Random Field Theory in fMRI

"SPM for Dummies"

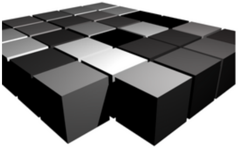
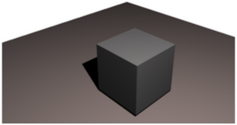
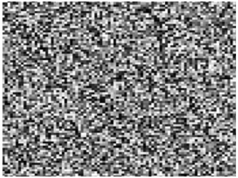
Jean-Etienne Poirrier

Cyclotron Research Center

January 11th, 2006

SPM, "*Statistical Parametric Mapping*"

1. Raw data collected as a group of voxels
2. Each voxel is independtly analysed
3. Creation of statistical "maps" coming from these independent statistical analysis: SPMs ("T maps" or "F maps")

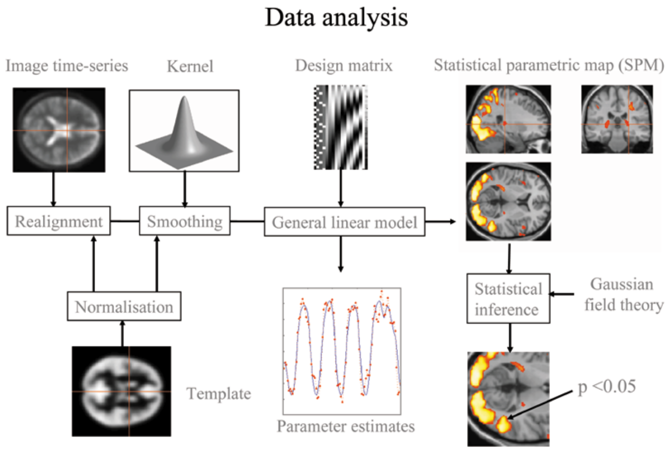




SPM analysis process

SPM software ...

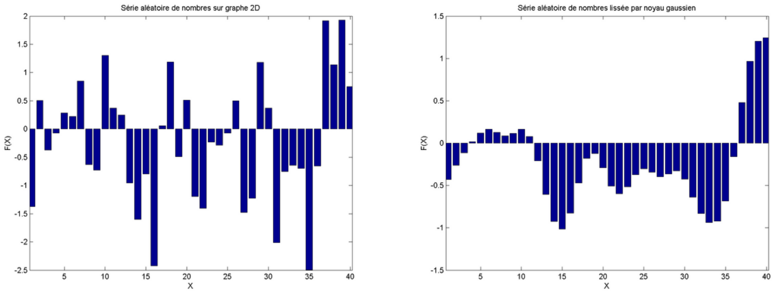
1. independently analyses variance for each voxel
2. creates t statistics for each voxel (data $\longrightarrow t$)
3. finds an equivalent Z score for t ($t \longrightarrow Z$)
4. shows t maps (SPM99) or Z maps (SPM96)
5. suggests a correction for the significance of t statistics (SPM99) or Z scores (SPM96) which take account of the multiple comparisons in the image

Data analysis process

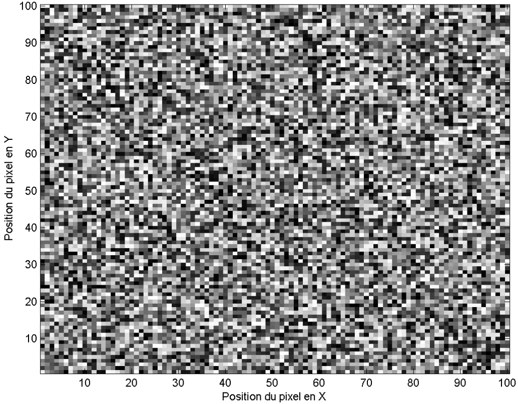


What does spatial smoothing?

Spatial smoothing reduces effect of high frequency variation in functional imaging data ("blurring sharp edges")



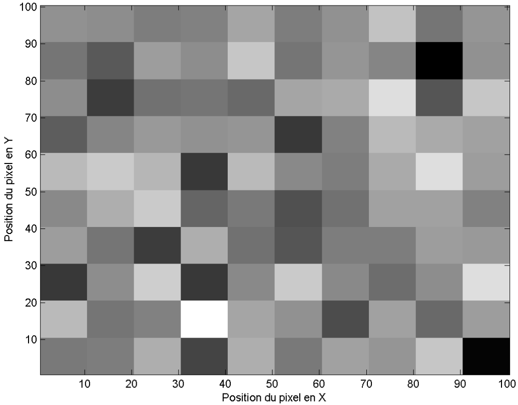
How to do a spatial smoothing?



