Testing for spatial heterogeneity in functional MRI using the multivariate general linear model

Robert Leech and Dennis Leech

No 938

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF
WARWICK

Testing for spatial heterogeneity in functional MRI using the multivariate general linear model

Robert Leech and Dennis Leech

Abstract—Much current research in functional magnetic resonance imaging (fMRI) employs multivariate machine learning approaches (e.g., support vector machines) to detect distributed spatial patterns from the temporal fluctuations of the neural signal. The aim of many studies is not classification, however, but investigation of multivariate spatial patterns, which pattern classifiers detect only indirectly. Here we propose a direct statistical measure for the existence of distributed spatial patterns (or spatial heterogeneity) applicable to fMRI datasets. We extend the univariate general linear model (GLM), typically used in fMRI analysis, to a multivariate case. We demonstrate that contrasting maximum likelihood estimations of different restrictions on this multivariate model can be used to estimate the extent of spatial heterogeneity in fMRI data. Under asymptotic assumptions inference can be made with reference to the χ^2 distribution. The test statistic is then assessed using simulated timecourses derived from real fMRI data followed by analyzing data from a real fMRI experiment. These analyses demonstrate the utility of the proposed measure of heterogeneity as well as considerations in its application. Measuring spatial heterogeneity in fMRI has important theoretical implications in its own right and may have potential uses for better characterising neurological conditions such as stroke and Alzheimer's disease.

Index Terms—Functional magnetic resonance imaging, multi-voxel pattern analysis, general linear model, spatial heterogeneity

I. INTRODUCTION

THE traditional approach to quantifying changes in brain activation with functional magnetic resonance imaging (fMRI) is predominantly univariate: a separate general linear model is fitted to time series data for each of the many tens of thousands of spatially distinct measurement locations (voxels). Given the relatively poor signal-to-noise of the measurable signal, Gaussian spatial smoothing across neighboring voxels is normally applied. However, this smoothing assumes that the detectable neural signal is spatially homogeneous: that is, approximately the same general linear model should fit adjacent voxels. Some multivoxel analysis techniques indicate that brain regions may be highly heterogeneous [1], [2], [3] and demonstrate huge improvements in detecting signal over the traditional univariate approach. Furthermore, whether areas of the brain are highly heterogeneous is important theoretically for understanding the roles of regions

R Leech is with The Computational, Cognitive and Clinical Neuroimaging Laboratory, The Division of Experimental Medicine, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0NN, UK

D Leech is in the Department of Economics, University of Warwick, Coventry, CV4 7AL

of cortex, with potential implications for better understanding neurological conditions such as Alzheimer's disease or stroke, although little work has been done to date in this field. However, the multivariate pattern classification techniques used to date (e.g., support vector machines and linear discriminant analysis) are indirect ways to measure the heterogeneity of the fMRI signal across brain regions.

In this paper, we propose a framework and test statistic to directly measure the extent to which there is a heterogeneous pattern of activation across neighboring voxels. The proposal takes as its starting point the information-based approach of Kriegeskorte and colleagues [4]. Instead of using multivoxel pattern classifiers to decode which state the brain is in as a proxy for the amount of information available in a region, Kriegeskorte and colleagues proposed using a variant on the Mahalanobis Distance. 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 [4], [5], [6], [7].

The Mahalanobis distance [8] is a multivoxel similarity measurement that unlike Euclidean distance is scale invariant and controls for covariance across data points. In fMRI datasets, the Mahalanobis distance can quantify 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, [9] 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 their covariance matrix across voxel

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

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

We start by demonstrating how the Mahalanobis distance relates to the multivariate general linear model. We then set up restrictions on the GLM to compare models under different assumptions about the spatial distribution of the fMRI signal. An assumption of spatial homogeneity is contrasted with

an assumption of spatial hetereogeneity resulting in a log-likilhood ratio statistic measuring the heterogeneity in a patch of cortex. This statistic measures the amount of distributed information across a number of voxels. It is then tested with synthetic and real fMRI data to demonstrate its utility and to explore limitations in its application.

II. THEORY

A. The statistical model

We start with the GLM as applied to fMRI datasets [10]. We assume the response by each voxel can be modelled by a linear regression model written as follows.

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

where y is the signal of the voxel, i is the voxel subscript, $i = 1, ..., n$, and t is the time subscript, $t = 1, ..., T$. β_{hi} represents the response of voxel i to the stimulus measured by regressor h and k is the number of regressors.

This can be written using matrix algebra as,

$$\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{u}_i \tag{3}$$

where $\boldsymbol{y}_i$ is the T element observation vector for voxel i, $\boldsymbol{X}$ the Txk input matrix, $\boldsymbol{u}_i$ 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, $E(\boldsymbol{u}_i) = \boldsymbol{0}$ for all i and $E(\boldsymbol{u}_i \boldsymbol{u}_j') = \sigma_{ij} \boldsymbol{I}_T$ for all i and j.

This system of equations can be written more compactly as:

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

where

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}, \quad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix},$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} & 0 & \dots & 0 \\ 0 & \boldsymbol{X} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{X} \end{pmatrix} = \boldsymbol{I}_n \otimes \boldsymbol{X},$$

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

and,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{pmatrix}.$$

It is well known (for example [11], [12]) that, when there is a common set of regressors, the correlation of the errors in different equations is irrelevant as far as coefficient estimation is concerned. The efficient estimator, equivalent to generalized least squares, is ordinary least squares applied to each equation separately. That is, the efficient estimator of the complete system is written:

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_i,$$

for all $i = 1, 2, ..., n$, and therefore,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}. \tag{5}$$

The covariance matrix of $\hat{\boldsymbol{\beta}}$ can be shown to be

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

Noting that,

$$\boldsymbol{Z}'\boldsymbol{Z} = [\boldsymbol{I}_n \otimes \boldsymbol{X}'][\boldsymbol{I}_n \otimes \boldsymbol{X}] = \boldsymbol{I}_n \otimes \boldsymbol{X}'\boldsymbol{X},$$

and hence that.

