Analysis of fMRI Time-Series Revisited - Again

K.J. Worsley K.J. Friston*

*Department of Mathematics and Statistics, McGill University, 805 Sherbrooke St. West, Montreal, Québec, Canada H3A 2K6; * Wellcome Department of Cognitive Neurology, MRC Cyclotron Unit, Hammersmith Hospital, DuCane Road, London W12 0HS, U.K.*

Friston *et al.* (1995) presented a method for detecting activations in fMRI time-series based on the general linear model and a heuristic analysis of the effective degrees of freedom. In this communication we present corrected results that replace those of the previous paper and solve the same problem without recourse to heuristic arguments. Specifically we introduce a proper and unbiased estimator for the error terms and provide a more generally correct expression for the effective degrees of freedom. The previous estimates of error variance were biased, and in some instances could have led to a 10-20% overestimate of Z values. Although the previous results are almost correct for the random regressors chosen for validation, the present theoretical results are exact for any covariate or waveform. We comment on some aspects of experimental design and data analysis, in the light of the theoretical framework discussed here.

Running title:
Analysis of fMRI Time-Series Revisited - Again

Addresses of corresponding author:
e-mail: keith@zaphod.math.mcgill.ca
ph: 1-514-398-3842
fax: 1-514-398-3899

1 Introduction

In this paper we review the approach considered in Friston *et al.* (1995) for the analysis of fMRI time-series. This previous paper had two shortcomings. Firstly the expression for the variance of the parameter estimates was inappropriate and biased. Secondly the subsequent analysis, using the effective degrees of freedom, was based on heuristic arguments which can, as we show below, be replaced with proper derivations.

This paper is divided into two parts. The first section presents theoretical results that are required to implement the extension of the general linear model described in Friston *et al.* (1995), in a more correct fashion. This section concludes with a re-analysis of the data presented in Friston *et al.* (1995) to compare the original and revised approaches.

Secondly we review some issues in experimental design and data analysis that depend directly on the mathematical theory presented here and on the the theory of Gaussian fields that is used in making statistical inferences about activation foci. The latter pertain to spatial smoothing, statistical power and the size of the underlying physiological activation. We demonstrate some of our points with further analyses of the data used in previous sections.

2 Extending the general linear model

2.1 Theory

The aim of Friston *et al.* (1995) is to estimate the parameter vector β of the linear model

$$\mathbf{X} = \mathbf{G}\beta + \mathbf{e}$$

where $\mathbf{X}$ represents the unsmoothed time-series, and the components of the error vector $\mathbf{e}$ are independent and normally distributed with mean 0 and variance σ^2 (equation (1), page 46). $\mathbf{X}$ is linearly smoothed by multiplying by a matrix $\mathbf{K}$ whose rows represent the hemodynamic response function, and "...the aim of this work is to extend the general linear model so that it can be applied to data with a stationary and known autocorrelation..." (page 45).

There is a large literature on this topic, and one of the best early references is Watson (1955). It can be shown that the optimum estimator of β that "...maximises variance in the signal frequencies relative to other frequencies" (bottom of page 45) is obtained by deconvoluting or un-smoothing the data by multiplying by $\mathbf{K}^{-1}$ and applying least squares to the uncorrelated data $\mathbf{X}$ (Gauss-Markov Theorem). Because this inversion is very sensitive to the correct specification of $\mathbf{K}$, Friston *et al.* chose to apply least squares to the smoothed observations instead, to obtain the following estimator of β

$$\mathbf{b} = (\mathbf{G}^{*T}\mathbf{G}^{*})^{-1}\mathbf{G}^{*T}\mathbf{K}\mathbf{X},$$

where $\mathbf{G}^{*} = \mathbf{K}\mathbf{G}$. Although not fully optimal, $\mathbf{b}$ is unbiased and the loss in efficiency is more than offset by the gain in robustness (see later). From here the derivation departs from that presented in Friston *et al.* (1995) and proceeds as follows:

- Equation (3), page 47, which gives an expression for the variance of $\mathbf{b}$, should be

$$\mathrm{Var}\{\mathbf{b}\} = \sigma^2(\mathbf{G}^{*T}\mathbf{G}^*)^{-1}\mathbf{G}^{*T}\mathbf{V}\mathbf{G}^*(\mathbf{G}^{*T}\mathbf{G}^*)^{-1},$$

  where $\mathbf{V} = \mathbf{K}\mathbf{K}^T$ (Watson, 1955, Seber, 1977, page 144).

