The General Linear Model (GLM)

Klaas Enno Stephan

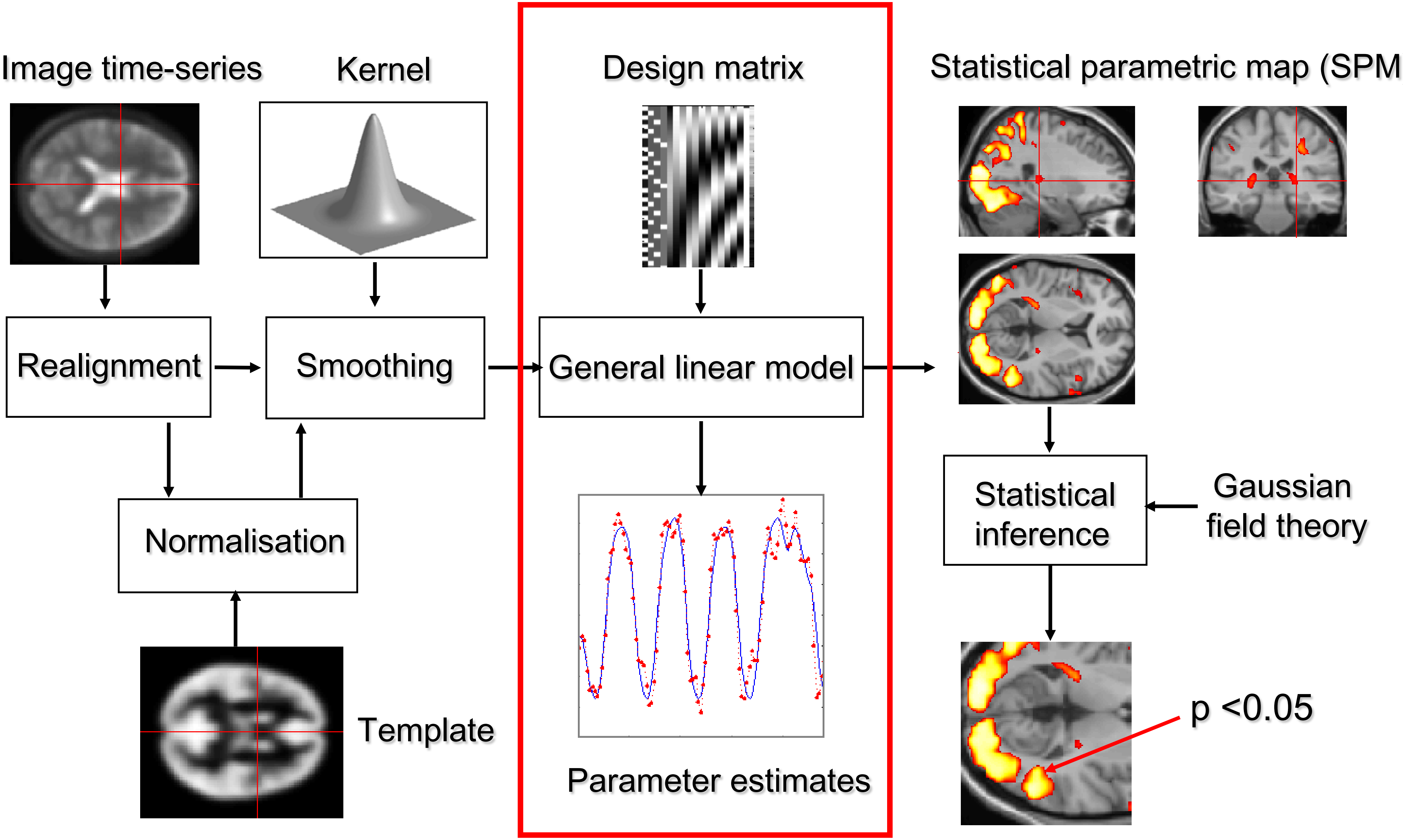
Translational Neuromodeling Unit (TNU)
Institute for Biomedical Engineering
University of Zurich & ETH Zurich

Wellcome Trust Centre for Neuroimaging
Institute of Neurology,
University College London

With many thanks to the FIL Methods group

Zurich SPM Course 2012
15-17 February

Overview of SPM



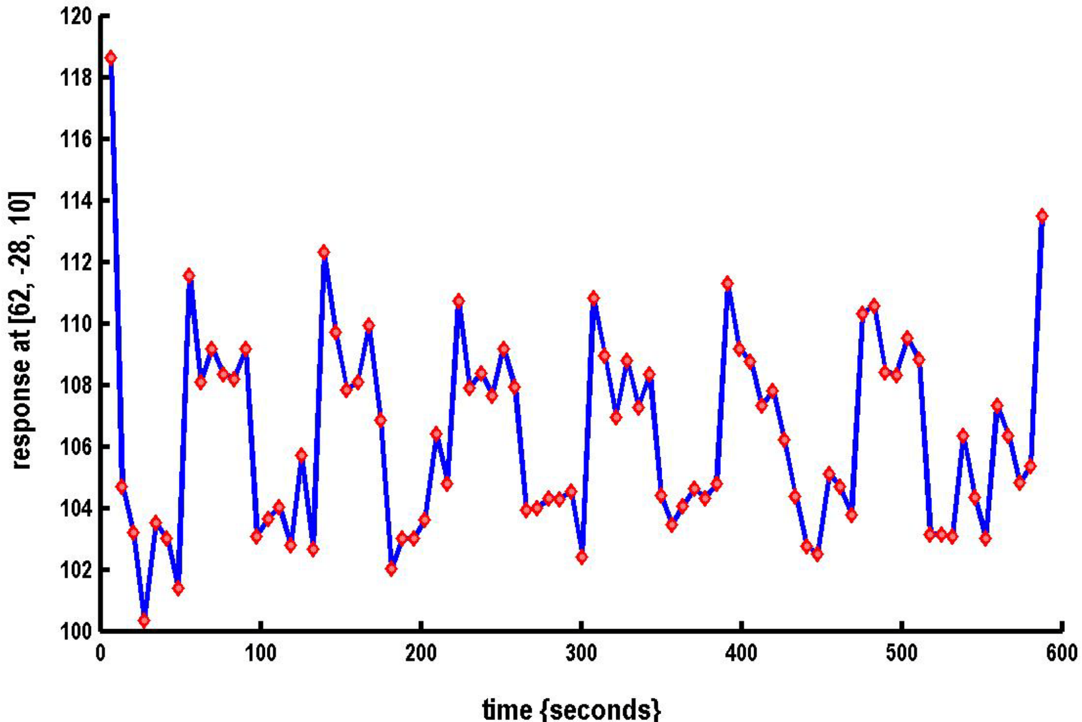
A very simple fMRI experiment

One session

Passive word listening versus rest

7 cycles of rest and listening

Blocks of 6 scans with 7 sec TR



Stimulus function

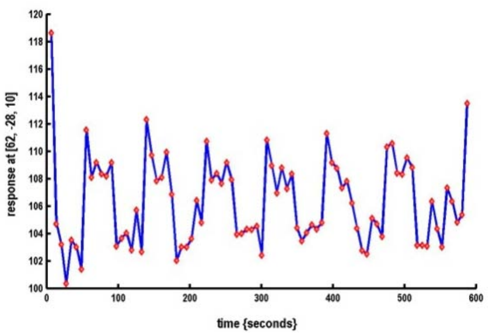
Question: Is there a change in the BOLD response between listening and rest?

