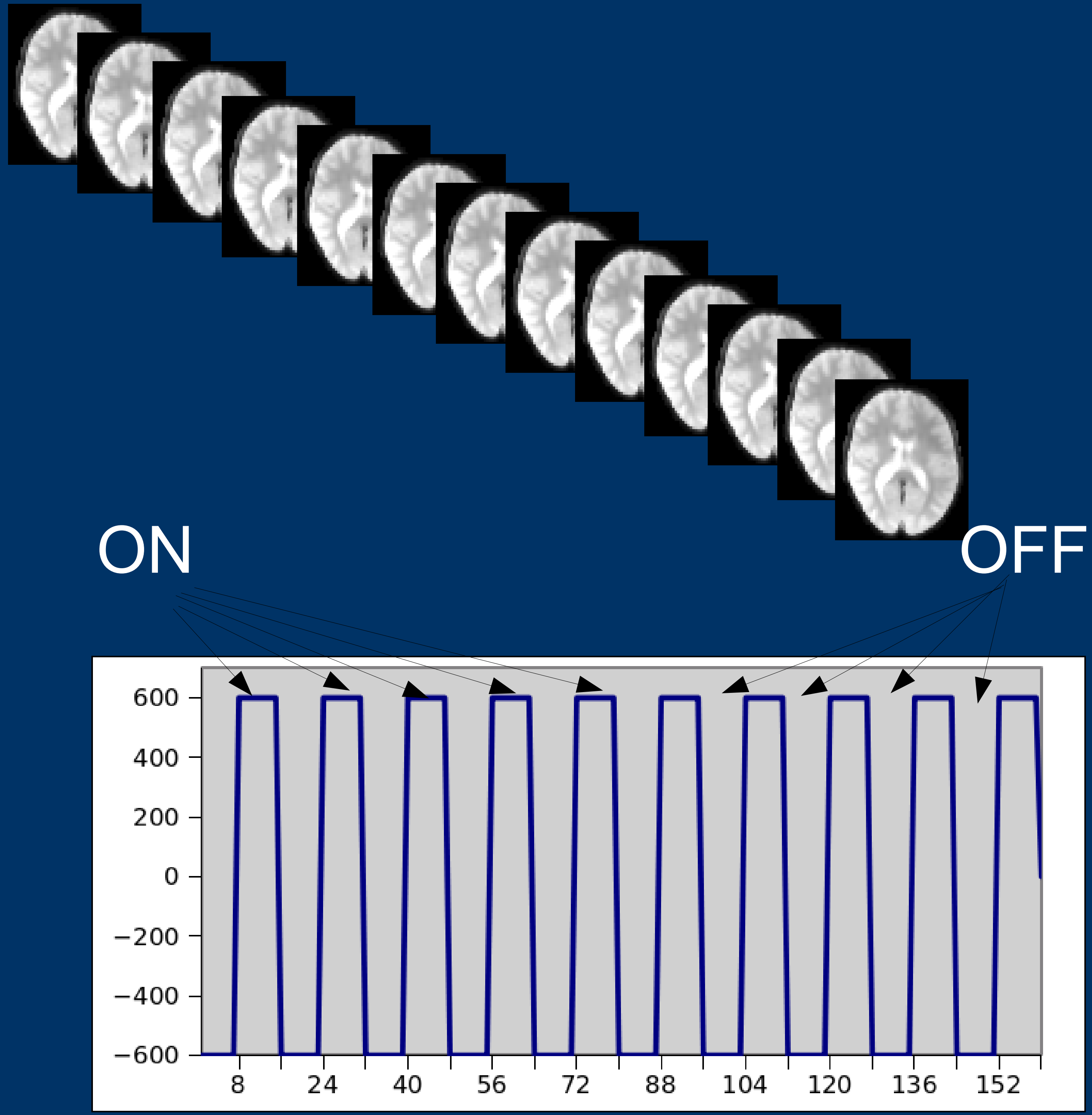
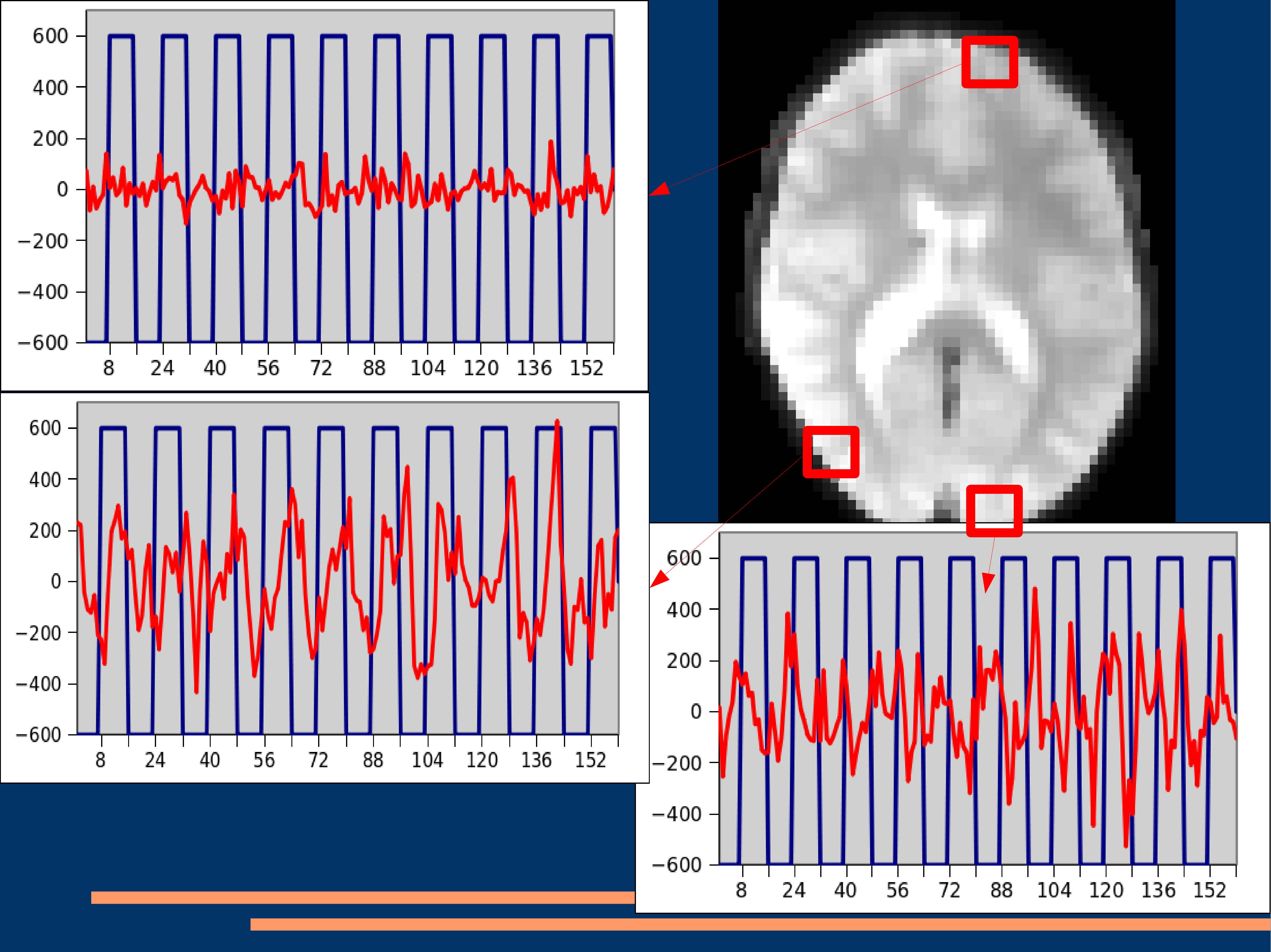
Data Analysis I: Single Subject



ON
OFF
600
400
200
0
−200
−400
−600
8
24
40
56
72
88
104
120
136
152



600
400
200
0
−200
−400
−600
8
24
40
56
72
88
104
120
136
152

The General Linear Model (GLM)

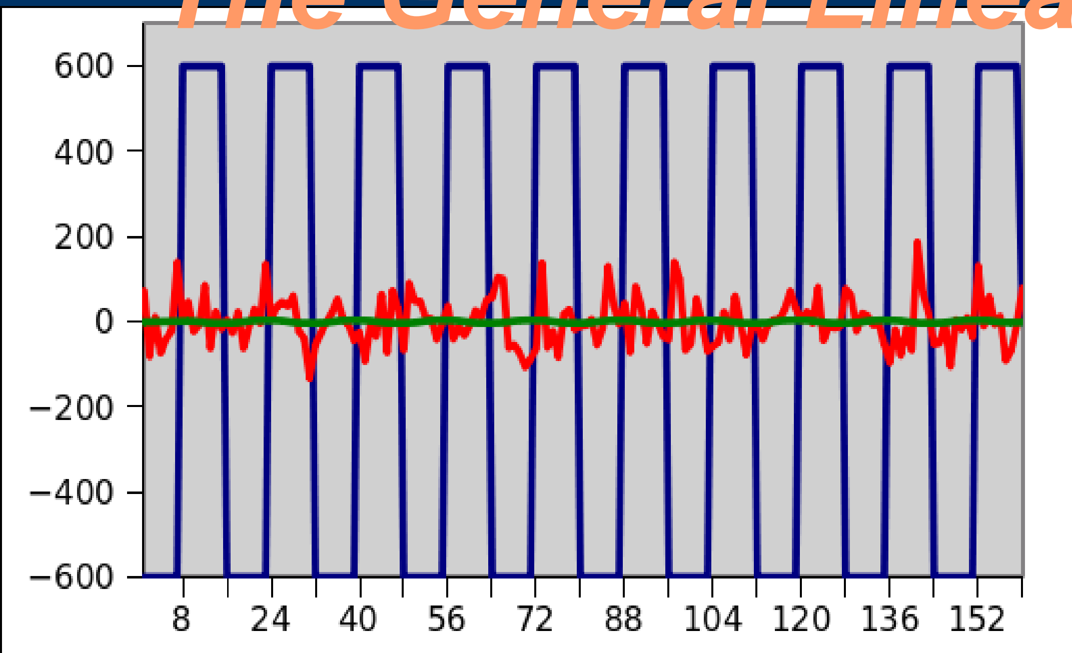
$$y = X \times \beta + \varepsilon$$

fMRI Signal	=	Design Matrix	x	Betas	+	Residuals
"our data"	=	*"what we CAN explain"*	x	*"how much of it we CAN explain"*	+	*"what we CANNOT explain"*

→ If a voxel "cares" about the task, then you should be able to predict its time-course from the onset and offset of the task.

→ β captures "how much" of 'y' your task/design is predicting.

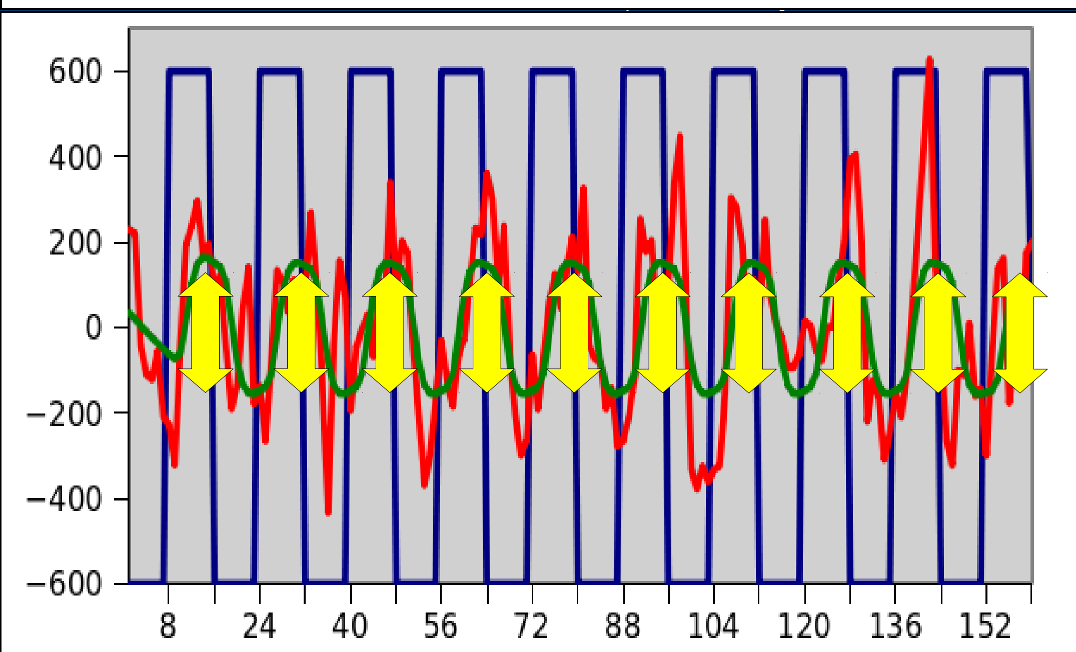
The General Linear Model (GLM)

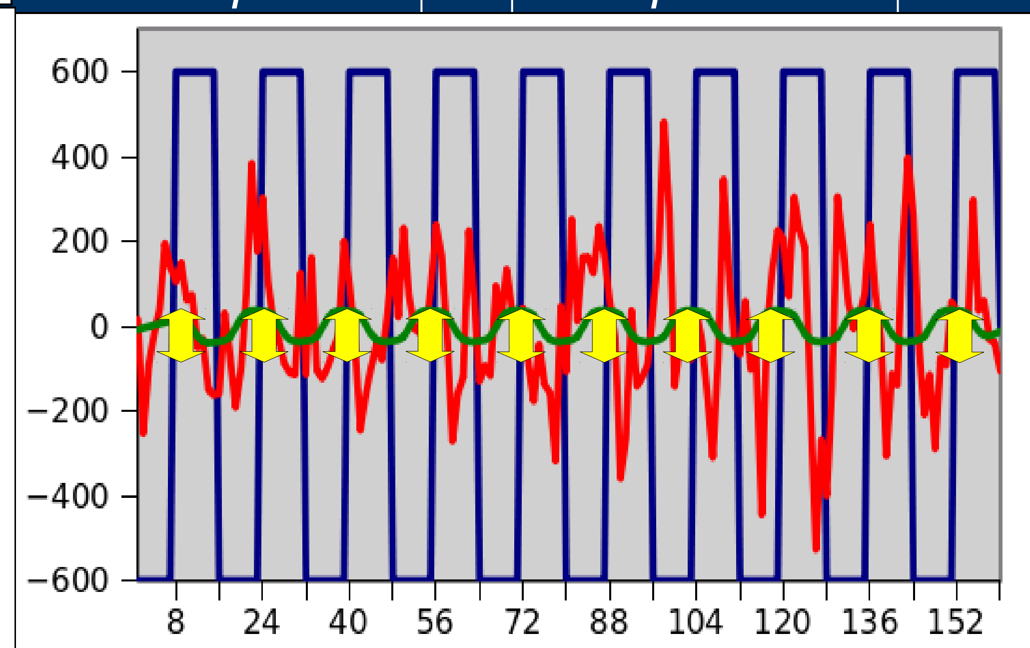


$x \quad \beta \quad + \quad \varepsilon$

x Betas + Residuals

x *“how much of it we CAN explain”* + *“what we CANNOT explain”*





The General Linear Model (GLM)

$$y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

y_i Reaction time

X_1 Age

β_0, β_1 are the model parameters

ε_i is the random error; how observation deviates from population mean

The General Linear Model (GLM)

$$y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

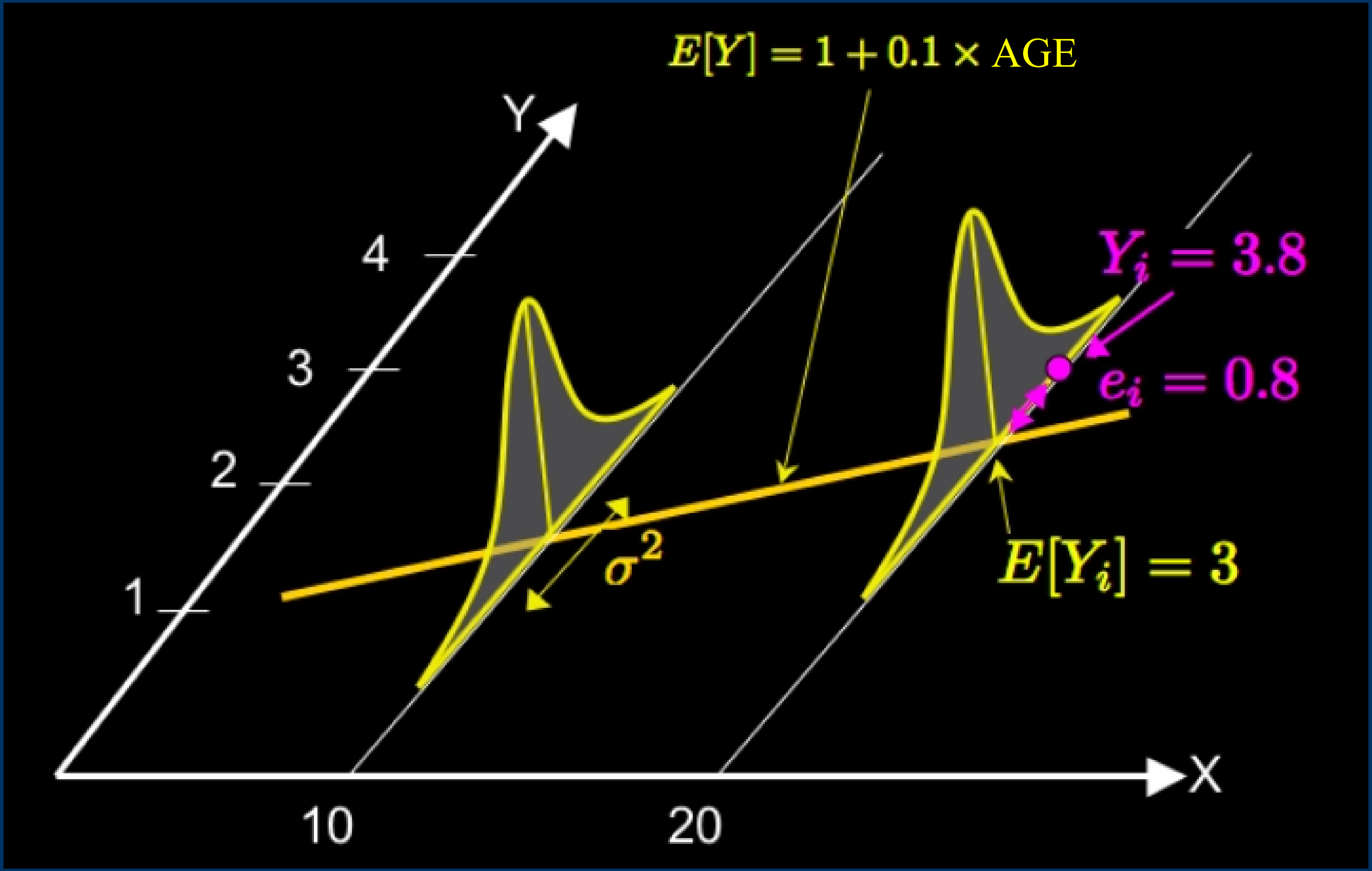
Fixed: $\beta_0 + \beta_1 X_1$

- Mean of Y_i ($E[Y_i]$)

Random: ε_i

- The variability of Y_i
- (needs to have specific properties)

The General Linear Model (GLM)



Source: J Mumford

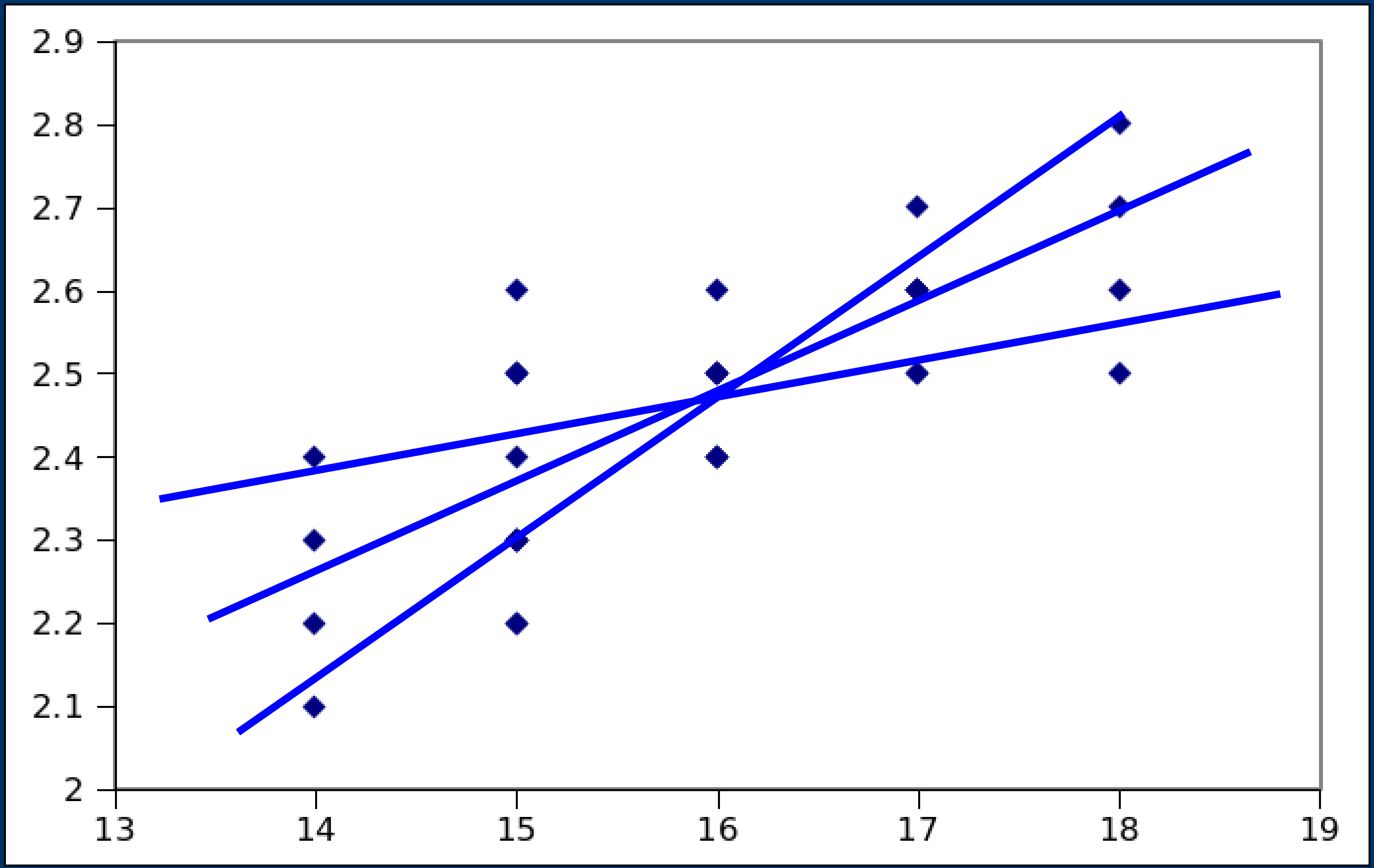
Statistical Analysis

•Conceptually the analysis is divided into 2 steps:

1. ***Estimation:*** the magnitude of the effect of interest (β)
2. ***Inference:*** assess significance of the estimated parameter

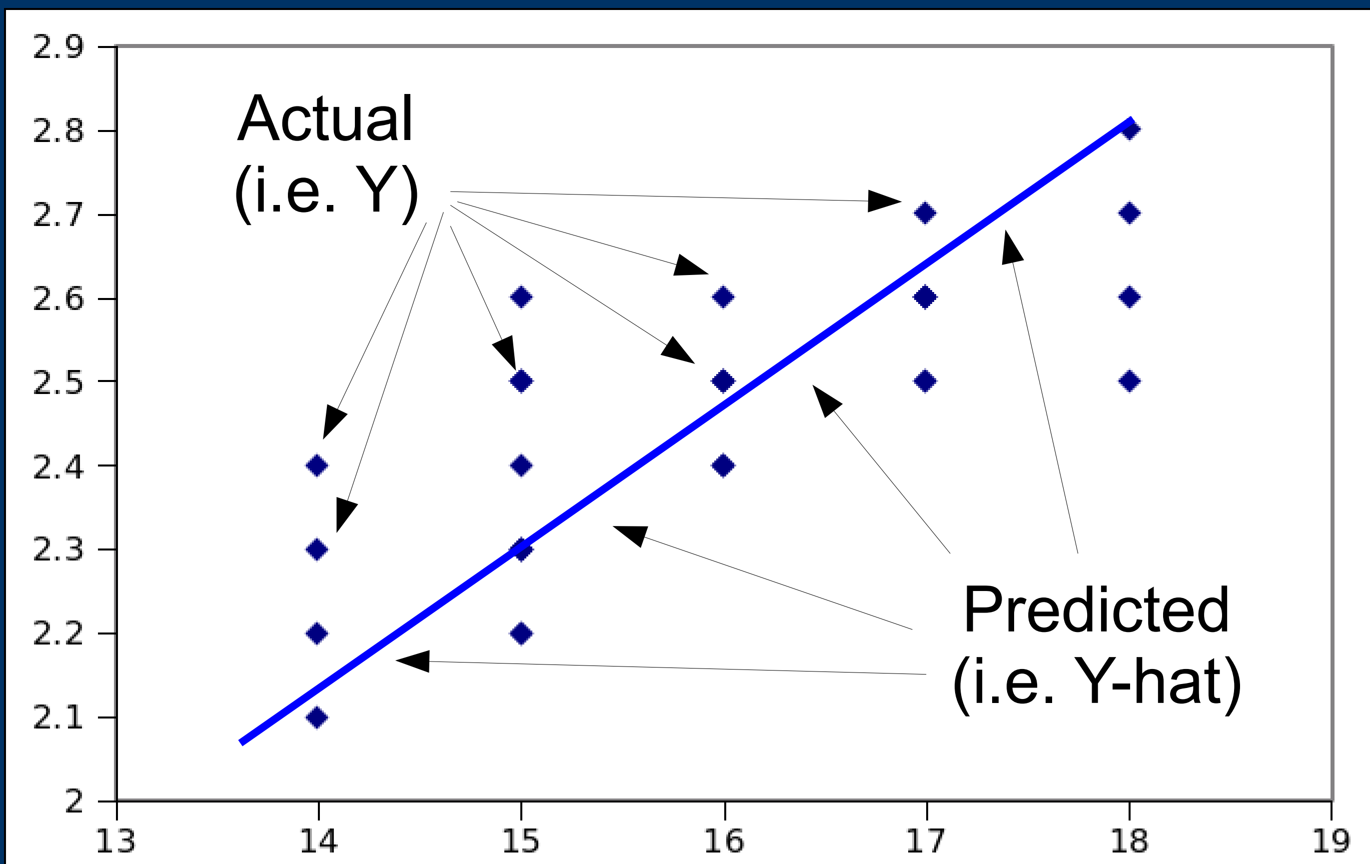
Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



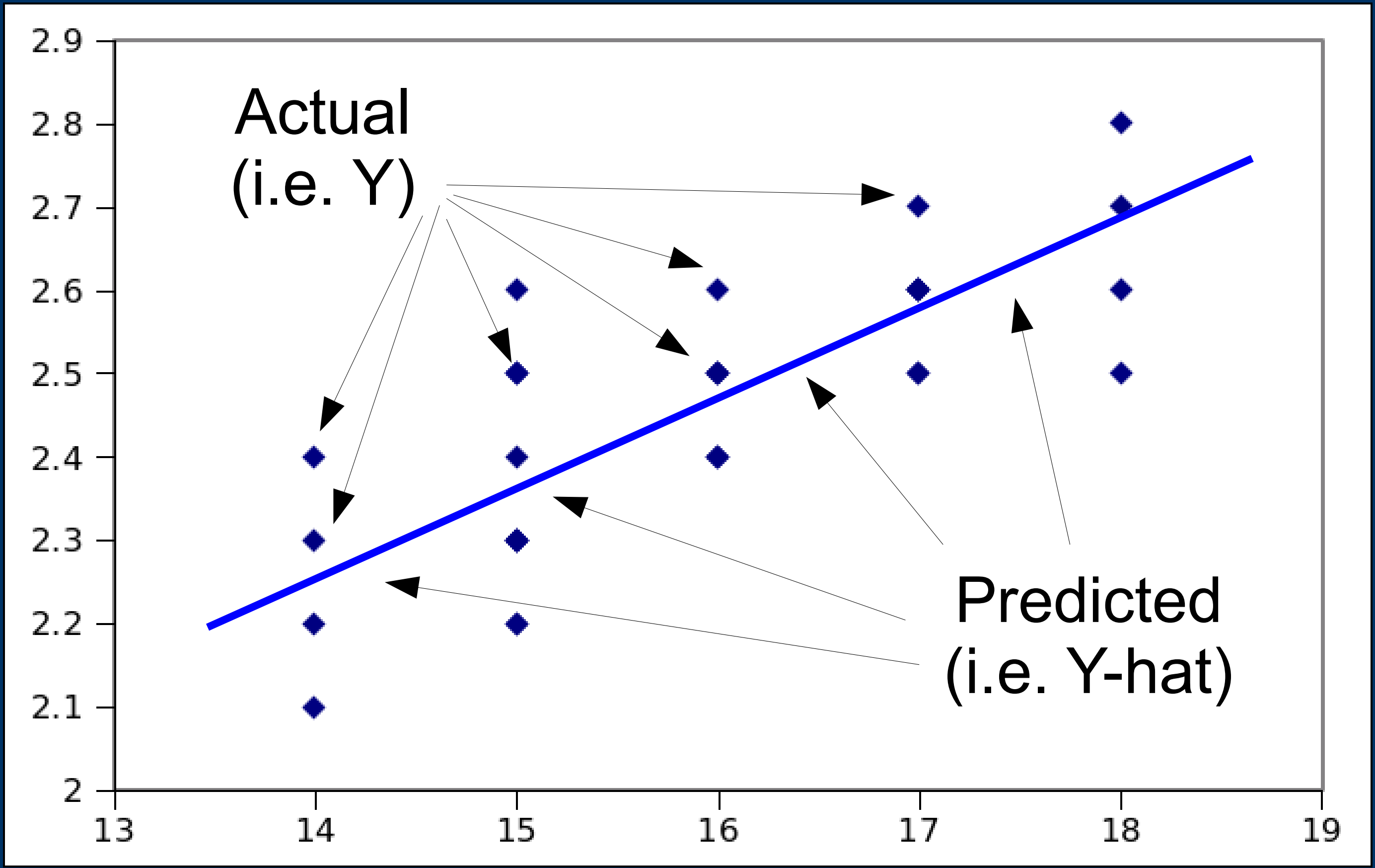
Minimize error:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_1)^2$$

Ordinary Least Squares (OLS)

e.g. What is the relationship tying years of MATLB experience and coding efficiency?



