

Comparing Functional (PET) Images: The Assessment of Significant Change

*‡K. J. Friston, *†C. D. Frith, *‡P. F. Liddle, and *R. S. J. Frackowiak

*MRC Cyclotron Unit and Department of Medicine (Neurology), Hammersmith Hospital, †Division of Psychiatry, Clinical Research Centre, Northwick Park Hospital, and ‡Academic Unit, Department for Rehabilitation Psychiatry, Westminster and Charing Cross Medical Schools, London, U.K.

Summary: Statistical parametric maps (SPMs) are potentially powerful ways of localizing differences in regional cerebral activity. This potential is limited by uncertainties in assessing the significance of these maps. In this report, we describe an approach that may partially resolve this issue. A distinction is made between using SPMs as images of change significance and using them to identify foci of significant change. In the first case, the SPM can be reported nonselectively as a single mathematical object with its omnibus significance. Alternatively, the SPM constitutes a large number of repeated measures over the brain. To reject the null hypothesis, that no change has

occurred at a specific location, a threshold adjustment must be made that accounts for the large number of comparisons made. This adjustment is shown to depend on the SPM's smoothness. Smoothness can be determined empirically and be used to calculate a threshold required to identify significant foci. The approach models the SPM as a stationary stochastic process. The theory and applications are illustrated using uniform phantom images and data from a verbal fluency activation study of four normal subjects. **Key Words:** Positron emission tomography—Statistics—Significance—Activation—Cerebral blood flow—Stochastic processes.

Statistical maps are an important alternative to region of interest (ROI) analysis in the localization of differences in positron emission tomography (PET) images. Statistical maps can be looked at in two ways. They represent images of change significance, as opposed to change magnitude, allowing brain regions to be compared qualitatively in terms of relative significance. Alternatively, they have the potential to localize change at a given level of significance.

The distinction between images of change size and change significance relates to regional differences in the variability of cerebral activity. Significance has two components: the size of the difference and the error variance associated with its measurement. Statistical parametric maps (SPMs) are

therefore functions of change size (e.g., subtraction images or change score images) and reliability of measuring change, or error variance. Qualitative interpretation and reporting of SPMs as single objects is justifiable given that the probability of obtaining the SPM, or a more unlikely one, is calculated. This sort of approach has been described for images of change scores by Fox et al. (1988) as part of "change distribution analysis." With SPMs, the estimation of this "omnibus significance" reduces to comparing the number of expected and observed pixels above a given threshold using the χ^2 statistic or with reference to the Poisson distribution (Collings, 1977), as previously described (Friston et al., 1990). This comparison is only valid if the size of the SPM is much greater than its resolution. If this were not the case, then the SPM's smoothness will affect the distribution of pixel values.

The omnibus significance relates to the collective profile of differences described by the SPM but does not relate to the significance of change at any single location. To report that a particular cortical area has increased significantly, irrespective of changes elsewhere, the null hypothesis of no

Received March 14, 1990; revised June 18, 1990; accepted January 10, 1991.

Address correspondence and reprint requests to Dr. K. J. Friston at MRC Cyclotron Unit, Hammersmith Hospital, DuCane Rd., London, W12 0HS, U.K.

Abbreviations used: FWHM, full width at half-maximum; PET, positron emission tomography; SPM, statistical parametric map.

change is being rejected for that position. The SPM in this case is being used as a set of repeated measures, each corresponding to a different location. The probability of a significant finding occurring by chance, in this set, must be approximated. That approximation is addressed in this article.

The assessment of significance is confounded by the large number of pixels and by image smoothness imposed by the PET technique. The smoothness present in statistical maps means a large number of comparisons are made that are *not* independent.

There is no satisfactory method that allows a proper adjustment to a threshold that accommodates both large pixel number and smoothness. This is a major obstacle to interpretation. We present a theory and method that makes an adjustment for multiple comparisons that is less severe than a Bonferroni correction, which treats every pixel as an independent test (which they are clearly not). This method depends on assumptions that are only tenable in the special case of SPMs.

The approaches described are empirically validated and then applied to a verbal fluency PET activation study, as an illustration of their use.

THEORY

A statistical map is a two-dimensional image process whose pixel values are a statistic that reflects a difference. For any comparison (pixel), under the null hypothesis, the distribution of the statistic approximates to an analytically determined distribution. This distribution allows the probability of getting an observed value of a statistic, or a greater value, to be calculated.

Statistical maps (e.g., of the t statistic from a comparison of means) can be transformed to the normal distribution (using a probability integral transformation) with no loss of information. Under the null hypothesis, such a map will have pixels whose mean is zero and whose standard distribution is unity. However, the image process is not completely specified until the smoothness is known. This can be estimated in terms of the spectral density, the autocorrelation function, or the variance of the image's gradients (see below). Furthermore, it is assumed that under the null hypothesis, this wide-sense stationary random process will be featureless, its autocorrelation function being a function only of the distance between two points. If this is the case, the process is said to be homogeneous and isotropic (Rosenfeld and Kak, 1976, p. 39). The concept of homogeneous random processes is a generalization of a one-dimensional stochastic process to higher dimensions. For a full dis-

cussion of stationary processes, see Cox and Miller (1980, pp. 272–366).

