Inference on SPMs: Random Field Theory & Alternatives

Thomas Nichols, Ph.D.
Department of Statistics &
Warwick Manufacturing Group
University of Warwick

Vancouver SPM Course
August 2010

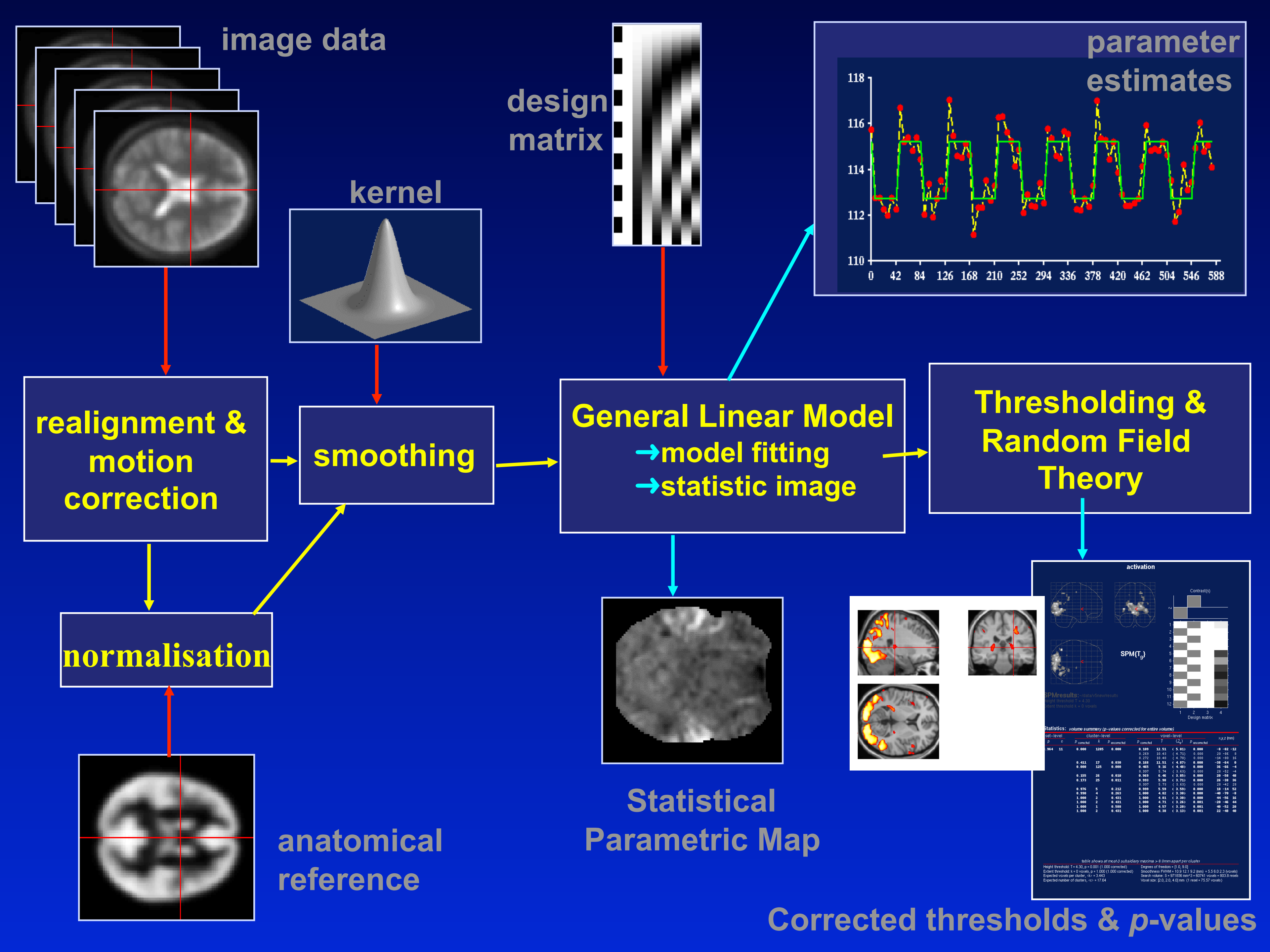
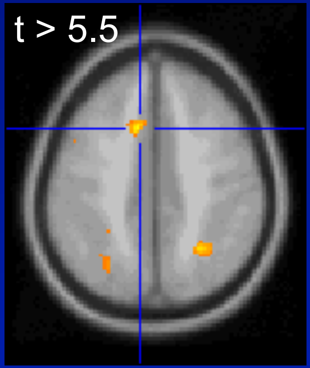
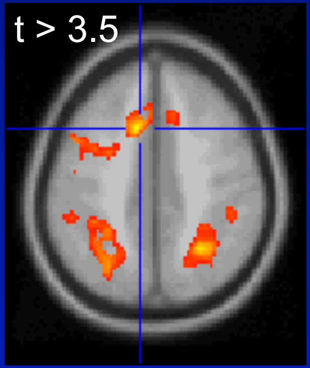
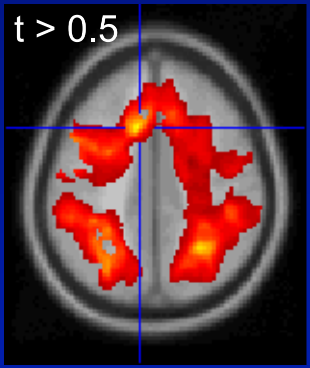


image data
design
matrix
parameter
estimates
kernel
realignment &
motion
correction
smoothing
General Linear Model
➜model fitting
➜statistic image
Thresholding &
Random Field
Theory
normalisation
Statistical
Parametric Map
anatomical
reference
Corrected thresholds & p-values

Assessing Statistic Images…

Assessing Statistic Images

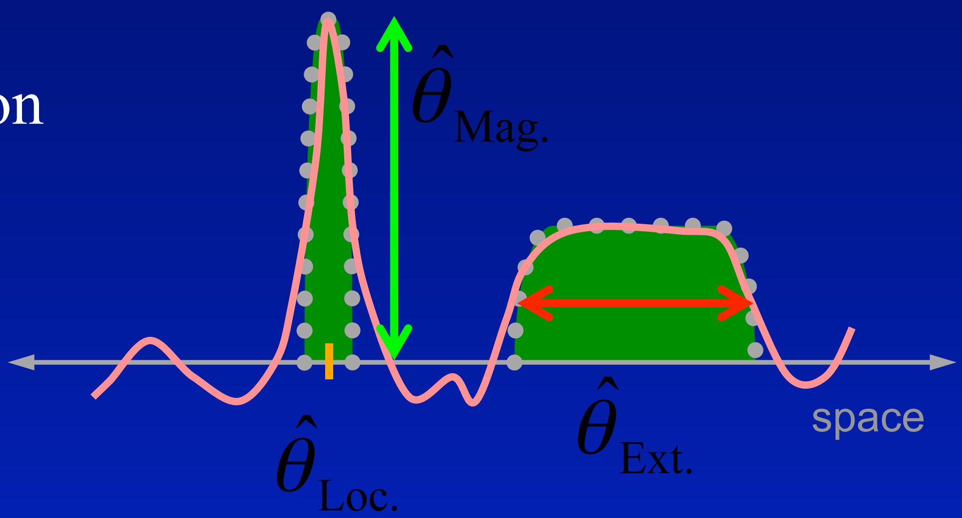
Where's the signal?

High Threshold	Med. Threshold	Low Threshold
		
Good Specificity Poor Power (risk of false negatives)		Poor Specificity (risk of false positives) Good Power

...but why threshold?!

Blue-sky inference: 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



Blue-sky inference: What we need

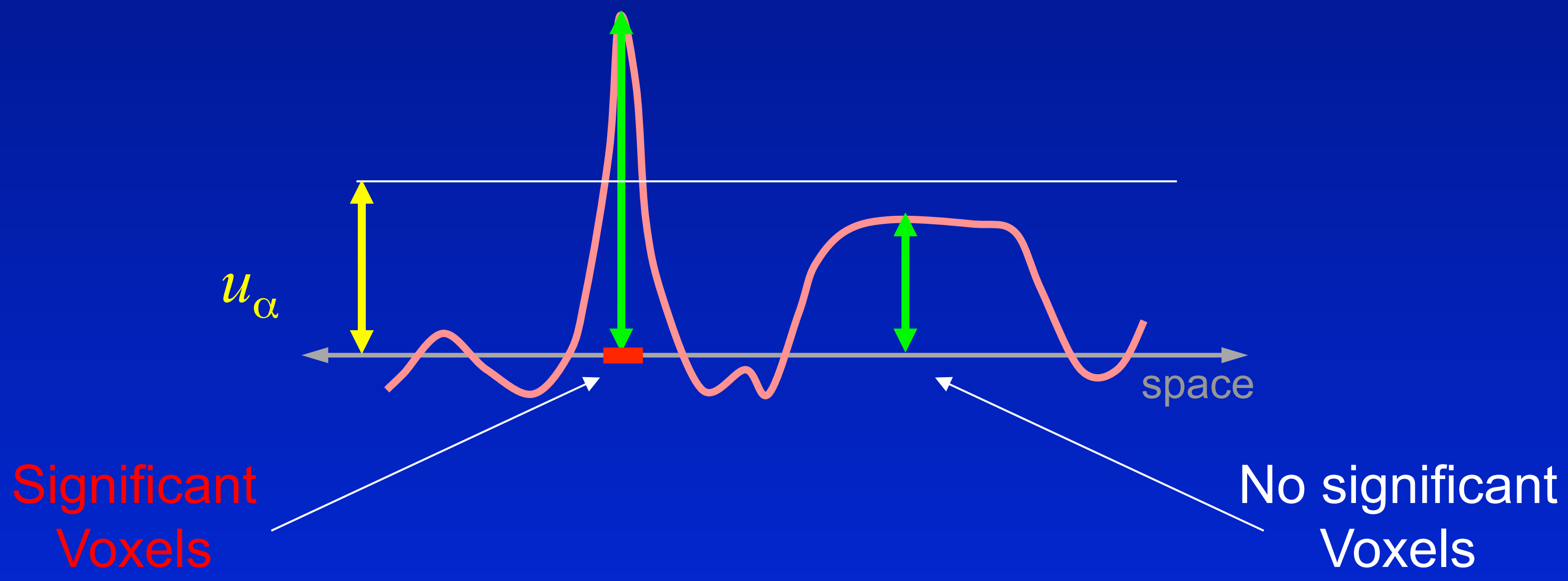
- Need an explicit spatial model
- No routine spatial modeling methods exist
  - High-dimensional mixture modeling problem
  - Activations don't look like Gaussian blobs
  - Need realistic shapes, sparse representation
    - Some work by Hartvig *et al.*, Penny *et al.*

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

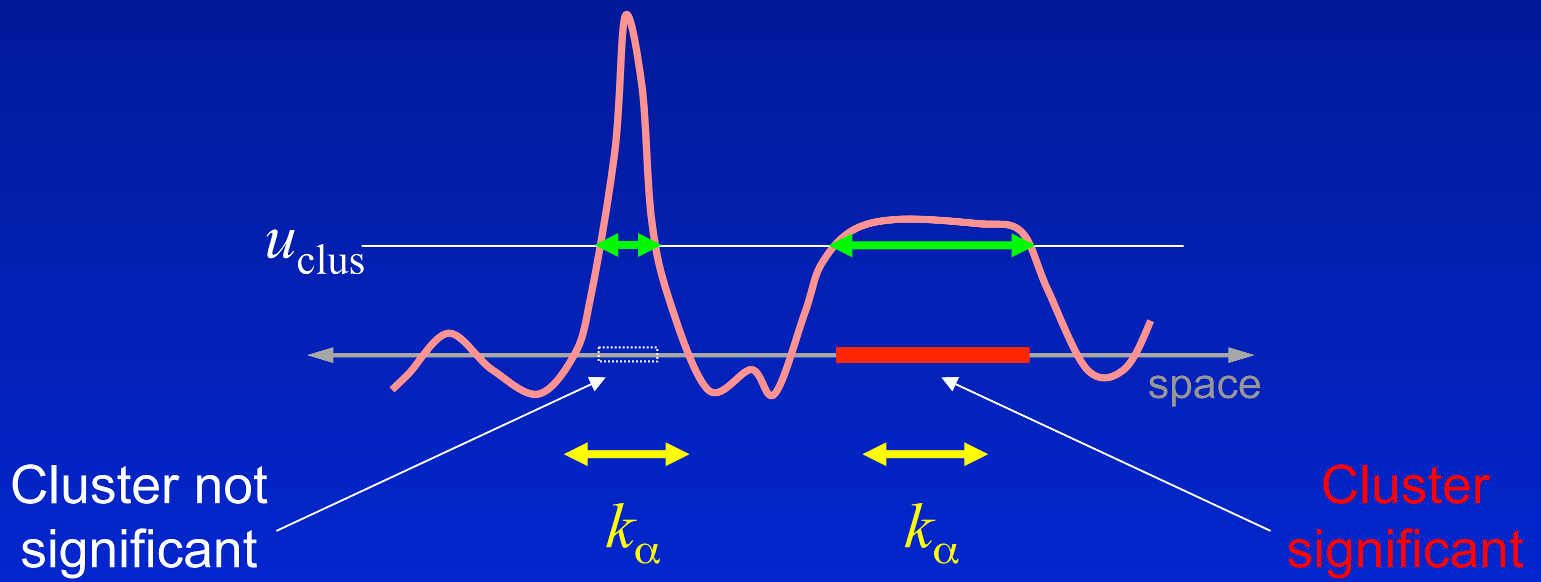
Voxel-level Inference

- Retain voxels above α-level threshold u_α
- Gives best spatial specificity
  - The null hyp. at a single voxel can be rejected