- This immediately implies that the test statistic for a particular linear compound $\mathbf{c}$ of the effects (equation (5), page 47) should read:

$$T = \mathbf{c}\mathbf{b}/(\mathbf{c}\epsilon^2(\mathbf{G}^{*T}\mathbf{G}^*)^{-1}\mathbf{G}^{*T}\mathbf{V}\mathbf{G}^*(\mathbf{G}^{*T}\mathbf{G}^*)^{-1}\mathbf{c}^T)^{1/2},$$

  where ϵ^2 is an unbiased estimator of σ^2.

- The estimator of σ^2 given in equation (6), page 47, is biased, and the correct unbiased estimator is obtained by dividing the residual sum of squares by its expectation, to give

$$\epsilon^2 = \mathbf{r}^T\mathbf{r}/\mathrm{trace}(\mathbf{R}\mathbf{V})$$

  where $\mathbf{r} = \mathbf{R}\mathbf{K}\mathbf{X}$ is the vector of residuals and $\mathbf{R}$ is the residual-forming matrix given by

$$\mathbf{R} = \mathbf{I} - \mathbf{G}^*(\mathbf{G}^{*T}\mathbf{G}^*)^{-1}\mathbf{G}^{*T}$$

  and $\mathbf{I}$ is the identity matrix (Seber, 1977, page 146). Note that the "effective degrees of freedom" ν defined in the next section is not used in this calculation. We shall see, however, that it does play a role in giving a better approximation to the null distribution of the T statistic defined above.

- Finally, the expression for the effective degrees of freedom is incorrect, though it can serve as a simple approximation. The original result was derived by removing the fitted values from the observations but ignoring the consequent changes in the covariance structure of the residuals. The correct degrees of freedom, which should replace equation (9), page 48, can be derived for any covariance structure without using Fourier methods [1]:

$$\mathrm{E}\{\epsilon^2\} = \sigma^2, \quad \mathrm{Var}\{\epsilon^2\} = 2\sigma^4\mathrm{trace}(\mathbf{R}\mathbf{V}\mathbf{R}\mathbf{V})/\mathrm{trace}(\mathbf{R}\mathbf{V})^2$$

  (Seber, 1977, page 16). The effective degrees of freedom is then

$$\nu = \frac{2\mathrm{E}\{\epsilon^2\}^2}{\mathrm{Var}\{\epsilon^2\}} = \frac{\mathrm{trace}(\mathbf{R}\mathbf{V})^2}{\mathrm{trace}(\mathbf{R}\mathbf{V}\mathbf{R}\mathbf{V})}.$$

  Unfortunately, there is no simple expression for ν for a Gaussian kernel, and in particular it cannot be factored into a part depending on the degrees of freedom of the model, and a part depending on the hemodynamic response function, as in equation (10), page 48.

[1] Note that there is a minus sign missing in front of ω_i^2 in the expression for $g(\omega_i)$ immediately above equation (10), page 48

2.2 Implications for the original analysis

In general the variance of the parameter estimates is underestimated by equation (3) but the estimator of the variance is overestimated by equation (6), so that the two tend to cancel each other out in the T statistic (5). It can be shown that they do cancel out almost exactly for the random regressors that were chosen for validating the methods, which explains why the biases were not obsereved. However for other non-random regressors these effects do not cancel and large discrepancies can occur.

The correct results were applied to the original example used for validation. There were 100 observations with 9 regressors: a constant, a linear trend, a square-wave of 10 values of +1 (on) followed by 10 values of -1 (off) repeated 5 times, and 6 random regressors generated from a standard Gaussian distribution. The same smoothness $s = \sqrt{8}/3 = 0.94$scans was used [2]:

- The standard deviations of the parameter estimates given by Friston *et al.* divided by the correct standard deviations were 0.86 for the constant term, 0.84 for the linear trend, 0.86 for the square wave, and 0.99±0.05 for the 6 random regressors. Thus the original results give nearly the correct answer (on average) for random regressors. However the standard deviations for the other terms, in particular the square-wave of most interest, are underestimated by 14%, so that the SPM$\{Z\}$ is over-estimated by 16%.

- The original effective degrees of freedom of ν=38.5 is almost correct; the actual value as calculated above is ν=36.2. This is not a serious error since ν is only used to define the null distribution of T, not to actually calculate T.

If we replace the 9 regressors in the previous example by the first 9 Fourier components, that is a constant term and sine and cosine terms of periods 100, 50, 33.3 and 25 scans, then the original standard deviations are too small by a factor of 0.75±0.01, and the correct degrees of freedom is reduced from 38.5 to 35.7. At the other extreme, a high frequency regressor with +1 and -1 for alternate scans gives a standard deviation that is almost twice as big as the correct value.

In summary, the original results give smaller standard deviations for low frequency regressors and the opposite for high frequency regressors; random regressors, which mix all frequencies uniformly, lie in between and the original results give almost the correct answer.