Two examples : simple smoothing by a mean and smoothing by a Gauss kernel

Simple smoothing by a mean

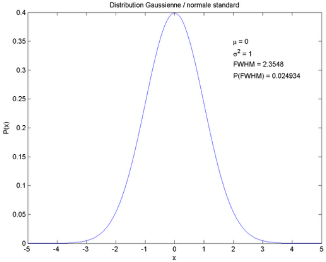
Remplacement of values in 10-pixels-side squares by the mean of all values in this square

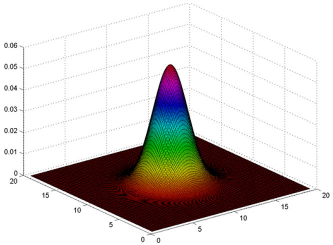


Gauss kernel

- Typically used in functional imaging, uses a form similar to normal distribution "bell curve"
- FWHM (Full Width at Half Maximum) $= \sigma \cdot \sqrt{8 \cdot \log 2}$

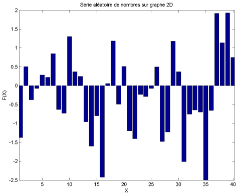
$$FG(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \exp^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$$

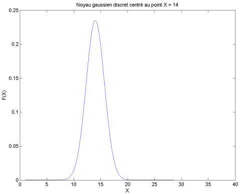


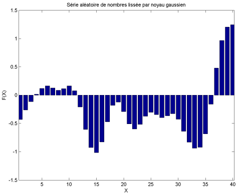


Smoothing by a Gaussian kernel - 2D

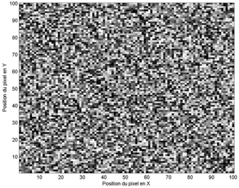
The Gauss kernel defines the form of the function successively used to compute the weighted average of each point (in relation with its neighbors)

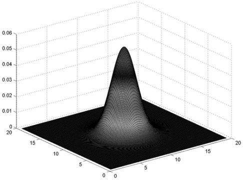


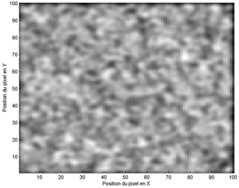




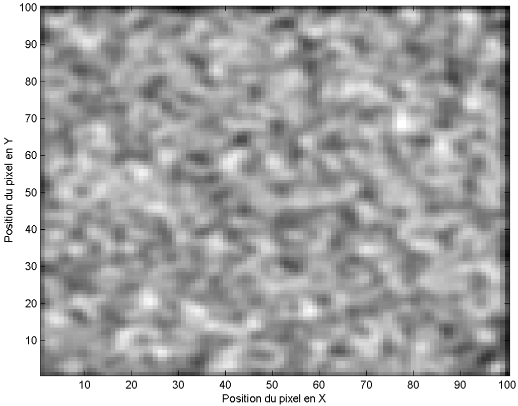
Smoothing by a Gaussian kernel - 3D







Smoothing by a Gaussian kernel - 3D



Why use spatial smoothing?

1. to increase signal-to-noise ratio
2. to enable averaging across subjects
3. to allow use of the RFT for thresholding

Spatial smoothing $\longrightarrow$ increases signal-to-nois ratio

- Depends on relative size of smoothing kernel and effets to be detected
- Matched filter theorem: smoothing kernel = expected signal
- Practically: FWHM $\geq 3\cdot$ voxel size
- May consider varying kernel size if interested in different brain regions (e.g. hippocampus -vs- parietal cortex)

Spatial smoothing → enables averaging across subjects

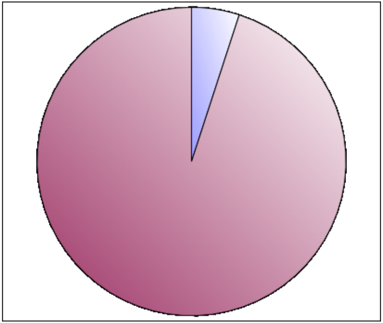
- Reduces influence of functional and/or anatomical differences between subjects
- Even after realignment and normalisation, residual between-subject variability may remain
- Smoothing data improves probability of identifying commonalities in activation between subjects (but trade-off with anatomical specificity)

Spatial smoothing $\longrightarrow$ allows use of RFT for thresholding

- Assumes error terms are roughly Gaussian form
- Requires FWHM to be substantially greater than voxel size
- Enables hypothesis testing and dealing with multiple comparison problem in functional imaging ...