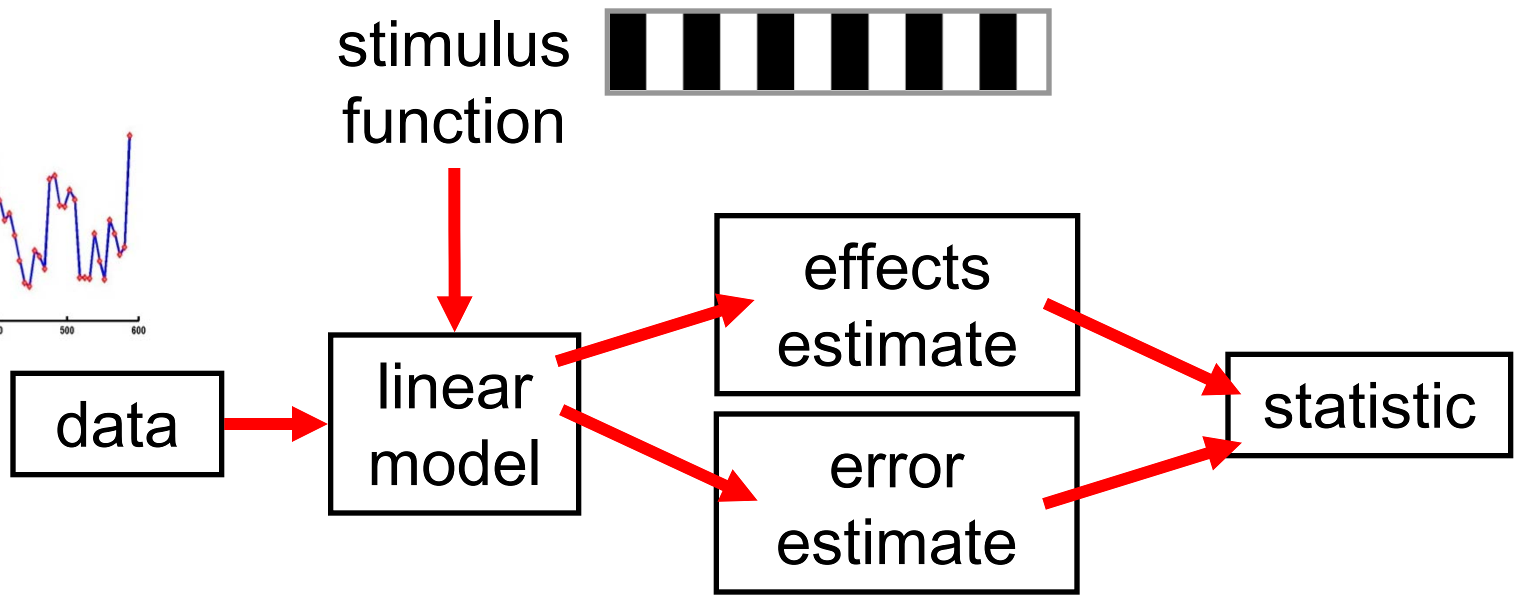
Modelling the measured data

Why? Make inferences about effects of interest

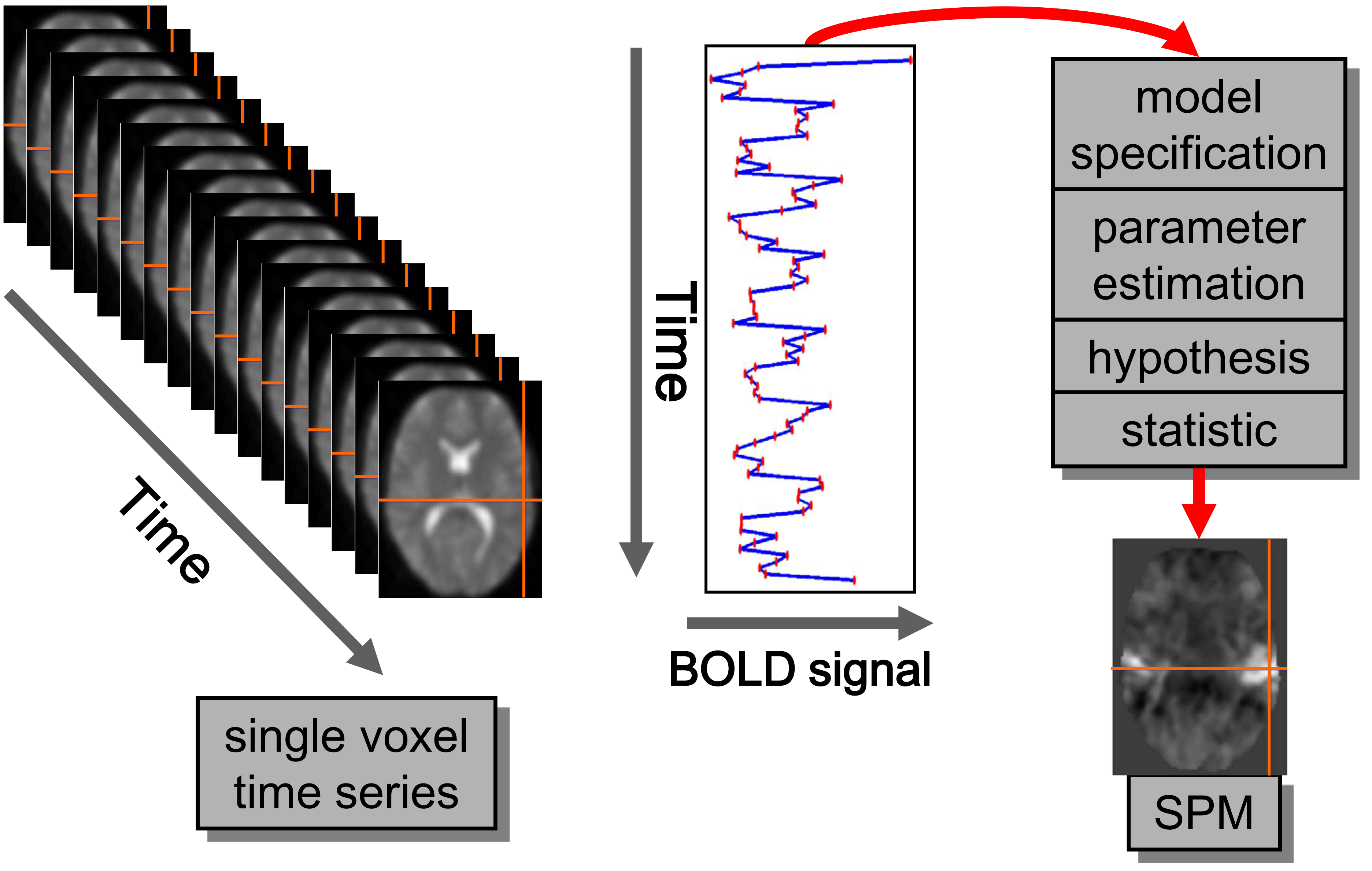
How?

