GLMsingle / denoising, ridge regression

presenter: Denis Schluppeck, 2025-03-18

eLife

TOOLS AND RESOURCES

Improving the accuracy of single-trial fMRI response estimates using GLMsingle

Jacob S Prince[1]*, Ian Charest[2,3], Jan W Kurzawski[4], John A Pyles[5], Michael J Tarr[6], Kendrick N Kay[7]

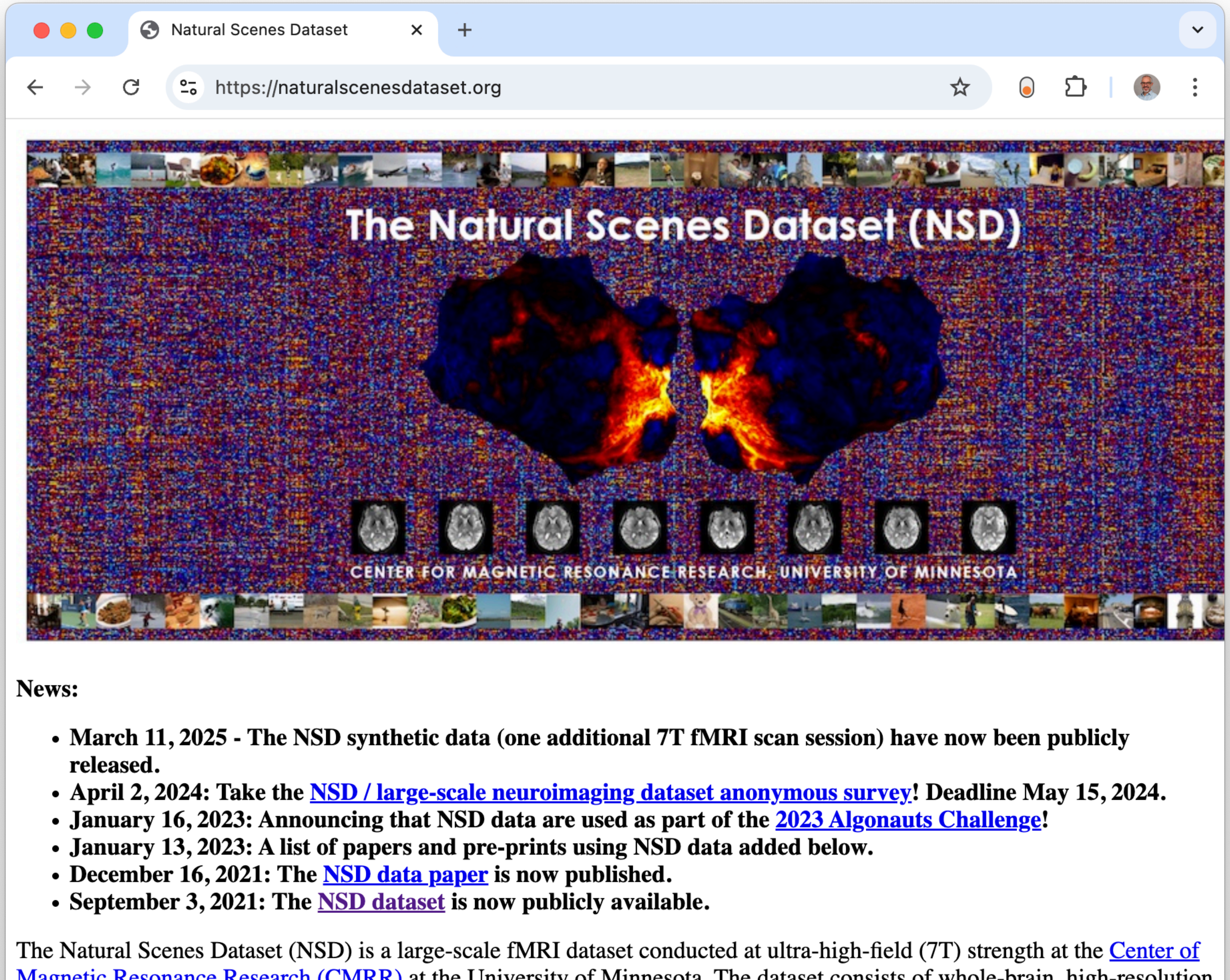
[1]Department of Psychology, Harvard University, Cambridge, United States; [2]Center for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, United Kingdom; [3]cerebrUM, Département de Psychologie, Université de Montréal, Montréal, Canada; [4]Department of Psychology, New York University, New York, United States; [5]Center for Human Neuroscience, Department of Psychology, University of Washington, Seattle, United States; [6]Department of Psychology, Neuroscience Institute, Carnegie Mellon University, Pittsburgh, United States; [7]Center for Magnetic Resonance Research (CMRR), Department of Radiology, University of Minnesota, Minneapolis, United States

NG data club
journal club edition

context

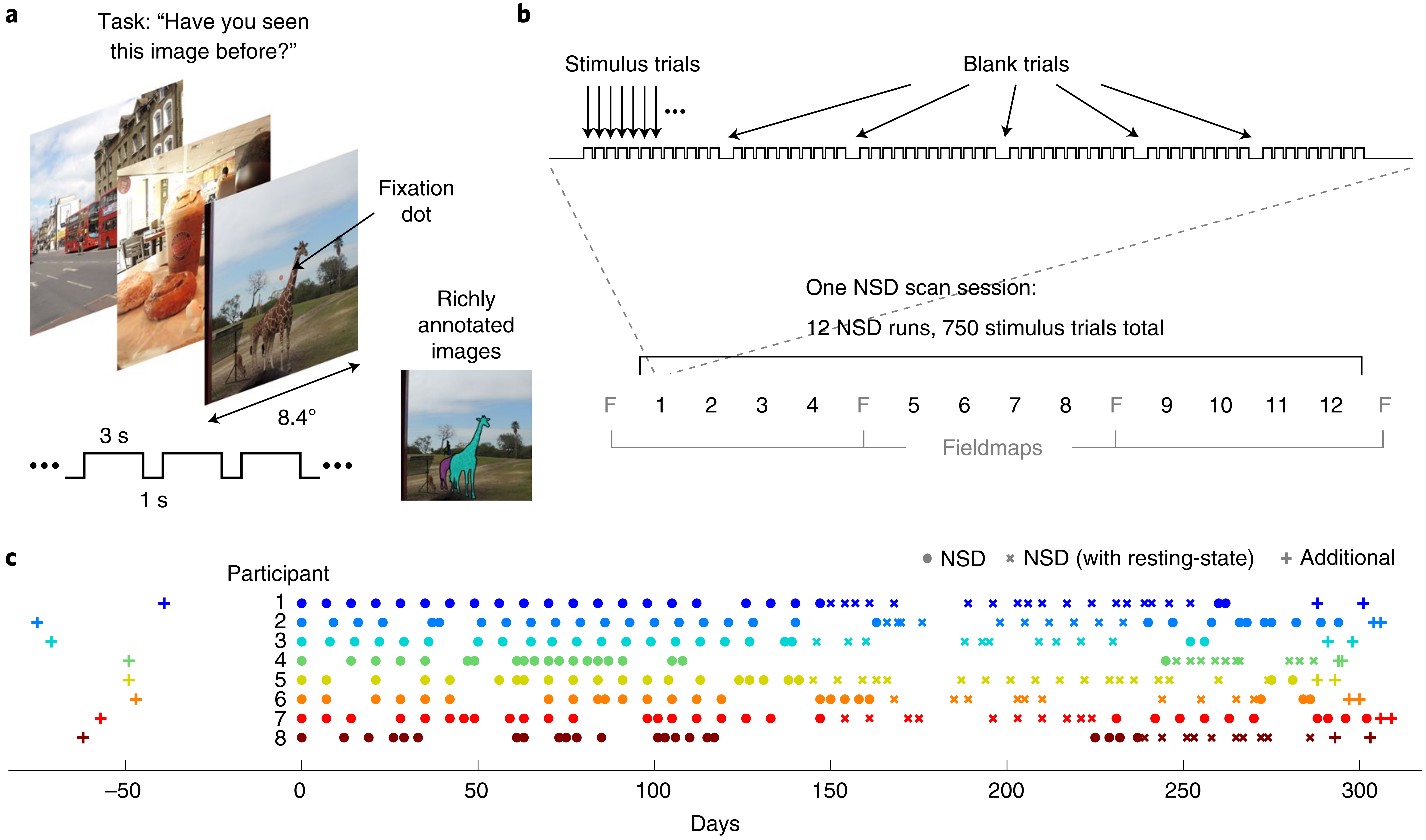
- quick intro: what does this relate to
  (fMRI, vision, objects, …) ?
- what's the proposal here
  (3 distinct ideas)
- is any of this relevant to other domains
  (discussion)

Background



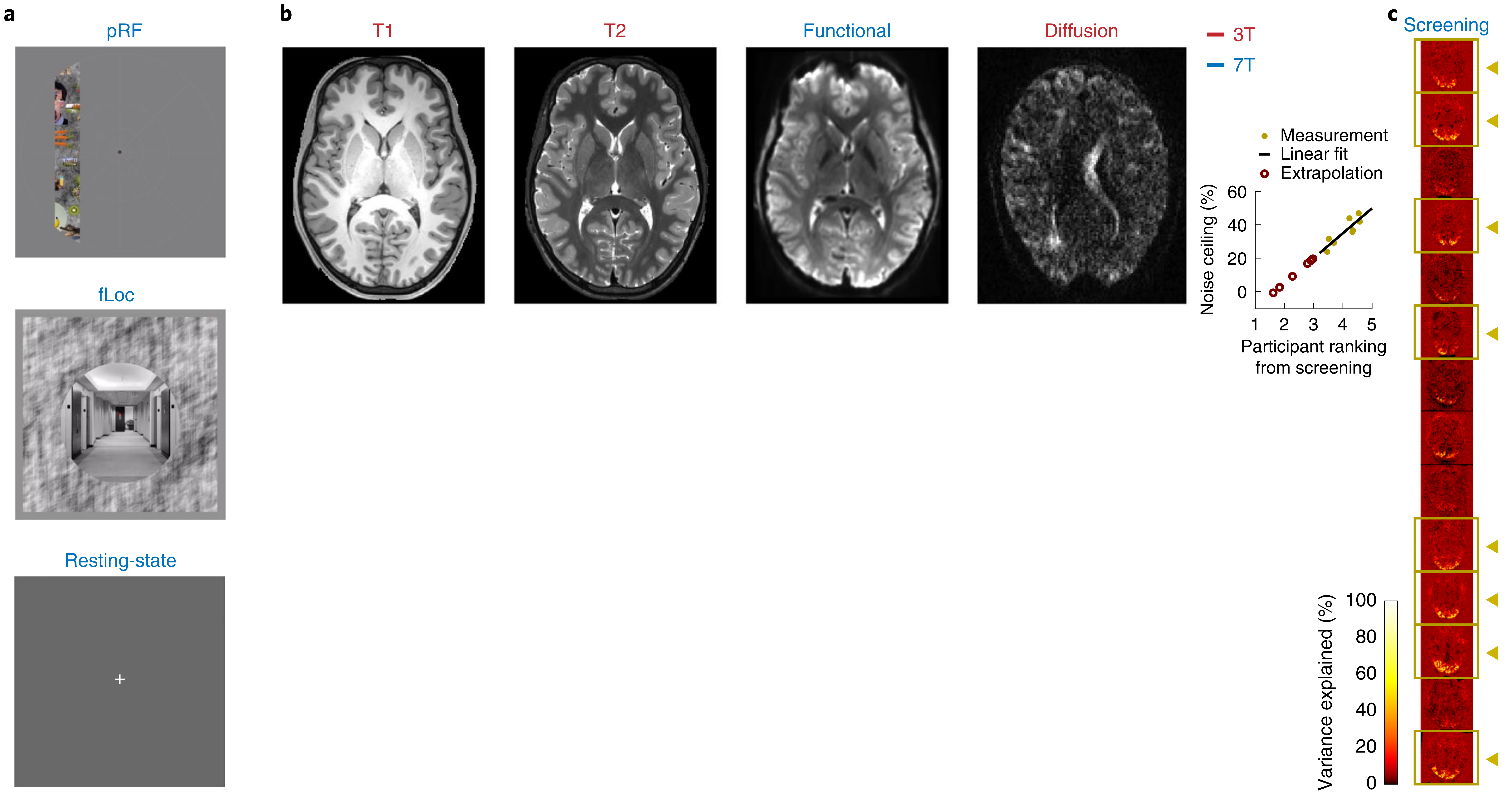
- NSD data set
- **eight participants** × 9k unique images + 1k shared images = **73k images**