Minimize error:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_1)^2$$

Least Squares Difference

- Minimize error

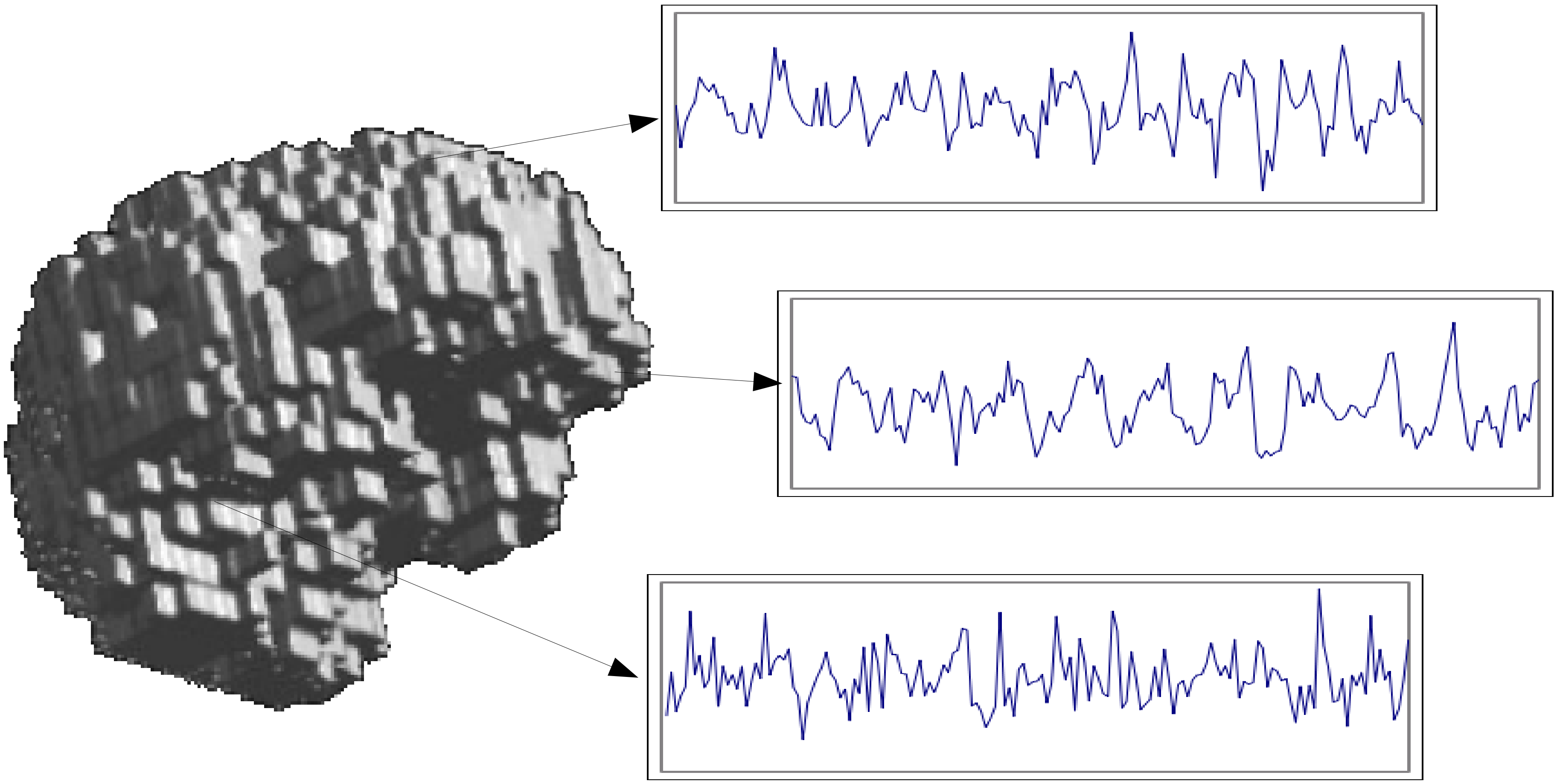
$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \beta_i X_i)^2$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}$$

$$\hat{\beta}_0 = \hat{Y} - \hat{\beta}_1 \bar{X}$$

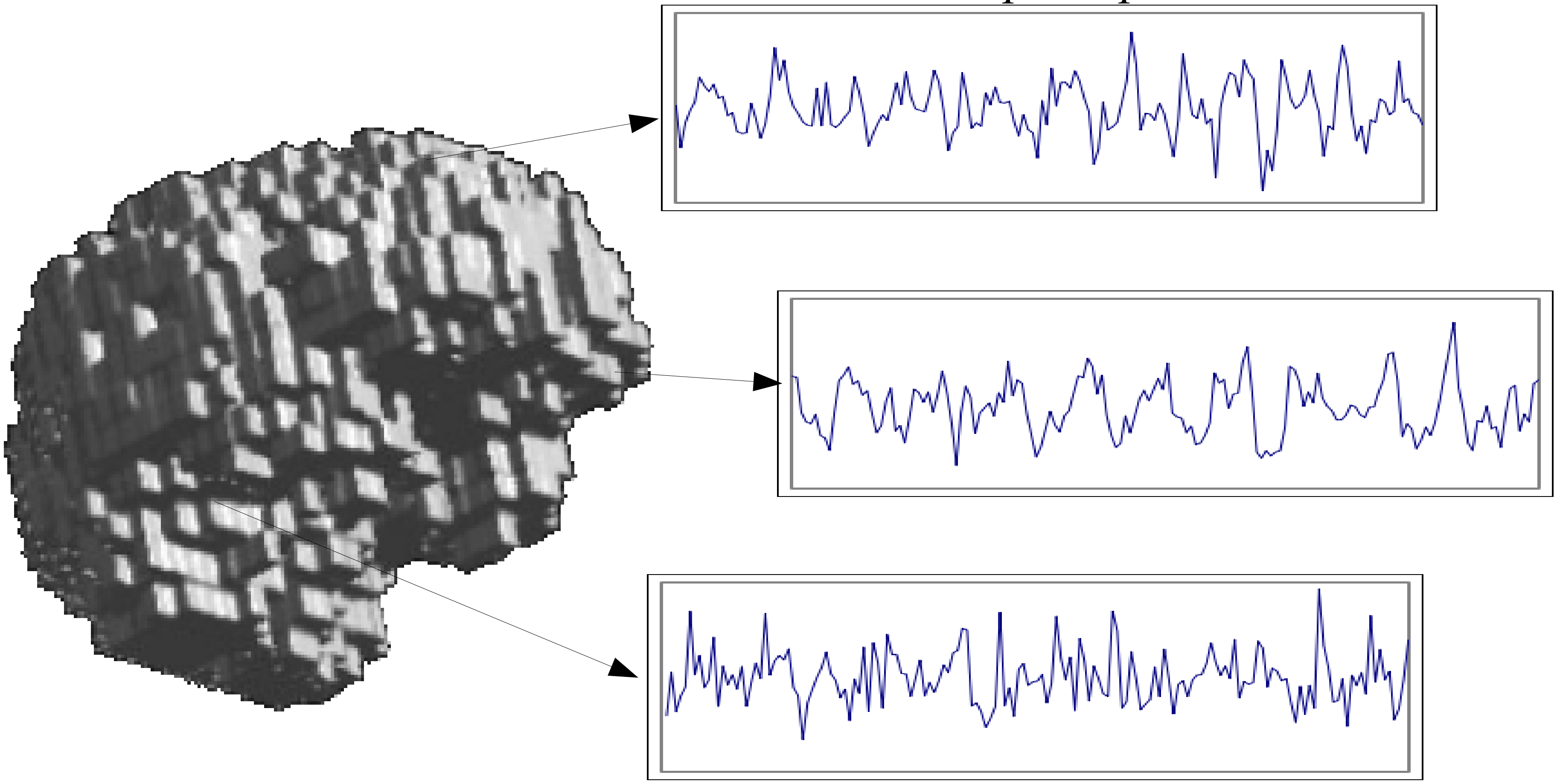
Massive Univariate Approach

$$y = \beta_0 + \beta_1 X_1 + \epsilon$$



Massive Univariate Approach

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon$$



Voxel-wise GLM

$$y_1 = \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{2,1} + \cdots + \beta_p X_{p,1} + \epsilon_1$$
$$y_2 = \beta_0 + \beta_1 X_{1,2} + \beta_2 X_{2,2} + \cdots + \beta_p X_{p,2} + \epsilon_2$$
$$y_3 = \beta_0 + \beta_1 X_{1,3} + \beta_2 X_{2,3} + \cdots + \beta_p X_{p,3} + \epsilon_3$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 X_{1,n} + \beta_2 X_{2,n} + \cdots + \beta_p X_{p,n} + \epsilon_n$$

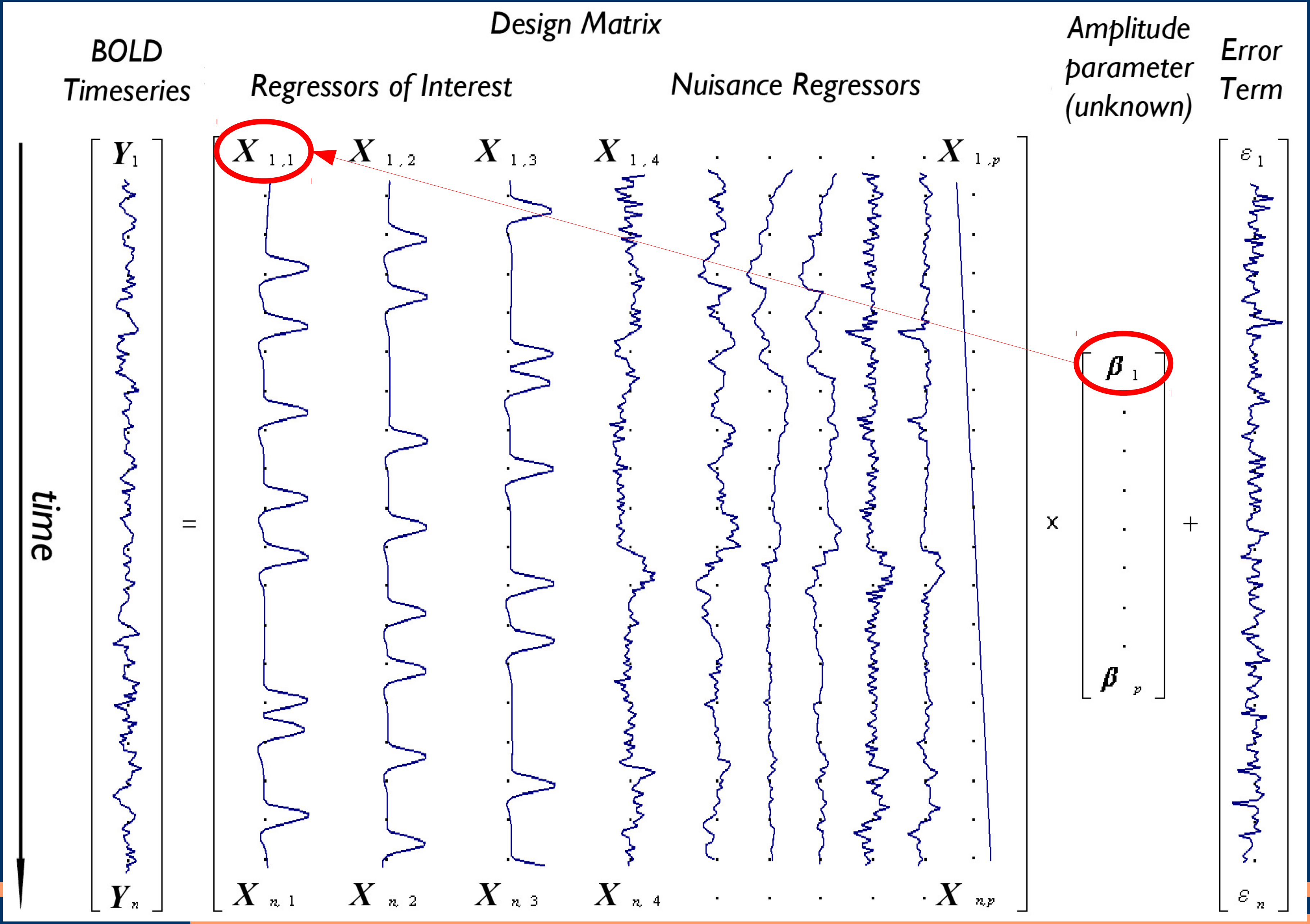
Voxel-wise GLM

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{2,1} & X_{3,1} & \cdots & X_{p,1} \\ X_{1,2} & X_{2,2} & X_{3,2} & \cdots & X_{p,2} \\ X_{1,3} & X_{2,3} & X_{3,3} & \cdots & X_{p,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1,n} & X_{2,n} & X_{3,n} & \cdots & X_{p,n} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}$$

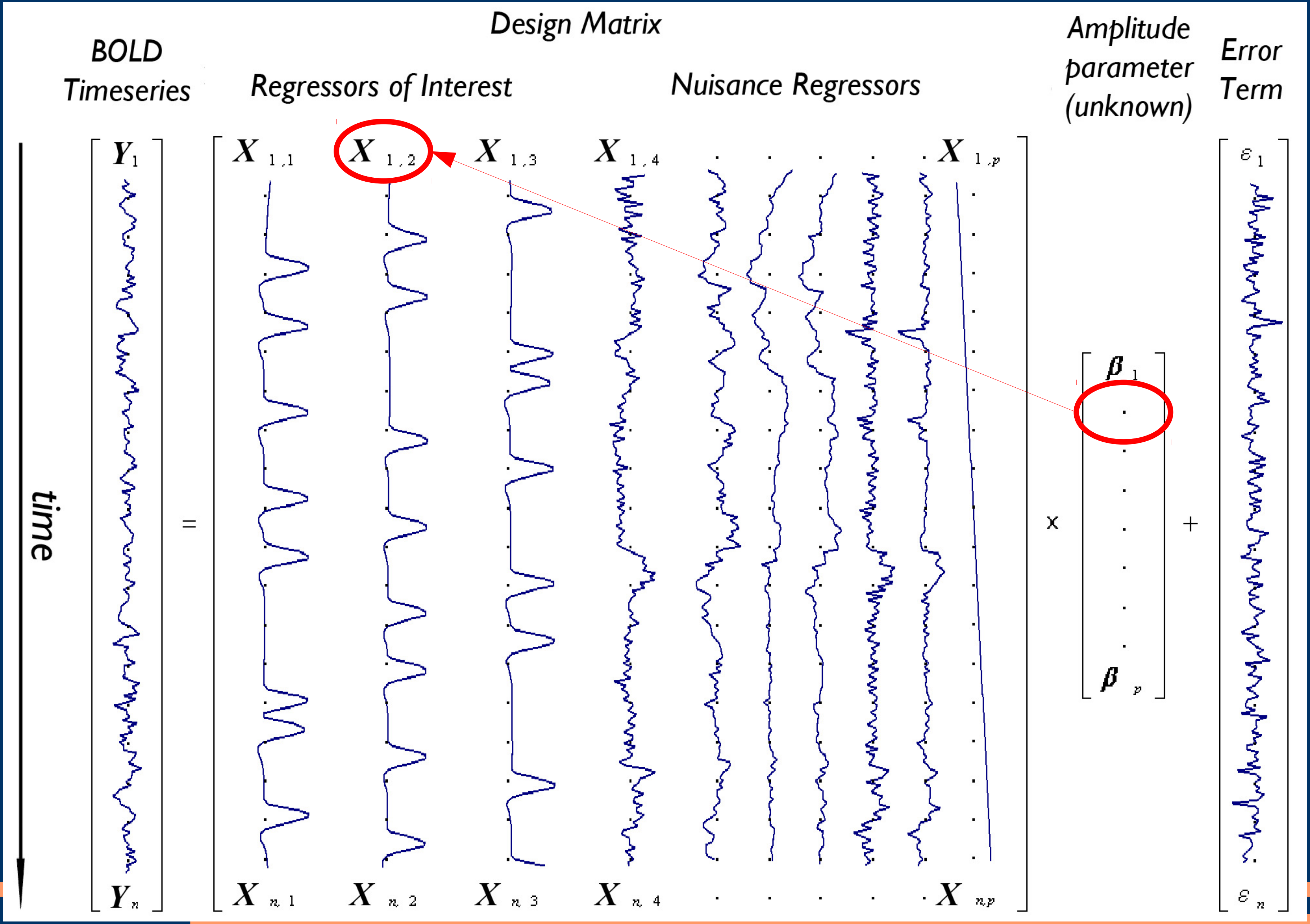
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

(nx1) (nxp) (px1) (nx1)

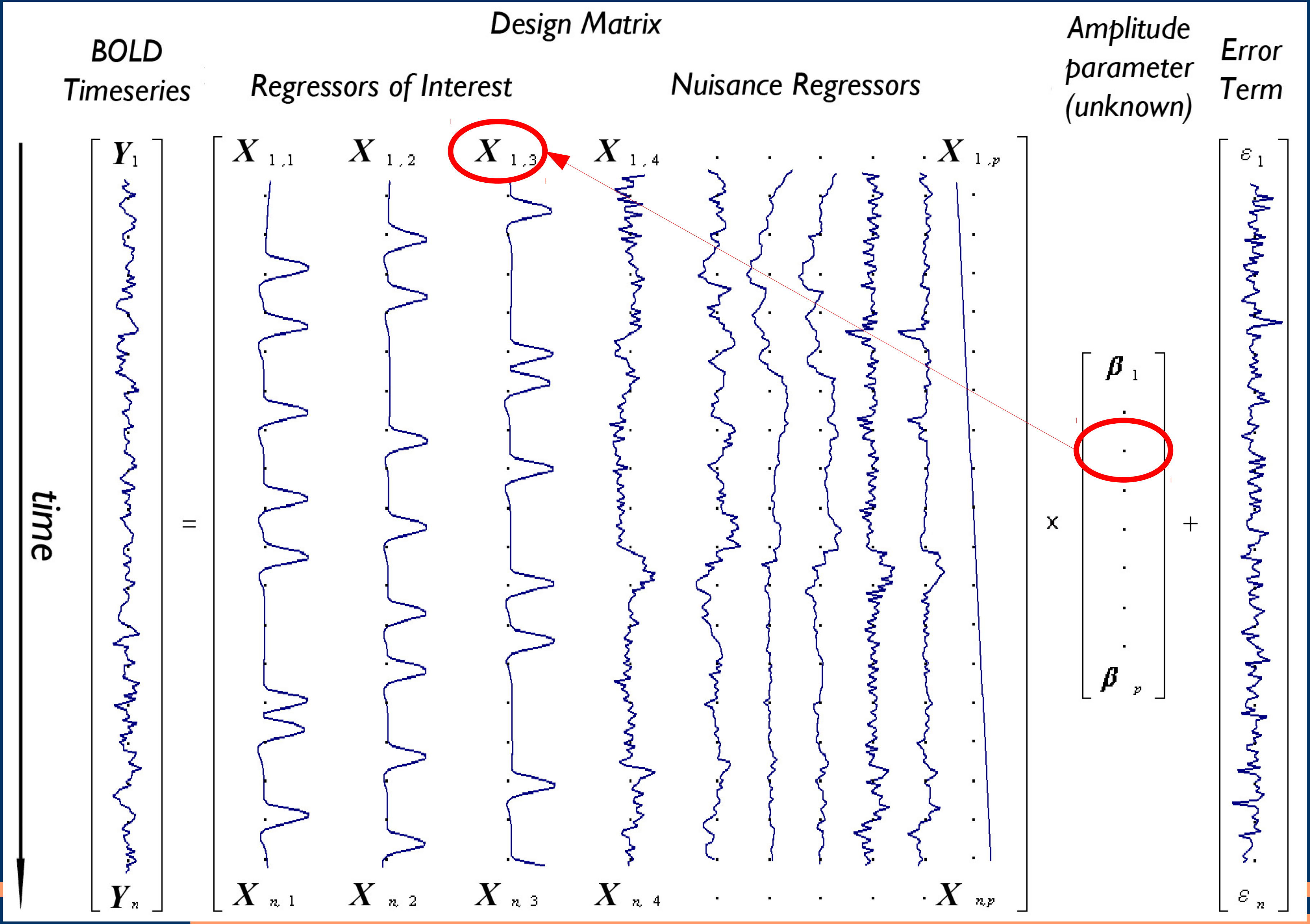
Voxel-wise GLM



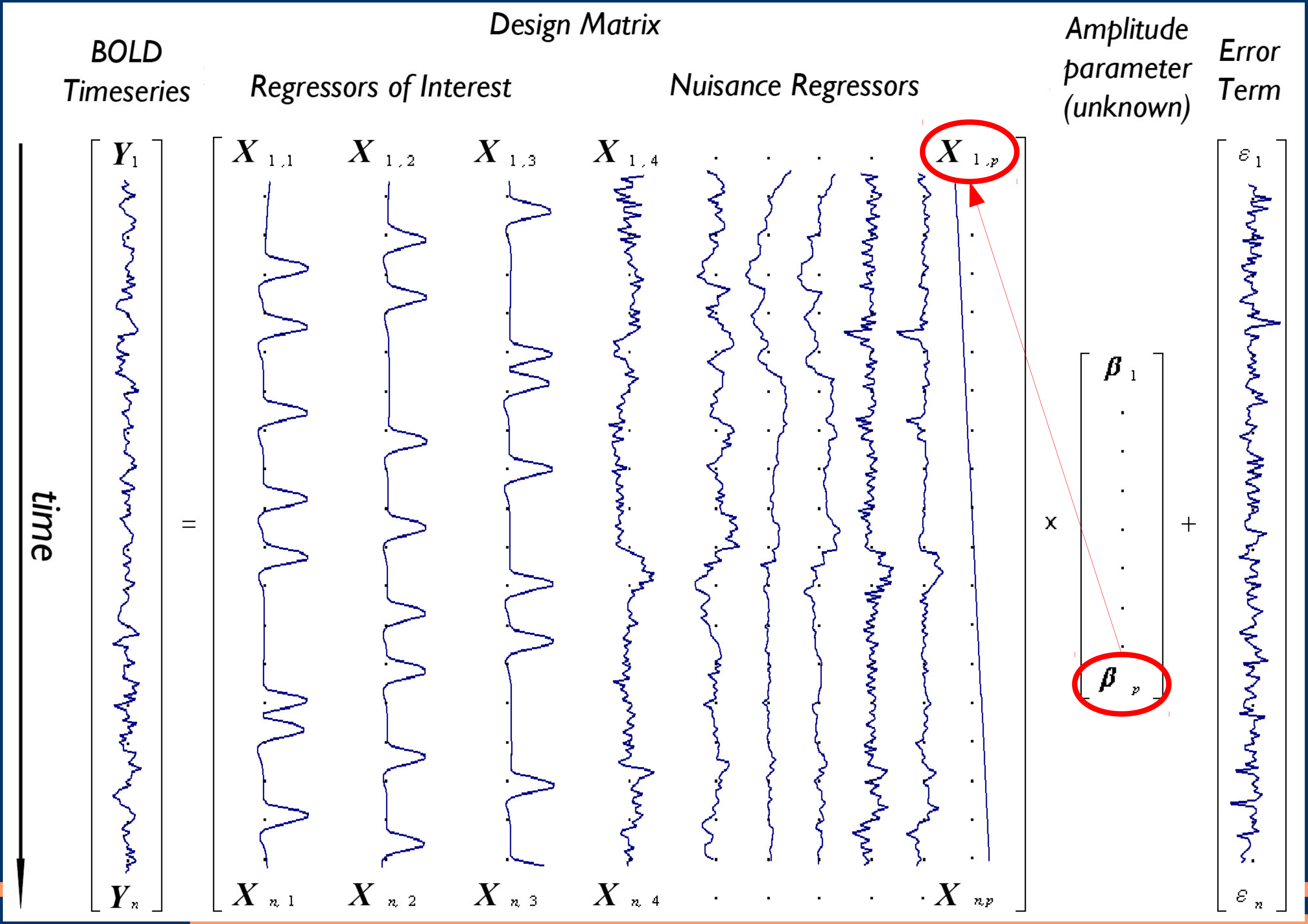
Voxel-wise GLM



Voxel-wise GLM



Voxel-wise GLM



Ordinary Least Squares (OLS)

$$Y = X\beta + \epsilon$$

$$Y = X\hat{\beta}$$

$$X^T Y = (X^T X)\hat{\beta}$$

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

Ordinary Least Squares (OLS)

$$Y = X\beta + \epsilon$$

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

$$\widehat{\boldsymbol{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (X^T X)^{-1}$$

$$t = \frac{\hat{\beta}}{\sqrt{\widehat{Var}(\beta)}} \qquad \sim T(N-p)$$

Ordinary Least Squares (OLS)

Effect of interest

Nature

$$t = \frac{\hat{\beta}}{\sqrt{\widehat{Var}(\hat{\beta})}} = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \sim T(N-p)$$

Subject, scanner, ...

Residuals Error (i.e., what we don't know)

Design Matrix

You

Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

Bias and Variance

low bias / low variance

Source: J Mumford

Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I)$$

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$E(\epsilon_i) = 0$$

Error has mean 0

Error is Homoschedastic

$$Var(\epsilon_i) = \sigma^2$$

Errors have same variance

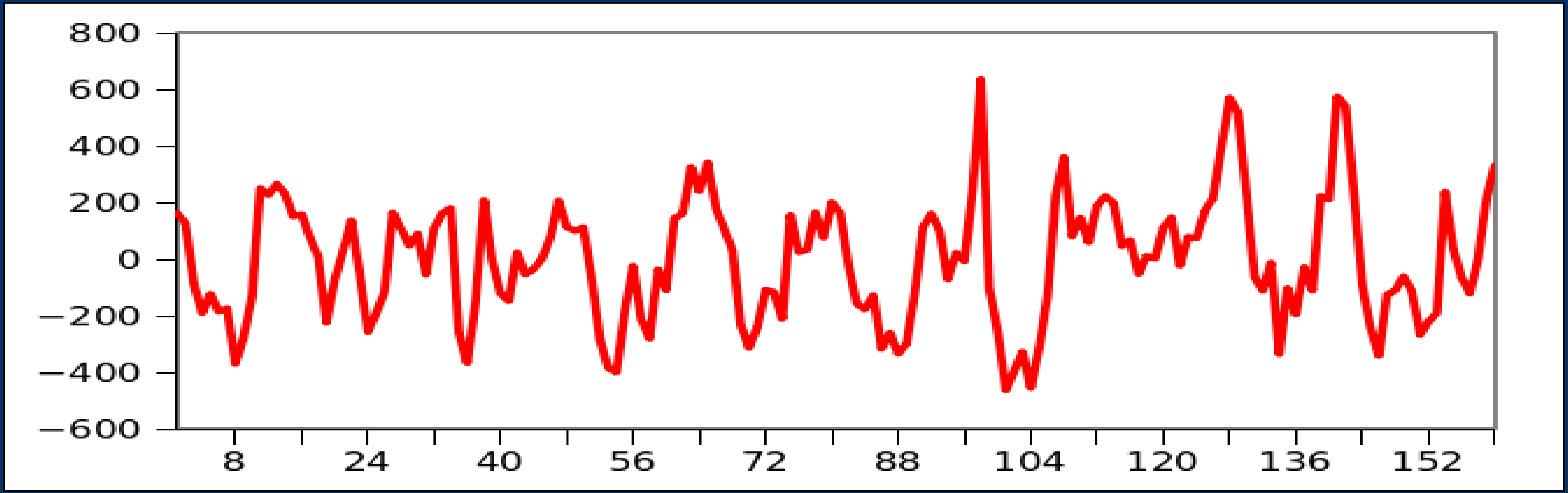
Errors are not Autocorrelated

$$Cov(\epsilon_i, \epsilon_j) = 0$$

Errors are independent of each other

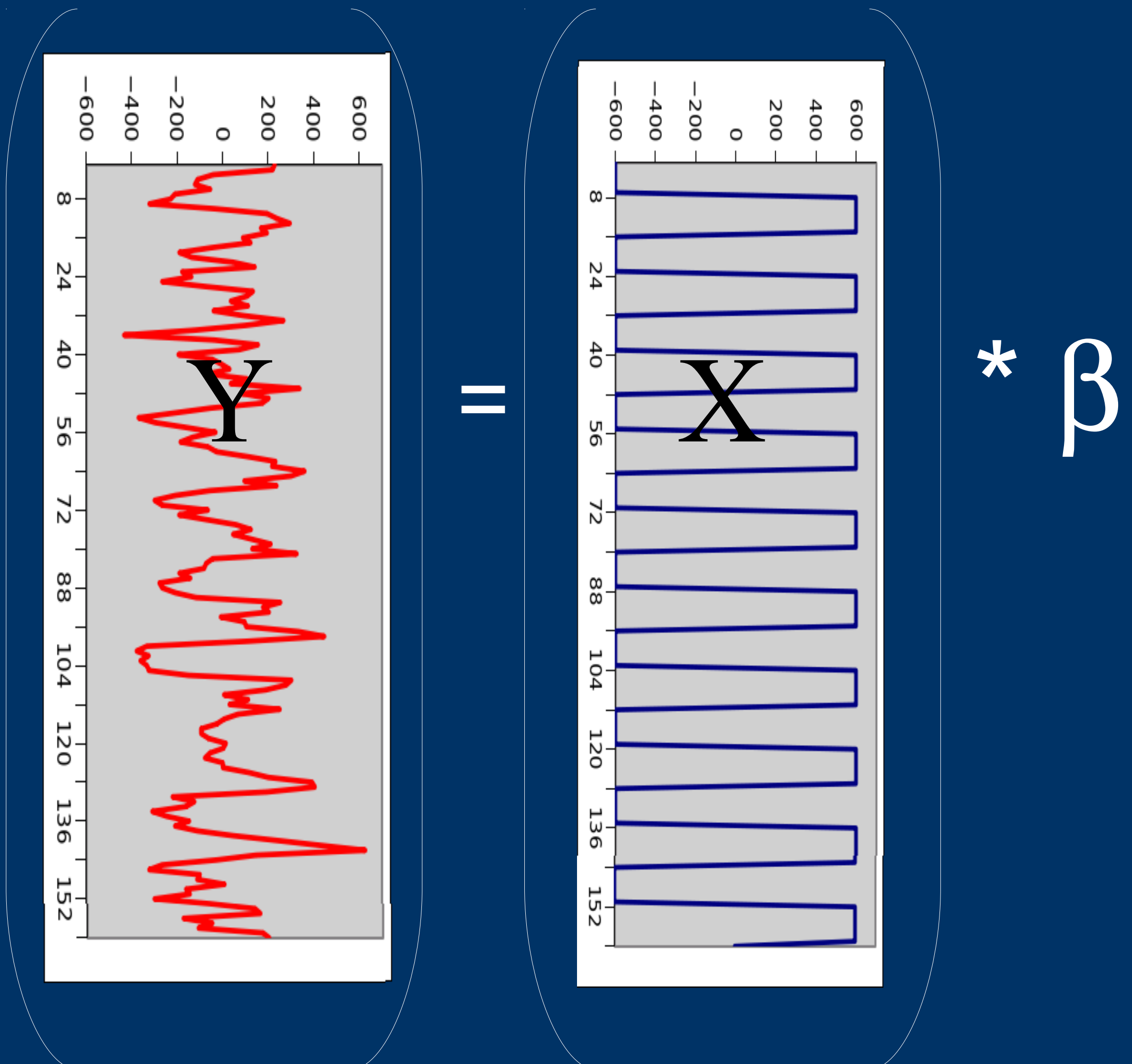
A simple experiment: the FFA

- Ss: 1 participant
- Design: 16s Faces, 16s Houses (block design)
- Take Voxel 15,19,18, in right posterior temporal cortex.

