

Parametric inference for functional information mapping

Dennis Leech, Robert Leech and Anna Simmonds

No 899

**WARWICK ECONOMIC RESEARCH PAPERS**

**DEPARTMENT OF ECONOMICS**

THE UNIVERSITY OF  
**WARWICK**

# Parametric inference for functional information mapping

Robert Leech

Division of Neuroscience and Mental Health  
Imperial College London

Anna Simmonds

MRC Clinical Sciences Center, Imperial College London

Dennis Leech

Department of Economics  
University of Warwick

## 1 abstract

An increasing trend in functional MRI experiments involves discriminating between experimental conditions on the basis of fine-grained spatial patterns extending across many voxels. Typically, these approaches have used randomized resampling to derive inferences. Here, we introduce an analytical method for drawing inferences from multivoxel patterns. This approach extends the general linear model to the multivoxel case resulting in a variant of the Mahalanobis distance statistic which can be evaluated on the  $\chi^2$  distribution. We apply this parametric inference to a single-subject fMRI dataset and consider how the approach is both computationally more efficient and more sensitive than resampling inference.

## 2 Introduction

A central aim of functional neuroimaging research is to locate brain regions that activate for a given task. For example, left superior temporal regions show preferential activation for intelligible speech over other types of sounds and as such are likely to play an important role in early spoken language processing e.g., (Scott et al. 2000). The traditional approach to finding areas of activation is, at its most basic, massively univariate; treating each of the tens of thousands of voxels in the brain as an independent experiment and fitting a separate statistical model to the time-course at each voxel. Normally, the data is spatially smoothed prior to statistical analysis, introducing some restricted combination of information across voxels and typically leading to a substantial increase in signal-to-noise.

In tandem with the search for activation differences - and gaining momentum in recent years - are approaches that search for information across multiple voxels. Analyzing for information involves assessing whether there is sufficient information to distinguish between two conditions across all the voxels in a given region (Haxby et al. 2001) or even across all the voxels in the brain (Hanson and Halchenko 2008), irrespective of the level of activation of a given voxel. By taking advantage of the fine spatial patterns across voxels, information-based approaches have become highly adept at detecting subtle differences in signal between conditions (Kay et al. 2008, Mitchell et al. 2008).

Kriegeskorte and colleagues (Kriegeskorte et al. 2006, Kriegeskorte and Bandettini 2007) proposed an information-based approach that attempts to reconcile the sensitivity of multivoxel analyses with the spatial localization of activation-based analyses. With this approach, a spherical searchlight was passed around the brain centered on each voxel, and a multivoxel distance statistic was calculated across all voxels within the searchlight. The Mahalanobis distance statistic was shown to be substantially more sensitive at detecting signals than standard univariate t-statistics or Euclidean distance measures. Variants of this statistic have been used successfully in several fMRI experiments within the visual domain (Kriegeskorte et al. 2006; 2007, Serences and Boynton 2007a;b, Stokes et al. 2009).

The Mahalanobis distance (Mahalanobis 1936) is a multivoxel similarity measurement that unlike Euclidean distance is scale invariant and controls for covariance across the datapoints. In fMRI datasets, the Mahalanobis distance can calculate the similarity between different experiment conditions, e.g., how dissimilar are visual and auditory processing for a given set of voxels. This statistic controls for error covariance across voxels; spatially correlated error is expected in fMRI datasets given spatial patterns of sources of noise affecting the BOLD signal, such as vasculature, movement artifacts etc. To calculate the Mahalanobis distance, Kriegeskorte et al. (2006) fitted a separate GLM analysis to the unsmoothed functional data for each voxel. The Mahalanobis distance was then calculated using the resulting vectors of beta weights and the residual error covariance matrix across voxels

$$\Delta = (\hat{\beta}^2 - \hat{\beta}^1)' \Sigma^{-1} (\hat{\beta}^2 - \hat{\beta}^1) \quad (1)$$

where  $\hat{\beta}^1$  and  $\hat{\beta}^2$  are vectors of estimated regression coefficients associated with two types of stimuli and  $\hat{\Sigma}$  is the error covariance matrix.