subjects in 7T scanner

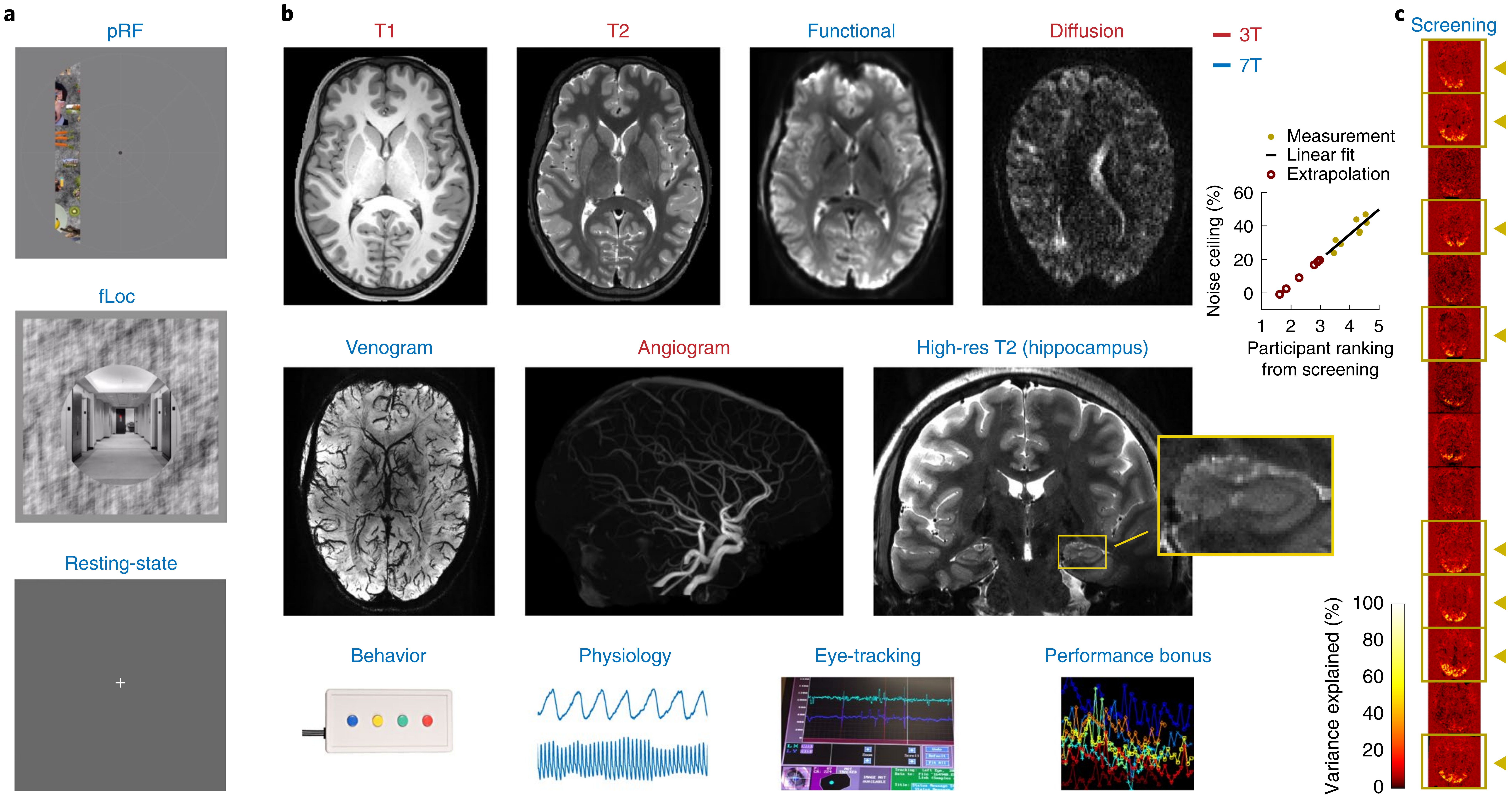


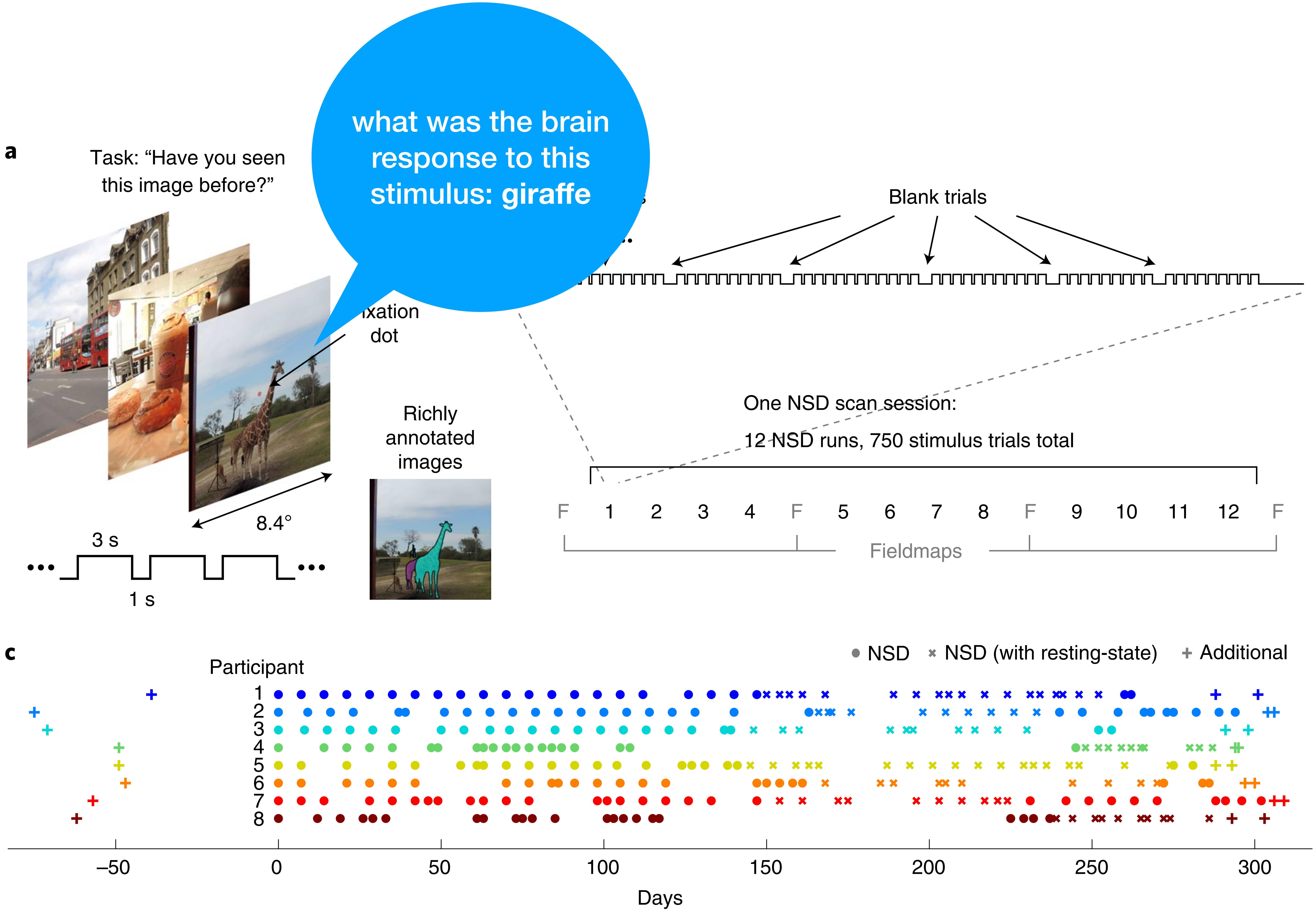
lots of data per participant

lots of data per participant

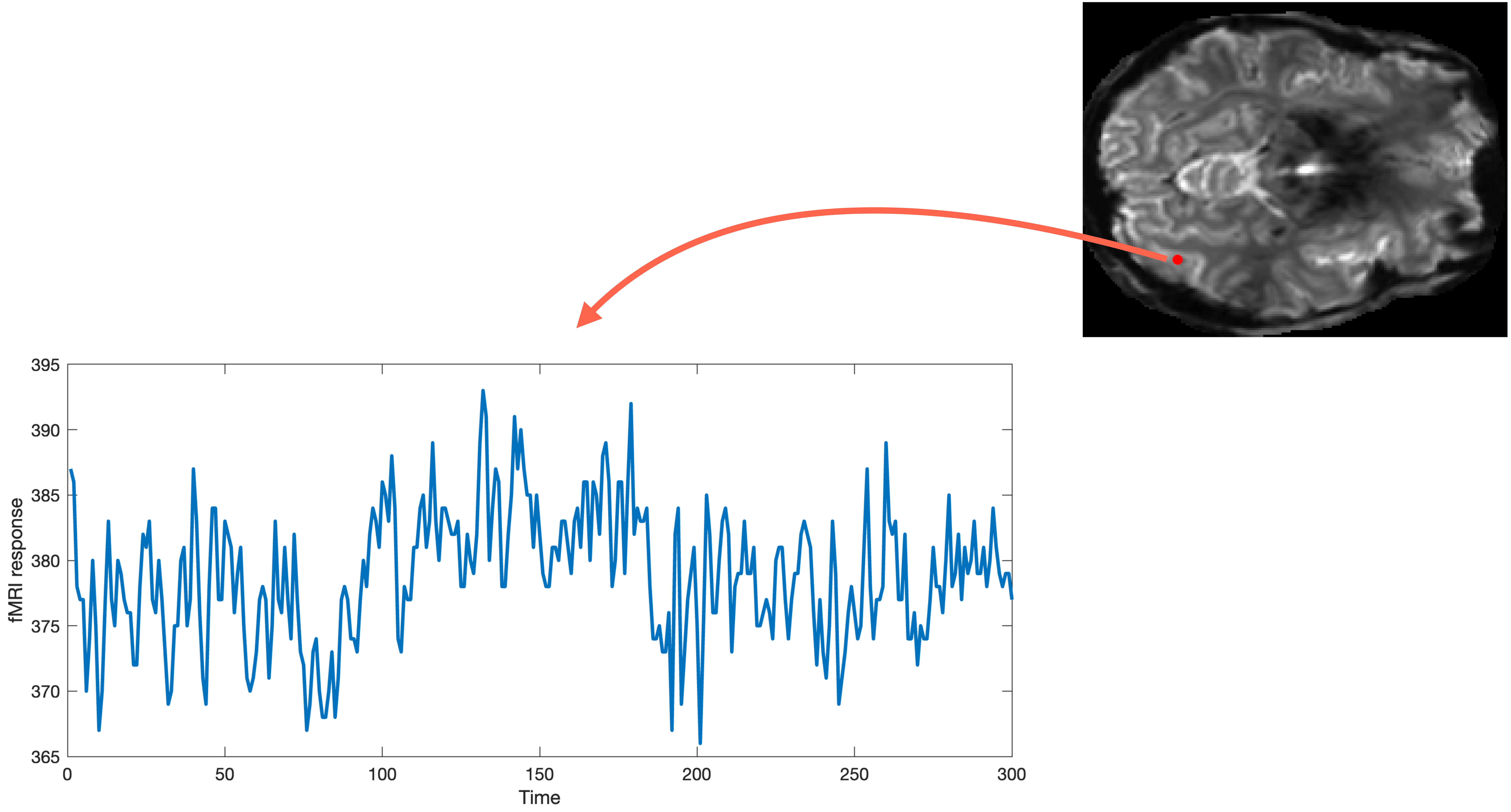


lots of data per participant





what was the brain response to this stimulus: giraffe
a
Task: “Have you seen this image before?”
Blank trials
ixation dot
Richly annotated images
8.4°
3 s
1 s
One NSD scan session:
12 NSD runs, 750 stimulus trials total
F 1 2 3 4 F 5 6 7 8 F 9 10 11 12 F
Fieldmaps
c
Participant
NSD
NSD (with resting-state)
Additional
1
2
3
4
5
6
7
8
–50
0
50
100
150
200
250
300
Days

