# Introduction à la statistique médicale

# Statistical Parametric Mapping short course

## Course 4:

## Multiple comparison problem & levels of inference

Christophe Phillips, Ir PhD
GIGA – CRC *In Vivo* Imaging &
GIGA – *In Silico* Medicine

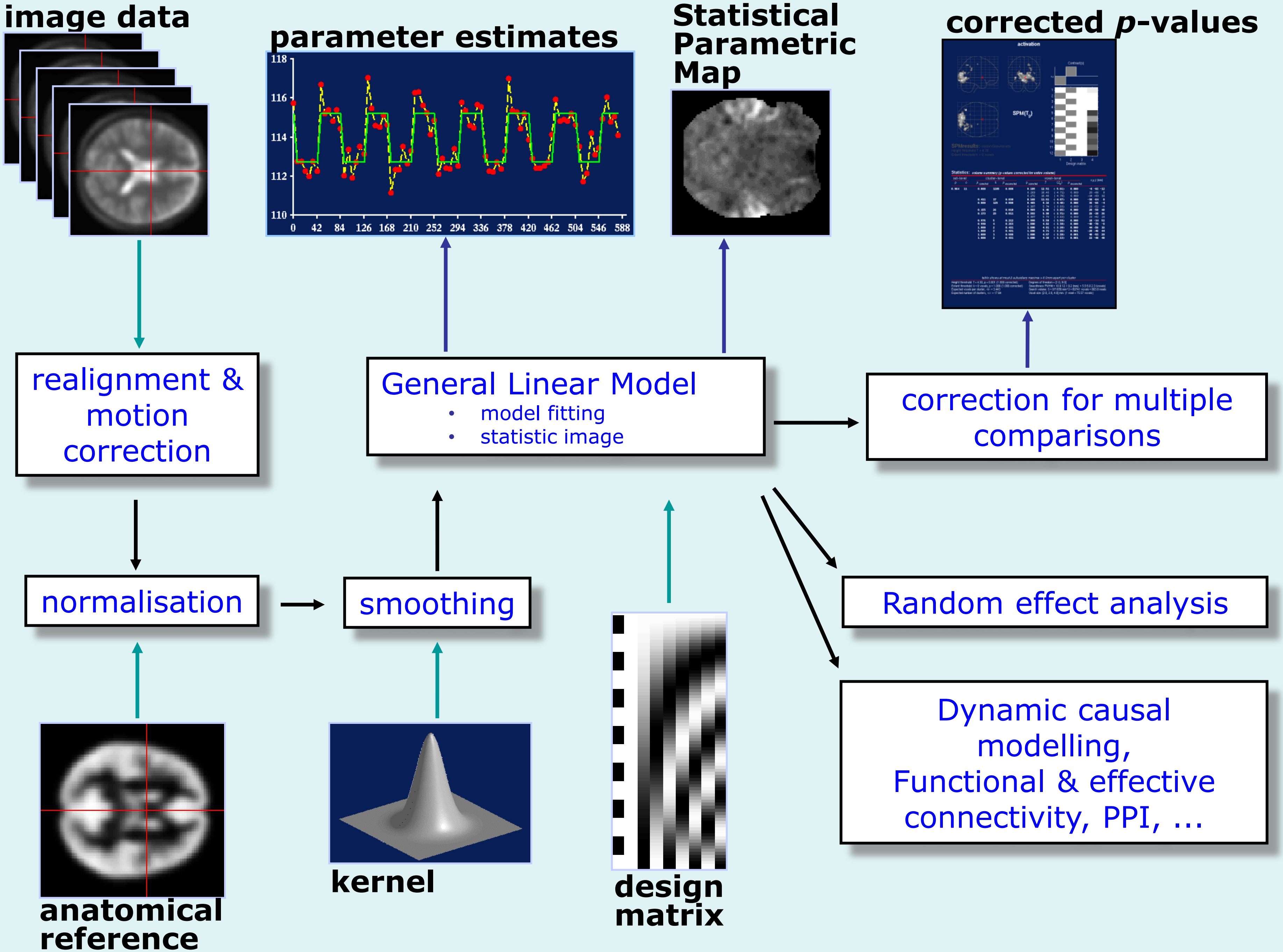
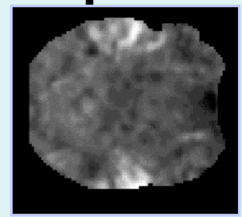
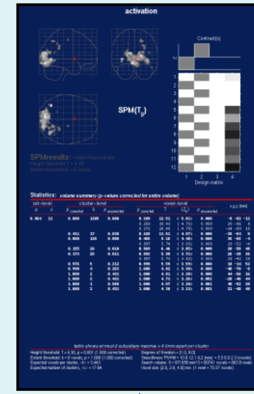


image data
parameter estimates
118
116
114
112
110
0 42 84 126 168 210 252 294 336 378 420 462 504 546 588
Statistical Parametric Map
corrected p-values
realignment & motion correction
General Linear Model
• model fitting
• statistic image
correction for multiple comparisons
normalisation
smoothing
Random effect analysis
Dynamic causal modelling, Functional & effective connectivity, PPI, ...
anatomical reference
kernel
design matrix

**Statistical Parametric Map**



**corrected *p*-values**

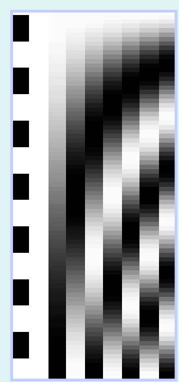


## General Linear Model

- model fitting
- statistic image

correction for multiple comparisons

Random effect analysis



**design matrix**

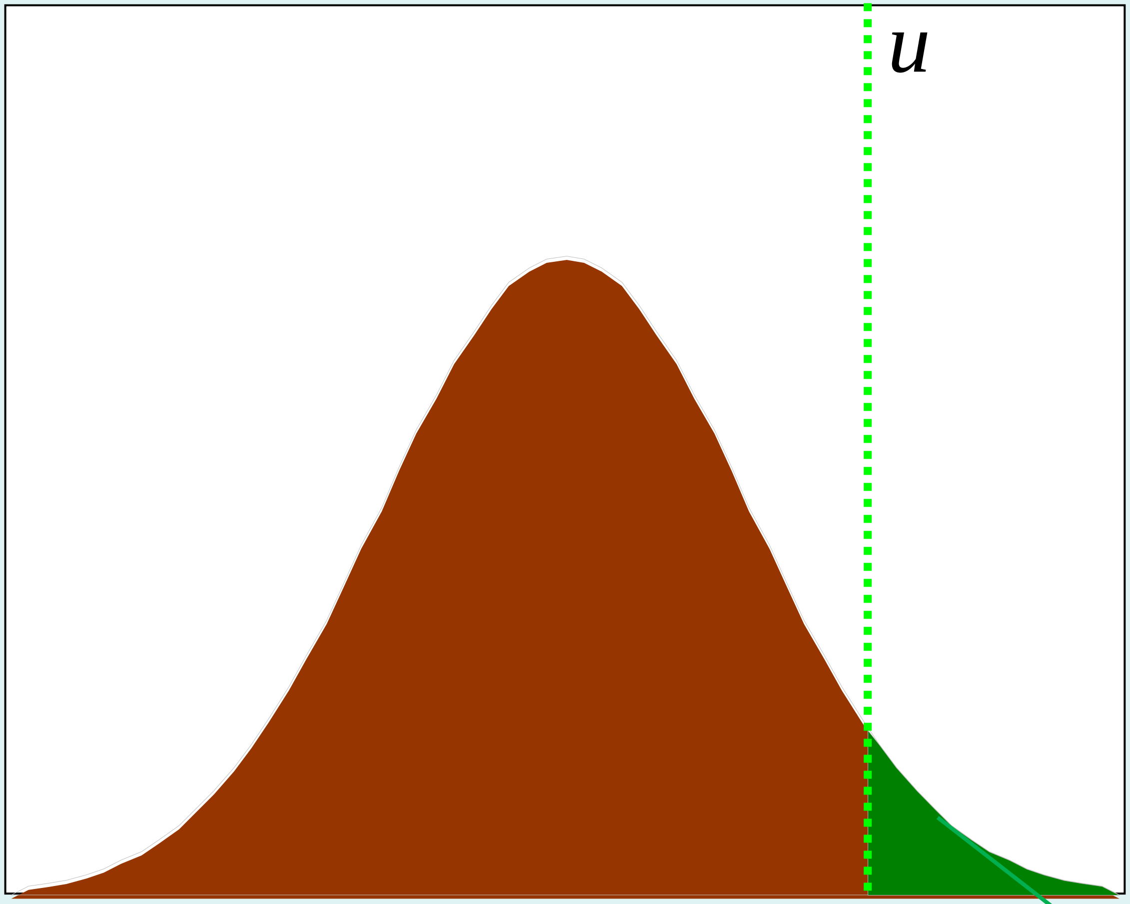
# Content

- **Introduction**
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- **Levels of inference in SPM**
- **Non-parametric permutation test**
- **Conclusion**

# Content

- **Introduction**
  - **Single voxel inference**
  - **Multiple comparison problem**
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- **Levels of inference in SPM**
- **Non-parametric permutation test**
- **Conclusion**

# Single voxel inference




Null distribution of test statistic T

**Null Hypothesis $H_0$**:
zero activation

Decision rule (*threshold*) $u$:
determines false positive rate $\alpha$

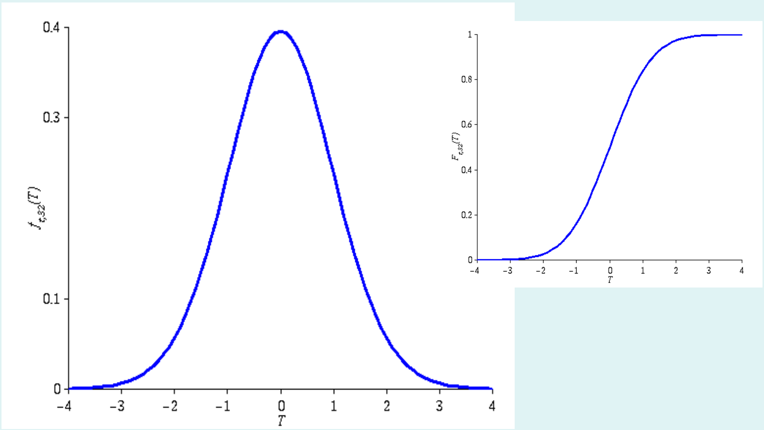
$\Rightarrow$ Choose $u$ to give acceptable $\alpha$ under $H_0$

$$\boldsymbol{\alpha = p(t > u | H_0)}$$

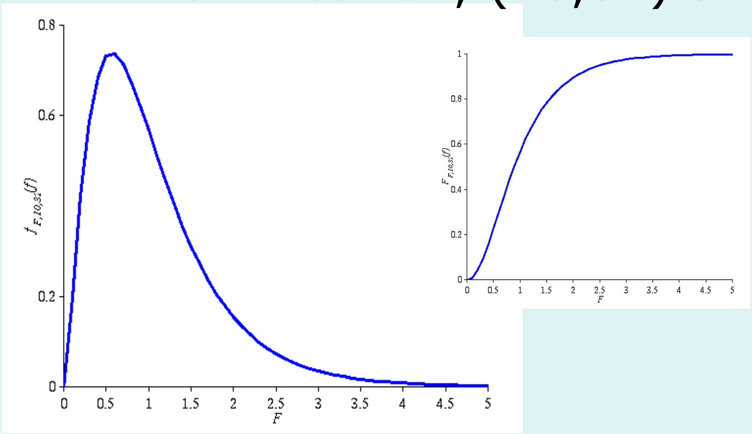
# Classical hypothesis testing…

- Null hypothesis *H*
  - test statistic
  - null distributions
- Hypothesis test
  - control Type I error
    - incorrectly reject *H*
  - test *level* $\alpha$
    - Pr("reject" $H \mid H) \leq \alpha$
- $p$ –value
  - min $\alpha$ at which *H* rejected
  - $\Pr(T \geq t \mid H)$
  - characterising *"surprise"*

$t$ –distribution, 32 df.




$F$ –distribution, (10,32) df.



# Sensitivity & specificity

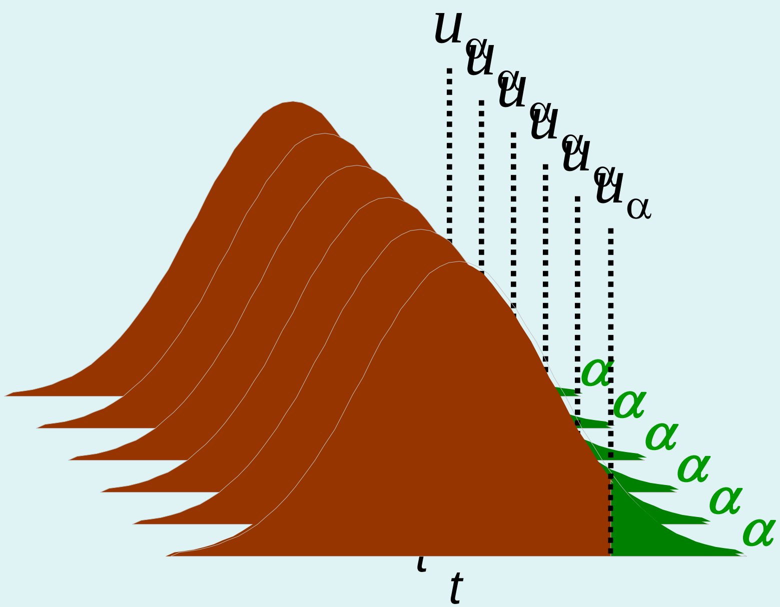
**ACTION**

| **TRUTH** | | Don't reject | Reject |
|---|---|---|---|
| | $H_0$ true | True Negative | False Positive |
| | $H_0$ false | False Negative | True Positive |

Sensitivity = TP/(TP+FN) = $\beta$
Specificity = TN/(TN+FP) = $1 - \alpha$

FP = Type I error or 'error'
FN = Type II error
$\alpha$ = p-value/FP rate/error rate/significance level
$\beta$ = power

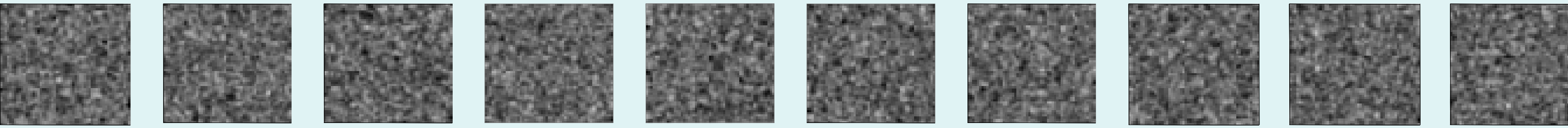
# Multiple tests



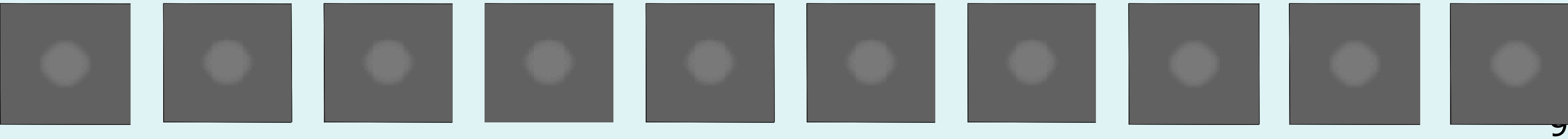

If we have 100000 voxels, $\alpha$=0.05
$\Rightarrow$ **5000 false positive voxels**.

This is clearly undesirable!
Need to define a *null hypothesis* for a *collection of tests*.

