Multiple comparison correction

Methods & models for fMRI data analysis
23 October 2018

Klaas Enno Stephan

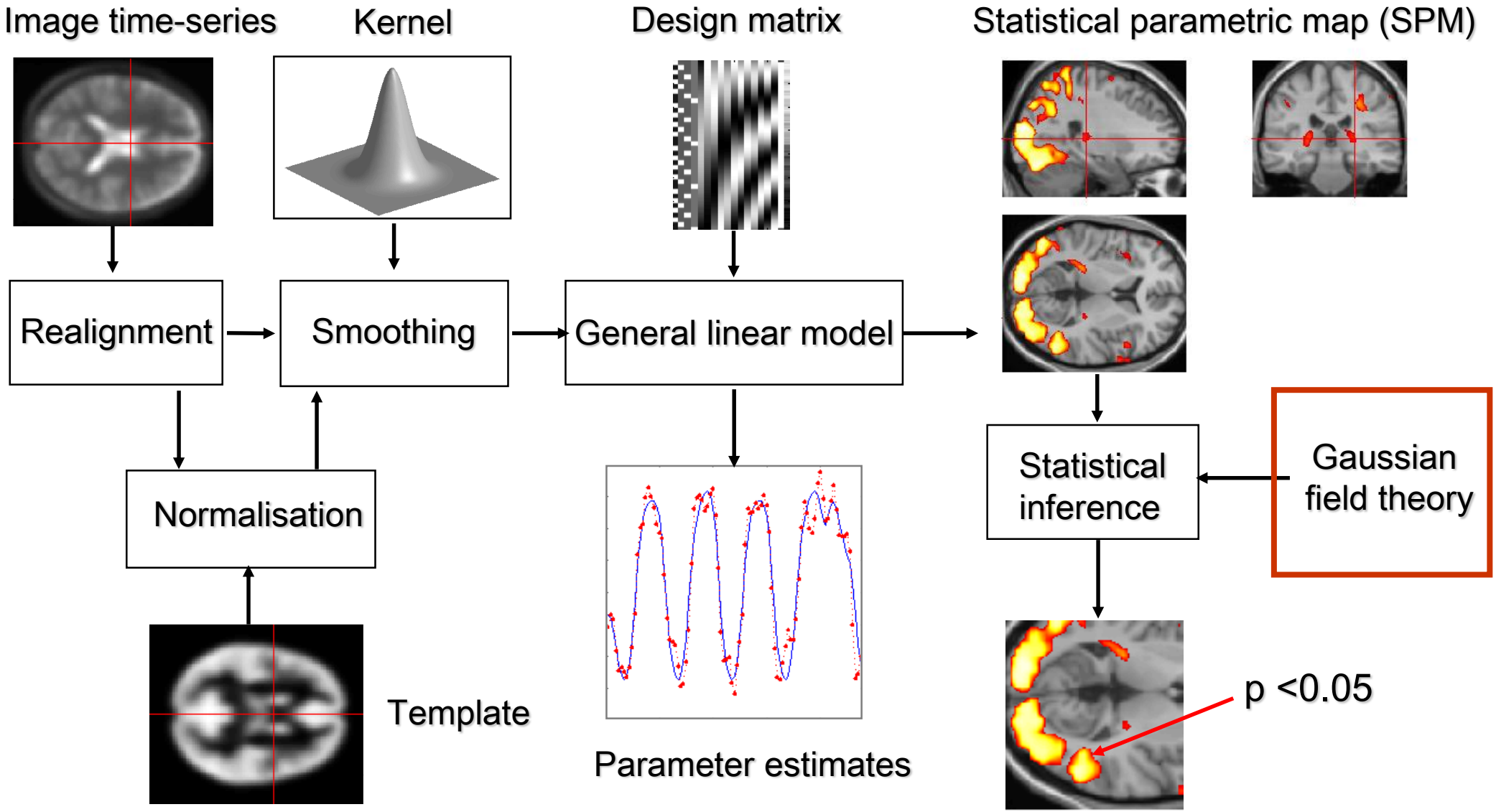
With many thanks for slides & images to:
FIL Methods group
Tom Nichols