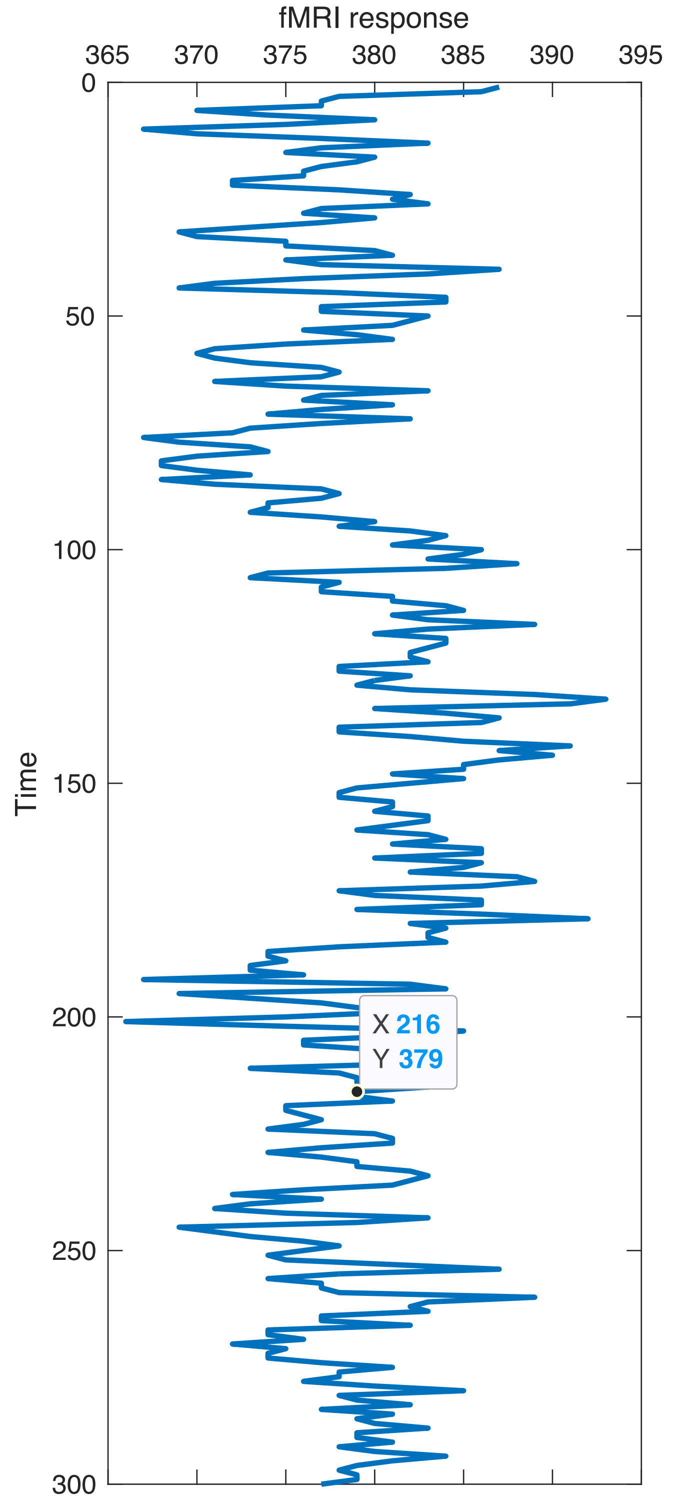

fMRI response
395
390
385
380
375
370
365
0
50
100
150
200
250
300
Time

enter: the GLM

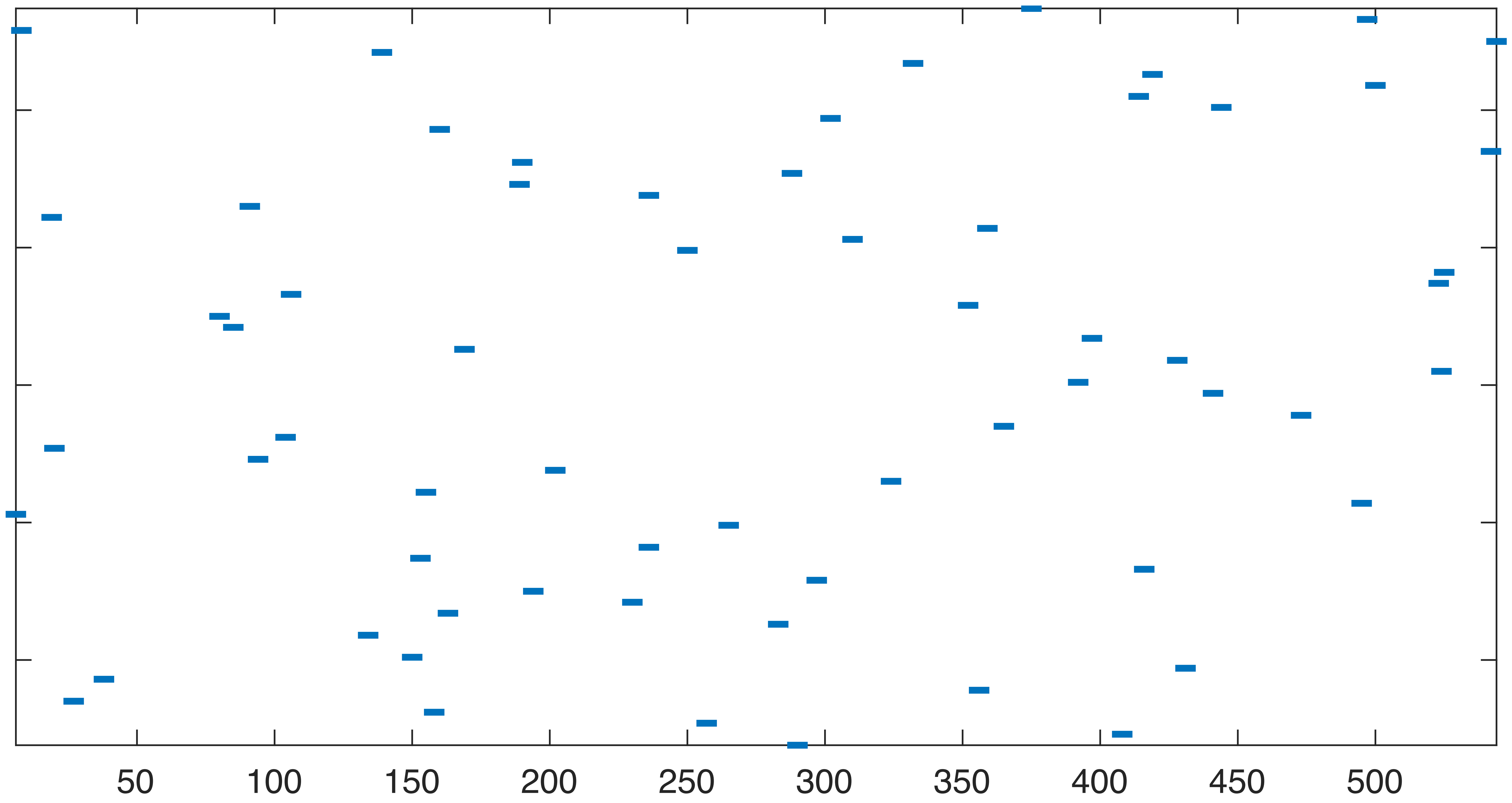
$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

enter: the GLM

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$



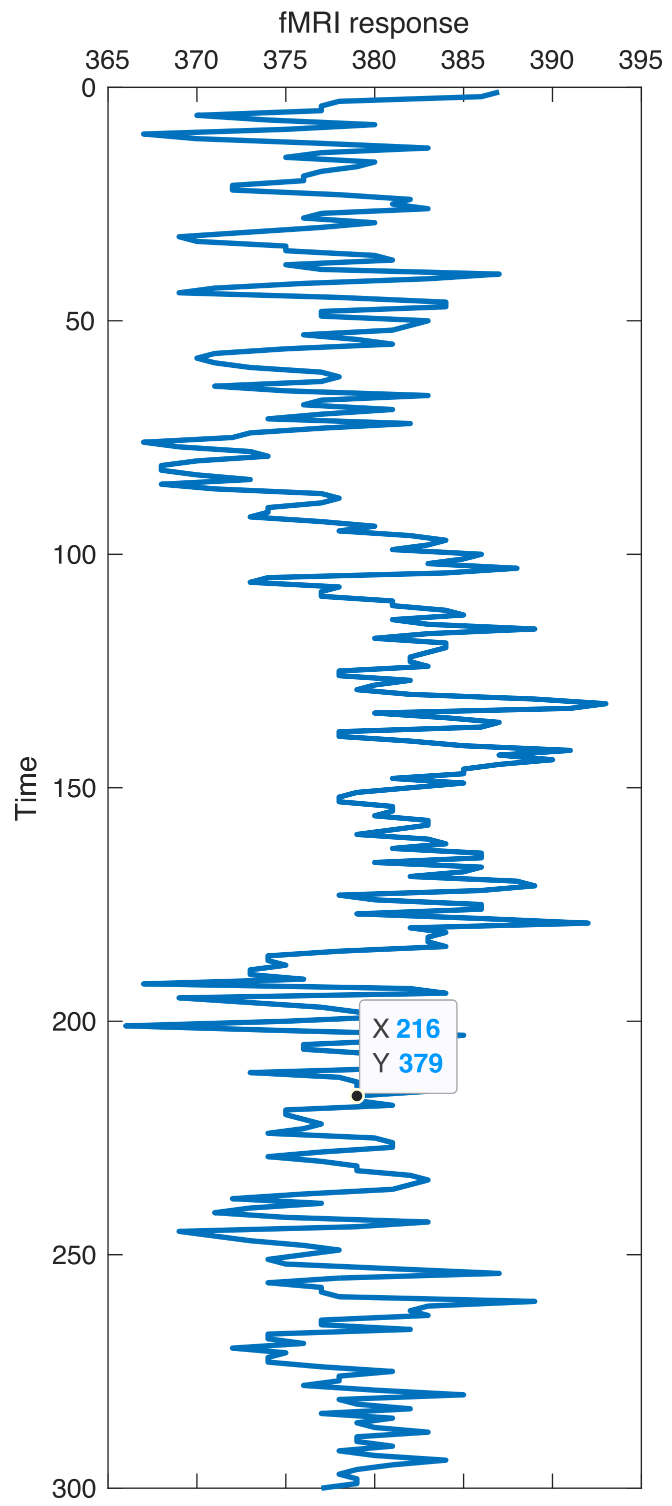
=



[betas]

enter: the GLM

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$



=

$$\begin{bmatrix} \text{betas} \end{bmatrix}$$

data = linear combination of effects

explanatory variable 1

explanatory variable 2

other explanatory variables

Measured data

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 0.72 \\ 0.90 \\ 0.65 \\ 0.20 \\ 0.01 \\ \cdot \\ \cdot \end{bmatrix} p_1 + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \end{bmatrix} p_2 + \dots$$

General Linear Model

Nx1 vector **NxL matrix** **Lx1 vector**

*Measured data (**y**)*

$$\begin{bmatrix} \textit{Measured time series} \end{bmatrix} = \begin{bmatrix} \textit{known matrix with more rows than columns } (\boldsymbol{X}) \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}$$

usually the case in fMRI experiments:

***N:** number of time points in the time series*
***L:** number of regressors in the design matrix*

General Linear Model

Nx1 vector **NxL matrix** **Lx1 vector**

$$\text{Measured data } (\mathbf{y}) \quad \begin{bmatrix} \rotatebox{90}{\textit{Measured time series}} \end{bmatrix} = \begin{bmatrix} \textit{known matrix} \\ \textit{with more rows} \\ \textit{than columns} \\ (\boldsymbol{X}) \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}$$

usually the case in fMRI experiments:

matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

N: *number of time points in the time series*
L: *number of regressors in the design matrix*

*How to solve for **p**?*

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

*Unlikely to find **exact** solution, because we have more equations than unknowns.*

$$\mathbf{p}_{\mathrm{opt}} = \mathbf{X}^{\#}\mathbf{y}$$

*... where $\mathbf{p}_{opt}$ are the (best) parameter estimates and **#** means pseudoinverse.*

*How to solve for **p**?*

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

*Unlikely to find **exact** solution, because we have more equations than unknowns.*

?

?

$$\mathbf{p}_{\mathrm{opt}} = \mathbf{X}^{\#}\mathbf{y}$$

... where $\mathbf{p}_{opt}$ are the (best) parameter estimates and # means pseudoinverse.