Null hypothesis in “classical” statistics

- Data $\longrightarrow$ statistical value
- Null hypothesis: the hypothesis that there is no effect
- Null distribution: distribution of statistic values we would expect if there is no effect
- Type I error rate: the chance we take that we are wrong when we reject the null hypothesis



Example

Degrees of freedom	α		
	0.05	0.02	0.01
30	2.042	2.457	2.750
40	2.021	2.423	2.704
60	2.000	2.390	2.660
120	1.980	2.358	2.617
∞	1.960	2.326	2.576

The multiple comparison problem

- 1 voxel $\longrightarrow$ is this voxel activation significantly different from zero?
- Many voxels $\longrightarrow$ huge amount of statistical values

How to "sort" them all? Where will our effect be?
Evidence against the null hypothesis: the whole observed *volume* of values is unlikely to have arison from a null distribution

From simple stats to functional imaging

Univariate statistics	→	Functional imaging
1 observed data	→	many voxels
1 statistical value	→	family of statistical values
type I error rate	→	family-wise error rate (FWE)
null hypothesis	→	family-wise null hypothesis

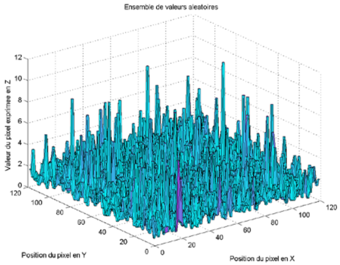
Test methods for the family-wise null hypothesis

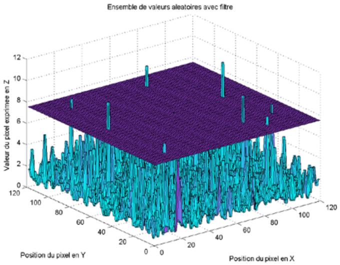
1. Height threshold
   - Maximum Test Statistic
     - Bonferroni correction
     - Random Field Theory
   - Maximum spatial extent of the test statistic
   - False Discovery Rate
2. Set-level inference
3. Cluster-level inference

Maximum Test Statistic methods

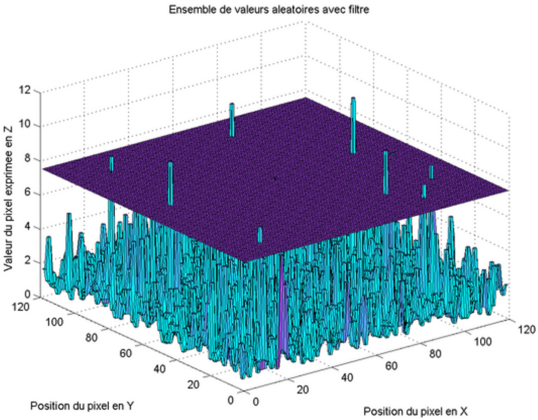
- Simple: choose locations where a test statistic Z (T, χ^2, ...) is large, i.e. to threshold the image of Z at a height z
- the problem is deferred: how to choose this threshold z to exclude false positives with a high probability (e.g. 0.95)?

Height thresholding

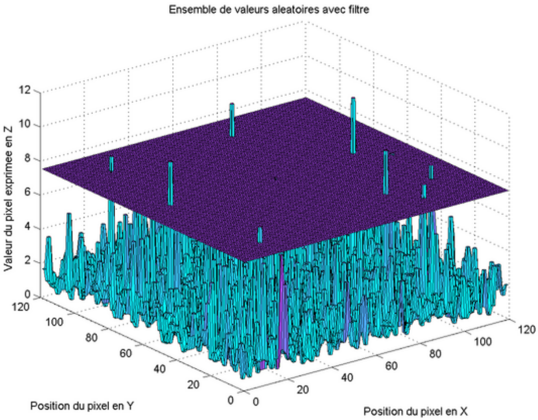




Height threshold and localising power



Height threshold and localising power



However, a height threshold that can control family-wise error must take into account the number of tests!

