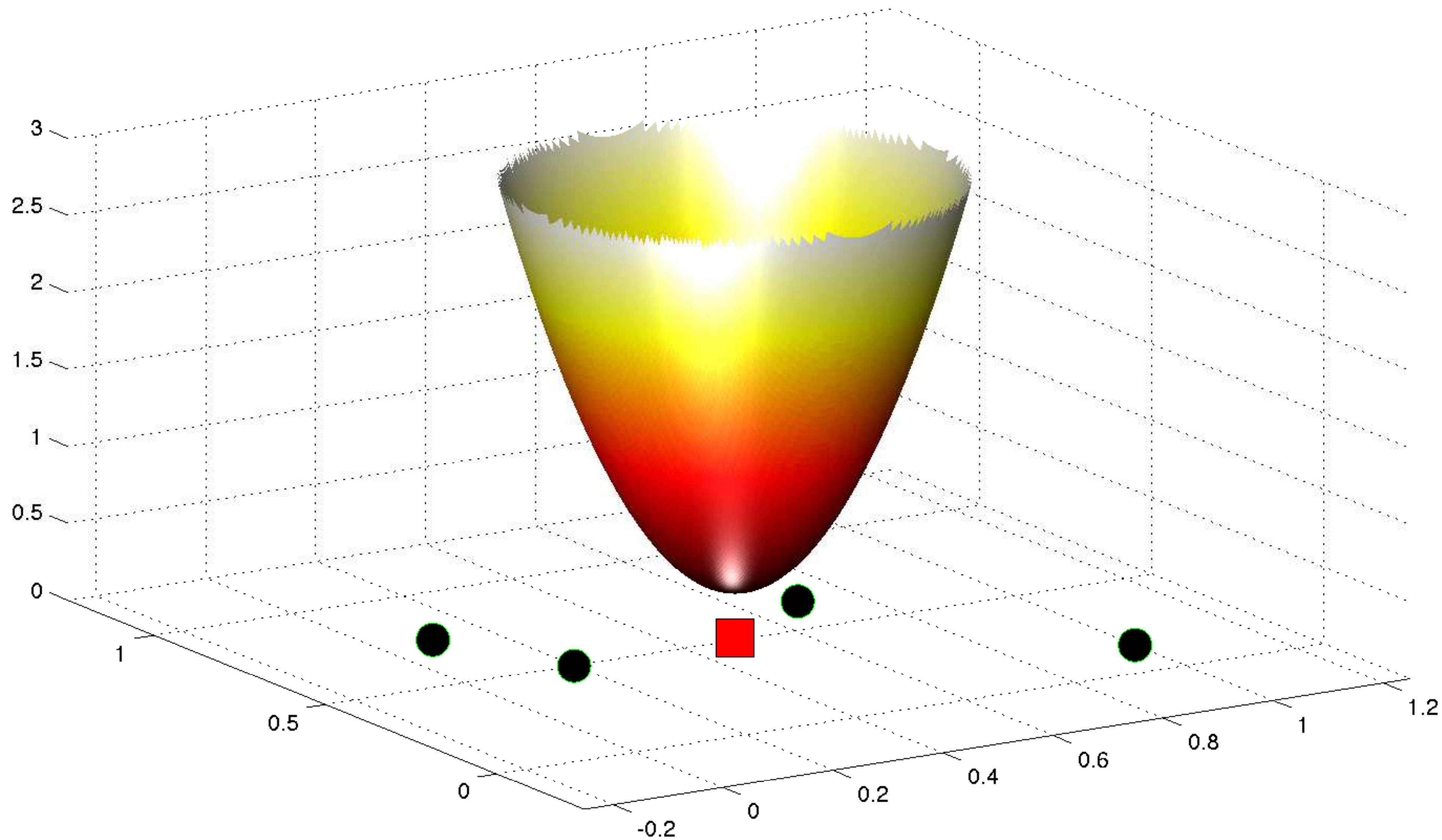
Their **mean** is $(x_1 + x_2 + x_3 + x_4)/4$.