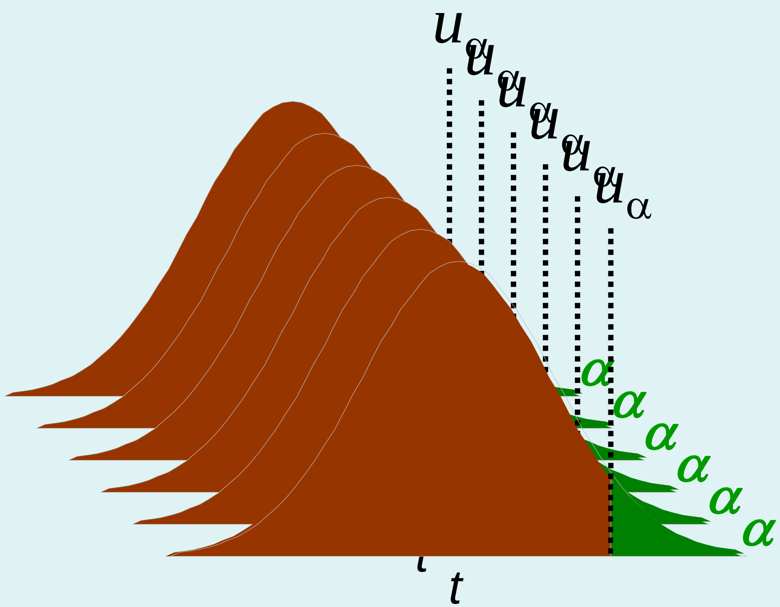
Noise



Signal



# Multiple tests

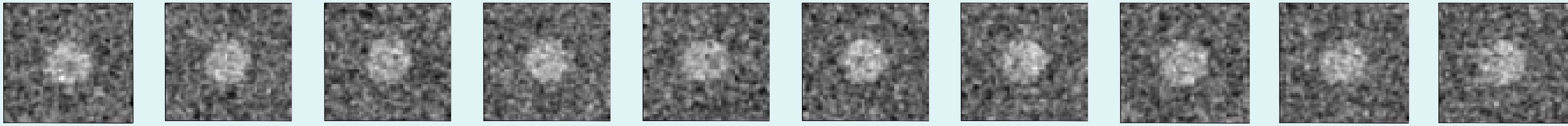



If we have 100000 voxels, $\alpha$=0.05

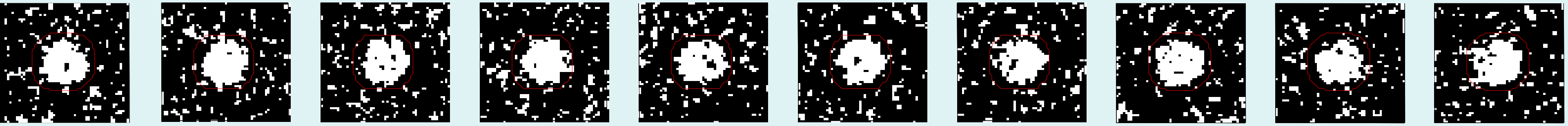
$\Rightarrow$ **5000 false positive voxels**.

This is clearly undesirable!
Need to define a *null hypothesis* for a *collection of tests*.

Noisy data



Use of 'uncorrected' *p*-value, $\alpha$ =0.1

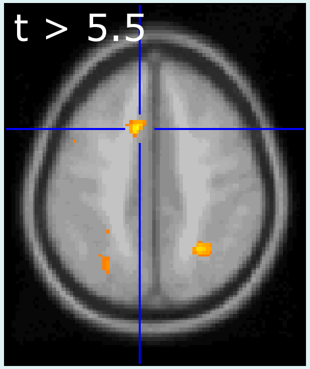
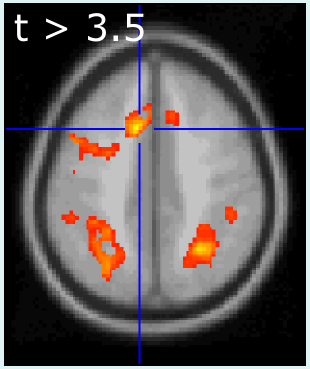
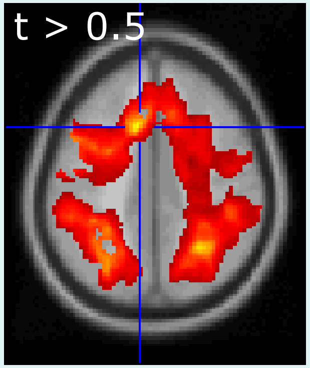


| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 11.3% | 11.3% | 12.5% | 10.8% | 11.5% | 10.0% | 10.7% | 11.2% | 10.2% | 9.5% |

*Percentage of Null Pixels that are False Positives*

# Assessing statistics images

Where's the signal?

| High Threshold | Med. Threshold | Low Threshold |
|---|---|---|
|  |  |  |
| Good Specificity<br><br>Poor Power<br>(risk of false negatives) | | Poor Specificity<br>(risk of false positives)<br><br>Good Power |

# Content

- Introduction
- **Family-wise error rate (FWER)**
  - **Family-wise Null hypothesis**
  - **Bonferroni correction**
  - **Random Field Theory**
- False discovery rate (FDR)
- Levels of inference in SPM
- Non-parametric permutation test
- Conclusion

# Family-Wise Null Hypothesis

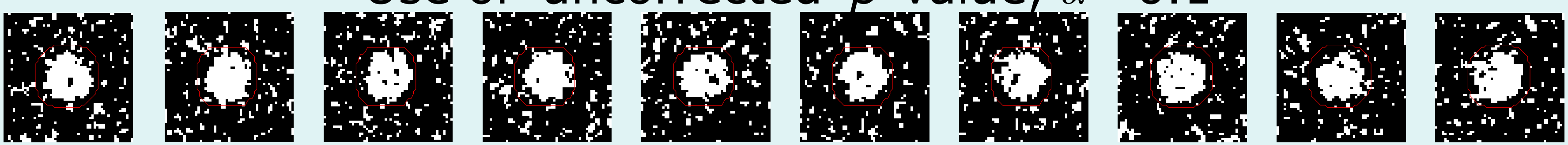
***Family-Wise Null Hypothesis:***
*Activation is zero everywhere*

If we reject a voxel null hypothesis at *any* voxel, we reject the family-wise Null hypothesis

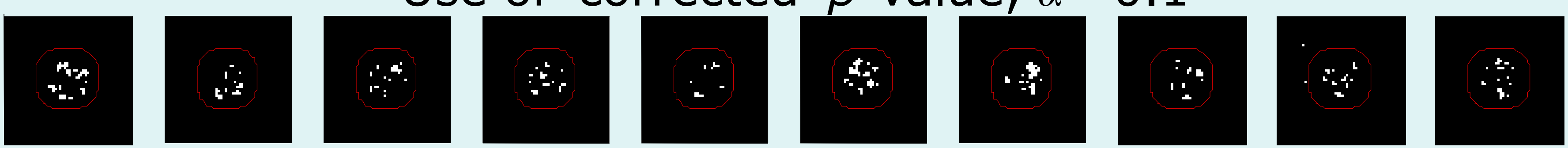
A FP ***anywhere*** in the image gives a **Family Wise Error** (FWE)

Family-Wise Error rate (FWER) = '*corrected*' *p*-value

Use of 'uncorrected' *p*-value, $\alpha$ =0.1



Use of 'corrected' *p*-value, $\alpha$ =0.1



FWE

# Bonferroni correction

The Family-Wise Error rate (FWER), $\alpha_{FWE}$, for a family of *N* tests follows the inequality:

$$\alpha_{FWE} \leq N\alpha$$

where $\alpha$ is the test-wise error rate.

Therefore, to ensure a particular FWER choose:

$$\alpha = \frac{\alpha_{FWE}}{N}$$

This correction does not require the tests to be independent but becomes very stringent if dependence.

# Bonferroni correction, example

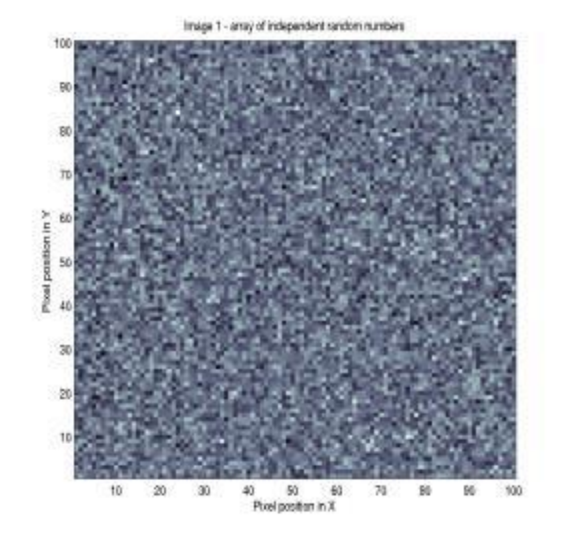
- Experiment with N = 100000 *independent* voxels and 40 d.f.
  - v = unknown corrected probability threshold,
  - find v such that family-wise error rate $\alpha = 0.05$
- **Bonferroni correction**:
  - probability that all tests are below the threshold,
  - use $v = \alpha / N$
  - here v=0.05/100000=0.0000005

    $\Rightarrow$ threshold $t = 5.77$
- **Interpretation**:

  Bonferroni procedure gives a corrected p-value,

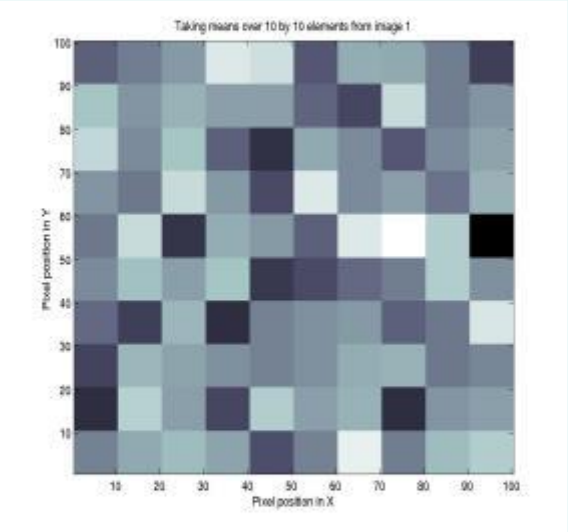
  i.e. for a *t* statistics = 5.77,
  - uncorrectd *p* value = 0.0000005
  - corrected *p* value = 0.05

# Bonferroni & independent observations




100 by 100 voxels.
**10000** independent measures
Fix the $P^{FWE} = 0.05$, $z$ threshold ?

Bonferroni:
$v = 0.05 / 10000 = 0.000005$
$\Rightarrow$ threshold $z = 4.42$

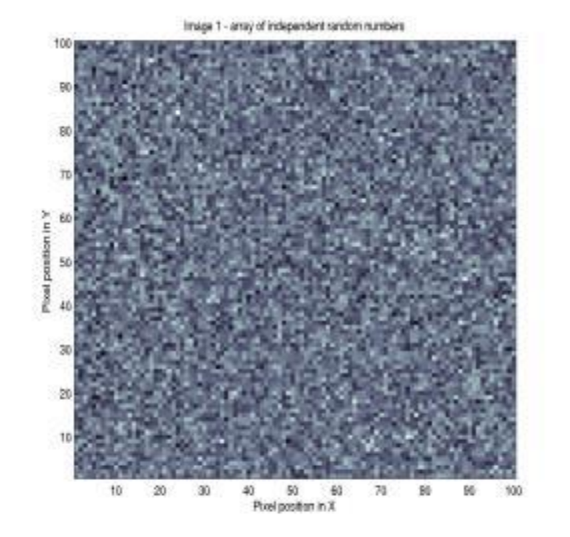


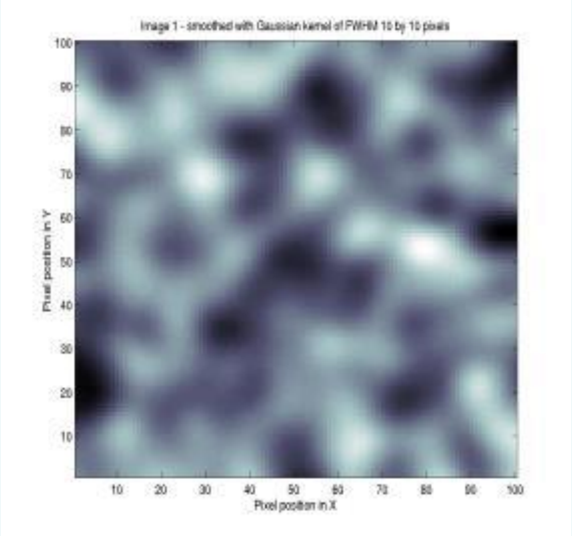

100 by 100 voxels.
**100** independent measures
Fix the $P^{FWE} = 0.05$, $z$ threshold ?

Bonferroni:
$v = 0.05 / 100 = 0.0005$
$\Rightarrow$ threshold $z = 3.29$

$v=\alpha/n_i$ where $n_i$ is the number of independent observations.

# Bonferroni & independent observations




100 by 100 voxels.
**10000** independent measures
Fix the $P^{FWE} = 0.05$, $z$ threshold ?

Bonferroni:

$v = 0.05 / 10000 = 0.000005$

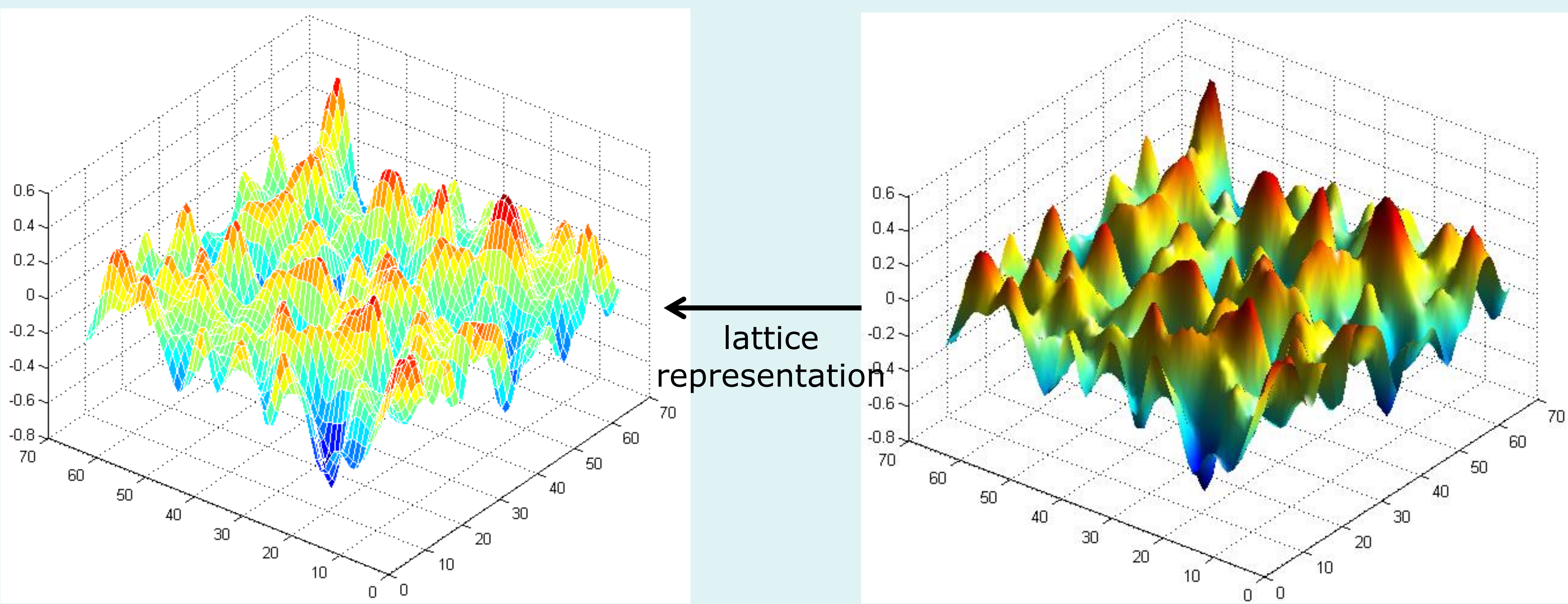
$\Rightarrow$ threshold $z = 4.42$

100 by 100 voxels.
**How many independent measures ???**

# Random Field Theory

⇒ Consider a statistic image as a discretisation of a continuous underlying random field.

⇒ Use results from continuous **random field theory**.