$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = \boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1},$$

and that,

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

we can rewrite (6) as

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = [\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}][\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})][\boldsymbol{I}_n \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}]$$
$$= \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{7}$$

This gives a simple expression for the covariance between any pair of estimated coefficients. Letting $\boldsymbol{W} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, we can write

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

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

All this assumes that the covariance matrix of equation errors, $\boldsymbol{\Sigma}$, is known. In practice it must be estimated consistently from the residuals, using,

$$\hat{\sigma}_{ij} = (\boldsymbol{y}_i - \boldsymbol{X}\hat{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_j - \boldsymbol{X}\hat{\boldsymbol{\beta}}_j)/T, \tag{9}$$

(or, $\hat{\sigma}_{ij} = (\boldsymbol{y}_i - \boldsymbol{X}\hat{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_j - \boldsymbol{X}\hat{\boldsymbol{\beta}}_j)/(T-k)$) and $\boldsymbol{\Sigma}$ replaced with $\hat{\boldsymbol{\Sigma}}$ throughout in the manner of FGLS.

B. Inference

Our focus is on testing a set of linear restrictions on coefficients across the separate voxel equations. We assume there are g restrictions written in the form $\boldsymbol{R\beta} = \boldsymbol{0}$ where $\boldsymbol{R}$ is a $g \times nk$ matrix of given constants of rank g. (We do not need to consider inhomogeneous restrictions here because units of measurement are arbitrary in the context.)

This framework allows us to devise tests for heterogeneity versus homogeneity (that is, constancy) of the effects of different stimuli across voxels, as well as to test if stimulae are significantly different from zero.

If the equation errors u_{it} are Gaussian then the efficient estimator of $\boldsymbol{\beta}$ given in equation (II-A) is normally distributed and the well known inferential procedures for generalized linear regression models are available based on the covariance matrix in equation (6), (7) or (8), see for example [13]. These hold asymptotically if $\boldsymbol{\Sigma}$ is estimated consistently.

We characterise the inferential problem as testing a null hypothesis of the general type

$$H_0 : \boldsymbol{R\beta} = \boldsymbol{0}. \qquad (10)$$

which can be approached in various ways, for example by a Wald test.

Letting $\boldsymbol{R\beta} = \boldsymbol{\delta}$ the null hypothesis of interest becomes $H_0 : \boldsymbol{\delta} = \boldsymbol{0}$. Now define $\hat{\boldsymbol{\delta}} = \boldsymbol{R\hat{\beta}}$. Then $\hat{\boldsymbol{\delta}}$ is asymptotically normally distributed with expectation $E(\hat{\boldsymbol{\delta}}) = \boldsymbol{0}$ and covariance matrix $V(\hat{\boldsymbol{\delta}}) = \boldsymbol{R}V(\hat{\boldsymbol{\beta}})\boldsymbol{R'}$. We call the test statistic Δ, given by

$$\Delta = \hat{\boldsymbol{\delta}}'[\boldsymbol{R}V(\hat{\boldsymbol{\beta}})\boldsymbol{R'}]^{-1}\hat{\boldsymbol{\delta}} \qquad (11)$$

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

Alternatively we employ the likelihood principle and likelihood ratio test (LR) based on the loglikelihood function

$$lnL = -\frac{Tn}{2}ln(2\pi) - \frac{1}{2}ln|\boldsymbol{\Omega}| - \frac{1}{2}Q \qquad (12)$$

where Q is the appropriate generalized sum of squared residuals,

$$Q = \boldsymbol{u'\Omega^{-1}u}. \qquad (13)$$

We can specify a LR test of a null hypothesis in the form of (10) as twice the difference between the maximized value of (12) under the null and alternate hypotheses, $LR = -2(lnL_0 - lnL_1)$. The test statistic is then

$$LR = Q_0 - Q_1, \qquad (14)$$

where Q_0 is the minimum of (13) under H_0 and Q_1 the unconstrained minimum (or, the minimum under the alternate H_1). This assumes $\boldsymbol{\Sigma}$ to be known or the same estimate of it used under both the null and alternate. Standard theory gives the statistic (14) an asymptotic χ^2 distribution with g degrees of freedom. (See [12].)

C. Testing heterogeneity across voxels

We now describe tests of spatial characteristics of the multivariate signal across voxels. In particular, we are interested in investigating heterogeneity of response coefficients. We describe a procedure for comparing the unrestricted model, allowing heterogeneity between voxels, with a restricted model, which we say has homogeneity between voxels.

We characterize the null hypothesis of homogeneity in terms of a linear function of the coefficients. For the ith voxel the function of interest is written $\boldsymbol{r'\beta_i}$ for suitable constants $\boldsymbol{r'} = (r_1, r_2, \cdots, r_k)$. The null hypothesis states that $\boldsymbol{r'\beta_i}$ is constant for all i. That is, $\boldsymbol{r'\beta_i} = \boldsymbol{r'\beta_j}$ for all $i \neq j$. Therefore, we can write the restrictions to be tested in the form: $H_0 : \boldsymbol{R\beta} = \boldsymbol{0}$

with

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{r'} & -\boldsymbol{r'} & \boldsymbol{0'} & \cdots & \boldsymbol{0'} \\ \boldsymbol{r'} & \boldsymbol{0'} & -\boldsymbol{r'} & \cdots & \boldsymbol{0'} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{r'} & \boldsymbol{0'} & \boldsymbol{0'} & \cdots & -\boldsymbol{r'} \end{pmatrix}.$$

An important case is where interest centers on the *difference* between two effects. If the two effects are assumed to be the first two coefficients, then $\boldsymbol{r'\beta_i} = \beta_{1i} - \beta_{2i}$ and $\boldsymbol{r'} = (1, -1, 0, \cdots 0)$.

Let the common difference be θ. Then the restricted model can be written

$$H_0 : \beta_{1i} - \beta_{2i} = \theta, \text{ constant for all } i.$$

The number of coefficients under the unrestricted model, H_1, is nk (the βs). Under the restrictions in H_0 there are $nk - n + 1$ coefficients (the nk β's less the n β_2's plus θ). Therefore the number of parameter restrictions under H_0 is $n - 1$.

The LR test requires estimating the model, and evaluating the log likelihood, separately under H_1 and H_0. We proceed as if $\boldsymbol{\Sigma}$ is known; it is estimated from the residuals of the unrestricted model. The unrestricted model is efficiently estimated as a set of ordinary least squares regressions. The restricted model can be estimated using a simple reparameterization as follows.