General Linear Model

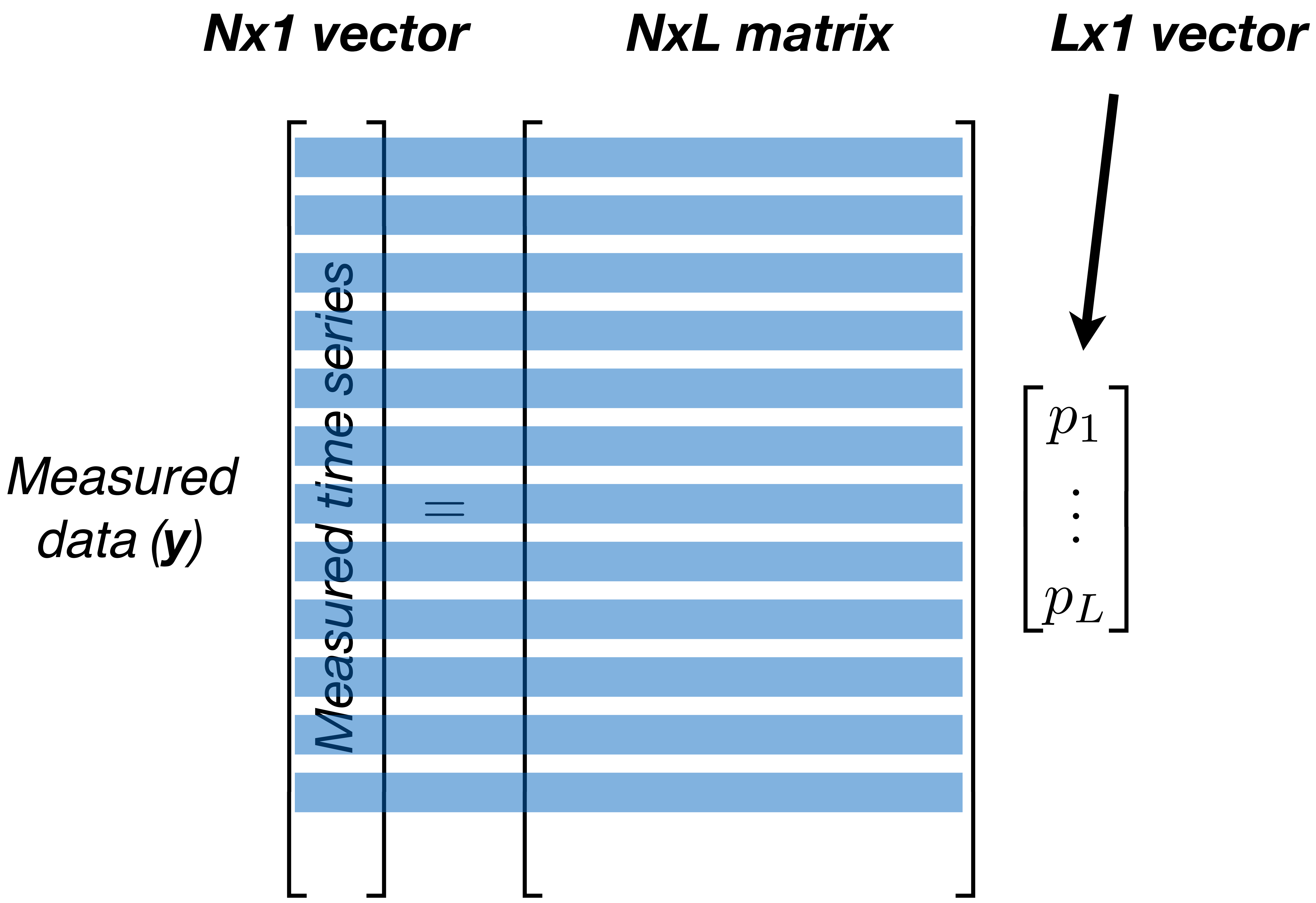
Nx1 vector **NxL matrix** **Lx1 vector**

*Measured data (**y**)*

$$\begin{bmatrix} \rotatebox{90}{\textit{Measured time series}} \end{bmatrix} = \begin{bmatrix} \quad & \quad \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}$$

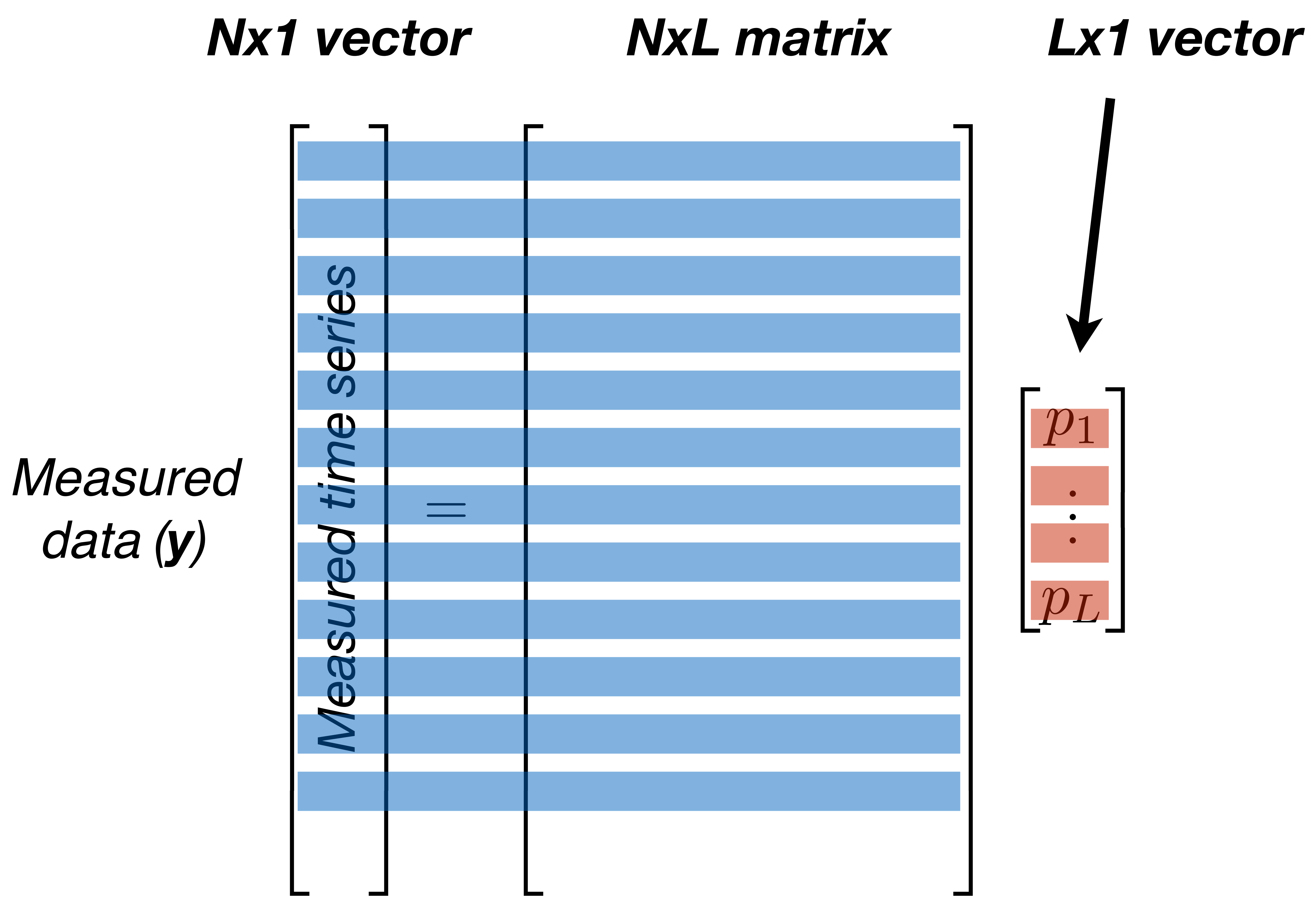
N: *number of time points in the time series*
L: *number of regressors in the design matrix*

General Linear Model



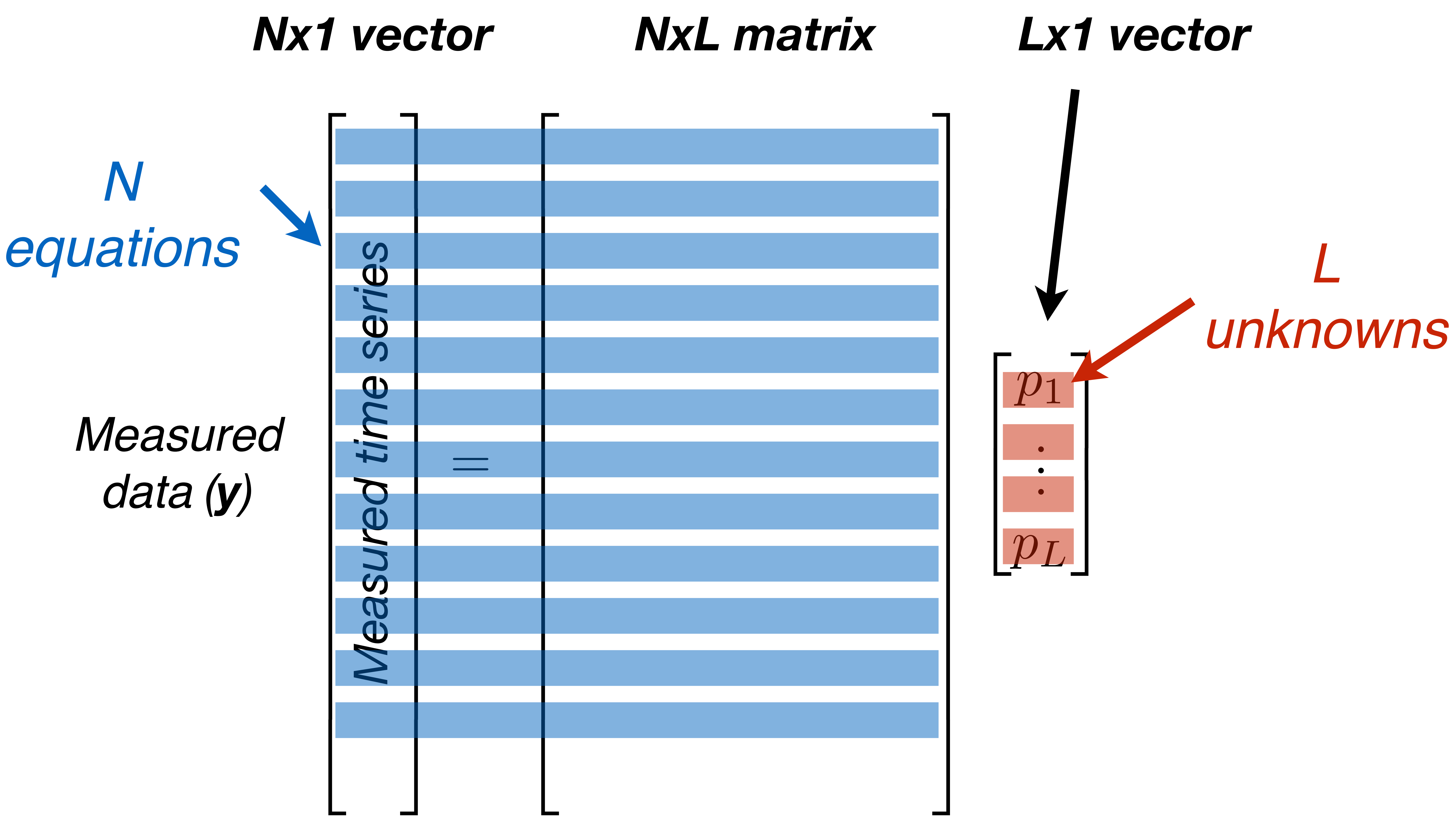
N: *number of time points in the time series*
L: *number of regressors in the design matrix*

General Linear Model



N: *number of time points in the time series*
L: *number of regressors in the design matrix*

General Linear Model



N: *number of time points in the time series*
L: *number of regressors in the design matrix*

Multiple parameters

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

*... where **y, p** are vectors and **X** is the known matrix. and [•]^T is transpose and [•]$^{-1}$ is the matrix inverse.*

Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p} \\ (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{p}_{\mathrm{opt}} \end{aligned}$$

... where ***y, p*** *are vectors*
and ***X*** *is the known matrix.*
and $[\bullet]^T$ *is transpose and*
$[\bullet]^{-1}$ *is the matrix inverse.*

Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p} \\ \underbrace{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}}_{\textbf{\textit{projection matrix}}}\mathbf{y} &= \mathbf{p}_{\mathrm{opt}} \end{aligned}$$

*... where **y, p** are vectors*
*and **X** is the known matrix.*
and $[\bullet]^T$ is transpose and
$[\bullet]^{-1}$ is the matrix inverse.

Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p} \\ \underbrace{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}}_{\text{projection matrix}}\mathbf{y} &= \mathbf{p}_{\mathrm{opt}} \end{aligned}$$

projection matrix

pseudoinverse (X)

*... where **y, p** are vectors and **X** is the known matrix. and [•]^T is transpose and [•]$^{-1}$ is the matrix inverse.*

```
>> p = pinv(X)*y

>> p = X \ y
```

Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p} \\ \underbrace{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}}_{\textit{projection matrix}}\mathbf{y} &= \mathbf{p}_{\mathrm{opt}} \end{aligned}$$

pseudoinverse (X)

*... where **y, p** are vectors and **X** is the known matrix. and $[\cdot]^T$ is transpose and $[\cdot]^{-1}$ is the matrix inverse.*

```
>> p = pinv(X)*y

>> p = X \ y
```

in matlab, this uses SVD under the hood...

Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{p} \\ (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} &= \mathbf{p}_{\mathrm{opt}} \end{aligned}$$

projection matrix

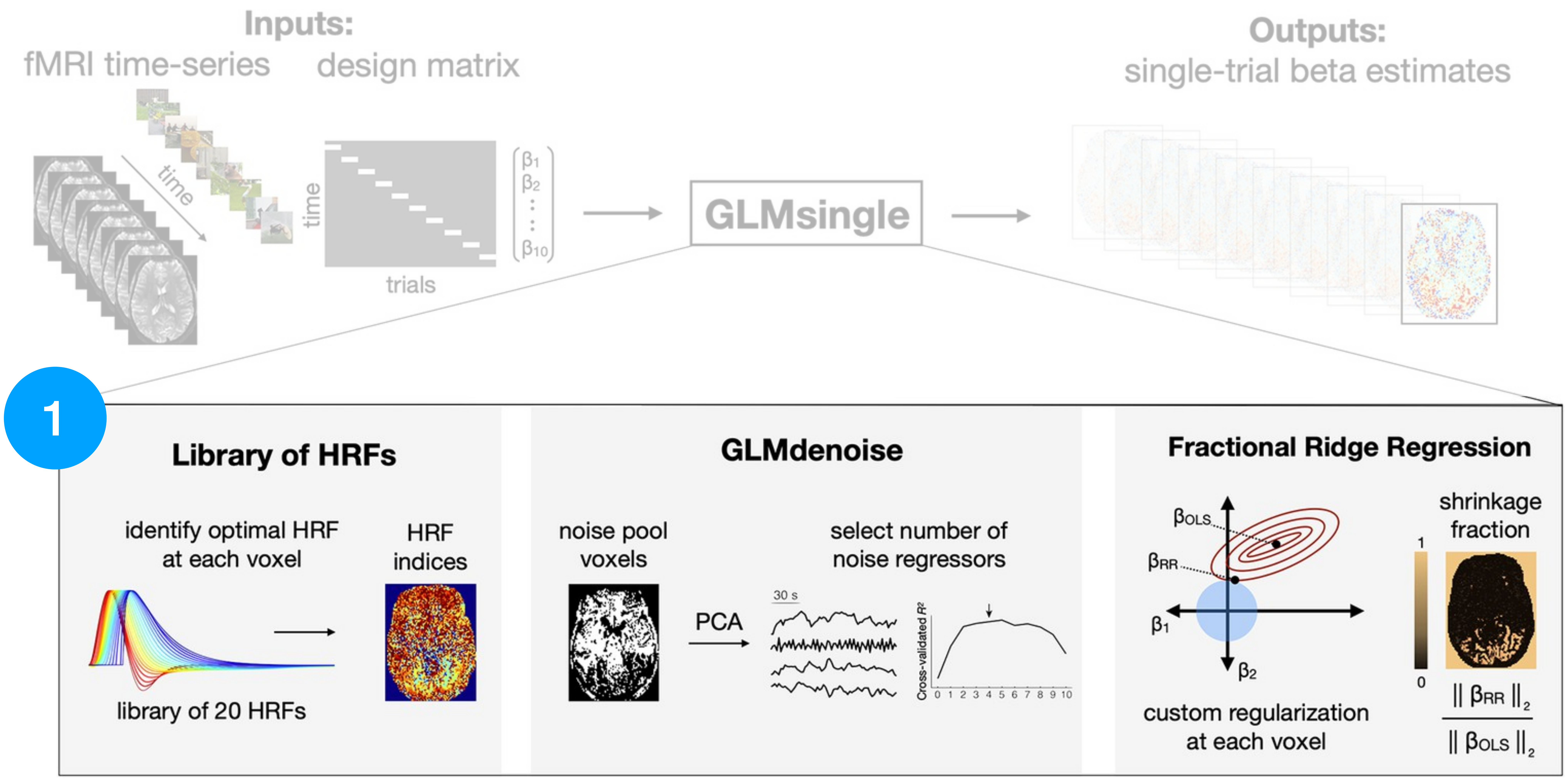
pseudoinverse (X)

in matlab, this uses SVD under the hood...

```
>> p = pinv(X)*y
>> p = X \ y
```

should give same answer, but uses different method (QR)

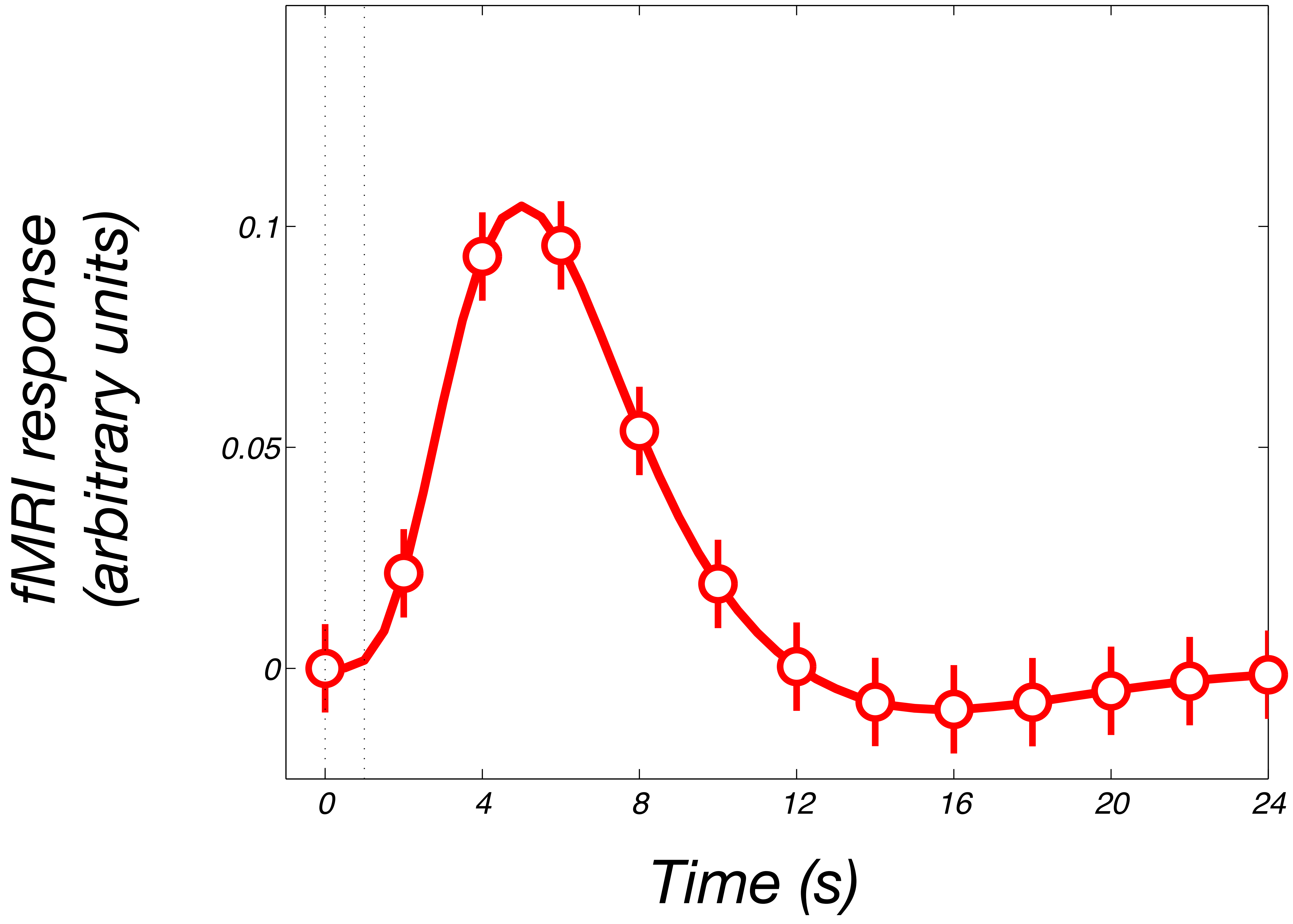
*... where **y, p** are vectors and **X** is the known matrix. and $[\bullet]^T$ is transpose and $[\bullet]^{-1}$ is the matrix inverse.*



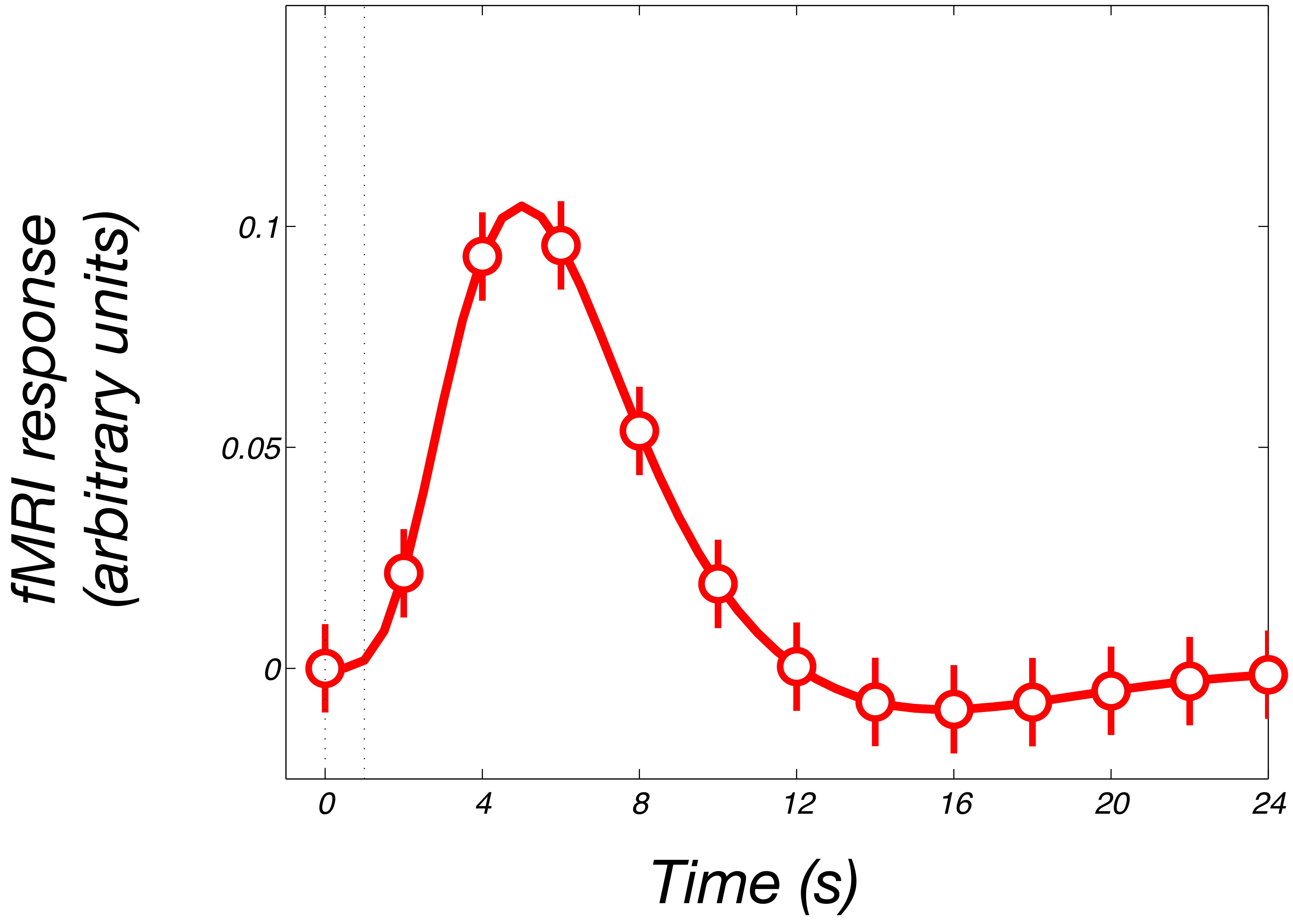
Inputs:
fMRI time-series
design matrix
time
time
trials
β1
β2
β10
GLMsingle
Outputs:
single-trial beta estimates
1
Library of HRFs
identify optimal HRF at each voxel
HRF indices
library of 20 HRFs
GLMdenoise
noise pool voxels
PCA
select number of noise regressors
30 s
Cross-validated R²
0 1 2 3 4 5 6 7 8 9 10
Fractional Ridge Regression
βOLS
βRR
β1
β2
custom regularization at each voxel
shrinkage fraction
1
0
|| βRR ||2 / || βOLS ||2