# RFT and Euler Characteristic

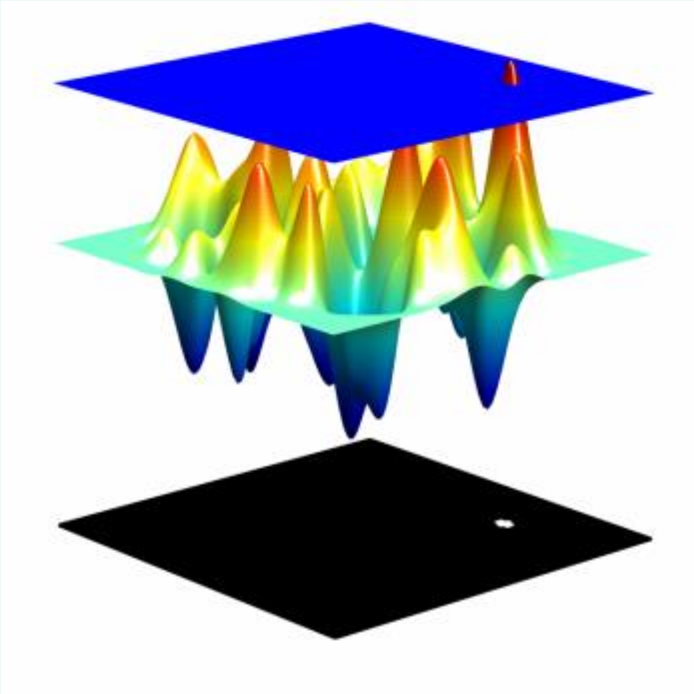
**Euler Characteristic** $\chi_u$:

- Topological measure

  $\chi_u$ = # blobs - # holes

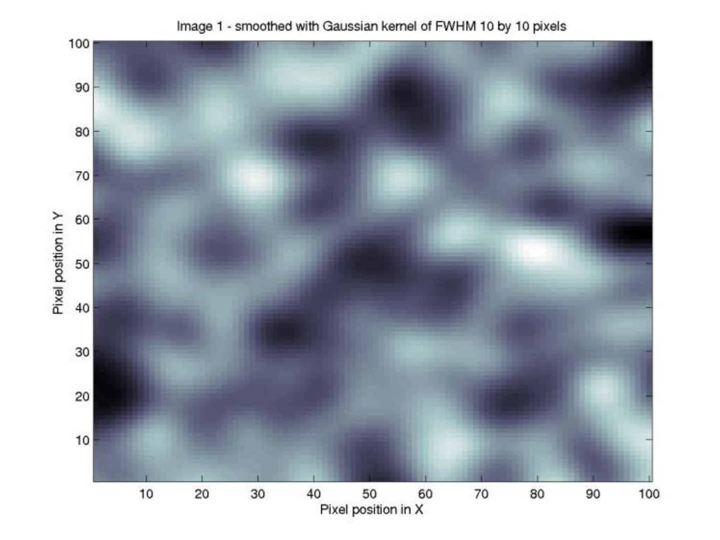
- at high threshold $u$:

  $\chi_u$ = # blobs



$$FWER = p(FWE) \approx E[\chi_u]$$

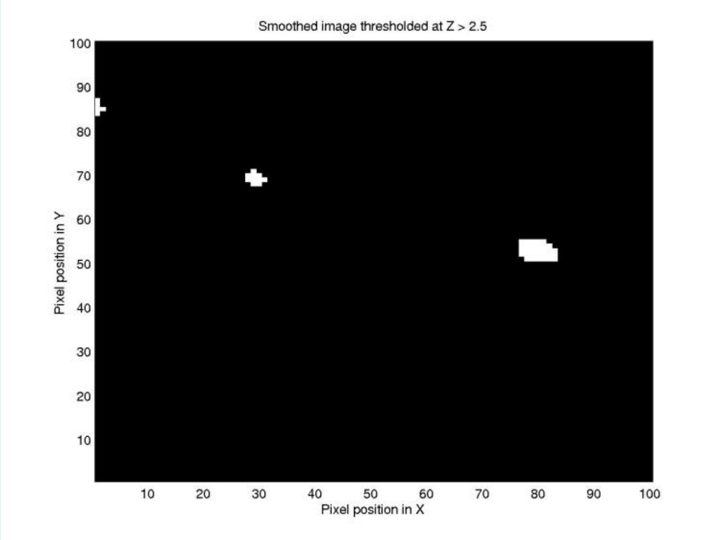
# Euler characteristic…



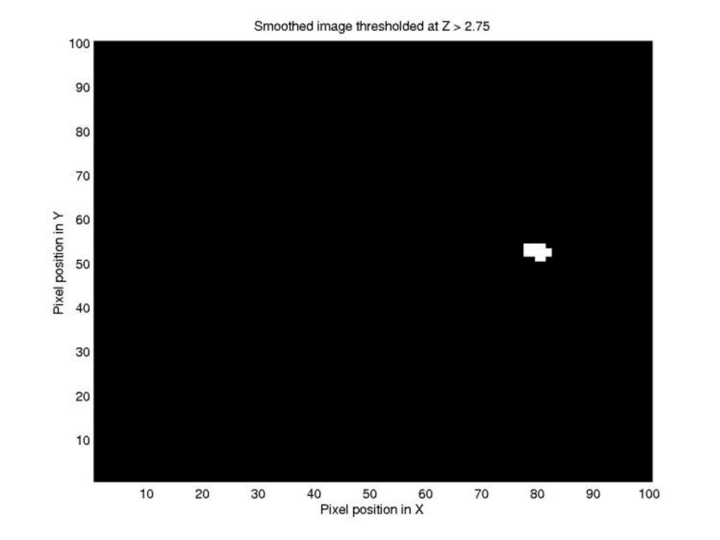

Threshold $z$-map at 2.50

Threshold $z$-map at 2.75

EC = 3




EC = 1



# Expected Euler Characteristic

*2D Gaussian Random Field*

$$E[\chi_u] = \lambda(\Omega)|\Lambda|^{1/2}\, u \exp(-u^2/2)/(2\pi)^{3/2}$$

Search volume

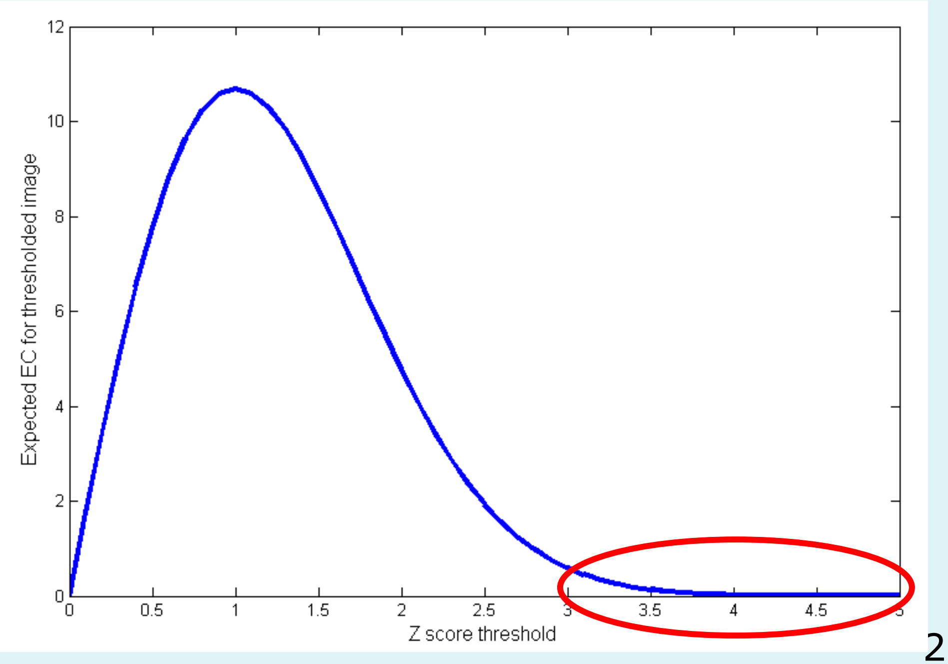
Roughness (1/smoothness)

Threshold

100 x 100 Gaussian Random Field with FWHM=10 smoothing

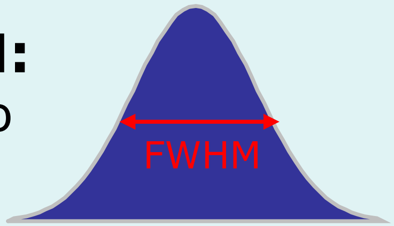
$\alpha_{FWE} = 0.05 \Rightarrow u_{RFT} = 3.8$

$(u_{BONF} = 4.42, u_{uncorr} = 1.64)$

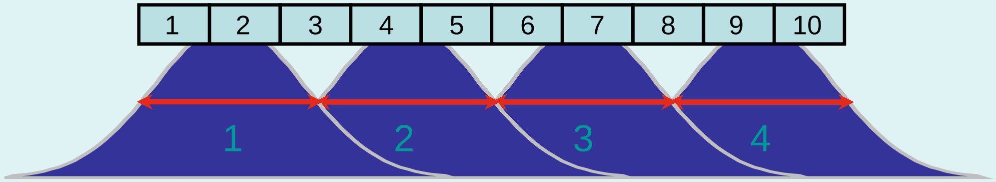


# Smoothness

**Smoothness parameterised in terms of FWHM:**
Size of Gaussian kernel required to smooth i.i.d. noise to have same smoothness as observed null (standardized) data.



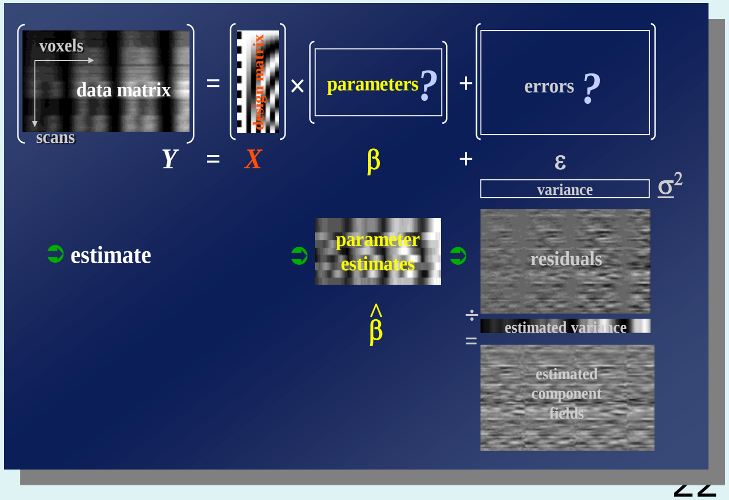

**RESELS (Resolution Elements):**
1 RESEL = $FWHM_x FWHM_y FWHM_z$
RESEL Count $R$ = volume of search region in units of smoothness




Eg: 10 voxels, 2.5 FWHM, 4 RESELS

The number of resels is similar, but not identical to the number independent observations.

**Smoothness estimated from spatial derivatives of standardised residuals:**
Yields an RPV image containing local roughness estimation.



# RFT intuition

Corrected *p*-value for statistic value *t*

$$\begin{aligned} p_c &= p(\max T > t) \\ &\approx E[\chi_t] \\ &\propto \lambda(\Omega)|\Lambda|^{1/2}\, t \exp(-t^2/2) \end{aligned}$$

- Statistic value *t* increases ?
  - $p_c$ decreases (better signal)
- Search volume increases ( $\lambda(\Omega)\uparrow$ ) ?
  - $p_c$ increases (more severe correction)
- Smoothness increases ( $|\Lambda|^{1/2}\downarrow$ ) ?
  - $p_c$ decreases (less severe correction)

# RFT, unified theory

## General form for expected Euler characteristic

• *t*, *F* & $\chi^2$ fields • restricted search regions • *D* dimensions •

$$E[\chi_u(\Omega)] = \sum_{d=0}^{D} R_d(\Omega)\rho_d(u)$$

$R_d(\Omega)$: *d*-dimensional Lipschitz-Killing curvatures of $\Omega$ ($\approx$ *intrinsic volumes*):

– *function of dimension,* space $\Omega$ and smoothness:

$R_0(\Omega) = \chi(\Omega)$ Euler characteristic of $\Omega$
$R_1(\Omega)$ = resel diameter
$R_2(\Omega)$ = resel surface area
$R_3(\Omega)$ = resel volume

$\rho_d(u)$ : *d*-dimensional EC density of the field

– *function of dimension and threshold, specific for RF type:*

E.g. Gaussian RF:

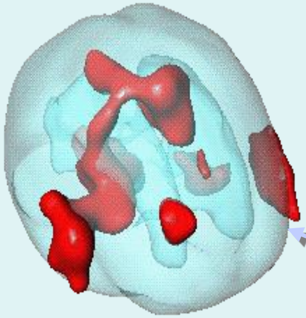
$\rho_0(u) = 1 - \Phi(u)$

$\rho_1(u) = (4 \ln 2)^{1/2} \exp(-u^2/2) / (2\pi)$

$\rho_2(u) = (4 \ln 2) \; u \; \exp(-u^2/2) / (2\pi)^{3/2}$

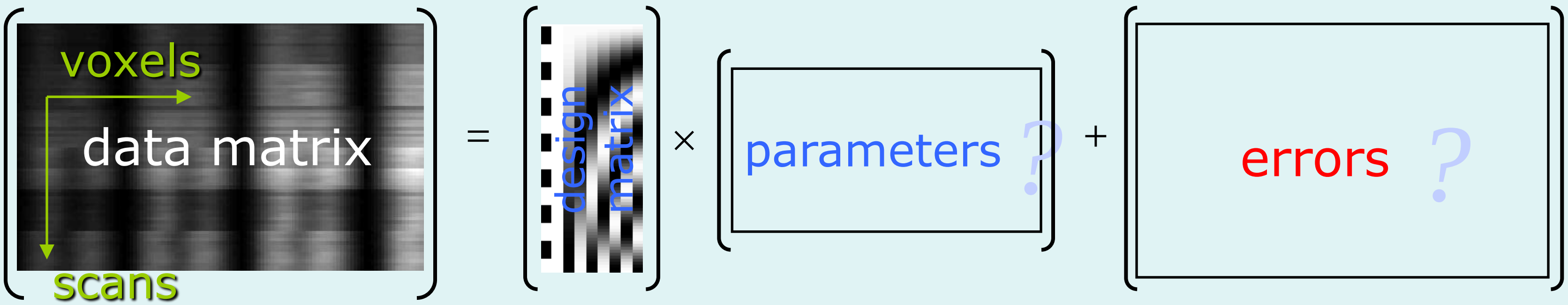
$\rho_3(u) = (4 \ln 2)^{3/2} (u^2 - 1) \exp(-u^2/2) / (2\pi)^2$

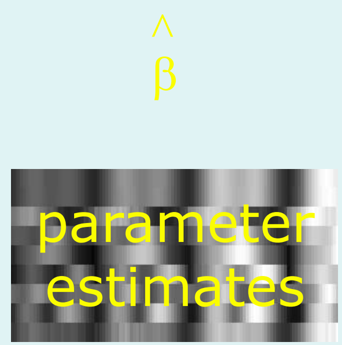
$\rho_4(u) = (4 \ln 2)^2 (u^3 - 3u) \exp(-u^2/2) / (2\pi)^{5/2}$



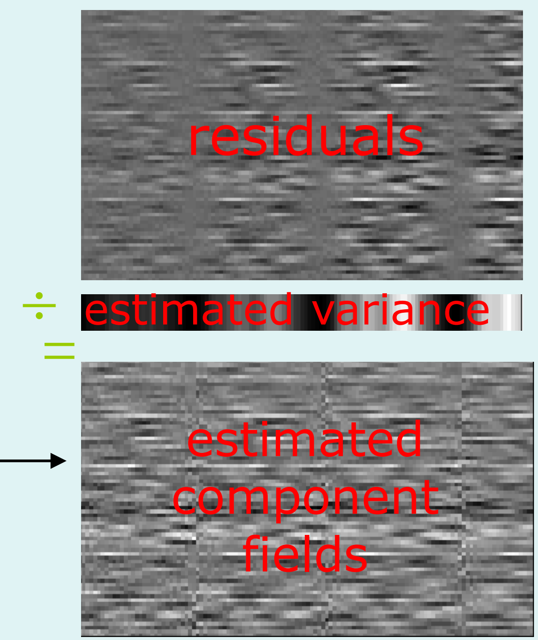
$\Omega$

# Estimated component fields




estimate



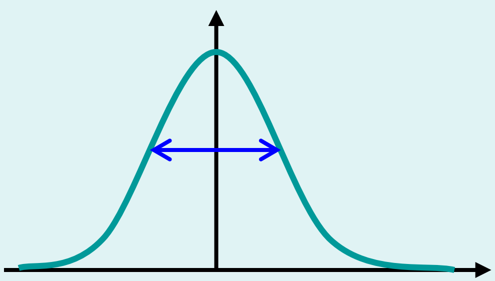

*Each row is*
*an estimated*
*component field*

# Smoothness, PRF, ResEls...

- Smoothness $\sqrt{|\Lambda|}$
  - variance-covariance matrix of partial derivatives *(possibly location dependent)*

$$\Lambda = \begin{pmatrix} \mathrm{var}\left[\frac{\partial e}{\partial x}\right] & \mathrm{cov}\left[\frac{\partial e}{\partial x}, \frac{\partial e}{\partial y}\right] & \mathrm{cov}\left[\frac{\partial e}{\partial x}, \frac{\partial e}{\partial z}\right] \\ \mathrm{cov}\left[\frac{\partial e}{\partial x}, \frac{\partial e}{\partial y}\right] & \mathrm{var}\left[\frac{\partial e}{\partial y}\right] & \mathrm{cov}\left[\frac{\partial e}{\partial y}, \frac{\partial e}{\partial z}\right] \\ \mathrm{cov}\left[\frac{\partial e}{\partial x}, \frac{\partial e}{\partial z}\right] & \mathrm{cov}\left[\frac{\partial e}{\partial y}, \frac{\partial e}{\partial z}\right] & \mathrm{var}\left[\frac{\partial e}{\partial z}\right] \end{pmatrix}$$

- Point Response Function PRF



- Full Width at Half Maximum FWHM. *Approximate the peak of the Covariance function with a Gaussian*

- Gaussian PRF
  - $\Sigma$ – kernel var/cov matrix
  - ACF $2\Sigma$
  - $\Lambda = (2\Sigma)^{-1}$

  $\Rightarrow$FWHM $f = \sigma\sqrt{(8\ln(2))}$
  - $\Sigma = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & f_z \end{bmatrix} \frac{1}{8\ln(2)}$

  *ignoring covariances*

  $\Rightarrow \sqrt{|\Lambda|} = (4\ln(2))^{3/2} / (f_x \times f_y \times f_z)$

- **Res**olution **El**ement (**ResEl**)
  - Resel dimensions $(f_x \times f_y \times f_z)$
  - $R_3(\Omega) = \lambda(\Omega) / (f_x \times f_y \times f_z)$

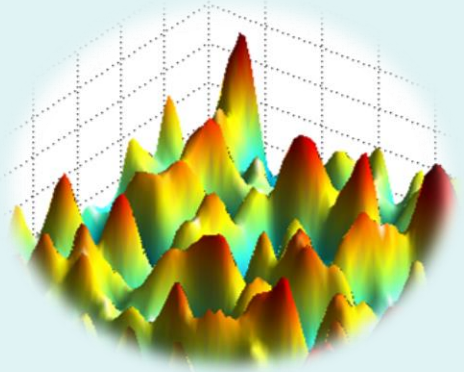
  *if strictly stationary*

$E[\chi(A_u)] = R_3(\Omega)\ (4\ln(2))^{3/2}\ (u^2 - 1)\ \exp(-u^2/2) / (2\pi)^2$

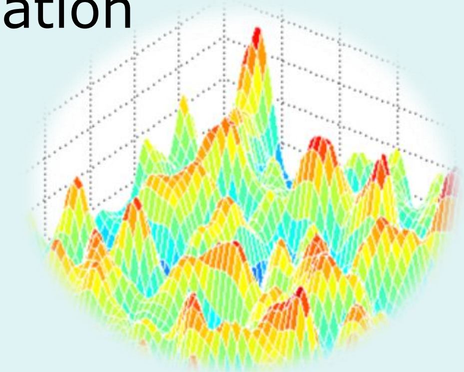
$\approx R_3(\Omega)\ (1 - \Phi(u))$

*for high thresholds u*

# RFT assumptions



- The statistic image is assumed to be a good lattice representation of an underlying random field with a multivariate Gaussian distribution.
- These fields are continuous, with an autocorrelation function twice differentiable at the origin.