Write $\beta_{1i} = \beta_{2i} + (\beta_{1i} - \beta_{2i})$. Under H_0, this becomes $\beta_{1i} = \beta_{2i} + \theta$.

Substituting into (2) and collecting terms in β_{2i} gives the restricted set of equations,

$$y_{it} = \theta x_{1t} + \beta_{2i}(x_{1t} + x_{2t}) + \sum_{h=3}^{k} \beta_{hi}x_{ht} + u_{it} \qquad (15)$$

The system of equations (15) is a set of seemingly unrelated regressions with a common set of regressors (a multivariate regression) with cross-equation restrictions. It can be written, in matrix notation, as

$$\boldsymbol{y} = \theta\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{Z}}\boldsymbol{\gamma} + \boldsymbol{u}, \qquad (16)$$

where we define

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_n \end{pmatrix}, \boldsymbol{\gamma}_i = \begin{pmatrix} \beta_{2i} \\ \beta_{3i} \\ \vdots \\ \beta_{ki} \end{pmatrix}, \ i = 1, ..., n,$$

and,

$$\tilde{\boldsymbol{Z}} = \begin{pmatrix} \tilde{\boldsymbol{X}} & 0 & \dots & 0 \\ 0 & \tilde{\boldsymbol{X}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{\boldsymbol{X}} \end{pmatrix} = \boldsymbol{I}_n \otimes \tilde{\boldsymbol{X}},$$

$\tilde{\boldsymbol{X}}$ is the $T(k-1)$ matrix, $\tilde{\boldsymbol{X}} = [\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_k]$, where $\boldsymbol{x}_i$ is the ith column of the design matrix $\boldsymbol{X}$,

$$\text{and } \tilde{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_1 \end{pmatrix} = \boldsymbol{\iota}_n \otimes \boldsymbol{x}_1, \text{where } = \boldsymbol{\iota}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

We can rewrite the system

$$\boldsymbol{y} - \theta\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{Z}}\boldsymbol{\gamma} + \boldsymbol{u}, \qquad (17)$$

which can also be written,

$$\boldsymbol{y}_i - \theta\boldsymbol{x}_1 = \tilde{\boldsymbol{X}}\boldsymbol{\gamma_i} + \boldsymbol{u_i}, i = 1, \dots n,$$

a system with common regressor set for given θ, and therefore the $\boldsymbol{\gamma}_i$s are estimated efficiently by OLS applied separately to each equation.

This is equivalent to estimating (17) by ordinary least squares giving,

$$\hat{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}'(\boldsymbol{y} - \theta\tilde{\boldsymbol{x}}). \qquad (18)$$

which is the maximum likelihood estimator of $\boldsymbol{\gamma}$ conditional on θ.

The likelihood function (12) is maximized when the generalized least squares criterion (13) is minimized. Write,

$$Q(\theta, \boldsymbol{\gamma}) = \boldsymbol{u}'\boldsymbol{\Omega}^{-1}\boldsymbol{u} = (\boldsymbol{y} - \theta\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{Z}}\boldsymbol{\gamma})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \theta\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{Z}}\boldsymbol{\gamma}).$$

Substituting (18) for $\boldsymbol{\gamma}$ in this, we concentrate the generalized sum of squared residuals function (13) with respect to γ and obtain a function of θ only, say $Q^*(\theta)$:

$$Q^*(\theta) = [\boldsymbol{y} - \theta\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{Z}}(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{y} - \theta\tilde{\boldsymbol{Z}}(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{x}}]'\boldsymbol{\Omega}^{-1}[\dots].$$

If we write $\boldsymbol{M} = \boldsymbol{I} - \tilde{\boldsymbol{Z}}(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}'$, we can define $\boldsymbol{y}^r = \boldsymbol{M}\boldsymbol{y}$, and $\tilde{\boldsymbol{x}}^r = \boldsymbol{M}\tilde{\boldsymbol{x}}$ as residuals from least squares regressions on $\tilde{\boldsymbol{Z}}$.

Hence,

$$Q^*(\theta) = [\boldsymbol{y}^r - \theta\tilde{\boldsymbol{x}}^r]'\boldsymbol{\Omega}^{-1}[\boldsymbol{y}^r - \theta\tilde{\boldsymbol{x}}^r], \qquad (19)$$

which is minimized by the maximum likelihood estimator.

The estimator of θ, which minimizes (19), and is also the ML estimator, is easily shown to be found by a bivariate generalized least squares regression of $\boldsymbol{y}^r$ on $\tilde{\boldsymbol{x}}^r$, giving:

$$\hat{\theta} = \frac{\tilde{\boldsymbol{x}}^{r'}\boldsymbol{\Omega}^{-1}\boldsymbol{y}^r}{\tilde{\boldsymbol{x}}^{r'}\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{x}}^r}. \qquad (20)$$

The LR criterion (14) is then simply evaluated.

III. SIMULATIONS

To test the proposed log-likelihood-ratio measure of heterogenity, we applied it to simulated data generated from an fMRI dataset. Using synthetic data allowed us to: (1) systematically vary the spatial characteristics of the signal; (2) test the validity of the asymptotic χ^2 distribution of the test statistic under different conditions (i.e., with different numbers of voxels and timepoints); and (3) investigate violations of the assumptions of the GLM, i.e., autocorrelation of error.

FMRI data was taken from a 32-year-old neurologically healthy male participant who provided informed consent according to local ethics procedures. Two-hundred and seventy-six echoplanar images of resting data (i.e., the participant lay in the scanner doing nothing) were acquired (i.e., 276 timepoints at each voxel). MRI data were obtained on a Philips Intera 3.0 Tesla scanner, using dual gradients, a phased array head coil, and sensitivity encoding with an undersampling factor of 2. Functional MR images were obtained using a T2-weighted, gradient-echo, echoplanar imaging (EPI) sequence with whole-brain coverage (TR=2s, echo time = 30ms; flip angle, 90). Thirty-two axial slices with a slice thickness of 3.25mm and an interslice gap of 0.75mm were acquired in ascending order (resolution, 2.19 x 2.19 x 4.00mm; field of view, 280 x 224 x 128 mm). Quadratic shim gradients were used to correct for magnetic field inhomogeneities within the anatomy of interest.