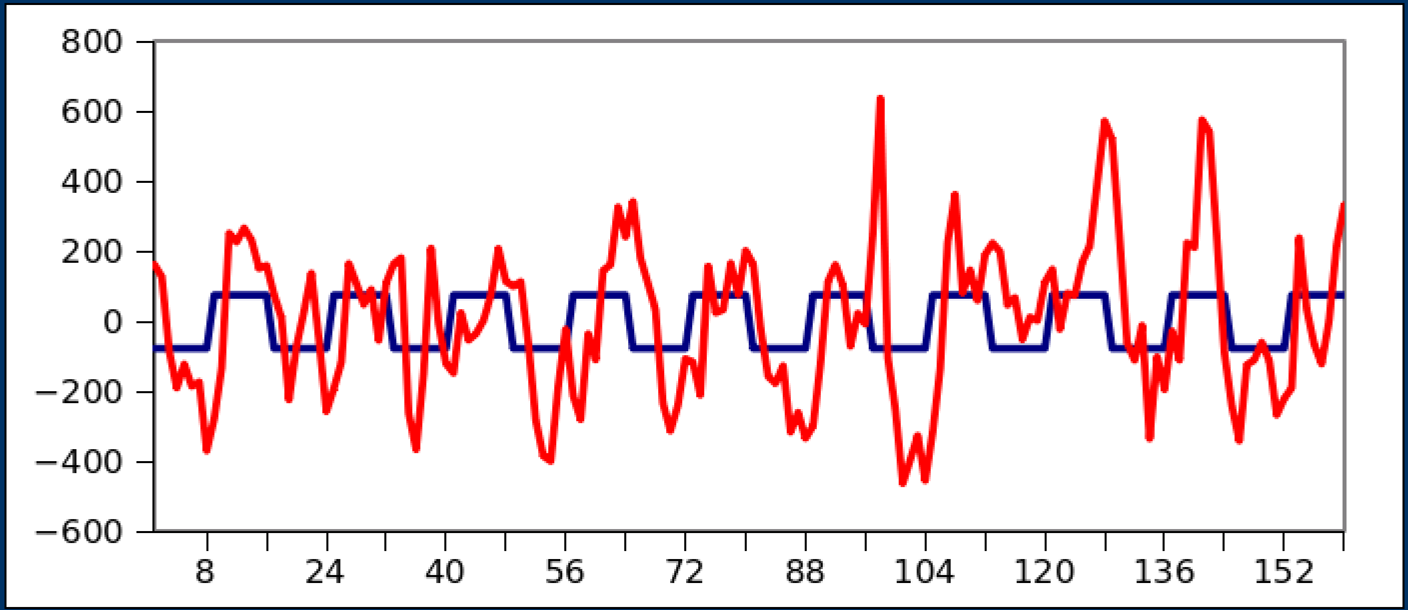


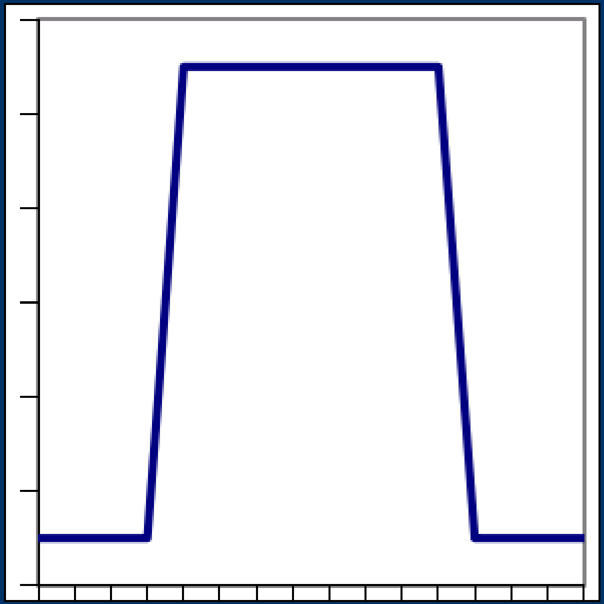
- Does this voxel “care” about faces?

Model I: Boxcar Regressor

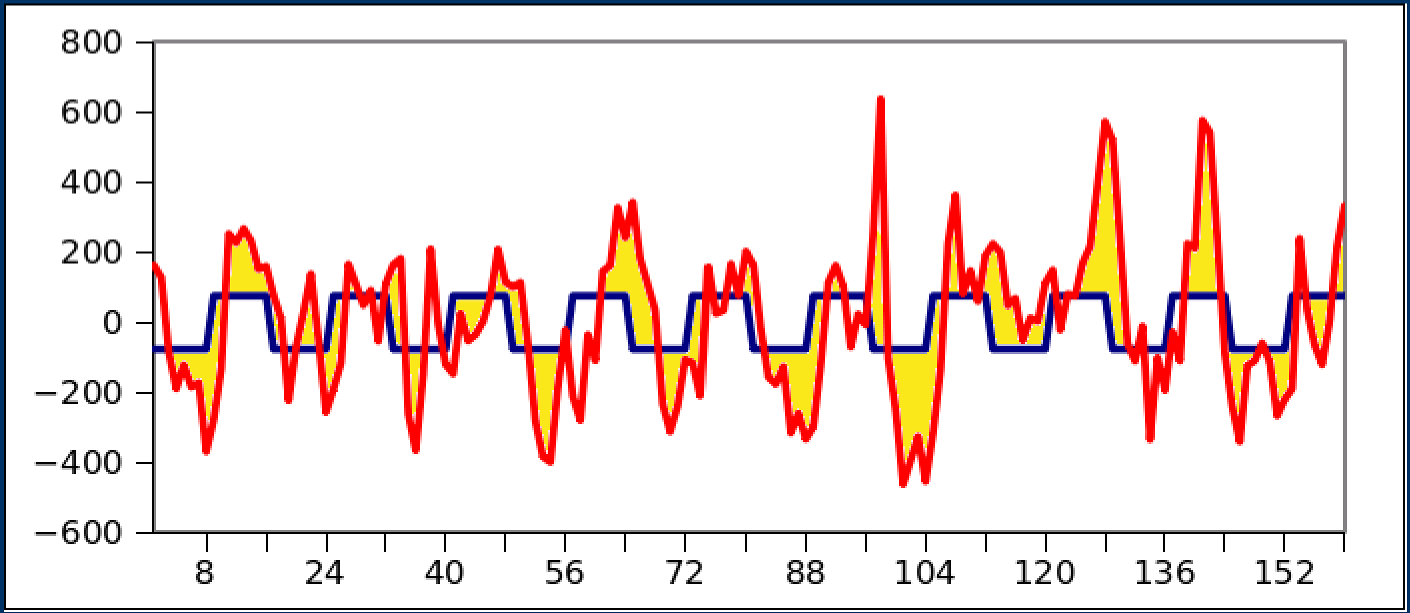


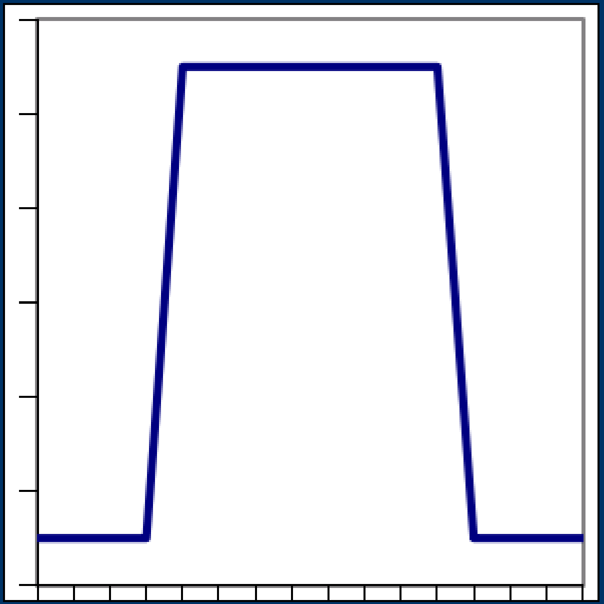
Model I: Boxcar Regressor





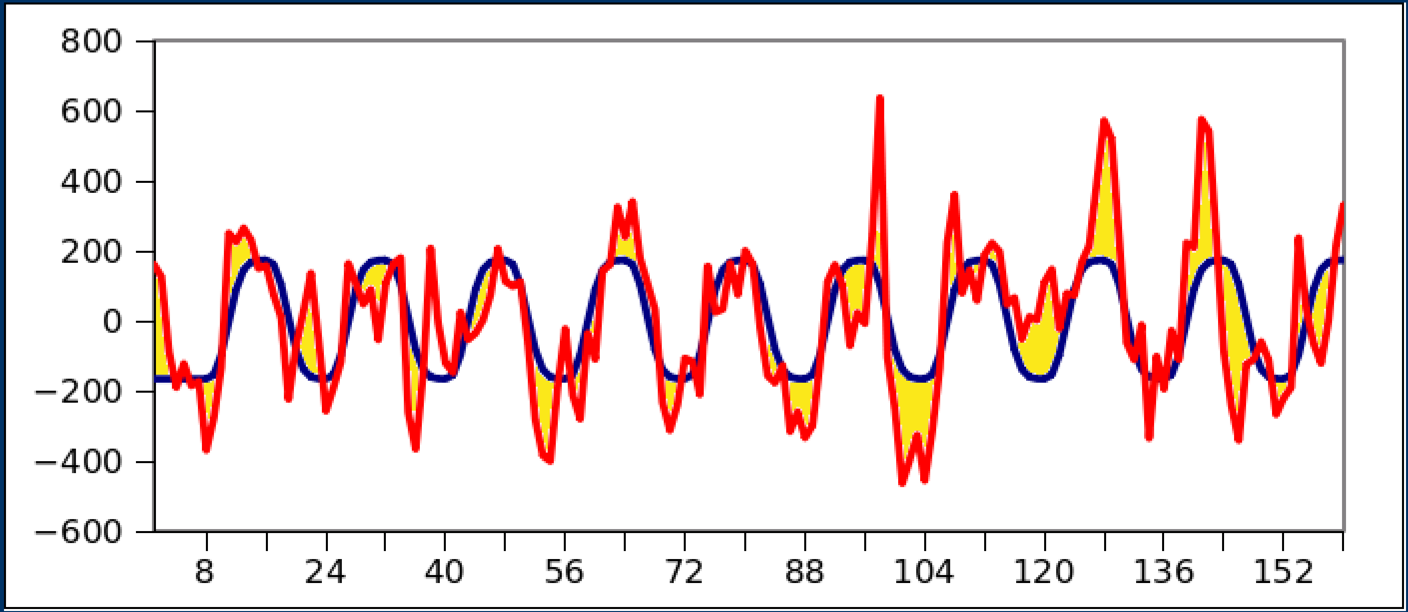
Model I: Boxcar Regressor

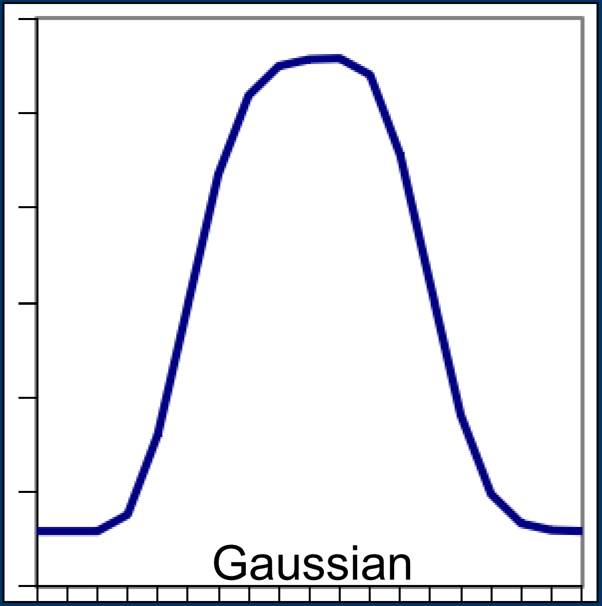




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{150.99}{202.44 \times 0.16} = 4.66$$

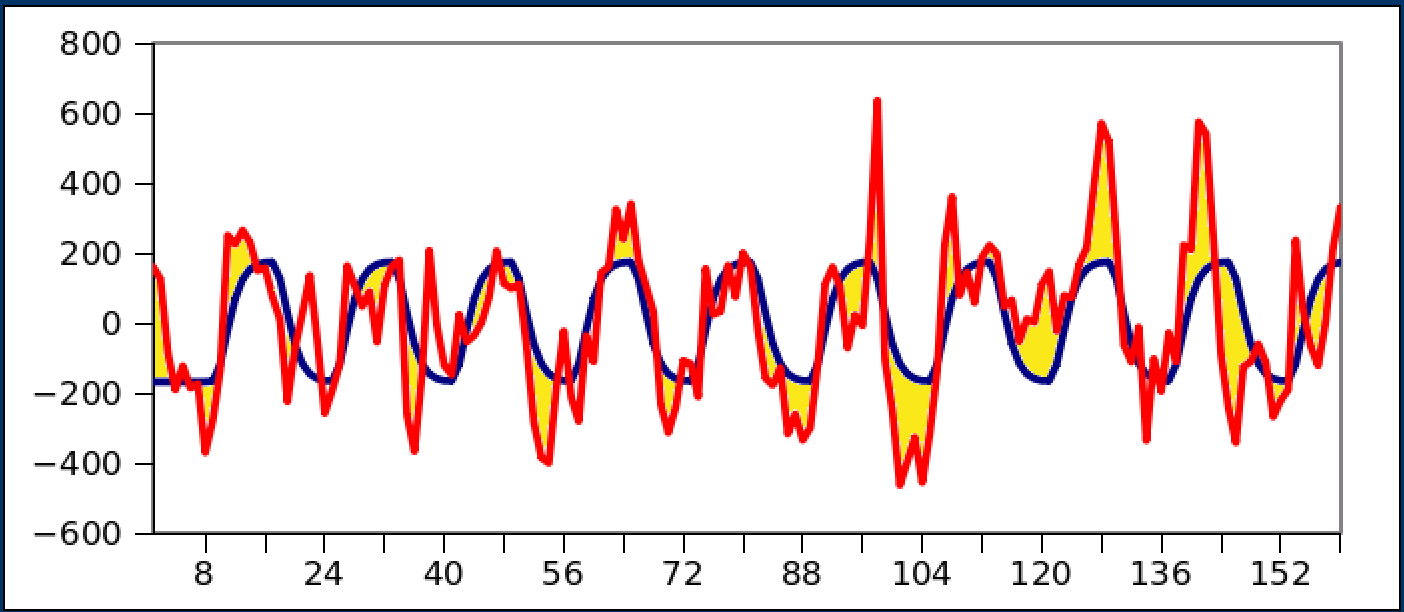
Model II: Convolved Boxcar Regressor

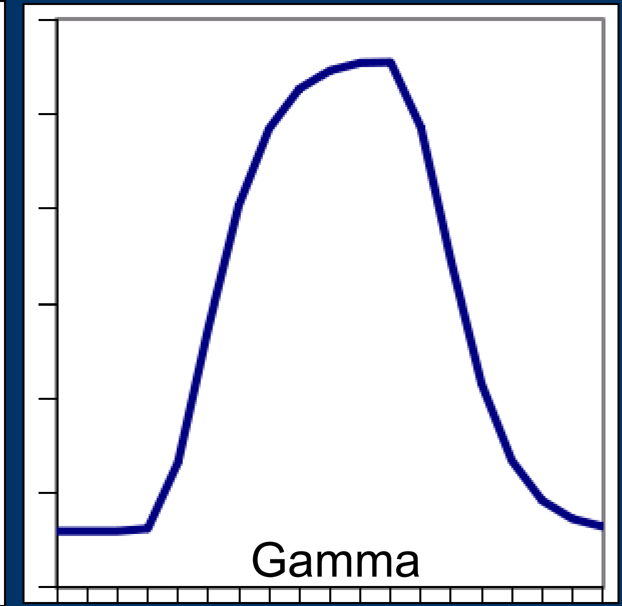




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \quad t(159) = \frac{267.02 \uparrow}{\downarrow 166.70 \times 0.16} = 10.01 \uparrow$$

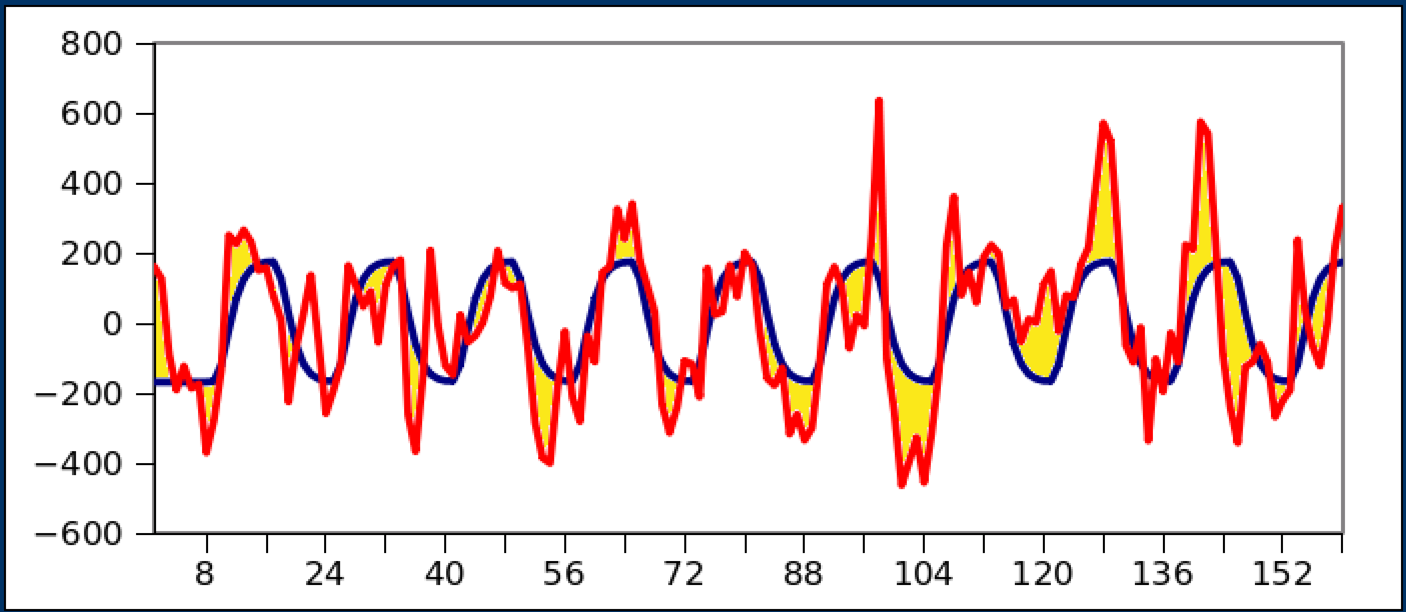
Model II: Convolved Boxcar Regressor

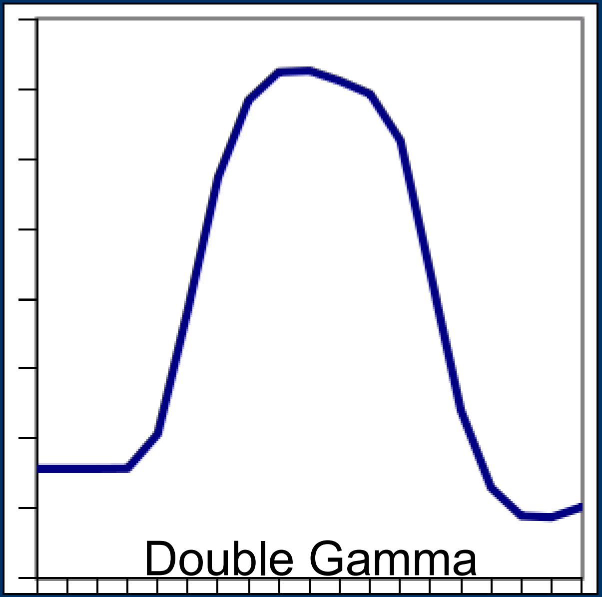




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{339.94}{166.31 \times 0.16} = 12.77$$

Model II: Convolved Boxcar Regressor





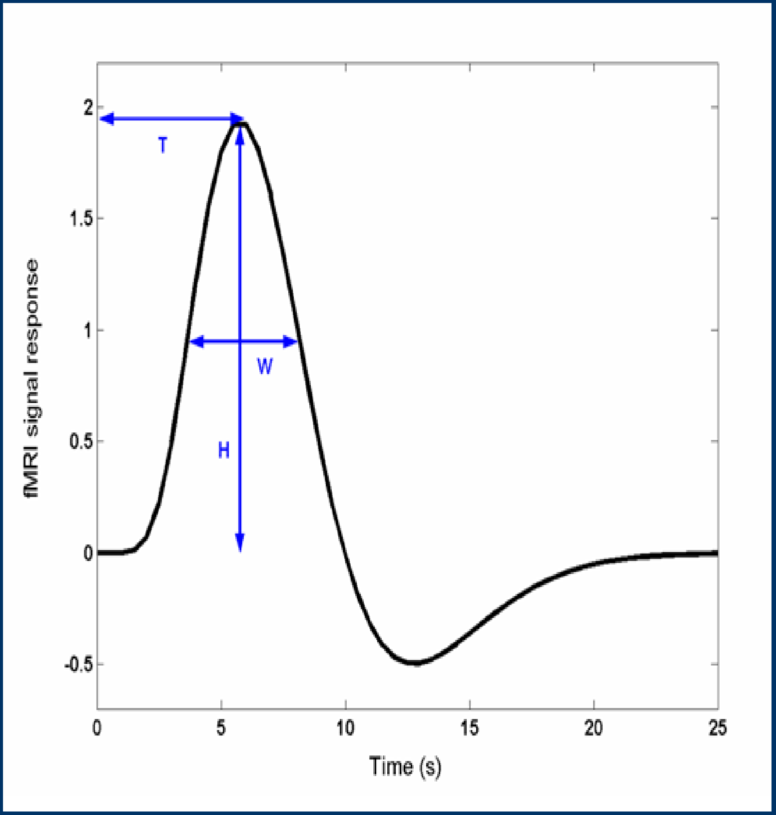
$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

What Convolution?

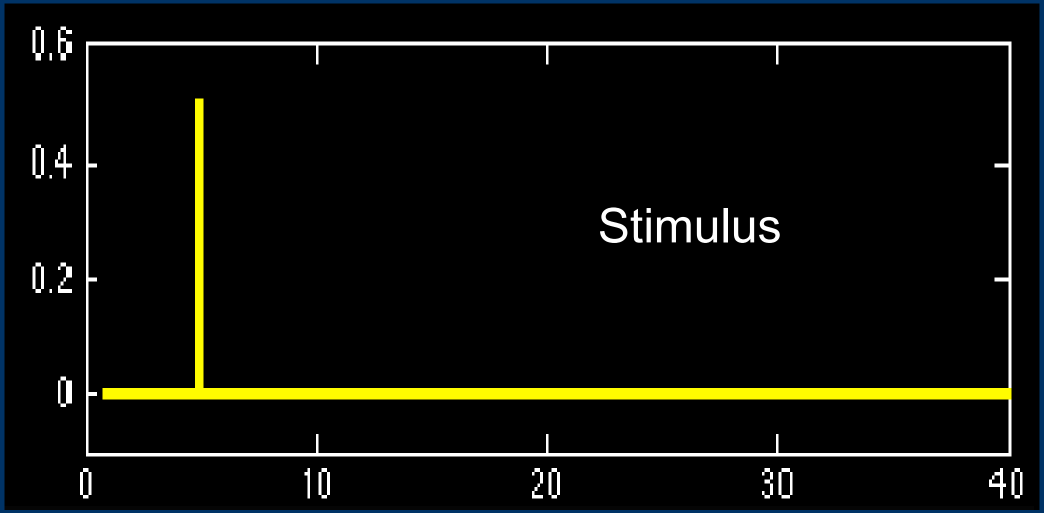
- Linear Time Invariant Basis Functions:
  - Canonical HRF
  - Canonical HRF + temp & disp derivs
  - Informed basis set
  - FIR/Smooth FIR

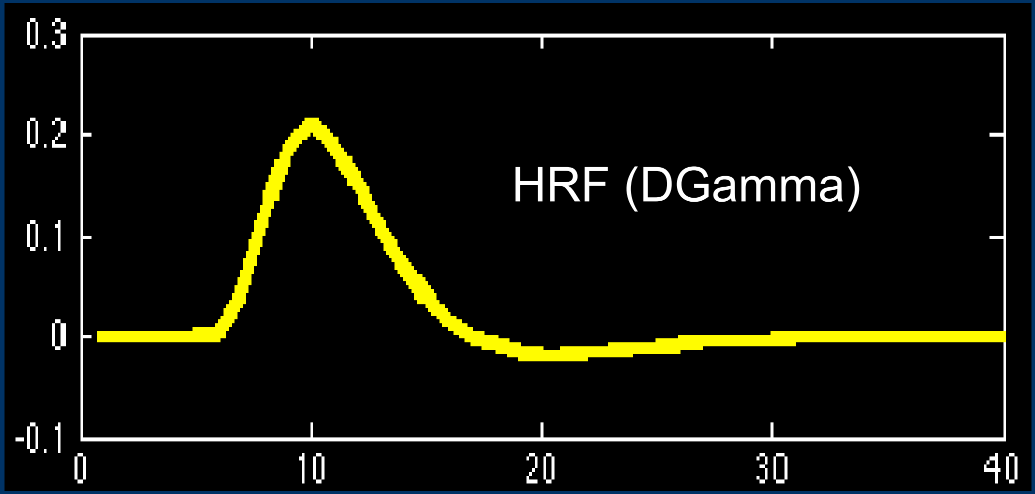
  …
- Non Linear:
  - Inverse Logit
  - Balloon model