Translational Neuromodeling Unit

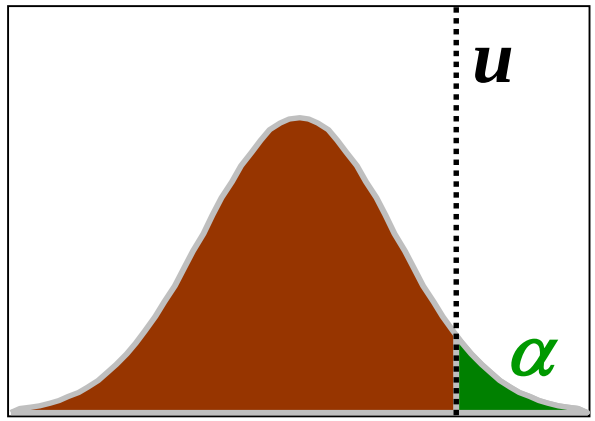
Universität Zürich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Overview of SPM



Inference at a single voxel



t distribution

NULL hypothesis
H_0: activation is zero

$\alpha = p(T > u \mid H_0)$

We can choose u to set a voxel-wise significance level of α.

p-value: probability of getting a value of the test statistic t , or a more extreme value, under the null hypothesis.

If the p-value is smaller than u, we reject the null hypothesis.

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

$$t = \frac{c^T \hat{\beta}}{\hat{std}(c^T \hat{\beta})} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T \left(X^T X\right)^{-1} c}} \sim t_{N-p}$$

Types of error

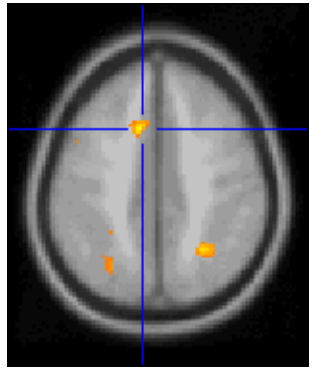
Test result	H_0 true	H_0 false
Reject H_0	False positive (FP) Type I error α	True positive (TP)
Failure to reject H_0	True negative (TN)	False negative (FN) Type II error β

specificity: 1-α
= TN / (TN + FP)
= proportion of actual negatives which are correctly identified

sensitivity (power): 1-β
= TP / (TP + FN)
= proportion of actual positives which are correctly identified

Assessing SPMs

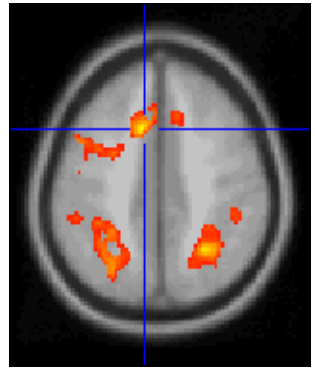
High Threshold



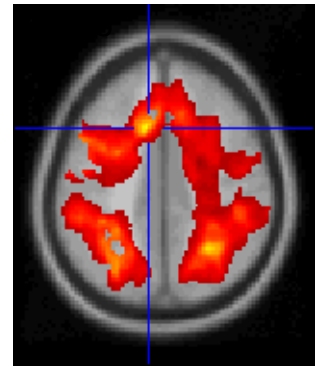
Good Specificity

Poor Power
(risk of false negatives)

Med. Threshold



Low Threshold

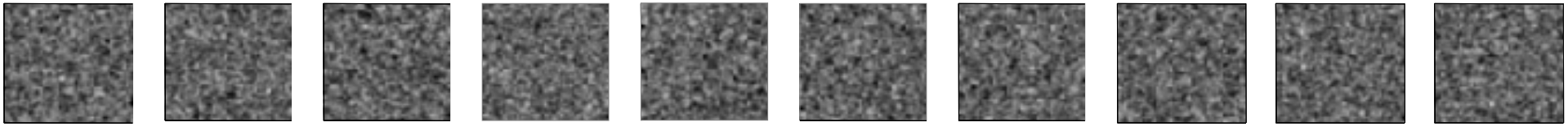


Poor Specificity
(risk of false positives)