Standard pre-processing of the functional data was conducted using FEAT (FMRI Expert Analysis Tool) Version 5.98. The data was motion-corrected [14] to correct for participant head movement and high-pass temporally filtered to remove linear trends and some physiological noise. No spatial smoothing was applied. The covariance matrix across neighboring voxels (centered on a single voxel randomly chosen from within superior temporal cortex) was calculated. This covariance matrix was used to generate random noise sampled from a multivariate normal distribution.

A. Simulating spatial heterogeneity

In the first set of simulations, we demonstrate that the test statistic has power with respect to multivariate heterogeneity. A design matrix was constructed with three columns (k=3),
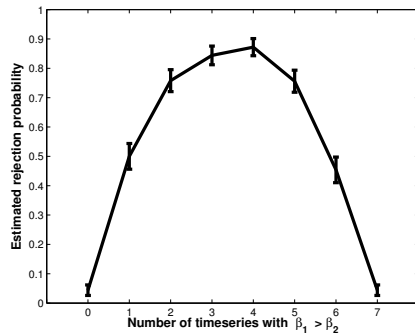
Fig. 1. Varying heterogeneity across the seven timeseries. At the extreme values of the x-axis all timeseries (voxels) have the same β coefficients: the homogeneous cases. In the center of the x-axis, half the voxels have $\beta_1 > \beta_2$ and half the reverse. The y-axis is the log-likelihood ratio comparing the unrestricted and restricted models. Mean and standard error of the estimated rejection probability across 500 simulations are shown.
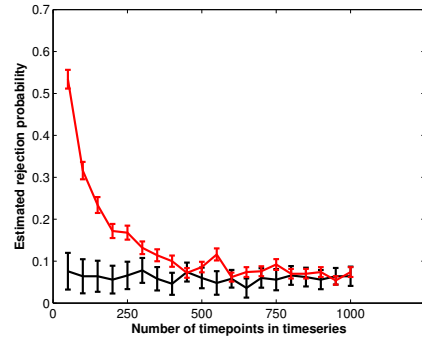

Fig. 2. The proportion of models that incorrectly reject the null hypothesis assuming the difference in log likelihoods is χ^2 distributed for different numbers of timepoints. The red line is the case with 33 timeseries and the black line with 7 timeseries. Values are the mean and standard error of 500 simulations.

consisting of two dichotomous variables x_{1t} and x_{2t} and a constant x_{3t}. At each timepoint either $x_{1t} = 1$ (and $x_{2t} = 0$), $x_{2t} = 1$ (and $x_{1t} = 0$) or, for a third of timepoints, both $x_{1t} = 0$ and $x_{2t} = 0$. In the first case, seven timeseries (corresponding to a central voxel and adjacent voxels) were generated by drawing 276 timepoints from a multivariate normal distribution with zero mean vector and a given real covariance matrix. The design matrix multiplied by assumed β values was added to the random noise. In the homogeneous case, all seven timeseries had the same signal (i.e., the betas associated with experimental condition one were either positive in all voxels or negative in all voxels). In the heterogenous case, half of the timeseries had positive β values for condition one and half had negative β values. Note the overall signal was the same in all situations (i.e., the absolute difference between condition 1 and condition 2 was constant).

Figure 1 shows the results of 500 simulations with different covariance matrices and randomly generated timeseries. The x-axis represents the number of timeseries with homogeneous signals. As the number of timeseries with β_1 positive and β_2 negative increases the heterogeneity across the timeseries increases. This is reflected in the increase in the estimated rejection probability. As such, the test statistic reflects the degree of heterogeneity in the β coefficients across the timeseries.

B. Asymptotic χ^2 assumption

The χ^2 properties of the test statistic only hold asymptotically, and previous work on seemingly unrelated regression suggests that as the covariance matrix increases in size, this assumption becomes untenable except with very large numbers of timepoints [15]. With smaller numbers of timepoints, overrejection of the null hypothesis becomes an issue.

We investigated the validity of the χ^2 assumption in two situations: 1) one with 7 timeseries (equivalent to a sphere of 1 voxel radius around a central voxel); and 2) one with 33 timeseries (equivalent to a sphere of 2 voxel radius). In both situations, multivariate random noise was generated with a real covariance structure. Since we are interested in assessing the overrejection rate of estimated χ^2 values, no signal was inserted into the background noise ($\beta = 0$). The number of timepoints was varied from 50 to 1500, and a hundred simulations and models were fitted with each number of timepoints.

Figure 2 shows the proportion of models that are judged to be significant at a $p < 0.05$ significance level. Given the absence of signal, this value should be approximately 0.05. We see that for the seven timeseries case, there is not substantial overrejection of the null hypothesis, even with only 50 timepoints. However, for the larger 33 timeseries scenario, incorrect overrejection of the null hypothesis occurs frequently; with only 50 timepoints, rejection of the null occurs half the time. Only after approximately 500 timepoints does the rejection rate approximate that expected given the null hypothesis. This finding highlights the problem of accurately estimating large covariance matrices without large numbers of datapoints. In these situations, the log-likelihood ratio does not adequately approximate the χ^2 distribution.

C. Autocorrelation of the residuals

One common problem in applying the GLM to fMRI timeseries is the presence of autocorrelation in the residuals. In general, this can lead to underestimation of the error variance and inefficient model estimation. We conducted simulations to investigate the effects of error autocorrelation on the proposed log-likelihood ratio statistic. Fifteen-order autocorrelation estimates were calculated using standard analysis software fsl [16]. Multivariate normally-distributed random data was then filtered using these autocorrelation estimates to simulate realistic autocorrelation in an fMRI dataset. Two situations were considered: one with a rapidly
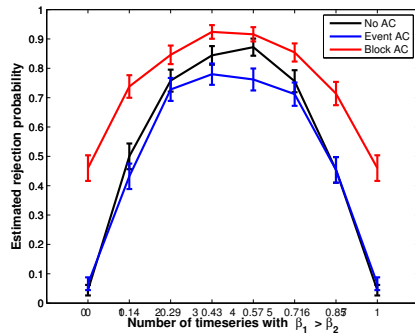
Fig. 3. The effects of autocorrelated residuals on the log-likelihood ratio statistics. The three lines correspond to either the situation with no autocorrelation of the residuals, or else realistic error autocorrelation, but with two different design matrices corresponding to standard experimental paradigms. Values are the mean estimated rejection probability and standard error of 500 simulations.

varying design matrix (with timepoints assigned to either condition one or condition two intermixed at random; equivalent to an event-related fMRI experimental design). The second was equivalent to a blocked fMRI design, with a continuous sequence of timepoints (one third of the total) corresponding to one condition, followed by another third corresponding to the other condition.