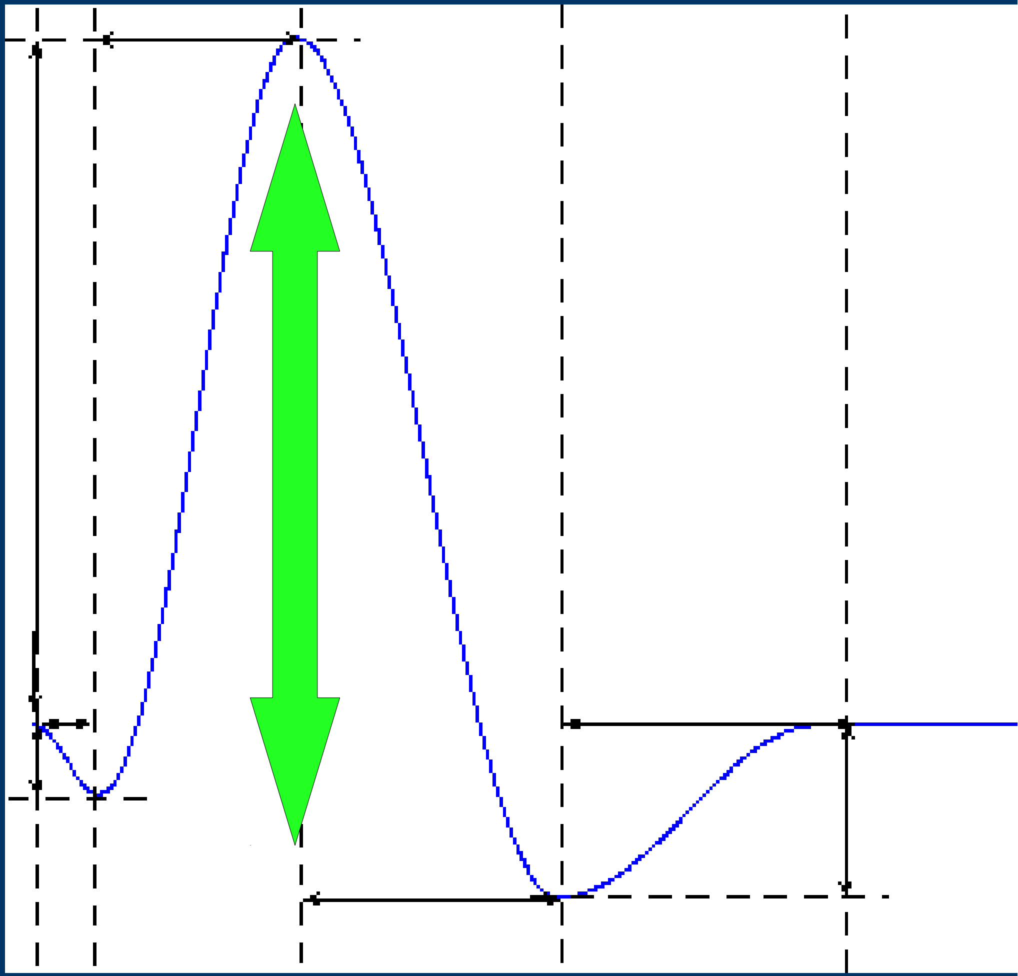
  …



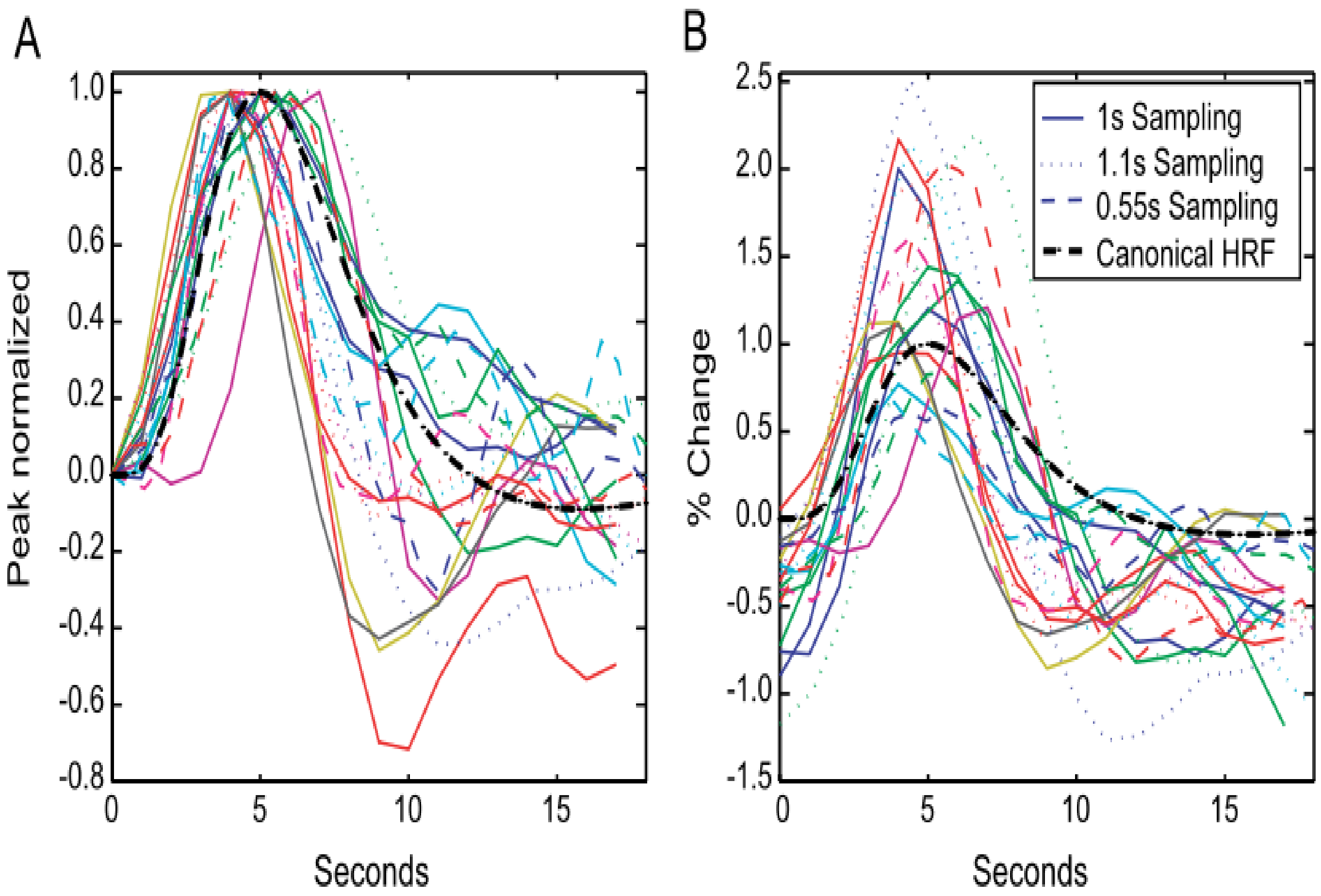
Canonical HRF



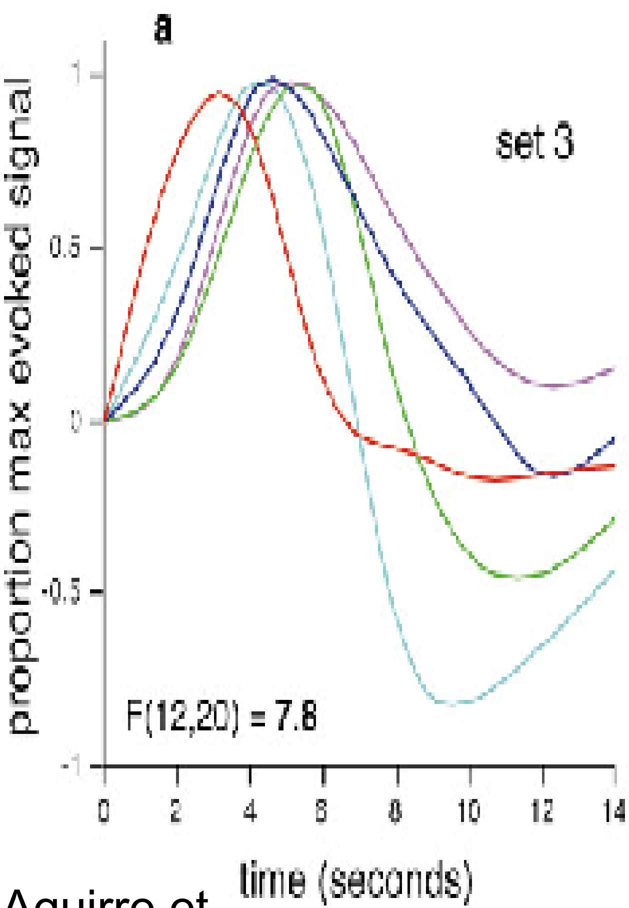




HRF Variability Across Subjects

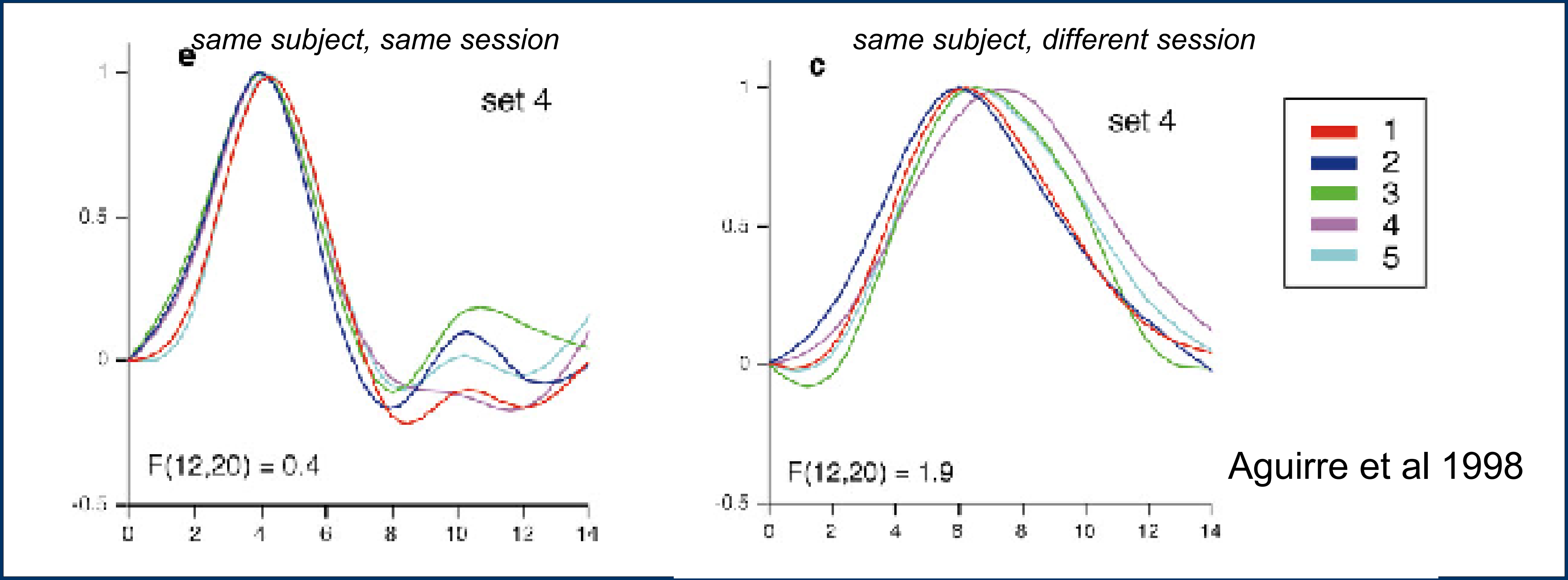


Handwerker et al 2004

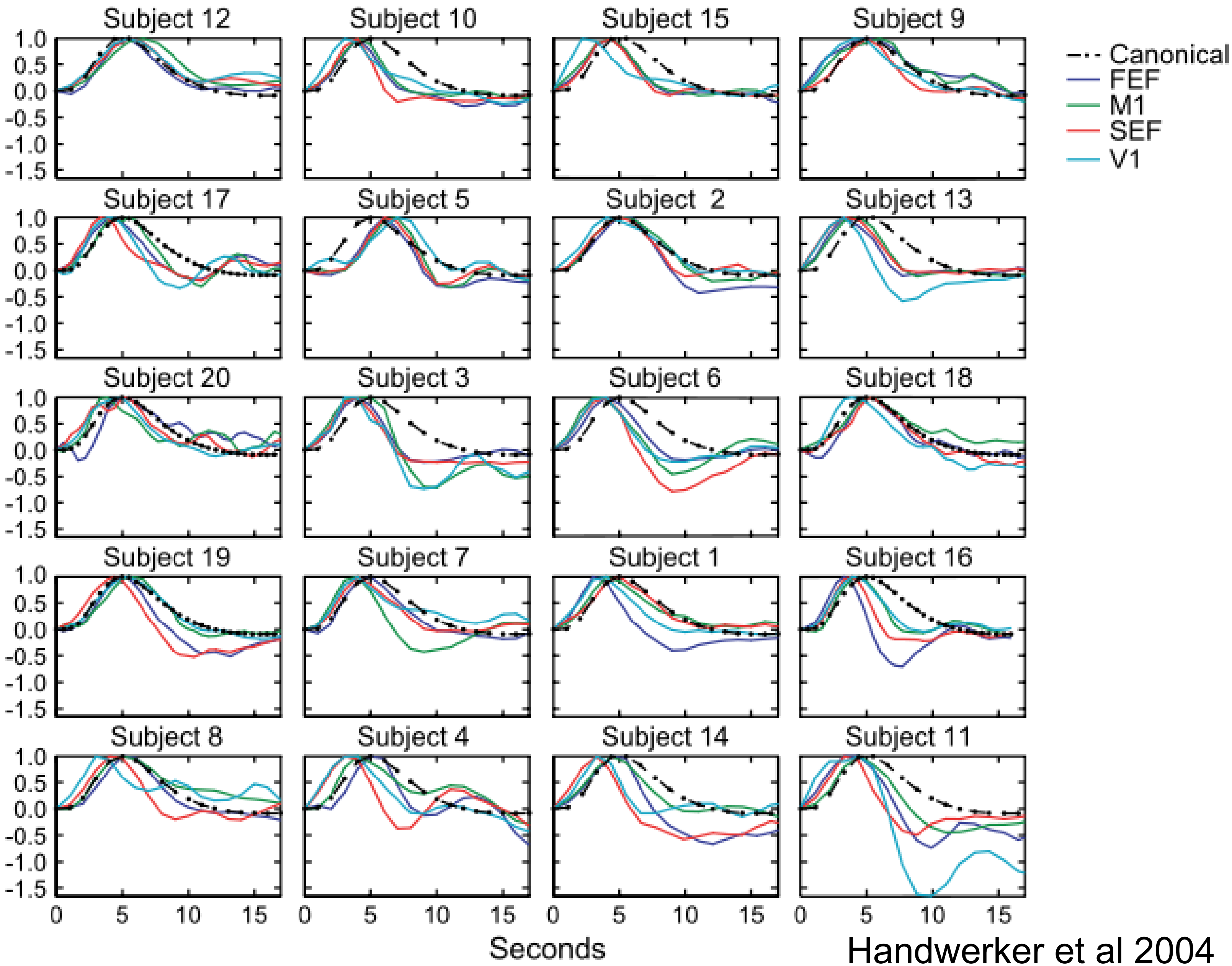


Aguirre et
Al 1998

HRF Variability Within Subjects

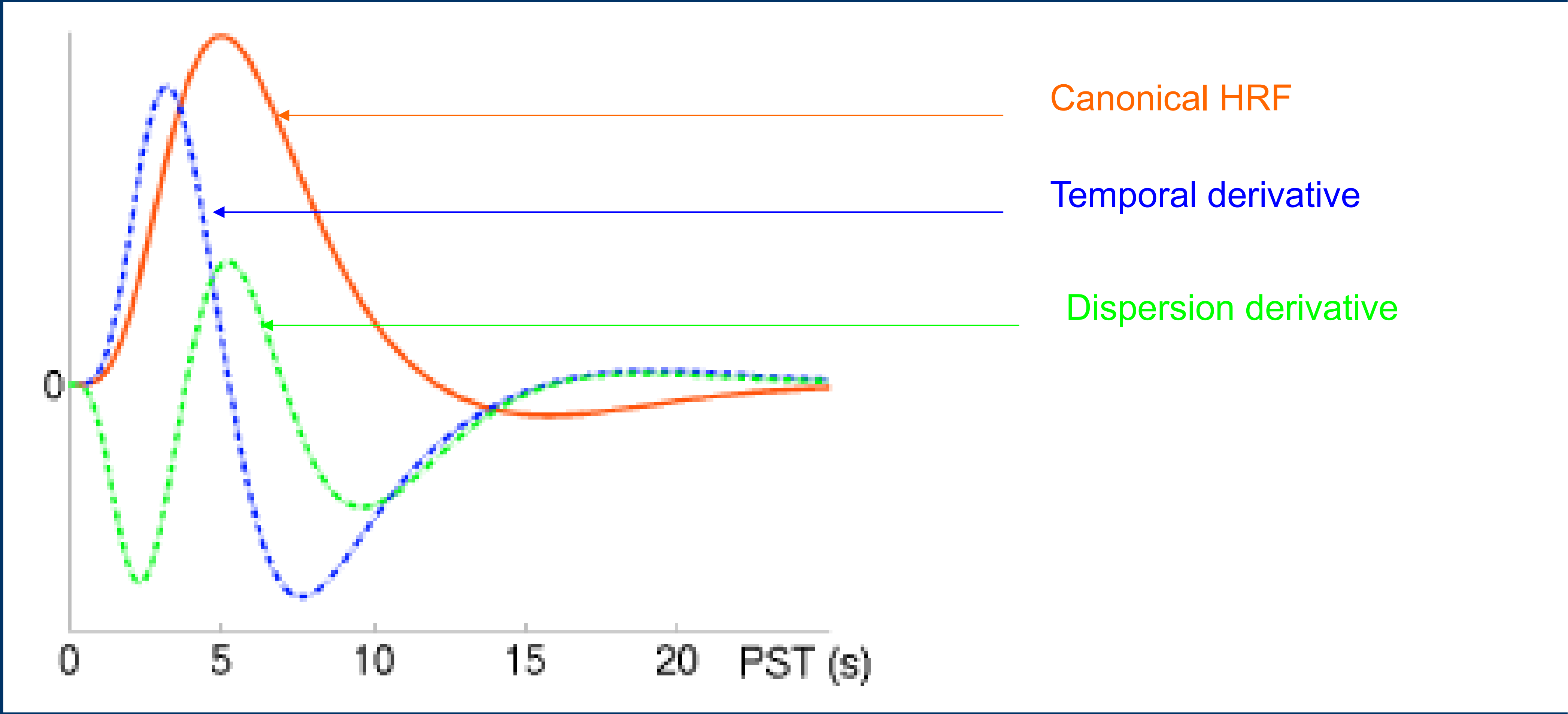


HRF Variability Across Brain Area



Handwerker et al 2004

Basis Functions I: HRF + Derivatives



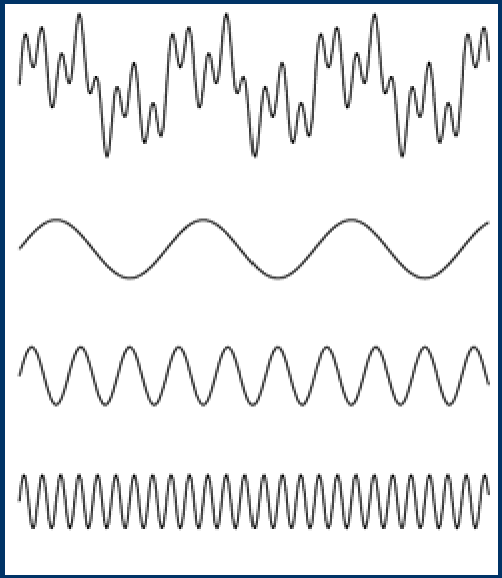
Basis Functions II: Fourier Basis Set

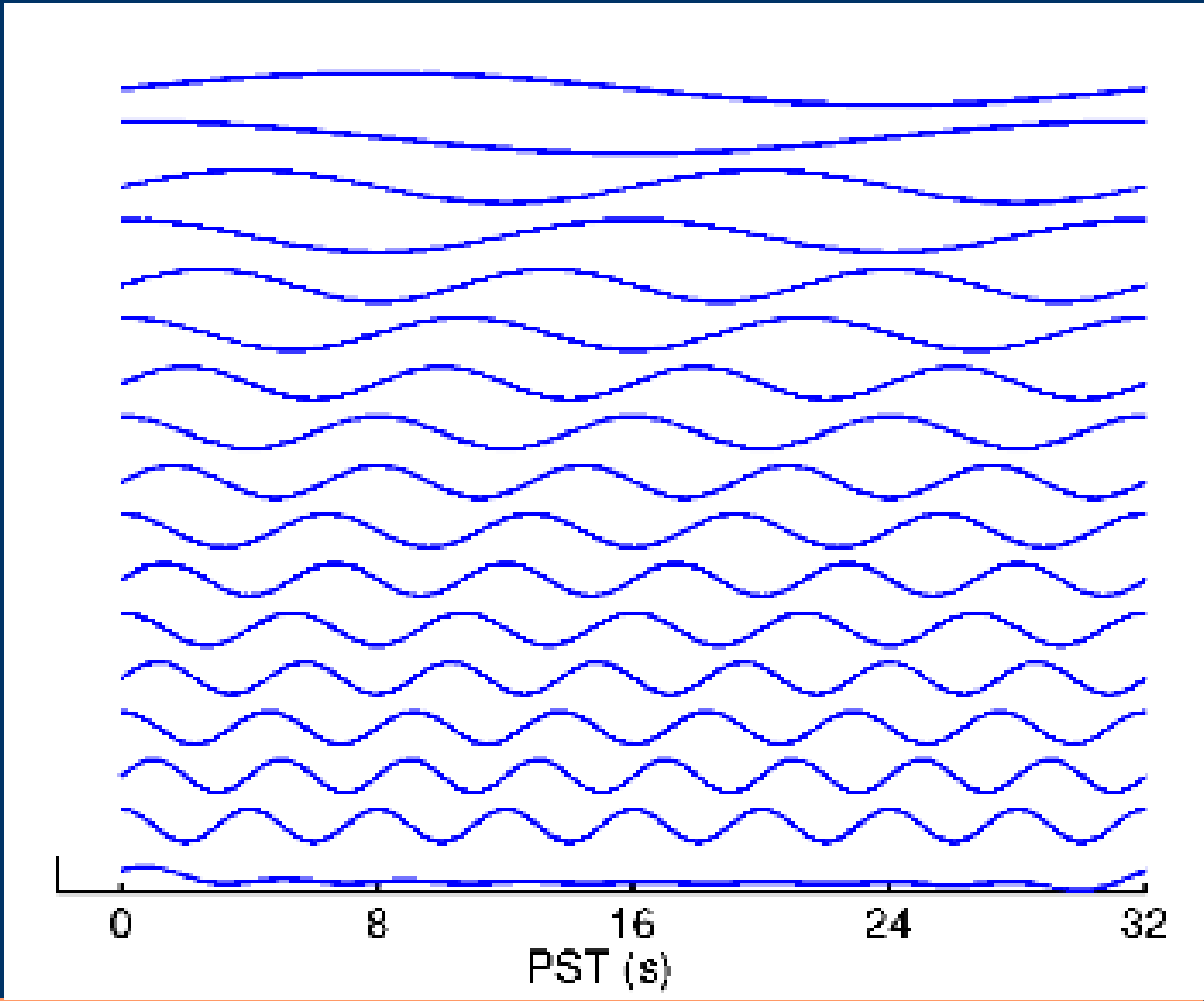
f(t)

h1(t)

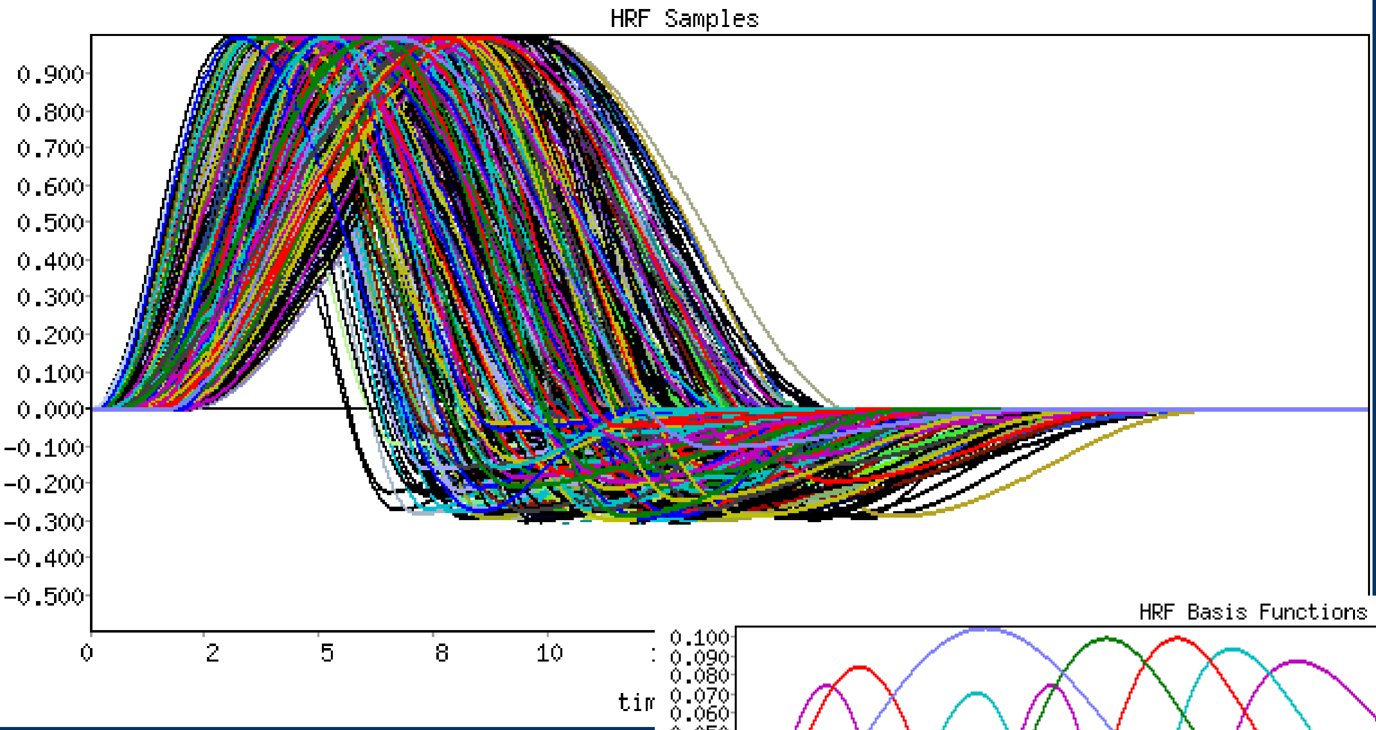
h2(t)

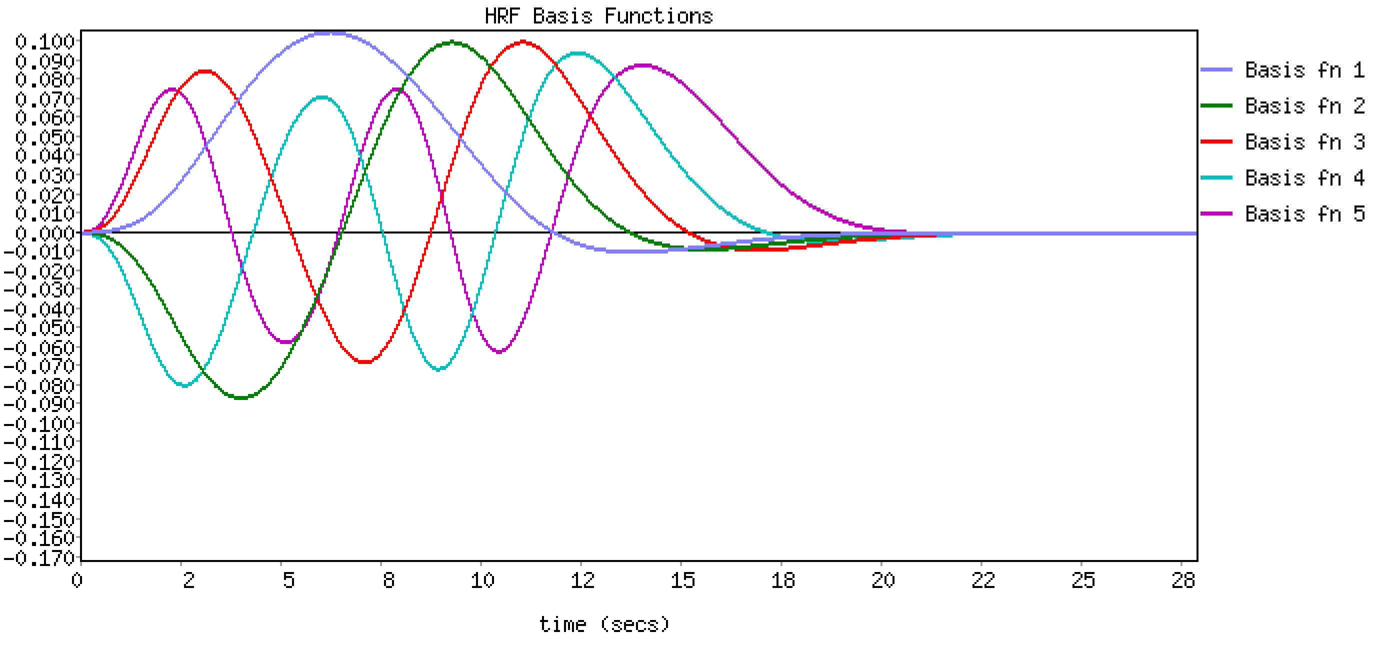
h3(t)



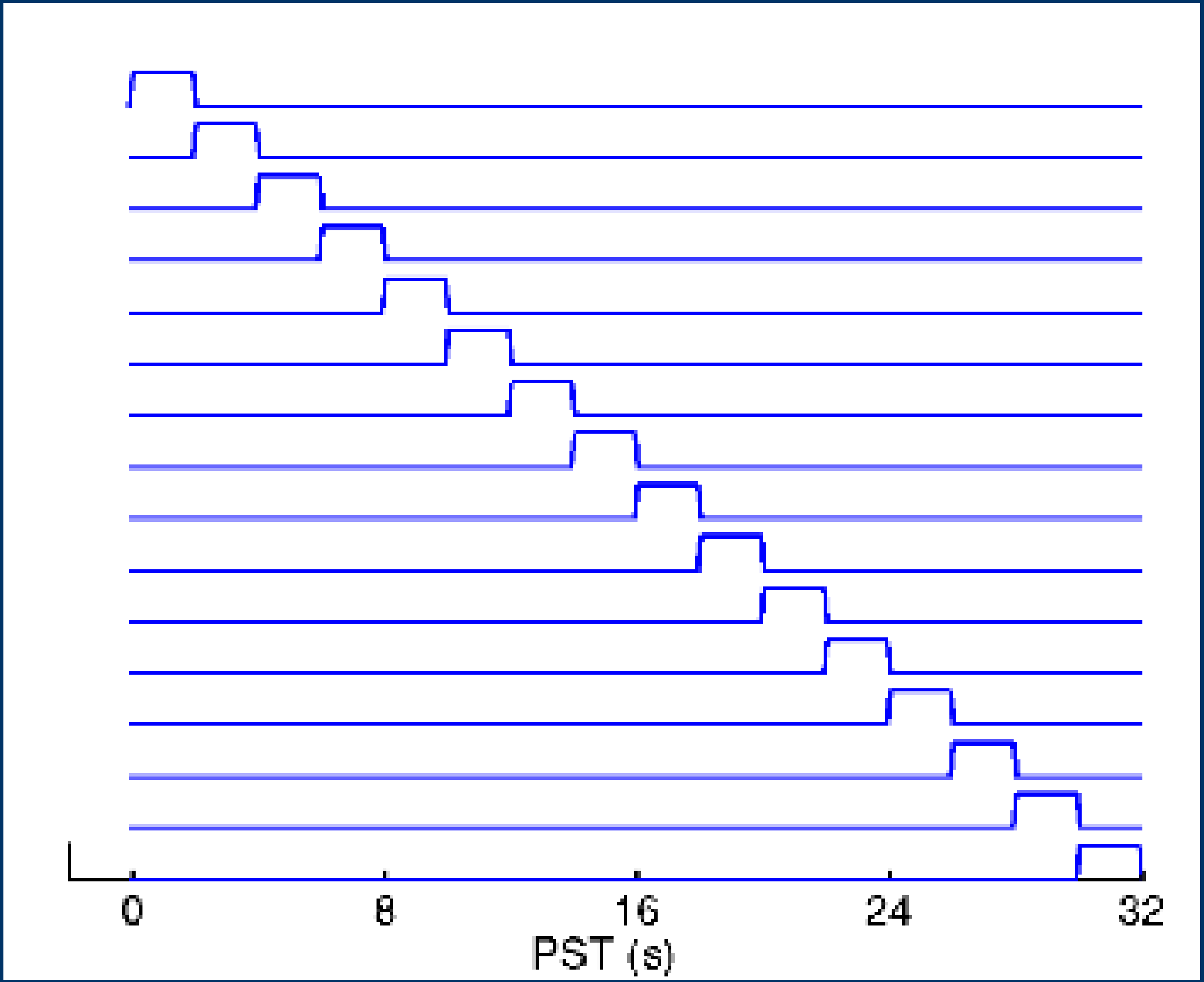


Basis Functions III: Informed Basis Functions (FLOBS)



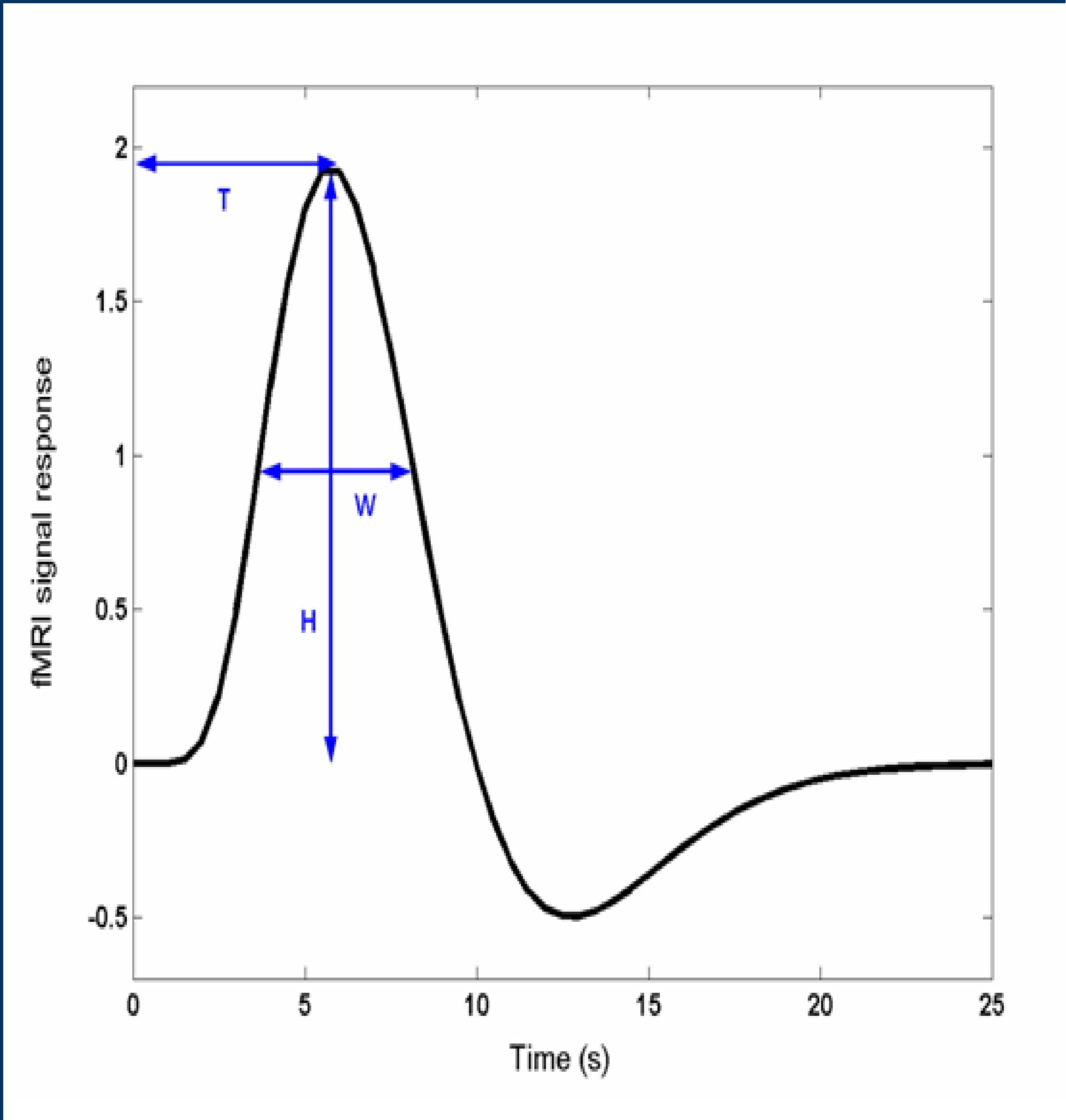


Basis Functions IV: FIR



Non Linear Approaches

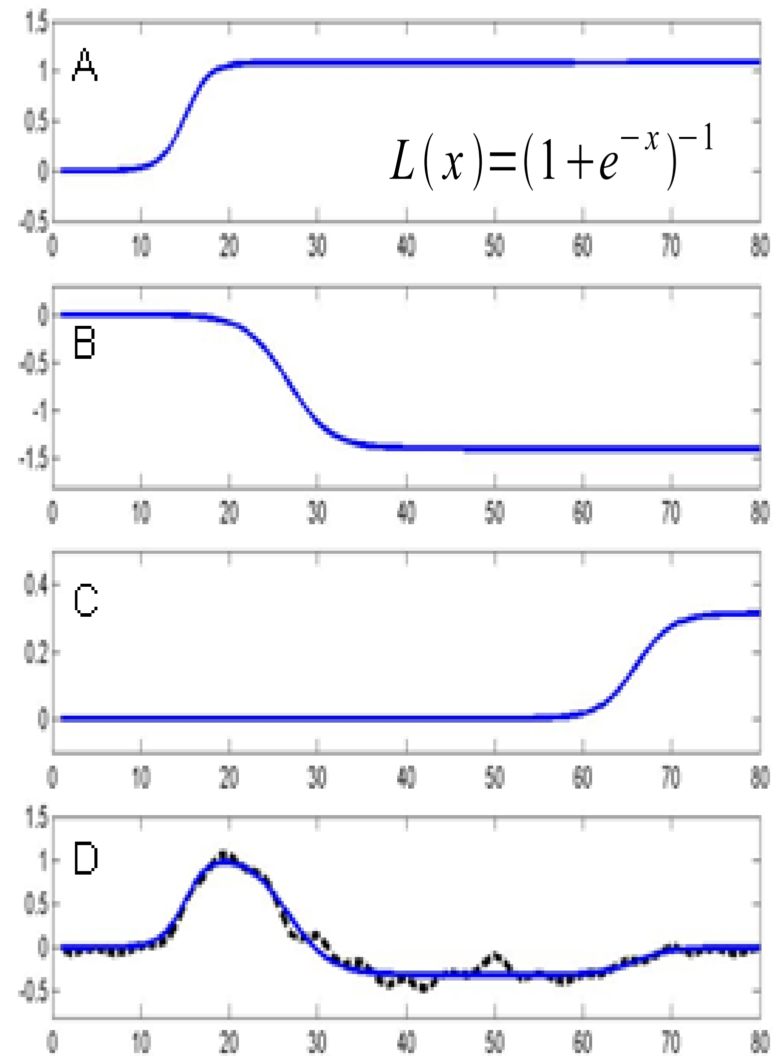
- Model all 3 parameters (T,W,H)



Slide from M Lindquist

Inverse Logit Model

- Superposition of three inverse logit (sigmoid) functions.
- Each function has three variable parameters dictating the amplitude, position and slope.