In previous information-based analyses, statistical inference has typically been achieved with resampling methods. In previous studies (Kriegeskorte et al. 2006; 2007), the labels associated with individual timepoints were randomly permuted 1000 times to generate a null-distribution of Mahalanobis distances for each voxel to compare with the unpermuted Mahalanobis distance. However, parametric methods have more power than resampling methods (assuming the required distributional assumptions are met for the parametric

test). Most importantly, for typical fMRI datasets, parametric techniques can be several orders of magnitude more computationally efficient than resampling methods. As such parametric inference for across-voxel Mahalanobis distance would be a useful tool for analyzing for information rather than activation differences in fMRI data.

In this paper we demonstrate how parametric inferences can be drawn based on the Mahalanobis distance. We set out how the Mahalanobis distance relates to the general linear model and how it can be evaluated analytically on the  $\chi^2$  distribution. We then apply the parametric Mahalanobis analysis to an example fMRI data. Finally, we compare parametric inference based on the Mahalanobis statistics with inference derived using resampling.

### 3 Theory

#### 3.1 The statistical model

We start with the GLM as applied to fMRI datasets (Friston et al. 1994). We assume the response by each voxel can be modelled by a regression model written as follows.

$$y_{it} = \sum_{h=1}^k \beta_{hi} x_{ht} + u_{it} \quad (2)$$

where  $i$  is the voxel subscript,  $i = 1, \dots, n$ , and  $t$  is the time subscript,  $t = 1, \dots, T$ .  $\beta_{hi}$  represents the response of voxel  $i$  to the stimulus measured by regressor  $h$ .

This can be written using matrix algebra as,

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{u}_i \quad (3)$$

where  $\mathbf{y}_i$  is the  $T$  element observation vector for voxel  $i$ ,  $\mathbf{X}$  the  $T \times k$  input matrix,  $\mathbf{u}_i$  the  $T$  element error vector, and  $\boldsymbol{\beta}_i$  the  $k$  element coefficient vector. There is a different regression model for each voxel but all models have a common regressor matrix.

The error for each voxel equation is assumed to have zero mean, be serially independent and homoscedastic, but correlated across equations.

That is,  $E(\mathbf{u}_i) = \mathbf{0}$  for all  $i$  and  $E(\mathbf{u}_i \mathbf{u}_j') = \sigma_{ij} \mathbf{I}_T$  for all  $i$  and  $j$ .

This can be written more compactly as:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{u}, E(\mathbf{u}) = \mathbf{0}, E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Omega}, \quad (4)$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}, \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} \mathbf{X} & 0 & \dots & 0 \\ 0 & \mathbf{X} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X} \end{pmatrix} = \mathbf{I}_n \otimes \mathbf{X},$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11}\mathbf{I}_T & \sigma_{12}\mathbf{I}_T & \dots & \sigma_{1n}\mathbf{I}_T \\ \sigma_{12}\mathbf{I}_T & \sigma_{22}\mathbf{I}_T & \dots & \sigma_{2n}\mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n}\mathbf{I}_T & \sigma_{2n}\mathbf{I}_T & \dots & \sigma_{nn}\mathbf{I}_T \end{pmatrix} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_{nn} \end{pmatrix}.$$

It is well known (Zellner 1962) that, when there is a common set of regressors, the correlation of the errors in different equations can be ignored in coefficient estimation and least squares applied separately to each equation. That is, the efficient estimator of the complete system is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}, \text{ and therefore } \hat{\boldsymbol{\beta}}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i. \quad (5)$$

The covariance matrix of  $\hat{\boldsymbol{\beta}}$  is

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}. \quad (6)$$

Noting that

$$\mathbf{Z}'\mathbf{Z} = [\mathbf{I}_n \otimes \mathbf{X}'][\mathbf{I}_n \otimes \mathbf{X}] = \mathbf{I}_n \otimes \mathbf{X}'\mathbf{X}, \text{ and hence that } (\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{I}_n \otimes (\mathbf{X}'\mathbf{X})^{-1},$$

and that

$$\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z} = [\mathbf{I}_n \otimes \mathbf{X}'][\boldsymbol{\Sigma} \otimes \mathbf{I}_T][\mathbf{I}_n \otimes \mathbf{X}] = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X}),$$

we can rewrite (6) as

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = [\mathbf{I}_n \otimes (\mathbf{X}'\mathbf{X})^{-1}][\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})][\mathbf{I}_n \otimes (\mathbf{X}'\mathbf{X})^{-1}] = \boldsymbol{\Sigma} \otimes \mathbf{W}, \quad (7)$$

where  $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{W}$ . From this we obtain the covariance between any pair of estimated coefficients as

$$\text{Cov}(\hat{\beta}_{gi}, \hat{\beta}_{hj}) = \sigma_{ij}w_{gh} \quad (8)$$

for all voxels  $i, j = 1, \dots, n$ , and all regression coefficients  $g, h = 1, \dots, k$ .

