◆ Human Brain Mapping 27:114–128(2006) ◆

Hypothesis Testing in Distributed Source Models for EEG and MEG Data

Lourens J. Waldorp,[1,2*] Hilde M. Huizenga,[1] Raoul P.P.P. Grasman,[1] Koen B.E. Böcker,[3] and Peter C.M. Molenaar[1]

[1]*Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*
[2]*Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands*
[3]*Department of Psychopharmacology, Utrecht University, Utrecht, The Netherlands*

Abstract: Hypothesis testing in distributed source models for the electro- or magnetoencephalogram is generally performed for each voxel separately. Derived from the analysis of functional magnetic resonance imaging data, such a statistical parametric map (SPM) ignores the spatial smoothing in hypothesis testing with distributed source models. For example, when intending to test a single voxel, actually an entire region of voxels is tested simultaneously. Because there are more parameters than observations, typically constraints are employed to arrive at a solution which spatially smooths the solution. If ignored, it can be concluded from the hypothesis test that there is activity at some location where there is none. In addition, an SPM on distributed source models gives the illusion of very high resolution. As an alternative, a multivariate approach is suggested in which a region of interest is tested that is spatially smooth. In simulations with MEG and EEG it is shown that clear hypothesis testing in distributed source models is possible, provided that there is high correspondence between what is intended to be tested and what is actually tested. The approach is also illustrated by an application to data from an experiment measuring visual evoked fields when presenting checkerboard patterns. *Hum Brain Mapp 27:114–128, 2006.* © 2005 Wiley-Liss, Inc.

Key words linear inverse; minimum norm; inverse problem; brain imaging; statistical map

INTRODUCTION

In the linear analysis of the electro- or magnetoencephalogram (EEG or MEG) it is often assumed that a great many sources are simultaneously active in the brain [Hämäläinen et al., 1993]. The brain model can accordingly be divided into many voxels, each of which contains a source (often a dipole). With a linear inverse method (e.g., minimum norm, or regularized inverse) the activity, amplitudes, or moments in each voxel can be estimated [Michel et al., 1999]. This approach, referred to as the distributed source model, is attractive because it has the advantage over a nonlinear analysis that the solution is easy to compute and corresponds to the intuitive idea of many small sources distributed across the cortex. A disadvantage, however, is that the number of parameters to estimate is larger than the number of sensors. Therefore, constraints have to be imposed which may spatially smooth out the estimate of activity [Grave de Peralta Menendez and Gonzalez Andino, 1998]; that is, it may introduce a bias. As a consequence, it is possible that an amplitude is estimated as nonzero, whereas in fact it is zero.

In order to obtain statistical verification, the estimates of activity in each voxel can be tested separately for significance. Dale et al. [2000] computed a statistical parametric map (SPM) of all voxels at each time sample. Since this SPM ignores the temporal correlations, Kiebel and Friston [2004a] introduced an SPM for EEG data, referred to as the mass univariate approach, based on the functional magnetic res-

*Correspondence to: Lourens J. Waldorp, University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB, The Netherlands. E-mail: waldorp@psy.uva.nl

Received for publication 21 December 2004; Accepted 5 May 2005
DOI: 10.1002/hbm.20170
Published online 20 July 2005 in Wiley InterScience (www.interscience.wiley.com).

onance imaging (fMRI) analyses. In addition to accounting for temporal correlations, a hierarchical approach was suggested that allows testing temporal patterns which are coupled to the design of the experiment [Kiebel and Friston, 2004b]. As with fMRI analyses, the multiple comparison problem arises in SPM for distributed source analysis, that is, testing a large number of possibly correlated voxels. In fMRI analyses this problem is solved using Gaussian random fields to compute the number of statistically independent regions [Worsley, 2001]. Kiebel and Friston [2004a] suggested using this method in the mass univariate approach to analyze EEG data using distributed source models. Barnes and Hillebrand [2003] computed for MEG data the spatial smoothness of a beamformer solution to take correlations between voxels into account when testing the amplitude of voxels. Another approach was offered by Bénar et al. [2005] in distributed source models, where EEG statistical maps were created by averaging *F*-values comprised of combinations of either one, two, or three sources. In order to account for the multiple comparisons, a nonparametric bootstrap procedure was used to get a distribution of *F*-values to set the threshold.

Besides the multiple comparison problem, hypothesis testing in distributed source models for EEG and MEG may result in incorrect decisions due to intrinsic bias in the estimate of activity of the voxels. Consider the following example. The orientations are known and only the amplitudes have been estimated. In the mass univariate approach each amplitude in the source region is tested separately for significance if there is no a priori region of interest (ROI). However, if the estimate of the amplitudes is biased, then a hypothesis test will indicate that some of the amplitudes deviate significantly from zero, whereas in fact the corresponding true amplitude is zero. Then an incorrect decision has been made due to the bias in the solution. Therefore, unbiased estimates are required for proper testing. In underdetermined systems (fewer observations than parameters), however, this is not possible since the solution of a linear inverse method in such cases is inherently biased [Grave de Peralta Menendez et al., 1997; Grave de Peralta Menendez and Gonzalez Andino, 1998]. As a consequence, often such simple hypotheses of whether there is activity in a single voxel cannot be tested.

Because of this bias, it is argued here that a mass univariate approach is inappropriate for hypothesis testing in distributed source models for EEG and MEG data. It is shown that when testing a single voxel in the mass univariate approach, actually a whole region is being tested. This is because in the estimate of the voxel activity in distributed source analysis, activity is spatially smoothed out [Grave de Peralta Menendez and Gonzalez Andino, 1998]. We will argue that a (spatially) multivariate approach is required, and that there should be a reasonable match between what is intended to be tested and what is actually being tested. It is furthermore demonstrated that a combination of a hypothesis (weighted average, see, e.g., Grave de Peralta Menendez and Gonzalez Andino [1998]) and a linear inverse method exists, such that the ROI that is intended to be tested is actually being tested. However, this combination is difficult to apply in practice, and so it is therefore suggested that the match between what is intended to be tested and what is actually being tested be made as good as possible. Several suggestions to accomplish this are given.

Three test statistics are defined to perform hypothesis testing, which are optimal in the sense that they can distinguish best between zero and nonzero activation. One of them can be used to determine whether there is activity in an ROI and two can be used to determine if there is a difference between conditions. It is shown with simulations that the distribution of the test statistics is correct (for white and colored noise) and that the tests can distinguish source areas which are reasonably close.

After a brief introduction of linear inverse methods, hypothesis testing in distributed source models is extensively discussed. An attempt is made to find a combination of a hypothesis and a linear inverse method such that reliable hypothesis testing can be performed. Next, the three test statistics are defined. The performance for some of the combinations of a hypothesis and a linear inverse method is investigated in a numerical example using both MEG and EEG. Finally, the method is applied to MEG visual evoked field data.

LINEAR INVERSE METHODS

Data from m sensors on independent trials $j=1,\ldots,n$ are collected in the m vector Y_j with average $\bar{Y} = (1/n)\sum_{j=1}^{n} Y_j$. A model for the mean $E\{Y_j\} = \mu$ is constructed such that hopefully $\mu = G\beta$, where G is the $m \times 3d$ lead field matrix [Hämäläinen et al., 1993] and β the $3d$ moment vector for d sources. The lead field matrix contains the locations for d dipoles, one in each voxel. The noise e_j present in the measurements is caused by nonstimulus related brain activity and sensor noise [Hämäläinen et al., 1993: De Munck et al., 1992], and is assumed to have mean zero and covariance matrix Σ for $j = 1,\ldots,n$. The noise can be either white $\Sigma = \sigma^2 I_m$ or colored and then Σ is a full matrix. The model for the averaged data is then:

$$\bar{Y} = G\beta + e. \tag{1}$$

The aim is to find a solution $\hat{\beta}$ with some linear inverse method. A linear inverse method refers to the matrix G^- such that:

$$\hat{\beta} = G^-\bar{Y} = G^-G\beta + G^-e. \tag{2}$$

If the solution is unbiased then a clear interpretation of hypothesis testing exists. Unbiasedness means that $E\{\hat{\beta}\} = \beta$ so that the true parameter is obtained in the mean. The concept of the resolution matrix $R = G^- G$ is convenient to determine unbiasedness [Grave de Peralta Menendez and Gonzalez Andino, 1998; Grave de Peralta Menendez et al., 1997]. If the noise has zero mean, unbiasedness means that

$E\{\hat{\beta}\} = R\beta = \beta$, and so R should be equal to the identity matrix I_{3d}. In underdetermined systems ($m \ll 3d$), however, the resolution matrix cannot be the identity matrix since it has at most rank m [Pringle and Rayner, 1971].

One way of obtaining a linear inverse method is to minimize the distance between the resolution matrix and the identity matrix, that is, minimize the Frobenius norm $\|W^{-1/2}(R - I_{3d})\|^2$,, where W is a weight matrix. If G has full row-rank, then the solution has the form $\hat{\beta} = W^{-1}G'(GW^{-1}G')^{-1}\bar{Y}$. The minimum norm (MN) method is obtained with $W = I_{3d}$ [Pringle and Rayner, 1971]. Many versions of the weighted MN are available. A relatively simple version is obtained with $W = \text{diag}(|g_1|^q, \ldots, |g_{3d}|^q)$, where g_i is a column of G [e.g. Pascual-Marqui et al., 1994, with $q = 2$]. This version is referred to here simply as the weighted MN (WMN). The weighted resolution optimization (WROP) method is obtained with $W_{ij}^{-1} = (d(x_i,x_j)^2 + \gamma) + (1 - \delta_{rs})\rho$, where $d(x_i,x_j)$ is the distance between two voxels, δ_{rs} is Kronecker's delta, and γ and ρ are constants [Grave de Peralta Menendez et al., 1997]. The MN is known to yield solutions which are superficial, whereas the weighted MN can obtain more deeper sources. The MN solution is very smeared spatially across the solution volume. The WROP and to some extent the weighted MN can produce solutions which are much less smeared spatially [Grave de Peralta Menendez et al., 1997]. In order to reduce the spatial smearing, it has recently been suggested [Grave de Peralta Menendez et al., 2004a] that the inverse of the resolution matrix can be used. The idea is that by using an approximate inverse of the resolution matrix, $E\{R^-\hat{\beta}\} = R^-R\beta$, which is approximately β. By perturbing (regularizing) the resolution matrix by a diagonal matrix ζD, a so-called update $(R + \zeta D)^{-1}G^+$ is used as the inverse method. The estimate is then $\hat{\beta} = (R + \zeta D)^{-1}G^+\bar{Y}$. As an example taken from Grave de Peralta Menendez et al. [2004a], the diagonal matrix could be $D = (R)$ with $\zeta = 10$, where $(\cdot)$ denotes the diagonal of a matrix, $R = G^+G$, and G^+ could be the MN. This linear inverse method with the MN will be denoted by inverse resolution (IR).