Threshold adjustment

In the special case of no smoothing, each pixel will represent an uncorrelated, independent measure. The probability that any pixel will have a value of \dagger or greater is given by p . For a Gaussian process, this is the appropriate integral under the normal distribution $[p(\$, \dagger)]$. For N pixels, an adjustment to \dagger is required to protect against false positives. A Bonferroni correction will ensure that the probability of obtaining at least one pixel above \dagger is less than p , where \dagger now corresponds to a probability of p/N for each pixel. N represents the number of independent measures or degrees of freedom. In this case, the event of interest is a single pixel exceeding \dagger ; however, in the presence of smoothing, pixel values will not be independent, suggesting that this would be an overcorrection. In the presence of smoothing, pixels that exceed \dagger will cluster together and constitute suprathreshold regions. The number of these regions will be less than the number of pixels above \dagger , as each region has many pixels. One approach is to reformulate the problem in terms of the event of interest. The event of interest (a false positive in the absence of change) can be defined as the centre of a contiguous region above \dagger . The number of false positives and the number of centres will be the same. The probability (Ω) that such an event will occur at any pixel once estimated can be adjusted using a Bonferroni correction to \dagger such that $\Omega = 0.05/N$. Intuitively, it might be thought that Ω falls as the number of pixels per false positive (or smoothness) rises. Indeed, this is the case and the precise relationship depends on smoothness. We will refer to the threshold adjustment for repeated measures in smooth processes as a *smoothness adjustment*.

Measuring smoothness

It is clear that smoothness must be estimated. This is a problem of parameter estimation, where smoothness can be reduced to a single parameter s (the actual smoothness will be denoted by s and the estimated value of s by \hat{s}). It is assumed that any "well-behaved" SPM could have been produced by convolving a completely random uncorrelated process with a Gaussian filter of width s . "Well-behaved" SPMs, under the null hypothesis, are defined as homogeneous and isotropic random fields with a Gaussian autocorrelation function. For a definition of these terms and a fuller discussion, see Rosenfeld and Kak (1976, p. 37). A number of approaches to estimating s can be envisaged. The approach chosen uses the measured variance of the

first partial derivatives of the image process (S_{θ}^2), where, in all dimensions (see the Appendix),

$$s = \sqrt{1/(2 \cdot S_{\theta}^2)} \quad (1)$$

If the size of a discrete pixel is small relative to the smoothness (s), then the partial derivative in one dimension is approximated by the difference in value between a pixel and its neighbour. The variance of the derivative can be approximated by calculating the variance of pooled interpixel differences along rows and columns of the SPM.

Given the assumption that the statistical process is stationary, isotropic, and Gaussian, for a threshold (\dagger), the probability of a false positive (Ω) is approximated by

$$\Omega \approx [32\pi \cdot s^2 e^{\dagger^2} p(\$, \dagger)]^{-1} \quad (2)$$

(see the Appendix for a definition of terms). This equality relates the smoothness s and threshold \dagger to the probability density of a false positive (Ω). To make a smoothness adjustment at $p = 0.05$, \dagger is found that satisfies $\Omega = 0.05/N$, where N is the size of the process, or number of pixels.

The above relationship is derived from theoretical considerations and the inherent assumptions require validation. In order to confirm that using discrete pixels as opposed to continuous variables has little effect on the above equalities, they were tested using simulated data. The assumptions made in generalizing the continuous variable model to discretized pixel values are that (a) pixel size is small relative to s and (b) s is small relative to the size of the SPM.

METHODS

Empirical validation of relationships

Estimating s from the variance of the first partial derivatives. A 256×256 image process was generated using a random (Gaussian) number generator. This matrix of random uncorrelated numbers was convolved with a series of Gaussian filters ($s = 2$ to 6 in 0.4 pixel steps). To simulate the effect of pixelation on this simulated continuous process, two secondary processes were derived from the 256×256 matrix. By taking the mean of 4 and of 16 contiguous pixels in the large process, 128×128 and 64×64 pixel processes were created. The partial derivative was measured in both dimensions (differences over rows and columns). The pooled variance of these differences was used to estimate s according to Eq. (1). This estimate was repeated 10 times and the mean compared to the known value. The results were presented graphically by regressing the observed s against the actual value of s . The percentage error across all residuals was recorded as a measure of precision for both pixel sizes.

False positive rates in smooth Gaussian processes. The ratio of false positives to pixel number was measured in

100 simulated (as above) Gaussian processes 128×128 pixels in size and normalized to unit variance. This ratio is an estimate of Ω —the probability per pixel of a false positive. The estimation was made for different values of smoothness, keeping threshold constant, and for different values of the threshold, keeping smoothness constant. The ranges used correspond to values typically encountered with real data. The range of smoothness s (determined by the width of convolution filter used in generating the simulated matrices) was 2 to 6 in 0.2 pixel steps, at a threshold (\dagger) of 1.96 and \dagger was varied from 1.8 to 4 in steps of 0.2 , at a smoothness of 4 pixels. The results of this simulation were presented graphically by plotting observed and predicted [Eq. (2)] false positives per pixel against smoothness (s) and threshold (\dagger).

Verbal fluency PET activation study

Subjects and activation paradigm. The subjects were four right-handed male volunteers (age 26 – 45 years) with no history of neurological or psychiatric symptoms. Each subject underwent six consecutive studies whilst performing different verbal fluency tasks. The details of these tasks and neurophysiological findings will be reported in a separate article. In brief, the first and last studies were under standardized relaxed conditions and constitute the baselines. The remaining four conditions included one task that involved listening to aurally presented words and nonwords.

Data acquisition. Scanning was performed using a PET scanner (CTI model 931-08/12, Knoxville, TN, U.S.A.) whose physical characteristics have been described (Spinks et al., 1988). Scans were reconstructed using a Hanning filter with a cutoff frequency of 0.5 giving a transaxial resolution of 8.5 mm full width at half-maximum (FWHM). The reconstructed images contained 128×128 pixels, each having a size of 2.05×2.05 mm.