## 3.2 Inference

We are interested in testing a set of linear restrictions on coefficients across the separate equations for each voxel. In general there are  $g$  homogeneous restrictions written as  $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$  where  $\mathbf{R}$  is a  $g \times nk$  matrix of given constants of rank  $g$ . In the next two subsections we use this framework to derive tests for the following restrictions: (1) that a coefficient of a particular stimulus is equal to zero in all voxels, and (2) that two effects are equal in all voxels.

If the equation errors  $u_{it}$  are gaussian then the efficient estimator of  $\boldsymbol{\beta}$  given in equation (5) is normally distributed and the well known (see for example Rao and Rao 1973) inferential procedures for generalised linear regression models are available based on the covariance matrix in equation (6), (7) or (8). This only holds exactly if  $\boldsymbol{\Sigma}$  is known and does not have to be estimated. The alternative approach, a Wald test adopted here, is to use large sample procedures based on the asymptotic normality of  $\hat{\boldsymbol{\beta}}$  (which does not depend on any particular assumption about the form of the distribution of the random errors) and the use of a consistent estimator of  $\boldsymbol{\Sigma}$ . (See for example Greene (2003).)

The null hypothesis of interest is written as

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}. \quad (9)$$

Letting  $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\delta}$  this can be written as  $H_0 : \boldsymbol{\delta} = \mathbf{0}$ . Now let  $\mathbf{R}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\delta}}$ . Then  $\hat{\boldsymbol{\delta}}$  is asymptotically normally distributed with expectation  $E(\hat{\boldsymbol{\delta}}) = \mathbf{0}$  and covariance matrix  $V(\hat{\boldsymbol{\delta}}) = \mathbf{R}V(\hat{\boldsymbol{\beta}})\mathbf{R}'$ . We call the test statistic  $\Delta$ , given by

$$\Delta = \hat{\boldsymbol{\delta}}'[\mathbf{R}V(\hat{\boldsymbol{\beta}})\mathbf{R}']^{-1}\hat{\boldsymbol{\delta}} \quad (10)$$

which has an asymptotic  $\chi^2$  distribution with  $g$  degrees of freedom if  $H_0$  is true.

### Testing a single coefficient value

Let us consider testing the null hypothesis that one particular coefficient is zero in all voxel equations. We assume that the coefficient is indexed number 1 in each voxel equation.

The null hypothesis is:

$$H_0 : \beta_{1i} = 0 \text{ for } i = 1, \dots, n, \quad (11)$$

which can be written in vector notation as

$$H_0 : \boldsymbol{\beta}^1 = \mathbf{0}$$

where

$$\boldsymbol{\beta}^1 = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1n} \end{pmatrix}.$$

These coefficients can also be written as linear functions of the complete coefficient vector as,

$$\boldsymbol{\beta}^1 = [\mathbf{I}_n \otimes \mathbf{e}'_1] \boldsymbol{\beta}$$

where  $\mathbf{e}'_1$  is a  $k$ -element vector whose first element is 1 and all others zeros: the first row of  $\mathbf{I}_k$ . Here we can write  $\mathbf{R} = \mathbf{I}_n \otimes \mathbf{e}'_1$ ,  $\boldsymbol{\delta} = \boldsymbol{\beta}^1$ .

We can thus obtain the covariance matrix of  $\hat{\boldsymbol{\beta}}^1$  from that of  $\hat{\boldsymbol{\beta}}$  and use it with the standard theory of testing linear restrictions. We know that

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\beta}}^1) &= [\mathbf{I}_n \otimes \mathbf{e}'_1] \mathbf{V}(\hat{\boldsymbol{\beta}}) [\mathbf{I}_n \otimes \mathbf{e}_1] = [\mathbf{I}_n \otimes \mathbf{e}'_1] [\boldsymbol{\Sigma} \otimes \mathbf{W}] [\mathbf{I}_n \otimes \mathbf{e}_1] \text{ (from (7))} \\ &= \boldsymbol{\Sigma} \otimes \mathbf{e}'_1 \mathbf{W} \mathbf{e}_1 = w_{11} \boldsymbol{\Sigma}. \end{aligned} \tag{12}$$

