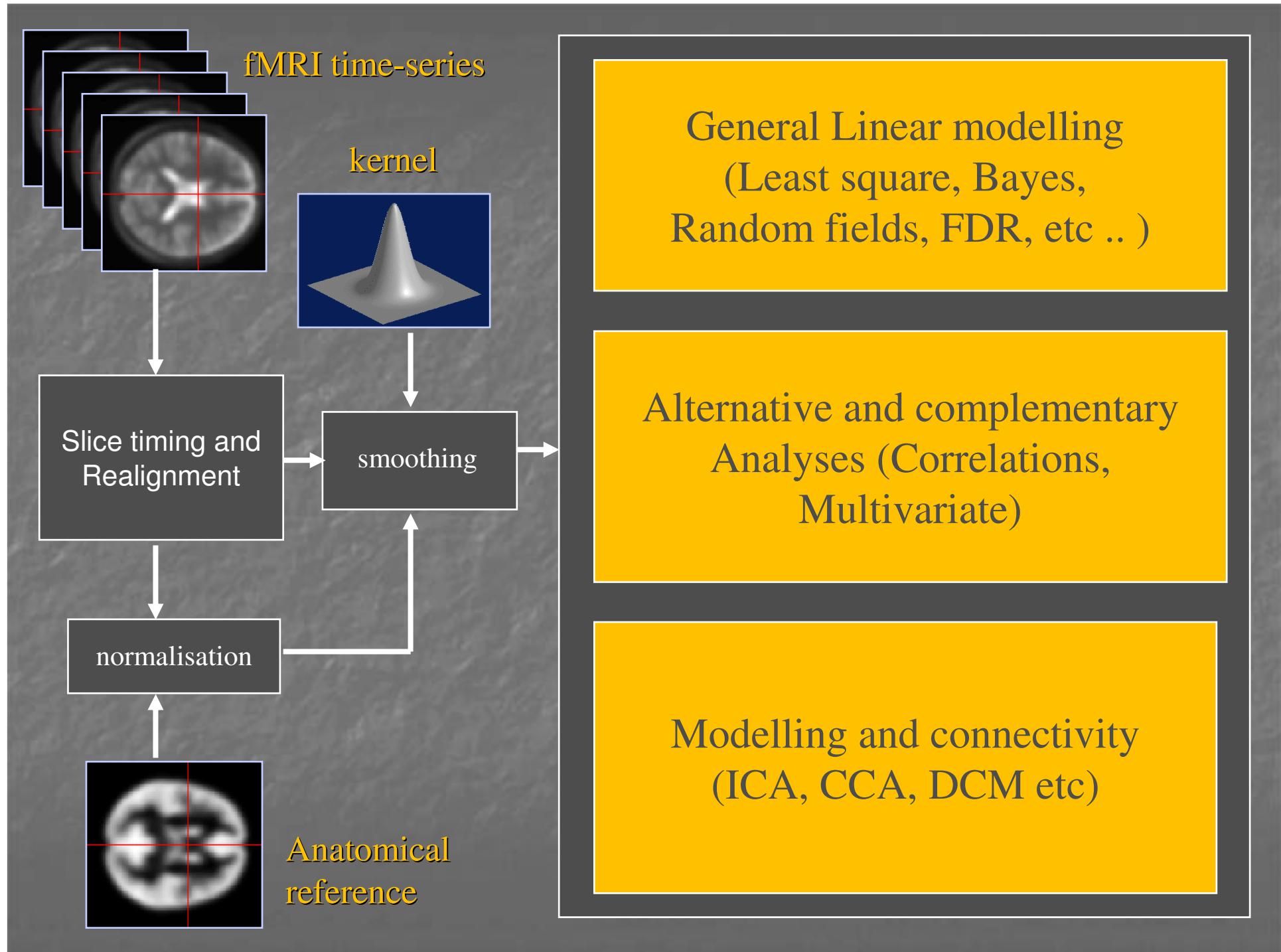


Functional MRI data analysis

Cyril Pernet, PhD



Overview

- Part I: General Linear Model

- What is linearity? / Why do we speak of models?

- A simple fMRI model / Contrasts

- Part II: Hierarchical modelling

- Definition / Fixed vs. Random Effects

- The summary statistics approach

- Part III: Statistical Inference

- Multiple comparisons correction procedures / Levels of inferences (set, cluster, voxel)

- Circularity issues

GLM

JB, Poline & M. Brett
The general linear model and fMRI: does love last forever?
Neuroimage, 15 (2012), p 871-880

GLM Overview

- What is linearity?
- Why do we speak of models?
- A simple fMRI model
- Contrasts

Linearity

- Means created by lines
- In maths it refers to equations or functions that satisfy 2 properties: additivity (also called superposition) and homogeneity of degree 1 (also called scaling)
- Additivity → $y = x_1 + x_2$ (output is sum of inputs)
- Scaling → $y = \beta x_1$ (output is proportional to input)

Examples

- $X = \text{randn}(10,1)$

- Linear correlation

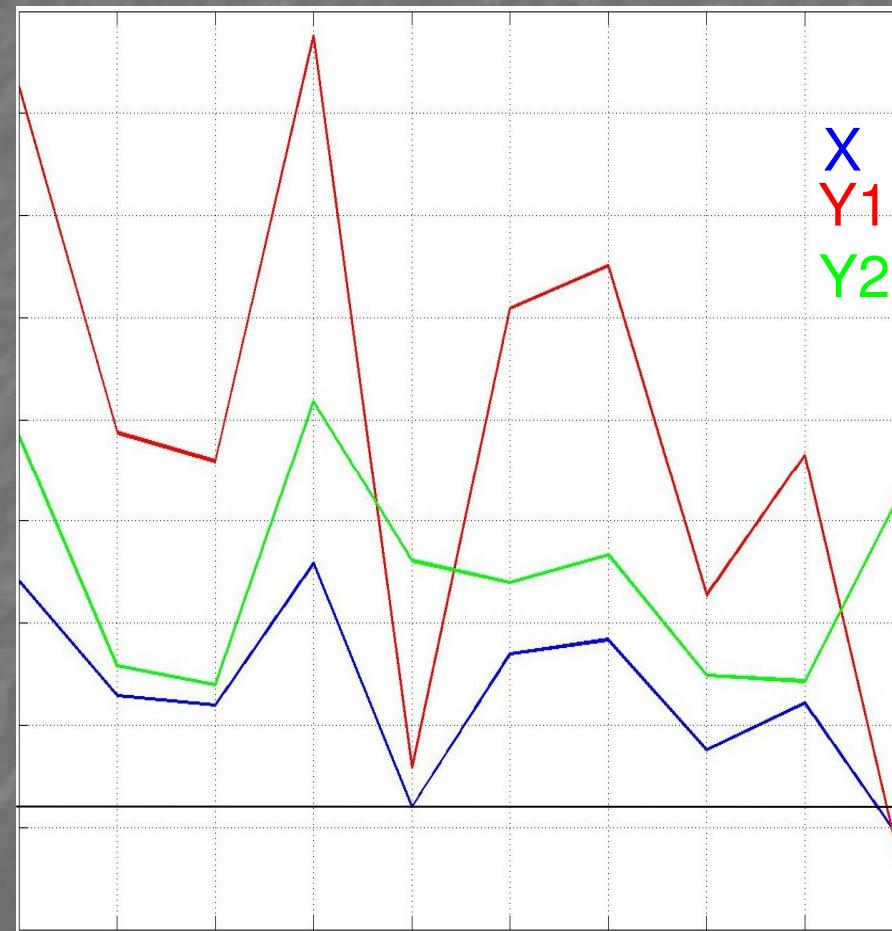
$$Y_1 = 3x + 2$$

Pearson $r = 1$

- Non linear correlation

$$Y_2 = \text{abs}(2x)$$

Pearson $r = 0.38$

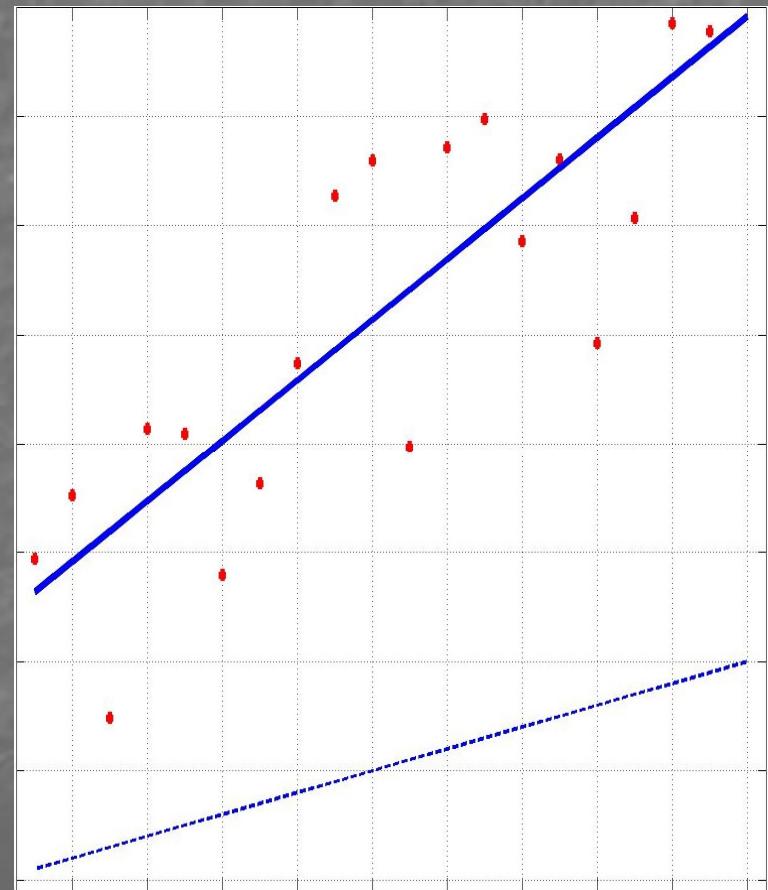


What is a linear model?

- An equation or a set of equations that models data and which corresponds geometrically to straight lines, plans, hyperplans and satisfy the properties of additivity and scaling.
- Simple regression: $y = \beta_1x_1 + \beta_2 + \varepsilon$
- Multiple regression: $y = \beta_1x_1 + \beta_2x_2 + \beta_3 + \varepsilon$
- One way ANOVA: $y = \mu + \alpha_i + \varepsilon$
- Repeated measure ANOVA: $y = u + S_i + \alpha_i + \varepsilon$
- ...