Motivation



Consider for each point the function $\|\cdot - x_i\|_2^2$

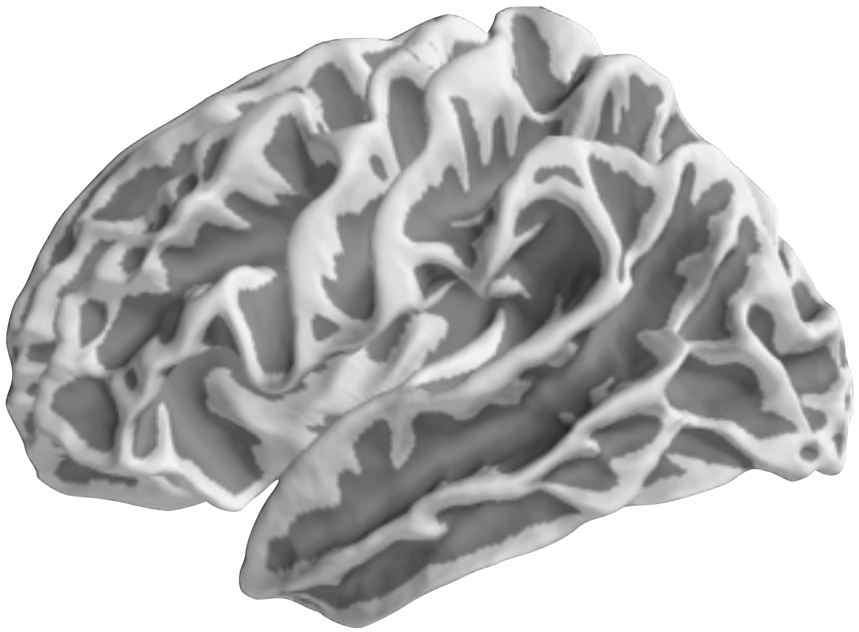
Motivation

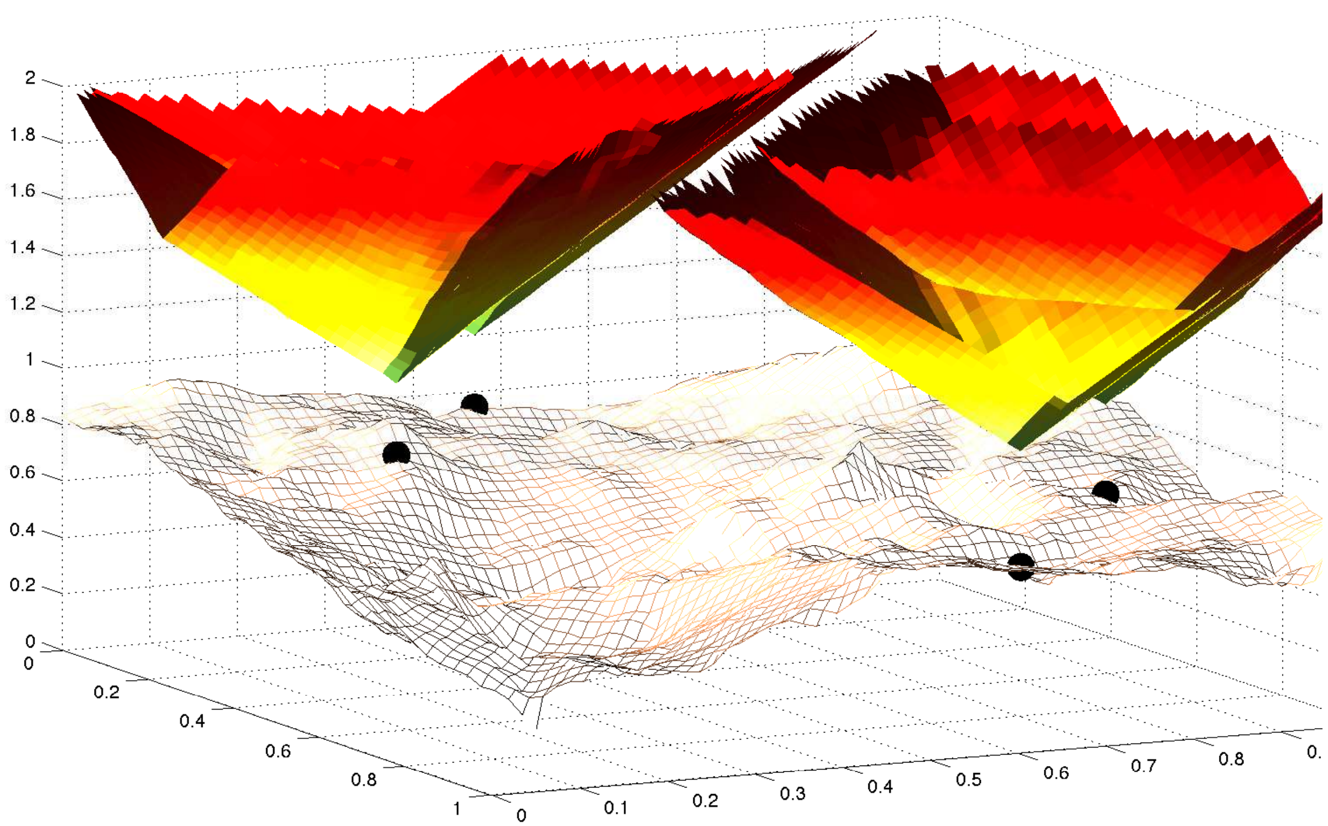


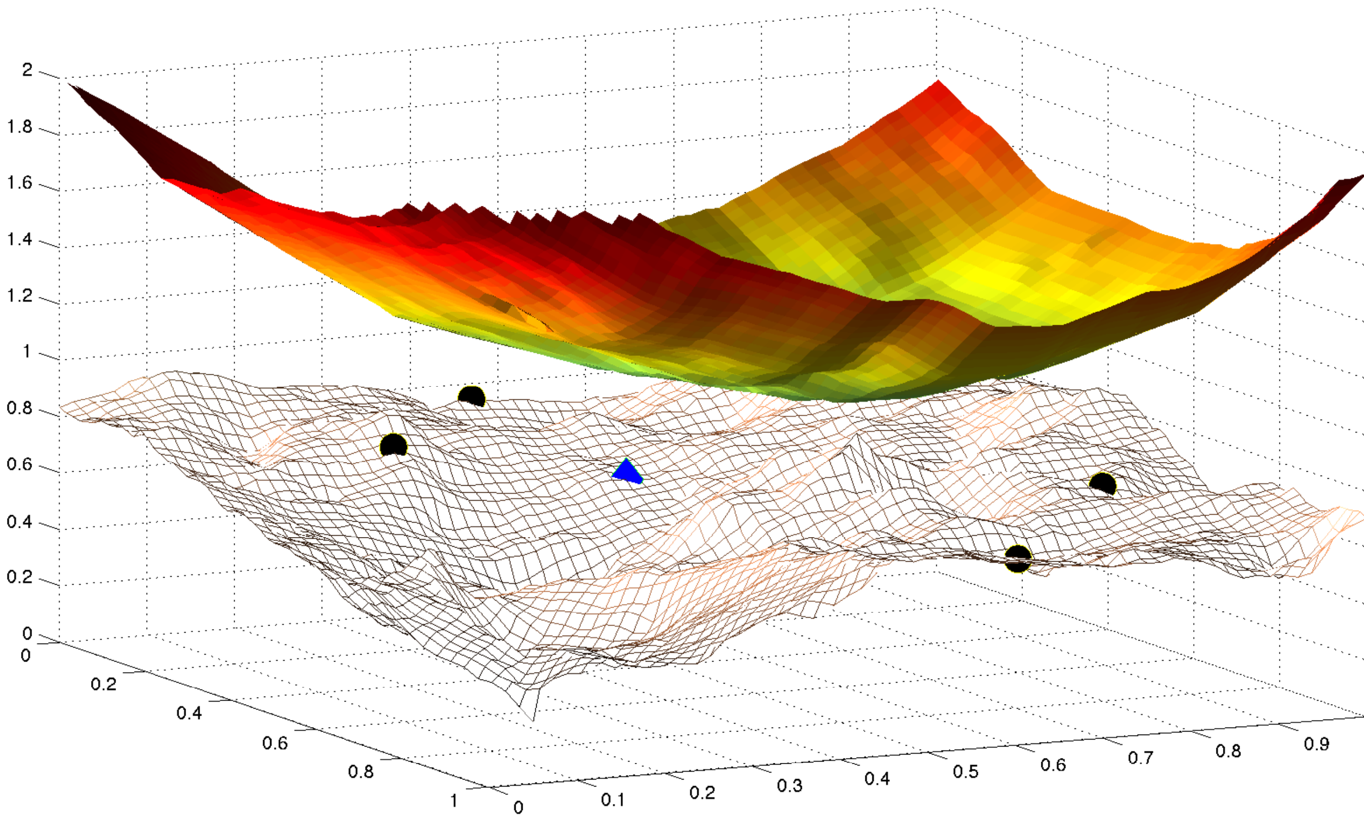
The **mean** is the $\operatorname{argmin} \frac{1}{4}\sum_{i=1}^{4} \|\cdot - x_i\|_2^2$.