Another way of obtaining a linear inverse method is by minimizing a combination of the least squares (LS) function and a penalty depending on the moment vector β and a regularization parameter λ. One possibility is to minimize the equation [Grave de Peralta Menendez and Gonzalez Andino, 1998]:

$$Q_{\beta,\lambda}(Y) = (\bar{Y} - G\beta)'V^{-1}(\bar{Y} - G\beta) + \lambda\beta'W\beta, \qquad (3)$$

where V is (usually) the covariance matrix of the residuals. The solution is of the form $\hat{\beta} = W^{-1}G'(GW^{-1}G' + \lambda V)^{-1}\bar{Y}$. An example of a regularized estimate is by Dale and Sereno [1993], where anatomical constraints about the cortical sheet are used to determine the orientations and W is determined by spatial source correlations. The regularization parameter λ can be determined by the L-curve method [Hansen, 1992] or the generalized cross validation (GCV) method [Xiang and Wahba, 1996]. The version with W as in WMN is referred to here as the regularized and weighted (RW) method.

HYPOTHESIS TESTING

As discussed in the Introduction, for hypothesis testing on the amplitudes to be clear the solution should be unbiased. Since this is not possible [Grave de Peralta Menendez and Gonzalez Andino, 1998], it should be determined to what extent the hypothesis corresponds to what is actually being tested. Consider the example of testing three independent moment components separately of each voxel, as discussed by Dale et al. [2000], which is a special case of the mass univariate approach. This can be described by testing the null hypothesis $H_0 : p'\hat{\beta} = 0$ for each of the three components separately of the d voxels, where $p' = (0,\ldots,0,1,0,\ldots,0)$ with a single 1 corresponding to the moment component of the required voxel. Since $E\{\hat{\beta}\} = R\beta$, what is actually being tested is whether $p'\hat{\beta} = p'R\beta = 0$, sometimes referred to as a weighted average [Menke, 1989]. It follows that the optimal weight vector p' would be when $p'R = p'$, since then the hypothesis corresponds exactly to what is being tested and so $E\{p'\hat{\beta}\} = p'\beta$. Generally, though, a weight vector with zeros and ones does not have this property with the resolution matrix. For the present example, Figure 1 shows how R "filters" the hypothesis for three different linear inverse methods (see Numerical Example for details): inverse resolution (IR) (Fig. 1b), weighted minimum norm (WMN) (Fig. 1c), and reweighted and regularized (RW) (Fig. 1d). It can be seen that for each of the linear inverse methods that instead of a single voxel a source area is actually being tested. Although the centers of the source areas of the WMN and RW correspond approximately to that of the single true voxel, the resolution matrix R induces weighted average that contains many more voxels than intended in the original p' indexing a single voxel.

From the previous section it is clear that $R \neq I_{3d}$ and so some other strategy is required to get $p'R = p'$. The aim is to choose a combination of a weight vector p' and a linear inverse method such that $p'R = p'$. If the weight vector p' is in the row-space of the lead field matrix G, then there exists an optimal choice of linear inverse method such that $p'R = p'$ exactly. This optimal linear inverse method satisfies the first of four conditions of the Moore-Penrose inverse. The four conditions are 1) $GG^-G = G$, 2) $G^-GG^- = G^-$, 3) $(GG^-)' = GG^-$, and 4) $(G^-G)' = G^-G$. A linear inverse method satisfying (1) is easily seen to yield an unbiased weighted average, also referred to as an estimable function [Pringle and Rayner, 1971]. If p' is in the row-space of G, then there is a q' such that $p' = q'G$, and so $E\{p'\hat{\beta}\} = p'R\beta = q'GG^-G\beta = p'\beta$, which is exactly what is required. A linear inverse method that satisfies the first three conditions is sometimes referred to as a 3-inverse [Pringle and Rayner, 1971] or as a {1,2,3}-inverse [Ben-Israel and Greville, 2003]. Linear inverse methods of the form $G^- = W^{-1}G'(GW^{-1}G')^{-1}$ are 3-inverses. Linear inverse methods of the form $G^- = (R + \zeta D)^{-1}G^+$ and $G^- = W^{-1}G'(GW^{-1}G' + \lambda V)^{-1}$, however, do not satisfy any condition.

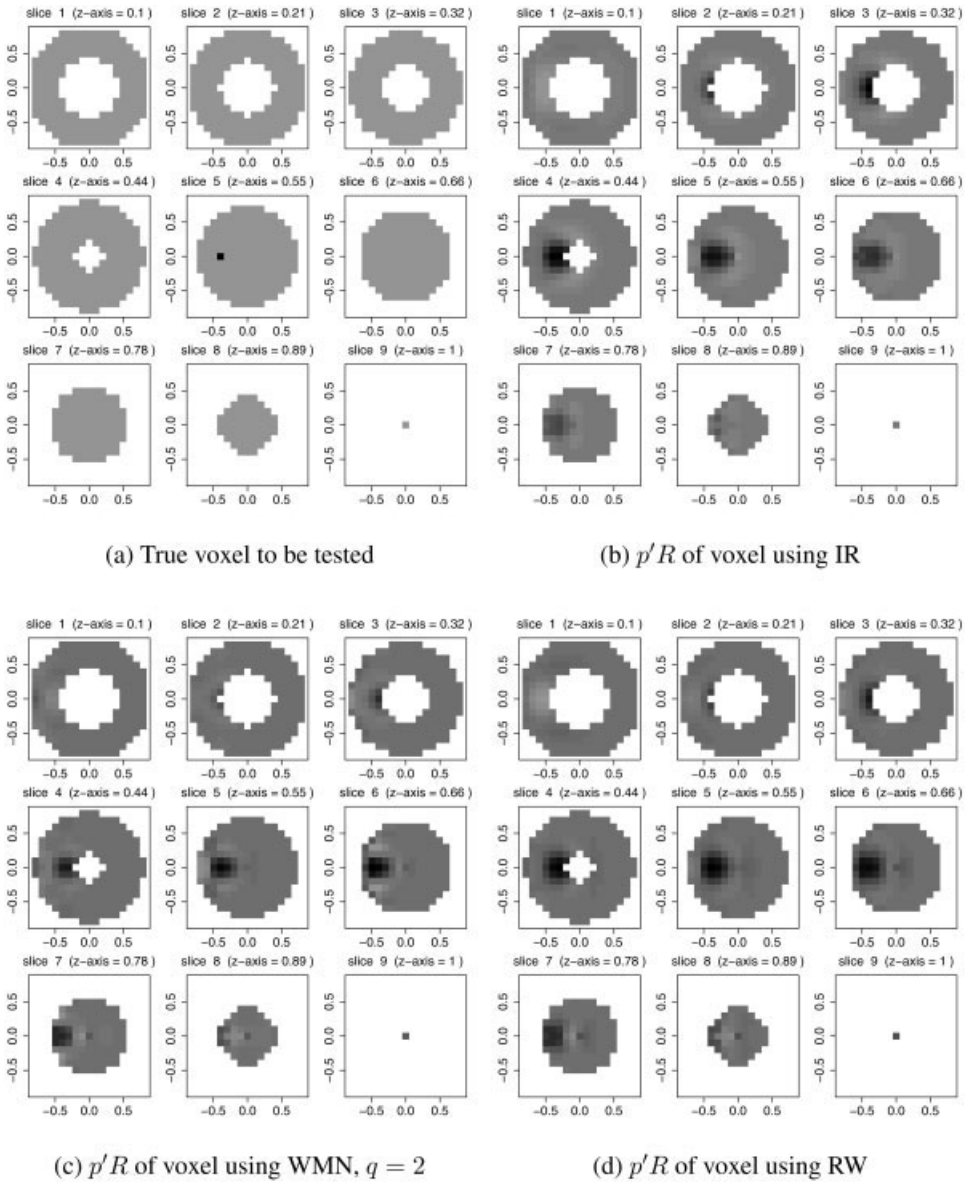
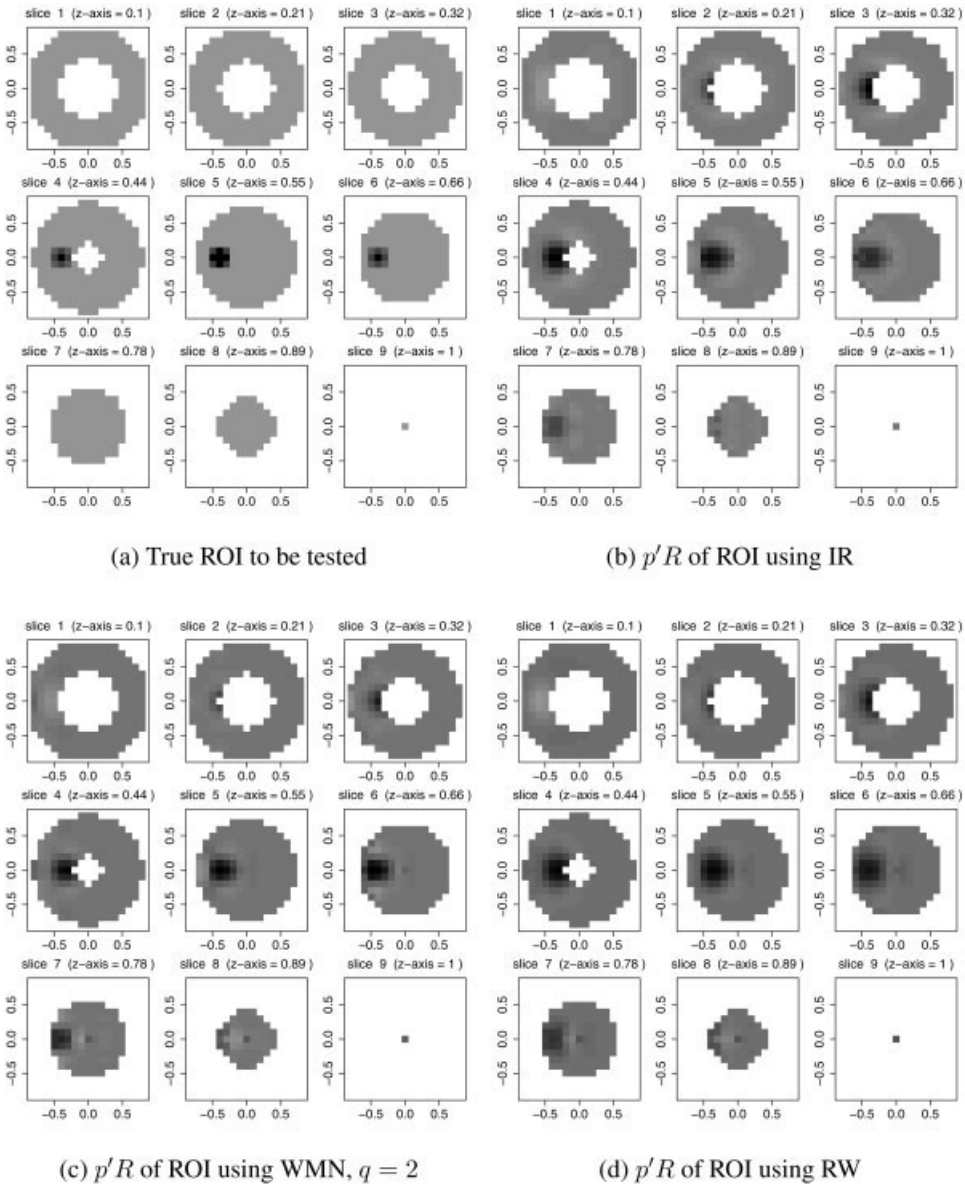


