Edinburgh Imaging
www.ed.ac.uk/edinburgh-imaging

Biennial SPM course 2017

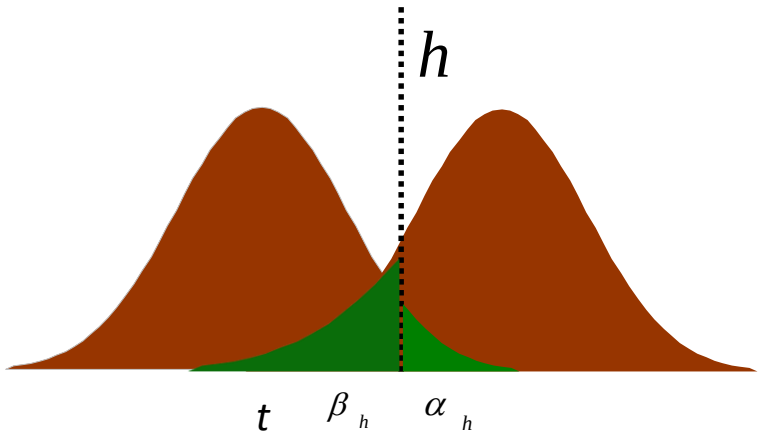
Statistical inferences in fMRI

Cyril Pernet, PhD
University of Edinburgh

Overview

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

Error at a single voxel



Decision:
H_0 , H_1: zero/non-zero activation

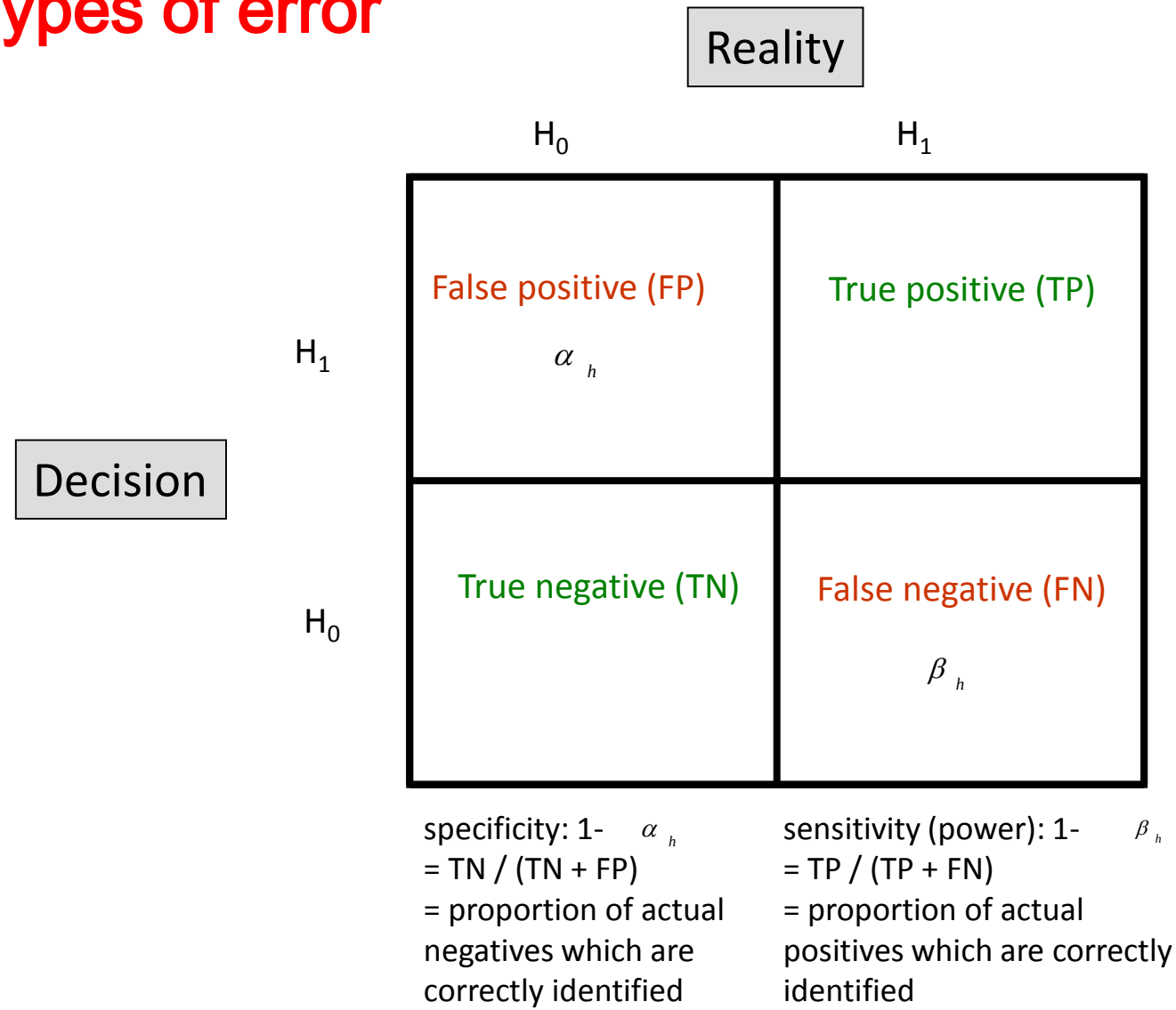
Decision rule (threshold) *h*,
determines related error rates α_h β_h