1. Decompose data into effects and error
2. Form statistic using estimates of effects and error

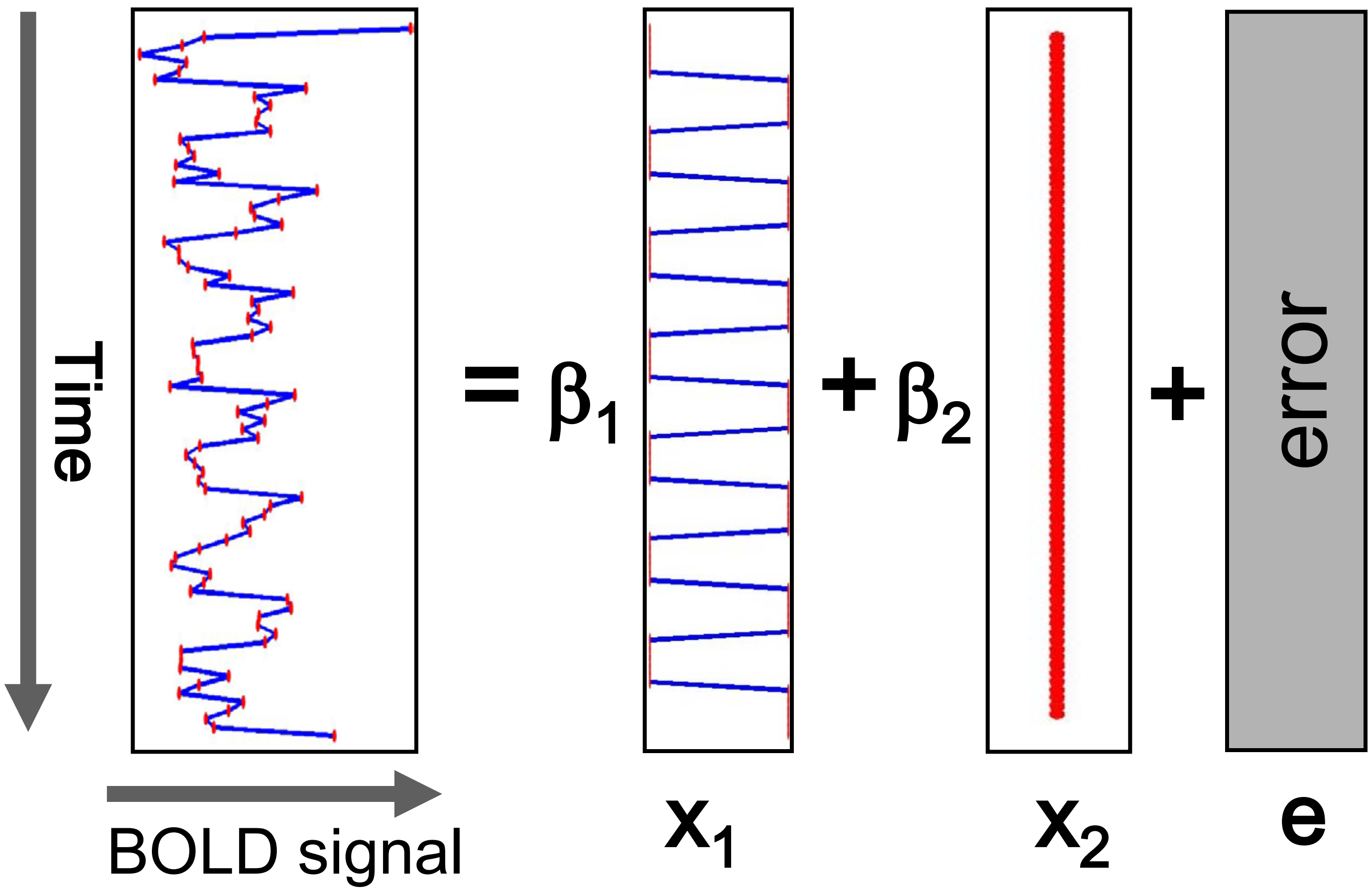




Voxel-wise time series analysis

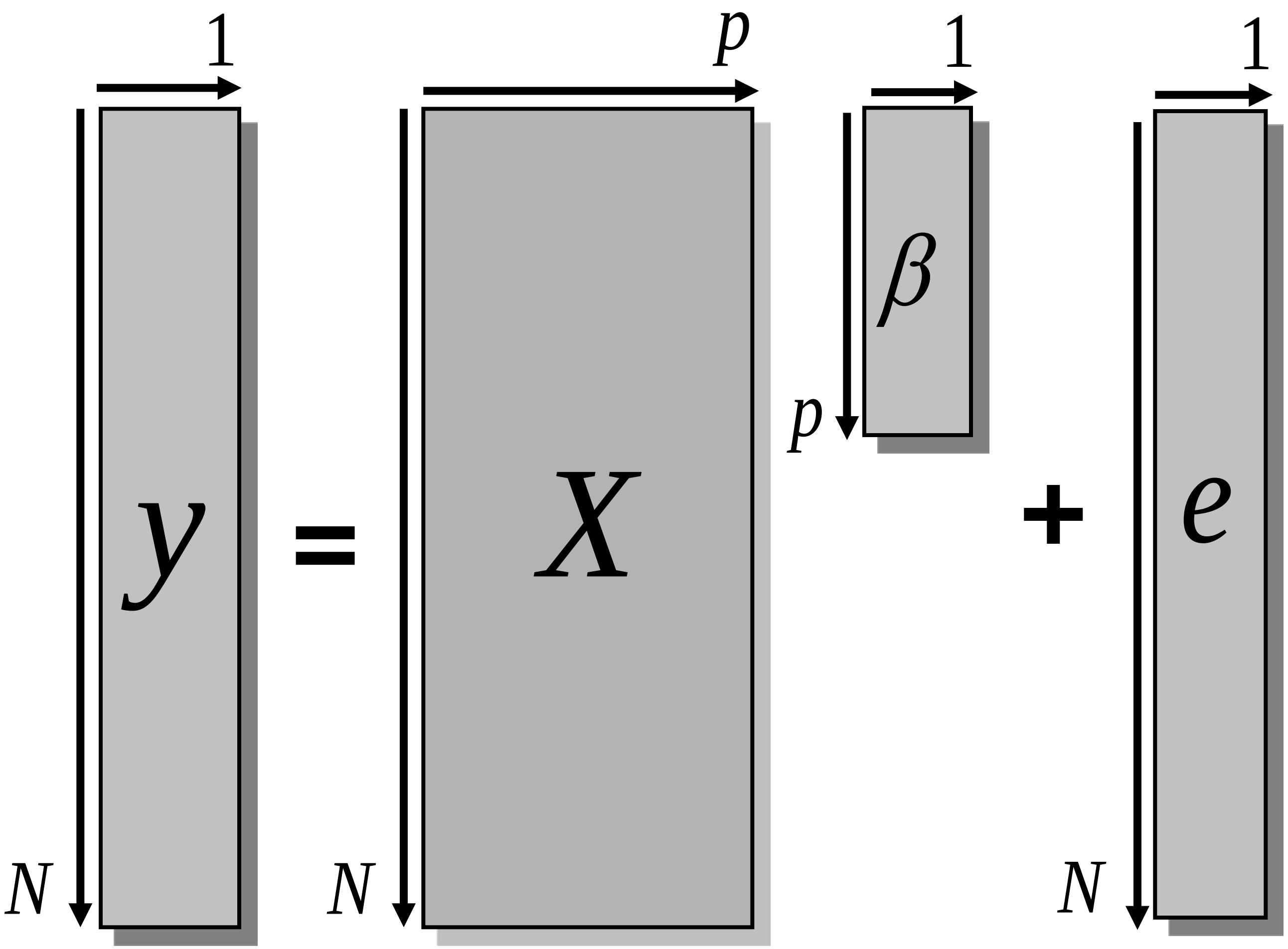


Single voxel regression model



$$y = x_1\beta_1 + x_2\beta_2 + e$$

Mass-univariate analysis: voxel-wise GLM



$$y = X\beta + e$$

$$e \sim N(0, \sigma^2 I)$$

Model is specified by

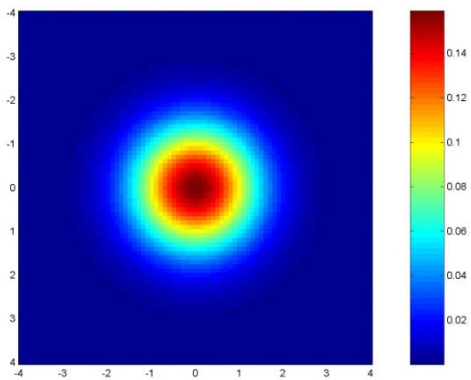
1. Design matrix X
2. Assumptions about e

N: number of scans
p: number of regressors

The design matrix embodies all available knowledge about experimentally controlled factors and potential confounds.

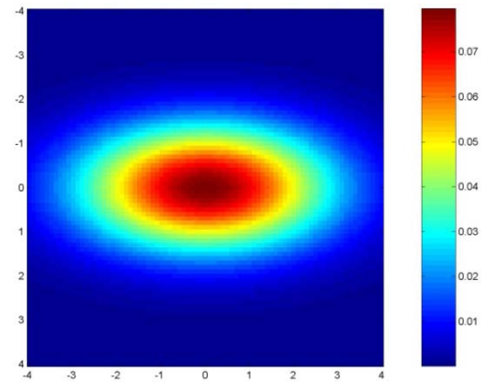
GLM assumes Gaussian "spherical" (i.i.d.) errors

sphericity = i.i.d.
error covariance is
scalar multiple of
identity matrix:
Cov(e) = $\sigma^2 I$



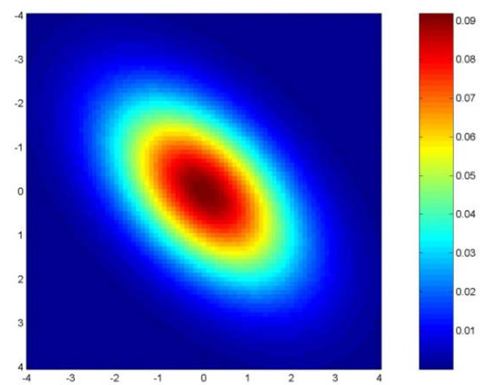
$$Cov(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Examples for non-sphericity:



$$Cov(e) = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

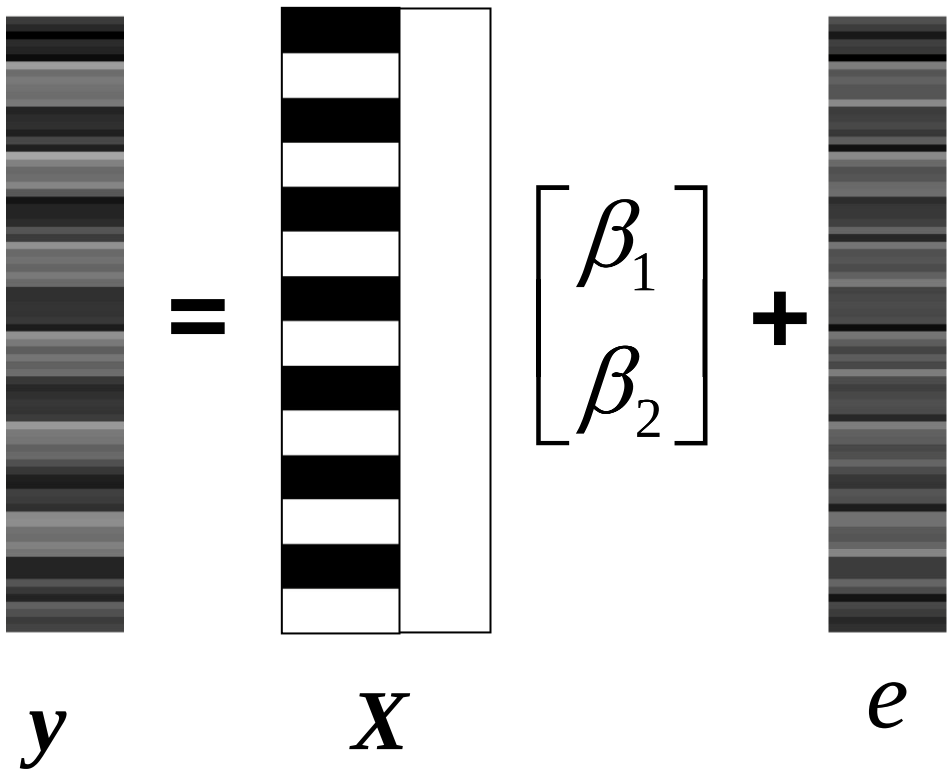
non-identity



$$Cov(e) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

non-independence

Parameter estimation



$$y = X\beta + e$$

Objective:
estimate parameters to minimize $\sum_{t=1}^{N} e_t^2$

Ordinary least squares estimation (OLS) (assuming i.i.d. error):