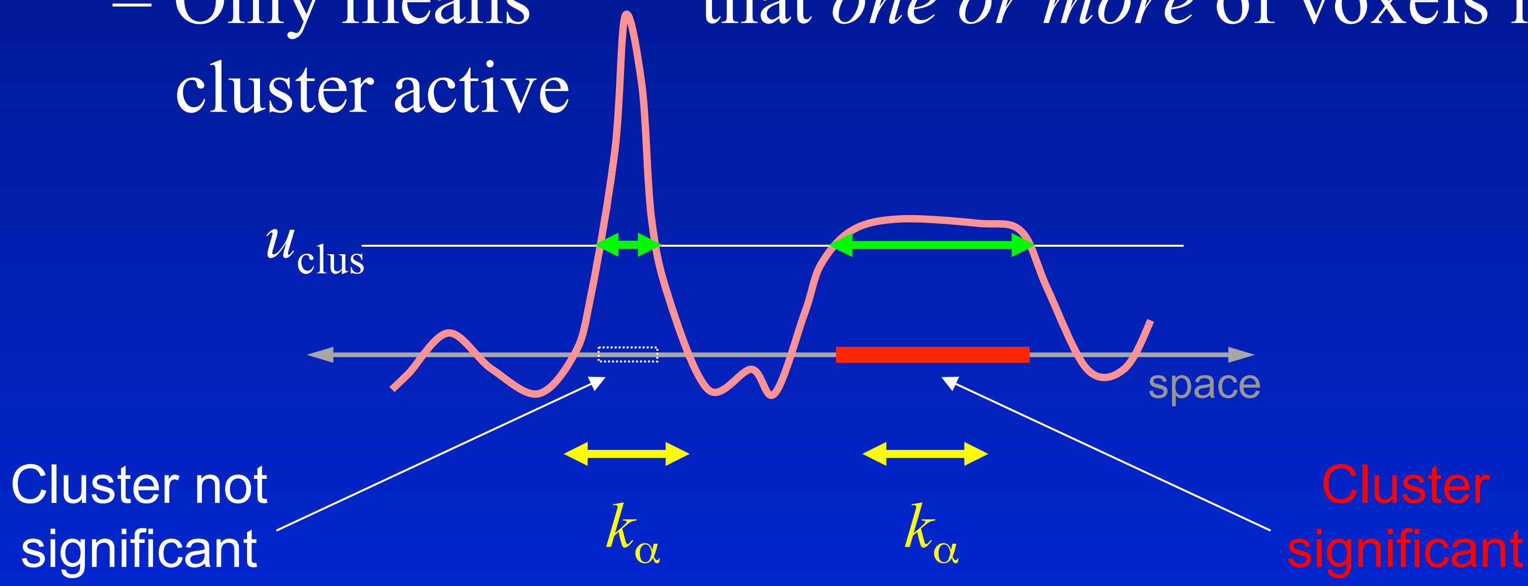
Cluster-level Inference

- Two step-process
  - Define clusters by arbitrary threshold u_{clus}
  - Retain clusters larger than α-level threshold k_α



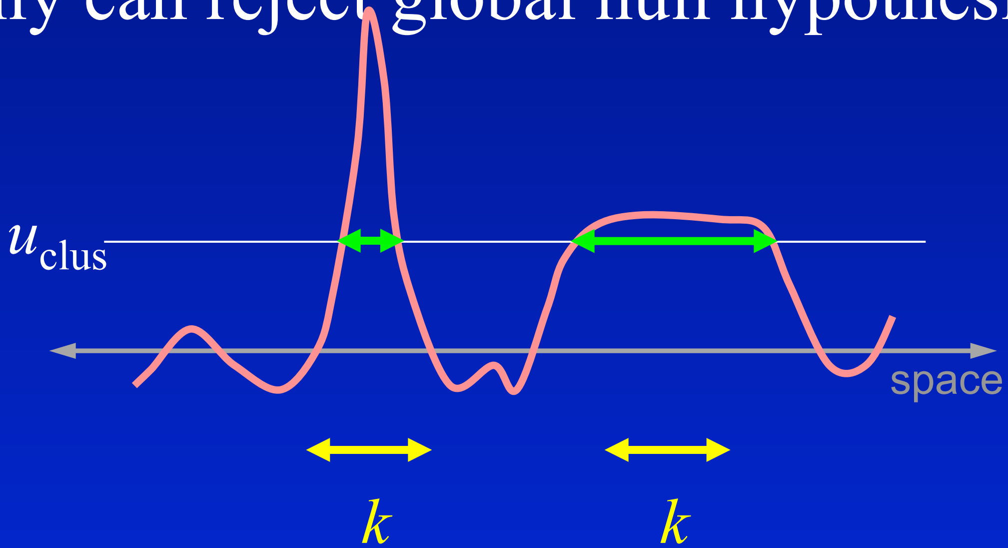
Cluster-level Inference

- Typically better sensitivity
- Worse spatial specificity
  - The null hyp. of entire cluster is rejected
  - Only means that *one or more* of voxels in cluster active



Set-level Inference

- Count number of blobs c
  - Minimum blob size k
- Worst spatial specificity
  - Only can reject global null hypothesis

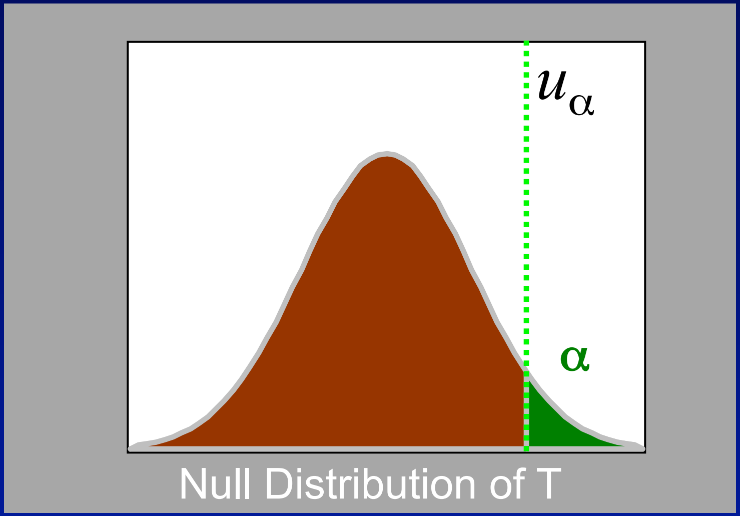


Here $c = 1$; only 1 cluster larger than k

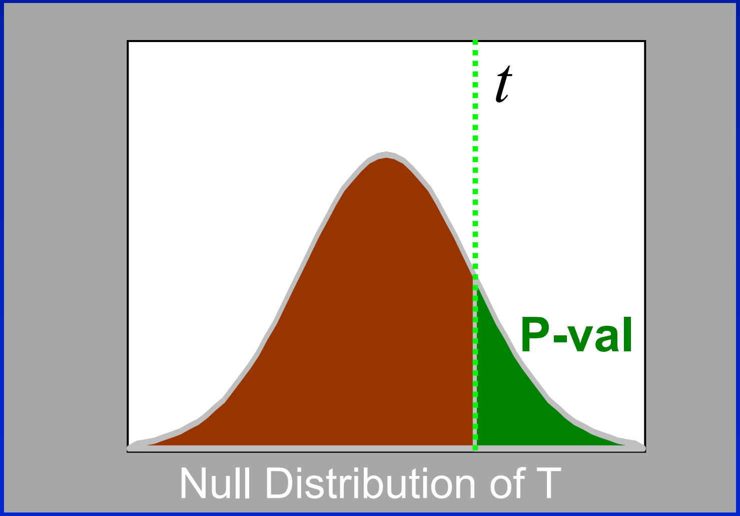
Multiple comparisons…

Hypothesis Testing

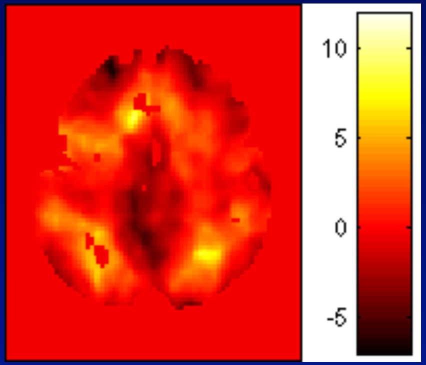
- Null Hypothesis H_0
- Test statistic T
  - t observed realization of T
- α level
  - Acceptable false positive rate
  - Level $\alpha = P(T>u_\alpha \mid H_0)$
  - Threshold u_α controls false positive rate at level α



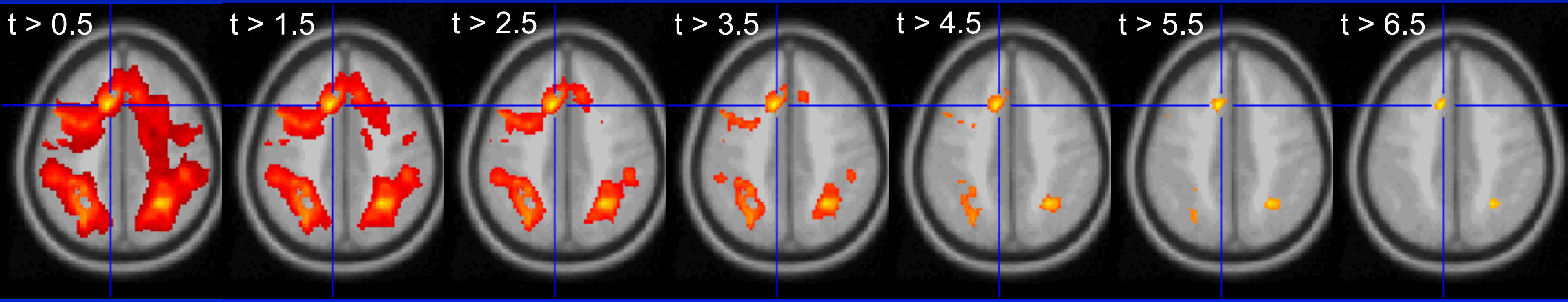
- P-value
  - Assessment of t assuming H_0
  - $P(T > t \mid H_0)$
    - Prob. of obtaining stat. as large or larger in a new experiment
  - P(Data|Null) *not* P(Null|Data)



Multiple Comparisons Problem



- Which of 100,000 voxels are sig.?
  - α=0.05 ⇒ 5,000 false positive voxels
- Which of (random number, say) 100 clusters significant?
  - α=0.05 ⇒ 5 false positives clusters



MCP Solutions: Measuring False Positives

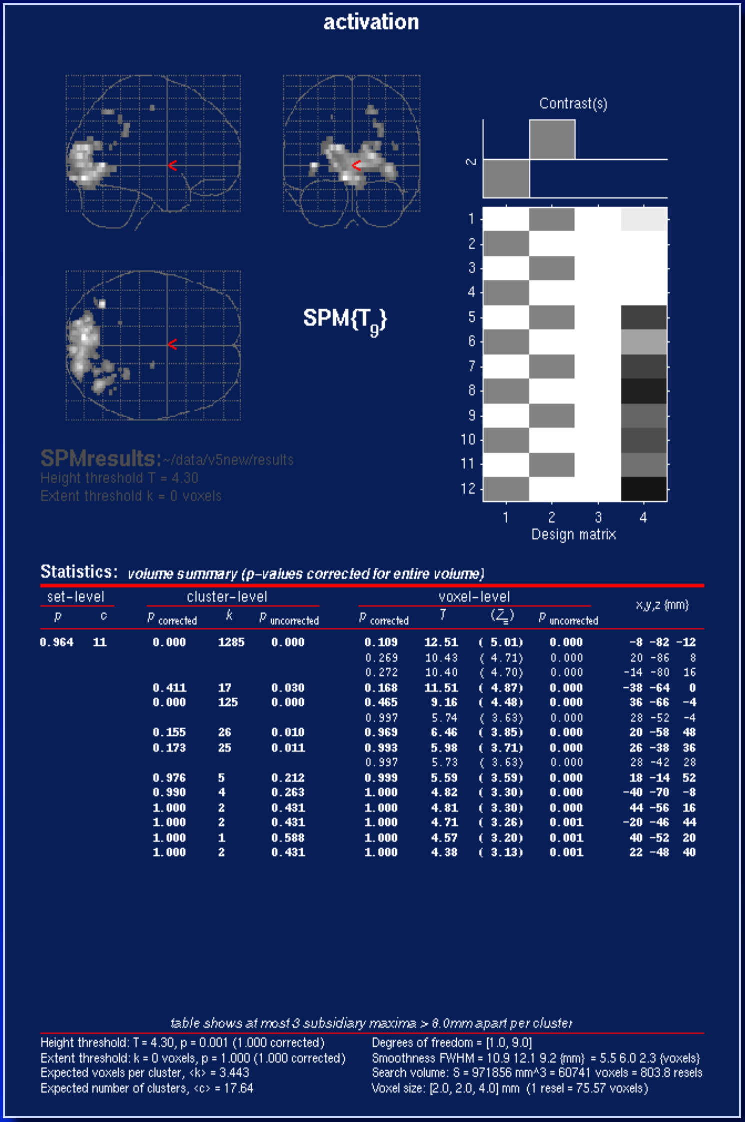
- Familywise Error Rate (FWER)
  - Familywise Error
    - Existence of one or more false positives
  - FWER is probability of familywise error
- False Discovery Rate (FDR)
  - FDR = E(V/R)
  - R voxels declared active, V falsely so
    - Realized false discovery rate: V/R