- ➢ The threshold chosen to define clusters is high enough such that the expected EC is a good approximation to the number of clusters.
- ➢ The lattice approximation is reasonable, which implies the smoothness is relatively large compared to the voxel size.
- ➢ The errors of the specified statistical model are normally distributed, which implies the model is not misspecified.

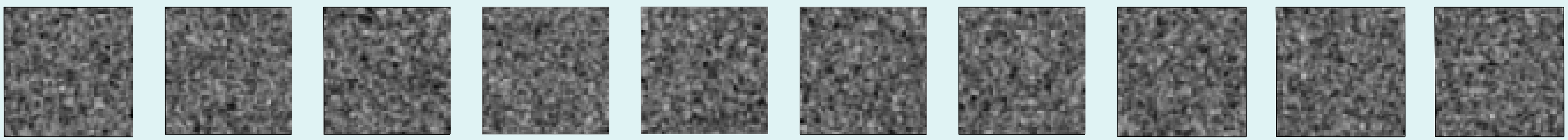
- Smoothness of the data is unknown and estimated: very precise estimate by pooling over voxels $\Rightarrow$ stationarity assumption.

# Content

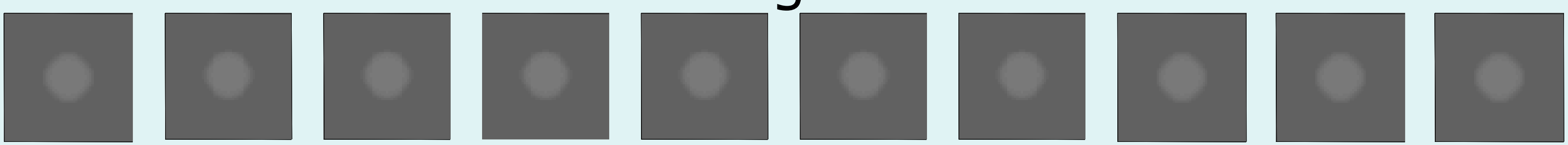
- **Introduction**
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- **Levels of inference in SPM**
- **Non-parametric permutation test**
- **Conclusion**

# FDR illustration:

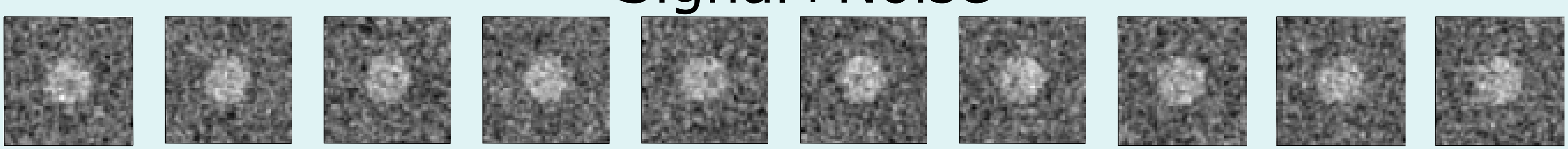
Noise



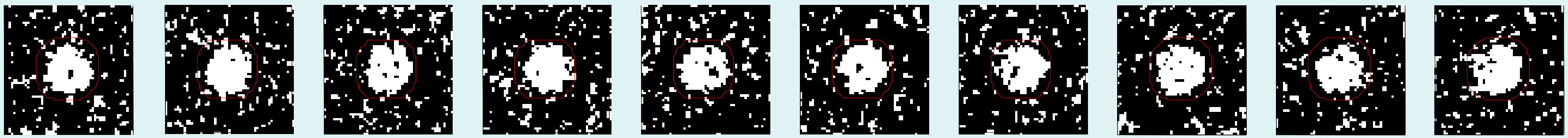
Signal



Signal+Noise



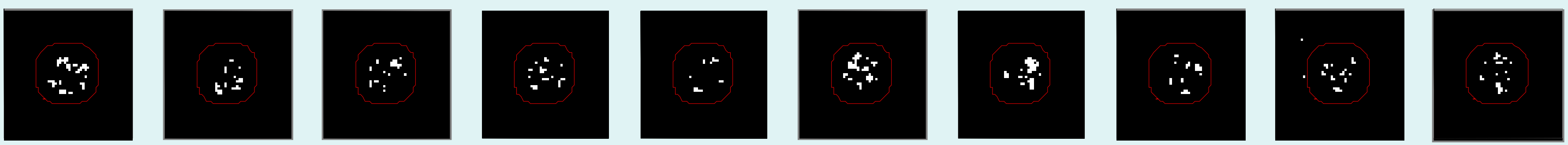
## Control of Per Comparison Rate at 10%



11.3% 11.3% 12.5% 10.8% 11.5% 10.0% 10.7% 11.2% 10.2% 9.5%

Percentage of Null Pixels that are False Positives

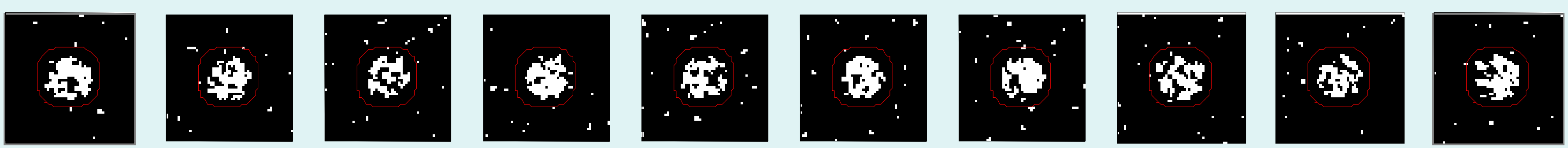
## Control of Familywise Error Rate at 10%



FWE

Occurrence of Familywise Error

## Control of False Discovery Rate at 10%



6.7% 10.4% 14.9% 9.3% 16.2% 13.8% 14.0% 10.5% 12.2% 8.7%

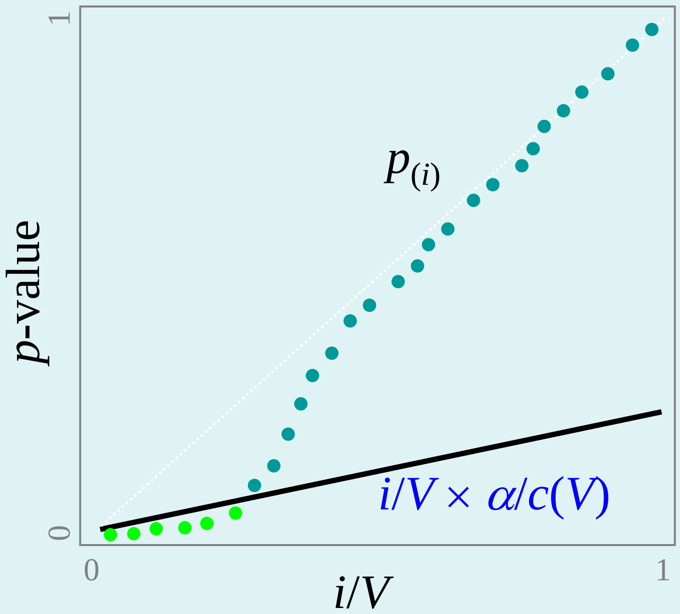
Percentage of Activated Pixels that are False Positives

# Benjamini & Hochberg Procedure

- Select desired limit $\alpha$ on E(FDR)
- Order p-values, $p_{(1)} \leq p_{(2)} \leq \ \ldots \leq p_{(V)}$
- Let $r$ be largest $i$ such that

$$p_{(i)} \leq \ i/V^{*}\alpha$$

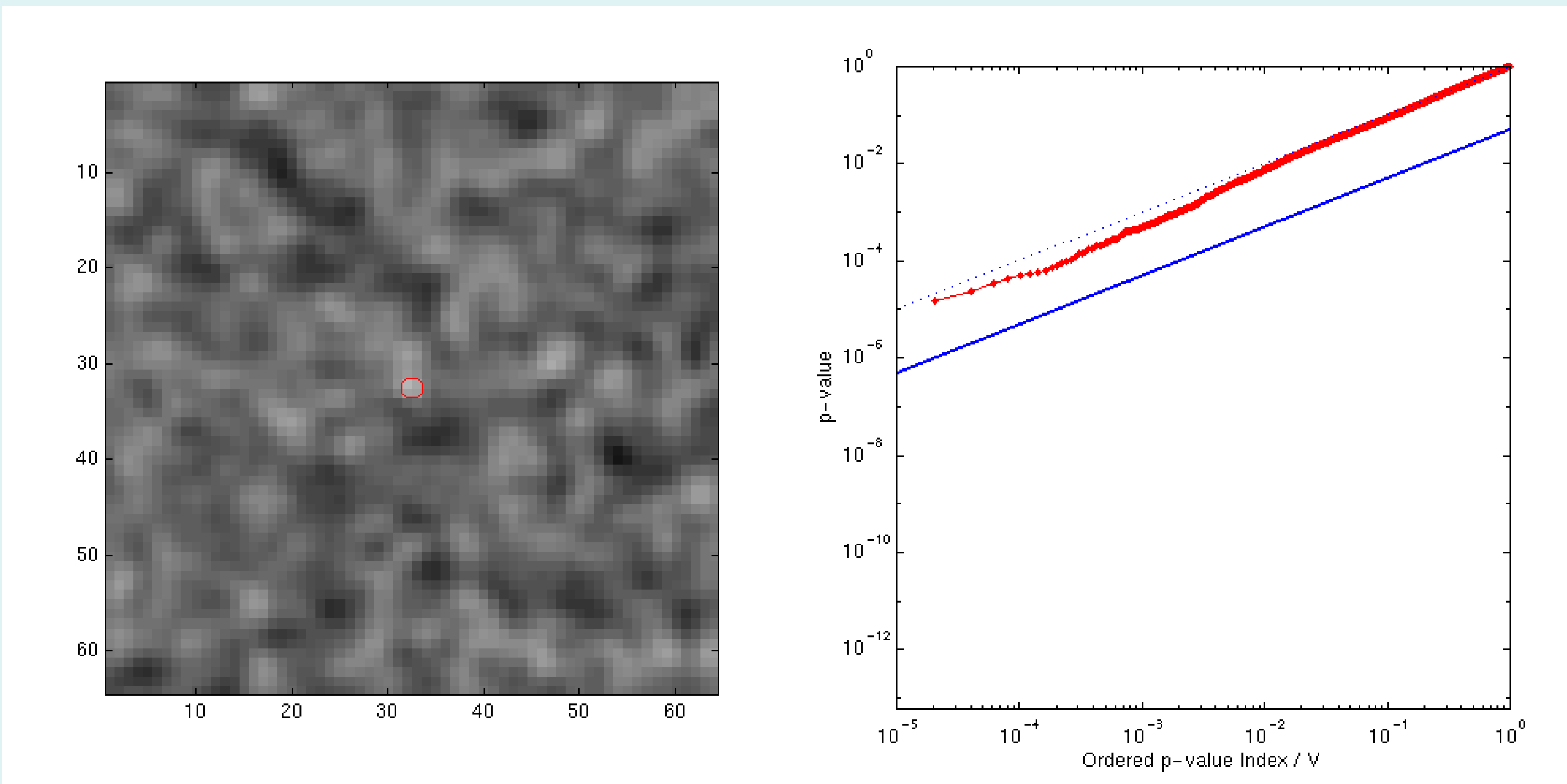
- Reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$.