haemodynamic response function

links event / event times + obverved data



haemodynamic response function

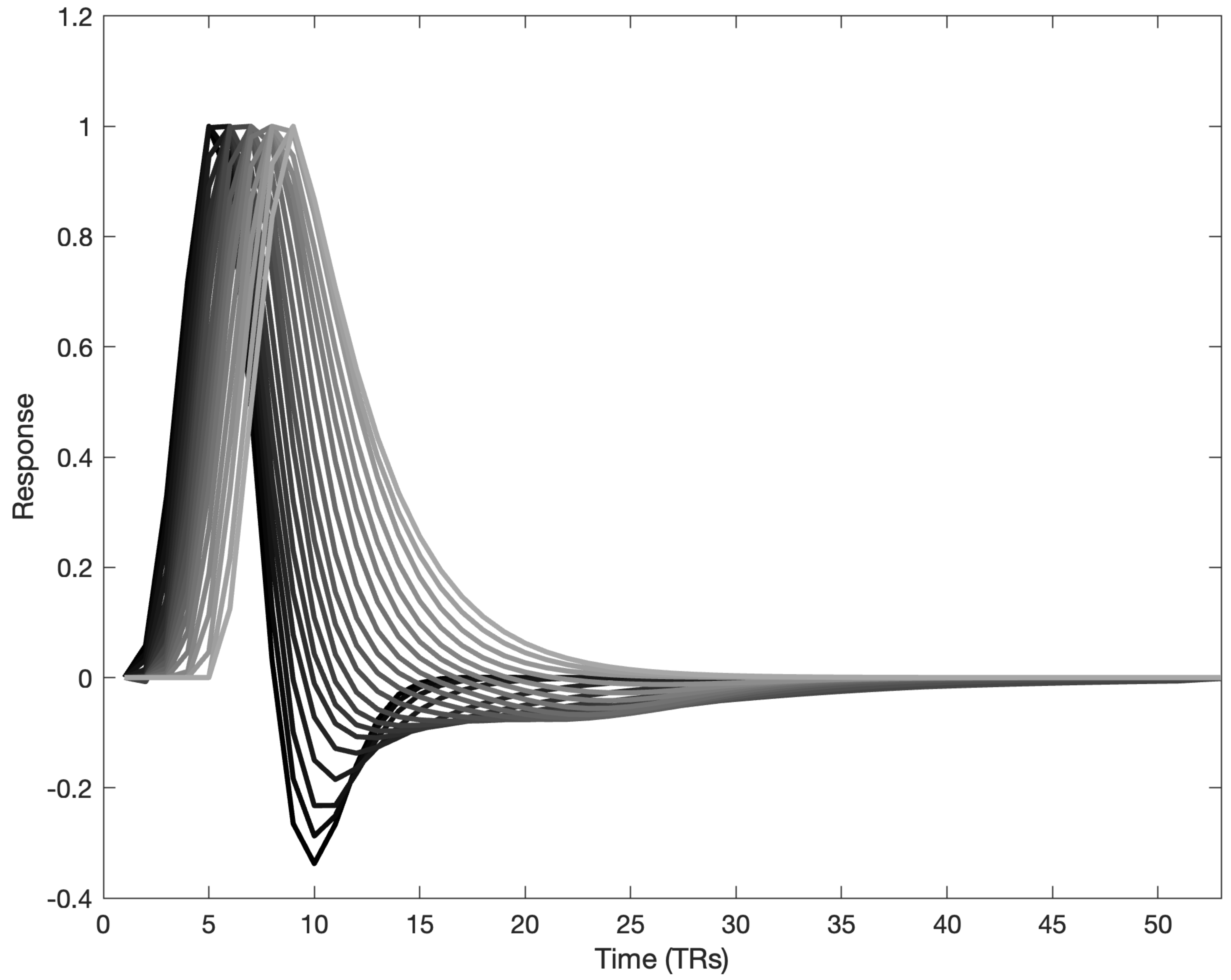


links event / event times + obverved data

shape usually assumed ("SPM", "Glover") – but it varies!