Figure 3 shows the performance of the blocked and the event-related experimental design in the presence of autocorrelation of the residuals. Inflation of type-II errors occurred for the blocked design; whereas, the event-related design was similar (slightly more conservative) than a model with no autocorrelation. This result highlights that correction of autocorrelated noise is an important step for the proposed statistic (as for many fMRI analyses), especially for blocked experimental designs. Standard techniques for prewhitening [16] non-white noise can be used in conjunction with the statistics proposed here.

IV. AN fMRI EXPERIMENT

Building on the simulation results, the measure of heterogeneity was applied to real fMRI data from a speech production and motor movement paradigm. Using real data allowed us to demonstrate that the approach is appropriate to detect distributed patterns in standard fMRI paradigms. In addition, using real data allowed us to briefly consider a number of practical issues about the application of the technique including: the types of distributed patterns that can be detected by the approach, the effect of different tissue types on the test statistic, multiple comparison correction and the consequences of spatial smoothing.

FMRI data was taken from the same 32-year-old healthy male participant as used to derive the simulation random datasets above. The participant either described black and white pictures presented visually using volitional, overt speech or moved their tongue. In the fMRI scanning session,

the subject produced overt propositional speech. Rest was included as a baseline condition.

One-hundred and forty four echoplanar images of data were acquired using the same scanning protocol as detailed above. The only difference was that a sparse fMRI design was used to minimize movement- and respiratory-related artifacts associated with speech studies (TR = 10 secs: acquisition time = 2 secs, giving 8 seconds for the subjects to speak during silence). Stimuli were presented visually using E-Prime software (Psychology Software Tools) run on an IFIS-SA system (In Vivo Corporation).

The data was preprocessed as detailed in the simulation section above. In addition, voxel timeseries were prewhitened and the autoregression model parameters used to filter the design matrix voxelwise [16]. The prewhitened data was used in subsequent analyses. The analysis was restricted to the precentral gyrus and supplementary motor area, cortical regions that are involved in oro-motor control of the tongue and articulators. These regions were defined anatomically using the Harvard-Oxford probabilistic atlas and transformed into the subject's native space.

For each voxel, separate general linear models were calculated for all voxels in the neighborhood, including speech production, tongue movements and a constant in the model. Each variable was convolved with a canonical double gamma function to account for hemodynamic delay. The test for spatial heterogeneity was then applied using either 1 or 2 voxel spheres. For the 1-voxel sphere case, the log-likilhood ratio statistic was evaluated with regard to the χ^2 distribution. As with most voxelwise fMRI analysis, there are problems with multiple comparison leading to an increase in overejection of the null-hypothesis. Here, false discovery rate [17] was used to adjust the p-value to control for multiple comparisons. Other correction methods could also be used (e.g., bonferroni correction or some form of cluster correction [18] depending on the type of analysis and question being asked), however, approaches based on random field theory [19] may be inappropriate due to the absence of Gaussian spatial smoothing as a preprocessing step.

Figure 4 shows the results of this analysis revealing a widespread pattern of heterogeneity, predominantly in medial and right-lateralized motor regions. Results from a single subject are sufficient to demonstrate significant heterogeneity after correcting for multiple comparisons. These results are substantially different from the more typical subtraction analysis following spatial averaging across adjacent voxels presented in Figure 5 for comparison purposes (restricted to the same ROI as the multivariate analysis). The subtraction analysis reveals fewer significant voxels and the results do not survive multiple comparison correction, and so are presented here at a liberal threshold. Unlike the heterogeneity analysis, the subtraction analysis is not right-lateralized.

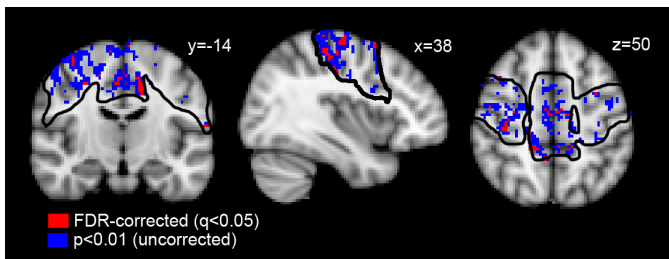Although the results from Figure 4 suggest there are regions

Fig. 4. Voxelwise analysis of heterogeneity in motor cortical regions overlaid on the MNI152 average brain. Voxels colored in red are significantly more heterogeneous than expected by chance $p < 0.05$, corrected for multiple comparisons using false discovery rate, FDR. For illustrative reasons, and because the results are only from a single subject, a more liberal threshold of $p < 0.01$ is also included (in blue). The outline of the extensive region of interest is displayed in black.
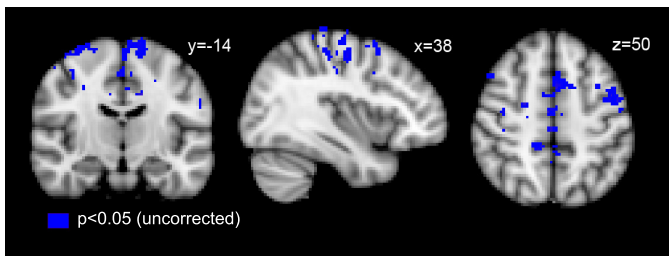

Fig. 5. A standard activation analysis (i.e., t-tests) following spatial averaging. The results are reported at the liberal p<0.05 level uncorrected for multiple comparisons, to highlight the different pattern from the heterogeneity analysis in Figure 4.

of primary somatosensory, motor and premotor cortex that possess significant heterogeneity in their task-evoked profile of activation, this could reflect multiple different underlying spatial distributions. Figure 6 presents four idealized patterns of fMRI data that give rise to homogeneous or heterogeneous activation profiles. The homogeneous case can arise either because: (a) the β vectors are all zero across the neighborhood of voxels (i.e., there is no signal); or (b) all voxels have equal β vectors. In contrast, significant heterogeneity can arise because: (c) the β vectors are different but non-zero; or (d) each voxel has a β vector (apart from the intercept and coefficients of variables not of direct interest) either equal to a given vector or zero; or a combination of (c) and (d). The distributed activation profiles such as (c) are presumably what many pattern classification approaches to fMRI analysis have utilized. Whereas the pattern described in (d) is likely to occur at the boundaries of cortical regions or where neighborhoods of voxels span white matter (with no signal) and grey matter.