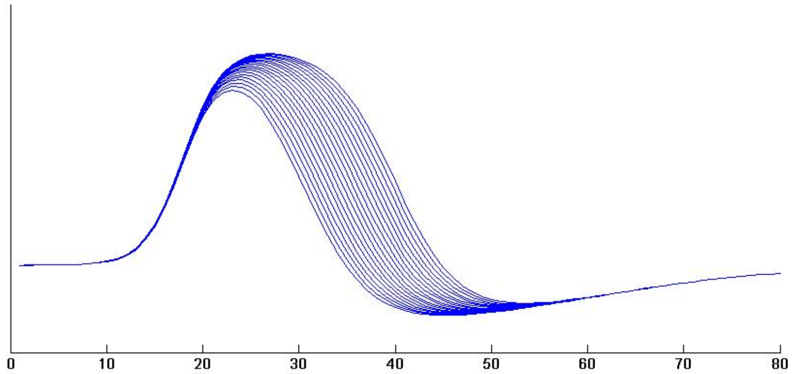
$$h(t|\theta) = \alpha_1 L\left((t-T_1)/D_1\right) + \alpha_2 L\left((t-T_2)/D_2\right) + \alpha_3 L\left((t-T_3)/D_3\right)$$

Lindquist & Wager (2007)

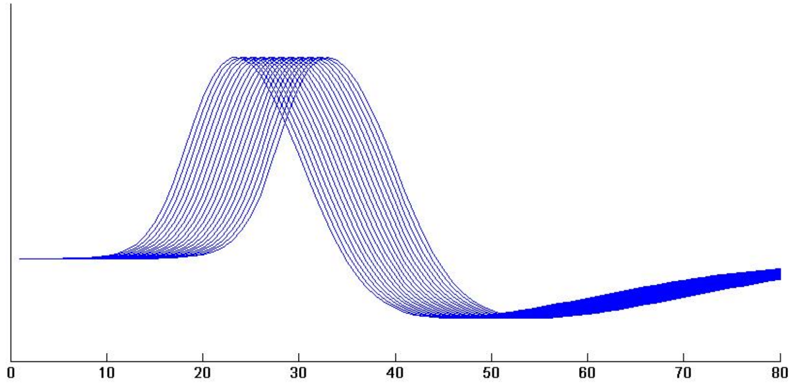
Slide from M Lindquist

Flexibility

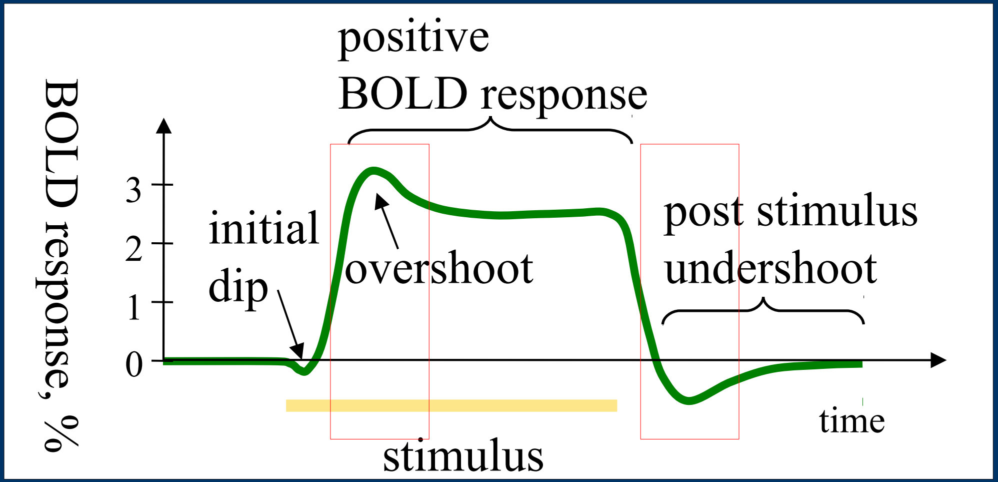
By shifting the position of the second IL function one can model differences in duration.

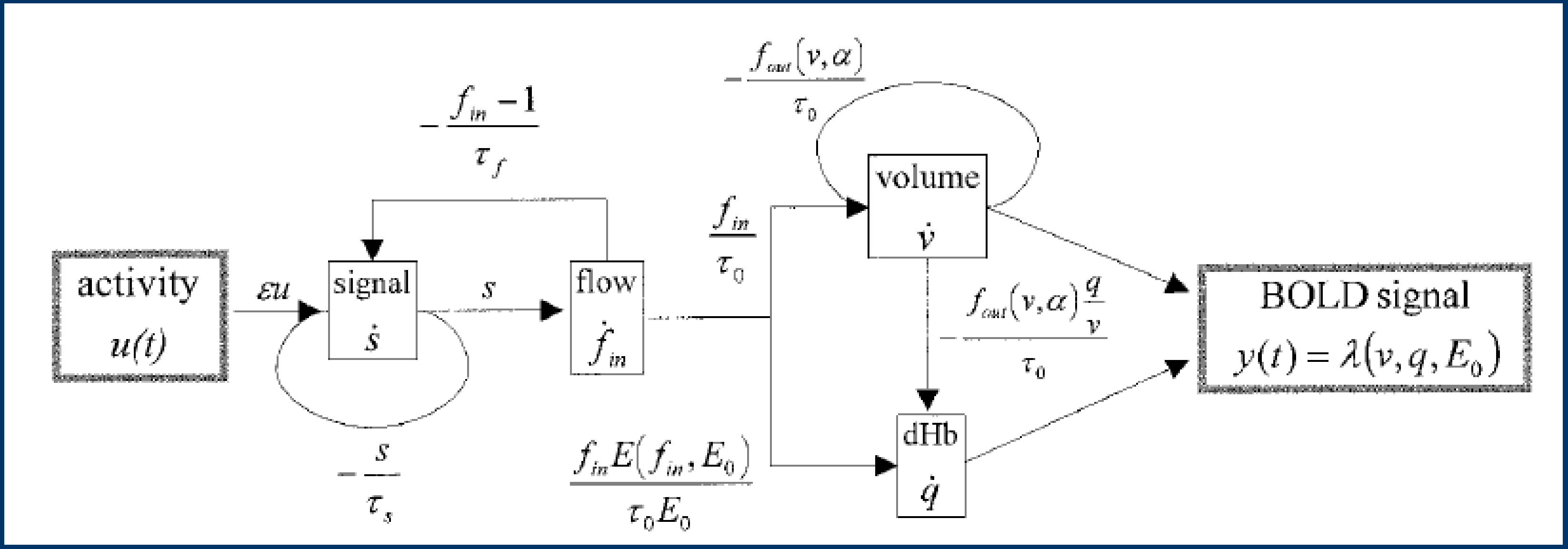


By shifting the position of all three IL functions one can model differences in onset.



Balloon Model





What Convolution? - Recap

- The problem:
  - Bias/Variance trade off
  - Assume known shape v estimate
  - **X-Matrix mis-specification**: Incorrect HRF will increase σ^2

What Convolution? - Recap

Canonical HRF:

- Pros: 1 parameter only, easy to interpret; easy to integrate at the second level.
- Cons: Bias in favour of a specific filter

- **Basis Sets Functions:**
  - Pros: Flexible, can capture greater variability, unexpected and subject specific differences.
  - Cons: many parameters to estimate (degrading dfs); difficult to interpret; how to aggregate at the 2nd level?
- **Non-Linear Systems:**
  - Potentially very flexible and may better capture parameters and vasculature
  - Computationally very demanding, parameters not obvious to interpret and aggregate across subjects (e.g., balloon model)

X-Matrix Assumptions

- Deterministic & known
- No regressor is a linear transformation of one (or more) regressors

Multicollinearity

- No column of the X matrix can be obtained by linear combination of one or more other columns.
- The issue is not whether there is/isn't multicollinearity, but rather, “how much” is there.
- Two main consequences:
  - Decreased efficiency of the design degrades power.
  - Parameter estimates can vary wildly, even in the wrong direction!

Source: J Mumford

Why is it a problem?

$$\begin{pmatrix} 12 \\ 12 \\ 12 \\ 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{pmatrix}$$

Source: J Mumford

Why is it a problem?

$$\begin{pmatrix} 12 \\ 12 \\ 12 \\ 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{pmatrix}$$

- There are an infinite # of solutions for β_1 and β_2

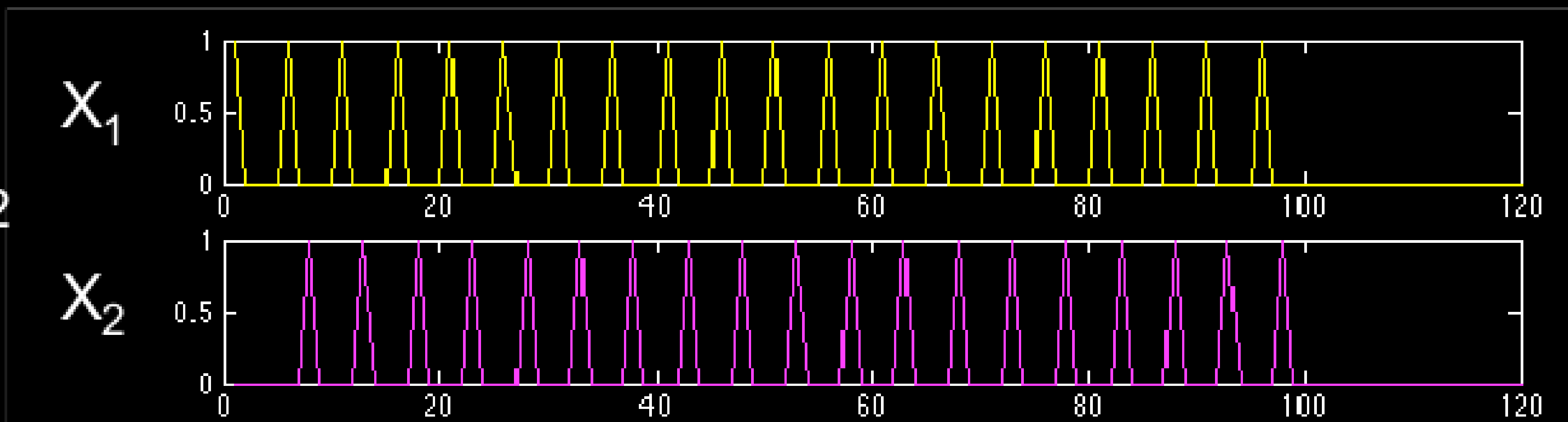
$\hat{\beta}_1 = 10, \hat{\beta}_2 = 2 \quad \hat{\beta}_1 = 100, \hat{\beta}_2 = -88$ …etc

Source: J Mumford

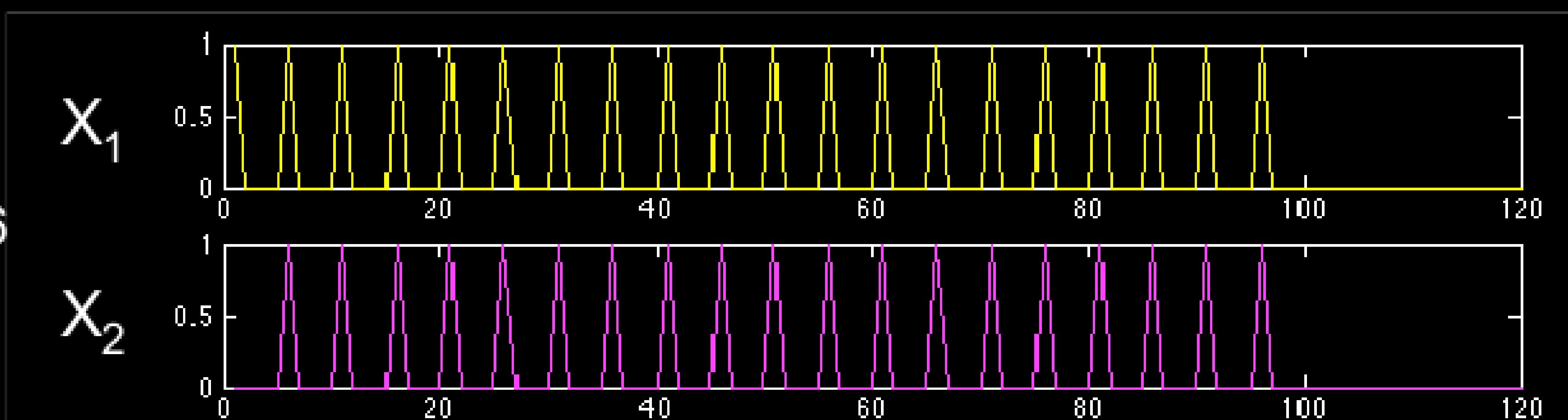
Collinearity illustration

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon \quad \beta_0 = 1, \beta_1 = 2, \beta_2 = 4$$

Cor= -0.2



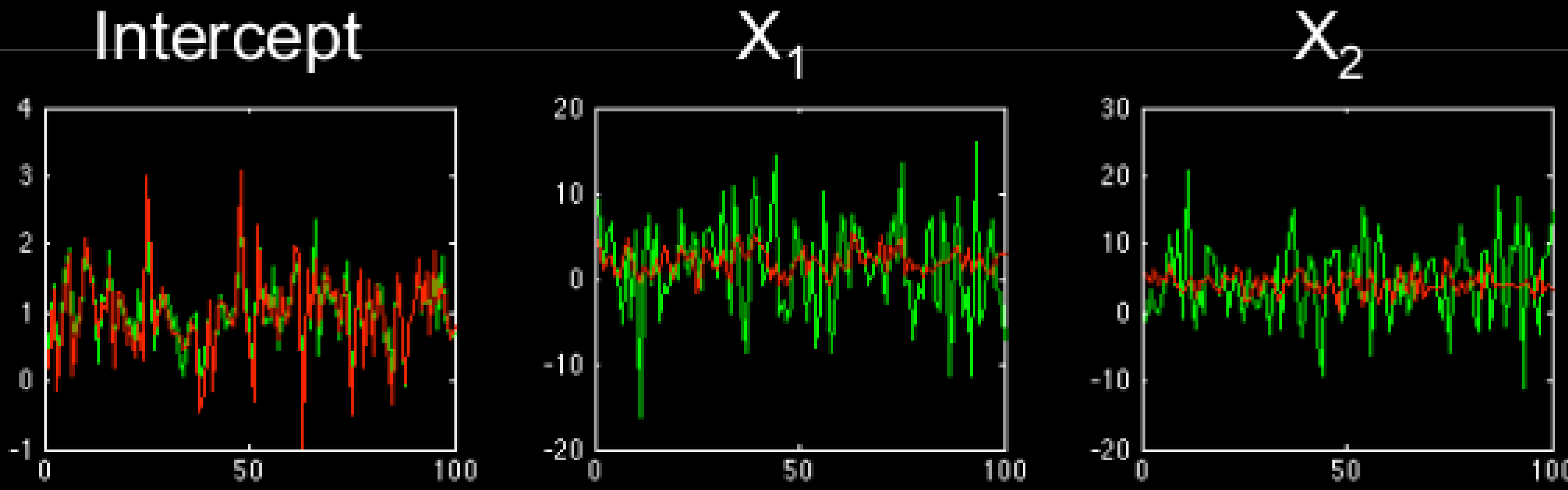
Cor=0.96



Source: J Mumford

$\hat{\beta}$

Highly variable over experiments

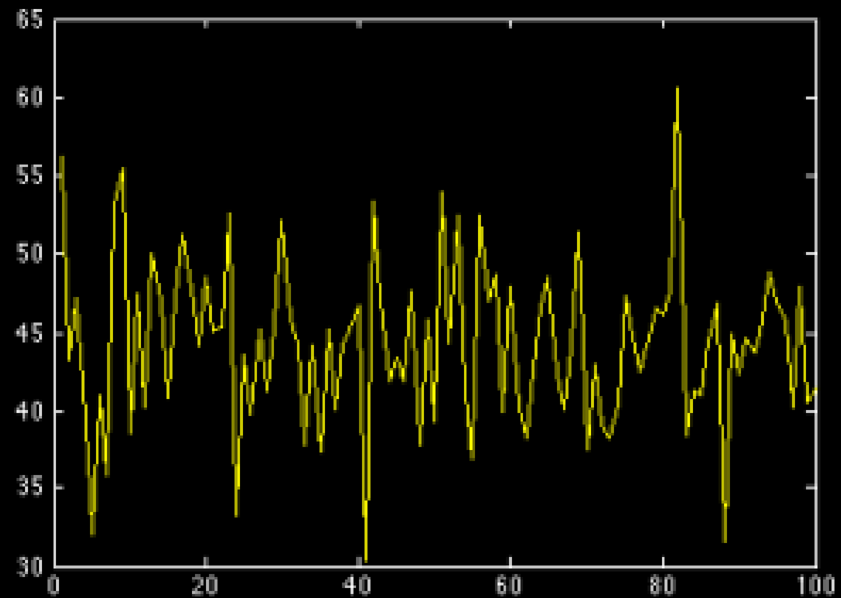


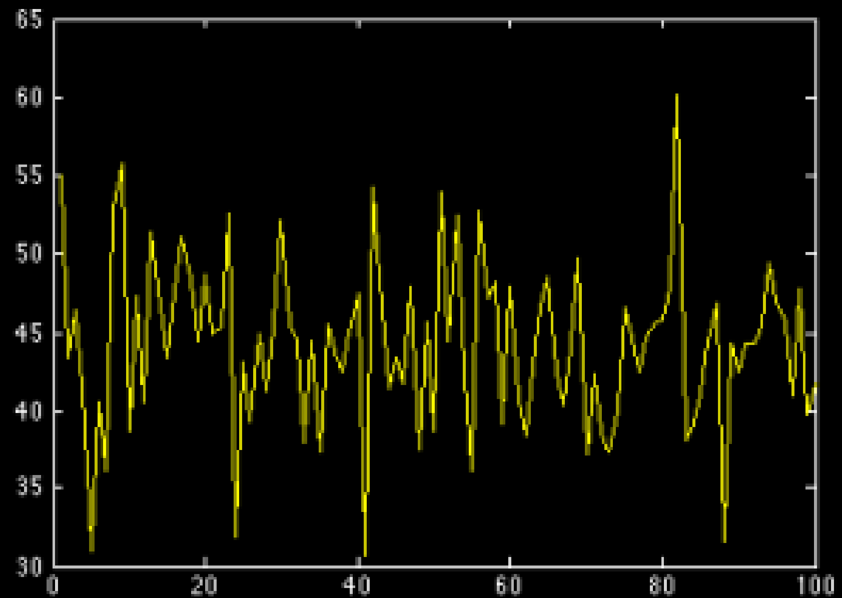
Source: J Mumford

Residuals don't change

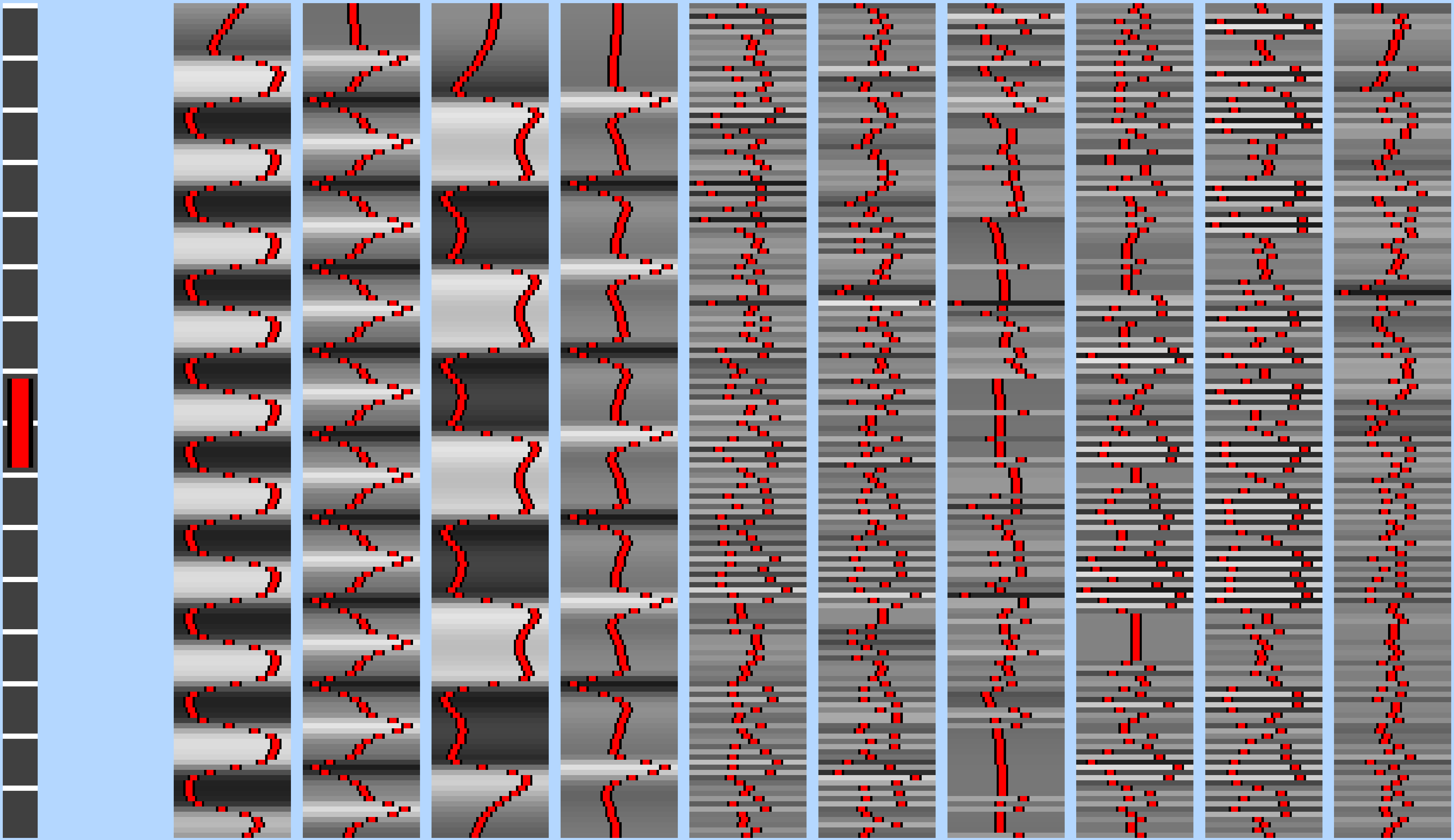
- The designs explain the same amount of variability

$\hat{\sigma}^2$





Contrasts



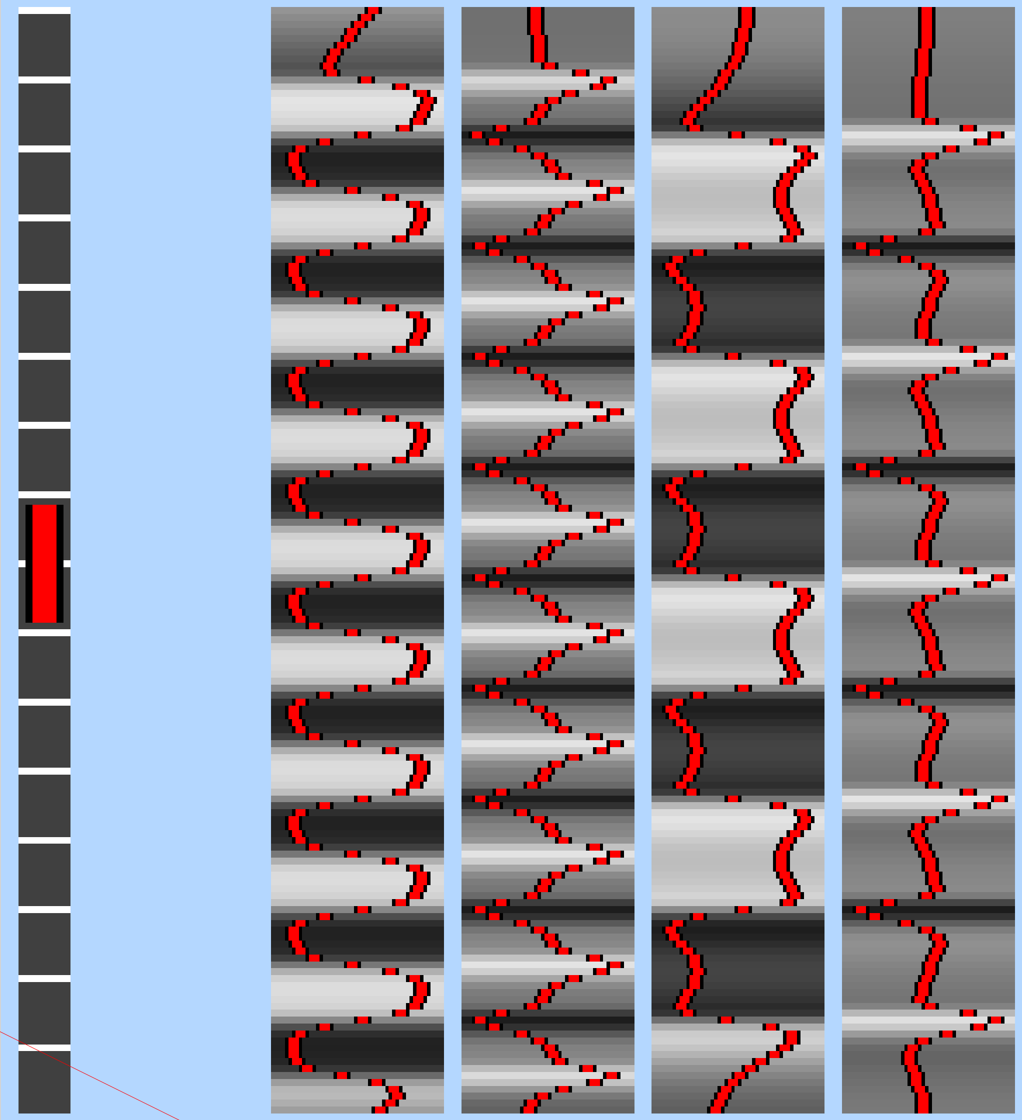
						conf	conf	conf	conf	conf	conf
C1	A	1	0	0	0	0	0	0	0	0	0
C2	B	0	0	1	0	0	0	0	0	0	0
C3	A > B	1	0	-1	0	0	0	0	0	0	0
C4	B > A	-1	0	1	0	0	0	0	0	0	0

Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

$$\hat{\beta} = \begin{array}{c} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \cdot \\ \hat{\beta}_p \end{array}$$

Model



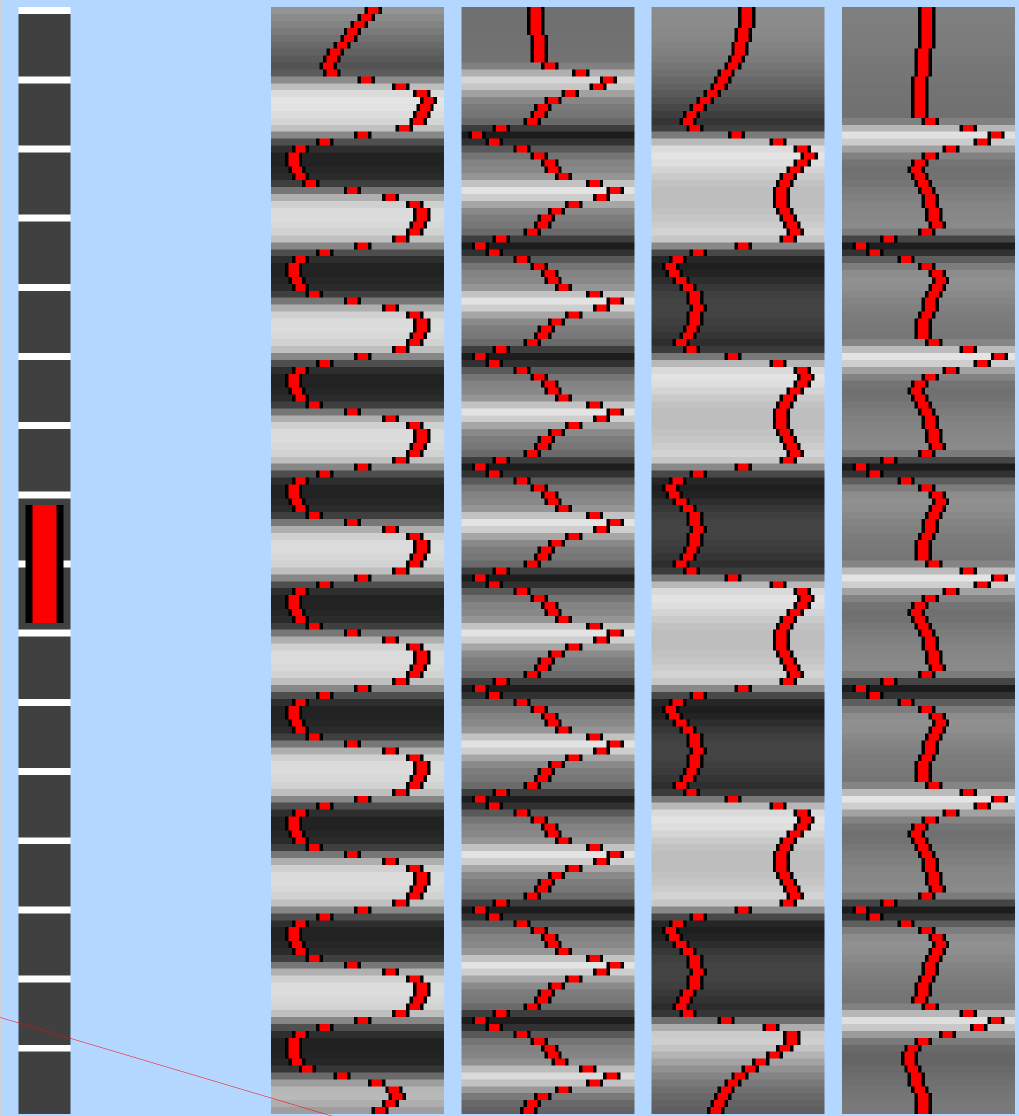
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \cdot \\ \hat{\beta}_p \end{matrix}$$