2.3 A re-analysis of the data

The data presented in Friston *et al.* (1995) were re-analysed using the original approach and the new results presented here. Briefly these data were acquired from a single subject performing one of two word generation tasks (word repetition and word generation). Each task was alternated in blocks of ten scans, where each scan was acquired every three seconds.

[2] Note that due to the assumption of a Poisson form for the hemodynamic response function (Friston *et al.* 1994a) the relationship between lag and smoothness is not scale-invariant and requires both to be calculated in seconds.

Figures 1a and 1b show the results of these two analyses in standard format. The SPM$\{Z\}$ is displayed as a maiximum intensity projection on the upper left, after thresholding at $p < 0.01$ (uncorrected). The design matrix is shown on the upper right (in image format) and shows the square-wave reference waveform, linear term, constant and global confounds used in the analysis. Tabular data on the activation foci are shown in the lower panel. Regional effects are characterized by the volume of each region (k), its significnace based on patial extent $P(n_{\max} > k)$, the highest Z value (Z), its significance based on $P(Z_{\max} > u)$ and the location of this primary maximum. We have also included up to three secondary maxima for each region and their associated significance based on the corrected p-value ($P(Z_{\max} > u)$) and the uncorrected p-value. We have only shown foci that survive a threshold of $p < 0.1$ (corrected) based on either peak height or spatial extent.

It can be seen that the highest SPM$\{Z\}$ is reduced (from 4.76 to 4.18) and similarly for all other Z values. The corrected p-values based on spatial extent show that the right prefrontal activation can no longer be considered significant ($p < 0.099$). The corrected p-values based on peak height are now all greater than 0.05. In short the square wave contrast used in this analysis highlights the dangers of the inappropriate and biased estimators adopted in Friston *et al.* (1995).

It is interesting to note that the p-values based on spatial extent appear to be more powerful than those based on peak height. We shall return to this point below.

3 Theoretical implications for experimental design and analysis

3.1 The effective degrees of freedom

Note that the effective degrees of freedom ν is not used to calculate the T statistic, but is used to give a better approximation to its null distribution. By analogy with the χ^2 approximation for quadratic forms (Satterthwaite, 1946), it seems likely that the null distribution of T can be well approximated by a t-distribution with ν degrees of freedom rather than a Gaussian distribution (Worsley *et al.*, 1995). Significance of peaks in SPM$\{T\}$ can be assessed directly from results for t fields (Worsley *et al.*, 1993), or, if the degrees of freedom is large ($\nu > 40$), SPM$\{T\}$ can be converted to SPM$\{Z\}$ and the Gaussian theory of Friston *et al.* (1991), Worsley *et al.* (1992) and Friston *et al.* (1994b) can be used as an approximation. If other sources of information about the error variance are avaliable, e.g. from other subjects or other voxels, then provided the underlying variances are equal these can be pooled to increase the effective degrees of freedom, thereby reducing the error in ϵ^2 and increasing sensitivity of the test.

3.2 Efficiency and optimal experimental design

As noted above, the least-squares estimator $\mathbf{b}$ is not the most efficient estimator of β; by this we mean that it does not have the smallest possible variance amongst all unbiased estimators of β. However we shall now show that it is almost fully efficient for most regressors of interest. A well-known theroem in linear models states that the least-squares estimator is

fully efficient if the regeressors are eigenvectors of the variance matrix $\mathbf{V}$, or in this case, eigenvectors of $\mathbf{K}$ (Seber, 1977, page 63). Now eigenvectors of $\mathbf{K}$ are time variables that are unchanged (in shape) by smoothing with $\mathbf{K}$. From signal detection theory we know that the time variables unaffected by a stationary linear filter are the Fourier sine and cosine functions themselves (provided the time interval is long enough). Thus the parameters of any regression model composed of Fourier sine and cosine functions are estimated with full efficiency by $\mathbf{b}$. Now many signals of neurobiological interest are Fourier sines and cosines, or very nearly so, such as the square-wave used in many on/off experimental designs. Thus we expect $\mathbf{b}$ to be very nearly fully efficient. Taking this a bit further, it provides a very strong argument for *designing* the experiment so that the signal is a sine or cosine function, thereby optimising the parameter estimator. The key requirement for this simple optimality is that the hemodynamic response must be *stationary* and *linear*; it can otherwise have any shape.

For the example in Friston *et al.* (1995), the loss of efficiency in using $\mathbf{b}$, as opposed to de-convoluting the data and applying least-squares, is small for the square-wave (12%), but not for the random regressors (27±2%). This is to be expected, since the square wave is almost unchanged by filtering with the hemodynamic response (see Figure 5, page 52), whereas the random regressors are, of course, altered considerably. If the random regressors are omitted, the loss in efficiency for the square wave is reduced to 5%. If we replace the 9 regressors in the previous example by the first 9 Fourier components, that is a constant term and sine and cosine terms of periods 100, 50, 33.3 and 25 scans, then the loss in efficiency is negligible: 0.2±0.2%. This is to be expected since the regressors are now eigenvectors of $\mathbf{K}$.