MCP Solutions: Measuring False Positives

- Familywise Error Rate (FWER)
  - Familywise Error
    - Existence of one or more false positives
  - FWER is probability of familywise error
- False Discovery Rate (FDR)
  - FDR = E(V/R)
  - R voxels declared active, V falsely so
    - Realized false discovery rate: V/R

FWE Multiple comparisons terminology…

- *Family* of hypotheses
  - $H^k \;\; k \in \Omega = \{1,\ldots,K\}$
  - $H^\Omega = \cap\, H^k$
- *Familywise* Type I error
  - *weak* control – *omnibus test*
    - Pr(“reject” H^Ω | H^Ω) ≤ α
    - *“anything, anywhere” ?*
  - *strong* control – *localising test*
    - Pr(“reject” H^W | H^W) ≤ α
      $\forall\;\; W: W \subseteq \Omega \;\&\; H^W$
    - *“anything, & where” ?*
- Adjusted *p*–values
  - test level at which reject H^k



set-level		cluster-level			voxel-level				x,y,z {mm}
p	c	p corrected	k	p uncorrected	p corrected	T	(Z≡)	p uncorrected	
0.964	11	0.000	1285	0.000	0.109	12.51	(5.01)	0.000	-8 -82 -12
					0.269	10.43	(4.71)	0.000	20 -86 8
					0.272	10.40	(4.70)	0.000	-14 -80 16
		0.411	17	0.030	0.168	11.51	(4.87)	0.000	-38 -64 0
		0.000	125	0.000	0.465	9.16	(4.48)	0.000	36 -66 -4
					0.997	5.74	(3.63)	0.000	28 -52 -4
		0.155	26	0.010	0.969	6.46	(3.85)	0.000	20 -58 48
		0.173	25	0.011	0.993	5.98	(3.71)	0.000	26 -38 36
					0.997	5.73	(3.63)	0.000	28 -42 28
		0.976	5	0.212	0.999	5.59	(3.59)	0.000	18 -14 52
		0.990	4	0.263	1.000	4.82	(3.30)	0.000	-40 -70 -8
		1.000	2	0.431	1.000	4.81	(3.30)	0.000	44 -56 16
		1.000	2	0.431	1.000	4.71	(3.26)	0.001	-20 -46 44
		1.000	1	0.588	1.000	4.57	(3.20)	0.001	40 -52 20
		1.000	2	0.431	1.000	4.38	(3.13)	0.001	22 -48 40

FWE MCP Solutions: Bonferroni

- For a statistic image T...
  - T_i i^{th} voxel of statistic image T
- ...use $\alpha = \alpha_0/V$
  - α_0 FWER level (e.g. 0.05)
  - V number of voxels
  - u_α α-level statistic threshold, $P(T_i \geq u_\alpha) = \alpha$
- By Bonferroni inequality...

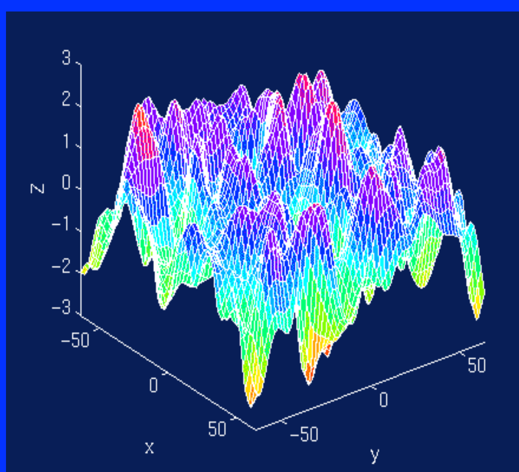
$$
\begin{aligned}
\text{FWER} &= P(\text{FWE}) \\
&= P(\cup_i \{T_i \geq u_\alpha\} \mid H_0) \\
&\leq \textstyle\sum_i P(T_i \geq u_\alpha \mid H_0) \\
&= \textstyle\sum_i \alpha \\
&= \textstyle\sum_i \alpha_0 / V = \alpha_0
\end{aligned}
$$

Conservative under correlation	
Independent:	*V* tests
Some dep.:	? tests
Total dep.:	1 test

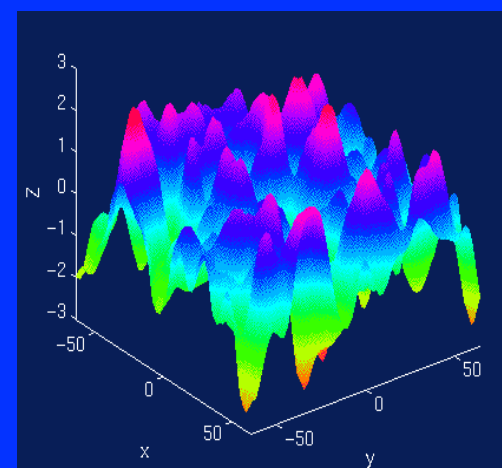
Random field theory…

SPM approach: Random fields…

- Consider statistic image as lattice representation of a continuous random field
- Use results from continuous random field theory



≈
lattice represtntation



FWER MCP Solutions: Controlling FWER w/ Max

- FWER & distribution of maximum

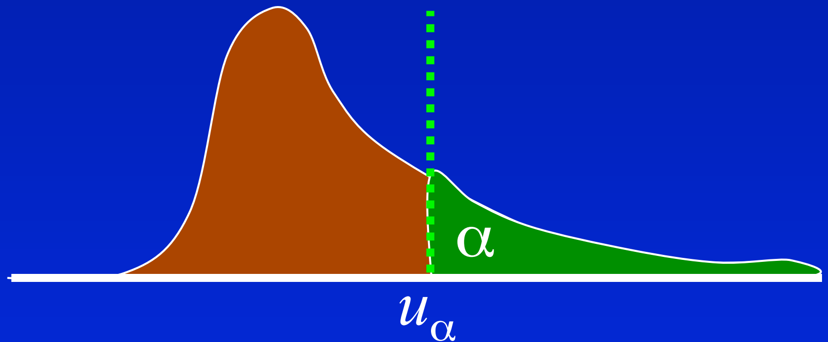
$$
\begin{aligned}
\text{FWER} &= \text{P(FWE)} \\
&= \text{P}(\cup_i \{T_i \geq u\} \mid H_o) \\
&= \text{P}(\max_i T_i \geq u \mid H_o)
\end{aligned}
$$

- 100(1-α)%ile of max dist^n controls FWER

$$\text{FWER} = \text{P}(\max_i T_i \geq u_\alpha \mid H_o) = \alpha$$

  – where

$$u_\alpha = F^{-1}_{\max} (1-\alpha)$$



FWER MCP Solutions: Random Field Theory

- Euler Characteristic χ_u
  - Topological Measure
    - #blobs - #holes
  - At high thresholds, just counts blobs
  - FWER = P(Max voxel ≥ u | H_o)

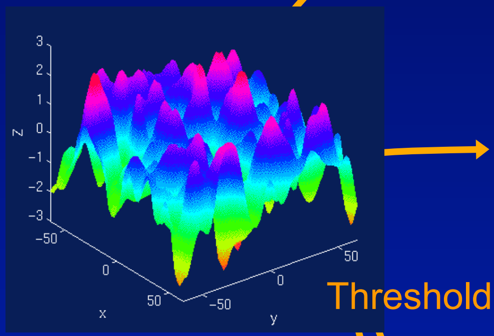
    = P(One or more blobs | H_o)

    ≈ $P(\chi_u \geq 1 \mid H_o)$

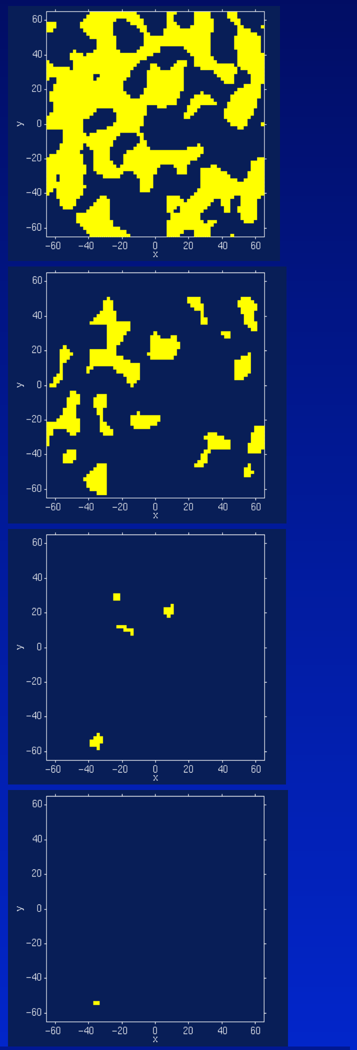
    ≈ $E(\chi_u \mid H_o)$

No holes

Never more than 1 blob



Random Field

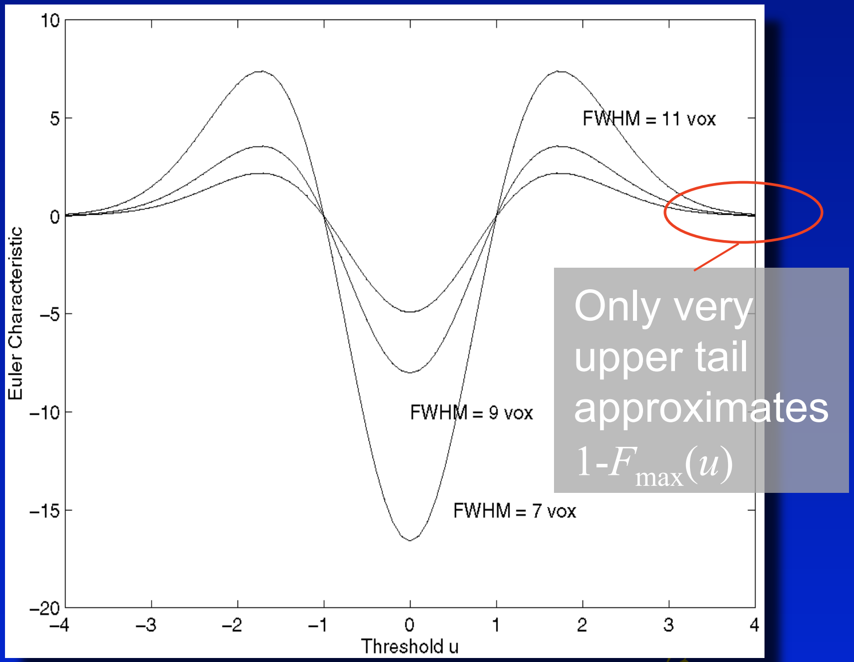


Suprathreshold Sets

RFT Details: Expected Euler Characteristic

$$E(\chi_u) \approx \lambda(\Omega)\ |\Lambda|^{1/2}\ (u^2 - 1)\ \exp(-u^2/2)\ /\ (2\pi)^2$$

- Ω → Search region $\Omega \subset \mathcal{R}^3$
- $\lambda(\Omega)$ → volume
- $|\Lambda|^{1/2}$ → roughness



- Assumptions
  - Multivariate Normal
  - Stationary*
  - ACF twice differentiable at 0

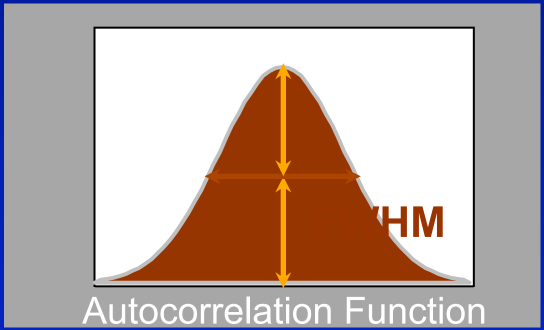
* Stationary
  - Results valid w/out stationary
  - More accurate when stat. holds

Random Field Theory Smoothness Parameterization

- $E(\chi_u)$ depends on $|\Lambda|^{1/2}$
  - Λ roughness matrix:

$$\Lambda = \mathbf{Var}\left(\frac{\partial G}{\partial(x,y,z)}\right)$$
$$= \begin{pmatrix} \mathbf{Var}\left(\frac{\partial G}{\partial x}\right) & \mathbf{Cov}\left(\frac{\partial G}{\partial x},\frac{\partial G}{\partial y}\right) & \mathbf{Cov}\left(\frac{\partial G}{\partial x},\frac{\partial G}{\partial z}\right) \\ \mathbf{Cov}\left(\frac{\partial G}{\partial y},\frac{\partial G}{\partial x}\right) & \mathbf{Var}\left(\frac{\partial G}{\partial y}\right) & \mathbf{Cov}\left(\frac{\partial G}{\partial y},\frac{\partial G}{\partial z}\right) \\ \mathbf{Cov}\left(\frac{\partial G}{\partial z},\frac{\partial G}{\partial x}\right) & \mathbf{Cov}\left(\frac{\partial G}{\partial z},\frac{\partial G}{\partial y}\right) & \mathbf{Var}\left(\frac{\partial G}{\partial z}\right) \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$$

