The General Linear Model

A talk for dummies, by dummies

Meghan Morley and Anne Urai

Where are we headed?

- A delicious analogy
- The General Linear Model equation
- What do the variables mean?
- How does this relate to fMRI?
- Minimizing error

Analogy: Reverse Cookery

Start with finished product and try to explain how it is made…

- You specify which *ingredients* to add (*X*)
- For each ingredient, GLM finds the *quantities* (β) that produce the best reproduction
- Then if you tried to make the cake with what you know about X and β then the error would be the difference between the original cake/ data and yours!

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

The General Linear Model

$$\boxed{y = X\beta + e}$$

The General Linear Model

Describes a response (*y*), such as the BOLD response in a voxel, in terms of all its contributing factors (*xβ*) in a linear combination, whilst also accounting for the contribution of error (*e*).

The General Linear Model

$$\boxed{y} = X\beta + e$$

Dependent variable

Describes a response

(such as the BOLD response in a single voxel, taken from an fMRI scan)

The General Linear Model

$$\boxed{y = \boxed{X}\beta + e}$$

Independent Variable

aka. Predictor

e.g. Experimental conditions

(Embodies all available knowledge about experimentally controlled factors and potential confounds)

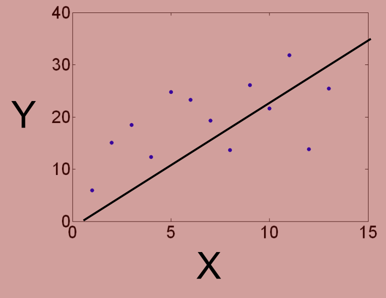
The General Linear Model

$$y = X\boxed{\beta} + e$$

Parameters (aka regression coefficient/beta weights)

Quantifies how much each predictor (X) independently influences the dependent variable (Y)

→ The slope of the line



The General Linear Model

$$\boxed{y = X\beta + \boxed{e}}$$

Error

Variance in the data (*y*) which is not explained by the linear combination of predictors (*x*)

Therefore…

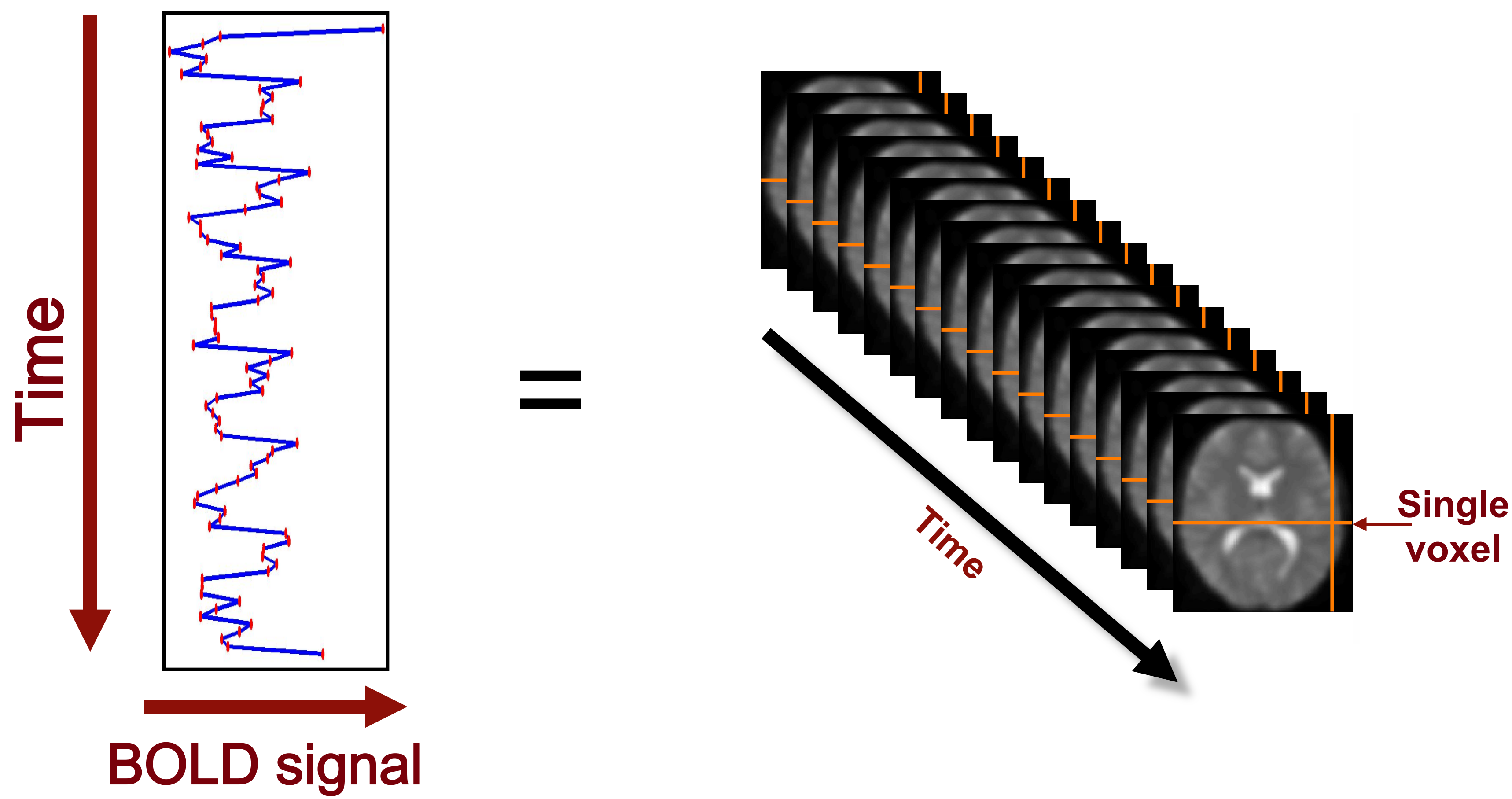
$$y = X\beta + e$$

$$\parallel$$

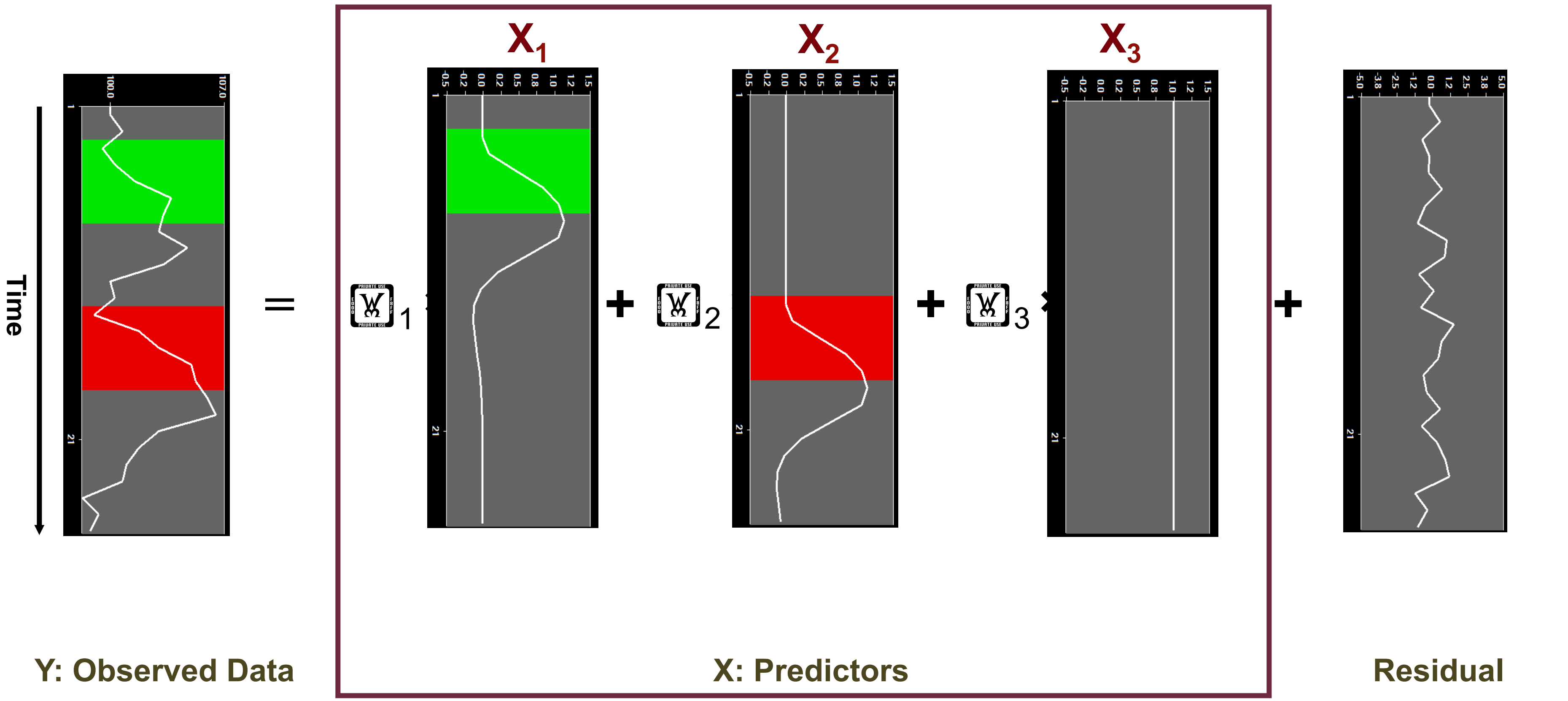
DV = (IV x Parameters) + Error

As we take samples of a response (*y*) *many times*, this equation actually represents a matrix…

