

A comparison of random field theory and permutation methods for the statistical analysis of MEG data

Dimitrios Pantazis,^a Thomas E. Nichols,^b Sylvain Baillet,^c and Richard M. Leahy^{a,*}

^aSignal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564, USA

^bDepartment of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA

^cNeurosciences Cognitives and Imagerie Cerebrale CNRS UPR640-LENA, Hospital de la Salpetriere, Paris, France

Received 1 March 2004; revised 11 June 2004; accepted 27 September 2004

We describe the use of random field and permutation methods to detect activation in cortically constrained maps of current density computed from MEG data. The methods are applicable to any inverse imaging method that maps event-related MEG to a coregistered cortical surface. These approaches also extend directly to images computed from event-related EEG data. We determine statistical thresholds that control the familywise error rate (FWER) across space or across both space and time. Both random field and permutation methods use the distribution of the maximum statistic under the null hypothesis to find FWER thresholds. The former methods make assumptions on the distribution and smoothness of the data and use approximate analytical solutions, the latter resample the data and rely on empirical distributions. Both methods account for spatial and temporal correlation in the cortical maps. Unlike previous nonparametric work in neuroimaging, we address the problem of nonuniform specificity that can arise without a Gaussianity assumption. We compare and evaluate the methods on simulated data and experimental data from a somatosensory-evoked response study. We find that the random field methods are conservative with or without smoothing, though with a 5 vertex FWHM smoothness, they are close to exact. Our permutation methods demonstrated exact specificity in simulation studies. In real data, the permutation method was not as sensitive as the RF method, although this could be due to violations of the random field theory assumptions.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Permutation; Random field; MEG; EEG; Familywise error rate

Introduction

Magnetoencephalography (MEG) is used to image electrical activity in the brain. Clusters of thousands of synchronously activated pyramidal cortical neurons are believed to be the main

generators of MEG signals. In particular, the currents associated with their large dendritic trunks, which are locally oriented in parallel and perpendicular to the cortical surface, are the primary source of the neuromagnetic fields outside the head (Hämäläinen et al., 1993). Imaging approaches to the MEG inverse problem exploit this concept by restricting the reconstruction to elemental sources (dipoles) oriented normally to the cortical surface (Dale and Serano, 1993). Consequently, a commonly used approach extracts a tessellated representation of the cerebral cortex from a coregistered MR image and solves the inverse problem for a current dipole located at each vertex of the tessellated surface. Since the position and orientation of the dipoles is fixed, image reconstruction is a linear problem and can be solved using standard techniques (Baillet et al., 2001; Hämäläinen et al., 1993; Katila and Karp, 1983; Phillips et al., 1997). However, the highly convoluted nature of the human cortex requires the use of many thousands of dipoles for an accurate representation of the cortical surface. The inverse problem becomes hugely under-determined and the resulting current density maps (CDMs) are of low resolution; interpretation is further confounded by the presence of additive noise exhibiting a highly nonuniform spatial correlation.

As with fMRI images, objective assessment of CDMs requires a principled approach to identifying regions of activation. The analysis of CDMs involves testing thousands of hypothesis (one per surface element) for statistically significant experimental effects. This raises the possibility of large numbers of false-positives simply as a result of multiple hypothesis testing. To effectively control the number of false-positives over all tests, we must therefore consider the multiple hypothesis-testing problem. Many false-positive measures have been proposed in this context, including familywise error rate, expected false discovery rate, per-comparison error rate and per-family error rate (Nichols and Hayasaka, 2003). The standard approach, and the one investigated in this paper, is to control the Familywise Error Rate (FWER), i.e., the chance of one or more false-positives under the null hypothesis.

The simplest approach to controlling the FWER is the Bonferroni correction method (Hochberg and Tamhane, 1987;

* Corresponding author.

E-mail address: leahy@sipi.usc.edu (R.M. Leahy).

Available online on ScienceDirect (www.sciencedirect.com).

Nichols and Hayasaka, 2003). This method produces conservative thresholds unless the tests are independent, a case that is rarely true in neuroimaging experiments and certainly not for the smooth images reconstructed from MEG data. Other methods that consider the spatial dependence of the data make inferences based on the global maximum distribution. The FWER is directly related to the maximum statistic; one or more voxels T_i will exceed the threshold u under the null hypothesis H_0 only if the maximum exceeds the threshold:

$$\begin{aligned} P(\text{FWER}) &= P(\cup_i T_i > u | H_0) = P\left(\max_i T_i > u | H_0\right) \\ &= 1 - F_{\max T | H_0}(u) = 1 - (1 - \alpha) = \alpha \end{aligned} \quad (1)$$

Therefore, we can control the FWER at level α , if we choose the threshold u to be in the $(1 - \alpha)$ 100th percentile of the maximum distribution.

Random Field (RF) theory methods approximate the upper tail of the maximum distribution F_{\max} using the expected value of the Euler characteristic of the thresholded image (Worsley et al., 1996). They are implemented in various software packages (SPM-<http://www.fil.ion.ucl.ac.uk>, VoxBo-<http://www.voxbo.org>, and FSL-<http://www.fmrib.ox.ac.uk/fsl> among others), and are typically used in PET and fMRI studies. However, RF theory relies on several assumptions including the following: the image has the same parametric distribution at each spatial location, the point spread function has two derivatives at the origin, sufficient smoothness to justify application of continuous RF theory, and a sufficiently high threshold for the asymptotic results to be accurate.

Resampling methods are a different approach to controlling the FWER that exploit the information contained in the data to estimate the empirical distribution of the maximum statistic. They do not assume parametric distributions, they adapt to underlying correlation patterns in the data, and are now computationally feasible. The two main categories are bootstrap-based, which allow for a general modeling framework, and permutation-based, which require some knowledge of exchangeability conditions under the null hypothesis. Here, we only consider permutation methods since they are exact, that is, they give precise control of the FWER, while bootstrap methods are only asymptotically exact. Furthermore, the permutation approach relies on a less restrictive exchangeability condition than the requirement in the bootstrap that samples are independent and identically distributed.

Permutation and RF theory methods have been applied widely in functional (Andrade et al., 2000; Nichols and Holmes, 2001; Worsley et al., 1992, 1996) and structural (Bullmore et al., 1999; Chung, 2001; Pantazis et al., 2004; Sowell et al., 1999a,b; Thompson et al., 2001, 2003) brain imaging. However, until recently, error rate control in MEG experiments has drawn little attention. Dale et al. (2000) normalized the CDMs using an estimate of the background noise variance at each cortical element. These normalized images follow a t distribution under the null hypothesis of Gaussian background noise. Thresholding of the resulting statistical maps was then used to detect significant activation, however, the multiple comparisons problem was not addressed. Barnes and Hillbrand (2003) presented an application of RF theory to MEG data but their method is specifically tailored to beamforming solutions rather than the general linear inverse

methods. Carbonell et al. (2004) used Hotelling's T^2 random fields to localize significant MEG/EEG activation in time, and then t statistics to achieve spatial localization. Permutation tests were applied by Blair and Karnisky (1994) for the analysis of EEG data as recorded on an array of electrodes, and by Pantazis et al. (2003) for the analysis of MEG data in reconstructed cortical maps of brain activation. An alternative permutation scheme, proposed by Singh et al. (2003), detects event-related synchronization or desynchronization components in an MEG study involving visual stimulation and a Linearly Constraint Minimum Variance (LCMV) beamformer applied to data decomposed into multiple frequency bands. The current work presents a novel general RF theory-based method to control FWER in MEG. Also, it extends the results in Pantazis et al. (2003) to extract thresholds for each time-point. Finally, we compare RF theory and permutation methods in terms of specificity, sensitivity, and possible limitations.

Methods

Our goal is to detect spatial and temporal regions of significant activity in MEG-based cortical maps while controlling familywise error rate. The methods to do this that we describe below also apply directly to cortical maps computed from EEG data, since the inverse imaging methods differ only in the forms of their lead field matrices (Baillet et al., 2001). In this section, we first describe our MEG data model. We then present two methods, the first based on RF theory and the second on permutations, for controlling the error rate in MEG experiments.