This assumes that the elements of  $\boldsymbol{\Sigma}$  are known. They can be estimated consistently from the residuals, using,

$$\hat{\sigma}_{ij} = (\mathbf{y}_i - \mathbf{X} \hat{\boldsymbol{\beta}}_i)' (\mathbf{y}_j - \mathbf{X} \hat{\boldsymbol{\beta}}_j) / T \tag{13}$$

We can therefore test the null hypothesis using the likelihood ratio statistic,

$$\Delta_1 = \hat{\boldsymbol{\beta}}^1' \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\beta}}^1 / w_{11} \tag{14}$$

which, for large  $T$ , is distributed as  $\chi^2$  with  $n$  degrees of freedom on  $H_0$ . We could also have obtained the covariance matrix of  $\hat{\boldsymbol{\beta}}^1$  directly from expression (8).

### Testing coefficient equality

A common hypothesis in fMRI data is to test if two coefficients are equal. This can be extended to the multivoxel case by testing the restriction that two coefficients are equal in every equation. That is that the effects of two simultaneous stimuli are equal in each voxel regardless of their signs or magnitudes. Thus, assuming the two stimulæ are regressors  $x_1$

and  $x_2$ , we test a set of linear restrictions in the null hypothesis

$$H_0 : \beta_{1i} = \beta_{2i} \text{ for } i = 1, 2, \dots, n. \quad (15)$$

We can write these restrictions, in vector notation, as:

$$H_0 : \boldsymbol{\beta}^1 - \boldsymbol{\beta}^2 = \mathbf{0}$$

where  $\boldsymbol{\beta}^2 = [\mathbf{I}_n \otimes \mathbf{e}'_2]\boldsymbol{\beta}$  and  $\mathbf{e}_2$  is defined, analogously to  $\mathbf{e}_1$ , as a vector of all zeros except for element 2 which equals 1.

$$\text{Letting } \delta_i = \beta_{1i} - \beta_{2i} \text{ for each } i \text{ and } \boldsymbol{\delta} = \boldsymbol{\beta}^1 - \boldsymbol{\beta}^2 = \begin{pmatrix} \beta_{11} - \beta_{21} \\ \beta_{12} - \beta_{22} \\ \vdots \\ \beta_{1n} - \beta_{2n} \end{pmatrix}, \text{ we can write}$$

$$\boldsymbol{\delta} = [\mathbf{I}_n \otimes (\mathbf{e}_1 - \mathbf{e}_2)']\boldsymbol{\beta}.$$

In this case we have  $\mathbf{R} = \mathbf{I}_n \otimes (\mathbf{e}_1 - \mathbf{e}_2)'$ .

Therefore the covariance matrix of  $\hat{\boldsymbol{\delta}}$  is

$$\begin{aligned} V(\hat{\boldsymbol{\delta}}) &= [\mathbf{I}_n \otimes (\mathbf{e}_1 - \mathbf{e}_2)'] [\boldsymbol{\Sigma} \otimes \mathbf{W}] [\mathbf{I}_n \otimes (\mathbf{e}_1 - \mathbf{e}_2)] \\ &= \boldsymbol{\Sigma} \otimes (\mathbf{e}_1 - \mathbf{e}_2)' \mathbf{W} (\mathbf{e}_1 - \mathbf{e}_2) = (w_{11} + w_{22} - 2w_{12}) \boldsymbol{\Sigma} \end{aligned} \quad (16)$$

The test statistic is then

$$\Delta_2 = \hat{\boldsymbol{\delta}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\delta}} / (w_{11} + w_{22} - 2w_{12}) \quad (17)$$

with asymptotic null distribution  $\chi^2$  with  $n$  degrees of freedom.

The statistic in equation (17) is the Mahalanobis distance measure between the vectors  $\hat{\boldsymbol{\beta}}^1 - \hat{\boldsymbol{\beta}}^2$  (Mahalanobis 1936).

### Computation by data transformation and reparameterisation

$\Delta_2$  in equation (17) can be computed easily using a simple data transformation before fitting the regression and then using  $\Delta_1$  to test one coefficient in the transformed regression.