Convention: Penalize complexity
Choose *h* to give acceptable α_h under H_0

$$t = \frac{\textit{contrast} \text{ of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

J Chumbley's slide 2015

Types of error



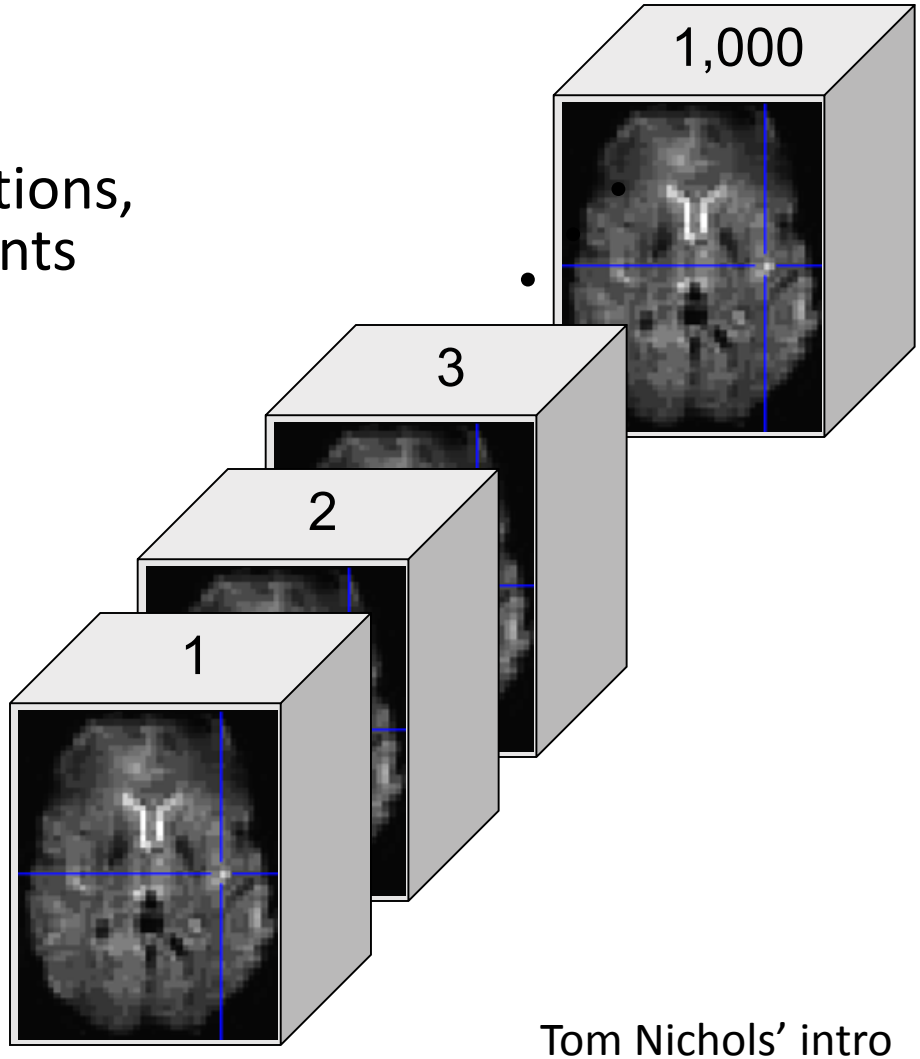
J Chumbley's slide 2015

Multiple comparisons correction

Avoiding false positives

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.

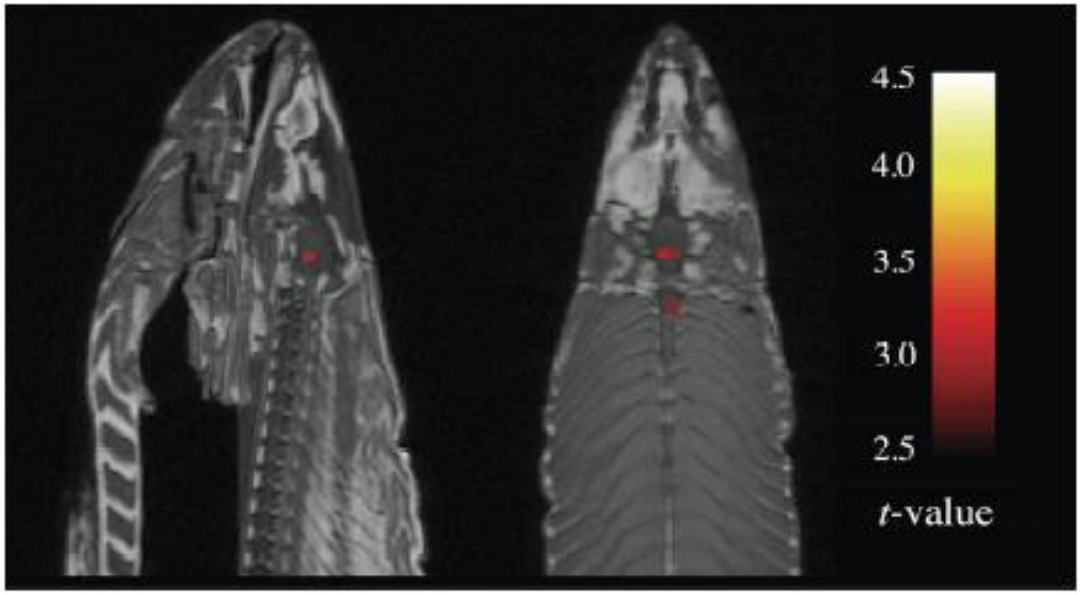
- Using extend threshold k without correction is not enough as it, by chance, can cluster as well.