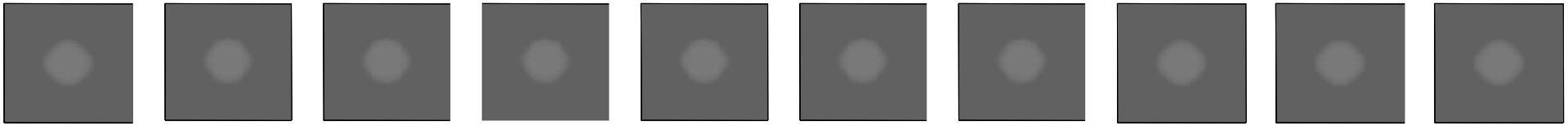
Good Power

Inference on images

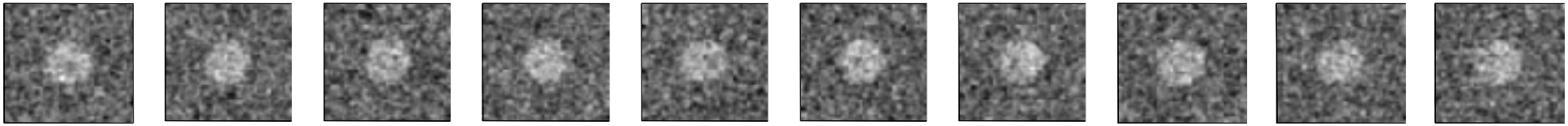
Noise



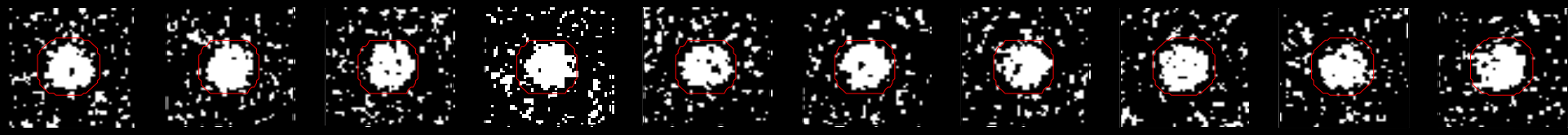
Signal



Signal+Noise



Use of ‘uncorrected’ p-value, α=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

Using an ‘uncorrected’ p-value of 0.1 will lead us to conclude on average that 10% of voxels are active when they are not.

This is clearly undesirable. To correct for this we can define a null hypothesis for images of statistics.

Family-wise null hypothesis

FAMILY-WISE NULL HYPOTHESIS:

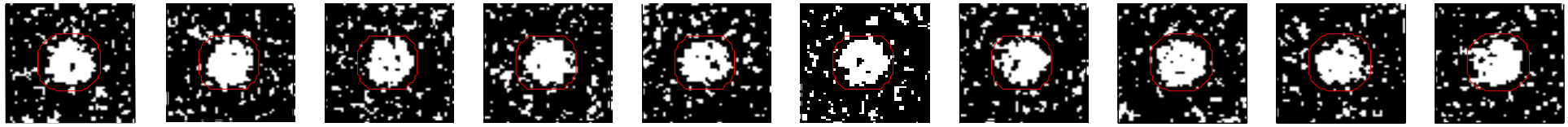
Activation is zero **everywhere**.

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

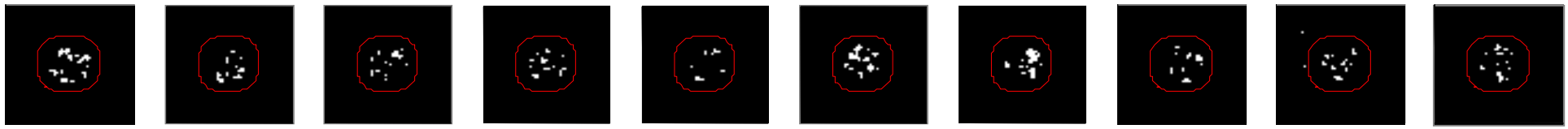
A false-positive **anywhere** in the image gives a Family Wise Error (FWE).

Family-Wise Error (FWE) rate = ‘corrected’ p-value

Use of ‘uncorrected’ p-value, α=0.1



Use of ‘corrected’ p-value, α=0.1



FWE

The Bonferroni correction

The family-wise error rate (FWE), α, for a family of N **independent** voxels is

$$\alpha = Nv$$

where v is the voxel-wise error rate.