Rewriting equation (2) for voxel  $i$ ,

$$y_{it} = \sum_{h=1}^k \beta_{hi} x_{ht} + u_{it}$$



as

$$y_{it} = \beta_{1i}x_{1t} + \beta_{2i}x_{2t} + \dots,$$

to highlight the two coefficients of interest, then adding and subtracting the same term on the RHS enables us to derive a new equation with a transformed coefficient and a transformed variable. Thus,

$$y_{it} = \beta_{1i}x_{1t} + \beta_{2i}x_{2t} - \beta_{2i}x_{1t} + \beta_{2i}x_{1t} + \dots = (\beta_{1i} - \beta_{2i})x_{1t} + \beta_{2i}(x_{1t} + x_{2t}) + \dots$$

Therefore the regression equation can be written:

$$y_{it} = \delta_i x_{1t} + \beta_{2i} s_t + \dots$$

and the coefficient on  $x_{1t}$  becomes  $\delta_i$  defined above as a result of the simple data transformation of replacing  $x_{2t}$  by a new variable  $s_t$  defined as the sum of  $x_{1t}$  and  $x_{2t}$ .

This is computationally very simple but does require running two regressions if we wish to test both sets of restrictions (11) and (15).

## 4 Example: single subject fMRI data

We applied this method to fMRI data acquired from a single subject. This allowed comparisons of the Mahalanobis distance statistic (i.e., 17) using different inference methods (both the analytic method described and using resampling).

### 4.1 fMRI experiment

The fMRI data was acquired from a pilot subject in an experiment investigating bilingual speech production. The subject was a 32-year old neurologically healthy, right-handed male and gave informed consent according to local ethics procedures. The subject viewed color and black and white pictures and was asked to describe the pictures in either his native (English) language or proficient non-native (Spanish) language. A third condition involved a picture prompting the subject to silently move his tongue. The rest baseline involved observing a color picture while remaining silent and still.

The fMRI data was acquired with a Philips Intera 3.0 Tesla scanner using dual gradients and a phased array head coil. Images were obtained using a T2\*-weighted, gradient-echo, echoplanar imaging (EPI) sequence with whole-brain coverage (TE, 30 ms; flip angle, 90°). Thirty-two axial slices were acquired with a slice thickness of 3.25 mm and an interslice gap of 0.75 mm (resolution 2.19 × 2.19 × 4.0 mm; field of view, 280 × 224 × 128 mm). Quadratic shim gradients corrected for inhomogeneities in the magnetic field. E-Prime

software (Psychology Software Tools) was used to present stimuli using the IFIS-SA system (In Vivo Corporation).

A sparse fMRI design was used to minimize movement- and respiratory-related artifact commonly associated with speech production. Each TR lasted 10 seconds including 2 seconds of image acquisition followed by a seven seconds of speech production and one second of rest. The subject was scanned twice with 72 TRs in each 12 minute run.

## 4.2 fMRI analysis

All images were registered to the middle functional image of the first run using 3dvolreg (AFNI toolbox, Cox et al. 1996). A single time series was created by concatenating the images from the first and second runs. No other preprocessing was carried out.

We restricted all our analysis to a broad anatomically defined region of interest, to reduce the computational requirements of resampling methods used to validate the parametric evaluation of the Mahalanobis distance. The region of interest was defined in MNI152 space using the probabilistic Harvard-Oxford anatomical atlas from fslview. The region of interest contained the inferior frontal gyrus (pars opercularis and pars triangularis); these regions have been reported in previous studies comparing speech production in different languages (e.g., Hernandez et al. 2001). To allow a spread of voxel activation patterns, the region of interest also included the superior frontal gyrus, less likely to be involved in bilingualism based on previous imaging studies. The region of interest was warped into the subject’s functional space and registered to the subject’s functional scans using FLIRT (Jenkinson and Smith 2001).

A general linear model was fitted to the unsmoothed time series for each voxel in matlab (Mathworks, 2007b). The GLM estimated nine parameters for each voxel: three parameters of interest (speaking in native language, non-native language and tongue movement), as well as three parameters for each run to account for mean, linear and quadratic trends. To account for the hemodynamic delay, the labels for each image (i.e., native language, non-native language, tongue movement and rest) were temporally shifted one timepoint (this is equivalent to convolving the dataset with the canonical hemodynamic function in AFNI).

### 4.2.1 Multivoxel parametric inference

Subsequent to fitting the GLM for each voxel, the second step involved passing a spherical ”searchlight” over all voxels in the regions of interest. The spherical ”searchlight” had a radius of 2 voxels, with a maximum extent of 33 voxels. Within each sphere, the Maha-

lanobis distance between native and non-native speech production was calculated using equation (17) which was evaluated with reference to the  $\chi^2$  distribution. The resulting p-value were then assigned to the center voxel.

