Inference on SPMs: Random Field Theory & Alternatives

Thomas Nichols, Ph.D.
Oxford Big Data Institute
Li Ka Shing Centre for Health Information and Discovery
Nuffield Department of Population Health
University of Oxford

FIL SPM Course
realignment & motion correction

smoothing

normalisation

realignement & motion correction

smoothing

normalisation

statistical parameter estimates

Thresholding & Random Field Theory

Statistical Parametric Map

Corrected thresholds & p-values

image data

design matrix

kernel

General Linear Model

model fitting

statistic image

kernel

Statistical Parametric Map

Thresholding & Random Field Theory

Corrected thresholds & p-values

image data

design matrix

kernel

General Linear Model

model fitting

statistic image

kernel

Statistical Parametric Map

Thresholding & Random Field Theory

Corrected thresholds & p-values

image data

design matrix

kernel

General Linear Model

model fitting

statistic image

kernel

Statistical Parametric Map

Thresholding & Random Field Theory

Corrected thresholds & p-values

image data

design matrix

kernel

General Linear Model

model fitting

statistic image

kernel

Statistical Parametric Map

Thresholding & Random Field Theory

Corrected thresholds & p-values
Assessing Statistic Images...
Assessing Statistic Images

Where’s the signal?

High Threshold

$t > 5.5$

Good Specificity

Poor Power
(risk of false negatives)

Med. Threshold

$t > 3.5$

Low Threshold

$t > 0.5$

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
  - We only have a univariate linear model at each voxel!
Real-life inference: What we get

- **Signal location**
  - Local maximum – *no inference*

- **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 $\alpha$-level threshold $u_\alpha$
- Gives best spatial specificity
  - The null hyp. at a single voxel can be rejected

![Diagram](image-url)
Cluster-level Inference

• Two step-process
  – Define clusters by arbitrary threshold $u_{\text{clus}}$
  – Retain clusters larger than $\alpha$-level threshold $k_\alpha$
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

![Diagram of cluster-level inference](image)
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 $T$
• $\alpha$ level
  – Acceptable false positive rate
  – Level $\alpha = P( T > u_\alpha \mid H_0 )$
  – Threshold $u_\alpha$ controls false positive rate at level $\alpha$
• 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
  – $P(\text{Data}\mid\text{Null})$ not $P(\text{Null}\mid\text{Data})$
Multiple Comparisons Problem

- Which of 100,000 voxels are sig.?
  - $\alpha=0.05 \Rightarrow 5,000$ false positive voxels

- Which of (random number, say) 100 clusters significant?
  - $\alpha=0.05 \Rightarrow 5$ 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 MCP Solutions: Bonferroni

- For a statistic image $T$...
  - $T_i$ $i^{th}$ voxel of statistic image $T$
- ...use $\alpha = \alpha_0 / V$
  - $\alpha_0$ FWER level (e.g. 0.05)
  - $V$ number of voxels
  - $u_\alpha$ $\alpha$-level statistic threshold, $P(T_i \geq u_\alpha) = \alpha$
- By Bonferroni inequality...
  
  \[
  \text{FWER} = P(\text{FWE}) = P(\bigcup_i \{T_i \geq u_\alpha\} | H_0) \leq \sum_i P( T_i \geq u_\alpha | H_0 ) = \sum_i \alpha = \sum_i \frac{\alpha_0}{V} = \alpha_0
  \]

  - 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
FWER MCP Solutions: Random Field Theory

- Euler Characteristic $\chi_u$
  - Topological Measure
  - #blobs - #holes
  - At high thresholds, just counts blobs
  - FWER = $P(\text{Max voxel } \geq u \mid H_o)$
    = $P(\text{One or more blobs } \mid H_o)$
    $\approx P(\chi_u \geq 1 \mid H_o)$
    $\approx E(\chi_u \mid H_o)$

- No holes
- Never more than 1 blob
RFT Details:
Expected Euler Characteristic

\[ E(\chi_u) \approx \lambda(\Omega) \ |\Lambda|^{1/2} (u^2 - 1) \exp(-u^2/2) / (2\pi)^2 \]
- \( \Omega \rightarrow \) Search region \( \Omega \subset \mathbb{R}^3 \)
- \( \lambda(\Omega) \rightarrow \) volume
- \( |\Lambda|^{1/2} \rightarrow \) roughness