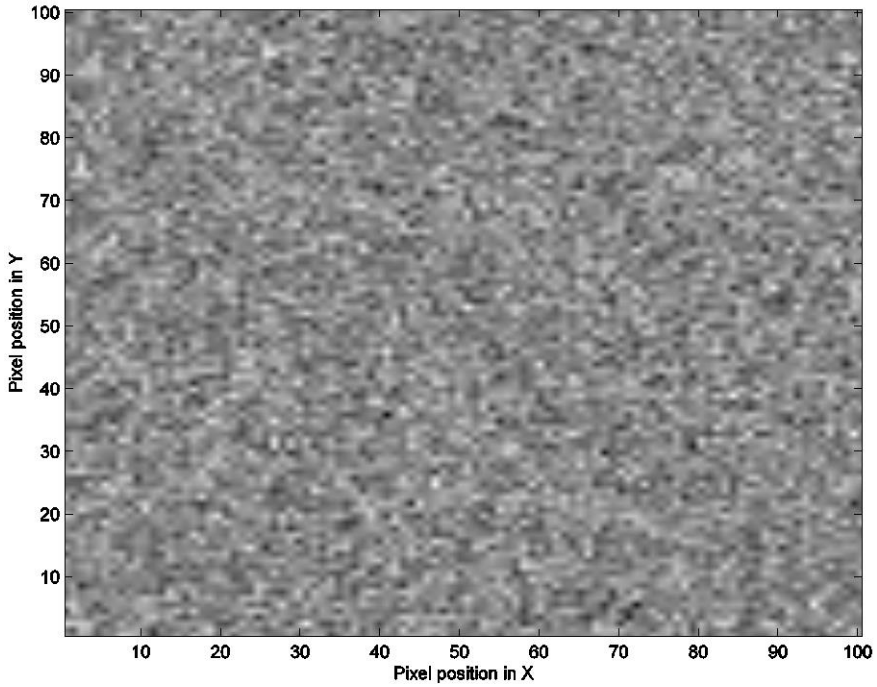
Therefore, to ensure a particular FWE, we can use

$$v = \alpha / N$$

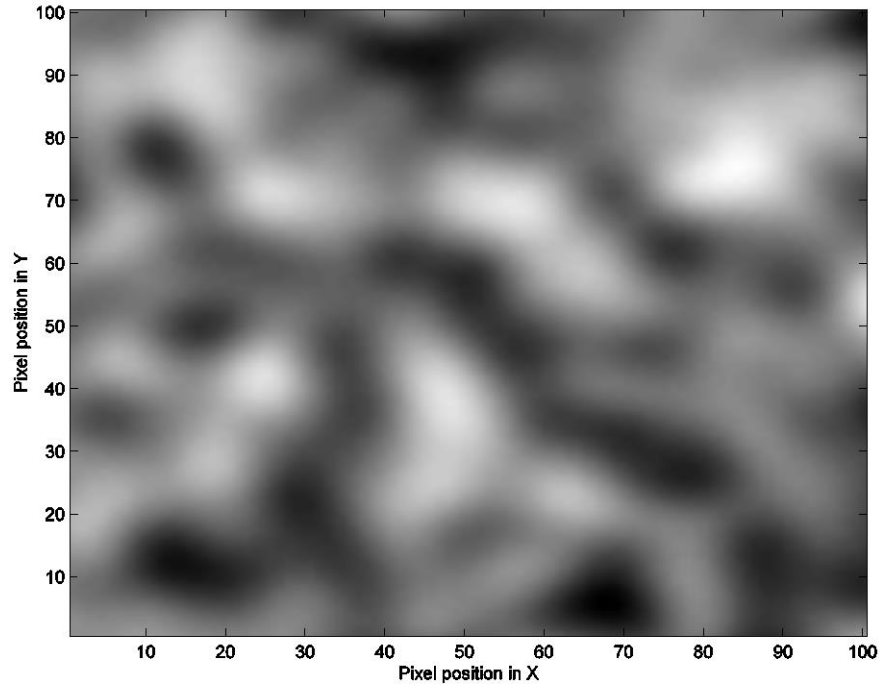
BUT ...