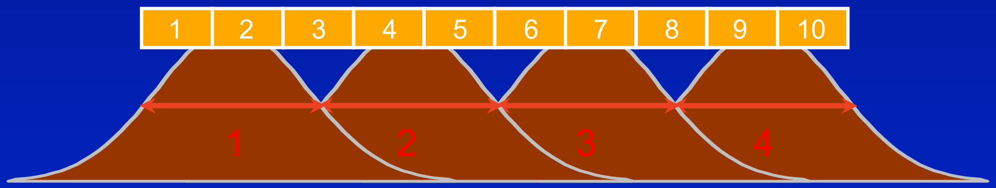
- Smoothness parameterized as Full Width at Half Maximum
  - FWHM of Gaussian kernel needed to smooth a white noise random field to roughness Λ



$$|\Lambda|^{1/2} = \frac{(4\log 2)^{3/2}}{\mathrm{FWHM}_x \mathrm{FWHM}_y \mathrm{FWHM}_z}.$$

Random Field Theory Smoothness Parameterization

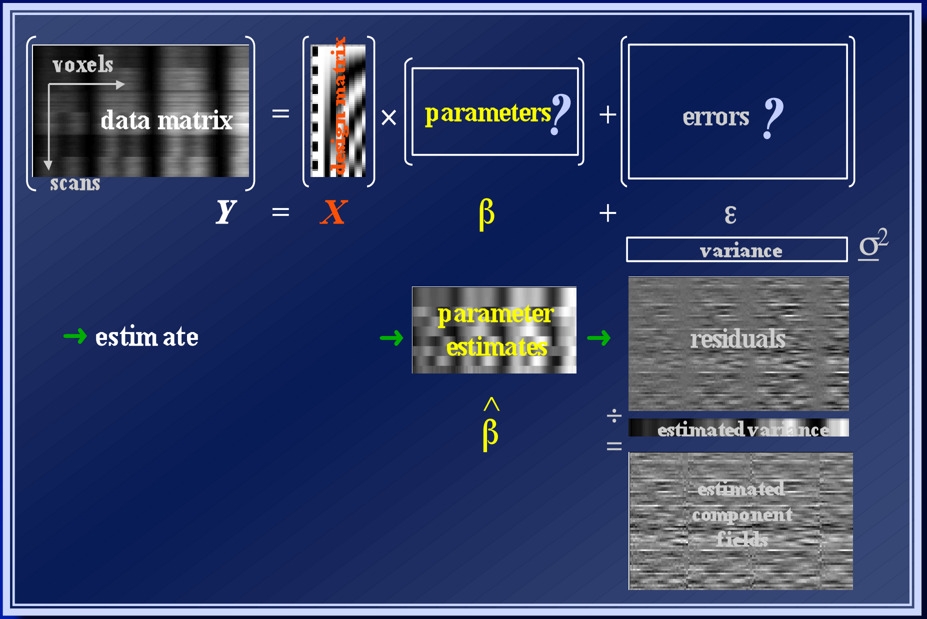
- RESELS
  - Resolution Elements
  - 1 RESEL = $FWHM_x \times FWHM_y \times FWHM_z$
  - RESEL Count R
    - $R = \lambda(\Omega) \sqrt{|\Lambda|} = (4\log 2)^{3/2} \lambda(\Omega) \; / \; (FWHM_x \times FWHM_y \times FWHM_z)$
    - Volume of search region in units of smoothness
    - Eg: 10 voxels, 2.5 FWHM 4 RESELS



- Beware RESEL misinterpretation
  - RESEL *are not* "number of independent 'things' in the image"
    - See Nichols & Hayasaka, 2003, Stat. Meth. in Med. Res.

Random Field Theory Smoothness Estimation

- Smoothness est'd from standardized residuals
  - Variance of gradients
  - Yields resels per voxel (RPV)
- RPV image
  - Local roughness est.
  - Can transform in to local smoothness est.
    - FWHM Img = (RPV Img)$^{-1/D}$
    - Dimension D, e.g. D=2 or 3



Random Field Intuition

- Corrected P-value for voxel value t

$$\begin{aligned} P^c &= \mathrm{P}(\max T > t) \\ &\approx \mathrm{E}(\chi_t) \\ &\approx \lambda(\Omega)\, |\Lambda|^{1/2}\, t^2 \exp(-t^2/2) \end{aligned}$$

- Statistic value t increases
  - P^c decreases (but only for large t)
- Search volume increases
  - P^c increases (more severe MCP)
- Roughness increases (Smoothness decreases)
  - P^c increases (more severe MCP)

RFT Details: Unified Formula

- General form for expected Euler characteristic
  - χ^2, F, & t fields • restricted search regions • D dimensions •

$$E[\chi_u(\Omega)] = \Sigma_d \, R_d(\Omega) \, \rho_d(u)$$

$R_d(\Omega)$: d-dimensional Minkowski functional of Ω

– function of dimension, space Ω and smoothness:

$R_0(\Omega) = \chi(\Omega)$ Euler characteristic of Ω

$R_1(\Omega)$ = resel diameter

$R_2(\Omega)$ = resel surface area

$R_3(\Omega)$ = resel volume

$\rho_d(\Omega)$: d-dimensional EC density of $Z(\underline{x})$

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

E.g. Gaussian RF:

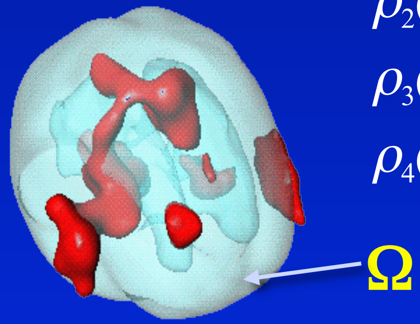
$\rho_0(u) = 1 - \Phi(u)$

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

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

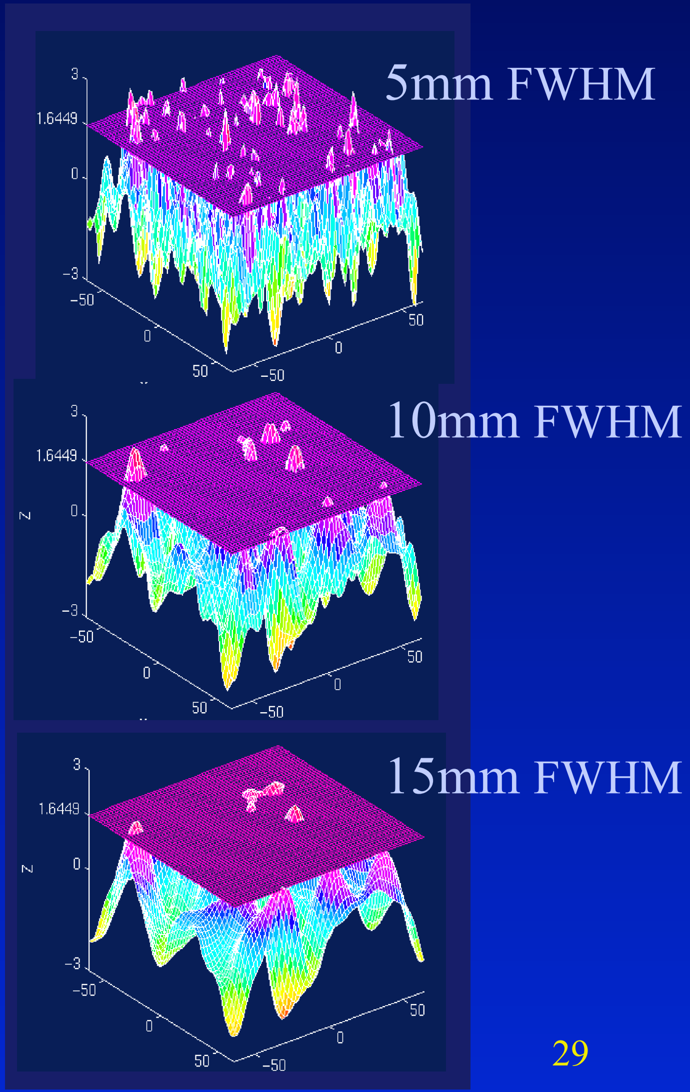
$\rho_3(u) = (4 \ln 2)^{3/2} (u^2 - 1) \quad \exp(-u^2/2) / (2\pi)^2$

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



Random Field Theory
Cluster Size Tests

- Expected Cluster Size
  - E(S) = E(N)/E(L)
  - S cluster size
  - N suprathreshold volume $\lambda(\{T > u_{\text{clus}}\})$
  - L number of clusters
- $E(N) = \lambda(\Omega)\ P(\ T > u_{clus}\ )$
- $E(L) \approx E(\chi_u)$
  - Assuming no holes



Random Field Theory Cluster Size Distribution

- Gaussian Random Fields (Nosko, 1969)

$$S^{2/D} \sim Exp\left(\left[\frac{E(N)}{\Gamma(D/2+1)E(L)}\right]^{-2/D}\right)$$

  - D: Dimension of RF

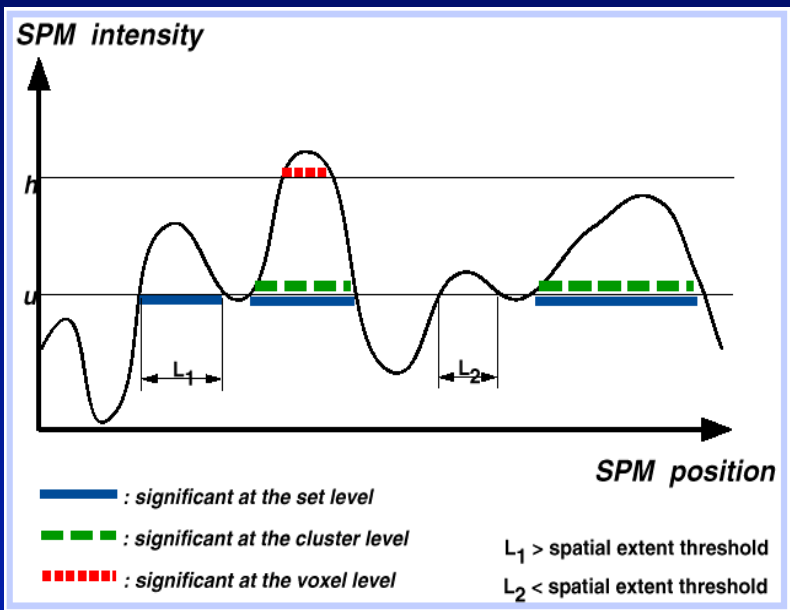
- t Random Fields (Cao, 1999)
  - B: Beta dist^n
  - U's: χ^2's
  - c chosen s.t. E(S) = E(N) / E(L)

$$S \sim cB^{1/2}\left[\frac{U_0^D}{\Pi_{b=0}^D U_b}\right]^{2/D}$$

Random Field Theory Cluster Size Corrected P-Values

- Previous results give uncorrected P-value
- Corrected P-value
  - Bonferroni
    - Correct for expected number of clusters
    - Corrected $P^c = E(L)\, P^{uncorr}$
  - Poisson Clumping Heuristic (Adler, 1980)
    - Corrected $P^c = 1 - \exp(-E(L)\, P^{uncorr})$

Review: Levels of inference & power



Sensitivity	Test based on	Parameters set by the user	Regional specificity
−	The intensity of a voxel	• Low pass filter	+
	The spatial extent above u or the spatial extent and the maximum peak height	• Low pass filter • intensity threshold u	
	The number of clusters above u with size greater than n	• Low pass filter • intensity thres. u • spatial threshold n	
+	The sum of square of the SPM or a MANOVA	• Low pass filter	−

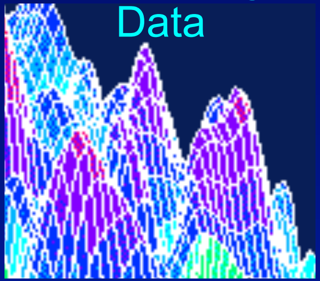
Set level…

Cluster level…

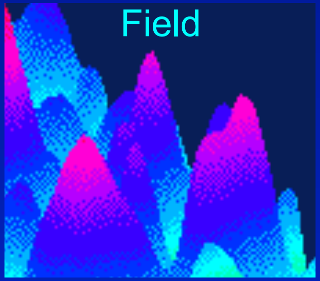
Voxel level…

Random Field Theory Limitations

- Sufficient smoothness
  - FWHM smoothness 3-4× voxel size (Z)
  - More like ~10× for low-df T images
- Smoothness estimation
  - Estimate is biased when images not sufficiently smooth
- Multivariate normality
  - Virtually impossible to check
- Several layers of approximations
- Stationary required for cluster size results