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!

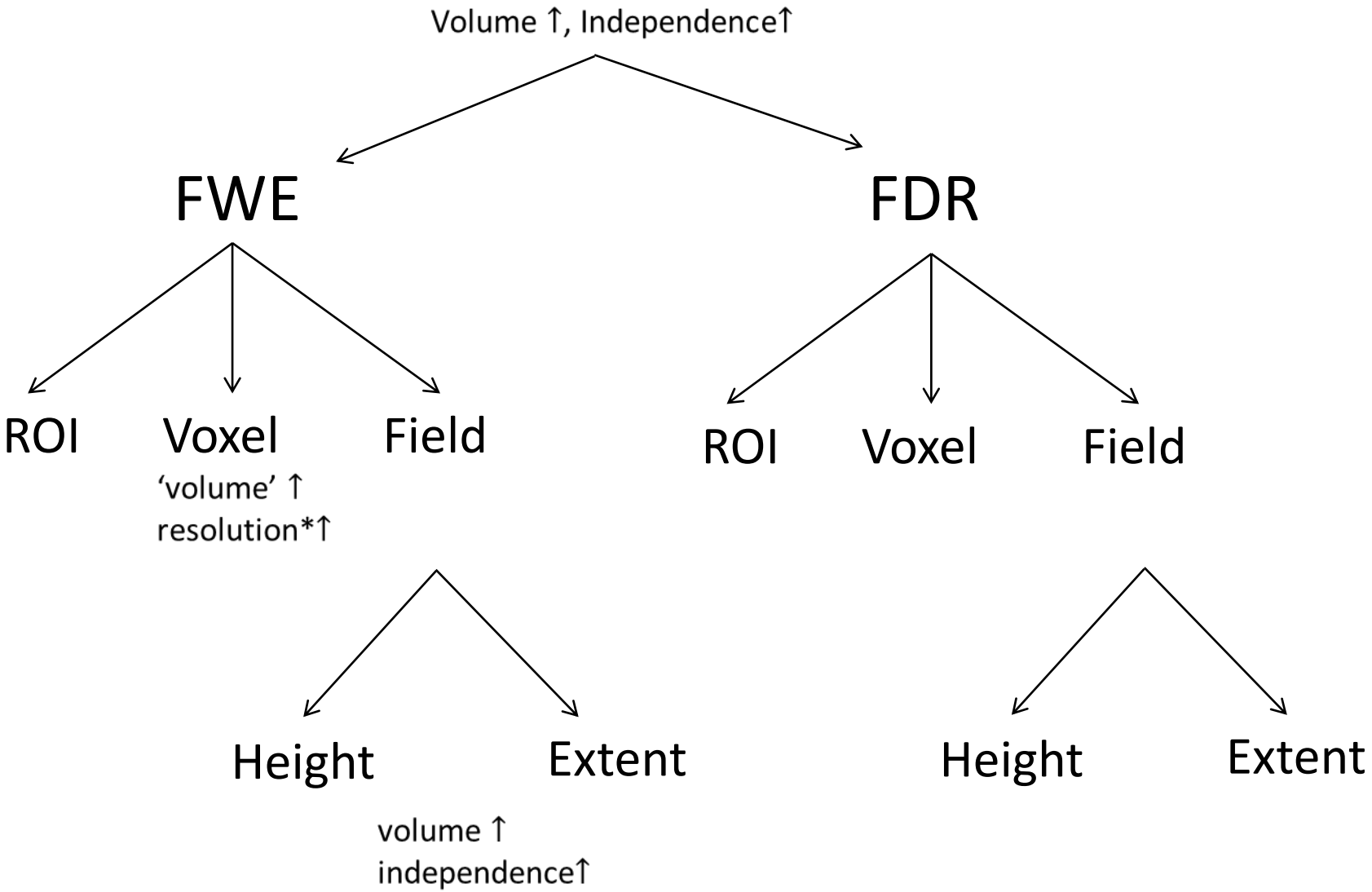
What Problem?



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(uncorrected) < 0.001, 3 voxel extent threshold.