Model

We assume that MEG data are collected as a set of J stimulus-locked event-related epochs (one per stimulus repetition), each consisting of a pre- and post-stimulus interval. Each epoch consists of an array of data M ($n_{\text{channels}} \times n_{\text{timepoints}}$) representing the measured magnetic field at each sensor as a function of time. The measurements M are linearly related with the brain activation X ($n_{\text{sources}} \times n_{\text{timepoints}}$) as:

$$M = GX + N \quad (2)$$

where G ($n_{\text{channels}} \times n_{\text{sources}}$) is the forward operator and N represents additive noise in the channel measurements. The lead field matrix G is dependent on the shape and conductivity of the head and can be computed using a simplified spherical head model, or more accurately using boundary or finite element methods that account for the true shape and conductivity within the head (Baillet et al., 2001; Mosher et al., 1999).

A cortical map can be computed for each epoch by applying a linear inverse method to produce an estimate of the temporal activity at each surface element i in the cortex. We write the reconstructed cortical maps as $\{X_{ij}\}$ where $i = 1, \dots, S$, $t = -N_0 + 1, \dots, N$, and $j = 1, \dots, J$ are indices in space, time, and epoch, respectively. We let $t = 1$ correspond to the stimulus event time so that there are N_0 pre-stimulus time points. We use the pre-stim data to estimate the baseline mean $\hat{\mu}_i$ and standard deviation $\hat{\sigma}_i$ at each spatial element i . We model the centered data $\equiv Y_{ij} = X_{ij} - \hat{\mu}_i$ as

$$Y_{ij} = \mu_{it} + \epsilon_{ij}, \quad t = 1, \dots, N \quad (3)$$

where μ_{it} is the spatiotemporal profile, and ϵ_{ij} is the zero mean random error. We assume independence across epochs, but not over time and space; that is, that ϵ_{ij} is independent of $\epsilon_{ij'}$ for $j \neq j'$ and all i, t .

We take the standard massively univariate approach and model each spatial location independently. For each spatial location, the model then amounts to a one-way ANOVA model with repeated measures. The standard linear modeling approach would be to find estimates $\hat{\mu}_{it}$ and test

$$H_0 : \{\mu_{it} = 0, t = 1, \dots, N\}, \quad (4)$$

for each location i . Estimation and inference on this hypothesis is complicated by the temporal dependence, as the optimal estimates require the temporal whitening of the data (and model). The temporal dependence can be difficult to estimate, and whitening with an inaccurate covariance can bias variance estimates and even increase the variance of mean estimates (Friston et al., 2000). Inference on the nulls in Eq. (4) with F statistics is also challenging. Without whitening, the null F distributions will be incorrect (though conservative approximations are available (Greenhouse and Geisser, 1959), while whitening with estimated covariances yields only approximately correct inferences.

Hence, we depart from standard statistical approaches in three important ways. First, instead of whitening, we use ordinary least squares with inference methods that are valid despite the dependence (the case for permutation), or that require only low-dimensional characterization of the dependence (the case for RF theory). While this implies that our parameter estimates do not have minimum variance, it makes our methods general and easy to implement. In particular, since we expect no experimental effect over epochs, the spatiotemporal profiles are estimated as the appropriate average over j in Eq. (5):

$$\hat{\mu}_{it} = \bar{Y}_{it\cdot}, \quad (5)$$

where the bar indicates an average over the dotted subscript.

Second, instead of testing the temporal-omnibus null hypothesis (Eq. (4)) for each location i , we test

$$H_0 : \mu_{it} = 0, \quad (6)$$

for each $t > 0$ and i . This allows temporal as well as spatial localization. The test t statistic for Eq. (6) is

$$T_{it} = \frac{\hat{\mu}_{it}}{\hat{\sigma}_i / \sqrt{J}}. \quad (7)$$

Lastly, we use the pre-stimulus variance estimate $\hat{\sigma}_i$ instead of a residual variance estimator. We may do this because we expect no systematic response in pre-stimulus time. We choose to do this since it is likely that the stimulus will increase the variance in the post-stimulus time period in a heterogeneous fashion; in the presence of a response, the residual variance estimator would overestimate the background noise level.

Random field theory method

Random field methods use the topology of the statistic image to estimate the maximum distribution. In this section, we describe their application to surface reconstructed MEG data.

Assumptions

Gaussian random field methods assume that the statistic image, or equivalently, the normalized errors, approximate a Gaussian random field. Based on our modeling, to control FWER over space, it is required that for each t and j the spatial process of the normalized errors

$$\left\{ \frac{\epsilon_{ij}}{\sqrt{\text{Var}(\epsilon_{ij})}} \right\}_i \quad (8)$$

is a sampled version of a continuous, mean zero, unit variance Gaussian random field. To control FWER over space and time,

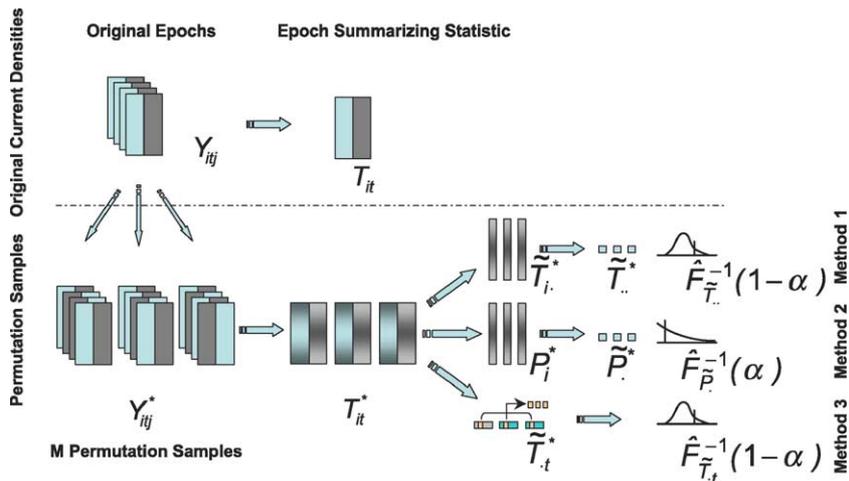


Fig. 1. Illustration of the summarizing procedures used to construct FWER-corrected thresholds: the original epochs Y_{ij} produce M permutation samples Y_{ij}^* by exchanging pre- and post-stimulus data. The epochs are then averaged and normalized to produce T_{it} and T_{it}^* . Three different summarizing methods can be used. Method 1: T_{it}^* are summarized in time ($\tilde{T}_{i\cdot}^*$) and space ($\tilde{T}_{\cdot t}^*$) to produce epochwise thresholds ($\hat{F}_{\tilde{T}_{i\cdot}^*}^{-1}(1-\alpha)$); method 2: T_{it}^* are summarized in time, converted into P -values (P_i^*) and summarized in space (\tilde{P}_i^*) to produce uniform specificity epochwise thresholds ($\hat{F}_{\tilde{P}_i^*}^{-1}(\alpha)$); method 3: T_{it}^* are summarized in space ($\tilde{T}_{\cdot t}^*$) to produce a threshold ($\hat{F}_{\tilde{T}_{\cdot t}^*}^{-1}(1-\alpha)$) for each time t .

Table 1
Summary statistics for three permutation methods

| | Time-summarizing | Space-summarizing |
|----------|--|--|
| Method 1 | $\tilde{T}_{it}^* = \max_{t > 0} T_{it}^* $ | $\tilde{T}_{i\cdot}^* = \max_i \tilde{T}_{i\cdot}^*$ |
| Method 2 | $P_i^* = p_i(\tilde{T}_{i\cdot}^*)$ | $\tilde{P}_{i\cdot}^* = \min_i \tilde{P}_{i\cdot}^*$ |
| Method 3 | | $\tilde{T}_{i\cdot}^* = \max_i T_{it}^* $ |

The permutation samples are T_{it}^* , with i the spatial index, and t the time index. The tilde indicates the maximum over the dotted subscript; $p_i(\cdot)$ is the permutation P -value function using only data from spatial location i .

the spatiotemporal process of normalized errors must satisfy this condition. While the core RF theory results also assume stationarity, there have been recent developments that relax this assumption as described below.

When the degrees of freedom are high, the t random fields will be well approximated by a Gaussian random field; Worsley et al. (1995) offers a rule of thumb of 120 degrees-of-freedom as being sufficient. Our t statistic T_{it} has $N_0 - 1$ degrees of freedom. As the number of epochs (J) grows, the central limit theorem implies that $\hat{\mu}_{it}$ will be Gaussian, even if ϵ_{ij} is not. Hence, if both N_0 and J are large, the random field methods should have their distributional assumptions satisfied.