The type of heterogeneity of interest to the experimenter - whether (c) or (d) - will depend on the experimental question being asked, and it is important to be able to distinguish between the different cases. For example, heterogeneity arising from absence of signal in white matter is unlikely to be theoretically interesting. It is possible to restrict the analyses to only voxels on the cortical surface or constrained to within grey matter in order to avoid this possibility. Similarly, when comparing heterogeneity across subjects or

patient groups that might differ in, for example, cortical atrophy differences in grey matter extent could be covaried out voxelwise. (This approach is well established with other fMRI analysis techniques e.g., [20].)

A second approach to identify situations such as (d) would be to pretest, for each voxel, the null hypothesis that its β vector (apart from the intercept and coefficients on nuisance variables) is zero by a suitable F test applied to the least squares regressions. Subsequently, only those voxels where the null is rejected are included in the multivariate analysis and used for estimating covariances. This pretest ensures that heterogeneity is calculated only across voxels that are all significantly activated in some way by the stimuli; as such it reduces the effect of neighboring voxels containing, for example, white matter, cerebrospinal fluid or grey matter from adjacent regions not affected by the stimuli, all of which could inflate measured heterogeneity. One drawback of this approach is that it runs the risk of falsely excluding voxels due to type II errors.

A third approach to distinguishing between patterns (c) and (d) involves using the log-likelihood estimates of the different idealized homogeneous patterns (a) and (b) in conjunction with the log-likelihoods of the heterogeneity cases. Contrasting the log-likihood of (b) with (a) provides a measure of the amount of signal across the voxels assuming the data are homogeneous (the average signal). In situation (d), i.e., heterogeneity arising from a subset of voxels having βs equal to zero and others β nonzero, as the nonzero β values increase then so will the measure of heterogeneity. However, in tandem with the measure of heterogeneity increasing, the average signal (i.e., the contrast of (a) and (b)) also rises. That is, in situation (d) the measures of heterogeneous and average signal will display a linear relationship, as the β values increase. This contrasts with the heterogeneous situation presented in (c), where the measure of heterogeneity varies independently of the measurement of average signal.

Distinguishing between situations (c) and (d) is likely to be done when comparing across subjects or conditions that might differ in their heterogeneity (e.g., different patient groups or during or under different task difficulty). In such a situation, both measures of heterogeneity and average signal would be calculated for each individual and then both measures entered simultaneously in a general linear model. Evidence for spatial heterogeneity described in (c) would be provided if there was a significant effect of heterogeneity controlling for the average signal. Alternatively, the likeilhood ratio statistic corresponding to the average signal could be used to scale the heterogeneous statistic.

As an illustration, we demonstrate this approach, in a single subject, by comparing heterogeneity in the left and right hemispheres. The data were registered into MNI 152 standard 2mm space, and the left hemisphere was flipped so that voxels in both hemispheres could be directly

Fig. 7. An illustration of the effects of typical spatial smoothing of 8mm FWHM on the measure of heterogeneity. The figure reports the difference in log-likelihood ratio statistics calculated for smoothed and unsmoothed data.
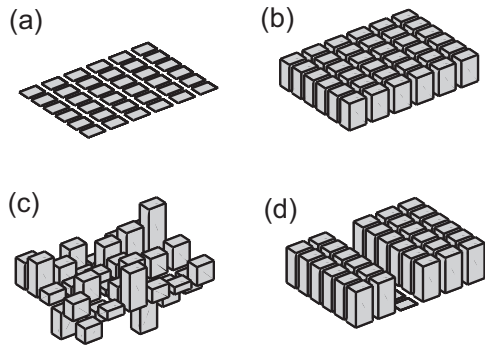

Fig. 6. An illustration of four different ideal spatial patterns: (a) The data is homogeneous because there is no signal. (b) The data is homogeneous because all voxels have the same signal. (c) The data is heterogeneous because voxels have different amounts of signal. (d) The data is heterogeneous because a subset of voxels have no signal.

compared. The likelihood ratio statistics for heterogeneity and for homogeneity were calculated for each voxel. The heterogeneity was scaled by the homogeneity statistic and the difference between left and right calculated. Under the null-hypothesis that heterogeneity is equivalent across the left and right hemispheres, the differences across the hemispheres should be zero centered. As such, parametric or nonparametric statistics can be used to show that there is significantly greater heterogeneity on the right than the left ($p < 0.001$, using a non-parametric sign test) and this heterogeneity is not the consequence of a subset of voxels having zero-signal or greater signal in a subset of voxels having greater signal on the right. It is worth emphasizing that there are multiple ways to compare across hemispheres; the approach here is in no means considered to be definite, but rather provides a simple way to illustrate how to disambiguate different spatial patterns underlying heterogeneity with the data from a single subject.