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

Detect an effect of *unknown* extent & location



J Chumbley's slide 2015

Bonferroni Correction

FWER is the prob. that any stats > u, is a FP
FWER is therefore also the prob. that the max stats > u is a FP

Bonferroni correction allows to keep the FWER at 5% by simply dividing alpha by the number of tests

$$P(T_i \geq u|H0) \leq \frac{\propto}{m}$$ *Find u to keep the FWER < α/m*

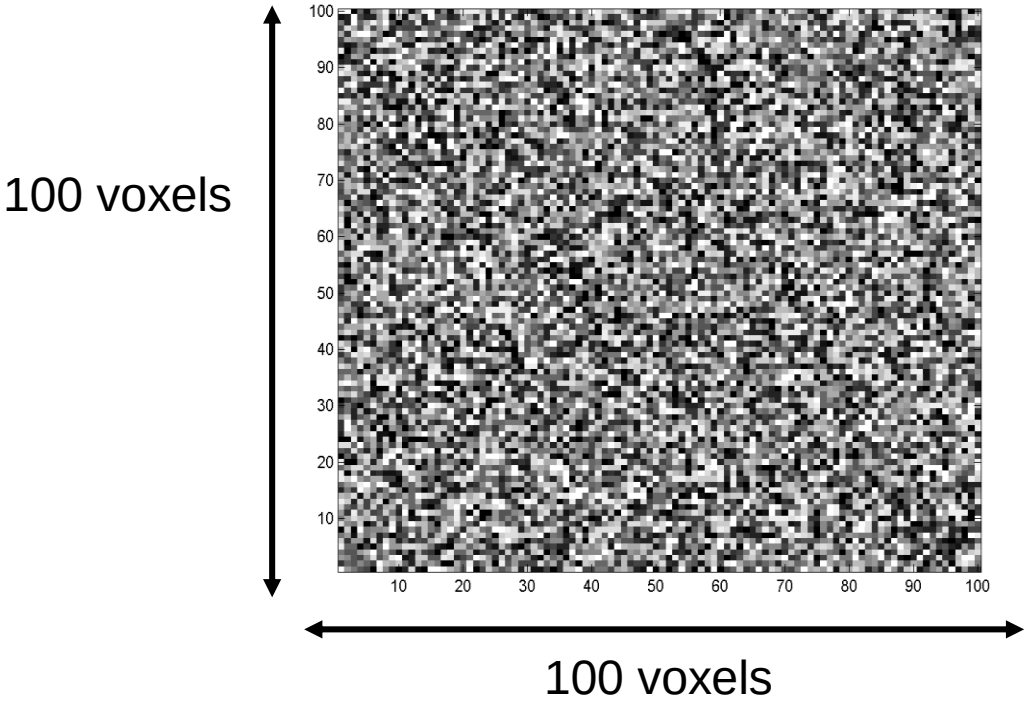
$$\text{FWER} = P(\bigcup_{i \in V}\{T_i \geq u\}|H_0) \leq \propto$$