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

A geometric perspective on the GLM

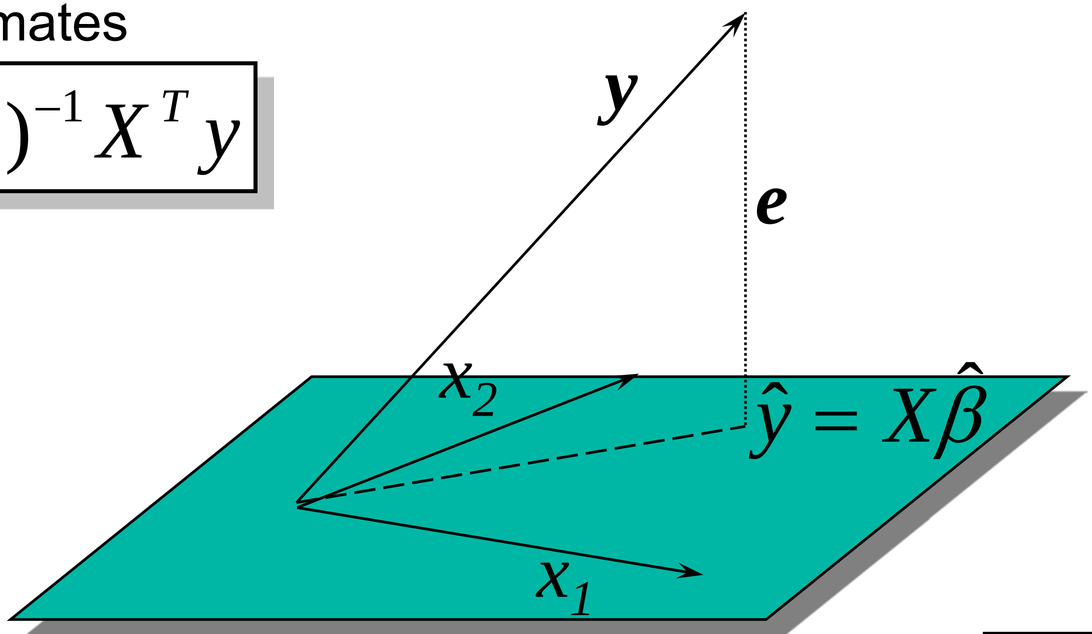
OLS estimates

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Residual forming matrix R

$$e = Ry$$

$$R = I - P$$



Design space defined by X

Projection matrix P

$$\hat{y} = Py$$

$$P = X(X^T X)^{-1} X^T$$

Deriving the OLS equation

$$X^T e = 0$$

$$X^T \left(y - X\hat{\beta} \right) = 0$$

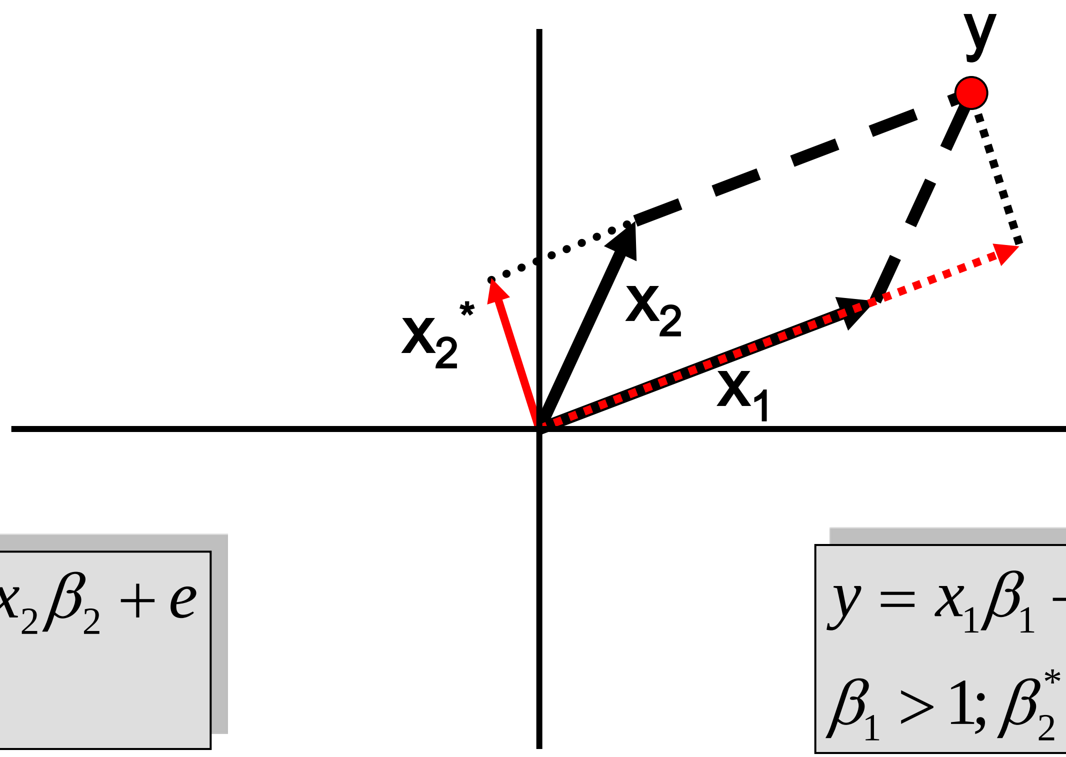
$$X^T y - X^T X \hat{\beta} = 0$$

$$X^T X \hat{\beta} = X^T y$$

$$\boxed{\hat{\beta} = \left(X^T X \right)^{-1} X^T y}$$

OLS estimate

Correlated and orthogonal regressors



$$y = x_1\beta_1 + x_2\beta_2 + e$$
$$\beta_1 = \beta_2 = 1$$

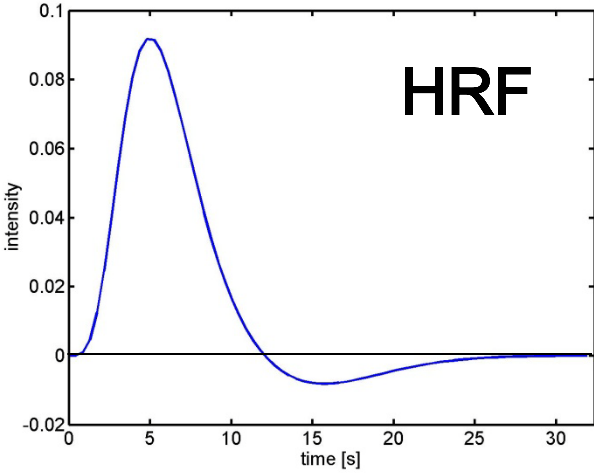
$$y = x_1\beta_1 + x_2^*\beta_2^* + e$$
$$\beta_1 > 1; \beta_2^* = 1$$

Correlated regressors = explained variance is shared between regressors

When x_2 is orthogonalized with regard to x_1, only the parameter estimate for x_1 changes, not that for x_2!

What are the problems of this model?

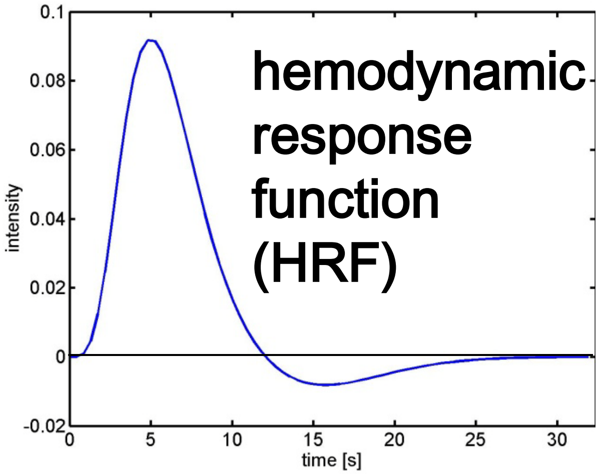
1. BOLD responses have a delayed and dispersed form.