Bonferroni Correction

- Simple method of setting the threshold above which values are unlikely to have arison by chance
- Based on probability rules

Mathematical expression

For one voxel (all values from a null distribution):

- Probability to be greater than the threshold: α
- Probability to be smaller than the threshold: $(1-\alpha)$

$\forall n$

Mathematical expression of Bonferroni correction

For a family of n values:

- Probability that *all* the n tests being less than α: $(1-\alpha)^n$
- Family-wise error rate, P^{FWE} : probability that one or more values will be greater than α
- $P^{FWE} = 1 - (1-\alpha)^n$
- Since α is small $(\Rightarrow \alpha^n \approx 0)$: $P^{FWE} \leq n \cdot \alpha$

$$\alpha = \frac{P^{FWE}}{n}$$

Notations

	1 value	family of values
Number of statistical values	n (100 000)	
Number of degree of freedom	(40)	
Error rate p	α	P^{FWE}
p corrected for the family?	no	yes

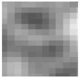
Bonferroni correction is not often applicable!

- Still used in some functional imaging analysis
- In other cases: **too conservative**
- Because most functional imaging data have some degree of spatial correlation (correlation between neighbouring statistic values). So:

$$\text{Number of independant values} < \text{number of voxels}$$

Spatial correlation

The value of any voxel in a functional image tends to be similar to those of neighbouring voxels. Some degree of spatial correlation is almost universally present in these data.



Multiple reasons :

- the way that the scanner collects and reconstructs the image (see point spread function)
- physiological signal
- spatial preprocessing applied to data before statistical analysis (realignment, spatial normalisation, resampling, smoothing)

Mathematical implication of spatial correlation

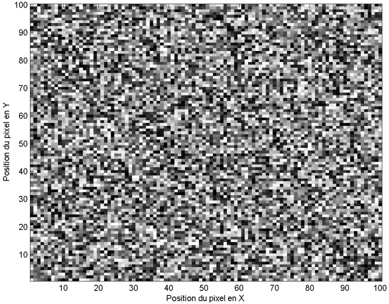
- Values of independent probability (Bonferroni) : $P^{FWE} = 1 - (1 - \alpha)^n$
- If $n \rightarrow$ number of independent observations n_i :

$$P^{FWE} = 1 - (1 - \alpha)^{n_i}$$

$$\alpha = \frac{P^{FWE}}{n_i}$$

Example (1/3)

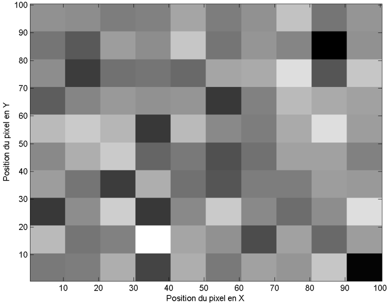
100 x 100 random number from a normal distribution



$= 10\ 000$ scores Z. For $P^{FWE} = 0.05$,
$\alpha_{\text{Bonferroni}} = \frac{0.05}{10000} = 0.000005 \rightarrow$ score Z of 4.42

Example (2/3)

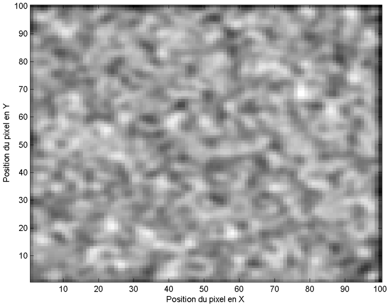
Previous image where we applied a simple spatial correlation (simple smoothing)



Still 10 000 scores Z but only 100 independent values! For $P^{FWE} = 0.05$, $\alpha_{\text{Bonferroni}} = \frac{0.05}{100} = 0.0005 \rightarrow$ score Z of 3.29

Example (3/3)

First image where we applied a complex spatial correlation (smoothing by a Gaussian kernel, FWHM of 10 pixels)



Still 10 000 scores Z but how many independent values? Probably less than 10 000 ; but *how many*? If we don't have n_i, how can we find P^{FWE} ?