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

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

Only very upper tail approximates \( 1 - F_{\max}(u) \)
Random Field Theory
Smoothness Parameterization

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

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

$$\Lambda = \text{Var} \left( \frac{\partial G}{\partial (x, y, z)} \right)$$

$$= \begin{pmatrix}
\text{Var} \left( \frac{\partial G}{\partial x} \right) & \text{Cov} \left( \frac{\partial G}{\partial x}, \frac{\partial G}{\partial y} \right) & \text{Cov} \left( \frac{\partial G}{\partial x}, \frac{\partial G}{\partial z} \right) \\
\text{Cov} \left( \frac{\partial G}{\partial y}, \frac{\partial G}{\partial x} \right) & \text{Var} \left( \frac{\partial G}{\partial y} \right) & \text{Cov} \left( \frac{\partial G}{\partial y}, \frac{\partial G}{\partial z} \right) \\
\text{Cov} \left( \frac{\partial G}{\partial z}, \frac{\partial G}{\partial x} \right) & \text{Cov} \left( \frac{\partial G}{\partial z}, \frac{\partial G}{\partial y} \right) & \text{Var} \left( \frac{\partial G}{\partial z} \right)
\end{pmatrix}$$

$$= \begin{pmatrix}
\lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\
\lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\
\lambda_{zx} & \lambda_{zy} & \lambda_{zz}
\end{pmatrix}$$

$$|\Lambda|^{1/2} = \frac{(4 \log 2)^{3/2}}{\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z}.$$
Random Field Theory
Smoothness Parameterization

- **RESELS**
  - Resolution Elements
  - 1 RESEL = FWHM\textsubscript{x} × FWHM\textsubscript{y} × FWHM\textsubscript{z}
  - RESEL Count \( R \)
    - \( R = \lambda(\Omega) \sqrt{|\Lambda|} = (4\log2)^{3/2} \lambda(\Omega) / (\text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z) \)
    - 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”
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

```matlab
spm_imcalc_ui('RPV.img', ...
  'FWHM.img','i1.^(-1/3)')
```
Random Field Intuition

• Corrected P-value for voxel value $t$
  
  $P^c = P(\text{max } T > t) 
  \approx E(\chi_t) 
  \approx \lambda(\Omega) |\Lambda|^{1/2} t^2 \exp(-t^2/2)$

• Statistic value $t$ increases
  – $P^c$ decreases (but only for large $t$)

• Search volume increases
  – $P^c$ increases (more severe MCP)

• Smoothness increases (roughness $|\Lambda|^{1/2}$ decreases)
  – $P^c$ decreases (less severe MCP)
RFT Details: Unified Formula

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

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

$R_d(\Omega)$: $d$-dimensional Minkowski functional of $\Omega$
- *function of dimension, space $\Omega$ and smoothness:

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

$\rho_d(\Omega)$: $d$-dimensional EC density of $Z(x)$
- *function of dimension and threshold, specific for RF type:

E.g. Gaussian RF:

$$\rho_0(u) = 1 - \Phi(u)$$
$$\rho_1(u) = (4 \ln 2)^{1/2} \exp(-u^2/2) / (2\pi)$$
$$\rho_2(u) = (4 \ln 2) \exp(-u^2/2) / (2\pi)^{3/2}$$
$$\rho_3(u) = (4 \ln 2)^{3/2} (u^2 -1) \exp(-u^2/2) / (2\pi)^2$$
$$\rho_4(u) = (4 \ln 2)^2 (u^3 -3u) \exp(-u^2/2) / (2\pi)^{5/2}$$
Random Field Theory
Cluster Size Tests

• Expected Cluster Size
  – $E(S) = E(N)/E(L)$
  – $S$ cluster size
  – $N$ suprathreshold volume
    $\lambda(\{T > u_{clus}\})$
  – $L$ number of clusters

• $E(N) = \lambda(\Omega) \cdot P( T > u_{clus} )$
• $E(L) \approx E(\chi_u)$
  – Assuming no holes
Random Field Theory
Limitations