$$\leq \sum P(T_i \geq u|H0)$$ *Boole's inequality*

$$\leq \sum_i \frac{\propto}{m} = \propto$$

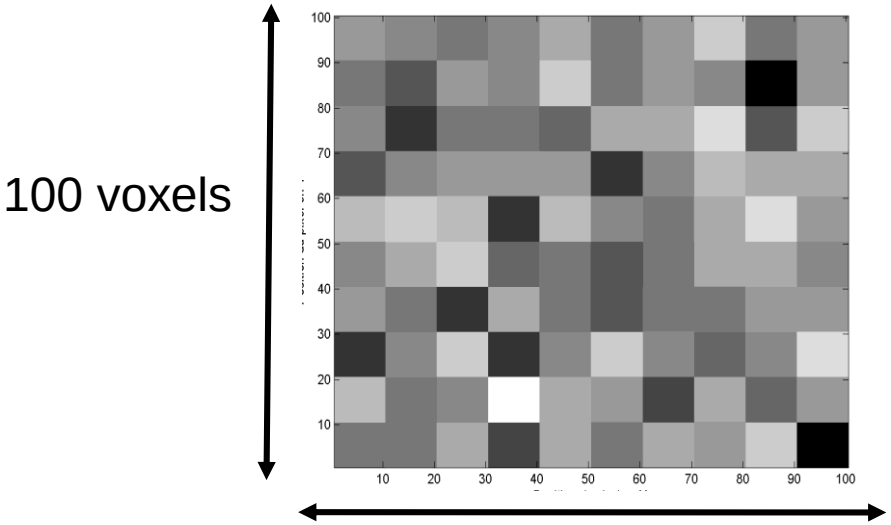
Bonferroni Correction

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



Bonferroni Correction

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



Solutions for MCP

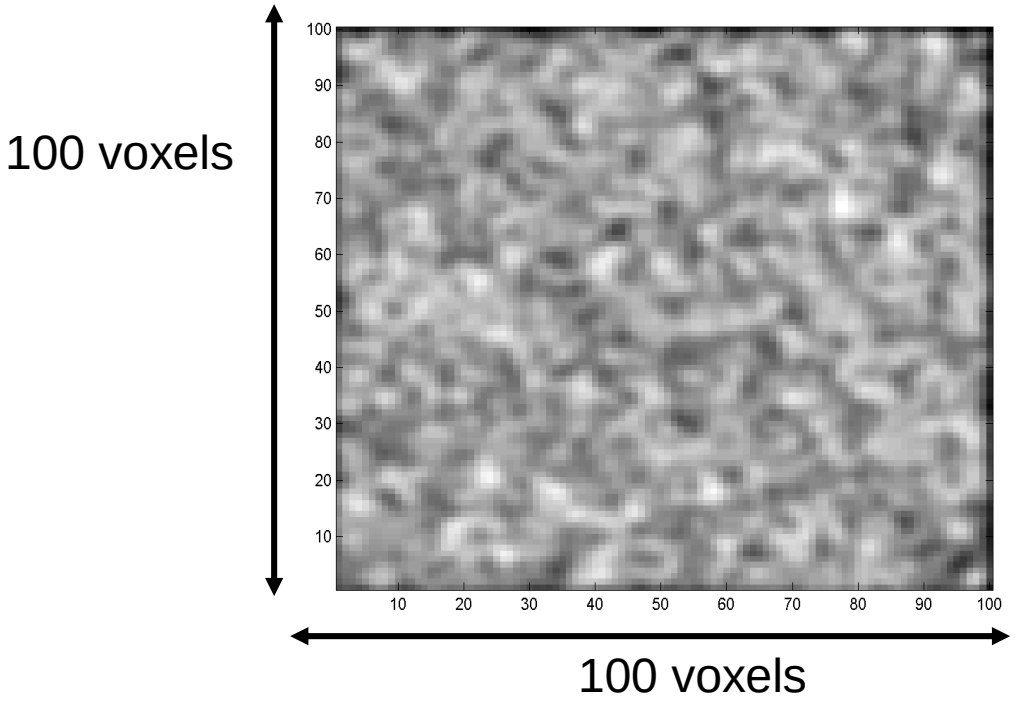
- An important feature of neuroimaging data is that we have a family of stat values that has topological features (Bonferroni for instance consider tests as independent)
- Why considering data as a smooth lattice? (Chumbley et al., 2009 NeuroImage 44)

➢ 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)

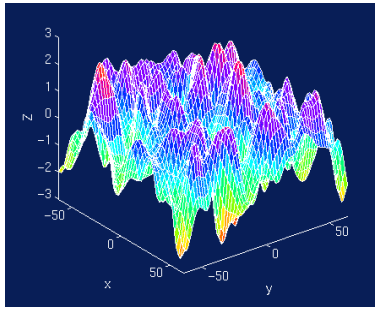
➢ Neural activity propagate locally through intrinsic/lateral connections and is distributed via extrinsic connections / Hemodynamic correlates are initiated by diffusing signals (e.g. NO)

Random Field Theory

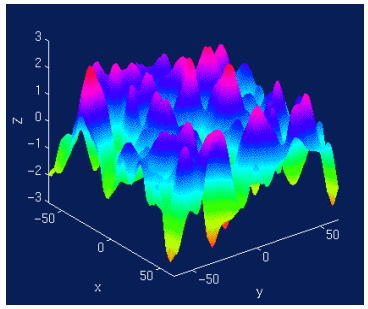
- 10000 Z-scores ; alpha = 5%
- Gaussian kernel smoothing –
- How many independent observations ?



RFT



“lattice approximation”

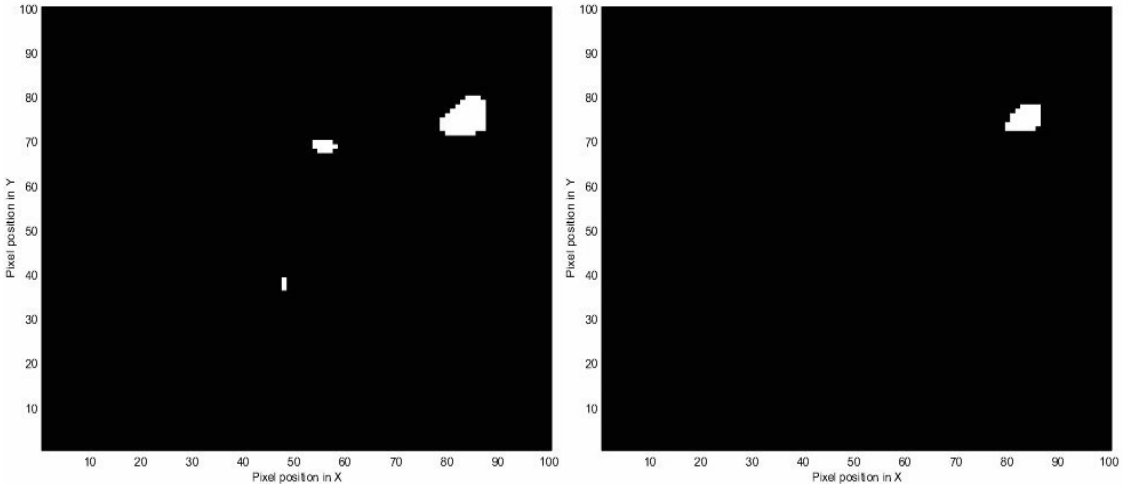


- RFT relies on theoretical results for smooth statistical maps (hence the need for smoothing), allowing to find a threshold in a set of data where it’s not easy to find the number of independent variables. Uses the expected Euler characteristic (EC density)