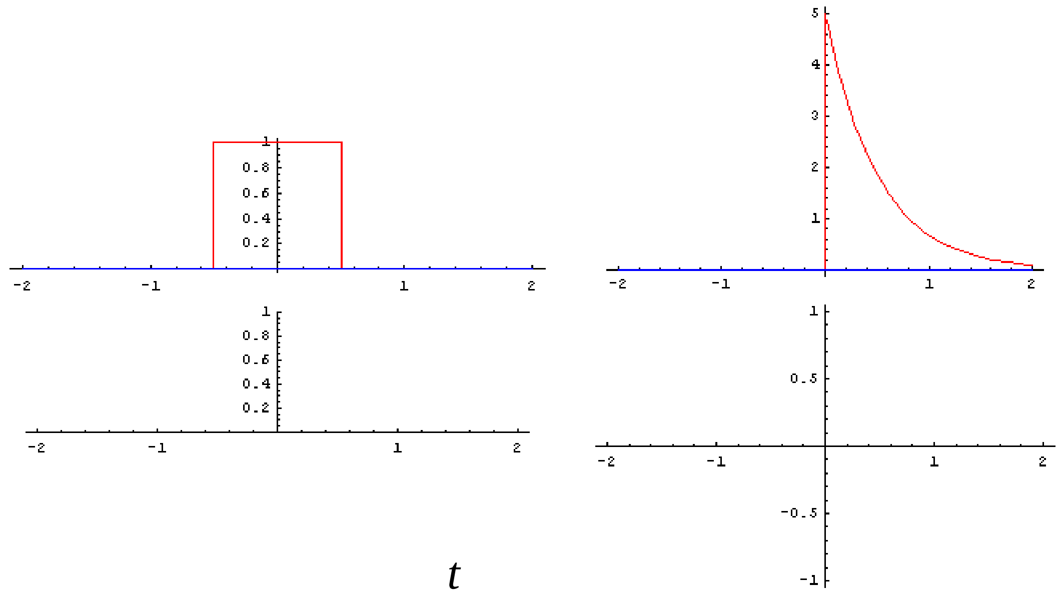


2. The BOLD signal includes substantial amounts of low-frequency noise.

3. The data are serially correlated (temporally autocorrelated) → this violates the assumptions of the noise model in the GLM

Problem 1: Shape of BOLD response
Solution: Convolution model





$$f \otimes g(t) = \int_0^t f(\tau) g(t-\tau) d\tau$$

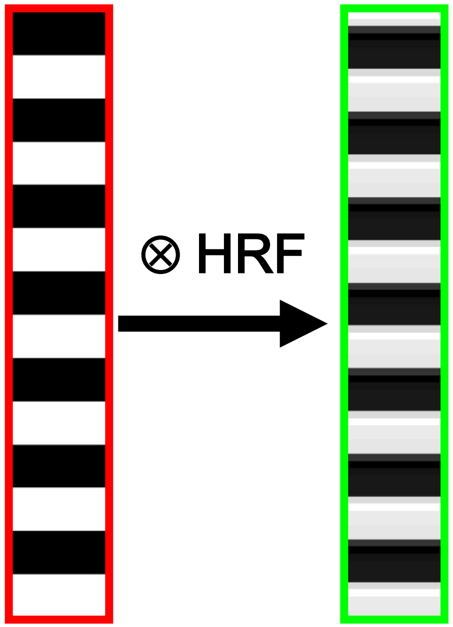
The response of a linear time-invariant (LTI) system is the convolution of the input with the system's response to an impulse (delta function).

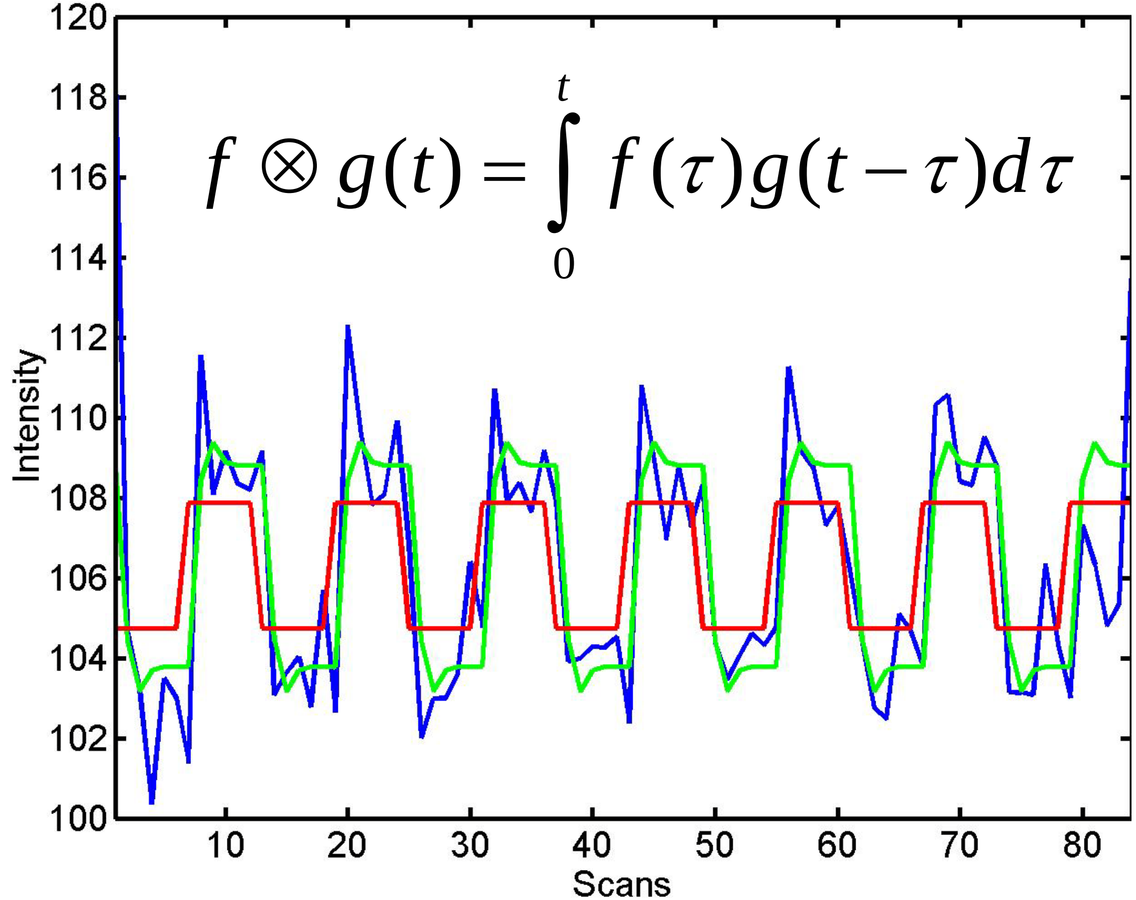
expected BOLD response

= input function ⊗ impulse response function (HRF)

Convolution model of the BOLD response

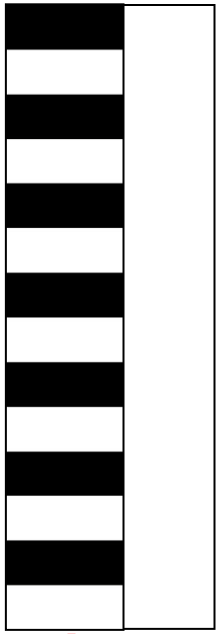
Convolve stimulus function with a canonical hemodynamic response function (HRF):

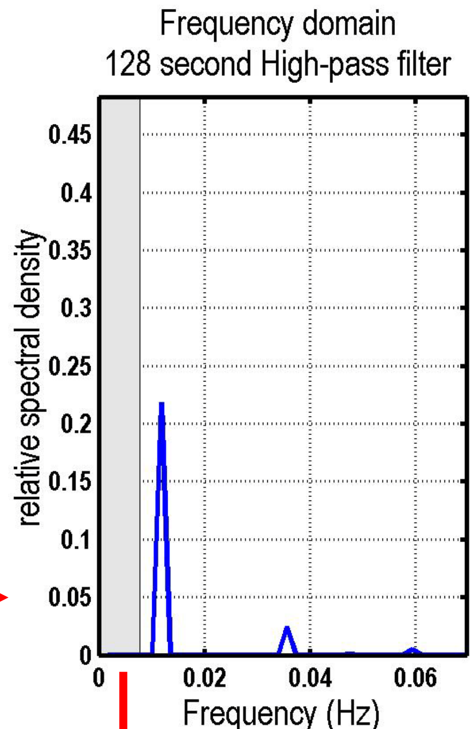




$$f \otimes g(t) = \int_0^t f(\tau)g(t-\tau)d\tau$$

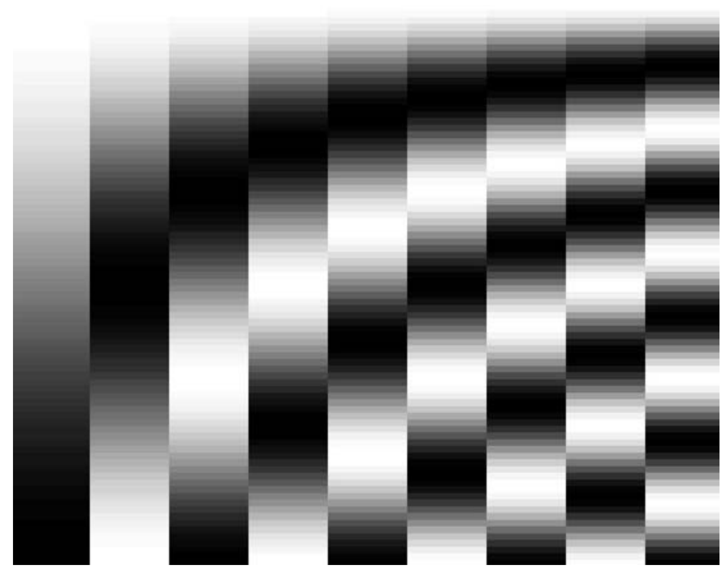
Problem 2: Low-frequency noise
Solution: High pass filtering