In summary the periodic presentation of blocked tasks or conditions, that is so prevalent in the fMRI literature, may well be an optimal experimental design from a purely theoretical perspective.

3.3 Spatial smoothing and the theory of Gaussian fields

The question addressed (if not answered) in this section relates to the best smoothing one should apply to the data before analysis. This is a complex area which involves a number of themes, some mathematical and some neurobiological. One knows from standard filtering theory that the 'best' smoothing filter or kernel is one that matches the objects to be identified. For example if cortical activations had, in general, a spatial extent of 4mm, then a 4mm smoothing would be chosen. There is however a constraint on the lower limit of smoothing that can be used: statistical inference in SPMs generally depends on the theory of Gaussian fields and implicitly assumes that the data are good lattice representations of a smooth Gaussian field. This only holds when the voxel size is appreciably smaller than smoothness. As pointed out in Friston *et al.* (1995) this is not a fundamental limitiation because the voxel size can always be reduced at acquisition. We recommend that smoothness should be at least twice voxel size before applying any results from the theory of Gaussian fields. If the smoothness approaches voxel size then the corrected p-values based on the theory of Gaussian fields will approach those based on a Bonnferroni correction and for very small values of the smoothness estimator the Gaussian field corrections can become *more* severe.

The second issue that now arises is whether to use corrected p-values that are based

on spatial extent or the Z maxima (or equivalently the Euler characteristic). Our previous theoretical analysis (Friston *et al.* 1994b) suggested that the power of tests based on spatial extent would increase with resolution or less smoothing (conversely for tests based on the Z maxima). This effect can be seen in Figures 1b, 1c and 1d; where the same data has been analysed (using the expressions in this paper) using a Gaussian kernel of 4, 8 and 16mm FWHM respectively for smoothing. By following the fate of nearly every activation focus (e.g. the left extrastriate region at -14, -72, 20mm) one can see that as smoothing is increased the p-value based on spatial extent decreases and that based on the maximal Z value increases. This is consistent with our theoretical predictions. It would, of course be nice to combine extent and height in the estimation of the p-value and this is the subject of current work.

The final issue condidered here is the optimal smoothing to use. If activations in the brain are highly focal then the best smoothing would be a minimal one. Conversely, if brain activations are diffuse and extend over many millimeters than a high degree of smoothing should be advised. The problem is that both sorts of activations may be prevalent. Consider the right prefrontal activation in Figure 1b (4mm smoothing). If we increase the smoothing to 8mm (Figure 1c) this activation disappears. Conversely the Z value for the extrastriate region increases when we go from 4mm to 8mm smoothing. Note also that all Z values decrease when we go from 8mm to 16mm smoothing. This suggests that the right prefrontal activation is more focal than the extrastriate and that all the activations are closer to 8mm in spatial extent than 16mm. More generally these anecdotal observations suggest that activations can be expressed over different scales in the same experiment. In this instance there is no 'best' filter in any generic sense and one has to accept *a priori* that the analysis will be most sensitive to activations with the same size as the smoothing kernel. One intriguing alternative is to search over smoothing or 'scale' space and apply suitable corrections using Gaussian field theory. This is again the subject of recent work (Poline and Mazoyer, 1994ab; Siegmund and Worsley, 1995).

Conclusion

In this short paper we have introduced some substantial revisions to earlier work that addressed the problem of statistical inference in temporally correlated fMRI time-series and have discussed some important issues and constraints that arise in the theoretical framework that has been developed to facilitate these inferences.

Note added in proof

The same comments and conclusions of this paper apply to two others recently published in this journal: Friston *et al.* (1995a,b). The first paper on characterising evoked hemodynamics with fMRI (Friston *et al.*, 1995a) is a straightforward linear models approach for distinguishing between early and late evoked hemodynamic responses (Eq. (3) of Friston *et al.*, 1995a; note that k should multiply t in the exponent, not divide it). The study tests first for a combined early and late response, then for a difference between early and late responses. Applying the corrected theory of this paper, the Z statistics of the first test, shown

in Figure 2 of the original paper, should be multipled by 0.757; this changes the maximum Z to 5.29×0.757=4.00 which is not significant ($P = 0.35$). The Z statistics of the second test (Figure 3) should be multiplied by 0.790, changing the maximum Z to 5.56×0.790=4.39 which is marginally significant ($P = 0.083$). The conclusions are the same as those in the original paper, though they cannot be made with the same degree of statistical confidence: there is no evidence for a combined response, but there is some evidence for a differential response between early and late components.