1

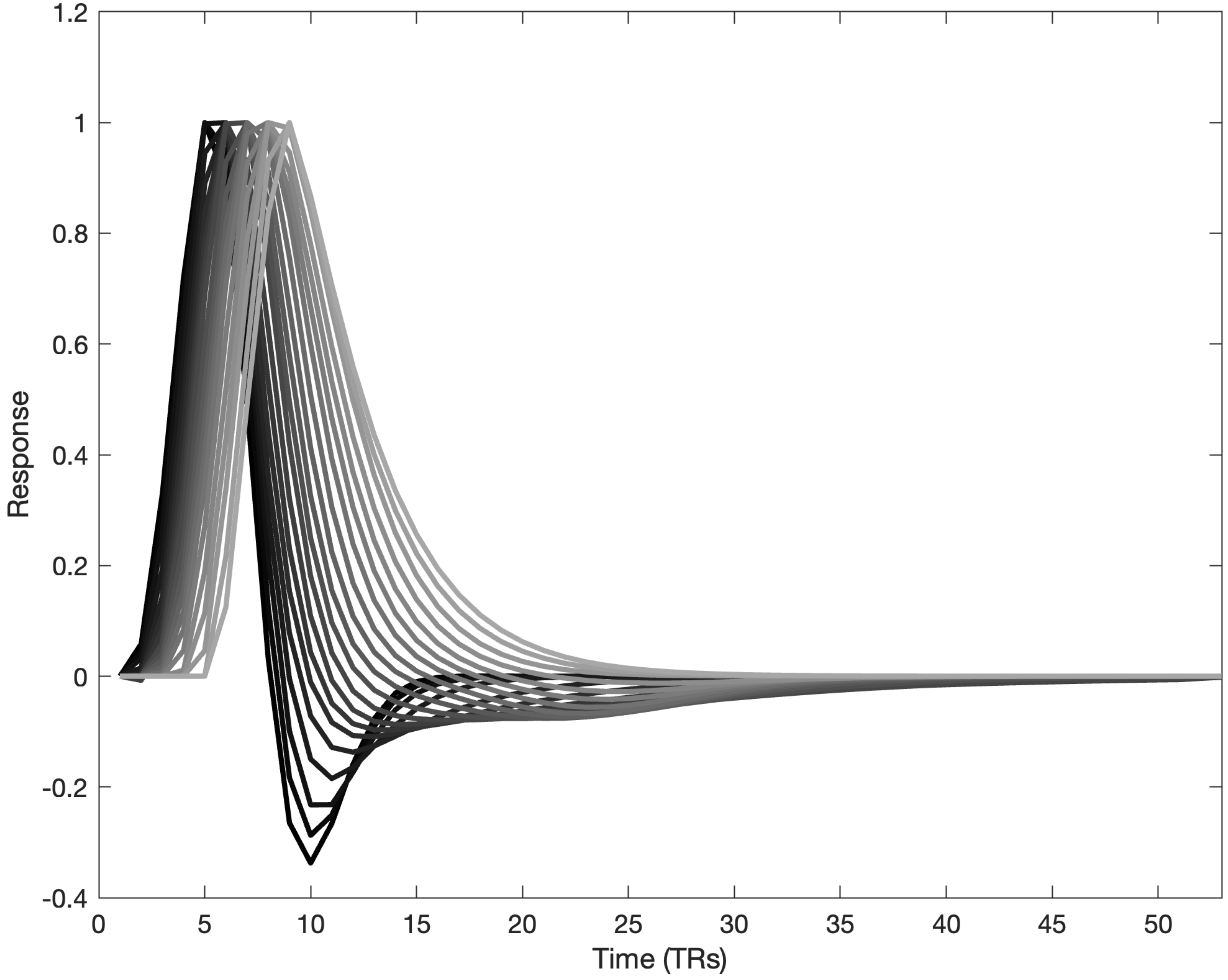
don’t assume: find the best



re-run model on data with one of 20 choices…

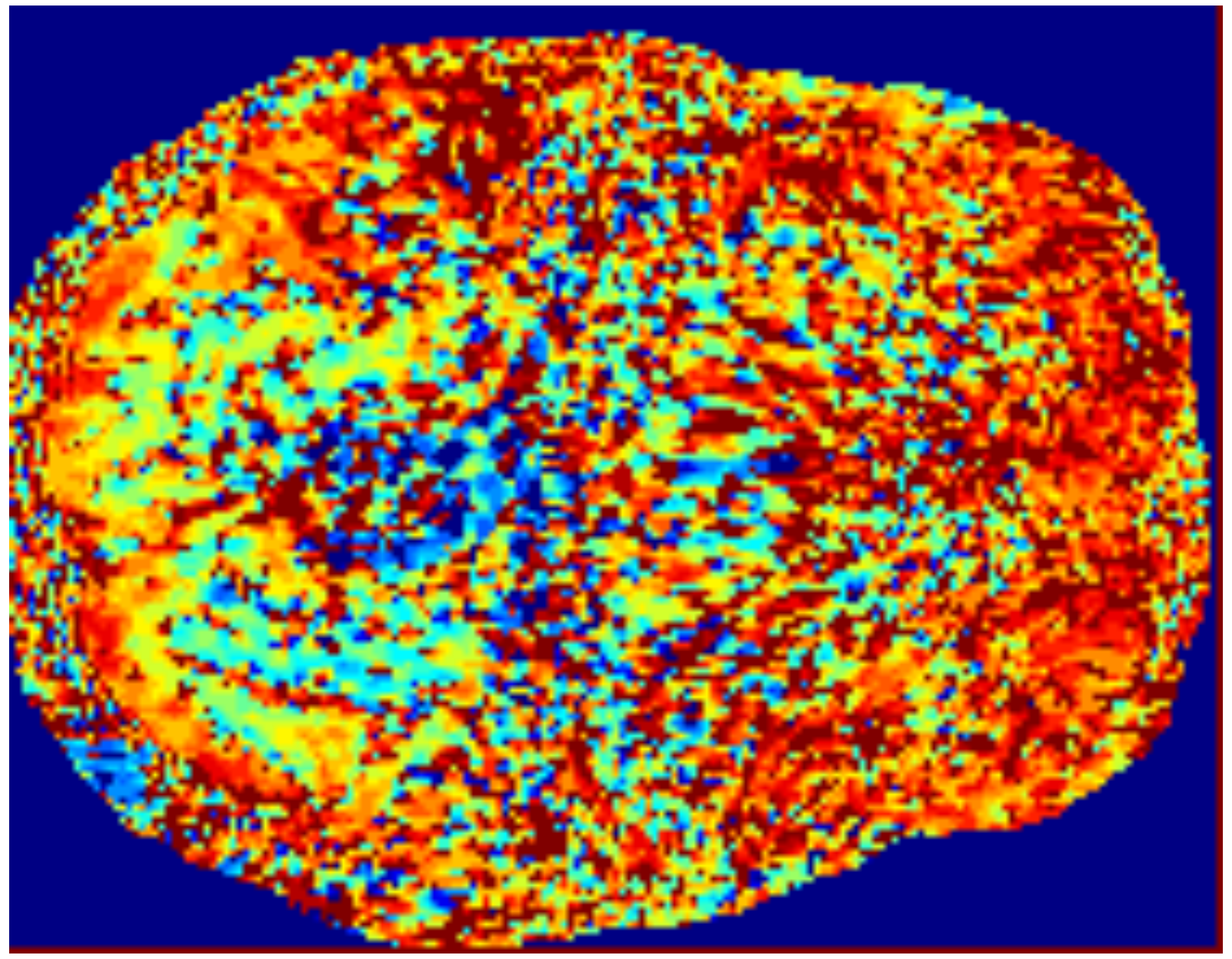
pick the one that maximises r^2

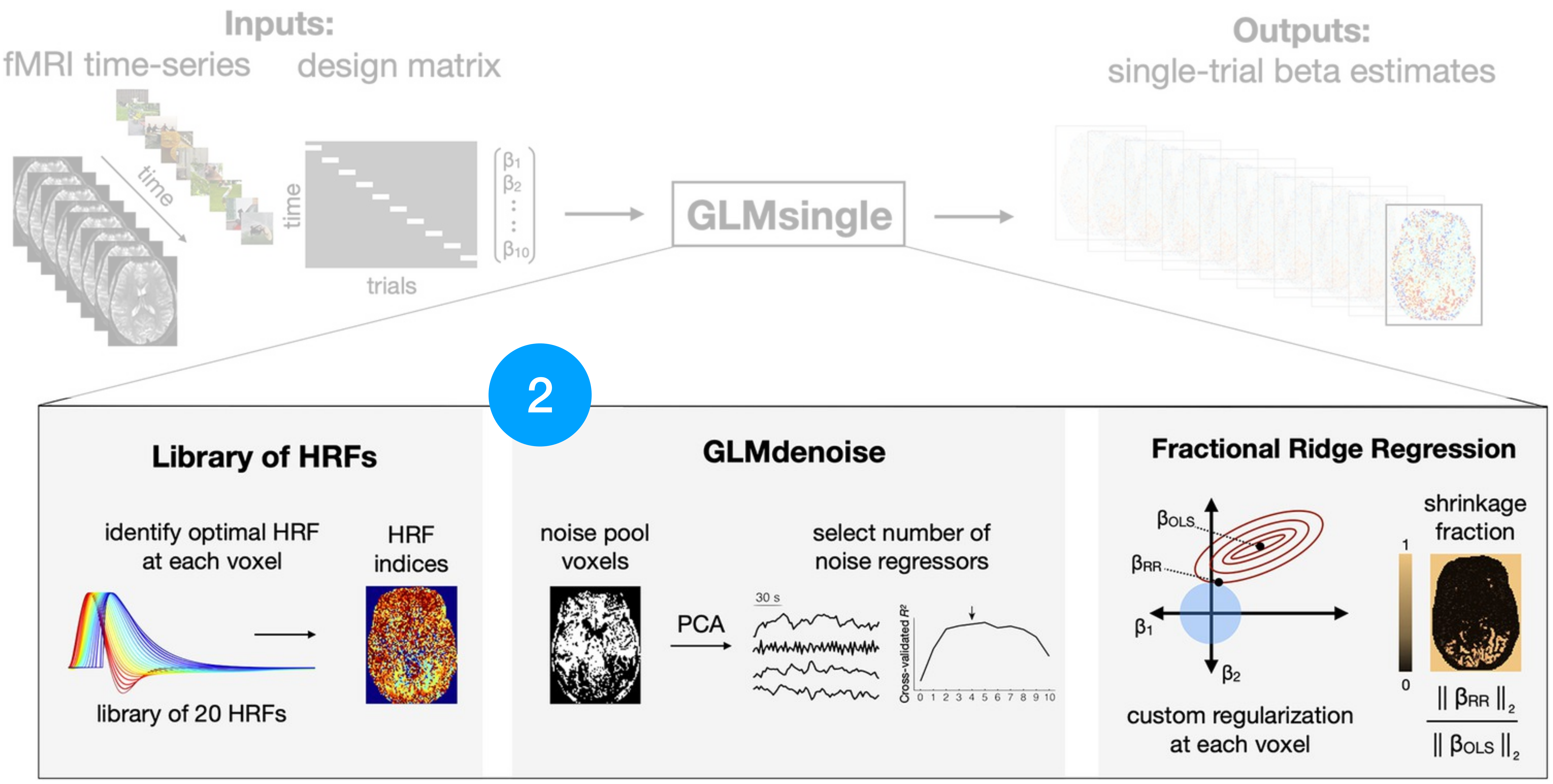
1 don’t assume: find the best



re-run model on data with one of 20 choices…

pick the one that maximises r^2



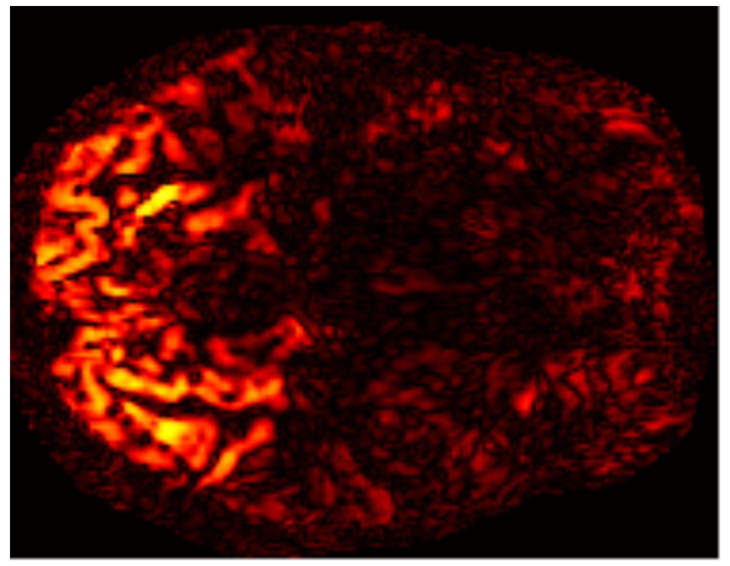

Inputs:
fMRI time-series
design matrix
time
time
trials
β1
β2
β10
GLMsingle
Outputs:
single-trial beta estimates
2
Library of HRFs
identify optimal HRF at each voxel
HRF indices
library of 20 HRFs
GLMdenoise
noise pool voxels
PCA
select number of noise regressors
30 s
Cross-validated R²
0 1 2 3 4 5 6 7 8 9 10
Fractional Ridge Regression
βOLS
βRR
β1
β2
custom regularization at each voxel
shrinkage fraction
1
0
|| βRR ||2 / || βOLS ||2

2

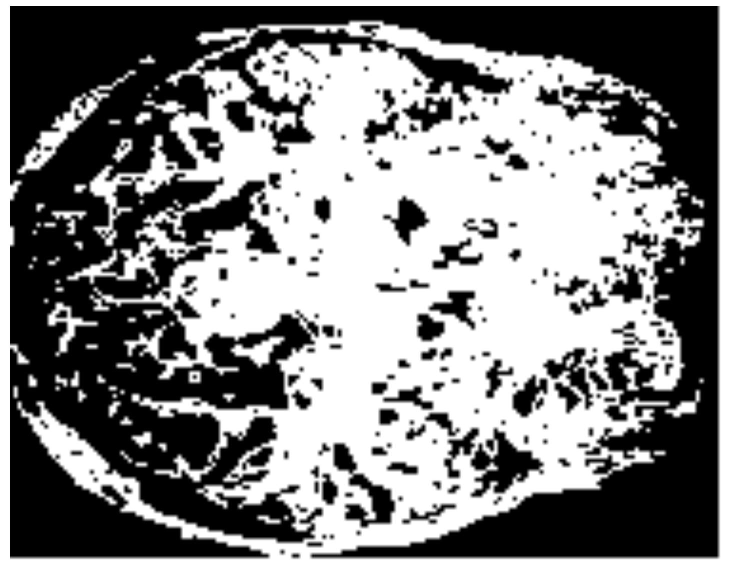
try to remove noise (PCA)

- treat all events as one (ON)… this leads to an ON-OFF design
- find voxels that have a low r^2 … this becomes a noise pool
- use PCA on the noise pool voxels to find common noise “time courses” (and include them in your model as nuisance regressors)

r^2



noise pool



2

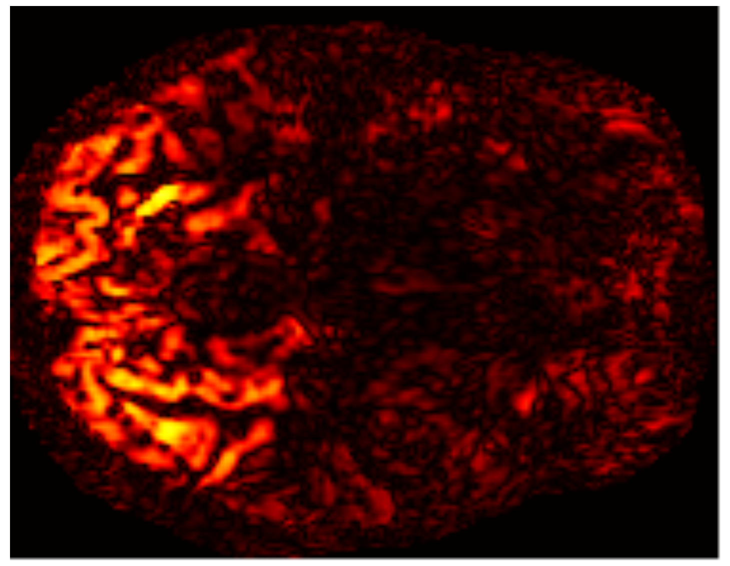
try to remove noise (PCA)

- treat all events as one (ON)… this leads to an ON-OFF design
- find voxels that have a low r^2 … this becomes a noise pool
- use PCA on the noise pool voxels to find common noise “time courses” (and include them in your model as nuisance regressors)

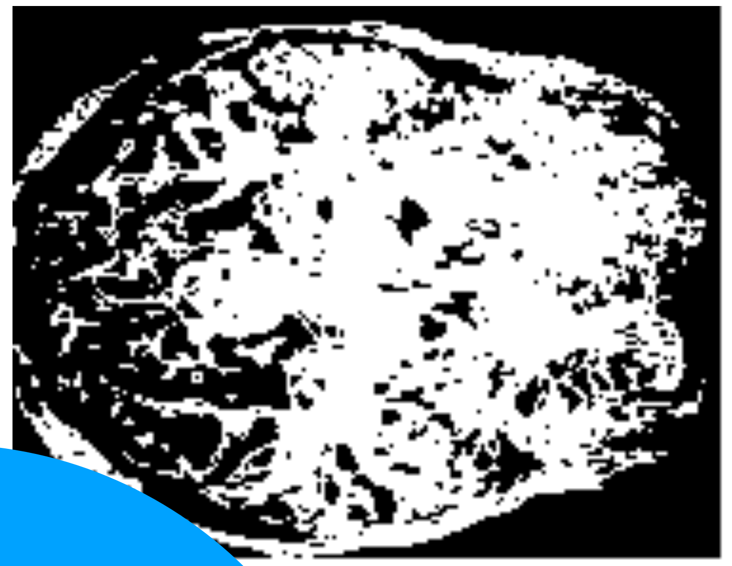
how many / which components to include?

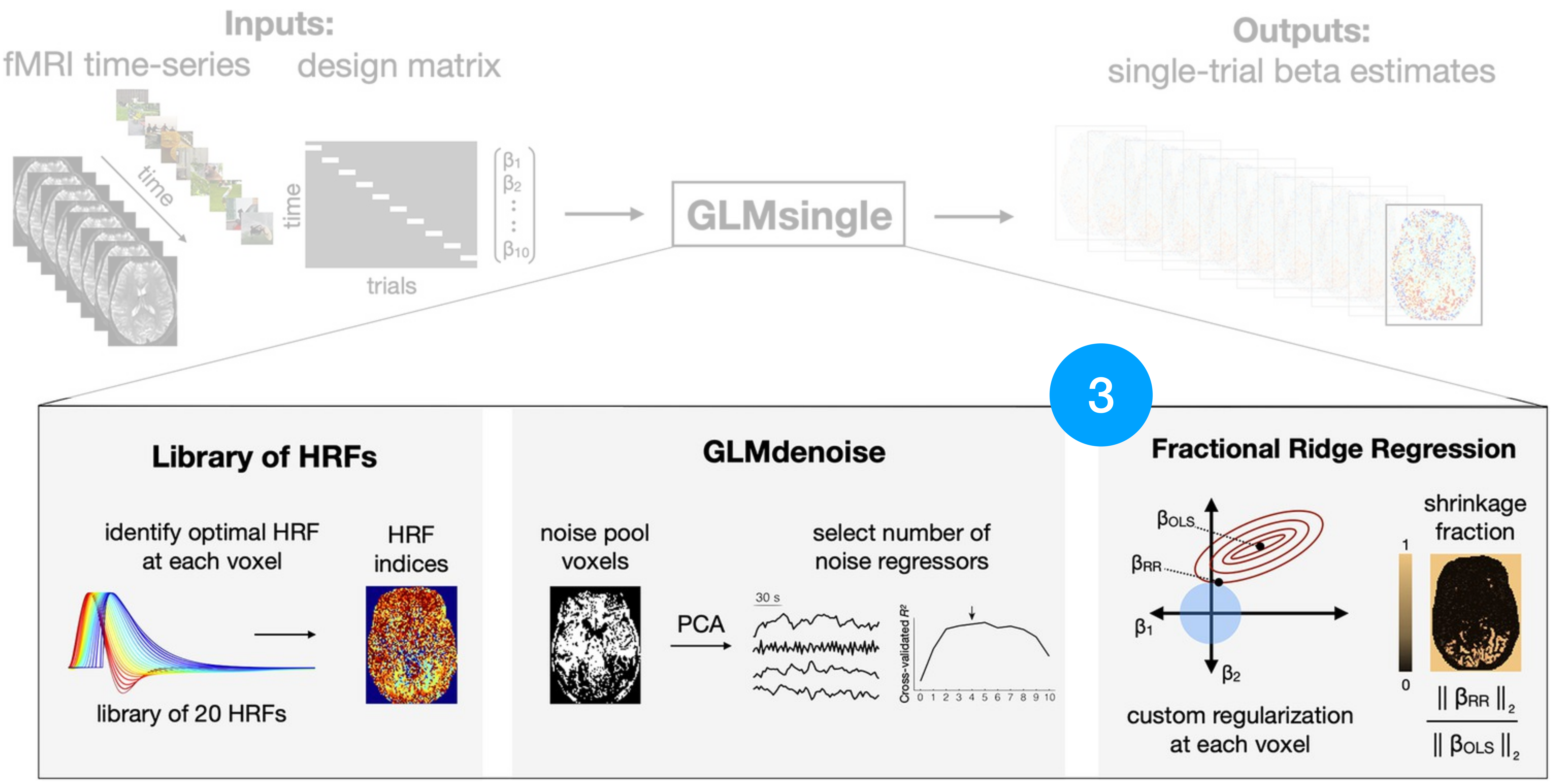
cross-validation

r^2



noise pool




Inputs:
fMRI time-series
design matrix
time
time
trials
β_1
β_2
β_{10}
GLMsingle
Outputs:
single-trial beta estimates
3
Library of HRFs
identify optimal HRF at each voxel
HRF indices
library of 20 HRFs
GLMdenoise
noise pool voxels
PCA
select number of noise regressors
30 s
Cross-validated R^2
0 1 2 3 4 5 6 7 8 9 10
Fractional Ridge Regression
β_{OLS}
β_{RR}
β_1
β_2
custom regularization at each voxel
shrinkage fraction
1
0
$\frac{\|\beta_{RR}\|_2}{\|\beta_{OLS}\|_2}$