Spatial and temporal smoothing

To apply the continuous RF theory results, the statistic images must be sufficiently smooth, since failure to satisfy this assumption typically leads to conservative thresholds. Nichols and Hayasaka (2003) found that 3 voxels FWHM smoothness is sufficient for 3-dimensional Gaussian images; more smoothness was required for low degrees-of-freedom t images. The smoothness required for surface reconstructed CDMs is unknown. The highly convoluted cerebral cortex causes T_{it} to be rather rough across i , especially on gyral crests and sulcal fundi where neighboring vertices have rapidly changing orientations. There-

fore, we consider spatially smoothing $\hat{\mu}_{it}$ to avoid conservative thresholds.

Smoothing in Euclidean space is typically performed by convolving the image $X(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^n$ with a Gaussian kernel of FWHM = $4(\ln 2)^{1/2} \sqrt{t}$ as follows:

$$X'(\mathbf{z}, t) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-(\mathbf{z}-\mathbf{y})^2/4t} X(\mathbf{y}) d\mathbf{y} \quad (9)$$

where $X'(\mathbf{z}, t)$ is the smoothed image at time t . Here, the time variable t determines the degree of smoothing by analogy to an isotropic diffusion process. This equation cannot be applied directly to cortical surfaces. To apply smoothing on arbitrary curved surfaces, we can reformulate Gaussian kernel smoothing as a solution of a diffusion equation on a Riemannian manifold (Chung, 2001):

$$\frac{\partial X'}{\partial t} = \Delta X' \quad (10)$$

with the initial condition $X'(\mathbf{z}, 0) = X(\mathbf{z})$. For an n -dimensional Euclidean space, $\Delta F = \partial^2 F / \partial x_1^2 + \dots + \partial^2 F / \partial x_n^2$; for an arbitrary Riemannian manifold (such as a cortical surface), Δ is called the Laplace–Beltrami operator and is given by:

$$\Delta F = \sum_{i,j} \frac{1}{|g|^{1/2}} \frac{\partial}{\partial u^i} \left(|g|^{1/2} g^{ij} \frac{\partial F}{\partial u^j} \right) \quad (11)$$

where $\mathbf{g} = (g_{ij})$ is a 2×2 matrix whose coefficients are the Riemannian metric tensors, u^1 and u^2 are the coefficients of a local parameterization of the cortical surface, and $(g^{ij}) = \mathbf{g}^{-1}$ (Arfken, 2000; Boothby, 2002).

If control of FWER's over time and space is desired, then smoothing in time may also be required. This can be accomplished with initial bandpass filtering of the data, or with temporal smoothing of the CDMs. One problem with temporal smoothing is

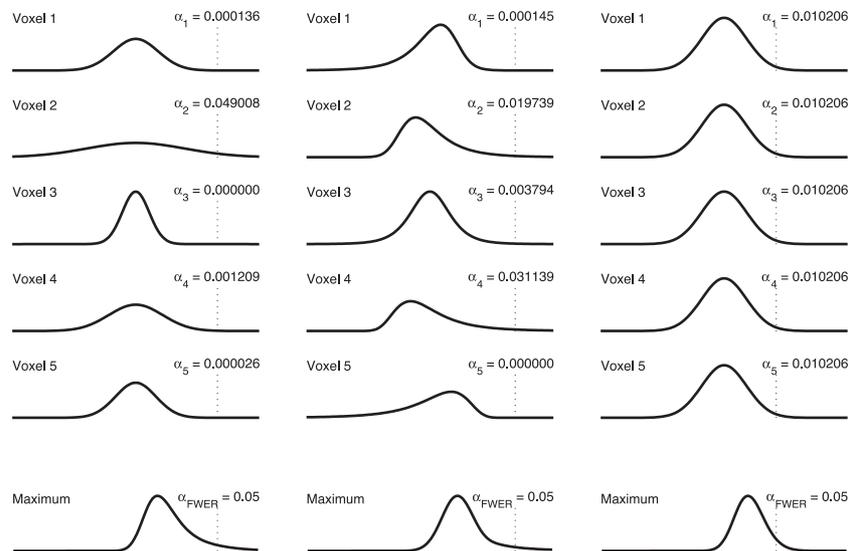


Fig. 2. Illustration of the impact of heterogeneous voxel null distributions on a 5% FWER threshold. Shown are null distributions of five surface elements and the distribution of their maximum in three cases: each having different variances, different skewed distributions, and one common distribution. The first case (left) demonstrates the variable false-positive rate when test statistics are not normalized (e.g., use raw CDM values $\hat{\mu}_{ij}$ instead of T_{ij}). The second case (center) demonstrates the impact of non-Gaussianity, even when variance is normalized, and motivates the use of P values to normalize T_{ij} into $p_i(T_{ij})$. The last case (right) shows that with homogeneous null distributions the false-positive rate at each surface element is homogeneous. Note that in all cases FWER is controlled at 5%.

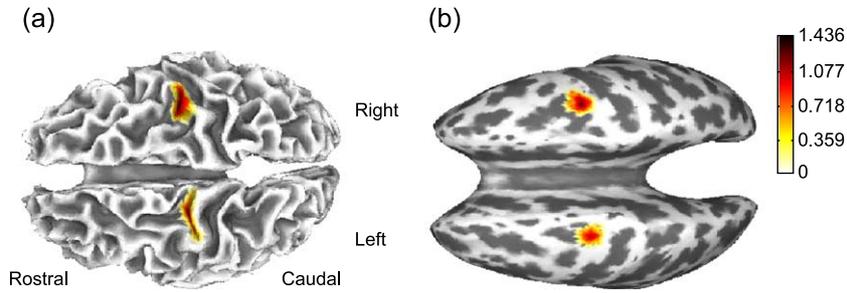


Fig. 3. Simulated source 1 (left hemisphere) and source 2 (right hemisphere) are shown mapped onto high resolution and smoothed versions of the cortical surface.

that the usual estimate of σ_i is only unbiased under independence; when temporal samples are correlated, a bias correction for $\hat{\sigma}_i$ can be found using the Saitherwaite correction (Satterthwaite (1946); Worsley and Friston (1995)).

Estimating the maximum distribution

As shown in Eq. (1), the FWER can be determined directly from the maximum distribution; specifically, the probability of the maximum exceeding u gives the FWER for threshold u . Adler (1981) demonstrated that the expected value of the Euler Characteristic is a good approximation of this probability when u is large. Worsley et al. (1996) provides an intuitive formula that unifies the results for all types of random fields:

$$P\left(\max_i T_{it} \geq u\right) \approx \sum_{d=0}^2 R_d(S) \rho_d(u) \tag{12}$$

This equation gives the FWER P -value for threshold u in a 2-dimensional random field (T_{it} for fixed t) in a search region S (cortical surface). The term $R_d(S)$ is the d -dimensional RESEL count, a unitless quantity that depends only on topological features of the CDMs in the search region S . The $\rho_d(u)$ term is the Euler Characteristic density that depends only on the threshold u and the type of statistical field (such as z , t , X^2 , and Hotelling’s T^2). In Eq. (12), the summation of the lower dimensional terms ($d = 0$ and $d = 1$) compensate for the case when the excursion set, that is, the regions of voxels in a field above a threshold u , touches the boundary.

It is straightforward to extend these results to a 3-dimensional random field that includes the time dimension (Worsley et al.,

1996). Let τ denote the number of RESELS computed in the time dimension, then:

$$P\left(\max_{it} T_{it} \geq u\right) \approx \sum_{d=0}^2 R_d(S) [\tau \rho_{d+1}(u) + \rho_d(u)] \tag{13}$$

In MEG we are interested in positive or negative deflections of current density, and hence make inferences on the absolute values $|T_{it}|$ using two-side FWER P -values:

$$P\left(\max_i |T_{it}| \geq u\right) \approx 2R_2(S) \rho_2(u) \tag{14}$$

for control of FWER over space only, and

$$P\left(\max_{it} |T_{it}| \geq u\right) \approx 2R_2(S) [\tau \rho_3(u) + \rho_2(u)] \tag{15}$$

for control of FWER over space and time. Note that we have dropped the $d = 0$ and $d = 1$ terms, as the cortical surface has a spherical topology and no edge. In the Appendix we give equations for $\rho_2(u)$, $\rho_3(u)$ and estimates for $R_2(S)$ and τ based on pre-stimulus data.