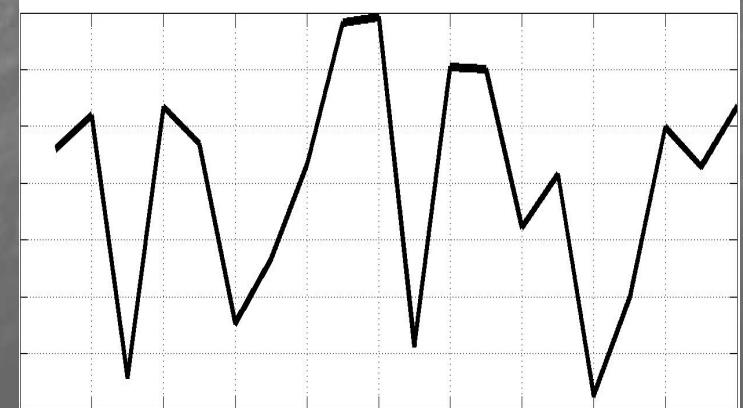
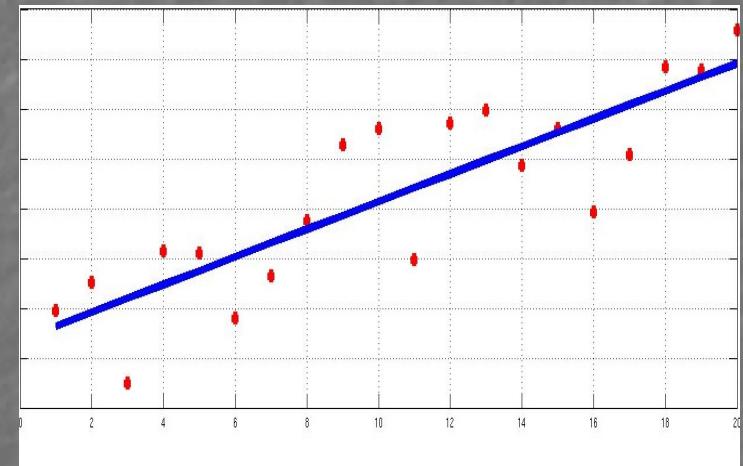
A regression is a linear model

- We have an experimental measure x (e.g. stimulus intensity from 0 to 20)
- We then do the expe and collect data y (e.g. RTs)
- Model: $y = \beta_1 x + \beta_2$
- Do some maths / run a software to find β_1 and β_2 $\hat{y} = 2.7x + 23.6$



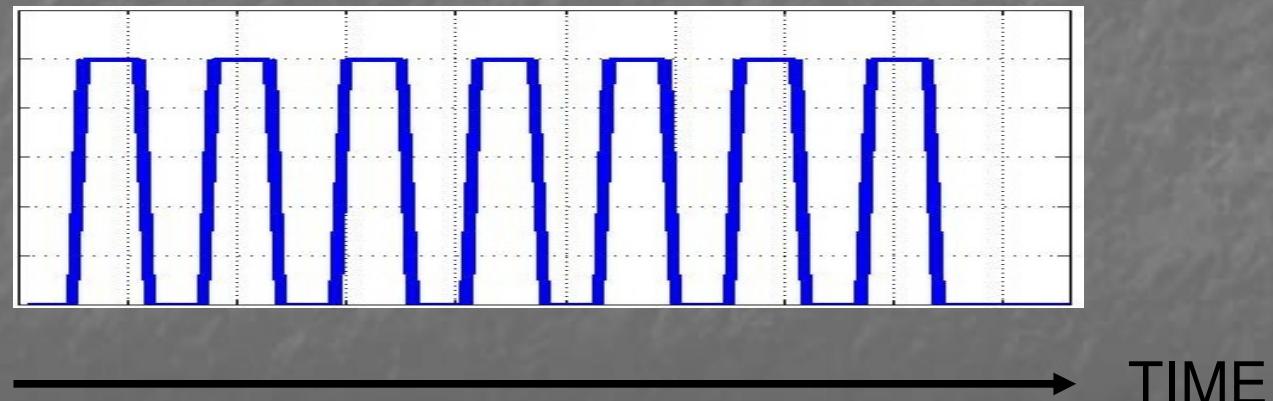
A regression is a linear model

- The error is the distance between the data and the model
- $F = (\text{SSeffect} / \text{df}) / (\text{SSerror} / \text{df_error})$
- SSeffect = `norm(model - mean(model)).^2;`
- SSerror = `norm(residuals).^2;`



FMRI experiment

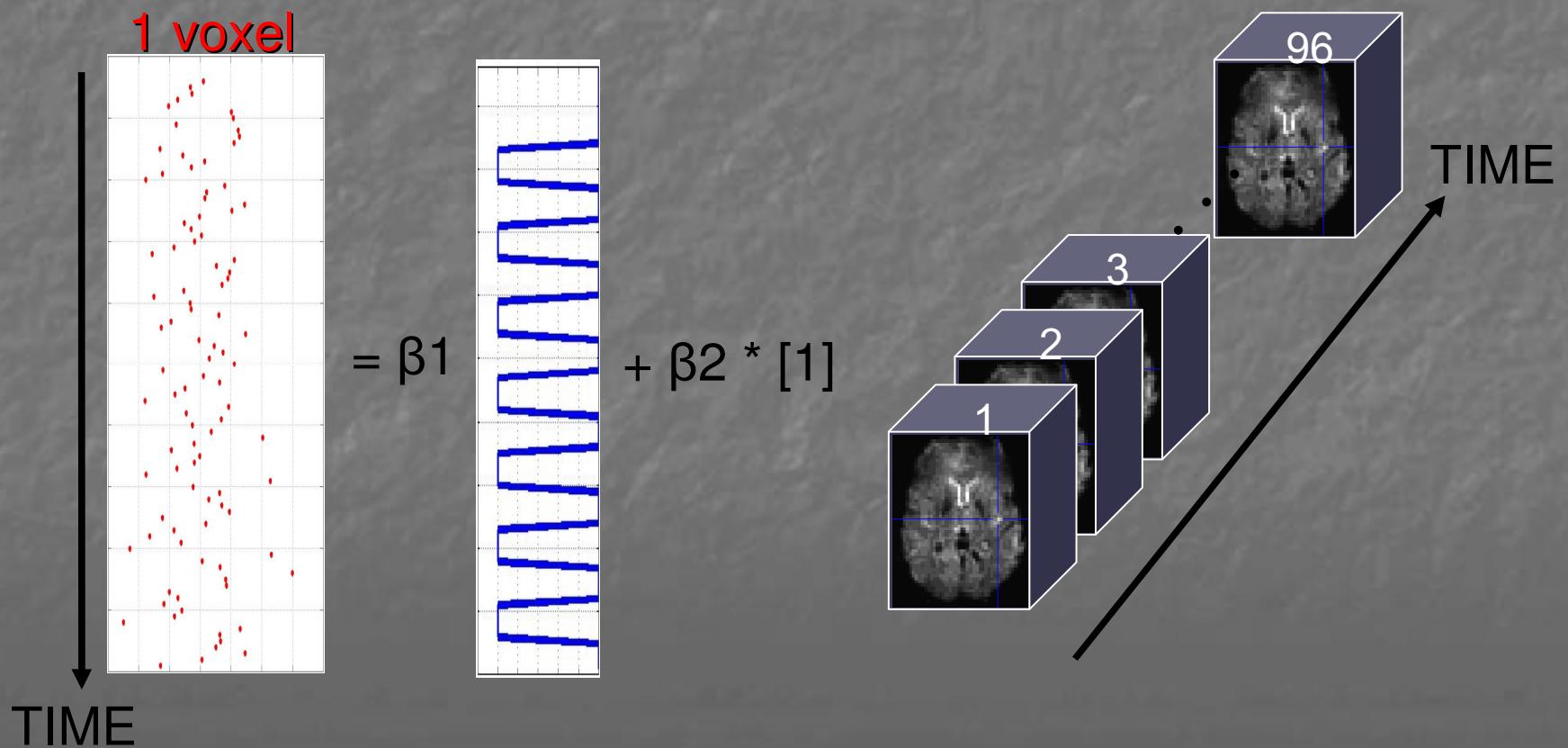
- *SPM data set*: which areas are activated by the presentation of bi-syllabic words presented binaurally (60 per minute)
- Experimental measure **x**: 7 blocks of 42 sec of stimulation



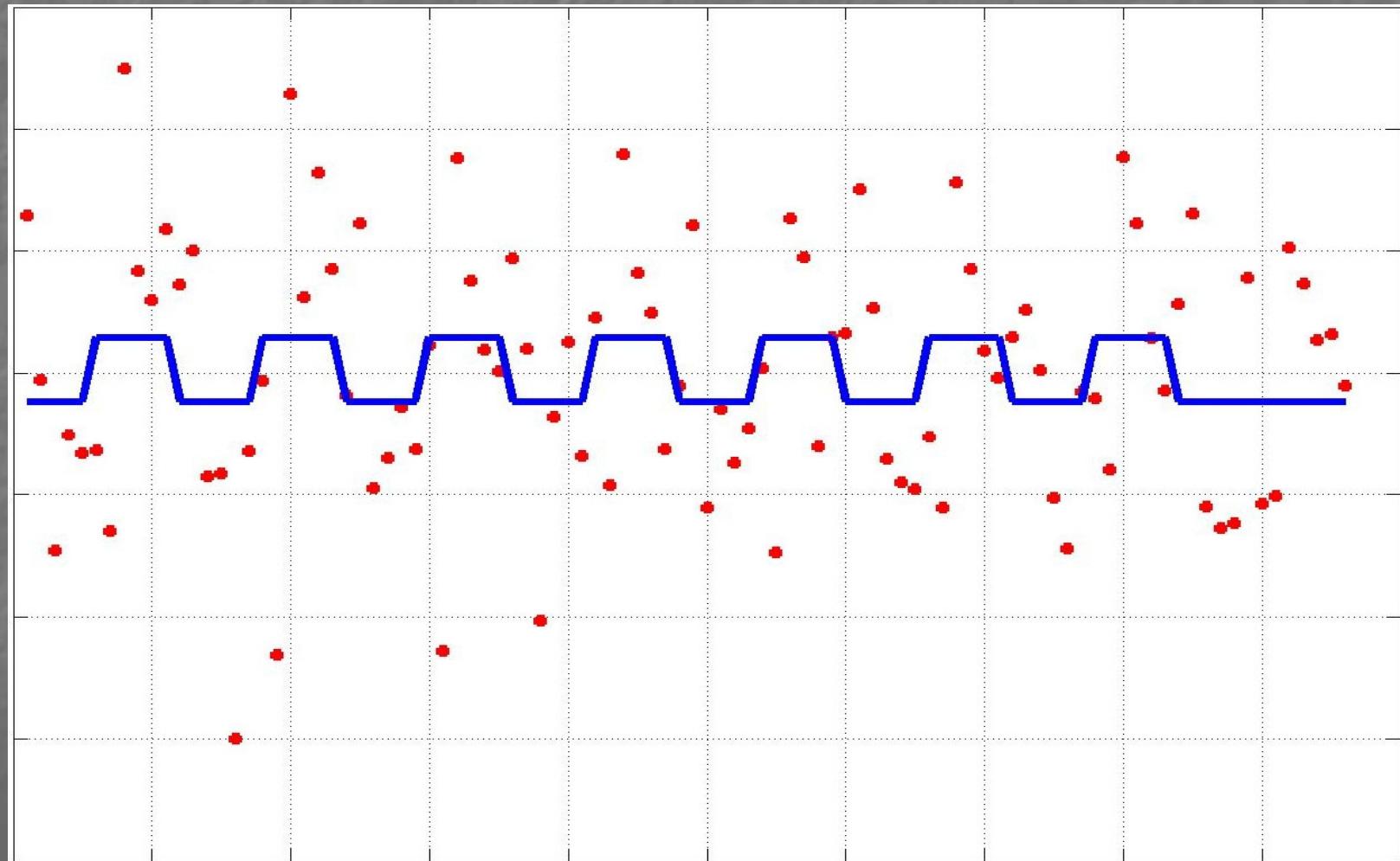
<http://www.fil.ion.ucl.ac.uk/spm/data/auditory/>