*JRSS-B* (1995) 57:289-300

# B&H: Varying Signal Extent

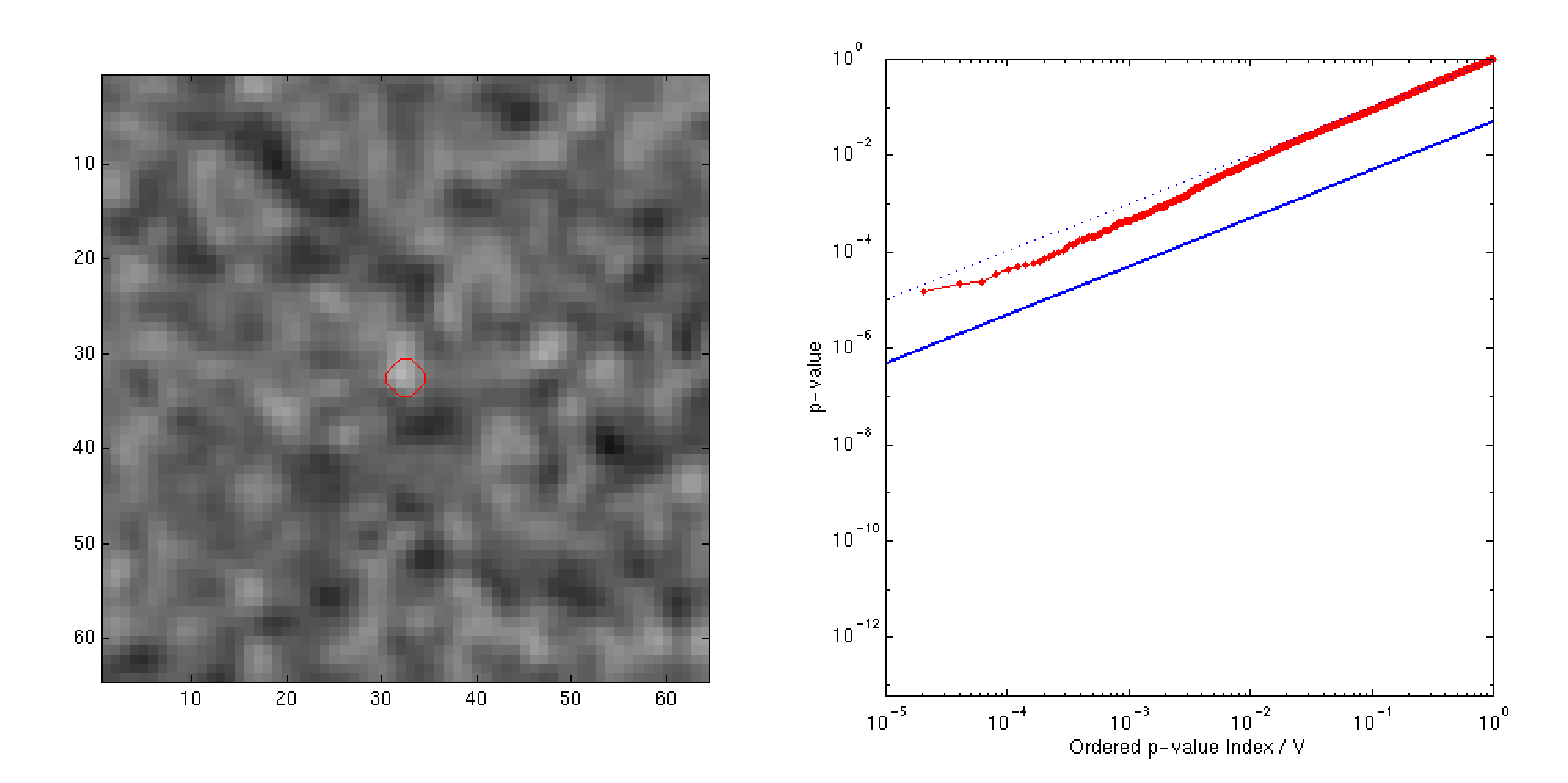
$p =$ $z =$




Signal Intensity 3.0 Signal Extent 1.0 Noise Smoothness 3.0

# B&H: Varying Signal Extent

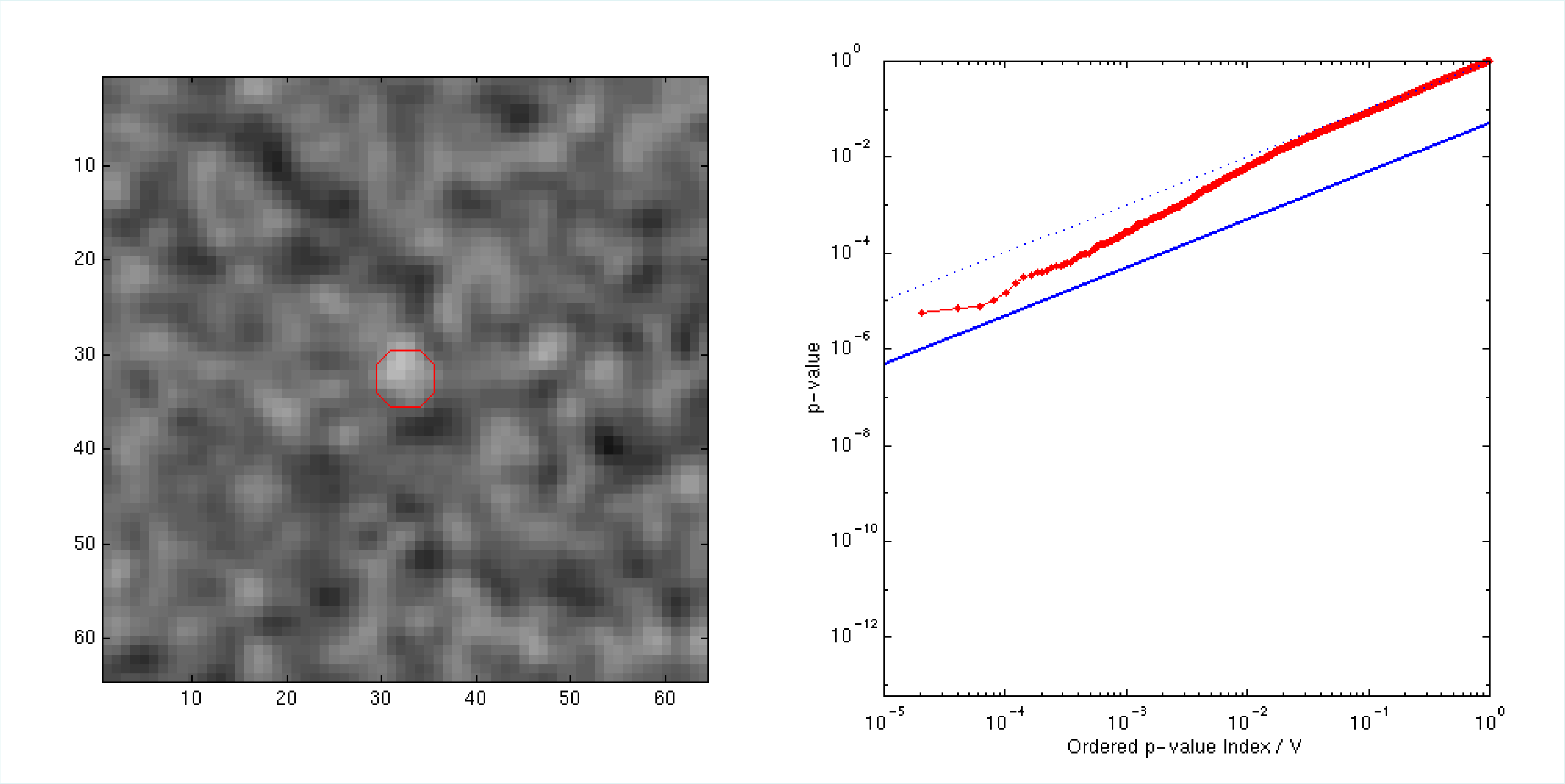
$p =$ $\qquad$ $z =$




Signal Intensity 3.0 Signal Extent 2.0 Noise Smoothness 3.0

# B&H: Varying Signal Extent

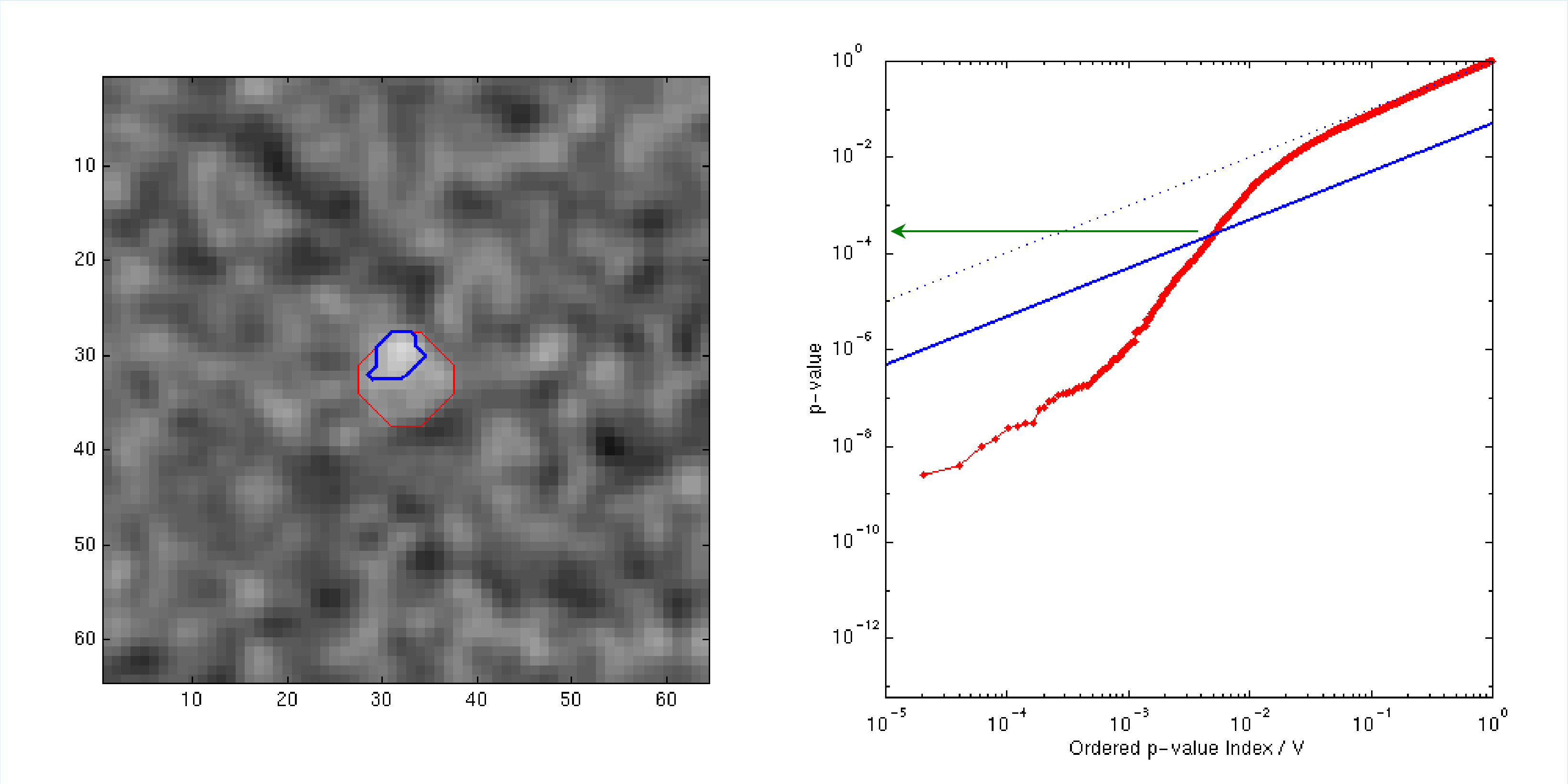
$p =$ $z =$




Signal Intensity 3.0 Signal Extent 3.0 Noise Smoothness 3.0

# B&H: Varying Signal Extent

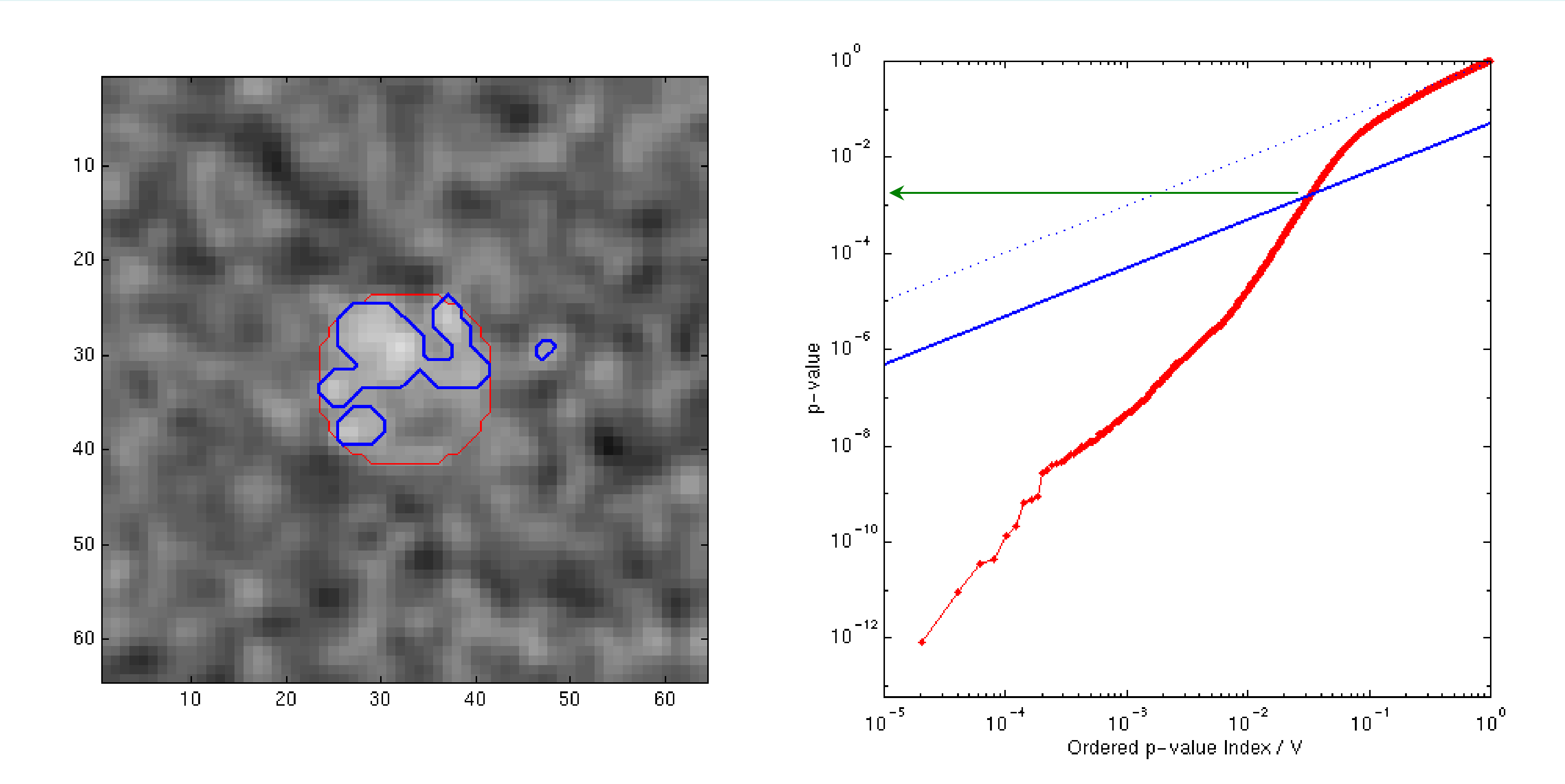
$p = 0.000252 \qquad z = 3.48$




Signal Intensity 3.0 Signal Extent 5.0 Noise Smoothness 3.0

# B&H: Varying Signal Extent

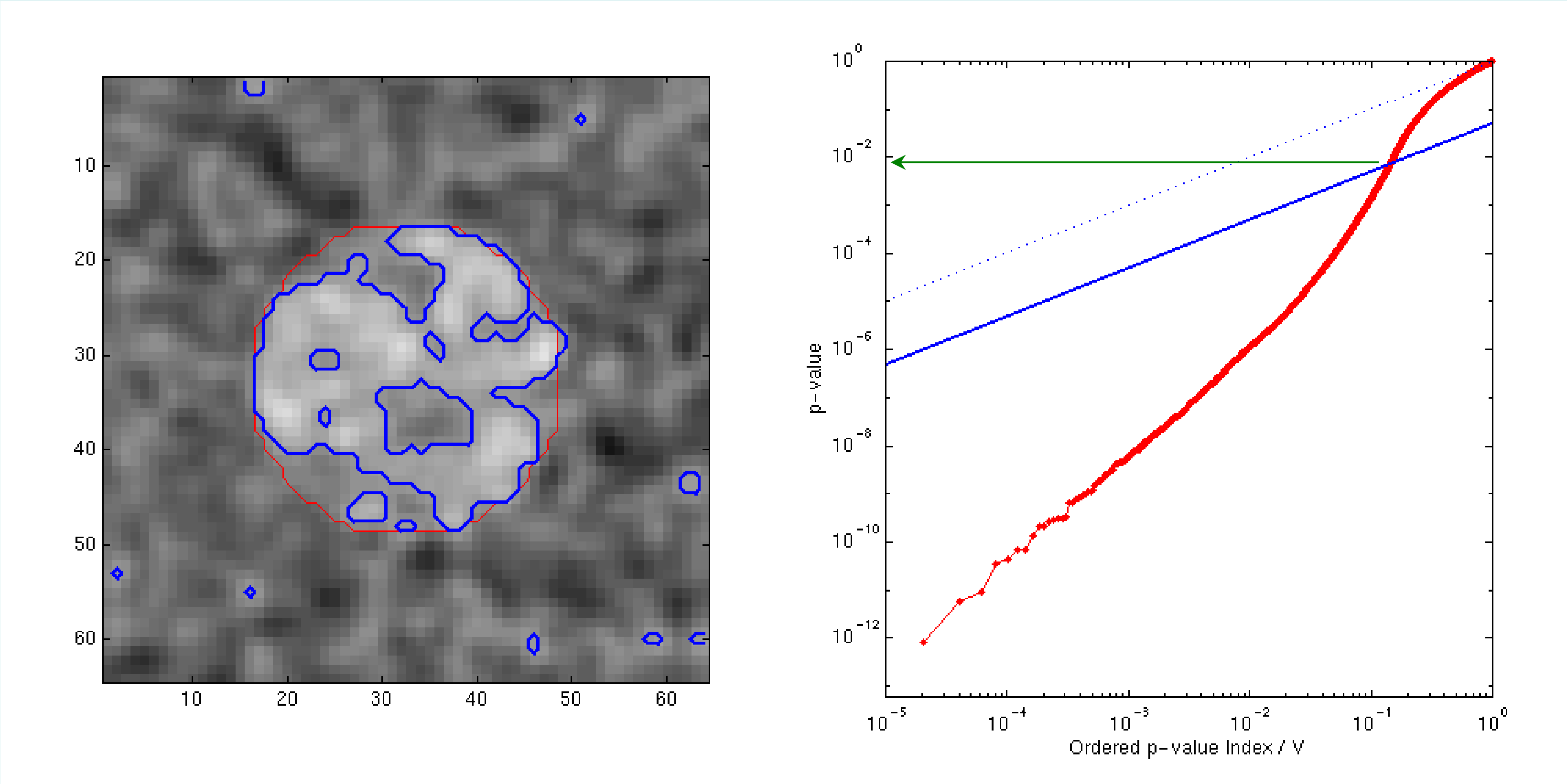
$p = 0.001628 \qquad z = 2.94$




Signal Intensity 3.0 Signal Extent 9.5 Noise Smoothness 3.0

# B&H: Varying Signal Extent

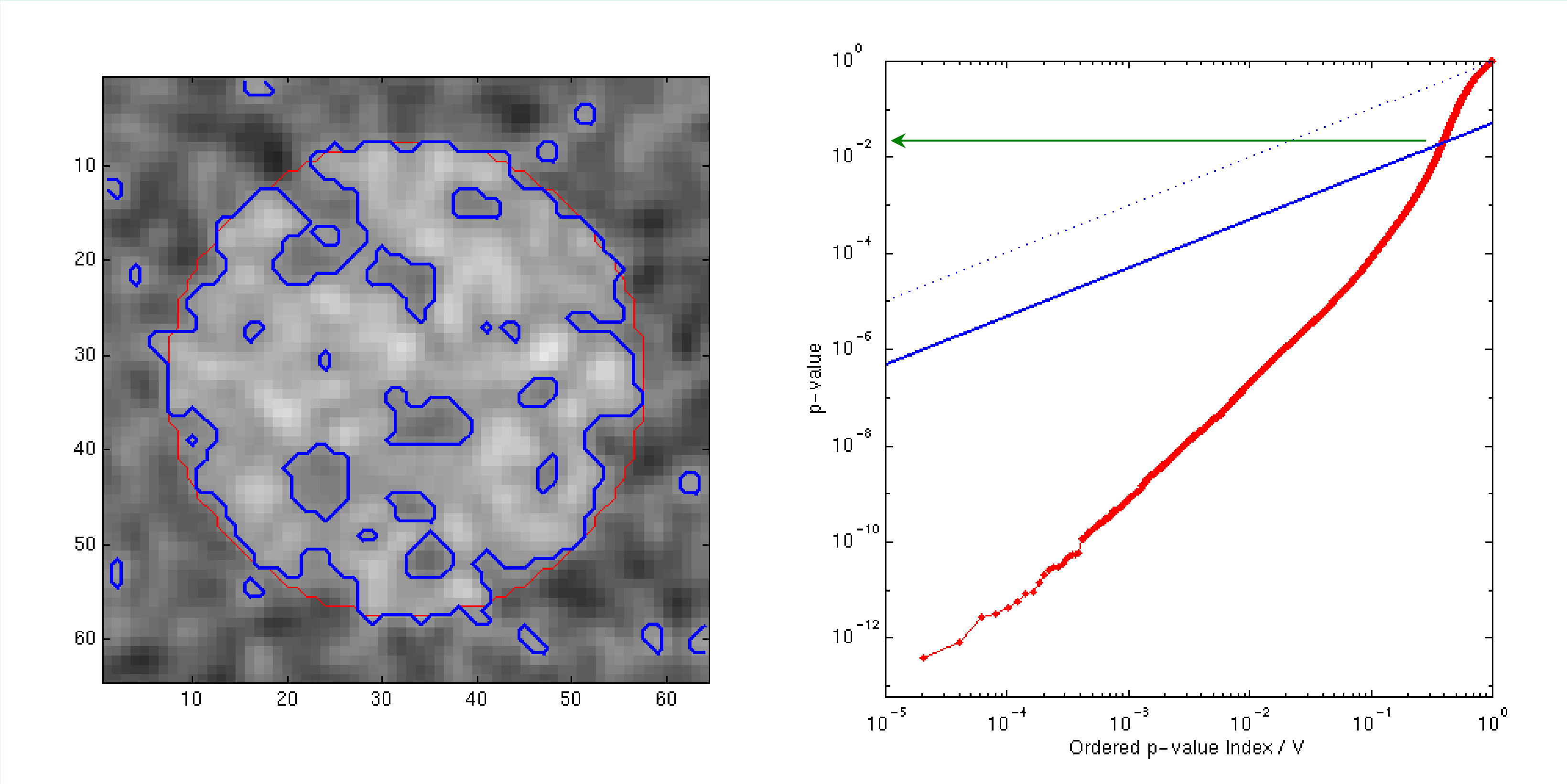
$p = 0.007157 \qquad z = 2.45$




Signal Intensity 3.0 Signal Extent 16.5 Noise Smoothness 3.0

# B&H: Varying Signal Extent

$p = 0.019274 \qquad z = 2.07$