Model



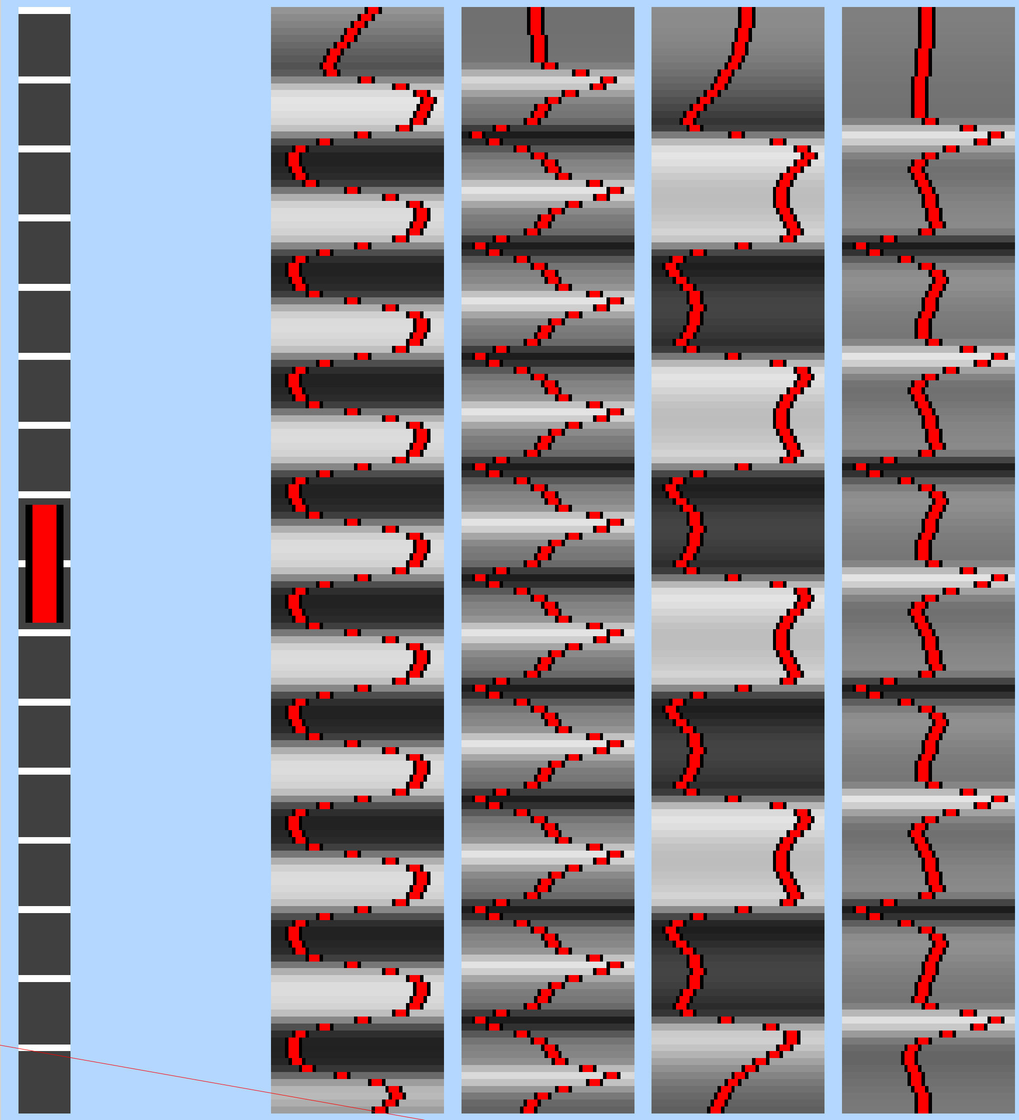
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix}$$

Model



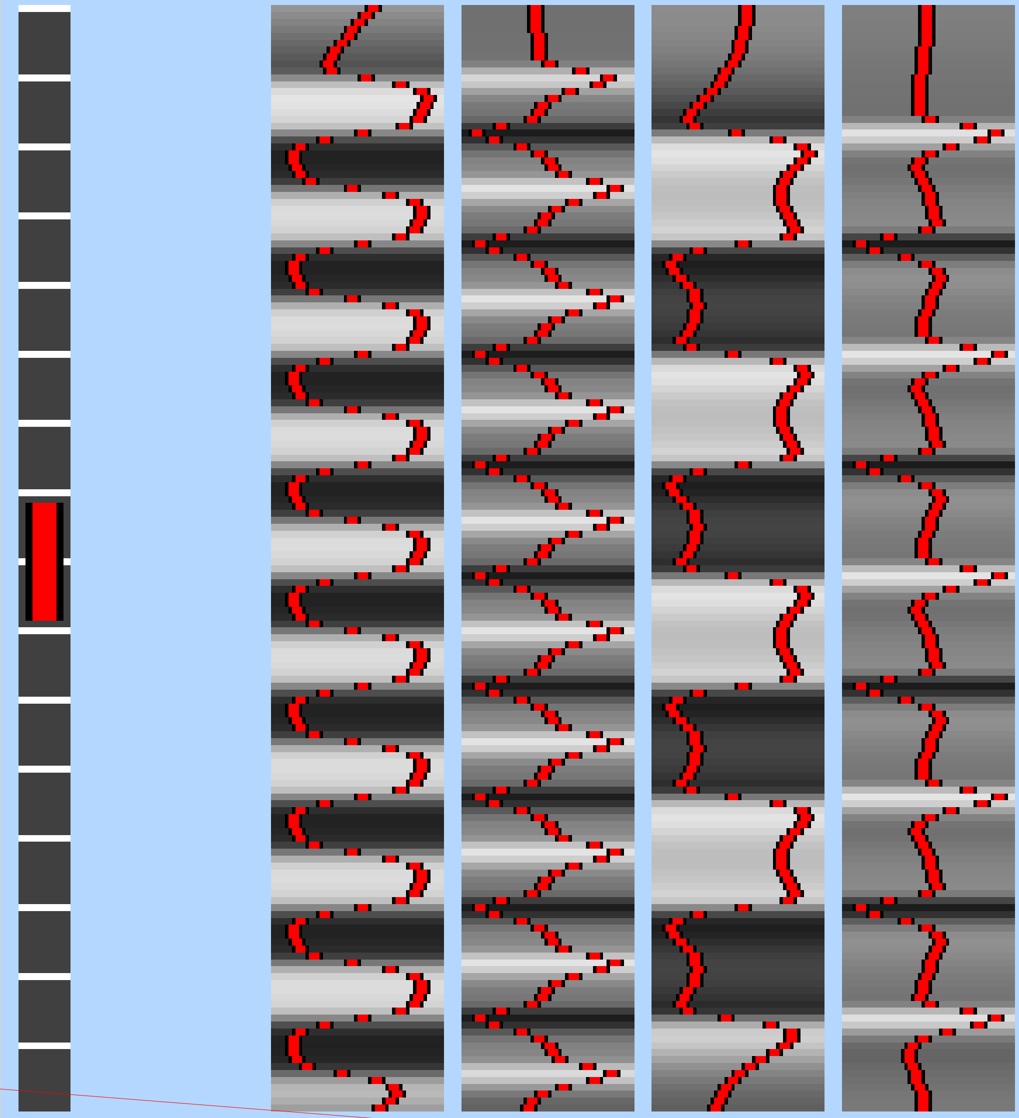
		Cond A	TD A	Cond B	TD B
C1	A	1	0	0	0
C2	B	0	0	1	0
C3	A > B	1	0	-1	0
C4	B > A	-1	0	1	0

Contrasts

$$t = \frac{\hat{\beta}}{\sqrt{\hat{Var}(\beta)}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix}$$

Model



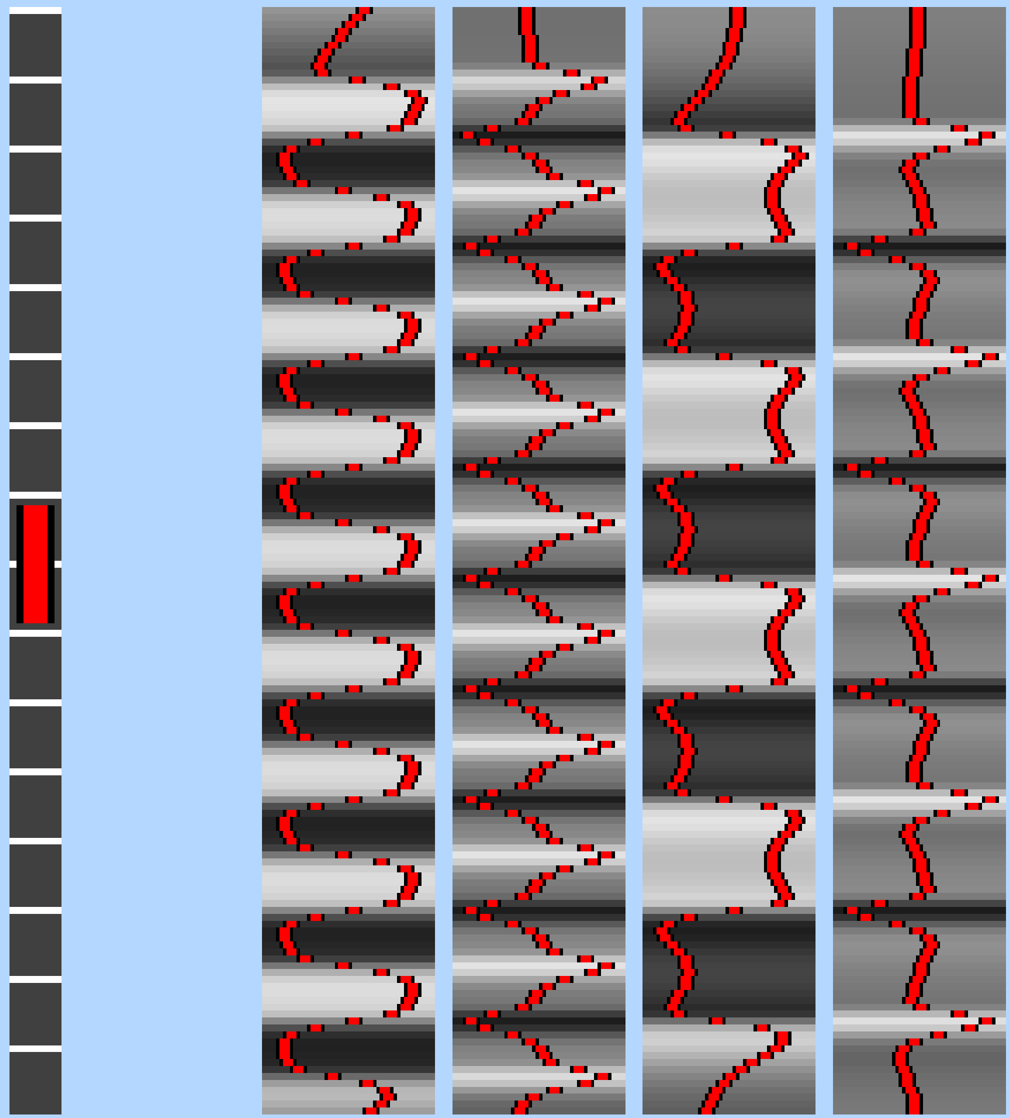
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{c^T \hat{\beta}}{\sqrt{Var(\widehat{c^T \beta})}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix} \qquad c = \begin{matrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ . \\ c_p \end{matrix}$$

Model



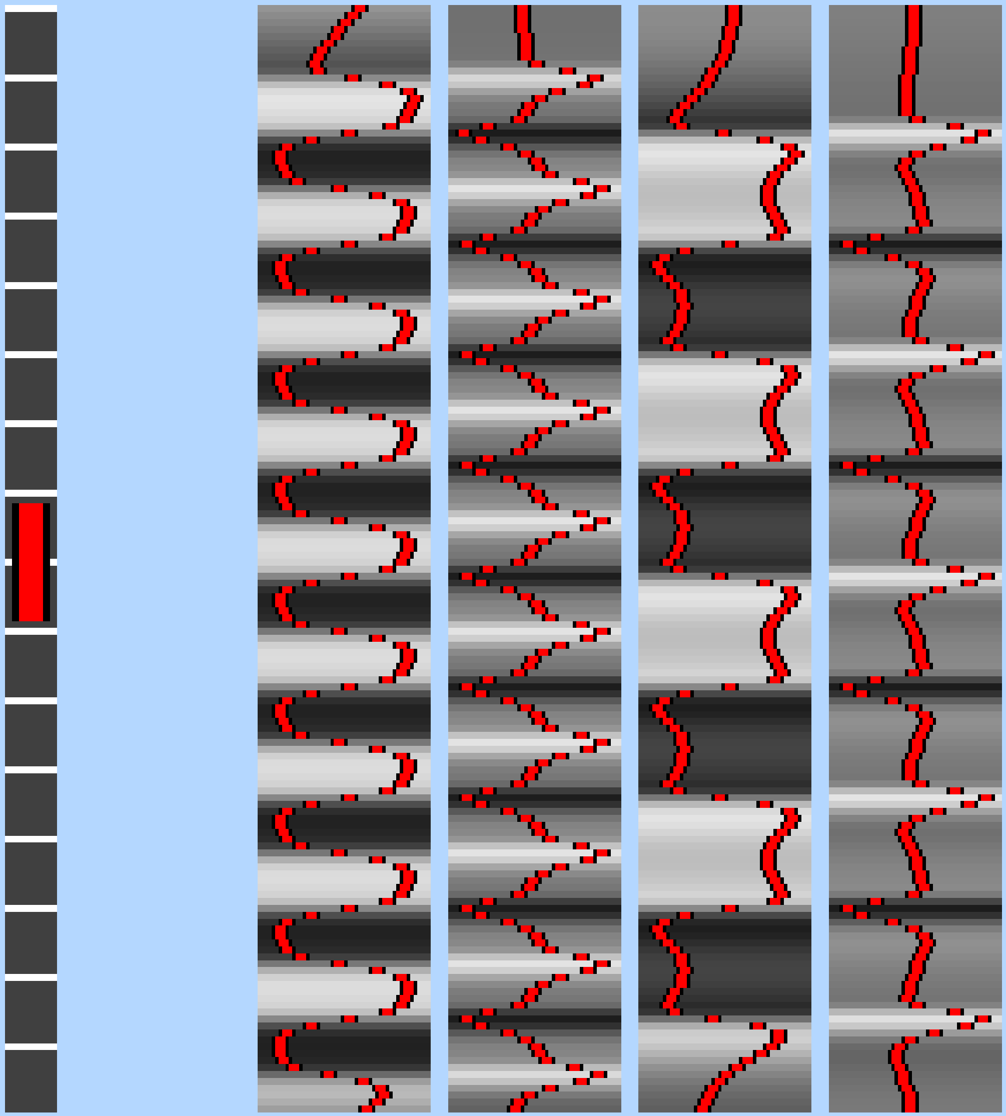
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

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

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix} \qquad c = \begin{matrix} 1 \\ 0 \\ 0 \\ 0 \\ . \\ 0 \end{matrix}$$

Model



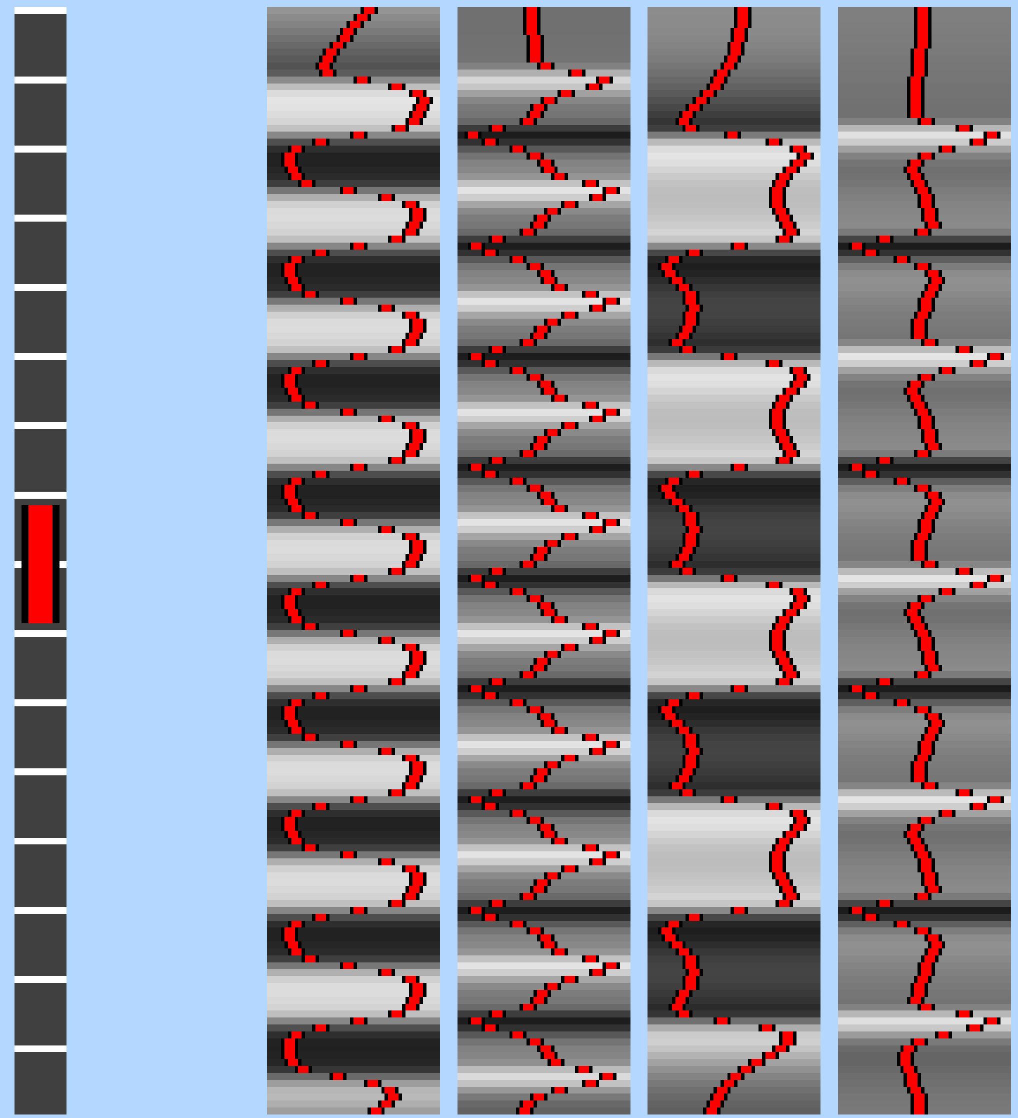
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{c^T \hat{\beta}}{\sqrt{Var(\widehat{c^T \beta})}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix} \qquad c = \begin{matrix} 0 \\ 0 \\ 1 \\ 0 \\ . \\ 0 \end{matrix}$$

Model



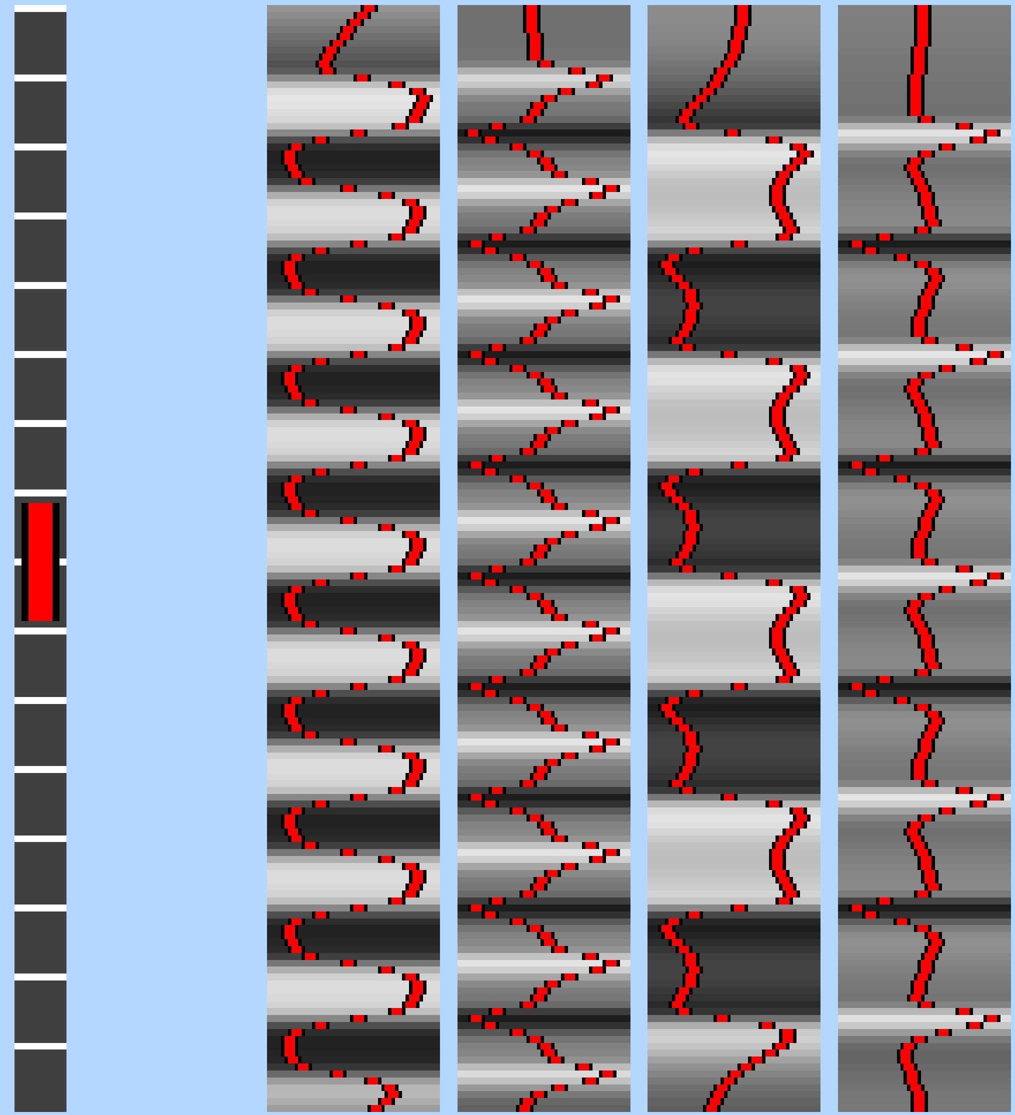
		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{c^T \hat{\beta}}{\sqrt{Var(\widehat{c^T \beta})}}$$

$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix} \qquad c = \begin{matrix} 1 \\ 0 \\ -1 \\ 0 \\ . \\ 0 \end{matrix}$$

Model

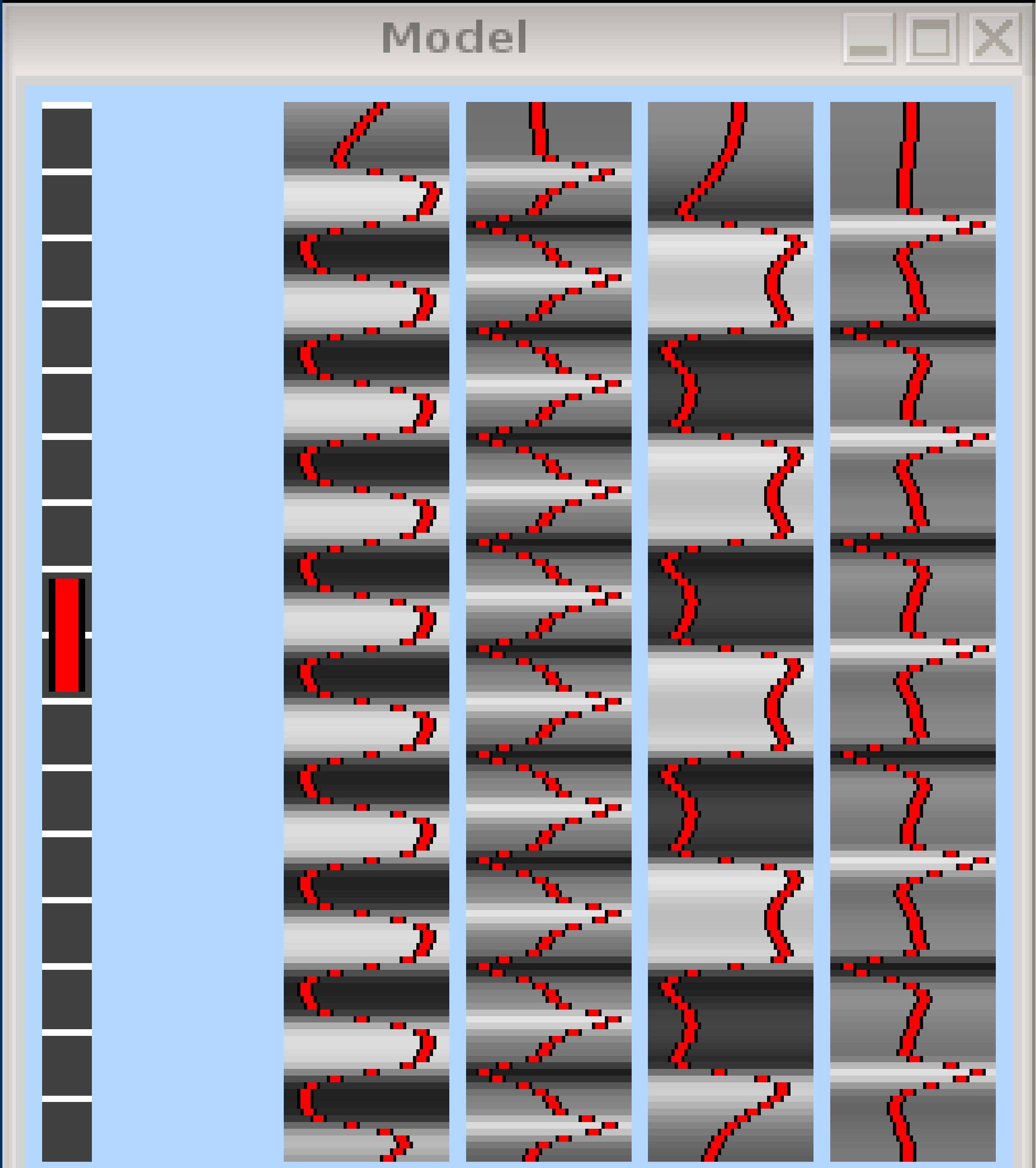


		Cond A	TD A	Cond B	TD B
c1	A	1	0	0	0
c2	B	0	0	1	0
c3	A > B	1	0	-1	0
c4	B > A	-1	0	1	0

Contrasts

$$t = \frac{c^T \hat{\beta}}{\sqrt{Var(\widehat{c^T \beta})}}$$

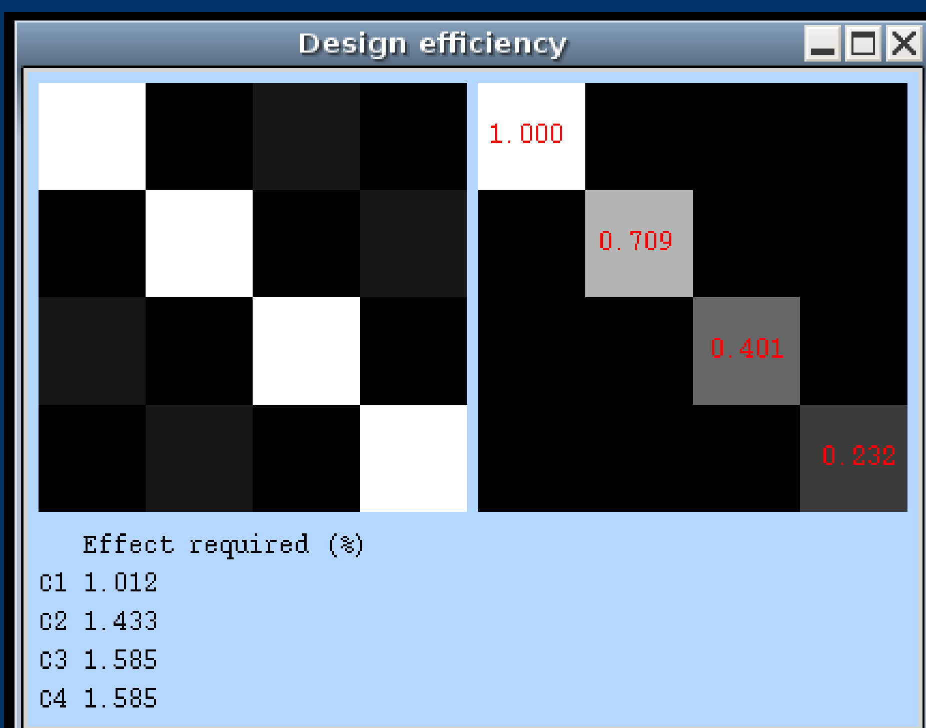
$$\hat{\beta} = \begin{matrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ . \\ \hat{\beta}_p \end{matrix} \qquad c = \begin{matrix} -1 \\ 0 \\ 1 \\ 0 \\ . \\ 0 \end{matrix}$$

