

Statistics and Brain Mapping

K.J. Worsley

*Department of Mathematics and Statistics, McGill University,
Montreal, Québec, Canada H3A 2K6*

Web: <http://www.math.mcgill.ca/~keith>

The scientific aim of brain mapping is to investigate which regions of the brain are used when a subject is involved in a particular task, such as cognitive tasks, sensory stimuli, or memory challenges, or how brain function differs between different groups of subjects, such as normals and schizophrenics. A second aim is the study of how brain anatomy differs between different groups of subjects. There are three different types of images for measuring brain *function*: cerebral blood flow (CBF) measured by positron emission tomography (PET) (Raichle, 1994), blood oxygenation level dependent (BOLD) response, measured by functional magnetic resonance imaging (fMRI) (Lange & Zeger, 1997), and neurotransmitter kinetics measured by dynamic PET (Dagher *et al.*, 1998). There are three different types of images for measuring changes in shape of brain structures, such as the cortex: binary structure masks (1=inside, 0=outside) from segmented MRI scans; normal displacements of the structure surface; and vector deformations required to warp the structure to an atlas (standard structure).

On the surface, this looks like a standard statistical problem of relating dependent variables $\mathbf{Y}(\mathbf{t})$ (function, anatomy) to explanatory variables $\mathbf{X}$ (tasks, groups). The novelty is that the dependent variables are 3D *images* with a value at each voxel $\mathbf{t} \in S \subset \mathbb{R}^3$. Usually S is the entire brain or part of it, or, in the case of surface normal displacements, the 2D manifold of the surface of the structure. The most widely used tool is the linear model which is applied separately to each voxel $\mathbf{t}$: $\mathbf{Y}(\mathbf{t}) = \mathbf{X}\beta(\mathbf{t}) + \epsilon(\mathbf{t})$ (Friston *et al.*, 1995). The errors $\epsilon(\mathbf{t})$ may be correlated for images taken on the same subject, either because of random effects, or because of temporal correlation, particularly for PET kinetics and fMRI data, where images are acquired every few seconds. This methodology has been found to be sufficiently flexible to treat image data from all the above six sources; the only slight generalisation required is a multivariate linear model for the three components of the vector deformations data (Cao & Worsley, 1999), logistic regression for binary structure masks (Taylor *et al.*, 1998), and nonlinear models for PET kinetics (Dagher *et al.*, 1998). Lange & Zeger (1997) have developed a more sophisticated model for fMRI data that incorporates spatial as well as temporal information. In most laboratories worldwide, this is done using the SPM (Statistical Parametric Mapping) software package (see Friston *et al.*, 1995). Eddy *et al.* (1996) have implemented a very careful analysis of fMRI data in the FIASCO software package.

The challenging part is the analysis of the resulting *images* of test statistics $Z(\mathbf{t})$ such as t -statistics for testing components of $\beta(\mathbf{t})$. Most experiments are designed to activate only a small region of the brain, so the main statistical problem is to detect those regions of the brain that show a local non-zero value of the selected components of $\beta(\mathbf{t})$. This can be done using the peak height of local maxima of $Z(\mathbf{t})$, or the maximum cluster size of all neighbouring voxels where $Z(\mathbf{t})$ exceeds a fixed high threshold. Methods have been developed recently for finding the P -value for such test statistics under the null hypothesis of no activations anywhere (Worsley, 1995; Cao, 1999); this is conservative for the more realistic case of a small number of isolated local activations. The next step is inference about the activation once it is detected.

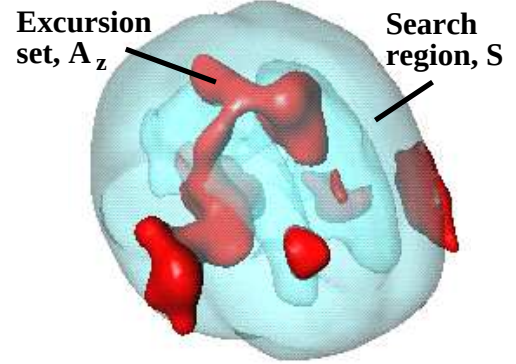
The main tool that has been used is a concept borrowed from topology, the Euler characteristic (EC) of the excursion set (see Worsley, 1996, for a non-technical overview). The

excursion set A_z is the set of all points $\mathbf{t} \in \mathbb{R}^D$ where $Z(\mathbf{t})$ exceeds a fixed threshold value z (see Figure), and the EC $\chi(S \cap A_z)$ counts the number of connected components of the excursion set, minus the number of ‘holes’. For high thresholds the holes disappear and the EC counts the number of peaks in the image above the threshold. For even higher threshold values near $Z_{\max} = \max_{\mathbf{t} \in S} Z(\mathbf{t})$, the EC takes the value 1 if the maximum is above the threshold, and 0 otherwise, so that the EC approximates the indicator function for the event $Z_{\max} \geq z$. Thus for high thresholds $E\{\chi(S \cap A_z)\}$ approximates $P\{Z_{\max} \geq z\}$ (Adler, 1981). The advantage of the EC is that a simple exact expression has been found for its expectation when no signal is present and $Z(\mathbf{t})$ is a sufficiently smooth isotropic random field:

$$P\{Z_{\max} \geq z\} \approx E\{\chi(S \cap A_z)\} = \sum_{d=0}^D \mu_d(S) \rho_d(z) \quad (1)$$

where $\mu_d(S)$ is the d -dimensional Minkowski functional of S and $\rho_d(z)$ is the d -dimensional EC density of $Z(\mathbf{t})$:

$$\rho_d(z) = E\{(Z \geq z) \det(-\ddot{Z}_d) \mid \dot{Z}_d = 0\} P\{\dot{Z}_d = 0\}$$



(dot notation with subscript d means differentiation with respect to the first d components of $\mathbf{t}$). The first and most important term ($d = D$) was found by Adler (1981); the remaining terms are a boundary correction for when A_z touches the boundary of S (Worsley, 1995). Results are available for the EC density of a variety of random fields: the Gaussian random field (Adler, 1981) χ^2 , t and F random fields (Worsley, 1994). For multivariate linear models, such as the deformations data, the appropriate test statistic is Hotelling’s T^2 . Even though this can be rescaled point-wise to an F -statistic, it is *not* an F -field; the EC density for Hotelling’s T^2 fields is available in Cao & Worsley (1999).

The approximation (1) appears to be very accurate for low P -values (those usually encountered in practice) and search regions S of almost any shape or size, even 2D manifolds such as the cortical surface which are embedded in 3D, so that it can be applied to the surface normal displacements data. In the case of $Z(\mathbf{t})$ a Gaussian random field and S convex, the approximation (1) is a sum of D terms in decreasing powers of z , plus $P\{Z \geq z\}$. It had been conjectured by Siegmund & Worsley (1995) that these were the first D terms in a power series expansion for $P\{Z_{\max} \geq z\}$. This was based on a completely different approach developed by David Siegmund and his co-workers that used volumes of tubes (Knowles & Siegmund, 1989; Sun, 1993). Here the random field is approximated by a finite Karhunen-Loève expansion, and the P -value for its maximum is then the volume of a particular tube about the search region, which can be evaluated using Weyl’s (1939) tube formula. The calculations are very laborious, but in every case looked at so far, the results give the same first D terms as the expected EC. In a remarkable result announced last spring, Robert Adler finally proved this conjecture (Adler, 1999). Using results of Piterbarg, he showed that the expected EC was even more accurate than previously thought: the error is exponentially (not just polynomially) smaller than the smallest term in the expected EC expansion.

What is intriguing about the proof of Adler’s result is that the EC itself is not directly involved, which suggests that perhaps there is a deeper reason for why it should be so good as an approximation. One direct connection is the work of Dan Naiman and Henry Wynn on truncated inclusion-exclusion formulas for the probability of a union of events (Naiman &

Wynn, 1992), which can be applied to maxima of discretely sampled random fields. Here the truncated inclusion-exclusion formula out to a depth D is exact provided the simplicial complex formed by the terms in the formula is an abstract tube. An abstract tube is a simplicial complex such that every sub-simplicial complex (the ‘excursion set’) has an EC of 1 (here the EC is defined by $\# \text{vertices} - \# \text{edges} + \# \text{faces} - \dots$). Thus the probability of the union is the expected EC of the whole abstract tube, which is of course the truncated inclusion-exclusion formula.

There are situations where the expected EC gives the exact result, that is where (1) is exact. This occurs for the problem of testing a multivariate normal mean of $\mathbf{Z} \in \mathbb{R}^n$ against an ordered alternative, that is $H_0 : \mu_1 = 0, \dots, \mu_n = 0$ against $H_1 : \mu_1 \leq \dots \leq \mu_n$ (Lin & Lindsay, 1997). Here the alternative is an n -dimensional convex cone; the P -value of the likelihood ratio test is proportional to the volume of a tube about the cone (Takemura & Kuriki, 1997). An alternative approach is to write the test statistic as the maximum of $Z(\mathbf{t}) = \mathbf{Z}'\mathbf{t}$ over all $\mathbf{t}$ in the cone ($\|\mathbf{t}\| = 1$); now it turns out that the EC is either 0 or 1 for *all* thresholds, so as in the case above, the expected EC *is* the P -value.

At present most PET, fMRI and binary structure mask images are smoothed to improve the local signal-to-noise ratio and to smooth out registration errors and anatomical variability. It is not hard to show that the optimal kernel should match the shape of the signal to be detected. Since we don’t know the width of the signal, Poline & Mazoyer (1994) proposed searching over kernel filter width as well as location in S , but only empirical results are given. Siegmund & Worsley (1995) modeled the process as a 4D non-stationary Gaussian random field and approached $P\{Z_{\max} \geq z\}$ from two different directions. The first found the expected EC of excursion sets of the 4D field, and the second took the volumes of tubes approach. The two methods again give the same terms in the asymptotic expansion of $P\{Z_{\max} \geq z\}$.