Motivation

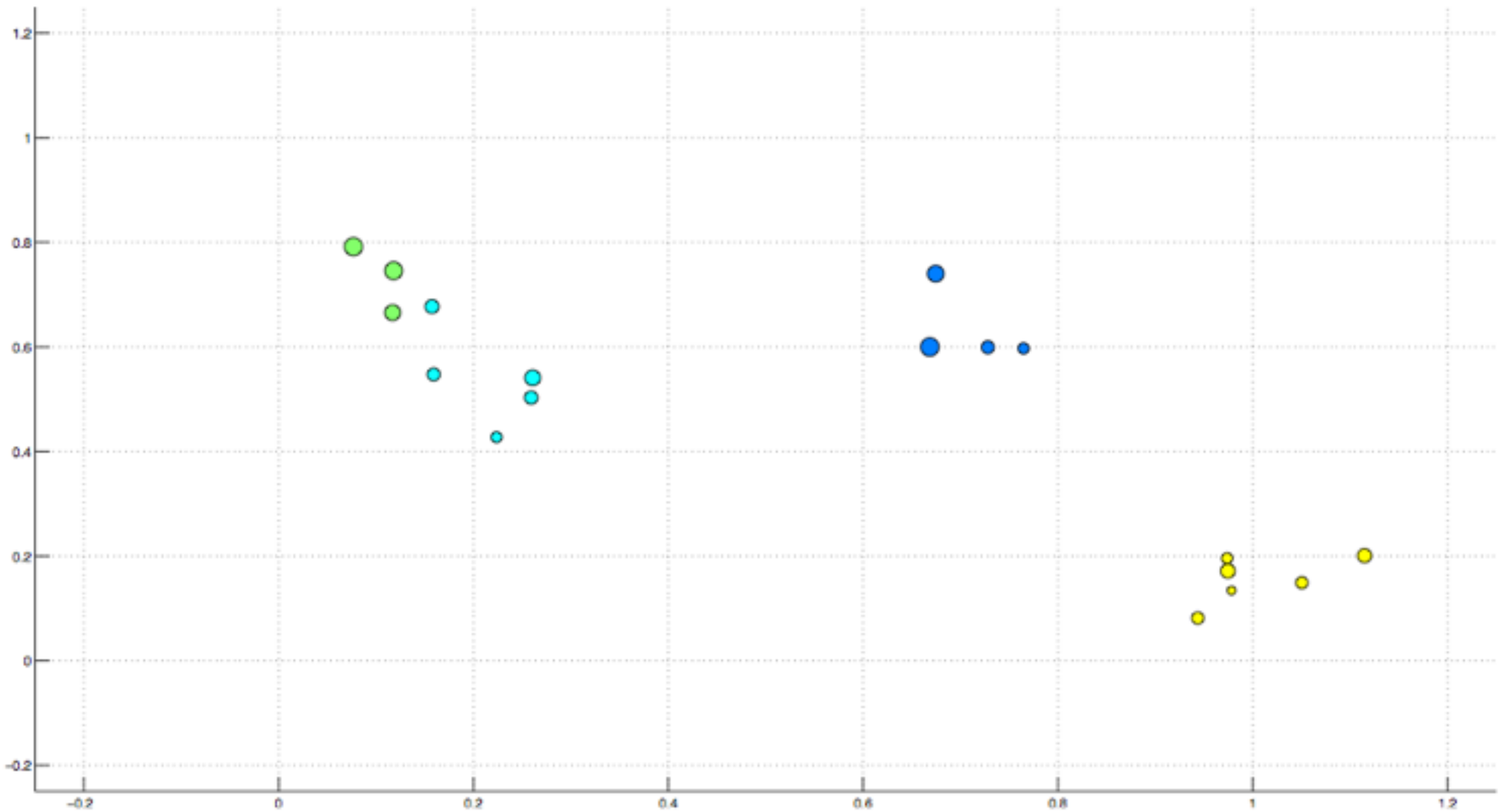
Now if the domain is not flat: you have a ground metric





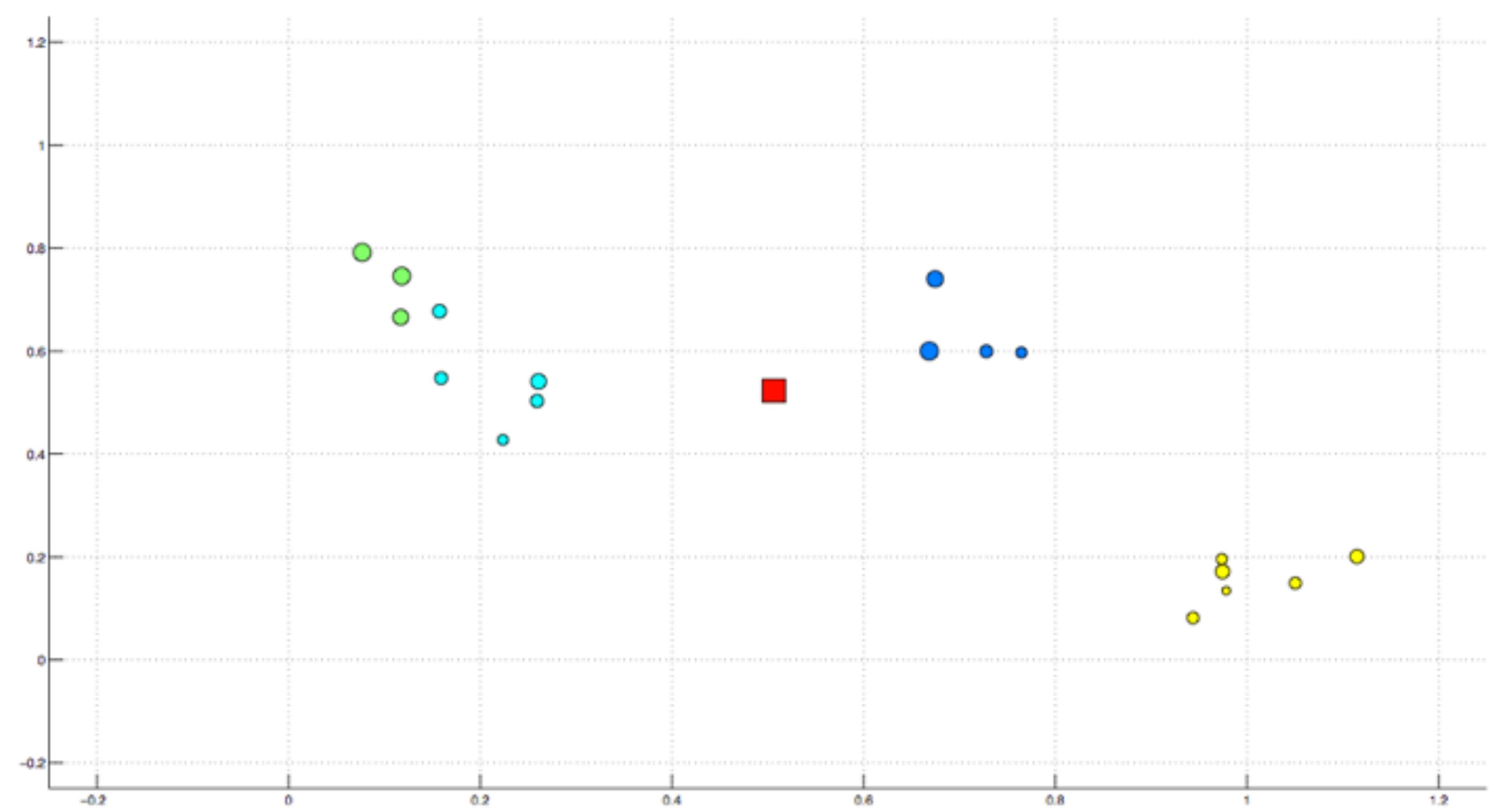


From points to probability measures



Assume that each datum is now an **empirical measure**.
What could be the mean of these 4 measures?