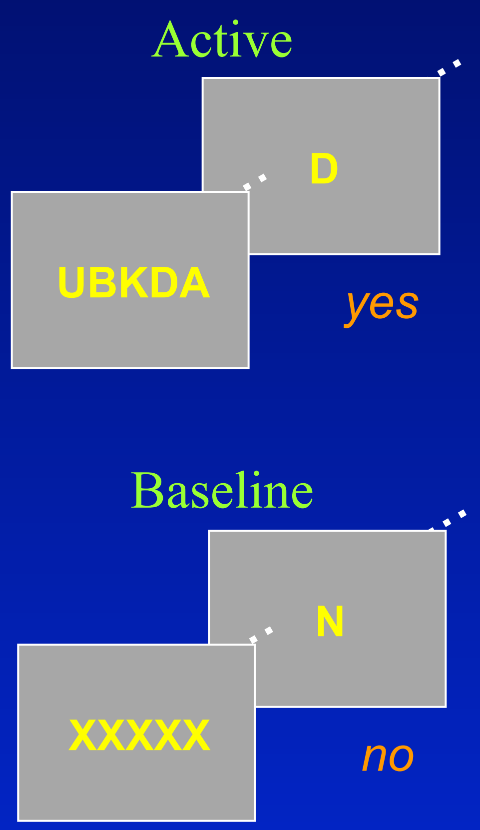


≈



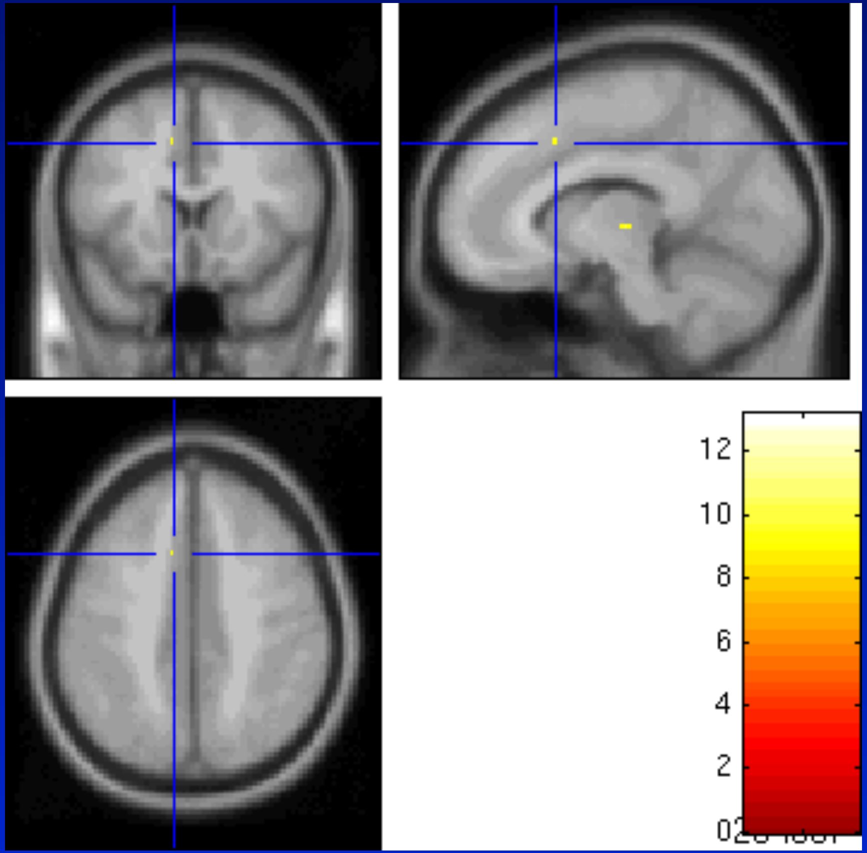
Real Data

- 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 *t* test

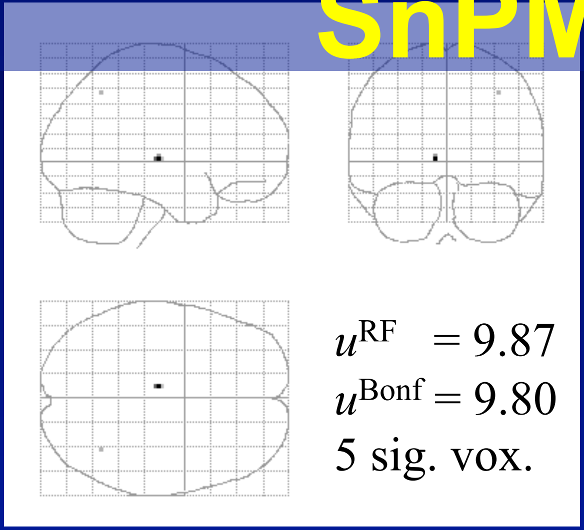


Real Data: RFT Result

- Threshold
  - $S = 110{,}776$
  - $2 \times 2 \times 2$ voxels $5.1 \times 5.8 \times 6.9$ mm FWHM
  - $u = 9.870$
- Result
  - 5 voxels above the threshold
  - 0.0063 minimum FWE-corrected p-value

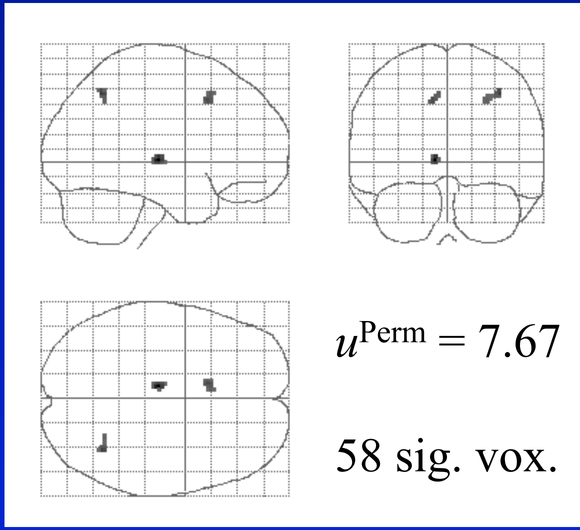


Real Data: SnPM Promotional

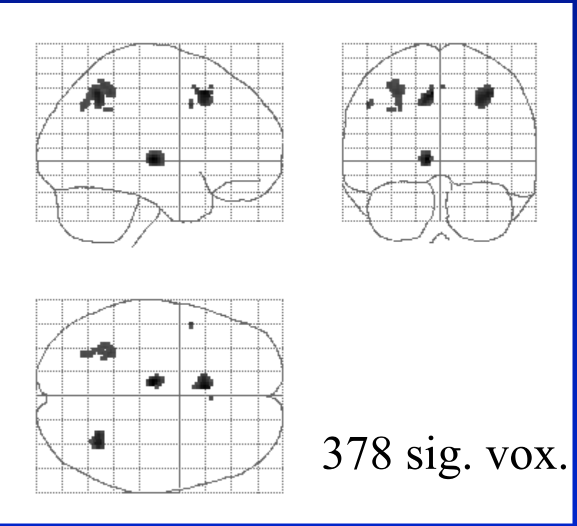


t_{11} Statistic, RF & Bonf. Threshold

- Nonparametric method more powerful than RFT for low DF
- "Variance Smoothing" even more sensitive
- FWE controlled all the while!



t_{11} Statistic, Nonparametric Threshold



Smoothed Variance *t* Statistic, Nonparametric Threshold

False Discovery Rate…

MCP Solutions: Measuring False Positives

- Familywise Error Rate (FWER)
  - Familywise Error
    - Existence of one or more false positives
  - FWER is probability of familywise error
- False Discovery Rate (FDR)
  - $FDR = E(V/R)$
  - R voxels declared active, V falsely so
    - Realized false discovery rate: V/R

False Discovery Rate

- For any threshold, all voxels can be cross-classified:

	Accept Null	Reject Null	
Null True	V_{0A}	V_{0R}	m_0
Null False	V_{1A}	V_{1R}	m_1
	N_A	N_R	V

- Realized FDR

$$rFDR = V_{0R}/(V_{1R}+V_{0R}) = V_{0R}/N_R$$

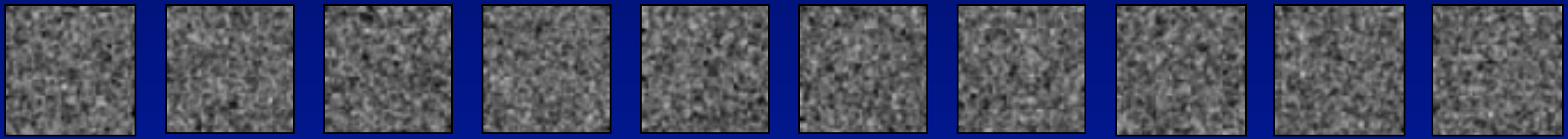
  – If $N_R = 0$, rFDR = 0

- But only can observe N_R, don't know V_{1R} & V_{0R}
  – We control the *expected* rFDR

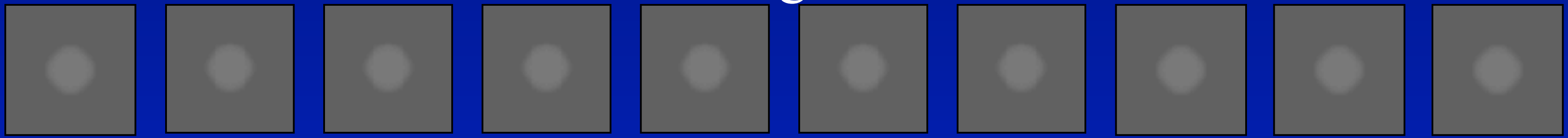
$$FDR = E(rFDR)$$

False Discovery Rate Illustration:

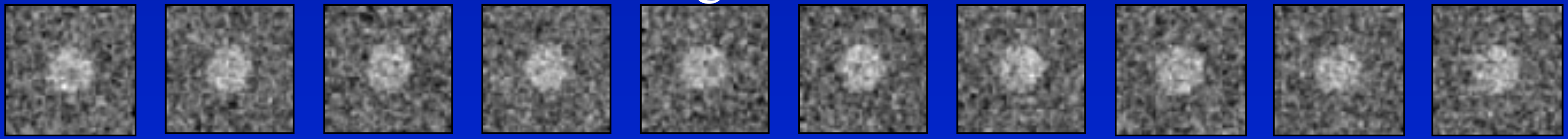
Noise



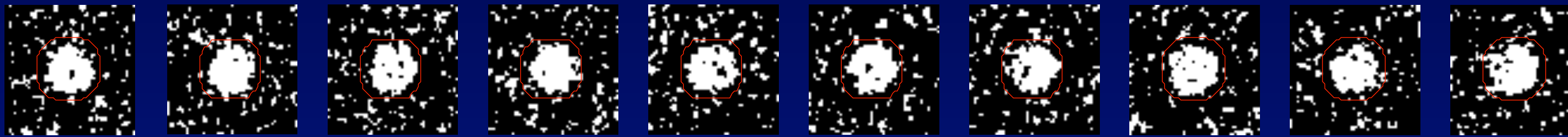
Signal



Signal+Noise



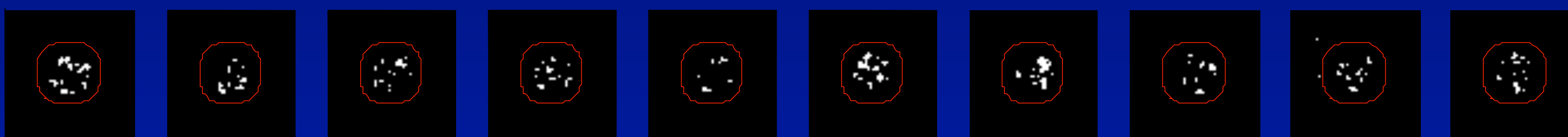
Control of Per Comparison Rate at 10%



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

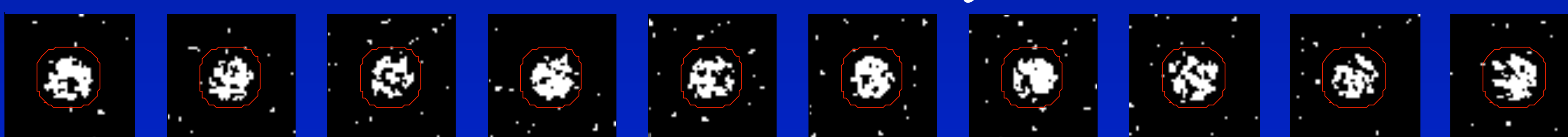
Control of Familywise Error Rate at 10%



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

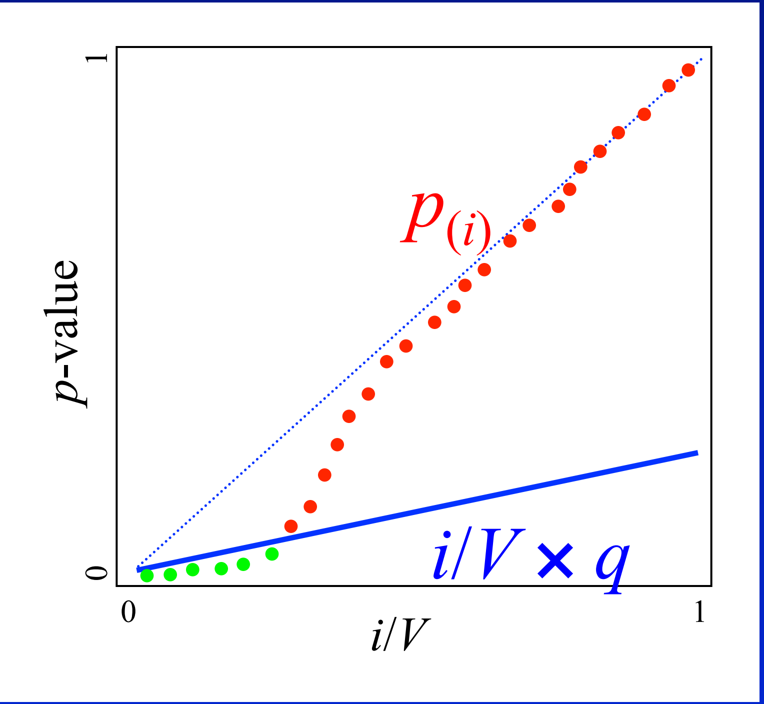
Benjamini & Hochberg Procedure

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

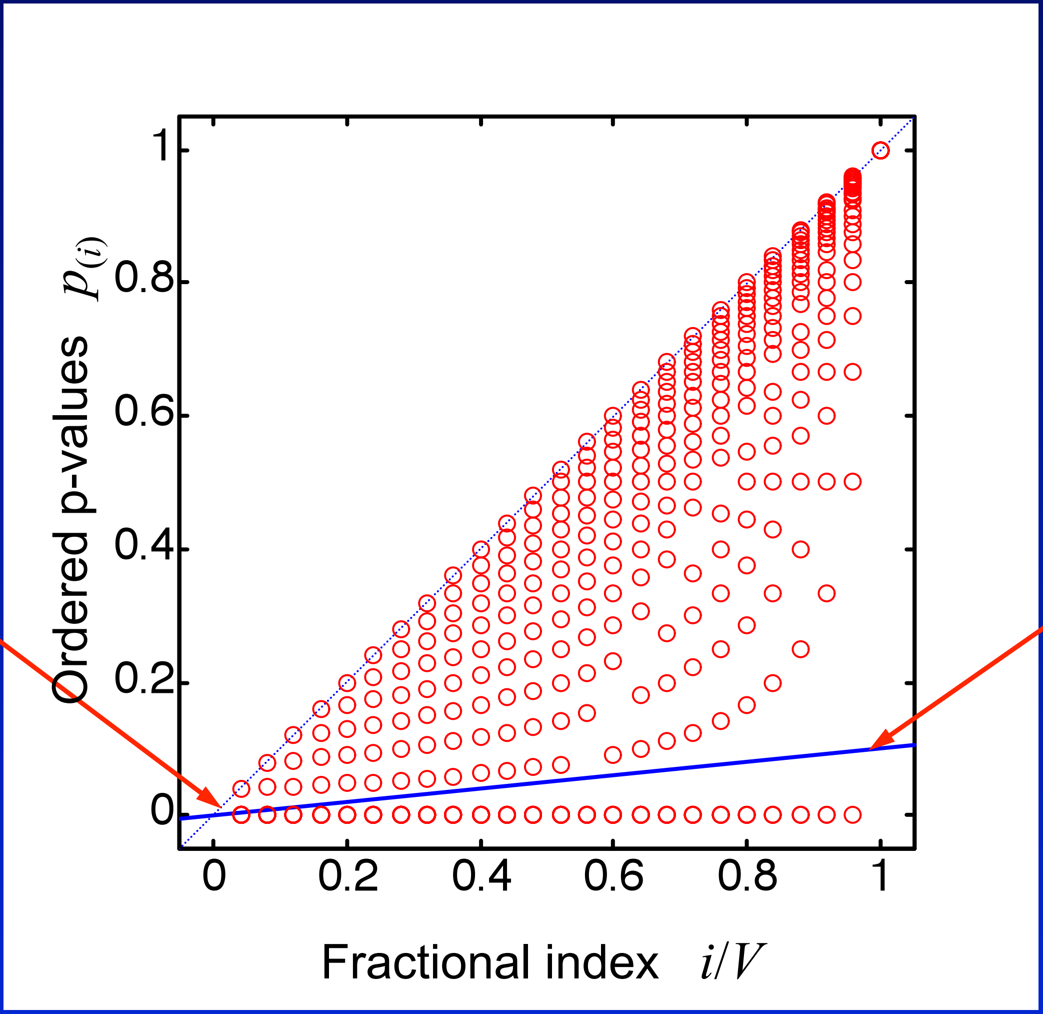
$$p_{(i)} \leq i/V \times q$$

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

JRSS-B (1995)
57:289-300



Adaptiveness of Benjamini & Hochberg FDR

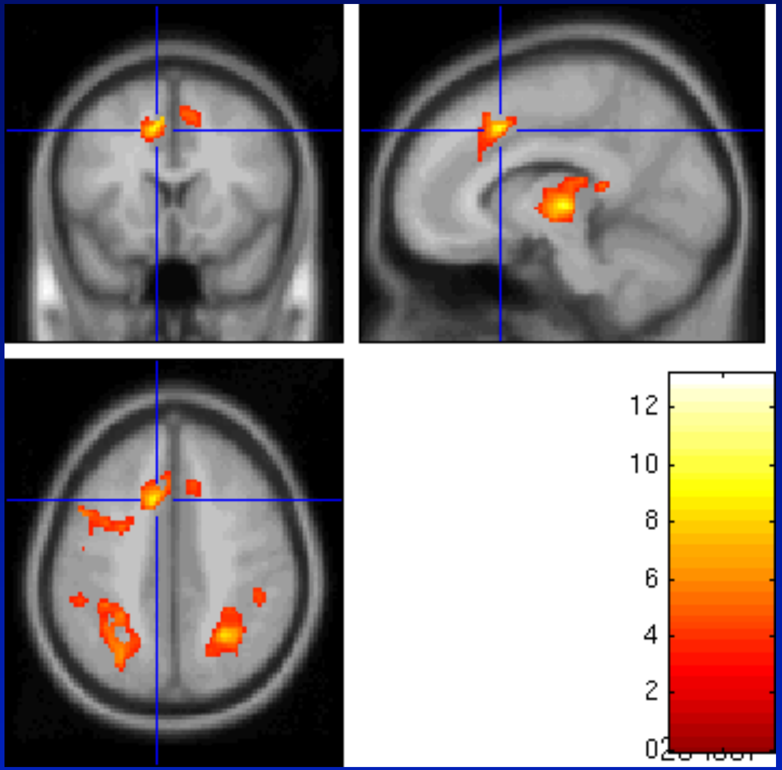


P-value threshold when no signal: α/V

P-value threshold when all signal: α

Real Data: FDR Example

- Threshold
  - Indep/PosDep
    $u = 3.83$
  - Arb Cov
    $u = 13.15$
- Result
  - 3,073 voxels above Indep/PosDep u
  - <0.0001 minimum FDR-corrected p-value



FDR Threshold = 3.83
3,073 voxels
FWER Perm. Thresh. = 9.87
7 voxels

FDR Changes

- Before SPM8
  - Only voxel-wise FDR
- SPM8
  - Cluster-wise FDR
  - Peak-wise FDR

Item Recognition data

Cluster-forming threshold P=0.001
    Cluster-wise FDR: 40 voxel cluster, PFDR 0.07
    Peak-wise FDR: t=4.84, PFDR 0.836
Cluster-forming threshold P=0.01
    Cluster-wise FDR: 1250 - 4380 voxel clusters, PFDR <0.001
    Cluster-wise FDR: 80 voxel cluster, PFDR 0.516
    Peak-wise FDR: t=4.84, PFDR 0.027

Benjamini & Hochberg Procedure Details

- Standard Result
  - Positive Regression Dependency on Subsets

    P($X_1 \geq c_1$, $X_2 \geq c_2$, ..., $X_k \geq c_k$ | $X_i = x_i$) is non-decreasing in x_i

    - Only required of null x_i's
      - Positive correlation between null voxels
      - Positive correlation between null and signal voxels
    - Special cases include
      - Independence
      - Multivariate Normal with all positive correlations
- Arbitrary covariance structure
  - Replace q by $q/c(V)$,

    $c(V) = \Sigma_{i=1,\ldots,V}\, 1/i \approx \log(V) + 0.5772$
  - Much more stringent

Benjamini & Yekutieli (2001). *Ann. Stat.* 29:1165-1188

Benjamini & Hochberg: Key Properties

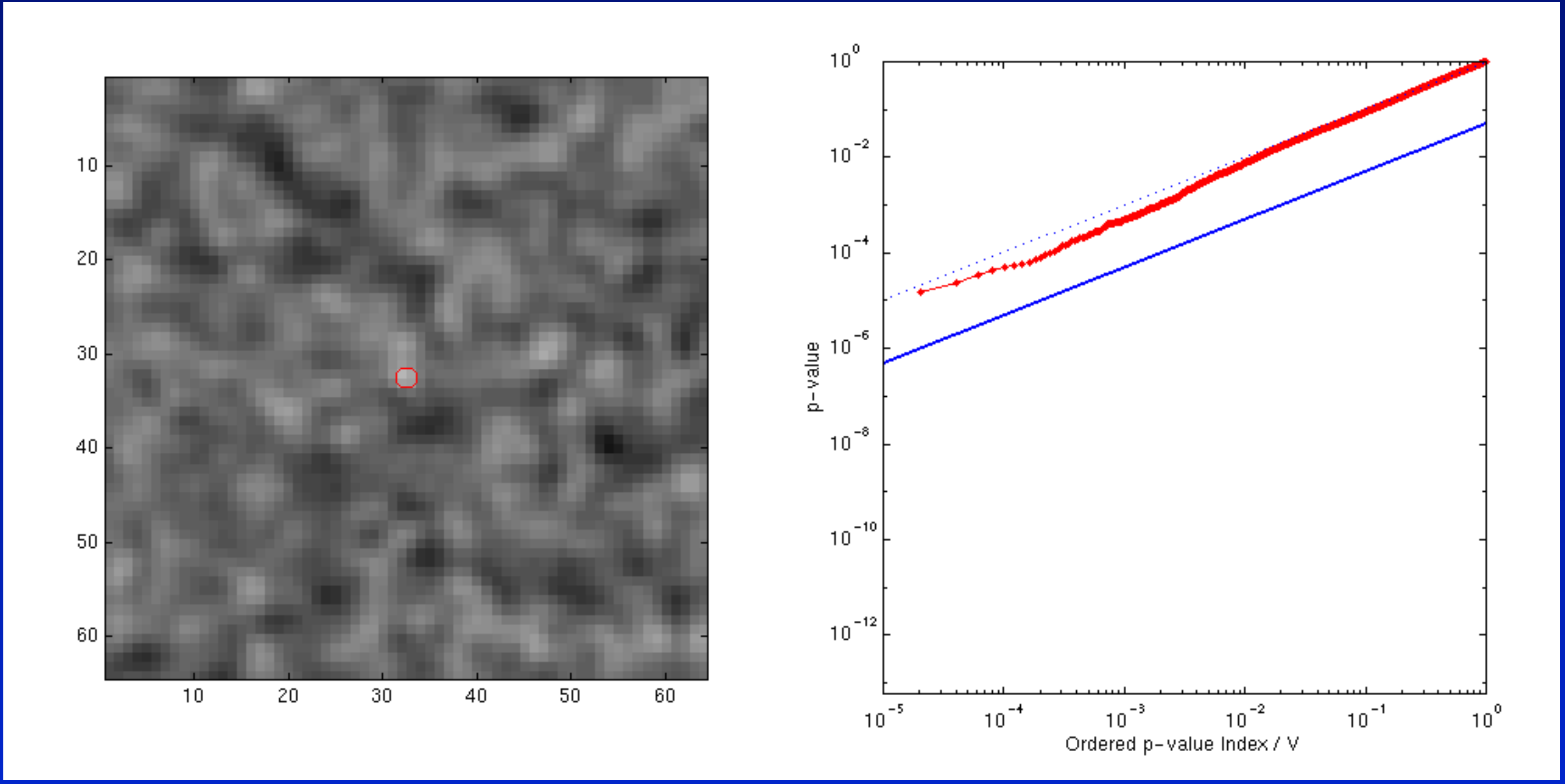
- FDR is controlled

  $$E(\mathrm{rFDR}) \leq q\, m_0/V$$

  - Conservative, if large fraction of nulls false
- Adaptive
  - Threshold depends on amount of signal
    - More signal, More small p-values, More $p_{(i)}$ less than $i/V \times q/c(V)$