Random Field Theory

- Recent body of mathematics defining theoretical results for smooth statistical maps
- Allows to find a threshold in a set of data where it's not easy (or even impossible) to find the number of independent variables
- Uses the expected Euler characteristic (EC density)

expected EC $\longrightarrow$ number of clusters above the threshold $\longrightarrow$ height threshold

1. Estimation of the smoothness
2. $\longrightarrow$ expected Euler characteristic
3. Calculation of the threshold

Smoothness & resels

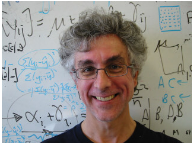
Smoothness

- unknown for SPMs because of the initial spatial correlation + treatments (→ see some slides after this one)
- known for our map of independent random number... "width of the smoothing kernel"

Smoothness & resels

Smoothness

- unknown for SPMs because of the initial spatial correlation + treatments (→ see some slides after this one)
- known for our map of independent random number... "width of the smoothing kernel"

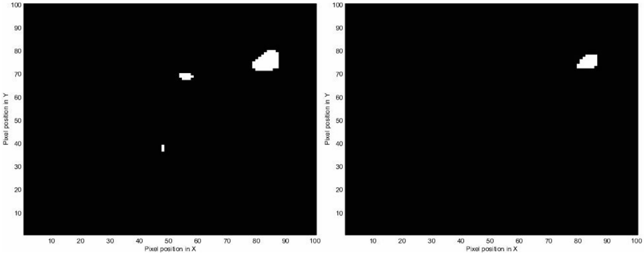


Resels (resolution elements)

- a measure of the number of "resolution elements"
- a bloc of values that is the same size as the FWHM
- the number of resels only depends on smoothness (FWHM) and the total number of pixels (voxels)

Euler characteristic

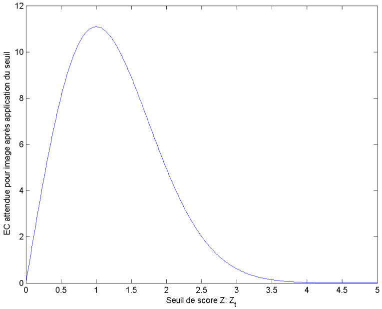
- property of an image after it has been thresholded
- can be seen as the number of blobs in an image after thresholding
- at high threshold, EC = 0 ou 1 $\Rightarrow$ mean or expected EC: $\mathrm{E}[EC] \approx P^{FWE}$



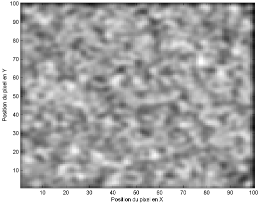
Expected Euler characteristic formula

$$E[EC] = R \cdot (4 \log_e 2) \cdot (2\pi)^{-\frac{2}{3}} \cdot Z_t \cdot e^{-\frac{1}{2} Z_t^2}$$

- 2 dimensions image
- R = number of resels
- Z_t = threshold of score Z



Euler Characteristic in our example



For 100 resels, the equation gives $\mathrm{E}[EC] = 0.049$ for a threshold Z of 3.8: the probability of getting one or more blobs where Z is greater than 3.8 is 0.049

α	number of resels in the image	Bonferroni		RFT
		threshold	score Z	score Z
0.05	100	$\frac{0.05}{100}$	3.3	
				3.8

P with RFT

$$P(\max Z > z) \approx \sum_{d=0}^{D} \text{Resels}_d \cdot \text{EC}_d(z)$$

- D, number of dimensions in the search region
- Resels_D, number of d-dimensional resels
- EC_d, d-dimensional Euler characteristics density

The left hand side of the equation is the exact expectation of the Euler characteristic of the region above threshold z.
This approximation is accurate for search regions of any size, even a single point, but it is best for search regions that are not too concave.

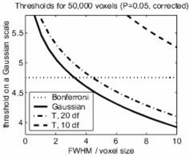
The size of search region

$$P(\max Z > z) \approx \sum_{d=0}^{D} \text{Resels}_d \cdot \text{EC}_d(z)$$

- Large search regions: the last term ($D = d$) is the most important. The number of resels is:

$$\text{Resels}_D = \frac{V}{\text{FWHM}^D}$$

- Small search regions: the lower dimensional terms ($d < D$) become important



E[EC] for a T statistic image

$$\mathrm{EC}_3(z) = \frac{(4\log_e 2)^{\frac{2}{3}}}{(2\pi)^2}\left(\frac{\nu-1}{\nu}z^2-1\right)\left(1+\frac{z^2}{\nu}\right)^{-\frac{1}{2}(\nu-1)}$$