FMRI experiment

- Collect the data : 96 fMRI volumes (RT=7s)
- Model: $y = \beta_1 x + \beta_2$

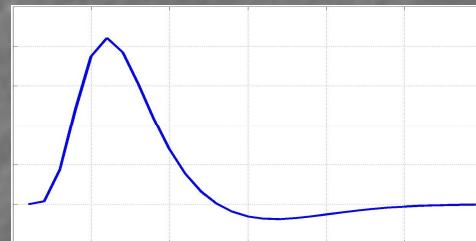


FMRI experiment

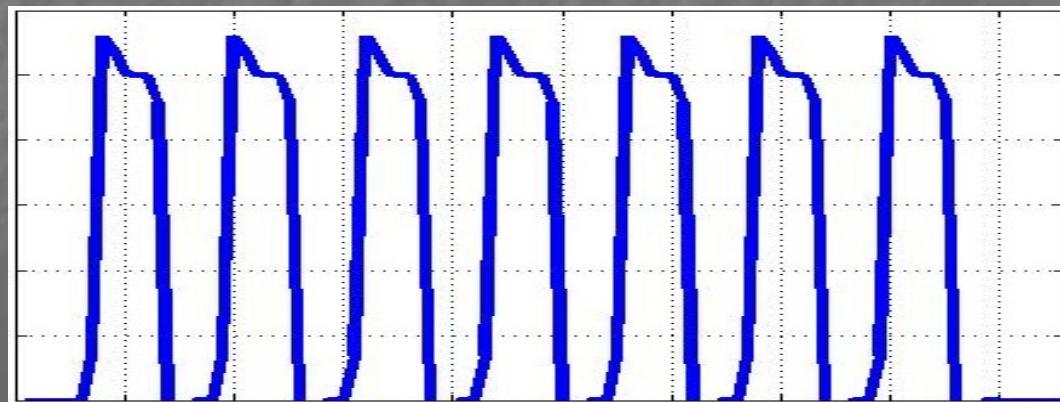


FMRI experiment

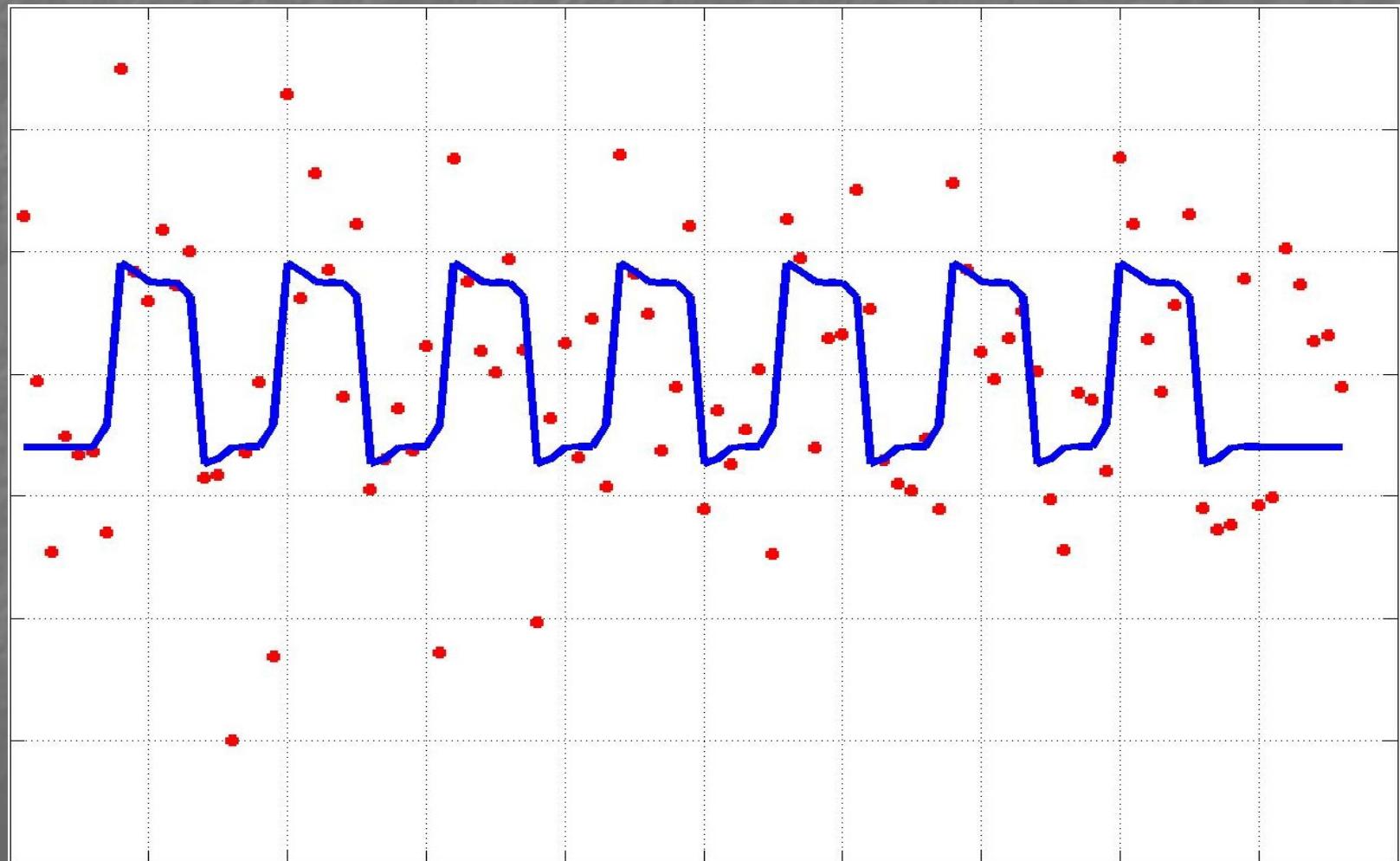
- A better model: we know the shape of the BOLD response



- Convolution by the hrf: $x \otimes \text{hrf}$



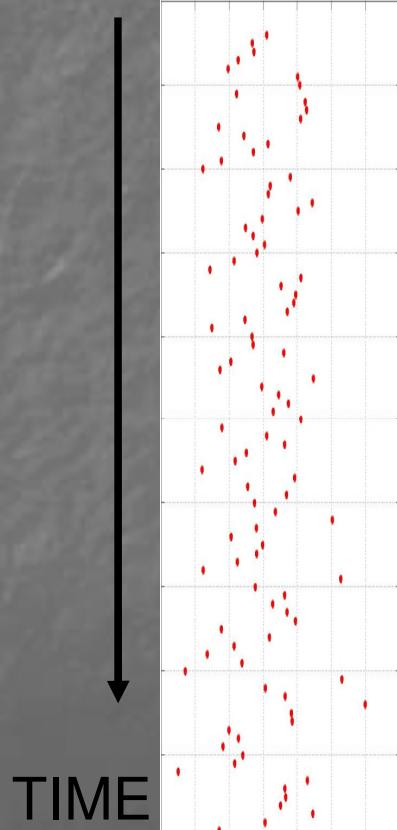
FMRI experiment



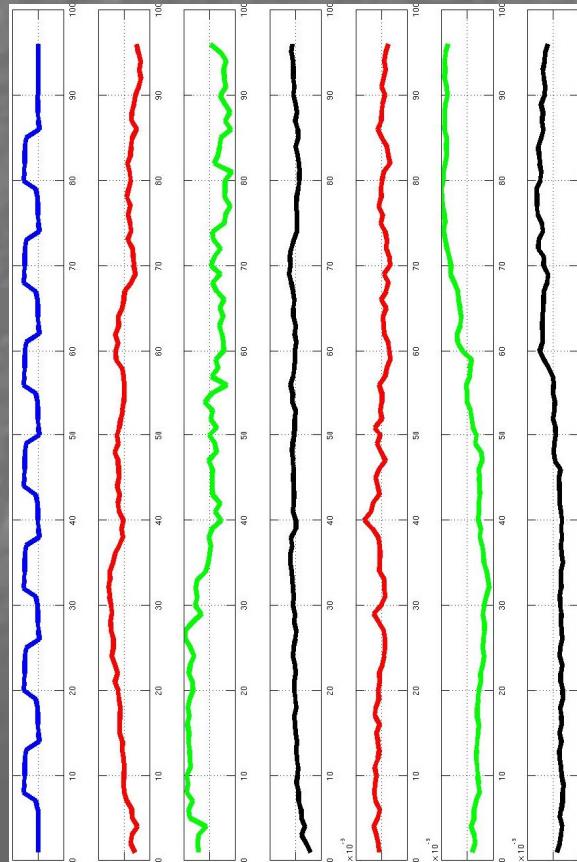
FMRI experiment

- An even better model: add motion parameters

1 voxel

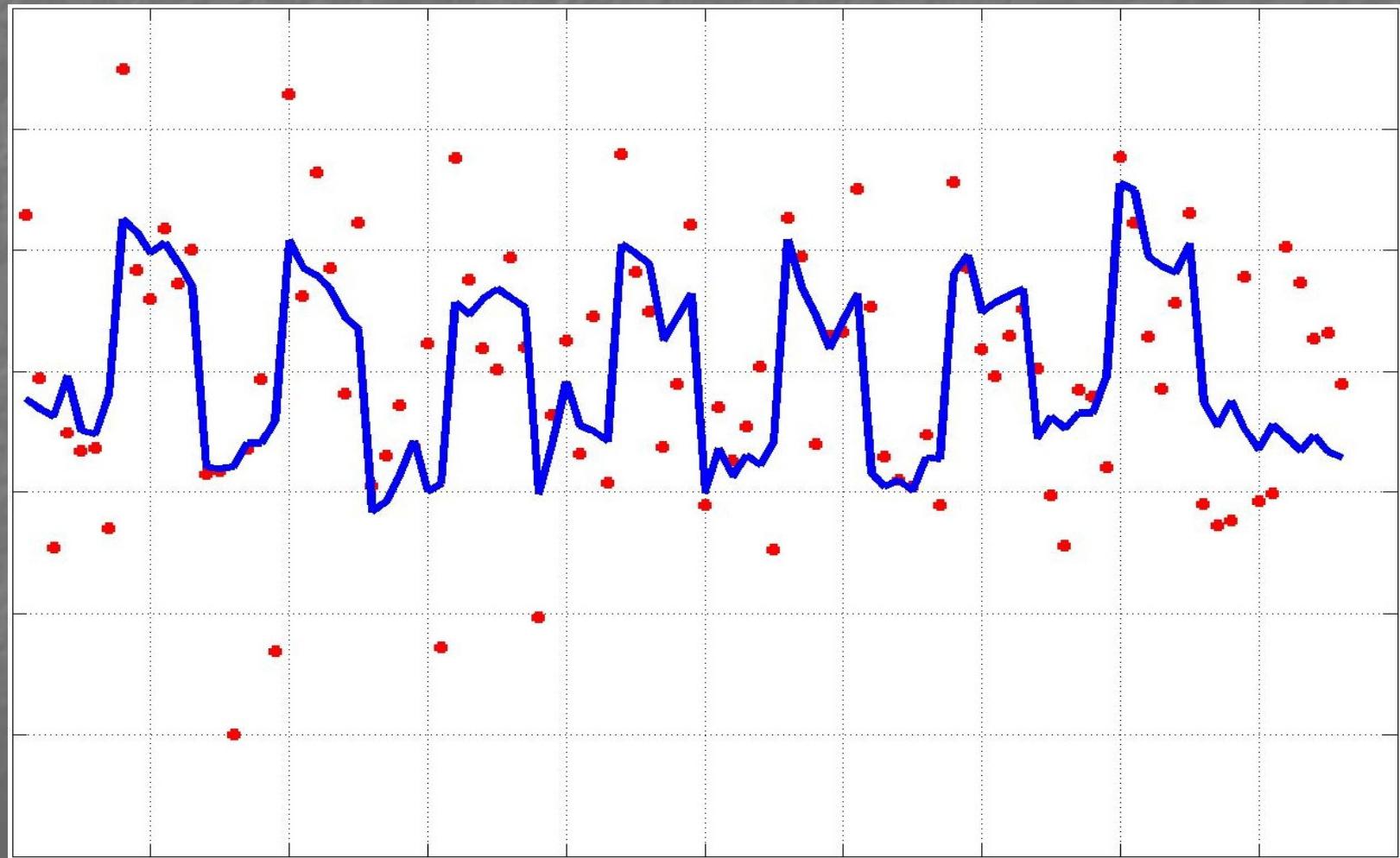


=



$$* [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5 \ \beta_6 \ \beta_7 \ \beta_8] + \beta_9 * [1]$$