Subjects inhaled $C^{15}O_2$ at a concentration of 6 MBq/ml and a flow rate of 500 ml/min through a standard oxygen face mask for a period of 2 min. Dynamic PET scans were collected for a period of 3.5 min starting 0.5 min prior to $C^{15}O_2$ delivery, according to a protocol described elsewhere (Lammertsma et al., 1990) used for measuring regional CBF (rCBF). For the present study, integrated counts per pixel for the 2 min buildup phase of activity during $C^{15}O_2$ inhalation were used.

Image analysis. Calculations and image matrix manipulations were performed in PRO MATLAB (Mathworks Inc., Sherborn, MA, U.S.A.).

The 15 original scan slices (6.75 mm interplane distance) were interpolated, and transformed into a 26 -plane standard stereotactic space. The intercommissural line was identified, according to a previously described method, directly from the primary PET image (Friston et al., 1989). When stereotactically normalized, 1 pixel in the transformed image represents 2 mm in the x and y dimensions according to the atlas of Talairach and Tournoux (1988). The brain image falls within the central 65×87 submatrix of the 128×128 image. The slices have an interplanar distance of 4 mm. Significant areas within the statistical parametric maps were displayed with the proportional "grid" and brain outline used by the atlas (Talairach and Tournoux, 1988). Each image was smoothed using a Gaussian filter (FWHM = 10 pixels). This increases the signal-to-noise ratio and accounts for variations in gyral anatomy.

Statistical analysis

Global variance was removed according to a previously described method (Friston et al., 1990) using analysis of covariance with global activity as covariate on a pixel-by-pixel basis. This analysis generates adjusted pixel means for each of the six conditions and the associated adjusted error variance required for comparison of these means. The listening condition was compared to the non-listening verbal fluency tasks by comparing the adjusted condition means with the t statistic (Wildt and Ahtola, 1978). The t value was transformed to the standard Gaussian distribution. The set of t values (one for each pixel) constitutes the statistical image (SPM).

The expected and observed distributions of the SPM pixel values over all pixels were plotted together for comparison.

s was estimated as for the simulation and substituted in Eq. (2). The threshold \dagger was chosen that satisfied Eq. (2) for $\Omega = 0.05/N$ (where N = the brain area in pixels for each slice). This is the smoothness adjusted threshold (per slice) used to detect significantly activated brain foci.

As a final empirical validation of these procedures, the whole analysis was repeated identically step for step using six groups of four, 26-slice, uniform phantom images with corresponding counts per pixel. The distribution of pixel values in the resulting SPM was displayed with the distribution expected and the number of false positives was recorded.

RESULTS

Validation of relationships

The results of the simulations testing the two relationships [Eq. (1) and (2)] are shown in Figs. 1 and 2. The agreement is evident. The percentage error in estimating s from the first derivatives of the sim-

ulated SPMs was 3.6% in the 128×128 pixel images and 6.4% in the 64×64 pixel images. Much of this error is attributable to a systematic underestimation of smoothness for high values of s . For 128×128 pixel processes, the estimation is good for s less than 4 pixels. This limit corresponds to a FWHM in the SPM of 18.8 mm. This is above the upper limit of smoothness typically encountered in this unit.

The estimated values of Ω compare well to those predicted by Eq. (2) and serve as a partial validation of the assumptions (see the Appendix) made in deriving this equality.

Activation study

Estimated smoothness (s) for the verbal fluency data corresponded to an effective resolution with FWHM of 13.21 mm. The expected and observed distributions are seen in Fig. 3, top. The measured distribution shows a degree of kurtosis, which may be a result of differences in brain states during the listening task compared to the remaining three fluency tasks. Over all 26 slices, the average adjusted threshold (\dagger) was 3.655 [Eq. (2), dotted line in Fig. 3, top]. There were two foci detected at this threshold. As expected, both were in the superior temporal regions, one on the left and one on the right. The left-sided focus was in Brodmann's area 42 according to the atlas of Talairach and Tournoux (1988). The slice 8 mm above the intercommissural line is displayed in Fig. 4 with the focus superimposed on the brain outline used by the atlas. For this slice, \dagger was 3.746. This corresponds to an uncorrected sig-