- 1 Estimation of the smoothness = number of resel (resolution element) = f(nb voxels, FWHM)
- 2 expected Euler characteristic = number of clusters above the threshold
- 3 Calculation of the threshold

Random Field Theory

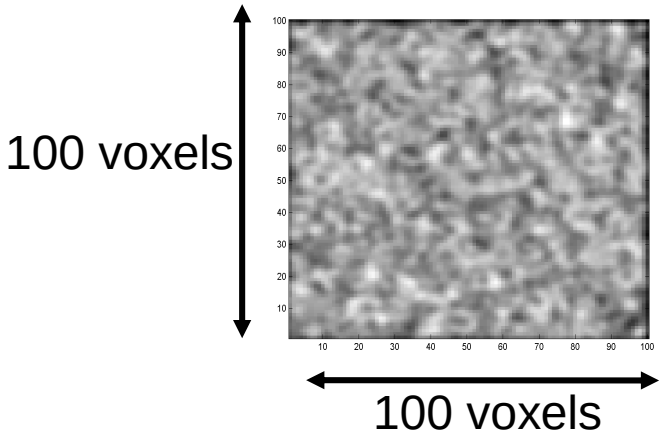
- The Euler characteristic can be seen as the number of blobs in an image after thresholding (p value that you select in SPM)
- At high threshold, EC = 0 or 1 per resel: $E[EC] \approx p^{FWE}$



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



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

- 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)

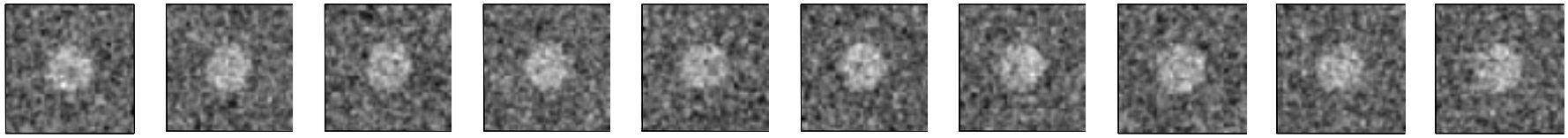
False Discovery Rate

- Whereas family wise approach corrects for any false positive, the FDR approach aim at correcting among positive results only.

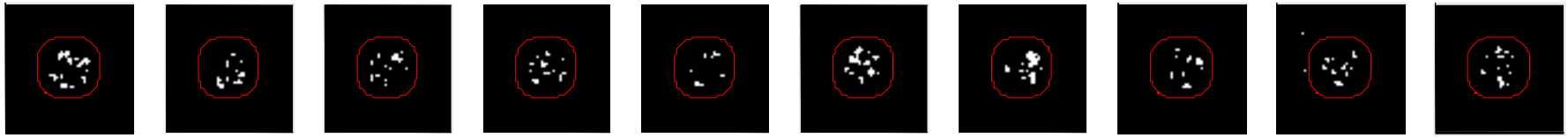
1. Run an analysis with alpha = x%
2. Sort the resulting positive data
3. Threshold to remove the false positives