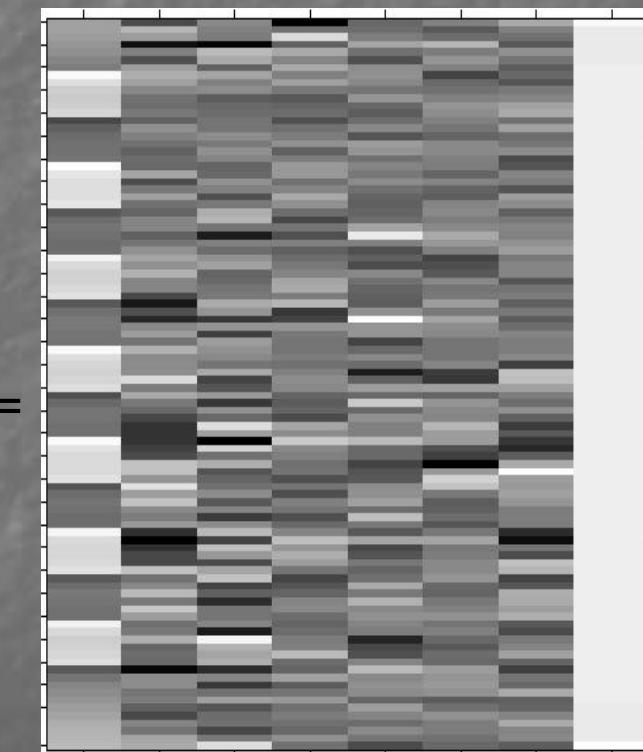
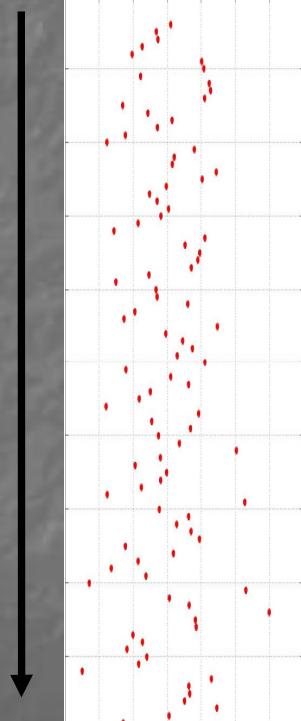
FMRI experiment



FMRI experiment

- Matrix formulation and SPM colour coding

1 voxel



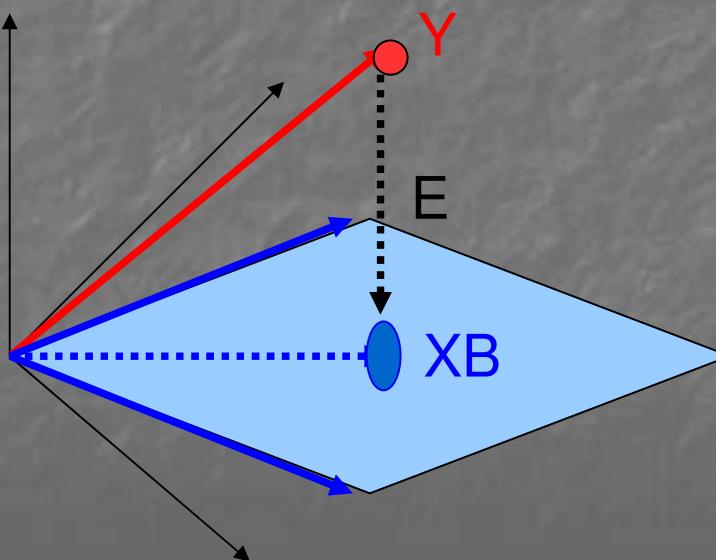
* [$\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5 \ \beta_6 \ \beta_7 \ \beta_8$]

BetaXXXX.hdr
BetaXXXX.img

$$\text{FMRI data (Y)} = \text{Design matrix (X = SPM.mat)} * \mathbf{B} + \mathbf{E} \text{ (ResMS)}$$

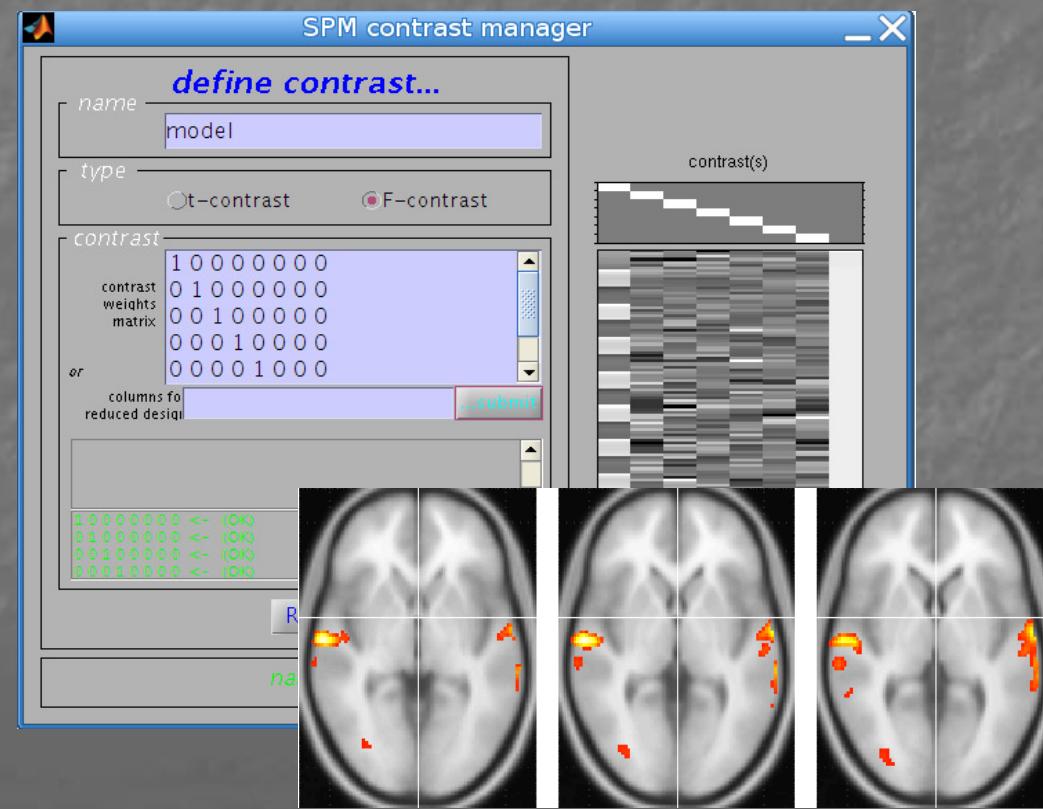
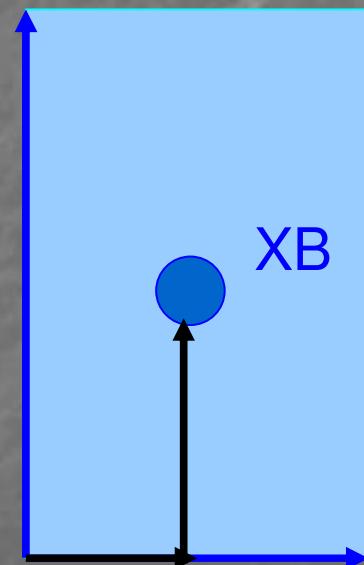
Model = \mathbb{R}^2

- Geometrical perspective
- \mathbf{Y} = 3 observations \mathbf{X} = 2 regressors
- $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$



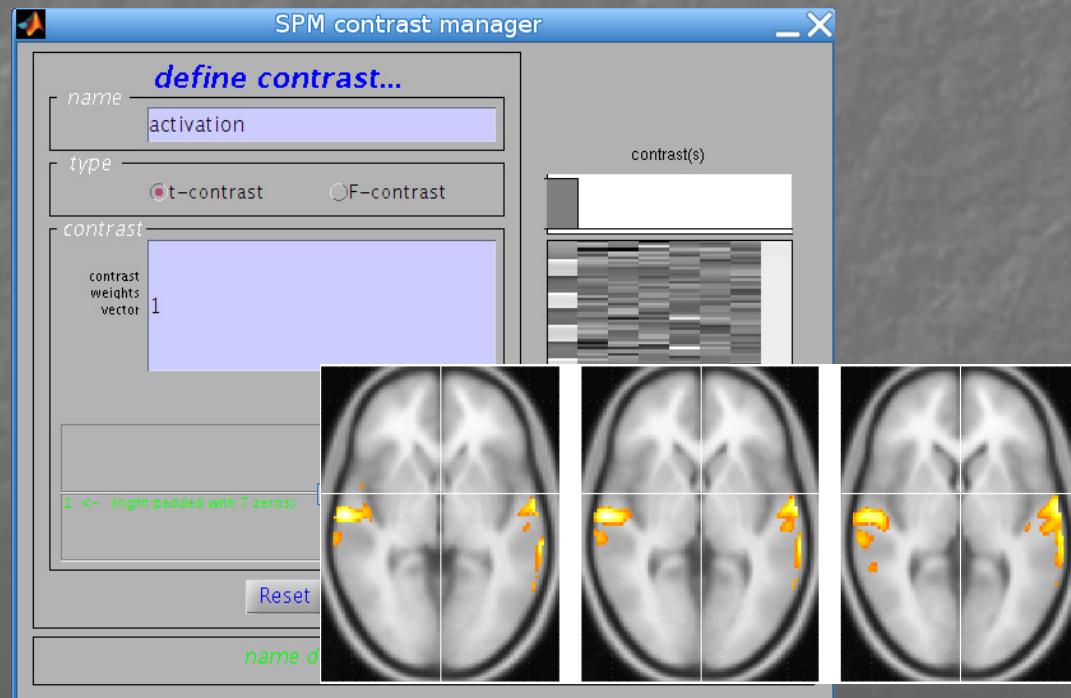
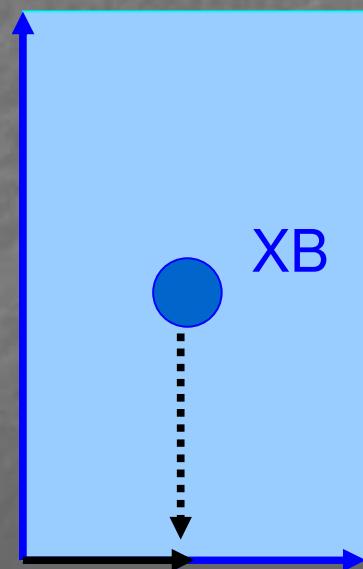
Model = R²

- Where does the model fit the data?
- F test for all regressors: $y = 1/2x_1 + 1/2x_2 + \varepsilon$



Contrast = effect to test

- Where does the regressor for activation only explain the data (given the model)
- $y = 1/2x_1 + \epsilon$ (the orientation of x_1 and value of β_1 are fixed by the model)



Summary

- Linear model: $y = \beta_1x_1 + \beta_2x_2$ (output = additivity and scaling of input)
- GLM: $Y = XB + E$ (matrix formulation, works for any statistics, express the data Y as a function of the design matrix X)
- Contrasts: F or t test for the effect of 1 or several regressors given the design matrix