From points to probability measures



■ = naive mean of *all* observations.

Mean of 4 measures = a point?

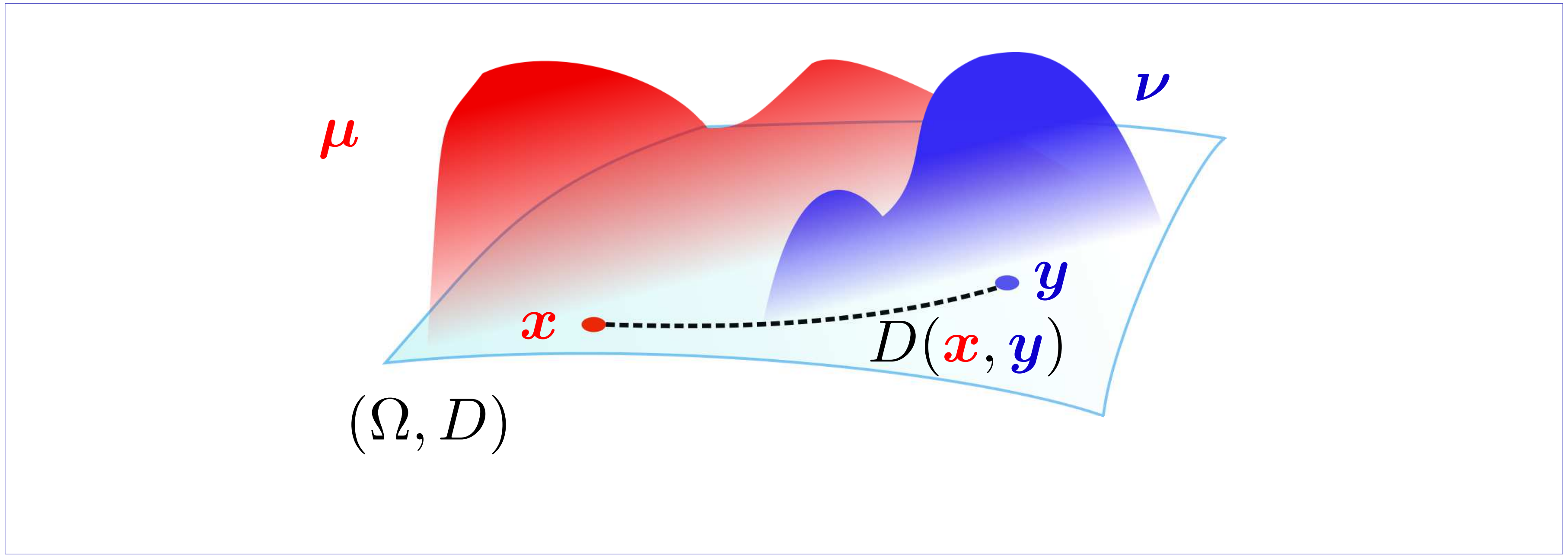
Should preserve the uncertainty
& take into account the metric

Problem formulation

Given a discrepancy function Δ between probabilities, compute **their mean:** $\operatorname{argmin} \sum_i \Delta(\cdot, \nu_i)$

Remark: If discrepancy is a squared Riemanian distance it's a Fréchet mean.