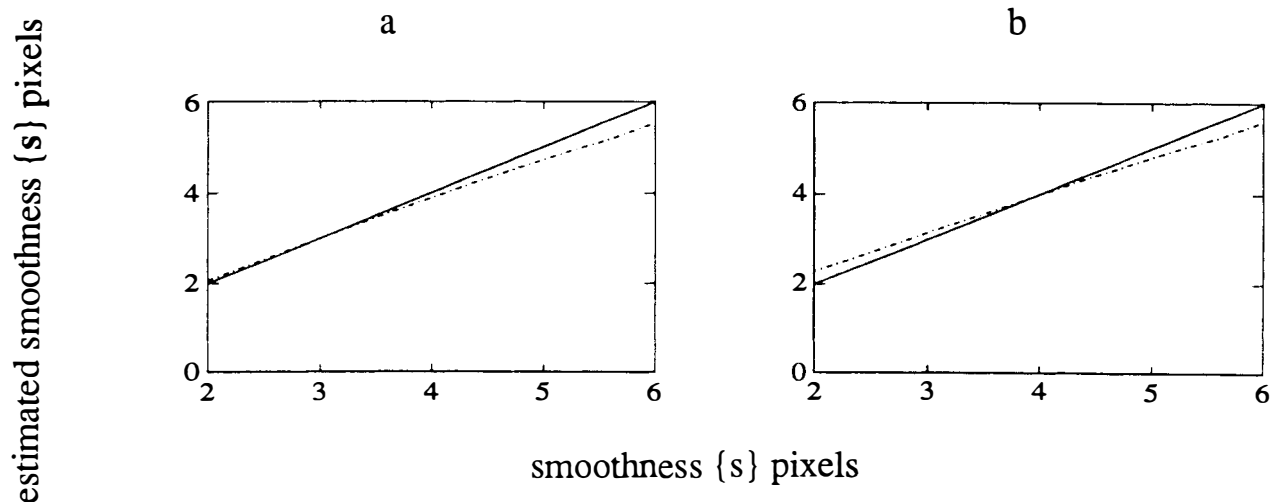


FIG. 1. Regression of estimated smoothness (\hat{s}) parameter on the actual value s . s is the width of a Gaussian convolution function that has been used to smooth a completely random, unsmooth image process to give the observed "smoothness." The estimation of \hat{s} is based on the variance of the image process's gradient over the whole plane. The data above were generated by smoothing a spectrally white Gaussian number field with a convolution function of known width and then estimating that width retrospectively using the measured variance of its gradients [Eq. (1) in the text]. In order to assess the effect of pixelating a continuous process, a 256×256 matrix was "contracted" to a 128×128 (a) and a 64×64 pixel matrix (b). This contraction involves taking the average of 4 and 16 pixel "blocks" in the large matrix, respectively.

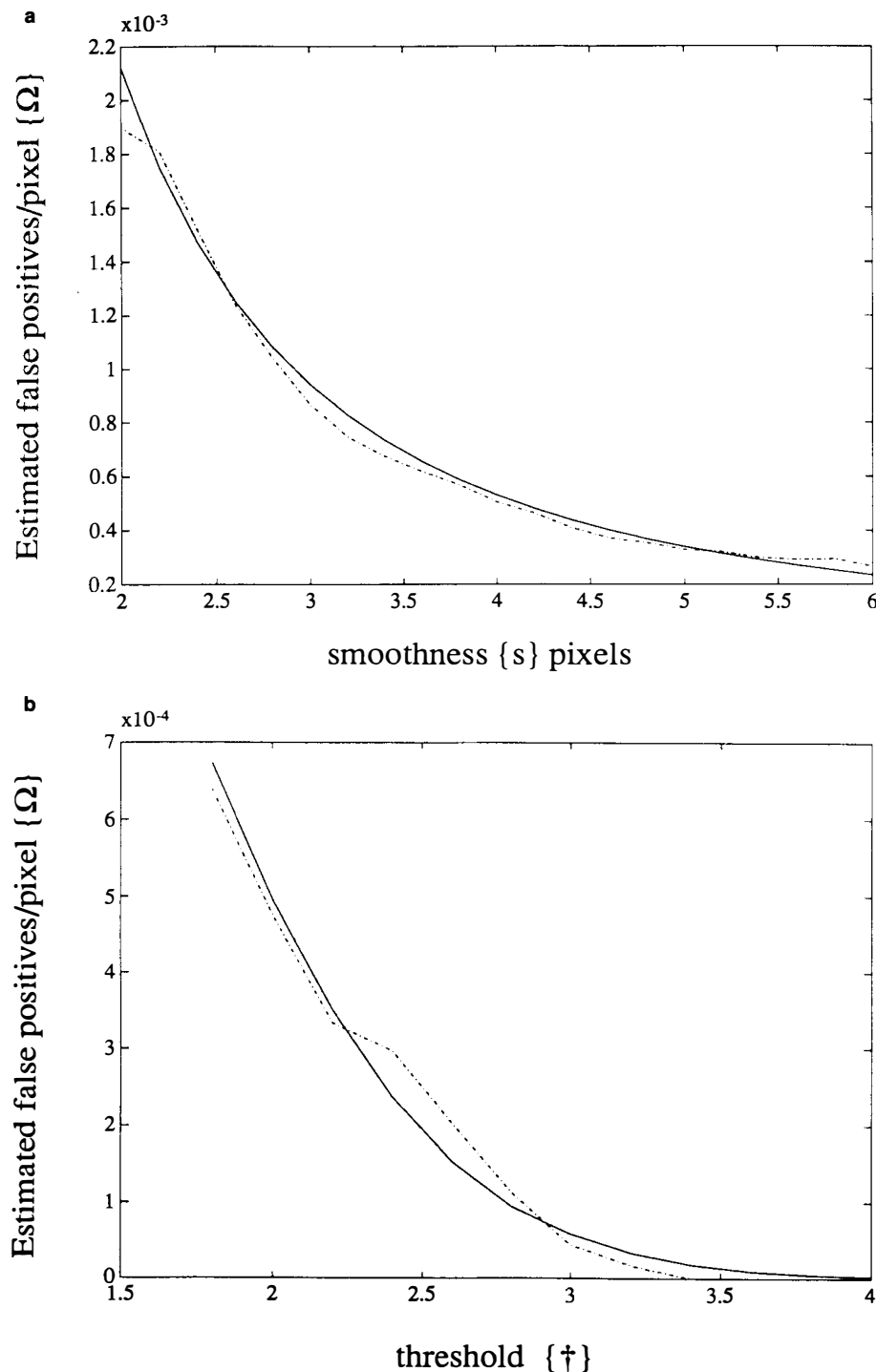


FIG. 2. Comparison of the expected and predicted number of false positives/pixel in a simulated Gaussian statistical parametric map. A false positive is defined as the centre of a collection of contiguous pixels with a value above a given threshold. To test the validity of the assumptions inherent in Eq. (2), the dependency of Ω (probability of a false positive) on s (smoothness) and \dagger (threshold), s and \dagger were varied independently and the number of false positives/pixel observed compared to that predicted. **a:** The x axis is the smoothness parameter s defined in the legend for Fig. 1. The y axis is the false-positive rate or false positives/pixel; this is an estimate of Ω . The variance of the simulated Gaussian statistical parametric map was unity in all cases. This means that the number of pixels above 1.96 was equivalent for different degrees of smoothing. Despite this, the number of false positives falls as s increases. The dotted line is the observed false-positive rate and the solid line that predicted analytically [from Eq. (2) in the text]. **b:** The same as for **a** but in this case \dagger has been varied with $s = 4$ pixels.