ν = number of degree of freedom

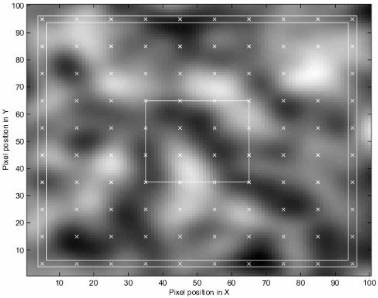
RFT and 3D functional imaging

- EC is the number of 3D blobs of Z scores above a certain threshold
- A resel is a cube of voxels of size (FWHM) in x, y et z
- The equation for E[EC] is different but still only depends on resels in image
- Equivalent results available for RF of t, F and χ^2 scores

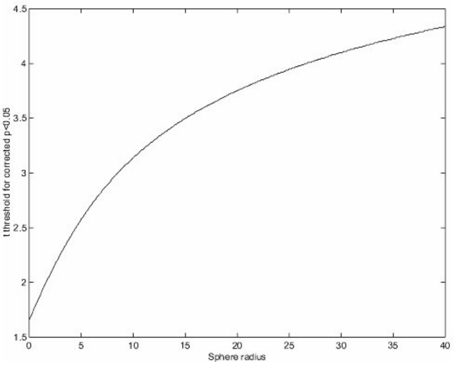
Smoothness of a statistic volume from functional imaging? Calculated using the residual values from the statistical analysis...

Shape and volume are important!

- Volume of resels >> size of a voxel : E[EC] only depends on the number of resels inside the volume considered
- Other cases : E[EC] depends on
  - the number of resels
  - the volume
  - the surface area and
  - the diametre of the search region



Shape and volume



Regional hypotheses

One never practically work on the whole brain volume

- Hypothetised region = 1 voxel → inference could be made using an uncorrected p-value
- Hypothetised = many voxels (≈ spheres or boxes) → inference must be made using a p-value that has been appropriately corrected

Underlying assumptions

1. The error fields are a reasonnable lattice approximation to an underlying random field with a multivariate Gaussian distribution
2. The error fields are continuous, with a twice-differentiable autocorrelation function (not necessarily Gaussian)

If the data have been sufficiently smoothed and the General Linear Model correctly specified (so that the errors are indeed Gaussian) then the RFT assumptions will be met.
Otherwize ...

When underlying assumptions are not met

Example: Random effect analysis with a small number of subjects
Solutions:

1. to reduce the voxel size by sub-sampling
2. other inference procedures:
    1. nonparametric framework (ch. 16)
    2. False Discovery Rate
    3. bayesian inference (ch. 17)

More on RFT

- Maximum spatial extend of the test statistic
- Searching in small regions
- Estimating the FWHM
- False Discovery Rate

Maximum spatial extend of the test statistic

Method based on the spatial extend of clusters of connected components of supra threshold voxels where$Z > z \approx 3$

Idea to approximate the shape of the image by a quadratic with a peak at the local maximum

For a Gaussian random field, the spatial extend S is...

$$S \approx cH^{\frac{D}{2}}$$

...

Searching in small regions

For small pre-defined search regions, the P-values for the maximum test statistic are very well estimated, except for the previous method $\longrightarrow$ Friston have proposed a method that avoids the awkward problem of pre-specifying a small region altogether.

1. thresholding of the image of test statistic z
2. pick the nearest peak to a point or region of interest
3. identification on spatial location $\longrightarrow$ no need to correct for searching over all peaks

Estimating the FWHM

1. The only 2 data-dependent component required: Resels_D et FWHM
2. FWHM often depends on the location $\rightarrow$ random field not isotropic
3. Estimating the FWHM separately at each voxel

$$\text{FWHM} = (4\log 2)^{\frac{1}{2}} \left|\mathbf{u}'\mathbf{u}\right|^{\frac{-1}{2D}}$$

$$\text{Resels}_D = \sum_{\text{volume}} \text{FWHM}^{-D} v$$

...