…the GLM matrix

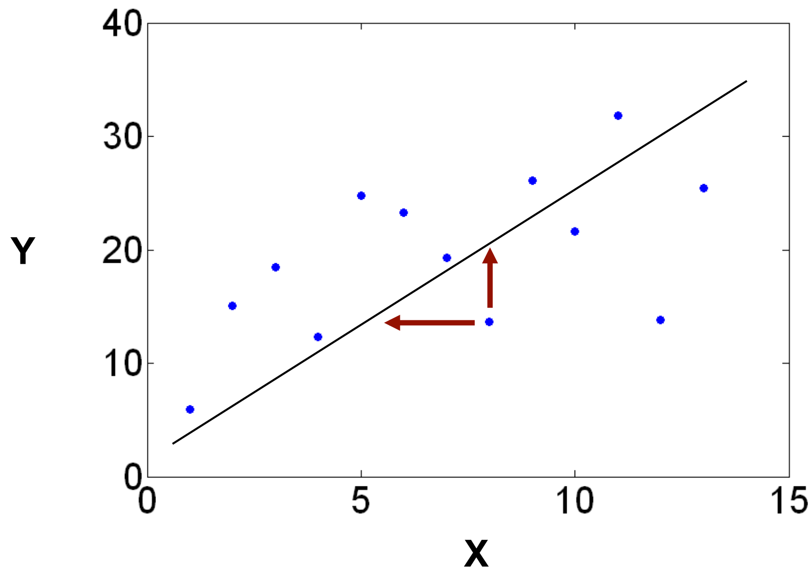


Each predictor (x) has an expected signal time course, which contributes to y



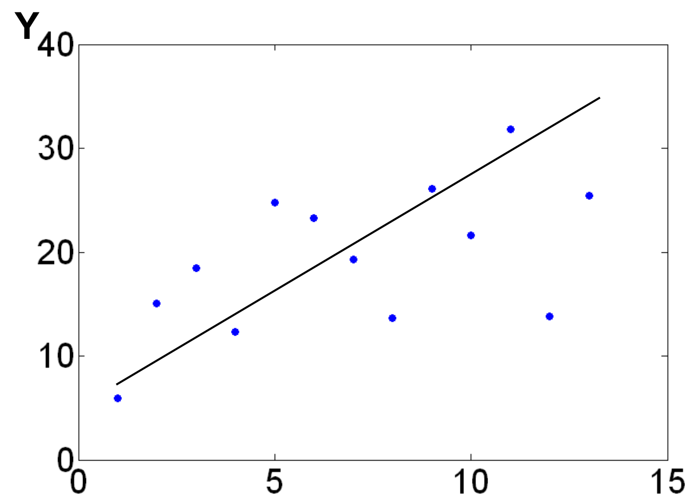
Parameters (β)

- <u>Beta is the slope of the regression line</u>
  - Quantifies a specific predictor's (x) **contribution to *y*.**
  - The parameter (β) chosen for a model should minimise the error (reducing the amount of variance in *y* which is left unexplained)



The design matrix does not account for all of *y*

- If we plot our observations (n) on a graph these will not fall in a straight line

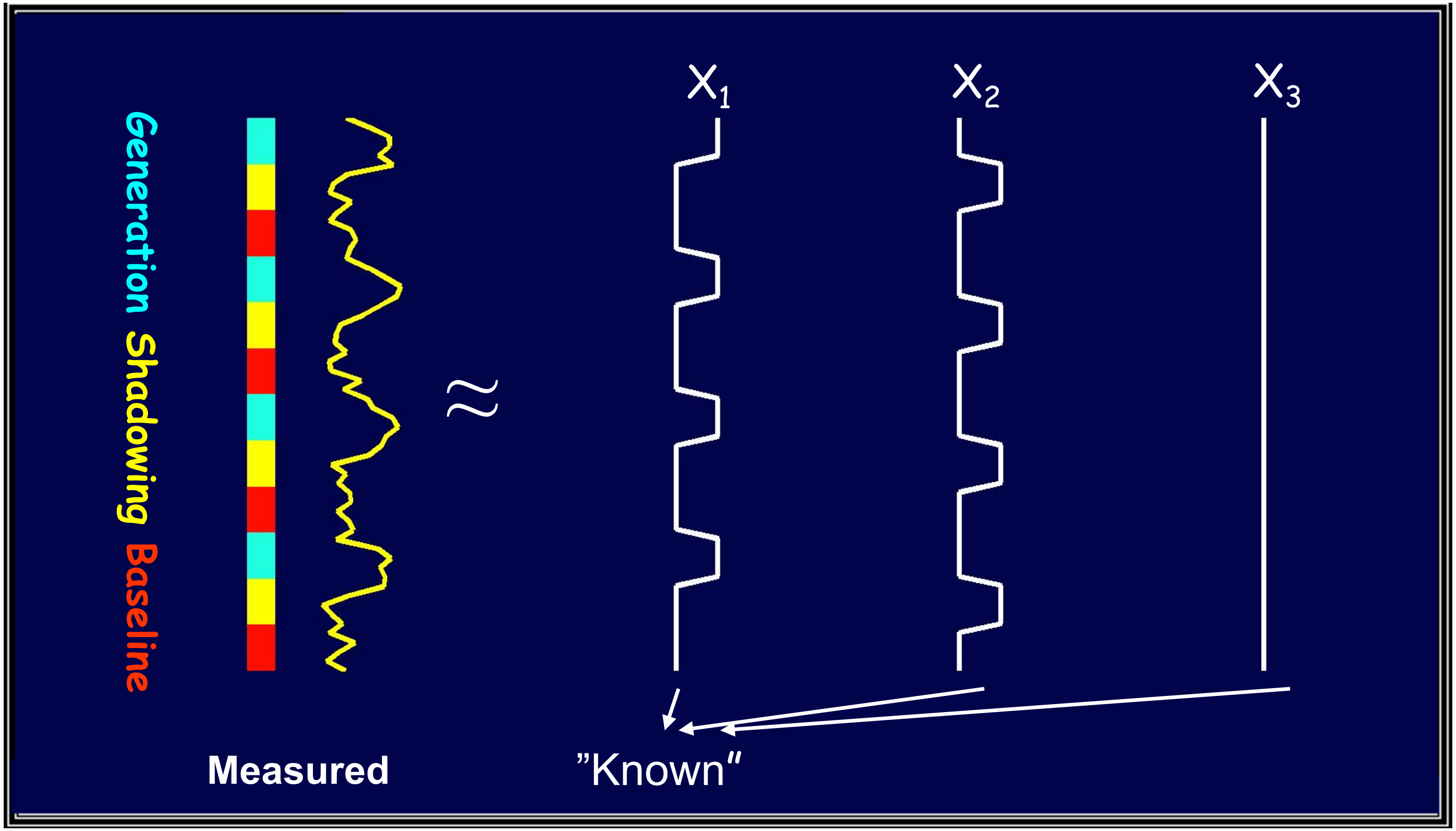


- This is a result of <u>uncontrolled influences</u> (other than *x*) on *y*

- This contribution to *y* is called the **error** (or residual)

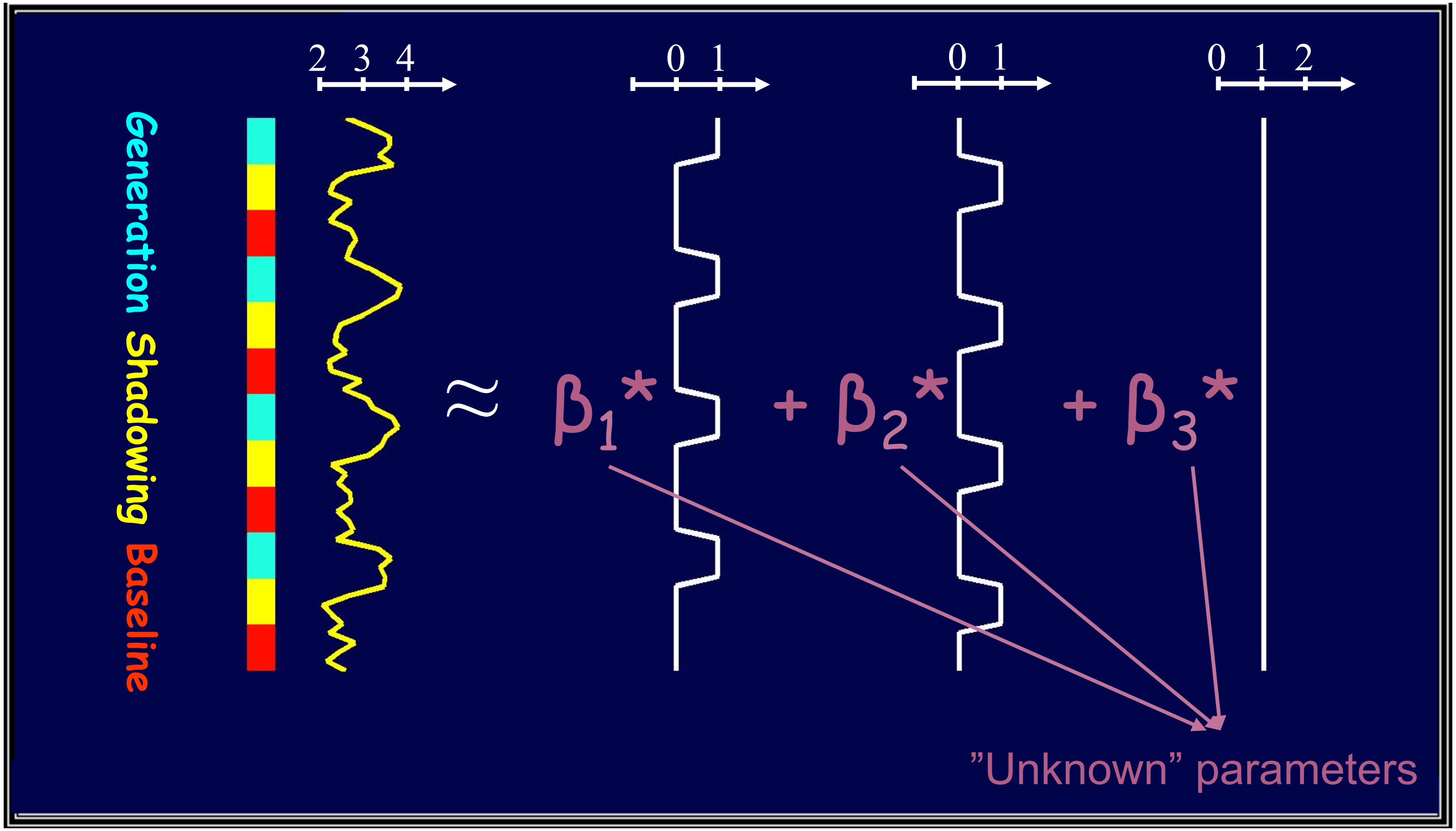
- Minimising the difference between the response predicted by the model ($\hat{y}$) and the actual response (*y*) **<u>minimises the error of the model</u>**

We have our set of hypothetical time-series: *x1*, *x2*, *x3*....