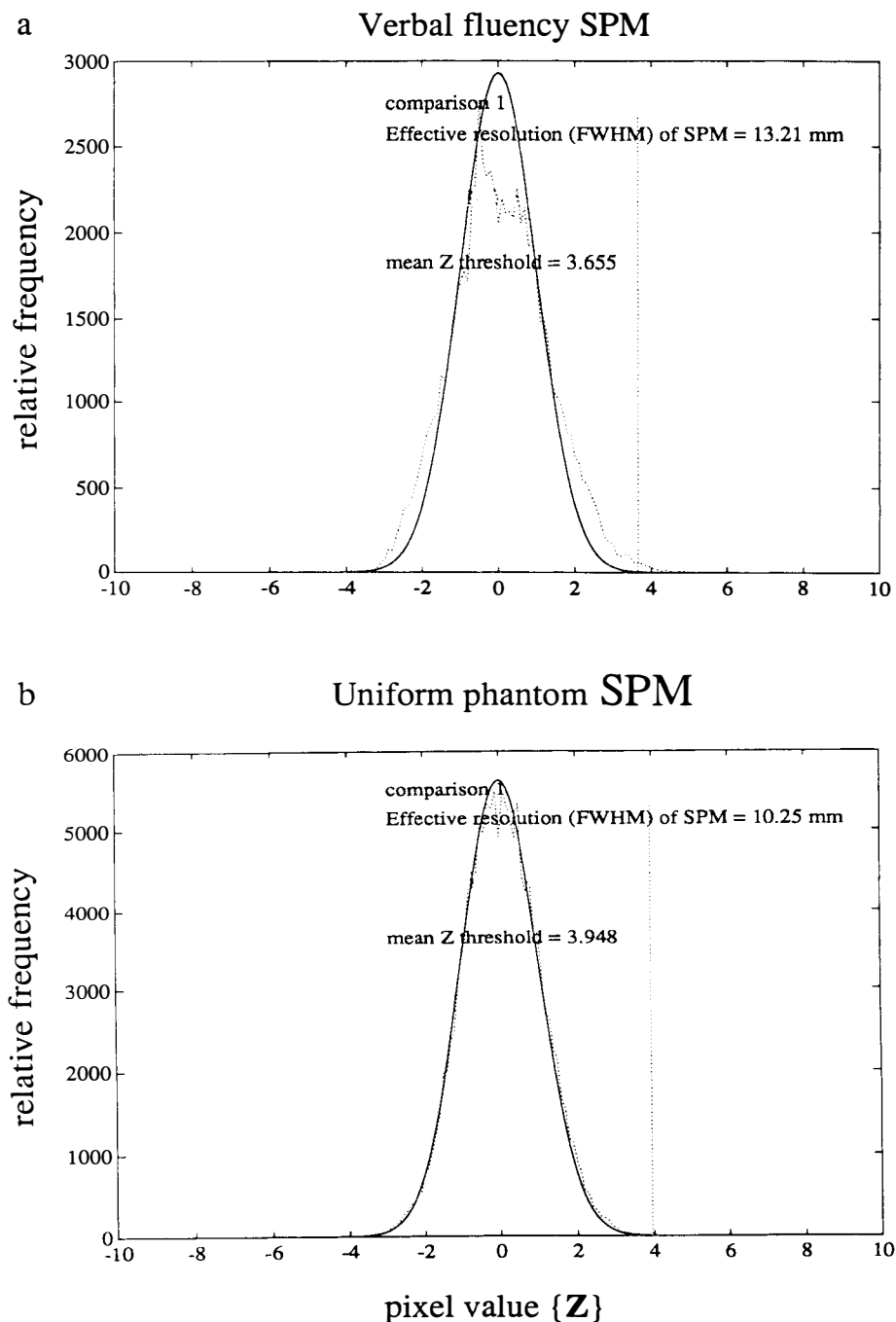
nificance of $p \approx 0.00014$. The greatest pixel value in this slice was 4.83, i.e., 4.83 standard deviations above the mean of 0.

Phantom study

The expected and measured distributions of pixel values in the SPM from the phantom study are seen in Fig. 3, bottom. The effective FWHM for this SPM was 10.25 mm. This is slightly smaller than for

the real data and may suggest that biological coherence in local activity may contribute (marginally) to smoothness in the final SPM. The dotted line is the average threshold across all planes ($= 3.948$). The observed distribution compares well with that predicted. There was one false positive 4 pixels in size on plane 12. A false positive would be expected every 20th plane. Twenty-six planes were analyzed in this study.

FIG. 3. Comparison of expected and observed distributions of pixel values in SPMs derived empirically from a verbal fluency activation study (a) and uniform phantoms (b). The t statistic was used to compare adjusted condition means from four conditions (one listening to aurally presented material and three nonlistening of the activation study). The t statistical parametric map (SPM) was rendered Gaussian by transformation and s estimated as in Fig. 1. An adjusted threshold was calculated for each plane by finding the values of \dagger that satisfied Eq. (2) for $\Omega = 0.05/N$. The dotted line is the standard Gaussian distribution with an SD of 1.