The immediately succeeding paper, Friston *et al.* (1995b), applies standard multivariate statistical approaches to the same problem. These methods are valid only for uncorrelated time observations, so the authors suggest a heuristic correction based on replacing the error degrees of freedom by the "effective" degress of freedom given by Eq. (9) of Friston *et al.*, (1995). Problems of biased error variance explained above appear to be avoided because error variance does not enter explicitly into the proposed test statistic, Wilk's Λ (Friston *et al.*, 1995b, Eq. (8)). However exactly the same biases occur, due to inappropriate correction for temporal correlation. The reason is straightforward. In the case of a single effect ($h = 1$ in the notation of Friston *et al.*, 1995b), and a single voxel (so that only one time series is available and $J = 1$), the test statistic $-(r - 1/2)\log(\Lambda)$ (Friston *et al.*, 1995b, Eq. (9)) is asymptotically identical to T^2 (Friston *et al.*, 1995, eq(5)). More precisely, $-(r - 1/2)\log(\Lambda) = (r - 1/2)\log(1 + T^2/r) \approx T^2$ for large r. Since T is biased, then $-(r - 1/2)\log(\Lambda)$ is also biased, and the test as originally proposed usually gives more false positives than claimed. The same applies to the test for canonical variates (Friston *et al.*, 1995b, Eq. (12)). Correcting these tests using the methods of this paper works in the simple case of $h = J = 1$, but simulation results show that such a correction does not appear to work in the more complex case of many voxels and their principal components. This will be the subject of a future communication. For the moment, P-values derived from Eq. (9) and Eq. (12) of Friston *et al.* (1995b) should be treated with extreme caution.

References

Friston, K.J., Frith, C.D., Liddle, P.F., and Frackowiak, R.S.J. 1991. Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, **11**:690-699.

Friston, K.J., Holmes, A.P., Poline, J-B., Grasby, B.J., Williams, C.R., Frackowiak, R.S.J., and Turner, R. 1995. Analysis of fMRI time-series revisited. *NeuroImage*, **2**:45-53.

Friston, K.J., Frith, C.D., Turner, R., and Frackowiak R.S.J. 1995a. Characterizing evoked hemodynamics with fMRI. *NeuroImage*, **2:** 157-165.

Friston, K.J., Frith, C.D., Frackowiak R.S.J., and Turner, R. 1995b. Characterizing dynamic brain responses with fMRI: a multivariate approach. *NeuroImage*, **2:** 166-172.

Friston, K.J., Jezzard, P., and Turner, R. 1994a. Analysis of functional MRI time series. *Human Brain Mapping*, **1**:153-171.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., and Evans, A.C. 1994b. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**:214-220.

Poline, J-B., and Mazoyer, B.M. 1994a. Enhanced detection in brain activation maps using a multifiltering approach. *Journal of Cerebral Blood Flow and Metabolism*, **14**:639-642.

Poline, J-B., and Mazoyer, B.M. 1994b. An analysis of individual brain activation maps using hierarchical description and mutiscale detection. *IEEE Transactions on Medical Imaging*, **13**:702-710.

Satterthwaite, F.E. 1946. An approximate distribution of estimates of variance components. *Biometrics*, **2**:110-114.

Seber, G.A.F. 1977. *Linear Regression Analysis*. Wiley, New York.

Siegmund, D.O., and Worsley, K.J. 1994. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, in press.

Talairach, J., and Tournoux, P. 1988. *Co-planar stereotactic atlas of the human brain: 3-Dimensional proportional system: an approach to cerebral imaging*. Georg Thieme Verlag, Stuttgart, New York.

Watson, G.S. 1955. Serial correlation in regression analysis. I. *Biometrika*, **42**:327-341.

Worsley, K.J., Evans, A.C., Marrett, S., and Neelin, P. 1992. A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**:900-918.

Worsley K.J., Evans, A.C., Marrett, S. & Neelin, P. 1993. Detecting and estimating the regions of activation in CBF activation studies in human brain. In *Quantification of Brain Function: Tracer kinetics and image analysis in brain PET* (K. Uemura, N. Lassen, T. Jones, and I. Kanno, Eds.), pp. 535-548. Excerpta Medica, Tokyo.