Permutation method

We use resampling methods to allow more flexible models and to avoid random field theory assumptions. The standard approach to permutation tests is to find units exchangeable under the null hypothesis. While epochs can be regarded as independent, permuting of epochs does not change the value of epoch-averaged statistics (although the bootstrap can make use of the epoch replicates). By collecting an equal duration of pre- and post-stimulus data ($N = N_0$), we can permute pre- and post-stim data since these intervals are interchangeable under the null hypothesis of no activation at any post-stimulus timepoint. Given J original epochs Y_{ij} , $j = \{1, \dots, J\}$, we can create $M \leq 2^J$ permutation samples Y_{ij}^* , each consisting of J new epochs (Fig. 1). The symbol (*) indicates that the values Y_{ij}^* are created by permutation.

Table 2
Thresholds for the sources illustrated in Figs. 3 and 4 for controlling FWER over space and time

| | Nominal FWER | Unsmoothed CDMs | Smoothed CDMs |
|----------------------|--------------|-----------------|---------------|
| Permutation method 1 | 0.050 | 5.236 | 5.151 |
| Permutation method 1 | 0.123 | 5.050 | – |
| Permutation method 2 | 0.123 | 4.998 | – |

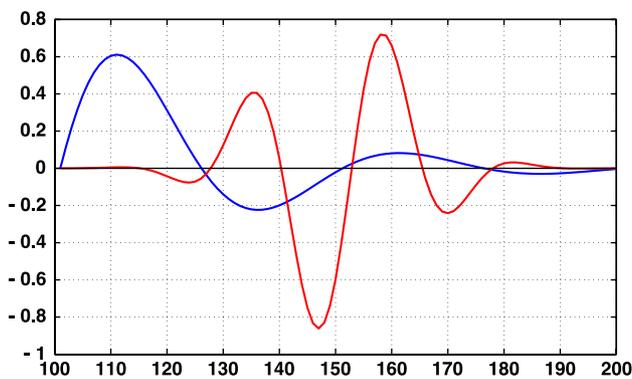


Fig. 4. Time-courses of simulated sources, blue for source 1 and red for source 2. The pattern of activation mimics a typical neuroimaging study where an early response to a stimulus propagates to another brain region giving a delayed component.

Table 3
Thresholds for the sources illustrated in Figs. 3 and 4 for controlling FWER over space only

| | Nominal FWER | Unsmoothed CDMs | Smoothed CDMs |
|----------------------|-----------------|--------------------|------------------|
| Permutation method 3 | 0.050 | 3.876 | 3.778 |
| Random field method | 0.050 | 4.451 | 4.076 |

Assumptions

Except in the special case of data from randomized experiments, permutation methods are not assumption-free. In order to permute the data under the null hypothesis, we must assume that the distribution of the errors is exchangeable with respect to epochs. That is, the multivariate distribution of $\{\epsilon_{ij}\}$ is invariant with respect to permutations of indices j , but is otherwise unspecified; independence across epochs is sufficient to satisfy this condition. We further assume that under the null hypothesis, the distribution of $\{\epsilon_{ij}\}$ is unaltered by exchanging the pre- and post-stim data (we require that $N = N_0$).

Permutation statistics

Both permutation and RF theory approaches use the maximum distribution to control the FWER over space for one time point, or over time and space simultaneously. To obtain thresholds that control FWER over space at a given time t , we compute the permutation distribution of the spatial maximum:

$$\tilde{T}_t^* = \max_i |T_{it}^*| \quad (16)$$

where the tilde indicates an extremum over the dotted subscript; here, for statistic values, this is a maximum. Later, when used with P -values, it will be a minimum. These M permutation samples \tilde{T}_t^* estimate the null distribution of \tilde{T}_t , written as $\hat{F}_{\tilde{T}_t}^{-1}$. The level α spatial FWER threshold is $\hat{F}_{\tilde{T}_t}^{-1}(1 - \alpha)$.

Thresholds that control FWER over time at a given spatial location i , or that control FWER over both time and space, are found by the corresponding permutation distribution:

$$\begin{aligned} \tilde{T}_i^* &= \max_{t>0} |T_{it}^*| \\ \tilde{T}_{::}^* &= \max_i \tilde{T}_i^* \end{aligned} \quad (17)$$

The $\tilde{T}_{::}^*$ are used to estimate the null distribution of $\tilde{T}_{::}$, written as $\hat{F}_{\tilde{T}_{::}}^{-1}$. The level α FWER threshold is $\hat{F}_{\tilde{T}_{::}}^{-1}(1 - \alpha)$. The above procedures are summarized as methods 1 and 3 in Table 1.

Achieving uniform spatial specificity

Permutation tests are always valid given the assumption of exchangeability under the null hypothesis. However, if the null distribution varies across space, different surface elements will have differing specificity, as illustrated in Fig. 2. To address the problem, we can perform an element by element normalization by converting the statistic values at each location to P -values, which are computed using permutations. We then control FWER with respect to the distribution of the minima of these P -values, rather than the maxima of the original statistic. In this way, we achieve uniform spatial specificity.

The method to achieve uniform specificity proceeds as follows. We first summarize by computing the maximum over time for each permutation sample at each location, to generate \tilde{T}_i^* as defined in Eq. (17). We then use these permutation statistics to estimate the null distribution $\hat{F}_{\tilde{T}_i}$ of \tilde{T}_i . Using this distribution, we then replace each permutation statistics \tilde{T}_i^* with its P -value:

$$P_i^* = p_i(\tilde{T}_i^*) = 1 - \hat{F}_{\tilde{T}_i}(\tilde{T}_i^*) \quad (18)$$

Effectively, this is equivalent to counting how many of the M permutation sample statistics \tilde{T}_i^* at location i are greater than or equal to the current permutation statistic, and dividing by M . For each i , the P_i^* have a uniform distribution under the null hypothesis, and hence are self-normalized. To find a threshold on the P -values that controls FWER, we first find the minimum P -value for each permutation sample:

$$\tilde{P}^* = \min_i \tilde{P}_i^* \quad (19)$$

We then use the M permutation values \tilde{P}^* to estimate the null distribution, $\hat{F}_{\tilde{P}^*}$, of \tilde{P}^* . By choosing a level α threshold on $\hat{F}_{\tilde{P}^*}$, that is, $(\hat{F}_{\tilde{P}^*}^{-1}(\alpha))$, and retaining only those locations at which the P -value is less than this threshold, we control the FWER at α . Furthermore, since under the null hypothesis the P -values will have the same uniform distribution at each location, this approach will guarantee uniform specificity. The above procedure is summarized in Table 1 (method 2).

One practical problem of this approach is the discreteness of the P -values P_i^* , which in turn causes \tilde{P}^* to be discrete. If many \tilde{P}^* have the smallest possible value ($1/M$), then small α level thresholds may be unattainable. For example, one Monte Carlo experiment with $M = 1000$ found that 30% of the permutations had a minimum P_i^* of value 0.001 and hence the smallest possible FWER threshold corresponded to $\alpha = 0.3$. Therefore, the P -value normalization approach, while it makes no assumptions on differing shapes of the local distributions, requires many permutations.

Thresholds

The three different permutation methods produce different types of thresholds. Method 1 produces a single statistic value threshold for all time and space: $\hat{F}_{\tilde{T}_{::}}^{-1}(1 - \alpha)$. Method 2 also produces a single P -value threshold for all time and space: $\hat{F}_{\tilde{P}^*}^{-1}(\alpha)$. This P -value can be re-expressed in terms of statistic values, which will yield different thresholds for each spatial location: $p_i^{-1}(\hat{F}_{\tilde{P}^*}^{-1}(\alpha))$. Method 3 produces one threshold for each time point: $\hat{F}_{\tilde{T}_t}^{-1}(1 - \alpha)$. The key difference between these thresholds is that methods 1 and 2 control FWER over all time and space, while method 3 controls FWER over space at a single time point.

Results

In this section, we evaluate the random field method and the three permutation methods given in Table 1 in terms of specificity, i.e., the ability of the methods to control false-positives, and sensitivity, a measure of how well the method can detect and localize true brain activation.

Simulation studies

A cortical surface was extracted from an MRI scan using BrainSuite, a brain surface extraction tool (Shattuck and Leahy,

¹ F_X denotes the cumulative density function (CDF) of the random variable X .