Optimal Transport

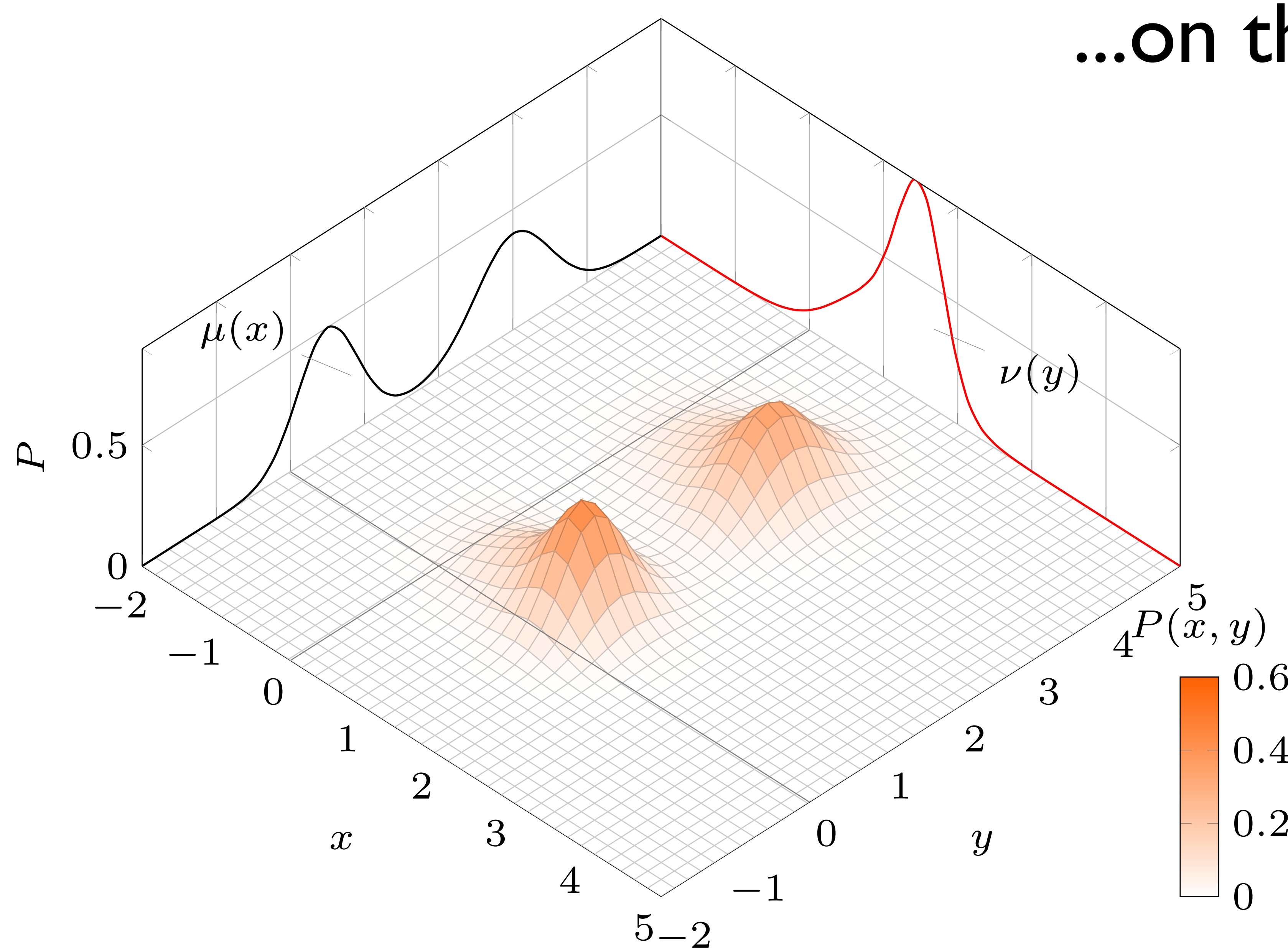


Optimal Transport distances rely **on 2 key concepts**:

- A **metric** $D : \Omega \times \Omega \to \mathbb{R}_+$;
- $\Pi(\mu, \nu)$: **joint probabilities** with marginals μ, ν.

Example of joint probabilities

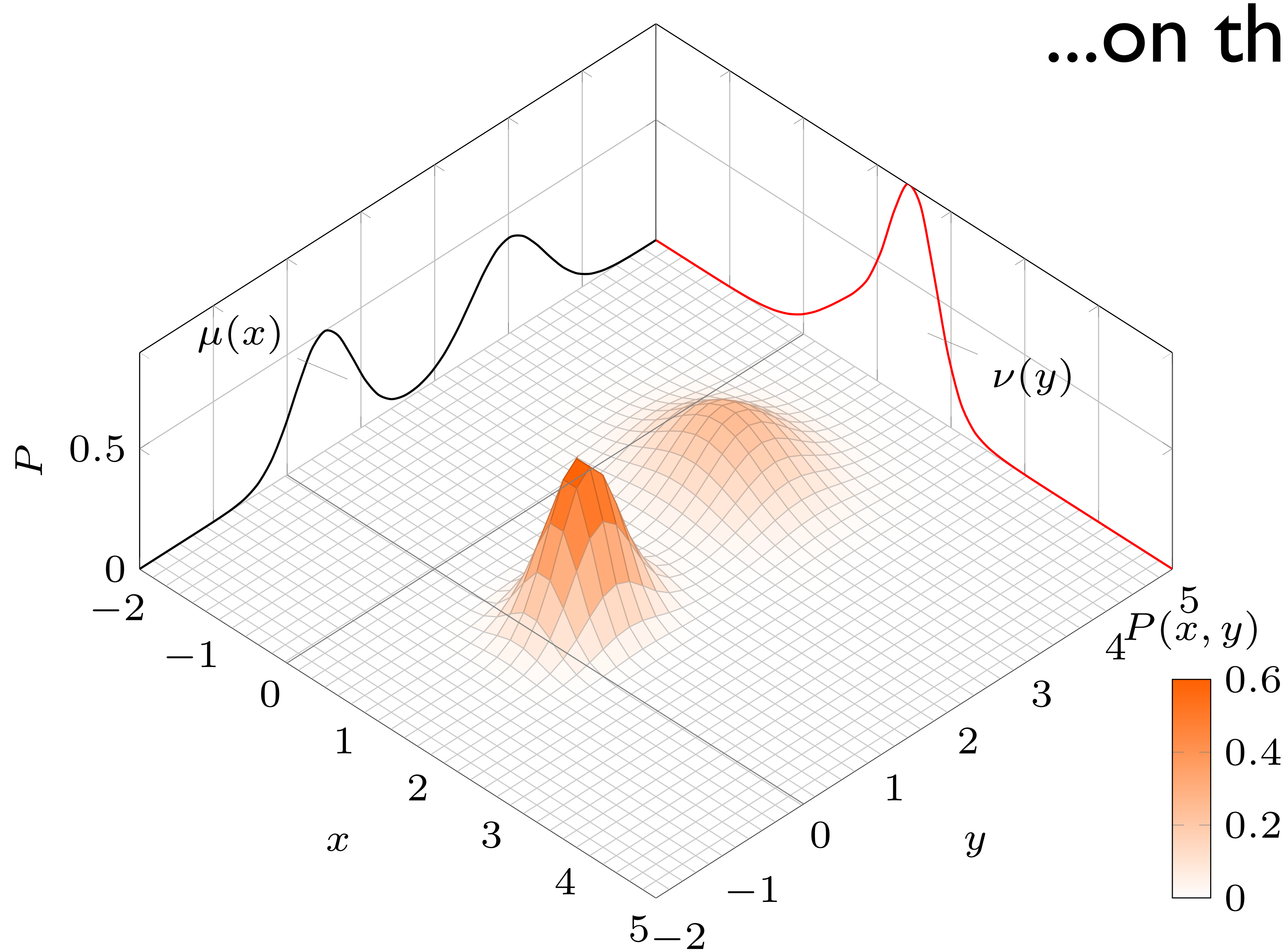
...on the real line



$\Pi(\mu, \nu) =$ probability measures on Ω^2 with marginals μ and ν.

Example of joint probabilities

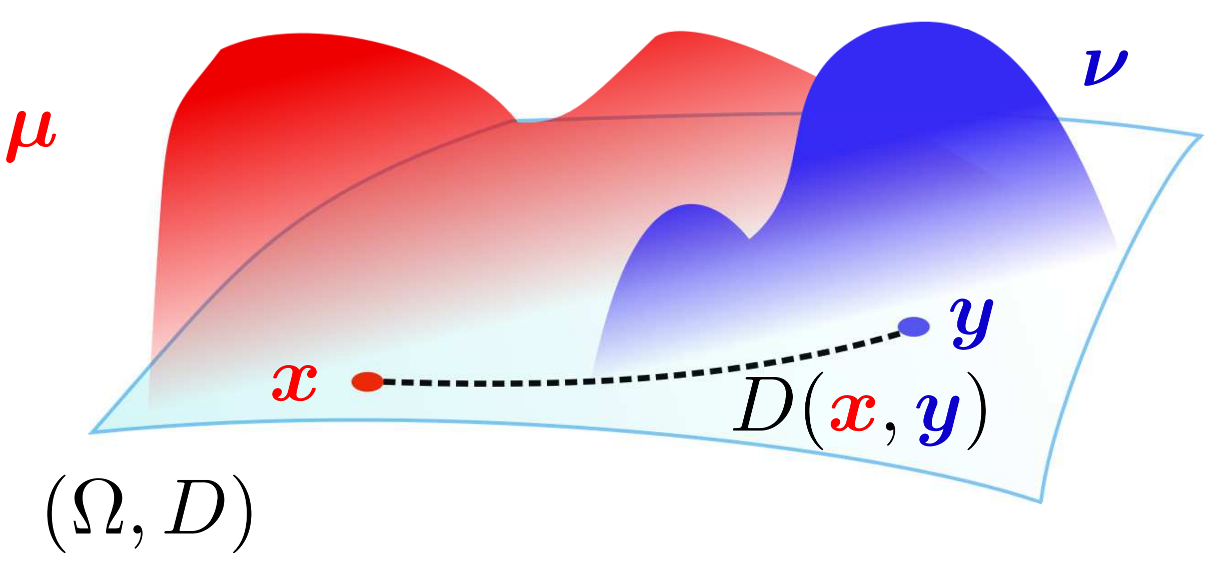
...on the real line



$$\Pi(\mu, \nu) = \text{probability measures on } \Omega^2$$

with marginals μ and ν.