Signal Intensity 3.0 Signal Extent 25.0 Noise Smoothness 3.0

# Benjamini & Hochberg: Properties

- Adaptive
  - Larger the signal, the lower the threshold
  - Larger the signal, the more false positives
    - False positives constant as fraction of rejected tests
    - Not a problem with imaging's sparse signals
- Smoothness OK
  - Smoothing introduces positive correlations

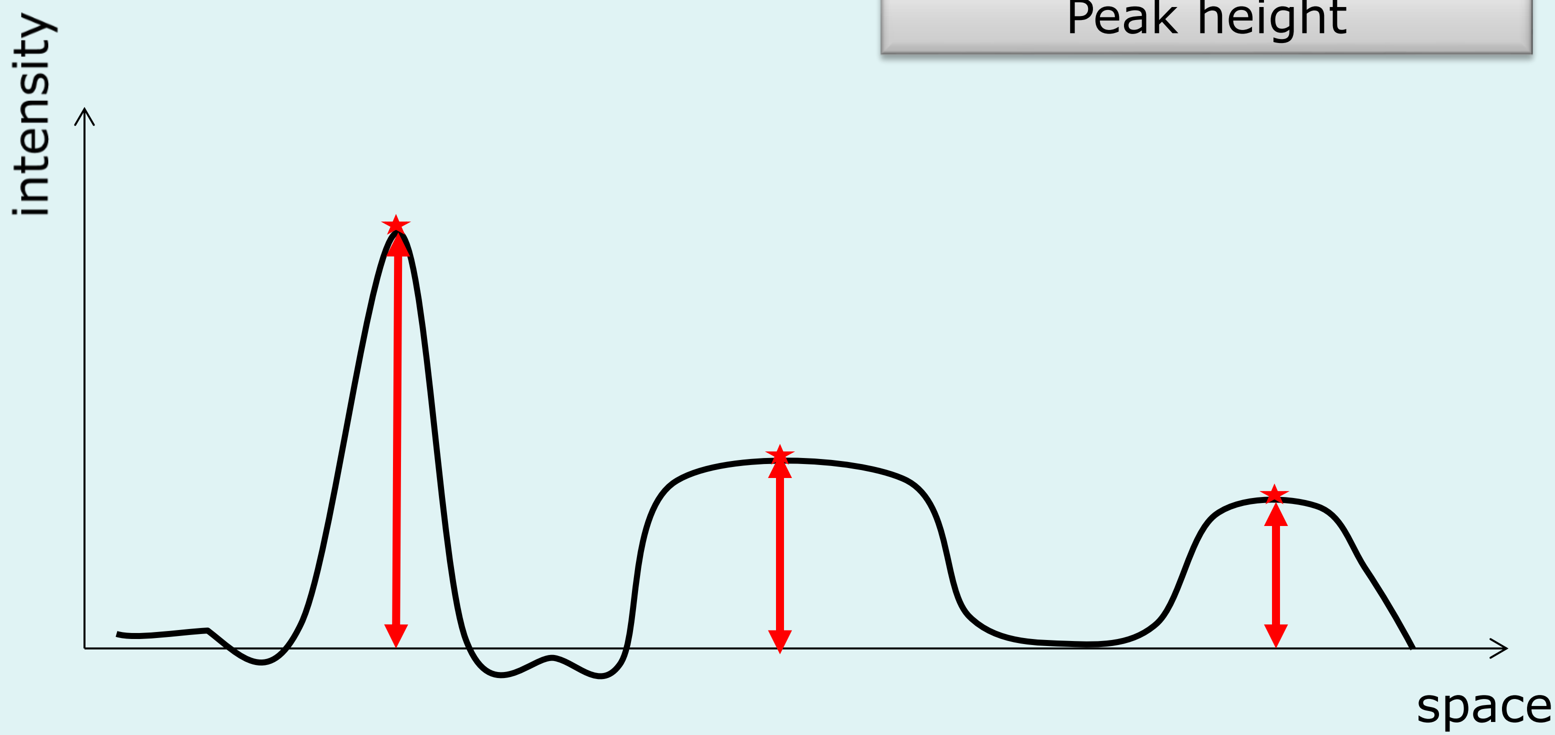
# Content

- **Introduction**
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- **Levels of inference in SPM**
  - **Topological inference**
  - **Small volume correction**
- **Non-parametric permutation test**
- **Conclusion**

# Topological inference

## Peak level inference

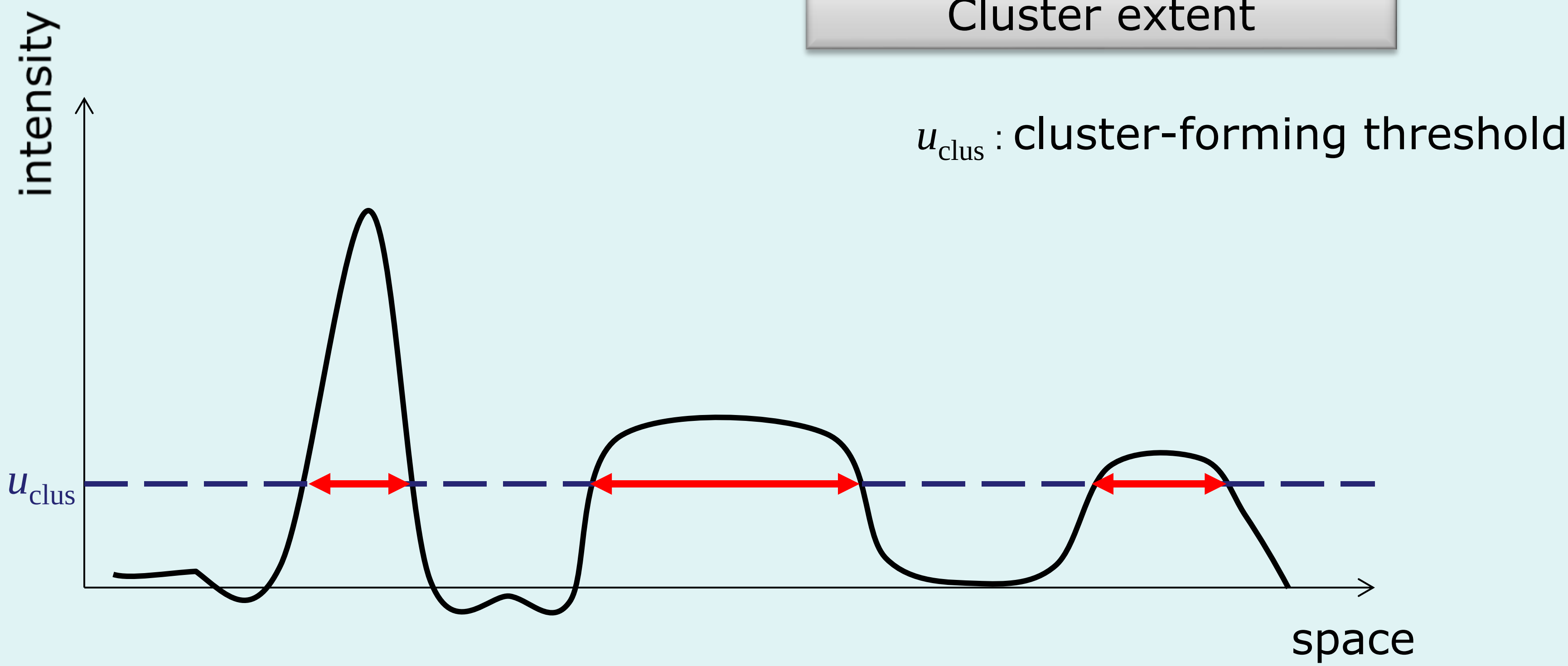
**Topological feature:**
Peak height



# Topological inference

## Cluster level inference

**Topological feature:**
Cluster extent

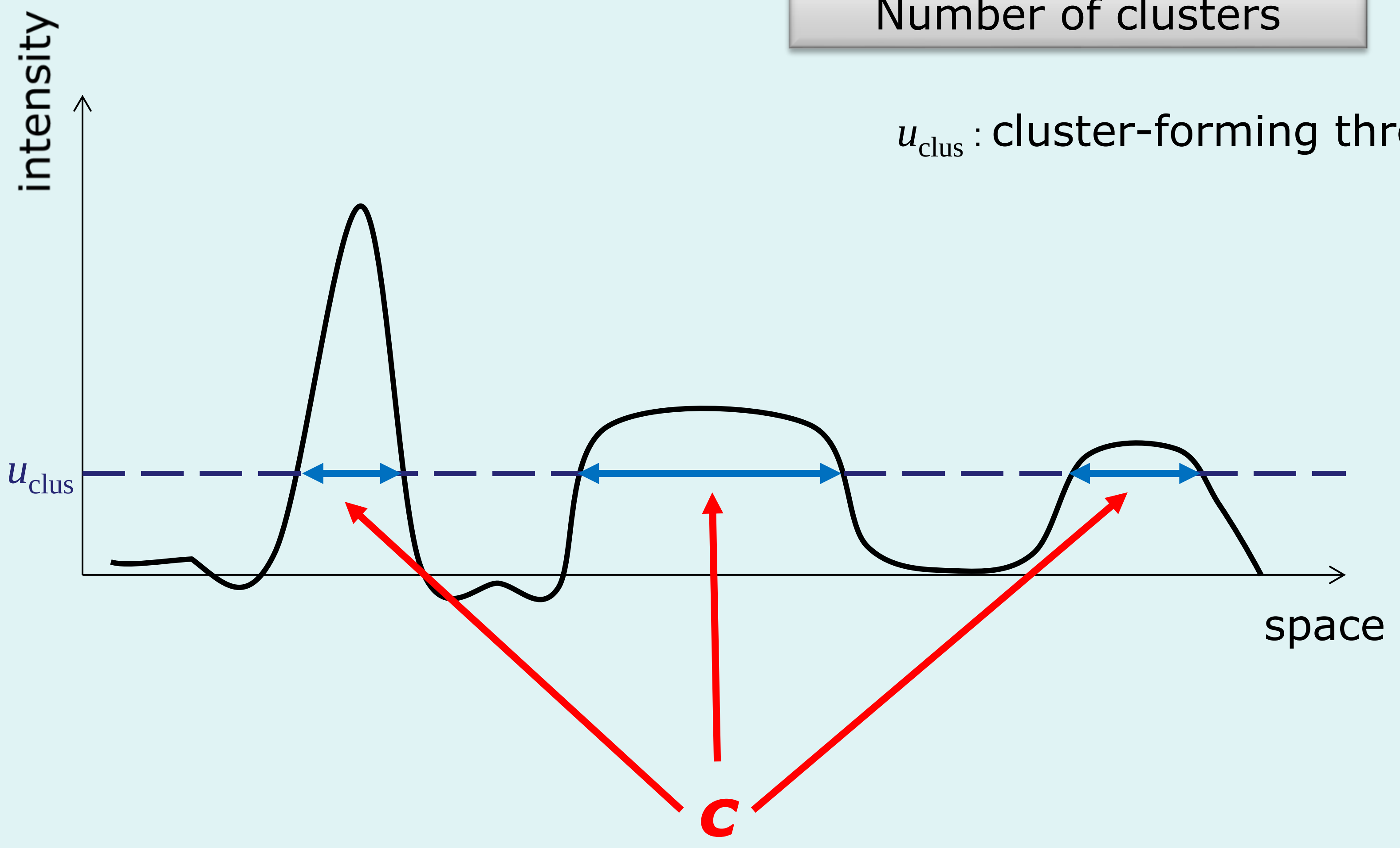



You MUST use a sufficiently high cluster-forming threshold $u_{clus}$, i.e. $\mathbf{p_{unc} < .001}$