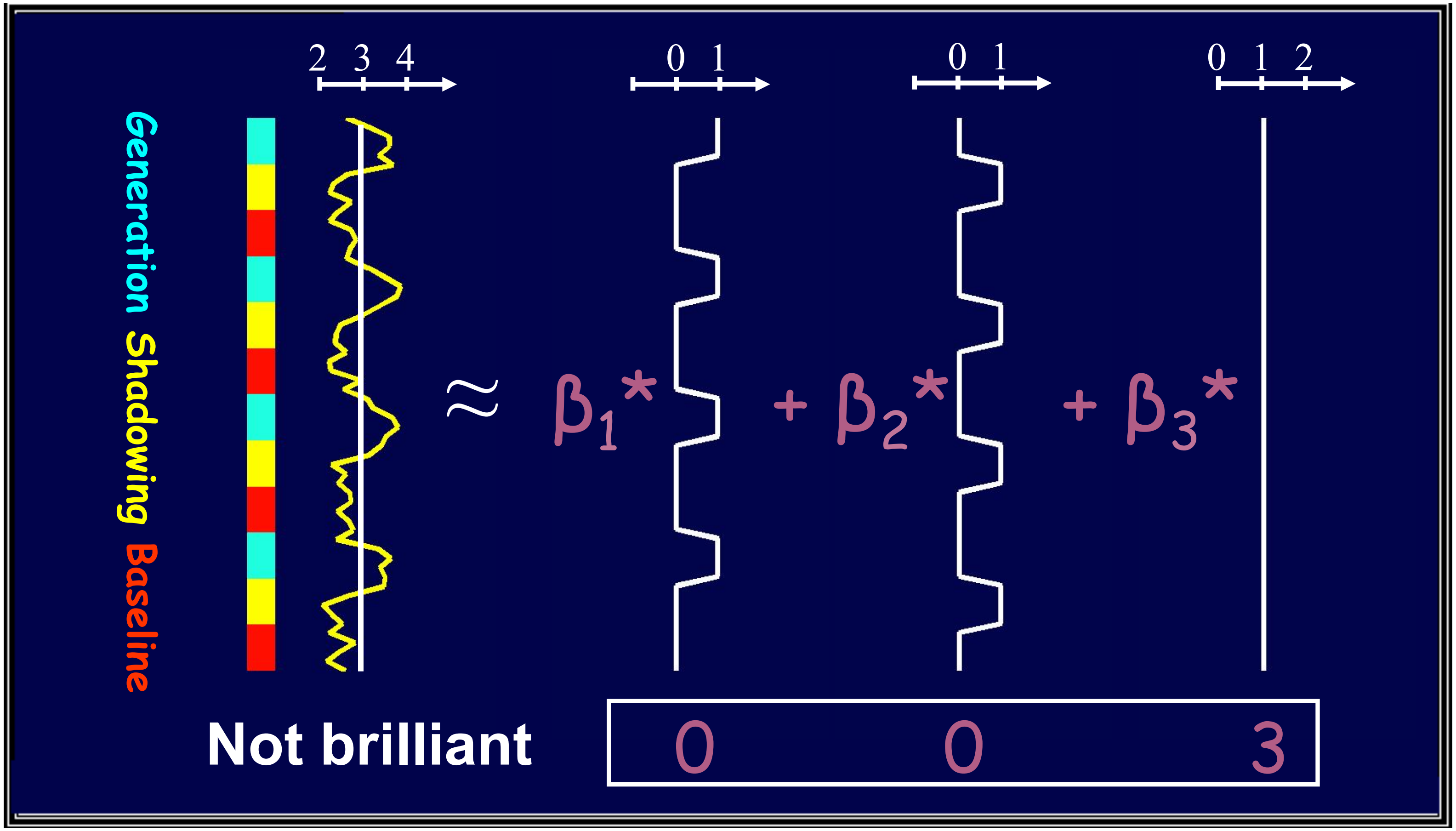
....and our data

We find the best parameter values by modelling…

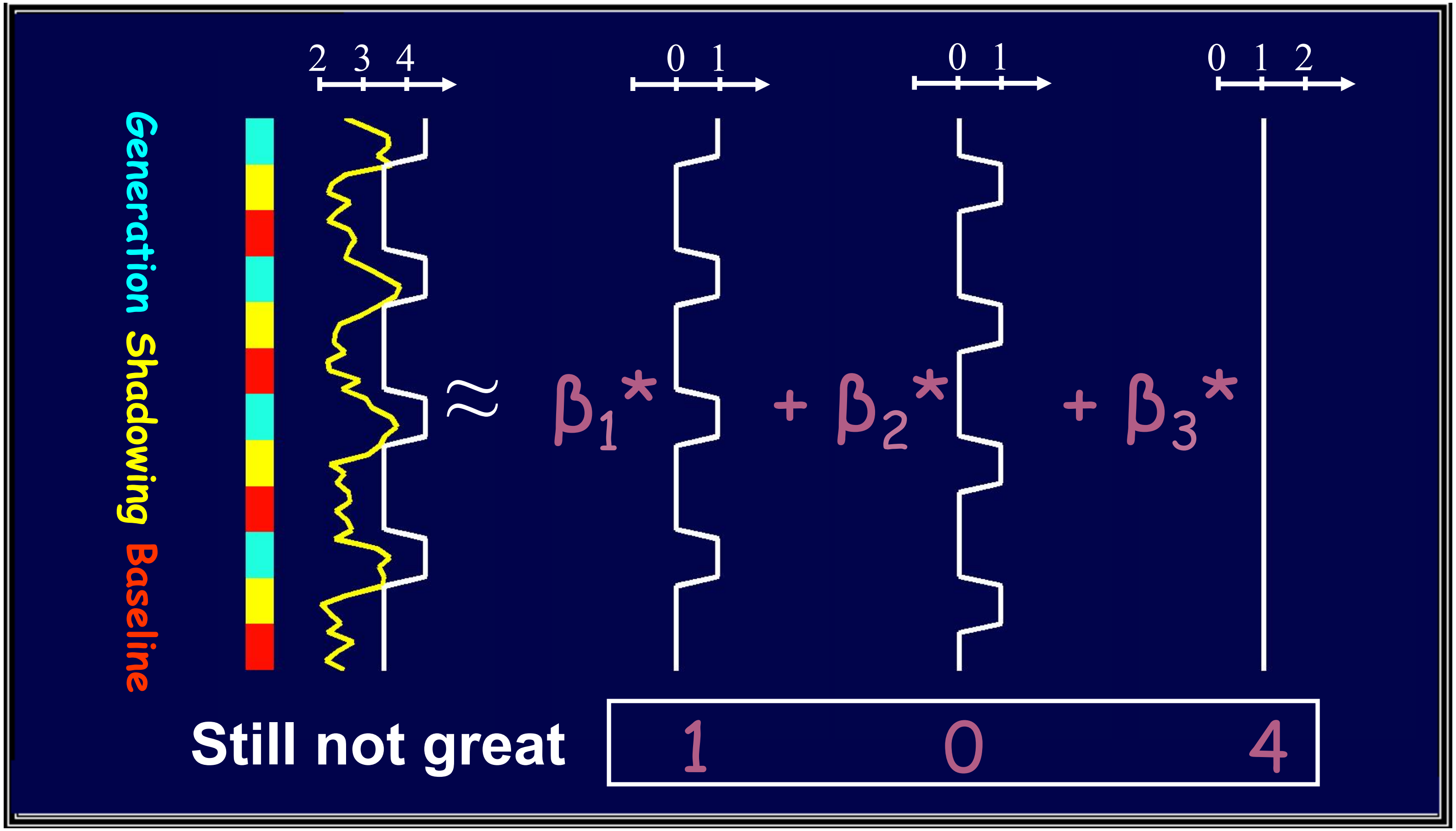


...the best parameter will miminise the error in the model

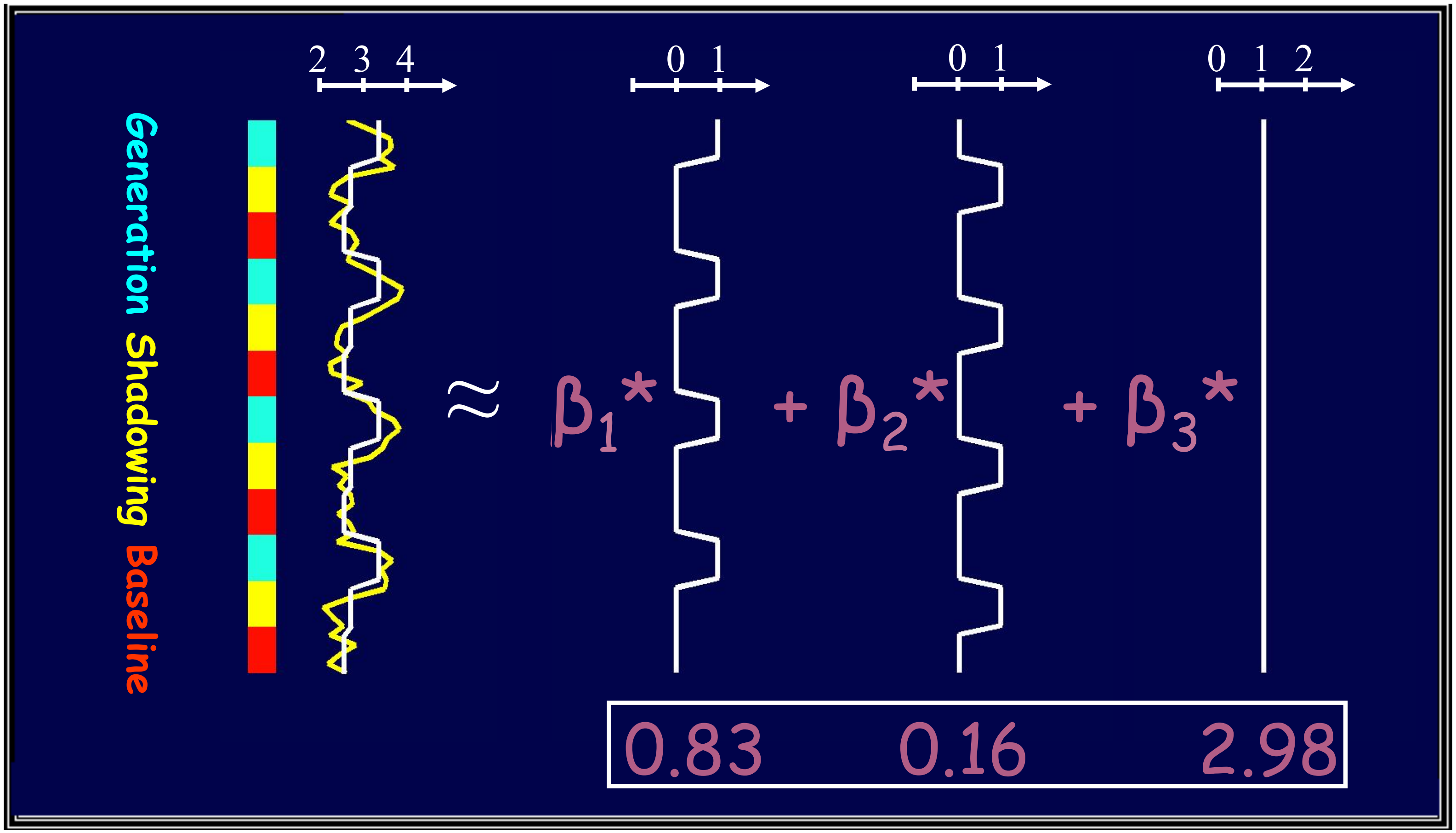