The Bonferroni correction

Independent voxels



Spatially correlated voxels



Bonferroni correction assumes independence of voxels

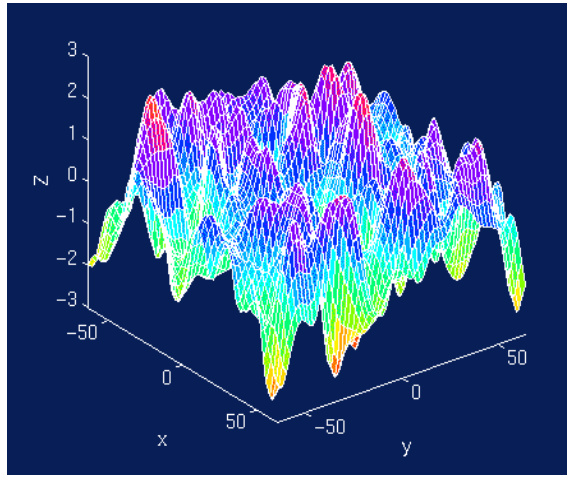
→ this is too conservative for brain images,
which always have a degree of smoothness

Smoothness (inverse roughness)

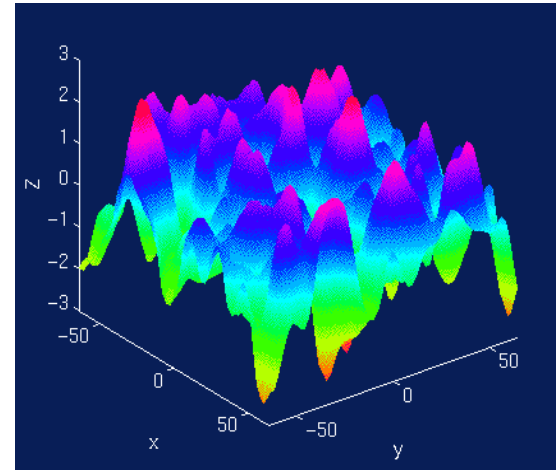
- roughness = 1/smoothness
- intrinsic smoothness
  - MRI signals are aquired in k-space (Fourier space); after projection on anatomical space, signals have continuous support
  - diffusion of vasodilatory molecules has extended spatial support
- extrinsic smoothness
  - resampling during preprocessing
  - matched filter theorem
    $\rightarrow$ deliberate additional smoothing to increase SNR
- described in resolution elements: "resels"
- resel = size of image part that corresponds to the FWHM (full width half maximum) of the Gaussian convolution kernel that would have produced the observed image if it had been applied to independent voxel values
- # resels is similar, but not identical to # independent observations
- can be computed from spatial derivatives of the residuals

Random Field Theory

- Consider a statistic image as a discretisation of a continuous underlying random field with a certain smoothness
- Use results from continuous random field theory



Discretisation
(“lattice
approximation”)



Euler characteristic (EC)