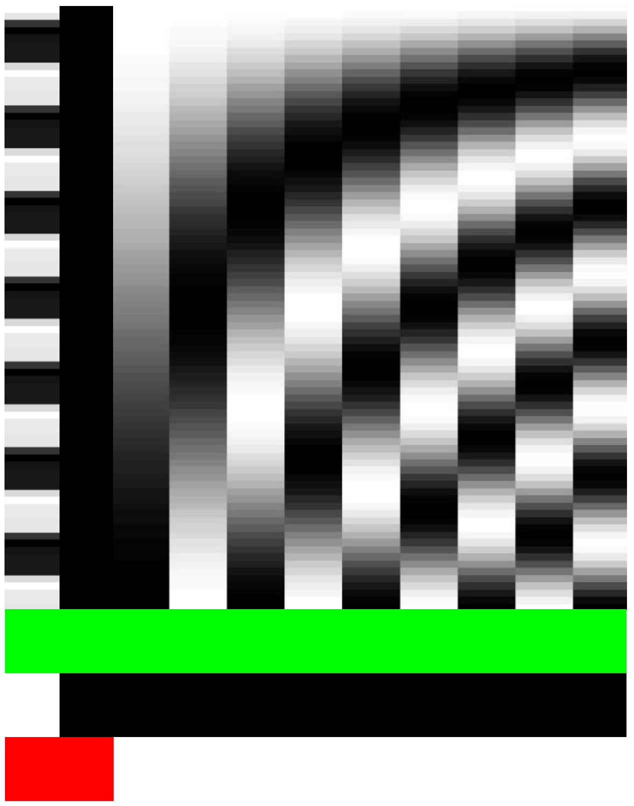
$$Sy = SX\beta + Se$$

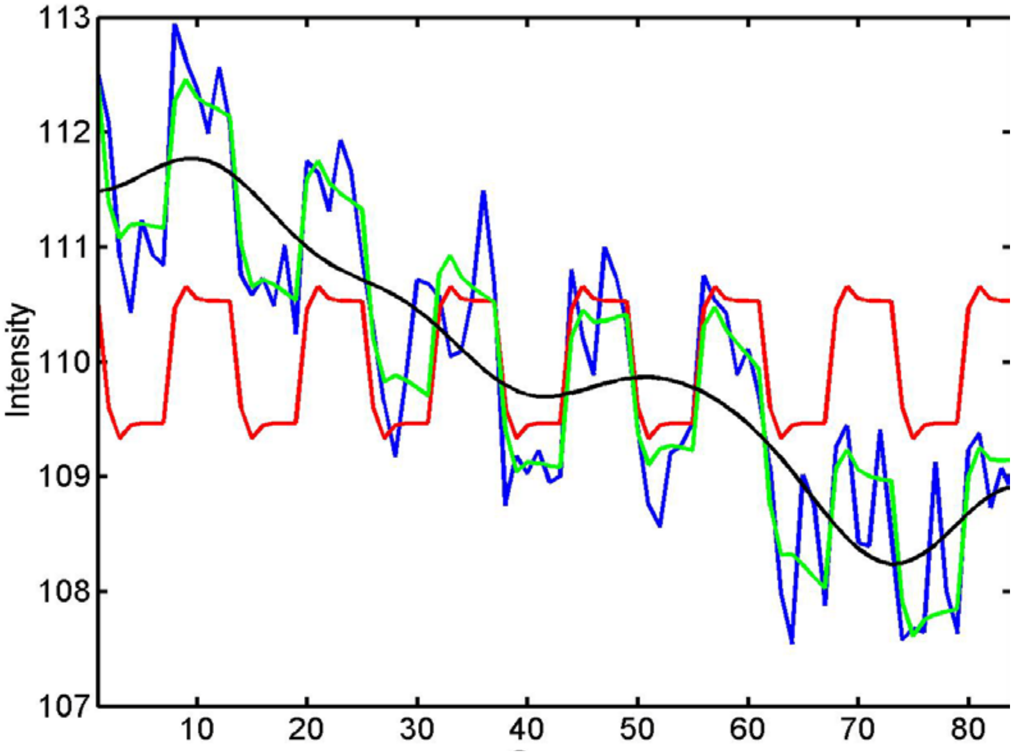
S = residual forming matrix of DCT set



discrete cosine transform (DCT) set

High pass filtering: example





blue = data

black = mean + low-frequency drift

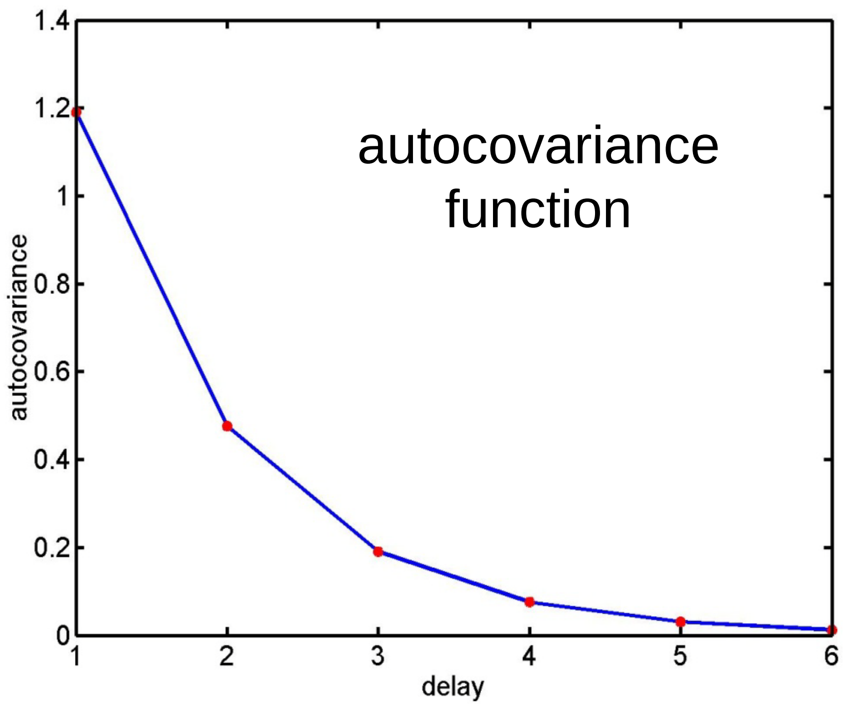
green = predicted response, taking into account low-frequency drift

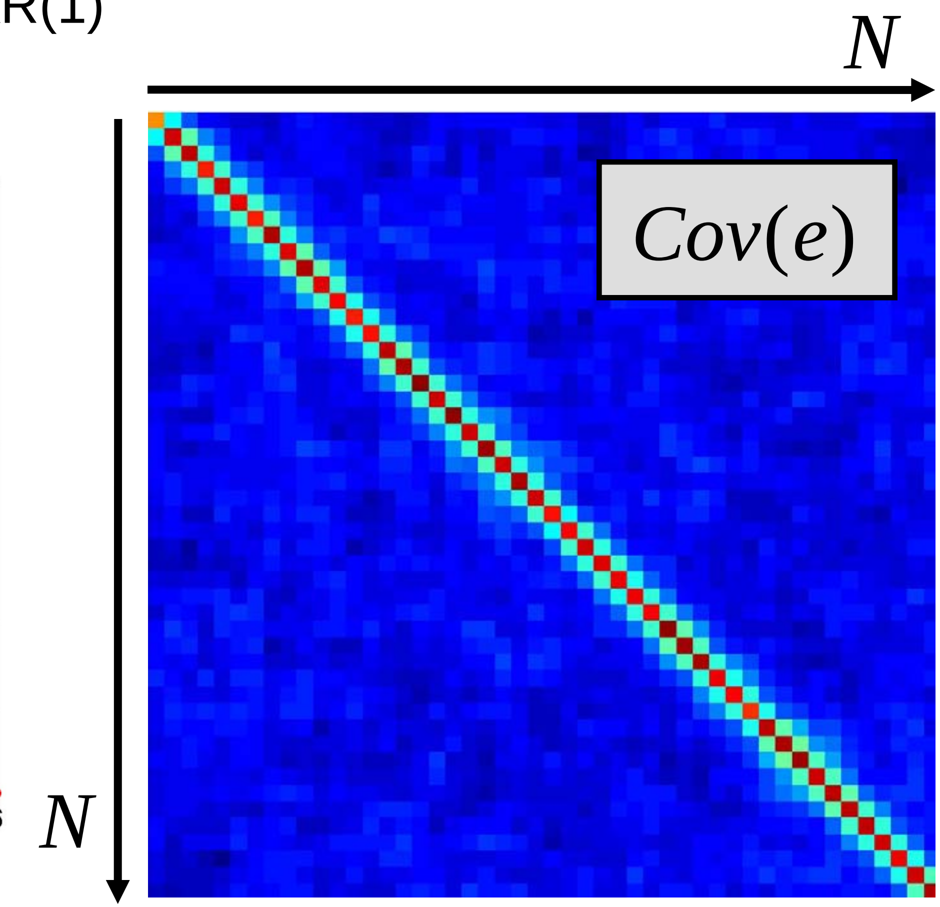
red = predicted response, NOT taking into account low-frequency drift

Problem 3: Serial correlations

$$e_t = a e_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0, \sigma^2)$$

1st order autoregressive process: AR(1)





Dealing with serial correlations

- ***Pre-colouring***: impose some known autocorrelation structure on the data (filtering with matrix W) and use Satterthwaite correction for df's.