Part II: Hierarchical modelling

W, Penny & K.J. Friston
Human Brain Function

HLM Overview

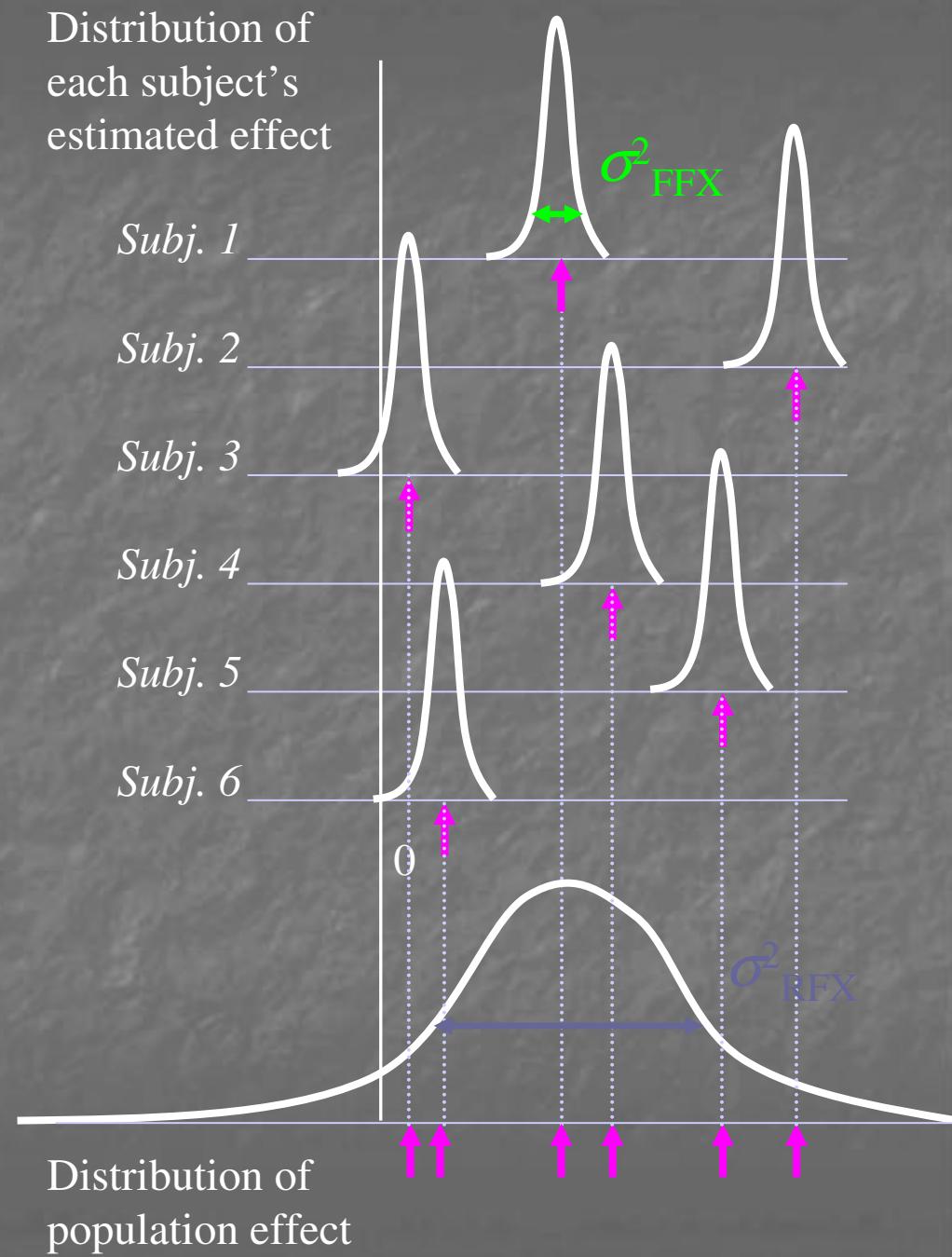
- Definition
- Fixed vs. Random effects
- Summary Statistics approach

Definition

- HLM allow variance in outcome variables (beta parameters) to be analysed at multiple hierarchical levels, whereas in simple linear and multiple linear regression all effects are modelled to occur at a single level. Thus, HLM are appropriate for use with nested data (subjects nested in a group).
- Multilevel models (also HLM, nested models, mixed models, random coefficient, random-effects models, random parameter models) are statistical models of parameters that vary at more than one level.

Fixed vs. Random Effects in fMRI

- Fixed Effects
 - Intra-subject variation suggests *all these subjects* different from zero
- Random Effects
 - Intersubject variation suggests *population* not very different from zero

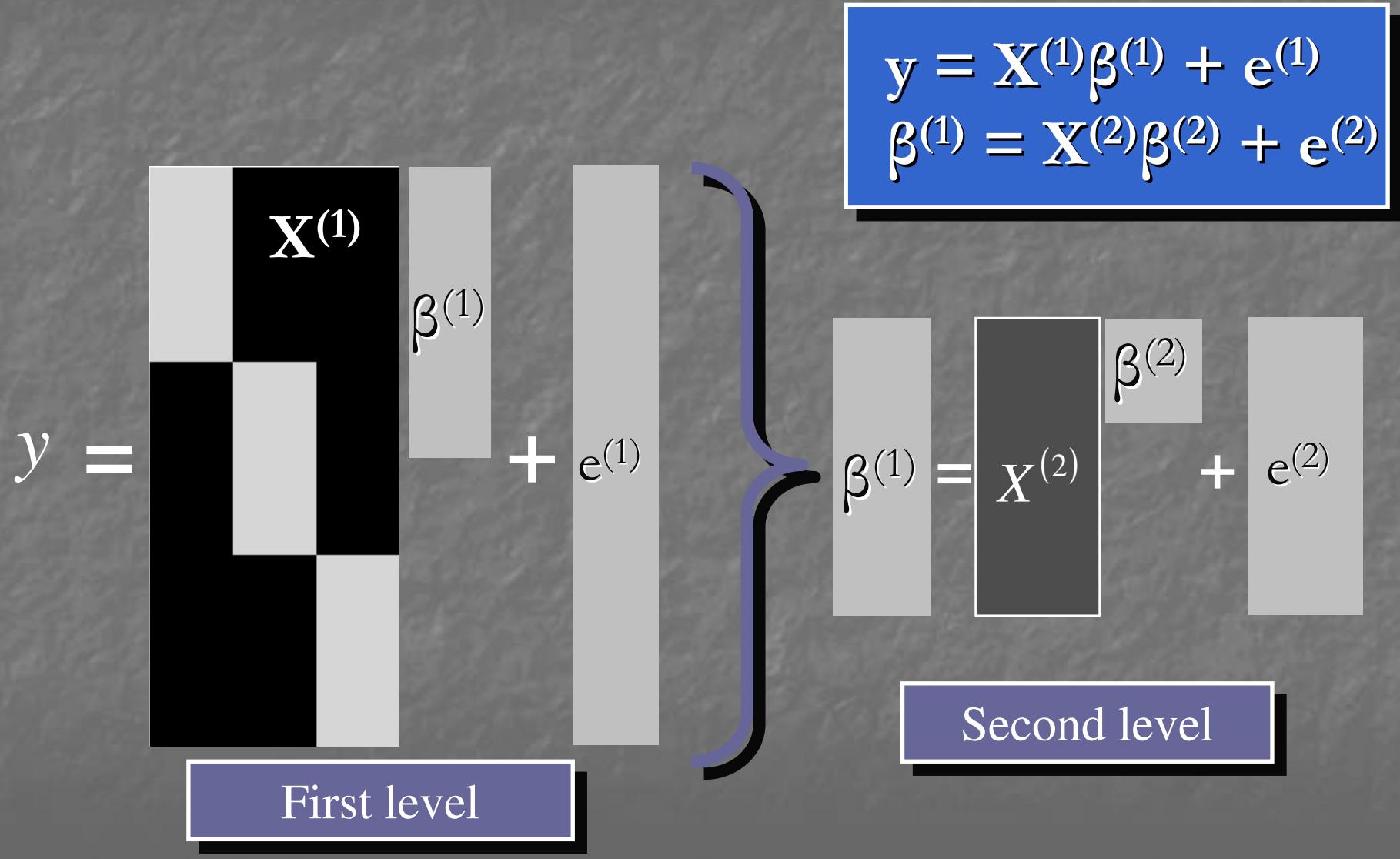


Some RFX Models in fMRI

- Holmes & Friston (HF)
 - Summary Statistic approach (contrasts only)
 - Holmes & Friston (HBM 1998). Generalisability, Random Effects & Population Inference. NI, 7(4 (2/3)):S754, 1999.
- Friston *et al.* (SPM2)
 - Empirical Bayesian approach
 - Friston et al. Classical and Bayesian inference in neuroimaging: theory. NI 16(2):465-483, 2002
 - Friston et al. Classical and Bayesian inference in neuroimaging: variance component estimation in fMRI. NI: 16(2):484-512, 2002.
- Beckmann et al. & Woolrich et al. (FSL)
 - Summary Statistics (contrast estimates *and* variance)
 - Beckmann, Jenkinson & Smith. General Multilevel linear modeling for group analysis in fMRI. NI 20(2):1052-1063 (2003)
 - Woolrich, Behrens et al. Multilevel linear modeling for fMRI group analysis using Bayesian inference. NI 21:1732-1747 (2004)

Summary Statistics approach

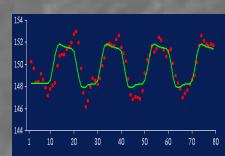
Summary Statistics approach



Summary Statistics approach

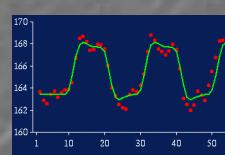
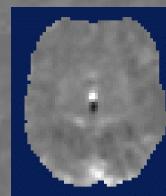
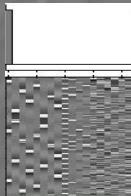
First level: model each subject

fMRI Data Design Matrix Contrast Img



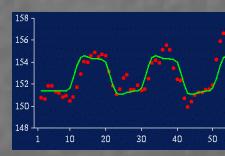
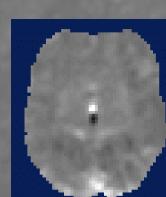
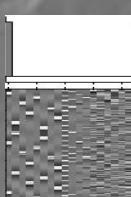
$$\hat{\alpha}_1$$

$$\hat{\sigma}_1^2$$



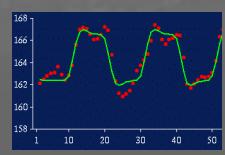
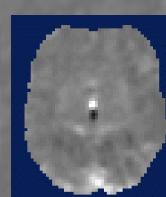
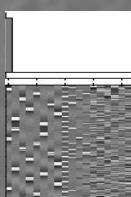
$$\hat{\alpha}_2$$

$$\hat{\sigma}_2^2$$



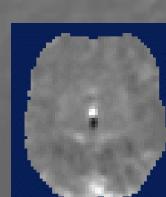
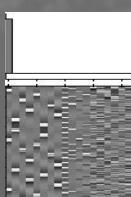
$$\hat{\alpha}_{11}$$

$$\hat{\sigma}_{11}^2$$

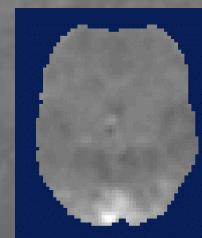
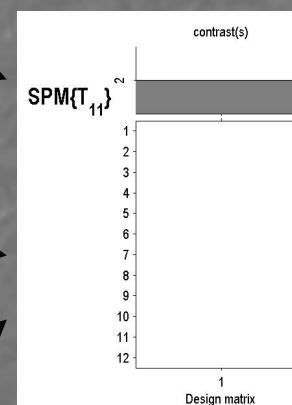


$$\hat{\alpha}_{12}$$

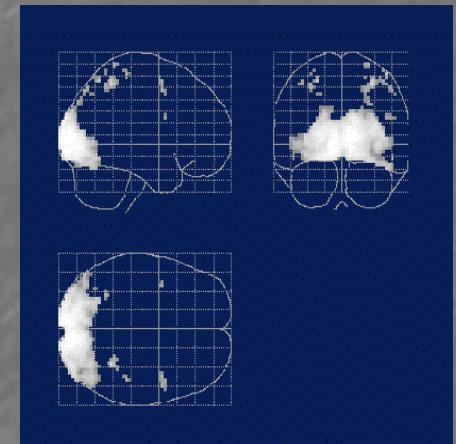
$$\hat{\sigma}_{12}^2$$



Second level: random effect



SPM(t)



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

Mixed model vs. Summary statistics

Mixed model

$$Y_{(ij)} = X\beta + Zv + e$$

The mean of Y is $X\beta$ and β is the fixed (experimental) effect
 Zv is the subject effect v is the random effect with mean = 0
 $\text{Var}(w_{\text{pop}})$ = between + within subjects

Summary Stat approach

$$Y_{(i)} = X\beta + e \text{ and } \text{Var}(e) = \sigma_w^2 / n$$

$$E(e) = 0 \text{ thus } \text{mean}(y_{(j)}) = w_{\text{pop}} + z_i + e_i$$

GLM assumptions:
Errors are iid centred on 0

Validity of the SS approach

- The method used in mixed effects analysis is Expectation Maximisation: estimates population mean effect as MEAN_{EM} and the variance of this estimate as VAR_{EM}
- For N subjects, n scans per subject and equal within-subject variance we have:
$$\text{VAR}_{\text{EM}} = \text{Var-between}/N + \text{Var-within}/Nn$$
- In this case, the Summary Stat approach gives the same results, on average:
$$\text{Avg}[\alpha] = \text{MEAN}_{\text{EM}} \text{ and } \text{Avg}[\text{Var}(\alpha)] = \text{VAR}_{\text{EM}}$$
- In other cases, with $N \sim 12$, and typical ratios of between-subject to within-subject variance found in fMRI, the Summary Stat approach will give very similar results to EM.