3

deal with correlated regressors

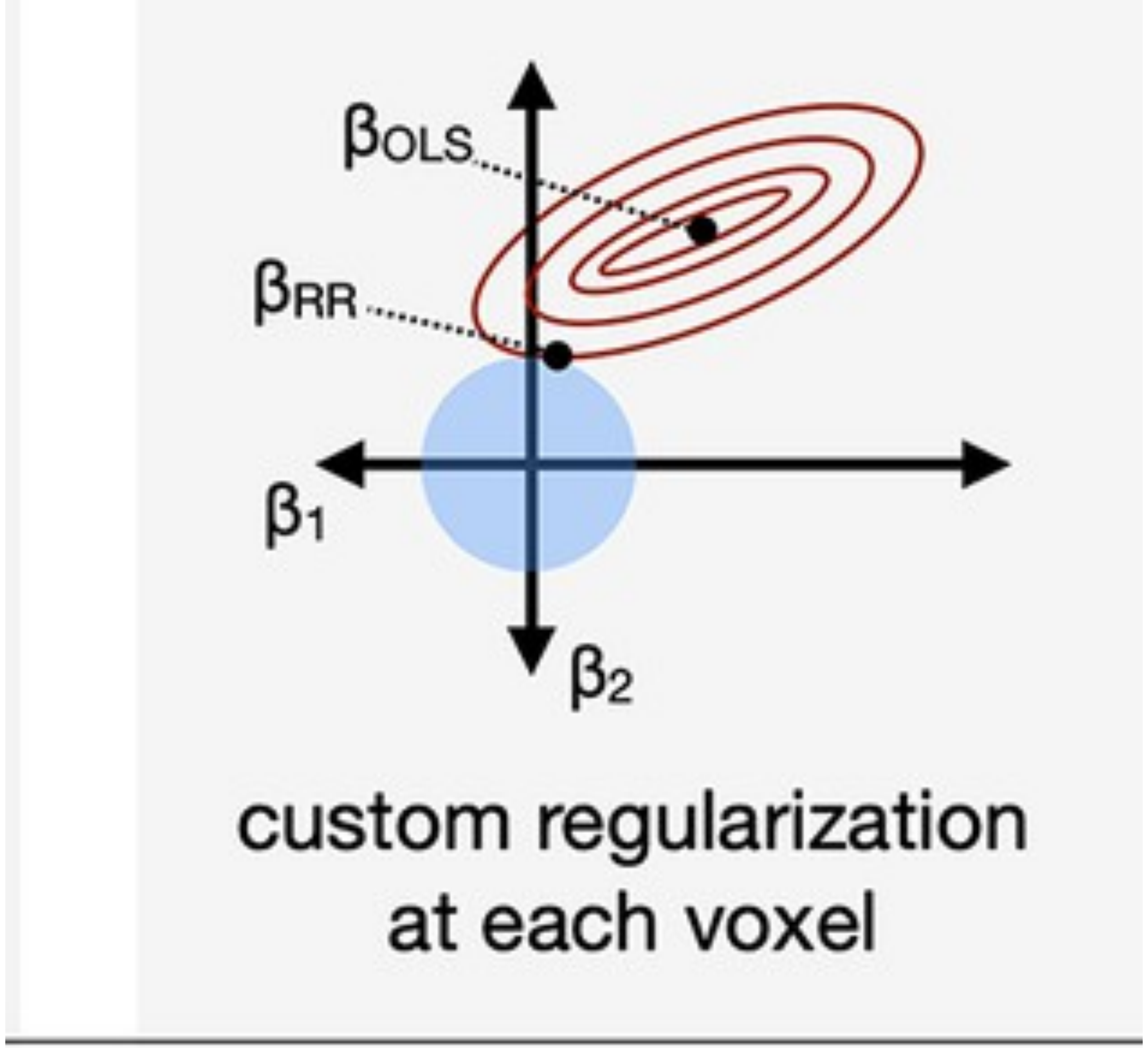
- *Make sure the design matrix makes sense!*
- *Is $X^T X$ always invertible? If not, why not?*
- *What is the interpretation for the values corresponding to each element of p_{opt}? Is the meaning of each value independent of the other elements?*

$$(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$$

what's the problem?

- when two regressors can explain similar parts of the data, then the problem is ill-posed
- noise can make estimates jump around a lot
- **solution: regularise the weights**

  (this means picking a set that fulfils certain constraints - punish large beta weights

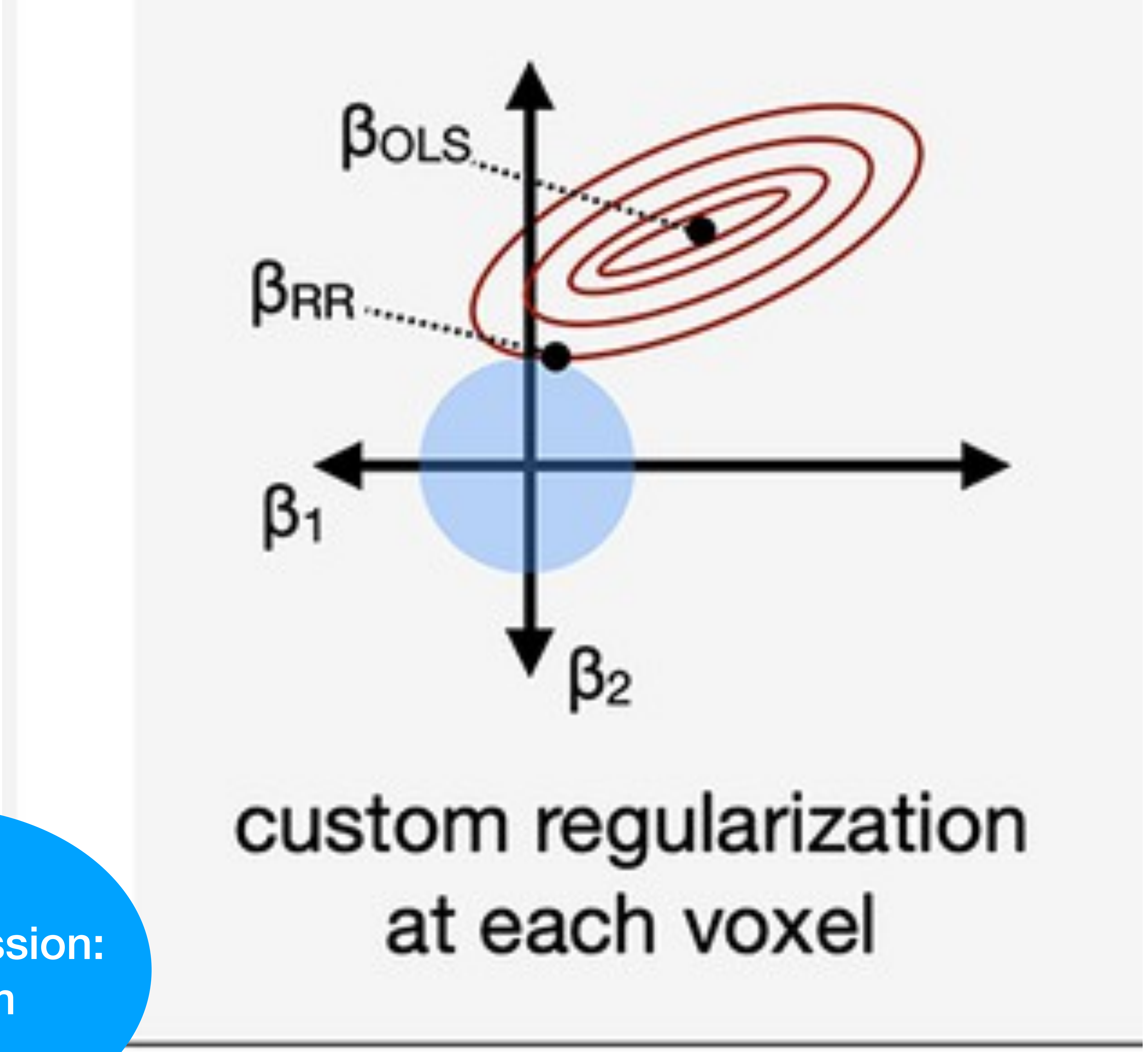


what's the problem?

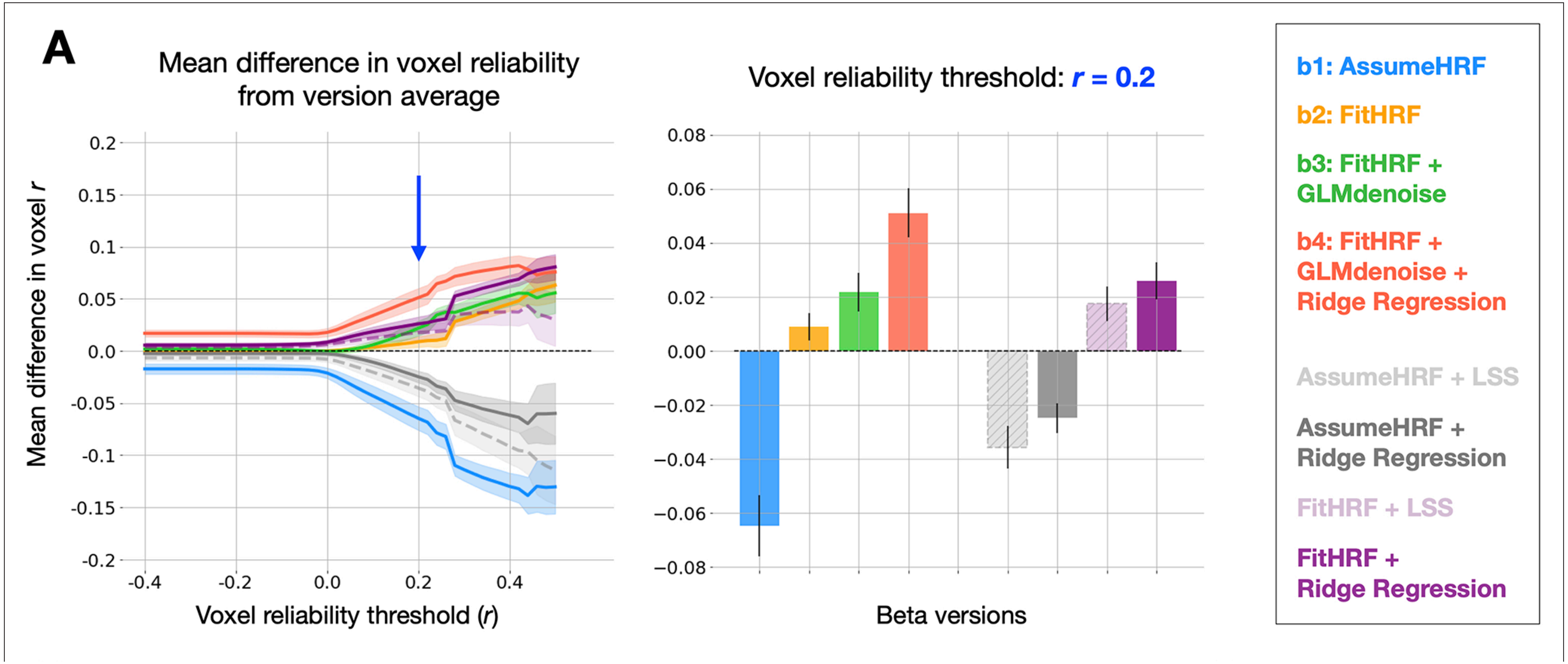
- when two regressors can explain similar parts of the data, then the problem is ill-posed
- noise can make estimates jump around a lot
- **solution: regularise the weights**

  (this means picking a set that fulfils certain constraints - punish large beta weights

ridge regression: L2 norm



reliability goes up



Discussion points

- do people analyse their timeseries data in similar ways (eye tracking? physiological data? EEG / event-related..)

- ONOFF idea - for a noise pool / data driven data cleaning?

- regularisation? ridge regression / LASSO, etc.