Figure 1.
Weights in p' of the intended and "filtered" $p'R$ hypothesis for MEG. The white area indicates that no activity is allowed, gray indicates where activity can occur, and dark gray/black indicates high activity. The regularization parameter for RW is $\lambda = 0.0001$, and the setting for IR is $\zeta D = 10\text{diag}(R)$. The true voxel is located at $(x, y, z) = (-0.4, 0, 0.5)$ with eccentricity 0.64 in the unit sphere.

Although this principle of an unbiased weighted average is convenient, in practice it is difficult to apply, since finding an appropriate weight vector p' in the row-space of G is difficult. The aim in general is to obtain as high a correspondence as possible between the intended hypothesis and what is actually being tested, that is, between the weighted averages $p'\beta$ and $p'R\beta$. In Figure 2 it can be seen that by defining an ROI instead of a single voxel, the correspondence between the intended weight vector p' and the actually tested weight vector $p'R$ is quite large. For comparison, the same ROI in p' and the "filtered" version $p'R$ are shown for EEG in Figure 3 (see Numerical Example for details). It can be seen that the ROI is spatially more smoothed out using EEG (Fig. 3) than using MEG (Fig. 2).

This high correspondence boils down to ruling out the possibility that in the weighted average there are contributions from sources other than the ROI. Such weighted averages are referred to as interpretable [see, e.g., Grasman, 2004, Ch. 6]. High correspondence (interpretability) can be achieved by either modifying the linear inverse method or the weight vector such that $p'R$ is approximately p'. Both the WROP [Grave de Peralta Menendez et al., 1997] and IR method [Grave de Peralta Menendez et al., 2004a] contain relatively few contributions from other locations and are therefore interpretable.

The weight vector p' (and thus the weighted average) is easiest to construct for the amplitudes as opposed to the entire moment vectors. The model in (1) only contains the moments but can be split up into orientation and amplitude parameters for each source. This is done by the transformation $\beta = B\alpha$, where $B = (\text{diag}(\alpha)^{-1} I_d \otimes 1'_3)\text{diag}(\beta)$,diag($\cdot$) creates a diagonal matrix of a vector, $\otimes$ denotes the Kro-



(a) True ROI to be tested

(b) $p'R$ of ROI using IR

(c) $p'R$ of ROI using WMN, $q = 2$

(d) $p'R$ of ROI using RW

Figure 2.
Same as in Figure 1 for an ROI for MEG. The center of the ROI is at $(x, y, z) = (-0.4, 0, 0.5)$ with eccentricity 0.64 and it has radius 0.2.

necker product, $1_3 = (1,1,1)'$, and α contains the d amplitudes. Then B is block-diagonal within each block the orientation vector $\beta_n^j = \beta_i/\alpha_i$ [Mosher et al., 1992]. With this reparameterization the model for the mean data is $G\beta = GB\alpha$. An ROI can be defined for the amplitudes and then made into a weight vector by $p = B\alpha_{ROI}$. If the orientations in B are unknown and no anatomical constraints for them are used [see, e.g., Dale and Sereno, 1993; Grave de Peralta Menendez et al., 2004b], then they can be estimated from the data first and then the weight vector can be created.

The object of defining an ROI in α_{ROI} is that its filtered version $p'R = \alpha'_{ROI}B'R$, using one of the linear inverse methods, should correspond as much as possible to $p' = \alpha'_{ROI}B'$. The ROI for can be defined in several ways. In Grave de Peralta Menendez and Gonzalez Andino [1998], for instance, it is suggested that the exponential function $w_k \exp[-d(x_i,x_j)^2/a]$ be used, where $d(x_i,x_j)$ is the Euclidean distance between two voxels x_i and x_j, a is a correlation parameter, and w_k is a scaling parameter. Another possibility is the inverse of the distances between voxels $1/d(x_i,x_j)^b$. In Grave de Peralta Menendez et al. [2004b], a variation of this is used by including a weighted average of neighboring voxels, which is a generalization of the function used in Pascual-Marqui et al. [1994]. These two functions can be very similar depending on the parameter a for the exponential function and on b in the inverse distances function. As an illustration, Figure 4 shows two ROI, one defined by the exponential function (Fig. 4a) and one defined by the inverse distance function with $b = 2$ (Fig. 4c). Their respective "filtered" versions $p'R$, using the inverse resolution (IR), in Figure 4b,d, show that the ROI corresponds approximately equally well with both functions. If it is desired to test a

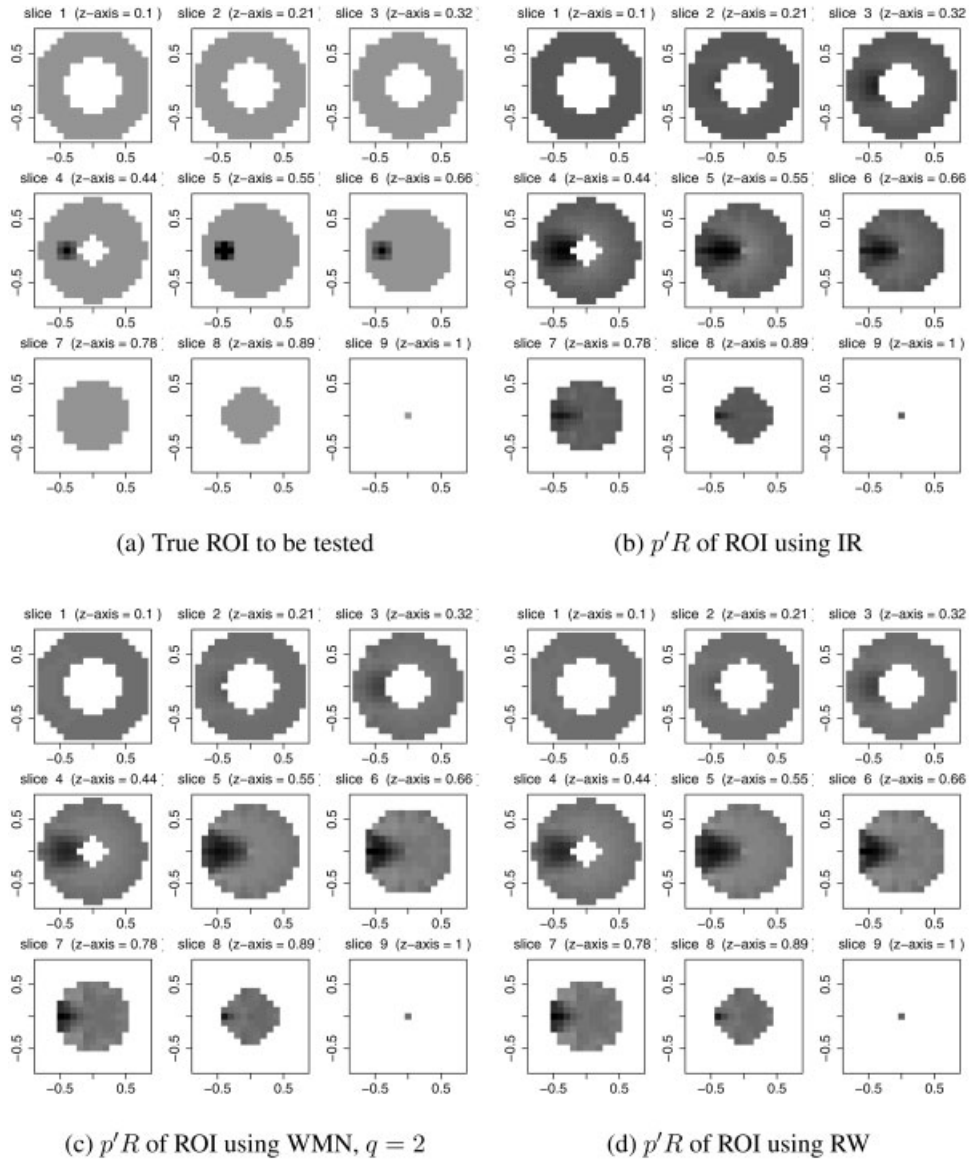


Figure 3.
The same ROI as in Figure 2 for EEG.