DISCUSSION

By modeling statistical maps as two-dimensional stationary stochastic processes and parameterizing their smoothness, it is possible to predict the expected frequency of certain features due to chance. By defining a false positive as the centre of a contiguous group of suprathreshold pixels, it is possible to derive an approximation relating the probability of a false positive, the threshold, and the image's

smoothness. This relationship can be used to adjust the threshold for any given pixel number and smoothness. The assumptions made are (a) the size of the statistical process is large relative to s (smoothness), (b) the statistical process is Gaussian, (c) the statistical process is homogeneous and isotropic with a Gaussian autocorrelation function, (d) foci above a threshold are ellipsoid, (e) the threshold is sufficiently high enough for d to hold as an approximation, and (f) pixel size is small relative

regions activated by listening

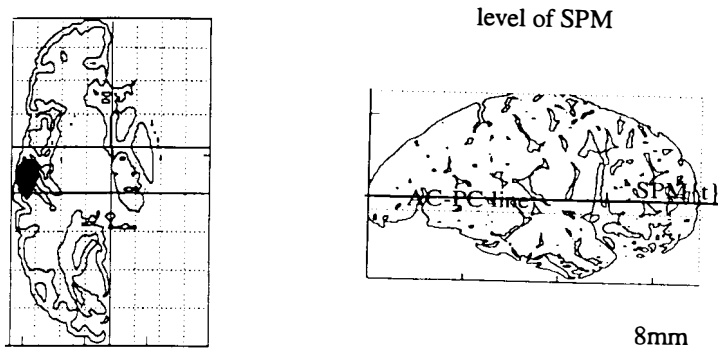


FIG. 4. Foci of pixels whose thresholds exceed \dagger (defined in the legend for Fig. 3). The focus displayed represents the extra-auditory area (Brodmann's area 42/41). This activation is attributable to listening. The focus is superimposed on the brain outline and proportion grid used by the atlas of Talairach and Tournoux (1988). The level of the SPM was 8 mm above the intercommissural line and is shown on the right lateral drawing.

smoothness adjustment to $p=0.05$

to s . These assumptions are only tenable under the null hypothesis of no change between the images compared.

Central to these methods is the estimation of s , the smoothness parameter. It is potentially the weak link in the chain and requires a large amount of data for precise determination. In the method described, s was determined retrospectively from the image's (approximated) partial derivatives. This estimation is unavoidable for proper characterization of statistical maps.

It has been shown that treating the image process as a stationary stochastic process can be useful. The property of SPMs that justifies this approach is their well-behaved spatial organization. This topographic attribute of SPMs makes them a special case in terms of statistical analysis. In particular, the null hypothesis must by implication have a topographic component. The question of interest is not simply "has change occurred?" but "is there change, and if so where?" Differences in the null hypothesis can have profound effects on the assess-

ment of significance. Consider three classes of null hypothesis (summarized in Table 1). In the first case, the null hypothesis explicitly includes a location. This topographically constrained null hypothesis is concerned, a priori, with detection of differences in a particular place and is rejected if the pixel value at that location corresponds to, say, $p < 0.05$ without any adjustment for repeated measures. Rejection of this null hypothesis is equivalent to concluding that activity at the *predetermined location* has changed.

The second class of null hypothesis has no topographic component and states that no differences in the profile of cerebral activity will be found. This hypothesis can be rejected if the collective distribution of pixel values, or a more "extreme" distribution, has a low ($p < 0.05$) probability of chance occurrence. If rejected, it is concluded that there is a difference *somewhere* in the brain and the pattern of cerebral activity has changed. The site of these differences remains unknown but the SPM can be reported in its entirety as a single result, with the

TABLE 1. Summary of different classes of null hypothesis and suggested assessment of significance

Research question	Null hypothesis	Criteria for rejection
The primary auditory cortex ($x = -50$, $y = -20$, $z = 12$ mm) responds to tones whose frequencies are modulated over time to a greater extent than it does to constant tones with the same frequencies and loudness	The SPM pixel value at -50 , -20 , 12 is zero	$p < 0.05$ at x , y , z
Are there any differences in the profile of cerebral activity between schizophrenics with auditory hallucinations and those without	Single null hypothesis: the number of pixels corresponding to $p < 0.05$ is not significantly greater than chance expectation, i.e., there is no omnibus difference in the profile of activity	Suprathreshold pixel number is greater than $0.05 \times$ total pixel number at $p < 0.05$
What cortical region is critically concerned with the emotional tone of spoken words	Many null hypotheses: each pixel is zero; the number of null hypotheses is equal to the number of pixels	Pixel value exceeds $p = 0.05$ adjusted as described to account for the number of null hypotheses tested

corresponding "omnibus" significance. Qualitative features suggesting regional specificity of differences can be observed descriptively. Clearly, to report all of the pixel values (above, say, $p = 0.05$) requires the results to be displayed as an image.

Finally, to say *activation has occurred at position* x, y, z , one is rejecting the null hypothesis that the change at that location is zero. This null hypothesis is one of a large set that have been simultaneously tested for all locations. If significant following an adjustment for multiple comparisons, a point is reported selectively with the stereotactic coordinates, irrespective of changes elsewhere.

There is a fundamental difference between a single null hypothesis relating to the distribution of statistical values across all pixels and a set of null hypotheses, one for each pixel. Note that it is possible to reject the second class of null hypothesis whilst being unable to reject the third in the same SPM. In other words, for a given level of significance, the topographic pattern of differences can be significant yet (using the described criteria) no single location can be said to be significant in its own right. This apparent paradox derives from using the same threshold for testing different hypotheses. For cognitive activations, this outcome may be understandable in physiological terms. The brain systems underlying complex behaviour are likely to be distributed and consequently the activation profile will be distributed across many pixels.