- ***Pre-whitening***:

  1. Use an enhanced noise model with multiple error covariance components, i.e. $e \sim N(0, \sigma^2 V)$ instead of $e \sim N(0, \sigma^2 I)$.

  2. Use estimated serial correlation to specify filter matrix W for whitening the data.

$$\boxed{Wy = WX\beta + We}$$

How do we define W?

- Enhanced noise model

$$e \sim N(0, \sigma^2 V)$$

- Remember linear transform for Gaussians

$$x \sim N(\mu, \sigma^2), y = ax$$
$$\Rightarrow y \sim N(a\mu, a^2\sigma^2)$$

- Choose W such that error covariance becomes spherical

$$We \sim N(0, \sigma^2 W^2 V)$$
$$\Rightarrow W^2 V = I$$
$$\Rightarrow W = V^{-1/2}$$

- **Conclusion: W is a simple function of V**
  $\Rightarrow$ so how do we estimate V?

$$Wy = WX\beta + We$$

Estimating *V*: Multiple covariance components

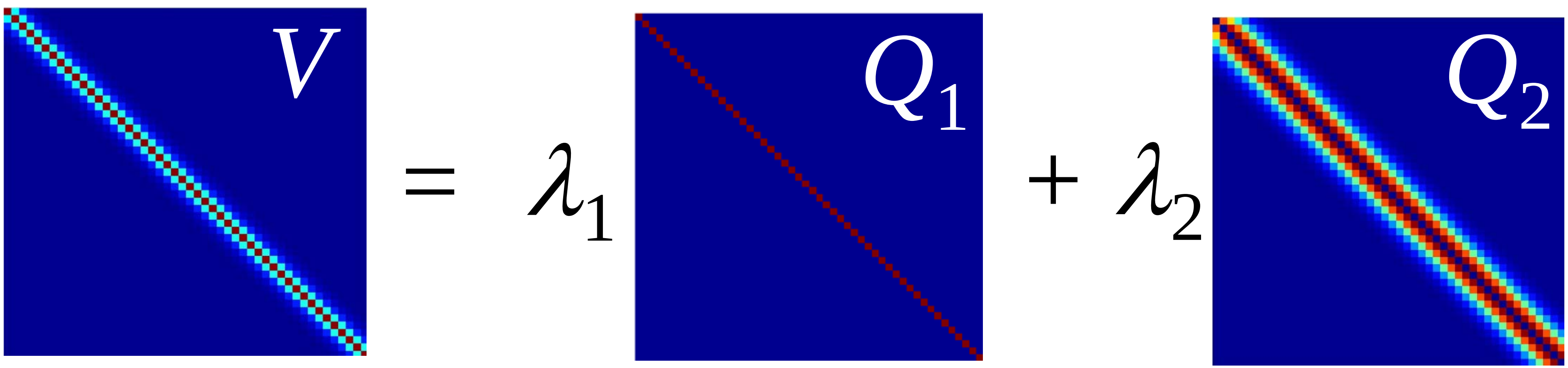
$$e \sim N(0, \sigma^2 V)$$

enhanced noise model

$$V \propto Cov(e)$$

$$V = \sum \lambda_i Q_i$$

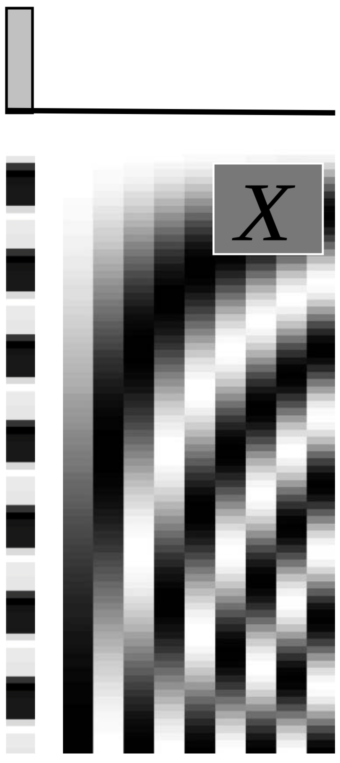
error covariance components Q and hyperparameters λ



Estimation of hyperparameters λ with ReML (restricted maximum likelihood).

Contrasts & statistical parametric maps

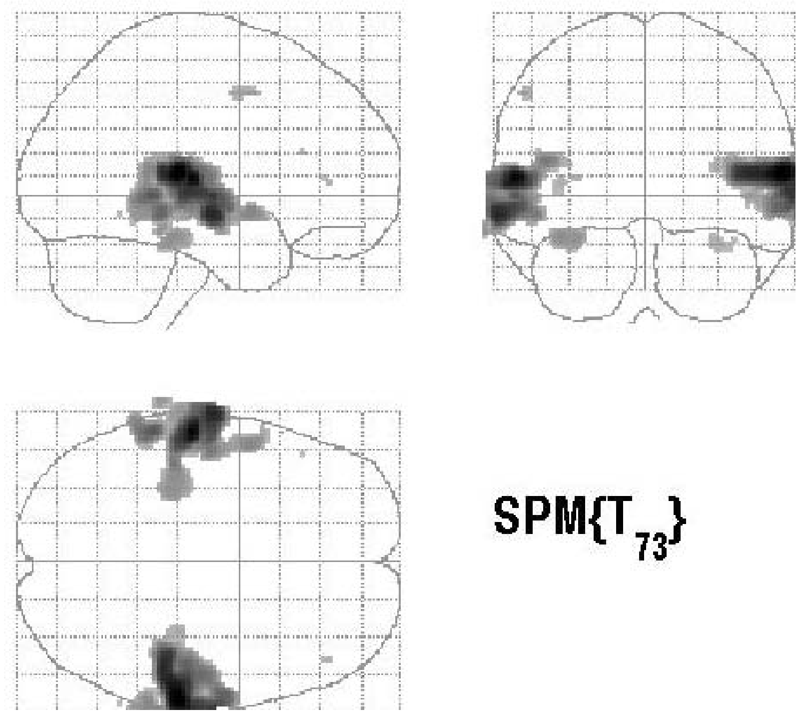
c = 1 0 0 0 0 0 0 0 0 0 0

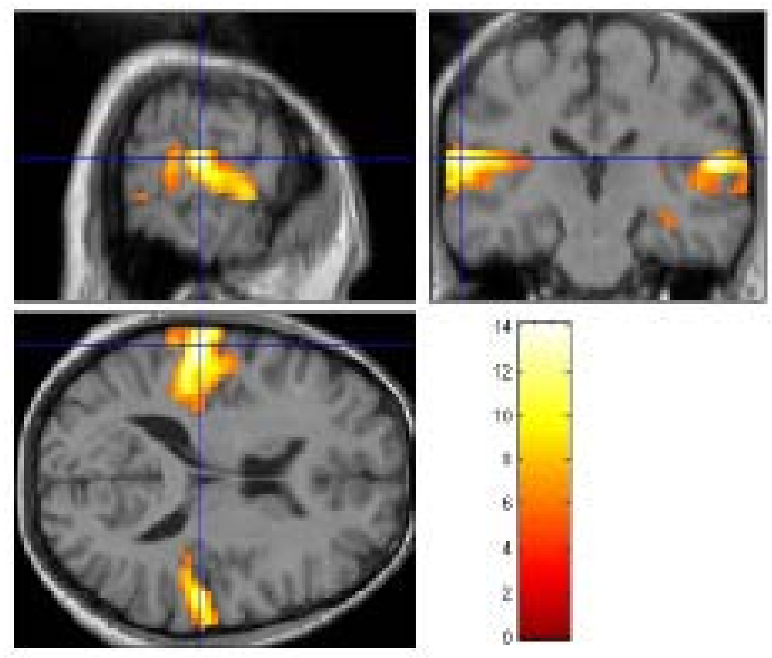


Q: activation during listening ?

Null hypothesis: $\beta_1 = 0$

$$t = \frac{c^T \hat{\beta}}{Std(c^T \hat{\beta})}$$



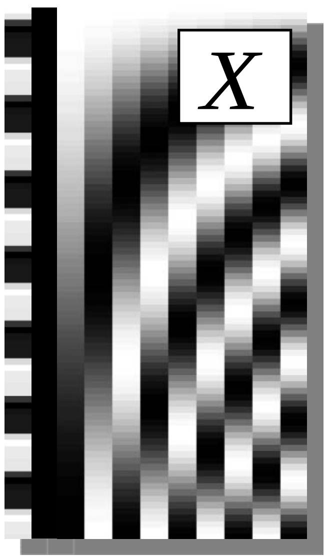


t-statistic based on ML estimates

$$Wy = WX\beta + We$$

$$\hat{\beta} = (WX)^{+}Wy$$

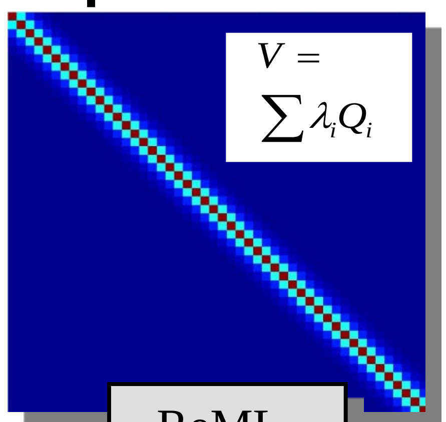
$$c = 1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0$$



$$t = \frac{c^T\hat{\beta}}{\hat{std}(c^T\hat{\beta})}$$

$$W = V^{-1/2}$$

$$\sigma^2 V = Cov(e)$$



ReML-estimates

$$\hat{std}(c^T\hat{\beta}) = \sqrt{\hat{\sigma}^2 c^T (WX)^{+} (WX)^{+T} c}$$

$$\hat{\sigma}^2 = \frac{\sum \left(Wy - WX\hat{\beta}\right)^2}{tr(R)}$$

$$R = I - WX(WX)^{+}$$

For brevity:

$$(WX)^{+} = (X^T WX)^{-1} X^T$$

Physiological confounds

- head movements
- arterial pulsations (particularly bad in brain stem)
- breathing
- eye blinks (visual cortex)
- adaptation effects, fatigue, fluctuations in concentration, etc.