Here, there is a lot of residual variance in *y* which is unexplained by the model (error)



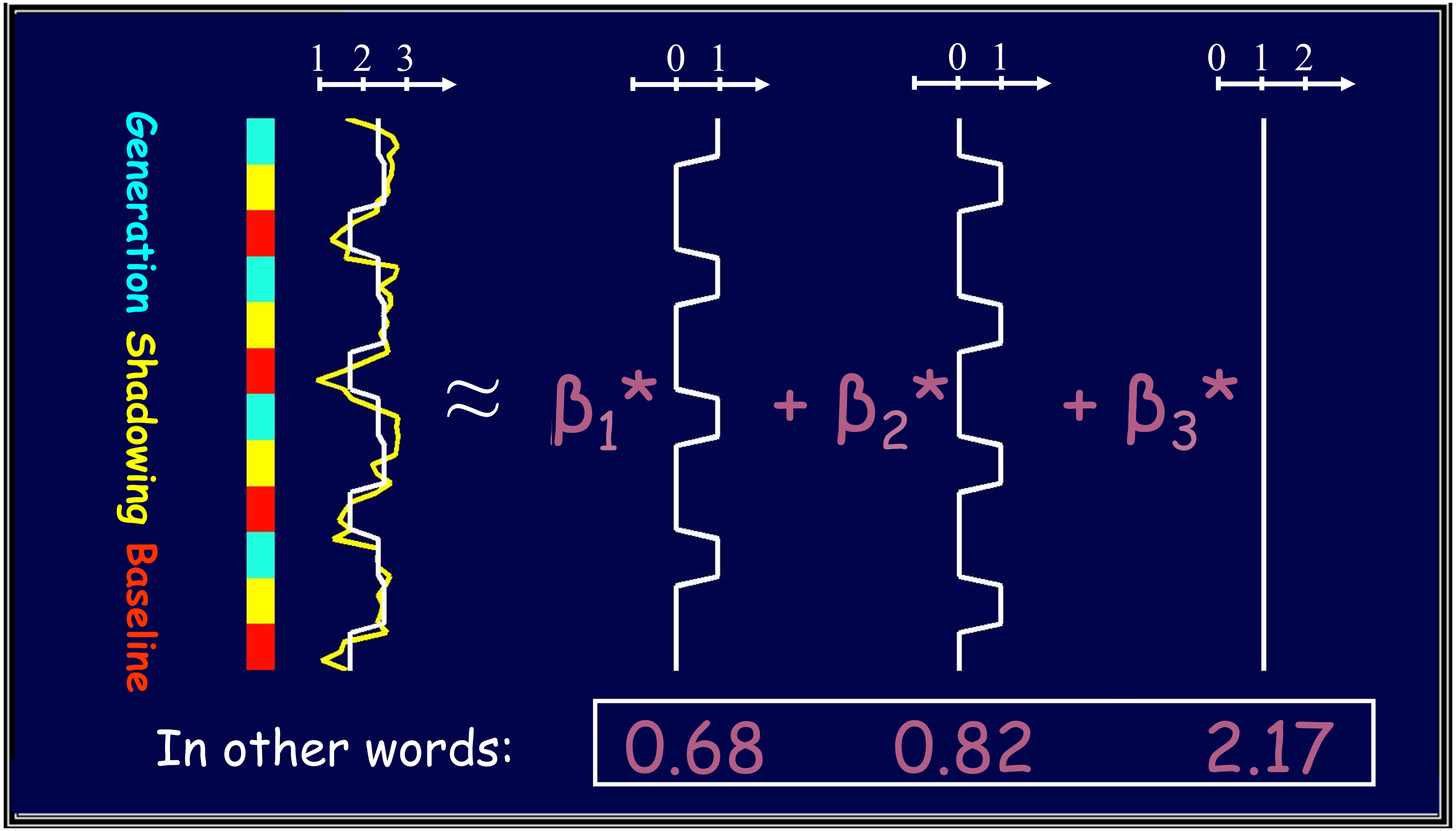
…and the same goes here



…but here we have a good fit, with minimal error



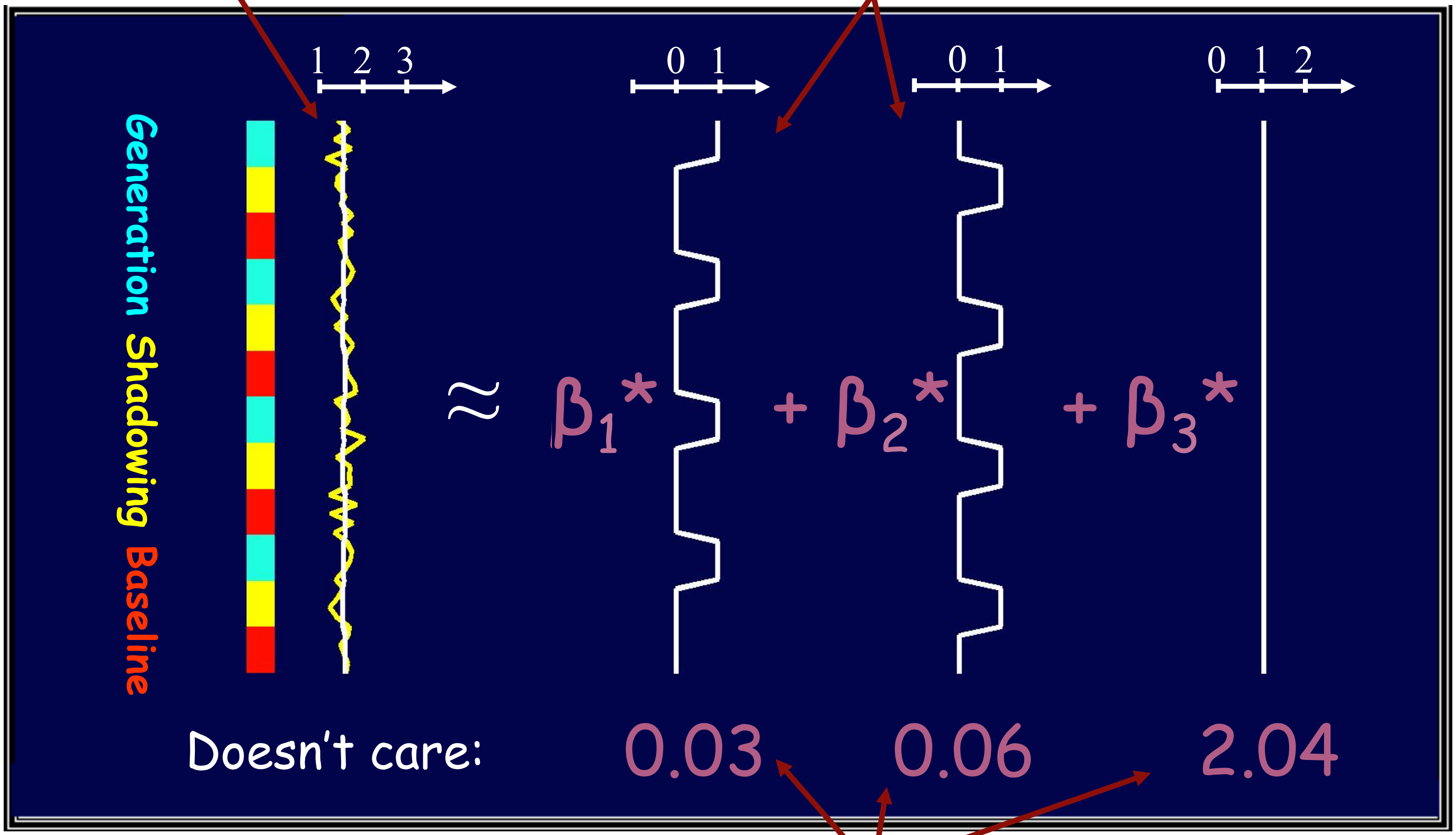
…and the same model can fit different data – just use different parameters