Optimal Transport



p-Wasserstein distance for $p \geq 1$ is:

$$W_p(\mu, \nu) = \left(\inf_{P \in \Pi(\mu,\nu)} \int \int_{\Omega \times \Omega} D(x,y)^p dP(x,y) \right)^{1/p}.$$

[Monge-Kantorovich, Kantorovich-Rubinstein, Wasserstein, Earth Mover's Distance, Mallows ...]

Optimal Transport in dimension d

$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu})$ can be cast as a linear program

1. $M_{\boldsymbol{XY}}^p \stackrel{\text{def}}{=} [D(\boldsymbol{x_i}, \boldsymbol{y_j})^p]_{ij} \in \mathbb{R}_+^{d\times d}$ ***(metric information)***

2. Transportation Polytope ***(joint probabilities)***

$$U(a,b) \stackrel{\text{def}}{=} \{T \in \mathbb{R}_+^{d\times d} \mid T\mathbf{1}_d = a,\, T^T\mathbf{1}_d = b\}.$$

Example:

$$T = \begin{bmatrix} .1 & 0 & .1 \\ .1 & 0 & .1 \\ .2 & .1 & .3 \end{bmatrix} \in U\left(\begin{bmatrix} .2 \\ .2 \\ .6 \end{bmatrix}, \begin{bmatrix} .4 \\ .1 \\ .5 \end{bmatrix}\right)$$

Optimal Transport in dimension d

$W_p^p(\boldsymbol{\mu}, \boldsymbol{\nu})$ can be cast as a linear program

Optimal transport problem reads:

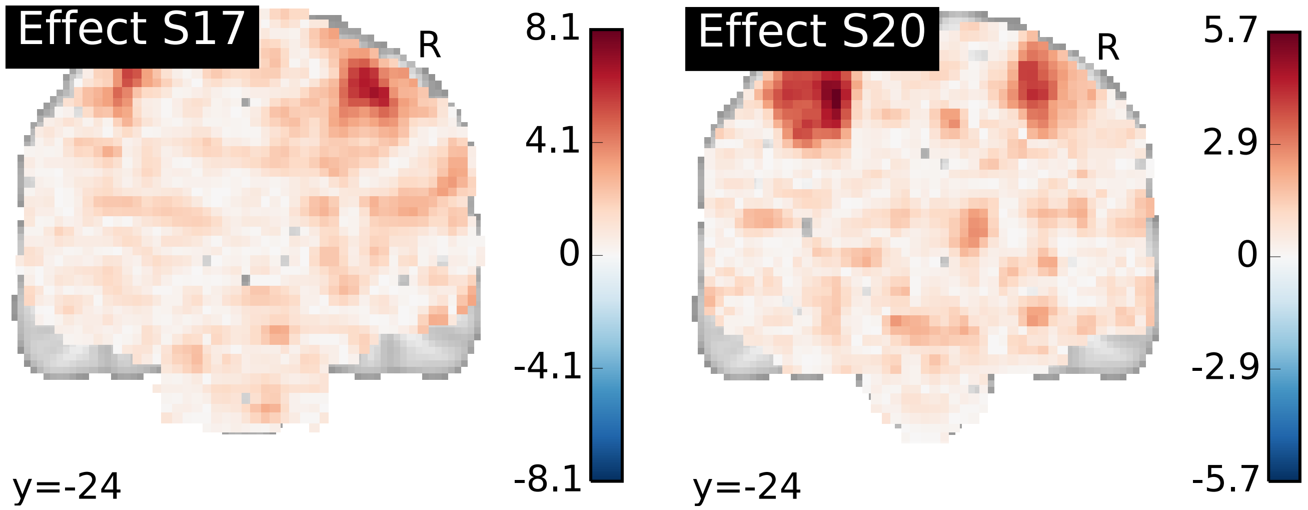
$$W_p^p(a, b) = \mathbf{OT}(a, b, M^p) \stackrel{\text{def}}{=} \min_{T \in U(a,b)} \langle T, M^p \rangle$$

$$\langle T, M^p \rangle = \sum_{i=1}^{d} \sum_{j=1}^{d} T_{ij} M_{ij}^p$$

T is the transport plan

Problem: No solution if

$$|a|_1 = \sum_{i=1}^{d} |a_i| \neq |b|_1$$



Need to add and remove mass

non-negative and non-normalized data

Add a virtual point ω whose distance to element i in Ω

$$D(i,\omega) = D(\omega,i) = \Delta_i$$

Scale each observation $b^j, 1 \leq j \leq n$ so that $|b^j|_1 \leq 1$

Map each a to $[a, |a|_1 - 1]$ (Kind of feature map)

Use as metric $\hat{M} = \begin{bmatrix} M & \Delta \\ \Delta^T & 0 \end{bmatrix} \in \mathbb{R}_+^{d+1 \times d+1}$

$$\operatorname*{argmin}_{u \in S_d} \frac{1}{N} \sum_{j=1}^{N} \mathbf{OT}\left(\left[\begin{smallmatrix} u \\ 1-|u|_1 \end{smallmatrix}\right], \left[\begin{smallmatrix} b^j \\ \beta^j \end{smallmatrix}\right], \hat{M}^p\right).$$

$S_d = \{u \in \mathbb{R}_+^d, |u|_1 \leq 1\}$ ... but a huge linear program

Smoothing to speed things up

Idea: Regularize cost with entropy *[Cuturi NIPS 2013]*

$$\mathbf{OT}_\lambda(a, b, M^p) \overset{\text{def}}{=} \min_{T \in U(a,b)} \langle T, M^p \rangle - \frac{1}{\lambda} H(T).$$