more distributed ROI that is extended along the cortex, then a combination of two or more spatially smooth ROI should be used. As can be seen in Figure 5, an ROI with hard edges (Fig. 5a) corresponds less well to the actual area being tested (Fig. 5b), than an ROI defined by a combination of two smooth ROI (Fig. 5c) and its actual test area (Fig. 5d). In general, the ROI should be defined according to some property of spatial smoothness, which reflects the intrinsic bias. Because of the spatial smoothing by the resolution matrix in $p'R\beta$, an ROI should accordingly be spatially smooth such that $p'\beta \approx p'R\beta$.

Test Statistics

If several ROI have been defined in $p_1,\ldots,p_r$ *with* $r < m$, then a multivariate test and subsequent univariate tests on each P_j individually can be performed. Usually the null hypothesis H_0 is that the weighted average is zero and the alternative H_1 is that it deviates from zero. Let $P' = (p_1,\ldots,p_r)'$ contain the r weight vectors, such that the general null hypothesis is $H_0 : P'R\beta - c = 0$ where $c = (c_1,\ldots,c_r)'$ is a specified constant, often zero. To test the hypothesis the estimate $\hat{\beta}$ is used in $P'\hat{\beta} = P'G^{-}\bar{Y}$. Since the variance of the average $\bar{Y}$ is $(1/n)\Sigma$ the variance of $P'\hat{\beta}$ is $(1/n)P'G^{-}\Sigma(G^{-})'P$. Generally, Σ is unknown and so an estimate should be available. An unbiased estimate of the noise covariance matrix Σ is [Muirhead, 1982]:

$$S = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})(Y_j - \bar{Y})'. \qquad (4)$$

This estimate requires that there are more trials than senors ($n > m$). It is to be tested whether the r vector $P'\hat{\beta}$

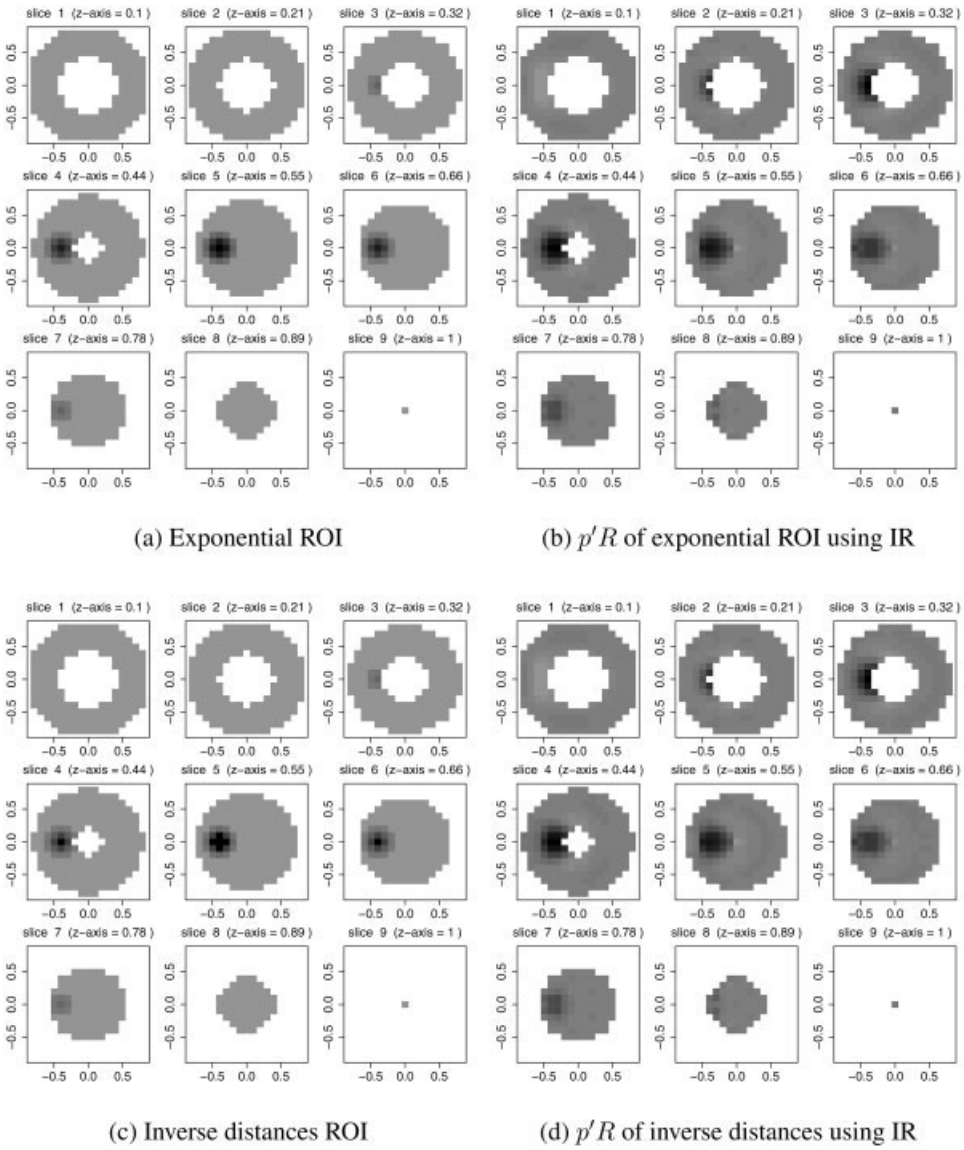


Figure 4.
In MEG, an ROI defined by the exponential function (**a**) and its filtered version (**b**) with the inverse resolution method (IR), and an ROI defined by the inverse squared distances function (**c**) and its filtered version (**d**). In both cases the center and radius of the areas are (−0.4, 0, 0.5) and 0.3, respectively.

– c equals zero given its variance estimate $(1/n)P'G^{-}S(G^{-})'P$. Since $P'\hat{\beta} = (PG^{-})\bar{Y}$ is a linear combination of the averaged data, Hotelling's T^2 is an obvious choice since it is best in distinguishing between the null and alternative hypothesis in this situation and has an exact distribution if the data are distributed as normal [Muirhead, 1982]. The test statistic T^2 is defined as [Bilodeau and Brenner, 1999]:

$$nT^2 = n(P'\hat{\beta} - c)'[P'G^{-}S(G^{-})'P]^{-1}(P'\hat{\beta} - c) \sim \frac{(n-1)r}{n-r}F(r,n-r), \quad (5)$$

where $F(r,n-r)$ denotes the F distribution with r and $n-r$ degrees of freedom. Each weighted average based on a single p'_j can also be tested individually with this test statistic.

To test whether there is any activity in the ROI, c can be set to zero, and then $nT^2 > K_r F_\alpha(r,n-r)$, with $K_r = (n-1)r/(n-r)$ is the overall test, indicating that there is some activity. Subsequently, $nT^2(p_j) > K_1 F_{\alpha/r}(1,n-1)$ indicates that in the ROI in p_j there are nonzero amplitudes. Note that a single hypothesis is tested at the Bonferroni corrected level α/r, to avoid incorrectly deciding that there is activity when in fact there is none.

Two conditions can also be compared. The ROI then indicates the region where a difference in activity is expected. Two assumptions should be satisfied for such an analysis: the sensor positions should be exactly the same in each condition, and the noise should be independently distributed between conditions. The first assumption is required