False Discovery Rate

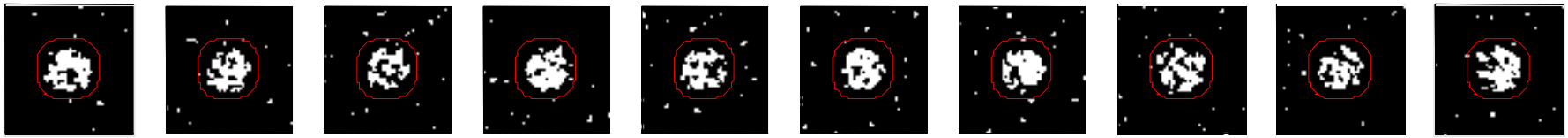
Signal+Noise



FEW correction



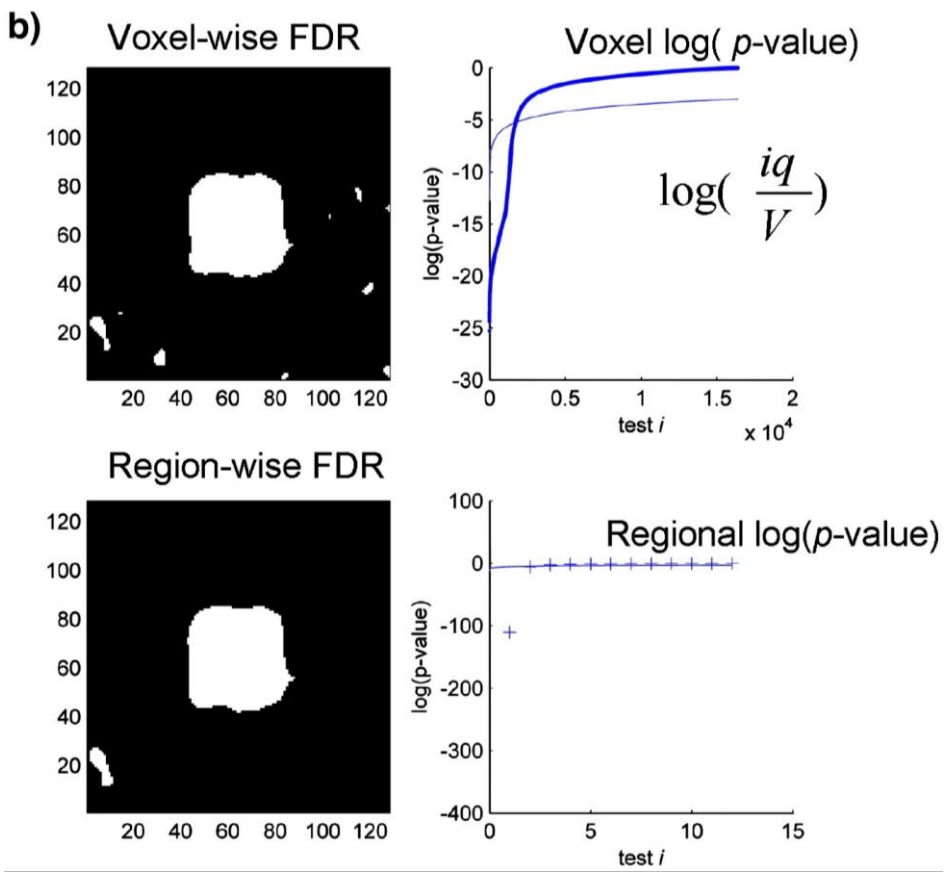
FDR correction



False Discovery Rate for clusters

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

Voxel, cluster and set

Levels of inference

- 3 levels of inference can be considered:

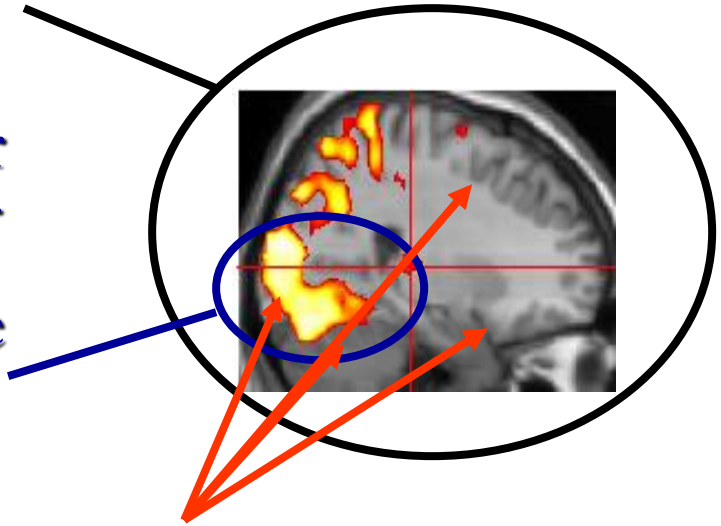
- Voxel level (prob associated at each voxel)
- Cluster level (prob associated to a set of voxels)
- Set level (prob associated to a set of clusters)

- 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 H0 for an omnibus test, i.e. there are some significant clusters of activation in the brain.

- Cluster level: we can reject H0 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 H0 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

Uncorrected **(**bad**)**

Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	$(Z_{=})$	p_{uncorr}	
0.000	22	0.000	0.000	2519	0.000	0.000	0.000	19.07	Inf	0.000	-36 -18 -10
						0.000	0.000	13.63	Inf	0.000	-36 -54 -12
						0.000	0.000	12.28	Inf	0.000	-46 -50 -2
		0.000	0.000	2340	0.000	0.000	0.000	18.16	Inf	0.000	38 -44 -26
						0.000	0.000	17.50	Inf	0.000	40 -16 -20
						0.000	0.000	13.48	Inf	0.000	14 -56 -26
		0.000	0.000	285	0.000	0.000	0.000	12.54	Inf	0.000	-16 -40 50
		0.000	0.000	276	0.000	0.000	0.000	10.91	Inf	0.000	44 -50 44
						0.000	0.000	7.93	Inf	0.000	38 -38 44
						0.000	0.000	6.76	6.71	0.000	48 -58 56
		0.000	0.000	218	0.000	0.000	0.000	10.67	Inf	0.000	-22 10 66
						0.000	0.000	7.84	7.76	0.000	-34 22 54
						0.004	0.097	5.47	5.45	0.000	-28 18 62
		0.000	0.000	76	0.000	0.000	0.000	10.18	Inf	0.000	12 10 76
		0.000	0.000	248	0.000	0.000	0.000	9.23	Inf	0.000	38 -52 14
						0.000	0.000	7.65	7.58	0.000	34 -54 22
						0.000	0.000	7.52	7.46	0.000	38 -60 6
		0.000	0.000	172	0.000	0.000	0.000	8.03	Inf	0.000	-20 -56 28
						0.000	0.000	7.99	Inf	0.000	-22 -58 20
		0.000	0.000	66	0.000	0.000	0.000	7.27	7.21	0.000	-8 -58 44
						0.000	0.000	6.55	6.51	0.000	-14 -62 38
		0.000	0.000	76	0.000	0.000	0.000	7.01	6.96	0.000	-36 -20 58
		0.000	0.000	40	0.000	0.000	0.000	6.70	6.66	0.000	44 2 58
						0.000	0.000	6.46	6.42	0.000	54 2 50
		0.000	0.000	92	0.000	0.000	0.000	6.63	6.58	0.000	60 -30 48
		0.000	0.000	68	0.000	0.000	0.000	6.54	6.50	0.000	46 -18 60
						0.001	0.025	5.73	5.71	0.000	50 -8 56
		0.000	0.001	16	0.001	0.000	0.004	6.08	6.05	0.000	46 -44 4
		0.000	0.000	23	0.000	0.000	0.011	5.90	5.87	0.000	48 -34 0