#### 4.2.2 Multivoxel nonparametric inference

We also assessed significance of the statistic in equation (17) using nonparametric randomized resampling methods. A null-distribution of test-statistics was calculated for each voxel by shuffling the labels (native, non-native, tongue) of each non-baseline block a 1000 times. Each of these 1000 randomized time courses was then analyzed as above. The distribution of randomized test statistics was then sorted and the rank order (divided by 1000) of the unpermuted test statistic provided a p-value.

#### 4.2.3 Univoxel parametric inference

We also aimed to replicate the findings of (Kriegeskorte et al. 2006) regarding the power of the information-based approach over more traditional activation-based analyses. To this end, we reran the GLM for each voxel, but following spatial averaging across all voxels within the spherical searchlight (to increase signal to noise). We subsequently calculated t-statistics at each voxel for the contrast comparing native language with non-native language.

### 4.3 Results

#### 4.3.1 Validating the parametric measure

To validate that the multivoxel parametric inference is comparable to the resampling approach proposed by (Kriegeskorte et al. 2006), we compared the p-values derived from equation (17) with p-values derived from resampling (following the method employed Kriegeskorte et al). Figure 1 compares the two inference methods at each voxel across the region of interest, indicating a strong correlation between the resulting p-values from the two methods (Pearson's  $r=0.90$ , Spearman's  $\rho = 0.96$ ). Figure 1 also indicates that the parametric inference method produces more significant p-values than the randomization method for every voxel in the region of interest. Thus, the two inference measures appear to agree on the spatial characteristics of the underlying signal, but the parametric approach appears to be more sensitive.

To illustrate this increased sensitivity, the parametric approach although more sensitive than the randomization approach, produces highly similar spatial patterns of voxels. Figure 2 contrasts the statistical maps derived using both inference methods. The randomization-derived map (in red) shows voxels thresholded at  $p<0.05$  (uncorrected for

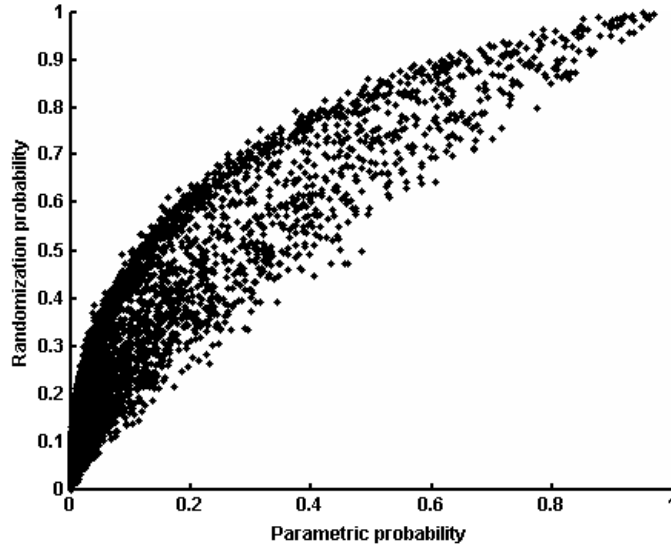


Figure 1: Scatter plot of parametric- and randomization-derived p-values for each voxel in the region of interest.

multiple comparisons). This map is nearly identical to the parametric-derived map that is thresholded at the far more stringent  $p < 0.003$ .

#### 4.3.2 Multivoxel parametric inference and univoxel t-statistic

Figure 3 demonstrates the substantial increase in signal detection for the multivoxel parametric inference over the more traditional univoxel t-statistics map. Figure 1A and Figure 1B both display voxels thresholded at the same p-value - determined by applying false discovery rate (FDR,  $q < 0.05$  Genovese et al. 2002) to the p-value map derived from the univoxel t-statistics. In total, there were 10226 voxels in the region of interest, of these 6367 voxels were superthreshold for the multivoxel parametric inference. For the equivalent univoxel t-statistics, only 1756 voxels survived FDR correction. There were 1515 (22.9%) superthreshold voxels common to both analyses, with 4852 (73.4%) unique to the multivoxel parametric analysis, and 241 (3.65%) unique to the t-statistic analysis. This pattern of results is qualitatively similar to Kriegeskorte et al. (2006), who found the Mahalanobis distance statistic revealed more significant differences across local neighborhoods of voxels than the smoothed univoxel t-statistic.