Acknowledgment: We would like to thank the editor for guidance in the presentation of this work and the reviewer for substantial help in developing these ideas. K.J.F. is funded by the Wellcome Trust. We wish to thank the staff of the Cyclotron Unit for valuable discussion and for making these studies possible and in particular Claire Taylor and Graham Lewington.

APPENDIX

Let $V\{\}$ denote variance, $C\{\}$ denote covariance, and $E\{\}$ the expectation. The following assumptions are made:

(a) Pixel size is sufficiently small for a Gaussian statistical map (X) to be an approximation of a continuous two-dimensional process $X(x,y)$ with spectral density $g_X(u,v)$.

(b) $X(x,y)$ is homogeneous and isotropic (Rosenfeld and Kak, 1976, p. 39) with unit variance (S_x^2). In terms of spectral density (see Cox and Miller, 1980, p. 313),

$$s_x^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_X(u,v) du dv = 1 \quad (A1)$$

(c) $X(x,y)$ is the process resulting from a convolution of a random uncorrelated process $Y(x,y)$ [of

uniform spectral density $g_Y(u,v)$] with a Gaussian convolution function $c(x,y)$. Indeed, this is how the simulated processes were created (see the Methods section). $c(x,y)$ is characterized by width s , where

$$c(x,y) = \frac{1}{(2\pi s^2)} \cdot e^{-(x^2+y^2)/2s^2} \quad (A2)$$

An alternative expression for function width is the FWHM. In this case, $s = \text{FWHM}/\sqrt{8\ln(2)}$. The transfer function associated with $c(x,y)$ is $lc(u,v)$. The transfer function is related to the Fourier transform of the convolution function (see Rosenfeld and Kak, 1976, p. 156 for a discussion of the relationship between the line spread function, or convolution filter and transfer function; see Cox and Miller, 1980, p. 320 for examples in one-dimensional processes).

(d) The dimensions of $X(x,y)$ (x and y) are large with respect to s in Eq. (A2). X is sufficiently large for edge effects to be ignored.

(e) Regions for which $X(x,y) > \dagger$ holds are elliptical with radii r_x and r_y . Because $X(x,y)$ is isotropic, it is assumed $E\{r_x\} = E\{r_y\}$ and $C\{r_x, r_y\} = 0$.

(f) \dagger is sufficiently large for (e) to be a reasonable approximation.

Derivations

From assumption (c), the spectral density of $X(x,y)$ is related to the spectral density of $Y(x,y)$ (a constant) and the transfer function of $c(x,y)$, where

$$g_X(u,v) = |lc(u,v)|^2 g_Y(u,v) \quad (A3)$$

where (see Cox and Miller, 1980, p. 322)

$$\begin{aligned} lc(u,v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(u \cdot x + v \cdot y)} c(x,y) dx dy \\ &= e^{-(u^2 + v^2) \cdot s^2 / 2} \end{aligned} \quad (A4)$$

Therefore,

$$\begin{aligned} g_X(u,v) &= g_X(u) \cdot g_X(v) \\ &= g_Y(u,v) \cdot e^{-(u^2 + v^2) \cdot s^2} \end{aligned} \quad (A5)$$

where $g_Y(u,v) = s^2/\pi$, to satisfy Eq. (A1).

The variance of the first partial derivative of $X(S_0^2)$ with respect to x (and y) is $V\{\partial(X)/\partial x\} = V\{\partial(X)/\partial y\}$ (because X is isotropic) and is given by (Cox and Miller, 1980, p. 321)

$$S_0^2 = S_X^2 \cdot \int_{-\infty}^{\infty} u^2 g_X(u) du = \frac{1}{2s^2} \quad (A6)$$

by Eqs. (A1) and (A5). Using an estimate of S_0^2 , this relationship is used to estimate smoothness s .

Expectation of false positives

A false positive at threshold \dagger is defined as the centre of a contiguous region for which $X(x,y) > \dagger$ is true.

Let $qx \Delta x$ be the probability of an upcross in x at $y = Y$ of the process $X(x, Y)$ at threshold \dagger in $(x, x + \Delta x)$. Let $p(a,b)$ be the joint probability density function of the process $X(x, Y)$ and its derivative $\partial X(x, Y)/\partial x$. $p(a,b)$ does not involve x because the image process is stationary. Furthermore, it can be shown for a stationary Gaussian process that the covariance of $X(x, Y)$ and $\partial X(x, Y)/\partial x$ is zero, i.e., $C\{X(x, Y), \partial X(x, Y)/\partial x\} = 0$. Hence, $p(a,b)$ has a bivariate normal distribution with mean $(0,0)$ and covariance matrix (see Cox and Miller, 1980, p. 294) of

$$\begin{pmatrix} 1 & 0 \\ 0 & 1/2s^2 \end{pmatrix}$$

The term $1/2s^2$ is the variance of the first partial derivative s^2_θ in Eq. (A6). The criteria for a threshold crossing in x is that the process $X(x, Y)$ intercepts the level \dagger with positive slope. Over a small distance Δx , the slope will be linear and the conditions to be satisfied will be $b > 0$ and $\dagger - b \cdot \Delta x < a < \dagger$. In the limit of $\Delta x \rightarrow 0$,

$$qx = \int_0^\infty b \cdot p(\dagger, b) db$$

where

$$p(a,b) = \frac{1}{2\pi} \cdot \frac{1}{s\theta} \cdot \exp\left(\frac{-a^2}{2} - \frac{b^2}{2s^2_\theta}\right)$$

Thus,

$$qx = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{2s^2}} \cdot e^{-\dagger^2/2} \tag{A7}$$

See Cox and Miller (1980, p. 295) for a general derivation but with $\dagger = 0$.