…as you can see

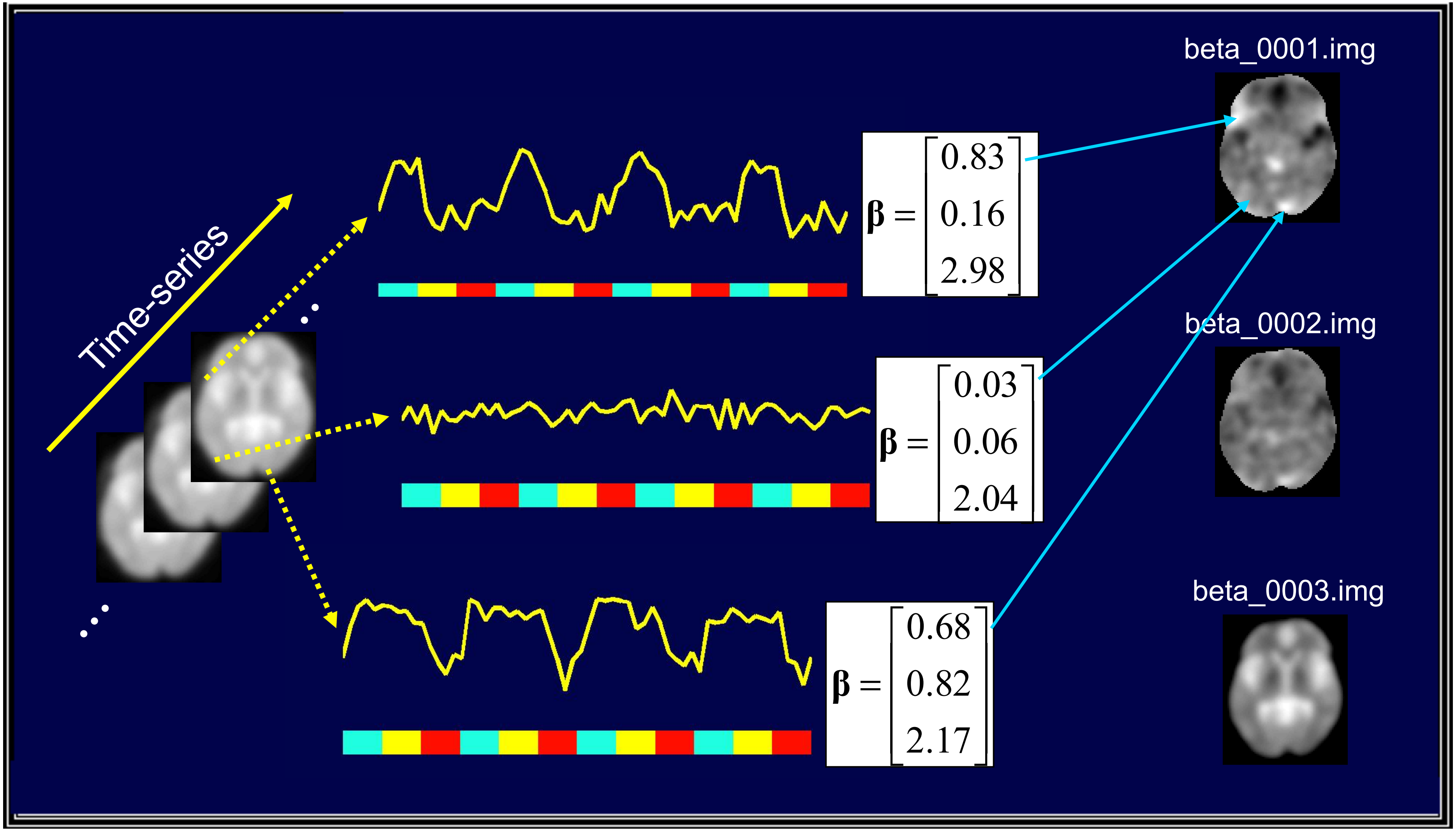
Different data (*y*)

The *same* predictors (*x*)



Different parameters ()

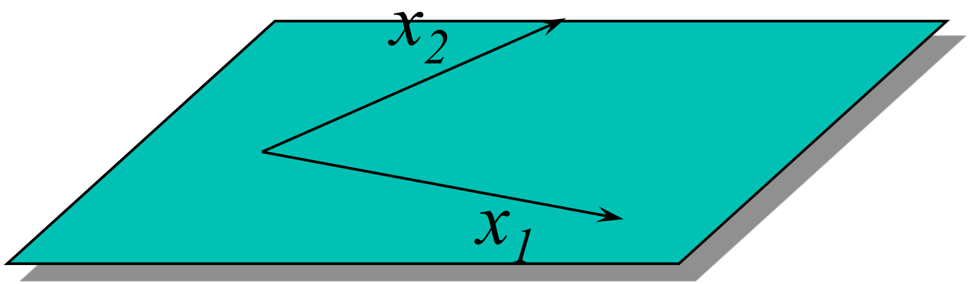
So the same model can be used across all voxels of the brain, just using different paramteres



Finding the optimal parameter can also be visualised in geometric space

- The $\hat{y}$ and x in our model are all **vectors of the same dimensionality** and so, lie within the same large space.

- The design matrix (x1, x2, x3…) defines a subspace; the design space (green panel).

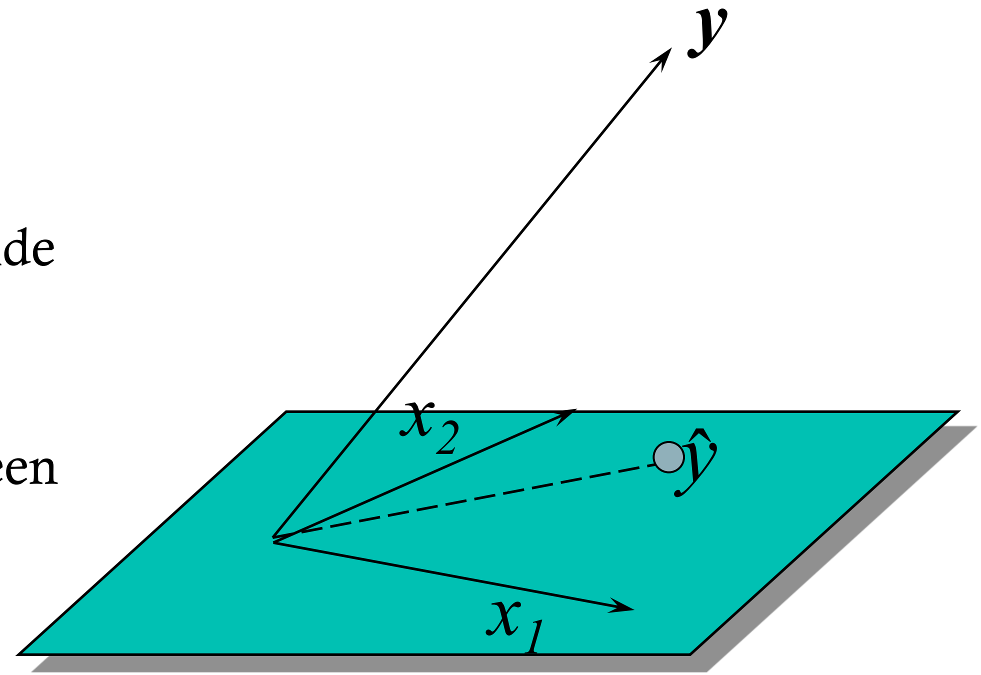


Design space defined by X

$$y = X\beta + e$$

Parameters determine the co-ordinates of the predicted response ($\hat{y}$) within this space

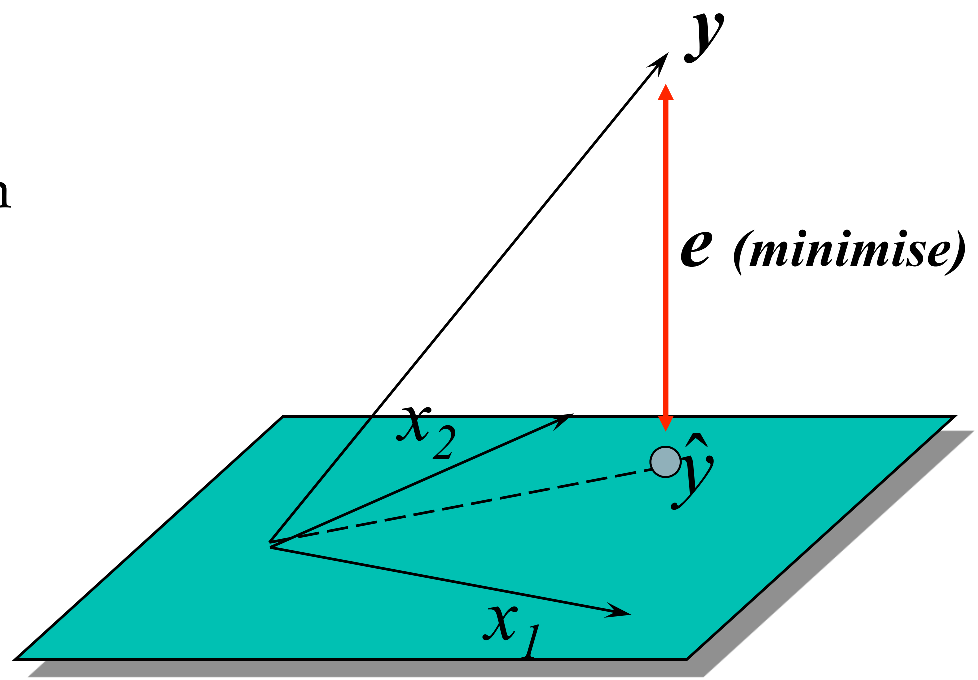
- The *actual* data (y) however, lie outside this space.
- So there is always a difference between the predicted $\hat{y}$ and actual y