Topological measure

threshold an image at u

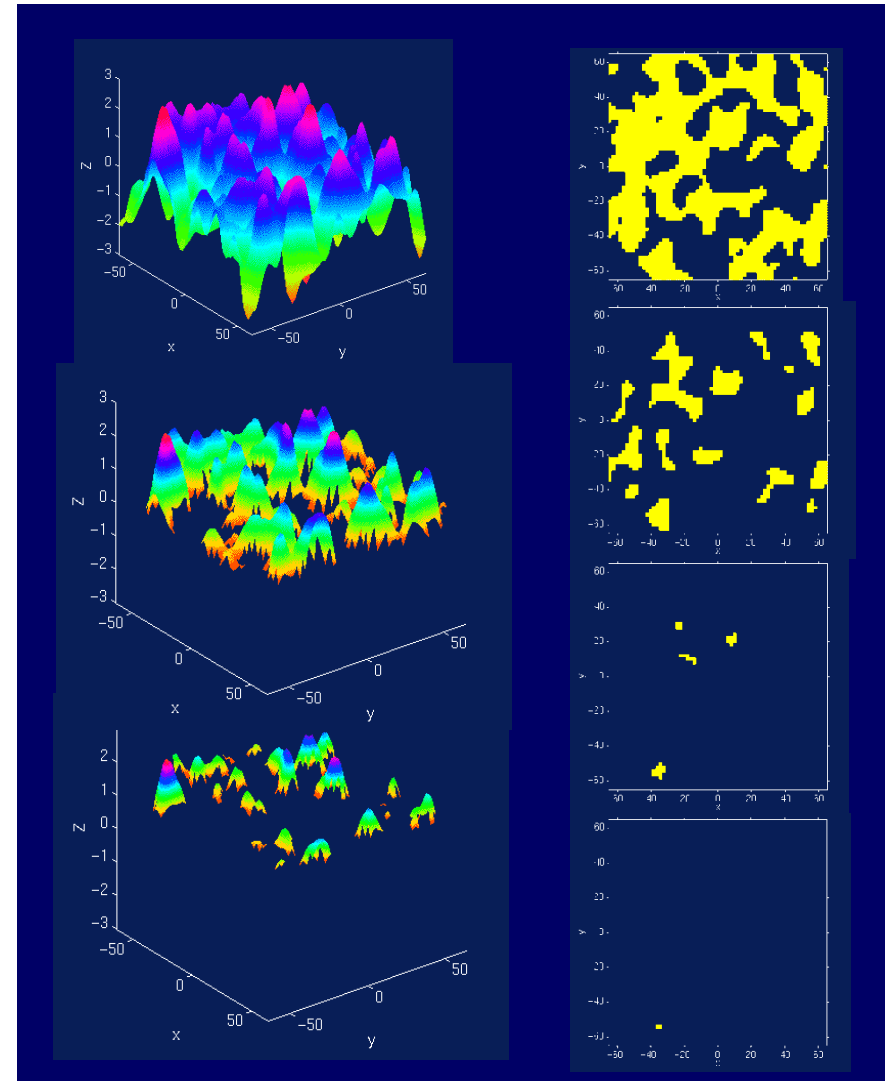
$\rightarrow$ EC $\propto$ # blobs

At high u:

p (blob) = E [EC],

therefore (under H_0):

FWE rate: α = E [EC]



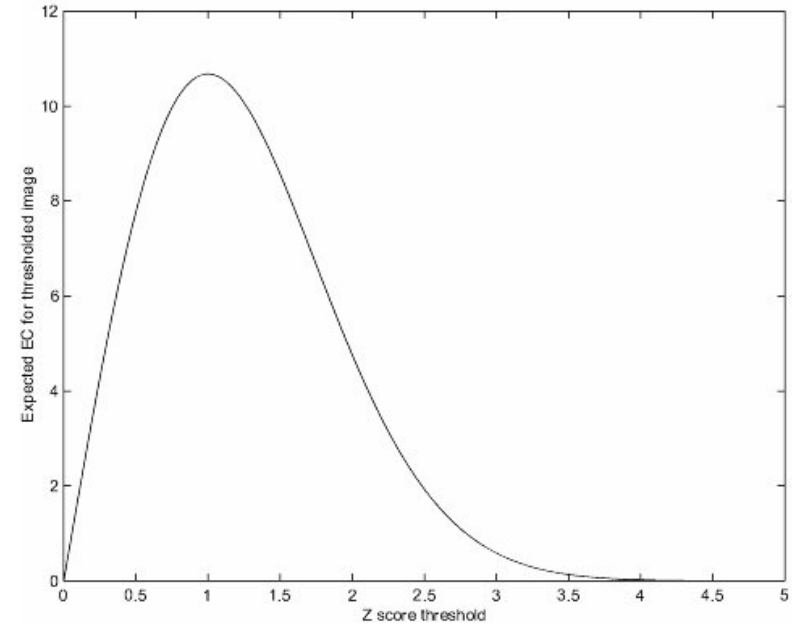
Euler characteristic (EC) for 2D images

$$\mathrm{E}[\mathrm{EC}] = R(4\log 2)(2\pi)^{-3/2} Z_T \exp(-0.5 Z_T^2)$$

R = number of resels
Z_T = Z value threshold

We can determine that Z threshold for which E[EC] = 0.05. At this threshold, every remaining peak represents a significant activation, corrected for multiple comparisons across the search volume.

Example: For 100 resels, E [EC] = 0.049 for a Z threshold of 3.8. That is, the probability of getting one or more blobs where Z is greater than 3.8, is 0.049.