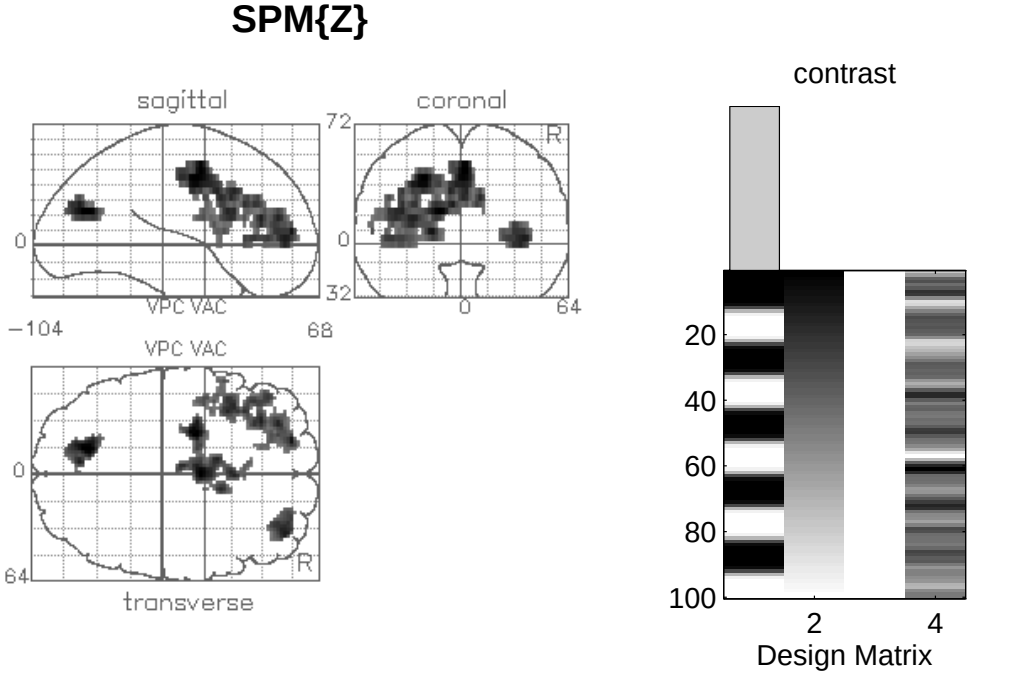
Worsley, K.J., Poline, J-B., Vandal, A.C., and Friston, K.J. 1995. Tests for distributed, non-focal brain activations. *NeuroImage*, submitted.

Acknowledgements

K.J.W. is supported by the Natural Sciences and Engineering Research Council of Canada, and the Fonds pour la Formation des Chercheurs et l'Aide à la Recherche de Québec. K.J.F. is funded by the Wellcome Trust.

Figure Legends

Figure 1. Upper left panel: Statistical parametric map of the T statistic (after transformation to a Z value) reflecting the significance of a compound of effects. The SPM is displayed in a standard format as a maximum intensity projection viewed from the back, the right hand side and the top of the brain. The anatomical space corresponds to the atlas of Talairach and Tournoux (1988). The SPM has been thresholded at 2.33 and the color scale is arbitrary. Upper right panel: Top - the contrast used for this SPM. The contrast is displayed above the appropriate effects (columns of the design matrix). Lower panel: Table of regional effects (activations or regional differences) characterized by the volume of each region (k), its significance based on patial extent $P(n_{\max} > k)$, the highest Z value (Z), its significance based on $P(Z_{\max} > u)$ and the location of this primary maximum. We have also included up to three secondary maxima for each region and their associated significance based on the corrected and uncorrected p-value.