Outlook: further challenges

- correction for multiple comparisons
- variability in the HRF across voxels
- slice timing
- limitations of frequentist statistics
  $\rightarrow$ Bayesian analyses
- GLM ignores interactions among voxels
  $\rightarrow$ models of effective connectivity

These issues are discussed in future lectures.

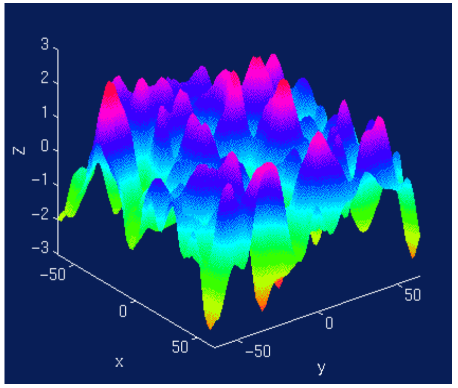
Correction for multiple comparisons

- Mass-univariate approach:
  We apply the GLM to each of a huge number of voxels (usually > 100,000).

- Threshold of $p<0.05 \rightarrow$ more than 5000 voxels significant by chance!

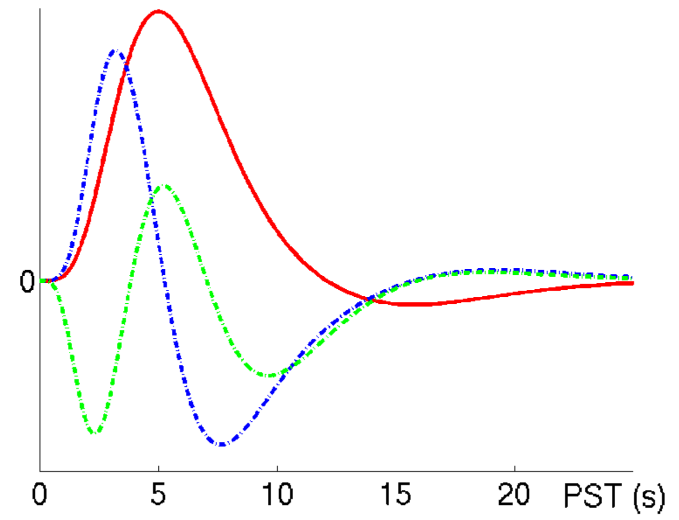
- Massive problem with multiple comparisons!

- Solution: Gaussian random field theory



Variability in the HRF

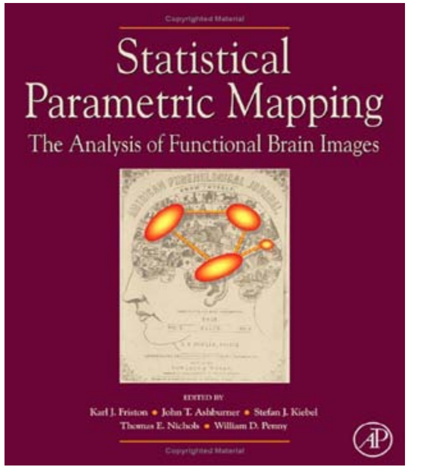
- HRF varies substantially across voxels and subjects
- For example, latency can differ by ± 1 second
- Solution: use multiple basis functions
- See talk on event-related fMRI



Summary

- Mass-univariate approach: same GLM for each voxel
- GLM includes all known experimental effects and confounds
- Convolution with a canonical HRF
- High-pass filtering to account for low-frequency drifts
- Estimation of multiple variance components (e.g. to account for serial correlations)

Bibliography



- Friston, Ashburner, Kiebel, Nichols, Penny (2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.

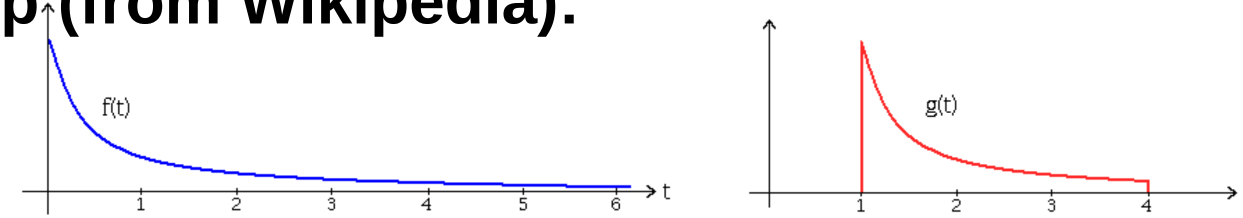
- Christensen R (1996) *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer.

- Friston KJ et al. (1995) Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 2: 189-210.

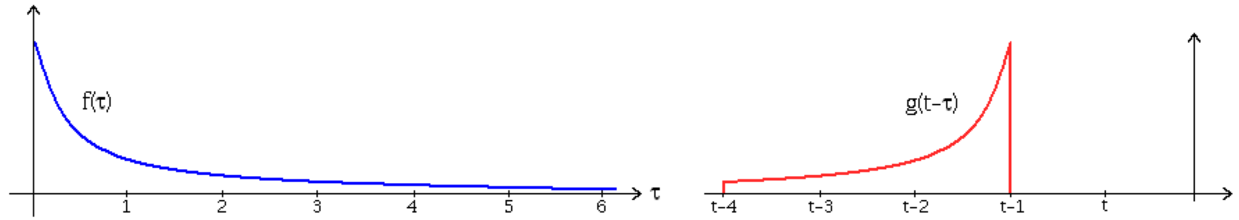
Supplementary slides

Convolution step-by-step (from Wikipedia):

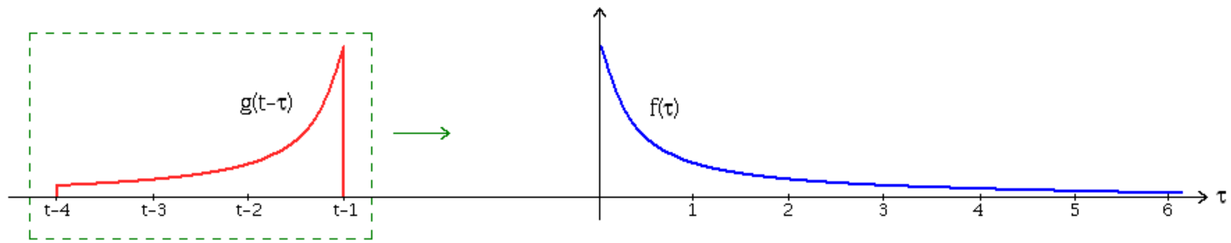
1. **Express each function in terms of a dummy variable τ.**



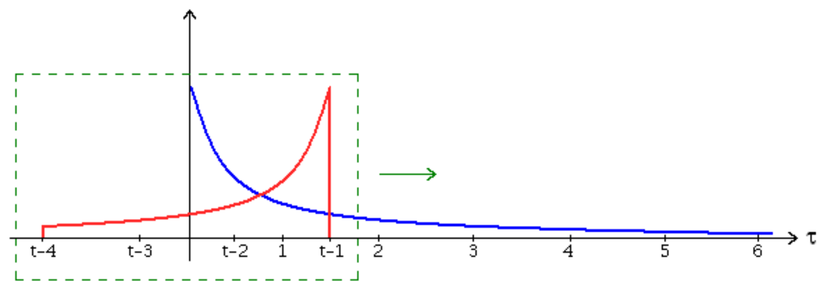
2. **Reflect one of the functions: $g(\tau) \rightarrow g(-\tau)$.**



3. **Add a time-offset, t, which allows $g(t-\tau)$ to slide along the τ-axis.**



4. Start t at -∞ and slide it all the way to +∞. Wherever the two functions intersect, find the integral of their product. In other words, compute a sliding, weighted-average of function $f(\tau)$, where the weighting function is $g(-\tau)$.



The resulting waveform (not shown here) is the convolution of functions f and g. If f(t) is a unit impulse, the result of this process is simply g(t), which is therefore called the impulse response.