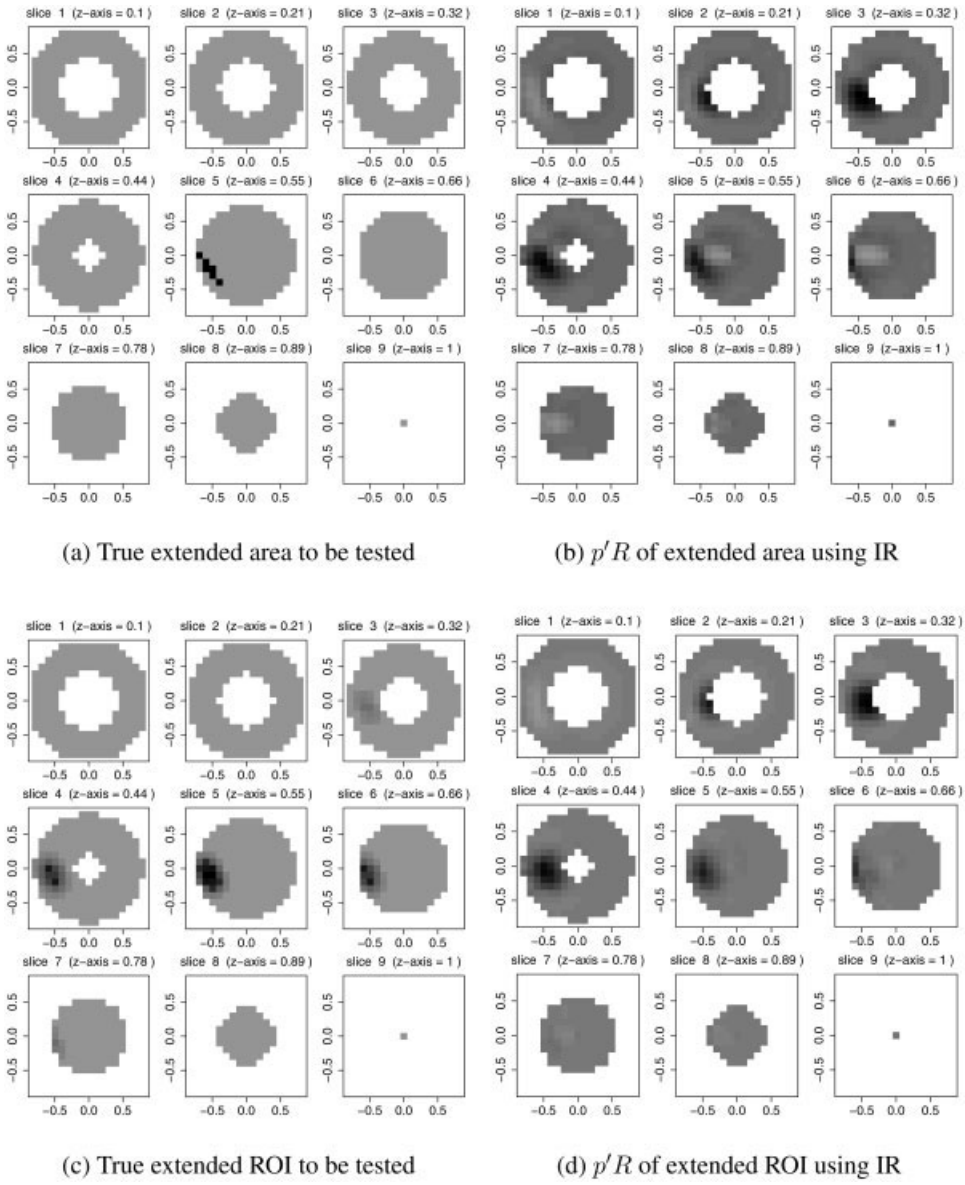


Figure 5.
In MEG, an extended area (**a**), with center approximately at (x, y, z) = (−0.7, −0.3, 0.5) and eccentricity 0.91, and an extended ROI (**c**), consisting of two ROI with centers (−0.6, 0, 0.5) and (−0.5,−0.3, 0.5) and radii 0.3. The extended regions actually tested are shown in (**b**) and (**d**), using the IR.

since the same lead field matrix is used for both conditions. This assumption is likely to be satisfied, for example, in experiments in which a participant is subjected to both conditions during a single test session. The second assumption is the same as the assumption of the independence of trials, which is likely to be satisfied in well-designed experiments. Two versions of the test between conditions are described: equal and unequal noise covariance matrices. (A procedure to determine if the noise covariance matrices are equal is described in Srivastava and Carter [1983].)

Both tests are essentially the same T^2 as described above and only require a different estimate of the noise covariance matrix. Let n_i, $\hat{\beta}_i$, and S_i denote the number of trials, the estimate, and noise covariance matrix, respectively, of condition i = 1,2. To test between conditions the null hypothesis $H_0 : P'R(\beta_1 - \beta_2) = 0$ is tested. If it is assumed that the noise covariance matrices are (approximately) equal for both conditions, then a pooled version of the noise covariance matrix should be used. An estimate of the pooled noise covariance matrix is [Anderson, 1958, p. 109]:

$$S_p = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1 + (n_2 - 1)S_2], \qquad (6)$$

Then the test with S_p is [Srivastava and Carter, 1983, p. 48]:

$$\frac{n_1 n_2}{n_1 + n_2} T_p^2 \sim \frac{(n_1 + n_2 - 2)r}{n_1 + n_2 - 1 - r} F(r, n_1 + n_2 - r - 1). \qquad (7)$$

If on the other hand the noise covariance matrices are unequal (Behrens-Fisher problem), the noise covariance matrix

is computed from transformed data. Assume that $n_1 < n_2$ and define for trial j [Anderson, 1958, p. 119]:

$$Z_j = Y_{1j} - \sqrt{\frac{n_1}{n_2}}Y_{2j} + \frac{1}{\sqrt{n_1 n_2}}\sum_{i=1}^{n_1} Y_{2i} - \frac{1}{n_2}\sum_{k=1}^{n_2} Y_{2k}. \quad (8)$$

The matrix $S_z = 1/(n_1 - 1)]\sum_{j=1}^{n_1}(Z_j - \bar{Z})(Z_j - \bar{Z})'$, where $\bar{Z} = (1/n_1)\sum_{j=1}^{n_1} Z_j$, is the noise covariance matrix. Then the test with S_z is:

$$n_1 T_z^{\,2} \sim \frac{(n_1 - 1)r}{n_1 - 1 - r}F(r, n_1 - 1 - r). \quad (9)$$

NUMERICAL EXAMPLE

The aim of this numerical example is to investigate the properties of the test statistics. Important properties of test statistics are its level and power. A test is called a level α test if the probability that the null hypothesis is rejected is at most α (typically 0.05) when the null hypothesis is true. The power of a test refers to the probability of rejecting the null hypothesis when it is false, which should be as high as possible.

In the present application a combination of level and power is important. It is desired that when the ROI does not overlap with the true source area that the null hypothesis of no activity is not rejected. But because the hypothesis actually being tested is not exactly the same as the ROI, the alternative hypothesis is true but is very close to the null hypothesis. When the alternative is true but is very close to the null hypothesis, such an alternative is called local [Van der Vaart, 1998]. So for $H_0 : P'R\beta = 0$ against local alternatives $H_1 : P'R\beta \approx 0$ the level and power are investigated and H_0 should only be rejected when the ROI overlaps with the true source area generating the data. In the hypothesis only a spatially smooth ROI is used, for this was seen to lead to more interpretable hypotheses (i.e., $p'\beta \approx p'R\beta$) than other kinds. Note that it makes no difference whether the true source area or the ROI is moved to obtain local alternatives, only their relative distance is important.

For the grid, only a ring in the upper half of the unit sphere is considered for analysis, and the inner sphere, where no activity is allowed, has radius 0.4. In the remaining volume there are 1,249 voxels defined. Both MEG and EEG data are generated with two different kinds of source areas defined in the unit sphere. The first is a relatively deep source area with center $x_{0s} = (-0.4, 0, 0.5)$, eccentricity of 0.64, and radius $\rho_0 = 0.2$, as displayed in Figure 2a. Its amplitude pattern is defined according to the inverse squared distances function, and is referred to as the smooth source area. The second source area is extended with hard edges, with center $x_{0e} = (-0.7, -0.3, 0.5)$ and eccentricity 0.91, as shown in Figure 5a. Its amplitude pattern consists of zeros and ones (hard edges) and the source area is referred to as the extended source area. All orientations of the sources were in the tangential plane [see Dogandžić and Nehorai, 2000], and the correct orientations were assumed known in determining the ROI and the estimate.

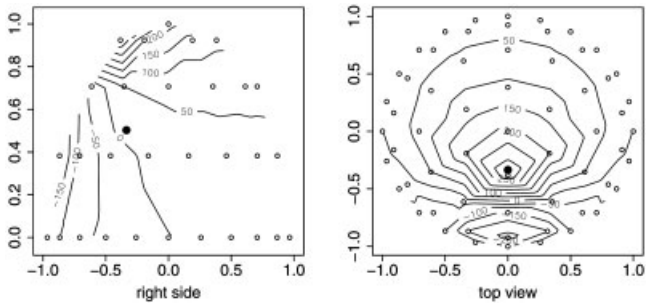


Figure 6.
The 61 sensor configuration (open dots) used for the simulation and the MEG signal from the true source area with center (solid dot) $x_{0s} = (-0.4, 0, 0.5)$.

MEG data are generated, with the above-mentioned source areas, on 61 axial gradiometers using a spherical shell as a head model. EEG data are generated similarly on 61 electrodes using a homogeneous single shell head model. The number of 61 sensors was chosen from the Neuromag system, which has 122 sensors at 61 different locations, used frequently [Liu et al., 2002], and to allow comparison of MEG with the often used caps measuring EEG with 61 electrodes. The sensor configuration is the same for the MEG and EEG simulation and is shown together with the topographical field plot for MEG in Figure 6. The noise added to the sensors is distributed as normal with zero mean and covariance matrix Σ. The noise can be either white, i.e., $\Sigma = \psi^2 I_m$, or colored and then Σ is set to have elements $\sigma_{ij} = \psi^2\exp(-d_{ij}/a)$, where d_{ij} is the distance between two sensors, a is a correlation parameter, and ψ^2 is a variance parameter which is adjusted such that the noise variance for the mean is 10% of the maximal signal $G\beta$. With $a = 1$ absolute correlations are obtained between (0.21, 1.00) with mean 0.34, which corresponds well to real MEG data with correlations between (0.02, 1.00) with mean 0.31. Data are generated for several number of trials $n = 80$, 100, 200, and 500.

The linear inverse methods are the minimum norm (MN), the weighted MN (WMN), the inverse resolution (IR), and the regularized and weighted inverse (RW), as described above. For IR, the settings were used as described in Linear Inverse Methods and the regularization parameter in RW is set to $\lambda = 0.0001$. The actual hypothesis $H_0 : P'R\beta$ for the spatially smooth area is shown for MEG in Figure 2 and for EEG in Figure 3 for the linear inverse methods IR, WMN, and RW. The actual hypothesis for the extended area is shown in Figure 5 for MEG.