Controlling FDR: Varying Signal Extent

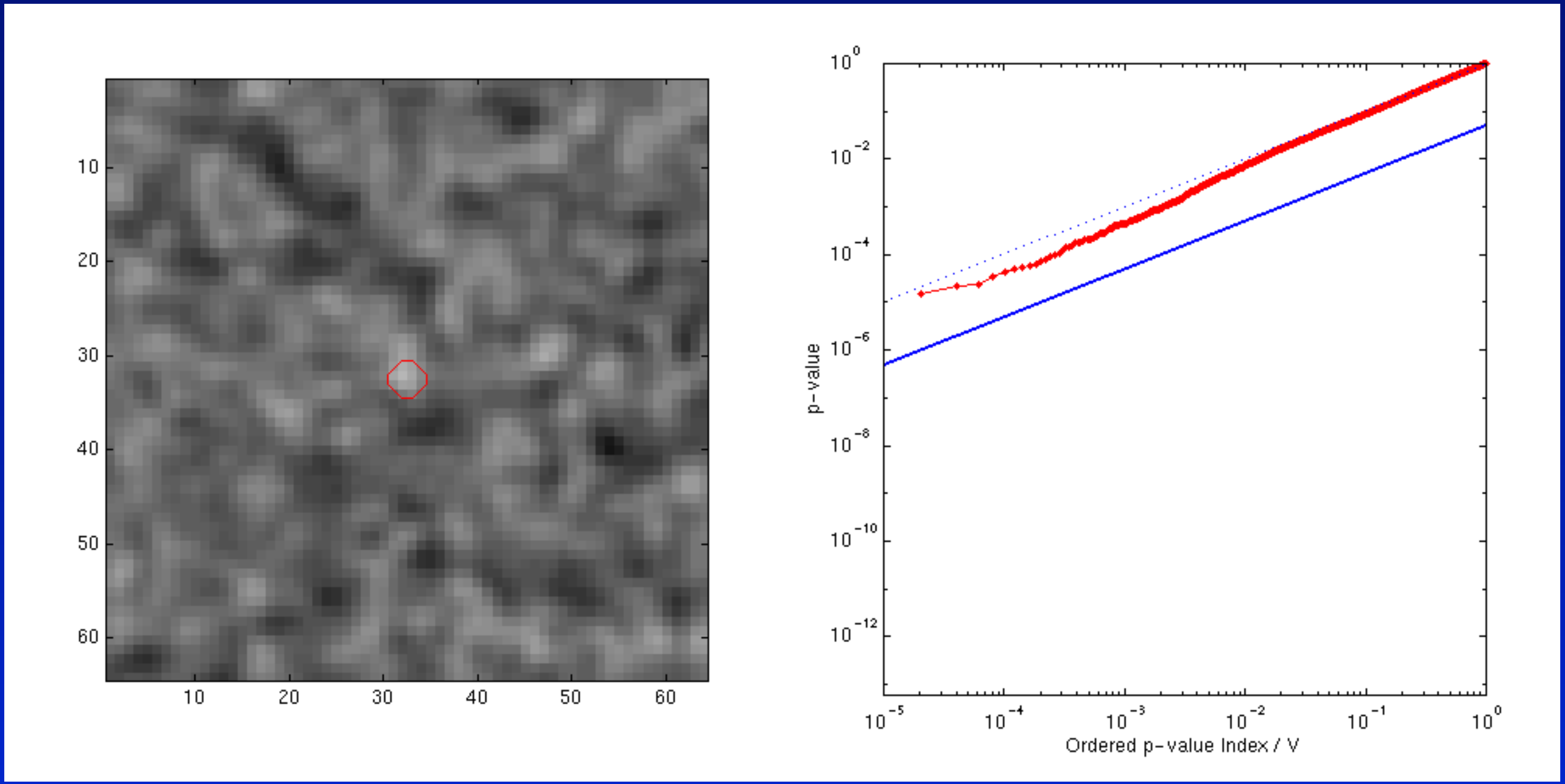
$p =$ $z =$



Signal Intensity 3.0 Signal Extent 1.0 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

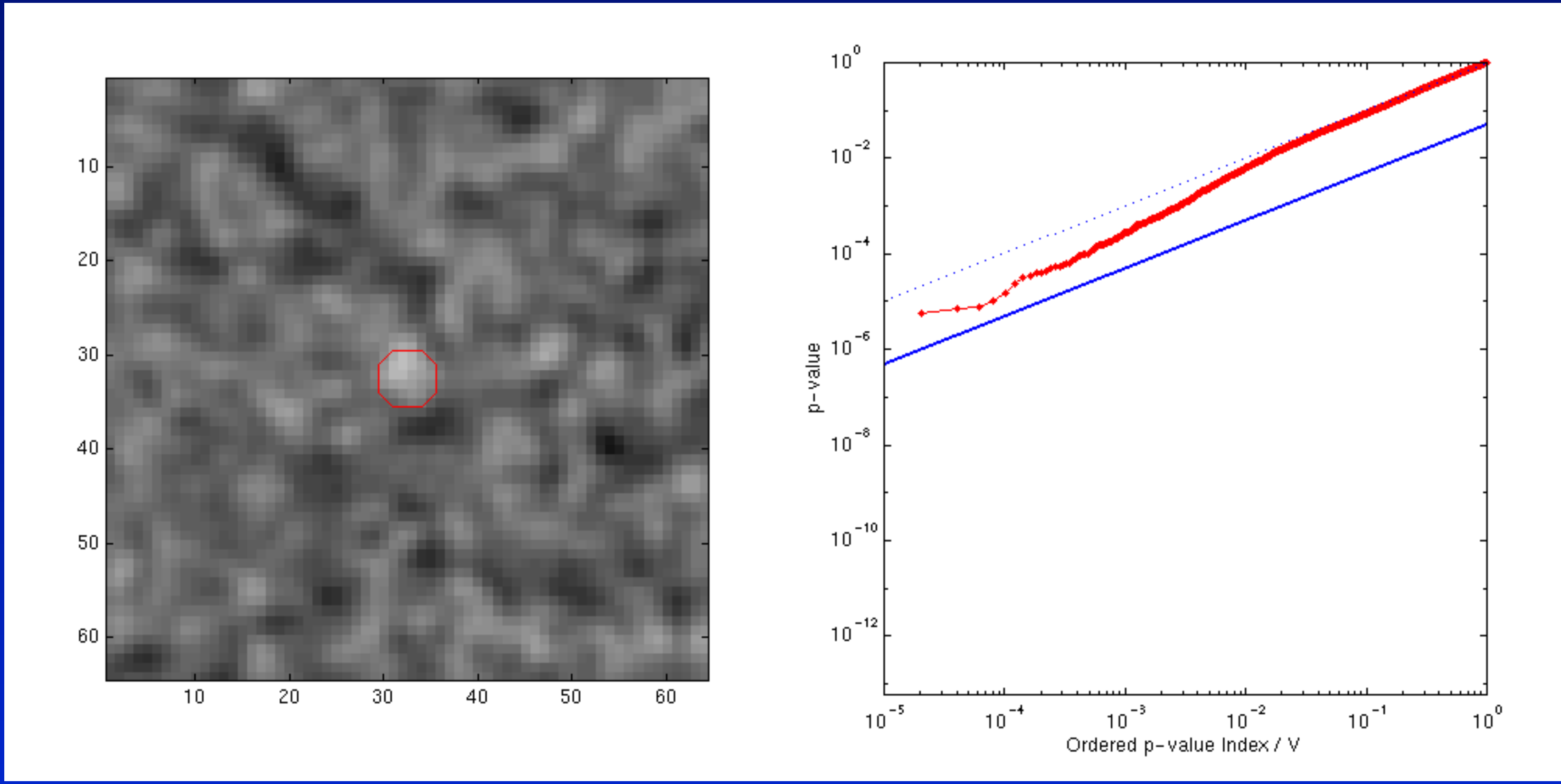
$p =$ $z =$



Signal Intensity 3.0 Signal Extent 2.0 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

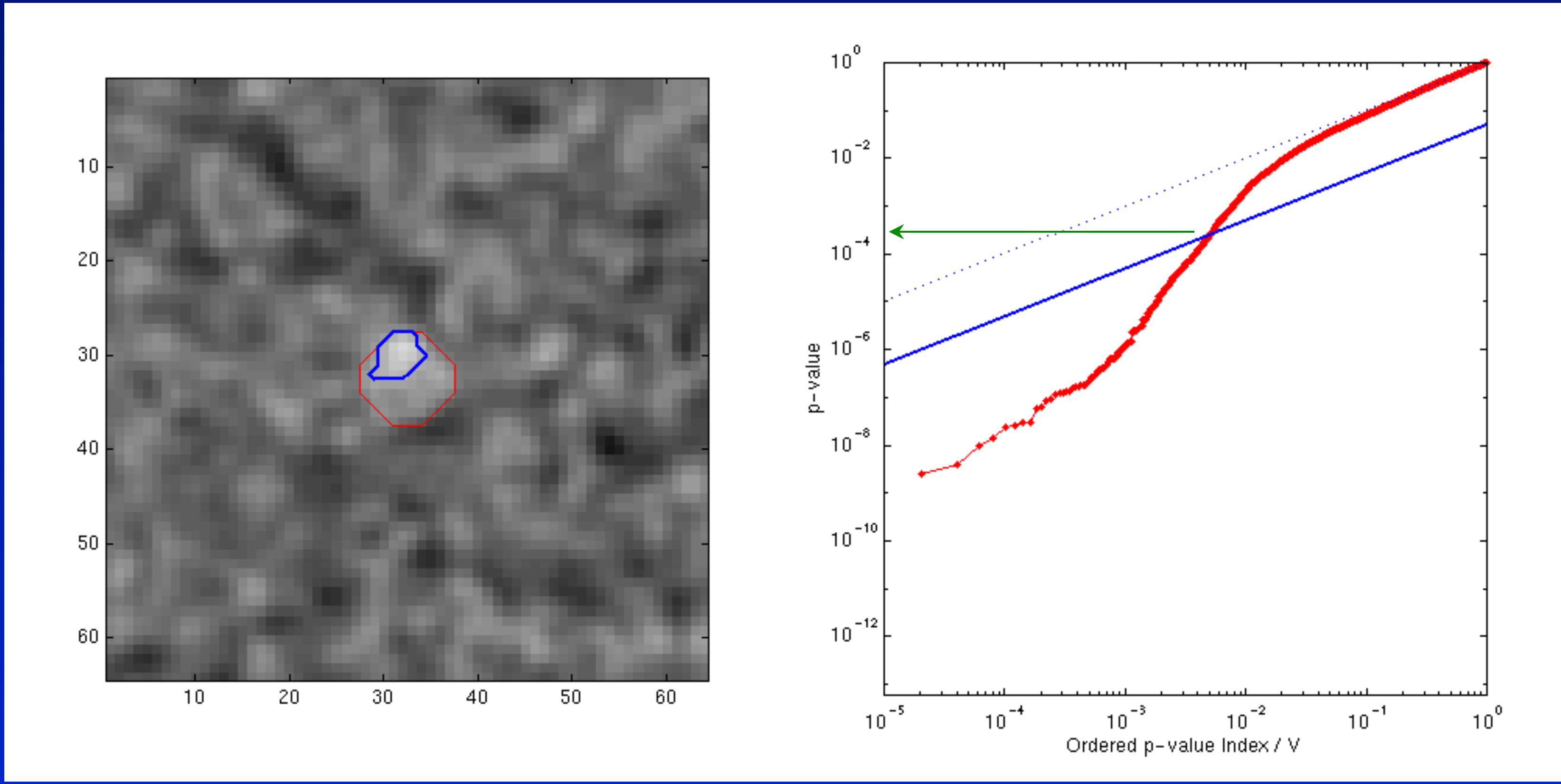
$p =$ $z =$



Signal Intensity 3.0 Signal Extent 3.0 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

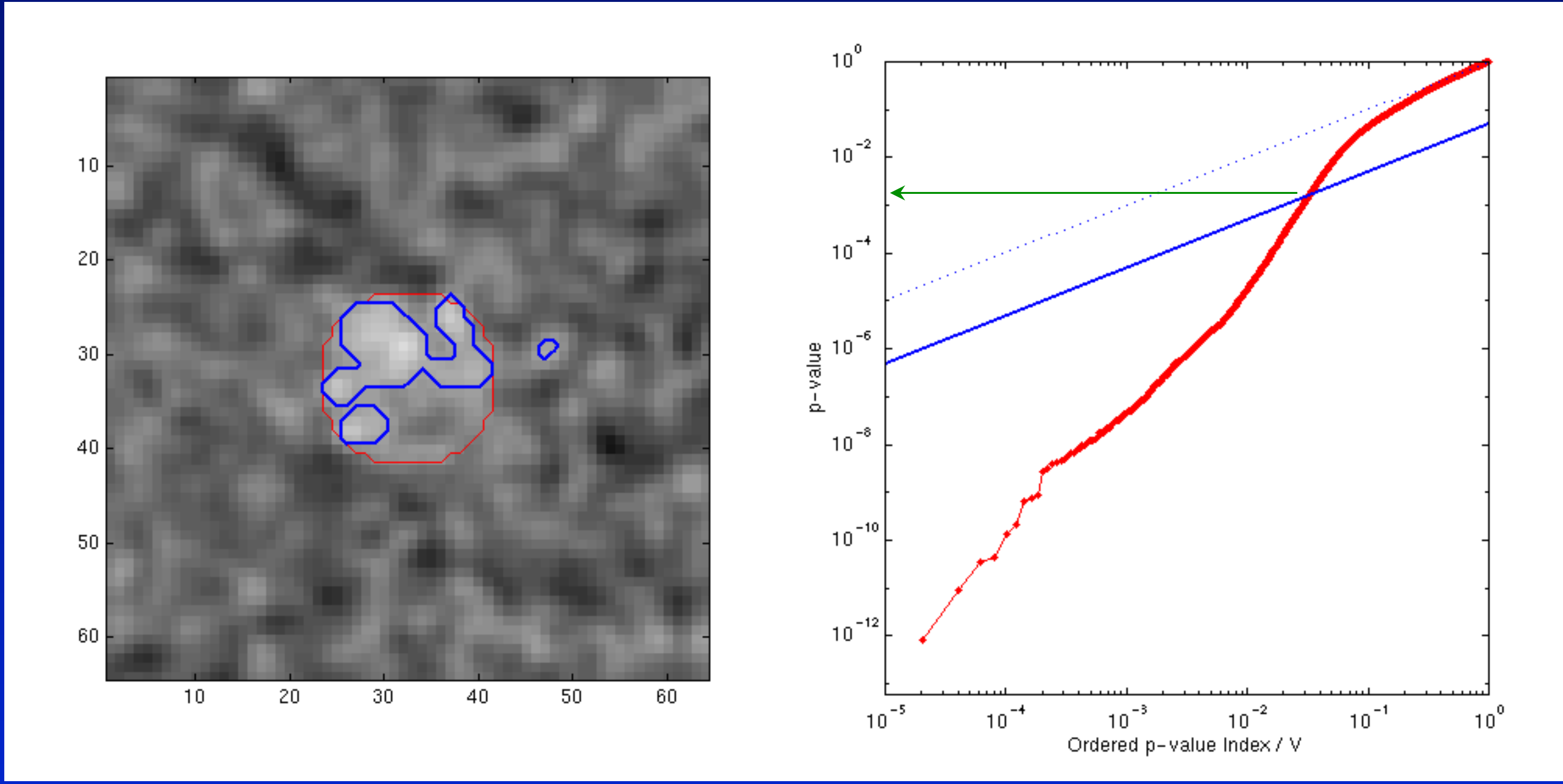
$p = 0.000252$ $z = 3.48$



Signal Intensity 3.0 Signal Extent 5.0 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