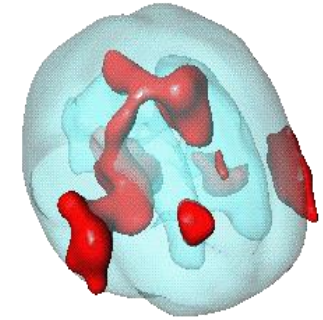
Expected EC values for an image of 100 resels

Euler characteristic (EC) for any image

- Computation of E[EC] can be generalized to volumes of any dimension, shape and size (Worsley et al. 1996).

- When we have an *a priori* hypothesis about where an activation should be, we can (and should) reduce the search volume:
  - mask defined by (probabilistic) anatomical atlases
  - mask defined by separate "functional localisers"
  - mask defined by orthogonal contrasts
  - (spherical) search volume around previously reported coordinates

➡ small volume correction (SVC)



Worsley et al. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping, 4, 58–83.

Computing EC wrt. search volume and threshold

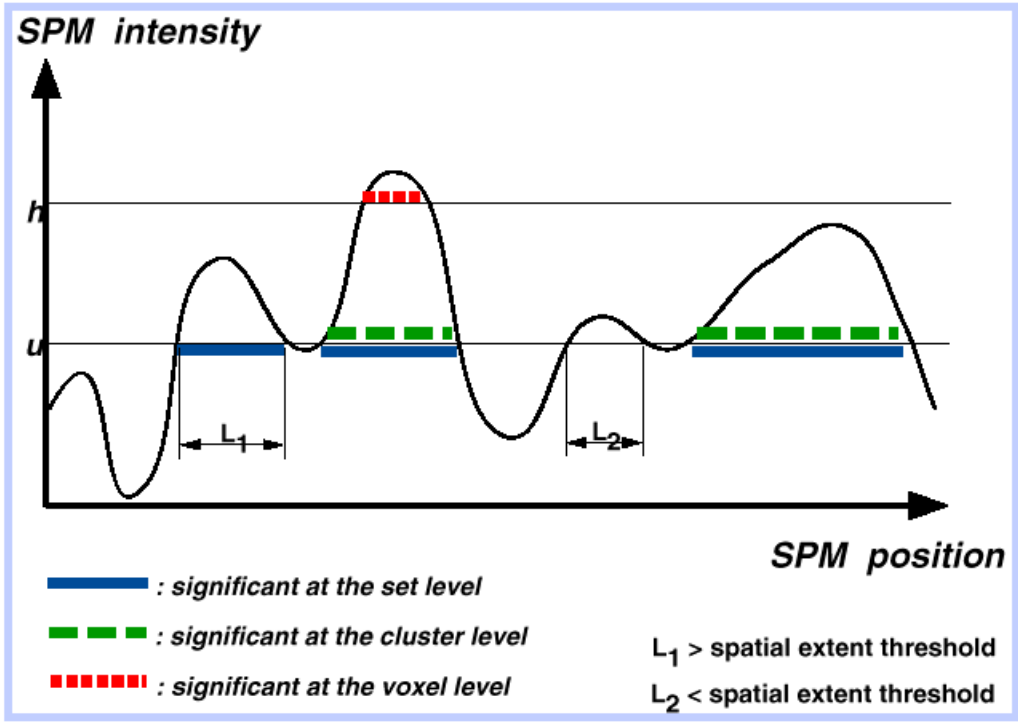
$$E(\chi_u) \approx \lambda(\Omega)\ |\Lambda|^{1/2}\ (u^2 - 1)\ \exp(-u^2/2)\ /\ (2\pi)^2$$

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

- Assumptions:
  - Multivariate normal
  - Stationary*
  - ACF twice differentiable at 0

* Stationarity
  - Results valid w/out stationarity
  - More accurate when stationarity holds

Height, cluster and set level tests



Sensitivity

Regional specificity

Height level test:
intensity of a voxel

Cluster level test:
spatial extent above u

Set level test:
number of clusters above u

Further reading

- Friston KJ, Frith CD, Liddle PF, Frackowiak RS. Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab. 1991 Jul;11(4):690-9.
- 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;4:58-73.

Thank you