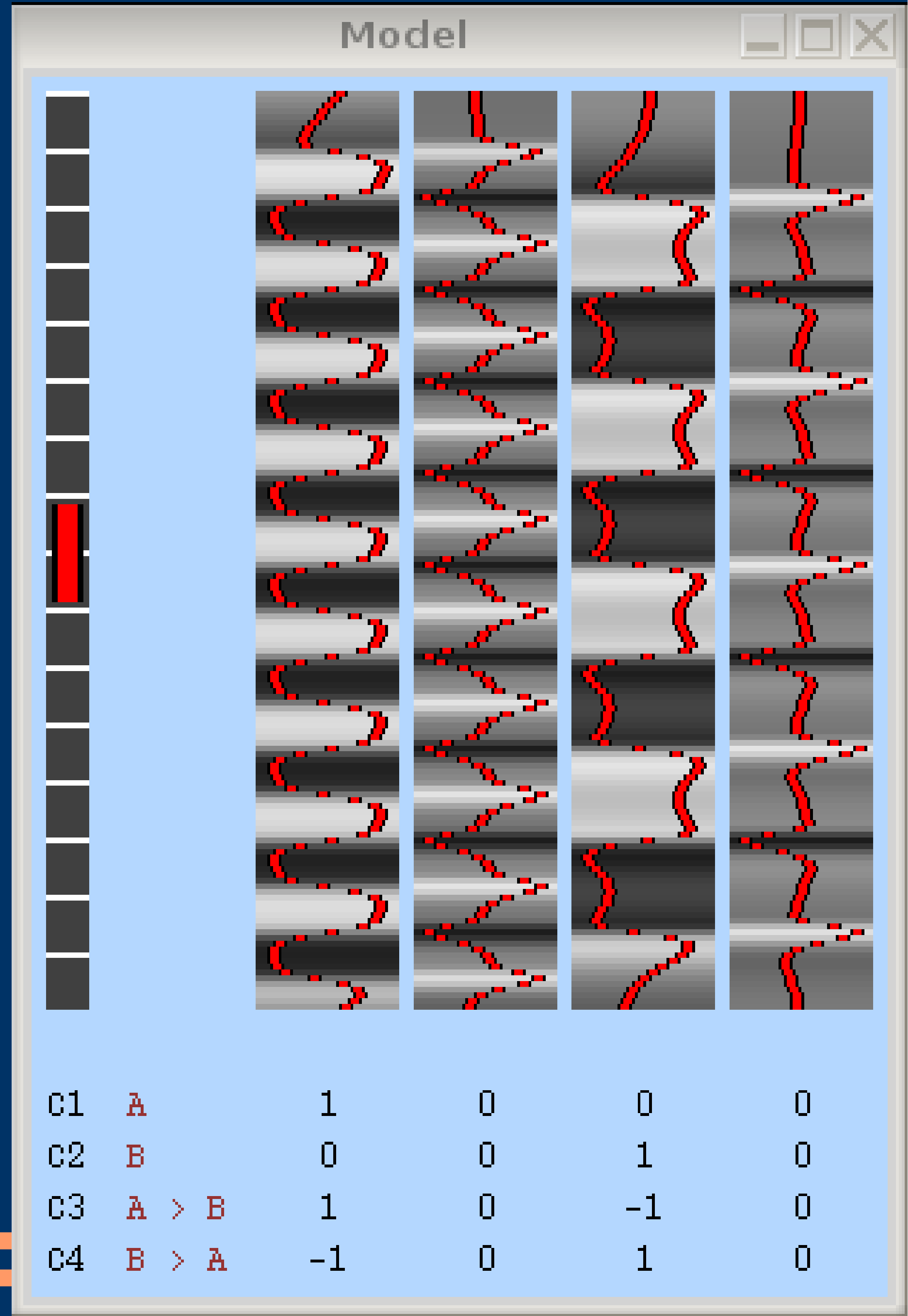


		Cond A	TD A	Cond B	TD B
C1	A	1	0	0	0
C2	B	0	0	1	0
C3	A > B	1	0	-1	0
C4	B > A	-1	0	1	0

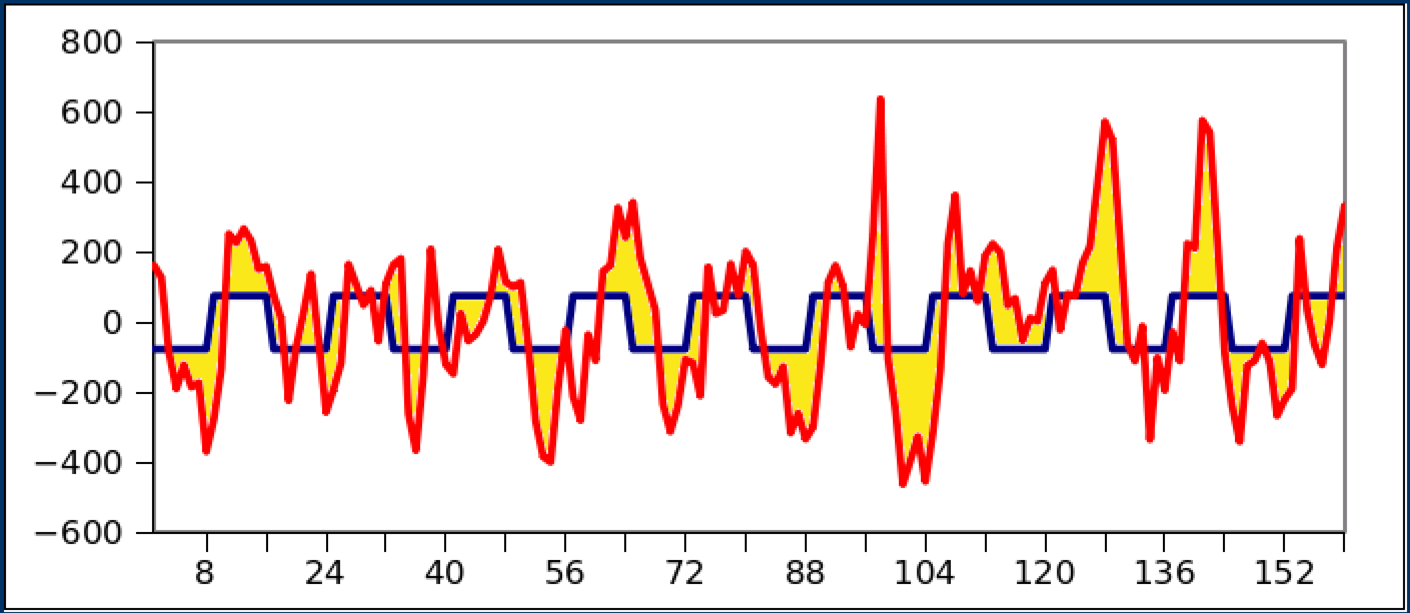
Contrasts

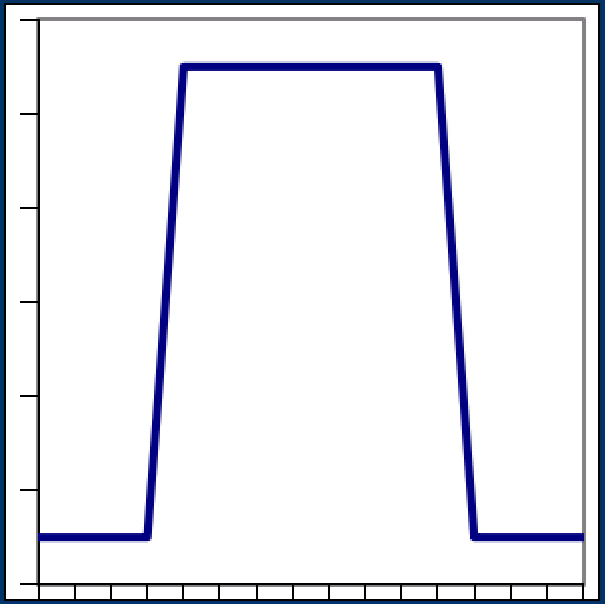
$$t = \frac{c^T \hat{\beta}}{\sqrt{\widehat{Var}(c^T \beta)}}$$





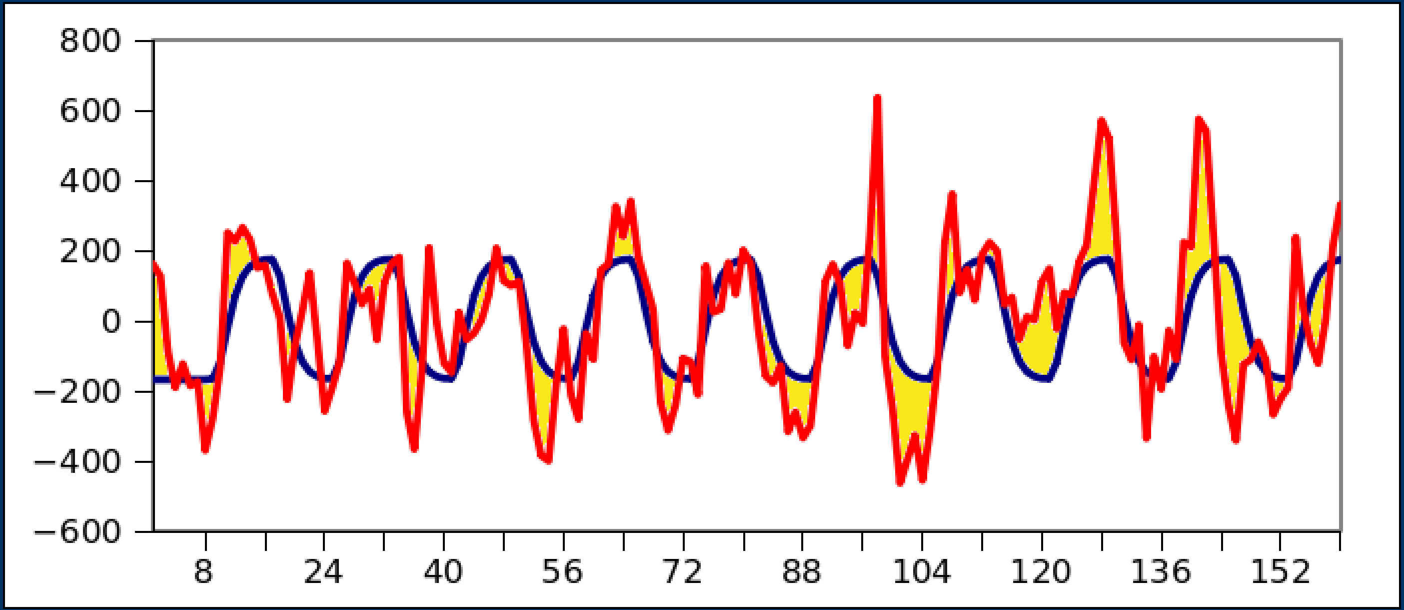
Model I: Boxcar Regressor

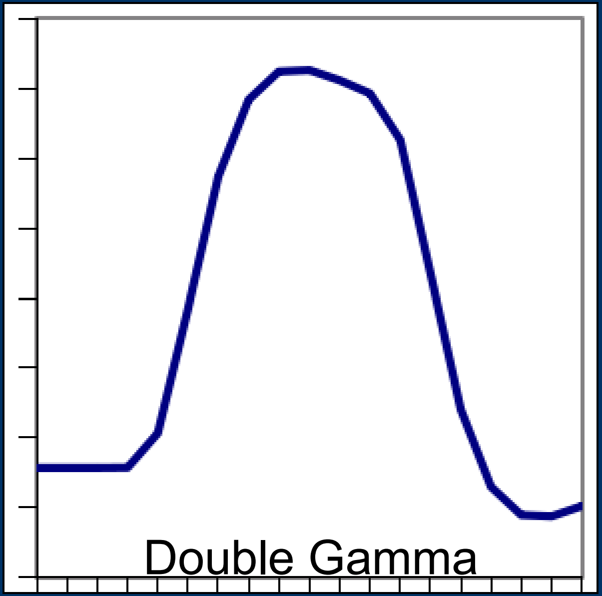




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{150.99}{202.44 \times 0.16} = 4.66$$

Model II: Convolved Boxcar Regressor





$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

Gauss Markov Assumptions

- Under the following assumptions, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I)$$

$$E(\epsilon_i) = 0$$

Error has mean 0

Error is Homoschedastic

$$Var(\epsilon_i) = \sigma^2$$

Errors have same variance

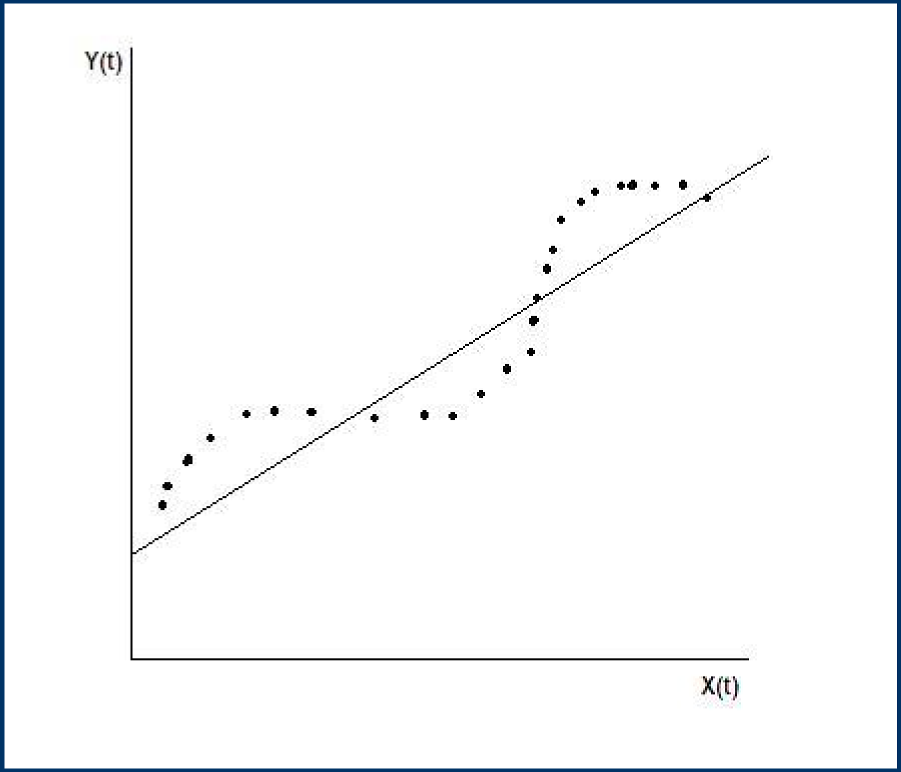
Errors are not Autocorrelated

$$Cov(\epsilon_i, \epsilon_j) = 0$$

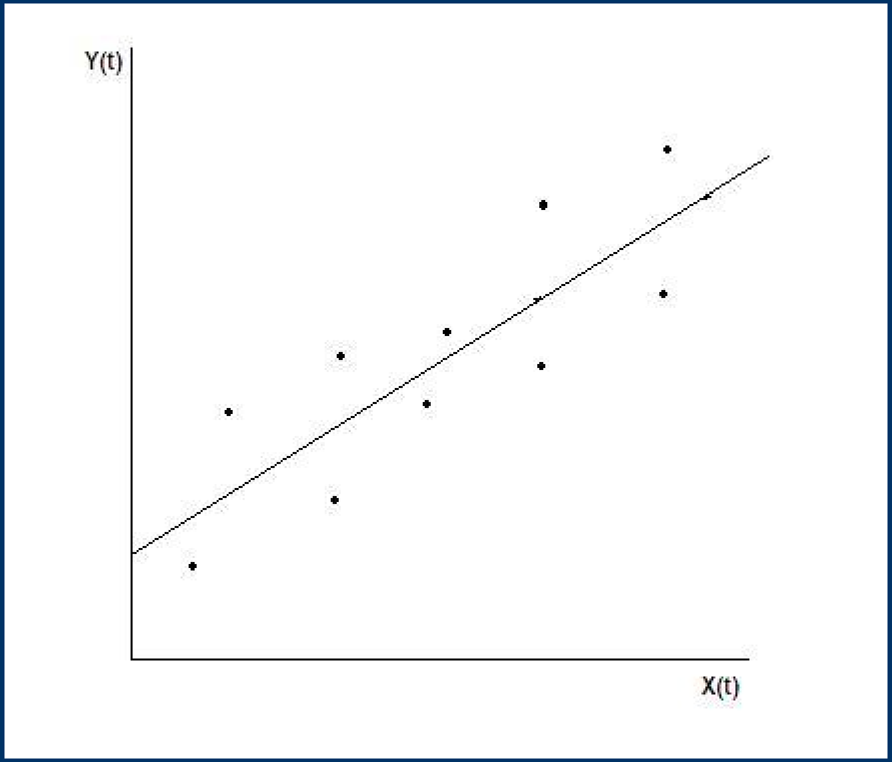
Errors are independent of each other

Temporal Structure

- If residuals are correlated, estimator is still unbiased but the variance estimate is not; and df are less than nominal.



$\text{Corr}(\varepsilon_t, \varepsilon_{t-1}) > 0$



$\text{Corr}(\varepsilon_t, \varepsilon_{t-1}) < 0$

Temporal Structure

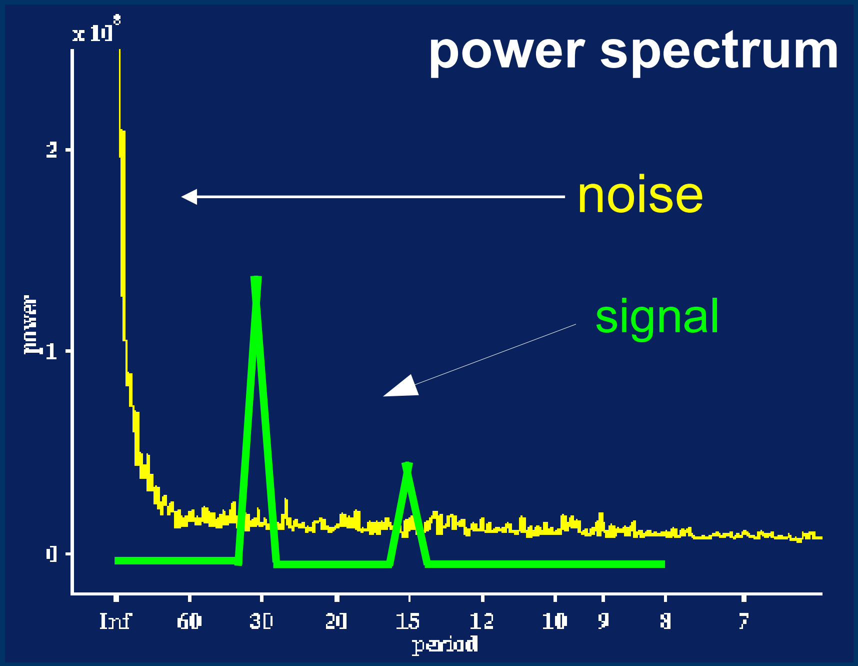
- If residuals are correlated, estimator is still unbiased but the variance estimate is not; and df are less than nominal.

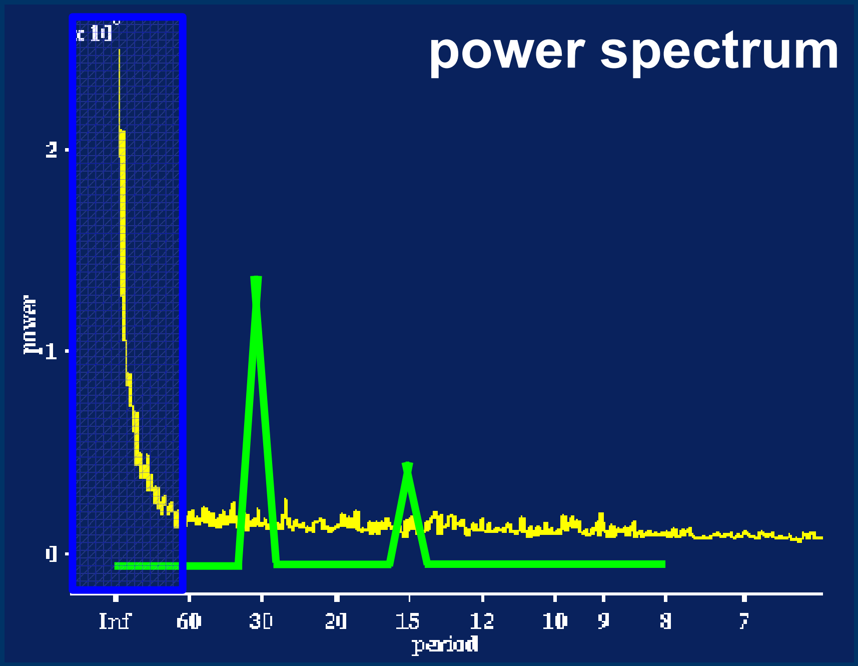
- This can affect your T-test either way:
  i. Positive autocorrelation: underestimate of variance, T-tests artificially inflated (actual α up to 0.16 for a nominal of 0.05).
  ii. Negative autocorrelation: overestimate of variance, T-test artificially reduced.

- Two strategies:
  i. High Pass filtering
  ii. Autocorrelation correction ("pre-whitening"; "colouring")

Low frequency noise

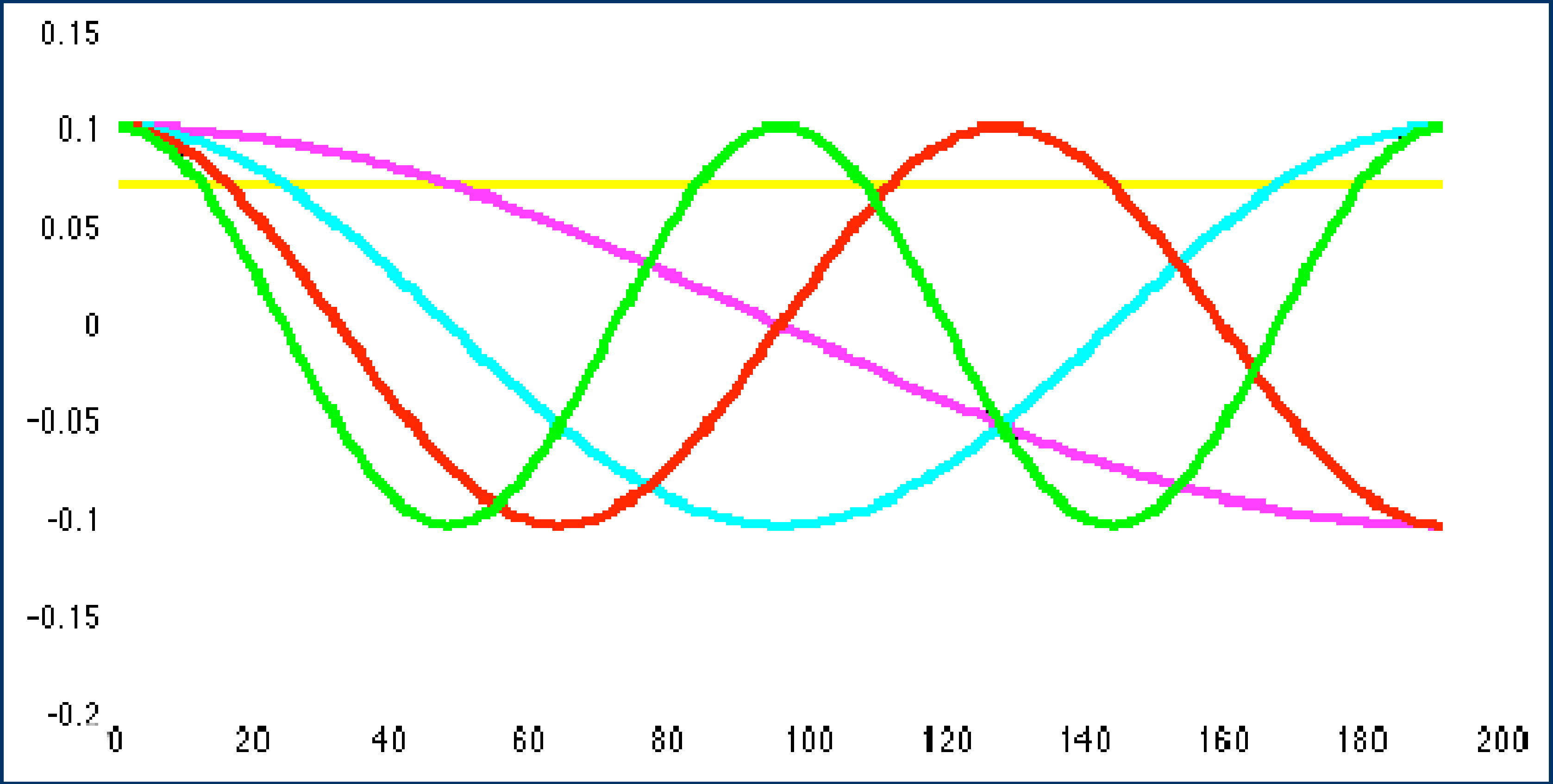
- Plenty low frequency noise: scanner drift, physiological (e.g., basal metabolism), psychological (e.g., learning, alertness changes), left-over from motion correction, ...





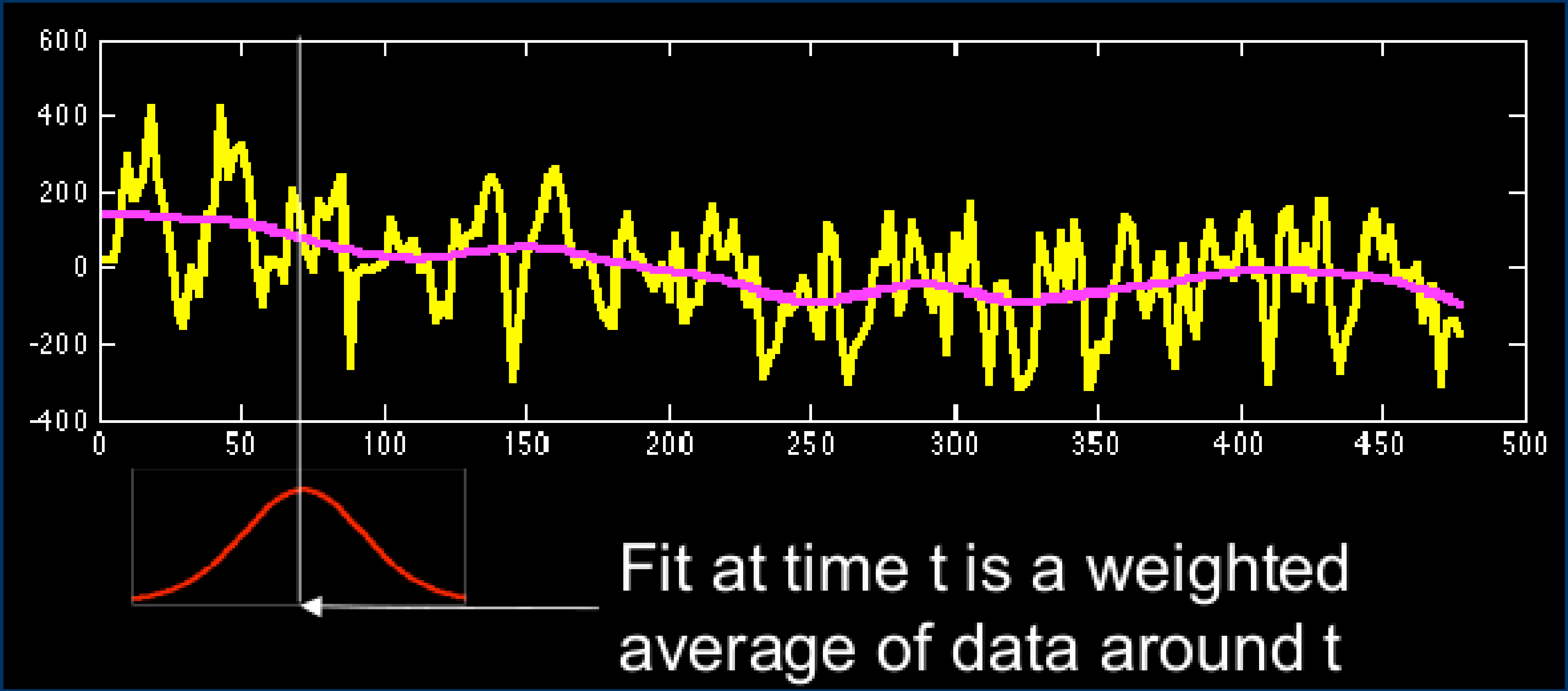
HP Filtering Strategy I: SPM

- Model low drifts to “soak up” their variance (using a discrete cosine transform basis set).