It is first investigated whether the correspondence under H_0 between the empirical cumulative distribution function (CDF) and the theoretical CDF is high. In order for H_0 to be (nearly) true the ROI is located far from the true source area. The distance between centers for the smooth hypothesized

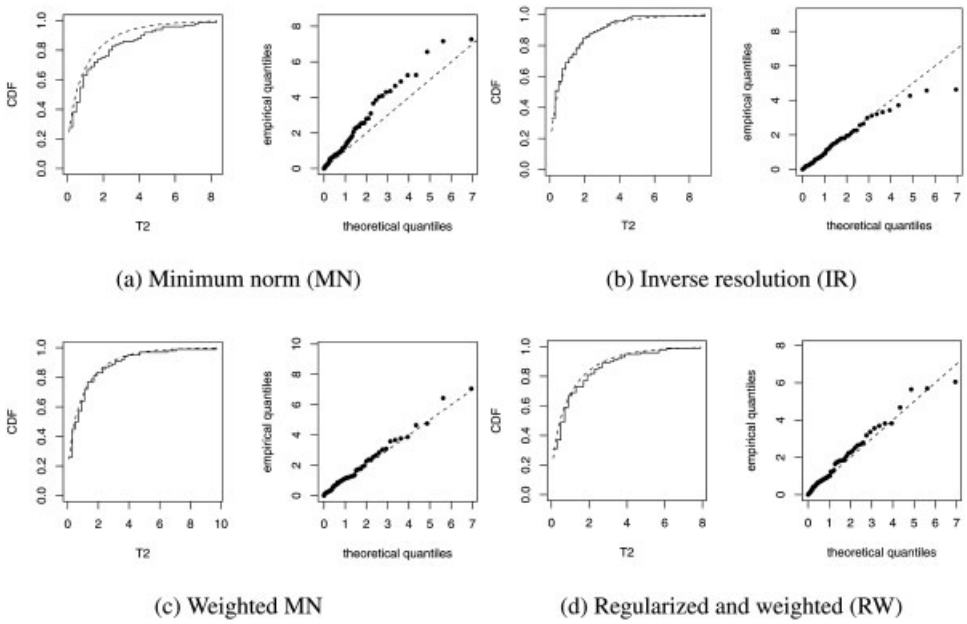


Figure 7.
CDF and q-q plots of T^2 of simulated MEG data using the smooth source area, with $n = 80$ and colored noise. **a:** Minimum norm (MN). **b:** Inverse resolution (IR). **c:** Weighted minimum norm (WMN). **d:** Regularized and weighted (RW). The empirical CDF is the solid line and the theoretical CDF the dashed line. In the q-q plot the dots are the empirical ordered values and the dashed line represents the theory.

ROI and true smooth source area is 1.27 and for the true extended source area and the smooth hypothesized ROI the distance is 1.48. Figure 7 shows for MEG and colored noise the cumulative distribution functions (CDFs) and quantile-quantile (q-q) plots. The CDF gives information on general correspondence while the q-q plot shows more clearly the correspondence for larger quantiles of the statistic, which is important for the level of the test. The correspondence between the empirical and theoretical distribution is quite good for IR (Fig. 7b) with $n = 80$ for colored noise, but diverges somewhat at the larger quantiles. This is also true for white noise, but is not shown. Hence, the statistic accounts very well for the correlations in the data. It can also be seen that for colored noise the correspondence between the empirical and theoretical distribution for IR and RW is good as well, but that for MN the correspondence is not very good. In a similar figure for EEG, Figure 8 shows that on the whole the correspondence between the empirical and theoretical distribution is not as good as for MEG, especially for the larger quantiles.

To investigate the level ($\alpha = 0.05$) of the test, again for H_0 to be (nearly) true the smooth ROI is used with its center at large distances from the center of the true smooth (1.27) and extended (1.48) source areas. In the left panel of Figure 9a it

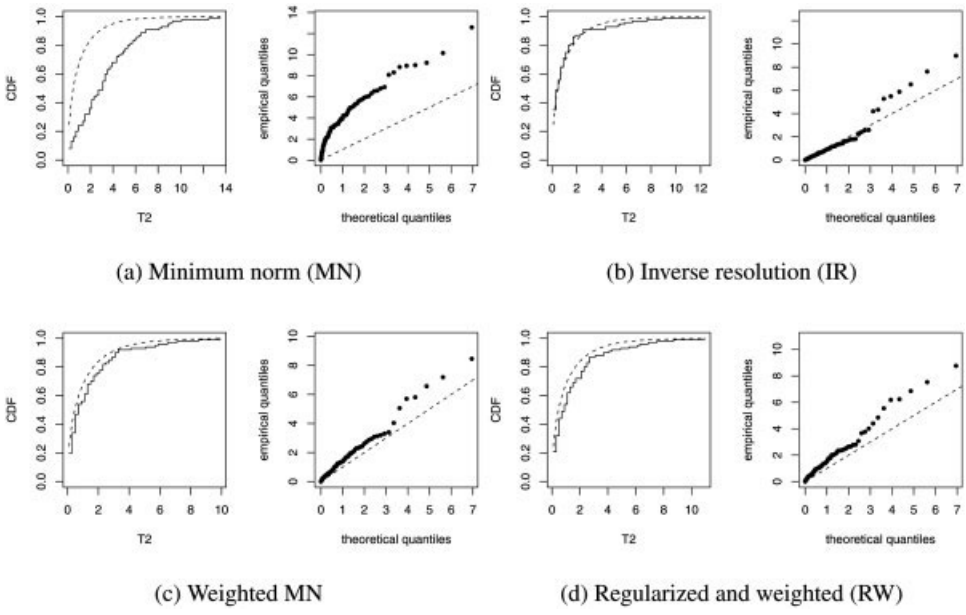


Figure 8.
CDF and q-q plots of T^2 of simulated EEG data with $n = 80$ and colored noise. **a:** Minimum norm (MN). **b:** Inverse resolution (IR). **c:** Weighted minimum norm (WMN). **d:** Regularized and weighted (RW). The empirical CDF is the solid line and the theoretical CDF the dashed line. In the q-q plot the dots are the empirical ordered values and the dashed line represents the theory.

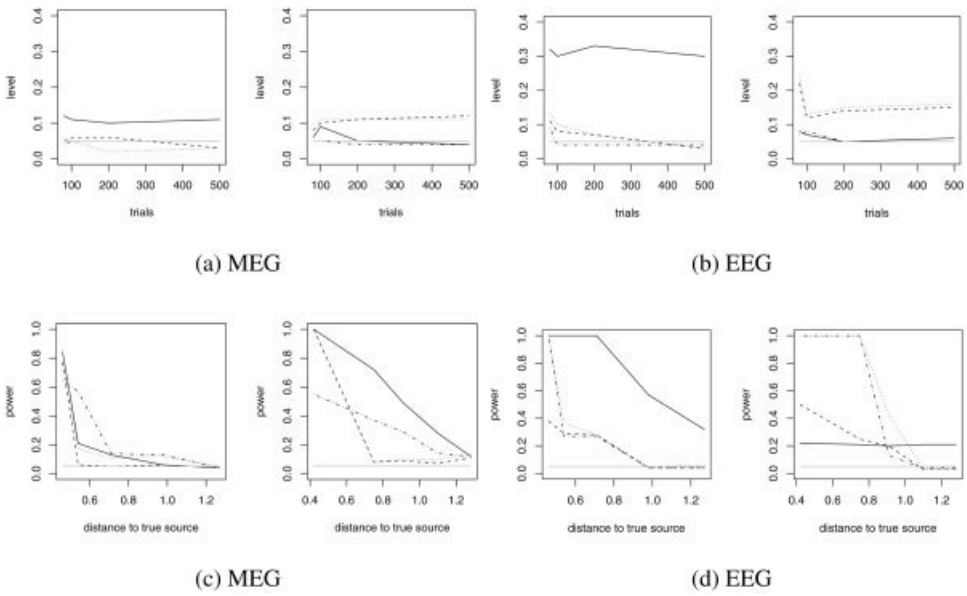


Figure 9.
The level (top) and power (bottom) of T^2 for MEG (left) and EEG (right), with $n = 80$ and colored noise. The linear inverse methods are the MN (solid line), IR (dashed-dotted line), WMN (dashed line), and RW (dotted line). In each panel the left graph is data from the smooth source area and the right panel is from the extended source area. The gray solid line is at the correct level of $\alpha = 0.05$.

can be seen that for MEG, colored noise, and the smooth source area the level with the MN is too high but that with IR, WMN, and RW it is ~0.05. For the extended source area, the right panel of Figure 9a shows that both the IR and MN are ~0.05, but that WMN and RW remain too high at ~0.10. A similar pattern is seen with EEG in Figure 9b, only more pronounced with MN at a very high level with the smooth source area, and together with the IR approximately correct with the extended source area. Since the extended source area is more superficial than the smooth source area (eccentricities of 0.64 and 0.91, respectively), the MN seems approximately to have the proper level. The IR has approximately the correct level for both types of sources and for MEG and EEG.

To test the level of the test statistics T^2_p and T^2_z the same smooth source area with center x_{s0} is defined for both conditions. The ROI for the hypothesis is again this smooth source area. The tests should not reject H_0 since the activity in the two conditions is the same. With T^2_z the correlation parameters are $a_1 = 1$ and $a_2 = 0.7$, which induces different covariance matrices for each condition. The level of the two tests is very similar to those presented in Figure 9a and are therefore not shown.