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 4.96, p = 0.000 (0.050)
Extent threshold: k = 0 voxels
Expected voxels per cluster, <k> = 1.102
Expected number of clusters, <c> = 0.05
FWEp: 4.959, FDRp: 5.735, FWEc: 1, FDRc: 5

Degrees of freedom = [1.0, 1728.0]
FWHM = 6.6 7.1 6.7 mm mm mm; 3.3 3.6 3.3 {voxels}
Volume: 1155888 = 144486 voxels = 3383.6 resels
Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 39.21 voxels)
Page 1

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

Using p=.001 this creates an excursion set with topology that satisfies RFT → FDR clusters size and height

Circularity issues in fMRI

Definition

- 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., 2009 *Nat. Neuroscience* 12

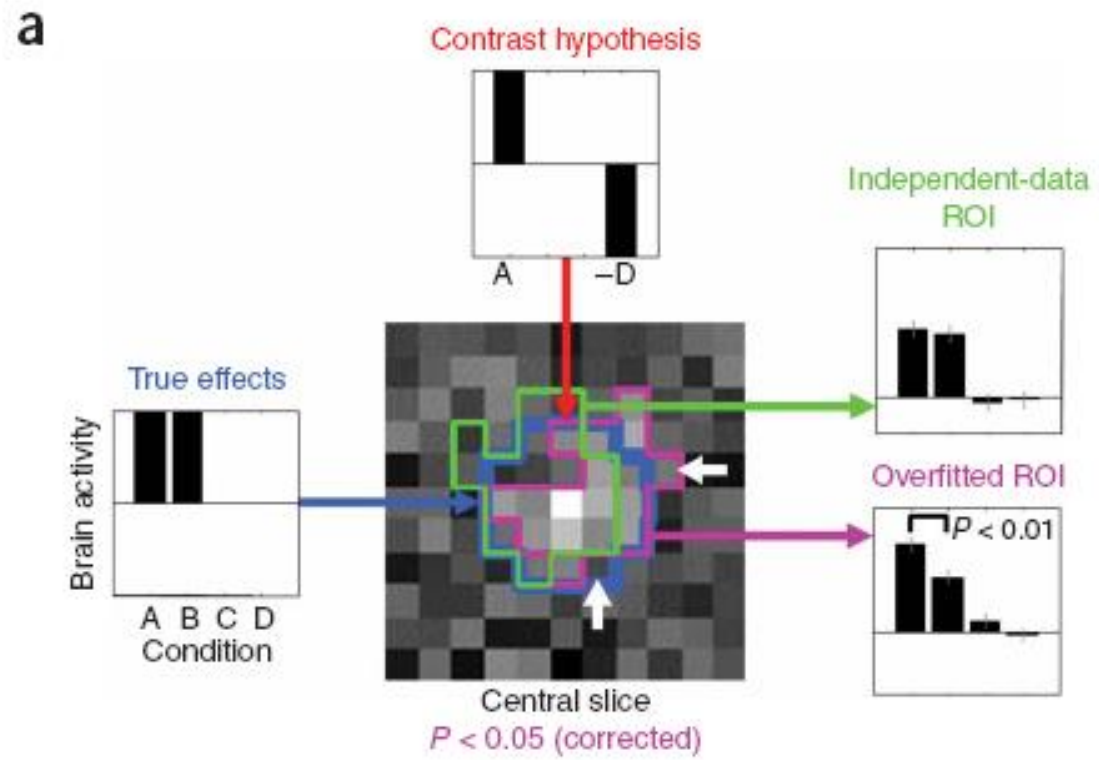
Circularity

- 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 a gp of subjects and measures RTs, then take 2 subgroups 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 $C_{selection}{}^{T}(X^{T}X)^{-1}C_{test}$)
3. Select using a subset of data, test with another one

Questions ?