False Discovery Rate

Procedure for controlling the expected proportion of false positives amongst those voxels declared positive

1. Calculate the uncorrected P-value for each voxel
2. Order them so that $P_1 \leq P_2 \leq P_3 \leq \cdots \leq P_N$
3. To control the FDR at α, find the largest value k so that:

$$P_k < \frac{\alpha k}{N}$$

- This procedure is conservative is the voxels are positively dependent
- The resulting threshold, corresponding to the value of Z for P_k depends on the amount of signal in the data (and not on the number of voxels or the smoothness)
- Interpretation is différent!

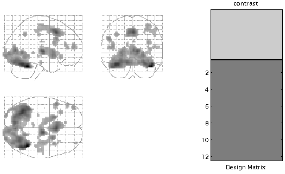
Which correction method to use?

- FWE (RFT) is the most "correct" method, but FDR may be more sensitive in some cases
- May be a good idea to use whatever method is employed in previous related studies, to increase comparability

Most important is **to decide on correction method a priori**, rather than subjectively adjusting thresholds to give desirable results!

Where can I find these values?

SnPM{Pseudo-t}



P values & statistics: .

cluster-level	voxel-level				x,y,z mm		
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	Pseudo-t	$p_{uncorrected}$			
2386	**0.0002**	**0.0023**	**14.37**	0.0002	**42**	**-48**	**-30**
	0.0002	0.0023	11.51	0.0002	45	-60	-27
	0.0002	0.0023	11.33	0.0002	-33	-60	-24
356	**0.0002**	**0.0023**	**10.57**	0.0002	**3**	**15**	**45**
367	**0.0002**	**0.0023**	**10.07**	0.0002	**12**	**-15**	**3**
	0.0005	0.0023	8.25	0.0002	-21	9	-3
	0.0056	0.0023	6.74	0.0002	-18	-6	9
92	**0.0005**	**0.0023**	**8.42**	0.0002	**24**	**6**	**-6**
65	**0.0010**	**0.0023**	**7.98**	0.0002	**-51**	**15**	**30**
	0.0112	0.0023	6.32	0.0002	-42	12	21
33	**0.0022**	**0.0023**	**7.31**	0.0002	**-42**	**54**	**15**
16	**0.0032**	**0.0023**	**7.16**	0.0002	**3**	**-18**	**-21**
21	**0.0034**	**0.0033**	**7.11**	0.0005	**36**	**9**	**24**
23	**0.0049**	**0.0023**	**6.81**	0.0002	**-33**	**-54**	**42**
35	**0.0054**	**0.0033**	**6.76**	0.0005	**-33**	**24**	**-9**
10	**0.0063**	**0.0023**	**6.68**	0.0002	**3**	**-18**	**24**
29	**0.0063**	**0.0023**	**6.68**	0.0002	**-48**	**39**	**12**
187	**0.0066**	**0.0023**	**6.67**	0.0002	**36**	**-12**	**60**
	0.0083	0.0023	6.57	0.0002	36	-21	63
	0.0127	0.0033	6.16	0.0005	54	-24	54

Height threshold:statistic u= 5.33 (0.0500 FWE)
Design: Multisubject, 1 scan per subj: 12(subj)
Perms: 4096 permutations of conditions, bhPerms=1
Pseudo-t: Variance smoothed with FWHM [8x8x8] mm

Degrees of freedom = n/a (variance smoothing)
Search vol: 1392120 cmm, 51560 voxels
Voxel size: [3.00, 3.00, 3.00] mm

Where can I find these values?

P values & statistics: .

cluster-level	voxel-level				x,y,z mm		
k	$p_{FWE\text{-}corr}$	$p_{FDR\text{-}corr}$	Pseudo-t	$p_{uncorrected}$			
2386	**0.0002**	**0.0023**	**14.37**	0.0002	**42**	**-48**	**-30**
	0.0002	0.0023	11.51	0.0002	45	-60	-27
	0.0002	0.0023	11.33	0.0002	-33	-60	-24
356	**0.0002**	**0.0023**	**10.57**	0.0002	**3**	**15**	**45**
367	**0.0002**	**0.0023**	**10.07**	0.0002	**12**	**-15**	**3**

Useful links

Useful links:

- The Human Brain Function book, chapters 14 and 15
- Website Introduction to SPM statistics & Thresholding with Random Field Theory (Matthew Brett, MRC - CBU)
- Website Image processing (computer vision) (David Jacobs, UMD - CS)
- Slides and images of this presentation are available on my website

I thank you for your attention!