Note: with different models X per subject better to use Bayes ...

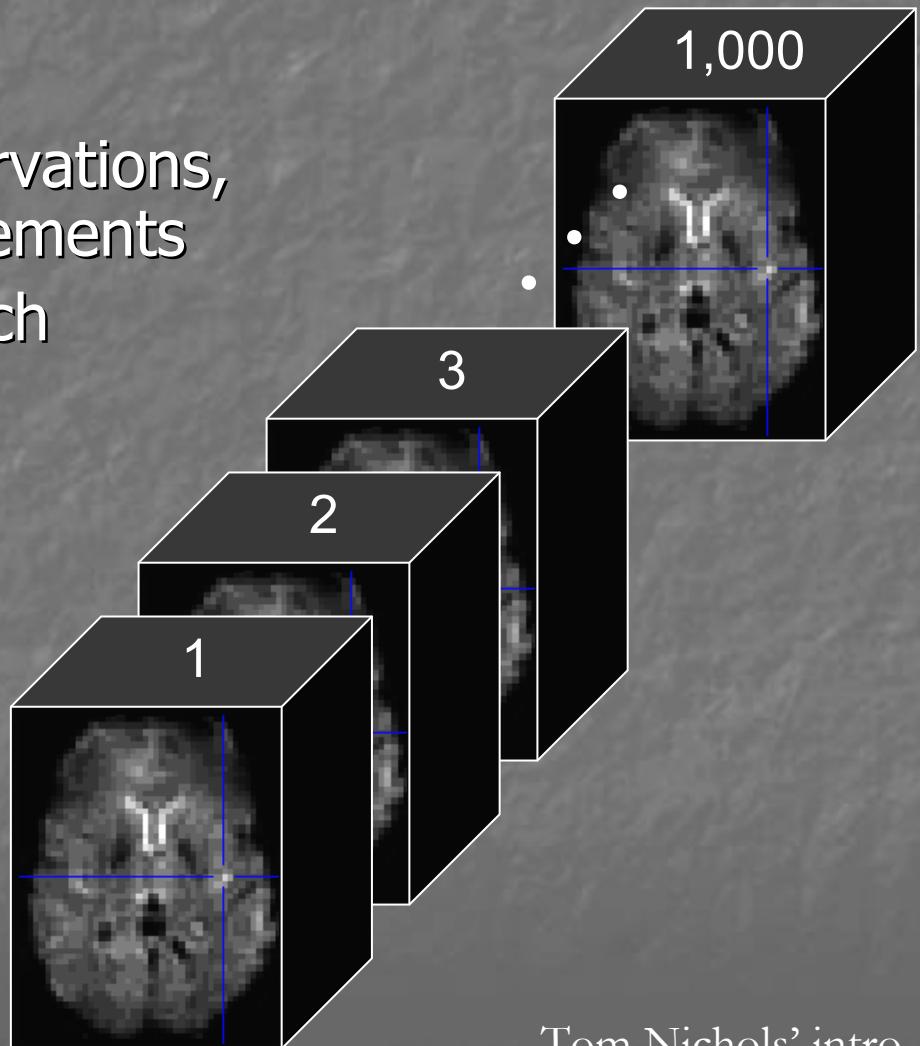
Part III: Statistical inferences

Stats Inference Overview

- Multiple comparisons correction procedures
- Levels of inferences (set, cluster, voxel)
- Circularity issues

What Problem?

- 4-Dimensional Data
 - 1,000 multivariate observations, each with $> 100,000$ elements
 - 100,000 time series, each with 1,000 observations
- Massively Univariate Approach
 - 100,000 hypothesis tests
- Massive MCP!



Tom Nichols' intro

What Problem?

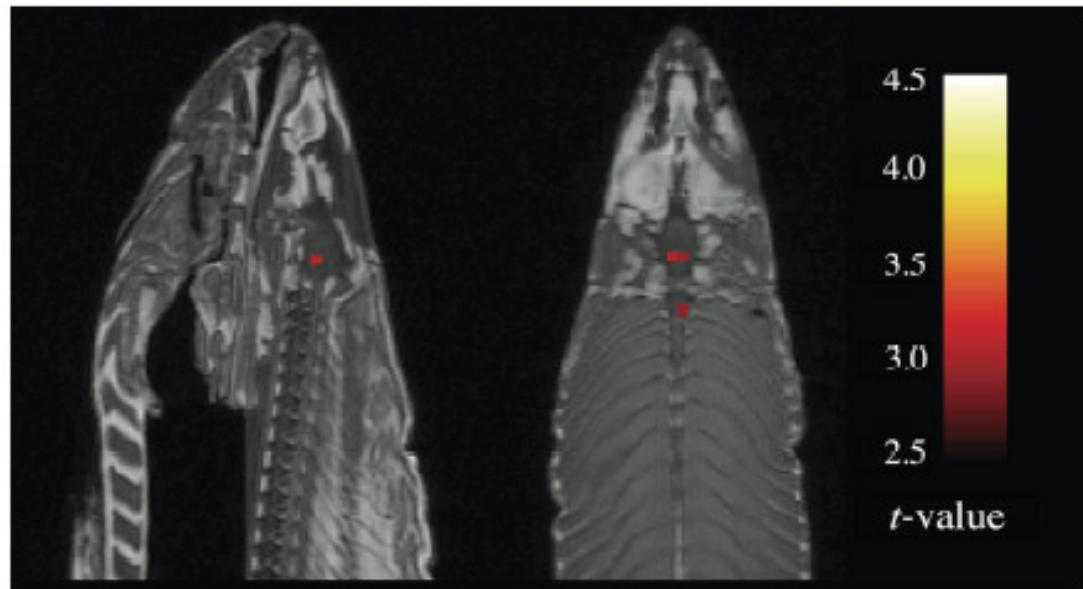
- Typical brain ~ 130000 voxels
- @ $p = .05$, it is expected = 6500 false positives!
- @ a more conservative value like $p = .001$ we still expect 130 false positives.

- Using extend threshold k without correction is not enough as it, by chance, can cluster as well.

What Problem?

- Bennet et al., 2009
- Task: take a decision about emotions on pictures
- Design: blocks of 12 sec activation/rest
- Analysis: standard data processing with SPM
- Subject: a dead salmon!

What Problem?



A t -contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, $p(\text{uncorrected}) < 0.001$, 3 voxel extent threshold.

- The cluster was 81mm^3 ! – after multiple comparison corrections all false activations were removed.

Solutions for MCP

- Height Threshold
 - Familywise Error Rate (FWER)
 - Chance of *any* false positives; Controlled by Bonferroni & Random Field Methods
 - False Discovery Rate (FDR)
 - Proportion of false positives *among* rejected tests
 - Bayes Statistics

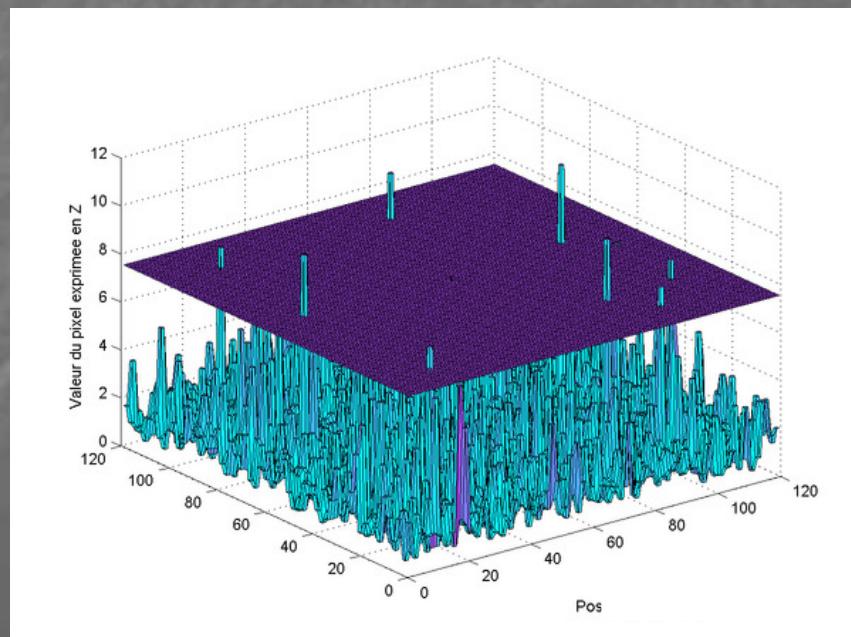
From single univariate to massive univariate

Univariate stat	Functional neuroimaging
1 observed data	Many voxels
1 statistical value	Family of statistical values
Type 1 error rate (chance to be wrong rejecting H0)	Family-wise error rate
Null hypothesis	Family-wise null hypothesis

Height Threshold

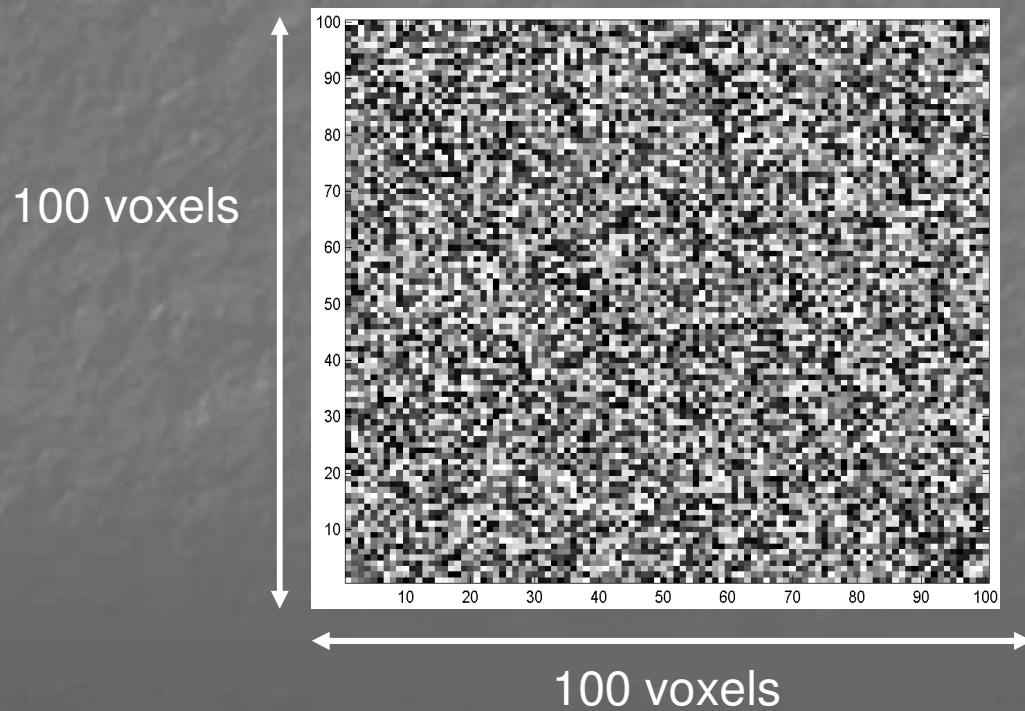
- Choose locations where a test statistic Z (T , F , ...) is large to threshold the image of Z at a height z
- The problem is how to choose this threshold z to exclude false positives with a high probability (e.g. 0.95)?