• Sufficient smoothness
  – FWHM smoothness $3-4 \times \text{voxel size (Z)}$
  – More like $\sim 10 \times$ 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
  – $u = 9.870$

• Result
  – 5 voxels above the threshold
  – 0.0063 minimum FWE-corrected $p$-value
Massive Null (resting-state) fMRI Evaluation

**Goal:** Evaluate AFNI, FSL & SPM *task* fMRI with *resting-state* fMRI data, using 4 designs, 3 million randomised analyses

**Outcome:**
- Voxel FWE OK (Conservative)
- Cluster FWE 0.001 OK
- Cluster FWE 0.01 Very Bad (Liberal)

**Why?** Spatial ACF not Gaussian, Nonstationarity smoothness

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates (2016). Eklund, TE Nichols, H Knutsson PNAS, 113(28), 7900-5
Real Data: SnPM Promotional

- Nonparametric method more powerful than RFT for low DF
- “Variance Smoothing” even more sensitive
- FWE controlled all the while!
- http://nisox.org/Software/SnPM

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

$u^{RF} = 9.87$
$u^{Bonf} = 9.80$
5 sig. vox.

$t_{11}$ Statistic, Nonparametric Threshold

$u^{Perm} = 7.67$
58 sig. vox.

Smoothed Variance $t$ Statistic, Nonparametric Threshold

378 sig. vox.
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:

<table>
<thead>
<tr>
<th></th>
<th>Accept Null</th>
<th>Reject Null</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null True</td>
<td>$V_{0A}$</td>
<td>$V_{0R}$</td>
</tr>
<tr>
<td>Null False</td>
<td>$V_{1A}$</td>
<td>$V_{1R}$</td>
</tr>
<tr>
<td></td>
<td>$N_A$</td>
<td>$N_R$</td>
</tr>
</tbody>
</table>

- Realized FDR

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

- 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%

Percentage of Null Pixels that are False Positives

Control of Familywise Error Rate at 10%

Occurrence of Familywise Error

Control of False Discovery Rate at 10%

Percentage of Activated Pixels that are False Positives
Benjamini & Hochberg Procedure

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

\[ p_{(i)} \leq \frac{i}{V} \times q \]

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

*JRSS-B (1995)*

57:289-300
Adaptiveness of Benjamini & Hochberg FDR

Ordered p-values $p^{(i)}$

P-value threshold when no signal: $\alpha/V$

P-value threshold when all signal: $\alpha$
Real Data: FDR Example

- Threshold
  - $u = 3.83$

- Result
  - 3,073 voxels above $u$
  - $<0.0001$ 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
  – Voxel-wise available: edit spm_defaults.m to read
defaults.stats.topoFDR = 0;
  – Note!
    • Both cluster- and peak-wise FDR depends on cluster-forming threshold!

Item Recognition data

Cluster-forming threshold P=0.001
  Peak-wise FDR: t=4.84, P_{FDR} 0.836
Cluster-forming threshold P=0.01
  Peak-wise FDR: t=4.84, P_{FDR} 0.027
Cluster FDR: Example Data

Level 5% **Voxel-FWE**

- Level 5% **Cluster-FWE**
  - \( P = 0.001 \) cluster-forming thresh
  - \( k_{\text{FWE}} = 241, 5 \text{ clusters} \)

- Level 5% **Cluster-FDR**
  - \( P = 0.001 \) cluster-forming thresh
  - \( k_{\text{FDR}} = 138, 6 \text{ clusters} \)

Level 5% **Voxel-FDR**

- Level 5% **Cluster-FWE**
  - \( P = 0.01 \) cluster-forming thresh
  - \( k_{\text{FWE}} = 1132, 4 \text{ clusters} \)

- Level 5% **Cluster-FDR**
  - \( P = 0.01 \) cluster-forming thresh
  - \( k_{\text{FDR}} = 1132, 4 \text{ clusters} \)
Conclusions

• Must account for multiplicity
  – Otherwise have a fishing expedition

• FWER
  – Very specific, not very sensitive

• FDR
  – Voxel-wise: Less specific, more sensitive
  – Cluster-, Peak-wise: Similar to FWER
References