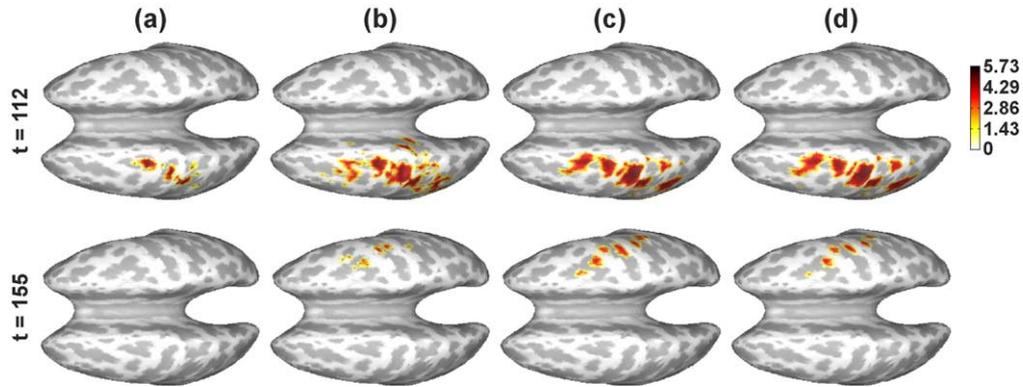


Fig. 5. Examples of significant activation maps for permutation and random field methods for two time instances, (a) permutation method 1 using unsmoothed CDMs, (b) permutation method 3 using unsmoothed CDMs, (c) permutation method 3 using smoothed CDMs, (d) random field using smoothed CDMs. The first method controls FWER over space and time, while the last three methods control FWER over space for one time point only.

2002). The surface extraction algorithm consists of skull and scalp removal, nonuniformity correction, tissue classification, and cortical surface topology correction. The surfaces produced by BrainSuite can be constrained to be topologically spherical, and are suitable for use in current source localization and visualization. The surfaces were then coregistered to the MEG sensor arrangement of a 151 channel CTF Systems Inc. Omega system using an affine transformation, based on three fiducial points (nasion, right/left preauricular). The resulting surface, which contained approximately 520,000 faces, was downsampled to produce a 15,000 face (7481 vertices) surface suitable for reconstruction purposes. CDMs were mapped back onto the higher resolution surface, as illustrated in Fig. 3a, as well as onto a smoothed version of the surface to assist in visualization, Fig. 3b. An orientation constraint was applied during reconstruction, using surface normals estimated from the original dense cortical surface. The forward model was calculated using an overlapping spheres model (Huang et al., 1998). We used a Tikhonov regularized linear inverse method (Tikhonov and Arsenin, 1977) with regularization parameter $\lambda = 4 \cdot 10^{-7}$ (Baillet et al., 2001).

Source simulation

We simulated two sources, one each in left and right primary motor cortex, as shown in Fig. 3. Each source consists of an

activated patch approximately 2 cm² in size. The time-courses simulate early and delayed responses to a stimulus (Fig. 4), which is a typical pattern in neuroimaging studies.

A total of 100 stimulus-locked event-related epochs (or trials) were simulated, each one consisting of 100 pre-stimulus and 100 post-stimulus time points. Each epoch is an array of data (151 channels \times 200 timepoints) representing the measured magnetic field at each sensor as a function of time. Gaussian i.i.d. noise with power 2000 times the averaged power of the signal was added to the channel measurements, making the reconstruction a difficult task. We then applied the Tikhonov regularized inverse operator (Tikhonov and Arsenin, 1977) to produce CDMs X_{ij} . We also generated spatially smoothed versions of the CDMs using diffusion smoothing as described in the RF theory section. We used an approximation of the Laplace–Beltrami operator given in Oostendorp and Oosterom (1988) with spatial filtering corresponded to a 23.5 mm FWHM. Since the mean distance between the vertices in our tessellated cortical surface was 4.5 mm, the spatial filtering was equivalent to approximately 5 vertices FWHM. Finally, we created $M = 1000$ permutation samples for permutation methods 1 and 3, and $M = 10,000$ permutation samples for permutation method 2 due to the discreteness concerns discussed above.

We tested smoothed and unsmoothed CDMs for significant activation using the random field and permutation methods. Spatiotemporal FWER-corrected thresholds (Table 2) are higher, as expected, than spatial FWER-corrected thresholds (Table 3).

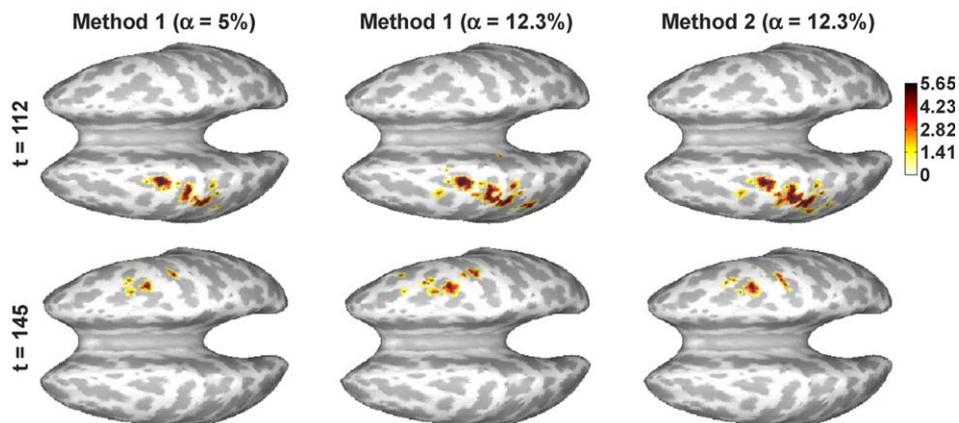


Fig. 6. Examples of significant activation maps for permutation methods 1 and 2 for two time instances. The lowest achieved threshold for method 2 is $\alpha = 0.123$ and the comparison for both methods is done at this level.

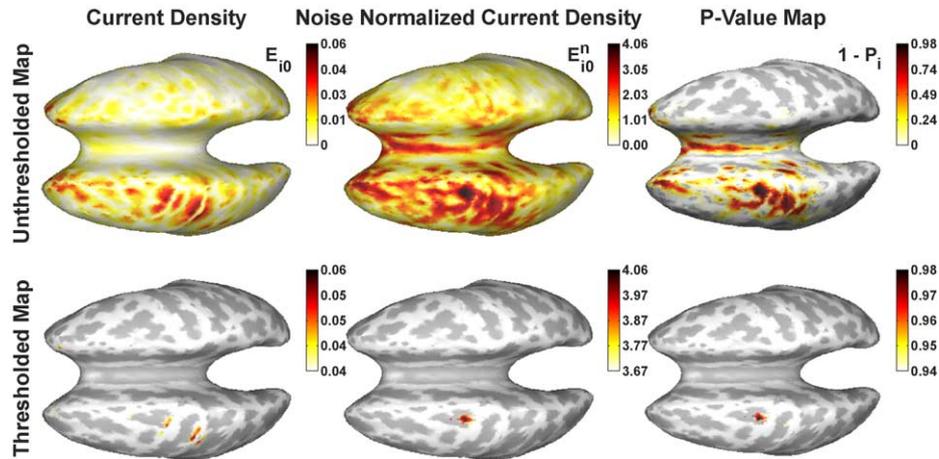


Fig. 7. Thresholded and unthresholded maps of the current density ($\hat{\mu}_{it}$), the noise normalized current density (T_{it}) and the $(1 - P)$ value map $p_i(T_{it})$ at $t = 113$.

The thresholds for permutation method 2 are given only for unsmoothed CDMs, because of the high computational cost of smoothing 10,000 permutations, and represent the average statistic value thresholds over all spatial locations, $E_i\{p_i^{-1}(\hat{F}_{\hat{P}}^{-1}(\alpha))\}$. For method 2, the smallest false positive rate possible with the empirical distribution of \hat{P} was 12.3%. Therefore, in order to compare methods 1 and 2, we repeated the analysis for method 1 with $\alpha = 12.3\%$ as listed in Table 2.

The design of the statistics T_{it} , involving post-stimulus data μ_{it} and pre-stimulus data σ_i , causes permutation methods to adapt to the activation pattern of the sources on the cortical surface. In particular, when a stimulus activates a source on the cortical surface, the corresponding time-course has higher amplitude and variability in the post-stimulus area than in the pre-stimulus area. As a result, the original data will have relatively high T_{it} (high μ_{it} and small σ_i) as compared to the permutation samples; as a source gets stronger, the significance of the original data increases. In our simulation experiments, the time-courses of the sources vary with time; therefore, we expect the FWER-corrected thresholds to adapt to this variation.