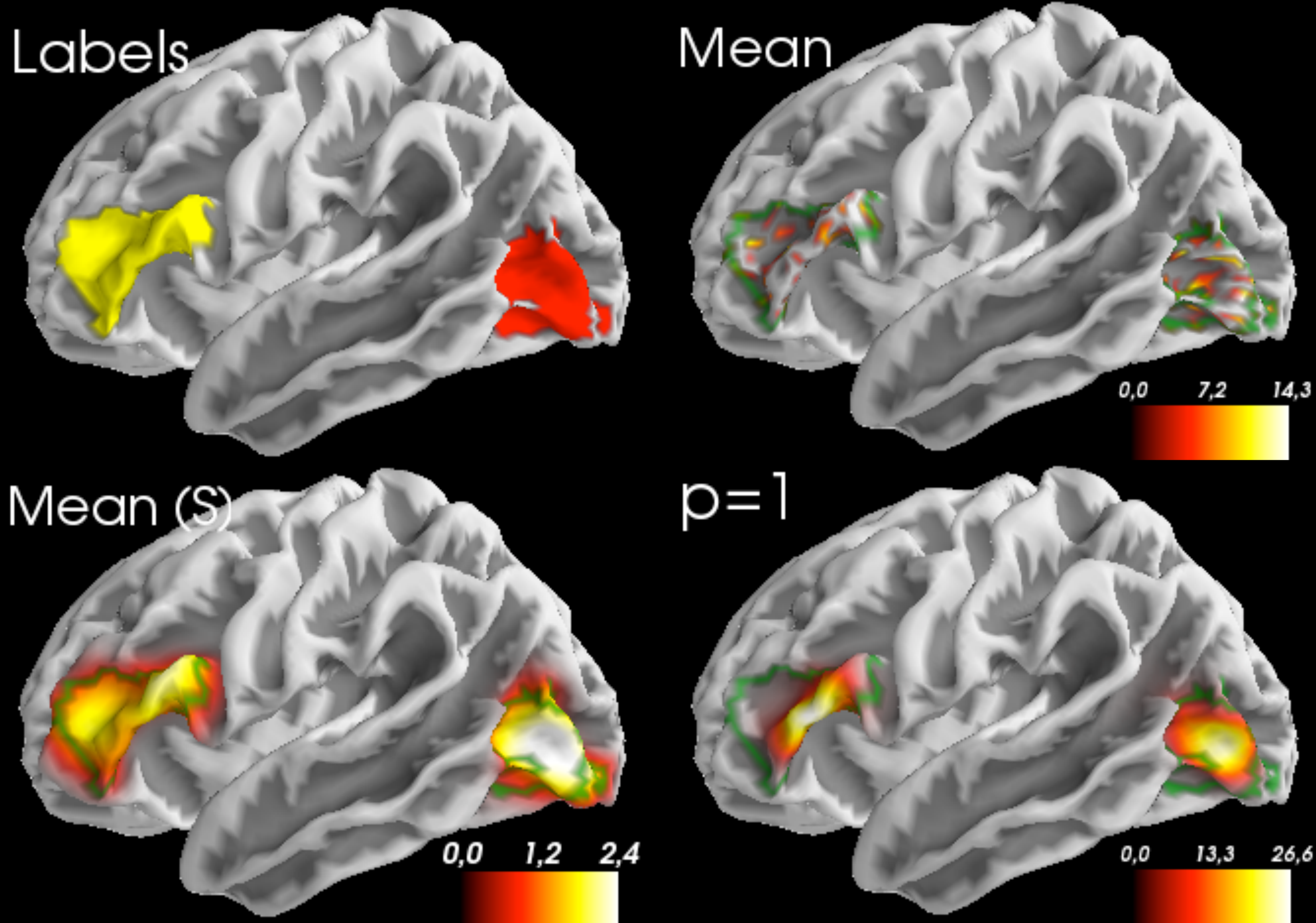
Strongly convex with unique minimum

Problem reads:

$$\underset{\substack{a \in S_d \\ |a|_1 = \rho}}{\operatorname{argmin}} \frac{1}{N} \sum_j \mathbf{OT}_\lambda(a, b^j, \hat{M}^p)$$

In practice: solved with an exponentiated gradient with projection in the dual (matrix-matrix computations and element wise multiplications which are GPGPU friendly)

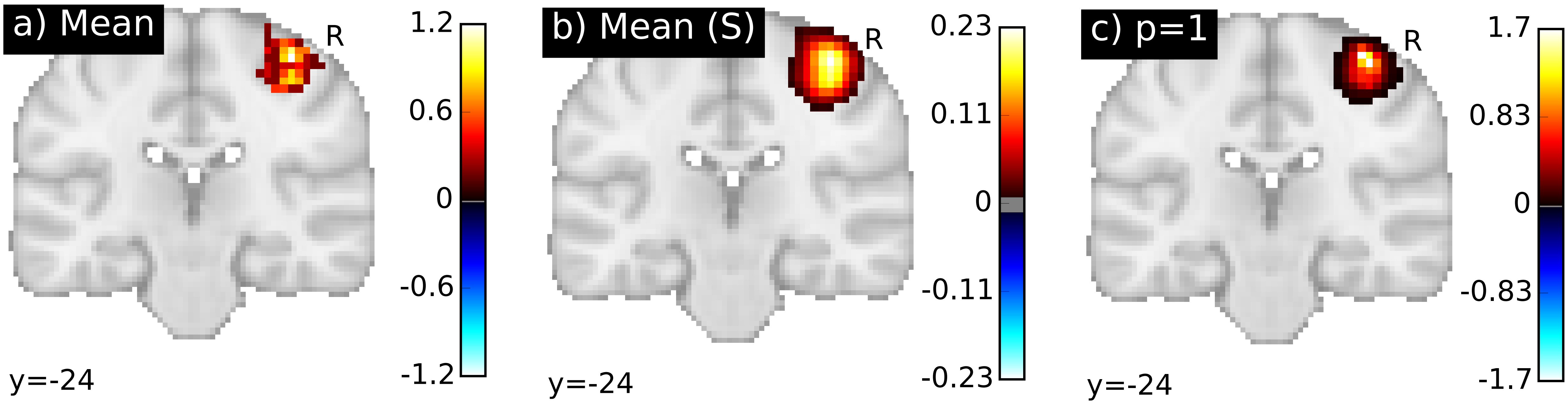
BA45 MT $n = 100 \quad d = 10242$



Results fMRI

- 20 subjects
- Left hand button press
- Averaging of standardized effect size

[Pinel et al. 2007]