To control for family wise error one must take into account the nb of tests



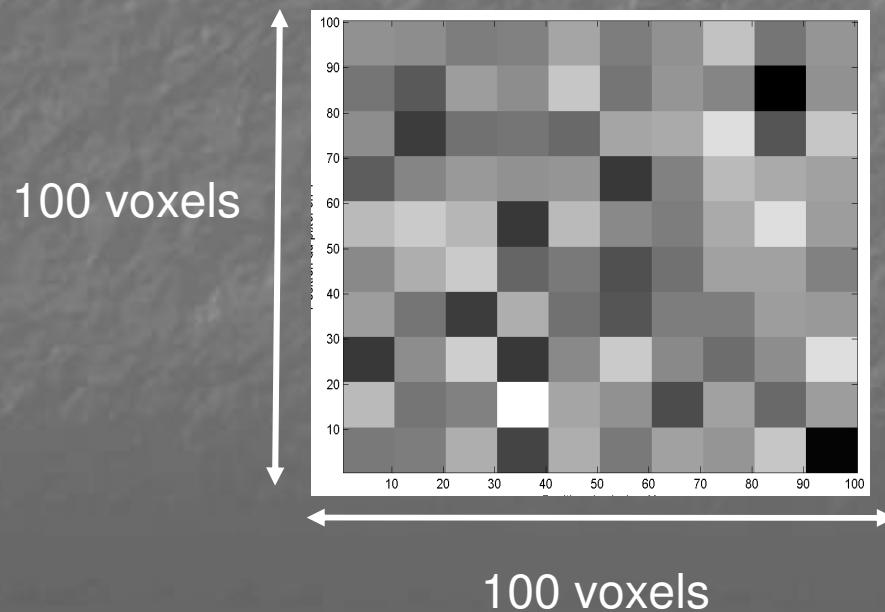
Bonferroni

- 10000 Z-scores ; alpha = 5%
- alpha corrected = .000005 ; z-score = 4.42



Bonferroni

- 10000 Z-scores ; alpha = 5%
- 2D homogeneous smoothing – 100 independent observations
- alpha corrected = .0005 ; z-score = 3.29

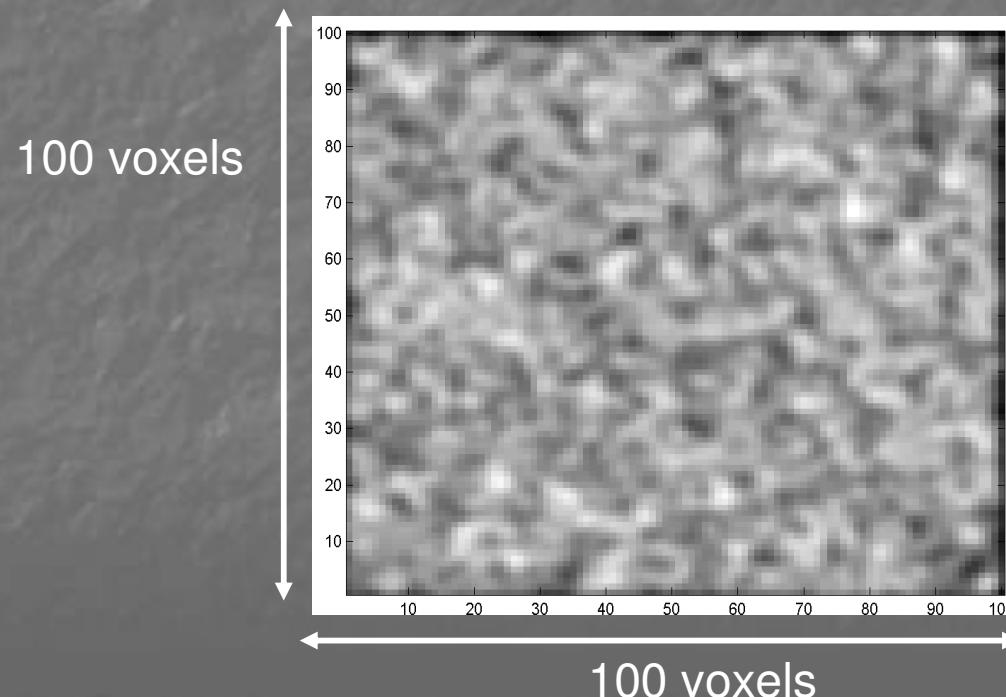


Solutions for MCP

- An important feature of neuroimaging data is that we have a family of stat values that has topological features (Bonferroni for instance consider tests as independent)
- Why considering data as a smooth lattice? (Chumbley et al., 2009 NeuroImage 44)
 - fMRI/PET are projection methods of data points onto the whole space – MEEG forms continuous functions in time and are smooth by the scalp (space)
 - Neural activity propagate locally through intrinsic/lateral connections and is distributed via extrinsic connections / Hemodynamic correlates are initiated by diffusing signals (e.g. NO)

Random Field Theory

- 10000 Z-scores ; alpha = 5%
- Gaussian kernel smoothing –
- How many independent observations ?

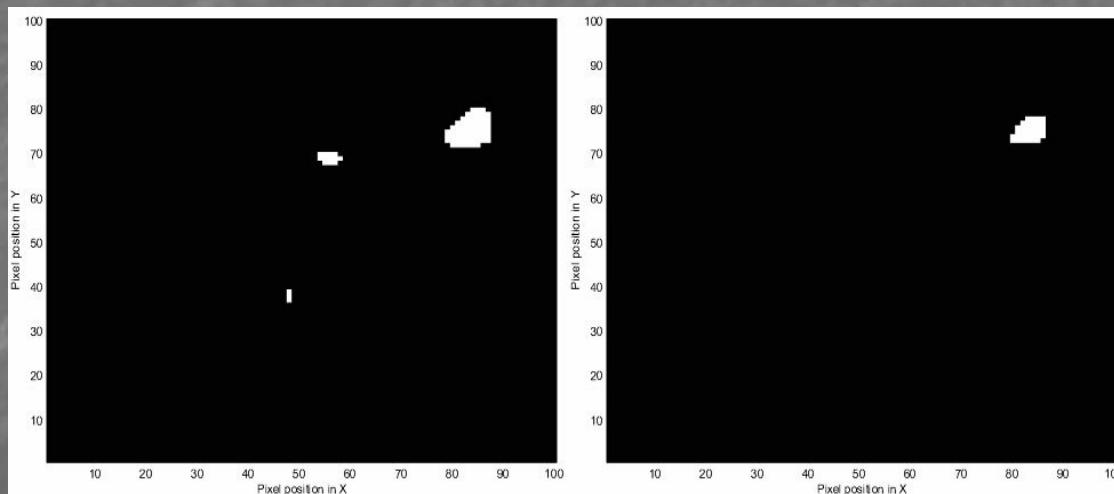


Random Field Theory

- RFT relies on theoretical results for smooth statistical maps (hence the need for smoothing), allowing to find a threshold in a set of data where it's not easy to find the number of independent variables. Uses the expected Euler characteristic (EC density)
- 1 Estimation of the smoothness = number of resel (resolution element) = $f(\text{nb voxels}, \text{FWHM})$
- 2 expected Euler characteristic = number of clusters above the threshold
- 3 Calculation of the threshold

Random Field Theory

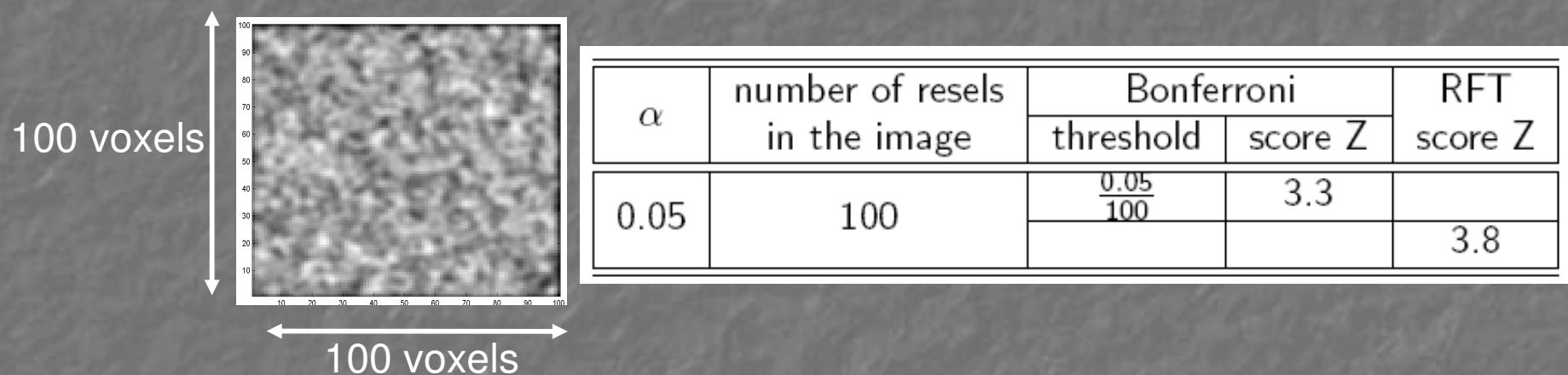
- The Euler characteristic can be seen as the number of blobs in an image after thresholding (p value that you select in SPM)
- At high threshold, $EC = 0$ or 1 per resel: $E[EC] \approx p^{FWE}$



$$E[EC] = R \cdot (4 \log_e 2) \cdot (2\pi)^{-2/3} \cdot Z_t \cdot e^{-1/2 Z_t^2} \text{ for a 2D image, more complicated in 3D}$$

Random Field Theory

- For 100 resels, the equation gives $E[EC] = 0.049$ for a threshold Z of 3.8, i.e. the probability of getting one or more blobs where Z is greater than 3.8 is 0.049



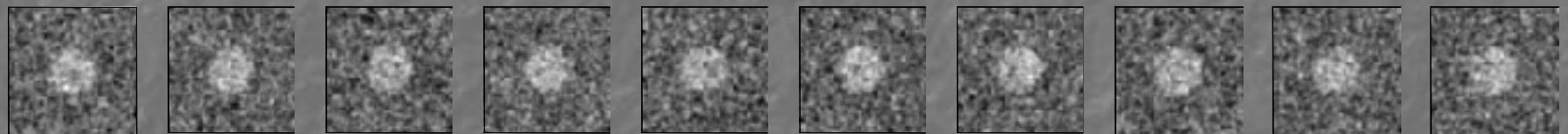
- If the resel size is much larger than the voxel size then $E[EC]$ only depends on the nb of resels otherwise it also depends on the volume, surface and diameter of the search area (i.e. shape and volume matter)

False discovery Rate

- Whereas family wise approach corrects for any false positive, the FDR approach aim at correcting among positive results only.
 1. Run an analysis with alpha = x%
 2. Sort the resulting (positive) data
 3. Threshold to remove the false positives