For permutation method 3, we estimate a different threshold $\hat{F}_{\hat{P}_i}^{-1}(1 - \alpha)$ for each timeslice after the stimulus $t > 0$ and Table 3 shows the mean of these thresholds (3.876 for unsmoothed CDMs and 3.778 for smoothed CDMs). The random field method is less sensitive with higher thresholds (4.451 and 4.076, respectively) and particularly conservative when unsmoothed CDMs are used.

Table 4
Noise-only simulation results for control of spatial and spatiotemporal FWER at nominal level $\alpha = 5\%$

| | Unsmoothed CDMs | | Smoothed CDMs | |
|------------------------------------|-----------------|---------------|---------------|---------------|
| | Threshold | Observed FWER | Threshold | Observed FWER |
| <i>Spatiotemporal FWER methods</i> | | | | |
| Permutation method 1 | 5.350 | 0.0600 | 5.245 | – |
| <i>Spatial FWER methods</i> | | | | |
| Permutation method 3 | 4.059 | 0.0480 | 3.980 | – |
| Random field method | 4.453 | 0.0139 | 4.081 | 0.0340 |

The Monte Carlo standard error for the spatiotemporal FWER is 0.0218; for the spatial FWER, it is 0.0022.

Fig. 5 shows examples of significant activation maps for the permutation and random field methods. Epochwise thresholds (Fig. 5a) are more stringent and thus less sensitive to signals than slice-wise thresholds (Fig. 5b). We are, however, more confident that a signal is truly present because the false-positives are controlled over all time slices. Spatial smoothing (Fig. 5c) produced a mild loss in resolution compared to the unsmoothed case (Fig. 5b). Finally, comparison of the permutation and random field results for the smoothed CDMs (Figs. 5c and d, respectively) indicates similar performance in this simulation.

We should comment here that permutation and random field tests do not address the limited resolution of MEG reconstruction methods. The MEG inverse problem is ill-posed and CDMs are of low resolution and tend to mislocalize source activation. If the inverse method identifies experimental variation in some region, permutation and random field tests will identify these regions regardless of the presence of an actual source at those locations. In most cases, CDM reconstructions of activation from a single sulcus or gyrus will tend to show activation in neighboring sulci or gyri, respectively. It is quite possible, as shown in Fig. 5, that reconstructions of activation in a single cortical area will exceed the determined threshold in multiple areas, and thus particular care is required in interpretation of CDMs, even after thresholding to control for FWER.

It is interesting to study the effect of the P -value transformation on the significant activation maps: if the effect is significant, it is an indication that Gaussianity assumptions, used by random fields and permutation methods 1 and 3, are violated. Fig. 6 shows that permutation methods 1 and 2 produce very similar results in simulations. This is expected, since in this case, the data was homogeneously Gaussian. The P -value transformation step only affects the solution if, under the null hypothesis, the surface elements have an inhomogeneous distribution (see also Fig. 2). However, for real data experiments, as we shall see in a

Table 5
Number of RESELS for simulated and real data

| | Unsmoothed CDMs | Smoothed CDMs |
|------------------------|-----------------|---------------|
| Source simulation data | 641.14 | 141.04 |
| Real data | 574.44 | 125.62 |

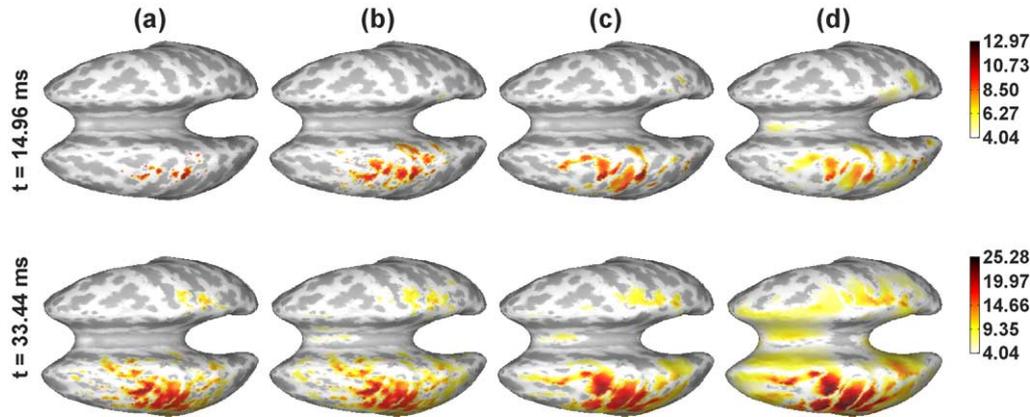


Fig. 8. Examples of significant activation maps for permutation and random field methods for real data. (a) Permutation method 1 using unsmoothed CDMs, (b) permutation method 3 using unsmoothed CDMs, (c) permutation method 3 using smoothed CDMs, (d) Random field controlling FWER over space only.

following section, the two methods produced different activation maps, an indication that the Gaussianity assumption may be violated.

We can display the unthresholded P -value maps of permutation method 2 by transforming the CDMs of the original data into P -values. Even though this does not address the multiple comparisons problem, it is interesting to compare the relative apparent localization properties of CDMs ($\hat{\mu}_{ii}$), noise-normalized CDMs (T_{ii}), and P value maps ($p_i(T_{ii})$). Such a result is shown in Fig. 7. Noise-normalized and $(1 - P)$ value maps are qualitatively similar, again this is because the data are Gaussian. However, $(1 - P)$ value maps offer a direct quantitative measure of significance. In the lower row of Fig. 7, all sources with P -value less than 0.05 are shown for the $(1 - P)$ value maps; CDMs and noise-normalized CDMs are thresholded subjectively, indicating the importance of some form of normalization of the CDMs, either by noise standard deviation or using P -values, before thresholding.

Noise simulation

In order to test all methods for specificity, we created MEG sensor data using only standard Gaussian noise and no signal. The data consisted of 100 epochs, each having 100 pre-stimulus and 100 post-stimulus time points. We then estimated the 5% threshold values for all methods, as given in Table 4.

We then repeated the above procedure 100 times and tested the simulated data ($n_1 = 100$ epochs for spatiotemporal FWER-corrected thresholds, $n_2 = 10,000$ slices for spatial FWER-Corrected thresholds) for activation. The approximate Monte Carlo standard errors for a true 0.05 rejection rate are $\sqrt{\alpha(1 - \alpha)/n_1} = 0.0218$ for spatiotemporal FWER-corrected thresholds and $\sqrt{\alpha(1 - \alpha)/n_2} = 0.00218$ for spatial FWER-corrected thresholds. The approximate 95% confidence intervals are (0.0073, 0.0927) and (0.0457, 0.0543), respectively.

The random field method is conservative, with or without spatial smoothing. Without smoothing the spatial FWER is 0.0139; with spatial smoothing of 5 vertices FWHM, the spatial FWER is 0.0340. Both are outside the 95% confidence limit, though the result with smoothed data is better.

It can be shown theoretically that permutation methods are always exact, that is, they will achieve the chosen FWER. Our experiments verify this: for unsmoothed data method 3 had a spatial FWER of 0.0480, while method 1 had a spatiotemporal FWER of 0.06, both inside the 95% confidence intervals.

Real data experiment

The effectiveness of the proposed algorithms was also investigated using data from a somatosensory experiment. Data

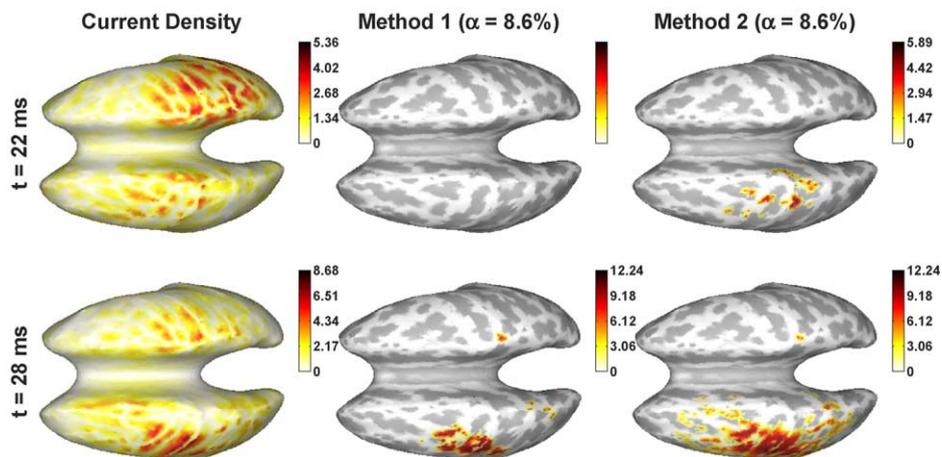


Fig. 9. Reconstruction and significant maps from permutation methods 1 and 2 for a FWER of $\alpha = 0.086$.