Khalil Shafie has taken this idea one step further (Shafie *et al.*, 1998). Instead of scaling the kernel by a scalar, he considered scaling it by a matrix, which allows for differential scaling along the axes and rotation of the kernel. This should be optimal at detecting ellipsoidal-countoured signals, at the cost of adding an extra $D(D+1)/2$ dimensions to the random field. He was successful in finding the expected EC in $D = 2$ dimensions; the calculations were just too heavy to attempt the 3D case. An interesting aspect of this work was that nearly all the algebra was done by computer.

One possible application of scale space is to thresholding the continuous wavelet transform. The scale space random field is just a continuous wavelet transform, and the above results give the null distribution of the maximum wavelet coefficient, which might be useful for signal detection in general.

REFERENCES

- Adler, R.J. (1981). *The Geometry of Random Fields*. Wiley, New York.
- Adler, R.J. (1999). On excursion sets, tube formulae, and maxima of random fields. *Annals of Applied Probability*, in press.
- Cao, J (1999). The size of the connected components of excursion sets of χ^2 , t and F fields. *Advances in Applied Probability*, in press.
- Cao, J. & Worsley, K.J. (1999). The geometry of the Hotelling’s T^2 random field with applications to the detection of shape changes. *Annals of Statistics*, in press.
- Dagher, A., Bleicher, C., Aston, J.A.D., Worsley, K.J. & Evans, A.C. (1998). Measuring neurotransmitter release with Positron Emission Tomography. *NeuroImage*, **7**:S27.

- Eddy, W.F., Fitzgerald, M., Genovese, C.R., Mockus, A. & Noll, D.C. (1996). Functional imaging analysis software - computational Olio. *Proceedings in Computational Statistics*, (A. Prat, Ed.), Physica-Verlag, Heidelberg, 39-49.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J-B., Frith, C.D. & Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**:189-210.
- Knowles, M. & Siegmund, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *International Statistical Review*, **57**:205-220.
- Lange, N. & Zeger, S.L. (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with Discussion). *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **14**:1-29.
- Lin, Y. & Lindsay, B.G. (1997). Projections on cones, chi-bar squared distributions and Weyl's formula. *Statistics & Probability Letters*, **32**:367-376.
- Naiman, D.Q. & Wynn, H.P. (1992). Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. *Annals of Statistics*, **20**:43-76.
- Poline, J-B. & Mazoyer, B.M. (1994). Enhanced detection in brain activation maps using a multifiltering approach. *Journal of Cerebral Blood Flow and Metabolism*, **14**:639-642.
- Raichle, M. E. (1994). Visualizing the mind. *Scientific American*, **July 1994**:58-64.
- Shafie, Kh., Worsley, K.J., Wolforth, M. & Evans, A.C. (1998). Rotation space: Detecting functional activation by searching over rotated and scaled filters. *NeuroImage*, **7**:S755.
- Siegmund, D.O. & Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**:608-639.
- Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Annals of Probability*, **21**:34-71.
- Takemura, A. & Kuriki, S. (1997). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise cone alternatives. *Annals of Statistics*, **25**:2368-2387.
- Taylor, J., Worsley, K.J., Zijdenbos, A.P., Paus, T. & Evans, A.C. (1998). Detecting anatomical changes using logistic regression of structure masks. *NeuroImage*, **7**:S753.
- Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2 , F and t fields. *Advances in Applied Probability*, **26**:13-42.
- Worsley, K.J. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability*, **27**:943-959.
- Worsley, K.J. (1996). The geometry of random images. *CHANCE*, **9(1)**:27-40.

RÉSUMÉ: Champs Aléatoires et la Cartographie Cérébrale

Les dernières années ont vu une forte croissance dans la recherche sur la cartographie cérébrale, commençant par la tomographie par émission des positones (TEP). Maintenant l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) est plus utilisée. À l'heure actuelle, le TEP sert à la cartographie des neurorecepteurs et ligands, plutôt qu'aux changements dans le flux sanguin en réponse à un stimuli, maintenant dominé par IRMf. Plusieurs nouvelles méthodes sont capables de reconstruire les images 3D de l'activité électrique provenant des mesures EEG du scalpe. Toutes les données prises à chaque instant ont la forme des images 3D du cerveau humain. Elles sont indépendantes dans le cas de TEP, corrélées dans le cas de IRMf, et continues pour EEG. Pendant ce temps, le sujet reçoit plusieurs stimuli, et le but statistique est de détecter les régions du cerveau qui y ont répondu, de faire l'inférence sur l'endroit, la réponse et son étendue. Ceci représente des défis statistiques qui ont été étudiés par les champs aléatoires.