The regions of $X(x,y)$ for which $X > \dagger$ are modeled as ellipses of radii rx and ry . The probability density of interest is the (uniform) probability that the centre of an ellipse will fall in the intervals $x:x + dx$, $y:y + dy$. Let this be Ω .

The conditional density, given an ellipse, of its radii is $p(rx, ry | \Omega = 1)$. By assumption (e), this will factorize into the marginal probability densities. Let $p(rx, ry | \Omega = 1) = p(rx) \cdot p(ry) \cdot p(rx, ry | \Omega = 1)$ will not be a function of x or y because ellipse size will not vary systematically with position. Therefore, the probability that the centre of an ellipse of

radii in the intervals $rx:rx + drx$, $ry:ry + dry$ will fall in $x:x + dx$, $y:y + dy$ is

$$f(x,y,rx,ry) = \Omega \cdot p(rx) \cdot p(ry) \tag{A8}$$

The probability of an upcross in $X(x, Y)$ over $x = 0 - x$ at \dagger is the approximate probability that the centre of an ellipse will fall in the interval $x = 0 - x$, $y = Y - ry - Y + ry$ (approximate because of the simplifying assumptions):

$$\begin{aligned} \int_0^x qx dx &\approx \int_0^x \int_{Y-ry}^{Y+ry} \int_0^\infty \int_0^\infty f(x,y,rx,ry) dx dy drx dry \\ &\approx \Omega \int_0^x dx \int_0^\infty drx \int_0^\infty 2 \cdot ry \cdot p(ry) dry \end{aligned}$$

Therefore,

$$qx \approx 2\Omega E\{ry\} \tag{A9}$$

The contribution of an elliptical region to the total expected area for which $X(x,y) > \dagger$ is true is $\pi rx \cdot ry$. This expectation of area is also determined by the probability that $X(x,y) > \dagger = p(X > \dagger)$. As X is Gaussian with unit variance, this is given by the integral under the normal distribution. Let $p(\S, \dagger)$ denote this integral:

$$p(\S, \dagger) = \int_\dagger^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Therefore, for a process of size x,y , the expectation of area is (assuming no overlap)

$$\int_0^x \int_0^y p(X > \dagger) dx dy = \int_0^x \int_0^y \int_0^\infty \int_0^\infty \pi \cdot rx \cdot ry \cdot f(x,y,rx,ry) dx dy drx dry$$

$$\begin{aligned} x \cdot y \cdot p(\S, \dagger) &= x \cdot y \cdot \Omega \cdot \pi \cdot \int_0^\infty \int_0^\infty rx \\ &\quad \cdot ry \cdot p(rx, ry) drx dry \end{aligned}$$

Given that $p(rx, ry) = p(rx) \cdot p(ry)$ and $E\{rx\} = E\{ry\}$, this equality reduces to

$$p(\S, \dagger) = \Omega \pi E\{ry\}^2 \tag{A10}$$

Eliminating $E\{ry\}$ from Eqs. (A9) and (A10) and rearranging for Ω , we have

$$\Omega \approx [32\pi \cdot s^2 \cdot e^{\dagger^2/2} \cdot p(\S, \dagger)]^{-1}$$

Ω is the probability of X [over $x:x + dx$, $y:y + dy$ for any x,y] taking a value greater than \dagger and being at the centre of an ellipsoid region where $X > \dagger$.

To make a Bonferroni correction at $p = 0.05$ for a process X of size N , we require \dagger to satisfy $\Omega = 0.05/N$.

REFERENCES

- Collings SN (1977) *Mathematical Statistics: Its Setting and Scope*. Milton Keynes, The Open University Press, pp 73–75
- Cox DR, Miller HD (1980) *The Theory of Stochastic Processes*. New York, Chapman and Hall, pp 272–336
- Fox PT, Mintun MA, Reiman EM, Raichle ME (1988) Enhanced detection of focal brain responses using intersubject averaging and change distribution analysis of subtracted PET images. *J Cereb Blood Flow Metab* 8:642–653
- Friston KJ, Passingham RE, Nutt JG, Heather JD, Sawle GV, Frackowiak RSJ (1989) Localization in PET images: Direct fitting of the intercommissural (AC–PC) line. *J Cereb Blood Flow Metab* 9:690–695
- Friston KJ, Frith CD, Liddle PF, Lammertsma AA, Dolan RD, Frackowiak RSJ (1990) The relationship between local and global changes in PET scans. *J Cereb Blood Flow Metab* 10:458–466
- Lammertsma AA, Cunningham VJ, Deiber MP, Heather JD, Bloomfield PM, Nutt JG, Frackowiak RSJ, Jones T (1990) Combination of dynamic and integral methods for generating reproducible functional CBF images. *J Cereb Blood Flow Metab* 10:675–686
- Rosenfeld A, Kak AC (1976) *Digital Image Processing*. New York, Academic Press
- Spinks TJ, Jones T, Gilardi MC, Heather JD (1988) Physical performance of the latest generation of commercial positron scanner. *IEEE Trans Nucl Sci* 35:721–725
- Talairach J, Tournoux P (1988) *A Co-planar Stereotaxic Atlas of a Human Brain*. Stuttgart, Thieme-Verlag
- Wildt AR, Ahtola OT (1978) *Analysis of Covariance*, Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 12. Beverly Hills, London, Sage Publications