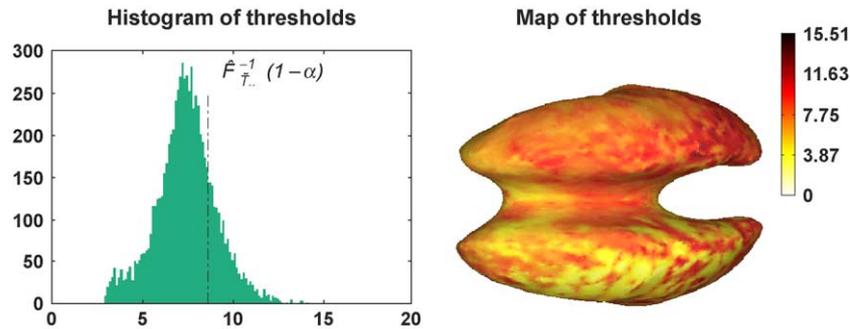


Fig. 10. Global threshold applied by permutation method 1 ($\hat{F}_{\bar{T}_*}^{-1}(1-\alpha)$) at level $\alpha = 0.086$, as compared to the histogram of the thresholds $p_i^{-1}(\hat{F}_{\bar{p}_*}^{-1}(\alpha))$ applied to each source i by method 2. Also, a map of the thresholds on the cortical surface is given on the right. Most of the individual thresholds are below $\hat{F}_{\bar{T}_*}^{-1}(1-\alpha)$.

were acquired using a CTF Systems Inc. Omega 151 MEG system. The somatosensory stimulation was an electrical square-wave pulse delivered randomly to the thumb, index, middle, and little finger of each hand of a healthy right-handed subject (Meunier et al., 2001). For the purposes of this study, only data from the right thumb was used. The recordings consist of 400 epochs each having 62 pre-stimulus and 62 post-stimulus time points.

Table 5 shows the number of estimated RESELS, $R_2(S)$, for the source simulation and real data experiments. Since both data have the same range of RESELS, our simulation experiments were performed under conditions consistent with our real experiment.

All methods identify significant activity in the left somatosensory cortex. Since the experiment involved right thumb stimulation, activation of area S1 (primary somatosensory cortex) in contralateral somatosensory cortex is expected (Kandel et al., 2000). Fig. 8 allows for similar inferences as presented in the simulation section for Fig. 5, that is, epoch-wise or spatiotemporal thresholds are more conservative than slice-wise thresholds (Fig. 8a vs. b) and spatial smoothing slightly reduces resolution (Fig. 8b vs. c). In this case, the RF threshold is lower than that for the permutation method 3 (Table 6), so that a broader area of activation is seen in the RF result compared to the permutation method (Fig. 8c vs. d). Since these are real data, we cannot know whether Fig. 8c or d is closer to truth; however, S1 activation, as revealed in other neuroimaging studies, is typically highly focal. Since the two methods performed similarly in simulations, one explanation is that the real data do not satisfy the distribution assumptions of RF theory.

Evidence of violated assumptions is supported by Fig. 9. This shows permutation method 2 to be more sensitive than method 1. Simulation experiments demonstrated that when the data are Gaussian, the two permutation methods exhibit very similar

performance; the discrepancies between the two maps of significant activation using methods 1 and 2 indicate a violation of this distributional assumption. At $t = 22$ ms, only method 2 detected significant activity. Further, it seems to correct the CDM, which shows the main activity in the ipsilateral hemisphere. As mentioned before, we expect activation of the left hemisphere (at least for the early component of the response) and this is supported by our results. For $t = 28$ ms, the same remarks for sensitivity are true. Finally, Fig. 10 shows the thresholds applied by each of the two methods. Again, due to discreteness, the lowest achieved FWER by method 2 is $\alpha = 0.086$.

An alternative explanation to the stringency of the permutation results is contamination of the empirical permutation null distributions. Permutations that are similar to the correctly labeled data will also yield relatively large statistic values, affecting the upper tail of the permutation distribution and shifting the threshold upwards. A possible solution to this problem is to use a step-down test in which the null distribution is 'purified' by removing spatial elements that test as significant and recomputing the permutation distribution of the maximum.

Note that our method can, to some extent, counteract the tendency to inflate the permutation distribution due to our use of a pre-stim standard deviation. While a strong signal can inflate a permuted statistic values, it will also induce more variability into the pre-stim period, inflate the estimated standard error, and hence, reduce permuted statistic values. In this data set, this effect apparently did not overcome the strong signal.

Conclusion

We have presented RF theory and permutation methods for processing of MEG data and extracting significant activation maps. They can be used with any linear or nonlinear cortical imaging method to obtain objective thresholds on statistic maps. The random field method demonstrated valid but conservative performance in our null simulation experiments; the observed FWER was 0.034 (vs. 0.05 nominal) for smoothed data, and worse for unsmoothed data. However, the method successfully identified the two simulated sources while rejecting false-positives on other surface elements. This suggests that the nonstationary smoothness estimation addresses the problem of highly variant spatial correlation of the noise.

Permutation method 3, which controls spatial FWER at each point in time, as does the RF method, achieved superior performance in the simulations, though there was little difference with

Table 6
Real data thresholds found with the different RF and permutation methods

| | Nominal FWER | Unsmoothed CDMs | Smoothed CDMs |
|------------------------------------|-----------------|--------------------|------------------|
| <i>Spatiotemporal FWER methods</i> | | | |
| Permutation method 1 | 0.050 | 9.346 | 8.925 |
| | 0.086 | 8.691 | 8.351 |
| Permutation method 2 | 0.086 | 7.473 | – |
| <i>Spatial FWER methods</i> | | | |
| Permutation method 3 | 0.050 | 6.597 | 6.387 |
| Random field method | 0.050 | 4.426 | 4.045 |

smoothed data. This is because in simulations with smoothed data, RF assumptions are satisfied, i.e., Gaussianity and sufficient smoothness, so that the method is almost exact and performs similarly to its non-parametric counterpart. In experimental somatosensory data, permutation method 3 found activated sources in the contralateral S1 area as expected, with fewer significant spatial elements than the random field method. We cannot say whether the random field method in this case is more sensitive or simply giving false-positives induced by violated assumptions.

Overall, the permutation methods are more flexible than RF based methods. We can work with spatially smoothed or unsmoothed CDMs depending on the desired tradeoff between SNR and spatial resolution, when RFs should only be used with sufficiently smoothed CDMs. In general, we prefer to smooth the data as little as possible; data are low resolution, and it seems undesirable to further reduce the resolution. Further, permutation methods can achieve uniform sensitivity by defining different thresholds per surface element via P -values. More importantly, they do not rely on distribution assumptions, making them suitable for real data experiments. Their major limitation is that we need equal pre- and post-stimulus regions to allow exchangeability. This issue can generally be addressed in planning the event-related MEG study.

Acknowledgments

This work was supported in part by grants from NIBIB (R01 EB002010) and NCRR (P41 RR013642) and in part by the Human Brain Project/Neuroinformatics research program funded by NIMH, NIA, and NIBIB.