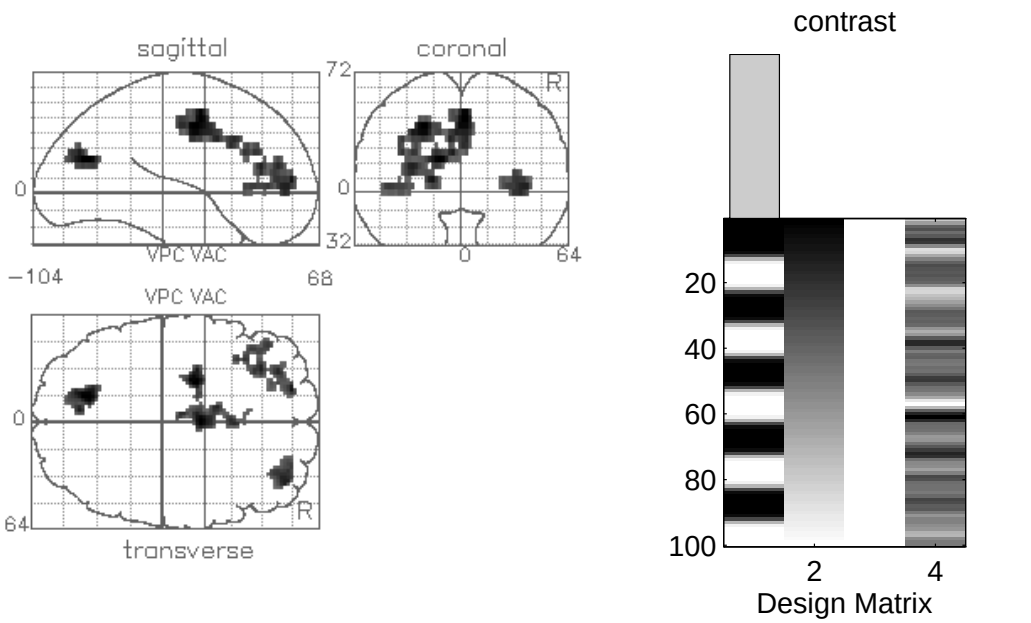
original estimators − 4mm smoothing

region	size {k}	P(n_{max} > k)	Z	P(Z_{max} > u) (Uncorrected)	{x,y,z mm}
1	**154**	**0.013**	**4.76**	**0.034** (0.000)	**−14 −72 20**
			4.12	0.438 (0.000)	−6 −76 24
2	**470**	**0.000**	**4.72**	**0.040** (0.000)	**2 −4 44**
			4.67	0.049 (0.000)	−24 −6 40
			3.98	0.711 (0.000)	−8 6 32
3	**596**	**0.000**	**4.31**	**0.214** (0.000)	**−18 50 8**
			4.07	0.525 (0.000)	−30 32 16
			3.95	0.800 (0.000)	−30 42 16
4	**121**	**0.044**	**4.09**	**0.484** (0.000)	**36 46 4**

Threshold = 2.33; Volume [S] = 40116 voxels; df = 41

FWHM = [8.3 8.1 8.9] mm (i.e. 1071 RESELS)

Figure 1a. Using the original estimators for the error variance and effective degrees of freedom as described in Friston *et al.* (1995).



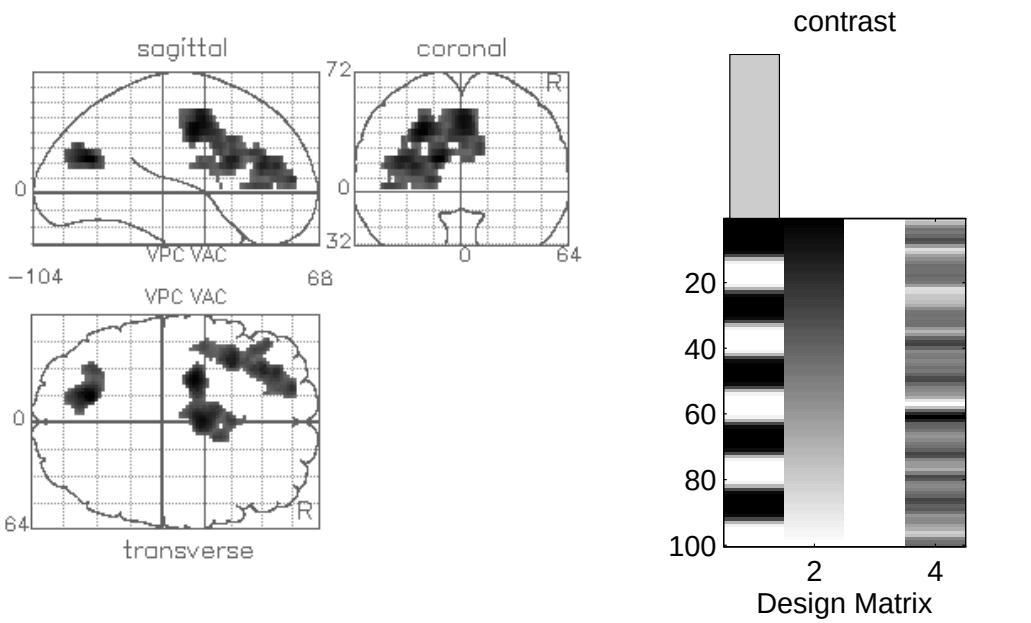
correct estimators − 4mm smoothing

region	size {k}	P(n max > k)	Z	P(Z max > u)	(Uncorrected)	{x,y,z mm}
1	**118**	**0.050**	**4.18**	**0.349**	(0.000)	**−14 −72 20**
			3.60	2.496	(0.000)	−6 −76 24
2	**217**	**0.002**	**4.15**	**0.390**	(0.000)	**2 −4 44**
			3.48	3.603	(0.000)	−8 6 32
			3.46	3.752	(0.000)	2 16 32
3	**101**	**0.095**	**4.10**	**0.462**	(0.000)	**−24 −6 40**
			2.99	12.732	(0.001)	−14 −6 36
4	**243**	**0.001**	**3.77**	**1.444**	(0.000)	**−18 50 8**
			3.55	2.865	(0.000)	−30 32 16
			3.44	3.937	(0.000)	−30 42 16
5	**100**	**0.099**	**3.57**	**2.692**	(0.000)	**36 46 4**

Threshold = 2.33; Volume [S] = 40116 voxels; df = 39

FWHM = [8.3 8.1 8.9] mm (i.e. 1069 RESELS)

Figure 1b. As for Figure 1a but using the correct estimators described herein.