We need to minimise the difference between predicted $\hat{y}$ and actual y

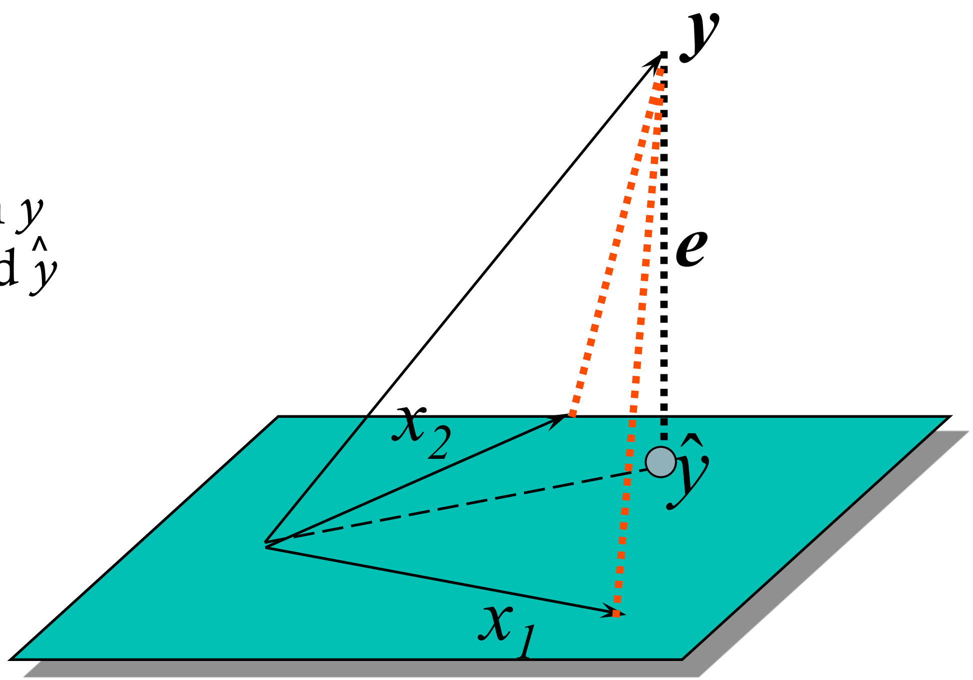
- So the GLM aims find the projection of y on the design space which minimises the error of the model

  (minimises the difference between predicted $\hat{y}$ and actual y)



The smallest error vector is orthogonal to the design space…

…So the best parameter will position y so that the error vector between y and $\hat{y}$ is orthogonal to the design space (minimising the error)



<u>How</u> do we find the parameter which produces minimal error?

The optimal parameter can be calculated using

<u>Ordinary Least Squares</u>

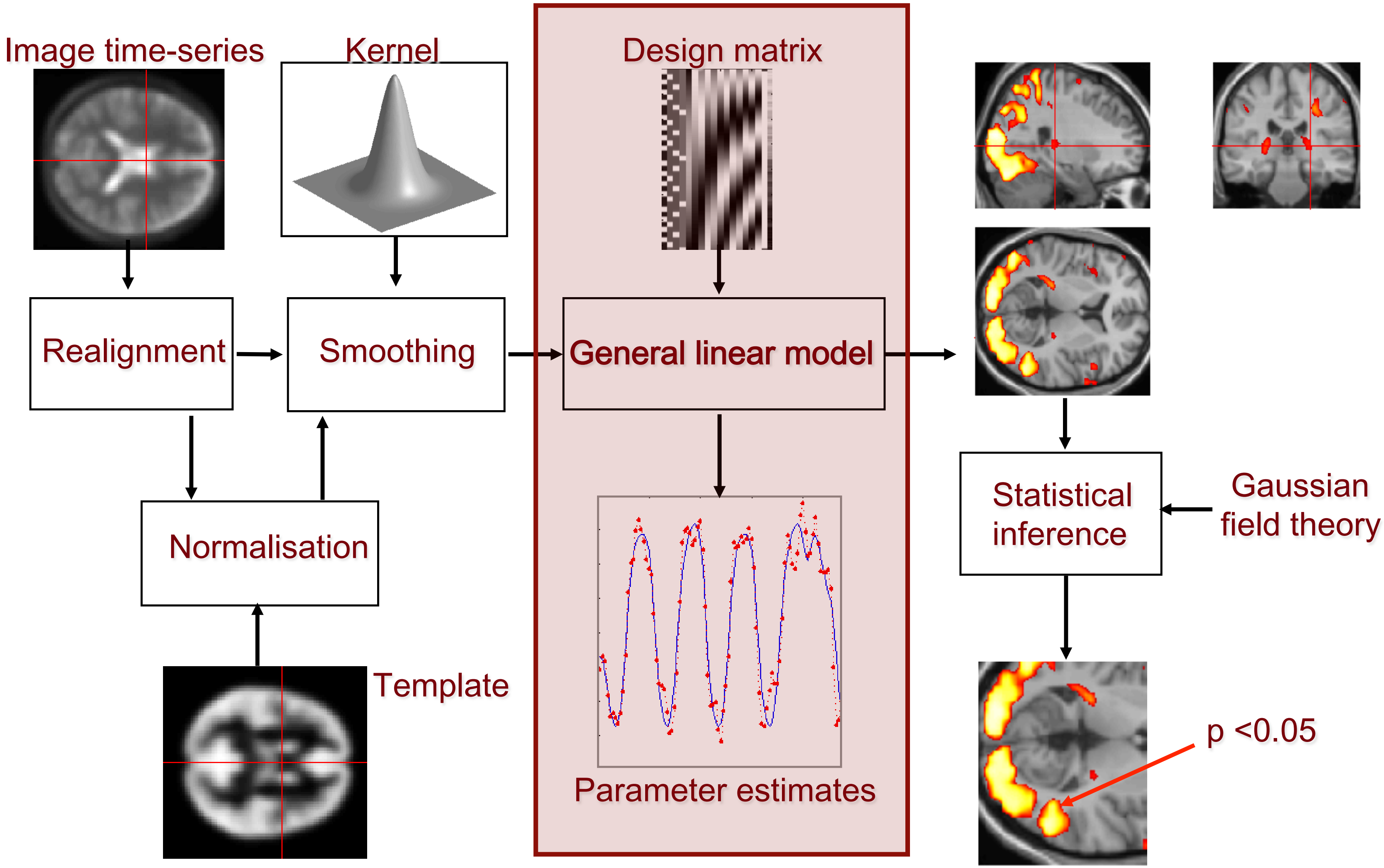
$$\hat{\beta} = (X^T X)^{-1} X^T y$$

A statistical method for **estimating unknown parameters from sampled data**

- minimizing the difference between predicted data and observed data.