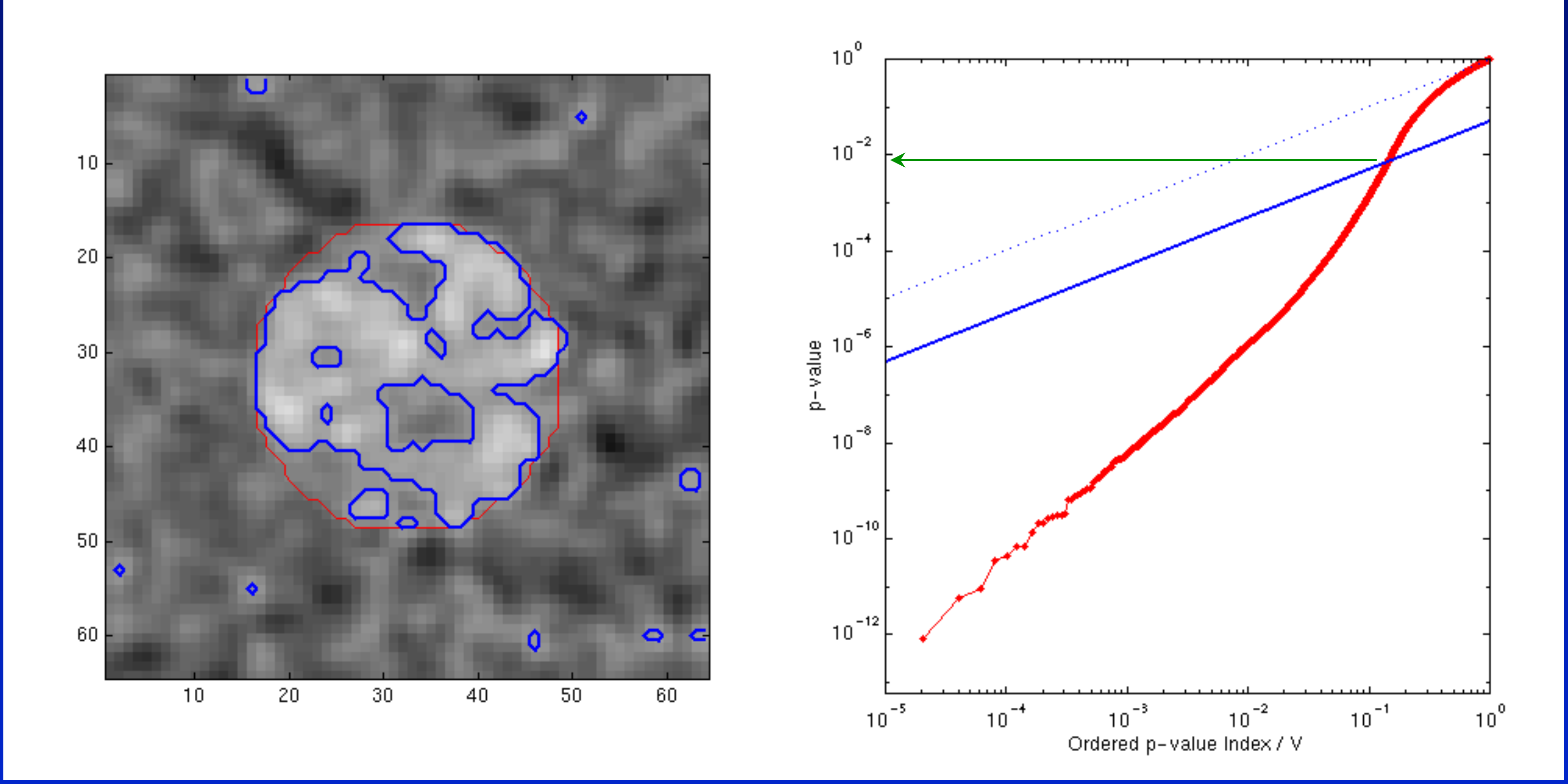
$p = 0.001628 \qquad z = 2.94$



Signal Intensity 3.0 Signal Extent 9.5 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

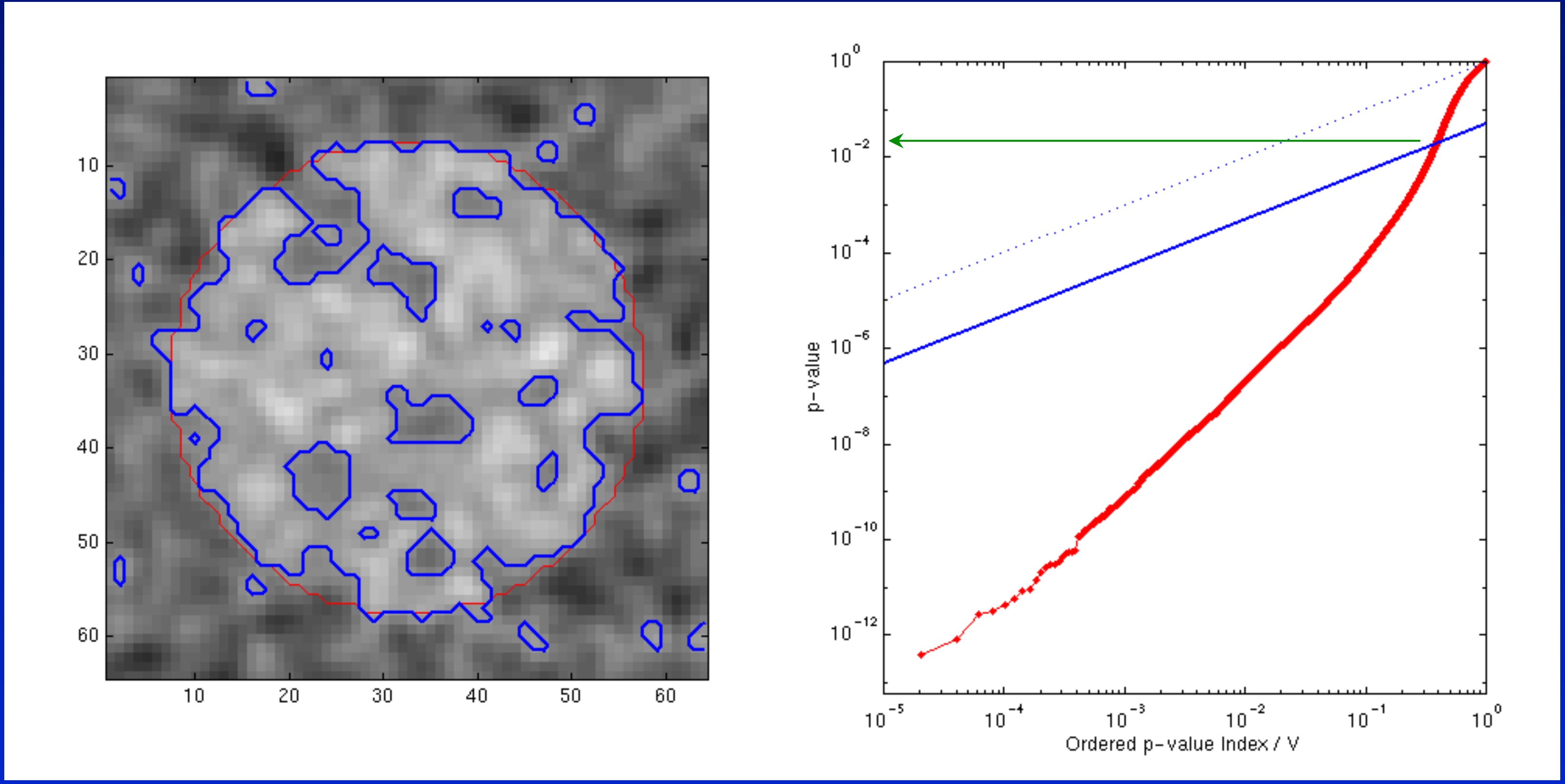
$p = 0.007157$ $z = 2.45$



Signal Intensity 3.0 Signal Extent 16.5 Noise Smoothness 3.0

Controlling FDR: Varying Signal Extent

$p = 0.019274 \qquad z = 2.07$

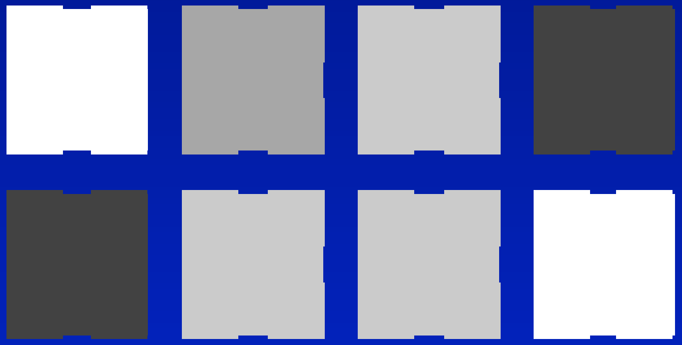


Signal Intensity 3.0 Signal Extent 25.0 Noise Smoothness 3.0

Controlling FDR: Benjamini & Hochberg

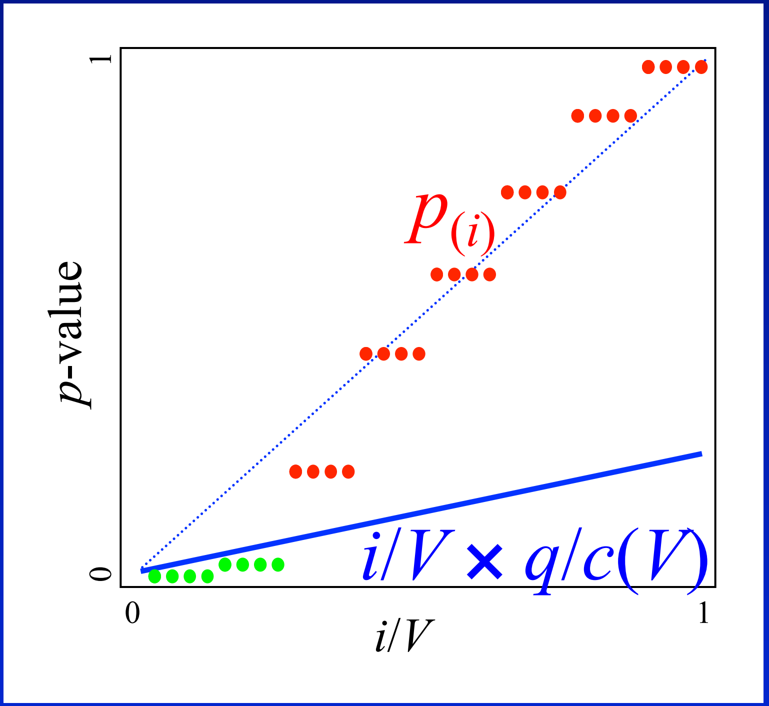
- Illustrating BH under dependence
  - Extreme example of positive dependence

8 voxel image



32 voxel image

(interpolated from 8 voxel image)



Conclusions

- Must account for multiplicity
  - Otherwise have a fishing expedition
- FWER
  - Very specific, not very sensitive
- FDR
  - Less specific, more sensitive
  - Sociological calibration still underway

References

- Most of this talk covered in these papers

TE Nichols & S Hayasaka, Controlling the Familywise Error Rate in Functional Neuroimaging: A Comparative Review. S*tatistical Methods in Medical Research*, 12(5): 419-446, 2003.

TE Nichols & AP Holmes, Nonparametric Permutation Tests for Functional Neuroimaging: A Primer with Examples. *Human Brain Mapping*, 15:1-25, 2001.

CR Genovese, N Lazar & TE Nichols, Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, 15:870-878, 2002.