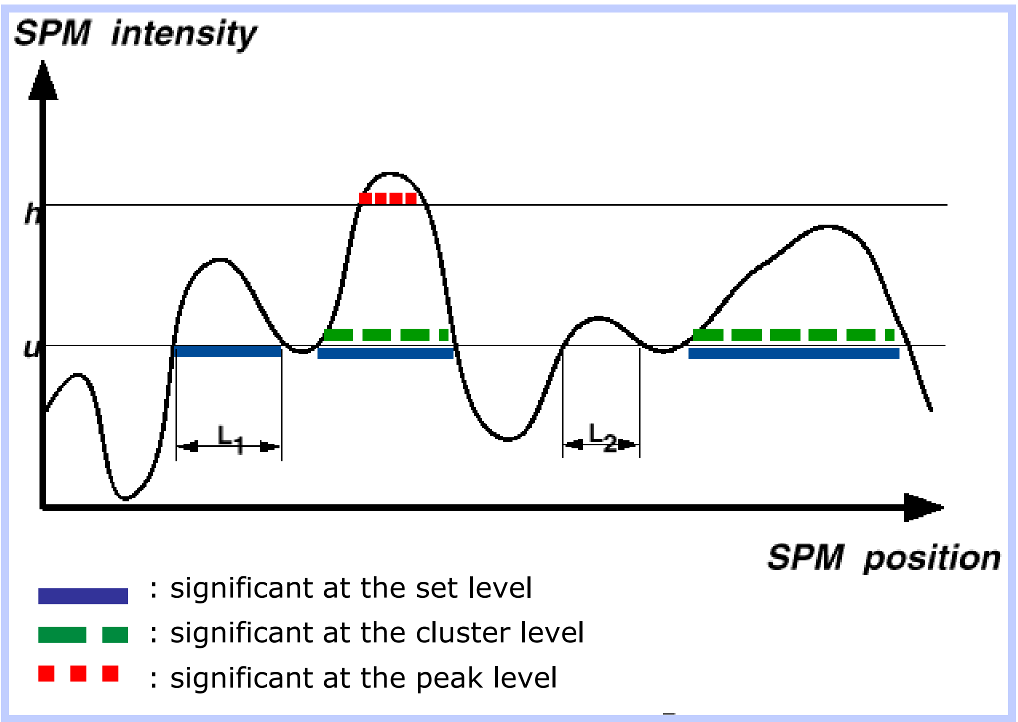
# Topological inference

## Set level inference

**Topological feature:**
Number of clusters



# Peak, cluster & set level inference




$L_1$ > spatial extent threshold
$L_2$ < spatial extent threshold

Sensitivity

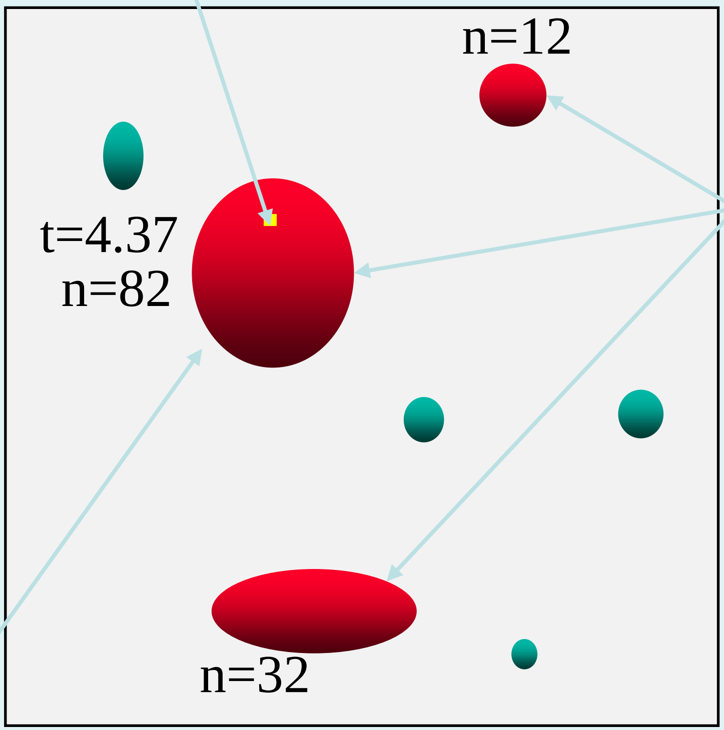
Regional specificity

**Peak level test:**
height of local maxima

**Cluster level test:**
spatial extent above u

**Set level test:**
number of clusters above u

# Levels of inference…




### Voxel-level

P(c ≥ 1 | n ≥ 0, t ≥ 4.37) = 0.048 (corrected)

P($t$ ≥ 4.37) = 1 - Φ{4.37} < 0.001 (uncorrected)

### Omnibus

P(c≥7 | n ≥ 0, t ≥ 3.09) = 0.031

### Set-level

P(c ≥ 3 | n ≥ 12, t ≥ 3.09) = 0.019

### Cluster-level

P(c ≥ 1 | n ≥ 82, t ≥ 3.09) = 0.029 (corrected)

P(n ≥ 82 | t ≥ 3.09) = 0.019 (uncorrected)

### Parameters

| | |
|---|---|
| u | - 3.09 |
| k | - 12 voxels |
| S | - $32^3$ voxels |
| *FWHM* | - 4.7 voxels |
| D | - 3 |

# Small volume correction

If one has some *a priori* idea of where an activation should be, one can pre-specify a small search space and make the appropriate correction instead of having to control for the entire search space

- mask defined by (probabilistic) anatomical atlases
- mask defined by separate "functional localisers"
- mask defined by orthogonal contrasts
- search volume around previously reported coordinates

With no prior hypothesis:
1. Test whole volume.
2. Identify SPM peak.
3. Then make a test assuming a single voxel.

# Small Volume Correction

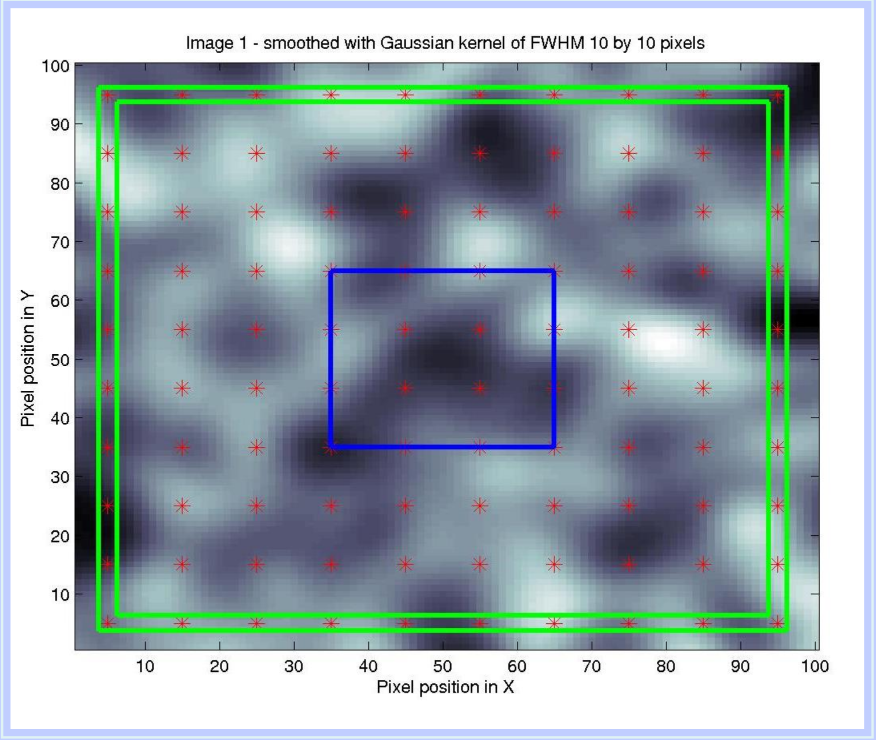
SVC = correction for multiple comparison in a user's defined volume 'of interest'.

Shape and size of volume become important for small or oddly shaped volume !

Example of SVC (900 voxels)

- compact volume: samples from maximum 16 resels
- spread volume: sample from up to 36 resels

$\Rightarrow$ threshold higher for spread volume than compact volume.



# Small volume correction, topology

TABLE 3. Representative examples of resel counts and critical values.

| Search region $V$ | Vol. (cc) | Resel counts $R_0(V)$ | $R_1(V)$ | $R_2(V)$ | $R_3(V)$ | $t$ for $P(M \geq t) =$ 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|---|
| Single voxel | 0 | 1 | 0 | 0 | 0 | 1.28 | 1.64 | 2.33 |
| Head Of Caudate | 7 | 0 | 6.18 | 4.63 | 0.65 | 2.75 | 3.02 | 3.55 |
| Putamen | 12 | 1 | 7.32 | 6.80 | 1.18 | 2.89 | 3.15 | 3.66 |
| Globus Pallidus | 3 | 0 | 4.03 | 2.29 | 0.24 | 2.49 | 2.78 | 3.35 |
| Thalamus | 11 | 1 | 4.94 | 5.14 | 1.13 | 2.79 | 3.05 | 3.59 |
| Anterior Cingulate Gyrus | 9 | 1 | 8.20 | 5.79 | 0.86 | 2.86 | 3.11 | 3.63 |
| Posterior Cingulate Gyrus | 6 | 1 | 5.32 | 3.85 | 0.58 | 2.70 | 2.97 | 3.51 |
| Cingulate Gyri | 15 | 0 | 12.89 | 9.63 | 1.44 | 3.03 | 3.27 | 3.77 |
| Superior Frontal Gyrus | 80 | 1 | 15.64 | 25.69 | 8.97 | 3.38 | 3.60 | 4.07 |
| Middle Frontal Gyrus | 57 | 1 | 14.89 | 21.14 | 6.23 | 3.31 | 3.53 | 4.00 |
| Inferior Frontal Gyrus | 37 | 1 | 11.22 | 14.25 | 4.06 | 3.17 | 3.41 | 3.89 |
| Precentral Gyrus | 32 | 1 | 12.30 | 14.23 | 3.40 | 3.16 | 3.40 | 3.88 |
| Frontal Gyri | 207 | 1 | 19.30 | 53.39 | 23.63 | 3.63 | 3.84 | 4.28 |
| Occipital Lobe | 65 | -1 | 10.68 | 23.11 | 7.17 | 3.32 | 3.55 | 4.02 |
| 4mm shell | 254 | 2 | 0.54 | 207.27 | 15.88 | 3.85 | 4.04 | 4.45 |
| Whole brain | 1294 | 1 | 20.43 | 107.09 | 153.42 | 4.05 | 4.23 | 4.63 |

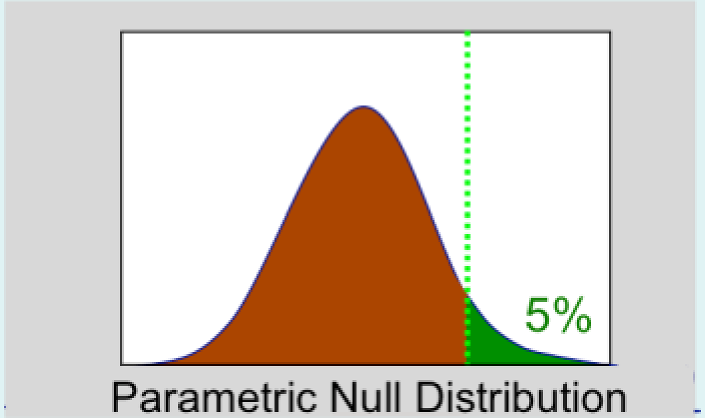
FWHM=20mm

# Content

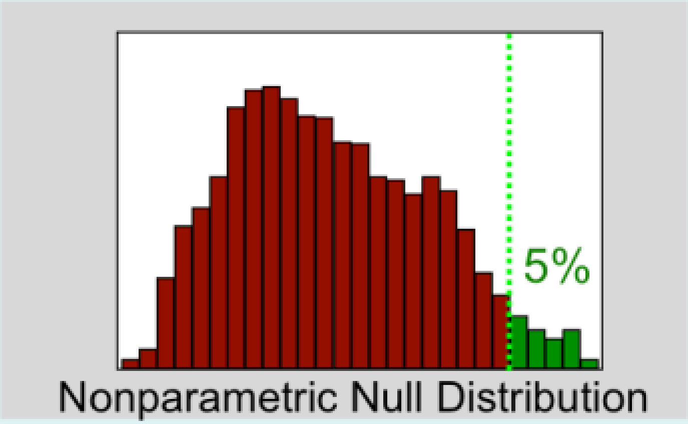
- Introduction
- Family-wise error rate (FWER)
- False discovery rate (FDR)
- Levels of inference in SPM
- **Non-parametric permutation test**
- Conclusion

# Non-parametric permutation test

- Parametric methods
  - Assume distribution of statistic under null hypothesis




- Nonparametric methods
  - Use *data* to find distribution of statistic under null hypothesis
  - Any statistic!



# Permutation Test : Toy Example

- Data from V1 voxel in visual stim. experiment
  A: Active, flashing checkerboard   B: Baseline, fixation
  6 blocks, ABABAB     Just consider block averages...

| A | B | A | B | A | B |
|---|---|---|---|---|---|
| **103.00** | **90.48** | **99.93** | **87.83** | **99.76** | **96.06** |

- Null hypothesis $H_o$
  - No experimental effect, A & B labels arbitrary
- Statistic
  - Mean difference

# Permutation Test : Toy Example

- Under $H_o$
  - Consider all equivalent relabelings

| | | | |
|---|---|---|---|
| **AAABBB** | **ABABAB** | **BAAABB** | **BABBAA** |
| **AABABB** | **ABABBA** | **BAABAB** | **BBAAAB** |
| **AABBAB** | **ABBAAB** | **BAABBA** | **BBAABA** |
| **AABBBA** | **ABBABA** | **BABAAB** | **BBABAA** |
| **ABAABB** | **ABBBAA** | **BABABA** | **BBBAAA** |

# Permutation Test : Toy Example

- Under $H_o$
  - Consider all equivalent relabelings
  - Compute all possible statistic values

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AAABBB** | **4.82** | **ABABAB** | **9.45** | **BAAABB** | **-1.48** | **BABBAA** | **-6.86** |
| **AABABB** | **-3.25** | **ABABBA** | **6.97** | **BAABAB** | **1.10** | **BBAAAB** | **3.15** |
| **AABBAB** | **-0.67** | **ABBAAB** | **1.38** | **BAABBA** | **-1.38** | **BBAABA** | **0.67** |
| **AABBBA** | **-3.15** | **ABBABA** | **-1.10** | **BABAAB** | **-6.97** | **BBABAA** | **3.25** |
| **ABAABB** | **6.86** | **ABBBAA** | **1.48** | **BABABA** | **-9.45** | **BBBAAA** | **-4.82** |

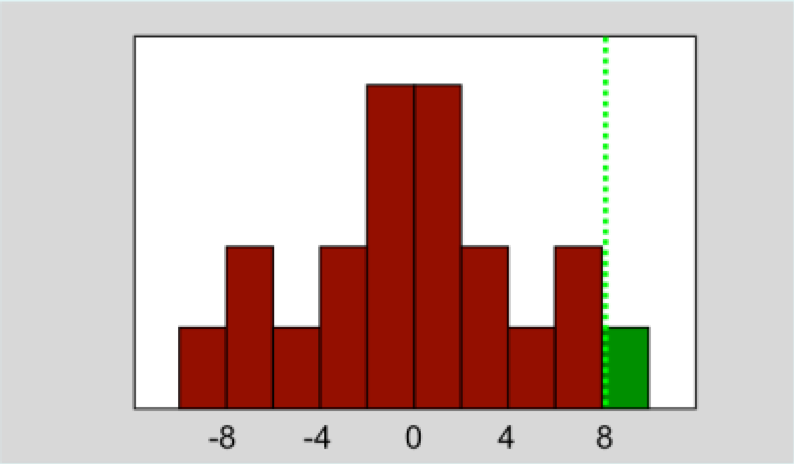
# Permutation Test : Toy Example

- Under $H_o$
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AAABBB** | **4.82** | **ABABAB** | **9.45** | **BAAABB** | **-1.48** | **BABBAA** | **-6.86** |
| **AABABB** | **-3.25** | **ABABBA** | **6.97** | **BAABAB** | **1.10** | **BBAAAB** | **3.15** |
| **AABBAB** | **-0.67** | **ABBAAB** | **1.38** | **BAABBA** | **-1.38** | **BBAABA** | **0.67** |
| **AABBBA** | **-3.15** | **ABBABA** | **-1.10** | **BABAAB** | **-6.97** | **BBABAA** | **3.25** |
| **ABAABB** | **6.86** | **ABBBAA** | **1.48** | **BABABA** | **-9.45** | **BBBAAA** | **-4.82** |

# Permutation Test : Toy Example

- Under $H_o$
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution



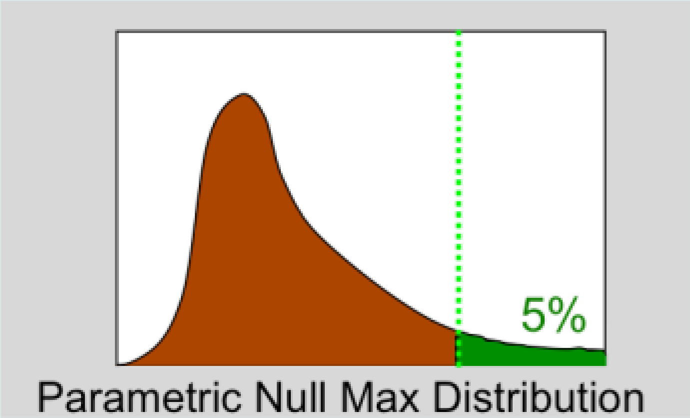
# Permutation Test : Toy Example

- Under $H_o$
  - Consider all equivalent relabelings
  - Compute all possible statistic values
  - Find 95%ile of permutation distribution

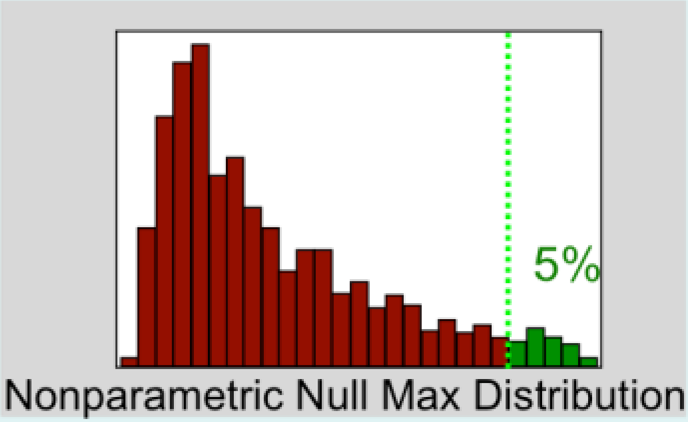
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AAABBB** | **4.82** | **ABABAB** | **9.45** | **BAAABB** | **-1.48** | **BABBAA** | **-6.86** |
| **AABABB** | **-3.25** | **ABABBA** | **6.97** | **BAABAB** | **1.10** | **BBAAAB** | **3.15** |
| **AABBAB** | **-0.67** | **ABBAAB** | **1.38** | **BAABBA** | **-1.38** | **BBAABA** | **0.67** |
| **AABBBA** | **-3.15** | **ABBABA** | **-1.10** | **BABAAB** | **-6.97** | **BBABAA** | **3.25** |
| **ABAABB** | **6.86** | **ABBBAA** | **1.48** | **BABABA** | **-9.45** | **BBBAAA** | **-4.82** |

# Controlling FWER: Permutation Test

- Parametric methods
  - Assume distribution of ***max*** statistic under null hypothesis




- Nonparametric methods
  - Use *data* to find distribution of ***max*** statistic under null hypothesis
  - Again, any max statistic!