For the power investigation, local alternatives are used. A local alternative or close ROI can be defined by the distance between the centers of the areas $d(x_1,x_0)$ and the size of the radius ρ. Only when the hypothesized ROI and the source area overlap but not just when the centers are the same should H_0 be rejected. The power is investigated in two ways: the center of an ROI with the same radius is moved closer to the true source area, and different sizes of radii are used in two different conditions. It is desired that at close distances between the true source area and the hypothesized ROI the power is high, but the power should drop off drastically when the distance increases. Note that since the radii of the smooth source areas are 0.2, the smooth ROI and source area overlap when the centers are 0.4 apart. In the hypothesis $P'R\beta$, when the ROI and true source area are spread out more (see Fig. 2), the areas overlap at even larger distances between the centers. In the left panel of Figure 9c it is seen that for MEG and colored noise with $n = 80$ and the smooth true source area, the power for local alternatives drops off to 0.05 quite rapidly with increasing distance between the centers for MN, IR, WMN, and RW alike. This means that T^2 can distinguish between H_0 and H_1 quite well even when the ROI interest and source area are close. However, when the true extended source area is used (right panel of Fig. 9c) it is seen that the power with MN remains large compared to IR and RW especially. With EEG (Fig. 9d), the power drops off slowly with MN with a smooth source area (left panel), and the power with MN remains at around 0.20 when the true source area is extended (right panel). The power with the other linear inverse methods drops off less rapidly as the distance increases than MEG.

For the power in two different conditions, two different radii are used for the true source areas at the same locations. In one of the conditions the radius is 0.2 and in the other condition it is 0.23. The difference between two conditions is tested at the true location x_{s0} with either T^2_p or T^2_z. It appeared that the power is at least 0.99 (not shown) for all number of trials and white or colored noise. So, any slight deviation from H_0, which states that the source areas are the same between conditions, is picked up immediately.

APPLICATION TO MEG DATA

The aim of this section is to demonstrate how the hypothesis testing method described above could be used with experimental data. The data are from an experiment, also described in Waldorp et al. [2002], in which the participant was presented with 400 checkerboard patterns with 4.8 cycles per degree spatial frequency and subtending 6.7 degrees

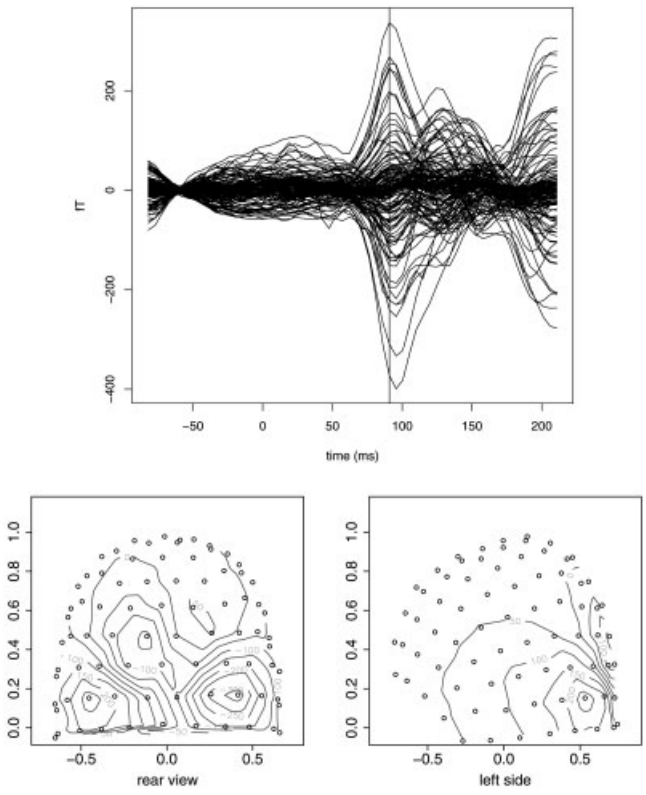


Figure 10.
Top: Averaged MEG data with the vertical line indicating the latency 91.2 ms used for analysis. Bottom: The CTF 151 sensor configuration with the MEG data at 91.2 ms.

of visual angle (full visual field). No task was performed by the subject. MEG was recorded with a CTF gradiometer system with 151 sensors for 300 ms of which 82 ms were before stimulus presentation. The data downsampled to 4.8 ms intervals and baseline corrected using data from –50 to 0 ms. The time sample at (latency of) 91.2 ms with the largest peaks is selected for analysis, which is plotted in Figure 10.

A grid is defined with 1,714 voxels in the volume between two spheres with the radius of the outer sphere 9.5 cm and of the inner sphere 3.8 cm. The linear inverse method IR is used, for it leads to interpretable hypotheses and in Grave de Peralta Menendez et al. [2004a] it was shown that it can reconstruct different types of sources. An estimate of the source distribution from the data with the IR linear inverse method can be seen in Figure 11a. From this estimate the orientations are obtained as described in the section Hypothesis Testing. An ROI is defined for the amplitudes and the estimate of the orientations is used to make the weight vector for hypothesis testing.

From the data in Figure 11a it appears that there might be two separate source areas in or near the occipital cortex. This corresponds roughly to previous results [Kenemans et al., 2000] where two areas have been reported in similar experiments with EEG. To investigate whether this is true, two ROIs are constructed with relatively small radii of two voxels. These ROIs had centers $x_1 = (-7, 2, 3)$ and $x_2 = (-7, -2, 3)$ and radius 2 cm. The original ROIs are displayed in Figure 11b and the actual ROIs being tested are shown in Figure 11c, obtained with the IR linear inverse method. If it appears that these ROIs do not seem to correspond to what is intended to be tested, then they can be changed in size and location. The overall test on these ROIs is $F(2,381) = 13.154$ with $P = 0.002$, indicating that there is activity in the ROI. Tests on the individual regions are $F(1,382) = 8.121$ with $P = 0.005$ for x_1 and $F(1,382) = 2.382$ with $P = 0.124$ for x_2, indicating that there is activity only in the ROI associated with x_1. When moving the center x_2 around in the same neighborhood for four different locations, it appears that at (–7, –3, 4) it is closest to being significant $F(1,382) = 3.636$ with $P = 0.057$. Since in total six tests were performed, the Bonferroni corrected significance level is $0.05/6 = 0.0083$. Hence, the ROI for x_2 with the smallest P-value is far from significant. In order to be sure that no spurious activity is detected, two additional ROIs are tested (Fig. 11d), which are slightly further from the area expected to contain activity, and have centers $x_3 = (-6, 3, 3)$ and $x_4 = (-6, -3, 3)$ also with a radius of two voxels. The overall test for these ROIs is $F(2,381) = 0.298$ with $P = 0.862$, indicating that there is no activity in these ROIs. From this analysis it can be concluded that there is one active region with center x_1 and a radius of two voxels present in the data. This is consistent with a similar study on selective attention in which EEG data were analyzed at approximately the same latency (89 ms after stimulus presentation) [Baas et al., 2002].

DISCUSSION

Testing each voxel separately in the mass univariate approach was seen not to correspond to actually testing a single voxel. For several different linear inverse methods it was seen that instead of testing a single voxel, a somewhat larger area is tested. This is because the constrained solutions of the amplitudes or moments are spatially smoothed out over an area, reflecting the intrinsic bias of the solution to an underdetermined system. This simply means that, also in hypothesis testing, the resolution in distributed source models is not in the order of the defined voxels, but in the order of a source area determined by the linear inverse method and underlying activity.

As an alternative to the mass univariate approach, a multivariate approach was suggested, in which an ROI is tested which to some extent corresponds to an area that is intended to be tested. It was shown in Figure 5 that a spatially smooth ROI leads to more interpretable hypotheses (i.e., the ROI corresponds highly to what is actually tested) than an ROI with crude edges with the linear inverse methods used here. In simulations it was seen that hypothesis testing with MEG leads to more accurate results than with EEG. Indeed, when moving an ROI closer to the true source area using MEG, the

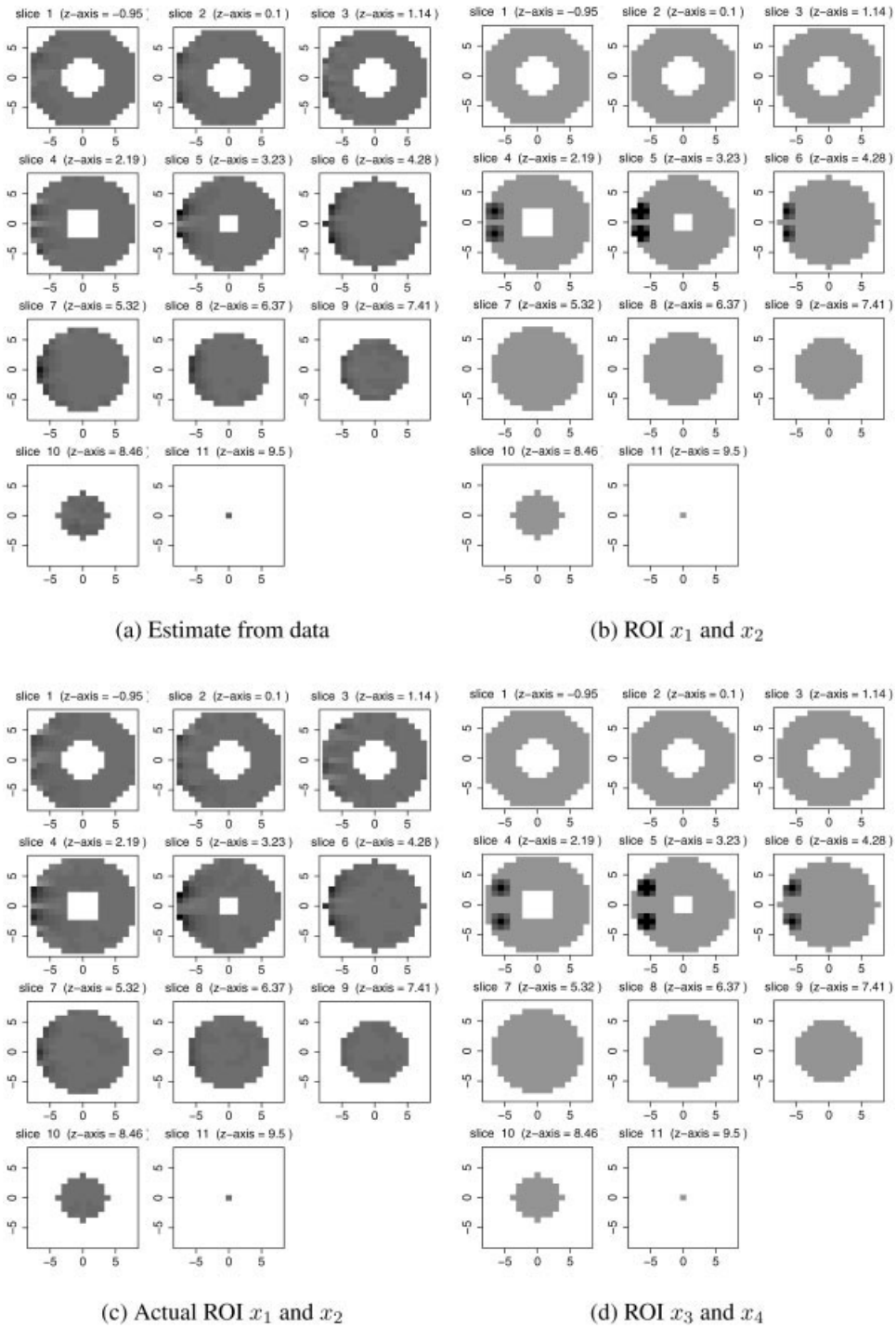


Figure 11.
Estimate of the data (**a**), two ROIs which are thought to be at the areas of activity (**b**), the actual ROI being tested (**c**), two other ROIs very near the areas of activity (**d**).