One typical step in fMRI analysis to improve signal detection is to spatially smooth the data, thereby changing information at higher spatial frequencies. Figure 7 illustrates the effects of spatial smoothing the data using 8mm full-width half max (FWHM) nonlinear spatial filtering [21]. Smoothing the data leads to a general reduction in the extent of spatial heterogeneity, as expected, given the loss of high-frequency spatial information. Note also, that this reduction in heterogeneity is unlikely to be due to averaging of data across different tissue types, because the nonlinear filter only smoothes across voxels of similar tissue (e.g., grey or white matter). This general reduction in heterogeneity is not, however, consistent over the whole brain, with some voxels displaying increases in their heterogeneity. This apparent local increase in heterogeneity, although at first counterintuitive, is theoretically possible, given that spatial smoothing typically improves signal to noise in fMRI analyses. As such, individual voxel activation patterns may be
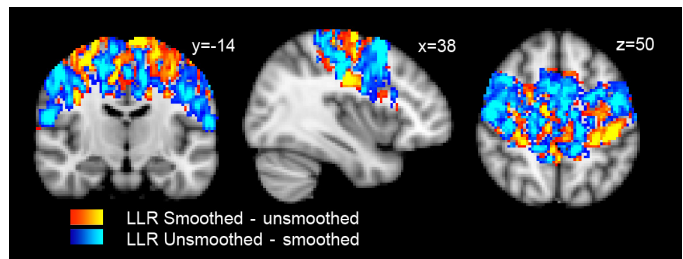
more reliably detected following smoothing. At boundaries between cortical regions or boundaries with white matter or cerebrospinal fluid, the local neigborhood consists of voxels that activate for a task as well as voxels that do not activate for the same task. As such some spatial smoothing as a pre-processing step may improve the signal-to-noise of the active voxels and reduce noise in non-active voxels. Subsequent spatial heterogeneity measures may therefore more reliably detect this boundary between active and non-active voxels leading to a greater measure of spatial heterogeneity.

V. DISCUSSION

In this paper, we have presented a measure of multivariate heterogeneity across multiple timeseries, that is designed to be able to assess the amount of multivoxel information available in functional MRI datasets. We extended the univariate GLM to a multivariate case, an example of seemingly unrelated regression, equivalent to a Mahalanobis distance measure previously used with fMRI. An unrestricted version of the multivariate GLM provides the maximum likelihood model across timeseries (allowing $\boldsymbol{\beta}$ values to be different for each timeseries). A restricted model assessed the homogeneous case that all timeseries had the same $\boldsymbol{\beta}$ coefficients (i.e., the maximum likelihood $\boldsymbol{\beta}$ coefficients that are constrained to be the same in all voxels). Comparing the restricted and unrestricted cases measures the amount of heterogeneity across timeseries and is asymptotically χ^2 distributed.

Simulations using synthetic data derived from an fMRI resting state experiment illustrated the approach, investigated its relationship to the theoretical asymptotic χ^2 distribution under a range of circumstances and considered the effects of auto-correlation on sensitivity and over-rejection rates. Subsequent analysis of speech production and oromotor movement fMRI data illustrated the use of the heterogeneity statistic using a spherical searchlight in an fMRI experiment. This analysis explored the effect of spatial smoothing using a typical Gaussian smoothing kernel resulting in broad but not uniform reduction in measured heterogeneity. The fMRI experiment was also used to consider ways of disambiguating different types of heterogeneous spatial pattern: such as distinguishing patterns that might occur at boundaries of cortical regions as distinct from the type of spatially complex, distributed heterogeneity that multivoxel pattern classifiers

might be using to substantially improve signal detection. This measure is intended to be applied in a searchlight approach, whereby patches of cortex (e.g,. spheres of voxels) are considered in turn, and the spatial heterogeneity of the neural pattern mapped out. The statistic could also be used to consider heterogeneity across multiple peak voxels within a single cluster of activation, to assess whether it is truly a single cluster. Future theoretical and empirical work is necessary to develop more sophisticated methods for discriminating between different types of heterogeneity and to uncover where these exist in the brain.

Many recent and high-profile fMRI studies [1], [22], [23], [24], [25], [26], [3] have taken advantage of multivariate pattern analysis techniques, however, until now, no specific statistical measure has been formulated to assess whether such approaches are picking up distributed spatial patterns across voxels. The measure of heterogeneity proposed here starts to fulfil this role, providing a statistically well-grounded quantification of multivoxel spatial heterogeneity.

There are limits to how much this measure of spatial heterogeneity can tell us about the spatial organization of the fMRI signal though. There is an ongoing debate as to whether multivariate pattern analysis can detect fine-scale (i.e., sub-voxel) signals [27], [28], [29]. Multivariate pattern classification suggests that classifiers can take advantage of information from orientation columns in visual cortex that is represented at a spatial resolution less than 1mm and so smaller than that of voxels [2]. However, spatial smoothing of the data can in some circumstances increase the success of pattern classifiers [27], leading some to argue that the classifiers may not be taking advantage of high-resolution spatial patterns. The measure of spatial heterogeneity explored here does demonstrate where there is spatial variation of the fMRI signal across voxels; a necessary condition for fine-scale pattern analysis. However, the presence of spatial heterogeneity does not on its own demonstrate the presence of fine-scale spatial patterns is present or not; i.e., spatial heterogeneity is not sufficient for inference about fine-scale spatial patterns. Instead, fine-scale information must be demonstrated more directly e.g., by relating the spatially heterogeneous pattern to successful stimulus decoding. Our exploration of spatial heterogeneity following spatial smoothing is relevant to the ongoing debate though, by demonstrating how the spatial heterogeneity changes following spatial smoothing, particularly by showing that heterogeneity can both increase and decrease.

The spatial distribution of patterns of neural activation has important theoretical implications in its own right. Spatially-distributed patterns of neural activation suggest complex underlying neural processing of a given task. Highly spatially homogeneous patterns of activation suggest underlying processing dedicated to a specific task or modality; this feeds into long-standing debates in cognitive neuroscience about regional neural specialization for a given task. The proposed heterogeneity statistic may also be relevant for better understanding various neurological conditions. Altered spatial heterogeneity has been previously reported as a hallmark of neural pathology in a range of conditions. Several studies have used simple correlation-based functional connectivity measures of spatial heterogeneity in fMRI 'resting state' studies (i.e., without explicit tasks) of various psychiatric and neurological disorders. For example, one study showed altered spatial heterogeneity in Alzheimer's disease [30] and another showed increased heterogeneity with attention-deficit hyperactivity disorder within the default network [31]. Although not using direct measures of heterogeneity, Saur and colleagues have shown in patients following stroke that studying multivariate patterns of activation is useful in the diagnosis and monitoring of recovery from disease [32]. From a theoretical perspective, spatial heterogeneity may provide a measure of redundancy in a given cortical region, which may in turn predict resilience from neurological insult. It is hypothesized that with increasing pathology (e.g., atrophy in Alzheimers disease) there will be changes not just in the overall activation level but also in the local spatial heterogeneity of the activation.