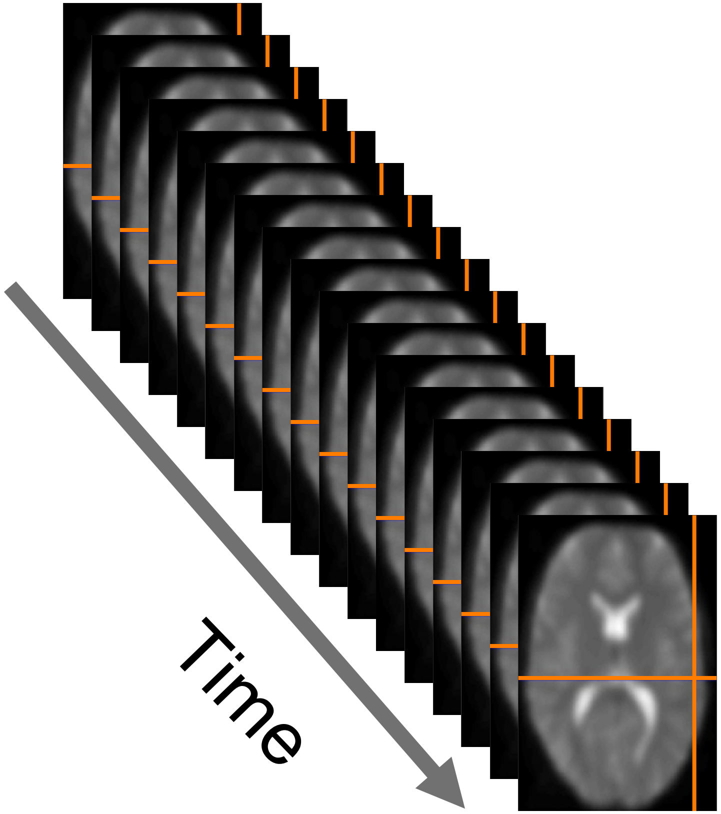
$$\sum_{t=1}^{N} e_t^2$$ **= minimum**

Overview of SPM

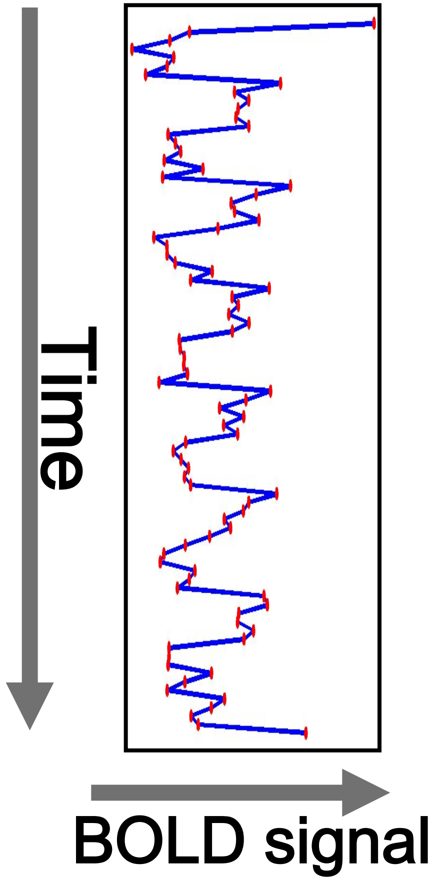


fMRI data

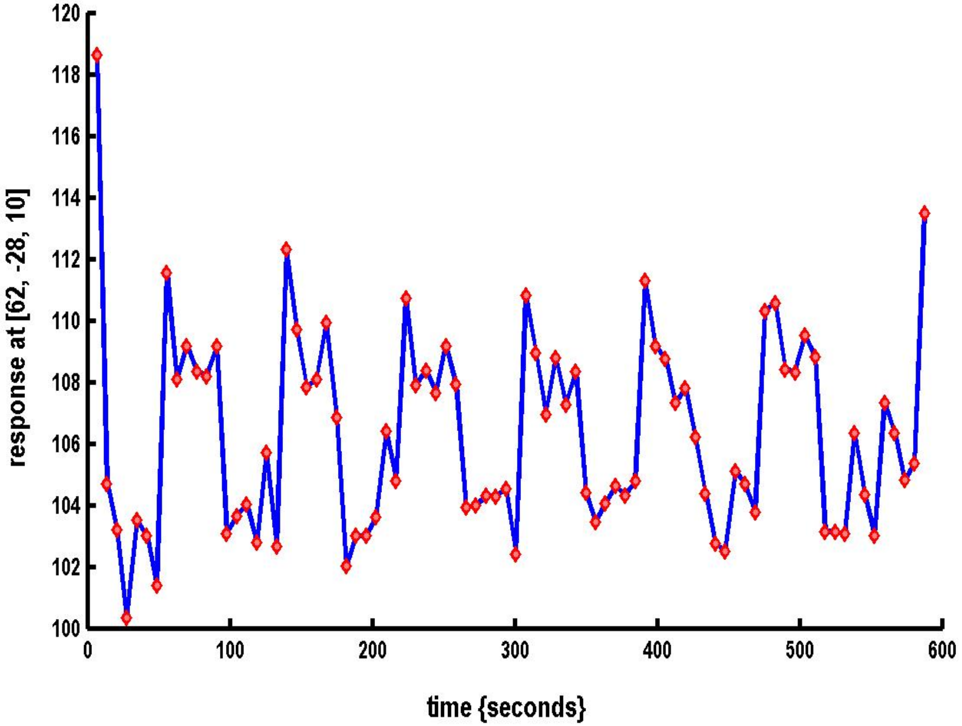
Single voxel analyzed across many different time points



Y = BOLD signal at each time point from that voxel



An fMRI experiment



Is there a change in the BOLD response between the two conditions?

Applying the GLM

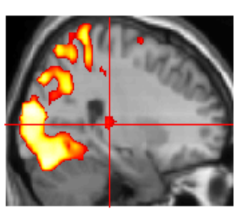
$$y = x_1\beta_1 + x_2\beta_2 + e$$

BOLD response at each time point at chosen voxel

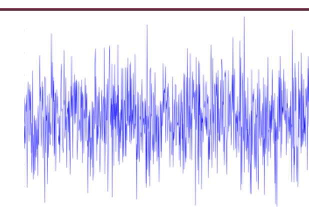
Predictors that explain the data

How much each predictor explains the data (Coefficient)

Variance in the data that cannot be explained by the predictors (noise)

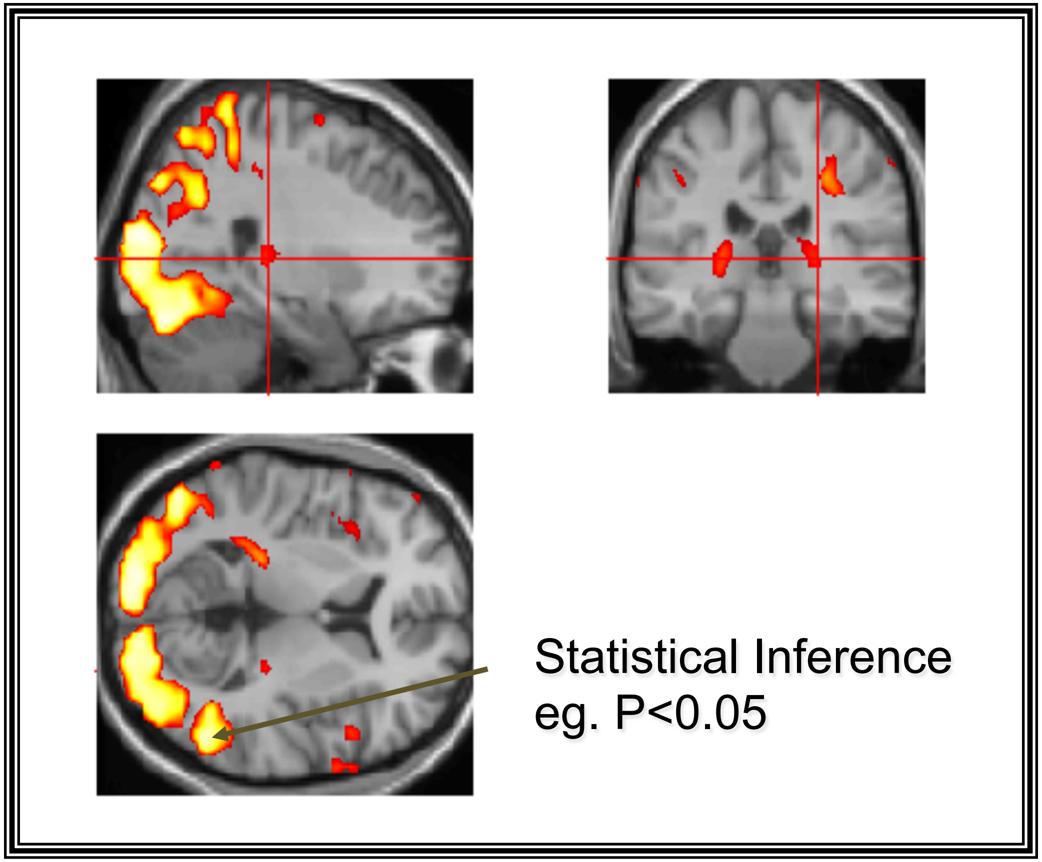


β = 0.44



Statistical Parametric Mapping

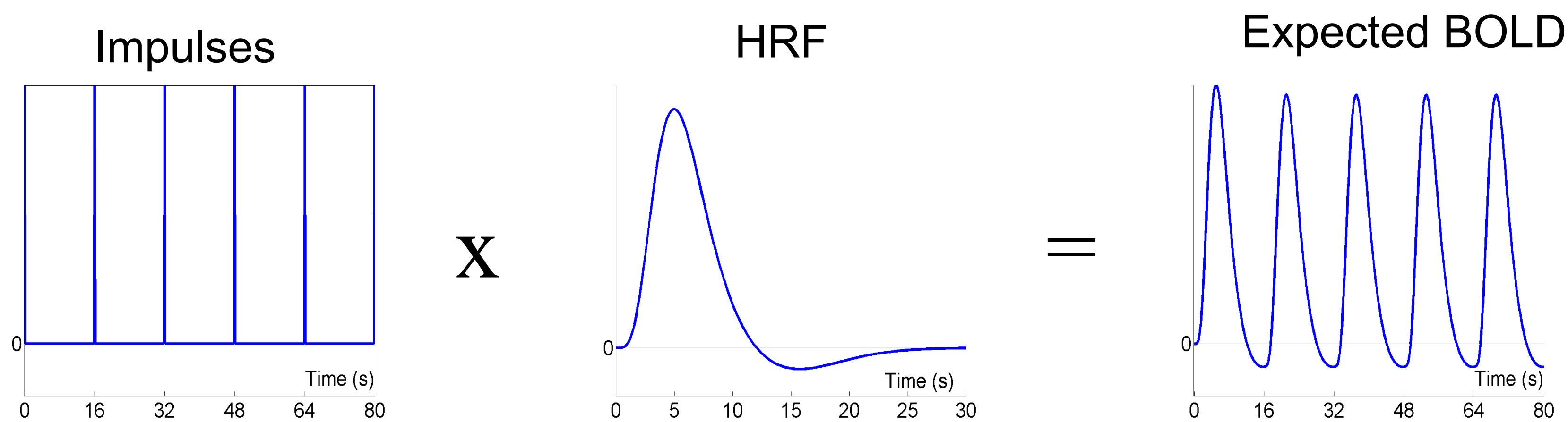
- Null hypothesis: your effect of interest explains none of your data.
- Is the task significally different?



Problems

1. BOLD signal is not a simple on/off
2. Low-frequency noise
3. Assumptions about the nature of the error
4. Physiological confounds