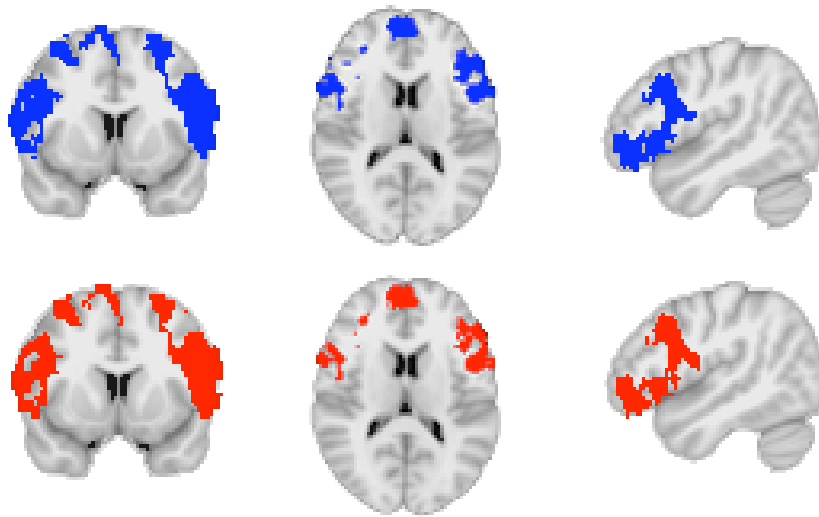


Figure 2: Comparison of parametric (in blue) and randomization (in red) derived p-values. The parametric map is thresholded at  $p < 0.003$ ; the randomization map is thresholded at  $p < 0.05$ . The statistical maps have been projected onto the MNI152  $2\text{mm} \times 2\text{mm} \times 2\text{mm}$  average brain

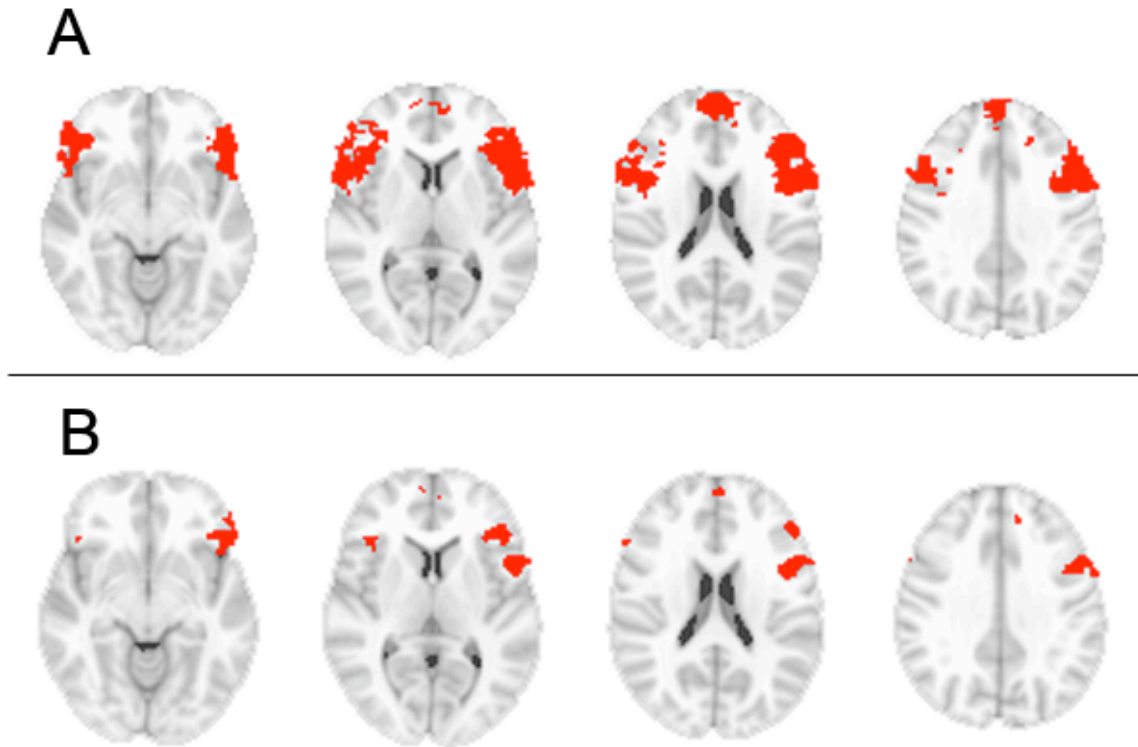


Figure 3: A: superthreshold voxels determined using multivoxel parametric inference calculated across 2 voxel radius spherical searchlights, on unsmoothed data, thresholded at  $p < 0.0086$ ; B: Superthreshold voxels associated with smoothed data and a univoxel t-test, thresholded at the same  $p < 0.0086$  (fdr corrected,  $q < 0.05$ ).