ROI and the true source area could be distinguished up to the point where they overlap. With EEG, on the other hand, the distance between the ROI and true source area remained somewhat larger before "setting of the alarm" of activity. This difference was more pronounced when an extended source area with crude edges was used. In addition, it appeared that the inverse resolution linear inverse method of Grave de Peralta Menendez et al. [2004a results in more interpretable hypotheses than other linear inverse methods.

Even though in the multivariate approach several voxels are tested simultaneously, the multiple comparison problem is still an issue, as was seen in the application to visual evoked field data. Only now it is at the level of the ROI instead of at the voxel level. Since probably not many ROI

will be tested, the Bonferroni correction, the Bonferroni-Holmes, or the false discovery rate could be used to avoid deciding too often that there is activity in the ROI. These procedures are adequate even when the ROIs are overlapping. This is because nowhere near as many tests are performed as when each voxel is tested separately.

The conclusions of the application to the visual evoked field data actually contradict a dipole analysis of the same data. It was found in Waldorp et al. [2002] by using several model selection methods that two dipoles give a good description of the data. In the present analysis, however, the distributed source model using the inverse resolution linear inverse method and multivariate hypothesis testing indicated that only one area is significantly active. Although it is far from trivial to compare the dipole analysis to the distributed source model approach, several general remarks can be made. First, in a distributed source model approach it is immediately clear that a source area is involved in generating the data, whereas in a dipole analysis a region of reliability has to be estimated separately by confidence intervals. Although such a source area cannot be interpreted as a probability of activity being present in each of the voxels, it indicates the extent of the activity. The multivariate approach suggested here corresponds to the resolution on the level of these source areas. Second, the constraints required to arrive at a solution directly and very clearly reflect the solution that is obtained. For example, in LORETA [Pascual-Marqui et al., 1994] the spatial smoothness constraint is explicitly stated and the spatially smooth solution obtained with LORETA reflects the constraint. In a dipole analysis, on the other hand, the activity of the dipole reflects in a more complicated way an aggregate of neuronal activity. Third, physiological constraints from other modalities can easily be used to obtain an estimate of activity. For instance, the shape and thickness of the cortical sheet can be used to constrain the solution to realistic areas [e.g., Phillips et al., 2002: Baillet and Garnero, 1997: Dale and Sereno, 1993], so that the obtained solution corresponds more closely to reality.

In the present analysis relatively simple linear inverse methods were employed. Since many variations exist, the method of multivariate hypothesis testing could easily be extended by using other linear inverse methods. If more information on the type of source expected is available, then a linear inverse method could be more tuned to this information. For example, if focal sources are expected, then possibly a linear inverse method tuned to focal sources could be used [e.g., Phillips et al., 1997]. The multivariate approach suggested herein is independent of the type of linear inverse method used; that is, for any type of linear inverse method there is intrinsic bias depending on both the type of source and constraints. If a priori knowledge of the source is known, as in Grave de Peralta Menendez et al. [2004b, then by using this information it is possible to obtain highly interpretable hypothesis tests when using the multivariate approach. The multivariate approach requires that hypothesis testing is done on a source area and not on each voxel separately.

ACKNOWLEDGMENTS

We thank the people at the VU Medical Center, MEG Center, Amsterdam, for helping Dr. K.B.E. Böcker collect the valuable data on visual evoked fields.

REFERENCES

Anderson T (1958): An introduction to multivariate statistical analysis. New York: John Wiley & Sons.

Baas J, Kenemans J, Mangun G (2002): Selective attention to spatial frequency: an ERP and source localization analysis. Clin Neurophysiol 113:1840–1854.

Baillet S, Garnero L (1997): A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans Biomed Eng 44:374–385.

Barnes G, Hillebrand A (2003): Statistical flattening of MEG beamformer images. Hum Brain Mapp 18:1–12.

Ben-Israel A, Greville T (2003): Generalized inverses: theory and applications. New York: Springer.

Bénar C-G, Gunn R, Grova C, Champagne B, Gotman J (2005): Statistical maps for EEG dipolar source localization. IEEE Trans Biomed Eng 52:401–413.

Bilodeau M, Brenner D (1999): Theory of multivariate statistics. New York: Springer.

Dale A, Liu A, Fischi B, Buckner R, Belliveau J, Lewine J, Halgren E (2000): Dynamic statistical parametric mapping: combining fMRI and MEG for high resolution imaging of cortical activity. Neuron 26:55–67.

Dale A, Sereno M (1993): Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J Cogn Neurosci 162–176.

De Munck J, Vijn P, da Silva FL (1992): A random dipole model for spontaneous brain activity. IEEE Transon Biomed Eng 39:791–804.

Dogandžić A, Nehorai A (2000): Estimating dipole responses in unknown spatially correlated noise with EEG/MEG arrays. IEEE Trans Signal Process 48:13–25.

Grasman R (2004): Sensor array signal processing and the neuroelectromagnetic inverse problem in functional connectivity analysis of the brain. Ph.D. thesis. Amsterdam, University of Amsterdam.

Grave de Peralta Menendez R, Hauk O, Gonzalez Andino S, Vogt H, Michel C (1997): Linear inverse solutions with optimal resolution kernels applied to electromagnetic tomography. Hum Brain Mapp 5:454–467.

Grave de Peralta Menendez R, Gonzalez Andino S (1998): A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. IEEE Trans Biomed Eng 45:440–448.

Grave de Peralta Menendez R, Murray M, Gonzalez Andino S (2004a): Improving the performance of linear inverse solutions by inverting the resolution matrix. IEEE Trans Biomed Eng 51:1680–1683.

Grave de Peralta Menendez R, Murray M, Michel C, Martuzzi R, Gonzalez Andino S (2004b): Electrical neuroimaging based on biophysical constraints. Neuroimage 21:527–539.

Hämäläinen M, Hari R, Ilmoniemi R, Knuutila J, Lounasmaa O (1993): Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65:413–497.

Hansen P (1992): Analysis of discrete ill-posed problems by means of the L-Curve. SIAM Rev 34:561–580.

Kenemans J, Baas J, Mangun G, Lijffijt M, Verbaten M (2000): On the processing of spatial frequencies as revealed by evoked-potential source modeling. Clin Neurophysiol 111:1113–1123.

Kiebel S, Friston K (2004a): Statistical parametric mapping for event-related potentials. I. Generic considerations. Neuroimage 22: 492–502.

Kiebel S, Friston K (2004b): Statistical parametric mapping for event-related potentials. II. A hierarchical temporal model. Neuroimage 22:503–520.

Liu A, Dale A, Belliveau J (2002): Monte Carlo simulation studies of EEG and MEG localization accuracy. Hum Brain Mapp 16:47–62.

Menke W (1989): Geophysical data analysis: discrete inverse theory. New York: Academic Press.

Michel C, Grave de Peralta R, Lantz G, Gonzalez Andino S, Spinelli L, Blanke O, Landis T, Seeck M (1999): Spatiotemporal EEG analysis and distributed source estimation in presurgical epilepsy evaluation. J Clin Neurophysiol 16:239–266.

Mosher J, Lewis P, Leahy R (1992): Multiple dipole modeling and localization from spatio-temporal meg data. IEEE Trans Biomed Eng 39:541–557.

Muirhead R (1982): Aspects of multivariate statistical theory. New York: John Wiley & Sons.

Pascual-Marqui R, Michel C, Lehmann D (1994): Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18:49–65.

Phillips J, Leahy R, Mosher J (1997): MEG-based imaging of focal neuronal current sources. IEEE Trans Med Imaging 16:338–348.

Phillips C, Rugg M, Friston K (2002): Anatomically informed basis functions for eeg source localization: combining functional and anatomical constraints. Neuroimage 16:678–695.

Pringle R, Rayner A (1971): Generalized inverse matrices with applications to statistics. London: Charles Griffin and Co.

Srivastava M, Carter E (1983): An introduction to applied multivariate statistics. New York: Elsevier Science.

Van der Vaart A (1998): Asymptotic statistics. New York: Cambridge University Press.

Waldorp L, Huizenga H, Grasman R, Böcker B, Munck J, Molenaar P (2002): Model selection in electromagnetic source analysis with an application to VEFs. IEEE Trans Biomed Eng 49:1121–1129.

Worsley K (2001): Statistical analysis of activation images. In: Jezzard, Matthews, Smith, editors. Functional magnetic resonance imaging: an introduction. New York: John Wiley & Sons. p 251–270.

Xiang D, Wahba G (1996): A generalized approximate cross validation for smoothing splines with non-Gaussian data. Stat Sin 6:675–692.