# Permutation Test & Exchangeability

- Exchangeability is fundamental
  - Def: Distribution of the data unperturbed by permutation
  - Under $H_0$, exchangeability justifies permuting data
  - Allows us to build permutation distribution

- Subjects are exchangeable
  - Under $H_o$, each subject's A/B labels can be flipped

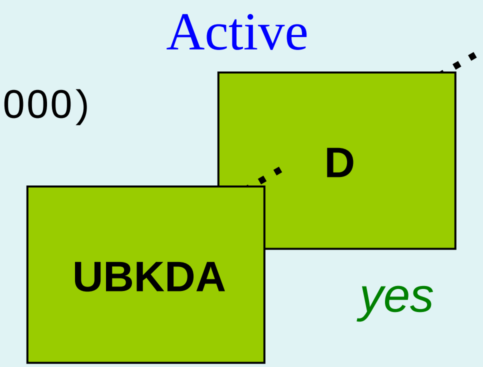
- Are fMRI scans exchangeable under $H_o$?
  - If no signal, can we permute over time?

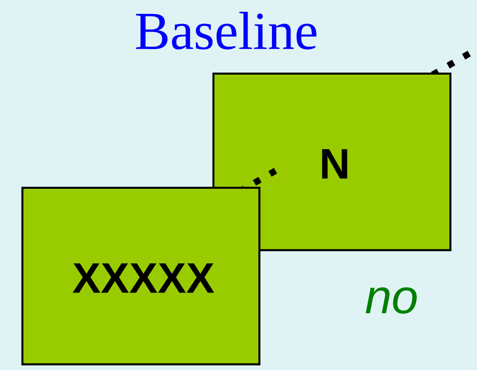
# Permutation Test & Exchangeability

- fMRI scans are *not* exchangeable
  - Permuting disrupts order, temporal autocorrelation
- *Intra*subject fMRI permutation test
  - Must decorrelate data, model before permuting
  - What is correlation structure?
    - Usually must use parametric model of correlation
  - E.g. Use wavelets to decorrelate
    - Bullmore et al 2001, HBM 12:61-78
- *Inter*subject fMRI permutation test
  - Create difference image for each subject
  - For each permutation, flip sign of some subjects

# Permutation Test : Example

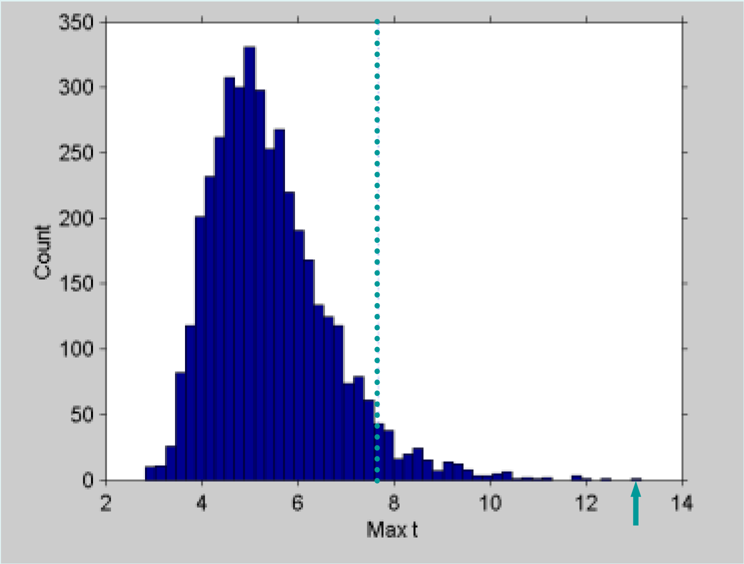
- fMRI Study of Working Memory
  - 12 subjects, block design Marshuetz et al (2000)
  - Item Recognition
    - Active: View five letters, 2s pause, view probe letter, respond
    - Baseline: View XXXXX, 2s pause, view Y or N, respond
- Second Level RFX
  - Difference image, A-B constructed for each subject
  - One sample, smoothed variance *t* test



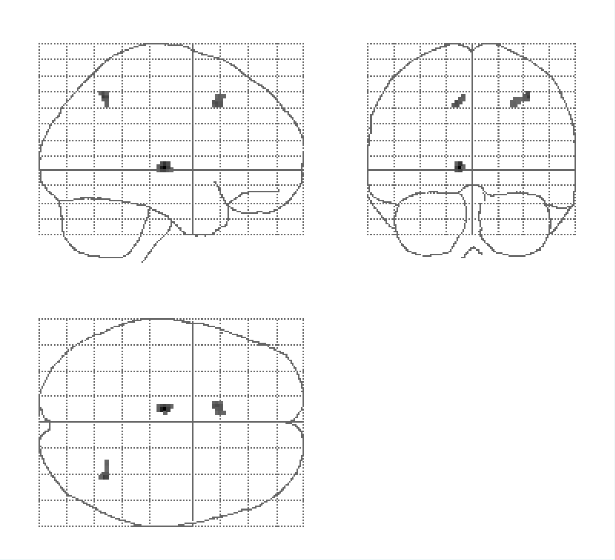



# Permutation Test : Example

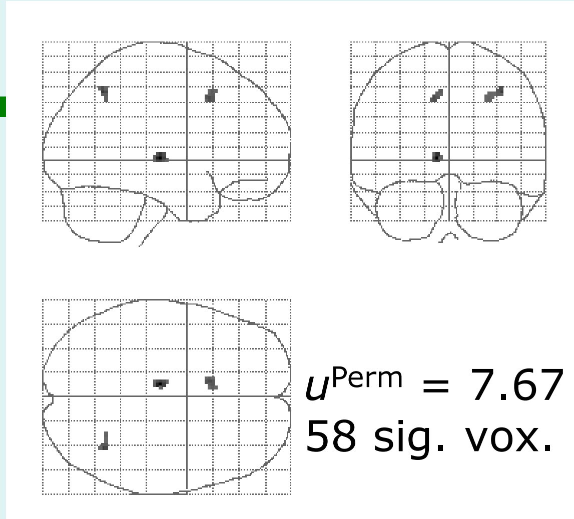
- Permute!
  - $2^{12}$ = 4,096 ways to flip 12 A/B labels
  - For each, note maximum of *t* image



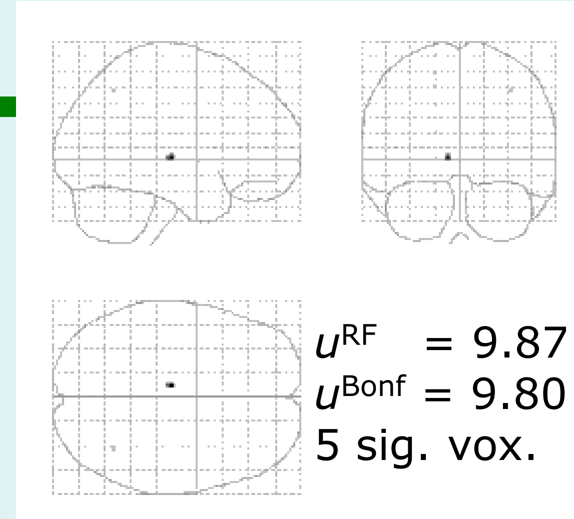

Permutation Distribution
Maximum *t*



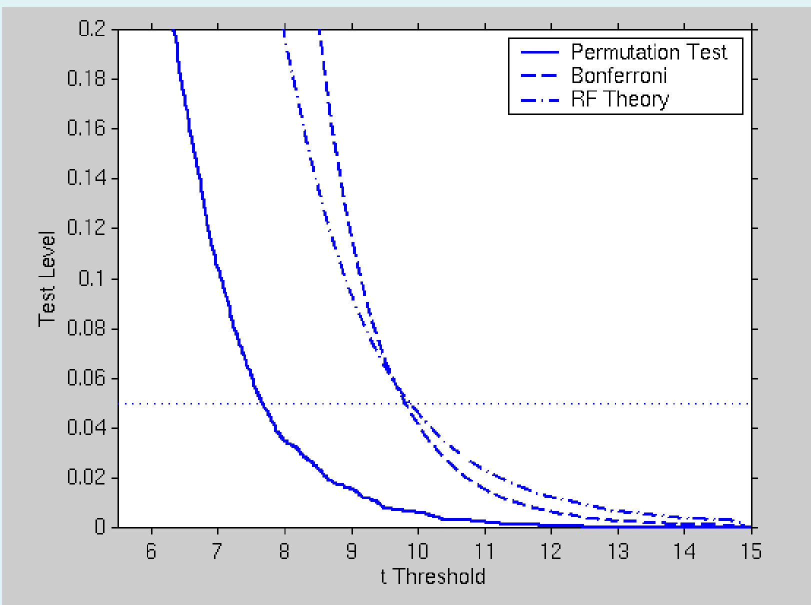
Maximum Intensity Projection
Thresholded *t*




$t_{11}$ Statistic, Nonparametric Threshold




$t_{11}$ Statistic, RF & Bonf. Threshold




Test Level vs. $t_{11}$ Threshold

- Compare with Bonferroni
  $\alpha$ = 0.05/110,776
- Compare with parametric RFT
  110,776 2×2×2mm voxels
  5.1×5.8×6.9mm FWHM

  smoothness
  462.9 RESELs

# Generalization: RFT vs Bonf. vs Perm.

| | | *t* Threshold (0.05 Corrected) | | |
|---|---|---|---|---|
| | df | RF | Bonf | Perm |
| Verbal Fluency | 4 | 4701.32 | 42.59 | 10.14 |
| Location Switching | 9 | 11.17 | 9.07 | 5.83 |
| Task Switching | 9 | 10.79 | 10.35 | 5.10 |
| Faces: Main Effect | 11 | 10.43 | 9.07 | 7.92 |
| Faces: Interaction | 11 | 10.70 | 9.07 | 8.26 |
| Item Recognition | 11 | 9.87 | 9.80 | 7.67 |
| Visual Motion | 11 | 11.07 | 8.92 | 8.40 |
| Emotional Pictures | 12 | 8.48 | 8.41 | 7.15 |
| Pain: Warning | 22 | 5.93 | 6.05 | 4.99 |
| Pain: Anticipation | 22 | 5.87 | 6.05 | 5.05 |

# RFT vs Bonf. vs Perm.

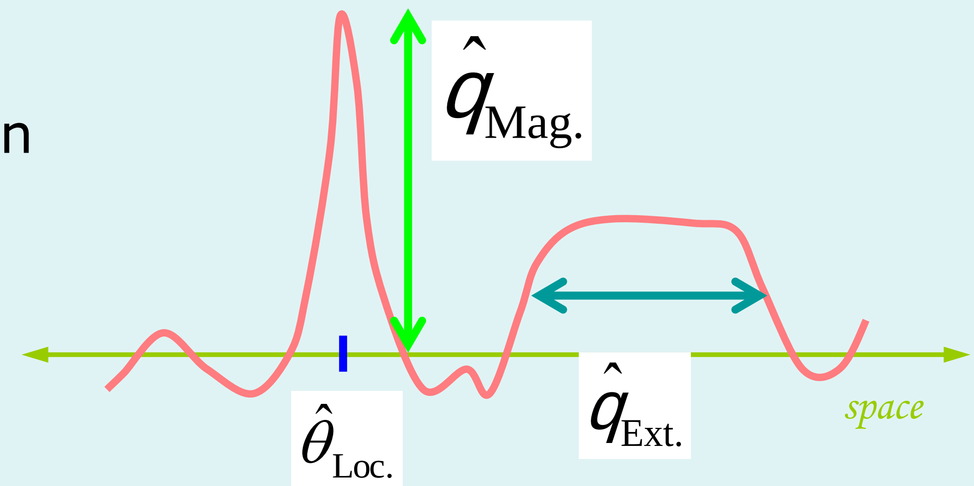
| | | No. Significant Voxels (0.05 Corrected) | | |
|---|---|---|---|---|
| | | *t* | | |
| | df | RF | Bonf | Perm |
| Verbal Fluency | 4 | 0 | 0 | 0 |
| Location Switching | 9 | 0 | 0 | 158 |
| Task Switching | 9 | 4 | 6 | 2241 |
| Faces: Main Effect | 11 | 127 | 371 | 917 |
| Faces: Interaction | 11 | 0 | 0 | 0 |
| Item Recognition | 11 | 5 | 5 | 58 |
| Visual Motion | 11 | 626 | 1260 | 1480 |
| Emotional Pictures | 12 | 0 | 0 | 0 |
| Pain: Warning | 22 | 127 | 116 | 221 |
| Pain: Anticipation | 22 | 74 | 55 | 182 |

# Content

- **Introduction**
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- **Levels of inference in SPM**
- **Non-parametric permutation test**
- **Conclusion**

# What we'd like

- Don't threshold, model the signal!
  - Signal location?
    - Estimates and CI's on (x,y,z) location
  - Signal magnitude?
    - CI's on % change
  - Spatial extent?
    - Estimates and CI's on activation volume
    - Robust to choice of cluster definition
- ...but this requires an explicit spatial model



# Real-life inference: What we get

- Signal location
  - Local maximum – *no inference*
  - Center-of-mass – *no inference*
    - Sensitive to blob-defining-threshold
- Signal magnitude
  - Local maximum intensity – P-values (& CI's)
- Spatial extent
  - Cluster volume – P-value, no CI's
    - Sensitive to blob-defining-threshold

# FWER vs. FDR

You <u>**MUST**</u> account for multiplicity

(Otherwise have a fishing expedition)

- FWER
  - Very specific, not very sensitive
- FDR
  - Less specific, more sensitive

(Sociological calibration still underway)

# Conclusion

- There is a **multiple testing problem** and corrections *must* be applied on p-values, possibly for the volume of interest only (see SVC).

- Inference is made about **topological features** (peak height, spatial extent, number of clusters). Use results from the **Random Field Theory**. Or **permutation tests**.

- **Control of FWER** (probability of a false positive anywhere in the image) for a space of any dimension and shape.

# References


Friston KJ, Frith CD, Liddle PF, Frackowiak RS. Comparing functional (PET) images: the assessment of significant change. Journal of Cerebral Blood Flow and Metabolism, 1991.

Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. Journal of Cerebral Blood Flow and Metabolism. 1992.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping,1996.

Chumbley J, Worsley KJ , Flandin G, and Friston KJ. Topological FDR for neuroimaging. NeuroImage, 2010.

Kilner J and Friston KJ. Topological inference for EEG and MEG data. Annals of Applied Statistics, 2010.

Flandin G and Friston KJ. Topological Inference. Brain Mapping: An Encyclopedic Reference, 2015.

Flandin G and Friston KJ. Analysis of family-wise error rates in statistical parametric mapping using random field theory. Human Brain Mapping, 2017.

- And now a little demo!