Appendix A. Euler characteristic and RESELS for Gaussian Random Fields

Worsley et al. (1996) gives the Euler characteristic density for a 2-dimensional Gaussian RF as

$$\rho_2(u) = \frac{4\log_e 2}{(2\pi)^{\frac{3}{2}}} e^{-u^2/2} u \quad (\text{A.1})$$

$$\rho_3(u) = \frac{(4\log_e 2)^{\frac{3}{2}}}{(2\pi)^2} e^{-u^2/2} (u^2 - 1) \quad (\text{A.2})$$

and for a t RF as

$$\rho_2(u) = \frac{4\log_e 2}{(2\pi)^{\frac{3}{2}}} \frac{\Gamma(\frac{v+1}{2})}{(\frac{v}{2})^{\frac{1}{2}} \Gamma(\frac{v}{2})} \left(1 + \frac{u^2}{v}\right)^{-\frac{1}{2}(v-1)} \quad (\text{A.3})$$

$$\rho_3(u) = \frac{(4\log_e 2)^{\frac{3}{2}}}{(2\pi)^2} \left(1 + \frac{u^2}{v}\right)^{-\frac{1}{2}(v-1)} \left(\frac{v-1}{v} u^2 - 1\right) \quad (\text{A.4})$$

The estimation of $R_2(S)$ is complicated by the strong but heterogeneous spatial correlation. A thorough treatment of such fields can be found in Worsley (2000) and Worsley et al. (1999); we only summarize the results here. The variance of T_{it} is already unity, so does not need to be normalized. For each surface element i , we define the time vector T_i with N_0 elements T_{it} , $t = -N_0 + 1, \dots, 0$. We assume the tessellated cortical surface

comprises K triangles Δ_k , $k = 1, \dots, K$ and let $T_{i_{0k}}$, $T_{i_{1k}}$, and $T_{i_{2k}}$ be the random field time vectors on the vertices of triangle k . Then, the number of RESELS, $R_2(S)$, on the 2-dimensional cortical surface is:

$$R_2(S) = \sum_{k=1}^K \frac{\det\left(\frac{1}{N_0-1} P_k^T P_k\right)^{\frac{1}{2}}}{2} (4\log_e 2)^{-1} \quad (\text{A.5})$$

$$P_k = [T_{i_{1k}} - T_{i_{0k}} \quad T_{i_{2k}} - T_{i_{0k}}] \quad (\text{A.6})$$

where P_k is an $N_0 \times 2$ matrix and $\frac{1}{N_0-1} P_k^T P_k$ corresponds to an averaging over time.

Finally, the number of time RESELS τ over a time interval T_0 is given by:

$$\tau = \frac{T_0}{\text{FWHM}_t} = \frac{T_0}{\sqrt{4\log_e 2}} \lambda \quad (\text{A.7})$$

where $\lambda = \text{Var}\left(\frac{\partial T_u}{\partial t}\right) = E_t\left\{\left(\frac{\partial T_u}{\partial t}\right)^2\right\}$ is the variance of the temporal derivative and is estimated by finite differences.

References

- Adler, R.J. (Ed.), 1981. The Geometry of Random Fields. Wiley, New York.
- Andrade, A., Kherif, F., Mangin, J.-F., Worsley, K.J., Paradis, A.L., Simon, O., Dehaene, S., Bihan, D.L., Poline, J.-B., 2000. Detection of fMRI activation using cortical surface mapping. Hum. Brain Mapp. 12, 79–93.
- Arfken, G.B. (Ed.), 2000. Mathematical Methods for Physicists (fifth ed.). Academic Press.
- Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 18, 14–30.
- Barnes, G.R., Hillbrand, A., 2003. Statistical flattening of MEG beamformer images. Hum. Brain Mapp. 18, 1–12.
- Blair, R.C., Karnisky, W., 1994. Distribution-free statistical analyses of surface and volumetric maps. In: Thatcher, R.W., Hallett, M., Zeffiro, T., Jony, E.R., Huerta, M. (Eds.), Functional Neuroimaging: Technical Foundation., pp. 19–28.
- Boothby, W. (Ed.), 2002. An Introduction to Differentiable Manifolds and Riemannian Geometry, Revised. Academic Press.
- Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999. Global, voxel, and cluster tests, by theory and permutation, 33 for a difference between two groups of structural MR images of the brain. IEEE Trans. Med. Imag. 18, 32–42.
- Carbonell, F., Galan, L., Valdes, P., Worsley, K., Biscay, R.J., Diaz-Comas, L., Bobes, M.A., Parra, M., 2004. Random field-union intersection tests for eeg/meg imaging. NeuroImage 22, 268–276.
- Chung, M.K., 2001. Statistical morphometry in neuroanatomy. PhD thesis, McGill University, Montreal.
- Dale, A.M., Serano, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. Cogn. Neurosci. 5, 162–176.
- Dale, A.M., Liu, A.K., Fischl, R.B., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67.
- Friston, K., Josephs, O., Zarahn, E., Rouquette, A.H.S., Poline, J.-B., 2000. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 12, 196–208.
- Greenhouse, S., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–112.
- Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., Lounasmaa, O., 1993. Magnetoencephalography. theory, instrumentation and applications to the noninvasive study of human brain function. Rev. Mod. Phys. 65, 413–448.

- Hochberg, Y., Tamhane, A.C. (Eds.), 1987. *Multiple Comparison Procedures*. Wiley.
- Huang, M.X., Mosher, J.C., Leahy, R.M., 1998. A sensor-weighted overlapping sphere head model and exhaustive head model comparison for MEG. *Phys. Med. Biol.*
- Kandel, E.R., Schwartz, J.H., Jessell, T.M. (Eds.), 2000. *Principles of Neural Science*. McGraw-Hill/Appleton and Lange.
- Katila, T., Karp, P., 1983. Magnetocardiography: morphology and multipole presentations. In: Williamson, S.J., Romani, G.L., Kaufman, L., Modena, I. (Eds.), *Proc. 16th Conf. Information Processing in Medical Imaging.*, pp. 237–263.
- Meunier, S., Garnero, L., Ducorps, A., Mazieres, L., Montcel, S.L.S.T.D., Renault, B., Vidailhet, M., 2001. Human brain mapping in dystonia reveals both endophenotypic traits and adaptative reorganization. *Ann. Neurol.* 47.
- Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse problems. *IEEE Trans. Biomed. Eng.* 46 (3), 245–259.
- Nichols, T.E., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* 12 (5), 419–446.
- Nichols, T.E., Holmes, A.P., 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.
- Oostendorp, T.F., Oosterom, A.V., 1988. Interpolation on a triangulated 3D surface. *J. Comput. Phys.* 80, 331–343.
- Pantazis, D., Nichols, T.E., Baillet, S., Leahy, R.M., 2003. Spatiotemporal localization of significant activation in MEG using permutation tests. In: Taylor, C., Noble, J.A. (Eds.), *Proc. 18th Conf. Information Processing in Medical Imaging.*, pp. 512–523. July.
- Pantazis, D., Leahy, R.M., Nichols, T.E., Styner, M., 2004. Statistical surface-based morphometry using a non-parametric approach. 2004 IEEE International Symposium on Biomedical Imaging. April.
- Phillips, J.W., Leahy, R.M., Mosher, J.C., 1997. MEG-based imaging of focal neuronal current sources. *IEEE Trans. Med. Imag.* 163, 338–348.
- Satterthwaite, F., 1946. An approximate distribution of estimates of variance components. *Biometrics* 2, 110–114.
- Shattuck, D.W., Leahy, R.M., 2002. Brainsuite: an automated cortical surface identification tool. *Med. Image Anal.* 6 (2), 129–142.
- Singh, K., Barnes, G.R., Hillebrand, A., 2003. Group imaging of task-related changes in cortical synchronization using nonparametric permutation testing. *NeuroImage* 19, 1589–1601.
- Sowell, E.R., Thompson, P.M., Holmes, C.J., Jernigan, T.L., Toga, A.W., 1999a. In vivo evidence for post-adolescent frontal and striatal maturation. *Nat. Neurosci.* (Sept./Oct.).
- Sowell, E.R., Thompson, P.M., Holmes, C.J., Jernigan, T.L., Toga, A.W., 1999b. Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage* 9, 587–597.
- Thompson, P.M., Cannon, T.D., Narr, K.L., Erp, T., Poutanen, V.-P., Huttunen, M., Lönnqvist, J., Standertskjöld-Nordenstam, C.-G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W., 2001. Genetic influences on brain structure. *Nat. Neurosci.* 4, 12 (December).
- Thompson, P.M., Hayashi, K.M., Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* 23 (3), 994–1005 (February).
- Tikhonov, A.N., Arsenin, V.Y. (Eds.), 1977. *Solutions to Ill-Posed Problems*. V.H. Winston and Sons.
- Worsley, K.J., 2000. Exceedence probabilities of local maxima of non-isotropic random fields, with an application to shape analysis via surface displacements. Unpublished report, Dept. of Mathematics, McGill University, Montreal, Quebec, Canada. Available online at <http://www.math.mcgill.ca/~keith>.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* 2, 173–181.
- Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional 36 statistical analysis for cbf activation studies in human brain. *J. Cereb. Blood Flow Metab.* 12, 900–918.
- Worsley, K., Marrett, S., Neelin, P., Vandal, A., Friston, K., Evans, A., 1995. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4, 58–73.
- Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4, 58–73.
- Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., Evans, A.C., 1999. Detecting changes in non-isotropic images. *Hum. Brain Mapp.* 8, 98–101.