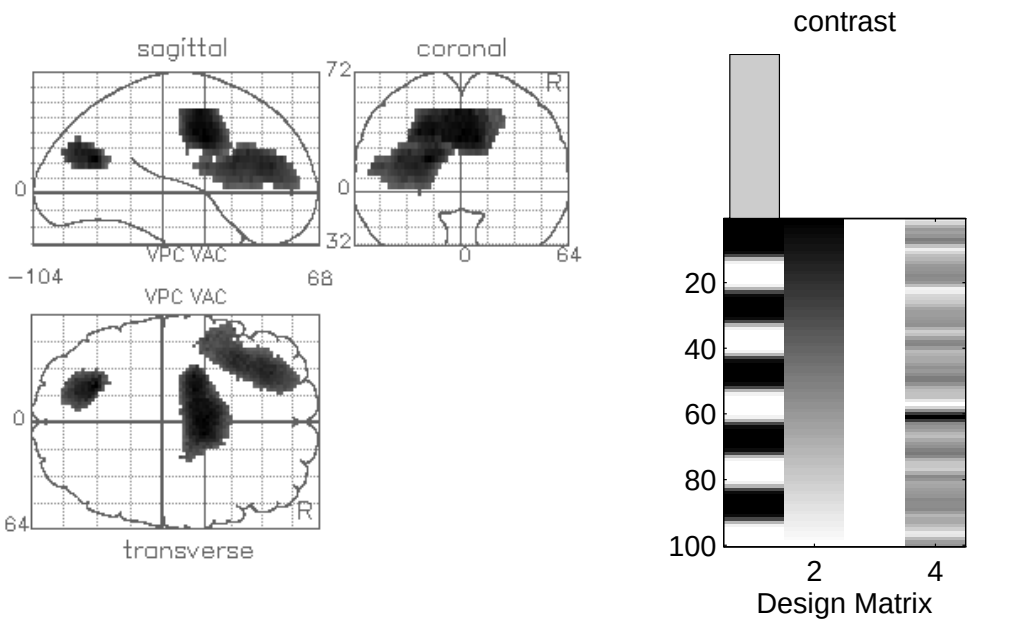
8mm smoothing

region	size {k}	P(n max > k)	Z	P(Z max > u)	(Uncorrected)	{x,y,z mm}
1	**232**	**0.106**	**4.22**	**0.097**	(0.000)	**−14 −70 20**
			3.15	2.742	(0.001)	−30 −68 20
2	**654**	**0.001**	**4.08**	**0.160**	(0.000)	**−24 −8 36**
			3.91	0.292	(0.000)	0 −2 44
			3.15	2.737	(0.001)	10 4 28
3	**575**	**0.002**	**3.68**	**0.607**	(0.000)	**−36 16 20**
			3.59	0.812	(0.000)	−28 34 16
			3.35	1.629	(0.000)	−18 50 8

Threshold = 2.33; Volume [S] = 41178 voxels; df = 39

FWHM = [12.6 12.6 12.4] mm (i.e. 337 RESELS)

Figure 1c. As for Figure 1b but increasing the spatial smoothing of the data from 4mm to 8mm.



16mm smoothing

region	size {k}	P(n max > k)	Z	P(Z max > u)	(Uncorrected)	{x,y,z mm}
1	**316**	**0.274**	**3.97**	**0.080**	(0.000)	**−20 −66 20**
			3.53	0.332	(0.000)	−18 −74 24
			3.14	0.964	(0.001)	−24 −78 28
2	**1345**	**0.004**	**3.93**	**0.092**	(0.000)	**0 −2 36**
			3.90	0.102	(0.000)	8 −2 32
			3.80	0.141	(0.000)	−18 −8 36
3	**1081**	**0.010**	**3.42**	**0.457**	(0.000)	**−28 38 16**
			3.40	0.480	(0.000)	−38 18 20
			3.30	0.639	(0.000)	−32 24 24

Threshold = 2.33; Volume [S] = 45485 voxels; df = 39

FWHM = [18.6 18.9 18.1] mm (i.e. 115 RESELS)

Figure 1d. As for Figure 1c but increasing the spatial smoothing of the data from 8mm to 16mm.