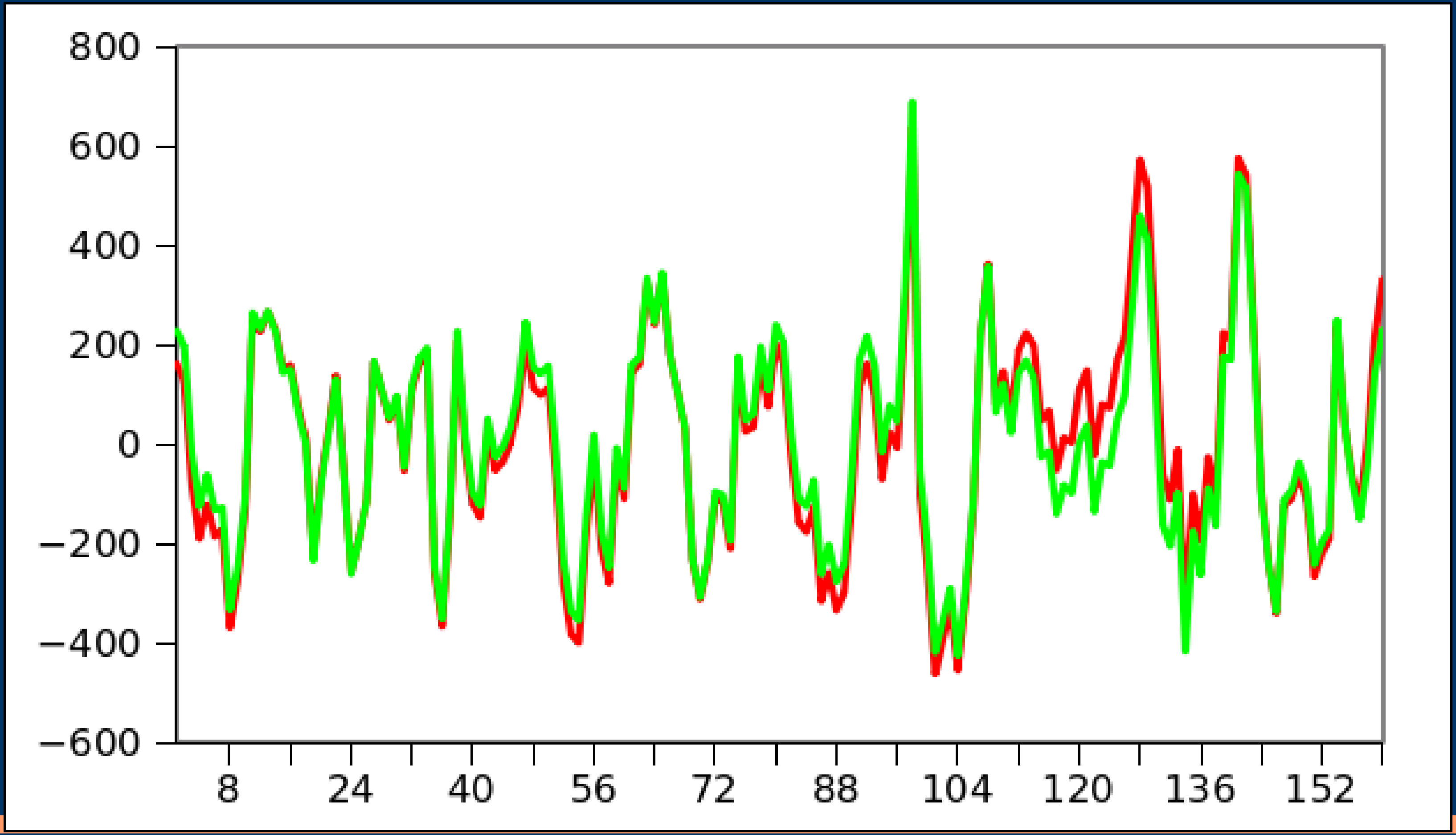
HP Filtering Strategy II: FSL

•Remove low drifts from the signal:
  i. Fit a Gaussian-weighted running line

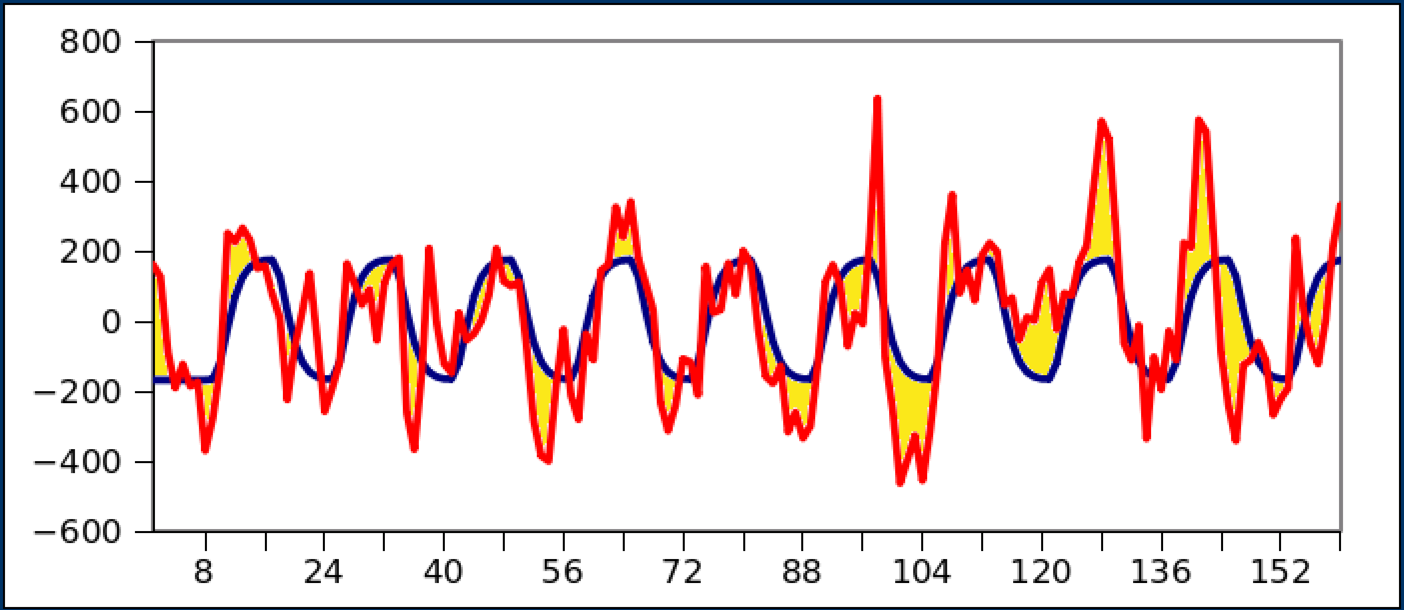


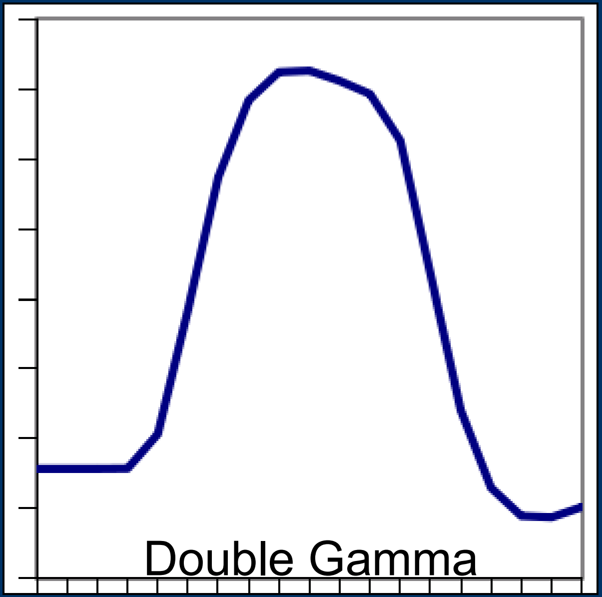
HP Filtering Strategy II: FSL

- Remove low drifts from the signal:
  i. Fit a Gaussian-weighted running line
  ii. Subtract from data (red is pre-HPF, green is post-HPF)



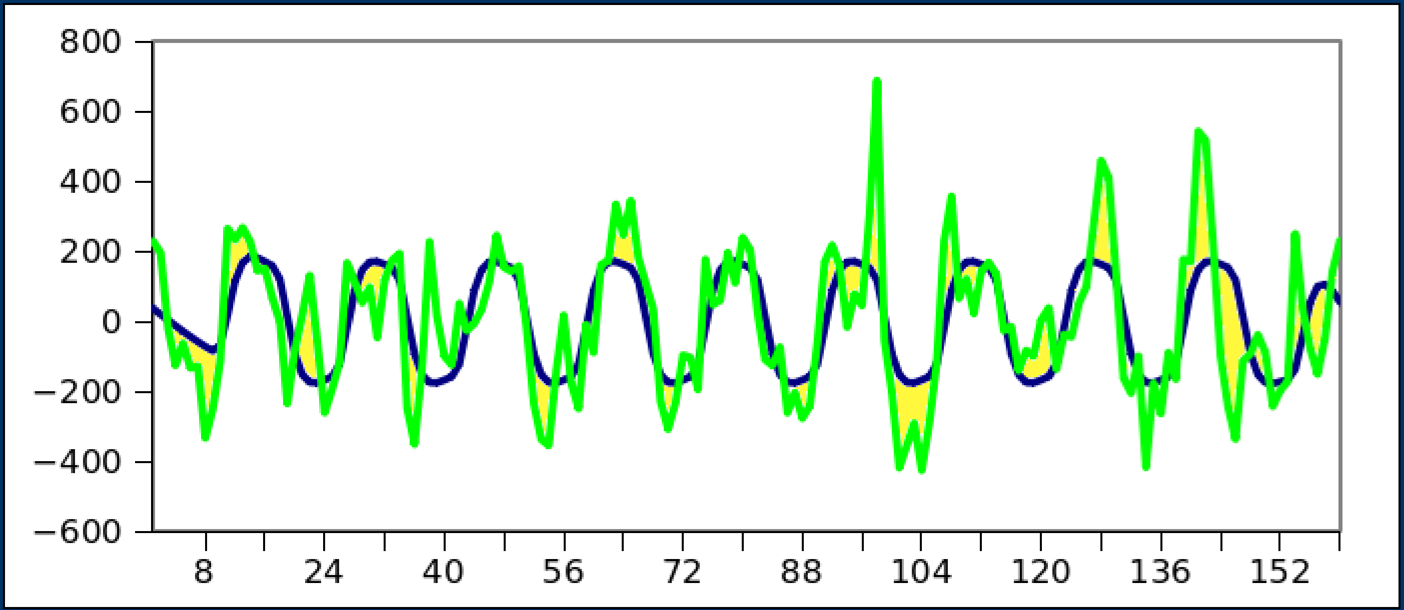
Model II: Convolved Boxcar Regressor

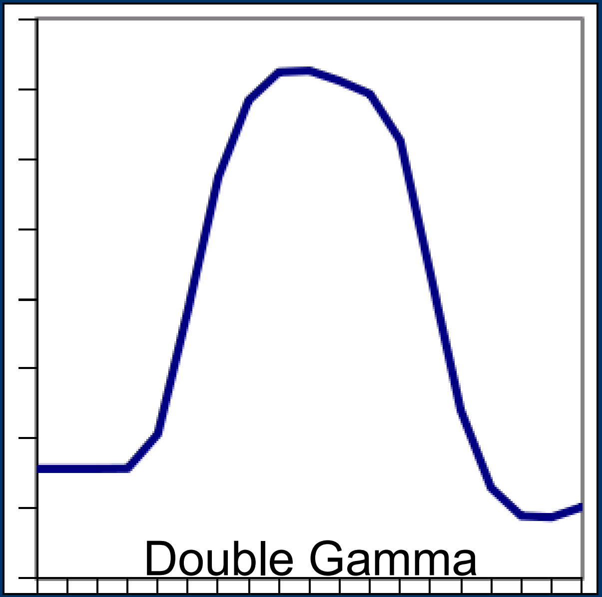




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{344.74}{168.67 \times 0.16} = 12.77$$

Model III: High Pass Filter





$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{349}{154.08 \times 0.21} = 11.05$$

Autocorrelation

GM Assumption:

$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I) \quad Cov(\epsilon_i, \epsilon_j) = 0$$

Errors are independent of each other

fMRI Data:

$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 V) \quad Cov(\epsilon_i, \epsilon_j) = \sigma^2 V$$

$$Y = X\beta + V\epsilon$$

Errors are **NOT** independent of each other

Autocorrelation

GM Assumption:

$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 I) \qquad Cov(\epsilon_i, \epsilon_j) = 0$$

Errors are independent of each other

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2 & \mathrm{cov}(\varepsilon_1,\varepsilon_2) & \dots & \mathrm{cov}(\varepsilon_1,\varepsilon_n) \\ \mathrm{cov}(\varepsilon_2,\varepsilon_1) & \sigma^2 & \dots & \mathrm{cov}(\varepsilon_2,\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{cov}(\varepsilon_n,\varepsilon_1) & \mathrm{cov}(\varepsilon_2,\varepsilon_n) & \dots & \sigma^2 \end{bmatrix}$$

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it:
- Given **V**, there exist a square (non singular) matrix **K** such that:
  i. $\mathbf{V} = \mathbf{K}\mathbf{K}^{\mathrm{T}}$
  ii. $e = K\varepsilon$ (where ε is a 'well behaved' residual $\sim N(0,\sigma^2 I)$)
- So, we can now clean-up (i.e., "whiten") our GLM

$$K^{-1}Y = K^{-1}X\beta + K^{-1}\epsilon$$

$$\begin{aligned} Cov(K^{-1}\epsilon) &= \sigma^2 K^{-1} V (K^T)^{-1} \\ &= \sigma^2 K^{-1} K K^T (K^T)^{-1} \\ &= \sigma^2 I \end{aligned}$$

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 1

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; but we don't know it.

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 2

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 3

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 8

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05 e06

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel):

$$r_{xx}(\tau) = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)/(N-\tau)$$

- Lag 9

[e01 e02 e03 e04 e05 e06 e07 e08 e09 e10 e11 e12]

[e01 e02 e03 e04 e05

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')
3. Smooth (temporally) the correlation estimate, for each voxel ('Tukey taper').

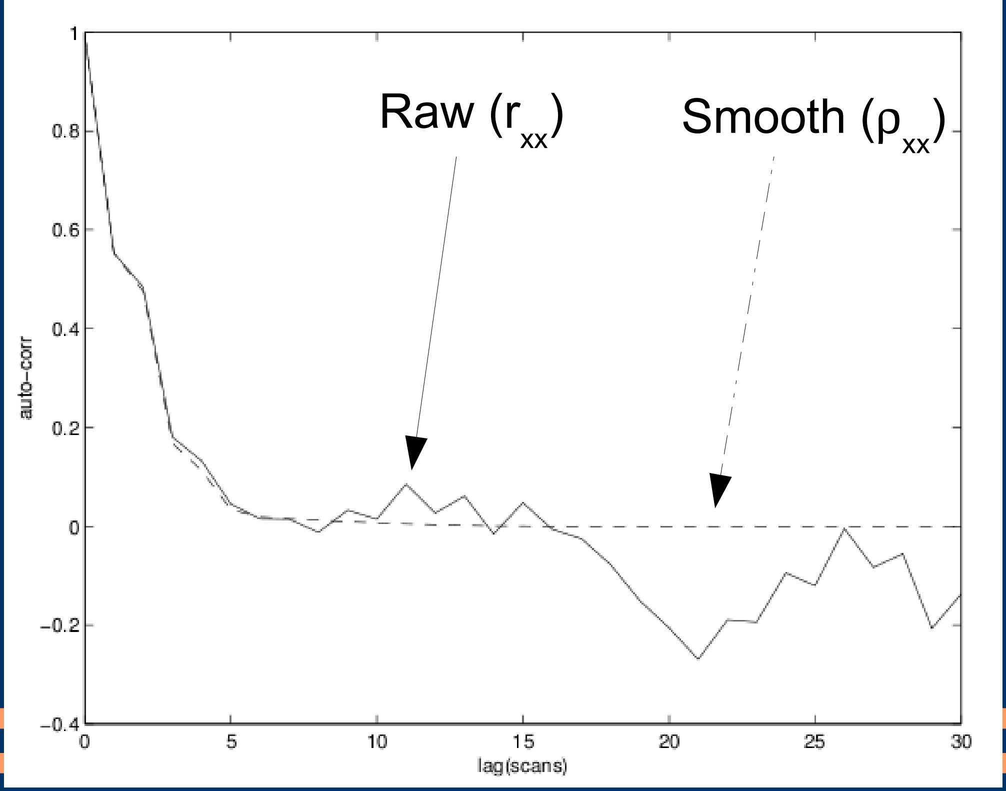
$$\hat{\rho}_{xx}(\tau) = \begin{cases} \frac{1}{2}\left(1 + cos\left(\frac{\pi\tau}{M}\right)\right) r_{xx}(\tau) & \text{if } \tau < M \\ 0 & \text{if } \tau \geq M \end{cases}$$

Woolrich et al, 2001 NeuroImage

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**

1. Estimate the autocorrelation structure (for every voxel)
2. For each lag, smooth estimates spatially ('regularization')
3. Smooth (temporally) the correlation estimate, for each voxel ('Tukey taper').



Woolrich et al, 2001 NeuroImage

Autocorrelation: Pre-whitening FSL

- If only we knew the structure of the serial correlation in the data (V) we could get rid of it; **but we don't know it.**
- Now that we have an estimate of the structure, we can run a second regression in which we "whiten" the residuals using the estimated **V**.
- The estimated β for the new model are BLUE (under the assumption that **V** was appropriately modelled)

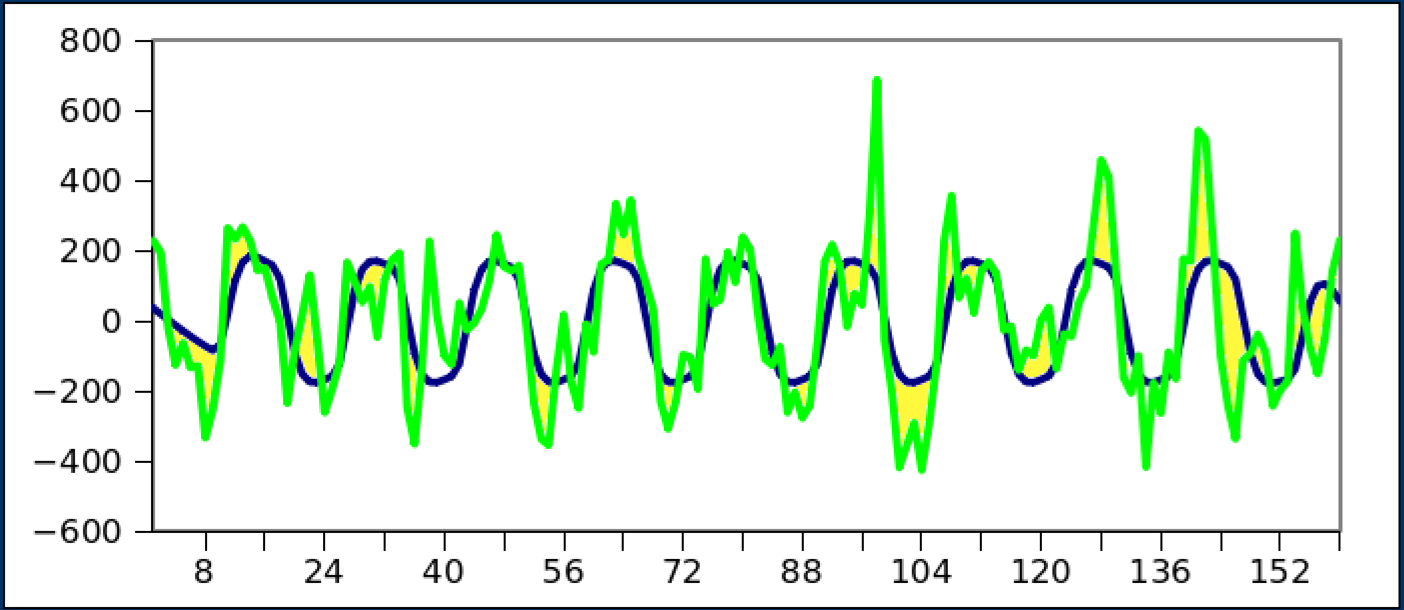
Autocorrelation: pre-whitening SPM

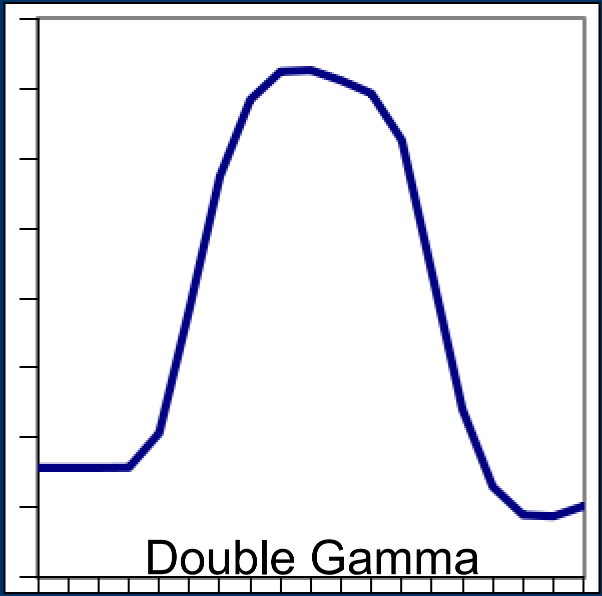
- Global estimation of the correlations: the correlation of the time series is averaged over voxels.
- Imposed AR(1) model with correlation 0.2 plus white noise.

$$\lambda_1 \begin{pmatrix} 1 & 0.2 & 0.04 & 0.008 & 0.0016 & 0.0003 \\ 0.2 & 1 & 0.2 & 0.04 & 0.008 & 0.0016 \\ 0.04 & 0.2 & 1 & 0.2 & 0.04 & 0.008 \\ 0.008 & 0.04 & 0.2 & 1 & 0.02 & 0.04 \\ 0.0016 & 0.008 & 0.04 & 0.2 & 1 & 0.2 \\ 0.0003 & 0.0016 & 0.008 & 0.04 & 0.2 & 1 \end{pmatrix} + \lambda_2 I$$

Only 2 parameters are estimated

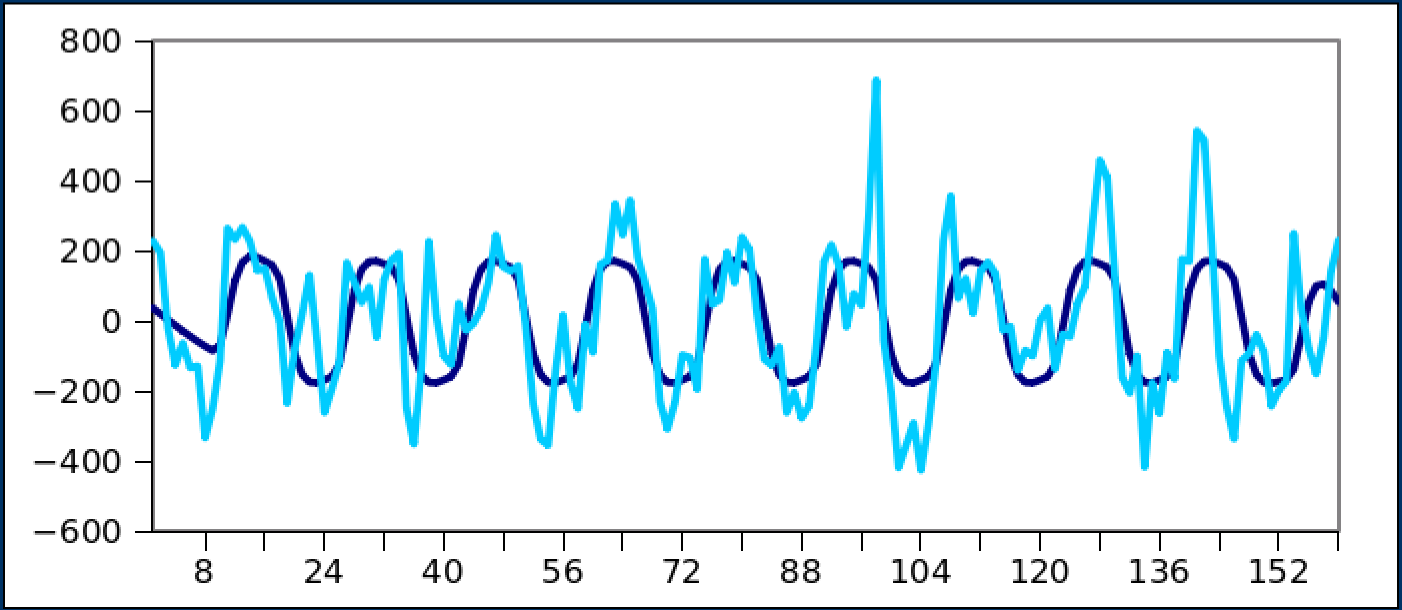
Model III: High Pass Filter

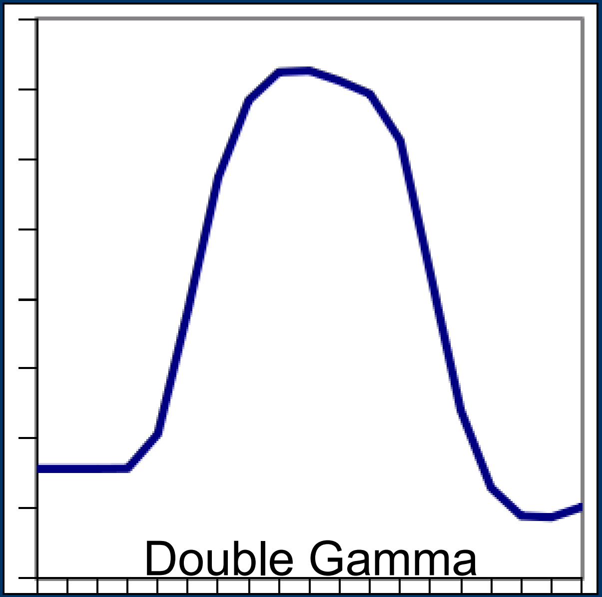




$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \qquad t(159) = \frac{349}{154.08 \times 0.21} = 10.78$$

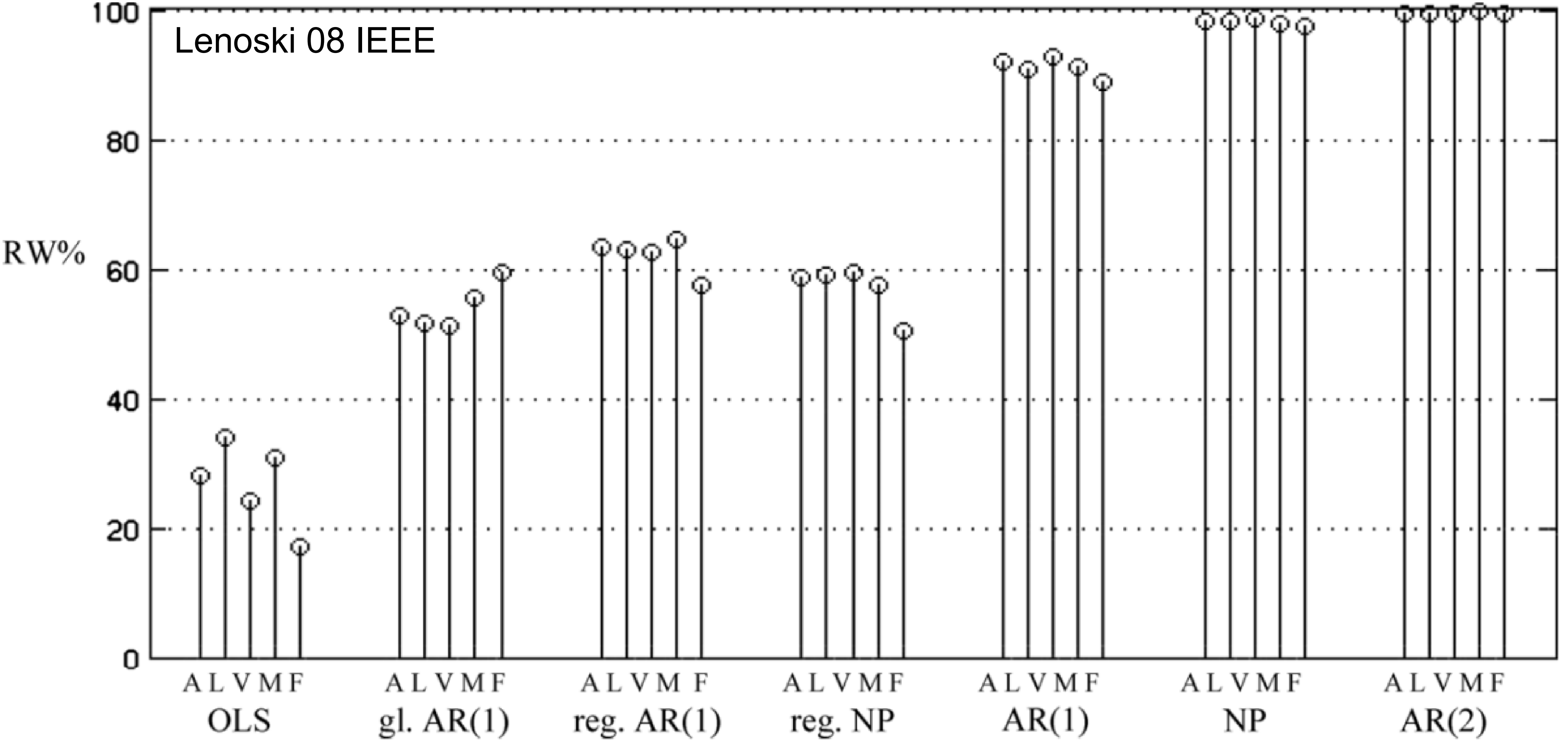
Model IV: Whitening





$$t(df) = \frac{(X^T X)^{-1} X^T Y}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}}} \quad t(159) = \frac{273.36}{130.20 \times 0.21} = 10.00$$

Whitening algorithms performance



Autocorrelation: Coloring

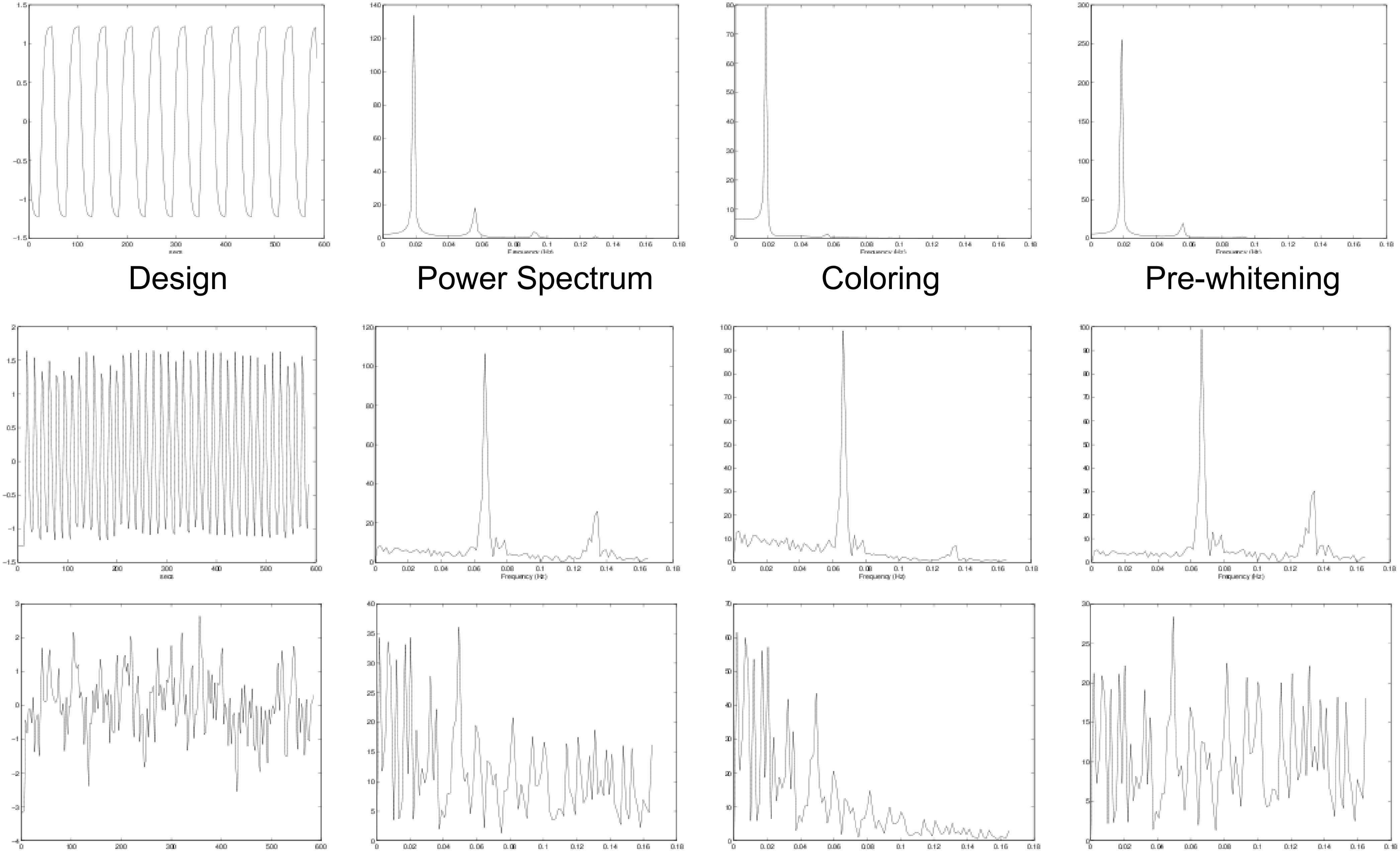
$$\epsilon \overset{i.i.d.}{\sim} N(0, \sigma^2 V) \quad Cov(\epsilon_i, \epsilon_j) = \sigma^2 V$$

$$Y = X\beta + V\epsilon$$

Errors are **NOT** independent of each other

- Impose a known correlation structure S
- If indeed $SVS^T \sim SS^T$ then the error is $N(0, \sigma^2 S)$
- So: $SY = SX\beta + S\varepsilon$
- Estimates are unbiased but no longer efficient (according to how well V is "swamped")

Coloring v. Whitening



Tomorrow…

- Setting up group fMRI models
- Email Jeanette your group model challenges and she may choose it to include in her workshop
  - **jeanette.mumford@gmail.com**