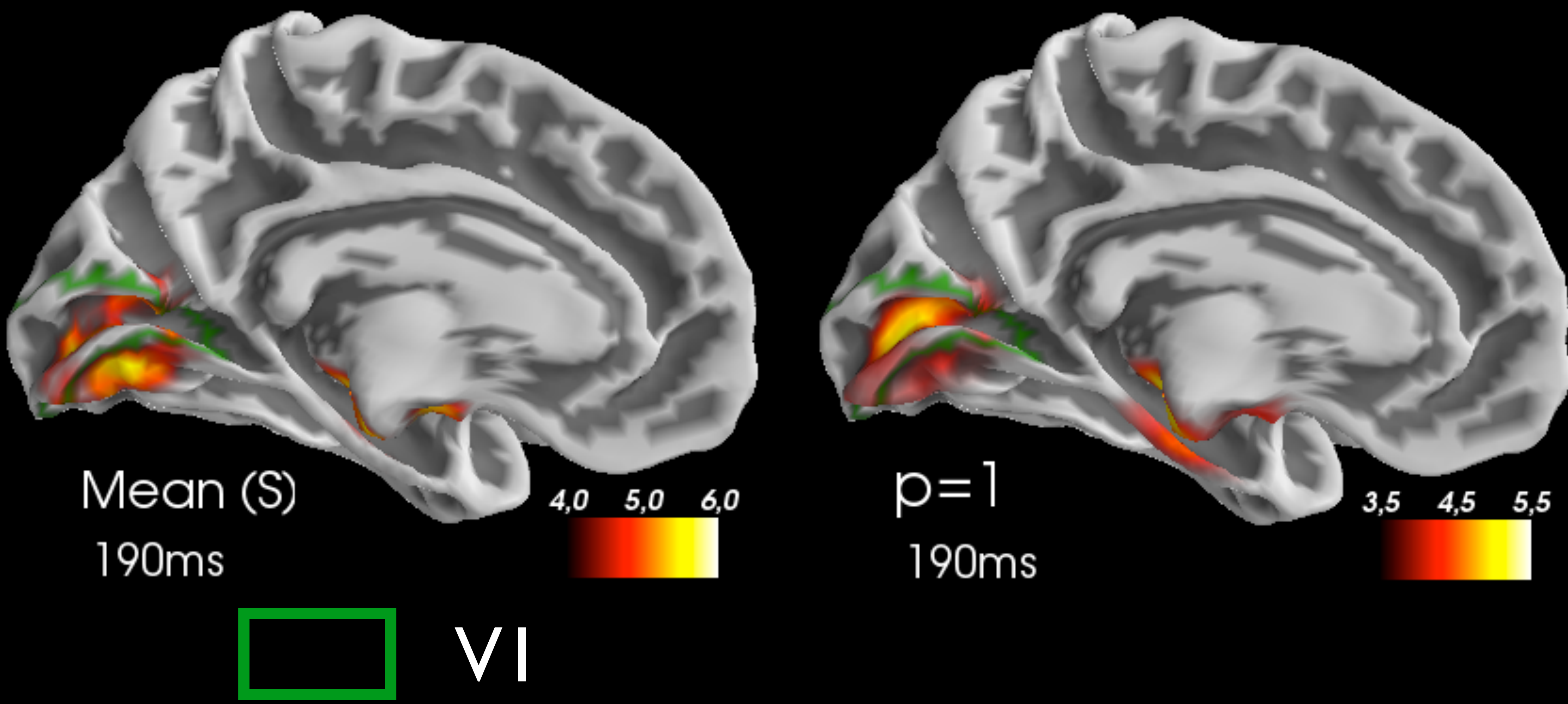
Sharp activation foci & less amplitude reduction

Results MEG

- 16 subjects

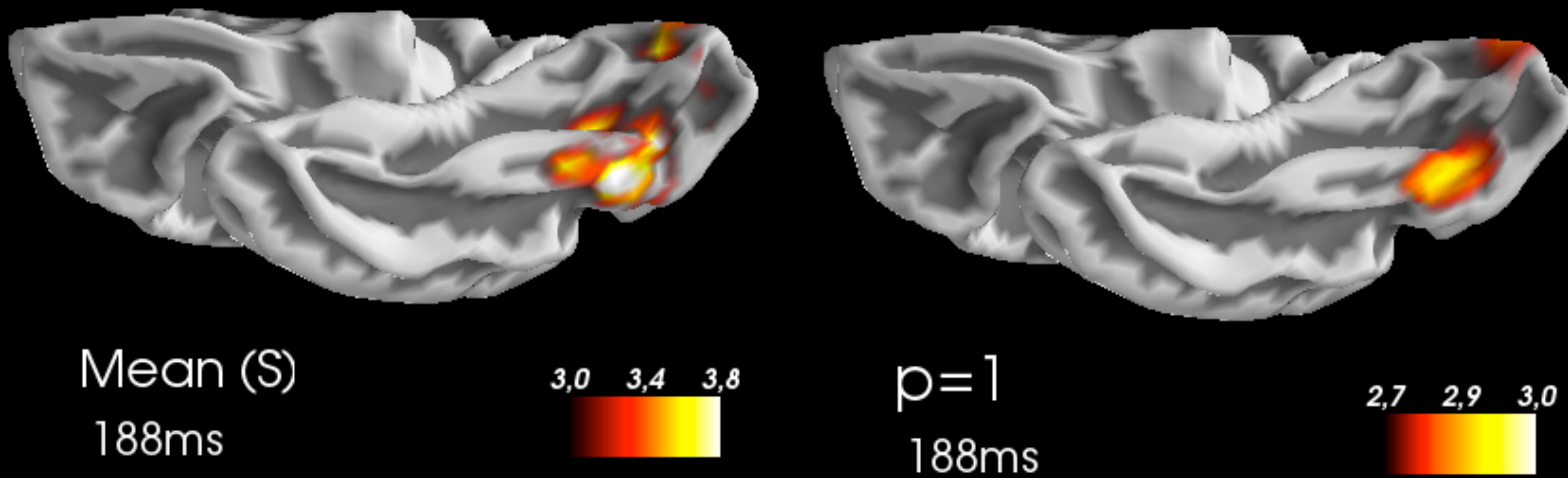
[Henson et al. 2011]

- Visual presentation of faces and scrambled faces
- Averaging of dSPM source estimates



Results MEG

- Contrast between faces and scrambled faces



With Tesla K40 GPU card (< a minute of computation)

“Philosophical” Conclusion

- The world of neuroimaging is full of challenging maths and computer science problems ...
- … look at the data to find the relevant ones
- … but don’t be scared if they are not well posed

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. ~ John Tukey"

Some refs:

Fabian Pedregosa, Michael Eickenberg, Philippe Ciuciu, Bertrand Thirion and Alexandre Gramfort, *Data-driven HRF estimation for encoding and decoding models,* Neuroimage 2015

Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, Bertrand Thirion, Seeing it all: Convolutional network layers map the function of the human visual system, Neuroimage 2016

Alexandre Gramfort, Gabriel Peyré, Marco Cuturi, Fast Optimal Transport Averaging of Neuroimaging Data, Proc. IPMI 2015

scikit learn

1 position to work on scikit-learn available !

Contact

http://alexandre.gramfort.net

GitHub : @agramfort *Twitter : @agramfort*