Convolution model

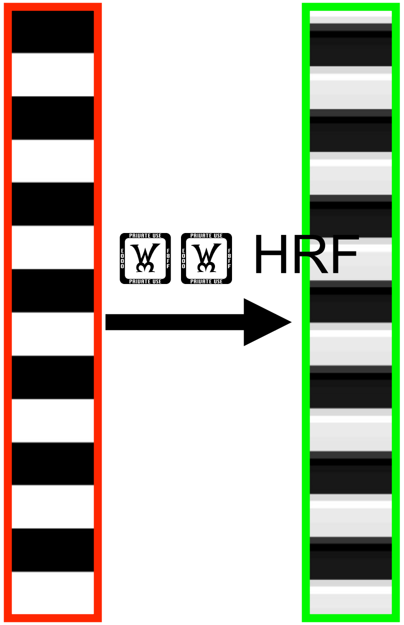


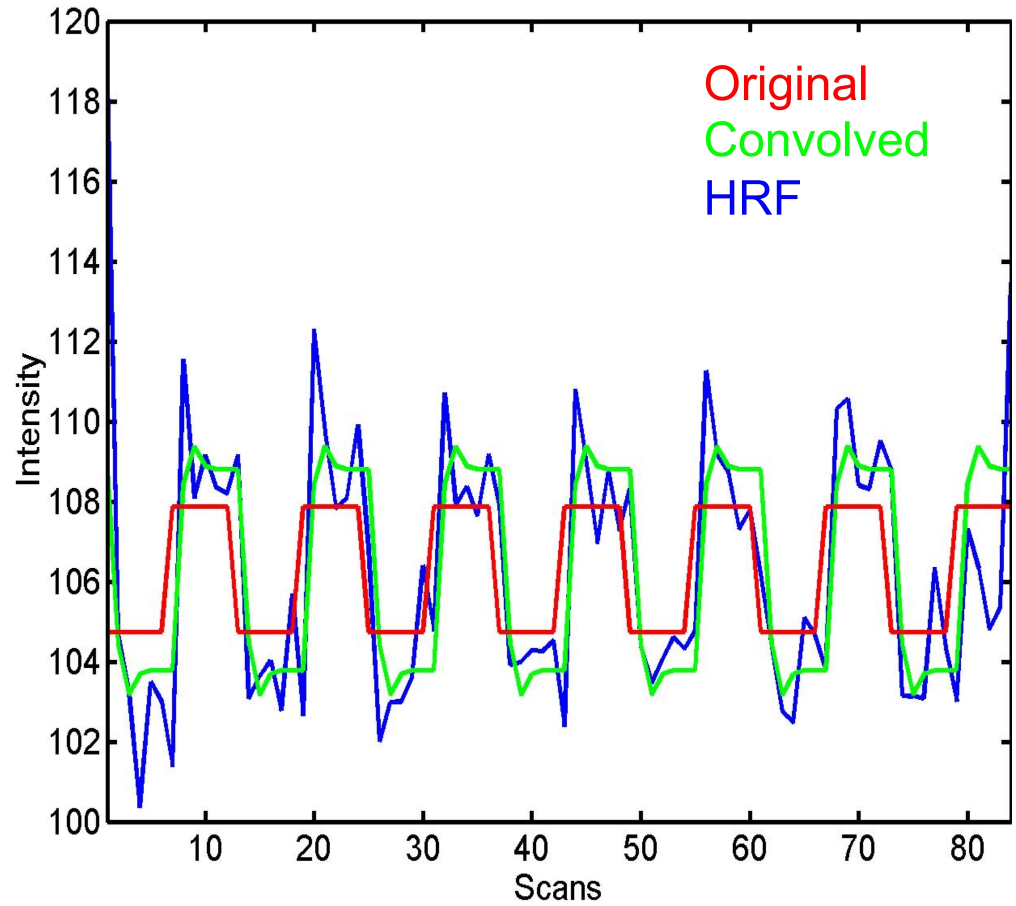
expected BOLD response

= input function x hemodynamic response function (HRF)

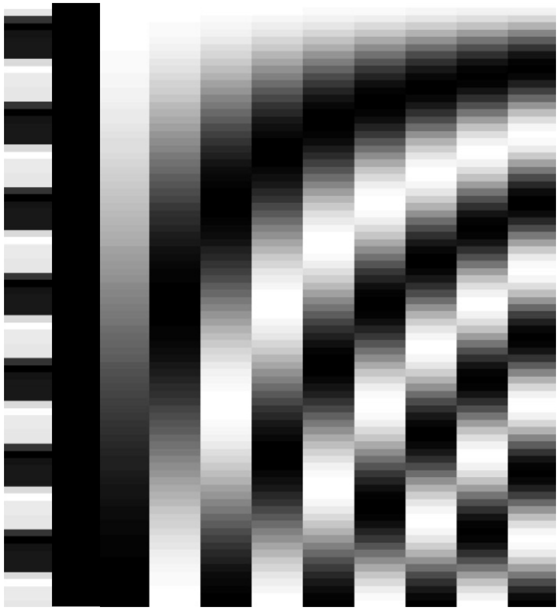
Convolution model

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



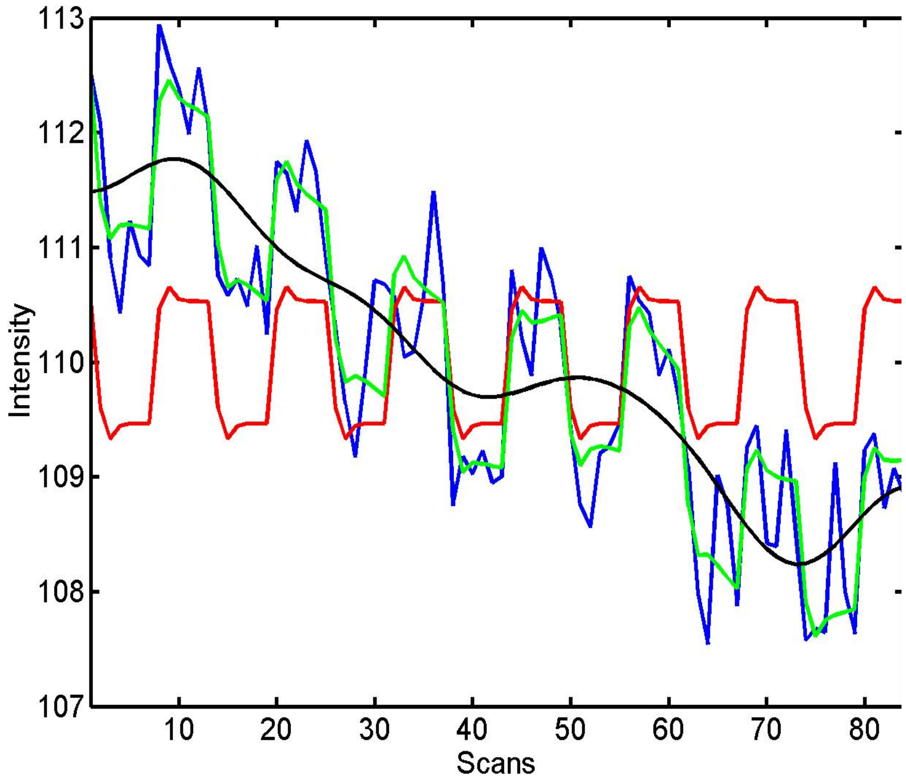


Adjusting for low frequencies



discrete cosine transform (DCT) set

Adjusting for low frequencies



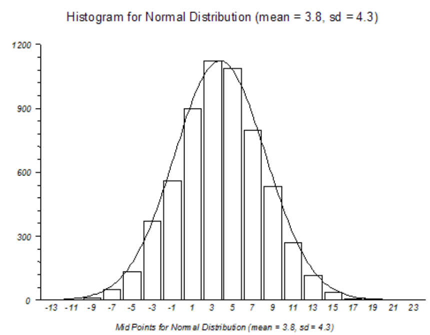
Blue = data

Black = mean + low-frequency drift

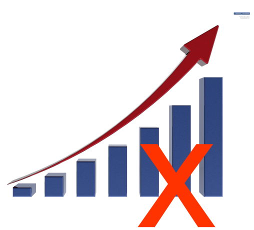
Green = predicted response, taking into account low-frequency drift

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

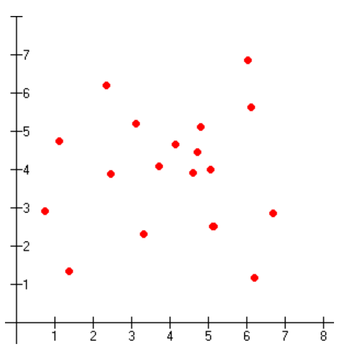
GLM Assumptions



1. Errors are **normally distributed - smoothing**



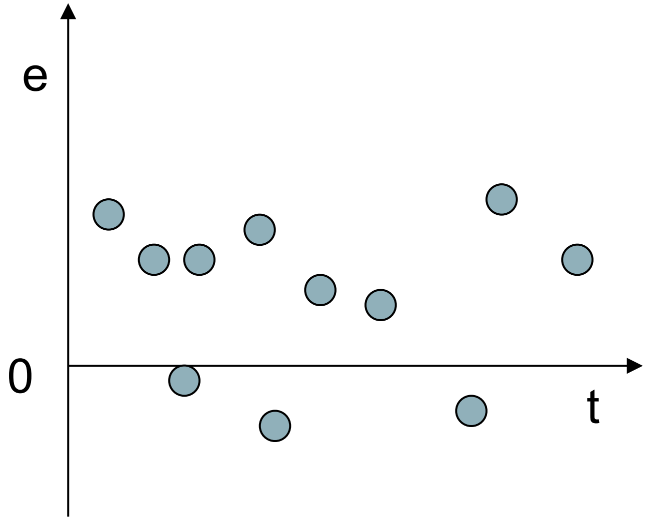
2. **Error is the same** in each & every measurement point



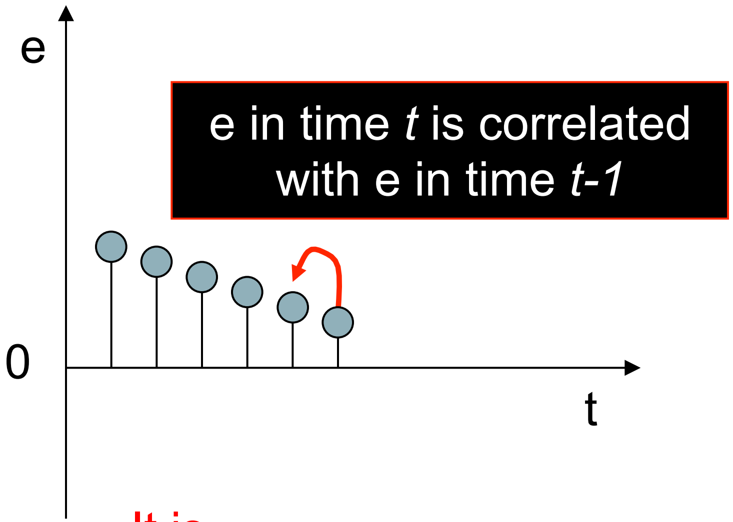
3. There is **no correlation between errors** at different time points/data points

Error is time-correlated

- The error at each time point is correlated to the error at the previous time point



It should be…



It is…

Autoregressive Model

- Temporal autocorrelation:
  in $y = X\beta + e$ over time
  $e_t = a e_{t-1} + \varepsilon$

- ‘Whitening’

- To compensate for inflated t-value

Physiological Confounds

- head movements
- arterial pulsations
- breathing
- eye blinks (visual cortex)
- adaptation affects, fatigue, changes in attention to task

To recap...

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

$$=$$

Response = (Predictors x Parameters...) + Error

Thanks to...

- Previous years MfD slides (2009-2010)
- Dr. Guillaume Flandin