The proposed approach is computationally efficient and potentially powerful; however, there are limitations to the asymptotic use of the log-likelihood ratio measure of heterogeneity for inference. In particular, the increase in the over-rejection rate when estimating larger covariance matrices. For the 33 voxel case, the number of timepoints necessary for the log-likelihood ratio statistic to approximate the χ^2 distribution is unrealistically large for many fMRI designs (that typically have fewer than 500 timepoints). This problem is known in the literature on multivariate regression with finite sample size and has been addressed using parametric bootstrapping to assess significance [15]. An alternative approach as discussed is to use the log-likelihood ratio measure of heterogeneity as a surrogate descriptive statistic to make comparisons across multiple participants, rather than to make formal inference. For example, data from two groups (patients and controls) could be transformed into a standard space, and the log-likelihood ratio at a given location could be calculated for each participant and compared across groups using parametric or non-parametric statistics. This approach is feasible for many fMRI designs and has been used previously for statistics whose distribution is not known [20].

References

[1] J. Haxby, M. Gobbini, M. Furey, A. Ishai, J. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[2] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature Neuroscience*, vol. 8, no. 5, pp. 679–685, 2005.

[3] J. Haynes, K. Sakai, G. Rees, S. Gilbert, C. Frith, and R. Passingham, "Reading hidden intentions in the human brain," *Current Biology*, vol. 17, no. 4, pp. 323–328, 2007.

[4] N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proceedings of the National Academy of Sciences*, vol. 103, no. 10, p. 3863, 2006.

[5] J. Serences and G. Boynton, "Feature-based attentional modulations in the absence of direct visual stimulation," *Neuron*, vol. 55, no. 2, pp. 301–312, 2007.

[6] ——, "The representation of behavioral choice for motion in human visual cortex," *Journal of Neuroscience*, vol. 27, no. 47, p. 12893, 2007.

[7] M. Stokes, R. Thompson, R. Cusack, and J. Duncan, "Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery," *Journal of Neuroscience*, vol. 29, no. 5, p. 1565, 2009.

[8] P. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Science of India*, vol. 12, no. 1, pp. 49–55, 1936.

[9] N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proceedings of the National Academy of Sciences*, vol. 103, no. 10, p. 3863, 2006.

[10] K. Friston, A. Holmes, K. Worsley, J. Poline, C. Frith, R. Frackowiak *et al.*, "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[11] A. Zellner, "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American Statistical Association*, vol. 57, no. 298, pp. 348–368, 1962.

[12] W. Greene, *Econometric analysis*. Prentice Hall Upper Saddle River, NJ, 2003.

[13] C. Rao, *Linear statistical inference and its applications*. Wiley New York, 1973.

[14] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.

[15] J. Dufour and L. Khalaf, *Simulation-based finite and large sample tests in multivariate regressions*. Université de Montréal, Centre de recherche et développement en économique, 2000.

[16] M. Woolrich, B. Ripley, M. Brady, and S. Smith, "Temporal autocorrelation in univariate linear modeling of FMRI data," *Neuroimage*, vol. 14, no. 6, pp. 1370–1386, 2001.

[17] C. Genovese, N. Lazar, and T. Nichols, "Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate* 1," *Neuroimage*, vol. 15, no. 4, pp. 870–878, 2002.

[18] S. Smith and T. Nichols, "Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference," *Neuroimage*, vol. 44, no. 1, pp. 83–98, 2009.

[19] K. Friston, K. Worsley, R. Frackowiak, J. Mazziotta, and A. Evans, "Assessing the significance of focal activations using their spatial extent," *Human brain mapping*, vol. 1, no. 3, pp. 210–220, 1993.

[20] N. Filippini, B. MacIntosh, M. Hough, G. Goodwin, G. Frisoni, S. Smith, P. Matthews, C. Beckmann, and C. Mackay, "Distinct patterns of brain activity in young carriers of the APOE-ε4 allele," *Proceedings of the National Academy of Sciences*, vol. 106, no. 17, p. 7209, 2009.

[21] S. Smith and J. Brady, "SUSAN—A new approach to low level image processing," *International journal of computer vision*, vol. 23, no. 1, pp. 45–78, 1997.

[22] D. Cox and R. Savoy, "Functional magnetic resonance imaging (fMRI)," *Neuroimage*, vol. 19, no. 2, pp. 261–270, 2003.

[23] M. Stokes, R. Thompson, R. Cusack, and J. Duncan, "Top-down activation of shape-specific population codes in visual cortex during mental imagery," *Journal of Neuroscience*, vol. 29, no. 5, p. 1565, 2009.

[24] Y. Miyawaki, H. Uchida, O. Yamashita, M. Sato, Y. Morito, H. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, no. 5, pp. 915–929, 2008.

[25] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns," *Neuroimage*, vol. 42, no. 4, pp. 1414–1429, 2008.

[26] N. Kriegeskorte, M. Mur, D. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. Bandettini, "Matching categorical object representations in inferior temporal cortex of man and monkey," *Neuron*, vol. 60, no. 6, pp. 1126–1141, 2008.

[27] O. de Beeck and P. Hans, "Probing the mysterious underpinnings of multi-voxel fMRI analyses," *NeuroImage*, 2009.

[28] ——, "Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses?" *NeuroImage*, vol. 49, no. 3, pp. 1943–1948, 2010.

[29] Y. Kamitani and Y. Sawahata, "Spatial smoothing hurts localization but not information: Pitfalls for brain mappers," *NeuroImage*, 2009.

[30] Y. Liu, K. Wang, C. Yu, Y. He, Y. Zhou, M. Liang, L. Wang, and T. Jiang, "Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: A review of resting-state fMRI studies," *Neuropsychologia*, vol. 46, no. 6, pp. 1648–1656, 2008.

[31] L. Uddin, A. Kelly, B. Biswal, D. Margulies, Z. Shehzad, D. Shaw, M. Ghaffari, J. Rotrosen, L. Adler, F. Castellanos *et al.*, "Network homogeneity reveals decreased integrity of default-mode network in ADHD," *Journal of neuroscience methods*, vol. 169, no. 1, pp. 249–254, 2008.

[32] D. Saur, O. Ronneberger, D. Kummerer, I. Mader, C. Weiller, and S. Kloppel, "Early functional magnetic resonance imaging activations predict language outcome after stroke," *Brain*, 2010.