## 5 Discussion

In this paper, we have demonstrated how multivoxel information-based inference can be accomplished parametrically by extending the general linear model. We have shown how an instantiation of the multivoxel Mahalanobis distance statistic proposed for fMRI by Kriegeskorte et al. (2006) can be evaluated with the  $\chi^2$  distribution. Parametric inference is a more powerful and more computationally efficient alternative to the resampling inference methods used previously. This is a potentially highly efficient and powerful approach to detect differences in blood oxygenation levels with high spatial frequency of the kind that are likely to occur in many cognitive and perceptual tasks.

The proposed approach has two principal benefits over randomization resampling inference methods used for multivariate statistics: 1) efficiency; and 2) sensitivity. Efficiency: the parametric inference involves fitting one model to each voxel, rather than fitting >1000 models to generate a null distribution. As such, this method is at least three orders of magnitude more computationally efficient than randomization inference, making information analyses more feasible for many fMRI experiments, including analyses involving the whole brain. Sensitivity: as with parametric approaches in general when distributional assumptions are met, we find that the parametric inference method detects significant voxels at a much lower thresholds when compared to randomization-based inference. Furthermore, with the example fMRI dataset analysed here we confirm previous studies, suggesting that the Mahalanobis distance method has considerably greater sensitivity to detect a signal than univoxel analyses following spatial smoothing.

The proposed multivoxel parametric inference can also be adapted for group fMRI analyses. The Mahalanobis distance can be calculated for each subject and then the statistical maps can be warped into standard space and compared across subjects. Since independent  $\chi^2$  statistics are additive, the statistical maps of  $\chi^2$  values can be combined across subjects to create a group statistic that can be evaluated on the  $\chi^2$  distribution with the summed degrees of freedom. This is a parametric version of the randomization approach taken by (Kriegeskorte et al. 2007).

One of the central attractions of the proposed inference method is its relationship to the class of general linear models commonly used for univoxel modeling of fMRI data (Friston et al. 1994). As such, the multivoxel parametric inference proposed here can be applied to any fMRI datasets that can be analyzed using the GLM. Furthermore, many of the advantages of the GLM can also be applied with multivoxel parametric inference. For example, pre-whitening of the data (Woolrich et al. 2001) to account for autocorrelation of errors in the fMRI time series can be applied as with single voxel analysis.

## References

- Cox, R. et al. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3):162–173.
- Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., Frackowiak, R., et al. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2(4):189–210.
- Genovese, C., Lazar, N., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878.
- Greene, W. (2003). *Econometric analysis*. Prentice Hall Upper Saddle River, NJ.
- Hanson, S. and Halchenko, Y. (2008). Brain reading using full brain support vector machines for object recognition: There is no "face" identification area. *Neural Computation*, 20:486–503.
- Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430.
- Hernandez, A., Dapretto, M., Mazziotta, J., and Bookheimer, S. (2001). Language switching and language representation in Spanish-English bilinguals: an fMRI study. *NeuroImage*, 14:510–520.
- Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156.
- Kay, K., Naselaris, T., Prenger, R., and Gallant, J. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185):352–355.
- Kriegeskorte, N. and Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4):649–662.
- Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(51):20600.
- Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10):3863–3868.
- Mahalanobis, P. (1936). On the generalized distance in statistics. In *Proceedings of the National Institute of Science of India*, volume 12, pages 49–55.



- Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., and Just, M. (2008). Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191.
- Rao, C. and Rao, C. (1973). Linear statistical inference and its applications.
- Scott, S., Blank, C., Rosen, S., and Wise, R. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12):2400.
- Serences, J. and Boynton, G. (2007a). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, 55(2):301–312.
- Serences, J. and Boynton, G. (2007b). The representation of behavioral choice for motion in human visual cortex. *Journal of Neuroscience*, 27(47):12893.
- Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *Journal of Neuroscience*, 29(5):1565.
- Woolrich, M., Ripley, B., Brady, M., and Smith, S. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage*, 14(6):1370–1386.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, pages 348–368.