False discovery Rate

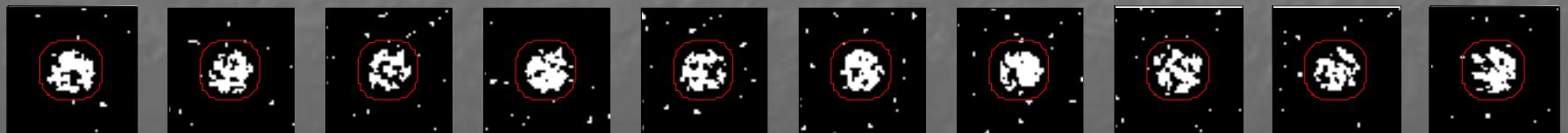
Signal+Noise



FEW correction



FDR correction

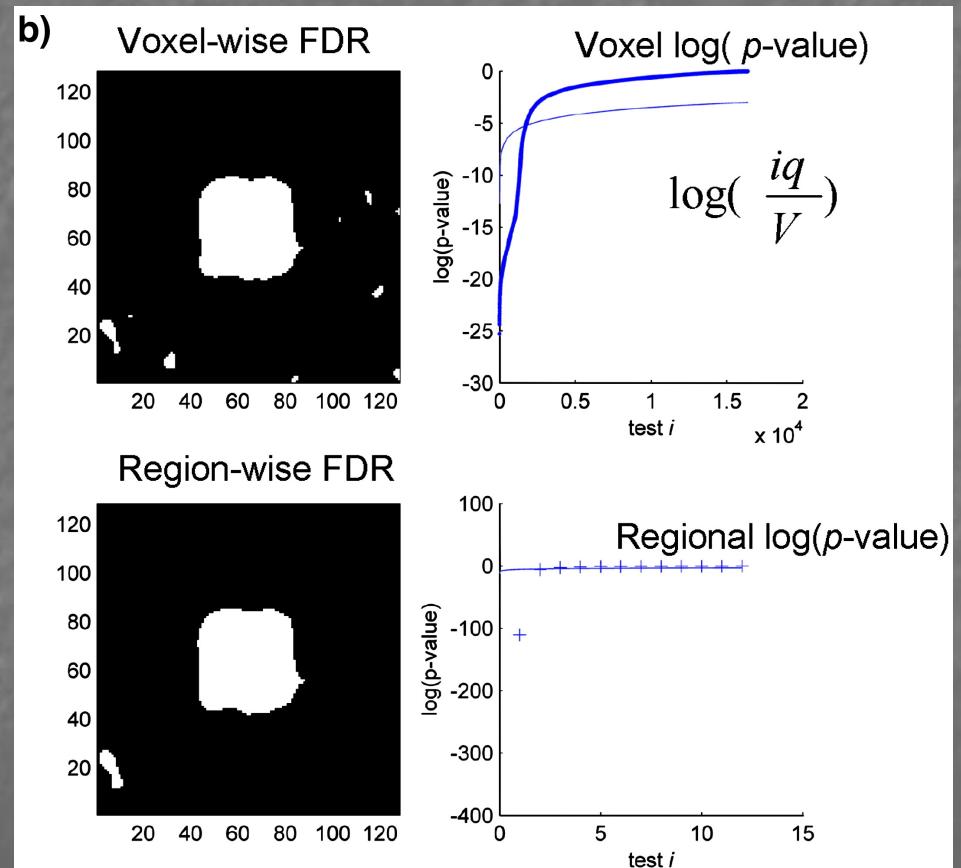


False discovery Rate

takes the spatial structure into account

Under H0 the nb of voxels per cluster is known → uncorrected p value for clusters → apply FDR on the clusters (volume-wise correction)

Assumes that the volume of each cluster is independent of the number of clusters

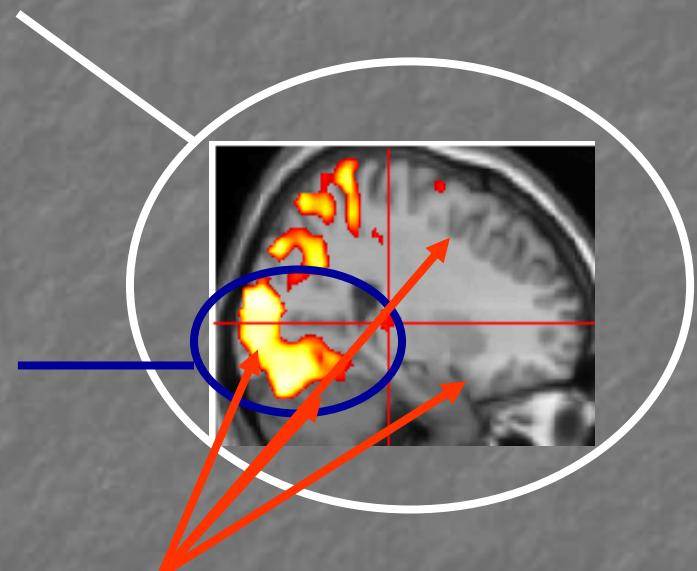


Levels of inference

- 3 levels of inference can be considered:
 - Voxel level (prob associated at each voxel)
 - Cluster level (prob associated to a set of voxels)
 - Set level (prob associated to a set of clusters)
- The 3 levels are nested and based on a single probability of obtaining c or more clusters (set level) with k or more voxels (cluster level) above a threshold u (voxel level):
 $P_w(u, k, c)$

Levels of inference

- Set level: we can reject H_0 for an omnibus test, i.e. there are some significant clusters of activation in the brain.
- Cluster level: we can reject H_0 for an area of a size k , i.e. a cluster of 'activated' voxels is likely to be true for a given spatial extend.
- Voxel level: we can reject H_0 at each voxel, i.e. a voxel is 'activated' if exceeding a given threshold



Levels of inference

- Each level of inference is valid, but the inferences are different – e.g. a set might be enough to check that subjects activated regions selected a priori for a connectivity analysis – clusters might be good enough if hypotheses are about the use of different brain areas between groups
- Both voxel and cluster levels need to address the multiple comparison problem. If the activated region is predicted in advance, the use of corrected p values is unnecessary and inappropriately conservative – a correction for the number of predicted regions (Bonferroni) is enough

Levels of inference

Statistics: p-values adjusted for search volume										
set-level		cluster-level				peak-level				
D	C	D _{FWE-corr}	q _{FDR-corr}	k _E	D _{uncorr}	D _{FWE-corr}	q _{FDR-corr}	T	(Z _E)	D _{uncorr}
0.226	21	0.000	0.000	508	0.000	0.070	0.101	7.94	4.90	0.000
		0.184	0.101	7.09	4.63	0.000	0.000	8	40	-8
		0.704	0.309	5.71	4.10	0.000	0.000	-4	38	-6
		0.959	0.364	5.00	3.78	0.000	0.000	-48	-66	24
		0.999	0.588	4.43	3.49	0.000	0.000	-44	-66	38
		0.187	0.101	7.08	4.62	0.000	0.000	-24	36	46
		0.198	0.101	7.03	4.61	0.000	0.000	-6	-50	30
		0.865	0.332	5.35	3.94	0.000	0.000	-6	-62	24
		0.910	0.358	5.21	3.88	0.000	0.000	-20	-50	28
		0.773	0.309	5.57	4.04	0.000	0.000	2	60	16
		0.818	0.326	5.46	3.99	0.000	0.000	-32	-22	22
		0.868	0.332	5.34	3.93	0.000	0.000	2	10	-8
		0.930	0.358	5.14	3.84	0.000	0.000	50	-70	36
		0.947	0.358	5.06	3.81	0.000	0.000	-2	-26	42
		0.964	0.364	4.97	3.76	0.000	0.000	12	-56	18
		0.989	0.475	4.74	3.65	0.000	0.000	-30	-50	8
		0.998	0.572	4.51	3.53	0.000	0.000	26	34	40
		1.000	0.781	4.07	3.29	0.001	0.001	20	38	34
		0.999	0.588	4.43	3.49	0.000	0.000	-16	64	10
		1.000	0.777	4.13	3.33	0.000	0.000	-36	-48	-2
		1.000	0.777	4.12	3.32	0.000	0.000	-12	62	20
		1.000	0.781	4.08	3.29	0.000	0.000	-30	-26	54
		1.000	0.936	3.87	3.17	0.001	0.001	44	-78	30
table shows 3 local maxima more than 8.0mm apart										
Height threshold: T = 3.73, p = 0.001 (1.000) Extent threshold: k = 0 voxels, p = 1.000 (1.000) Expected voxels per cluster, <k> = 10.115 Expected number of clusters, <> = 17.44 FWEp: 8.202, FDRp: Inf, FWEc: 133, FDRC: 133										
Degrees of freedom = [1.0, 15.0] FWHM: 10.8 10.6 10.0 mm mm mm; 5.4 5.3 5.0 {voxels} Volume: 1265568 = 158196 voxels = 1018.0 resels Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 142.91 voxels) Page 1										

RFT (Gaussian Random Fields)
 -> Prob of cluster
 -> Prob of voxel

Using p=.001 this creates an excursion set
 Prob clusters of that size
 Prob peak that height
 → after FDR correction

Uncorrected (bad)

Circularity

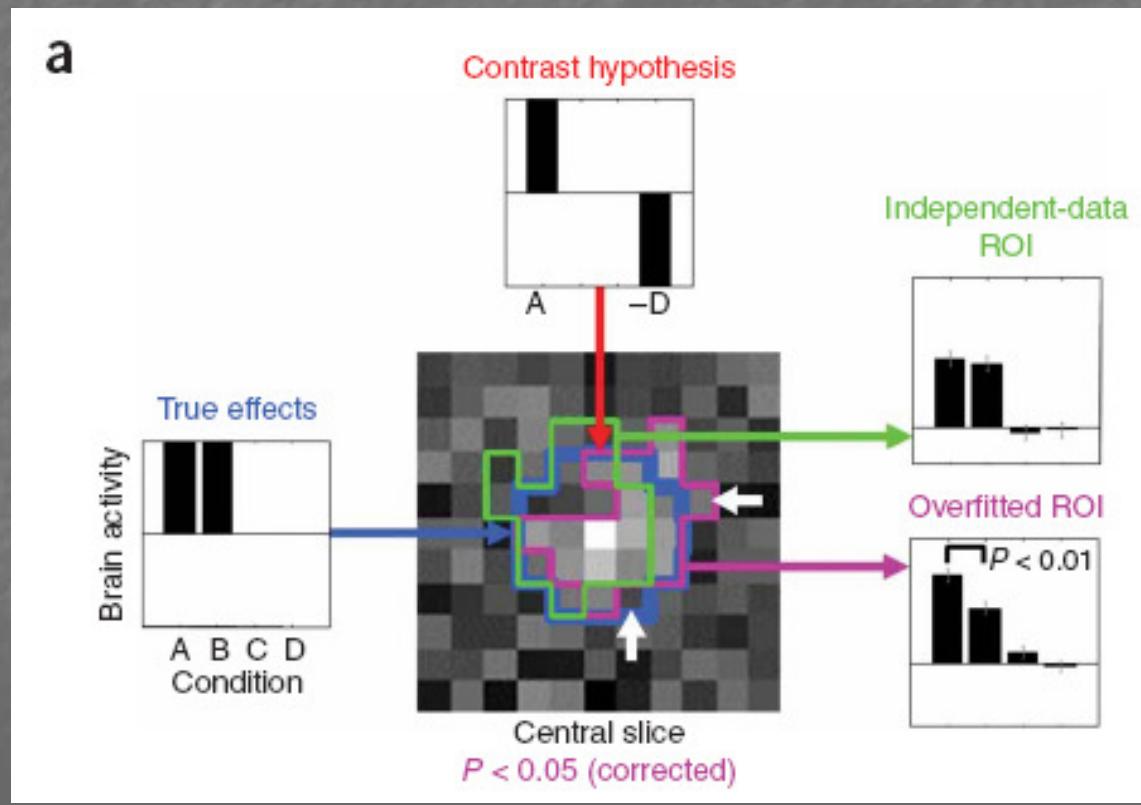
- Refers to the problem of selecting data for analysis
- How data (areas usually) are selected, analysed and sorted is key to avoid circularity
- Put forward by Vul et al. 2009, *Perspectives on Psychological Science*. 4
- Better explained in Kriegeskorte et al. *Nat. Neuroscience* 12 (2009) & *JCBF* 30 (2010).

Circularity

- Double dipping pblm: “data are first analyzed to select a subset and then the subset is reanalyzed to obtain the results. In this context, assumptions and hypotheses determine the selection criterion and selection can, in turn, distort the results.”
- Take several groups of subjects and test RTs differences, then take 2 subgroups only from the same subjects and re-do some analysis?? → increases the diff.
- Take fMRI data and get activated areas, extract ROI and re-do some analyses??

Circularity

- Selection and tests must be independent – non independence create spurious effects



Circularity

- Independence of the selection and tests
 1. Anatomic ROI, analysis of fMRI
 2. SPM, minimal requirement is orthogonality of the contrasts (e.g. find regions using $A+B>0$ $C=[1 \ 1]$ and test A vs B $C=[1 \ -1]$) but if N_A and N_B are different there is still a bias when testing $A-B$ (across subjects independence is ensured by $C_{selection}^T(X^TX)^{-1}C_{test}$)
 3. Select using a subset of data, test with another one